

IM&F시리즈 18

## 금융공학 VIII

Bayesian Methods  
for Finance and Economics

최병선 지음

*SIRFE*

[표지사진]

**The Second Wave, Arizona**

2018. 04. 19.

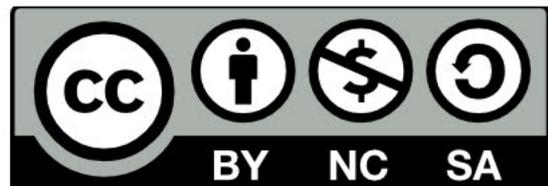
F4.5 and 1/1250sec

Canon EOS 70D with Tamron 18-270 F/3.5-6.3

이 저작물에는 크리에이티브 커먼즈 저작자표시-비영리 4.0 국제 라이선스가 적용 되어 있습니다. 이 라이선스의 설명을 보고 싶으시면 다음 웹사이트를 참조하세요.

<http://creativecommons.org/licenses/by-nc/4.0/>

 **creative  
commons**



# 머릿글

In these matters, the only certainty is that nothing is certain.

Pliny the Elder [78]

본서를 읽고 있는 독자는 고전통계학(classical statistics)이나 베이지안통계학(Bayesian statistics)에 대한 지식이 조금은 있으리라 생각한다. 특히 고전통계학을 어느 정도 알고 있으리라 생각한다. 고전통계학이 본서에서 다루는 베이지안통계학과 근본적으로 다른 이유는 확률을 보는 관점의 차이에서 시작된다.

옛날부터 인간은 확률에 관심이 많았다. 그 주된 이유는 도박이 확률과 밀접한 관계가 있기 때문이다. 문헌에 의하면 그리스시대 올림픽의 어떤 경기에서는 도박금액이 몇 백만볼에 이르기도 했다고 한다. 또한 고대 중국에서 도박이 성행했다는 것은 널리 알려진 이야기이다. 도박에서 이기는 비율을 계산하는 과정에서 초기 확률이론의 체계가 잡혔다. Luca Paccioli는 1494년 그의 저서 *Summa de Arithmetica, Geometrica, Proportioni et Proportionalità*에서 problem of points(또는 division of the stakes)라는 확률문제를 제시했고, 200년이 지난 17세기 Blaise Pascal이 Pierre de Fermat에 보낸 편지에서 해답을 제시했다. 이 과정에서 확률이론이 발전하기 시작했다.

확률을 보는 관점에 따라, 확률에 관한 이론을 고전이론(classical theory), 상대빈도론(relative frequency theory), 주관론(subjective theory, subjectivity theory) 등으로 나눌 수 있다. 첫째, 고전이론에서는 동일한 가능성을 지닌 모든 경우(event)들의 개수(number of all equipossible cases)  $N$ 에 대한 원하는 경우의 개수(number of favorable cases)  $M$ 의 이론적(theoretical) 비율(ratio)로 확률을 정의한다. 여기서 유의할 점은  $N$ 이나  $M$ 은 실험에서 사용된 샘플 개수나 실험에서 선택된 샘플 개수가 아니라 사건의 개수라는 것이다. 둘째, 상대빈도론에서는 확률을 실제 세계에서 발생하는 사건들 사이 경험적 비율(empirical rate)이라고 설명한다. 즉, 확률을 주어진 모집단(population)에서 어떤 사건이 발생할 실제 상대빈도로 정의한다. 예를 들어,  $n$ 번 주사위를 던져서  $m$ 번 눈 '1'이 나오면 눈 '1'이 나올 확률은  $m/n$ 이다. 물론, 사건이 발생하는 환경이 적당하다면, 대수법칙(law of the large numbers)에 의해서 상대빈도  $m/n$ 은 고전이론적 확률  $M/N$ 으로 수렴할 것이다. 셋째, 주관론에서 확률은 각 개인이 특정한 순간에 주어진 사건에 대해서 갖는 믿음의 정도(degree of belief)를 나타낸다. 수리경제학자이기도 했던 Frank P. Ramsey (1931)[82]는 John Maynard Keynes가 그의 저서 *Treatise on Probability*[58]에서 강조한 확률의 객관성에 대해서 비판했다. Ramsey는

Keynes가 논리학(logics)에서 확률과 물리학(physics)에서 확률에는 차이가 있음을 간과하고 있다고 주장했다. Ramsey는 확률은 믿음의 정도라고 주장하며, 그 믿음의 정도를 어떻게 비교하고 측정할 수 있는가를 설명하고자 했다. 안타깝게도 27세에 요절한 천재 Ramsey 뒤를 이어 Bruno de Finetti와 Leonard J. Savage 등에 의해서 주관적 확률론이 발전하였다. 주관적 확률론자들은 객관적 확률이란 존재하지 않으며, 확률은 개인적이며 주관적 믿음을 나타내는 정도라고 주장한다.

본서에서 다루는 베이저안통계학은 주관적 확률을 바탕으로 구축된 학문이다. 즉, 주관적 확률로서 다양한 모형들의 타당성을 나타내고, 이 확률을 가중치로 해서 관찰값들이 나타내는 평균모형을 구한다. 이 평균모형을 실증분석과 의사결정에 이용함으로써, 모형이 가지고 있는 불확실성을 축소하려는 것이 베이저안통계분석의 목표이다. 본저자는 데이터가 넘쳐나는 오늘날에는 고전통계학을 적용할 수 있는 모형들은 한계가 있고, 그들을 이용해서 하는 통계적 추론도 실용적이지 못하다고 생각한다. 어떤 이론의 전개과정이 수리적으로 정교하다고 해도 그 이론에서 전제하는 가정이 현실적이지 못한 경우에는 그 분석결과를 믿을 수 없는 것이 당연하다. 따라서 본저자는 베이저안통계학이 고전통계학의 영역을 잠식해 들어가는 것은 필연적 결과라고 생각한다. 실제로 베이저안통계학이 얼마나 우리 생활 깊숙이 들어왔나를 알고 싶으면, 다음 책들을 읽어볼 것을 권한다.

McGrayne, S.B. (2012) *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant From Two Centuries of Controversy*, Yale University Press.

Silver, N. (2012) *The Signal and the Noise: Why So Many Predictions Fail-But Some Don't*, Penguin Press.

베이저안통계학은 우리가 사용할 수 있는 모형의 다양성을 제공한다. 컴퓨터과·공학이 급격하게 발전하는 오늘날에는 과거에는 다룰 수 없었던 모형을 다룰 수 있게 한다. 즉, 컴퓨터과·공학의 발전은 베이저안통계학에서 제공하는 모형의 다양성을 실제로 즐기는 것이 가능한 환경을 만들고 있다. 본서에서는 이러한 추세를 반영해서 재무학과 경제학을 위한 베이저안통계기법들을 설명한다. 본서는 계량경제학에서 베이저안통계학의 위치, 베이저안통계학의 수리적 배경, 베이저안통계모형, MCMC법(Markov Chain Monte Carlo method)을 이해하는데 필요한 Markov체인이론, Gibbs샘플러, 그리고 Metropolis-Hastings 법으로 구성되어 있다.

본서를 만드는 데 많은 사람들의 도움이 있었다. 우선 원고를 세심하게 읽고 코멘트를 해준 부경대학교 이성백교수, 교정을 보아준 김원세박사, 김찬수군, 백승헌군에게 감사한다.

또한 2017년도 2학기 수리금융경제학특수연구를 수강하면서 교정에 참여해준 학생들에게 감사한다. 본서는 L<sup>A</sup>T<sub>E</sub>X으로 작성한 것으로, 김찬수군이 제2.5절의 표들을 L<sup>A</sup>T<sub>E</sub>X으로 입력해 주었고, 또한 L<sup>A</sup>T<sub>E</sub>X스타일파일과 표지를 만들었다. 다시 한번 김찬수군에게 감사한다. 마지막으로 본서를 출간을 허락해주신 서울대학교 금융경제연구원 김영식원장님께 감사드리며, 출판과정에서 실무를 도와준 김남재군에게 고마움을 표한다.<sup>1</sup>

대한민국과 조선민주주의인민공화국의 영원한 평화를 기원하며 ...

최병선

2018년 11월 15일

*We are so much at the mercy of chance that Chance is our god.*

Pliny the Elder, Roman Philosopher and Statesman [78]

---

<sup>1</sup>현대문화사 (hun1249@hanmail.net) 에서 본서의 POD판을 구하실 수 있습니다.



# 차례

차례	i
<b>제 1 장 베이저안통계학과 계량경제학</b>	<b>1</b>
제 1.1 절 베이저안통계분석이란 무엇인가? . . . . .	1
제 1.2 절 Bayes 정리 . . . . .	4
1.2.1 Bayes 정리와 사후확률분포 . . . . .	4
1.2.2 유용한 확률분포들 . . . . .	21
제 1.3 절 다변량정규확률분포 . . . . .	42
1.3.1 적률생성함수 . . . . .	42
1.3.2 정규확률벡터 . . . . .	44
1.3.3 조건부확률밀도함수 . . . . .	48
제 1.4 절 Wishart 확률분포 . . . . .	58
1.4.1 Wishart 확률분포의 정의 . . . . .	58
1.4.2 역Wishart 확률분포 . . . . .	61
1.4.3 표본확률분포 . . . . .	63
1.4.4 최우추정량 . . . . .	66
1.4.5 신뢰구간 . . . . .	68
1.4.6 손실함수와 위험함수 . . . . .	70
1.4.7 미니맥스결정과정 . . . . .	74
1.4.8 James-Stein 추정벡터 . . . . .	75
1.4.9 Hotelling $T^2$ 통계량 . . . . .	78
제 1.5 절 교환가능성 . . . . .	84
<b>제 2 장 베이저안통계학의 기초</b>	<b>91</b>

제2.1절	리스크분석과 베이지안통계학 . . . . .	91
2.1.1	조건부리스크가치 CVaR . . . . .	91
2.1.2	CVaR의 평가 . . . . .	94
제2.2절	신용리스크의 베이지안통계분석 . . . . .	105
2.2.1	파산확률과 선형회귀모형 . . . . .	105
2.2.2	파산확률의 베이지안추정 . . . . .	111
제2.3절	정규확률분포와 베이지안통계학 . . . . .	121
2.3.1	정규확률분포의 평균 . . . . .	121
2.3.2	정규확률분포의 분산 . . . . .	125
2.3.3	정규확률분포의 평균과 분산 . . . . .	128
제2.4절	다변량정규확률분포와 베이지안통계학 . . . . .	133
2.4.1	다변량정규확률분포 . . . . .	134
2.4.2	다변량정규확률분포의 평균벡터 . . . . .	142
2.4.3	다변량정규확률분포의 분산공분산행렬 . . . . .	145
2.4.4	다변량정규확률분포의 평균벡터과 분산공분산행렬 . . . . .	146
제2.5절	사전확률분포 . . . . .	152
2.5.1	공액사전확률분포 . . . . .	152
2.5.2	무정보사전확률분포 . . . . .	166
2.5.3	조건부공액 . . . . .	172
제2.6절	사후확률분포와 통계적 추론 . . . . .	177
2.6.1	사후주변확률분포 . . . . .	177
2.6.2	점추정 . . . . .	178
2.6.3	구간추정 . . . . .	186
2.6.4	가설검정 . . . . .	188
2.6.5	예측 . . . . .	191

**제 3 장 베이지안통계모형들** **193**

제3.1절	선형모형의 베이지안통계분석 . . . . .	193
제3.2절	계층모형과 베이지안통계분석 . . . . .	224
제3.3절	다변량선형회귀모형 . . . . .	249
3.3.1	고전적 통계추론 . . . . .	249

3.3.2	베이지안 통계추론 . . . . .	281
<b>제 4 장</b>	<b>Markov 체인</b>	<b>289</b>
제 4.1 절	베이지안분석과 Markov 체인 . . . . .	289
제 4.2 절	Markov 체인의 정의 . . . . .	290
제 4.3 절	통계모형과 Markov 성 . . . . .	294
4.3.1	ARMA 모형과 Markov 성 . . . . .	294
4.3.2	Markov 스위칭 ARMA 모형과 Markov 성 . . . . .	295
4.3.3	상태공간모형과 Markov 성 . . . . .	297
4.3.4	확률미분방정식과 Markov 성 . . . . .	297
제 4.4 절	Markov 체인 예제 . . . . .	298
제 4.5 절	Chapman-Kolmogorov 식 . . . . .	315
제 4.6 절	상태의 분류 . . . . .	330
4.6.1	불변분포 . . . . .	330
4.6.2	최초방문시점 . . . . .	336
4.6.3	과도상태, 재귀상태와 흡수상태 . . . . .	346
4.6.4	도달확률과 도달횟수 . . . . .	349
제 4.7 절	상호도달가능과 동치류 . . . . .	354
제 4.8 절	단한집합과 재귀성 . . . . .	367
제 4.9 절	에르고딕 . . . . .	378
제 4.10 절	시간가역성 . . . . .	387
제 4.11 절	연속형 상태공간 . . . . .	408
<b>제 5 장</b>	<b>Gibbs 샘플러</b>	<b>411</b>
제 5.1 절	Gibbs 샘플러의 정의 . . . . .	411
제 5.2 절	Gibbs 샘플러의 수렴성 . . . . .	446
제 5.3 절	베이지안통계분석과 Gibbs 샘플러 . . . . .	461
5.3.1	정규확률분포와 Gibbs 샘플러 . . . . .	463
5.3.2	시계열의 변환점과 Gibbs 샘플러 . . . . .	469
5.3.3	선형회귀모형과 Gibbs 샘플러 . . . . .	481
5.3.4	구조변화가 있는 회귀모형과 Gibbs 샘플러 . . . . .	490
5.3.5	프로빗모형, 로지스틱회귀모형, 토빗모형과 Gibbs 샘플러 . . . . .	492

5.3.6	SUR모형과 Gibbs샘플러 . . . . .	526
5.3.7	상태공간모형과 Gibbs샘플러 . . . . .	552
<b>제 6 장</b>	<b>Metropolis-Hastings샘플러</b>	<b>555</b>
제6.1절	MH샘플러 . . . . .	555
6.1.1	MH샘플러 . . . . .	555
6.1.2	다변량 MH샘플러 . . . . .	568
6.1.3	Metropolis샘플러 . . . . .	584
6.1.4	확률보행Metropolis샘플러 . . . . .	592
6.1.5	독립샘플러 . . . . .	598
6.1.6	AR-MH샘플러 . . . . .	601
6.1.7	자기회귀샘플러 . . . . .	602
제6.2절	불변분포의 존재성 . . . . .	606
제6.3절	베이지안통계분석과 MH샘플러 . . . . .	608
6.3.1	베이지안통계분석을 위한 MH샘플러 . . . . .	608
6.3.2	MH샘플의 수렴성 . . . . .	612
6.3.3	사후확률분포를 이용한 추론과 모형선택 . . . . .	614
제6.4절	MH샘플러와 베이지안통계모형 . . . . .	616
6.4.1	MH샘플러를 사용한 이항확률모형분석 . . . . .	616
6.4.2	MH샘플러를 사용한 로지스틱회귀모형분석 . . . . .	620
6.4.3	MH샘플러를 사용한 AR모형분석과 ARX모형분석 . . . . .	632
6.4.4	MH샘플러를 사용한 자기회귀오차모형분석 . . . . .	662
<b>부록 A</b>		<b>669</b>
A.1	특성함수 . . . . .	669
A.2	중심극한정리 . . . . .	700
<b>참고 문헌</b>		<b>703</b>

## 제 1 장

# 베이저안통계학과 계량경제학

When my information changes,  
I alter my conclusions.  
What do you do, sir?

John Maynard Keynes  
(but controversial)

### 제1.1 절 베이저안통계분석이란 무엇인가?

계량경제학은 관측된 데이터를 바탕으로 경제학적인 결론을 유도하는 학문이다. 오늘날 계량경제학적 기법의 큰 부분을 베이저안통계학이 담당하고 있다. 베이저안통계학 (Bayesian statistics)을 주로 사용하는 사람을 베이저안(Bayesian) 이라고 하고, 고전통계학(classical statistics)을 주로 사용하는 사람을 도수주의자(度數主義者, frequentist)라 부른다. 실증 분석을 하는데 베이저안통계학이 유효하다는 것이 널리 받아들여지고 있지만, 아직까지도 베이저안통계학보다는 고전통계학이 계량경제학의 실증연구에 더 많이 사용되고 있다. 그 이유는 계량경제학을 사용해서 실증연구를 하는 연구자들에게 베이저안통계학이 익숙하지 않고 또한 상대적으로 어렵다는 오해에서 비롯된 것이다. 베이저안통계학에서는 모수를 확률변수로 가정하므로, 모수를 미지의 상수로 가정하는 고전통계학보다 수리적으로 더 어려운 면이 있다. 더구나, 고전통계학을 바탕으로 하는 계량경제학적 실증분석에서는 다루기 쉬운 상용소프트웨어가 널리 사용된 것에 반해서, 베이저안통계학을 편리하게 적용하기 위한 상용소프트웨어는 그리 많지 않았다.

표준화된 계량경제학적 기법을 이용해서 실증분석을 하는 경우에는 기존 소프트웨어를 이용할 수 있어서 편리하다. 그러나, 분석하고자 하는 문제나 분석에 사용되는 데이터에 따라서는 기존 계량경제학적 기법을 이용할 수 없는 경우가 있고, 이러한 경우에는 실증분석을

하기 위해서 당연히 새로운 방법을 개발할 필요가 있다. 모형이 아주 복잡하거나 아주 많은 계산을 필요로 하기 때문에 과거에는 분석하고자 하는 시도조차 되지 않았던 계량경제학 문제들 중에는 최근 들어 베이저안통계기법을 사용해서 실증분석을 할 수 있게 된 것들이 많다.

데이터를 바탕으로 경제나 금융시장을 실증분석하기 위해서는 먼저 데이터가 생성되었다고 생각되는 특정한 모형을 상정(想定)한다. 이러한 모형은 경제학과 재무이론에 근거하여 주가와 같은 금융변수들의 움직임을 설명하는 수식들로 이루어진다. 실제 관측된 데이터를 이용해서 이 상정된 모형을 추정한다. 또한, 추정된 모형을 바탕으로 가정된 모형 자체를 검증하고, 미래를 예측하고, 리스크관리와 같은 의사결정을 수행한다. 그러나, 데이터와 모형을 바탕으로 하는 실증분석과 의사결정에는 아주 중요한 가정이 전제되어 있다. 그것은 ‘우리는 데이터를 올바르게 기술할 수 있는 모형을 알고 있다’는 것이다. 잘못된 모형을 바탕으로 하는 실증분석과 의사결정은 큰 의미가 없으므로, 우리는 경제와 금융시장의 움직임을 조금이라도 더 정확하게 설명할 수 있도록 모형을 구축하려는 노력을 계속하고 있다. 또한, 데이터를 바탕으로 모형의 타당성을 검증하는 다양한 통계적 수법이 개발되고 있다. 그러나, 본저자는 완벽하게 경제와 금융시장의 움직임을 기술할 수 있는 모형은 존재하지 않는다고 생각한다. 또한, 인간이 그러한 경제모형을 찾아낼 수 있으리라 생각하지도 않는다. 그런 모형을 찾아낸다면 더 이상 경제학이나 재무학을 공부할 필요도 없을 것이다. 우리가 현재 의사결정에 사용하고 있는 경제모형이나 재무모형은 오차가 큰 근사모형일 수도 있다고 가정하는 것이 타당할 것이다. 본저자는 아주 큰 데이터세트를 사용하는 것은 현상을 올바르게 나타내는 모형을 선택하는 것에 대한 필요조건도 충분조건도 아니라고 생각한다. 데이터세트 크기가 너무 작아도 문제지만 큰 데이터세트라고 해도 이 데이터세트의 질과 이를 분석하는 능력에 따라 현실과는 동떨어진 결론을 내리기가 십상이기 때문이다.

베이저안통계분석의 기본적 철학은 다음과 같다.

*All models are wrong, but some are useful.*

이러한 철학을 바탕으로 확률로서 모형의 타당성을 나타내고, 이 확률을 가중치로 하는 평균모형을 구한다. 이 평균모형을 실증분석과 의사결정에 이용하므로써, 모형의 불확실성을 보완하는 것이 베이저안통계분석의 목표이다. 원래 이 문장은 George E.P. Box [20]에 의한 것이고, 이 문장은 후에 다음과 같이 정제된다. [21, p. 424]

*Essentially, all models are wrong, but some are useful.*

그러나 본저자는 다음 문장이 현재와 미래를 더 잘 나타낸다고 생각한다.

All models are wrong, and increasingly you can succeed without them.

이 문장은 Google의 리서치디렉터인 Peter Norvig이 한 것으로 알려져 있으나, 본인은 자기가 한 말이 아니라고 주장하는 것으로 알려져 있다. 이에 대한 자세한 내용은 다음 웹사이트를 참조하라.

<https://blog.alexlenail.me/models-1b5f443fddd3>

베이저안통계분석을 다음과 같이 다섯 단계로 나눌 수 있다.

**첫째**, 경제나 금융시장의 어떤 움직임을 설명하는데 필요한 모형 후보들을 모은다.

**둘째**, 모형에 대한 신뢰 정도를 사전확률분포(prior probability distribution)로 표현한다.

**셋째**, 관측된 데이터를 우도함수(likelihood function)로 나타낸다.

**넷째**, 이 사전확률분포와 우도함수를 Bayes정리에 대입해서 사후확률분포(posterior probability distribution)를 구한다.

**다섯째**, 사후확률분포를 가중치로하는 평균모형인 예측분포(predictive distribution)를 구하고, 이를 바탕으로 의사결정을 수행한다.

베이저안통계학을 바탕으로 하는 계량경제분석은 1960년대 초부터 시작되었다. 당시 연구는 계량경제학적 문제를 해결하기 위해서, 베이저안통계기법을 개발하고 적용하는 것이 중심이었다. 즉, 경제현상을 기술하고 모형화하고, 경제이론에 따라 가설을 세우고, 이 가설을 검정하거나 예측을 시행함으로써, 다양한 경제문제를 계량경제학적으로 기술하고 이에 대한 해법을 제시하였다. 베이저안통계학을 바탕으로 하는 계량경제학에 관한 연구는 1960년대에 급속히 발전하였고, 1970년에는 베이스계량경제학에 관한 NBER-NSF 세미나가 처음으로 시카고대학에서 개최되었다. 그러나, 1990년대 초까지도 계량경제학에서 베이저안통계학을 사용하는 것은 주로 이론적 측면이었고, 베이저안통계학을 사용한 실증연구는 많지 않았다. 그렇게 된 첫 번째 이유는 모수의 사전확률분포라는 개념에는 주관성이 따를 수밖에 없기 때문이다. 즉, 베이저안통계분석에서는 사전확률분포와 우도함수가 필요하며, 사전확률분포는 주관적으로 선택된다. 통계학계의 주류를 이루는 도수주의자들은 객관성을 중요시하고, 따라서 주관적으로 선택되는 사전확률분포를 받아들이지 않는다. 두 번째 이유는 베이저안통계학에서 사후확률분포나 사후적률(posterior moment)을 구하기 위해서 다중적분을 해야하는데, 이 경우에 피적분함수가 복잡해서 적분계산이 곤란한 경우가 많았기 때문이다. 이 두 번째 이유는 컴퓨터과 공학의 발전에 힘입어서 쉽게 극복할 수 있다. 베이저안통계학

의 이론적 발전과 컴퓨터학의 발전에 따른 계산능력의 급속한 발달은 베이저안통계기법을 적용할 수 있는 문제의 범위를 급격히 넓히고 있다. 최근 Python과 R에서는 베이저안통계기법에 관한 모듈이 많이 제공되고 있다. 베이저안통계분석을 하기 위한 R의 패키지들로는 BUGS(Baysian Inference Using Gibbs Sampling) 계열의 WinBUGS와 OpenBUGS, 그리고 JAGS(Just Another Gibbs Sampler), Stanislaw Ulam의 이름을 딴 Stan 등이 널리 사용되고 있다. 또한, Python 패키지로는 BayesPy, PyPI, PyMC 등이 있다.

본서의 원고를 작성하는데 참고한 문헌들은 다음과 같다. 기초적 베이저안통계학에 관해서는 Leonard & Hsu [67]과 Hoff [55], 계산베이저안통계학 (computational Bayesian statistics)에 관해서는 Bolstad [19], MCMC법 (Markov Chain Monte Carlo method)을 사용해서 시계열모형에 관한 주제를 주로 다루고 있는 Bauwens & Lubrano & Richard [16], 계량경제분석에서 나타나는 다양한 모형들에 적용하는 베이저안통계기법에 관해서는 Geweke [46]와 Koop & Poirier & Tobias [62], 재무이론에 나타나는 베이저기법들을 잘 설명하고 있는 Rachev & Hsu & Bagasheva & Fabozzi [80], 베이저계량경제학의 최근 경향을 담고 있는 Geweke & Koop & Van Dijk [47]와 재무학에서 나타나는 몬테카를로시뮬레이션에 관해 자세히 설명하고 있는 Brandimarte [22] 등을 주로 참조하였다.

## 제1.2절 Bayes 정리

### 1.2.1 Bayes 정리와 사후확률분포

베이저안통계분석에서는 모수를 확률변수로 간주한다. 이후, 관찰된 데이터를 나타내는 관찰벡터를  $\mathbf{y}$  그리고 모수벡터를  $\boldsymbol{\theta}$ 로 표기하자. 만약 결측데이터(missing data)가 존재하면, 이 결측데이터도 모수벡터  $\boldsymbol{\theta}$ 에 포함된다고 하자.

통계적 추론을 하기 위해서는  $(\mathbf{y}, \boldsymbol{\theta})$ 에 관한 결합확률밀도함수  $f(\mathbf{y}, \boldsymbol{\theta})$ 가 필요하다. 다음 식이 성립한다.

$$f(\mathbf{y}, \boldsymbol{\theta}) = f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \quad (1.2.1)$$

즉, 결합확률밀도함수  $f(\mathbf{y}, \boldsymbol{\theta})$ 는 사전확률분포의 확률밀도함수인 사전확률밀도함수  $\pi(\boldsymbol{\theta})$ 와 우도함수  $f(\mathbf{y}|\boldsymbol{\theta})$ 의 곱이다. 사전확률분포는 데이터를 관측하기 전에 가지고 있는  $\boldsymbol{\theta}$ 에 관한 사전 정보를 나타내고, 우도함수는  $\boldsymbol{\theta}$ 가 주어진 조건 하에서 관찰벡터  $\mathbf{y}$ 의 확률밀도함수를

나타낸다. 데이터  $\mathbf{y}$ 의 확률밀도함수  $f(\mathbf{y})$ 는 다음 식을 만족한다.

$$f(\mathbf{y}) = \int f(\mathbf{y}, \boldsymbol{\theta}) d\boldsymbol{\theta} \quad (1.2.2)$$

식 (1.2.1)과 식 (1.2.2)에서 알 수 있듯이, 다음 식들이 성립한다.

$$f(\mathbf{y}) = \int f(\mathbf{y}, \boldsymbol{\theta}) d\boldsymbol{\theta} = \int f(\mathbf{y} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (1.2.3)$$

식 (1.2.3)에서 알 수 있듯이,  $f(\mathbf{y} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta})$ 를  $\boldsymbol{\theta}$ 의 지지대에서 적분한 것이  $f(\mathbf{y})$ 이다. 식 (1.2.3)의 우변항을 정규화상수(normalizing constant)라고 부른다. 정규화상수는  $\boldsymbol{\theta}$ 에 의존하지 않음을 기억하라. 관찰벡터  $\mathbf{y}$ 가 주어진 조건 하에서 모수벡터  $\boldsymbol{\theta}$ 의 확률밀도함수  $\pi(\boldsymbol{\theta} | \mathbf{y})$ 는 다음 식들을 만족한다.

$$\pi(\boldsymbol{\theta} | \mathbf{y}) = \frac{f(\mathbf{y} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta})}{f(\mathbf{y})} = \frac{f(\mathbf{y} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta})}{\int f(\mathbf{y} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}} \quad (1.2.4)$$

여기서 첫 번째 등호는 조건부확률밀도함수의 성질에 의해서, 그리고 두 번째 등호는 식 (1.2.3)에 의해서 성립한다. 식 (1.2.4)를 Bayes 정리(Bayes theorem)라 부른다. 식 (1.2.4)의  $\pi(\boldsymbol{\theta} | \mathbf{y})$ 는 모수벡터  $\boldsymbol{\theta}$ 의 사후확률분포의 확률밀도함수, 즉 사후확률밀도함수이다. 베이저안통계분석의 첫 번째 단계는 이 사후확률밀도함수  $\pi(\boldsymbol{\theta} | \mathbf{y})$ 를 구하는 것이다. 식 (1.2.4)를 다음과 같이 쓸 수 있다.

$$\pi(\boldsymbol{\theta} | \mathbf{y}) \propto f(\mathbf{y} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \quad (1.2.5)$$

식  $\int \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} = 1$ 을 이용해서, 식 (1.2.5)로부터 식 (1.2.4)를 쉽게 유도할 수 있다.

사후확률분포를 잘 이용하는 것이 베이저안통계분석의 핵심이다. 즉, 베이저안통계분석은 사후확률분포를 이용해서 적률이나 사분위수와 같은 통계량들을 추정하고 이를 사용해서 가설검정하는 것을 목표로 한다. 이러한 통계량들을 다음과 같이 모수벡터의 함수  $h(\boldsymbol{\theta})$ 의 사후기대값으로 표현할 수 있다.

$$E(h(\boldsymbol{\theta}) | \mathbf{y}) = \frac{\int h(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) f(\mathbf{y} | \boldsymbol{\theta}) d\boldsymbol{\theta}}{\int \pi(\boldsymbol{\theta}) f(\mathbf{y} | \boldsymbol{\theta}) d\boldsymbol{\theta}} \quad (1.2.6)$$

앞에서도 언급했듯이, 식 (1.2.4)의 우변 또는 식 (1.2.6)의 우변에서 분모는 사후확률밀도함수를 적분해서 1이 되게 하는 정규화상수이다. 일반적으로 정규화상수를 해석적으로 구하는 것은 쉬운 일이 아니다. 특히, 모수벡터  $\boldsymbol{\theta}$ 가 고차원인 경우에는 더욱 그러하다. 앞에서도

언급했듯이, 이 문제가 베이지안통계분석을 하는데 발생하는 난점들 중 하나였다. 이 난점을 피하기 위해서, 과거에는 단순한 베이지안통계모형을 사용하는 편법을 사용해왔다. 그러나, 오늘날에는 MCMC를 사용해서 이 적분문제를 수치적으로 쉽게 해결할 수 있으므로, 다양한 베이지안통계모형들을 분석에 사용할 수 있다.

**예제 1.2.1** 확률변수들  $x_1, x_2, \dots, x_n$ 이 성공확률이  $p$ 인 Bernoulli 확률분포를 따른다고 하고, 모수  $p$ 가 주어진 조건 하에서 이 확률변수들이 독립이라고 가정하자. 이 Bernoulli 확률변수들의 합  $y \doteq \sum_{i=1}^n x_i$ 는 모수들이  $n$ 과  $p$ 인 이항확률분포  $Binomial(n, p)$ 를 따른다. 즉,  $y$ 의 확률질량함수는 다음과 같다.

$$f(y|p) = \binom{n}{y} p^y [1-p]^{n-y} \quad (1)$$

모수  $p$ 의 사전확률분포가 베타확률분포  $Beta(\alpha, \beta)$ 라고 가정하자. 여기서  $\alpha$ 와  $\beta$ 는 미리 알고 있는 상수들이다. 즉, 사전확률밀도함수는 다음과 같다.

$$\pi(p|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1} [1-p]^{\beta-1} \quad (2)$$

여기서 베타함수는 다음과 같음을 상기하라.

$$B(a, b) \doteq \int_0^1 x^{a-1} [1-x]^{b-1} dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad (a > 0, b > 0) \quad (3)$$

확률변수  $y$ 와 모수  $p$ 의 결합확률밀도함수  $f(y, p|\alpha, \beta)$ 는  $y$ 의 우도함수  $f(y|p)$ 와  $p$ 의 사전확률밀도함수  $\pi(p|\alpha, \beta)$ 의 곱이다. 즉, 다음 식들이 성립한다.

$$f(y, p|\alpha, \beta) = f(y|p)\pi(p|\alpha, \beta) = \binom{n}{y} p^y [1-p]^{n-y} \frac{1}{B(\alpha, \beta)} p^{\alpha-1} [1-p]^{\beta-1} \quad (4)$$

식 (4)에서 각 변을  $p$ 에 관해 적분하면, 다음 식들이 성립함을 알 수 있다.

$$f(y|\alpha, \beta) = \int_0^1 f(y, p|\alpha, \beta) dp = \binom{n}{y} \frac{1}{B(\alpha, \beta)} B(y+\alpha, n-y+\beta) \quad (5)$$

식 (5)의 우변과 같이 정의되는 함수를 베타이항확률분포 (beta-binomial distribution)에 해당하는 확률밀도함수라고 한다.

식 (1.2.4)에서 알 수 있듯이,  $p$ 의 사후확률밀도함수는 다음과 같다.

$$\pi(p|\alpha, \beta, y) = \frac{f(y, p|\alpha, \beta)}{f(y|\alpha, \beta)} = \frac{1}{B(y + \alpha, n - y + \beta)} p^{y+\alpha-1} [1 - p]^{n-y+\beta-1} \quad (6)$$

이 사후확률밀도함수  $\pi(p|\alpha, \beta, y)$ 는 모수들을  $\alpha_* \doteq y + \alpha$ 와  $\beta_* \doteq n - y + \beta$ 로 하는 베타확률분포  $Beta(\alpha_*, \beta_*)$ 의 확률밀도함수이다. 여기서 유의할 점은 사후확률밀도함수

$\pi(p|y) = \pi\left(p \mid \sum_{i=1}^n x_i\right)$ 는 관찰값들  $x_1, x_2, \dots, x_n$ 에 직접 의존하는 것이 아니라 이들의 합  $y = \sum_{i=1}^n x_i$ 에 의존한다는 것이다. 즉,  $y$ 는 충분통계량(sufficient statistic)이다.

베이저안통계분석의 개요를 파악하기 위해서, 다음 R프로그램 BayesianRatio101.R을 실행해 보자.

```

1 # -----
2 # Filename: BayesianRatio101.R
3 # Prior, Likelihood and Posterior by Bayesian Method
4 # Programmed by CBS
5 # -----
6 setEPS()
7 p <- seq(0,1,length=301)
8 ( alpha <- 5); ( beta <- 10 )
9 ( n <- 16 ); ( y <- 9 )
10 # Plotting
11 plot.new()
12 frame()
13 postscript('BayesianRatio101a.eps') # Start to save figure
14 curve(dbeta(x,alpha,beta), from=0, to=1,
15       xlab="p", ylab="PDF", lty=3, lwd=4, ylim=c(0,6), col=4 )
16 curve(dbeta(x,y+1,n-y+1), add=TRUE, lty=2, lwd=4, col=2)
17 curve(dbeta(x,alpha+s,beta+f), add=TRUE, lty=1, lwd=4, col=1)
18 legend(0.7,5,c("Prior","Likelihood","Posterior"),
19       lty=c(3,2,1), lwd=c(4,4,4), col=c(4,2,1) )
20 dev.off() # End to save figure
21 # Cumulative Distribution: # Pr(x <= 0.5)
22 ( PrFreq <- pbeta(0.5,alpha,beta) )
23 ( PrBayes <- pbeta(0.5,alpha+n,beta+y) )
24 # Percentiles: Pr( 0.05 <= x <= 0.95)
25 ( PercenFreq <- qbeta(c(0.05,0.95),alpha,beta) )
26 ( PercenBayes <- qbeta(c(0.05,0.95),alpha+n,beta+y) )
27
28 # Simulation
29 set.seed(1)
30 ( K = 1024 ) # Number of observations
31 RNpri = rbeta(K,alpha,beta)
32 RNpost = rbeta(K,alpha+s,beta+f)
33 # Sample Cumulative Distribution: # Pr(x <= 0.5)
34 ( SamplePrFreq <- sum(RNpri <= 0.5)/K )
35 ( SamplePrBayes <- sum(RNpost <= 0.5)/K )
36 # Sample Percentiles: Pr( 0.05 <= x <= 0.95)
37 ( SamplePercenFreq <- quantile(RNpri,c(0.05,0.95)) )
38 ( SamplePercenBayes <- quantile(RNpost,c(0.05,0.95)) )
39 # Plotting
40 plot.new()
41 frame()

```

```

42| postscript('BayesianRatio101b.eps') # Start to save figure
43| par(mfrow=c(2,1))
44| hist(RNpri, xlab="p", main="Prior", xlim=c(0,1))
45| abline(v=SamplePercenFreq, col="red", lty=2, lwd=4)
46| hist(RNpost, xlab="p", main="Posterior", xlim=c(0,1))
47| abline(v=SamplePercenBayes, col="red", lty=2, lwd=4)
48| dev.off() # End to save figure
49| # -----

```

이 R 프로그램 BayesianRatio101.R에서 사전확률분포는  $Beta(5, 10)$  이고, 관찰값들  $x_1, x_2, \dots, x_{16}$  의 합은  $y = 9$ 이다.

이 R 프로그램 BayesianRatio101.R을 실행하면, 그림 1.2.1이 출력된다. 그림 1.2.1에서 청색 점선은 사전확률밀도함수를, 적색 긴점선은 우도함수를, 그리고 흑색 실선은 사후확률밀도함수를 나타낸다. 그림 1.2.1에서 알 수 있듯이, 사전확률밀도함수에 우도함수의 정보를 추가한 사후확률밀도함수는 사전확률밀도함수와 우도함수의 중간에 위치한다. 또한, 다음 식들이 성립함을 알 수 있다.

$$\Pr(p \leq 0.5 \mid \alpha = 5, \beta = 10) = 0.9102173 \quad (7)$$

$$\Pr(p \leq 0.5 \mid \alpha = 5, \beta = 10, y = 9) = 0.3746293 \quad (8)$$

사전확률분포의 평균은  $\frac{\alpha}{\alpha+\beta} = \frac{5}{5+10} = 0.3333333$ 이다. 즉, 사전확률의 반 정도가  $p = 1/3$  아래에 몰려있다. 특히, 사전확률이  $\Pr(p \leq 0.5 \mid \alpha = 5, \beta = 10)$ 이 0.9102173이라는 것은 사전확률의 대부분  $p = 0.5$  이하에 몰려있음을 의미한다. 그러나, 16개 관찰값들 중 9개가 성공이므로  $p$ 의 사후확률은 좀 더 큰 쪽으로 이동한다. 즉, 사후확률  $\Pr(p \leq 0.5 \mid \alpha = 5, \beta = 10, y = 9)$ 이 0.3746293이므로 사후확률의 반 이상이  $p > 0.5$ 에 몰려있다. 다음 식들이 성립함을 알 수 있다.

$$\Pr(p \leq 0.05 \mid \alpha = 5, \beta = 10) = 0.1527176 \quad (9)$$

$$\Pr(p \leq 0.05 \mid \alpha = 5, \beta = 10, y = 9) = 0.3958057 \quad (10)$$

$$\Pr(p \leq 0.95 \mid \alpha = 5, \beta = 10) = 0.5400054 \quad (11)$$

$$\Pr(p \leq 0.95 \mid \alpha = 5, \beta = 10, y = 9) = 0.6527542 \quad (12)$$

즉, 사전확률분포의 5-백분위수는 0.1527176이고 사후확률분포의 5-백분위수는 0.3958057이고, 사전확률분포의 95-백분위수는 0.5400054이고 사후확률분포의 95-백분위수는 0.6527542이다. 이 식들에서 알 수 있듯이, 사후확률분포의 (5-백분위수, 95-백분위수)는 사전확률분포의

(5-백분위수, 95백분위수)보다 더 오른쪽으로 이동하였으며, 백분위수의 범위 또한 좁아졌다.

이 R 프로그램 BayesianRatio101.R을 실행하면, 사전확률밀도함수  $\pi(p|\alpha, \beta)$ 와 사후확률밀도함수  $\pi(p|\alpha, \beta, y)$  각각에서 1024개 난수들을 생성하여 변수 RNpri와 RNpost에 저장한다. 이 난수들이 다음 식들을 만족함을 알 수 있다.

$$\widehat{\Pr}(p \leq 0.5 | \alpha = 5, \beta = 10) = 0.9072266 \quad (13)$$

$$\widehat{\Pr}(p \leq 0.5 | \alpha = 5, \beta = 10, y = 9) = 0.6386719 \quad (14)$$

사전확률추정값  $\widehat{\Pr}(p \leq 0.5 | \alpha = 5, \beta = 10)$ 이 0.9072266라는 것은 사전확률질량의 대부분이  $p = 0.5$  이하에 몰려있음을 의미한다. 그러나, 16개 관찰값들 중 9개가 성공이므로  $p$ 의 사후확률질량은 좀 더 큰 쪽으로 이동함을 알 수 있다, 즉, 사후확률추정값  $\widehat{\Pr}(p \leq 0.5 | \alpha = 5, \beta = 10, y = 9)$ 이 0.6386719이므로 사전확률질량의 많은 부분이  $p > 0.5$ 로 이동했다는 것을 알 수 있다. 다음 식들이 성립한다.

$$\widehat{\Pr}(p \leq 0.05 | \alpha = 5, \beta = 10) = 0.1544059 \quad (15)$$

$$\widehat{\Pr}(p \leq 0.05 | \alpha = 5, \beta = 10, y = 9) = 0.3514593 \quad (16)$$

$$\widehat{\Pr}(p \leq 0.95 | \alpha = 5, \beta = 10) = 0.5415539 \quad (17)$$

$$\widehat{\Pr}(p \leq 0.95 | \alpha = 5, \beta = 10, y = 9) = 0.6047515 \quad (18)$$

즉, 추정된 사전확률분포의 5-백분위수는 0.1544059이고 추정된 사후확률분포의 5-백분위수는 0.3514593이고, 추정된 사전확률분포의 95-백분위수는 0.5415539이고 추정된 사후확률분포의 95-백분위수는 0.6047515이다. 이 값들에서 알 수 있듯이, 추정된 사후확률분포의 90% 신뢰구간인 (5-백분위수, 95백분위수)는 추정된 사전확률분포의 90% 신뢰구간인 (5-백분위수, 95백분위수)보다 더 오른쪽으로 이동하였으며 또한 신뢰구간의 길이가 짧아졌다. 여기서 유의할 점은 관찰점들의 개수가 작은 경우에 추정된 사전확률분포의 신뢰구간이 모수공간을 벗어날 수도 있다는 것이다. 즉, 이 문제에서 추정된 사전확률분포의 신뢰구간은 0을 포함할 수도 있다. 그러나, 사전확률분포가 모수공간 내에서 정의되며 사후확률분포의 지지대는 사전확률분포의 지지대에 포함되므로 사후확률분포의 신뢰구간이 모수공간을 벗어나지 않는다. 이 R 프로그램 BayesianRatio101.R을 실행하면, 그림 1.2.2가 출력된다. 그림 1.2.2에서 상단 그래프에는 추정된 사전확률분포의 히스토그램과 90% 신뢰구간이 그려져 있고, 하단 그래프에는 추정된 사후확률분포의 히스토그램과 90% 신뢰구간이 그려져 있다. ■

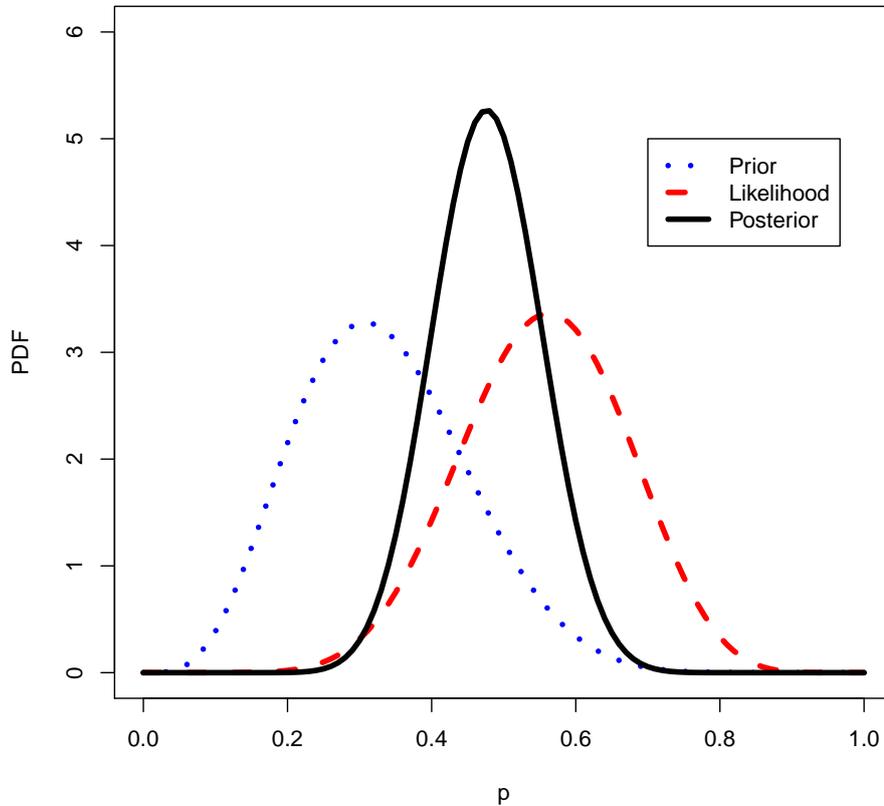


그림 1.2.1. 사전확률밀도함수, 우도함수와 사후확률밀도함수

**예제 1.2.2** 예제 1.2.1을 다시 생각해보자. 즉, 서로 독립인 확률변수들  $x_1, x_2, \dots, x_n$  이 성공확률이  $p$ 인 Bernoulli 확률분포를 따른다고 하자. 모수  $p$ 의 사전확률분포가 베타확률분포  $Beta(\alpha, \beta)$  라고 가정하자. 여기서  $\alpha$ 와  $\beta$ 는 미리 알고 있는 상수들이다. 예제 1.2.1에서 알 수 있듯이,  $p$ 의 사후확률밀도함수는 다음과 같다.

$$\pi(p|y) = \frac{f(y, p)}{f(y)} = \frac{1}{B(y + \alpha, n - y + \beta)} p^{y+\alpha-1} [1 - p]^{n-y+\beta-1} \quad (1)$$

이 사후확률밀도함수  $\pi(p|y)$ 는 모수들을  $\alpha_* \doteq y + \alpha$ 와  $\beta_* \doteq n - y + \beta$ 로 하는 베타확률분포  $Beta(\alpha_*, \beta_*)$ 의 확률밀도함수이다. 꼭 필요한 경우가 아니면, 이후 상수들  $\alpha$ 와  $\beta$ 가 주어졌다는 조건을 사전확률밀도함수, 우도함수 그리고 사후확률밀도함수에 명시하지 않기로 하자.

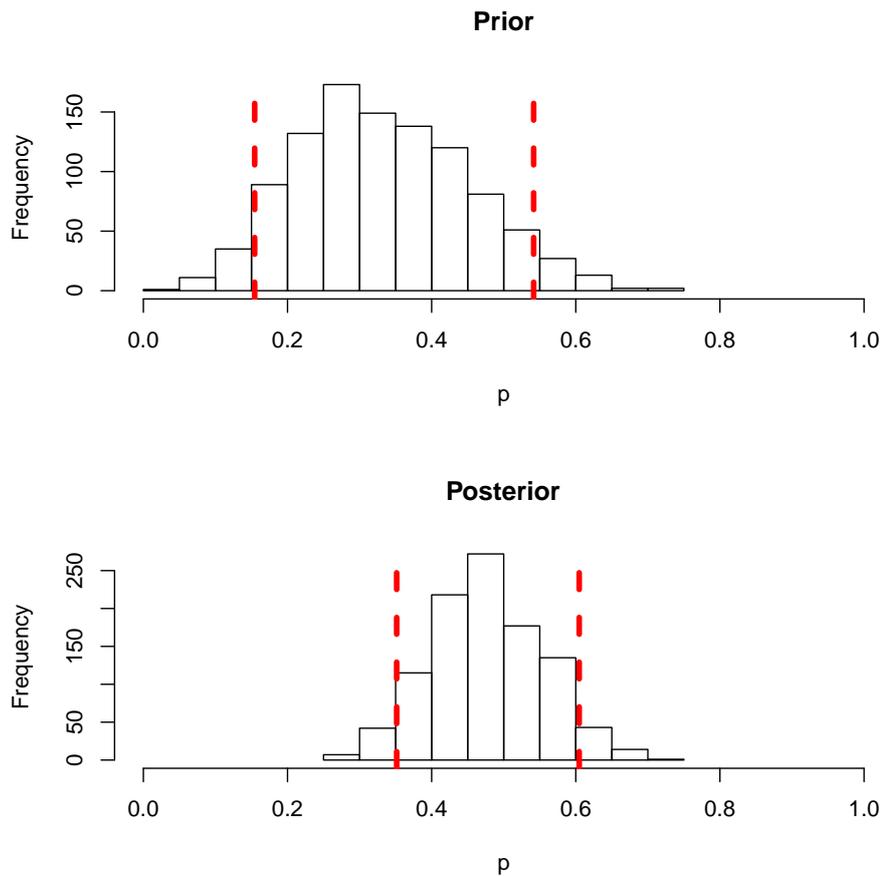


그림 1.2.2. 사전히스토그램과 사후히스토그램

이  $p$ 의 사후확률분포의 평균, 즉 사후평균  $p_*$ 는 다음과 같다.

$$p_* = \frac{\alpha_*}{\alpha_* + \beta_*} = \frac{y + \alpha}{\alpha + \beta + n} \tag{2}$$

식 (2)를 다음과 같이 쓸 수 있다.

$$p_* = \frac{n}{\alpha + \beta + n} \cdot \frac{1}{n} \sum_{i=1}^n x_i + \frac{\alpha + \beta}{\alpha + \beta + n} \cdot \frac{\alpha}{\alpha + \beta} \tag{3}$$

식 (3)에서 알 수 있듯이, 이 사후평균은 우도함수인 이항확률분포의 표본평균  $\frac{1}{n} \sum_{i=1}^n x_i$ 와 사전확률분포인 베타확률분포의 평균  $\alpha/[\alpha + \beta]$ 의 가중평균이며, 이 때 가중값들(weights)은 베타확률분포의 모수들  $\alpha$ 와  $\beta$ , 그리고 표본크기  $n$ 에 의해 결정된다. 만약  $\alpha$ 와  $\beta$ 를 고정시키면서 표본크기  $n$ 을 증가시키면, 표본평균  $\frac{1}{n} \sum_{i=1}^n x_i$ 에 보다 큰 무게를 두게 된다.

사전확률  $p$ 와 사후확률  $p_*$ 를 계산하기 위해서, 다음 R프로그램 BayesianRatio102.R을

실행해 보자.

```

1 # -----
2 # Filename: BayesianRatio102.R
3 # Prior, Likelihood and Posterior by Bayesian Method
4 # Programmed by CBS
5 # -----
6 setEPS()
7 p <- seq(0,1,length=501)
8 ( alpha <- 5); ( beta <- 10 )
9 prior.p <- dbeta(p,alpha,beta)
10 ( n1 <- 16 ); ( y1 <- 3 )
11 posterior1.p <- dbeta(p,alpha+y1,beta+n1-y1)
12 ( n2 <- 16 ); ( y2 <- 3 )
13 posterior2.p <- dbeta(p,alpha+y1+y2,beta+n1-y1+n2-y2)
14 ( n3 <- 16 ); ( y3 <- 3 )
15 posterior3.p <- dbeta(p,alpha+y1+y2+y3,beta+n1-y1+n2-y2+n3-y3)
16 # Plotting
17 plot.new()
18 frame()
19 postscript('BayesianRatio102.eps') # Start to save figure
20 plot(p,prior.p,type="l",xlab="p",ylab="Prior",lty=1,lwd=3,col="black",
21      ylim = c(0,8))
22 lines(p,posterior1.p,lty=2,lwd=3,col=2)
23 lines(p,posterior2.p,lty=3,lwd=3,col=3)
24 lines(p,posterior3.p,lty=4,lwd=3,col=4)
25 legend(0.6,4.5,c("Prior","Posterior 1","Posterior 2","Posterior 3"),
26        lty=c(1:4),lwd=c(3,3,3,3),col=c(1:4))
27 dev.off() # End to save figure
28 # -----

```

이 R 프로그램 BayesianRatio102.R에서 사전확률분포는  $Beta(5, 10)$  이고, 첫 번째 실험에서 관찰값들  $x_1, x_2, \dots, x_{16}$  의 합은  $y_1 = 3$ 이다. 첫 번째 실험 결과, 사후확률분포는  $Beta(8, 23)$ 이다. 이 사후확률분포  $Beta(8, 23)$ 가 두 번째 실험의 사전확률분포이다. 이 두 번째 실험에서 관찰값들  $x_{17}, x_{18}, \dots, x_{32}$  의 합은  $y_2 = 3$ 이다. 따라서 두 번째 실험 결과, 사후확률분포는  $Beta(11, 36)$ 이다. 이 사후확률분포  $Beta(11, 36)$ 가 세 번째 실험의 사전확률분포이다. 이 세 번째 실험에서 관찰값들  $x_{33}, x_{34}, \dots, x_{48}$  의 합은  $y_3 = 3$ 이다. 따라서, 세 번째 실험 결과, 사후확률분포는  $Beta(14, 49)$ 이다. 이 R 프로그램을 실행하면, 그림 1.2.3이 출력된다. 그림 1.2.3에서 흑색 실선은 사전확률밀도함수를, 적색 긴점선은 첫 번째 사후확률밀도함수를, 녹색 점선은 두 번째 사후확률밀도함수를, 그리고 청색 반직선은 세 번째 사후확률밀도함수를 나타낸다. 사후확률밀도함수가 점점 좌측으로 쏠리며 분산이 감소함을 알 수 있다.

다음 Python 프로그램 BayesianRatio102.py을 실행하면, R 프로그램 BayesianRatio102.R을 실행한 것과 동일한 결과를 얻는다.

```
1 # BayesianRatio102.py
```

```

2 """
3 Filename : BayesianRatio102.py
4 Prior, Likelihood and Posterior by Bayesian Method
5 Programmed by C.Han based on the R program by CBS
6 """
7
8 from scipy.stats import beta
9 import numpy as np
10 import matplotlib.pyplot as plt
11
12 p = np.linspace(0, 1, 501)
13
14 a, b = 5, 10
15 prior_p = [beta.pdf(i, a, b) for i in p]
16
17 n1, y1 = 16, 3
18 posterior1_p = [beta.pdf(i, a + y1, b + n1 - y1) for i in p]
19
20 n2, y2 = 16, 3
21 posterior2_p = [beta.pdf(i, a + y1 + y2, b + n1 - y1 + n2 - y2) for i in p]
22
23 n3, y3 = 16, 3
24 posterior3_p = [beta.pdf(i, a + y1 + y2 + y3, b + n1 - y1 + n2 - y2 + n3 - y3)
25                 for i in p]
26
27 plt.xlabel('p')
28 plt.ylabel('PDF')
29 plt.plot(p, prior_p, label = 'Prior', color = 'black', linestyle = '-')
30 plt.plot(p, posterior1_p, label = 'Posterior 1', color = 'blue', linestyle = '--')
31 plt.plot(p, posterior2_p, label = 'Posterior 2', color = 'red', linestyle = '-.')
32 plt.plot(p, posterior3_p, label = 'Posterior 3', color = 'green', linestyle = ':')
33 plt.legend(loc = 5)
34 plt.savefig('BayesianRatio102.png')
35 # End of program
36 # -----

```



**예제 1.2.3** 베이저안접근법에서는 과거 데이터를 사용해서 예측값의 확률밀도함수를 구할 수 있다. 이를 예증하기 위해서, 예제 1.2.2에서 다른 모형을 다시 살펴보자.

관찰값들  $x_1, x_2, \dots, x_n$  이 주어졌을 때  $x_{n+1}$  의 확률밀도함수를 구해보자. 다음 식들이

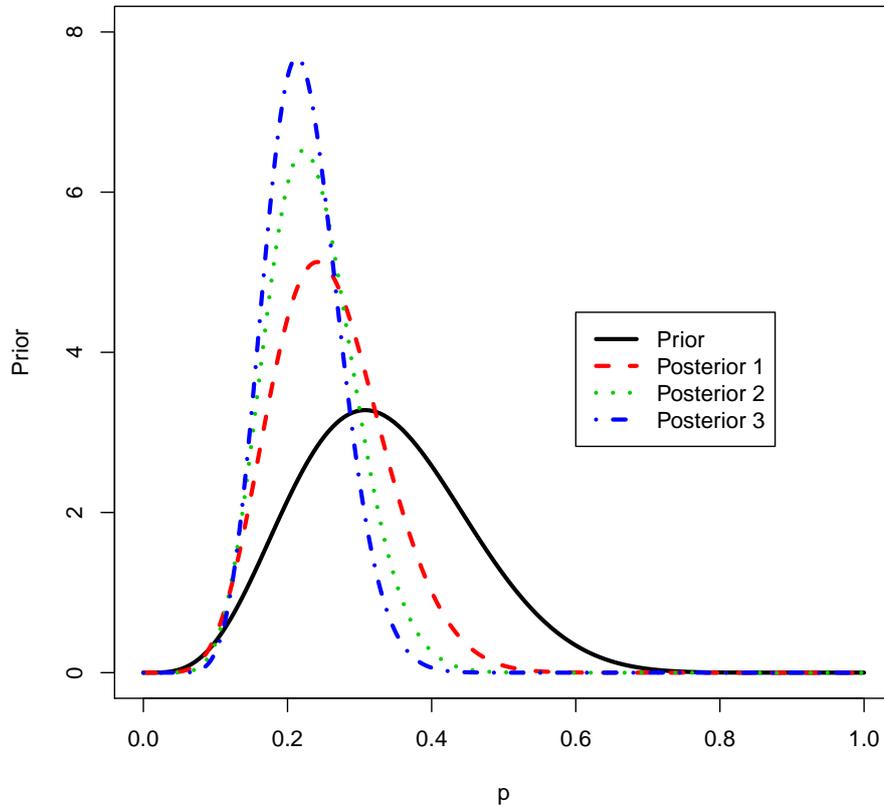


그림 1.2.3. 사전확률밀도함수와 사후확률밀도함수들

성립한다.

$$\begin{aligned}
 & f(x_{n+1} = 1 | x_1, x_2, \dots, x_n) \\
 &= \int_0^1 f(x_{n+1} = 1 | p, x_1, x_2, \dots, x_n) \pi(p | x_1, x_2, \dots, x_n) dp \\
 &= \int_0^1 f(x_{n+1} = 1 | p) \pi(p | x_1, x_2, \dots, x_n) dp \\
 &= \int_0^1 p \pi(p | x_1, x_2, \dots, x_n) dp \tag{1}
 \end{aligned}$$

여기서 두 번째 등호는  $p$ 가 주어졌을 때  $x_1, x_2, \dots, x_n, x_{n+1}$ 이 서로 독립이어서 성립한다. 식 (1)에서 알 수 있듯이, 다음 식들이 성립한다.

$$f(x_{n+1} = 1 | x_1, x_2, \dots, x_n) = E(p | x_1, x_2, \dots, x_n) = p_* = \frac{\alpha + \sum_{i=1}^n x_i}{\alpha + \beta + n} \tag{2}$$

여기서 세 번째 등호는 예제 1.2.2의 식 (2)에 의해서 성립한다. 식 (2)에서 알 수 있듯이, 다음 식이 성립한다.

$$f(x_{n+1} = 0 | x_1, x_2, \dots, x_n) = 1 - p_* = \frac{\beta + n - \sum_{i=1}^n x_i}{\alpha + \beta + n} \quad (3)$$

만약  $\alpha$ 를 성공횟수로  $\beta$ 를 실패횟수를 간주하면, 직관적으로 보아 식 (2)와 식 (3)은 당연한 것이다.

두 번째 예로서, 관찰값들  $x_1, x_2, \dots, x_n$ 이 주어졌을 때 예측변수  $z \doteq \sum_{j=1}^N x_{n+j}$ 의 확률밀도함수를 구해보자. 다음 식들이 성립한다.

$$\begin{aligned} f(z | x_1, x_2, \dots, x_n) &= \int_0^1 f(z | p, x_1, x_2, \dots, x_n) \pi(p | x_1, x_2, \dots, x_n) dp \\ &= \int_0^1 f(z | p) \pi\left(p \mid \sum_{i=1}^n x_i\right) dp \end{aligned} \quad (4)$$

여기서 두 번째 등호는  $p$ 가 주어졌을 때  $x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_{n+N}$ 이 서로 독립이고 또한  $\sum_{i=1}^n x_i$ 는 충분통계량이어서 성립한다. 예제 1.2.1에서 알 수 있듯이, 다음 식들이 성립한다.

$$f(z | p) = \binom{N}{z} p^z [1 - p]^{N-z} \quad (5)$$

$$\pi\left(p \mid \sum_{i=1}^n x_i\right) = \frac{\Gamma(\alpha + \beta + n)}{\Gamma\left(\alpha + \sum_{i=1}^n x_i\right) \Gamma\left(\beta + n - \sum_{i=1}^n x_i\right)} p^{\alpha + \sum_{i=1}^n x_i - 1} [1 - p]^{\beta + n - \sum_{i=1}^n x_i - 1} \quad (6)$$

식 (5)와 식 (6)을 식 (4)에 대입하면, 다음 식들을 얻는다.

$$\begin{aligned}
 f(z|x_1, x_2, \dots, x_n) &= \binom{N}{z} \frac{\Gamma(\alpha + \beta + n)}{\Gamma\left(\alpha + \sum_{i=1}^n x_i\right) \Gamma\left(\beta + n - \sum_{i=1}^n x_i\right)} \\
 &\quad \cdot \int_0^1 p^{z+\alpha+\sum_{i=1}^n x_i-1} [1-p]^{N-z+\beta+n-\sum_{i=1}^n x_i-1} dp \\
 &= \binom{N}{z} \frac{1}{B\left(\alpha + \sum_{i=1}^n x_i, \beta + n - \sum_{i=1}^n x_i\right)} \\
 &\quad \cdot B\left(z + \alpha + \sum_{i=1}^n x_i, N - z + \beta + n - \sum_{i=1}^n x_i\right) \quad (7)
 \end{aligned}$$

식 (7)에 기술한 예측확률을 구하는 예로서, 다음 R프로그램 BayesianRatio103.R을 실행해 보자.

```

1 # -----
2 # Filename: BayesianRatio103.R
3 # Predicted Probabilities
4 # Programmed by CBS
5 # -----
6 setEPS()
7 ( alpha <- 5 ); ( beta <- 10 )
8 ( n <- 64 ); ( y <- 23 ) # y = sum x
9 ( N = 8 )
10 ( z <- c(0:N) )
11 ( predict.p <- choose(N,z)*beta(z+alpha+y,N-z+beta+n-y)
12     /beta(alpha+y,beta+n-y) )
13 # Plotting
14 plot.new()
15 frame()
16 postscript('BayesianRatio103.eps') # Start to save figure
17 plot(z, predict.p,type="h", pch=15, col="red", xlab="z",lwd=10,
18     ylab="probability", ylim = c(0,0.3))
19 axis(1, side=3, line=4, font.axis=2, font.lab=2, cex.lab=1.3, cex.axis=1.2)
20 dev.off() # End to save figure
21 # -----

```

이 R프로그램 BayesianRatio103.R에서 사전확률분포는  $Beta(5, 10)$  이고, 관찰값들  $x_1, x_2, \dots, x_{64}$ 의 합은  $y = \sum_{i=1}^{64} x_i = 23$ 이다. 새로운 관찰값들의 합  $z = \sum_{i=1}^8 x_{64+i}$ 의 지지대는  $\{0, 1, \dots, 8\}$ 이다. 이 R프로그램 BayesianRatio103.R을 실행하면, 예측변수  $z$ 의 사후확률밀도함수가 출력된다. 이 사후확률밀도함수가 그림 1.2.4에 그려져 있다. ■

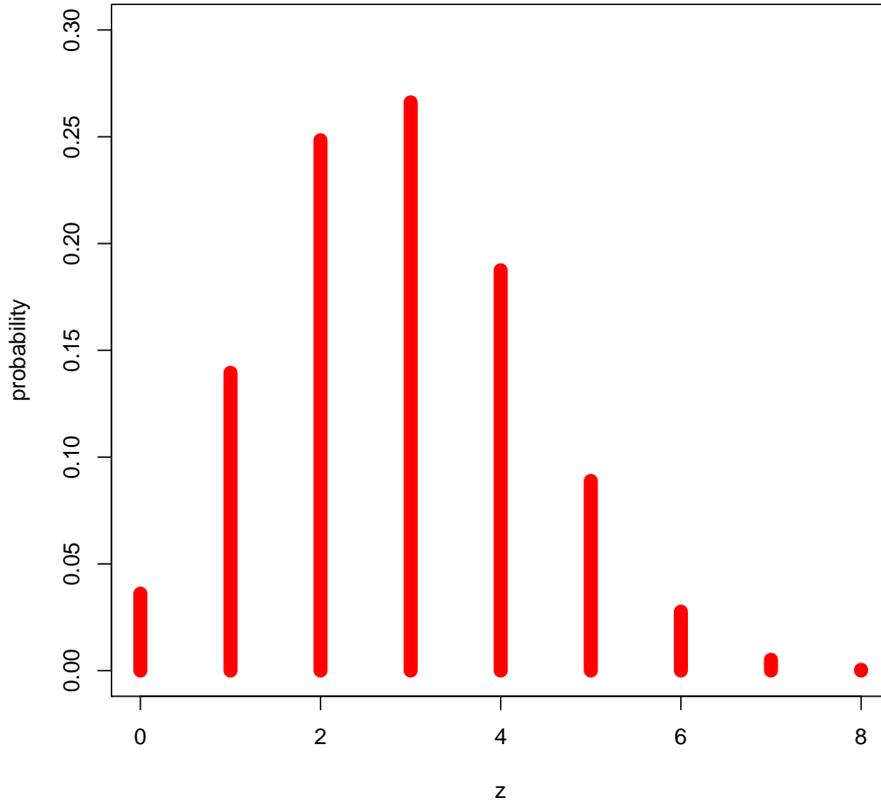


그림 1.2.4. 예측변수의 사후확률밀도함수

**예제 1.2.4** 예제 1.2.3에서 해석적으로 구한 예측변수  $z$ 의 사후확률밀도함수를 몬테카를로 방법으로 구해보자.

예제 1.2.3의 식 (4)에서 알 수 있듯이, 관찰값들  $x_1, x_2, \dots, x_n$ 이 주어졌을 때 예측변수  $z \doteq \sum_{j=1}^N x_{n+j}$ 의 조건부확률밀도함수는 다음과 같다.

$$f(z|x_1, x_2, \dots, x_n) = \int_0^1 f(z|p) \pi(p|y) dp \tag{1}$$

여기서  $y = \sum_{i=1}^n x_i$ 이다. 식 (1)을 다음과 같이 쓸 수 있다.

$$f(z|x_1, x_2, \dots, x_n) = E_{p|y}(f(z|p)) \tag{2}$$

예제 1.2.3의 식 (5)에서 알 수 있듯이, 다음 식이 성립한다.

$$f(z|p) = \binom{N}{z} p^z [1-p]^{N-z} \quad (3)$$

식 (2)에서 알 수 있듯이, 약대수법칙에 의해 다음 식이 성립한다.

$$f(z|x_1, x_2, \dots, x_n) = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=1}^K f(z|p_k) \quad (4)$$

여기서  $p_1, p_2, \dots$  는 사후확률밀도함수  $\pi(p|y)$ 에서 추출된 난수들이다. 예제 1.2.1에서 알 수 있듯이,  $\pi(p|y)$ 는 모수들을  $\alpha_* = y + \alpha$ 와  $\beta_* = n - y + \beta$ 로 하는 베타확률분포  $Beta(\alpha_*, \beta_*)$ 의 확률밀도함수이다. 즉, 다음 식이 성립한다.

$$\pi(p|y) = \frac{1}{B(y + \alpha, n - y + \beta)} p^{y+\alpha-1} [1-p]^{n-y+\beta-1} \quad (5)$$

식 (4)에 기술한 예측변수의 사후확률밀도함수, 즉 예측확률을 구하기 위해서, 다음 R 프로그램 BayesianRatio104.R을 실행해 보자.

```

1 # -----
2 # Filename: BayesianRatio104.R
3 # Predicted Probabilities Through Monte Carlo Method
4 # Programmed by CBS
5 # -----
6 setEPS()
7 ( alpha <- 5 ); ( beta <- 10 )
8 ( n <- 64 ); ( y <- 23 ) # y = sum x
9 ( N = 8)
10 ( z <- c(0:N) )
11 ( K <- 10 )
12 f_z_p <- matrix(0,K,N)
13 # Simulations
14 p_y <- matrix(rbeta(K,alpha+y,n-y+beta), ncol = 1, nrow = K)
15 for(kk in 1:K){
16     f_z_p[kk,] <- dbinom(z,size=K,prob=p_y[kk,]) }
17 f_z_p
18 ( predicted.p <- colMeans(f_z_p) )
19 sum(predicted.p)
20 # Plotting
21 plot.new()
22 frame()
23 postscript('BayesianRatio104.eps') # Start to save figure
24 plot(z, predicted.p,type="h", pch=16, col="red", xlab="z",
25     ylab="probability", ylim=c(0,0.3), lwd=10 )
26 dev.off() # End to save figure
27 # -----

```

이 R 프로그램 BayesianRatio104.R에서 사전확률분포는  $Beta(5, 10)$ 이고, 첫 번째 실험

에서 관찰값들  $x_1, x_2, \dots, x_{64}$  의 합은  $y = 23$ 이다. 예측값들의 합  $z = \sum_{i=1}^8 x_{64+i}$  의 지지대는  $\{0, 1, \dots, 8\}$ 이다.

이 R 프로그램 BayesianRatio104.R을 실행하면, 식 (4)를 바탕으로 한 몬테카를로법을 통해서 예측변수  $z$ 의 사후확률밀도함수가 추정된다. 이 사후확률밀도함수가 그림 1.2.5에 출력된다. 그림 1.2.5는 그림 1.2.4와 비슷함을 알 수 있다. ■

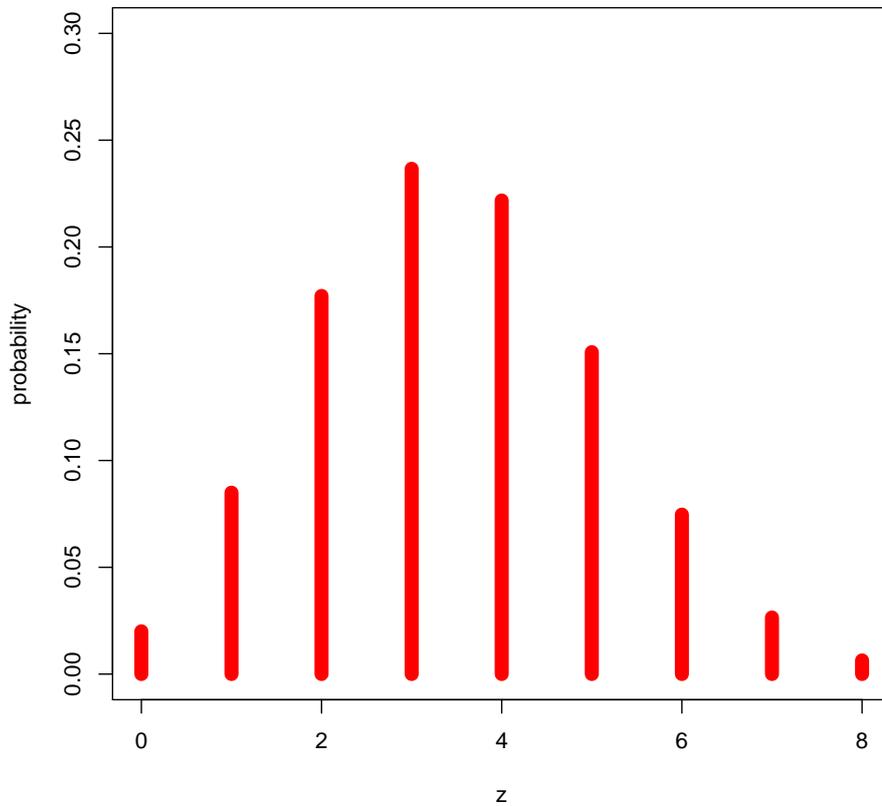


그림 1.2.5. 시뮬레이션과 예측확률 1

**예제 1.2.5** 예제 1.2.4와는 다른 몬테카를로법을 사용해서, 예제 1.2.3에서 해석적으로 구한 예측확률을 구해보자.

예제 1.2.3의 식 (4)에서 알 수 있듯이, 관찰값들  $x_1, x_2, \dots, x_n$ 이 주어졌을 때 예측변수  $z = \sum_{j=1}^N x_{n+j}$ 의 확률밀도함수는 다음과 같다.

$$f(z|x_1, x_2, \dots, x_n) = \int_0^1 f(z|p) \pi(p|y) dp \tag{1}$$

여기서  $y = \sum_{i=1}^n x_i$  이다. 다음 식이 성립한다.

$$f_{z|y}(a|y) = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M 1_{\{z_m\}}(a), \quad (a = 0, 1, \dots, N) \quad (2)$$

여기서  $z_m$  은 확률밀도함수  $f(z|p_m)$ , 즉 이항확률분포  $Binomial(N, p_m)$ 에서 생성된 난수이고,  $p_m$  은 사후확률밀도함수  $\pi(p|y)$ , 즉 베타확률분포  $Beta(y + \alpha, n - y + \beta)$ 에서 생성된 난수이다.

식 (2)를 바탕으로 예측확률을 구하기 위해서, 다음 R프로그램 BayesianRatio105.R을 실행해 보자.

```

1 # -----
2 # Filename: BayesianRatio105.R
3 # Predicted Probabilities Through Monte Carlo Method
4 # Programmed by CBS
5 # -----
6 setEPS()
7 ( alpha <- 5 ); ( beta <- 10 )
8 ( n <- 64 ); ( y <- 23 ) # y = sum x
9 ( N = 8 )
10 ( z <- c(0:N) )
11 ( K <- 1024 )
12 M <- (N+1)*K
13 # Simulations
14 p_y <- rbeta(M, alpha+y, n-y+beta)
15 z_p <- rbinom(M, N, p_y)
16 ( z_p_freq = table(z_p)/NK )
17 ( sum(z_p_freq) )
18 # Plotting
19 plot.new()
20 frame()
21 postscript('BayesianRatio105a.eps') # Start to save figure
22 plot(p_y ~ z_p, col=1, pch=21, xlab="z", ylab="Posterior Probability",
23      xlim=c(0, N) )
24 dev.off() # End to save figure
25 plot.new()
26 frame()
27 postscript('BayesianRatio105b.eps') # Start to save figure
28 plot(z_p_freq, type="h", col="red", xlab="z", ylab="probability", xlim=c(0, N),
29      lwd=10)
30 dev.off() # End to save figure
31
32 # Plotting
33 plot.new()
34 frame()
35 postscript('BayesianRatio105a.eps') # Start to save figure
36 plot(p_y ~ z_p, col=1, pch=21, xlab="z", ylab="Posterior Probability",
37      xlim=c(0, N) )
38 dev.off() # End to save figure
39 plot.new()
40 frame()
41 postscript('BayesianRatio105b.eps') # Start to save figure
42 plot(histt, type="h", col="red", xlab="z", ylab="probability", xlim=c(0, N),

```

```

43 |     lwd=10)
44 | dev.off()           # End to save figure
45 | # -----
    
```

이 R 프로그램 BayesianRatio105.R에서 사전확률분포는  $Beta(5, 10)$  이고, 첫 번째 실험에서 관찰값들  $x_1, x_2, \dots, x_{64}$  의 합은  $y = 23$  이다. 새로운 관찰값들의 합  $z = \sum_{i=1}^8 x_{64+i}$  의 지지대는  $\{0, 1, \dots, 8\}$  이다.

이 R 프로그램 BayesianRatio105.R을 실행하면, 식 (2)를 바탕으로 한 몬테카를로실험을 하기 위해서  $M = [N + 1]K$  개 난수벡터들  $\{p_m, z_m\}$  을 생성한다. 이 난수벡터들의 산점도가 그림 1.2.6에 그려져 있다. 그림 1.2.6에서 각  $z$  값에 대한 사후확률이 어떻게 분포하는지를 짐작할 수 있다. R명령어 table을 사용해서 난수들  $\{z_m\}$  의 분할표(contingency table)를 만든 다음, 이를 바탕으로 그린 난수들  $\{z_m\}$  의 히스토그램이 그림 1.2.7에 그려져 있다. 그림 1.2.7이 확률변수  $z$  의 예측확률, 즉 추정된 사후확률밀도함수이다. 그림 1.2.7는 그림 1.2.4와 아주 비슷함을 알 수 있다. ■

### 1.2.2 유용한 확률분포들

이 소절에서는 베이지안통계분석에 유용한 확률분포들에 대해 살펴보자. 이 확률분포들에 대한 자세한 내용은 최병선 [11]을 참조하라.

**정의 1.2.1: 이산형 일양확률분포**

지지대가  $\{m, m + 1, \dots, n\}$  인 이산형 일양확률분포의 확률질량함수는 다음과 같다.

$$p(x) = \frac{1}{n - m + 1} 1_{\{m, m+1, \dots, n\}}(x)$$

이 이산형 일양확률분포의 평균, 분산 그리고 특성함수는 각각 다음과 같다.

$$E(x) = \frac{n + m}{2}, \quad Var(x) = \frac{[n - m + 1]^2 - 1}{12} \tag{1.2.7}$$

$$E(e^{i\theta x}) = \frac{e^{im\theta} - e^{i[n+1]\theta}}{[n - m + 1][1 - e^{i\theta}]} \tag{1.2.8}$$

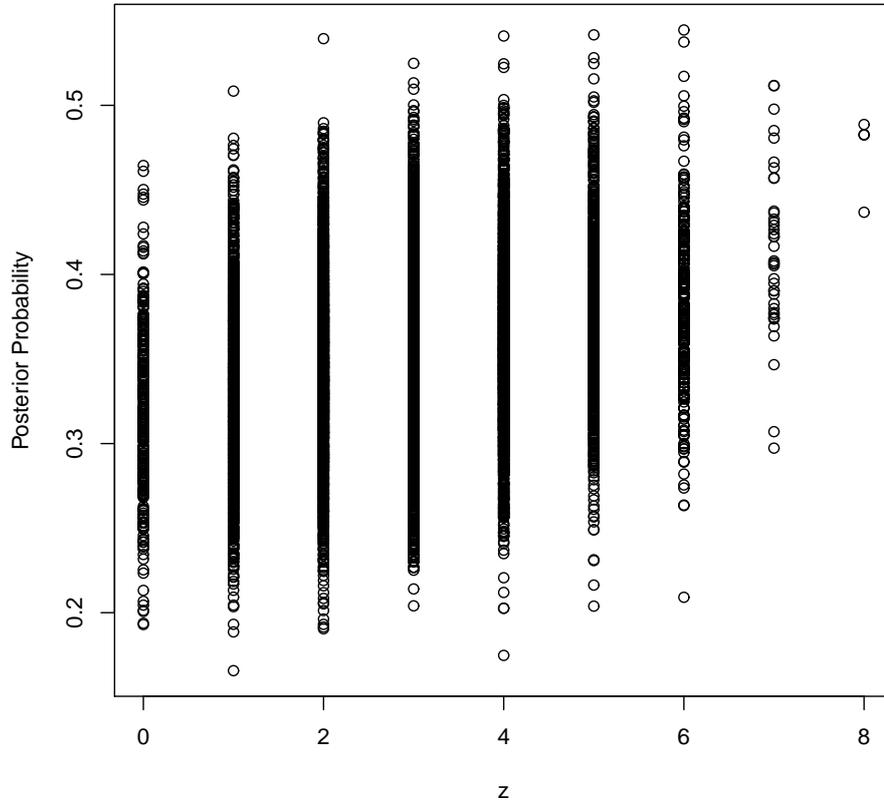


그림 1.2.6. 난수들의 산점도

**정의 1.2.2: Bernoulli 확률분포**

지지대가  $\{0, 1\}$  인 Bernoulli 확률분포의 확률질량함수는 다음과 같다.

$$p(x) = \begin{cases} p, & (x = 1) \\ 1 - p, & (x = 0) \end{cases}$$

이 Bernoulli 확률분포의 평균, 분산 그리고 특성함수는 각각 다음과 같다.

$$E(x) = p, \quad Var(x) = p[1 - p] \tag{1.2.9}$$

$$E(e^{i\theta x}) = 1 - p + pe^{i\theta} \tag{1.2.10}$$

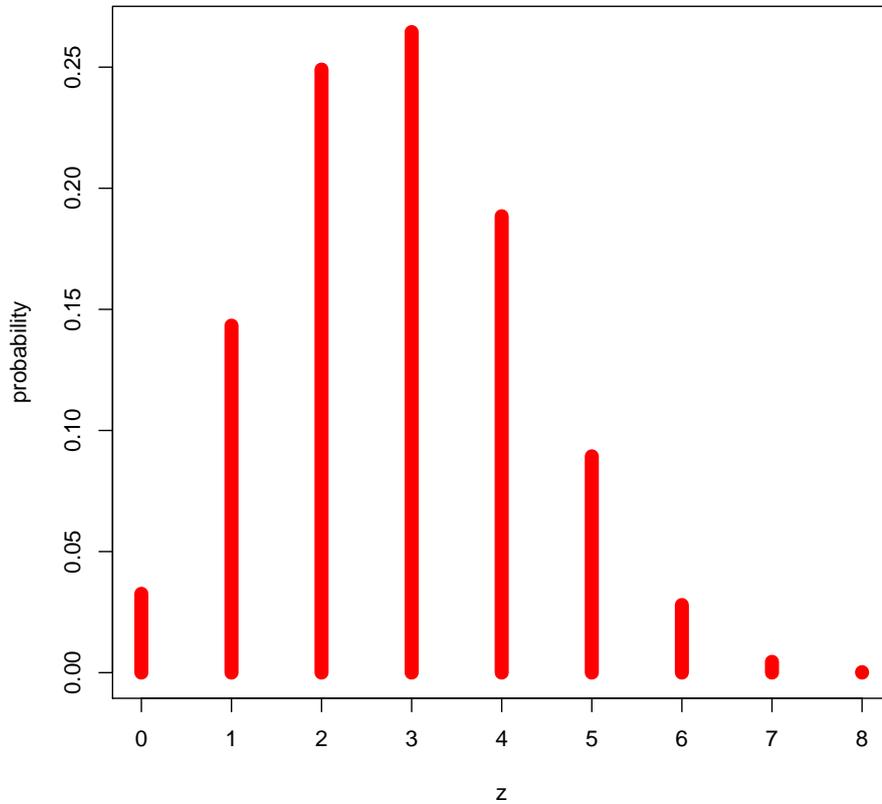


그림 1.2.7. 예측확률의 히스토그램

**정의 1.2.3: 이항확률분포**

서로 독립이고 성공확률이  $p$ 인 Bernoulli 확률변수들  $v_1, v_2, \dots, v_n$ 의 합  $x \doteq \sum_{i=1}^n v_i$ 를 이항확률변수라 부르고, 이에 해당하는 이항확률분포 (binomial probability distribution)를  $Binominal(n, p)$ 로 표기하자. 이항확률분포의 확률질량함수는 다음과 같다.

$$f(x) = \binom{n}{x} p^x [1 - p]^{n-x} 1_{\{0,1,\dots,n\}}(x)$$

이 이항확률분포의 평균, 분산 그리고 특성함수는 각각 다음과 같다.

$$E(x) = np, \quad Var(x) = np[1 - p] \tag{1.2.11}$$

$$E(e^{i\theta x}) = [1 - p + pe^{i\theta}]^n \tag{1.2.12}$$

**정의 1.2.4: 기하확률분포**

성공확률이  $p$ 인 Bernouille시행에서 처음으로 성공할 때까지 필요한 실패횟수  $x$ 는 기하확률분포 (geometric probability distribution)를 따른다고 한다. 이 확률변수의 확률질량함수는 다음과 같다.

$$p(x) = p[1 - p]^{x-1} 1_{\{1, 2, \dots\}}(x)$$

이 기하확률질량함수 (geometric mass function)에 해당하는 기하확률분포함수는 다음과 같다.

$$F(x) = \{1 - [1 - p]^{x-1}\} 1_{\{1, 2, \dots\}}(x) \quad (1.2.13)$$

이 기하확률분포의 평균, 분산 그리고 특성함수는 각각 다음과 같다.

$$E(x) = \frac{1 - p}{p}, \quad Var(x) = \frac{1 - p}{p^2} \quad (1.2.14)$$

$$E(e^{i\theta x}) = \frac{p}{1 - [1 - p]e^{i\theta}} \quad (1.2.15)$$

**정의 1.2.5: Pascal 확률분포**

성공확률이  $p$ 인 Bernouille시행에서 제  $r$ 번째 성공을 할 때까지 필요한 실험횟수  $x$ 는 Pascal 확률분포를 따른다고 한다. 이 확률변수의 확률질량함수는 다음과 같다.

$$p(x) = \binom{x-1}{r-1} p^r [1 - p]^{x-r} 1_{\{r, r+1, \dots\}}(x)$$

이 Pascal 확률분포의 평균, 분산 그리고 특성함수는 각각 다음과 같다.

$$E(x) = \frac{r}{p}, \quad Var(x) = \frac{r[1 - p]}{p^2} \quad (1.2.16)$$

$$E(e^{i\theta x}) = \left\{ \frac{pe^{i\theta}}{1 - [1 - p]e^{i\theta}} \right\}^r \quad (1.2.17)$$

Pascal 확률분포는 다음에 정의되는 음이항확률분포와 동등(equivalent)하다.

**정의 1.2.6: 음이항확률분포**

성공확률이  $p$  인 Bernouille시행에서 제  $r$  번째 성공을 할 때까지 필요한 실패횟수  $x$  는 음이항확률분포 (negative binomial distribution) 를 따른다고 한다. 이 확률변수  $x$  의 확률질량함수는 다음과 같다.

$$p(x) = \binom{r+x-1}{r-1} p^r [1-p]^x 1_{\{0,1,\dots\}}(x)$$

다음 식이 성립함을 쉽게 증명할 수 있다.

$$\begin{aligned} \binom{r+x-1}{r-1} &= \frac{(r+x-1)!}{(r-1)!x!} = \frac{r[r+1]\cdots[r+x-1]}{x!} \\ &= [-1]^x \frac{[-r][-r-1]\cdots[-r-x+1]}{x!} = [-1]^x \frac{(-r)!}{(-r-x)!x!} = [-1]^x \binom{-r}{x} \end{aligned} \quad (1.2.18)$$

식 (1.2.18) 을 이용하면, 이 음이항확률분포의 평균, 분산 그리고 특성함수가 다음과 같음을 증명할 수 있다.

$$E(x) = \frac{r[1-p]}{p}, \quad Var(x) = \frac{r[1-p]}{p^2} \quad (1.2.19)$$

$$E(e^{i\theta x}) = \left\{ \frac{p}{1-[1-p]e^{i\theta}} \right\}^r \quad (1.2.20)$$

**정의 1.2.7: Poisson 확률분포**

서로 독립적인 사건들이 발생하는 시간간격이 지수확률분포를 따를 때, 어떤 시간구간에서 이 사건이 발생하는 횟수는 Poisson 확률분포를 따른다고 한다. 단위시간당 사건이 발생하는 평균횟수인 모수가  $\lambda (> 0)$  인 Poisson 확률분포의 확률질량함수는 다음과 같다.

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!} 1_{\{0,1,\dots\}}(x)$$

이 Poisson 확률분포의 평균, 분산 그리고 특성함수는 다음과 같다.

$$E(x) = \lambda, \quad Var(x) = \lambda \tag{1.2.21}$$

$$E\left(e^{i\theta x}\right) = \exp\left(\lambda\left[e^{i\theta} - 1\right]\right) \tag{1.2.22}$$

**정의 1.2.8: 초기하확률분포**

다음 확률질량함수를 갖는 확률변수를 초기하확률변수 (hypergeometric random variable) 라 부른다.

$$p(x) = \frac{\binom{K}{x} \binom{M-K}{N-x}}{\binom{M}{N}} 1_{\{\max\{0, K-M+N\}, \dots, \min\{K, N\}\}}(x)$$

여기서  $M \in \mathbb{Z}_{\geq 0}$ ,  $0 \leq K \leq M$ 이고  $0 \leq N \leq M$ 이다.

이 초기하확률분포의 평균과 분산은 각각 다음과 같다.

$$E(x) = N \frac{K}{M}, \quad Var(x) = N \frac{K[M-K][M-N]}{M^2[M-1]} \tag{1.2.23}$$

이항정리를 확장하면, 다음 식을 얻는다.

$$\left[ \sum_{j=1}^k a_j \right]^n = \sum_i \frac{n!}{\prod_i x_i!} \prod_{i=1}^k a_i^{x_i} \tag{1.2.24}$$

식 (1.2.24)를 이용해서, 다음과 같이 다항확률분포 (multinomial probability distribution) 를 정의한다.

**정의 1.2.9: 다항확률분포**

다음 확률질량함수를 갖는 확률변수를 다항확률변수 (multinomial random variable) 라

부른다.

$$p(x_1, \dots, x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}, \quad \left( x_1 \geq 0, \dots, x_k \geq 0, \sum_{i=1}^k x_i = n \right)$$

여기서  $p_1 \geq 0, \dots, p_k \geq 0$ 는 식  $\sum_{i=1}^k p_i = 1$ 을 만족한다.

이 다항확률분포의 평균, 분산 그리고 공분산은 각각 다음과 같다.

$$E(x_i) = np_i, \quad Var(x_i) = np_i[1 - p_i], \quad Cov(x_i, x_j) = -np_i p_j, \quad (i \neq j) \quad (1.2.25)$$

또한 특성함수는 다음과 같다.

$$E \left( \exp \left( i \sum_{l=1}^k \theta_l x_l \right) \right) = \left[ \sum_{l=1}^k p_l \exp(i\theta_l) \right]^n \quad (1.2.26)$$

**정의 1.2.10: 연속형 일양확률분포**

지지대가  $(a, b)$ 인 일양확률분포  $Uniform(a, b)$ 의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{b - a} 1_{(a,b)}(x)$$

이 연속형 일양확률변수의 확률분포함수  $F(x)$ 는 다음과 같다.

$$F(x) = \frac{x - a}{b - a} 1_{(a,b)}(x) + 1_{[b, \infty)}(x) \quad (1.2.27)$$

이 일양확률분포의 평균, 분산 그리고 특성함수는 각각 다음과 같다.

$$E(x) = \frac{a + b}{2}, \quad Var(x) = \frac{[b - a]^2}{12} \quad (1.2.28)$$

$$E(e^{i\theta x}) = \frac{e^{i\theta b} - e^{i\theta a}}{i\theta[b - a]} \quad (1.2.29)$$

**정의 1.2.11: 정규확률분포**

모수벡터가  $[\mu, \sigma]$ 인 정규확률분포 (normal probability distribution)의 확률밀도함수는 다음과 같다.

$$n(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{[x - \mu]^2}{2\sigma^2}\right)$$

이 정규확률분포의 확률분포함수를  $N(x; \mu, \sigma^2)$ 로 표기하자. 즉, 다음 식이 성립한다.

$$N(x; \mu, \sigma^2) = \int_{-\infty}^x n(z; \mu, \sigma^2) dz$$

확률변수  $x$ 가 이 정규확률분포를 따른다는 것을 다음과 같이 표기하자.

$$x \stackrel{d}{\sim} \mathcal{N}(\mu, \sigma^2)$$

이 정규확률분포의 평균은  $\mu$ 이고 분산은  $\sigma^2$ 이다. 평균이 0이고 분산이 1인 정규확률분포를 표준정규확률분포 (standard normal probability distribution)라 하고, 이에 해당하는 확률밀도함수와 확률분포함수를 각각  $n(x)$ 와  $N(x)$ 로 표기하자.

**정의 1.2.12: 지수확률분포**

크기모수 (scale parameter)가  $\mu$ 인 지수확률분포 (exponential probability distribution)의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\mu} \exp\left(-\frac{x}{\mu}\right) 1_{[0, \infty)}(x)$$

크기모수의 역수  $1/\mu$ 를 비율 (rate)이라고 한다.

이 지수확률변수의 확률분포함수는 다음과 같다.

$$F(x) = \left[1 - \exp\left(-\frac{x}{\mu}\right)\right] 1_{[0, \infty)}(x) \quad (1.2.30)$$

이 지수확률분포의 평균과 분산은 각각 다음과 같다.

$$E(x) = \mu, \quad Var(x) = \mu^2 \tag{1.2.31}$$

또한, 적률생성함수 (moment generating function) 와 특성함수 (characteristic function) 는 각각 다음과 같다.

$$m_x(t) = [1 - \mu t]^{-1}, \quad \left(t < \frac{1}{\mu}\right), \quad E\left(e^{i\theta x}\right) = [1 - i\mu\theta]^{-1} \tag{1.2.32}$$

지수확률분포는 대기행렬모형 (queueing model) 에서 자주 사용된다.

서로 독립인 확률변수들  $v_1, v_2, \dots, v_m$  이 평균이  $1/\beta$  인 지수확률분포를 따르면, 확률변수  $x = \sum_{i=1}^m v_i$  는 형태모수 (shape parameter) 가  $m$  이고, 비율 (rate) 이  $\beta$  이며, 크기모수 (scale parameter) 가  $1/\beta$  인 Erlang 확률분포를 따른다고 한다. 이 Erlang 확률변수의 적률생성함수와 특성함수는 각각 다음과 같다.

$$m_x(t) = \left[1 - \frac{t}{\beta}\right]^{-m}, \quad (t < \beta), \quad E\left(e^{i\theta x}\right) = \left[1 - i\frac{\theta}{\beta}\right]^{-m} \tag{1.2.33}$$

이 적률생성함수 또는 특성함수를 바탕으로 다음과 같은 정의를 할 수 있다.

**정의 1.2.13: Erlang 확률분포**

형태모수가 자연수  $m$  이고 크기모수가  $1/\beta$  인 Erlang 확률변수의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\Gamma(m)} \beta^m x^{m-1} \exp(-\beta x) 1_{[0, \infty)}(x)$$

부분적분을 사용하면, 이 Erlang 확률분포의 확률분포함수가 다음과 같음을 알 수 있다.

$$F(x) = \left[1 - \exp(-x\beta) \sum_{i=0}^{m-1} \frac{1}{\Gamma(i+1)} [\beta x]^i\right] 1_{[0, \infty)}(x) \tag{1.2.34}$$

이 Erlang 확률분포의 평균과 분산은 각각 다음과 같다.

$$E(x) = \frac{m}{\beta}, \quad Var(x) = \frac{m}{\beta^2} \tag{1.2.35}$$

Erlang 확률분포도 대기행렬모형에서 자주 사용된다.

Erlang 확률분포에서 형태모수 (shape parameter)  $m$  을 자연수에서 실수  $\alpha (> 0)$  로 확장한 것을 감마확률분포 (gamma probability distribution) 라 한다.

**정의 1.2.14: 감마확률분포**

형태모수가 실수  $\alpha (> 0)$  이고 크기모수가  $1/\beta$  인 감마확률변수의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\Gamma(\alpha)} \beta^\alpha x^{\alpha-1} \exp(-\beta x) 1_{[0, \infty)}(x)$$

감마확률변수  $x$  가 이 감마확률분포를 따르는 것을 다음과 같이 표기하자.

$$x \stackrel{d}{\sim} \text{Gamma}(\alpha, \beta)$$

이 감마확률분포의 적률생성함수와 특성함수는 각각 다음과 같다.

$$m_x(t) = \left[1 - \frac{t}{\beta}\right]^{-\alpha}, \quad (t < \beta), \quad E\left(e^{i\theta x}\right) = \left[1 - i\frac{\theta}{\beta}\right]^{-\alpha} \quad (1.2.36)$$

여기서 하불완비감마함수 (lower incomplete gamma function)  $\gamma(s, z)$  와 상불완비감마함수 (upper incomplete gamma function)  $\Gamma(s, z)$  는 각각 다음과 같이 정의된다.

$$\gamma(s, z) \doteq \int_0^z t^{s-1} e^{-t} dt, \quad \Gamma(s, z) \doteq \int_z^\infty t^{s-1} e^{-t} dt \quad (1.2.37)$$

이 감마확률분포의 확률분포함수는 다음과 같다.

$$F_\Gamma(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \gamma(\alpha, \beta x) 1_{[0, \infty)}(x) \quad (1.2.38)$$

또한, 평균과 분산은 각각 다음과 같다.

$$E(x) = \frac{\alpha}{\beta}, \quad \text{Var}(x) = \frac{\alpha}{\beta^2} \quad (1.2.39)$$

**정의 1.2.15: 일반지수확률분포**

일반지수확률분포 (generalized exponential probability distribution) 의 확률밀도함수는 다음과 같다.

$$f_{GE}(x; \alpha, \beta) = \frac{\alpha}{\beta} \exp\left(-\frac{x}{\beta}\right) \left[1 - \exp\left(-\frac{x}{\beta}\right)\right]^{\alpha-1} 1_{[0, \infty)}(x)$$

여기서  $\alpha$ 와  $\beta$ 는 양수들이다.

이 일반지수확률분포의 확률분포함수는 다음과 같다.

$$F_{GE}(x; \alpha, \beta) = \left[1 - \exp\left(-\frac{x}{\beta}\right)\right]^{\alpha} 1_{[0, \infty)}(x) \quad (1.2.40)$$

또한, 이 일반지수확률분포함수의 역함수는 다음과 같다.

$$F_{GE}^{-1}(x; \alpha, \beta) = -\beta \ln\left(1 - x^{1/\alpha}\right) 1_{[0, 1)}(x) \quad (1.2.41)$$

**정의 1.2.16: 카이제곱확률분포**

서로 독립인 표준정규확률변수들  $v_1, v_2, \dots, v_\nu$  의 제곱합  $x \doteq \sum_{i=1}^{\nu} v_i^2$  를 자유도가  $\nu$  인 카이제곱확률변수라 부르고, 이에 해당하는 카이제곱확률분포를  $\chi_\nu^2$  로 표기하자.

자유도가  $\nu$  인 카이제곱확률변수 (Chi-squared random variable) 의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{2^{\nu/2} \Gamma(\frac{\nu}{2})} x^{[\nu-2]/2} \exp\left(-\frac{x}{2}\right) 1_{[0, \infty)}(x) \quad (1.2.42)$$

자유도가  $\nu$  인 카이제곱확률변수는 크기모수가 2이고 형태모수가  $\nu/2$  인 감마확률분포를 따른다. 따라서, 식 (1.2.39)에서 알 수 있듯이, 이 카이제곱확률분포의 평균과 분산은 각각

다음과 같다.

$$E(x) = \nu, \quad Var(x) = 2\nu \quad (1.2.43)$$

또한, 식 (1.2.36)에서 알 수 있듯이, 이 카이제곱확률분포의 적률생성함수와 특성함수는 각각 다음과 같다.

$$m_x(t) = [1 - 2t]^{-\frac{\nu}{2}}, \quad \left(t < \frac{1}{2}\right), \quad E\left(e^{i\theta x}\right) = [1 - 2i\theta]^{-\frac{\nu}{2}} \quad (1.2.44)$$

**정의 1.2.17:  $t$  확률분포**

서로 독립인 표준정규확률변수  $z$ 와 자유도가  $\nu$ 인 카이제곱확률변수  $y$ 에 대해서 확률변수  $x \doteq \frac{z}{\sqrt{y/\nu}}$ 를 자유도가  $\nu$ 인  $t$  확률변수라고 부르고, 이에 해당하는  $t$  확률분포를  $t_\nu$ 로 표기하자.

확률분포  $t_\nu$ 의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})} \left[1 + \frac{x^2}{\nu}\right]^{-[\nu+1]/2} \quad (1.2.45)$$

또한, 평균과 분산은 각각 다음과 같다.

$$E(x) = 0, \quad (\nu > 1), \quad Var(x) = \frac{\nu}{\nu - 2}, \quad (\nu > 2) \quad (1.2.46)$$

**정의 1.2.18:  $F$  확률분포**

자유도가  $\nu_1$ 인 카이제곱확률변수  $u$ 와 자유도가  $\nu_2$ 인 카이제곱확률변수  $v$ 가 서로 독립이라고 하고, 다음과 같은 확률변수를 정의하자.

$$x \doteq \frac{u/\nu_1}{v/\nu_2}$$

이  $x$ 를 자유도가  $(\nu_1, \nu_2)$ 인  $F$  확률변수라고 부르고, 이에 해당하는 확률분포를  $F_{\nu_1, \nu_2}$ 로

표기하자.

자유도가  $(\nu_1, \nu_2)$  인  $F$  확률변수의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{[\nu_1/\nu_2]^{\nu_1/2}}{B(\frac{\nu_1}{2}, \frac{\nu_2}{2})} x^{\nu_1-2/2} \left[1 + \frac{\nu_1}{\nu_2} x\right]^{-[\nu_1+\nu_2]/2} \quad (1.2.47)$$

또한, 평균과 분산은 각각 다음과 같다.

$$E(x) = \frac{\nu_2}{\nu_2 - 2}, \quad (\nu_2 > 2), \quad Var(x) = \frac{2\nu_2^2[\nu_1 + \nu_2 - 2]}{\nu_1[\nu_2 - 2]^2[\nu_2 - 4]}, \quad (\nu_2 > 4) \quad (1.2.48)$$

그러나,  $F$  확률분포의 적률생성함수는 존재하지 않고, 특성함수는 [77]를 참조하라.

**정의 1.2.19: 베타확률분포**

모수들이  $a(> 0)$ 와  $b(> 0)$  인 베타확률분포  $Beta(a, b)$  의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{B(a, b)} x^{a-1} [1-x]^{b-1} 1_{[0,1]}(x)$$

모수들이  $(a, b)$  인 베타확률분포의 평균과 분산은 각각 다음과 같다.

$$E(x) = \frac{a}{a+b}, \quad Var(x) = \frac{ab}{[a+b]^2[a+b+1]} \quad (1.2.49)$$

또한, 적률생성함수와 특성함수는 각각 다음과 같다.

$$m_x(t) = 1 + \sum_{k=1}^{\infty} \left[ \prod_{l=0}^{k-1} \frac{a+l}{a+b+l} \right] \frac{t^k}{k!} \quad (1.2.50)$$

$$E(e^{i\theta x}) = 1 + \sum_{k=1}^{\infty} \left[ \prod_{l=0}^{k-1} \frac{a+l}{a+b+l} \right] \frac{[i\theta]^k}{k!} \quad (1.2.51)$$

**정의 1.2.20: Weibull 확률분포**

크기모수 (scale parameter) 가  $a(> 0)$  이고 형태모수 (shape parameter) 가  $b(> 0)$  인

Weibull 확률분포의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{b}{a} \left[ \frac{x}{a} \right]^{b-1} \exp \left( - \left[ \frac{x}{a} \right]^b \right) 1_{[0, \infty)}(x)$$

모수들이  $(a, b)$  인 Weibull 확률분포의 확률분포함수는 다음과 같다.

$$F(x) = \left[ 1 - \exp \left( - \left[ \frac{x}{a} \right]^b \right) \right] 1_{[0, \infty)}(x) \quad (1.2.52)$$

이 Weibull 확률분포의 평균과 분산은 각각 다음과 같다.

$$E(x) = \frac{a}{b} \Gamma \left( \frac{1}{b} \right), \quad Var(x) = \frac{a^2}{b^2} \left\{ 2b \Gamma \left( \frac{2}{b} \right) - \left[ \Gamma \left( \frac{1}{b} \right) \right]^2 \right\} \quad (1.2.53)$$

또한, 적률생성함수와 특성함수는 각각 다음과 같다.

$$m_x(t) = \sum_{k=0}^{\infty} \Gamma \left( 1 + \frac{k}{b} \right) \frac{a^k t^k}{k!}, \quad (b \geq 1) \quad (1.2.54)$$

$$E \left( e^{i\theta x} \right) = \sum_{k=0}^{\infty} \Gamma \left( 1 + \frac{k}{b} \right) \frac{a^k [i\theta]^k}{k!} \quad (1.2.55)$$

#### 정의 1.2.21: Rayleigh 확률분포

크기모수(scale parameter)가  $\sigma$  인 Rayleigh 확률분포의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{x}{\sigma^2} \exp \left( - \frac{x^2}{2\sigma^2} \right) 1_{[0, \infty)}(x)$$

Rayleigh 확률분포는 Weibull 확률분포의 특수한 경우이다. 이 Rayleigh 확률분포의 확률분포함수는 다음과 같다.

$$F(x) = \left[ 1 - \exp \left( - \frac{x^2}{2\sigma^2} \right) \right] 1_{[0, \infty)}(x) \quad (1.2.56)$$

이 Rayleigh 확률분포의 평균과 분산은 각각 다음과 같다.

$$E(x) = \sqrt{\frac{\pi}{2}}\sigma, \quad Var(x) = \frac{4-\pi}{2}\sigma^2 \quad (1.2.57)$$

또한, 적률생성함수와 특성함수는 각각 다음과 같다.

$$m_x(t) = 1 + \sqrt{\frac{\pi}{2}}\sigma t \exp\left(\frac{\sigma^2}{2}t^2\right) \left[1 + \operatorname{erf}\left(\frac{\sigma t}{\sqrt{2}}\right)\right] \quad (1.2.58)$$

$$E\left(e^{i\theta x}\right) = 1 - \sqrt{\frac{\pi}{2}}\sigma\theta \exp\left(-\frac{\sigma^2}{2}\theta^2\right) \left[\operatorname{erfi}\left(\frac{\sigma\theta}{\sqrt{2}}\right) - i\right] \quad (1.2.59)$$

여기서 오차함수들 erf와 erfi는 각각 다음과 같다.

$$\operatorname{erf}(x) \doteq \frac{2}{\sqrt{\pi}} \int_0^x \exp(-z^2) dz \quad (1.2.60)$$

$$\operatorname{erfi}(x) \doteq -i \operatorname{erf}(ix) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(z^2) dz \quad (1.2.61)$$

**정의 1.2.22: 일반화Pareto확률분포**

크기모수(scale parameter)가  $b$ 인 일반화Pareto확률분포의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\sigma} \left\{ 1 + k \left[ \frac{x - \theta}{\sigma} \right] \right\}^{-1-1/k}$$

만약  $k > 0$ 이면, 일반화Pareto확률분포 (generalized Pareto distribution)의 확률밀도 함수  $f(x)$ 는 구간  $x > \theta$ 에서 정의된다. 만약  $k < 0$ 이면,  $f(x)$ 는 구간  $\theta < x < -\sigma/k$ 에서 정의된다. 만약  $k > 0$ 이고  $\theta = \sigma/k$ 이면, 이 일반화Pareto확률분포는 Pareto확률분포이다. 일반화Pareto확률분포의 확률밀도함수는 다음과 같다.

$$F(x) = 1 - \left\{ 1 + k \left[ \frac{x - \theta}{\sigma} \right] \right\}^{-1/k}, \quad (k \neq 0) \quad (1.2.62)$$

이 일반화Pareto확률분포의 평균과 분산은 각각 다음과 같다.

$$E(x) = \theta + \frac{\sigma}{1-k}, \quad Var(x) = \frac{\sigma^2}{[k-1]^2[1-2k]}, \quad \left(k < \frac{1}{2}\right) \quad (1.2.63)$$

이 일반화Pareto 확률분포함수의 역함수는 다음과 같다.

$$F^{-1}(x) = \theta + \frac{\sigma}{k} \left\{ [1 - x]^{-k} - 1 \right\} \quad (1.2.64)$$

식 (1.2.64)와 역함수법을 사용해서, 일반화Pareto 난수를 생성할 수 있다.

**정의 1.2.23: 일반화극한값확률분포**

형태모수가  $k (\neq 0)$  이고 위치모수가  $\theta$  그리고 크기모수가  $\sigma (> 0)$  인 일반화극한값확률분포 (generalized extreme value probability distribution) 의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\sigma} \left\{ 1 + k \left[ \frac{x - \theta}{\sigma} \right] \right\}^{-1-1/k} \exp \left( - \left\{ 1 + k \left[ \frac{x - \theta}{\sigma} \right] \right\} \right) \mathbf{1} \left( k \left[ \frac{x - \theta}{\sigma} \right] > -1 \right)$$

만약 형태모수가  $k = 0$  이면, 일반화극한값확률분포의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\sigma} \exp \left( - \exp \left( - \frac{x - \theta}{\sigma} \right) - \frac{x - \theta}{\sigma} \right)$$

형태모수가  $k \neq 0$  이고 위치모수가  $\theta$  그리고 크기모수가  $\sigma (> 0)$  인 일반화극한값확률분포의 확률분포함수는 다음과 같다.

$$F(x) = \exp \left( - \left\{ 1 + k \left[ \frac{x - \theta}{\sigma} \right] \right\}^{-1/k} \right) \mathbf{1} \left( 1 + k \left[ \frac{x - \theta}{\sigma} \right] > 0 \right) \quad (1.2.65)$$

또한, 평균과 분산은 각각 다음과 같다.

$$E(x) = \theta - \frac{\sigma}{k} + \frac{\sigma}{k} g_1, \quad Var(x) = \frac{\sigma^2}{k^2} [g_2 - g_1^2] \quad (1.2.66)$$

여기서  $g_j \doteq \Gamma(1 + jk)$ , ( $j = 1, 2$ ) 이다. 일반화극한값확률분포의 확률밀도함수에 극한  $k \rightarrow 0$  를 취한 확률분포를 극한값확률분포라 부른다. 극한값확률분포함수는 다음과 같다.

$$F(x) = \exp \left( - \exp \left( - \frac{x - \theta}{\sigma} \right) \right) \quad (1.2.67)$$

또한, 평균과 분산은 각각 다음과 같다.

$$E(x) = \theta + \gamma\sigma, \quad Var(x) = \frac{1}{6}\pi^2\sigma^2 \quad (1.2.68)$$

여기서  $\gamma$ 는 Euler 상수이다.

**정의 1.2.24: 비중심카이제곱확률분포**

서로 독립인 확률변수들  $v_1, v_2, \dots, v_\nu$ 가 각각 다음과 같은 확률분포를 따른다고 하자.

$$v_i \stackrel{d}{\sim} N(\mu_i, \sigma_i^2), \quad (i = 1, 2, \dots, \nu)$$

다음과 같은 확률변수와 상수를 정의하자.

$$x \doteq \sum_{i=1}^{\nu} \left[ \frac{v_i}{\sigma_i} \right]^2, \quad \delta \doteq \sum_{i=1}^{\nu} \left[ \frac{v_i}{\sigma_i} \right]^2$$

확률변수  $x$ 는 자유도가  $\nu$ 이고 비중심모수(noncentrality)가  $\delta$ 인 비중심카이제곱확률분포(noncentral chi-square probability distribution)  $\chi_\nu^2(\delta)$ 를 따른다고 한다. 문헌에 따라서는  $\delta/2$ 를 비중심모수라고 부르기도 한다.

자유도가  $\nu$ 이고 비중심모수가  $\delta$ 인 비중심카이제곱확률분포의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{2} \left[ \frac{x}{\delta} \right]^{[\nu-2]/4} \exp\left(-\frac{x+\delta}{2}\right) I_{[\nu-2]/2}(\sqrt{\delta x}) \quad (1.2.69)$$

여기서 제1종 변형 Bessel 함수  $I_a(y)$ 는 다음과 같다.

$$I_a(y) \doteq \left[ \frac{y}{2} \right]^a \sum_{j=0}^{\infty} \frac{1}{j! \Gamma(a+j+1)} \left[ \frac{y^2}{4} \right]^j \quad (1.2.70)$$

이 비중심카이제곱확률분포의 평균과 분산은 각각 다음과 같다.

$$E(x) = \nu + \delta, \quad Var(x) = 2\nu + 4\delta \quad (1.2.71)$$

**정의 1.2.25: 비중심  $t$  확률분포**

서로 독립인 표준정규확률변수  $z$ 와 자유도가  $\nu$ 인 카이제곱확률변수  $y$ 에 대해서 다음 확률변수를 정의하자.

$$x \doteq \frac{z + \delta}{\sqrt{y/\nu}}$$

확률변수  $x$ 는 자유도가  $\nu$ 이고 비중심모수가  $\delta$ 인 비중심  $t$  확률분포  $t_\nu(\delta)$ 를 따른다고 한다.

이  $t_\nu(\delta)$  확률변수의 평균과 분산은 각각 다음과 같다.

$$E(x) = \delta \sqrt{\frac{\nu}{2}} \frac{\Gamma([\nu - 1]/2)}{\Gamma(\nu/2)}, \quad (\nu > 1) \quad (1.2.72)$$

$$Var(x) = \frac{\nu[1 + \delta^2]}{\nu - 2} - E(x)^2, \quad (\nu > 2) \quad (1.2.73)$$

**정의 1.2.26: 비중심  $F$  확률분포**

자유도가  $\nu_1$ 이고 비중심모수가  $\delta$ 인 비중심카이제곱확률변수  $u$ 와 자유도가  $\nu_2$ 인 카이제곱확률변수  $v$ 가 서로 독립이라고 하고, 다음 확률변수를 정의하자.

$$x \doteq \frac{u/\nu_1}{v/\nu_2}$$

확률변수  $x$ 를 자유도들이  $(\nu_1, \nu_2)$ 이고 비중심모수가  $\delta$ 인 비중심  $F$  확률변수라고 부르고, 이에 해당하는 확률분포를  $F_{\nu_1, \nu_2}(\delta)$ 로 표기하자.

비중심모수가  $\delta$ 인 비중심  $F$  확률변수의 확률밀도함수는 다음과 같다.

$$f(x) = \sum_{j=0}^{\infty} e^{-\delta/2} \left[ \frac{\delta}{2} \right]^j \frac{[\nu_1/\nu_2]^{\nu_1/2+j}}{B(\frac{\nu_1}{2}, \frac{\nu_2}{2} + j)} x^{[\nu_1-2+2j]/2} \left[ 1 + \frac{\nu_1}{\nu_2} x \right]^{-[\nu_1+\nu_2]/2-j} \quad (1.2.74)$$

또한, 평균과 분산은 각각 다음과 같다.

$$E(x) = \frac{[\nu_1 + \delta]\nu_2}{\nu_1[\nu_2 - 2]}, \quad (\nu_2 > 2) \tag{1.2.75}$$

$$Var(x) = \frac{2\{[\nu_1 + \delta]^2 + [\nu_1 + 2\delta][\nu_2 - 2]\}}{[\nu_2 - 2]^2[\nu_2 - 4]} \frac{\nu_2^2}{\nu_1^2}, \quad (\nu_2 > 4) \tag{1.2.76}$$

**정의 1.2.27: 로지스틱확률분포**

위치모수가  $\mu$ 이고 형태모수가  $\sigma$ 인 로지스틱확률분포 (logistic probability distribution)의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{\exp(-\frac{x-\mu}{\sigma})}{\sigma[1 + \exp(-\frac{x-\mu}{\sigma})]^2}$$

위치모수가  $\mu$ 이고 형태모수가  $\sigma$ 인 로지스틱확률분포의 확률분포함수는 다음과 같다.

$$F(x) = \frac{1}{1 + \exp(-\frac{x-\mu}{\sigma})} \tag{1.2.77}$$

로지스틱확률분포함수는 성장곡선으로서 S형이다. 이 로지스틱확률분포의 평균과 분산은 각각 다음과 같다.

$$E(x) = \mu, \quad Var(x) = \frac{\pi^2}{3}\sigma^2\mu^2 \tag{1.2.78}$$

또한, 적률생성함수와 특성함수는 각각 다음과 같다.

$$m_x(t) = e^{\mu t} B(1 - \sigma t, 1 + \sigma t) \quad (\sigma t \in (-1, 1)) \tag{1.2.79}$$

$$E(e^{i\theta x}) = e^{i\mu\theta} \frac{\pi\sigma\theta}{\sinh(\pi\sigma\theta)} \tag{1.2.80}$$

**정의 1.2.28: 역정규확률분포**

위치모수가  $\mu(> 0)$ 이고 형태모수가  $\sigma(> 0)$ 인 역정규확률분포 (inverse Gaussian prob-

ability distribution)의 확률밀도함수는 다음과 같다.

$$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda}{2\mu^2 x}[x - \mu]^2\right) 1_{\{x>0\}}$$

위치모수가  $\mu$ 이고 형태모수가  $\sigma$ 인 역정규확률분포의 확률분포함수는 다음과 같다.

$$F(x) = \left[ N\left(\sqrt{\frac{\lambda}{x}}\left[\frac{x}{\mu} - 1\right]\right) + \exp\left(\frac{2\lambda}{\mu}\right) N\left(-\sqrt{\frac{\lambda}{x}}\left[\frac{x}{\mu} + 1\right]\right) \right] 1_{\{x>0\}} \quad (1.2.81)$$

이 역정규확률분포는 Brown운동에서 정해진 레벨에 도달하는데 걸리는 시간의 확률분포이다. 이 역정규확률분포의 평균과 분산은 각각 다음과 같다.

$$E(x) = \mu, \quad Var(x) = \frac{\mu^3}{\lambda} \quad (1.2.82)$$

또한, 적률생성함수와 특성함수는 각각 다음과 같다.

$$m_x(t) = \exp\left(\frac{\lambda}{\mu}\right) \left[ 1 - \sqrt{1 - \frac{2\mu^2 t}{\lambda}} \right] \quad (1.2.83)$$

$$E\left(e^{i\theta x}\right) = \exp\left(\frac{\lambda}{\mu}\right) \left[ 1 - \sqrt{1 - i\frac{2\mu^2 \theta}{\lambda}} \right] \quad (1.2.84)$$

#### 정의 1.2.29: Cauchy 확률분포

서로 독립인 표준정규확률변수들  $v_1$  과  $v_2$  에 대해서 다음과 같은 확률변수를 정의하자.

$$x \doteq \frac{v_1}{v_2}$$

확률변수  $x$  를 Cauchy 확률변수라고 부른다.

Cauchy 확률변수의 확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2} \quad (1.2.85)$$

또한, 확률분포함수와 그 역함수는 각각 다음과 같다.

$$F(x) = \frac{1}{\pi} \tan^{-1} x + \frac{1}{2} \tag{1.2.86}$$

$$F^{-1}(x) = \tan\left(\pi x - \frac{\pi}{2}\right) = \cot(\pi x) \tag{1.2.87}$$

또한, 특성함수는 다음과 같다.

$$E\left(e^{i\theta x}\right) = \exp(-|\theta|) \tag{1.2.88}$$

Cauchy 확률변수의 평균은 존재하지 않는다. 따라서, 분산도 존재하지 않는다.

확률변수  $x$ 가 감마확률분포를 따를 때 확률변수  $1/x$ 가 따르는 확률분포를 역감마확률분포 (inverse gamma distribution)라고 한다. 역감마확률분포의 정형적 정의는 다음과 같다.

**정의 1.2.30: 역감마확률분포**

주어진 상수들  $\alpha(> 0)$ 와  $\beta(> 0)$ 에 대해서 다음 확률밀도함수를 살펴보자.

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right), \quad (x > 0)$$

이 확률변수  $x$ 는 역감마확률분포를 따른다고 하고, 다음과 같이 표기하자.

$$x \stackrel{d}{\sim} \mathcal{IG}(\alpha, \beta)$$

만약 조건  $\alpha > 1$ 이 성립하면, 역감마확률분포의 평균은 다음과 같다.

$$E(x) = \frac{\beta}{\alpha - 1}, \quad (\alpha > 1) \tag{1.2.89}$$

만약 조건  $\alpha > 2$ 가 성립하면, 역감마확률분포의 분산은 다음과 같다.

$$Var(x) = \frac{\beta^2}{[\alpha - 1]^2[\alpha - 2]}, \quad (\alpha > 2) \tag{1.2.90}$$

감마확률분포,  $F$  확률분포, 대수정규확률분포 등과 같이 비음인 실수들의 집합  $\mathbb{R}_{\geq 0}$ 를 지지대로 갖는 확률분포들에 비해 일반적으로 역감마확률분포의 꼬리가 더 두껍다.

**정의 1.2.31: 척도화역카이제곱확률분포**

확률변수  $x$ 가 역감마확률분포  $IG(\nu/2, \nu S/2)$ 를 따르면, 확률변수  $x$ 는 자유도가  $\nu$ 인 척도화역카이제곱확률분포 (scaled inverse-chi-squared distribution)를 따른다고 하고 다음과 같이 표기하자.

$$x \stackrel{d}{\sim} Inv - \chi^2(\nu, S)$$

**제 1.3 절 다변량정규확률분포****1.3.1 적률생성함수**

확률벡터  $\mathbf{x}$ 의 적률생성함수 (moment generating function: mgf)  $M_{\mathbf{x}}(\mathbf{t}) \doteq E(\exp(\mathbf{t}^T \mathbf{x}))$ 를 생각해보자. 예를 들어 3변량 확률벡터  $\mathbf{x} = [x_1, x_2, x_3]^T$ 의 적률생성함수는 다음과 같다.

$$M_{\mathbf{x}}(t_1, t_2, t_3) = E(\exp(x_1 t_1 + x_2 t_2 + x_3 t_3)) \quad (1.3.1)$$

확률변수들  $x_1$ 과  $x_2$ 가 독립적이기 위한 필요충분조건은 다음과 같다.

$$M_{x_1, x_2}(t_1, t_2) = M_{x_1}(t_1)M_{x_2}(t_2) \quad (1.3.2)$$

이 성질을 이용해서, 확률벡터들  $\mathbf{x}_1$ 과  $\mathbf{x}_2$ 가 독립적이기 위한 필요충분조건이 다음과 같음을 쉽게 증명할 수 있다.

$$M_{\mathbf{x}_1, \mathbf{x}_2}(\mathbf{t}_1, \mathbf{t}_2) = M_{\mathbf{x}_1}(\mathbf{t}_1)M_{\mathbf{x}_2}(\mathbf{t}_2) \quad (1.3.3)$$

확률벡터들  $\mathbf{x}_1$ 과  $\mathbf{x}_2$ 가 독립이라고 가정하고, 상수가 아닌 함수들  $\mathbf{g}_1$ 과  $\mathbf{g}_2$ 에 대해서 다음 확률벡터들과 벡터들을 정의하자.

$$\mathbf{y} \doteq \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \doteq \begin{bmatrix} \mathbf{g}_1(\mathbf{x}_1) \\ \mathbf{g}_2(\mathbf{x}_2) \end{bmatrix}, \quad \mathbf{t} \doteq \begin{bmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \end{bmatrix} \quad (1.3.4)$$

다음 식들이 성립한다.

$$\begin{aligned}
 M_{\mathbf{y}}(\mathbf{t}) &= E(\exp(\mathbf{t}^T \mathbf{y})) = E(\exp(\mathbf{t}_1^T \mathbf{y}_1 + \mathbf{t}_2^T \mathbf{y}_2)) \\
 &= E(\exp(\mathbf{t}_1^T \mathbf{g}_1(\mathbf{x}_1)) \exp(\mathbf{t}_2^T \mathbf{g}_2(\mathbf{x}_2))) \\
 &= \int \int \exp(\mathbf{t}_1^T \mathbf{g}_1(\mathbf{x}_1)) \exp(\mathbf{t}_2^T \mathbf{g}_2(\mathbf{x}_2)) f_{\mathbf{x}_1}(\mathbf{x}_1) f_{\mathbf{x}_2}(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\
 &= \int \exp(\mathbf{t}_1^T \mathbf{g}_1(\mathbf{x}_1)) \left\{ \int \exp(\mathbf{t}_2^T \mathbf{g}_2(\mathbf{x}_2)) f_{\mathbf{x}_2}(\mathbf{x}_2) d\mathbf{x}_2 \right\} f_{\mathbf{x}_1}(\mathbf{x}_1) d\mathbf{x}_1 \\
 &= M_{\mathbf{g}_1(\mathbf{x}_1)}(\mathbf{t}_1) M_{\mathbf{g}_2(\mathbf{x}_2)}(\mathbf{t}_2) \\
 &= M_{\mathbf{y}_1}(\mathbf{t}_1) M_{\mathbf{y}_2}(\mathbf{t}_2) \tag{1.3.5}
 \end{aligned}$$

여기서 네 번째 등호는 확률벡터들  $\mathbf{x}_1$  과  $\mathbf{x}_2$  의 독립성에 의해서 성립한다. 식 (1.3.5) 에서 알 수 있듯이, 확률벡터들  $\mathbf{y}_1$  와  $\mathbf{y}_2$  는 독립이다. 다음 식들이 성립한다.

$$M_{\mathbf{Ax}}(\mathbf{t}) = E(\exp(\mathbf{t}^T \mathbf{Ax})) = E(\exp([\mathbf{A}^T \mathbf{t}]^T \mathbf{x})) = M_{\mathbf{x}}(\mathbf{A}^T \mathbf{t}) \tag{1.3.6}$$

$$M_{\mathbf{x}+\mathbf{c}}(\mathbf{t}) = E(\exp(\mathbf{t}^T [\mathbf{x} + \mathbf{c}])) = \exp(\mathbf{t}^T \mathbf{c}) M_{\mathbf{x}}(\mathbf{t}) \tag{1.3.7}$$

정규확률변수  $y \stackrel{d}{\sim} \mathcal{N}(\mu, \sigma^2)$  의 적률생성함수는  $M_y(t) = \exp(\mu t + \frac{1}{2} \sigma^2 t^2)$  이다. 적률생성함수가 존재하는 경우에는 적률생성함수와 확률분포는 일대일 대응관계에 있으므로, 다음과 같이 정규확률변수를 정의할 수 있다.

**정의 1.3.1**

적률생성함수가  $\exp(\mu t + \frac{1}{2} \sigma^2 t^2)$  인 단변량확률변수를 평균이  $\mu$  이고 분산이  $\sigma^2$  인 정규확률변수라고 한다.

확률변수  $U$  의 지지대 (support) 가  $\{u_0\}$  라고 하자. 이 확률변수의 적률생성함수는 다음과 같다.

$$M_U(t) = E(\exp(Ut)) = \sum_u e^{ut} p(u) = \exp(u_0 t) p(u_0) = \exp(u_0 t) \tag{1.3.8}$$

여기서 확률밀도함수  $p(\cdot)$  는 식  $p(u) = 1_{\{u_0\}}(u)$  를 만족한다. 이 적률생성함수는 정규확률분포의 적률생성함수  $\exp(\mu t + \frac{1}{2} \sigma^2 t^2)$  에서 평균이  $\mu = u_0$  이고 분산이 0 인 형태이다. 즉,

$U \stackrel{d}{\sim} \mathcal{N}(u_0, 0)$ 이다. 따라서,  $U$ 는 정규확률변수 중에서 분산이 0인 퇴화확률변수 (degenerate random variable)이다. 이를 특이 (singular) 정규확률변수라고 한다. 예를 들어, 식  $\mathbf{a}^T \Sigma \mathbf{a} = 0$ 를 만족하는 벡터  $\mathbf{a} \neq \mathbf{0}$ 에 대해서 확률변수  $u \doteq \mathbf{a}^T \mathbf{y}$ 를 생각해보자. 다음 식들이 성립한다.

$$\text{Var}(u) = \text{Var}(\mathbf{a}^T \mathbf{y}) = \mathbf{a}^T \Sigma \mathbf{a} = 0 \quad (1.3.9)$$

확률변수  $u$ 는 정규확률분포를 따르지만, 분산이 0인 특이확률변수이다. 특이확률변수는 Dirac 델타함수와 밀접한 관계가 있다.

### 1.3.2 정규확률벡터

서로 독립인 표준정규확률변수들  $z_1, z_2, \dots, z_p$ 로 이루어진 정규확률벡터  $\mathbf{z} = [z_1, z_2, \dots, z_p]^T$ 의 평균벡터와 분산공분산행렬은 각각 다음과 같다.

$$E(\mathbf{z}) = \mathbf{0}, \quad \text{Var}(\mathbf{z}) = I_p \quad (1.3.10)$$

정규확률벡터  $\mathbf{z}$ 의 확률밀도함수는 다음과 같다.

$$f_{\mathbf{z}}(\mathbf{z}) = \frac{1}{[2\pi]^{p/2}} \exp\left(-\frac{1}{2}\mathbf{z}^T \mathbf{z}\right) \quad (1.3.11)$$

또한, 적률생성함수는 다음과 같다.

$$M_{\mathbf{z}}(\mathbf{t}) = \prod_{j=1}^p M_{Z_j}(t_j) = \prod_{j=1}^p \exp\left(\frac{1}{2}t_j^2\right) = \exp\left(\frac{1}{2}\sum_{j=1}^p t_j^2\right) = \exp\left(\frac{1}{2}\mathbf{t}^T \mathbf{t}\right) \quad (1.3.12)$$

여기서  $\mathbf{t} \doteq [t_1, t_2, \dots, t_p]^T$ 이다.

비음정치행렬 (non-negative definite matrix)  $\Sigma$ 와 벡터  $\boldsymbol{\mu} \in \mathbb{R}^p$ 에 대해서 벡터  $\mathbf{y} = \Sigma^{1/2}\mathbf{z} + \boldsymbol{\mu}$ 를 정의하자. 다음 식들이 성립한다.

$$E(\mathbf{y}) = \boldsymbol{\mu} \quad (1.3.13)$$

$$\text{Var}(\mathbf{y}) = \Sigma^{1/2}\text{Var}(\mathbf{z})\Sigma^{1/2} = \Sigma^{1/2} I_p \Sigma^{1/2} = \Sigma \quad (1.3.14)$$

벡터  $\mathbf{y}$ 의 각 원소는 서로 독립인 단변량정규확률변수들의 선형결합이다. 단변량정규확률변수들의 선형결합도 단변량정규확률변수이므로, 확률벡터  $\mathbf{y}$ 의 각 원소는 단변량정규확률분포를

따른다. 벡터  $\mathbf{y} = \Sigma^{1/2}\mathbf{z} + \boldsymbol{\mu}$ 의 적률생성함수는 다음과 같다.

$$\begin{aligned}
 M_{\mathbf{y}}(\mathbf{t}) &= M_{\Sigma^{1/2}\mathbf{z} + \boldsymbol{\mu}}(\mathbf{t}) = \exp(\mathbf{t}^T \boldsymbol{\mu}) M_{\Sigma^{1/2}\mathbf{z}}(\mathbf{t}) \\
 &= \exp(\mathbf{t}^T \boldsymbol{\mu}) M_{\mathbf{z}}\left(\left(\Sigma^{1/2}\right)^T \mathbf{t}\right) = \exp(\mathbf{t}^T \boldsymbol{\mu}) M_{\mathbf{z}}(\Sigma^{1/2}\mathbf{t}) \\
 &= \exp(\mathbf{t}^T \boldsymbol{\mu}) \exp\left(\frac{1}{2}(\Sigma^{1/2}\mathbf{t})^T (\Sigma^{1/2}\mathbf{t})\right) \\
 &= \exp\left(\mathbf{t}^T \boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^T \Sigma \mathbf{t}\right) \tag{1.3.15}
 \end{aligned}$$

따라서 다음과 같이 다변량정규확률벡터를 정의할 수 있다.

**정의 1.3.2**

적률생성함수가  $\exp(\mathbf{t}^T \boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^T \Sigma \mathbf{t})$ 인 다변량확률벡터를 평균벡터가  $\boldsymbol{\mu}$ 이고 분산공분산 행렬이  $\Sigma$ 인 다변량정규확률벡터라고 하고, 다음과 같이 표기하자.

$$\mathbf{y} \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$$

만약  $\Sigma$ 가 대각행렬  $\text{diag}(\sigma_{11}, \sigma_{22}, \dots, \sigma_{pp})$ 이면, 식 (1.3.15)에서 알 수 있듯이 다음 식이 성립한다.

$$M_{\mathbf{y}}(\mathbf{t}) = \prod_{i=1}^p \exp\left(t_i \mu_i + \frac{1}{2} t_i^2 \sigma_{ii}\right) \tag{1.3.16}$$

따라서, 식 (1.3.2)에서 알 수 있듯이  $y_1, y_2, \dots, y_p$ 는 서로 독립이다. 즉, 정규성가정 하에서 독립성과 비상관성은 동치이다.

만약  $\mathbf{z} \stackrel{d}{\sim} \mathcal{N}_p(\mathbf{0}, I_p)$ 이고, 또한 만약  $\Sigma$ 가 양정치행렬 (positive definite matrix)이면, 확률벡터  $\mathbf{y} = \Sigma^{1/2}\mathbf{z} + \boldsymbol{\mu}$ 의 확률밀도함수  $f_{\mathbf{y}}(\mathbf{y})$ 는 다음과 같다.

$$f_{\mathbf{y}}(\mathbf{y}) = f_{\mathbf{z}}(\mathbf{y}) \left| \frac{\partial \mathbf{z}}{\partial \mathbf{y}} \right| = f_{\mathbf{z}}\left(\Sigma^{-1/2}[\mathbf{y} - \boldsymbol{\mu}]\right) \left| \Sigma^{-1/2} \right| \tag{1.3.17}$$

즉, 다음 식이 성립한다.

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{1}{[2\pi]^{p/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}[\mathbf{y} - \boldsymbol{\mu}]^T \Sigma^{-1}[\mathbf{y} - \boldsymbol{\mu}]\right) \tag{1.3.18}$$

따라서 다음 식이 성립한다.

$$\mathbf{y} \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\mu}, \Sigma) \quad (1.3.19)$$

행렬  $A$ 는 랭크가  $r$ 인  $r \times p$  행렬이라 하면, 확률벡터  $\mathbf{u} \doteq A\mathbf{y}$ 의 적률생성함수는 다음과 같다.

$$\begin{aligned} M_{\mathbf{u}}(\mathbf{t}) &= M_{A\mathbf{y}}(\mathbf{t}) = M_{\mathbf{y}}(A^T \mathbf{t}) \\ &= \exp\left([A^T \mathbf{t}]^T \boldsymbol{\mu} + \frac{1}{2}[A^T \mathbf{t}]^T \Sigma [A^T \mathbf{t}]\right) \\ &= \exp\left(\mathbf{t}^T [A\boldsymbol{\mu}] + \frac{1}{2}\mathbf{t}^T [A\Sigma A^T] \mathbf{t}\right) \end{aligned} \quad (1.3.20)$$

정의 1.3.2에서 알 수 있듯이, 다음 식이 성립한다.

$$\mathbf{u} \stackrel{d}{\sim} \mathcal{N}_r(A\boldsymbol{\mu}, A\Sigma A^T) \quad (1.3.21)$$

벡터  $\mathbf{c} \in \mathbb{R}^p$ 에 대해서 확률벡터  $\mathbf{v} \doteq \mathbf{y} + \mathbf{c}$ 의 적률생성함수는 다음과 같다.

$$\begin{aligned} M_{\mathbf{v}}(\mathbf{t}) &= M_{\mathbf{y}+\mathbf{c}}(\mathbf{t}) = \exp(\mathbf{t}^T \mathbf{c}) M_{\mathbf{y}}(\mathbf{t}) \\ &= \exp(\mathbf{t}^T \mathbf{c}) \exp\left(\mathbf{t}^T \boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^T \Sigma \mathbf{t}\right) \\ &= \exp\left(\mathbf{t}^T [\boldsymbol{\mu} + \mathbf{c}] + \frac{1}{2}\mathbf{t}^T \Sigma \mathbf{t}\right) \end{aligned} \quad (1.3.22)$$

따라서 다음 식이 성립한다.

$$\mathbf{v} \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\mu} + \mathbf{c}, \Sigma) \quad (1.3.23)$$

식 (1.3.21)과 식 (1.3.23)에서 알 수 있듯이, 다음 정리가 성립한다.

#### 정리 1.3.1

확률벡터  $\mathbf{y} \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ , 벡터  $\mathbf{d} \in \mathbb{R}^p$ 와 랭크가  $r$ 인 행렬  $A \in \mathbb{R}^{r \times p}$ 에 대해서 다음 식이

성립한다.

$$A\mathbf{y} + \mathbf{d} \stackrel{d}{\sim} \mathcal{N}_r(A\boldsymbol{\mu} + \mathbf{d}, A\Sigma A^T)$$

**예제 1.3.1** 다음과 같은 확률변수들을 살펴보자.

$$y_1 \stackrel{d}{\sim} \mathcal{N}(1, 2), \quad y_2 \stackrel{d}{\sim} \mathcal{N}(2, 4), \quad y_3 \stackrel{d}{\sim} \mathcal{N}(6, 3) \quad (1)$$

확률변수들  $y_1, y_2, y_3$ 는 서로 독립이라고 가정하고, 확률변수들  $w_1 = y_1 + y_2$  과  $w_2 = y_2 + y_3$ 의 결합확률분포를 구해보자. 정리 1.3.1에서 알 수 있듯이, 다음 식들이 성립한다.

$$\mathbf{w} = A\mathbf{y} \stackrel{d}{\sim} \mathcal{N}(A\boldsymbol{\mu}, A\Sigma A^T) \quad (2)$$

여기서 벡터들과 행렬은 각각 다음과 같다.

$$\mathbf{w} \doteq \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \quad \mathbf{y} \doteq \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \quad A \doteq \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad (2)$$

확률벡터  $\mathbf{w}$ 의 평균벡터는 다음과 같다.

$$A\boldsymbol{\mu} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix} = \begin{bmatrix} 3 \\ 8 \end{bmatrix} \quad (3)$$

또한, 확률벡터  $\mathbf{w}$ 의 분산공분산행렬은 다음과 같다.

$$A\Sigma A^T = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ 4 & 7 \end{bmatrix} \quad (4)$$

■

**예제 1.3.2** 정규확률벡터  $\mathbf{x} \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 에 대해서 다음 식이 성립한다.

$$\Sigma^{-\frac{1}{2}}[\mathbf{x} - \boldsymbol{\mu}] \stackrel{d}{\sim} \mathcal{N}(\mathbf{0}, I_p) \quad (1)$$

즉,  $\Sigma^{-\frac{1}{2}}[\mathbf{x} - \boldsymbol{\mu}]$ 의 원소들은 서로 독립인 표준정규확률변수들이다. 따라서 다음 식들이 성립한다.

$$\left(\Sigma^{-\frac{1}{2}}[\mathbf{x} - \boldsymbol{\mu}]\right)^T \Sigma^{-\frac{1}{2}}[\mathbf{x} - \boldsymbol{\mu}] = [\mathbf{x} - \boldsymbol{\mu}]^T \Sigma^{-1}[\mathbf{x} - \boldsymbol{\mu}] \stackrel{d}{\sim} \chi_p^2 \quad (2)$$

■

### 1.3.3 조건부확률밀도함수

정규확률벡터  $\mathbf{y} \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 에 대해서 다음과 같이 분할행렬들을 정의하자.

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \mathbf{t} = \begin{bmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \end{bmatrix} \quad (1.3.24)$$

여기서  $\mathbf{y}_1, \boldsymbol{\mu}_1, \mathbf{t}_1 \in \mathbb{R}^q$  이고  $\mathbf{y}_2, \boldsymbol{\mu}_2, \mathbf{t}_2 \in \mathbb{R}^{p-q}$  이고,  $\Sigma_{11} \in \mathbb{R}^{q \times q}$ ,  $\Sigma_{12} \in \mathbb{R}^{q \times [p-q]}$ ,  $\Sigma_{21} \in \mathbb{R}^{[p-q] \times q}$ ,  $\Sigma_{22} \in \mathbb{R}^{[p-q] \times [p-q]}$  이고, 또한  $\Sigma_{21} = \Sigma_{12}^T$  이다.

만약  $\Sigma_{21} = \Sigma_{12}^T = O$ 이면, 다음 식들이 성립한다.

$$\begin{aligned} M_{\mathbf{y}}(\mathbf{t}) &= M_{\mathbf{y}}([\mathbf{t}_1^T, \mathbf{t}_2^T]^T) \\ &= \exp\left([\mathbf{t}_1^T, \mathbf{t}_2^T] \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} + \frac{1}{2}[\mathbf{t}_1^T, \mathbf{t}_2^T] \begin{pmatrix} \Sigma_{11} & O \\ O & \Sigma_{22} \end{pmatrix} \begin{bmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \end{bmatrix}\right) \\ &= \exp\left(\mathbf{t}_1^T \boldsymbol{\mu}_1 + \mathbf{t}_2^T \boldsymbol{\mu}_2 + \frac{1}{2}[\mathbf{t}_1^T \Sigma_{11} \mathbf{t}_1 + \mathbf{t}_2^T \Sigma_{22} \mathbf{t}_2]\right) \\ &= \exp\left(\mathbf{t}_1^T \boldsymbol{\mu}_1 + \frac{1}{2} \mathbf{t}_1^T \Sigma_{11} \mathbf{t}_1\right) \exp\left(\mathbf{t}_2^T \boldsymbol{\mu}_2 + \frac{1}{2} \mathbf{t}_2^T \Sigma_{22} \mathbf{t}_2\right) \end{aligned} \quad (1.3.25)$$

즉, 다음 식이 성립한다.

$$M_{\mathbf{y}}(\mathbf{t}) = M_{\mathbf{y}_1}(\mathbf{t}_1) M_{\mathbf{y}_2}(\mathbf{t}_2) \quad (1.3.26)$$

결론적으로, 만약  $\Sigma_{21} = \Sigma_{12}^T = O$ 이면, 식 (1.3.3)에서 알 수 있듯이 확률벡터들  $\mathbf{y}_1$  과  $\mathbf{y}_2$  는 서로 독립이다. 또한, 그 역도 성립한다.

행렬  $A = [I_q, O_{q \times [p-q]}]$  에 대해서 다음 식들이 성립한다.

$$\mathbf{y}_1 = A\mathbf{y}, \quad \boldsymbol{\mu}_1 = A\boldsymbol{\mu}, \quad A\Sigma A^T = \Sigma_{11} \quad (1.3.27)$$

따라서 다음 식들이 성립한다.

$$M_{\mathbf{y}_1}(\mathbf{t}_1) = \exp\left(\mathbf{t}_1^T A\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}_1^T A\Sigma A^T \mathbf{t}_1\right) = \exp\left(\mathbf{t}_1^T \boldsymbol{\mu}_1 + \frac{1}{2}\mathbf{t}_1^T \Sigma_{11} \mathbf{t}_1\right) \quad (1.3.28)$$

정의 1.3.2에서 알 수 있듯이, 다음 식이 성립한다.

$$\mathbf{y}_1 \stackrel{d}{\sim} \mathcal{N}_q(\boldsymbol{\mu}_1, \Sigma_{11}) \quad (1.3.29)$$

즉, 다변량정규확률벡터의 주변확률분포도 다변량정규확률분포이다. 따라서, 확률벡터  $\mathbf{y}_1$  의 결합확률밀도함수는 다음과 같다.

$$f_{\mathbf{y}_1}(\mathbf{y}_1) = \frac{1}{(2\pi)^{\frac{q}{2}} |\Sigma_{11}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} [\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma_{11}^{-1} [\mathbf{y}_1 - \boldsymbol{\mu}_1]\right) \quad (1.3.30)$$

**예제 1.3.3** 확률변수들  $x$ 와  $y$ 가 서로 독립인 표준정규확률변수들이면, 확률변수  $z = |x|\text{sign}(y)$ 도 표준정규확률변수이다. 즉, 이 주변확률분포들이 단변량정규확률분포들이라 하더라도 이들의 결합확률분포는 다변량정규확률분포가 아니다. ■

**예제 1.3.4** 서로 독립이고 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 를 따르는 확률변수들  $x_1, x_2, \dots, x_n$ 으로 이루어진 정규확률벡터  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ 의 평균벡터는  $\boldsymbol{\mu}\mathbf{1}$ 이고 분산공분산행렬은

$\sigma^2 I_n$ 이다. 표본평균  $\bar{x} \doteq \frac{1}{n} \sum_{i=1}^n x_i$ 에 대해서, 다음과 같은 확률벡터와 행렬을 정의하자.

$$\mathbf{y} \doteq \begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \\ \sqrt{n}\bar{x} \end{bmatrix} \doteq \begin{bmatrix} \mathbf{y}_1 \\ y_{n+1} \end{bmatrix}, \quad A \doteq \begin{bmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1 - \frac{1}{n} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \end{bmatrix} \doteq \begin{bmatrix} A_1 \\ \frac{1}{\sqrt{n}} \mathbf{1}^T \end{bmatrix} \quad (1)$$

식  $\mathbf{y} = A\mathbf{x}$ 가 성립함을 자명하다. 따라서 다음 식이 성립한다.

$$\mathbf{y} \stackrel{d}{\sim} \mathcal{N}_{p+1} \left( \begin{bmatrix} \mu A_1 \mathbf{1} \\ \sqrt{n}\mu \end{bmatrix}, \sigma^2 \begin{bmatrix} A_1 A_1^T & O \\ O & 1 \end{bmatrix} \right) \quad (2)$$

식 (2)에서 알 수 있듯이, 확률변수  $y_{n+1} = \sqrt{n}\bar{x}$ 와 확률벡터  $\mathbf{y}_1 = [x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}]^T$ 는 서로 독립이다. 따라서, 확률변수  $\bar{x}$ 와 확률변수  $\sum_{i=1}^n [x_i - \bar{x}]^2$ 는 서로 독립이다. 또한 다음 식이 성립한다.

$$\frac{\sqrt{n}[\bar{x} - \mu]}{\sigma} \stackrel{d}{\sim} \mathcal{N}(0, 1) \quad (3)$$

■

**예제 1.3.5** 예제 1.3.4를 다시 살펴보자. 서로 독립이고 표준정규확률분포  $\mathcal{N}(0, 1)$ 을 따르는 확률변수들  $x_1, x_2, \dots, x_n$ 으로 이루어진 정규확률벡터  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ 의 평균벡터는  $\mathbf{0}$ 이고 분산공분산행렬은  $I_n$ 이다. 다음과 같이 행렬  $Q$ 를 정의하자.

$$Q \doteq \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{-3}{\sqrt{12}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \frac{1}{\sqrt{[n-1]n}} & \frac{1}{\sqrt{[n-1]n}} & \frac{1}{\sqrt{[n-1]n}} & \frac{1}{\sqrt{[n-1]n}} & \cdots & \frac{1}{\sqrt{[n-1]n}} & \frac{1-n}{\sqrt{[n-1]n}} \end{bmatrix} \quad (1)$$

다음 식이 성립함을 쉽게 증명할 수 있다.

$$QQ^T = I_n \tag{2}$$

즉,  $Q$ 는 직교행렬 (orthonormal matrix)이다. 벡터  $\mathbf{y} = [y_1, y_2, \dots, y_n]^T \doteq Q\mathbf{x}$ 는 다음 식을 만족한다.

$$\mathbf{y} \stackrel{d}{\sim} \mathcal{N}_n(\mathbf{0}, I_n) \tag{3}$$

따라서, 다음 식들이 성립한다.

$$\sqrt{n}\bar{x} = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i = y_1 \stackrel{d}{\sim} \mathcal{N}(0, 1) \tag{4}$$

$$\sum_{i=1}^n [x_i - \bar{x}]^2 = \sum_{i=1}^n x_i^2 - [\sqrt{n}\bar{x}]^2 = \sum_{i=1}^n y_i^2 - y_1^2 = \sum_{i=2}^n y_i^2 \stackrel{d}{\sim} \chi_{n-1}^2 \tag{5}$$

또한,  $\sqrt{n}\bar{x} = y_1$ 과  $\sum_{i=1}^n [x_i - \bar{x}]^2 = \sum_{i=2}^n y_i^2$ 는 독립이다. 따라서, 다음 식이 성립한다.

$$\frac{\sqrt{n}\bar{x}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n [x_i - \bar{x}]^2}} \stackrel{d}{\sim} t_{n-1} \tag{7}$$

식 (7)을 다음과 같이 확장할 수 있다. 서로 독립이고 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 을 따르는 확률변수들  $x_1, x_2, \dots, x_n$ 에 대해서, 다음 식이 성립한다.

$$\frac{\sqrt{n}[\bar{x} - \mu]}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n [x_i - \bar{x}]^2}} \stackrel{d}{\sim} t_{n-1} \tag{8}$$

■

다음 보조정리의 증명은 최병선 [3, pp. 43-4]를 참조하라.

**보조정리 1.3.1**

행렬  $A$ 를 다음과 같이 분할하자.

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

여기서  $A_{11}$  과  $A_{22}$ 는 정방행렬들이다. 만약 행렬  $A_{11}$  이나 행렬  $A_{22}$ 가 비정칙이면, 다음과 같이 Schur상보행렬들 (Schur complements)을 정의하자.

$$A_{11.2} \doteq A_{11} - A_{12}A_{22}^{-1}A_{21}, \quad A_{22.1} \doteq A_{22} - A_{21}A_{11}^{-1}A_{12}$$

다음 명제들이 성립한다.

a) 만약  $A_{22}$ 가 정칙이면, 다음 식이 성립한다.

$$A = \begin{bmatrix} I & A_{12}A_{22}^{-1} \\ O & I \end{bmatrix} \begin{bmatrix} A_{11.2} & O \\ O & A_{11} \end{bmatrix} \begin{bmatrix} I & O \\ A_{22}^{-1}A_{21} & I \end{bmatrix}$$

b) 만약  $A_{11}$ 이 정칙이면, 다음 식이 성립한다.

$$A = \begin{bmatrix} I & O \\ A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & O \\ O & A_{22.1} \end{bmatrix} \begin{bmatrix} I & A_{11}^{-1}A_{12} \\ O & I \end{bmatrix}$$

c) 만약  $A$ 와  $A_{22}$ 가 정칙이면, 다음 식이 성립한다.

$$A^{-1} = \begin{bmatrix} A_{11.2}^{-1} & -A_{11.2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11.2}^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}A_{11.2}^{-1}A_{12}A_{22}^{-1} \end{bmatrix}$$

d) 만약  $A$ 와  $A_{11}$ 이 정칙이면, 다음 식이 성립한다.

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}A_{22.1}^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22.1}^{-1} \\ -A_{22.1}^{-1}A_{21}A_{11}^{-1} & A_{22.1}^{-1} \end{bmatrix}$$

e) 만약  $A_{11}$  과  $A_{22}$ 가 정칙이면, 다음 식들이 성립한다.

$$A_{11.2}^{-1} = A_{11}^{-1} + A_{11}^{-1}A_{12}A_{22.1}^{-1}A_{21}A_{11}^{-1}$$

$$A_{22.1}^{-1} = A_{22}^{-1} + A_{22}^{-1}A_{21}A_{11.2}^{-1}A_{12}A_{22}^{-1}$$

f) 만약  $A_{11}$  이 정칙이면, 다음 식이 성립한다.

$$\det(A) = \det(A_{11}) \det(A_{22.1})$$

g) 만약  $A_{22}$  이 정칙이면, 다음 식이 성립한다.

$$\det(A) = \det(A_{22}) \det(A_{11.2})$$

식 (1.3.18)과 식 (1.3.30)에서 알 수 있듯이, 다음 식들이 성립한다.

$$f_{\mathbf{y}_2|\mathbf{y}_1}(\mathbf{y}_2) = \frac{f_{\mathbf{y}}(\mathbf{y})}{f_{\mathbf{y}_1}(\mathbf{y}_1)} = \frac{\frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}[\mathbf{y} - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{y} - \boldsymbol{\mu}]\right)}{\frac{1}{(2\pi)^{\frac{q}{2}} |\Sigma_{11}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}[\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma_{11}^{-1} [\mathbf{y}_1 - \boldsymbol{\mu}_1]\right)} \quad (1.3.31)$$

dl 함수  $f_{\mathbf{y}_2|\mathbf{y}_1}(\mathbf{y}_2)$ 는 확률벡터  $\mathbf{y}_1$ 이 주어진 조건 하에서 확률벡터  $\mathbf{y}_1$ 의 조건부확률밀도함수 (conditional probability density)이다. 보조정리 1.3.1에서 알 수 있듯이, 다음 식이 성립한다.

$$\begin{aligned} & [\mathbf{y} - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{y} - \boldsymbol{\mu}] - [\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma_{11}^{-1} [\mathbf{y}_1 - \boldsymbol{\mu}_1] \\ &= [\mathbf{y}_2 - \boldsymbol{\mu}_2 - \Sigma_{21} \Sigma_{11}^{-1} [\mathbf{y}_1 - \boldsymbol{\mu}_1]]^T \Sigma_{22.1}^{-1} [\mathbf{y}_2 - \boldsymbol{\mu}_2 - \Sigma_{21} \Sigma_{11}^{-1} [\mathbf{y}_1 - \boldsymbol{\mu}_1]] \end{aligned} \quad (1.3.32)$$

식 (1.3.32)을 유도하는 자세한 과정은 예제 1.3.7을 참조하라.

다음 확률벡터를 정의하자.

$$\mathbf{z} \doteq \mathbf{y}_2 - \boldsymbol{\mu}_2 - \Sigma_{21} \Sigma_{11}^{-1} [\mathbf{y}_1 - \boldsymbol{\mu}_1] \quad (1.3.33)$$

다음 식이 성립한다.

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{z} \end{bmatrix} = \begin{bmatrix} I_q & O \\ -\Sigma_{21} \Sigma_{11}^{-1} & I_{p-q} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \quad (1.3.34)$$

다음 식들이 성립한다.

$$\begin{aligned}
 \text{Var}(\mathbf{y}_1, \mathbf{z}) &= \text{Var}(\mathbf{y}_1, \mathbf{y}_2 - \boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}[\mathbf{y}_1 - \boldsymbol{\mu}_1]) \\
 &= \text{Var}(\mathbf{y}_1, \mathbf{y}_2) - \text{Var}(\mathbf{y}_1, \boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}[\mathbf{y}_1 - \boldsymbol{\mu}_1]) \\
 &= \Sigma_{12} - \Sigma_{11}^{-1}\Sigma_{11}\Sigma_{12} \\
 &= O
 \end{aligned} \tag{1.3.35}$$

즉, 확률벡터들  $\mathbf{y}_1$  과  $\mathbf{z}$  는 서로 독립이다. 따라서, 다음 식들이 성립한다.

$$E(\mathbf{z} | \mathbf{y}_1) = E(\mathbf{z}) = E(\mathbf{y}_2 - \boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}[\mathbf{y}_1 - \boldsymbol{\mu}_1]) = \mathbf{0} \tag{1.3.36}$$

식 (1.3.33)와 식 (1.3.36)에서 알 수 있듯이, 다음 식들이 성립한다.

$$E(\mathbf{y}_2 | \mathbf{y}_1) = E(\mathbf{z} | \mathbf{y}_1) + \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}[\mathbf{y}_1 - \boldsymbol{\mu}_1] = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}[\mathbf{y}_1 - \boldsymbol{\mu}_1] \tag{1.3.37}$$

확률벡터들  $\mathbf{y}_1$  과  $\mathbf{z}$  가 서로 독립이므로, 다음 식이 성립한다.

$$\text{Var}(\mathbf{z} | \mathbf{y}_1) = \text{Var}(\mathbf{z}) \tag{1.3.38}$$

식 (1.3.33)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\text{Var}(\mathbf{y}_2 | \mathbf{y}_1) = \text{Var}(\mathbf{z} + \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}[\mathbf{y}_1 - \boldsymbol{\mu}_1] | \mathbf{y}_1) = \text{Var}(\mathbf{z} | \mathbf{y}_1) \tag{1.3.39}$$

다음 식들이 성립한다.

$$\begin{aligned}
 \text{Var}(\mathbf{z}) &= \text{Var}(\mathbf{y}_2 - \boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}[\mathbf{y}_1 - \boldsymbol{\mu}_1]) \\
 &= \text{Var}(\mathbf{y}_2 - \boldsymbol{\mu}_2) - 2\text{Cov}(\mathbf{y}_2 - \boldsymbol{\mu}_2, \Sigma_{21}\Sigma_{11}^{-1}[\mathbf{y}_1 - \boldsymbol{\mu}_1]) + \text{Var}(\Sigma_{21}\Sigma_{11}^{-1}[\mathbf{y}_1 - \boldsymbol{\mu}_1]) \\
 &= \Sigma_{22} - 2\Sigma_{21}[\Sigma_{21}\Sigma_{11}^{-1}]^T + \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \\
 &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \Sigma_{22 \cdot 1}
 \end{aligned} \tag{1.3.40}$$

식 (1.3.38)~식 (1.3.40)에서 알 수 있듯이, 다음 식이 성립한다.

$$\text{Var}(\mathbf{y}_2 | \mathbf{y}_1) = \Sigma_{22 \cdot 1} \tag{1.3.41}$$

지금까지 내용을 정리하면, 다음과 같다.

**정리 1.3.2**

확률벡터  $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 에 대해서 확률벡터  $\mathbf{y}_1$ 의 주변확률분포와 확률벡터  $\mathbf{y}_1$ 이 주어진 조건 하에서 확률벡터  $\mathbf{y}_2$ 의 조건부확률분포는 각각 다음과 같다.

$$\mathbf{y}_1 \stackrel{d}{\sim} \mathcal{N}_q(\boldsymbol{\mu}_1, \Sigma_{11})$$

$$\mathbf{y}_2 | \mathbf{y}_1 \stackrel{d}{\sim} \mathcal{N}_{p-q}(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}[\mathbf{y}_1 - \boldsymbol{\mu}_1], \Sigma_{22 \cdot 1})$$

**예제 1.3.6** 확률벡터  $\mathbf{y} \stackrel{d}{\sim} \mathcal{N}_3(\boldsymbol{\mu}, \Sigma)$ 를 살펴보자. 여기서 벡터들과 행렬은 각각 다음과 같다.

$$\mathbf{y} \doteq \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \quad \boldsymbol{\mu} \doteq \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \quad \Sigma \doteq \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & 0 \\ \rho^2 & 0 & 1 \end{bmatrix} \tag{1}$$

식 (1)에서 알 수 있듯이, 조건부확률벡터  $[y_1, y_2]^T | y_3$ 의 조건부확률분포함수는 다음과 같다.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} | y_3 \stackrel{d}{\sim} \mathcal{N}_2 \left( \begin{bmatrix} \mu_1 + \rho^2 [y_3 - \mu_3] \\ \mu_2 \end{bmatrix}, \begin{bmatrix} 1 - \rho^4 & \rho \\ \rho & 1 \end{bmatrix} \right) \tag{2}$$

■

**예제 1.3.7** 확률벡터  $\mathbf{y} \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 의 확률밀도함수는 다음과 같다.

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{1}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} \exp \left( -\frac{1}{2} Q(\mathbf{y}_1, \mathbf{y}_2) \right) \tag{1}$$

여기서  $Q(\mathbf{y}_1, \mathbf{y}_2) \doteq [\mathbf{y} - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{y} - \boldsymbol{\mu}]$ 이다. 식 (1.3.24)와 마찬가지로 확률벡터  $\mathbf{y}$ 를 분해하기로 하자. 또한, 행렬  $\Sigma$ 의 역행렬을 다음과 같이 표기하자.

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \doteq \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \quad (2)$$

다음 식들이 성립한다.

$$\Sigma_{12}^T = \Sigma_{21}, \quad (\Sigma^{12})^T = \Sigma^{21} \quad (3)$$

다음 식들이 성립한다.

$$\begin{aligned} Q(\mathbf{y}_1, \mathbf{y}_2) &= \begin{bmatrix} \mathbf{y}_1 - \boldsymbol{\mu}_1 \\ \mathbf{y}_2 - \boldsymbol{\mu}_2 \end{bmatrix}^T \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 - \boldsymbol{\mu}_1 \\ \mathbf{y}_2 - \boldsymbol{\mu}_2 \end{bmatrix} \\ &= [\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma^{11} [\mathbf{y}_1 - \boldsymbol{\mu}_1] + 2[\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma^{12} [\mathbf{y}_2 - \boldsymbol{\mu}_2] + [\mathbf{y}_2 - \boldsymbol{\mu}_2]^T \Sigma^{22} [\mathbf{y}_2 - \boldsymbol{\mu}_2] \end{aligned} \quad (4)$$

보조정리 1.3.1에서 정의했듯이, 행렬  $\Sigma_{11.2}$ 와 행렬  $\Sigma_{22.1}$ 는 다음과 같다.

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \quad (5)$$

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \quad (6)$$

보조정리 1.3.1에서 알 수 있듯이, 다음 식들이 성립한다.

$$\Sigma^{11} = \Sigma_{11.2}^{-1} = \Sigma_{11}^{-1} + \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22.1}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \quad (7)$$

$$\Sigma^{22} = \Sigma_{22.1}^{-1} = \Sigma_{22}^{-1} + \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11.2}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \quad (8)$$

$$\Sigma^{12} = -\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22.1}^{-1} = (\Sigma^{21})^T \quad (9)$$

다음 벡터와 이차형식들을 정의하자.

$$\boldsymbol{\mu}_{2.1} \doteq \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}[\mathbf{y}_1 - \boldsymbol{\mu}_1] \quad (10)$$

$$Q_1(\mathbf{y}_1) = [\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma_{11}^{-1}[\mathbf{y}_1 - \boldsymbol{\mu}_1] \quad (11)$$

$$Q_{2.1}(\mathbf{y}_2 | \mathbf{y}_1) = [\mathbf{y}_2 - \boldsymbol{\mu}_{2.1}]^T \Sigma_{22.1}^{-1}[\mathbf{y}_2 - \boldsymbol{\mu}_{2.1}] \quad (12)$$

식 (7)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\begin{aligned} & [\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma^{11}[\mathbf{y}_1 - \boldsymbol{\mu}_1] \\ &= [\mathbf{y}_1 - \boldsymbol{\mu}_1]^T [\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}][\mathbf{y}_1 - \boldsymbol{\mu}_1] \\ &= Q_1(\mathbf{y}_1) + [\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}[\mathbf{y}_1 - \boldsymbol{\mu}_1] \end{aligned} \quad (13)$$

식 (8)과 식 (9)에서 알 수 있듯이, 다음 식들이 성립한다.

$$[\mathbf{y}_2 - \boldsymbol{\mu}_2]^T \Sigma^{22}[\mathbf{y}_2 - \boldsymbol{\mu}_2] = [\mathbf{y}_2 - \boldsymbol{\mu}_2]^T \Sigma_{22.1}^{-1}[\mathbf{y}_2 - \boldsymbol{\mu}_2] \quad (14)$$

$$2[\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma^{12}[\mathbf{y}_2 - \boldsymbol{\mu}_2] = -2[\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1}[\mathbf{y}_2 - \boldsymbol{\mu}_2] \quad (15)$$

식 (10)과 식 (12)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\begin{aligned} & Q_{2.1}(\mathbf{y}_2 | \mathbf{y}_1) \\ &= [\mathbf{y}_2 - \boldsymbol{\mu}_{2.1}]^T \Sigma_{22.1}^{-1}[\mathbf{y}_2 - \boldsymbol{\mu}_{2.1}] \\ &= [\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}[\mathbf{y}_1 - \boldsymbol{\mu}_1] \\ &\quad - 2[\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1}[\mathbf{y}_2 - \boldsymbol{\mu}_2] + [\mathbf{y}_2 - \boldsymbol{\mu}_2]^T \Sigma_{22.1}^{-1}[\mathbf{y}_2 - \boldsymbol{\mu}_2] \end{aligned} \quad (16)$$

식 (13)과 식 (16)에서 알 수 있듯이, 다음 식이 성립한다.

$$\begin{aligned} & Q_{2.1}(\mathbf{y}_2 | \mathbf{y}_1) + Q_1(\mathbf{y}_1) \\ &= [\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma^{11}[\mathbf{y}_1 - \boldsymbol{\mu}_1] \\ &\quad - 2[\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22.1}^{-1}[\mathbf{y}_2 - \boldsymbol{\mu}_2] + [\mathbf{y}_2 - \boldsymbol{\mu}_2]^T \Sigma_{22.1}^{-1}[\mathbf{y}_2 - \boldsymbol{\mu}_2] \end{aligned} \quad (17)$$

식 (14)와 식 (15)를 식 (17)에 대입하면, 다음 식이 성립한다.

$$\begin{aligned}
 & Q_{2.1}(\mathbf{y}_2 | \mathbf{y}_1) + Q_1(\mathbf{y}_1) \\
 &= [\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma^{11} [\mathbf{y}_1 - \boldsymbol{\mu}_1] + 2[\mathbf{y}_1 - \boldsymbol{\mu}_1]^T \Sigma^{12} [\mathbf{y}_2 - \boldsymbol{\mu}_2] + [\mathbf{y}_2 - \boldsymbol{\mu}_2]^T \Sigma^{22} [\mathbf{y}_2 - \boldsymbol{\mu}_2] \\
 &= [\mathbf{y} - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{y} - \boldsymbol{\mu}]
 \end{aligned} \tag{18}$$

식 (4)와 식 (18)에서 알 수 있듯이, 다음 식이 성립한다.

$$Q(\mathbf{y}_1, \mathbf{y}_2) = Q_1(\mathbf{y}_1) + Q_{2.1}(\mathbf{y}_2 | \mathbf{y}_1) \tag{19}$$

Schur보완의 성질에서 알 수 있듯이, 다음 식이 성립한다.

$$|\Sigma| = |\Sigma_{11}| |\Sigma_{22.1}| \tag{20}$$

식 (19)와 식 (20)을 이용하면, 식 (1)을 다음과 같이 쓸 수 있다.

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{q}{2}} |\Sigma_{11}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} Q_1(\mathbf{y}_1)\right) \cdot \frac{1}{(2\pi)^{\frac{[p-q]}{2}} |\Sigma_{22.1}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} Q_{2.1}(\mathbf{y}_2 | \mathbf{y}_1)\right) \tag{21}$$

즉, 다음 식이 성립함을 증명할 수 있다.

$$f_{\mathbf{y}}(\mathbf{y}) = f_{\mathbf{y}_1}(\mathbf{y}_1) f_{\mathbf{y}_2 | \mathbf{y}_1}(\mathbf{y}_2) \tag{22}$$

■

## 제 1.4 절 Wishart 확률분포

### 1.4.1 Wishart 확률분포의 정의

다변량정규확률분포  $\mathcal{N}_p(\mathbf{0}, \Sigma)$ 를 따르는 서로 독립인 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ 의 결합확률 밀도함수는 다음과 같다.

$$f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m | \Sigma) = \frac{1}{[2\pi]^{pm/2} |\Sigma|^{m/2}} \exp\left(-\frac{1}{2} \sum_{j=1}^m \mathbf{x}_j^T \Sigma^{-1} \mathbf{x}_j\right) \tag{1.4.1}$$

여기서  $m > p$ 이다. 다음 식들이 성립한다.

$$\begin{aligned} \sum_{j=1}^m \mathbf{x}_j^T \Sigma^{-1} \mathbf{x}_j &= \text{Tr} \left( \sum_{j=1}^m \mathbf{x}_j^T \Sigma^{-1} \mathbf{x}_j \right) = \sum_{j=1}^m \text{Tr} (\mathbf{x}_j^T \Sigma^{-1} \mathbf{x}_j) \\ &= \sum_{j=1}^m \text{Tr} (\Sigma^{-1} \mathbf{x}_j \mathbf{x}_j^T) = \text{Tr} \left( \Sigma^{-1} \sum_{j=1}^m \mathbf{x}_j \mathbf{x}_j^T \right) \end{aligned} \quad (1.4.2)$$

식 (1.4.2)에서 알 수 있듯이, 식 (1.4.1)의 결합확률밀도함수를 다음과 같이 쓸 수 있다.

$$f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m | \Sigma) \propto |\Sigma|^{-m/2} \exp \left( -\frac{1}{2} \text{Tr} \left( \Sigma^{-1} \sum_{j=1}^m \mathbf{x}_j \mathbf{x}_j^T \right) \right) \quad (1.4.3)$$

식 (1.4.3)에서 알 수 있듯이,  $\sum_{j=1}^m \mathbf{x}_j \mathbf{x}_j^T$ 는  $\Sigma$ 의 충분통계량이다.

다변량선형회귀모형을 다루기 위해서는 다음과 같이 정의되는 Wishart 확률분포를 알아야 한다.

#### 정의 1.4.1: Wishart 확률분포

다변량정규확률분포  $\mathcal{N}_p(\mathbf{0}, \Sigma)$ 를 따르는 서로 독립인 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ 에 대해서 행렬  $W = \sum_{i=1}^m \mathbf{x}_i \mathbf{x}_i^T$ 는 크기행렬 (scale matrix)이  $\Sigma$ 이고 자유도가  $m$ 인 Wishart 확률분포를 따른다고 한다. 이  $W$ 를 모수들을  $(\Sigma, m)$ 으로 하는 Wishart 확률행렬이라 부르고, 다음과 같이 표기하자.

$$W \stackrel{d}{\sim} \mathcal{W}_p(\Sigma, m)$$

여기서  $W$ 는  $p \times p$  차원 행렬로서 거의 모든 점에서 양정치이다.

만약  $\Sigma = I_p$ 이면, Wishart 확률분포의 표준형이라 한다. 다음 식이 성립한다.

$$W_1(\sigma^2, m) = \sigma^2 \chi_m^2 \quad (1.4.4)$$

따라서, Wishart 확률분포는 카이제곱확률분포를 다변량으로 확장한 것이다.

식 (1.4.3)으로부터 확률행렬  $W$ 의 확률밀도함수가 다음과 같음을 알 수 있다.

$$\pi(W) = \frac{1}{C_W} |W|^{[m-p-1]/2} \exp\left(-\frac{1}{2}\text{Tr}(\Sigma^{-1}W)\right) \quad (1.4.5)$$

여기서  $C_W$ 는 다음과 같다.

$$C_W \doteq 2^{mp/2} |\Sigma|^{m/2} \left[ \pi^{p[p-1]/4} \prod_{i=1}^p \Gamma\left(\frac{m}{2} - \frac{i-1}{2}\right) \right] \doteq 2^{mp/2} |\Sigma|^{m/2} \Gamma_p\left(\frac{m}{2}\right) \quad (1.4.6)$$

여기서 다변량감마함수  $\Gamma_p(m)$ 은 다음과 같이 정의된다.

$$\Gamma_p(m) \doteq \pi^{p[p-1]/4} \prod_{i=1}^p \Gamma\left(m - \frac{i-1}{2}\right) \quad (1.4.7)$$

식 (1.4.5)는  $W$ 가 양정치인 경우 성립하고, 그렇지 않으면 확률밀도함수값은 0이다. 식 (1.4.5)를 유도하는 자세한 과정은 Anderson [15, 제7.2절]을 참조하라.

다음 명제들이 성립한다. 이에 대한 자세한 내용은 Anderson [15, 제7장]을 참조하라.

#### 명제 1.4.1

확률분포  $W_p(\Sigma, m)$ 을 따르는 Wishart 확률행렬  $W$ 와  $r \times p$  행렬  $A$ 에 대해서 다음 식들이 성립한다.

- a)  $E(W) = m\Sigma$
- b)  $\text{Mode}(W) = [m - p - 1]\Sigma, \quad (m \geq p + 1)$
- c)  $AWA^T \stackrel{d}{\sim} \mathcal{W}_r(A\Sigma A^T, m)$

#### 명제 1.4.2

만약  $W_1$ 과  $W_2$ 가 서로 독립이고 또한  $W_i$ 가 확률분포  $\mathcal{W}_p(\Sigma, m_i)$ 를 따르면, 다음 식이 성립한다.

$$W_1 + W_2 \stackrel{d}{\sim} \mathcal{W}_p(\Sigma, m_1 + m_2)$$

**명제 1.4.3**

확률분포  $\mathcal{W}_p(\Sigma, m)$  를 따르는 Wishart 확률행렬  $W$  에 대해서, 행렬  $W$  와  $\Sigma$  를 다음과 같이 분해하자.

$$W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

여기서  $W_{11}$  과  $\Sigma_{11}$  은  $p_1 \times p_1$  행렬들이다. 다음 행렬들을 정의하자.

$$W_{11 \cdot 2} \doteq W_{11} - W_{12}W_{22}^{-1}W_{21}$$

$$\Sigma_{11 \cdot 2} \doteq \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

다음 식들이 성립한다.

- a) 확률행렬  $W_{11 \cdot 2}$  는  $(W_{12}, W_{22})$  와 독립이다.
- b)  $W_{11 \cdot 2} \stackrel{d}{\sim} \mathcal{W}_{p_1}(\Sigma_{11 \cdot 2}, m - [p - p_1])$
- c)  $W_{22} \stackrel{d}{\sim} \mathcal{W}_{p-p_1}(\Sigma_{22}, m)$

**1.4.2 역Wishart 확률분포**

확률행렬  $W \stackrel{d}{\sim} \mathcal{W}_p(\Sigma, m)$  에 대해서, 역행렬  $V = W^{-1}$  의 확률밀도함수는 다음과 같다.

$$\pi(V) = \frac{1}{C_V} |V|^{-[m+p+1]/2} \exp\left(-\frac{1}{2} \text{Tr}(\Sigma^{-1}V^{-1})\right) \quad (1.4.8)$$

여기서 조건  $m > p - 1$  이 성립해야 하고,  $C_V$  는 다음과 같다.

$$C_V \doteq 2^{mp/2} |\Sigma|^{m/2} \left[ \pi^{p[p-1]/4} \prod_{i=1}^p \Gamma\left(\frac{m}{2} - \frac{i-1}{2}\right) \right] = 2^{mp/2} |\Sigma|^{m/2} \Gamma_p\left(\frac{m}{2}\right) \quad (1.4.9)$$

식 (1.4.8) 를 유도하는 자세한 과정은 Anderson [15, 제7.7절] 을 참조하라. 이에 해당하는 확률분포를 모수들을  $(\Sigma^{-1}, m)$  으로 하는 역Wishart 확률행렬 (inverse Wishart probability distribution) 이라 부르고,  $\mathcal{IW}_p(\Sigma^{-1}, m)$  으로 표기하자. 즉, 다음 명제가 성립한다.

$$W \stackrel{d}{\sim} \mathcal{W}_p(\Sigma, m) \Leftrightarrow W^{-1} \stackrel{d}{\sim} \mathcal{IW}_p(\Sigma^{-1}, m) \quad (1.4.10)$$

**명제 1.4.4**

확률분포  $\mathcal{IW}_p(\Psi, m)$ 를 따르는 역Wishart 확률행렬  $V$ 에 대해서, 다음 식들이 성립한다.

$$\text{a) } E(V) = \frac{1}{m-p-1}\Psi, \quad (m > p + 1)$$

$$\text{b) } \text{Mode}(V) = \frac{1}{m+p+1}\Psi$$

증명. 식  $\Psi^{-1} = CC^T$ 를 만족하는 가역행렬  $C$ 에 대해서 확률행렬  $U \doteq C^{-1}V^{-1}(C^{-1})^T$ 는 다음 식을 만족한다.

$$U \stackrel{d}{\sim} \mathcal{W}_p(I_p, m) \quad (1)$$

다음 식들이 성립한다.

$$E(V) = E\left((C^{-1})^T U^{-1} C^{-1}\right) = (C^{-1})^T E(U^{-1}) C^{-1} \quad (2)$$

행렬  $U$ 에서 제 $i$ 행과 제 $j$ 행을 바꾸는 동시에 제 $i$ 열과 제 $j$ 열을 바꾸어도 식 (1)이 성립한다. 이 대칭성에서 알 수 있듯이, 행렬  $E(U^{-1})$ 의 대각원소들이 동일하고 또한 비대각원소들도 동일하다. 즉, 다음 식을 만족하는 상수들  $k_1$ 과  $k_2$ 가 존재한다.

$$E(U^{-1}) = k_1 I_p + k_2 \mathbf{1}_p \mathbf{1}_p^T \quad (3)$$

임의의 직교행렬  $Q$ 에 대해서 다음 식이 성립한다.

$$QUQ^T \stackrel{d}{\sim} \mathcal{W}_p(I_p, m) \quad (4)$$

즉,  $E\left((QUQ^T)^{-1}\right)$ 는 직교행렬  $Q$ 에 의존하지 않는다. 따라서, 다음 식들이 성립한다.

$$E\left((QUQ^T)^{-1}\right) = QE(U^{-1})Q^T = E(U^{-1}) \quad (5)$$

식 (3)과 식 (5)에서 알 수 있듯이, 식  $k_2 = 0$ 가 성립한다. 따라서, 다음 식이 성립한다.

$$E(U^{-1}) = k_1 I_p \quad (6)$$

행렬  $U^{-1} = [u^{ij}]$ 의 대각원소는  $(\chi_{m-p+1}^2)^{-1}$ 를 따른다. 따라서, 다음 식들이 성립한다.

$$E(u^{ii}) = E\left((\chi_{m-p+1}^2)^{-1}\right) = \frac{1}{m-p-1} \quad (7)$$

식 (6)과 식 (7)에서 알 수 있듯이, 다음 식이 성립한다.

$$E(U^{-1}) = \frac{1}{m-p-1} I_p \quad (8)$$

따라서 성질 (a)가 성립한다.

성질 (b)의 증명은 독자들에게 숙제로 남겨둔다. 참고로 모드(mode)를 구하기 위해서는 당연히 다변량미적분을 사용하면 된다. 그러나, Schur분해를 먼저 적용한 다음 미적분을 적용하는 것이 쉬울 것이다. ■

### 1.4.3 표본확률분포

다변량정규확률분포  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 를 따르며 서로 독립인 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ 과  $N \times N$  직교행렬  $C = [c_{ij}]$ 에 대해서 다음 확률벡터들을 살펴보자.

$$\mathbf{y}_i \doteq \sum_{j=1}^N c_{ij} \mathbf{x}_j, \quad (i = 1, 2, \dots, N) \quad (1.4.11)$$

확률벡터  $\mathbf{y}_i$ 가 다변량정규확률분포를 따르고, 확률벡터  $\mathbf{y}_i$ 의 평균벡터는 다음과 같다.

$$E(\mathbf{y}_i) = \sum_{j=1}^N c_{ij} \boldsymbol{\mu} \doteq \boldsymbol{\nu}_i, \quad (i = 1, 2, \dots, N) \quad (1.4.12)$$

확률벡터  $\mathbf{y}_i$ 와 확률벡터  $\mathbf{y}_j$ 의 분산공분산행렬은 다음과 같다.

$$\begin{aligned}
 & E\left([\mathbf{y}_i - \boldsymbol{\nu}_i][\mathbf{y}_j - \boldsymbol{\nu}_j]^T\right) \\
 &= E\left(\left\{\sum_{a=1}^N c_{ia}[\mathbf{x}_a - \boldsymbol{\nu}_a]\right\}\left\{\sum_{b=1}^N c_{jb}[\mathbf{x}_b - \boldsymbol{\nu}_b]\right\}^T\right) \\
 &= \sum_{a=1}^N \sum_{b=1}^N c_{ia}c_{jb}E\left([\mathbf{x}_a - \boldsymbol{\nu}_a][\mathbf{x}_b - \boldsymbol{\nu}_b]^T\right) \\
 &= \sum_{a=1}^N \sum_{b=1}^N c_{ia}c_{jb}\delta_{a,b}\Sigma \\
 &= \left[\sum_{a=1}^N c_{ia}c_{ja}\right]\Sigma
 \end{aligned} \tag{1.4.13}$$

여기서  $\delta_{i,j}$ 는 Kronecker 델타함수이다. 행렬  $C$ 가 직교행렬이므로, 식 (1.4.13)에서 알 수 있듯이 다음 식들이 성립한다.

$$E\left([\mathbf{y}_i - \boldsymbol{\nu}_i][\mathbf{y}_j - \boldsymbol{\nu}_j]^T\right) = \delta_{i,j}\Sigma, \quad (i, j = 1, 2, \dots, N) \tag{1.4.14}$$

다음 식들이 성립한다.

$$\begin{aligned}
 \sum_{i=1}^N \mathbf{y}_i \mathbf{y}_i^T &= \sum_{i=1}^N \left[ \sum_{a=1}^N c_{ia} \mathbf{x}_a \right] \left[ \sum_{b=1}^N c_{ib} \mathbf{x}_b \right]^T = \sum_{a=1}^N \sum_{b=1}^N \mathbf{x}_a \mathbf{x}_b^T \left[ \sum_{i=1}^N c_{ia} c_{ib} \right] \\
 &= \sum_{a=1}^N \sum_{b=1}^N \mathbf{x}_a \mathbf{x}_b^T \delta_{a,b} = \sum_{a=1}^N \mathbf{x}_a \mathbf{x}_a^T
 \end{aligned} \tag{1.4.15}$$

여기서 세 번째 등호는  $C$ 가 직교행렬이어서 성립한다.

임의의 벡터  $\mathbf{b}$ 에 대해서 다음 식들이 성립한다.

$$\begin{aligned}
 & \sum_{i=1}^N [\mathbf{x}_i - \mathbf{b}][\mathbf{x}_i - \mathbf{b}]^T \\
 &= \sum_{i=1}^N ([\mathbf{x}_i - \bar{\mathbf{x}}] + [\bar{\mathbf{x}} - \mathbf{b}])([\mathbf{x}_i - \bar{\mathbf{x}}] + [\bar{\mathbf{x}} - \mathbf{b}])^T \\
 &= \sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}][\mathbf{x}_i - \bar{\mathbf{x}}]^T + \left[ \sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}] \right] [\bar{\mathbf{x}} - \mathbf{b}]^T \\
 &\quad + [\bar{\mathbf{x}} - \mathbf{b}] \left[ \sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}] \right]^T + \sum_{i=1}^N [\bar{\mathbf{x}} - \mathbf{b}][\bar{\mathbf{x}} - \mathbf{b}]^T
 \end{aligned} \tag{1.4.16}$$

즉, 다음 식이 성립한다.

$$\sum_{i=1}^N [\mathbf{x}_i - \mathbf{b}] [\mathbf{x}_i - \mathbf{b}]^T = \sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}] [\mathbf{x}_i - \bar{\mathbf{x}}]^T + N [\bar{\mathbf{x}} - \mathbf{b}] [\bar{\mathbf{x}} - \mathbf{b}]^T \quad (1.4.17)$$

첫 번째 열이  $[1/\sqrt{N}, 1/\sqrt{N}, \dots, 1/\sqrt{N}]$  인 직교행렬 (orthonormal matrix)  $Q = [q_{ij}]$  를 구성하자. 예를 들어, 나머지 열들을  $\mathbf{e}_2^T, \mathbf{e}_3^T, \dots, \mathbf{e}_N^T$  으로 채운 다음, Gram-Schmidt 정리를 적용해서 직교행렬  $Q$  를 만들 수 있다. 다음 벡터들을 정의하자.

$$\mathbf{z}_1 = \sum_{j=1}^N q_{1,j} \mathbf{x}_j = \sum_{j=1}^N \frac{1}{\sqrt{N}} \mathbf{x}_j = \sqrt{N} \bar{\mathbf{x}} \quad (1.4.18)$$

$$\mathbf{z}_i = \sum_{j=1}^N q_{i,j} \mathbf{x}_j, \quad (i = 2, 3, \dots, N) \quad (1.4.19)$$

행렬  $Q$  의 직교성에서 알 수 있듯이,  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$  은 서로 독립인  $p$  변량 정규확률벡터들이다. 다음 식들이 성립한다.

$$E(\mathbf{z}_1) = E(\sqrt{N} \bar{\mathbf{x}}) = \sqrt{N} \boldsymbol{\mu} \quad (1.4.20)$$

$$E(\mathbf{z}_i) = \sqrt{N} \sum_{j=1}^N q_{i,j} q_{i,j} E(\mathbf{x}_j) = \sqrt{N} \boldsymbol{\mu} \sum_{j=1}^N q_{i,j} q_{i,j} = \mathbf{0}, \quad (i = 2, 3, \dots, N) \quad (1.4.21)$$

식 (1.4.21) 은  $Q$  가 직교행렬이어서 성립한다. 다음 식들이 성립한다.

$$\mathbf{z}_1 \stackrel{d}{\sim} \mathcal{N}_p(\sqrt{N} \boldsymbol{\mu}, \Sigma) \quad (1.4.22)$$

$$\mathbf{z}_i \stackrel{d}{\sim} \mathcal{N}_p(\mathbf{0}, \Sigma), \quad (i = 2, 3, \dots, N) \quad (1.4.23)$$

식 (1.4.22) 에서 알 수 있듯이, 다음 식이 성립한다.

$$\sqrt{N} [\bar{\mathbf{x}} - \boldsymbol{\mu}] \stackrel{d}{\sim} \mathcal{N}_p(\mathbf{0}, \Sigma) \quad (1.4.24)$$

식 (1.4.17) 에서 알 수 있듯이, 다음 식이 성립한다.

$$\sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}] [\mathbf{x}_i - \bar{\mathbf{x}}]^T = \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T - N \bar{\mathbf{x}} \bar{\mathbf{x}}^T = \sum_{i=1}^N \mathbf{z}_i \mathbf{z}_i^T - N \bar{\mathbf{x}} \bar{\mathbf{x}}^T \quad (1.4.25)$$

여기서 두 번째 등호는 식 (1.4.15) 에 의해서 성립한다. 식 (1.4.18) 과 식 (1.4.25) 에서 알 수

있듯이, 다음 식이 성립한다.

$$\sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}] [\mathbf{x}_i - \bar{\mathbf{x}}]^T = \sum_{i=2}^N \mathbf{z}_i \mathbf{z}_i^T \quad (1.4.26)$$

식 (1.4.22)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\sqrt{N} [\bar{\mathbf{x}} - \boldsymbol{\mu}] = \mathbf{z}_1 - \sqrt{N} \boldsymbol{\mu} \stackrel{d}{\sim} \mathcal{N}_p(\mathbf{0}, \Sigma) \quad (1.4.27)$$

식 (1.4.23)과 식 (1.4.26)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}] [\mathbf{x}_i - \bar{\mathbf{x}}]^T = \sum_{i=2}^N \mathbf{z}_i \mathbf{z}_i^T \stackrel{d}{\sim} \mathcal{W}_p(\Sigma, N-1) \quad (1.4.28)$$

지금까지 내용을 정리하면, 다음과 같다.

#### 정리 1.4.1

다변량정규확률분포  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 를 따르며 서로 독립인 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ 에 대해서 다음 성질들이 성립한다.

- a)  $\sqrt{N} [\bar{\mathbf{x}} - \boldsymbol{\mu}] \stackrel{d}{\sim} \mathcal{N}_p(\mathbf{0}, \Sigma)$
- b)  $\sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}] [\mathbf{x}_i - \bar{\mathbf{x}}]^T \stackrel{d}{\sim} \mathcal{W}_p(\Sigma, N-1)$
- c) 확률벡터  $\sqrt{N} [\bar{\mathbf{x}} - \boldsymbol{\mu}]$ 와 확률행렬  $\sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}] [\mathbf{x}_i - \bar{\mathbf{x}}]^T$ 은 서로 독립이다.

#### 1.4.4 최우추정량

다변량정규확률분포  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 를 따르는 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ 의 우도함수는 다음과 같다.

$$\begin{aligned} L(\boldsymbol{\mu}, \Sigma) &\doteq \prod_{i=1}^N \frac{1}{|\Sigma|^{\frac{1}{2}} [2\pi]^{\frac{p}{2}}} \exp\left(-\frac{1}{2} [\mathbf{x}_i - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{x}_i - \boldsymbol{\mu}]\right) \\ &= |\Sigma|^{-N/2} [2\pi]^{-Np/2} \exp\left(-\frac{1}{2} \sum_{i=1}^N [\mathbf{x}_i - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{x}_i - \boldsymbol{\mu}]\right) \end{aligned} \quad (1.4.29)$$

또한, 대수우도함수는 다음과 같다.

$$\begin{aligned} l(\boldsymbol{\mu}, \Sigma) &\doteq \ln L(\boldsymbol{\mu}, \Sigma) \\ &= -\frac{Np}{2} \ln(2\pi) - \frac{N}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^N [\mathbf{x}_i - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{x}_i - \boldsymbol{\mu}] \end{aligned} \quad (1.4.30)$$

다음 식들이 성립한다.

$$\begin{aligned} &\sum_{i=1}^N [\mathbf{x}_i - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{x}_i - \boldsymbol{\mu}] \\ &= \sum_{i=1}^N \{[\mathbf{x}_i - \bar{\mathbf{x}}] + [\bar{\mathbf{x}} - \boldsymbol{\mu}]\}^T \Sigma^{-1} \{[\mathbf{x}_i - \bar{\mathbf{x}}] + [\bar{\mathbf{x}} - \boldsymbol{\mu}]\} \\ &= \sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}]^T \Sigma^{-1} [\mathbf{x}_i - \bar{\mathbf{x}}] + N[\bar{\mathbf{x}} - \boldsymbol{\mu}]^T \Sigma^{-1} [\bar{\mathbf{x}} - \boldsymbol{\mu}] \end{aligned} \quad (1.4.31)$$

행렬  $\Sigma$ 가 양정치이므로,  $\Sigma^{-1}$ 도 양정치행렬이다. 따라서, 다음 식이 성립한다.

$$\sum_{i=1}^N [\mathbf{x}_i - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{x}_i - \boldsymbol{\mu}] \geq \sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}]^T \Sigma^{-1} [\mathbf{x}_i - \bar{\mathbf{x}}] \quad (1.4.32)$$

다음과 같이 추정행렬  $\hat{\Sigma}$ 를 정의하자.

$$\hat{\Sigma} \doteq \frac{1}{N} \sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}] [\mathbf{x}_i - \bar{\mathbf{x}}]^T \quad (1.4.33)$$

식 (1.4.32)에서 알 수 있듯이, 다음 식이 성립한다.

$$l(\boldsymbol{\mu}, \Sigma) \leq l(\bar{\mathbf{x}}, \Sigma) = -\frac{Np}{2} \ln(2\pi) - \frac{N}{2} \ln |\Sigma| - \frac{N}{2} \text{Tr} \left( \Sigma^{-1} \hat{\Sigma} \right) \quad (1.4.34)$$

다음 식들이 성립한다.

$$\begin{aligned} l(\bar{\mathbf{x}}, \hat{\Sigma}) &= -\frac{Np}{2} \ln(2\pi) - \frac{N}{2} \ln |\hat{\Sigma}| - \frac{N}{2} \text{Tr} \left( \hat{\Sigma}^{-1} \hat{\Sigma} \right) \\ &= -\frac{Np}{2} \ln(2\pi) - \frac{N}{2} \ln |\hat{\Sigma}| - \frac{N^2}{2} \end{aligned} \quad (1.4.35)$$

따라서, 다음 식이 성립한다.

$$l(\bar{\mathbf{x}}, \hat{\Sigma}) - l(\bar{\mathbf{x}}, \Sigma) = \frac{N}{2} \left[ \text{Tr}(\Sigma^{-1} \hat{\Sigma}) - N - \ln |\Sigma^{-1} \hat{\Sigma}| \right] \quad (1.4.36)$$

행렬  $\Sigma^{-1} \hat{\Sigma}$ 가 양정치일 확률이 1이다. 행렬  $\Sigma^{-1} \hat{\Sigma}$ 의 고유값들을  $\eta_1, \eta_2, \dots, \eta_N$ 으로 표기하면, 다음 식들이 성립한다.

$$l(\bar{\mathbf{x}}, \hat{\Sigma}) - l(\bar{\mathbf{x}}, \Sigma) = \frac{N}{2} \sum_{i=1}^N [\eta_i - 1 - \ln \eta_i] \geq 0 \quad (1.4.37)$$

여기서 부등호는 식  $\ln x \leq x - 1$ 에 의해서 성립한다. 식 (1.4.34)과 식 (1.4.37)에서 알 수 있듯이, 다음 식이 성립한다.

$$l(\boldsymbol{\mu}, \Sigma) \leq l(\bar{\mathbf{x}}, \hat{\Sigma}) \quad (1.4.38)$$

따라서,  $\bar{\mathbf{x}}$ 와  $\hat{\Sigma}$ 는 각각  $\boldsymbol{\mu}$ 와  $\Sigma$ 의 최우추정량이다.

정리 1.4.1에서 알 수 있듯이,  $\bar{\mathbf{x}}$ 는 불편추정량이나  $\hat{\Sigma}$ 는 편의추정량이다. 다음 추정량을 정의하자.

$$S \doteq \frac{1}{N-1} \sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}] [\mathbf{x}_i - \bar{\mathbf{x}}]^T \quad (1.4.39)$$

정리 1.4.1에서 알 수 있듯이,  $S$ 는 불편추정량이다.

### 1.4.5 신뢰구간

#### 정리 1.4.2

정규확률벡터  $\mathbf{y} \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\nu}, \Omega)$ 와 양정치행렬  $\Omega$ 에 대해서,  $\mathbf{y}^T \Omega^{-1} \mathbf{y}$ 는 비중심모수가  $\boldsymbol{\nu}^T \Omega^{-1} \boldsymbol{\nu}$ 이고 자유도가  $p$ 인 비중심카이제곱확률분포(noncentral chi-squared distribution)를 따른다. 만약  $\boldsymbol{\nu} = \mathbf{0}$ 이면, 이 확률분포는 중심카이제곱확률분포(central chi-squared distribution)이다.

증명. 행렬  $\Omega$ 가 양정치행렬이므로, Schur분해에 의해서 식  $C \Omega C^T = I_p$ 를 만족하는 가역행렬  $C$ 가 존재한다. 확률벡터  $\mathbf{z} \doteq C \mathbf{y}$ 는 다변량정규확률분포를 따르며, 평균벡터는  $E(\mathbf{z}) = C \boldsymbol{\nu}$

이고 분산공분산행렬은 다음과 같다.

$$E\left([z - C\boldsymbol{\nu}][z - C\boldsymbol{\nu}]^T\right) = CE\left([\mathbf{y} - \boldsymbol{\nu}][\mathbf{y} - \boldsymbol{\nu}]^T\right)C^T = C\Omega C^T = I_p \quad (1)$$

즉, 다음 식이 성립한다.

$$\mathbf{z} \stackrel{d}{\sim} \mathcal{N}_p(C\boldsymbol{\nu}, I_p) \quad (2)$$

따라서  $\mathbf{z}^T \mathbf{z}$ 는 비중심모수가  $\boldsymbol{\nu}^T C^T C \boldsymbol{\nu} = \boldsymbol{\nu}^T \Omega^{-1} \boldsymbol{\nu}$ 이고 자유도가  $p$ 인 비중심카이제곱확률분포를 따른다. 다음 식들이 성립한다.

$$\mathbf{y}^T \Omega^{-1} \mathbf{y} = \mathbf{z}^T (C^T)^{-1} \Omega^{-1} C^{-1} \mathbf{z} = \mathbf{z}^T (C\Omega C^T)^{-1} \mathbf{z} = \mathbf{z}^T \mathbf{z} \quad (3)$$

따라서  $\mathbf{y}^T \Omega^{-1} \mathbf{y}$ 는 비중심모수가  $\boldsymbol{\nu}^T \Omega^{-1} \boldsymbol{\nu}$ 이고 자유도가  $p$ 인 비중심카이제곱확률분포를 따른다. 만약  $\boldsymbol{\nu} = \mathbf{0}$ 이면, 비중심모수는 0이다. ■

**예제 1.4.1** 정규확률벡터  $\mathbf{x} \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 에 대해서 다음 식이 성립한다.

$$\Sigma^{-\frac{1}{2}} [\mathbf{x} - \boldsymbol{\mu}] \stackrel{d}{\sim} \mathcal{N}(\mathbf{0}, I_p) \quad (1)$$

즉,  $\Sigma^{-\frac{1}{2}} [\mathbf{x} - \boldsymbol{\mu}]$ 의 원소들은 서로 독립인 표준정규확률변수들이다. 따라서 다음 식들이 성립한다.

$$[\mathbf{x} - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{x} - \boldsymbol{\mu}] = \left(\Sigma^{-\frac{1}{2}} [\mathbf{x} - \boldsymbol{\mu}]\right)^T \left(\Sigma^{-\frac{1}{2}} [\mathbf{x} - \boldsymbol{\mu}]\right) \stackrel{d}{\sim} \chi_p^2 \quad (2)$$

■

#### 따름정리 1.4.1

다변량정규확률분포  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 를 따르는 서로 독립인 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ 에 대해서  $N[\bar{\mathbf{x}} - \boldsymbol{\mu}_0]^T \Sigma^{-1} [\bar{\mathbf{x}} - \boldsymbol{\mu}_0]$ 은 비중심모수가  $N[\boldsymbol{\mu} - \boldsymbol{\mu}_0]^T \Sigma^{-1} [\boldsymbol{\mu} - \boldsymbol{\mu}_0]$ 이고 자유도가  $p$ 인 비중심카이제곱확률분포를 따른다.

**예제 1.4.2** 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ 가 서로 독립이고 다변량정규확률분포  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 를 따른다고 가정하고, 확률변수  $y \doteq N[\bar{\mathbf{x}} - \boldsymbol{\mu}]^T \Sigma^{-1} [\bar{\mathbf{x}} - \boldsymbol{\mu}]$ 의 특성함수를 구해보자.

다음 식들이 성립한다.

$$\begin{aligned} \phi_y(\theta) &= E(\exp(i\theta y)) = E\left(\exp\left(i\theta N[\bar{\mathbf{x}} - \boldsymbol{\mu}]^T \Sigma^{-1} [\bar{\mathbf{x}} - \boldsymbol{\mu}]\right)\right) \\ &= \int \exp\left(i\theta N[\bar{\mathbf{x}} - \boldsymbol{\mu}]^T \Sigma^{-1} [\bar{\mathbf{x}} - \boldsymbol{\mu}]\right) \\ &\quad \cdot [2\pi]^{-p/2} \left|\frac{1}{N}\Sigma\right|^{-1/2} \exp\left(-\frac{1}{2}N[\bar{\mathbf{x}} - \boldsymbol{\mu}]^T \Sigma^{-1} [\bar{\mathbf{x}} - \boldsymbol{\mu}]\right) d\mathbf{x} \\ &= [2\pi]^{-p/2} \left|\frac{1}{N}\Sigma\right|^{-1/2} \int \exp\left(-\frac{1}{2}N[\bar{\mathbf{x}} - \boldsymbol{\mu}]^T [\Sigma^{-1} - 2i\theta\Sigma^{-1}] [\bar{\mathbf{x}} - \boldsymbol{\mu}]\right) d\mathbf{x} \end{aligned} \quad (1)$$

따라서 다음 식이 성립한다.

$$\begin{aligned} \phi_y(\theta) &= \left|\frac{1}{N}[\Sigma^{-1} - 2i\theta\Sigma^{-1}]^{-1}\right|^{1/2} \left|\frac{1}{N}\Sigma\right|^{-1/2} \cdot [2\pi]^{-p/2} \left|\frac{1}{N}[\Sigma^{-1} - 2i\theta\Sigma^{-1}]^{-1}\right|^{-1/2} \\ &\quad \cdot \int \exp\left(-\frac{1}{2}N[\bar{\mathbf{x}} - \boldsymbol{\mu}]^T [\Sigma^{-1} - 2i\theta\Sigma^{-1}] [\bar{\mathbf{x}} - \boldsymbol{\mu}]\right) d\mathbf{x} \end{aligned} \quad (2)$$

다변량정규확률밀도함수를 지지대(support)에서 적분하면 1이라는 사실을 이용하면, 식 (2)를 다음과 같이 쓸 수 있다.

$$\phi_y(\theta) = |I_p - 2i\theta I_p|^{-1/2} = [1 - 2i\theta]^{-p/2} \quad (3)$$

함수  $[1 - 2i\theta]^{-p/2}$ 는 자유도가  $p$ 인 카이제곱확률분포의 특성함수이다. 즉, 다음 식이 성립한다.

$$N[\bar{\mathbf{x}} - \boldsymbol{\mu}]^T \Sigma^{-1} [\bar{\mathbf{x}} - \boldsymbol{\mu}] \stackrel{d}{\sim} \chi_p^2 \quad (4)$$

■

#### 1.4.6 손실함수와 위험함수

모수벡터  $\boldsymbol{\theta}$ 가 모수공간  $\Omega$ 에 속하는 확률분포  $P_{\boldsymbol{\theta}}$ 와 이 확률분포를 따르는 확률벡터  $\mathbf{x}$ 를 살펴보자. 모수벡터  $\boldsymbol{\theta}$ 를 추정하는 결정과정(decision procedure)  $\boldsymbol{\delta}(\mathbf{x})$ 의 손실함수(loss function)

$L(\boldsymbol{\theta}, \boldsymbol{\delta}(\mathbf{x}))$ 는 비음함수 (nonnegative function)이다. 이 결정과정  $\boldsymbol{\delta}(\mathbf{x})$ 을 평가하기 위해서 다음과 같이 정의되는 위험함수 (risk function)를 사용한다.

$$R(\boldsymbol{\theta}, \boldsymbol{\delta}) \doteq E_{\boldsymbol{\theta}}(L(\boldsymbol{\theta}, \boldsymbol{\delta}(\mathbf{x}))) \quad (1.4.40)$$

예를 들어, 오차제곱  $\|\boldsymbol{\theta} - \boldsymbol{\delta}(\mathbf{x})\|^2$ 은 손실함수이고 평균제곱오차 (mean squared error)  $E_{\boldsymbol{\theta}}(\|\boldsymbol{\theta} - \boldsymbol{\delta}(\mathbf{x})\|^2)$ 는 위험함수이다. 다음 식이 성립하면, 결정과정  $\boldsymbol{\delta} = \boldsymbol{\delta}(\mathbf{x})$ 는 결정과정  $\boldsymbol{\delta}^* = \boldsymbol{\delta}(\mathbf{x}^*)$ 만큼 좋다고 (good) 한다.

$$R(\boldsymbol{\theta}, \boldsymbol{\delta}) \leq R(\boldsymbol{\theta}, \boldsymbol{\delta}^*), \quad (\boldsymbol{\theta} \in \Omega) \quad (1.4.41)$$

만약 식 (1.4.41)이 성립하는 동시에 다음 식을 만족하는  $\boldsymbol{\theta}_0 \in \Omega$ 가 적어도 하나 존재하면, 결정과정  $\boldsymbol{\delta}(\mathbf{x})$ 는 결정과정  $\boldsymbol{\delta}(\mathbf{x}^*)$ 보다 더 좋다고 (better) 한다.

$$R(\boldsymbol{\theta}_0, \boldsymbol{\delta}) < R(\boldsymbol{\theta}_0, \boldsymbol{\delta}^*) \quad (1.4.42)$$

만약  $\boldsymbol{\delta}^*(\mathbf{x})$ 보다 더 좋은 결정과정  $\boldsymbol{\delta}(\mathbf{x})$ 가 존재하면, 결정과정  $\boldsymbol{\delta}^*(\mathbf{x})$ 는 선택불가능 (inadmissible)이라 한다. 만약  $\boldsymbol{\delta}^*(\mathbf{x})$ 가 선택불가능이 아니면, 선택가능 (admissible)이라 한다. 결정과정들의 모임 (class)  $\mathcal{D}$ 에 속하지 않는 어떤 결정과정에 대해  $\mathcal{D}$ 에 속하는 더 좋은 결정과정이 존재하면, 모임  $\mathcal{D}$ 는 완전하다고 (complete) 한다. 만약 완전한 모임  $\mathcal{D}$ 의 완전한 진부모임 (proper complete subclass)을 포함하지 않으면,  $\mathcal{D}$ 가 최소완전하다고 (minimal complete) 한다. 만약 최소완전한 모임  $\mathcal{D}$ 가 존재하면,  $\mathcal{D}$ 는 선택가능 결정과정들의 모임이다.

모수벡터  $\boldsymbol{\theta}$ 의 사전확률밀도함수를  $\rho(\boldsymbol{\theta})$ 라 하면, 결정과정  $\boldsymbol{\delta}(\mathbf{x})$ 의 평균손실 (average loss)은 다음과 같다.

$$r(\rho, \boldsymbol{\delta}) = E_{\rho}(R(\boldsymbol{\theta}, \boldsymbol{\delta})) = E_{\rho}(E_{\boldsymbol{\theta}}(L(\boldsymbol{\theta}, \boldsymbol{\delta}(\mathbf{x})))) \quad (1.4.43)$$

주어진 사전확률밀도함수  $\rho(\boldsymbol{\theta})$ 에 대해서 평균손실  $r(\rho, \boldsymbol{\delta})$ 를 최소화하는 결정과정  $\boldsymbol{\delta}(\mathbf{x})$ 를 Bayes 결정과정이라 하고, Bayes 결정과정의 평균손실을 Bayes 위험이라 한다. 일반적인 조건 하에서 Bayes 결정과정은 선택가능하고, 또한 선택가능한 결정과정은 Bayes 결정과정이거나

Bayes 결정과정의 극한이다. 다음 식들이 성립한다.

$$f(\mathbf{x}) = \int_{\Omega} f(\mathbf{x} | \boldsymbol{\theta}) \rho(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (1.4.44)$$

$$g(\boldsymbol{\theta} | \mathbf{x}) \doteq \frac{f(\mathbf{x} | \boldsymbol{\theta}) \rho(\boldsymbol{\theta})}{f(\mathbf{x})} \quad (1.4.45)$$

식 (1.4.44) 와 식 (1.4.45) 를 이용해서, 결정과정  $\boldsymbol{\delta}(\mathbf{x})$  의 평균손실 (average loss) 을 다음과 같이 쓸 수 있다.

$$\begin{aligned} r(\rho, \boldsymbol{\delta}) &= \int_{\Omega} \int L(\boldsymbol{\theta}, \boldsymbol{\delta}(\mathbf{x})) f(\mathbf{x} | \boldsymbol{\theta}) \rho(\boldsymbol{\theta}) d\mathbf{x} d\boldsymbol{\theta} \\ &= \int \left[ \int_{\Omega} L(\boldsymbol{\theta}, \boldsymbol{\delta}(\mathbf{x})) g(\boldsymbol{\theta} | \mathbf{x}) d\boldsymbol{\theta} \right] f(\mathbf{x}) d\mathbf{x} \\ &= \int E_{\boldsymbol{\theta} | \mathbf{x}} (L(\boldsymbol{\theta}, \boldsymbol{\delta}(\mathbf{x}))) f(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (1.4.46)$$

주어진  $\mathbf{x}$  에 대해서  $E_{\boldsymbol{\theta} | \mathbf{x}} (L(\boldsymbol{\theta}, \boldsymbol{\delta}(\mathbf{x})))$  는 모수벡터  $\boldsymbol{\theta}$  의 사후기대값이다. 식 (1.4.46) 에서 알 수 있듯이, 각  $\mathbf{x}$  에 대해서 사후기대값  $E_{\boldsymbol{\theta} | \mathbf{x}} (L(\boldsymbol{\theta}, \boldsymbol{\delta}(\mathbf{x})))$  를 최소화하는 결정과정이 평균손실  $r(\rho, \boldsymbol{\delta})$  를 최소화하는 결정과정이다.

양정치행렬  $Q$  에 대해서 다음과 같은 손실함수를 살펴보자.

$$L(\boldsymbol{\theta} - \mathbf{d}(\mathbf{x})) = [\boldsymbol{\theta} - \mathbf{d}(\mathbf{x})]^T Q [\boldsymbol{\theta} - \mathbf{d}(\mathbf{x})] \quad (1.4.47)$$

다음 식이 성립한다.

$$\begin{aligned} E_{\boldsymbol{\theta} | \mathbf{x}} (L(\boldsymbol{\theta}, \mathbf{d}(\mathbf{x}))) &= E_{\boldsymbol{\theta} | \mathbf{x}} \left( [\boldsymbol{\theta} - E(\boldsymbol{\theta} | \mathbf{x})]^T Q [\boldsymbol{\theta} - E(\boldsymbol{\theta} | \mathbf{x})] \right) \\ &\quad + [E(\boldsymbol{\theta} | \mathbf{x}) - \mathbf{d}(\mathbf{x})]^T Q [E(\boldsymbol{\theta} | \mathbf{x}) - \mathbf{d}(\mathbf{x})] \end{aligned} \quad (1.4.48)$$

따라서, 각  $\mathbf{x}$  에 대해  $\mathbf{d}(\mathbf{x})$  가 사후평균벡터일때, 즉  $\mathbf{d}(\mathbf{x}) = E(\boldsymbol{\theta} | \mathbf{x})$  일때 사후기대값  $E_{\boldsymbol{\theta} | \mathbf{x}} (L(\boldsymbol{\theta}, \boldsymbol{\delta}(\mathbf{x})))$  가 최소가 된다.

### 정리 1.4.3

서로 독립인 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  이 다변량정규확률분포  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$  를 따르고, 모수

벡터  $\boldsymbol{\mu}$ 의 사전확률분포가  $\mathcal{N}(\boldsymbol{\nu}, \Phi)$ 라고 하면,  $\boldsymbol{\mu}$ 의 사후확률분포는 다음과 같다.

$$\boldsymbol{\mu}(\mathbf{x}_1, \dots, \mathbf{x}_N) \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\mu}|X, \Sigma|X)$$

여기서  $X \doteq (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ 이고, 사후평균벡터  $\boldsymbol{\mu}|X$ 와 사후분산공분산행렬  $\Sigma|X$ 는 각각 다음과 같다.

$$\begin{aligned}\boldsymbol{\mu}|X &\doteq \Psi \left[ \Psi + \frac{1}{N}\Sigma \right]^{-1} \bar{\mathbf{x}} + \frac{1}{N}\Sigma \left[ \Psi + \frac{1}{N}\Sigma \right]^{-1} \boldsymbol{\nu} \\ \Sigma|X &\doteq \Psi - \Psi \left[ \Psi + \frac{1}{N}\Sigma \right]^{-1} \Psi\end{aligned}$$

증명. 평균벡터  $\bar{\mathbf{x}}$ 는 모수벡터  $\boldsymbol{\mu}$ 의 충분통계량이므로 관찰점들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  대신에 평균벡터  $\bar{\mathbf{x}}$ 를 사용할 수 있다. 다음 식들이 성립한다.

$$\bar{\mathbf{x}} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\nu}, \quad \boldsymbol{\nu} \stackrel{d}{\sim} \mathcal{N}\left(\mathbf{0}, \frac{1}{N}\Sigma\right) \quad (1)$$

또한  $\boldsymbol{\mu}$ 와  $\boldsymbol{\nu}$ 는 서로 독립이다. 따라서 다음 식이 성립한다.

$$\begin{bmatrix} \boldsymbol{\mu} \\ \bar{\mathbf{x}} \end{bmatrix} \stackrel{d}{\sim} \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\nu} \\ \boldsymbol{\nu} \end{bmatrix}, \begin{bmatrix} \Psi & \Psi \\ \Psi & \Psi + \frac{1}{N}\Sigma \end{bmatrix}\right) \quad (2)$$

정리 1.3.2에서 알 수 있듯이, 다음 식이 성립한다.

$$\boldsymbol{\mu}|\bar{\mathbf{x}} \stackrel{d}{\sim} \mathcal{N}\left(\boldsymbol{\nu} + \Psi \left[ \Psi + \frac{1}{N}\Sigma \right]^{-1} [\bar{\mathbf{x}} - \boldsymbol{\nu}], \Psi - \Psi \left[ \Psi + \frac{1}{N}\Sigma \right]^{-1} \Psi\right) \quad (3)$$

다음 식이 성립한다.

$$I - \Psi \left[ \Psi + \frac{1}{N}\Sigma \right]^{-1} = \frac{1}{N}\Sigma \left[ \Psi + \frac{1}{N}\Sigma \right]^{-1} \quad (4)$$

식 (4)에서 알 수 있듯이, 다음 식이 성립한다.

$$\boldsymbol{\nu} + \Psi \left[ \Psi + \frac{1}{N}\Sigma \right]^{-1} [\bar{\mathbf{x}} - \boldsymbol{\nu}] = \Psi \left[ \Psi + \frac{1}{N}\Sigma \right]^{-1} \bar{\mathbf{x}} + \frac{1}{N}\Sigma \left[ \Psi + \frac{1}{N}\Sigma \right]^{-1} \boldsymbol{\nu} \quad (5)$$

■

**따름정리 1.4.2**

서로 독립인 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  이 다변량정규확률분포  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$  를 따르고, 모수 벡터  $\boldsymbol{\mu}$  의 사전확률분포가  $\mathcal{N}(\boldsymbol{\nu}, \Phi)$  라고 하자. 손실함수가  $[\boldsymbol{\theta} - \mathbf{d}(\mathbf{x})]^T Q [\boldsymbol{\theta} - \mathbf{d}(\mathbf{x})]$  라고 하면, 평균벡터  $\boldsymbol{\mu}$  의 Bayes 결정과정은 다음과 같다.

$$\boldsymbol{\mu}|X = \Psi \left[ \Psi + \frac{1}{N} \Sigma \right]^{-1} \bar{\mathbf{x}} + \frac{1}{N} \Sigma \left[ \Psi + \frac{1}{N} \Sigma \right]^{-1} \boldsymbol{\nu}$$

**1.4.7 미니맥스결정과정**

다음 식을 만족하면,  $\mathbf{d}_0(\mathbf{x})$  를 미니맥스결정과정 (minimax decision procedure) 이라 한다.

$$\sup_{\boldsymbol{\theta}} R(\boldsymbol{\theta}, \mathbf{d}_0) \stackrel{d}{=} \inf_{\boldsymbol{\delta}} \sup_{\boldsymbol{\theta}} R(\boldsymbol{\theta}, \boldsymbol{\delta}) \quad (1.4.49)$$

다음 정리에서 알 수 있듯이, 평균벡터  $\bar{\mathbf{x}}$  는 미니맥스결정과정이다.

**정리 1.4.4**

서로 독립인 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  이 다변량정규확률분포  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$  를 따르고, 모수 벡터  $\boldsymbol{\mu}$  의 사전확률분포가  $\mathcal{N}(\boldsymbol{\nu}, \Phi)$  라고 하자. 손실함수가  $[\boldsymbol{\theta} - \mathbf{d}(\mathbf{x})]^T Q [\boldsymbol{\theta} - \mathbf{d}(\mathbf{x})]$  일때 표본평균벡터  $\bar{\mathbf{x}}$  는 평균벡터  $\boldsymbol{\mu}$  의 미니맥스결정과정이다.

증명. Ferguson [40, Section 2.11]에 의하면, 다음 정리가 성립한다.

적절한  $\rho$  에 대한 Bayes 결정과정  $\boldsymbol{\delta}_\rho$  가 임의의 양수  $\epsilon$  에 대해서 다음 식을 만족한다고 하자.

$$r(\rho, \boldsymbol{\delta}_0) \leq r(\rho, \boldsymbol{\delta}_\rho) + \epsilon$$

이러한 조건 하에서  $R(\boldsymbol{\theta}, \boldsymbol{\delta}_0)$  가 상수이면,  $\boldsymbol{\delta}_0$  는 미니맥스결정과정이다.

다음 식들이 성립한다.

$$\begin{aligned} R(\boldsymbol{\mu}, \bar{\boldsymbol{x}}) &= E\left([\bar{\boldsymbol{x}} - \boldsymbol{\mu}]^T Q [\bar{\boldsymbol{x}} - \boldsymbol{\mu}]\right) \\ &= E\left(\text{Tr}\left(Q [\bar{\boldsymbol{x}} - \boldsymbol{\mu}] [\bar{\boldsymbol{x}} - \boldsymbol{\mu}]^T\right)\right) = \frac{1}{N} \text{Tr}(Q\Sigma) \end{aligned} \quad (1)$$

즉,  $R(\boldsymbol{\mu}, \bar{\boldsymbol{x}})$ 는 상수이다.

결정과정  $d(\bar{\boldsymbol{x}})$ 를 다음과 같이 정의하자.

$$d(\bar{\boldsymbol{x}}) \doteq \Psi \left[ \Psi + \frac{1}{N} \Sigma \right]^{-1} \bar{\boldsymbol{x}} + \frac{1}{N} \Sigma \left[ \Psi + \frac{1}{N} \Sigma \right]^{-1} \boldsymbol{\nu} \quad (2)$$

결정과정  $d(\bar{\boldsymbol{x}})$ 의 평균위험은 다음과 같다.

$$\begin{aligned} &E_{\bar{\boldsymbol{x}}}\left(E_{\boldsymbol{\mu}}\left(\text{Tr}\left(Q [d(\bar{\boldsymbol{x}}) - \boldsymbol{\mu}] [d(\bar{\boldsymbol{x}}) - \boldsymbol{\mu}]^T \mid \bar{\boldsymbol{x}}\right)\right)\right) \\ &= E_{\bar{\boldsymbol{x}}}\left(E_{\boldsymbol{\mu}}\left(\text{Tr}\left(Q \left\{ \Psi - \Psi \left[ \Psi + \frac{1}{N} \Sigma \right]^{-1} \Psi \right\}\right)\right)\right) \\ &= \text{Tr}\left(Q \left\{ I - \Psi \left[ \Psi + \frac{1}{N} \Sigma \right]^{-1} \right\} \Psi\right) \\ &= \frac{1}{N} \text{Tr}\left(Q \Sigma \left[ \Psi + \frac{1}{N} \Sigma \right]^{-1} \Psi\right) \\ &= \frac{1}{N} \text{Tr}\left(Q \Sigma \left[ I + \frac{1}{N} \Sigma \Psi^{-1} \right]^{-1}\right) \end{aligned} \quad (3)$$

따라서 다음 식이 성립한다.

$$\lim_{\Psi^{-1} \rightarrow \mathbf{0}} E_{\bar{\boldsymbol{x}}}\left(E_{\boldsymbol{\mu}}\left(\text{Tr}\left(Q [d(\bar{\boldsymbol{x}}) - \boldsymbol{\mu}] [d(\bar{\boldsymbol{x}}) - \boldsymbol{\mu}]^T \mid \bar{\boldsymbol{x}}\right)\right)\right) = \frac{1}{N} \text{Tr}(Q\Sigma) \quad (4)$$

Ferguson에 기술된 정리와 식 (4)에서 알 수 있듯이, 표본평균벡터  $\bar{\boldsymbol{x}}$ 는 평균벡터  $\boldsymbol{\mu}$ 의 미니맥스결정과정이다. ■

#### 1.4.8 James-Stein 추정벡터

함수  $f(\cdot)$ 가  $C^1$ 급이고, 또한 다음 식을 만족한다고 가정하자.

$$\int_{-\infty}^{\infty} |f'(x)| n(x; \theta, 1) dx < \infty \quad (1.4.50)$$

여기서  $n(\cdot; \mu, \sigma^2)$  는  $\mathcal{N}(\mu, \sigma^2)$  의 확률밀도함수이다. 다음 식들이 성립한다.

$$\begin{aligned} & \int_{-\infty}^{\infty} [f(x) - f(\theta)][x - \theta]n(x; \theta, 1)dx \\ &= \int_{\theta}^{\infty} [f(x) - f(\theta)][x - \theta]n(x; \theta, 1)dx + \int_{-\infty}^{\theta} [f(x) - f(\theta)][x - \theta]n(x; \theta, 1)dx \\ &= \int_{\theta}^{\infty} \int_{\theta}^x f'(y)[x - \theta]n(x; \theta, 1)dydx - \int_{-\infty}^{\theta} \int_x^{\theta} f'(y)[x - \theta]n(x; \theta, 1)dydx \quad (1.4.51) \end{aligned}$$

식 (1.4.51) 에 Fubini정리를 적용하면, 다음 식이 성립한다.

$$\begin{aligned} & \int_{-\infty}^{\infty} [f(x) - f(\theta)][x - \theta]n(x; \theta, 1)dx \\ &= \int_{\theta}^{\infty} \int_y^{\infty} f'(y)[x - \theta]n(x; \theta, 1)dx dy - \int_{-\infty}^{\theta} \int_{-\infty}^y f'(y)[x - \theta]n(x; \theta, 1)dx dy \quad (1.4.52) \end{aligned}$$

다음 식이 성립한다.

$$\int_y^{\infty} [x - \theta]n(x; \theta, 1)dx = \int_y^{\infty} \left[ -\frac{dn(x; \theta, 1)}{dx} \right] dx = n(y; \theta, 1) \quad (1.4.53)$$

식 (1.4.53) 을 식 (1.4.52) 에 대입하면, 다음 식을 얻는다.

$$\int_{-\infty}^{\infty} [f(x) - f(\theta)][x - \theta]n(x; \theta, 1)dx = \int_{-\infty}^{\infty} f'(y)n(y; \theta, 1)dy \quad (1.4.54)$$

다음 식이 성립한다.

$$f(\theta) \int_{-\infty}^{\infty} [x - \theta]n(x; \theta, 1)dx = 0 \quad (1.4.55)$$

식 (1.4.54) 와 식 (1.4.55) 에서 알 수 있듯이, 다음 식이 성립한다.

$$\int_{-\infty}^{\infty} f(x)[x - \theta]n(x; \theta, 1)dx = \int_{-\infty}^{\infty} f'(y)n(y; \theta, 1)dy \quad (1.4.56)$$

다변량정규확률분포  $\mathcal{N}_p(\boldsymbol{\mu}, I_p)$  를 따르는 서로 독립인 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  에 대해서 표본평균벡터  $\mathbf{y}$  는 다음과 같다.

$$\mathbf{y} \doteq \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \quad (1.4.57)$$

여기서  $p \geq 3$ 이다. 모수벡터  $\boldsymbol{\mu}$ 의 추정벡터로  $\mathbf{y}$ 와 함께 다음과 같이 정의되는 James-Stein 추정벡터  $\mathbf{m}(\mathbf{y})$ 를 고려해보자.

$$\mathbf{m}(\mathbf{y}) \doteq \left[ 1 - \frac{p-2}{\|\mathbf{y} - \boldsymbol{\nu}\|^2} \right] [\mathbf{y} - \boldsymbol{\nu}] + \boldsymbol{\nu} \quad (1.4.58)$$

여기서  $\boldsymbol{\nu}$ 는 임의의 고정된 (fixed) 벡터이다. 이 James-Stein 추정벡터  $\mathbf{m}(\mathbf{y})$ 는 고정된  $\boldsymbol{\nu}$ 를 향해서  $\mathbf{y}$ 를 축소한 추정벡터이다. 다음 식들이 성립한다.

$$\begin{aligned} & E_{\boldsymbol{\mu}} (\|\mathbf{m}(\mathbf{y}) - \boldsymbol{\mu}\|^2) \\ &= E_{\boldsymbol{\mu}} \left( \left\| [\mathbf{y} - \boldsymbol{\nu}] - \frac{p-2}{\|\mathbf{y} - \boldsymbol{\nu}\|^2} [\mathbf{y} - \boldsymbol{\nu}] + \boldsymbol{\nu} - \boldsymbol{\mu} \right\|^2 \right) \\ &= E_{\boldsymbol{\mu}} \left( \left\| [\mathbf{y} - \boldsymbol{\mu}] - \frac{p-2}{\|\mathbf{y} - \boldsymbol{\nu}\|^2} [\mathbf{y} - \boldsymbol{\nu}] \right\|^2 \right) \\ &= E_{\boldsymbol{\mu}} (\|\mathbf{y} - \boldsymbol{\mu}\|^2) + E_{\boldsymbol{\mu}} \left( \frac{[p-2]^2}{\|\mathbf{y} - \boldsymbol{\nu}\|^2} \right) - 2[p-2] E_{\boldsymbol{\mu}} \left( \frac{[\mathbf{y} - \boldsymbol{\mu}]^T [\mathbf{y} - \boldsymbol{\nu}]}{\|\mathbf{y} - \boldsymbol{\nu}\|^2} \right) \end{aligned} \quad (1.4.59)$$

따라서 다음 식들이 성립한다.

$$\begin{aligned} & E_{\boldsymbol{\mu}} (\|\mathbf{y} - \boldsymbol{\mu}\|^2) - E_{\boldsymbol{\mu}} (\|\mathbf{m}(\mathbf{y}) - \boldsymbol{\mu}\|^2) \\ &= 2[p-2] E_{\boldsymbol{\mu}} \left( \frac{[\mathbf{y} - \boldsymbol{\mu}]^T [\mathbf{y} - \boldsymbol{\nu}]}{\|\mathbf{y} - \boldsymbol{\nu}\|^2} \right) - E_{\boldsymbol{\mu}} \left( \frac{[p-2]^2}{\|\mathbf{y} - \boldsymbol{\nu}\|^2} \right) \\ &= 2[p-2] E_{\boldsymbol{\mu}} \left( \sum_{i=1}^p f(y_i) [y_i - \mu_i] \right) - E_{\boldsymbol{\mu}} \left( \frac{[p-2]^2}{\|\mathbf{y} - \boldsymbol{\nu}\|^2} \right) \end{aligned} \quad (1.4.60)$$

여기서  $f(y_i)$ 는 다음과 같다.

$$f(y_i) \doteq \frac{y_i - \nu_i}{\sum_{j=1}^p [y_j - \nu_j]^2} = \frac{y_i - \nu_i}{\|\mathbf{y} - \boldsymbol{\nu}\|^2} \quad (1.4.61)$$

이 함수의 도함수는 다음과 같다.

$$f'(y_i) = \frac{1}{\sum_{j=1}^p [y_j - \nu_j]^2} - \frac{2[y_i - \nu_i]^2}{\left\{ \sum_{j=1}^p [y_j - \nu_j]^2 \right\}^2} = \frac{1}{\|\mathbf{y} - \boldsymbol{\nu}\|^2} - \frac{2[y_i - \nu_i]^2}{\|\mathbf{y} - \boldsymbol{\nu}\|^4} \quad (1.4.62)$$

식 (1.4.61)과 식 (1.4.62)를 식 (1.4.56)에 대입하면, 다음 식을 얻는다.

$$\begin{aligned} E_{\boldsymbol{\mu}} \left( \sum_{i=1}^p f(y_i)[y_i - \mu_i] \right) &= E_{\boldsymbol{\mu}} \left( \sum_{i=1}^p f'(y_i) \right) \\ &= E_{\boldsymbol{\mu}} \left( \frac{\sum_{i=1}^p 1}{\|\mathbf{y} - \boldsymbol{\nu}\|^2} - \frac{\sum_{i=1}^p 2[y_i - \nu_i]^2}{\|\mathbf{y} - \boldsymbol{\nu}\|^4} \right) = E_{\boldsymbol{\mu}} \left( \frac{p-2}{\|\mathbf{y} - \boldsymbol{\nu}\|^2} \right) \end{aligned} \quad (1.4.63)$$

식 (1.4.63)을 식 (1.4.60)에 대입하면, 다음 식들이 성립함을 알 수 있다.

$$E_{\boldsymbol{\mu}} (\|\mathbf{y} - \boldsymbol{\mu}\|^2) - E_{\boldsymbol{\mu}} (\|\mathbf{m}(\mathbf{y}) - \boldsymbol{\mu}\|^2) = E_{\boldsymbol{\mu}} \left( \frac{[p-2]^2}{\|\mathbf{y} - \boldsymbol{\nu}\|^2} \right) > 0 \quad (1.4.64)$$

즉, James-Stein 추정벡터의 기대손실함수 (expected loss function)  $E_{\boldsymbol{\mu}} (\|\mathbf{m}(\mathbf{y}) - \boldsymbol{\mu}\|^2)$ 가 최소제곱추정벡터의 기대손실함수  $E_{\boldsymbol{\mu}} (\|\mathbf{y} - \boldsymbol{\mu}\|^2)$ 보다 작다.

#### 1.4.9 Hotelling $T^2$ 통계량

다음과 같이 평균벡터  $\bar{\mathbf{x}}$ 의 표본공분산행렬 (sample covariance matrix)을 정의하자.

$$\hat{\Sigma}_{\bar{\mathbf{x}}} \doteq \frac{1}{N} S = \frac{1}{N[N-1]} \sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}][\mathbf{x}_i - \bar{\mathbf{x}}]^T \quad (1.4.65)$$

이 표본공분산행렬을 이용해서 다음과 같이 Hotelling  $T^2$  통계량을 정의하자.

$$T^2 = [\bar{\mathbf{x}} - \boldsymbol{\mu}]^T \hat{\Sigma}_{\bar{\mathbf{x}}}^{-1} [\bar{\mathbf{x}} - \boldsymbol{\mu}] \quad (1.4.66)$$

다음과 같은 벡터들을 정의하자.

$$\mathbf{y} = \sqrt{N} [\bar{\mathbf{x}} - \boldsymbol{\mu}_0], \quad \boldsymbol{\nu} = \sqrt{N} [\boldsymbol{\mu} - \boldsymbol{\mu}_0] \quad (1.4.67)$$

다음 식이 성립한다.

$$T^2 = \mathbf{y}^T S^{-1} \mathbf{y} \quad (1.4.68)$$

다음 식을 만족하는 가역행렬  $P$ 가 존재한다.

$$P\Sigma P^T = I_p \quad (1.4.69)$$

다음 벡터들과 행렬을 정의하자.

$$\mathbf{y}_* = [y_{i*}] = P\mathbf{y}, \quad \boldsymbol{\nu}_* = P\boldsymbol{\nu}, \quad \mathbf{z}_{i*} = P\mathbf{z}_i, \quad (i = 1, 2, \dots, N), \quad S_* \doteq PSP^T \quad (1.4.70)$$

다음 식들이 성립한다.

$$T^2 = \mathbf{y}_*^T S_*^{-1} \mathbf{y}_* \quad (1.4.71)$$

$$\boldsymbol{\nu}_*^T \boldsymbol{\nu}_* = \boldsymbol{\nu}^T \Sigma^{-1} \boldsymbol{\nu} \quad (1.4.72)$$

$$\mathbf{z}_{i*} \stackrel{d}{\sim} \mathcal{N}(\mathbf{0}, I_p), \quad (i = 2, 3, \dots, N) \quad (1.4.73)$$

제1.4.3소절에서 알 수 있듯이,  $\mathbf{y}_*, \mathbf{z}_{2*}, \dots, \mathbf{z}_{N*}$ 는 서로 독립이고, 또한 다음 식들이 성립한다.

$$\mathbf{y}_* \stackrel{d}{\sim} \mathcal{N}(\boldsymbol{\nu}_*, I_p) \quad (1.4.74)$$

$$\sum_{i=2}^N \mathbf{z}_{i*} \mathbf{z}_{i*}^T \stackrel{d}{\sim} \mathcal{W}(I_p, N-1) \quad (1.4.75)$$

식 (1.4.74)에서 알 수 있듯이, 다음 식이 성립한다.

$$\mathbf{y}_*^T \mathbf{y}_* \stackrel{d}{\sim} \chi_p^2(\boldsymbol{\nu}_*^T \boldsymbol{\nu}_*) \quad (1.4.76)$$

행렬  $Q = [q_{ij}]$ 의 첫 번째 행을 다음과 같이 선택하자.

$$q_{1j} \doteq \frac{1}{\mathbf{y}_*^T \mathbf{y}_*} y_{j*} \quad (1.4.77)$$

나머지  $[p-1]$  열들은  $Q$ 가 직교행렬이 되도록 선택한다. 행렬  $Q$ 는 확률벡터  $\mathbf{y}_*$ 의 함수이다.

다음 벡터와 행렬들을 정의하자.

$$\mathbf{u} = [u_i] = Q\mathbf{y}_*, \quad B = [b_{ij}] = [N-1]QS_*Q^T, \quad B^{-1} = [b^{ij}] \quad (1.4.78)$$

다음 식들이 성립한다.

$$u_1 = \sum_{j=1}^p q_{1j} y_{j*} = \sqrt{\mathbf{y}_*^T \mathbf{y}_*} \quad (1.4.79)$$

$$u_i = \sum_{j=1}^p q_{ij} y_{j*} = \sqrt{\mathbf{y}_*^T \mathbf{y}_*} \sum_{j=1}^p q_{ij} q_{1j} = 0, \quad (i = 2, 3, \dots, p) \quad (1.4.80)$$

따라서, 다음 식들이 성립한다.

$$\frac{1}{N-1} T^2 = \mathbf{u}^T B^{-1} \mathbf{u} = \begin{bmatrix} u_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}^T \begin{bmatrix} b^{11} & b^{12} & \dots & b^{1p} \\ b^{21} & b^{22} & \dots & b^{2p} \\ \vdots & \vdots & & \vdots \\ b^{p1} & b^{p2} & \dots & b^{pp} \end{bmatrix} \begin{bmatrix} u_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = u_1^2 b^{11} \quad (1.4.81)$$

다음과 같이 행렬  $B$ 를 분해하자.

$$B = \begin{bmatrix} b_{11} & \mathbf{b}_1^T \\ \mathbf{b}_1 & B_{22} \end{bmatrix} \quad (1.4.82)$$

다음 식들이 성립한다.

$$\frac{1}{b^{11}} = b_{11} - \mathbf{b}_1^T B_{22}^{-1} \mathbf{b}_1 = b_{11 \cdot 2, \dots, p} \quad (1.4.83)$$

식 (1.4.83)을 식 (1.4.81)에 대입하면, 다음 식들이 성립함을 알 수 있다.

$$\frac{1}{N-1} T^2 = \mathbf{u}^T B^{-1} \mathbf{u} = \frac{1}{b_{11 \cdot 2, \dots, p}} \mathbf{y}_*^T \mathbf{y}_* \quad (1.4.84)$$

지금부터는  $\mathbf{y}_*^T \mathbf{y}_*$ 와  $b_{11 \cdot 2, \dots, p}$ 는 서로 독립이며 각각은 카이제곱확률분포를 따른다는 것을 증명하기로 하자. 우선 다음 보조정리를 인용하자. 이 보조정리는 Anderson [15, p. 141]에서 인용한 것이다.

#### 보조정리 1.4.1

서로 독립인 확률벡터들  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m$ 이 다음 식들을 만족한다고 하자.

$$\mathbf{y}_i \stackrel{d}{\sim} \mathcal{N}_r(\Gamma \mathbf{w}_i, \Omega), \quad (i = 1, 2, \dots, m)$$

다음 행렬들을 정의하자.

$$H \doteq \sum_{i=1}^m \mathbf{w}_i \mathbf{w}_i^T, \quad G \doteq \sum_{i=1}^m \mathbf{y}_i \mathbf{w}_i^T H^{-1}$$

$$C \doteq \sum_{i=1}^m [\mathbf{y}_i - G \mathbf{w}_i] [\mathbf{y}_i - G \mathbf{w}_i]^T = \sum_{i=1}^m \mathbf{y}_i \mathbf{y}_i^T$$

확률벡터들  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$ 이 서로 독립이면 동일한 확률분포  $\mathcal{N}(\mathbf{0}, \Omega)$ 를 따른다고 하자. 행렬  $H$ 가 가역이라고 가정하면, 확률행렬  $C$ 의 결합확률분포와  $\sum_{i=1}^m \mathbf{u}_i \mathbf{u}_i^T$ 의 결합확률분포는 같고, 또한 확률행렬  $C$ 와 확률행렬  $G$ 는 서로 독립이다.

다음과 같이 확률벡터들을 정의하자.

$$\mathbf{v}_i = Q \mathbf{z}_{i*}, \quad (i = 2, 3, \dots, N) \quad (1.4.85)$$

확률벡터들  $\mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_N$ 은 서로 독립이며 결합확률분포  $\mathcal{N}(\mathbf{0}, I_p)$ 를 따른다. 확률행렬  $Q$ 가 주어진 조건 하에서 확률행렬  $B$ 의 조건부확률분포와 동일한 조건 하에서 확률행렬  $\sum_{i=2}^N \mathbf{v}_i \mathbf{v}_i^T$ 의 조건부확률분포는 같다. 보조정리 1.4.1에서 알 수 있듯이, 다음 식이 성립한다.

$$b_{11 \cdot 2, \dots, p} | Q \stackrel{d}{=} \sum_{i=1}^{[N-1]-[p-1]} w_i^2 \quad (1.4.86)$$

여기서  $w_1, w_2, \dots, w_{N-p}$ 는 서로 독립인 표준정규확률변수들이다. 조건부 확률변수  $b_{11 \cdot 2, \dots, p} | Q$ 의 확률분포는  $Q$ 에 의존하지 않는다. 또한, 확률행렬  $Q$ 는 확률벡터  $\mathbf{y}_*$ 의 함수이다. 따라서, 확률벡터  $\mathbf{y}_*$ 와  $b_{11 \cdot 2, \dots, p}$ 는 서로 독립이다. 또한, 식 (1.4.86)에서 알 수 있듯이, 다음 식이 성립한다.

$$b_{11 \cdot 2, \dots, p} \stackrel{d}{\sim} \chi_{N-p}^2 \quad (1.4.87)$$

식 (1.4.76)에서 알 수 있듯이, 다음 식이 성립한다.

$$\mathbf{y}_*^T \mathbf{y}_* \stackrel{d}{\sim} \chi_p^2 (\boldsymbol{\nu}^T \Sigma^{-1} \boldsymbol{\nu}) \quad (1.4.88)$$

지금까지 내용을 정리하면 다음과 같다.

#### 정리 1.4.5

서로 독립인 확률벡터들  $\mathbf{y}, \mathbf{z}_2, \mathbf{z}_3, \dots, \mathbf{z}_N$  이 다음과 같은 확률분포들을 따른다고 하자.

$$\mathbf{y} \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\nu}, \Sigma), \quad \mathbf{z}_i \stackrel{d}{\sim} \mathcal{N}_p(\mathbf{0}, \Sigma), \quad (i = 2, 3, \dots, N)$$

확률행렬  $S$ 와 확률변수  $T^2$ 를 각각 다음과 같이 쓸 수 있다.

$$S \stackrel{d}{=} \frac{1}{N-1} \sum_{i=2}^N \mathbf{z}_i \mathbf{z}_i^T, \quad T^2 \stackrel{d}{=} \mathbf{y}^T S^{-1} \mathbf{y}$$

또한 다음 식이 성립한다.

$$\frac{N-p}{p[N-1]} T^2 \stackrel{d}{\sim} F_{p, N-p}(\boldsymbol{\nu}^T \Sigma^{-1} \boldsymbol{\nu})$$

#### 따름정리 1.4.3

서로 독립인 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  이 다음과 같은 확률분포들을 따른다고 하자.

$$\mathbf{x}_i \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\mu}, \Sigma), \quad (i = 1, 2, \dots, N)$$

이러한 조건 하에서 다음 식이 성립한다.

$$\frac{N-p}{p[N-1]} N [\bar{\mathbf{x}} - \boldsymbol{\mu}_0]^T S^{-1} [\bar{\mathbf{x}} - \boldsymbol{\mu}_0] \stackrel{d}{\sim} F_{p, N-p} \left( N [\boldsymbol{\mu} - \boldsymbol{\mu}_0]^T \Sigma^{-1} [\boldsymbol{\mu} - \boldsymbol{\mu}_0] \right)$$

따름정리 1.4.3의 조건 하에서 다음 식이 성립한다.

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N-1} \sum_{i=2}^N \mathbf{z}_i \mathbf{z}_i^T = \Sigma \quad (1.4.89)$$

정리 1.4.5와 식 (1.4.89)에서 알 수 있듯이, 다음 정리가 성립한다.

## 정리 1.4.6

서로 독립인 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots$  가 다음과 같은 확률분포들을 따른다고 하자.

$$\mathbf{x}_i \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\mu}, \Sigma), \quad (i = 1, 2, \dots)$$

확률벡터  $\bar{\mathbf{x}}_N$ , 확률행렬  $S_N$  그리고 통계량  $T_N^2$  를 다음과 같이 정의하자.

$$\begin{aligned}\bar{\mathbf{x}}_N &\doteq \sum_{i=1}^N \mathbf{x}_i \\ S_N &\doteq \frac{1}{N-1} \sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}_N][\mathbf{x}_i - \bar{\mathbf{x}}_N]^T \\ T_N^2 &\doteq N [\bar{\mathbf{x}}_N - \boldsymbol{\mu}_0]^T S_N^{-1} [\bar{\mathbf{x}}_N - \boldsymbol{\mu}_0]\end{aligned}$$

만약  $\boldsymbol{\mu} = \boldsymbol{\mu}_0$  이면, 다음 식이 성립한다.

$$T_N^2 \stackrel{d}{\rightarrow} \chi_p^2$$

**예제 1.4.3** 벡터  $\boldsymbol{\nu} \in \mathbb{R}^p$  와 가역행렬  $B \in \mathbb{R}^{p \times p}$  에 대해서 다음 식을 만족하는  $\lambda \neq 0$  가 존재한다.

$$\det(\boldsymbol{\nu}\boldsymbol{\nu}^T - \lambda B) = 0 \quad (1)$$

이 근을  $\lambda_1$  이라 하면, 다음 식을 만족하는 벡터  $\boldsymbol{\eta}_1 (\neq \mathbf{0})$  이 존재한다.

$$\boldsymbol{\nu}\boldsymbol{\nu}^T \boldsymbol{\beta}_1 = \lambda_1 B \boldsymbol{\beta}_1 \quad (2)$$

식 (2) 의 각 변의 좌측에  $\boldsymbol{\nu}^T B^{-1}$  를 곱하면, 다음 식들이 성립함을 알 수 있다.

$$\boldsymbol{\nu}^T B^{-1} \boldsymbol{\nu}\boldsymbol{\nu}^T \boldsymbol{\beta}_1 = \lambda_1 \boldsymbol{\nu}^T B^{-1} B \boldsymbol{\beta}_1 = \lambda_1 \boldsymbol{\nu}^T \boldsymbol{\beta}_1 \quad (3)$$

식  $\boldsymbol{\nu}^T \boldsymbol{\beta}_1 \neq \mathbf{0}$  에 의해서,  $\lambda_1 = \boldsymbol{\nu}^T B^{-1} \boldsymbol{\nu}$  임을 알 수 있다. 만약  $\boldsymbol{\nu} = \sqrt{N} [\bar{\mathbf{x}}_N - \boldsymbol{\mu}_0]$  라 하고  $B = S_N$  이라 놓으면, 다음 식이 성립한다.

$$\det(N [\bar{\mathbf{x}}_N - \boldsymbol{\mu}_0][\bar{\mathbf{x}}_N - \boldsymbol{\mu}_0]^T - T_N^2 S_N) = 0. \quad (4)$$



## 제 1.5 절 교환가능성

교환가능성 (exchangeability) 은 베이시안통계학에서 중요한 개념이다. 이 절에서는 교환가능성과 de Finetti 정리를 살펴보자. 교환가능성에 관한 자세한 내용은 Lauritzen [65]을 참조하라.

**예제 1.5.1** 어떤 항아리에  $b$  개 검은 공과  $w$  개 흰 공이 들어있다고 하자. 이 항아리에서 임의로 공 하나를 뽑은 다음, 뽑힌 공과 더불어 동일한 색깔을 갖는 공  $c$  개를 다시 항아리에 넣기로 하자. 제  $i$  번째 뽑기에서 검은 공이 나오면  $x_i = 1$  이라고 하고, 흰 공이 나오면  $x_i = 0$  라 하자. 다음 식들이 성립함을 알 수 있다.

$$\Pr(1, 1, 0, 1) = \frac{b}{b+w} \cdot \frac{b+c}{b+w+c} \cdot \frac{w}{b+w+2c} \cdot \frac{b+2c}{b+w+3c} \quad (1)$$

$$\Pr(1, 0, 1, 1) = \frac{b}{b+w} \cdot \frac{w}{b+w+c} \cdot \frac{b+c}{b+w+2c} \cdot \frac{b+2c}{b+w+3c} \quad (2)$$

식 (1) 과 식 (2) 에서 알 수 있듯이, 다음 식이 성립한다.

$$\Pr(1, 1, 0, 1) = \Pr(1, 0, 1, 1) \quad (3)$$

같은 방법으로 다음 식들이 성립함을 증명할 수 있다.

$$\Pr(0, 1, 1, 1) = \Pr(1, 0, 1, 1) = \Pr(1, 1, 0, 1) = \Pr(1, 1, 1, 0) \quad (4)$$

즉, 다음 식이 성립한다.

$$\Pr(x_1, x_2, x_3, x_4) = \Pr(x_{\pi(1)}, x_{\pi(2)}, x_{\pi(3)}, x_{\pi(4)}) \quad (5)$$

여기서  $[\pi(1), \pi(2), \pi(3), \pi(4)]$  는  $[1, 2, 3, 4]$  의 순열 (permutation) 이다. 식 (5) 가 만족되면, 확률변수열  $x_1, x_2, x_3, x_4$  가 교환가능 (exchangeable) 이라고 한다. 확률변수열  $x_1, x_2, x_3, x_4$  는 식 (5) 를 만족하나, 이 확률변수열은 독립이 아니고 또한 Markov 체인도 아니다. ■

**정의 1.5.1: 교환가능성**

확률변수열  $x_1, x_2, \dots$ 가 각  $n(= 2, 3, \dots)$ 에 대해서 다음 식을 만족하면, 이 확률변수열은 교환가능하다고 한다.

$$[x_1, x_2, \dots, x_n] \stackrel{d}{=} [x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)}]$$

여기서  $[\pi(1), \pi(2), \dots, \pi(n)]$ 은  $[1, 2, \dots, n]$ 의 임의의 순열이다.

**예제 1.5.2** 만약 유한변수열  $x_1, x_2, \dots, x_n$ 이 교환가능하면, 확률벡터  $\mathbf{x}_n = [x_1, x_2, \dots, x_n]^T$ 의 평균벡터는  $\mu \mathbf{1}_n$ 이고 분산공분산행렬  $\Sigma_n$ 은 다음과 같은 Toeplitz 행렬로 표기할 수 있다.

$$\Sigma_n = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho & \rho \\ \rho & 1 & \rho & \cdots & \rho & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho & \rho & \rho & \cdots & 1 & \rho \\ \rho & \rho & \rho & \cdots & \rho & 1 \end{bmatrix} \quad (1)$$

여기서  $\mu$ 와  $\rho$ 는 상수들이다. 다음 식들이 성립한다.

$$n\text{Var}(X_1) + n[n-1]\text{Var}(X_1, X_2) = \text{Var}\left(\sum_i^n X_i\right) \geq 0 \quad (2)$$

식 (2)에서 알 수 있듯이, 다음 식이 성립한다.

$$\rho \geq -\frac{1}{n-1} \quad (3)$$

만약 무한변수열  $x_1, x_2, \dots$ 이 교환가능하면, 모든  $n(\geq 2)$ 에 대해서 식 (3)이 성립해야한다. 즉, 다음 식이 성립한다.

$$\rho \geq 0 \quad (4)$$

만약 각  $n$ 에 대해서 확률벡터  $\mathbf{x}_n$ 이 다변량정규분포를 따르면, 이 확률벡터의 결합확률분포는 평균벡터  $\mu \mathbf{1}_n$ 과 분산공분산행렬  $\Sigma_n$ 에 의해서 결정되므로, 조건 (4)가 만족되면 확률변수열  $x_1, x_2, \dots$ 은 교환가능하다. ■

정의 1.5.1에서 알 수 있듯이, 확률변수열  $x_1, x_2, \dots$  가 서로 독립이면 교환가능하다. 그러나, 예제 1.5.1에서 알 수 있듯이, 그 역은 성립하지 않는다.

모수  $\theta$ 가 주어진 조건 하에서  $x_1, x_2, \dots$  이 서로 독립이면,  $x_1, x_2, \dots$  가 조건부독립 (conditional independence)이라고 한다. 조건부독립성과 교환가능성은 밀접한 관계가 있다.

**예제 1.5.3** 제  $i$  번째 사람이 미국 공화당을 좋아하면  $x_i = 1$  이라고 하고, 그렇지 않으면  $x_i = 0$  라 하자. 또한, 다음 식들이 성립한다고 가정하자.

$$\Pr(x_i = 1) = \theta = \Pr(x_i = 0) \quad (1)$$

고전통계학자, 즉 도수주의자의 관점에서는 모수  $\theta$ 가 미지인 상수(unknown constant)라고 가정하므로, 확률변수열  $x_1, x_2, \dots$  이 서로 독립이라고 가정하는 데는 별 문제가 없다. 그러나, 베이시안통계학자 관점에서 보면 이러한 독립성가정은 혼란을 일으킬 수도 있다. 모수  $\theta$ 가 주어진 조건 하에서  $x_1, x_2, \dots$  이 서로 독립이면, 즉 조건부독립이면, 다음 식이 성립한다.

$$\Pr(x_i = a_i | \theta, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \theta^{a_i} [1 - \theta]^{1-a_i} \quad (2)$$

따라서, 모수  $\theta$ 가 주어진 조건 하에서 결합밀도함수는 다음과 같다.

$$\begin{aligned} \Pr(x_1 = a_1, x_2 = a_2, \dots, x_n = a_n | \theta) \\ = \prod_{i=1}^n \theta^{a_i} [1 - \theta]^{1-a_i} = \theta^{\sum_{i=1}^n a_i} [1 - \theta]^{n - \sum_{i=1}^n a_i} \end{aligned} \quad (3)$$

식 (3)에서 알 수 있듯이, 비조건부 결합밀도함수는 다음과 같다.

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= \int_0^1 \Pr(x_1, x_2, \dots, x_n | \theta) \pi(\theta) d\theta \\ &= \int_0^1 \theta^{\sum_{i=1}^n x_i} [1 - \theta]^{n - \sum_{i=1}^n x_i} \pi(\theta) d\theta \end{aligned} \quad (4)$$

식 (4)에서 알 수 있듯이, 비조건부 확률밀도함수  $f(x_1, x_2, \dots, x_n)$ 는  $\sum_{i=1}^n x_i$ 에 의존하나 확률변수열  $x_1, x_2, \dots, x_n$ 의 순서에 의존하지 않는다. 즉, 이 확률변수열  $x_1, x_2, \dots$ 은 교환가능하다. 결론적으로, 확률변수열  $x_1, x_2, \dots$ 가 모수  $\theta$ 가 주어진 조건 하에서 독립이며 동일한 확률분포를 따르면, 이 확률변수열은 교환가능하다.

모수  $\theta$ 의 사전확률분포가 지지대  $(0, 1)$ 인 일양확률분포라고 하자. 즉, 식  $\pi(\theta) = 1_{(0,1)}(\theta)$ 가 성립한다고 하자. 식 (4)에서 알 수 있듯이, 다음 식들이 성립한다.

$$f(x_1, x_2, \dots, x_n) = \int_0^1 \theta^{\sum_{i=1}^n x_i} [1 - \theta]^{n - \sum_{i=1}^n x_i} d\theta = B\left(\sum_{i=1}^n x_i + 1, n + 1 - \sum_{i=1}^n x_i\right) \quad (5)$$

식 (5)에 식  $n = 1$ 을 대입하면, 다음 식을 얻는다.

$$f(x_1) = B(x_1 + 1, 2 - x_1) \quad (6)$$

식 (5)와 식 (6)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= B\left(\sum_{i=1}^n x_i + 1, n + 1 - \sum_{i=1}^n x_i\right) \\ &\neq \prod_{i=1}^n B(x_i + 1, 2 - x_i) = \prod_{i=1}^n f(x_i) \end{aligned} \quad (7)$$

즉, 확률변수열  $x_1, x_2, \dots$ 는 모수  $\theta$ 가 주어진 조건 하에서 독립이나 비조건부로 독립이 아니다. ■

예제 1.5.3에서 알 수 있듯이, 확률변수열  $x_1, x_2, \dots$ 가 모수  $\theta$ 가 주어진 조건 하에서 독립이며 동일한 확률분포를 따르면 이 확률변수열은 교환가능하다. 그 역은 성립하는 것일까? De Finetti [29]의 다음 정리가 그 답을 제시하고 있다.

**정리 1.5.1: De Finetti 정리**

Bernouille 확률변수열  $x_1, x_2, \dots$ 가 교환가능하기 위한 필요충분조건은 모든 자연수  $n$ 에 대해서 결합확률질량함수  $p(x_1, x_2, \dots, x_n)$ 이 다음 식을 만족하는 확률분포함수  $F$ 가 존재하는 것이다.

$$p(x_1, x_2, \dots, x_n) = \int_0^1 \theta^{\sum_{i=1}^n x_i} [1 - \theta]^{n - \sum_{i=1}^n x_i} dF(\theta)$$

정리 1.5.1에 관해서는 Diaconis & Freedman [31], [32]을 참조하라. 정리 1.5.1로부터

다음 성질들이 성립함을 알 수 있다.

$$F(y) = \Pr \left( \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i \leq y \right) \tag{1.5.1}$$

$$\Pr \left( x_1 = a_1, x_2 = a_2, \dots, x_n = a_n \mid \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i = \theta \right) = \theta^{\sum_{i=1}^n a_i} [1 - \theta]^{n - \sum_{i=1}^n a_i} \tag{1.5.2}$$

식 (1.5.2)에서 알 수 있듯이, 확률변수  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i = \theta$ 가 주어진 조건 하에서 확률변수열  $x_1, x_2, \dots$ 은 독립이며 또한 모수가  $\theta$ 인 Bernoulli 확률분포를 따른다.

**예제 1.5.4** 예제 1.5.1을 다시 생각해보자. Black & Kendall [18]은 다음 식을 유도하였다.

$$\Pr \left( \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i \leq y \right) = \int_0^y \frac{1}{B\left(\frac{b}{c}, \frac{w}{c}\right)} x^{b/c-1} [1-x]^{w/c-1} dx \tag{1}$$

즉, 극한확률변수  $y \doteq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i$ 는 베타확률분포  $Beta\left(\frac{b}{c}, \frac{w}{c}\right)$ 를 따른다. De Finetti 정리와 식 (1.5.2)에서 알 수 있듯이, 확률변수  $y = \theta$ 가 주어진 조건 하에서 확률변수열  $x_1, x_2, \dots$ 은 독립이고, 모수가  $\theta$ 인 Bernoulli 확률분포를 따른다. ■

Hewitt & Savage [54]는 de Finetti 정리를 다음과 같이 확장하였다.

**정리 1.5.2: Hewitt-Savage 정리**

확률변수열  $x_1, x_2, \dots$ 가 교환가능하면, 모든 자연수  $n$ 에 대해서 다음 식을 만족하는 확률측도  $\mu$ 가 존재한다.

$$\Pr(x_1 \in A_1, x_2 \in A_2, \dots, x_n \in A_n) = \int Q(A_1) Q(A_2) \cdots Q(A_n) \mu(dQ)$$

정리 1.5.2로부터 다음 식들이 성립함을 알 수 있다.

$$\mu(A) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n 1_A(x_i) \tag{1.5.3}$$

$$\Pr(x_1 \in A_1, x_2 \in A_2, \dots, x_n \in A_n \mid \mu = Q) = Q(A_1) Q(A_2) \cdots Q(A_n) \tag{1.5.4}$$

식 (1.5.4)에서 알 수 있듯이, 확률측도  $\mu = Q$ 가 주어진 조건 하에서 확률변수열  $x_1, x_2, \dots$ 은 독립이고 확률측도  $Q$ 를 따른다.



## 제 2 장

# 베이지안통계학의 기초

### 제 2.1 절 리스크분석과 베이지안통계학

베이지안통계이론을 체계적으로 다루기 전에 베이지안통계기법을 적용하는 첫 번째 예제로서 조건부리스크가치 (conditional value at risk; CVaR) 를 추정하는 문제를 살펴보자. 베이지안통계학의 관점에서 보면, 이 절에서 다루는 문제는 수리적으로 매우 간단한 것이다.

#### 2.1.1 조건부리스크가치 CVaR

재무를 담당하고 있는 실무자는 항상 직면하는 다양한 리스크 (risk) 에 대해 대응해야한다. 실무자가 고려하는 리스크중에서 중요한 것으로는 자산가격의 예측불가능한 변동에 의한 시장리스크 (market risk), 거래상대방 (counterparty) 의 예측불가능한 채무불이행에 의한 신용리스크 (credit risk), 온라인 결제시스템의 오작동 같은 예측불가능한 업무상 오차에 의한 오퍼레이셔널리스크 (operational risk) 등이 있다.

리스크관리에서 중요한 점은 어떤 정량적 지표 (measure) 를 사용해서 이론적으로 리스크를 정의하고 실제로 리스크를 계측하는 것이다. 리스크사상 (risk event)  $A$  가 발생하면 손실액이  $x$  이고 그 사건이 발생할 확률이  $\Pr(A)$  이면, 정량적 리스크계측을 위한 리스크지표 (risk measure 또는 risk metric)  $R_M$  을 다음과 같이 정의한다.

$$R_M \doteq x\Pr(A) \tag{2.1.1}$$

리스크사상이 발생하였을 때 손실액은 예측불가능한 경우가 대부분이다. 예를 들어, 시장에서 주가가 폭락하는 경우 구체적으로 주가가 얼마만큼 하락할지 예측하는 것은 거의 불가능하다.

또한, 채무기업이 디폴트하였을 때 회수가 가능한 채권액이나, 시스템 오작동에 의한 손실액을 확실하게 예측하는 것도 불가능하다. 따라서, 리스크관리에서는 리스크사상이 발생할 확률뿐만 아니라 예상되는 손실액도 확률적으로 다루어야 한다. 리스크사상의 발생확률과 예상되는 손실액을 평가하기 위해서는, 리스크사상 발생과 손실액 변동을 나타내는 특정 시스템을 상정(想定)할 필요가 있다. 이러한 시스템을 리스크모형이라 부른다. 즉, 리스크모형은 리스크사상의 발생과 손실액을 확률변수들로 간주하고 그 확률분포들을 기술한다. 리스크모형 구축을 위한 이론적인 틀과 통계적 방법이 활발하게 연구되고 있으며, 반복되는 경제위기 그리고 이에 따른 BIS규제 강화 등에 의해서 실무에서도 리스크모형을 바탕으로 하는 리스크관리의 중요성이 강조되고 있다. 정량적인 리스크관리 방법에 관해서는 McNeil & Frey & Embrechts [74]를 참조하라.

채무기업의 부도가능성에 관해서는 식 (2.1.1) 처럼 한 종류 리스크사상으로 충분히 리스크지표를 나타낼 수 있다. 그러나, 오퍼레이셔널리스크는 업무상 과오의 종류에 따라 손실액이 변하므로 여러 리스크사상들을 상정해야한다. 리스크사상들  $A_1, A_2, \dots$  가 출현가능하다고 하고, 제  $n$  번째 리스크사상  $A_n$  이 일어날 확률을  $\Pr(A_n)$  그리고 리스크사상  $A_n$  이 발생했을 때 손실액을  $x_n$  이라 하자. 즉, 손실액이 이산형 확률변수라고 하자. 이 리스크사상들을 고려하는 리스크지표는 다음과 같다.

$$R_M \doteq \sum_{n=1}^{\infty} x_n \Pr(A_n) \quad (2.1.2)$$

실제로는 단순히 손실액 그 자체를 쓰는 것이 아니라 손실액을 변환하는 (transforming) 경우도 많다. 예를 들어, 은행에 보통예금을 했을 때 은행이 디폴트되면 최고 5천만원까지 국가가 보상을 책임진다. 함수  $L(x)$  가 손실액  $x$  를 변환하는 규칙이라고 하자. 이  $L(x)$  를 손실함수 (loss function) 라고 부른다. 이 경우에 리스크지표는 다음과 같다.

$$R_M \doteq \sum_{n=1}^{\infty} L(x_n) \Pr(A_n) \quad (2.1.3)$$

만약 손실액이 확률밀도함수  $f(x)$  를 갖는 연속형 확률변수이면, 리스크지표를 다음과 같이 쓸 수 있다.

$$R_M \doteq \int_{-\infty}^{\infty} L(x) f(x) dx \quad (2.1.4)$$

식 (2.1.3)과 식 (2.1.4)를 다음과 같이 쓸 수 있다.

$$R_M = E(L(x)) \tag{2.1.5}$$

통계분석이나 계량경제분석 관점에서 보면, 리스크모형의 주된 실체는 특정된 확률분포이다. 즉, 식 (2.1.3)의 리스크지표에서 손실액의 모형은 확률질량함수  $\{\Pr(A_n) \mid n = 1, 2, \dots\}$  이고, 식 (2.1.4)에서 손실액의 모형은 확률밀도함수  $f(x)$ 이다. 리스크모형을 바탕으로 하는 리스크평가에서는 손실액의 변동을 적절하게 설명할 수 있는 확률분포함수를 선택하는 것이 가장 중요한 작업이다. 본서에서는 손실액이 연속형 확률변수인 경우만을 다루기로 하자.

널리 사용되는 리스크지표는 다음과 같다. 리스크지표에 관한 자세한 내용은 Ruiz [89]를 참조하라.

**정의 2.1.1: 리스크지표**

a) Current Exposure (CE):  $CE \doteq \max\{x, 0\}$

b) Expected Positive Exposure (EPE):  $EPE \doteq \int_{-\infty}^{\infty} \max\{x, 0\}f(x)dx$

c) Expected Negative Exposure (ENE):  $EPE \doteq \int_{-\infty}^{\infty} \min\{x, 0\}f(x)dx$

d) Potential Future Exposure (PFE):  $\int_{-\infty}^{PFE_{\beta}} f(x)dx = \beta$

e) Credit Expected Short Fall (CESF):  $CESF \doteq \int_{PFE_{\beta}}^{\infty} xf(x)dx$

f) Conditional Value at Risk (CVaR):  $CVaR \doteq CESF/[1 - \beta]$

정의에서 알 수 있듯이,  $PFE_{\beta}$ 는 손실액의 누적확률이  $\beta$ 가 되는 지점의 손실액이다. 즉, 다음 식이 성립한다.

$$\Pr(x \leq PFE_{\beta}) = \beta \tag{2.1.6}$$

즉,  $PFE_{\beta}$ 는 유명한 리스크지표인 VaR(value at risk)이다. 다음 함수를 정의하자.

$$f_{\beta}(x) \doteq \frac{1}{1 - \beta}f(x)1_{[PFE_{\beta}, \infty)}(x) \tag{2.1.7}$$

다음 식들이 성립한다.

$$\int_{-\infty}^{\infty} f_{\beta}(x)dx = \int_{PFE_{\beta}}^{\infty} \frac{1}{1-\beta} f(x)dx = \frac{1}{1-\beta} \int_{PFE_{\beta}}^{\infty} f(x)dx = \frac{1-\beta}{1-\beta} = 1 \quad (2.1.8)$$

즉,  $f_{\beta}(x)$ 는 절단확률밀도함수(truncated probability density function)이다. 따라서, 다음 식들이 성립한다.

$$CVaR = \int_{-\infty}^{\infty} x \frac{1}{1-\beta} f(x) 1_{[PFE_{\beta}, \infty)}(x) dx = \frac{1}{1-\beta} \int_{PFE_{\beta}}^{\infty} x f(x) dx \quad (2.1.9)$$

식 (2.1.9)에서 알 수 있듯이, CVaR의 손실함수는 다음과 같다.

$$L(x) = \frac{x}{1-\beta} 1_{[PFE_{\beta}, \infty)}(x) \quad (2.1.10)$$

식 (2.1.9)를 다음과 같이 쓸 수 있다.

$$CVaR = E(x \mid x \geq PFE_{\beta}) \quad (2.1.11)$$

즉, CVaR는 손실액이 확률분포의  $100\beta\%$ 를 상회하는 조건 하에서 손실액의 기대값이다. 확률  $\beta$ 를 강조하는 의미로  $\beta$ -CVaR라고 표기하기도 한다.

일반적으로  $\beta$ 로는 0.99 또는 0.999와 같이 1에 가까운 값이 사용된다. 이 경우에  $PFE_{\beta}$ 가 손실액분포의 오른쪽에 위치하게 되고, 이  $PFE_{\beta}$  값을 넘는 커다란 손실을 입는 확률  $1-\beta$ 는 작은 값이 된다. 따라서, 손실액이  $PFE_{\beta}$  값을 넘는 경우는 거의 발생하지 않지만, 그러한 일이 발생하면 큰 손실을 가져오는 리스크사상을 CVaR로 나타낼 수 있다. 즉, CVaR는 잘 발생하지는 않지만 일단 발생하면 커다란 손실을 입게되는 상황에서 예상되는 손실액이다.

### 2.1.2 CVaR의 평가

식 (2.1.11)에 기술한 CVaR를 이용하기 위해서는 손실액의 확률밀도함수  $f(x)$ 를 정해야한다. 앞에서 언급했듯이, 확률밀도함수  $f(x)$ 는 손실액모형의 실제이다. 확률밀도함수  $f(x)$ 를 선택하는 것은 어렵기 때문에, 보통 손실액의 변동을 설명할 수 있을 것으로 보이는 확률밀도함수  $f(x)$ 의 후보를 좁히는 일부부터 시작한다. 이 소절에서는 손실액  $x$ 가 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 를 따른다고 가정하자. 일반적으로 신용리스크나 오퍼레이셔널리스크 평가에서 손실액은

비음이라고 가정한다. 따라서, 지지대가 음수를 포함하는 정규확률분포는 손실액의 확률분포로서 적당한 것이 아니다. 이러한 경우에 손실액이 대수정규확률분포를 따른다고 하는 것이 더 타당할 것이다. 그러나, 여기서는 설명을 간단히 하기 위해서 손실액  $x$ 가 정규확률분포를 따른다고 가정하자. 식 (2.1.9)의  $f(x)$ 를 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 의 확률밀도함수라고 하면,  $f_\beta(x)$ 는 절단정규확률분포(truncated normal distribution)의 확률밀도함수이다. 따라서, 식 (2.1.11)의 CVaR는 절단정규확률분포의 평균과 같으므로, 다음 식이 성립함을 알 수 있다.

$$CVaR = \mu + \sigma \frac{n\left(\frac{PFE_\beta - \mu}{\sigma}\right)}{1 - N\left(\frac{PFE_\beta - \mu}{\sigma}\right)} \tag{2.1.12}$$

여기서  $n(\cdot)$ 와  $N(\cdot)$ 는 각각 표준정규확률변수의 확률밀도함수와 확률분포함수이다. 식 (2.1.6)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\beta = \Pr(x \leq PFE_\beta) = N\left(\frac{PFE_\beta - \mu}{\sigma}\right) \tag{2.1.13}$$

즉, 다음 식이 성립한다.

$$PFE_\beta = \mu + \sigma N^{-1}(\beta) \tag{2.1.14}$$

이  $PFE_\beta$ 가 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 하에서 VaR이다. 식 (2.1.14)를 식 (2.1.12)에 대입하면, 다음 식을 얻는다.

$$CVaR = \mu + \sigma \frac{n(N^{-1}(\beta))}{1 - \beta} \tag{2.1.15}$$

**예제 2.1.1** 손실액  $x$ 가 정규확률분포를 따른다는 가정 하에서 CVaR와 VaR를 그리기 위해, 다음 MATLAB 프로그램 CVaR101.m을 실행해 보자.

```

1 % -----
2 % Filename: CVaR101.m
3 % CVar and VaR
4 % Programmed by CBS
5 % -----
6 clear all, clf, close all
7 beta = 0.8999:0.0001:0.9999;
8 mu = 10;
9 sigma = 3;
10 invN = norminv(beta,0,1);
11 VaR = mu + sigma*invN;
12 CVaR = mu + sigma./(1-beta).*normpdf(invN);
13 % Plotting
14 subplot(2,1,1)
15 plot(beta,CVaR,'k-', 'LineWidth',1.5)
16 set(gca, 'fontsize',11, 'fontweigh', 'bold')
    
```

```

17 xlabel('\bf \beta', 'FontSize', 12');
18 ylabel('\bf CVaR', 'FontSize', 12');
19 subplot(2,1,2)
20 plot(beta, VaR, 'r-', 'LineWidth', 1.5)
21 set(gca, 'fontSize', 11, 'fontweight', 'bold')
22 xlabel('\bf \beta', 'FontSize', 12');
23 ylabel('\bf VaR', 'FontSize', 12');
24 saveas(gcf, 'CVaR101', 'eps')
25 save('CVaR101', 'beta', 'VaR', 'CVaR')
26 % End of program
27 % -----

```

이 MATLAB 프로그램 CVaR101.m을 실행하면, 손실액이 정규확률분포  $\mathcal{N}(10, 3^2)$ 를 따른다는 가정 하에  $\beta$ 가 0.8999에서 0.9999까지 변화할 리스크지표들 CVaR와 VaR의 그래프들을 그린다. 이 그래프들이 그림 2.1.1에 그려져 있다. 그림 2.1.1의 상단 그래프에서 흑색 실선은 CVaR를 그린 것이고, 하단 그래프의 적색 실선은 VaR를 그린 것이다. ■

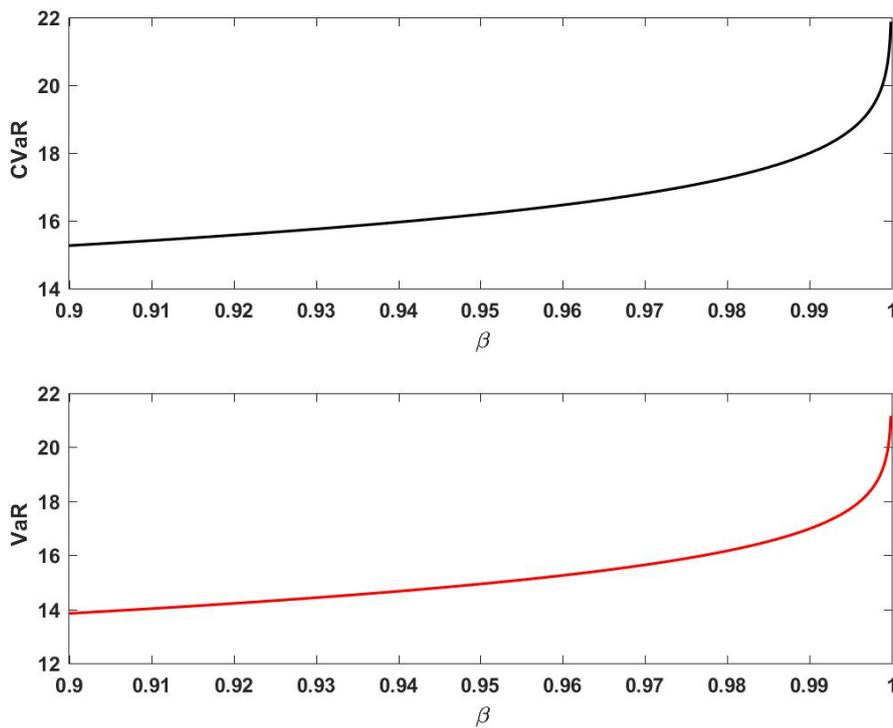


그림 2.1.1. CVaR와 VaR

CVaR를 사용해서 리스크를 계측하기 위해서는 손실액 확률분포인 정규확률분포의 모수벡터  $\theta \doteq [\mu, \sigma^2]$ 가 필요하다. 첫째로,  $\theta_1 \doteq [\mu_1, \sigma_1^2]$ 와  $\theta_2 \doteq [\mu_2, \sigma_2^2]$ 가 선택가능한

모수벡터라고 하자. 또한, 다음 식들이 성립한다고 가정하자.

$$\mu_1 > \mu_2, \quad \sigma_1^2 > \sigma_2^2 \tag{2.1.16}$$

식 (2.1.12)와 식 (2.1.16)에서 알 수 있듯이, CVaR로 리스크를 측정하면 손실액의 확률분포로는  $\mathcal{N}(\mu_2, \sigma_2^2)$ 보다  $\mathcal{N}(\mu_1, \sigma_1^2)$ 이 더 위험을 잘 나타낸다. 현실적으로 어떤 손실액 확률분포가 맞는지에 관한 확증은 없지만, 각 확률분포가 진짜 확률분포일 가능성이 다음과 같다고 가정하자.

$$\Pr(\boldsymbol{\theta} = \boldsymbol{\theta}_1) = \pi = 1 - \Pr(\boldsymbol{\theta} = \boldsymbol{\theta}_2) \tag{2.1.17}$$

즉, 확률  $\pi$ 의 크기로 확률분포에 대한 신용도가 결정된다고 하자. 베이저안통계학적 관점에서 말한다면, 식 (2.1.17)의  $\pi$ 는 모수벡터  $\boldsymbol{\theta}$ 의 사전확률이다. 만약  $\pi$ 가 0 또는 1에 가깝다면, CVaR를 계산하는데 두 확률분포들 중에서 신용도가 높은 확률분포를 사용한다. 그러나, 만약  $\pi$ 가 0.5에 가깝다면, 두 확률분포들 중 하나만 사용하는 것은 타당하지가 않다. 식 (2.1.5)에서 알 수 있듯이, 손실함수의 기대값으로 리스크지표를 정의한다. 즉, 리스크사상의 발생 확률 뿐 아니라 손실액이 불확실하기 때문에, 평균적인 손실함수를 리스크지표로 사용한다. 그러나, 손실액 확률분포의 모수벡터를 알지 못하는 경우에는 손실액에 관한 불확실성 뿐 아니라 손실액의 확률분포에 관한 불확실성도 존재한다. 따라서, 다음과 같이 확률분포에 대한 신용도를 가중값으로 하는 리스크지표의 기대값을 구하여, 손실액의 확률분포에 관한 불확실성을 리스크지표 CVaR에 반영시킨다.

$$CVaR = \pi CVaR(\boldsymbol{\theta}_1) + [1 - \pi] CVaR(\boldsymbol{\theta}_2) \tag{2.1.18}$$

여기서  $CVaR(\boldsymbol{\theta})$ 는 모수벡터가  $\boldsymbol{\theta}$ 인 경우 CVaR이다. 손실액이 정규확률분포를 따르는 경우 CVaR를 살펴보자. 먼저 선택가능한 모수벡터들이  $\boldsymbol{\theta}_1 \doteq [\mu_1, \sigma_1^2]$ 과  $\boldsymbol{\theta}_2 \doteq [\mu_2, \sigma_2^2]$ 인 경우, 식 (2.1.15)를 식 (2.1.18)에 대입하면 다음 식을 얻는다.

$$CVaR = \pi \left[ \mu_1 + \sigma_1 \frac{n(N^{-1}(\beta))}{1 - \beta} \right] + [1 - \pi] \left[ \mu_2 + \sigma_2 \frac{n(N^{-1}(\beta))}{1 - \beta} \right] \tag{2.1.19}$$

둘째로, 선택가능한 모수벡터들이  $\boldsymbol{\theta}_1 \doteq [\mu_1, \sigma_1^2], \boldsymbol{\theta}_2 \doteq [\mu_2, \sigma_2^2], \dots, \boldsymbol{\theta}_N \doteq [\mu_N, \sigma_N^2]$ 이고, 각 모수벡터가 실현될 확률이 다음과 같다고 하자.

$$\Pr(\boldsymbol{\theta} = \boldsymbol{\theta}_n) = \pi_n, \quad (n = 1, 2, \dots, N) \tag{2.1.20}$$

베이시안통계학 관점에서 말한다면, 식 (2.1.20)의  $\{\pi_n\}$ 은 모수벡터  $\theta$ 의 사전확률밀도함수이다. 이 경우 리스크지표 CVaR는 다음과 같다.

$$CVaR = \sum_{j=1}^N \pi_j \left[ \mu_j + \sigma_j \frac{n(N^{-1}(\beta))}{1-\beta} \right] \quad (2.1.21)$$

셋째로, 선택가능한 모수벡터들의 집합이  $\{\theta \doteq [\mu, \sigma^2] \mid \mu \in R, \sigma^2 > 0\}$ 이고, 각 모수벡터가 실현될 신용도, 즉 확률밀도함수  $\pi(\theta)$ 가 다음과 같다고 하자.

$$\pi(\theta) = \pi(\mu, \sigma^2) \quad (2.1.22)$$

베이시안통계학 관점에서 말한다면, 식 (2.1.22)의  $\pi(\theta)$ 는 모수벡터  $\theta = [\mu, \sigma^2]$ 의 사전확률밀도함수이다. 이 경우에 리스크지표 CVaR는 다음과 같다.

$$CVaR = \int_{-\infty}^{\infty} \int_0^{\infty} \left[ \mu + \sigma \frac{n(N^{-1}(\beta))}{1-\beta} \right] \pi(\mu, \sigma^2) d\sigma^2 d\mu \quad (2.1.23)$$

식 (2.1.21)과 식 (2.1.23)에서 알 수 있듯이, 정규성가정 하에서 다음 식이 성립한다.

$$CVaR = E(\mu) + E(\sigma) \frac{n(N^{-1}(\beta))}{1-\beta} \quad (2.1.24)$$

즉, 식 (2.1.15)에서  $\mu$ 와  $\sigma$  대신 각각  $E(\mu)$ 와  $E(\sigma)$ 를 대입하면, 식 (2.1.24)를 얻는다.

지금까지 다룬 내용을 일반화하기로 하자. 만약 손실액의 확률밀도함수가  $f(x|\theta) = f(x|\mu, \sigma^2)$ 이고 모수벡터들  $\{\theta = [\mu, \sigma^2] \mid \mu \in R, \sigma^2 > 0\}$ 의 실현될 신용도, 즉 확률밀도함수가  $\pi(\theta) = \pi(\mu, \sigma^2)$ 라고 하자. 이 경우에 리스크지표 CVaR를 다음과 같이 쓸 수 있다.

$$CVaR = \int_{-\infty}^{\infty} \int_0^{\infty} \left[ \int_{-\infty}^{\infty} x \frac{1}{1-\beta} f(x|\mu, \sigma^2) 1_{[PFE_{\beta, \infty})}(x) dx \right] \pi(\mu, \sigma^2) d\sigma^2 d\mu \quad (2.1.25)$$

다음과 같이  $PFE_{\beta}(\mu, \sigma^2)$ 를 정의하자.

$$\int_{-\infty}^{PFE_{\beta}(\mu, \sigma^2)} f(x|\mu, \sigma^2) dx = \beta \quad (2.1.26)$$

손실액이  $PFE_\beta(\mu, \sigma^2)$  를 초과하는 손실액의 조건부확률밀도함수는 다음과 같다.

$$f_\beta(x|x > PFE_\beta(\mu, \sigma^2), \mu, \sigma^2) = \frac{1}{1-\beta} f(x|\mu, \sigma^2) 1_{[PFE_\beta, \infty)}(x) \quad (2.1.27)$$

다음 함수를 정의하자.

$$f_\beta(x|x > PFE_\beta) \doteq \int_{-\infty}^{\infty} \int_0^{\infty} f_\beta(x|x > PFE_\beta(\mu, \sigma^2), \mu, \sigma^2) \pi(\mu, \sigma^2) d\sigma^2 d\mu \quad (2.1.28)$$

식 (2.1.28) 의  $f_\beta(x|x > PFE_\beta)$  는  $PFE_\beta(\mu, \sigma^2)$  를 넘는 손실액을 모수벡터의 신용도  $\pi(\mu, \sigma^2)$  를 가중치로 하는 가중평균으로 해석할 수 있다. 베이지안통계분석에서는 모수벡터의 신용도로 확률분포들을 가중평균한 것을 예측분포(predictive distribution) 라고 부른다. 예측분포는 모형의 불확실성을 반영한 확률분포로서 예측이나 불확실성 하에서 의사결정을 수행할 때 사용된다. 식 (2.1.28) 을 식 (2.1.25) 에 대입하면, 다음 식을 얻는다.

$$CVaR = \int_{-\infty}^{\infty} x f_\beta(x|x > PFE_\beta) dx \quad (2.1.29)$$

식 (2.1.29) 에서 알 수 있듯이, CVaR는 100β%가 넘는 손실액의 예측분포  $f_\beta(x|x > PFE_\beta)$  의 기대값으로 해석할 수 있다.

**예제 2.1.2** 식 (2.1.24) 의 CVaR를 구해보자. 이 예제를 읽는데 수리적으로 어려움을 느끼는 독자는 제2.1절을 숙독한 다음 다시 이 예제를 생각해볼 것을 권한다.

사전확률분포가 다음과 같다고 하자.

$$\mu|\sigma^2 \stackrel{d}{\sim} \mathcal{N}\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right), \quad \sigma^2 \stackrel{d}{\sim} \text{IG}\left(\frac{\nu_0}{2}, \frac{\nu_0 S_0}{2}\right) \quad (1)$$

다음 식들이 성립한다.

$$\pi(\mu|\sigma^2) = \sqrt{\frac{\kappa_0}{2\pi\sigma^2}} \exp\left(-\frac{\kappa_0}{2\sigma^2} [\mu - \mu_0]^2\right) \quad (2)$$

$$\pi(\sigma^2) = \frac{1}{\Gamma\left(\frac{\nu_0}{2}\right)} \left[\frac{\nu_0 S_0}{2}\right]^{\nu_0/2} [\sigma^2]^{-[\nu_0/2+1]} \exp\left(-\frac{\nu_0 S_0}{2\sigma^2}\right) \quad (3)$$

$$\pi(\mu, \sigma^2) = \pi(\mu|\sigma^2) \pi(\sigma^2) \quad (4)$$

모수벡터  $[\mu_0, \kappa_0, \nu_0, \nu_0 S_0]$  는 사전확률분포의 형상을 정하는 모수벡터로서 하이퍼모수벡터

(hyper-parameter vector)라고 부른다. 식 (3)에서 알 수 있듯이,  $\sigma^2$ 의 사전확률분포의 평균과 분산은 각각 다음과 같다.

$$E(\sigma^2) = \frac{\nu_0 S_0}{\nu_0 - 2} \quad (5)$$

$$Var(\sigma^2) = \frac{2\nu_0^2 S_0^2}{[\nu_0 - 2]^2 [\nu_0 - 4]} \quad (6)$$

관찰된 손실액들을  $x_1, x_2, \dots, x_n$ 이라 하고, 이 관찰값들의 관찰벡터를  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ 이라 하자. 이 관찰값들은 서로 독립이며 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 를 따른다고 하자. 우도함수는 다음과 같다.

$$L(\mu, \sigma^2 | \mathbf{x}) = [2\pi\sigma^2]^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n [x_j - \mu]^2\right) \quad (7)$$

다음 식이 성립함을 쉽게 알 수 있다.

$$\sum_{j=1}^n [x_j - \mu]^2 = \sum_{j=1}^n [x_j - \bar{x}]^2 + n[\bar{x} - \mu]^2 \quad (8)$$

식 (8)을 식 (7)에 대입하면, 우도함수와 다음과 같음을 알 수 있다.

$$L(\mu, \sigma^2 | \mathbf{x}) = [2\pi\sigma^2]^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n [x_j - \bar{x}]^2\right) \exp\left(-\frac{n}{2\sigma^2} [\mu - \bar{x}]^2\right) \quad (9)$$

식 (1)과 식 (9)를 Bayes정리에 적용하면, 사후확률밀도함수가 다음과 같음을 알 수 있다.

$$\begin{aligned} & \pi(\mu, \sigma^2 | \mathbf{x}) \\ & \propto L(\mu, \sigma^2 | \mathbf{x}) \pi(\mu, \sigma^2) \\ & \propto [2\pi\sigma^2]^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n [x_j - \bar{x}]^2\right) \exp\left(-\frac{n}{2\sigma^2} [\mu - \bar{x}]^2\right) \\ & \quad \cdot \sqrt{\frac{\kappa_0}{2\pi\sigma^2}} \exp\left(-\frac{\kappa_0}{2\sigma^2} [\mu - \mu_0]^2\right) \frac{1}{\Gamma\left(\frac{\nu_0}{2}\right)} \left[\frac{\nu_0 S_0}{2}\right]^{\nu_0/2} [\sigma^2]^{-[\nu_0/2+1]} \exp\left(-\frac{\nu_0 S_0}{2\sigma^2}\right) \quad (10) \end{aligned}$$

다음 식이 성립한다.

$$n[\mu - \bar{x}]^2 + \kappa_0[\mu - \mu_0]^2 = \kappa_*[\mu - \mu_*]^2 + \frac{n\kappa_0}{n + \kappa_0}[\bar{x} - \mu_0]^2 \quad (11)$$

여기서  $\kappa_*$  과  $\mu_*$  은 각각 다음과 같다.

$$\kappa_* \doteq \kappa_0 + n, \quad \mu_* \doteq \frac{\kappa_0 \mu_0 + n \bar{x}}{\kappa_0 + n} \quad (12)$$

식 (11)을 식 (10)에 대입하면, 사후확률밀도함수가 다음과 같음을 알 수 있다.

$$\pi(\mu, \sigma^2 | \mathbf{x}) \propto \frac{1}{\sqrt{\sigma^2}} \exp\left(-\frac{\kappa_*}{2\sigma^2} [\mu - \mu_*]^2\right) [\sigma^2]^{-[\nu_*/2+1]} \exp\left(-\frac{S_* \nu_*}{2\sigma^2}\right) \quad (13)$$

여기서  $\nu_*$  과  $S_*$  는 각각 다음과 같다.

$$\nu_* \doteq n + \nu_0, \quad S_* \doteq \frac{1}{\nu_*} \left\{ \sum_{j=1}^n [x_j - \bar{x}]^2 + \frac{n\kappa_0}{n + \kappa_0} [\bar{x} - \mu_0]^2 + \nu_0 S_0 \right\} \quad (14)$$

식 (13)에서 알 수 있듯이, 사후확률밀도함수  $\pi(\mu, \sigma^2 | \mathbf{x})$  를 다음과 같이 쓸 수 있다.

$$\pi(\mu | \sigma^2, \mathbf{x}) = \sqrt{\frac{\kappa_*}{2\pi\sigma^2}} \exp\left(-\frac{\kappa_*}{2\sigma^2} [\mu - \mu_*]^2\right) \quad (15)$$

$$\pi(\sigma^2 | \mathbf{x}) = \frac{1}{\Gamma(\frac{\nu_*}{2})} \left[\frac{S_* \nu_*}{2}\right]^{\nu_*/2} [\sigma^2]^{-[\nu_*/2+1]} \exp\left(-\frac{S_* \nu_*}{2\sigma^2}\right) \quad (16)$$

$$\pi(\mu, \sigma^2 | \mathbf{x}) = \pi(\mu | \sigma^2, \mathbf{x}) \pi(\sigma^2 | \mathbf{x}) \quad (17)$$

즉, 다음 식들이 성립한다.

$$\mu | (\sigma^2, \mathbf{x}) \stackrel{d}{\sim} \mathcal{N}\left(\mu_*, \frac{\sigma^2}{\kappa_*}\right), \quad \sigma^2 | \mathbf{x} \stackrel{d}{\sim} \mathcal{IG}\left(\frac{\nu_*}{2}, \frac{S_* \nu_*}{2}\right) \quad (18)$$

식 (1)과 식 (18)에서 알 수 있듯이, 사전확률밀도함수와 사후확률밀도함수는 같은 형태를 취한다. 따라서, 이 사전확률분포는 자연공액사전확률분포 (natural conjugate prior distribution) 이다. 식 (18)에서 알 수 있듯이,  $\sigma^2$  의 사후확률분포의 평균과 분산은 각각 다음과 같다.

$$E(\sigma^2 | \mathbf{x}) = \frac{\nu_* S_*}{\nu_* - 2} \quad (19)$$

$$Var(\sigma^2 | \mathbf{x}) = \frac{2\nu_*^2 S_*^2}{[\nu_* - 2]^2 [\nu_* - 4]} \quad (20)$$

첫째로, 사전CVaR를 구하기로 하자. 식 (4)에서 알 수 있듯이,  $\mu$  의 사전확률밀도함수는

다음과 같다.

$$\pi(\mu) = \int \pi(\mu, \sigma^2) d\sigma^2 \propto \left\{ 1 + \frac{1}{\nu_0} \left[ \frac{\mu - \mu_0}{\sqrt{S_0/\kappa_0}} \right]^2 \right\}^{-[\nu_0+1]/2} \quad (21)$$

식 (21)에서 알 수 있듯이, 다음 식이 성립한다.

$$E(\mu) = \mu_0 \quad (22)$$

식 (3)에서 알 수 있듯이, 다음 식이 성립한다.

$$E(\sigma) = \frac{\Gamma(\frac{\nu_0-1}{2})}{\Gamma(\frac{\nu_0}{2})} \sqrt{\frac{S_0\nu_0}{2}} \quad (23)$$

식 (22)와 식 (23)을 식 (2.1.24)에 대입하면, 다음과 같은 사전CVaR를 얻는다.

$$CVaR_0 = \mu_0 + \frac{\Gamma(\frac{\nu_0-1}{2})}{\Gamma(\frac{\nu_0}{2})} \sqrt{\frac{S_0\nu_0}{2}} \frac{n(N^{-1}(\beta))}{1-\beta} \quad (25)$$

둘째로, 사후CVaR를 구하기로 하자. 제2.1절에서 알 수 있듯이,  $\mu|\mathbf{x}$ 의 확률밀도함수는 다음과 같다.

$$\pi(\mu|\mathbf{x}) \propto \left\{ 1 + \frac{1}{\nu_*} \left[ \frac{\mu - \mu_*}{\sqrt{S_*/\kappa_*}} \right]^2 \right\}^{-[\nu_*+1]/2} \quad (26)$$

식 (26)에서 알 수 있듯이, 다음 식이 성립한다.

$$E(\mu|\mathbf{x}) = \mu_* \quad (27)$$

식 (16)에서 알 수 있듯이, 다음 식이 성립한다.

$$E(\sigma|\mathbf{x}) = \frac{\Gamma(\frac{\nu_*-1}{2})}{\Gamma(\frac{\nu_*}{2})} \sqrt{\frac{S_*\nu_*}{2}} \quad (28)$$

식 (27)과 식 (28)를 식 (2.1.24)에 대입하면, 다음과 같은 사후CVaR를 얻는다.

$$CVaR_* = \mu_* + \frac{\Gamma(\frac{\nu_*-1}{2})}{\Gamma(\frac{\nu_*}{2})} \sqrt{\frac{S_*\nu_*}{2}} \frac{n(N^{-1}(\beta))}{1-\beta} \quad (29)$$



**예제 2.1.3** 예제 2.1.2의 결과를 사용해서 CVaR를 구하는 실례를 다루기로 하자. 어떤 자산의 시가하락률에 대한 CVaR가 주어지면, 다음 식을 사용해서 시가감손(市價減損)의 CVaR를 구할 수 있다.

$$\text{시가감손의 CVaR} = (\text{자산의 시가}) \times (\text{시가하락률의 CVaR}) \quad (1)$$

보유자산의 시가감손에 관한 CVaR를 구할 때는 자산의 시가변화율에  $-1$ 을 곱한 자산의 시가하락률의 확률분포에 대한 CVaR를 구하는 것이 편리하다.

사전CVaR와 사후CVaR를 계산하기 위해서, 다음 MATLAB 프로그램 CVaR102.m을 실행해 보자.

```

1 % -----
2 % Filename: CVaR102.m
3 % Application of CVaR
4 % Programmed by CBS
5 % -----
6 clear all, clf, close all
7 % Prior
8 mu0 = (5:5:50)';
9 mu0dim = length(mu0);
10 kappa0 = 10;
11 nu0 = 5;
12 lambda0 = 100:100:1200;
13 lambda0dim = length(lambda0)
14 S0 = lambda0/nu0;
15 % prior CVaR
16 beta = 0.99;
17 invN = norminv(beta,0,1);
18 mu0E = mu0*ones(1,lambda0dim);
19 sigmaOE = ones(mu0dim,1)*gamma((nu0-1)/2)./gamma(nu0/2)*sqrt(lambda0/2);
20 VaR0 = mu0E + sigmaOE*invN
21 CVaR0 = mu0E + sigmaOE./(1-beta).*normpdf(invN)
22 % Plotting
23 contour(lambda0,mu0,CVaR0)
24 set(gca,'fontsize',11,'fontweigh','bold')
25 colorbar
26 xlabel('\bf \lambda_{0}','FontSize',12);
27 ylabel('\bf \mu_{0}','FontSize',12,'rotation',0);
28 saveas(gcf,'CVaR102a','eps')
29 figure
30 contour(lambda0,mu0,VaR0)
31 set(gca,'fontsize',11,'fontweigh','bold')
32 colorbar
33 xlabel('\bf \lambda_{0}','FontSize',12);
34 ylabel('\bf \mu_{0}','FontSize',12,'rotation',0);
35 saveas(gcf,'CVaR102b','eps')
36 % Observations

```

```

37 rand('twister',5489); % Use twister or state for seed.
38 n = 64; % No of Observations
39 x = random('Normal',15,4^2,[1,n]);
40 meanx = mean(x)
41 sigmax = std(x)
42 % posterior
43 muS = (kappa0*mu0 + n*meanx)./(kappa0 + n);
44 kappaS = kappa0 + n;
45 nuS = nu0 + n;
46 lambdaS = (n-1)*sigmax^2 + ones(mu0dim,1)*lambda0 ...
47 + (kappa0*n)/(kappa0 + n)*(meanx-mu0).^2*ones(1,lambda0dim);
48 SS = lambdaS./nuS;
49 % posterior CVaR
50 beta = 0.99;
51 invN = norminv(beta,0,1);
52 muSE = muS*ones(1,lambda0dim);
53 sigmaSE = gamma((nuS-1)/2)./gamma(nuS/2)*sqrt(lambdaS/2);
54 VaRS = muSE + sigmaSE*invN
55 CVaRS = muSE + sigmaSE./(1-beta).*normpdf(invN)
56 % Plotting
57 figure
58 contour(lambda0,mu0,CVaRS)
59 set(gca,'fontsize',11,'fontweigh','bold')
60 colorbar
61 xlabel('\bf \lambda_{0}','FontSize',12);
62 ylabel('\bf \mu_{0}','FontSize',12,'rotation',0);
63 saveas(gcf,'CVaR102c','eps')
64 figure
65 contour(lambda0,mu0,VaRS)
66 set(gca,'fontsize',11,'fontweigh','bold')
67 colorbar
68 xlabel('\bf \lambda_{0}','FontSize',12);
69 ylabel('\bf \mu_{0}','FontSize',12,'rotation',0);
70 saveas(gcf,'CVaR102d','eps')
71 save('CVaR102','VaR0','CVaR0','VaRS','CVaRS')
72 % End of program
73 % -----

```

이 MATLAB 프로그램 CVaR102.m에서는 다음과 같은 사전확률분포를 가정한다.

$$\mu | \sigma^2 \stackrel{d}{\sim} \mathcal{N}\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right), \quad \sigma^2 \stackrel{d}{\sim} \mathcal{IG}\left(\frac{5}{2}, \frac{\nu_0 S_0}{2}\right) \quad (1)$$

여기서 각 모수는 다음과 같다.

$$\mu_0 = 5, 10, \dots, 45, 50, \quad \lambda_0 = 100, 200, \dots, 900, 1000 \quad (2)$$

$$\kappa_0 = 10, \quad \nu_0 = 5, \quad S_0 = \frac{\lambda_0}{\nu_0} \quad (3)$$

또한, 경계확률은  $\beta = 0.99$ 이다. 예제 2.1.3의 식 (25)에서 알 수 있듯이, 사전 CVaR는 다음과

같다.

$$CVaR_0 = \mu_0 + \frac{2\sqrt{2}}{3\sqrt{\pi}} \frac{n(N^{-1}(0.99))}{0.01} \sqrt{\lambda_0} \quad (4)$$

주어진  $\{(\lambda_0, \mu_0)\}$ 에 대해서 계산한 사전CVaR값들의 등가곡선(level curve)이 그림 2.1.2에 그려져 있다. 그림 2.1.2에서 알 수 있듯이,  $\lambda_0$ 가 커질수록 그리고  $\mu_0$ 가 커질수록 사전CVaR도 커진다. 이 사전CVaR는 20에서 90사이를 움직인다.

이 MATLAB프로그램 CVaR102.m에서는 정규확률분포에서 손실액의 실현값들을 64개 발생시킨다. 이들의 표본평균은 12.3013이고 표본분산은 15.1560<sup>2</sup>이다. 예제 2.1.3의 식 (29)에서 알 수 있듯이, 사후CVaR는 다음과 같다.

$$CVaR_* = \mu_* + \frac{\Gamma(34)}{\Gamma(34.5)} \sqrt{\frac{69}{2}} \frac{n(N^{-1}(0.99))}{0.01} \sqrt{S_*} \quad (5)$$

주어진  $\{(\lambda_0, \mu_0)\}$ 에 대해서 계산한 사후CVaR값들의 등가곡선이 그림 2.1.3에 그려져 있다. 그림 2.1.3에서 알 수 있듯이,  $\lambda_0$ 가 커질수록 그리고  $\mu_0$ 가 커질수록 사후CVaR도 커진다. 사전CVaR과는 달리,  $\lambda_0$ 가 크게 변화해도 사후CVaR는 크게 변하지 않는다. 즉,  $\mu_0$ 가 고정되어있는 경우  $S_0$ 가 크게 변화해도 사후CVaR는 크게 변하지 않는다. 더구나, 관찰값들의 개수가  $n = 64$ 로 비교적 크기 때문에 사후CVaR가 취하는 범위는 사전CVaR가 취하는 범위보다 작아진다. 이것은 관찰값들이 제공하는 정보에 의해 사전확률분포가 수정된 결과로 사전정보의 영향이 작아져서 사후확률분포가 한 곳으로 집중되기 때문이다. ■

## 제2.2절 신용리스크의 베이지안통계분석

베이지안통계분석법을 적용하는 두 번째 예제로서 파산확률을 계측하는 문제를 살펴보자.

### 2.2.1 파산확률과 선형회귀모형

신용리스크(credit risk)란 돈을 빌린 사람이 빌린 돈을 갚을 수 없게 되는 채무불이행 리스크로서 디폴트리스크(default risk)라고도 부른다. 기업의 경영상태가 악화되어 은행으로부터 빌린 용자와 같은 채무를 갚을 수 없게 되는 것이 디폴트의 전형적인 예이다. 그러나, 디폴트는 기업에만 일어나는 것은 아니다. 미국의 서브프라임모기지론사태는 신용도가 낮은 가계를

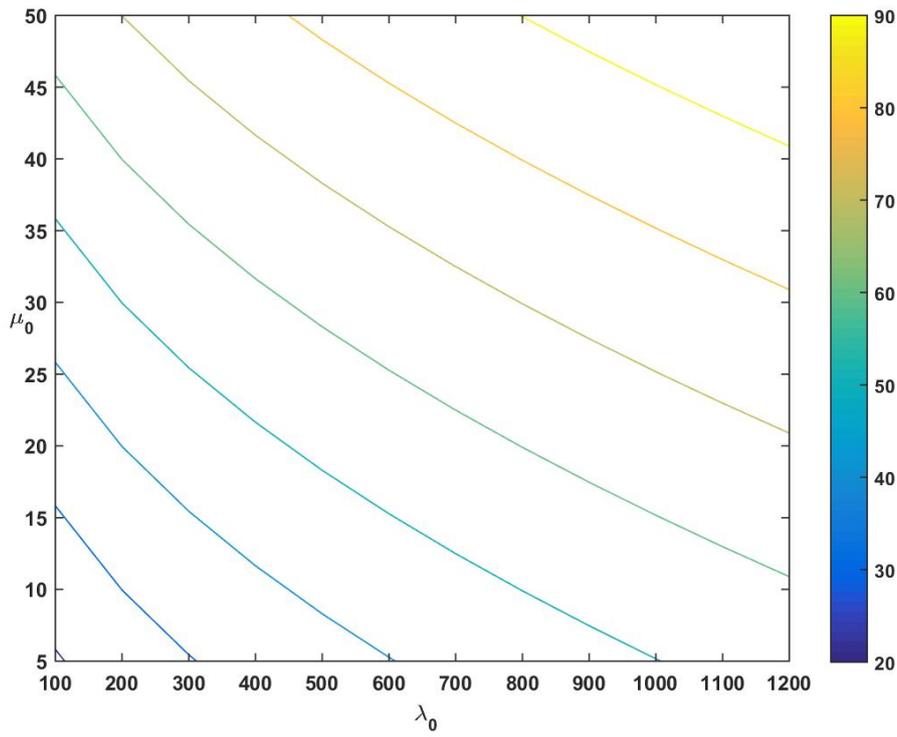


그림 2.1.2. 사전CVaR

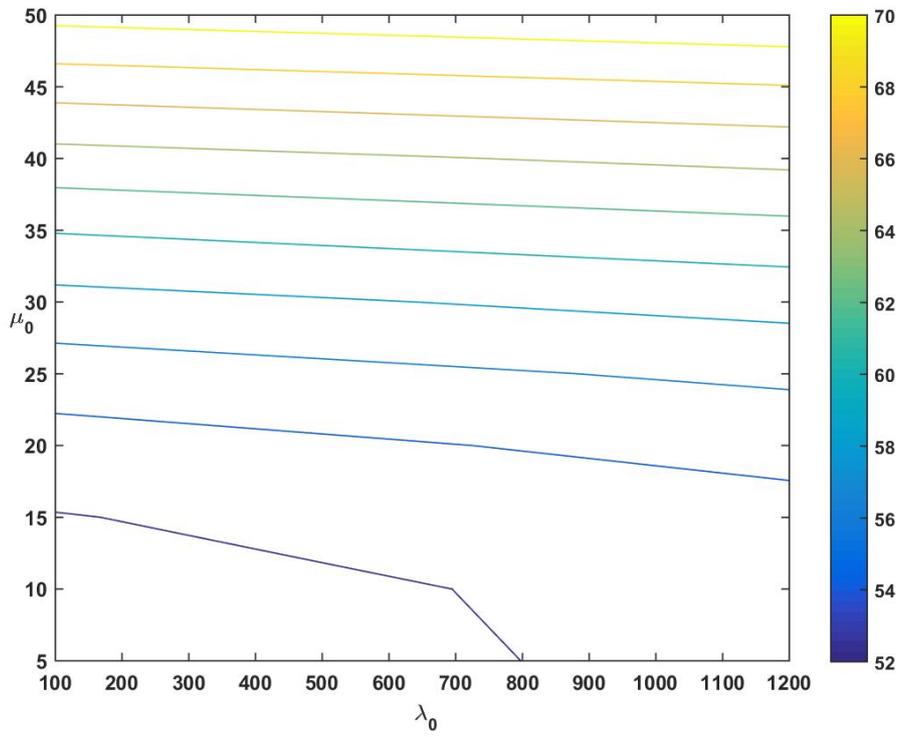


그림 2.1.3. 사후CVaR

위한 서브프라임모기지론이 남발되어 결국에는 회수불능 상태가 되는 경우가 아주 많이 발생해서 일어난 사태이다. 채무자가 외국정부이거나 외국의 공적기관이거나 또는 채무에 대해 외국정부나 외국의 공적기관이 그 원리금 지불에 대해 보증한 경우, 그 외국정부가 최종적인 채무상환의무를 지게된다. 이 경우에 국가주권자(sov​er​eign)에는 이 채무상환에 관계된 리스크가 따른다. 이러한 리스크를 소버린리스크(sov​er​eign risk)라 부른다. 즉, 이러한 채무가 최종적으로 회수불능 상태가 되면, 정부가 디폴트된 것이다. 유럽 경제위기 당시 그리스와 같이 재정기반이 취약한 국가가 국채를 포함한 채무를 갚지 못하는 소버린리스크가 있었다. 해외투용자에 관계된 가장 넓은 의미의 위험을 컨트리리스크(country risk)라고도 하는데, 컨트리리스크보다는 소버린리스크라는 용어가 더 많이 사용되고 있다.

신용리스크분석의 중심은 기본적으로 채무자의 디폴트확률을 평가하는 것이다. 디폴트 확률의 평가법은 다양하지만, 이 절에서는 회귀모형을 사용해서 디폴트확률을 평가하고자 한다. 디폴트확률을 종속변수로 하는 회귀모형을 사용하는 목적은 채무자의 재무상황 및 경기동향에 관한 경제지표 등과 디폴트확률 사이의 관계를 설명하는 것이다. 그리고 과거 데이터를 사용해서 추정한 디폴트확률의 회귀모형을 사용해서, 새로운 채무자의 디폴트확률을 예측하고자 한다. 이 절에서는 베이지안방법을 사용해서 디폴트확률의 회귀모형을 추정하고, 추정된 모형을 사용해서 디폴트확률을 예측하고자 한다.

기본적으로 은행의 비지니스모형은 단기금융시장에서 낮은 이자율로 자금을 조달해서 높은 이자율로 기업 등에 용자를 해주고 차익을 벌어들이는 것이다. 그러나, 용자를 받은 기업이 디폴트되면 채무를 모두 회수할 수 있다는 보장이 없다. 은행의 조달이자율을  $R$ 이라고 하고, 기업에 대한 대출이자율을 조달이자율  $R$ 에 이자율스프레드  $S$ 를 더한  $R + S$ 라고 하자. 이 기업의 디폴트확률을  $\pi$  그리고 디폴트 발생시에 채권의 회수율을  $\rho(\in (0, 1))$ 라고 하자. 디폴트확률이  $\pi$ 인 기업에 대해 은행이  $D$ 만큼 용자를 하면, 회수되는 채권액은 다음과 같다.

$$\text{회수금액} = \begin{cases} [1 + R + S]D, & (\text{파산하지 않는 경우}) \\ \rho[1 + R + S]D, & (\text{파산하는 경우}) \end{cases} \quad (2.2.1)$$

따라서, 기대회수액은 다음과 같다.

$$\begin{aligned} \text{기대회수액} &= [1 - \pi][1 + R + S]D + \pi\rho[1 + R + S]D \\ &= [1 - \pi + \pi\rho][1 + R + S]D \end{aligned} \quad (2.2.2)$$

은행이 용자금  $D$ 를 조달이자율  $R$ 로 조달하였으므로, 다음 식이 성립하지 않으면 은행은

순이익을 기대할 수 없다.

$$[1 - \pi + \pi\rho][1 + R + S]D \geq [1 + R]D \quad (2.2.3)$$

채무기업의 디폴트확률  $\pi$ , 조달이자율  $R$ , 그리고 회수율  $\rho$ 는 은행이 받아들여야 하는 값들이다. 따라서, 은행은 식 (2.2.3)을 만족하는 이자율스프레드  $S$ 를 채택해야 한다. 식 (2.2.3)을 다음과 같이 쓸 수 있다.

$$S \geq \frac{\pi}{[1 - \pi + \pi\rho]}[1 + R][1 - \rho] \quad (2.2.4)$$

디폴트확률  $\pi$ 와 회수율  $\rho$ 가 모두 작은 양수들이라는 가정 하에 식 (2.2.4)를 다음과 같이 쓸 수 있다.

$$S \geq \frac{\pi}{1 - \pi}[1 + R][1 - \rho] \quad (2.2.5)$$

어떤 사건이 발생할 확률  $p$ 에 대해  $OR(p) \doteq p/[1 - p]$ 를 오즈비(odds ratio)라고 한다. 따라서, 디폴트확률  $\pi$ 에 대해  $OR(\pi) = \pi/[1 - \pi]$ 를 디폴트오즈비(default odds ratio)라고 한다. 이 디폴트오즈비  $OR(\pi)$ 는  $\pi$ 의 증가함수이다. 은행이 기업에 용자를 할 경우, 식 (2.2.5)에서 알 수 있듯이 조달이자율에 더해야 하는 최소스프레드  $S_m$ 은 다음과 같다.

$$S_m \doteq OR(\pi)[1 + R][1 - \rho] \quad (2.2.6)$$

디폴트확률  $\pi$ 가 커지면 디폴트오즈비  $OR(\pi)$ 도 커지므로, 식 (2.2.6)은 기업이 디폴트할 가능성이 높을수록 이자율스프레드를 높게 설정해야함을 알 수 있다. 이는 현실과 부합되는 결과이다.

식 (2.2.6)을 사용해서 이자율스프레드를 정할 때 필요한 것은 디폴트확률  $\pi$ , 조달이자율  $R$  그리고 회수율  $\rho$ 이다. 조달이자율  $R$ 은 단기금융시장에서 결정되는 것으로 은행은 대출 전에 이 값을 이미 알고 있다. 그러나, 디폴트확률  $\pi$ 와 회수율  $\rho$ 는 은행이 정확하게는 모르는 값들로 보는 것이 타당하다. 회수율  $\rho$ 에 대해서는 최악의 상황인 회수율  $\rho = 0$ 를 상정해서 가장 보수적으로 이자율스프레드를 결정할 수 있다. 이 경우에 식 (2.2.6)을 이용해서 이자율스프레드를 구하기 위해서는 디폴트오즈비만 알면 충분하다.

지금부터 디폴트오즈비를 추정하는 방법을 살펴보자. 용자받는 기업의 디폴트오즈비를 추정하기 위해서 우선 유사한 기업의 디폴트오즈비에 관한 데이터를 모은다. 식 (2.2.6)에서

알 수 있듯이, 다음 식을 사용해서 시장으로부터 디폴트오즈비를 계산할 수 있다.

$$\text{디폴트오즈비} = \frac{\text{회사채스프레드}}{[1 + \text{안전이자율}][1 - \text{회수율}]} \quad (2.2.7)$$

여기서 안전이자율은 현실적으로 가능한 무위험이자율을 뜻하고, 회사채스프레드는 회사채 이자율과 안전이자율의 차이이다.

만약 유사한 기업의 회사채가 시장에서 거래되고 있다면, 식 (2.2.7)을 사용해서 용자받는 기업의 디폴트오즈비를 추정할 수 있는 데이터를 얻을 수 있을 것이다. 또한 회사채스프레드 대신에 CDS(credit default swap)의 프리미엄(premium)을 사용할 수도 있다. CDS란 채권자가 디폴트했을 때 회사채의 상환액 등을 보장받는 대신에 수수료인 프리미엄을 지불하는 계약이다. CDS를 매입하면, 용자받는 기업이 디폴트되는 경우에 매입한 채권의 전액 환수가 보장된다. 이 경우에는 용자받는 기업의 디폴트 유무와 관계없이 기대회수액은 채권액  $[1 + R + S]D$  자체가 된다. 그러나, 채권자가 추가로 CDS프리미엄을 매입하기 위한 전체 조달비용은 조달이자율  $R$ 에 CDS프리미엄  $P$ 를 더한  $R + P$ 이다. 따라서, CDS를 보유하는 경우에는 다음 식이 성립하지 않으면 은행은 순이익을 기대할 수 없다.

$$[1 + R + S]D \geq [1 + R + P]D \quad (2.2.8)$$

식 (2.2.8)은 다음 식과 동치이다.

$$S \geq P \quad (2.2.9)$$

식 (2.2.7)과 식 (2.2.9)에서 알 수 있듯이, 허용가능한 최대 CDS프리미엄은 다음과 같다.

$$\text{최대CDS프리미엄} = \text{디폴트오즈비} \times [1 + \text{조달이자율}][1 - \text{회수율}] \quad (2.2.10)$$

이 경우에 다음 식을 사용해서 시장으로부터 디폴트오즈비를 계산할 수 있다

$$\text{디폴트오즈비} = \frac{\text{CDS프리미엄}}{[1 + \text{안전이자율}][1 - \text{회수율}]} \quad (2.2.11)$$

즉, 회사채스프레드를 대신해서 CDS프리미엄을 사용해서 디폴트오즈비를 추정한다. 회사채스프레드나 CDS프리미엄을 대신해서 신용평가기관이 공표한 신용등급 별 디폴트확률의 추정값을 이용해서 디폴트오즈비를 계산할 수도 있다. 그러나, 이 경우에는 개별 기업의 디폴트가능성을 분석할 수는 없다. 신용평가기관이 공표한 신용등급 별 디폴트확률추정값은

같은 신용등급을 갖는 기업들을 하나의 그룹으로 분석할 때 사용할 수 있다. 그러나, 현실 세계에는 동일한 기업은 존재하지 않는다. 같은 업종에서 동일한 등급을 받는 기업들이라해도 디폴트확률은 같지 않다. 따라서, 비슷한 기업의 디폴트오즈비에 관한 데이터셋로부터 표본평균 등을 사용해서 융자받는 기업의 디폴트오즈비를 추정하는 것은 좋은 방법이라고 할 수 없다. 대안으로 기업의 디폴트확률을 결정하는 다양한 요인들을 골라서, 이 요인들을 사용해서 디폴트오즈비를 나타내는 모형을 살펴보기로 하자. 예를 들어, 디폴트오즈비모형으로 다음과 같은 선형회귀모형을 생각해보자.

$$\ln \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + \epsilon, \quad \epsilon \stackrel{d}{\sim} \mathcal{N}(0, \sigma^2) \quad (2.2.12)$$

편의상, 다음 확률변수를 정의하자.

$$y \doteq \ln \frac{\pi}{1-\pi} \quad (2.2.13)$$

확률변수  $y$ 는 디폴트오즈비의 자연로그값이므로 대수디폴트오즈비라고 하자.

선형회귀모형 (2.2.12)의 특징을 살펴보자. 첫째, 다음 식이 성립한다.

$$\frac{\pi}{1-\pi} = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + \epsilon) \quad (2.2.14)$$

따라서, 이 회귀모형에서 디폴트오즈비는 반드시 양수이다. 둘째, 설명변수들  $x_1, x_2, \dots, x_K$ 는 디폴트확률을 결정하는 요인들이다. 각  $k(= 1, 2, \dots, K)$ 에 대해서,  $\beta_k$ 가 양수인 경우에는 요인  $x_k$ 가 커지면 디폴트오즈비는 상승한다. 반대로,  $\beta_k$ 가 음수이면 요인  $x_k$ 가 커지면 디폴트오즈비는 하락한다. 즉,  $\beta_k$ 는 디폴트오즈비에 요인  $x_k$ 가 미치는 영향의 방향을 결정하는 모수이다. 셋째, 종속변수  $y$ 가 포함한 정보 중에서 설명변수들  $x_1, x_2, \dots, x_K$ 로 설명할 수 없는 부분을 오차항  $\epsilon$ 으로 표현한다. 오차항  $\epsilon$ 은 확률변수이므로 식 (2.2.14)의 디폴트오즈비도 확률변수가 된다. 따라서, 식 (2.2.14)의 디폴트오즈비를 그대로 식 (2.2.6)의 디폴트오즈비  $OR(\pi)$ 로 대체해서 이자율스프레드 평가에 사용하는 것은 타당하지 않다. 넷째, 식 (2.2.12)에서 알 수 있듯이, 대수디폴트오즈비의 기대값은 설명변수들  $x_1, x_2, \dots, x_K$ 의 선형함수이다. 즉, 다음 식이 성립한다.

$$E(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K \quad (2.2.15)$$

다섯째, 식 (2.2.14)에서 알 수 있듯이, 다음 식이 성립한다.

$$\frac{\pi}{1-\pi} = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K) \exp(\epsilon) \quad (2.2.16)$$

정규확률분포  $\mathcal{N}(0, \sigma^2)$ 를 따르는 확률변수  $z$ 의 적률생성함수  $M_z(t)$ 는 다음과 같다.

$$M_z(t) = \exp\left(\frac{1}{2}\sigma^2 t^2\right) \quad (2.2.17)$$

식 (2.2.16)과 식 (2.2.17)에서 알 수 있듯이, 다음 식이 성립한다.

$$E\left(\frac{\pi}{1-\pi}\right) = \exp\left(\beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + \frac{1}{2}\sigma^2\right) \quad (2.2.18)$$

식 (2.2.18)의 기대디폴트오즈비는 비확률적이므로, 이 기대디폴트오즈비를 식 (2.2.6)의 디폴트오즈비  $OR(\pi)$ 로 사용할 수 있다. 식 (2.2.18)에서  $\beta_0$ 와  $\sigma^2$ 는 상수항들이다. 따라서, 데이터로부터  $\beta_0 + \sigma^2/2$ 를 추정한다해도 다른 제약조건이 부과되지 않으면  $\beta_0$ 와  $\sigma^2$  각각을 추정할 수는 없다. 따라서, 다음과 같은 선형회귀모형을 사용하기로 하자.

$$\ln \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + \epsilon, \quad \epsilon \stackrel{d}{\sim} \mathcal{N}(0, 1) \quad (2.2.19)$$

즉,  $\sigma^2 = 1$ 이라고 가정하자. 식 (2.2.18)과 식 (2.2.19)에서 알 수 있듯이, 다음 식이 성립한다.

$$E\left(\frac{\pi}{1-\pi}\right) = \exp\left(\beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + \frac{1}{2}\right) \quad (2.2.20)$$

### 2.2.2 파산확률의 베이지안추정

식 (2.2.20)의 기대디폴트오즈비를 식 (2.2.6)의 디폴트오즈비로 대체하기 위해서는, 식 (2.2.19)의 설명변수들  $x_1, x_2, \dots, x_K$ 를 먼저 결정해야 한다. 설명변수들  $x_1, x_2, \dots, x_K$ 로 무엇을 사용하는게 좋을까? 기업의 디폴트확률은 해당 기업의 제품에 대한 수요나 경쟁사의 능력 등 미시경제적 동향과 한국과 세계의 경기변동이나 환율 등 거시경제적 동향에 따라 변화한다. 당연히 융자를 받는 기업의 실적에 영향을 줄 가능성이 높은 경제적 요인이 설명변수의 후보이다. 소비자의 기호 변화나 다른 회사와 경쟁이 격화됨에 따라 기업의 매출이 감소하여, 이 기업이 디폴트로 몰리게 될수도 있다. 또한, 원화가치가 급격히 상승하거나 세계경제의 급속한 침체에 따라 수출에 의존하는 기업이 디폴트할 수도 있다. 그러나, 이러한 경제적

요인들이 모든 기업의 디폴트를 유발하지는 않는다. 현금을 많이 보유한 기업이라면 일시적인 이익감소를 이겨낼 수 있으나, 재무능력이 약한 기업은 경기 하락의 영향을 피하지 못하고 디폴트할 수도 있다. 따라서 기업의 재무상태를 나타내는 변수 역시 회귀모형의 설명변수가 될 수 있다. 기업의 업적은 경영자의 능력에 크게 좌우되므로, 경영자의 자질을 수치화한 것을 설명변수로 사용할 수 있다. 이렇게 경제적 요인과 재무적 요인 등을 설명변수들로 간주해서 디폴트오즈비를 파악하려고 하는 것이 회귀모형 (2.2.19)를 사용하는 목적이다.

다음과 같은 세 단계들을 거쳐 용자 받는 기업의 디폴트오즈비와 이자율스프레드를 평가할 수 있다.

- (1단계) 시장에서 모든 데이터를 바탕으로 회귀모형 (2.2.19)의 회귀계수들  $\beta_0, \beta_1, \dots, \beta_K$ 를 추정한다.
- (2단계) 현재의 경제상황을 나타내는 설명변수값들과 용자를 받는 기업의 경영상황을 나타내는 설명변수값들을 추정된 선형회귀모형에 대입해서 용자를 받는 기업의 디폴트오즈비 추정값을 구한다.
- (3단계) 이 디폴트오즈비 추정값을 디폴트오즈비로 사용하여, 이자율스프레드를 구한다.

회귀모형의 추정에 사용할 수 데이터를 얻을 수 있는 기업이  $N$ 개라고 하고, 제  $n$  번째 기업에 관한 대수디폴트오즈비를  $y_n$ , 설명변수들의 관찰값들을  $x_{n1}, x_{n2}, \dots, x_{nK}$ 라 하자. 즉, 추정하고자 하는 선형회귀모형은 다음과 같다.

$$y_n = \beta_0 + \beta_1 x_{n1} + \dots + \beta_K x_{nK} + \epsilon_n, \quad (n = 1, 2, \dots, N) \quad (2.2.21)$$

여기서  $\epsilon_1, \epsilon_2, \dots, \epsilon_N$  서로 독립이며 표준정규확률분포  $\mathcal{N}(0, 1)$ 을 따른다. 선형회귀모형 (2.2.21)을 행렬형태로 나타내기 위해서, 다음 벡터들과 행렬을 정의하자.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1K} \\ 1 & x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{NK} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_N \end{bmatrix} \quad (2.2.22)$$

식 (2.2.21)과 식 (2.2.22)에서 알 수 있듯이, 이 선형회귀모형을 다음과 같이 쓸 수 있다.

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \stackrel{d}{\sim} \mathcal{N}_N(\mathbf{0}, I) \quad (2.2.23)$$

회귀계수벡터  $\boldsymbol{\beta}$ 의 사전확률분포가 다음과 같다고 가정하자.

$$\boldsymbol{\beta} \stackrel{d}{\sim} \mathcal{N}_{K+1}(\boldsymbol{\beta}_0, \Sigma_0) \quad (2.2.24)$$

즉, 회귀계수벡터  $\boldsymbol{\beta}$ 의 사전확률밀도함수는 다음과 같다.

$$p(\boldsymbol{\beta}) = [2\pi]^{-[K+1]/2} |\Sigma_0|^{-1/2} \exp\left(-\frac{1}{2} [\boldsymbol{\beta} - \boldsymbol{\beta}_0]^T \Sigma_0^{-1} [\boldsymbol{\beta} - \boldsymbol{\beta}_0]\right) \quad (2.2.25)$$

대수디폴트오즈비들의 벡터  $\mathbf{y}$ 의 확률밀도함수는 다음과 같다.

$$p(\mathbf{y} | X, \boldsymbol{\beta}) \propto \exp\left(-\frac{1}{2} [\mathbf{y} - X\boldsymbol{\beta}]^T [\mathbf{y} - X\boldsymbol{\beta}]\right) \quad (2.2.26)$$

이 확률밀도함수  $p(\mathbf{y} | X, \boldsymbol{\beta})$ 를  $\boldsymbol{\beta}$ 의 함수로 본 것이 우도함수  $L(\boldsymbol{\beta} | \mathbf{y}, X)$ 이다.

식 (2.2.25)와 식 (2.2.26)을 Bayes정리를 대입하면, 회귀계수벡터  $\boldsymbol{\beta}$ 의 사후확률밀도함수가 다음과 같음을 알 수 있다.

$$\begin{aligned} p(\boldsymbol{\beta} | \boldsymbol{\beta}_0, \Sigma_0, X, \mathbf{y}) &\propto p(\mathbf{y} | X, \boldsymbol{\beta}) p(\boldsymbol{\beta}) \\ &\propto \exp\left(-\frac{1}{2} \left\{ [\mathbf{y} - X\boldsymbol{\beta}]^T [\mathbf{y} - X\boldsymbol{\beta}] + [\boldsymbol{\beta} - \boldsymbol{\beta}_0]^T \Sigma_0^{-1} [\boldsymbol{\beta} - \boldsymbol{\beta}_0] \right\}\right) \end{aligned} \quad (2.2.27)$$

식 (2.2.27)을 간단히 쓰기 위해서, 다음 벡터들과 행렬을 정의하자.

$$\mathbf{u} \doteq \begin{bmatrix} \mathbf{y} \\ \boldsymbol{\beta}_0 \end{bmatrix}, \quad Z \doteq \begin{bmatrix} X \\ I_{K+1} \end{bmatrix}, \quad \Omega \doteq \begin{bmatrix} I_N & O \\ O & \Sigma_0 \end{bmatrix} \quad (2.2.28)$$

식 (2.2.27)을 다음과 같이 쓸 수 있다.

$$p(\boldsymbol{\beta} | \boldsymbol{\beta}_0, \Sigma_0, X, \mathbf{y}) \propto \exp\left(-\frac{1}{2} [\mathbf{u} - Z\boldsymbol{\beta}]^T \Omega^{-1} [\mathbf{u} - Z\boldsymbol{\beta}]\right) \quad (2.2.29)$$

식 (2.2.29)를 식 (2.2.26)과 비교해보면, 식 (2.2.23)에서 알 수 있듯이 식 (2.2.29)는 다음

선형회귀모형의 우도함수이다.

$$\mathbf{u} = Z\boldsymbol{\beta} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \stackrel{d}{\sim} \mathcal{N}_{N+K+1}(\mathbf{0}, \Omega) \quad (2.2.30)$$

선형회귀모형 (2.2.30)에서  $\boldsymbol{\beta}$ 의 일반화최소제곱추정량 (generalized least squares estimator: GLS추정량)은 다음과 같다.

$$\boldsymbol{\beta}_* = [Z^T \Omega^{-1} Z]^{-1} Z^T \Omega^{-1} \mathbf{u} = [X^T X + \Sigma_0^{-1}]^{-1} [X^T \mathbf{y} + \Sigma_0^{-1} \boldsymbol{\beta}_0] \quad (2.2.31)$$

식 (2.2.31)에서 알 수 있듯이, 다음 식이 성립한다.

$$Z^T \Omega^{-1} [\mathbf{u} - Z\boldsymbol{\beta}_*] = \mathbf{0} \quad (2.2.32)$$

식 (2.2.32)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\begin{aligned} & [\mathbf{u} - Z\boldsymbol{\beta}]^T \Omega^{-1} [\mathbf{u} - Z\boldsymbol{\beta}] \\ &= \{[\mathbf{u} - Z\boldsymbol{\beta}_*] + Z[\boldsymbol{\beta}_* - \boldsymbol{\beta}]\}^T \Omega^{-1} \{[\mathbf{u} - Z\boldsymbol{\beta}_*] + Z[\boldsymbol{\beta}_* - \boldsymbol{\beta}]\} \\ &= [\boldsymbol{\beta}_* - \boldsymbol{\beta}]^T Z^T \Omega^{-1} Z [\boldsymbol{\beta}_* - \boldsymbol{\beta}] + [\mathbf{u} - Z\boldsymbol{\beta}_*]^T \Omega^{-1} [\mathbf{u} - Z\boldsymbol{\beta}_*] \end{aligned} \quad (2.2.33)$$

식 (2.2.33)을 식 (2.2.29)에 대입하면, 다음 식을 얻는다.

$$\begin{aligned} p(\boldsymbol{\beta} \mid \boldsymbol{\beta}_0, \Sigma_0, X, \mathbf{y}) &\propto \exp\left(-\frac{1}{2} [\boldsymbol{\beta}_* - \boldsymbol{\beta}]^T Z^T \Omega^{-1} Z [\boldsymbol{\beta}_* - \boldsymbol{\beta}]\right) \\ &\quad \cdot \exp\left(-\frac{1}{2} [\mathbf{u} - Z\boldsymbol{\beta}_*]^T \Omega^{-1} [\mathbf{u} - Z\boldsymbol{\beta}_*]\right) \end{aligned} \quad (2.2.34)$$

식 (2.2.34)의 우변에서 두 번째 항은 관측값과 상수의 함수이므로 이 식을 다음과 같이 쓸 수 있다.

$$p(\boldsymbol{\beta} \mid \boldsymbol{\beta}_0, \Sigma_0, X, \mathbf{y}) \propto \exp\left(-\frac{1}{2} [\boldsymbol{\beta}_* - \boldsymbol{\beta}]^T \Sigma_*^{-1} [\boldsymbol{\beta}_* - \boldsymbol{\beta}]\right) \quad (2.2.35)$$

여기서  $\Sigma_*$ 는 다음과 같다.

$$\Sigma_* \doteq [Z^T \Omega^{-1} Z]^{-1} = [X^T X + \Sigma_0^{-1}]^{-1} \quad (2.2.36)$$

식 (2.2.35)에서 알 수 있듯이,  $\beta$ 의 사후확률분포는 다음과 같다.

$$\beta \mid (\beta_0, \Sigma_0, X, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}_{K+1}(\beta_*, \Sigma_*) \quad (2.2.37)$$

식 (2.2.37)에 기술한 회귀계수벡터  $\beta$ 의 사후확률분포를 사용해서, 융자를 받는 기업의 디폴트오즈비에 대한 베이지안추정을 하자. 여기서 유의해야 할 것은 융자를 받는 기업은 식 (2.2.37)에 기술된 사후확률분포를 구할 때 사용한  $N$ 개 회사들에 속할 필요가 없다는 점이다. 융자를 받는 기업의 디폴트오즈비를  $\tilde{\pi}/[1 - \tilde{\pi}]$ 로 그리고 이 디폴트오즈비에 영향을 주는 요인들의 벡터를  $\tilde{\mathbf{x}}$ 로 표기하면, 식 (2.2.20)에 기술한 기대디폴트오즈비는 다음과 같다.

$$\tilde{O} = E \left( \frac{\tilde{\pi}}{1 - \tilde{\pi}} \right) = \exp \left( \tilde{\mathbf{x}}^T \beta + \frac{1}{2} \right) \quad (2.2.38)$$

베이지안통계분석에서는 사후확률분포의 평균으로 미지의 모수벡터를 추정하는 것이 일반적 이므로, 식 (2.2.37)에 기술된 평균벡터, 즉 식 (2.2.31)의  $\beta_*$ 를 식 (2.2.38)에 대입해서 다음과 같은 디폴트오즈비의 추정량을 얻을 수 있을 것 같다.

$$\tilde{O}^* = \frac{\tilde{\pi}}{1 - \tilde{\pi}} = \exp \left( \tilde{\mathbf{x}}^T \beta_* + \frac{1}{2} \right) \quad (2.2.39)$$

유의할 점은 식 (2.2.39)의  $\tilde{O}^*$ 는 베이지안추정량이 아니라는 것이다. 사후확률분포의 평균을 모수에 대한 베이지안추정량으로 삼으면,  $\tilde{O}$ 의 베이지안추정량으로는  $\tilde{O}$  자체의 사후확률분포  $p(\tilde{O} \mid \beta_0, \Sigma_0, X, \mathbf{y})$ 의 평균을 사용해야 한다. 즉, 디폴트오즈비의 베이지안추정량은 다음과 같다.

$$\tilde{O}_* \doteq E \left( \exp \left( \tilde{\mathbf{x}}^T \beta + \frac{1}{2} \right) \mid \beta_0, \Sigma_0, X, \mathbf{y} \right) \quad (2.2.40)$$

다변량정규확률분포  $\mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$ 를 따르는 확률벡터  $\mathbf{v}$ 의 적률생성함수는 다음과 같다.

$$M_{\mathbf{v}}(\mathbf{t}) = E(\exp(\mathbf{t}^T \mathbf{v})) = \exp \left( \mathbf{t}^T \boldsymbol{\mu} + \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t} \right) \quad (2.2.41)$$

식 (2.2.37)과 식 (2.2.41)을 바탕으로 식 (2.2.40)를 계산하면, 디폴트오즈비의 베이지안추정량이 다음과 같음을 알 수 있다.

$$\tilde{O}_* = \exp \left( \tilde{\mathbf{x}}^T \beta_* + \frac{1}{2} \tilde{\mathbf{x}}^T \Sigma_* \tilde{\mathbf{x}} + \frac{1}{2} \right) \quad (2.2.42)$$

분산공분산행렬  $\Sigma_*$ 가 양정치행렬이므로, 각  $\mathbf{x} \neq \mathbf{0}$ 에 대해서 다음 식들이 성립한다.

$$\tilde{O}_* = \exp\left(\tilde{\mathbf{x}}^T \boldsymbol{\beta}_* + \frac{1}{2} \tilde{\mathbf{x}}^T \Sigma_* \tilde{\mathbf{x}} + \frac{1}{2}\right) > \exp\left(\tilde{\mathbf{x}}^T \boldsymbol{\beta}_* + \frac{1}{2}\right) = \tilde{O}^* \quad (2.2.43)$$

즉, 디폴트오즈비의 베이저안추정량  $\tilde{O}_*$ 는 식 (2.2.40)에 정의된 디폴트오즈비의 추정량  $\tilde{O}^*$ 보다 크다. 식 (2.2.43)는 일종의 Jensen부등식임을 상기하라. 식 (2.2.37)에서 알 수 있듯이,  $\tilde{\mathbf{x}}^T \Sigma_* \tilde{\mathbf{x}}$ 는  $\tilde{\mathbf{x}}^T \boldsymbol{\beta}$ 의 사후분산이다. 만약 추정오차가 없다면, 즉 참인  $\boldsymbol{\beta}$ 를 알고 있다면, 식  $\tilde{\mathbf{x}}^T \Sigma_* \tilde{\mathbf{x}} = 0$ 가 성립한다. 따라서, 디폴트오즈비의 베이저안추정량  $\tilde{O}_*$ 는  $\tilde{\mathbf{x}}^T \boldsymbol{\beta}$ 의 추정오차를 고려해서 용자를 받는 기업의 디폴트 가능성을 높게 평가하고 있다고 해석할 수 있다. 즉,  $\tilde{O}^*$ 는  $\boldsymbol{\beta}$ 의 추정오차를 고려하지 않기 때문에 기업의 디폴트 가능성을 작게 평가한다고 할 수 있을 것이다.

**예제 2.2.1** 인공적 재무데이터를 만들고, 이를 사용해서 파산확률의 회귀모형을 추정하고 기대파산오즈비를 추정해보자. 우선 다음과 같이 설명변수들을 생성하자.

$$x_1 = \frac{\text{전년대비 유동자산 증가액} - \text{전년대비 유동부채 증가액}}{\text{자산합계}} \quad (1)$$

$$x_2 = \frac{\text{당기순이익}}{\text{자산합계}}, \quad x_3 = \frac{\text{자기자본}}{\text{부채합계}}, \quad x_4 = \frac{\text{매출액}}{\text{자산합계}} \quad (2)$$

따라서, 설명변수들 개수는  $K = 4$ 이다. 또한, 다음 선형회귀모형을 사용해서  $N = 50$ 개의 대수파산오즈비  $y$ 를 생성하자.

$$y = \pi - 7x_1 - x_2 - 15x_3 - 2x_4 + \epsilon, \quad \epsilon \stackrel{d}{\sim} \mathcal{N}(0, 1) \quad (3)$$

다음과 같은 회귀계수벡터  $\boldsymbol{\beta}$ 의 사전확률분포를 생각해보자.

$$\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_4]^T \stackrel{d}{\sim} \mathcal{N}(\mathbf{0}, I_5) \quad (4)$$

회귀계수벡터  $\boldsymbol{\beta}$ 의 사후확률분포를 구하기 위해서, 다음 MATLAB 프로그램 DefaultProbability101.m을 실행해 보자.

```

1 % -----
2 % Filename: DefaultProbability101.m
3 % Bayesian Default Probability
4 % Programmed by CBS
5 % -----
6 clear all, clf, close all

```

```

7 N = 50;
8 rand('twister',5489)
9 x1 = -0.06 + 0.15*randn(N,1);
10 x2 = -0.01 + 0.10*randn(N,1);
11 x3 = 0.25 + 0.45*randn(N,1);
12 x4 = 2.00 + 2.85*randn(N,1);
13 epsilon = randn(length(x1),1);
14 y = pi - 7*x1 - 1*x2 - 15*x3 - 2*x4 + epsilon;
15 X = [ ones(length(x1),1), x1, x2, x3, x4 ];
16 [ N K1 ] = size(X);
17 % Frequentist
18 piF = 1./(1+exp(-y));
19 OddsF = piF./(1-piF);
20 betaF = X\y; % betaF = (X'*X)^(-1)*X'*y
21 % Bayesian
22 beta0 = zeros(K1,1);
23 Sigma0 = eye(K1);
24 InvSigma0 = inv(Sigma0);
25 SigmaB = inv(X'*X+InvSigma0)
26 betaB = SigmaB*(X'*y+InvSigma0*beta0)
27 for nn=1:N
28     xn = X(nn,:)';
29     yB(nn) = xn'*betaB;
30     OddsB(nn) = exp(xn'*betaB+1/2*xn'*SigmaB*xn+1/2);
31 end
32 piB = 1./(1+exp(-yB));
33 OddsB = OddsB';
34 % Plotting
35 NN = 1:N;
36 plot(NN,piF,'k-',NN,piB,'r--','LineWidth',1.5)
37 legend('True Prob','Bayesian Prob','location','NW')
38 set(gca,'fontsize',11,'fontweigh','bold','ylim',[-0.05 1.25])
39 xlabel('\bf Company No','FontSize',12');
40 ylabel('\bf Default Probability','FontSize',12');
41 saveas(gcf,'DefaultProbability101','eps')
42 save('DefaultProbability101','piF','piB')
43 % End of program
44 % -----

```

이 MATLAB 프로그램 DefaultProbability101.m을 실행하면, 회귀계수벡터  $\beta$ 의 사후확률분포가 다음과 같음을 알 수 있다.

$$\beta \mid (\beta_0, \Sigma_0, X, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}(\beta_*, \Sigma_*) \quad (5)$$

여기서  $\beta_*$ 와  $\Sigma_*$ 는 각각 다음과 같다.

$$\beta_* = \begin{bmatrix} 2.7110 \\ -3.5960 \\ 0.2238 \\ -13.1940 \\ -1.9733 \end{bmatrix}, \quad \Sigma_* = \begin{bmatrix} 0.0314 & 0.0262 & -0.0054 & -0.0216 & -0.0034 \\ 0.0262 & 0.4532 & 0.0180 & 0.0360 & -0.0006 \\ -0.0054 & 0.0180 & 0.6864 & 0.0279 & -0.0009 \\ -0.0216 & 0.0360 & 0.0279 & 0.1104 & 0.0013 \\ -0.0034 & -0.0006 & -0.0009 & 0.0013 & 0.0020 \end{bmatrix} \quad (6)$$

또한, 진짜디폴트확률과 사후디폴트확률을 그린 그래프가 그림 2.2.1에 그려져 있다. 그림 2.2.1에서 알 수 있듯이, 추정된 사후디폴트확률이 진짜디폴트확률을 잘 나타냄을 알 수 있다.

다음 Python 프로그램 DefaultProbability101.py를 실행하면, 이 MATLAB 프로그램 DefaultProbability101.m을 실행한 것과 동일한 결과를 얻는다.

```

1 # DefaultProbability101.py
2 """
3 Filename : DefaultProbability101.py
4 Bayesian Default Probability
5 Programmed by C.Han based on the R program by CBS
6 """
7
8 import numpy as np
9 import matplotlib.pyplot as plt
10
11 N = 50
12 np.random.seed(200)
13 x1 = -0.06 + 0.15*np.random.normal(0, 1, N)
14 x2 = -0.01 + 0.10*np.random.normal(0, 1, N)
15 x3 = 0.25 + 0.45*np.random.normal(0, 1, N)
16 x4 = 2.00 + 2.85*np.random.normal(0, 1, N)
17 epsilon = np.random.normal(0, 1, N)
18 y = np.pi - 7*x1 - 1*x2 - 15*x3 - 2*x4 + epsilon
19 X = np.array([np.ones(N), x1, x2, x3, x4]).T
20 K = X.shape[1]
21
22 # Frequentist
23 piF = 1/(1 + np.exp(-y))
24 OddsF = piF/(1 - piF)
25 betaF = np.linalg.lstsq(X, y)[0] # betaF = (X'*X)^(-1)*X'*y
26
27 # Bayesian
28 beta0 = np.zeros(K)
29 Sigma0 = np.eye(K)
30 InvSigma0 = np.linalg.inv(Sigma0)
31 SigmaB = np.linalg.inv(X.T@X + InvSigma0.T@beta0)
32 betaB = SigmaB@(X.T@y + InvSigma0@beta0)
33
34 yB = np.zeros(N)
35 OddsB = np.zeros(N)
36 for nn in range(N):

```

```

37     x_n = X[nn]
38     yB[nn] = x_n.T@betaB
39     OddsB[nn] = np.exp(x_n.T@betaB + (1/2)*x_n.T@SigmaB@x_n + 1/2)
40     piB = 1/(1 + np.exp(-yB))
41
42 # Plot
43 plt.plot(piF, label = 'True Prob.', color = 'navy', linestyle = '-')
44 plt.plot(piB, label = 'Bayesian Prob.', color = 'red', linestyle = '--')
45 plt.xlabel('Company No')
46 plt.ylabel('Default Probability')
47 plt.legend(loc = 2)
48 plt.savefig('DefaultProbability101.png')
49
50 # End of program
51 # -----

```

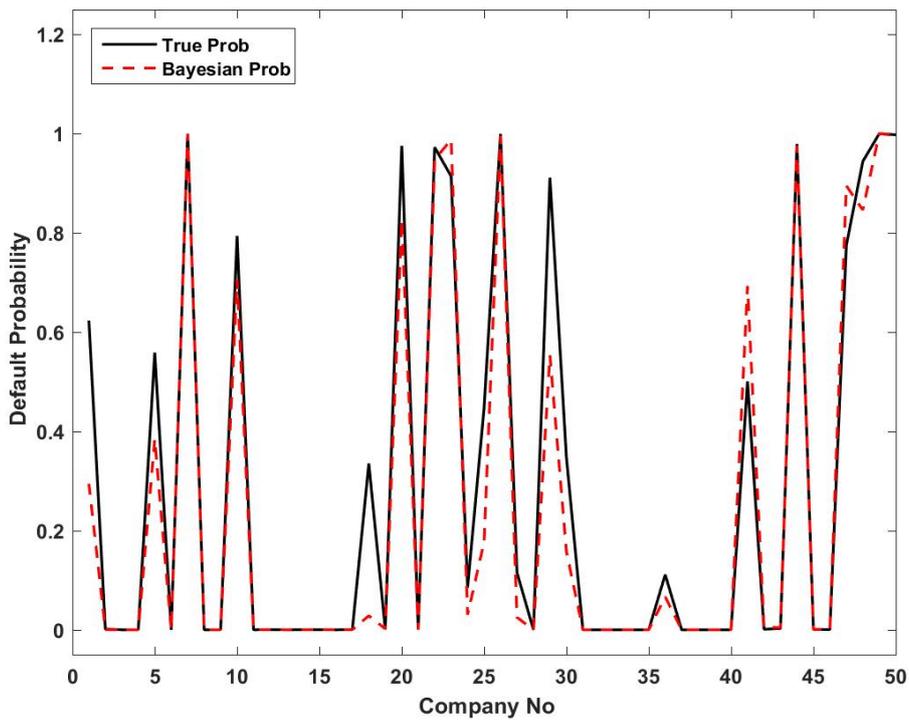


그림 2.2.1. 디폴트확률

**예제 2.2.2** 로지스틱회귀모형 (logistic regression model) 을 사용해서 기업의 디폴트확률을 추정하는 문제를 살펴보자. 식 (2.2.14)에서 오차항  $\epsilon$ 이 0이라고 가정하면, 디폴트확률은 다음과 같다.

$$\pi = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_K x_K)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_K x_K)} \quad (1)$$

유사한 기업들의 디폴트확률들이 주어진다고 가정하면, 식 (1)을 사용해서 계수들  $\beta_0, \beta_1, \dots, \beta_K$ 를 추정할 수 있다. 이 모형을 사용해서 해당 기업의 디폴트확률을 평가할 수도 있다. 식 (1)의 모형을 로지스틱회귀모형이라고 부른다. 로지스틱회귀모형에서는 디폴트오즈비의 데이터를 이용하지 않고, 다음과 같은 확률변수를 이용한다.

$$y_n = \begin{cases} 1, & \text{(제 } n \text{ 번째 기업이 파산)} \\ 0, & \text{(제 } n \text{ 번째 기업이 파산하지 않음)} \end{cases} \quad (2)$$

식 (2.2.22)에서 정의된 행렬  $X$ 와 벡터들  $\mathbf{y}$ 와  $\boldsymbol{\beta}$ 를 사용하자. 또한, 행렬  $X$ 의 제  $n$  번째 행을 행벡터  $\mathbf{x}_n$ 으로 표기하자. 식 (1)에서 알 수 있듯이, 제  $n$  번째 기업의 디폴트확률은 다음과 같다.

$$\pi_n = \frac{\exp(\mathbf{x}_n \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_n \boldsymbol{\beta})} \quad (3)$$

따라서,  $y_n$ 의 조건부확률분포는 다음과 같다.

$$p(y_n | \boldsymbol{\beta}, \mathbf{x}_n) = \pi_n^{y_n} [1 - \pi_n]^{1-y_n} \quad (4)$$

주어진  $\boldsymbol{\beta}$ 에 대해서 기업의 디폴트가 독립적으로 일어난다고 가정하면, 우도함수는 다음과 같다.

$$L(\boldsymbol{\beta}) \doteq L(\boldsymbol{\beta} | \mathbf{y}, X) = p(\mathbf{y} | \boldsymbol{\beta}, X) = \prod_{n=1}^N p(y_n | \boldsymbol{\beta}, \mathbf{x}_n) = \prod_{n=1}^N \pi_n^{y_n} [1 - \pi_n]^{1-y_n} \quad (5)$$

사전확률밀도함수를  $\pi(\boldsymbol{\beta})$ 라고 하면, 사후확률밀도함수는 다음과 같다.

$$\pi(\boldsymbol{\beta} | \mathbf{y}, X) \propto \pi(\boldsymbol{\beta}) p(\mathbf{y} | \boldsymbol{\beta}, X) = \pi(\boldsymbol{\beta}) \prod_{n=1}^N \pi_n^{y_n} [1 - \pi_n]^{1-y_n} \quad (6)$$

이 로지스틱회귀모형에서는  $\boldsymbol{\beta}$ 의 사후확률밀도함수를 해석적으로 구할 수 없다.

물론 MCMC를 써서 사후확률분포의 평균이나 디폴트확률을 구할 수 있다. 이러한 MCMC분석은 뒤에서 다루게 될 것이다. 대수디폴트오즈비의 회귀모형과는 다르게 로지스틱회귀모형의 추정에서는 회사채스프레드나 CDS프리미엄을 필요로 하지 않는다. 단지, 경제데이터와 과거에 디폴트한 기업과 현재에도 존속하는 기업에 관한 재무데이터가 있으면 로지스틱회귀모형을 추정할 수 있다. 따라서, 회사채나 CDS에 관한 데이터를 구할 수 없는 중소기업의 디폴트가능성 평가에서 로지스틱회귀모형이 널리 이용되고 있다.

로지스틱회귀모형의 추정에 대해서는 제3.1절을 참조하라. ■

## 제2.3절 정규확률분포와 베이저안통계학

고전통계학에서와 마찬가지로 베이저안통계학에서도 정규확률분포가 자주 사용된다. 고전통계학에서 정규확률분포가 널리 사용되는 이유는 다양하다. 관측값들이 근사적으로 정규확률분포를 따르고, 중심극한정리의 결과가 정규확률분포로 귀착되며, 1차적률과 2차적률이 주어진 조건 하에서 최대엔트로피확률분포가 정규확률분포이기 때문이다. 그러한 이유 이외에도 베이저안통계학에서 정규확률분포가 널리 사용되는 이유는 다양한 사전확률분포와 정규확률분포를 바탕으로 하는 우도함수를 조합해서 사후확률분포를 유도할 수 있기 때문이다. 특히, 모형을 수치적으로 추정하는 것이 곤란한 경우에도 정규확률분포 하에서는 해석적으로 사후확률분포를 구할 수도 있기 때문이다. 사후확률분포가 대칭이고 단봉(unimodal)인 경우에는 정규확률분포로 모형화할 수 있는 경우가 많다. 지금부터는 베이저안통계학에서 정규확률분포가 어떠한 역할을 하는지를 살펴보자.

### 2.3.1 정규확률분포의 평균

정규확률분포  $\mathcal{N}(\mu, \sigma^2)$  를 따르는 확률변수들  $x_1, x_2, \dots, x_n$  이 서로 독립이라고 가정하자. 이 소절에서는 분산  $\sigma^2$  를 알고 있다고 가정하고 평균  $\mu$  를 추정하기로 하자.

평균  $\mu$  의 사전확률분포가 다음과 같다고 가정하자.

$$\mu \stackrel{d}{\sim} \mathcal{N}(\mu_0, \sigma_0^2) \quad (2.3.1)$$

여기서  $\mu_0$  와  $\sigma_0^2$  는 알려진 상수들이다. 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$  의 확률밀도함수를  $n(\cdot | \mu, \sigma^2)$  로 표기하면,  $\mu$  의 사전확률밀도함수  $\pi(\mu)$  는 다음과 같다.

$$\pi(\mu) = n(\mu | \mu_0, \sigma_0^2) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{[\mu - \mu_0]^2}{2\sigma_0^2}\right) \quad (2.3.2)$$

관찰벡터  $\mathbf{x} \doteq [x_1, x_2, \dots, x_n]^T$  의 우도함수  $L(\mu | \mathbf{x})$  는 다음과 같다.

$$L(\mu | \mathbf{x}) = f(\mathbf{x} | \mu) = \prod_{j=1}^n n(x_j | \mu, \sigma^2) = \frac{1}{[\sqrt{2\pi}\sigma]^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n [x_j - \mu]^2\right) \quad (2.3.3)$$

식 (2.3.2)와 식 (2.3.3)을 Bayes정리에 대입하면,  $\mu$ 의 사후확률밀도함수가 다음과 같음을 알 수 있다.

$$\begin{aligned}\pi(\mu | \mathbf{x}) &\propto \pi(\mu)f(\mathbf{x} | \mu) = \pi(\mu) \prod_{j=1}^n f(x_j | \mu) \\ &\propto \exp\left(-\frac{1}{2} \left\{ \frac{1}{\sigma_0^2} [\mu - \mu_0]^2 + \frac{1}{\sigma^2} \sum_{j=1}^n [x_j - \mu]^2 \right\}\right) \\ &\propto \exp\left(-\frac{1}{2} \left\{ \frac{1}{\sigma_0^2} [\mu - \mu_0]^2 + \frac{n}{\sigma^2} [\bar{x} - \mu]^2 \right\}\right)\end{aligned}\quad (2.3.4)$$

다음 식이 성립함을 쉽게 알 수 있다.

$$\frac{1}{\sigma_0^2} [\mu - \mu_0]^2 + \frac{n}{\sigma^2} [\bar{x} - \mu]^2 = \frac{1}{\sigma_*^2} [\mu - \mu_*]^2 \quad (2.3.5)$$

여기서  $\bar{x}$ ,  $\mu_*$ 와  $\sigma_*^2$ 는 각각 다음과 같다.

$$\bar{x} \doteq \frac{1}{n} \sum_{j=1}^n x_j \quad (2.3.6)$$

$$\mu_* \doteq \frac{\bar{x}n/\sigma^2 + \mu_0/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2} = \frac{n/\sigma^2}{n/\sigma^2 + 1/\sigma_0^2} \bar{x} + \frac{1/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2} \mu_0 \quad (2.3.7)$$

$$\sigma_*^2 \doteq \frac{1}{n/\sigma^2 + 1/\sigma_0^2} \quad (2.3.8)$$

식 (2.3.4) ~ 식 (2.3.8)에서 알 수 있듯이, 평균  $\mu$ 의 사후확률분포는 다음과 같다.

$$\mu | \mathbf{x} \stackrel{d}{\sim} \mathcal{N}(\mu_*, \sigma_*^2) \quad (2.3.9)$$

정규확률분포라는 가정 하에서  $\bar{x}$ 는  $\mu$ 의 충분통계량이다. 이 충분통계량을 바탕으로 사후확률분포를 구하기로 하자. 다음 식들이 성립한다.

$$\bar{x} | \mu \stackrel{d}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \quad \mu \stackrel{d}{\sim} \mathcal{N}(\mu_0, \sigma_0^2) \quad (2.3.10)$$

Bayes정리에서 알 수 있듯이, 다음 식들이 성립한다.

$$\pi(\mu | \bar{x}) \propto f(\bar{x} | \mu)\pi(\mu) \propto \exp\left(-\frac{1}{2} \left\{ \frac{1}{\sigma_0^2} [\mu - \mu_0]^2 + \frac{n}{\sigma^2} [\bar{x} - \mu]^2 \right\}\right) \quad (2.3.11)$$

식 (2.3.11)은 식 (2.3.4)와 동치이다.

지금까지 내용을 다음과 같이 설명할 수 있다.

**첫째,**  $1/\sigma_0^2$ 를 사전정밀도 (prior precision)라 부르고,  $n/\sigma^2$ 를 데이터정밀도 (data precision)라 부른다. 이들을 사전가정 (prior assumption)에서 발생하는 불확실성과 관측된 데이터에서 발생하는 불확실성을 설명하는 데 사용한다. 또한,  $1/\sigma_*^2$ 가 이 정밀도들을 결합한 사후정밀도 (posterior precision)이다. 이 정밀도들은 베이저안통계학에서 중요한 역할을 한다.

**둘째,** 표본크기  $n$ 이 커짐에 따라,  $\mu_*$ 는 데이터의 표본평균  $\bar{x}$ 와 분산  $\sigma^2$ 에 더 많은 영향을 받으며 사전평균  $\mu_0$ 는 점차 중요하지 않게 된다. 식 (2.3.7)~식 (2.3.9)에서 알 수 있듯이, 표본크기  $n$ 이 아주 크면 사후확률분포는 점근적으로  $\mathcal{N}(\bar{x}, \sigma^2/n)$ 을 따른다. 이는 고전통계학에서 결과와 동일하다. 이 결과를 베이저안중심극한정리라 부른다.

**셋째,** 사전확률분포의 분산  $\sigma_0^2$ 가  $\infty$ 인 경우에도 사후확률분포는 근사적으로 정규확률분포  $\mathcal{N}(\bar{x}, \sigma^2/n)$ 을 따른다. 즉, 사전확률분포의 분산이 커지면, 사전확률분포에 포함되어 있는 정보가 사후확률분포에 미치는 영향은 감소한다.

**넷째,** 여기서는 사전확률분포와 사후확률분포가 모두 정규확률분포들이다. 이렇게 사전확률분포와 사후확률분포가 같은 확률분포족이 되는 사전확률분포의 확률분포족을 자연공액사전확률분포족 (natural conjugate prior distribution family) 또는 단순히 공액사전확률분포족 (conjugate prior distribution)이라 한다. 다시 말해서 우도함수  $f(\mathbf{x} | \boldsymbol{\theta})$ 의 모임 (family)  $\mathcal{F}$ 와 모수벡터  $\boldsymbol{\theta}$ 의 사전확률밀도함수  $\pi(\boldsymbol{\theta})$ 의 모임  $\mathcal{P}$ 에 대해서 다음 식이 성립하면, 모임  $\mathcal{P}$ 는 모임  $\mathcal{F}$ 에 대해서 공액 (conjugate)이라 한다.

$$f(\mathbf{x} | \boldsymbol{\theta}) \in \mathcal{F} \text{ and } \pi(\boldsymbol{\theta}) \in \mathcal{P} \Rightarrow \pi(\boldsymbol{\theta} | \mathbf{x}) \in \mathcal{P}$$

사전확률분포가 공액사전확률분포가 아닌 경우에는 해석적으로 사후확률분포를 쉽게 구할 수 없다. 이 경우에는 사후확률분포를 구하기 위해서 수치적분법이나 몬테카를로법을 적용할 필요가 있다.

**예제 2.3.1** 정규확률분포의 분산을 아는 경우 평균에 대한 베이저안통계분석을 하기 위해서, 다음 R프로그램 Normal\_Mean101.R을 실행해 보자.

```
1 # -----
2 # Filename: Normal_Mean101.R
```

```

3 # Conjugate Prior: Normal Distribution
4 # Programmed by CBS
5 # -----
6 # Prior
7 ( mu0 <- 1); ( sigma0 <- 1.1 )
8 Nmu <- 128
9 mu.gd <- seq(-4,6,length=Nmu)
10 priorPDF <- matrix(0,Nmu,1)
11 for (ii in 1:Nmu){
12     priorPDF[ii] <- dnorm(mu.gd[ii],mean=mu0,
13                           sd=sigma0,log = FALSE)
14 }
15 priorPDF
16 # Likelihood
17 ( mux <- 0.5); ( sigmax <- 1.3 ); ( nx <- 16)
18 set.seed(3)
19 x <- mux + sigmax*rnorm(nx,0,1)
20 xbar <- mean(x); s2 = var(x)
21 likelihood <- dnorm(x, mean=mux, sd=sigmax, log=FALSE)
22
23 # Posterior
24 mu.gd <- seq(-2,4,length=Nmu)
25 ( muStar <- (mu0/sigma0^2+nx*xbar/sigmax^2)/(nx/sigmax^2+1/sigma0^2) )
26 ( sigma2Star <- 1/(nx/sigmax^2+1/sigma0^2) )
27 posteriorPDF <- matrix(0,Nmu,1)
28 for (ii in 1:Nmu){
29     posteriorPDF[ii] <- dnorm(mu.gd[ii],mean=muStar,
30                               sd=sqrt(sigma2Star), log = FALSE)
31 }
32
33 # Plotting
34 setEPS()
35 plot.new()
36 par(mfrow=c(1,1))
37 postscript('Normal_Mean101.eps') # Start to save figure
38 plot(mu.gd,posteriorPDF,type="l", col="red", xlab="mean",
39       ylab="PDF", ylim=c(0,1.5), lwd=3 )
40 lines(mu.gd,priorPDF,lty=2,lwd=3,col =1)
41 legend(2,1,c("Prior","Posterior"),
42        lty=c(3,1),lwd=c(3,3),col=c(1,2) )
43 dev.off() # End to save figure
44 # -----

```

이 R 프로그램 Normal\_Mean101.R에서 사전확률분포는 정규확률분포  $\mathcal{N}(1, 1.1^2)$  이고, 관찰값들  $x_1, x_2, \dots, x_{16}$  을 정규확률분포  $\mathcal{N}(0.5, 1.3^2)$  에서 발생시킨다. 따라서, 사후확률분포는 정규확률분포  $\mathcal{N}(0.303507, 0.09714489)$  이다.

이 R 프로그램 Normal\_Mean101.R을 실행하면, 그림 2.3.1이 그려진다. 그림 2.3.1에서 흑색 긴점선은 사전확률밀도함수를, 그리고 적색 실선은 사후확률밀도함수를 나타낸다.

이 R 프로그램 Normal\_Mean101.R 대신에 다음 MATLAB 프로그램 Normal\_Mean101M.m 을 실행해도 동일한 결과를 얻는다.

```

1 % -----
2 % Filename: Normal_Mean101M.m

```

```

3 % Normal Mean Problem
4 % Programmed by CBS
5 % -----
6 clear all, clf, close all, clc
7 Nobs = 16;
8 xx_mu = linspace(-2,4,301);
9 rng(1,'twister');
10 muX = 0.5;
11 sigmaX = 1.3;
12 x = normrnd(muX,sigmaX,[Nobs,1]);
13 likelihoodX = normpdf(xx_mu,mean(x),sigmaX/sqrt(Nobs));
14 mu0 = 1;
15 sigma0 = 1.1;
16 pdfPrior = normpdf(xx_mu,mu0,sigma0);
17 dum1 = Nobs/sigmaX^2
18 dum0 = 1/sigma0^2
19 mean(x)
20 postMean = dum1/(dum1 + dum0)*mean(x) + dum0/(dum1 + dum0)*mu0
21 postSD = 1/sqrt(dum1 + dum0)
22 pdfPosterior = normpdf(xx_mu,postMean,postSD);
23 plot(xx_mu,likelihoodX,'g-',xx_mu,pdfPrior,'r--', ...
24      xx_mu,pdfPosterior,'b-.','LineWidth',2)
25 set(gca,'fontsize',11,'fontweigh','bold')
26 legend('Likelihood','Prior','Posterior','location','NW')
27 xlabel('\theta','FontSize',12');
28 saveas(gcf,'Normal_Mean101M','jpg')
29 % End of program
30 % -----

```



### 2.3.2 정규확률분포의 분산

정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 를 따르는 확률변수들  $x_1, x_2, \dots, x_n$ 이 서로 독립이라고 가정하자. 여기서서는 평균  $\mu$ 는 알고 있다고 가정하고 분산  $\sigma^2$ 를 추정하기로 하자.

분산  $\sigma^2$ 의 사전확률분포가 역감마확률분포  $\mathcal{IG}(\alpha, \beta)$ 를 따른다고 하자. 앞에서 설명했듯이, 역감마확률분포란 확률변수  $x$ 가 감마확률분포를 따를 때 확률변수  $1/x$ 가 따르는 확률분포를 의미한다. 따라서, 주어진 상수들  $\alpha(> 0)$ 와  $\beta(> 0)$ 에 대해서 분산  $\sigma^2$ 의 사전확률밀도함수를 다음과 같이 쓸 수 있다.

$$\pi(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} [\sigma^2]^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma^2}\right), \quad (\sigma^2 > 0) \quad (2.3.12)$$

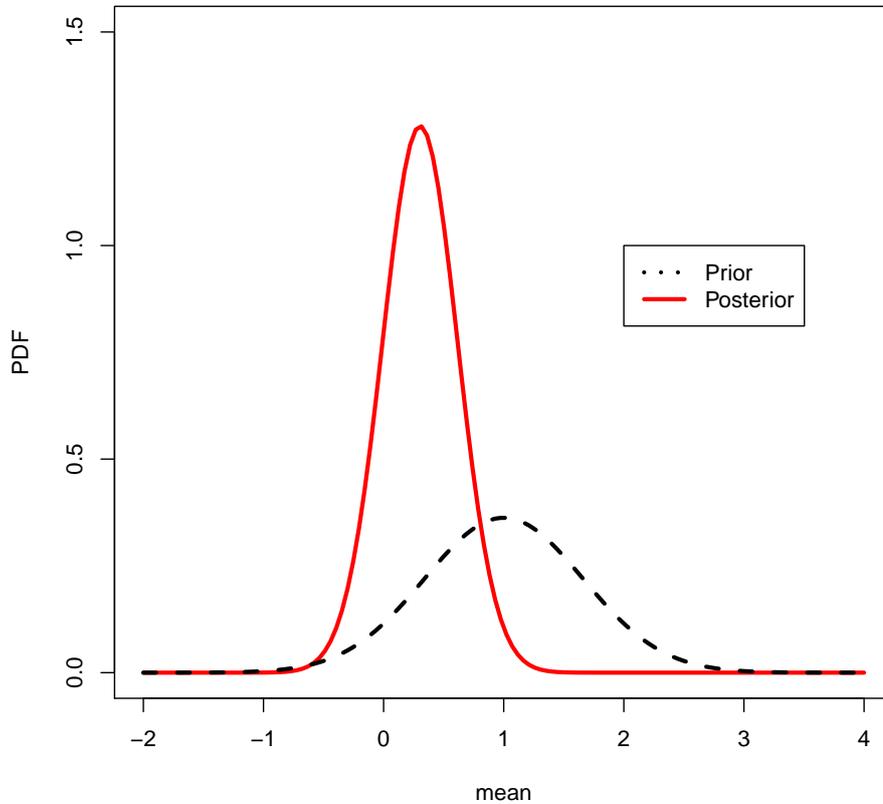


그림 2.3.1. 정규확률분포의 평균

조건부확률변수  $x_j | \sigma^2$ 의 확률밀도함수  $f(x_j | \sigma^2)$ 는 다음과 같다.

$$f(x_j | \sigma^2) = [2\pi\sigma^2]^{-\frac{1}{2}} \exp\left(-\frac{[x_j - \mu]^2}{2\sigma^2}\right) \quad (2.3.13)$$

따라서, 관찰벡터  $\mathbf{x} \doteq [x_1, x_2, \dots, x_n]^T$ 에 대한 모수  $\sigma^2$ 의 우도함수  $L(\sigma^2 | \mathbf{x})$ 는 다음과 같다.

$$L(\sigma^2 | \mathbf{x}) = f(\mathbf{x} | \sigma^2) \propto [\sigma^2]^{-\frac{n}{2}} \exp\left(-\frac{n}{2\sigma^2} \left\{ \frac{1}{n} \sum_{j=1}^n [x_j - \mu]^2 \right\}\right) \quad (2.3.14)$$

이 우도함수를 최대화하는  $\sigma^2$ 의 추정량  $S_1$ 은 다음과 같다.

$$S_1 = \frac{1}{n} \sum_{j=1}^n [x_j - \mu]^2 \quad (2.3.15)$$

이  $S_1$ 은  $\sigma^2$ 에 대한 충분통계량이다.

식 (2.3.12)와 식 (2.3.14)를 Bayes 정리에서 대입하면, 분산  $\sigma^2$ 의 사후확률밀도함수가 다음과 같음을 알 수 있다.

$$\begin{aligned}\pi(\sigma^2 | \mathbf{x}) &\propto \pi(\sigma^2)f(\mathbf{x} | \sigma^2) \\ &= [\sigma^2]^{-n/2} \exp\left(-\frac{nS_1}{2\sigma^2}\right) \frac{\beta^\alpha}{\Gamma(\alpha)} [\sigma^2]^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma^2}\right)\end{aligned}\quad (2.3.16)$$

식 (2.3.16)을 다음과 같이 쓸 수 있다.

$$\pi(\sigma^2 | \mathbf{x}) \propto [\sigma^2]^{-\{\alpha+n/2\}+1} \exp\left(-\frac{1}{\sigma^2} \left[\beta + \frac{nS_1}{2}\right]\right)\quad (2.3.17)$$

즉, 분산  $\sigma^2$ 의 사후확률분포는 다음과 같다.

$$\sigma^2 | \mathbf{x} \stackrel{d}{\sim} \text{IG}\left(\alpha + \frac{n}{2}, \beta + \frac{nS_1}{2}\right)\quad (2.3.18)$$

따라서, 이 정규확률모형의 분산  $\sigma^2$ 에 대한 공액사전확률분포는 역감마확률분포임을 알 수 있다.

**예제 2.3.2** 정규확률분포의 평균을 아는 경우 분산의 사후확률분포를 구하기 위해서, 다음 R 프로그램 Normal\_Variance101.R을 실행해 보자.

```

1 # -----
2 # Filename: Normal_Variance101.R
3 # Conjugate Prior: Normal Distributions
4 # Programmed by CBS
5 # -----
6 # Prior
7 ( alpha = 3 ); ( beta = 2.1 )
8 Nsigma2 <- 128
9 sigma2.gd <- seq(0.1,10,length=Nsigma2)
10 priorPDF <- matrix(0,Nsigma2)
11 for (jj in 1:Nsigma2){
12   priorPDF[jj] <- 1/gamma(alpha)*(beta)^(alpha)*
13     sigma2.gd[jj]^(-alpha-1)*
14     exp(-beta/(sigma2.gd[jj])) # Use dgamma
15 }
16
17 # Likelihood
18 ( mux <- 1.2); ( sigmax <- 1.3 ); (nx <- 16)
19 set.seed(3)
20 x <- mux + sigmax*rnorm(nx,0,1)
21 xbar <- mean(x);
22 S1 = (x-mux) %*% (x-mux)/nx
23 likelihood <- dnorm(x, mean = mux, sd = sigmax, log = FALSE)
24
25 # Posterior

```

```

26 ( alphaStar = alpha+nx/2 )
27 ( betaStar <- beta+nx*S1/2 )
28
29 posteriorPDF <- matrix(0,Nsigma2)
30 for (jj in 1:Nsigma2){
31   posteriorPDF[jj] <- 1/gamma(alphaStar)*(betaStar)^(alphaStar)*
32     sigma2.gd[jj]^(-alphaStar-1)*
33     exp(-betaStar/(sigma2.gd[jj])) # Use dgamma
34 }
35
36 # Plotting
37 setEPS()
38 plot.new()
39 par(mfrow=c(1,1))
40 postscript('Normal_Variance101.eps') # Start to save figure
41 plot(sigma2.gd,posteriorPDF,type="l", col="red", xlab="variance",
42       ylab="PDF", ylim=c(0,1.5), lwd=3 )
43 lines(sigma2.gd,priorPDF,lty=2,lwd=3,col=1)
44 legend(2,1,c("Prior","Posterior"),
45        lty=c(3,1),lwd=c(3,3),col=c(1,2) )
46 dev.off() # End to save figure
47 # -----

```

이 R 프로그램 Normal\_Variance101.R에서는 사전확률분포는 역감마확률분포  $IG(3, 2.1)$  이고, 관찰값들  $x_1, x_2, \dots, x_{16}$  을 정규확률분포  $\mathcal{N}(1.2, 1.3^2)$  에서 발생시킨다. 따라서, 사후확률분포는 역감마확률분포  $IG(11, 9.975049)$  이다.

이 R 프로그램 Normal\_Variance101.R을 실행하면, 그림 2.3.2이 그려진다. 그림 2.3.2에서 흑색 긴점선은 사전확률밀도함수를, 그리고 적색 실선은 첫 번째 사후확률밀도함수를 나타낸다. ■

### 2.3.3 정규확률분포의 평균과 분산

정규확률분포  $\mathcal{N}(\mu, \sigma^2)$  를 따르는 확률변수들  $x_1, x_2, \dots, x_n$  이 서로 독립이라고 가정하고, 평균  $\mu$  와 분산  $\sigma^2$  를 모두 추정하기로 하자.

추정해야 할 모수들은 평균  $\mu$  와 분산  $\sigma^2$  이므로, 사전확률밀도함수는  $\pi(\mu, \sigma^2) = \pi(\sigma^2)\pi(\mu | \sigma^2)$  이다. 분산  $\sigma^2$  의 사전확률분포는 역감마확률분포이고,  $\sigma^2$  가 주어진 조건 하에서 평균  $\mu$  의 조건부확률분포가 정규확률분포라고 하자. 즉, 다음 식들이 성립한다고

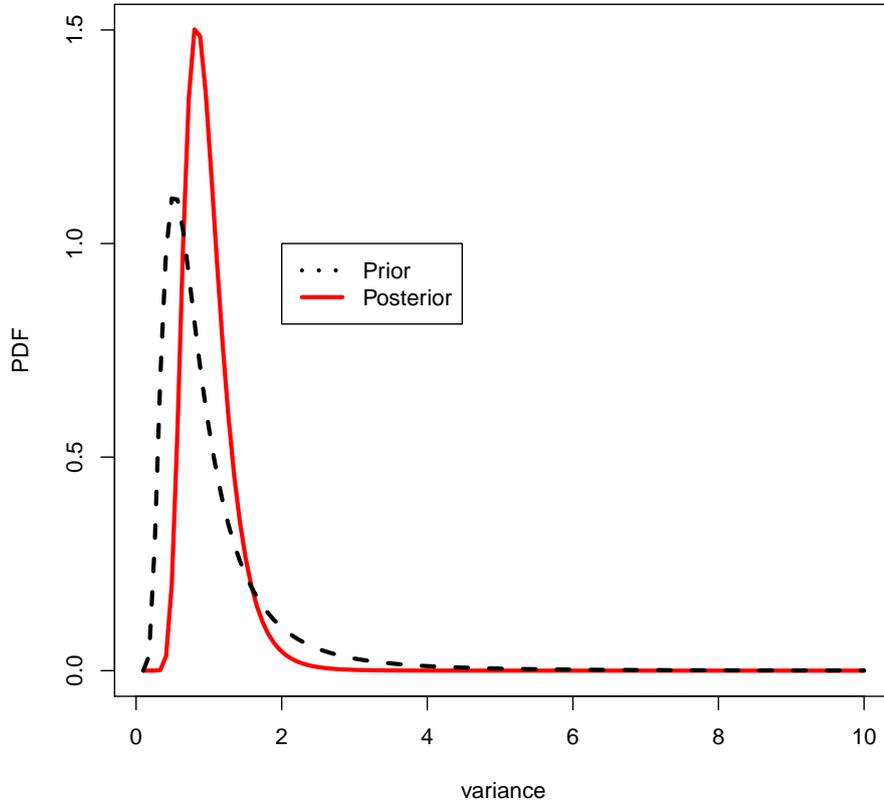


그림 2.3.2. 정규확률분포의 분산

가정하자.

$$\sigma^2 \stackrel{d}{\sim} \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\nu_0 S_0}{2}\right) \quad (2.3.19)$$

$$\mu \mid \sigma^2 \stackrel{d}{\sim} \mathcal{N}\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right) \quad (2.3.20)$$

만약  $\nu_0$ 가 자연수이면,  $\nu_0 S_0 / \sigma^2$ 는 자유도가  $\nu_0$ 인 카이제곱확률분포를 따르고,  $\mu$ 의 주변 확률분포는 자유도가  $\nu_0$ 인  $t$  확률분포를 따른다. 앞서서도 언급했듯이, 역감마확률분포  $\mathcal{IG}(\nu_0/2, \nu_0 S_0/2)$ 를 자유도가  $\nu_0$ 인 척도화역카이제곱확률분포 (scaled inverse- $\chi^2$  distribution)라고 부르며  $Inv - \chi^2(\nu_0, S_0)$ 로 표기한다. 식 (2.3.19)에서 알 수 있듯이, 분산  $\sigma^2$ 의 사전확률밀도함수  $\pi(\sigma^2)$ 는 다음과 같다.

$$\pi(\sigma^2) = \frac{1}{\Gamma(\frac{\nu_0}{2})} \left[ \frac{\nu_0 S_0}{2} \right]^{\nu_0/2} [\sigma^2]^{-\nu_0/2-1} \exp\left(-\frac{1}{\sigma^2} \frac{\nu_0 S_0}{2}\right) \quad (2.3.21)$$

식 (2.3.20)과 식 (2.3.21)에서 알 수 있듯이, 사전결합확률밀도함수는 다음과 같다.

$$\pi(\mu, \sigma^2) = \pi(\sigma^2)\pi(\mu | \sigma^2) \propto \frac{1}{\sigma} \left[ \frac{1}{\sigma^2} \right]^{\frac{\nu_0}{2}+1} \exp \left( -\frac{1}{2\sigma^2} \left\{ \nu_0 S_0 + \kappa_0 [\mu - \mu_0]^2 \right\} \right) \quad (2.3.22)$$

이에 해당하는 사전결합확률분포를  $\mathcal{N} - \mathcal{IG}(\mu_0, \kappa_0; \frac{\nu_0}{2}, \frac{\nu_0 S_0}{2})$  또는  $\mathcal{N} - \text{Inv } \chi^2(\mu_0, \kappa_0; \nu_0, S_0)$ 로 표기하자. 여기서 모수들은 각각 평균  $\mu$ 의 위치와 척도 그리고 분산  $\sigma^2$ 의 자유도와 척도를 나타낸다.

조건부확률변수  $x_j | (\mu, \sigma^2)$ 의 확률밀도함수  $f(x_j | \mu, \sigma^2)$ 는 다음과 같다.

$$f(x_j | \mu, \sigma^2) = [2\pi\sigma^2]^{-\frac{1}{2}} \exp \left( -\frac{[x_j - \mu]^2}{2\sigma^2} \right) \quad (2.3.23)$$

따라서, 관찰벡터  $\mathbf{x} \doteq [x_1, x_2, \dots, x_n]^T$ 에 대한  $[\mu, \sigma^2]$ 의 우도함수  $L(\mu, \sigma^2 | \mathbf{x}) = f(\mathbf{x} | \mu, \sigma^2)$ 는 다음과 같다.

$$L(\mu, \sigma^2 | \mathbf{x}) = [2\pi\sigma^2]^{-n/2} \exp \left( -\frac{1}{2\sigma^2} \sum_{j=1}^n [x_j - \mu]^2 \right) \quad (2.3.24)$$

즉, 다음 식이 성립한다.

$$f(\mathbf{x} | \mu, \sigma^2) \propto [\sigma^2]^{-\frac{n}{2}} \exp \left( -\frac{1}{2\sigma^2} \left\{ [n-1]s^2 + n[\bar{x} - \mu]^2 \right\} \right) \quad (2.3.25)$$

여기서  $\bar{x}$ 와  $s^2$ 는 각각 다음과 같다.

$$\bar{x} \doteq \frac{1}{n} \sum_{i=1}^n x_i, \quad s^2 \doteq \frac{1}{n-1} \sum_{i=1}^n [x_i - \bar{x}]^2 \quad (2.3.26)$$

평균  $\mu$ 와 분산  $\sigma^2$ 의 사전결합확률분포가 식 (2.3.22)에 해당하는  $\mathcal{N} - \mathcal{IG}(\mu_0, \kappa_0; \frac{\nu_0}{2}, \frac{\nu_0 S_0}{2})$ 이고 또한 우도함수가 식 (2.3.25)에 해당하는 다변량정규확률분포를 따르므로, Bayes정리에 서 알 수 있듯이 평균  $\mu$ 와 분산  $\sigma^2$ 의 사후결합확률밀도함수는 다음과 같다.

$$\pi(\mu, \sigma^2 | \mathbf{x}) \propto \frac{1}{\sigma} [\sigma^2]^{-\nu_*/2-1} \exp \left( -\frac{1}{2\sigma^2} \left\{ \nu_* S_* + \kappa_* [\mu - \mu_*]^2 \right\} \right) \quad (2.3.27)$$

여기서 각 모수는 다음과 같다.

$$\mu_* \doteq \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{x} \quad (2.3.28)$$

$$\kappa_* \doteq \kappa_0 + n, \quad \nu_* \doteq \nu_0 + n \quad (2.3.29)$$

$$\nu_* S_* \doteq \nu_0 S_0 + [n - 1]s^2 + \frac{\kappa_0 n}{\kappa_0 + n} [\bar{x} - \mu_0]^2 \quad (2.3.30)$$

식 (2.3.27)에서 알 수 있듯이, 사후결합확률분포는 다음과 같다.

$$(\mu, \sigma^2) \mid \mathbf{x} \stackrel{d}{\sim} \mathcal{N} - Inv \chi^2(\mu_*, \kappa_*; \nu_*, S_*) \quad (2.3.31)$$

식 (2.3.28)에서 알 수 있듯이, 사후평균  $\mu_*$ 는 사전평균  $\mu_0$ 와 표본평균  $\bar{x}$ 의 가중평균이고, 가중값은 정밀도들에 의해서 결정된다. 식 (2.3.29)에서 알 수 있듯이, 사후자유도  $\nu_*$ 는 사전자유도  $\nu_0$ 에 표본의 크기  $n$ 을 더한 것이다. 식 (2.3.30)에서 알 수 있듯이,  $\nu_* S_*$ 는 사후제곱합으로서 사전제곱합  $\nu_0 S_0$ , 표본오차제곱합  $[n - 1]s^2$ , 그리고 표본평균  $\bar{x}$ 와 사전평균  $\mu_0$ 의 차이에서 발생하는 불확실성을 나타내는  $\frac{\kappa_0 n}{\kappa_0 + n} [\bar{x} - \mu_0]^2$ 을 더한 것이다.

분산  $\sigma^2$ 가 주어진 조건 하에서  $\mu$ 의 사후조건부확률밀도함수  $\pi(\mu \mid \sigma^2, \mathbf{x})$ 는 분산  $\sigma^2$ 를 상수로 하는 사후결합확률밀도함수에 비례한다. 식 (2.3.9)에서 알 수 있듯이, 이에 해당하는 사후확률분포는 다음과 같다.

$$\mu \mid (\sigma^2, \mathbf{x}) \stackrel{d}{\sim} \mathcal{N}\left(\mu_*, \frac{\sigma^2}{\kappa_*}\right) \quad (2.3.32)$$

식 (2.3.31)에서 알 수 있듯이,  $\sigma^2$ 의 사후주변확률분포  $\pi(\sigma^2 \mid \mathbf{x})$ 는 다음과 같이 자유도가  $\nu_*$ 인 척도화역카이제곱확률분포이다.

$$\sigma^2 \mid \mathbf{x} \stackrel{d}{\sim} Inv \chi^2(\nu_*, S_*) \quad (2.3.33)$$

식 (2.3.31)의 사후결합확률밀도함수를  $\sigma^2$ 에 관해 적분하면,  $\mu$ 에 대한 사후주변확률밀도함수가 다음 식을 만족함을 알 수 있다.

$$\pi(\mu \mid \mathbf{x}) \propto \left\{ 1 + \frac{1}{\nu_*} \left[ \frac{\mu - \mu_*}{\sqrt{S_*/\kappa_*}} \right]^2 \right\}^{-[\nu_*+1]/2} \quad (2.3.34)$$

즉,  $\frac{\mu - \mu_*}{\sqrt{S_*/\kappa_*}}$  의 사후주변확률분포는 자유도가  $\nu_*$  인  $t$  확률분포이다. 즉, 다음 식이 성립한다.

$$\mu \mid \mathbf{x} \stackrel{d}{\sim} t_{\nu_*} \left( \mu \mid \mu_*, \frac{S_*}{\kappa_*} \right) \quad (2.3.35)$$

**예제 2.3.3** 정규확률분포의 평균과 분산에 대한 베이저안통계분석을 하기 위해서, 다음 R 프로그램 Normal\_MeanVariance101.R을 실행해 보자.

```

1 # -----
2 # Filename: Normal_MeanVariance101.R
3 # Conjugate Prior: Normal Distributions
4 # Programmed by CBS
5 # -----
6 setEPS()
7 # Prior
8 ( mu0 <- 0); ( kappa0 <- 1 )
9 ( nu0 <- 1.1); ( S0 <- 2.4^2 )
10 ( shape0 <- nu0/2 );
11 ( rate0 <- nu0*S0/2 );
12 Nmu <- 128; Nsigma2 <- 128
13 mu.gd <- seq(-4,4,length=Nmu)
14 sigma2.gd <- seq(0.1,15,length=Nsigma2)
15 priorPDF <- matrix(0,Nmu,Nsigma2)
16 for (jj in 1:Nsigma2){
17   prior.sigma2 <- 1/gamma(shape0)*(rate0)^(shape0)*
18     sigma2.gd[jj]^(-shape0-1)*
19     exp(-rate0/sigma2.gd[jj])
20     # dgamma(x, shape, rate = 1, scale = 1/rate, log = FALSE)
21   for (ii in 1:Nmu){
22     prior.mu <- dnorm(mu.gd[ii],mean=mu0,
23       sd=sqrt(sigma2.gd[jj]/kappa0),log = FALSE)
24     priorPDF[ii,jj] <- prior.sigma2*prior.mu
25   }
26 }
27
28 # Contour Plotting of Prior
29 plot.new()
30 postscript('Normal_MeanVariance101.eps') # Start to save figure
31 par(mfrow=c(2,2))
32 contour(mu.gd,sigma2.gd,priorPDF,nlev=7,lty=1,lwd=3,col="blue",
33   xlab=expression(mu),ylab=expression(sigma^2),main="Prior Contour")
34
35
36 # Likelihood
37 ( mux <- 1.2); ( sigmax <- 1.3 ); ( nx <- 16)
38 set.seed(3)
39 x <- mux + sigmax*rnorm(nx,0,1)
40 xbar <- mean(x); s2 = var(x)
41 likelihood <- dnorm(x, mean = mux, sd = sigmax, log = FALSE)
42
43 # Posterior
44 ( mu.gd <- seq(0,2,length=Nmu) )
45 ( sigma2.gd <- seq(0,4,length=Nsigma2) )
46 ( muStar <- (kappa0*mu0+nx*xbar)/(kappa0+nx) );
47 ( kappaStar <- kappa0+nx )
48 ( nuStar <- nu0+nx);

```

```

49 ( SStar <- (nu0*S0+(nx-1)*s2+kappa0*nx/(kappa0+nx)*(xbar-nu0)^2)/nuStar )
50 ( shapeStar <- nuStar/2 );
51 ( rateStar <- nuStar*SStar/2 );
52 posteriorPDF <- matrix(0,Nmu,Nsigma2)
53 for (jj in 1:Nsigma2){
54   posterior.sigma2 <- 1/gamma(shapeStar)*(rateStar)^(shapeStar)*
55     sigma2.gd[jj]^(-shapeStar-1)*
56     exp(-rateStar/sigma2.gd[jj])
57   for (ii in 1:Nmu){
58     posterior.mu <- dnorm(mu.gd[ii],mean=muStar,
59       sd=sqrt(sigma2.gd[jj]/kappaStar),log = FALSE)
60     posteriorPDF[ii,jj] <- posterior.mu*posterior.sigma2
61   }
62 }
63
64 # Monte Carlo Simulations and Contour Plotting of Posterior
65 NoSim <- 1024
66 ( sigma2.Sim <- 1/rgamma(NoSim,shape=shapeStar,rate=rateStar) )
67   # rgamma(n, shape, rate = 1, scale = 1/rate)
68 ( mu.Sim <- rnorm(NoSim,muStar,sqrt(sigma2.Sim/kappaStar)) )
69 plot(mu.Sim,sigma2.Sim,ylab=quote(sigma^{2}),xlab=quote(mu),col="red",
70   xlim=c(0.0,2.0),ylim=c(0,6))
71 contour(mu.gd,sigma2.gd,posteriorPDF,nlev=7,lty=1,lwd=3,add=TRUE,
72   main="Posterior Contour")
73 hist(mu.Sim,xlab="mean",border="blue",col="green",breaks=12)
74 hist(sigma2.Sim,,xlab="variance",border="blue",col="pink",breaks=12)
75 dev.off() # End to save figure
76 # -----

```

이 R프로그램 Normal\_MeanVariance101.R에서 가정한 사전확률분포는  $\mathcal{N}-IG\left(0, -1; \frac{1.1}{2}, \frac{1.1 \times 2.4^2}{2}\right)$  이고, 관찰값들  $x_1, x_2, \dots, x_{16}$  을 정규확률분포  $\mathcal{N}(1.2, 1.3^2)$  에서 발생시킨다. 따라서, 사후 확률분포는 다음과 같다.

$$\mu | \mathbf{x} \stackrel{d}{\sim} \mathcal{N}-IG\left(0.9358595, 21; \frac{21.1}{2}, \frac{21.1 \times 1.232237}{2}\right) \quad (1)$$

이 R프로그램 Normal\_MeanVariance101.R을 실행하면, 그림 2.3.3이 그려진다. 그림 2.3.3의 좌측상단 그래프에서 청색 실선은 사전확률밀도함수의 등가곡선들(contours)을, 우측상단 그래프에서 적색 점들은 관찰점들을 그리고 흑색 실선은 사후확률밀도함수의 등가곡선들을, 좌측하단 그래프는 평균 관찰값들의 히스토그램을, 그리고 우측하단 그래프는 분산 관찰값들의 히스토그램을 보여준다. ■

## 제2.4절 다변량정규확률분포와 베이esian통계학

이 절에서는 다변량정규확률분포의 베이esian통계분석에 대해서 살펴보자.

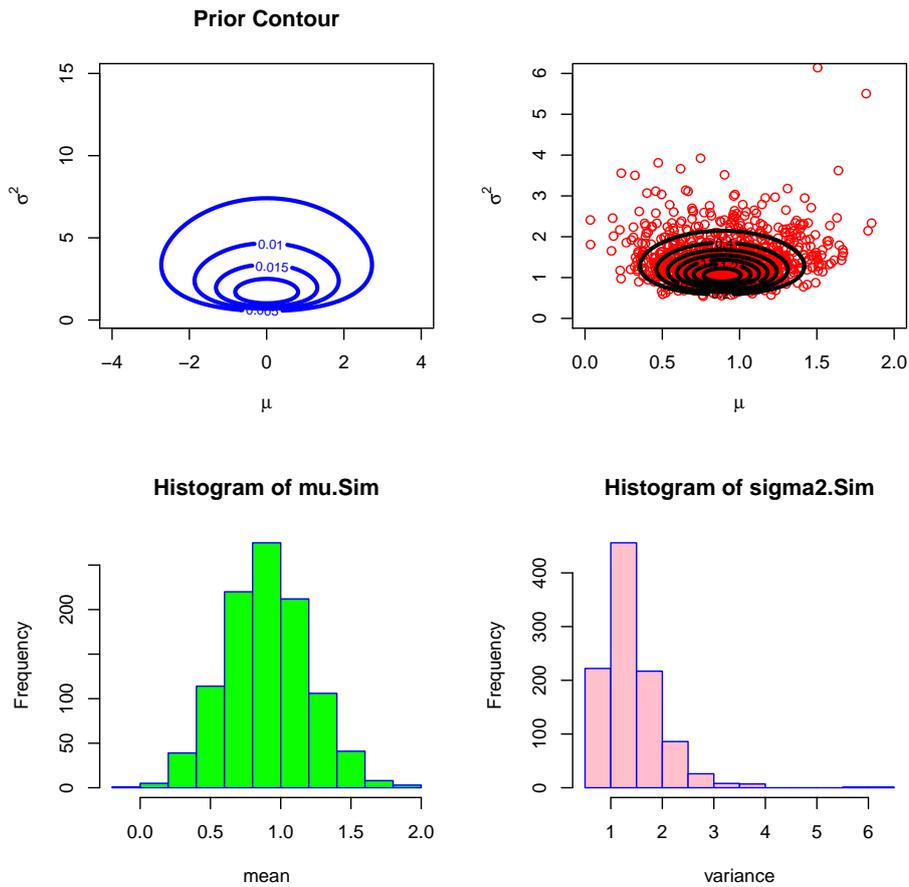


그림 2.3.3. 정규확률분포의 평균과 분산

### 2.4.1 다변량정규확률분포

이 소절에서는 다변량정규확률분포에 관한 몇 가지 재미있는 문제들을 살펴보자.

**예제 2.4.1** 다변량정규확률분포를 사용해서 다음과 같은  $n$ 차원 초구(hyper-ball)의 부피를 구해보자.

$$V_n(a) = \int \cdots \int_{\sum_{k=1}^n x_k^2 \leq a^2} dx_1 \cdots dx_n \tag{1}$$

첫째,  $n = 2$ 인 경우에는 다음 식들이 성립한다.

$$V_2(a) = \iint_{x_1^2 + x_2^2 \leq a^2} dx_1 dx_2 = \int_0^a 2\pi r dr = \pi a^2 \tag{2}$$

둘째,  $n = 3$ 인 경우에는 다음 식들이 성립한다.

$$V_3(a) = \iiint_{x_1^2+x_2^2+x_3^2 \leq a^2} dx_1 dx_2 dx_3 = \int_0^a 4\pi r^2 dr = \frac{4\pi}{3} a^3 \quad (3)$$

셋째, 자연수  $n$ 에 대해서 다음 함수를 정의하자.

$$f(x_1, \dots, x_n) = \prod_{k=1}^n \exp\left(-\frac{1}{2}x_k^2\right) \quad (4)$$

다음 식들이 성립한다.

$$\int_{\mathbb{R}^n} f(x_1, \dots, x_n) dV = \prod_{k=1}^n \left[ \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}x_k^2\right) dx_k \right] = [2\pi]^{n/2} \quad (5)$$

여기서  $dV$ 는  $n$ 차원 부피요소(volume element)이다. 함수  $f$ 는 원점을 중심으로 하는 회전성에 대해 불변(invariance)이다. 따라서, 적분  $\int_{\mathbb{R}^n} f dV$ 를 다음과 같이 쓸 수 있다.

$$\int_{\mathbb{R}^n} f(x_1, \dots, x_n) dV = \int_0^{\infty} \int_{S^{n-1}(r)} \exp\left(-\frac{1}{2}r^2\right) dA dr \quad (6)$$

여기서  $S^{n-1}(r)$ 은 반경이  $r$ 인  $[n-1]$ 차원 초구의 표면(sphere)이고,  $dA$ 는 면적요소(area element)이다. 만약 반경이  $r$ 인  $[n-1]$ 차원 초구의 표면적을  $A_{n-1}(r)$ 이라 하면, 다음 식이 성립한다.

$$A_{n-1}(r) = r^{n-1} A_{n-1}(1) \quad (7)$$

식 (6)과 식 (7)에서 알 수 있듯이, 다음 식이 성립한다.

$$\int_{\mathbb{R}^n} f dV = \int_0^{\infty} \exp\left(-\frac{1}{2}r^2\right) r^{n-1} A_{n-1}(1) dr \quad (8)$$

즉,  $r^{n-1} A_{n-1}(1)$ 은 변수변환  $x_n^2 = r^2 - \sum_{k=1}^{n-1} x_k^2$ 에 해당하는 Jacobian이다. 식 (5)와 식 (8)에서 알 수 있듯이, 다음 식이 성립한다.

$$A_{n-1}(1) \int_0^{\infty} \exp\left(-\frac{1}{2}r^2\right) r^{n-1} dr = [2\pi]^{n/2} \quad (9)$$

식 (9)에 변수변환  $z = r^2/2$ 를 적용하면, 다음 식을 얻는다.

$$A_{n-1}(1)2^{[n-2]/2} \int_0^\infty \exp(-z) z^{n/2-1} dz = [2\pi]^{n/2} \quad (10)$$

즉, 다음 식이 성립한다.

$$A_{n-1}(1) = \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)} \quad (11)$$

따라서 우리가 구하고자 하는  $n$ 차원 초구의 부피  $V_n(a)$ 은 다음과 같다.

$$V_n(a) = \int_0^a \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)} r^{n-1} dr \quad (12)$$

즉, 다음 식들이 성립한다.

$$V_n(a) = \frac{2\pi^{n/2}}{n\Gamma\left(\frac{n}{2}\right)} a^n = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)} a^n \quad (13)$$

■

**예제 2.4.2** 다음 행렬  $\Sigma$ 가 분산공분산행렬이기 위한, 즉 양정치이기 위한 필요충분조건을 구하고, 이 조건을 만족하는 물체의 부피를 구하라.

$$\Sigma = \begin{bmatrix} a & b_1 & b_2 & \cdots & b_{n-1} & b_n \\ b_1 & a & 0 & \cdots & 0 & 0 \\ b_2 & 0 & a & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ b_{n-1} & 0 & 0 & \cdots & a & 0 \\ b_n & 0 & 0 & \cdots & 0 & a \end{bmatrix} \quad (1)$$

행렬  $\Sigma$ 가 양정치이기 위한 필요충분조건은 다음과 같다.

$$a > 0, \tag{2}$$

$$a - \begin{bmatrix} b_1 & b_2 & \cdots & b_{n-1} & b_n \end{bmatrix} \begin{bmatrix} a & 0 & \cdots & 0 & 0 \\ 0 & a & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a & 0 \\ 0 & 0 & \cdots & 0 & a \end{bmatrix}^{-1} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} > 0 \tag{3}$$

식 (3)은 Schur 보완 (complement)의 성질이다. 식 (3)을 다음과 같이 쓸 수 있다.

$$\sum_{k=1}^n b_k^2 < a^2 \tag{4}$$

식 (4)는 반경이  $a$ 인 초구를 나타낸다. 예제 2.4.1에서 알 수 있듯이, 이 초구의 부피는  $\frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} a^n$ 이다. ■

**예제 2.4.3** 행렬들  $A_{n \times n}, B_{n \times p}$ 와  $C_{p \times n}$ 에 대해서 다음 식이 성립함을 증명하기로 하자.

$$[A + BC]^{-1} = A^{-1} - A^{-1}B [I + CA^{-1}B]^{-1} CA^{-1} \tag{1}$$

여기서 역행렬들  $A^{-1}$ 와  $[A + BC]^{-1}$ 가 존재한다고 가정하자.

다음 연립방정식을 생각해보자.

$$\begin{bmatrix} A & B \\ C & -I \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \tag{2}$$

연립방정식 (2)의 해는 다음과 같다.

$$\mathbf{y} = C\mathbf{x} \tag{3}$$

$$\mathbf{x} = A^{-1}[\mathbf{b} - B\mathbf{y}] \tag{4}$$

식 (3)과 식 (4)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\mathbf{y} = [I + CA^{-1}B]^{-1} CA^{-1}\mathbf{b} \quad (5)$$

$$\mathbf{x} = A^{-1}\mathbf{b} - A^{-1}B [I + CA^{-1}B]^{-1} CA^{-1}\mathbf{b} \quad (6)$$

$$\mathbf{x} = [A + BC]^{-1}\mathbf{b} \quad (7)$$

식 (6)과 식 (7)에서 알 수 있듯이, 임의의 벡터  $\mathbf{b}$ 에 대해서 다음 식이 성립한다.

$$[A + BC]^{-1}\mathbf{b} = \left\{ A^{-1} - A^{-1}B [I + CA^{-1}B]^{-1} CA^{-1} \right\} \mathbf{b} \quad (8)$$

식 (8)에서 행렬  $\mathbf{b}$  대신에 단위행렬  $\mathbf{e}_k$ , ( $k = 1, 2, \dots, n$ )를 대입하면, 다음 식이 성립함을 알 수 있다.

$$[A + BC]^{-1} = A^{-1} - A^{-1}B [I + CA^{-1}B]^{-1} CA^{-1} \quad (8)$$

■

**예제 2.4.4** 제1.4.4소절에서 다루었던 다변량정규확률분포 문제를 다시 살펴보자. 즉, 다변량정규확률분포  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 를 따르는 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ 이 서로 독립이라고 가정하고, 평균벡터  $\boldsymbol{\mu}$ 와 분산공분산행렬  $\Sigma$ 의 최우추정량을 구해보자.

관찰벡터들  $X \doteq [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ 이 주어졌을 때,  $[\boldsymbol{\mu}, \Sigma]$ 의 우도함수  $L(\boldsymbol{\mu}, \Sigma)$ 는 다음과 같다.

$$L(\boldsymbol{\mu}, \Sigma) = \frac{1}{\sqrt{2\pi}^{Np} |\Sigma|^{N/2}} \exp \left( -\frac{1}{2} \sum_{j=1}^N [\mathbf{x}_j - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{x}_j - \boldsymbol{\mu}] \right) \quad (1)$$

제1.4.4소절에서 알 수 있듯이, 식 (1)을 다음과 같이 쓸 수 있다.

$$L(\boldsymbol{\mu}, \Sigma) = \frac{1}{\sqrt{2\pi}^{Np} |\Sigma|^{N/2}} \exp \left( -\frac{1}{2} N [\bar{\mathbf{x}} - \boldsymbol{\mu}]^T \Sigma^{-1} [\bar{\mathbf{x}} - \boldsymbol{\mu}] \right) \exp \left( -\frac{1}{2} \text{Tr} (A \Sigma^{-1}) \right) \quad (2)$$

여기서  $A$ 는 다음과 같다.

$$A = \sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}] [\mathbf{x}_i - \bar{\mathbf{x}}]^T \quad (3)$$

분산공분산행렬  $\Sigma$ 가 양정치이므로, 식 (2)에서 알 수 있듯이 다음 식이 성립한다.

$$L(\boldsymbol{\mu}, \Sigma) \leq L(\bar{\mathbf{x}}, \Sigma) \quad (4)$$

즉, 평균벡터  $\boldsymbol{\mu}$ 의 최우추정량은 다음과 같다.

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} \quad (5)$$

따라서 평균벡터  $\boldsymbol{\mu}$ 의 최우추정량  $\bar{\mathbf{x}}$ 이 주어졌을 때 대수우도함수는 다음과 같다.

$$l(\bar{\mathbf{x}}, \Sigma) = \ln L(\bar{\mathbf{x}}, \Sigma) = -\frac{Np}{2} \ln(2\pi) - \frac{N}{2} \ln |\Sigma| - \frac{1}{2} \text{Tr}(A\Sigma^{-1}) \quad (6)$$

식 (6)의 양변을 미분하면, 다음 식을 얻는다.

$$\frac{\partial l(\bar{\mathbf{x}}, \Sigma)}{\partial \Sigma} = -\frac{N}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} A \Sigma^{-1} \quad (7)$$

분산공분산행렬  $\Sigma$ 의 최우추정량은 다음과 같다.

$$\hat{\Sigma} = \frac{1}{N} A = \frac{1}{N} \sum_{i=1}^N [\mathbf{x}_i - \bar{\mathbf{x}}] [\mathbf{x}_i - \bar{\mathbf{x}}]^T \quad (8)$$

식 (8)을 좀 더 자세히 유도하기 위해서, 다음 함수를 정의하자.

$$h(G) = -N \ln \det(G) - \text{Tr}(G^{-1}A) \quad (9)$$

함수  $h$ 의 최대값을 구하기 위해서, 행렬  $A$ 를  $EE^T$ 로 분해하고 행렬  $H = E^T G^{-1} E$ 를 정의

하자. 다음 식들이 성립한다.

$$G = EH^{-1}E^T \quad (10)$$

$$\det(G) = \det(H^{-1}) \det(EE^T) = \frac{\det(EE^T)}{\det(H)} \quad (11)$$

$$\text{Tr}(G^{-1}A) = \text{Tr}(G^{-1}EE^T) = \text{Tr}(E^T G^{-1}E) = \text{Tr}(H) \quad (12)$$

따라서 식 (9)를 다음과 같이 쓸 수 있다.

$$h(G) = -N \ln \det(A) + N \ln \det(H) - \text{Tr}(H) \quad (13)$$

대칭행렬  $H$ 의 Cholesky분해를  $H = TT^T$ 라고 하자. 여기서  $T = [t_{ij}]$ 는 하삼각행렬이다. 식 (13)을 다음과 같이 쓸 수 있다.

$$h(G) = -N \ln \det(A) + \sum_{i=1}^p [N \ln t_{ii}^2 - t_{ii}^2] - \sum_{i>j} t_{ij}^2 \quad (14)$$

식 (14)에서 알 수 있듯이, 다음 식이 성립한다.

$$h(G) \leq -N \ln \det(A) + \sum_{i=1}^p [N \ln t_{ii}^2 - t_{ii}^2] \quad (15)$$

여기서 등호는 식  $t_{ij} = 0, (i > j)$ 에서 성립한다. 또한 미분을 사용해서 다음 식이 성립함을 증명할 수 있다.

$$N \ln t_{ii}^2 - t_{ii}^2 \leq N \ln N - N \quad (16)$$

여기서 등호는 식  $t_{ii}^2 = N$ 에서 성립한다. 식 (15)와 식 (16)에서 알 수 있듯이, 다음 식이 성립한다.

$$h(G) \leq -N \ln \det(A) + pN \ln N - pN = h\left(\frac{1}{N}A\right) \quad (17)$$

식 (17)에서 분산공분산행렬  $\Sigma$ 의 최우추정량이  $\hat{\Sigma}$ 임을 확인할 수 있다. ■

**예제 2.4.5** 예제 2.4.4를 다시 생각해보자. 대수우도함수  $l(\mathbf{x}, \Sigma)$ 는 볼록항(convex term)

$-\ln \det(\Sigma)$  을 포함하고 있으므로 대수우도함수  $l(\hat{\boldsymbol{x}}, \Sigma)$  가 오목함수 (concave function) 인지 쉽게 알 수 없다. 이 예제에서는 조건  $\Sigma \preceq 2\hat{\Sigma}$  하에서 대수우도함수  $l(\boldsymbol{\mu}, \Sigma)$  가  $(\boldsymbol{\mu}, \Sigma)$  에 대해서 오목함수임을 증명하기로 하자. 즉, 대수우도함수  $l(\boldsymbol{\mu}, \Sigma)$  가 최적화 2차조건을 만족함을 증명하기로 하자.

우선 주어진 조건 하에서 다음 함수가 오목임을 보이자

$$f(\Sigma) = -\ln \det(\Sigma) - \text{Tr}(\Sigma^{-1}\hat{\Sigma}) \tag{1}$$

함수  $f$  의 그레디언트 (gradient) 와 Hessian 행렬은 각각 다음과 같다.

$$\nabla f(\Sigma) = -\Sigma^{-1} + \Sigma^{-1}\hat{\Sigma}\Sigma^{-1} \tag{2}$$

$$\begin{aligned} \nabla^2 f(\Sigma) [V] &\doteq \left. \frac{d}{dt} \nabla f(\Sigma + tV) \right|_{t=0} \\ &= \Sigma^{-1}V\Sigma^{-1} - \Sigma^{-1}V\Sigma^{-1}\hat{\Sigma}\Sigma^{-1} - \Sigma^{-1}\hat{\Sigma}\Sigma^{-1}V\Sigma^{-1} \end{aligned} \tag{3}$$

각  $V$  에 대해서 다음 식들이 성립한다.

$$\begin{aligned} &\text{Tr}(V\nabla^2 f(\Sigma) [V]) \\ &= \text{Tr}(V\Sigma^{-1}V\Sigma^{-1}) - 2\text{Tr}(V\Sigma^{-1}V\Sigma^{-1}\hat{\Sigma}\Sigma^{-1}) \\ &= \text{Tr}\left(\left[\Sigma^{-1/2}V\Sigma^{-1/2}\right]^2 \left[I - 2\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2}\right]\right) \end{aligned} \tag{4}$$

조건  $\Sigma \preceq 2\hat{\Sigma}$  을 다음과 같이 쓸 수 있다.

$$2\Sigma^{-1/2}\hat{\Sigma}\Sigma^{-1/2} \succeq I \tag{5}$$

식 (4) 와 식 (5) 에서 알 수 있듯이, 모든  $V$  에 대해서 다음 식들이 성립한다.

$$\left. \frac{d^2}{dt^2} f(\Sigma + tV) \right|_{t=0} = \text{Tr}(V\nabla^2 f(\Sigma) [V]) \leq 0 \tag{6}$$

식 (3) 과 식 (6) 에서 알 수 있듯이, 다음 식들이 성립한다.

$$\nabla^2 f(\Sigma) [V] = \left. \frac{d}{dt} \nabla f(\Sigma + tV) \right|_{t=0} \preceq O \tag{7}$$

식 (7)에서 알 수 있듯이, 대수우도함수  $l(\boldsymbol{\mu}, \Sigma)$ 가 최적화 2차조건을 만족한다. ■

### 2.4.2 다변량정규확률분포의 평균벡터

다변량정규확률분포  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 를 따르는 확률변수들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 이 서로 독립이라고 가정하자. 이 소절에서는 분산공분산행렬  $\Sigma \in \mathbb{R}^{p \times p}$ 를 알고 있다고 가정하고 평균벡터  $\boldsymbol{\mu} \in \mathbb{R}^p$ 를 추정하기로 하자.

평균벡터  $\boldsymbol{\mu}$ 의 사전확률분포가 다음과 같다고 가정하자.

$$\boldsymbol{\mu} \stackrel{d}{\sim} \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0) \quad (2.4.1)$$

여기서  $\boldsymbol{\mu}_0 \in \mathbb{R}^p$ 와  $\Sigma_0 \in \mathbb{R}^{p \times p}$ 는 알고있는 벡터와 행렬이다. 다변량정규확률분포  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 의 확률밀도함수를  $n(\cdot | \boldsymbol{\mu}, \Sigma)$ 로 표기하면,  $\boldsymbol{\mu}$ 의 사전확률밀도함수  $\pi(\boldsymbol{\mu})$ 는 다음과 같다.

$$\pi(\boldsymbol{\mu}) = n(\boldsymbol{\mu} | \boldsymbol{\mu}_0, \Sigma_0) = \frac{1}{\sqrt{2\pi}^p |\Sigma_0|^{1/2}} \exp\left(-\frac{1}{2}[\boldsymbol{\mu} - \boldsymbol{\mu}_0]^T \Sigma_0^{-1} [\boldsymbol{\mu} - \boldsymbol{\mu}_0]\right) \quad (2.4.2)$$

관찰벡터  $\mathbf{x} \doteq [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_n^T]^T$ 의 우도함수  $f(\mathbf{x} | \boldsymbol{\mu})$ 는 다음과 같다.

$$\begin{aligned} f(\mathbf{x} | \boldsymbol{\mu}) &= \prod_{j=1}^n n(\mathbf{x}_j | \boldsymbol{\mu}, \Sigma) \\ &= \frac{1}{\sqrt{2\pi}^{np} |\Sigma|^{n/2}} \exp\left(-\frac{1}{2} \sum_{j=1}^n [\mathbf{x}_j - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{x}_j - \boldsymbol{\mu}]\right) \end{aligned} \quad (2.4.3)$$

식 (2.4.2)와 식 (2.4.3)을 Bayes정리에 대입하면, 평균벡터  $\boldsymbol{\mu}$ 의 사후확률밀도함수가 다음과 같음을 알 수 있다.

$$\begin{aligned} \pi(\boldsymbol{\mu} | \mathbf{x}) &\propto \pi(\boldsymbol{\mu}) f(\mathbf{x} | \boldsymbol{\mu}) = \pi(\boldsymbol{\mu}) \prod_{j=1}^n f(\mathbf{x}_j | \boldsymbol{\mu}) \\ &\propto \exp\left(-\frac{1}{2}[\boldsymbol{\mu} - \boldsymbol{\mu}_0]^T \Sigma_0^{-1} [\boldsymbol{\mu} - \boldsymbol{\mu}_0] - \frac{1}{2} \sum_{j=1}^n [\mathbf{x}_j - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{x}_j - \boldsymbol{\mu}]\right) \end{aligned} \quad (2.4.4)$$

다음 식들이 성립한다.

$$\begin{aligned}
 & -\frac{1}{2}[\boldsymbol{\mu} - \boldsymbol{\mu}_0]^T \Sigma_0^{-1} [\boldsymbol{\mu} - \boldsymbol{\mu}_0] - \frac{1}{2} \sum_{j=1}^n [\mathbf{x}_j - \boldsymbol{\mu}]^T \Sigma^{-1} [\mathbf{x}_j - \boldsymbol{\mu}] \\
 &= -\frac{1}{2} \boldsymbol{\mu}^T \Sigma_0^{-1} \boldsymbol{\mu} + \boldsymbol{\mu}^T \Sigma_0^{-1} \boldsymbol{\mu}_0 - \frac{n}{2} \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} + \sum_{j=1}^n \boldsymbol{\mu}^T \Sigma^{-1} \mathbf{x}_j + \text{constant} \\
 &= -\frac{1}{2} \boldsymbol{\mu}^T [\Sigma_0^{-1} + n\Sigma^{-1}] \boldsymbol{\mu} + \boldsymbol{\mu}^T \left[ \Sigma_0^{-1} \boldsymbol{\mu}_0 + \Sigma^{-1} \sum_{j=1}^n \mathbf{x}_j \right] + \text{constant} \\
 &= -\frac{1}{2} \left\{ \boldsymbol{\mu} - [\Sigma_0^{-1} + n\Sigma^{-1}]^{-1} \left[ \Sigma_0^{-1} \boldsymbol{\mu}_0 + \Sigma^{-1} \sum_{j=1}^n \mathbf{x}_j \right] \right\}^T [\Sigma_0^{-1} + n\Sigma^{-1}] \\
 & \quad \left\{ \boldsymbol{\mu} - [\Sigma_0^{-1} + n\Sigma^{-1}]^{-1} \left[ \Sigma_0^{-1} \boldsymbol{\mu}_0 + \Sigma^{-1} \sum_{j=1}^n \mathbf{x}_j \right] \right\} + \text{constant} \tag{2.4.5}
 \end{aligned}$$

다음 식들이 성립한다.

$$[\Sigma_0^{-1} + n\Sigma^{-1}]^{-1} = \frac{1}{n} \Sigma \left[ \Sigma_0 + \frac{1}{n} \Sigma \right]^{-1} \Sigma_0 = \frac{1}{n} \Sigma_0 \left[ \Sigma_0 + \frac{1}{n} \Sigma \right]^{-1} \Sigma \tag{2.4.6}$$

따라서 다음 식이 성립한다.

$$\begin{aligned}
 & [\Sigma_0^{-1} + n\Sigma^{-1}]^{-1} \left[ \Sigma_0^{-1} \boldsymbol{\mu}_0 + \Sigma^{-1} \sum_{j=1}^n \mathbf{x}_j \right] \\
 &= \Sigma \left[ \Sigma_0 + \frac{1}{n} \Sigma \right]^{-1} \frac{1}{n} \boldsymbol{\mu}_0 + \Sigma_0 \left[ \Sigma_0 + \frac{1}{n} \Sigma \right]^{-1} \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j \tag{2.4.7}
 \end{aligned}$$

식 (2.4.7)을 식 (2.4.5)에 대입하면, 식 (2.4.4)를 다음과 같이 쓸 수 있다.

$$\pi(\boldsymbol{\mu} | X) \propto \exp \left( -\frac{1}{2} [\boldsymbol{\mu} - \boldsymbol{\mu}_*]^T \Sigma_*^{-1} [\boldsymbol{\mu} - \boldsymbol{\mu}_*] \right) \tag{2.4.8}$$

여기서  $X, \bar{\mathbf{x}}, \boldsymbol{\mu}_*$  와  $\Sigma_*$  는 각각 다음과 같다.

$$X \doteq [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \quad (2.4.9)$$

$$\bar{\mathbf{x}} \doteq \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j \quad (2.4.10)$$

$$\boldsymbol{\mu}_* \doteq \Sigma_0 \left[ \Sigma_0 + \frac{1}{n} \Sigma \right]^{-1} \bar{\mathbf{x}} + \frac{1}{n} \Sigma \left[ \Sigma_0 + \frac{1}{n} \Sigma \right]^{-1} \boldsymbol{\mu}_0 \quad (2.4.11)$$

$$\Sigma_* \doteq [\Sigma_0^{-1} + n\Sigma^{-1}]^{-1} \quad (2.4.12)$$

식 (2.4.8) 에서 알 수 있듯이, 평균벡터  $\boldsymbol{\mu}$  의 사후확률분포는 다음과 같다.

$$\boldsymbol{\mu} | X \stackrel{d}{\sim} \mathcal{N}_p(\boldsymbol{\mu}_*, \Sigma_*) \quad (2.4.13)$$

지금까지 내용을 다음과 같이 설명할 수 있다.

**첫째,**  $\Sigma_0^{-1}$  를 사전정밀도 (prior precision) 라 부르고,  $n\Sigma^{-1}$  를 데이터정밀도 (data precision) 라 부른다. 이 항들은 각각 사전가정 (prior assumption) 에서 발생하는 불확실성과 관측된 데이터에서 발생하는 불확실성을 나타낸다. 식 (2.4.8) 에서 알 수 있듯이,  $\Sigma_*^{-1}$  가 이 정밀도들을 결합한 사후정밀도 (posterior precision) 이다.

**둘째,** 표본크기  $n$  이 커짐에 따라,  $\boldsymbol{\mu}_*$  는 데이터의 표본평균벡터  $\bar{\mathbf{x}}$  와 분산공분산행렬  $\Sigma$  에 더 많은 영향을 받으며 사전평균벡터  $\boldsymbol{\mu}_0$  는 점차 중요하지 않게 된다. 식 (2.4.11) ~ 식 (2.4.13) 에서 알 수 있듯이, 표본크기  $n$  이 아주 크면 사후확률분포는 근사적으로  $\mathcal{N}(\bar{\mathbf{x}}, \frac{1}{n}\Sigma)$  을 따른다. 이는 고전통계학에서 얻은 결과와 동일하다. 이 결과를 베이저안 중심극한정리라 부른다.

**셋째,** 사전확률분포의 분산공분산행렬  $\det(\Sigma_0)$  가  $\infty$  인 경우에도 사후확률분포는 근사적으로 다변량정규확률분포  $\mathcal{N}(\bar{\mathbf{x}}, \frac{1}{n}\Sigma)$  을 따른다. 즉, 사전확률분포의 분산공분산행렬  $\Sigma_*$  의 행렬식  $\det(\Sigma_*)$  가 커지면, 사전확률분포에 포함되어 있는 정보가 사후확률분포에 미치는 영향은 감소한다.

**넷째,** 사전확률분포와 사후확률분포가 모두 다변량정규확률분포들이다. 즉, 이 사전확률분포는 공액사전확률분포족이다.

### 2.4.3 다변량정규확률분포의 분산공분산행렬

다변량정규확률분포  $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ 를 따르는 확률변수들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 이 서로 독립이라고 가정하자. 여기서는 평균벡터  $\boldsymbol{\mu}$ 는 알고 있다고 가정하고 분산공분산행렬  $\Sigma$ 를 추정하기로 하자. 편의상, 각  $i$ 에 대해서 확률벡터  $\mathbf{y}_i \doteq \mathbf{x}_i - \boldsymbol{\mu}$ 를 정의하자. 확률벡터들  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ 은 서로 독립이고 다음과 같은 다변량정규확률분포를 따른다.

$$\mathbf{y}_j \stackrel{d}{\sim} \mathcal{N}_p(\mathbf{0}, \Sigma), \quad (j = 1, 2, \dots, n) \quad (2.4.14)$$

이 확률벡터들의 결합확률밀도함수는 다음과 같다.

$$f(\mathbf{y}_1, \dots, \mathbf{y}_n | \Sigma) \propto |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} \sum_{j=1}^n \mathbf{y}_j^T \Sigma^{-1} \mathbf{y}_j\right) \quad (2.4.15)$$

다음 식들이 성립한다.

$$\begin{aligned} \sum_{j=1}^n \mathbf{y}_j^T \Sigma^{-1} \mathbf{y}_j &= \text{Tr} \left( \sum_{j=1}^n \mathbf{y}_j \Sigma^{-1} \mathbf{y}_j^T \right) = \sum_{j=1}^n \text{Tr}(\mathbf{y}_j^T \Sigma^{-1} \mathbf{y}_j) \\ &= \sum_{j=1}^n \text{Tr}(\Sigma^{-1} \mathbf{y}_j \mathbf{y}_j^T) = \text{Tr} \left( \Sigma^{-1} \sum_{j=1}^n \mathbf{y}_j \mathbf{y}_j^T \right) = \text{Tr}(\Sigma^{-1} S) \end{aligned} \quad (2.4.16)$$

여기서  $S$ 는 다음과 같다.

$$S \doteq \sum_{j=1}^n \mathbf{y}_j \mathbf{y}_j^T \quad (2.4.17)$$

식 (2.4.17)을 식 (2.4.15)에 대입하면, 결합확률밀도함수가 다음과 같음을 알 수 있다.

$$f(\mathbf{y}_1, \dots, \mathbf{y}_n | \Sigma) \propto |\Sigma|^{-n/2} \exp\left(-\frac{1}{2} \text{Tr}(\Sigma^{-1} S)\right) \quad (2.4.18)$$

분산공분산행렬  $\Sigma$ 의 사전확률분포가 다음과 같다고 하자.

$$\Sigma \stackrel{d}{\sim} \mathcal{IW}_p(\Sigma_0, m_0) \quad (2.4.19)$$

여기서 조건  $m_0 > p$ 가 성립해야 한다. 따라서 분산공분산행렬  $\Sigma$ 의 역행렬인  $\Psi = \Sigma^{-1}$ 의

사전확률분포는 다음과 같다.

$$\Psi \stackrel{d}{\sim} \mathcal{IW}_p(\Sigma_0^{-1}, m_0) \quad (2.4.20)$$

식 (2.4.19)에서 알 수 있듯이, 분산공분산행렬  $\Sigma$ 의 사전확률밀도함수는 다음과 같다.

$$\pi(\Sigma) \propto |\Sigma|^{-[m_0+p+1]/2} \exp\left(-\frac{1}{2}\text{Tr}(\Sigma_0\Sigma^{-1})\right) \quad (2.4.21)$$

식 (2.4.18)과 식 (2.4.21)에서 알 수 있듯이,  $\Sigma$ 의 사후확률밀도함수는 다음과 같다.

$$\pi(\Sigma | \mathbf{y}_1, \dots, \mathbf{y}_n) \quad (2.4.22)$$

$$\propto \Sigma^{-[m_0+p+1]/2} \exp\left(-\frac{1}{2}\text{Tr}(\Sigma_0\Sigma^{-1})\right) \Sigma^{-n/2} \exp\left(-\frac{1}{2}\text{Tr}(\Sigma^{-1}S)\right) \quad (2.4.23)$$

$$= |\Sigma|^{-[m_0+p+n+1]/2} \exp\left(-\frac{1}{2}\text{Tr}((\Sigma_0 + S)\Sigma^{-1})\right) \quad (2.4.24)$$

따라서, 다음 식들이 성립한다.

$$\Sigma | (\mathbf{x}_1, \dots, \mathbf{x}_n) \stackrel{d}{\sim} \mathcal{IW}_p(\Sigma_0 + S, m_0 + n) \quad (2.4.25)$$

$$\Psi | (\mathbf{x}_1, \dots, \mathbf{x}_n) \stackrel{d}{\sim} \mathcal{W}_p((\Sigma_0 + S)^{-1}, m_0 + n) \quad (2.4.26)$$

즉, 분산공분산행렬  $\Sigma$ 의 사후확률분포는 역Wishart 확률분포이고, 정밀도행렬  $\Psi = \Sigma^{-1}$ 의 사후확률분포는 Wishart 확률분포이다. 또한, 다음 식들이 성립한다.

$$E(\Psi | \mathbf{x}_1, \dots, \mathbf{x}_n) = E(\Sigma^{-1} | \mathbf{x}_1, \dots, \mathbf{x}_n) = [m_0 + n](\Sigma_0 + S)^{-1} \quad (2.4.27)$$

#### 2.4.4 다변량정규확률분포의 평균벡터와 분산공분산행렬

이 소절에서는 다변량정규확률분포  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 를 따르는 확률변수들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ 이 서로 독립이라고 가정하고, 평균벡터  $\boldsymbol{\mu}$ 와 분산공분산행렬  $\Sigma$ 를 모두 추정하기로 하자. 우선 다음 명제를 살펴보자.

##### 명제 2.4.1

만약  $A \stackrel{d}{\sim} \mathcal{W}_p(\Sigma, n)$  이고, 또한 만약  $\Sigma \stackrel{d}{\sim} \mathcal{IW}_p(\Psi, m)$  이면,  $A$ 가 주어졌을 때  $\Sigma$ 의 조건부확률분포는  $\mathcal{IW}_p(A + \Psi, n + m)$ 이다.

증명. 확률행렬들  $A$ 와  $\Sigma$ 의 결합확률밀도함수는 다음과 같다.

$$\begin{aligned} \pi(A, \Sigma) &= \pi(A | \Sigma) \pi(\Sigma) \\ &= \frac{1}{C_{A, \Sigma}} |A|^{[n-p-1]/2} |\Sigma|^{-[n+m+p+1]/2} \exp\left(-\frac{1}{2} \text{Tr}([A + \Psi]\Sigma^{-1})\right) \end{aligned} \quad (1)$$

여기서  $A$ 와  $\Sigma$ 는 양정치행렬들이고  $C_{A, \Sigma}$ 는 다음과 같다.

$$C_{A, \Sigma} \doteq 2^{[n+m]p/2} |\Psi|^{-m/2} \Gamma_p\left(\frac{n}{2}\right) \Gamma_p\left(\frac{m}{2}\right) \quad (2)$$

따라서 다음 식들이 성립한다.

$$\begin{aligned} \pi(A) &= \int_{\Sigma} \pi(A, \Sigma) d\Sigma \\ &= \frac{1}{C_{A, \Sigma}} |A|^{[n-p-1]/2} \int_{\Sigma} |\Sigma|^{-[n+m+p+1]/2} \exp\left(-\frac{1}{2} \text{Tr}(\Sigma^{-1}[A + \Psi])\right) d\Sigma \end{aligned} \quad (3)$$

역Wishart 확률밀도함수로부터, 다음 식이 성립함을 알 수 있다.

$$\int_V |V|^{-[m+p+1]/2} \exp\left(-\frac{1}{2} \text{Tr}(\Sigma^{-1}V^{-1})\right) dV = 2^{mp/2} |\Sigma|^{m/2} \Gamma_p\left(\frac{m}{2}\right) \quad (4)$$

식 (4)를 이용해서, 식 (3)을 다음과 같이 쓸 수 있다.

$$\pi(A) = 2^{[n+m]p/2} |A + \Psi|^{-[n+m]/2} \Gamma_p\left(\frac{n+m}{2}\right) \frac{1}{C_{A, \Sigma}} |A|^{[n-p-1]/2} \quad (5)$$

식 (2)와 식 (5)에서 알 수 있듯이, 다음 식이 성립한다.

$$\pi(A) = \frac{\Gamma_p\left(\frac{n+m}{2}\right)}{\Gamma_p\left(\frac{n}{2}\right) \Gamma_p\left(\frac{m}{2}\right)} |\Psi|^{m/2} |A|^{[n-p-1]/2} |A + \Psi|^{-[n+m]/2} \quad (6)$$

식 (1)과 식 (6)에서 알 수 있듯이,  $A$ 가 주어진 조건 하에서  $\Sigma$ 의 조건부확률밀도함수는 다음과

같다.

$$\begin{aligned}
 \pi(\Sigma|A) &= \frac{\pi(A, \Sigma)}{\pi(A)} \\
 &= \frac{\frac{1}{C_{A, \Sigma}} |A|^{[n-p-1]/2} |\Sigma|^{-[n+m+p+1]/2} \exp\left(-\frac{1}{2} \text{Tr}([A + \Psi]\Sigma^{-1})\right)}{\frac{\Gamma_p\left(\frac{n+m}{2}\right)}{\Gamma_p\left(\frac{n}{2}\right)\Gamma_p\left(\frac{m}{2}\right)} |\Psi|^{m/2} |A|^{[n-p-1]/2} |A + \Psi|^{-[n+m]/2}} \\
 &= \frac{1}{2^{[n+m]p/2} \Gamma_p\left(\frac{n+m}{2}\right)} |A + \Psi|^{[n+m]/2} |\Sigma|^{-[n+m+p+1]/2} \exp\left(-\frac{1}{2} \text{Tr}([A + \Psi]\Sigma^{-1})\right) \quad (7)
 \end{aligned}$$

식 (7)은 확률분포  $\mathcal{IW}_p(A + \Psi, n + m)$ 의 확률밀도함수이다. ■

**따름정리 2.4.1**

만약  $nS \stackrel{d}{\sim} \mathcal{W}_p(\Sigma, n)$ 이고,  $\Sigma \stackrel{d}{\sim} \mathcal{IW}_p(\Psi, m)$ 이면,  $S$ 가 주어졌을 때  $\Sigma$ 의 조건부확률 분포는  $\mathcal{IW}_p(nS + \Psi, n + m)$ 이다.

분산공분산행렬  $\Sigma$ 의 베이저안추정량은  $E(\Sigma|S)$ 이다. 또한 식 (2.4.26)과 식 (2.4.27)에서 알 수 있듯이,  $\Psi = \Sigma^{-1}$ 의 사후확률분포는 다음과 같다.

$$\Sigma^{-1} | S \stackrel{d}{\sim} \mathcal{W}_p([nS + \Psi]^{-1}, n + m) \quad (2.4.28)$$

식 (2.4.28)에서 알 수 있듯이, 다음 명제가 성립한다.

**명제 2.4.2**

행렬  $G$ 와 행렬  $H$ 가 양정치라고 하자. 만약  $nS \stackrel{d}{\sim} \mathcal{W}_p(\Sigma, n)$ 이고,  $\Sigma \stackrel{d}{\sim} \mathcal{IW}_p(\Psi, m)$ 이고, 또한 손실함수(loss function)가  $\text{Tr}([D - \Sigma]G[D - \Sigma]H)$ 이면,  $\Sigma$ 의 Bayes 추정행렬은 다음과 같다.

$$\frac{1}{n + m - p - 1} [nS + \Psi]$$

다변량정규확률분포  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ 를 따르는 확률변수들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ 이 서로 독립이라고 가정하자. 지금부터 평균벡터  $\boldsymbol{\mu} \in \mathbb{R}^p$ 와 분산공분산행렬  $\Sigma \in \mathbb{R}^{p \times p}$ 를 추정하기로 하자.

평균벡터  $\boldsymbol{\mu}$ 와 분산공분산행렬  $\Sigma$ 의 사전확률분포가 다음과 같다고 가정하자.

$$\boldsymbol{\mu} | \Sigma \stackrel{d}{\sim} \mathcal{N}_p \left( \boldsymbol{\mu}_0, \frac{1}{K} \Sigma \right) \quad (2.4.29)$$

$$\Sigma \stackrel{d}{\sim} \mathcal{IW}_p(\Sigma | \Psi, m) \quad (2.4.30)$$

여기서  $\boldsymbol{\mu}_0 \in \mathbb{R}^p$ 와  $\Psi \in \mathbb{R}^{p \times p}$ 는 알려진 벡터와 행렬이고,  $K$ 와  $m$ 은 상수들이다. 평균벡터  $\boldsymbol{\mu}$ 의 추정량과 분산공분산행렬의 추정량  $S$ 를 각각 다음과 같이 표기하자.

$$\bar{\boldsymbol{x}} = \frac{1}{N} \sum_{i=1}^N \boldsymbol{x}_i \quad (2.4.31)$$

$$S = \frac{1}{N-1} \sum_{i=1}^N [\boldsymbol{x}_i - \bar{\boldsymbol{x}}][\boldsymbol{x}_i - \bar{\boldsymbol{x}}]^T \quad (2.4.32)$$

평균벡터  $\boldsymbol{\mu} | \Sigma$ 의 사전조건부확률밀도함수  $\pi(\boldsymbol{\mu} | \Sigma)$ 는 다음과 같다.

$$\pi(\boldsymbol{\mu} | \Sigma) = n_p \left( \boldsymbol{\mu} \mid \boldsymbol{\mu}_0, \frac{1}{K} \Sigma \right) = \frac{K^{p/2}}{\sqrt{2\pi^p} |\Sigma|^{1/2}} \exp \left( -\frac{K}{2} [\boldsymbol{\mu} - \boldsymbol{\mu}_0]^T \Sigma^{-1} [\boldsymbol{\mu} - \boldsymbol{\mu}_0] \right) \quad (2.4.33)$$

사전분산공분산행렬  $\Sigma$ 의 확률밀도함수는 다음과 같다.

$$\pi(\Sigma) = \frac{|\Psi|^{m/2}}{2^{mp/2} \Gamma_p \left( \frac{m}{2} \right)} |\Sigma|^{-[m+p+1]/2} \exp \left( -\frac{1}{2} \text{Tr} (\Sigma^{-1} \Psi) \right) \quad (2.4.34)$$

관찰벡터들  $X \doteq [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N]$ 의 우도함수  $f(X | \boldsymbol{\mu}, \Sigma)$ 는 다음과 같다.

$$f(X | \boldsymbol{\mu}, \Sigma) = \frac{1}{\sqrt{2\pi}^{Np} |\Sigma|^{N/2}} \exp \left( -\frac{1}{2} \sum_{j=1}^N [\boldsymbol{x}_j - \boldsymbol{\mu}]^T \Sigma^{-1} [\boldsymbol{x}_j - \boldsymbol{\mu}] \right) \quad (2.4.35)$$

다음 식들이 성립한다.

$$\begin{aligned} & \sum_{j=1}^N [\boldsymbol{x}_j - \boldsymbol{\mu}]^T \Sigma^{-1} [\boldsymbol{x}_j - \boldsymbol{\mu}] \\ &= \sum_{j=1}^N [\boldsymbol{x}_j - \bar{\boldsymbol{x}}]^T \Sigma^{-1} [\boldsymbol{x}_j - \bar{\boldsymbol{x}}] + N[\bar{\boldsymbol{x}} - \boldsymbol{\mu}]^T \Sigma^{-1} [\bar{\boldsymbol{x}} - \boldsymbol{\mu}] \\ &= \text{Tr} (A \Sigma^{-1}) + N[\bar{\boldsymbol{x}} - \boldsymbol{\mu}]^T \Sigma^{-1} [\bar{\boldsymbol{x}} - \boldsymbol{\mu}] \end{aligned} \quad (2.4.36)$$

여기서  $A$ 는 다음과 같은 확률행렬이다.

$$A \doteq [N-1]S = \sum_{j=1}^N [\mathbf{x}_j - \bar{\mathbf{x}}][\mathbf{x}_j - \bar{\mathbf{x}}]^T \quad (2.4.37)$$

식 (2.4.35)를 다음과 같이 쓸 수 있다.

$$\begin{aligned} f(X | \boldsymbol{\mu}, \Sigma) &= \frac{1}{\sqrt{2\pi}^p \left|\frac{1}{N}\Sigma\right|^{1/2}} \exp\left(-\frac{1}{2}N[\bar{\mathbf{x}} - \boldsymbol{\mu}]\Sigma^{-1}[\bar{\mathbf{x}} - \boldsymbol{\mu}]^T\right) \\ &\cdot \frac{1}{N^{p/2}\sqrt{2\pi}^{[N-1]p} |\Sigma|^{[N-1]/2}} \exp\left(-\frac{1}{2}\text{Tr}(A\Sigma^{-1})\right) \end{aligned} \quad (2.4.38)$$

식 (2.4.38)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\bar{\mathbf{x}} | (\boldsymbol{\mu}, \Sigma) \stackrel{d}{\sim} \mathcal{N}_p\left(\boldsymbol{\mu}, \frac{1}{N}\Sigma\right) \quad (2.4.39)$$

$$A | (\boldsymbol{\mu}, \Sigma) \stackrel{d}{\sim} \mathcal{W}_p(\Sigma, N-1) \quad (2.4.40)$$

또한  $\bar{\mathbf{x}} | (\boldsymbol{\mu}, \Sigma)$ 와  $A | (\boldsymbol{\mu}, \Sigma)$ 는 서로 독립이다. 식 (2.4.40)에서 알 수 있듯이, 조건부확률행렬  $A | (\boldsymbol{\mu}, \Sigma)$ 의 확률밀도함수는 다음과 같다.

$$f(A | \boldsymbol{\mu}, \Sigma) = \frac{1}{2^{[N-1]p/2} |\Sigma|^{[N-1]/2} \Gamma_p\left(\frac{N-1}{2}\right)} |A|^{[N-p-2]/2} \exp\left(-\frac{1}{2}\text{Tr}(\Sigma^{-1}A)\right) \quad (2.4.41)$$

식 (2.4.33), 식 (2.4.34), 식 (2.4.39) 그리고 식 (2.4.41)에서 알 수 있듯이,  $(\boldsymbol{\mu}, \Sigma, \bar{\mathbf{x}}, A)$ 의 결합확률밀도함수는 다음과 같다.

$$\begin{aligned} \pi(\boldsymbol{\mu}, \Sigma, \bar{\mathbf{x}}, A) &= \pi(\boldsymbol{\mu} | \Sigma)\pi(\Sigma)f(\bar{\mathbf{x}} | \boldsymbol{\mu}, \Sigma)f(A | \boldsymbol{\mu}, \Sigma) \\ &= \frac{K^{p/2}N^{p/2} |\Psi|^{m/2}}{\pi^p 2^{[N+m+1]p/2} \Gamma_p\left(\frac{m}{2}\right) \Gamma_p\left(\frac{N-1}{2}\right)} |\Sigma|^{-[N+m+p+2]/2} |A|^{[N-p-2]/2} \\ &\cdot \exp\left(-\frac{1}{2}K[\boldsymbol{\mu} - \boldsymbol{\mu}_0]^T \Sigma^{-1}[\boldsymbol{\mu} - \boldsymbol{\mu}_0] - \frac{1}{2}\text{Tr}(\Sigma^{-1}\Psi)\right) \\ &\cdot \exp\left(-\frac{1}{2}N[\bar{\mathbf{x}} - \boldsymbol{\mu}]^T \Sigma^{-1}[\bar{\mathbf{x}} - \boldsymbol{\mu}] - \frac{1}{2}\text{Tr}(\Sigma^{-1}A)\right) \end{aligned} \quad (2.4.42)$$

다음 식들이 성립한다.

$$\begin{aligned}
 & K[\boldsymbol{\mu} - \boldsymbol{\mu}_0]^T \Sigma^{-1} [\boldsymbol{\mu} - \boldsymbol{\mu}_0] + \text{Tr}(\Sigma^{-1} \Psi) + N[\bar{\mathbf{x}} - \boldsymbol{\mu}]^T \Sigma^{-1} [\bar{\mathbf{x}} - \boldsymbol{\mu}] + \text{Tr}(\Sigma^{-1} A) \\
 &= [N + K] \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} - 2[N\bar{\mathbf{x}} + K\boldsymbol{\mu}_0]^T \Sigma^{-1} \boldsymbol{\mu} + N\bar{\mathbf{x}}^T \Sigma^{-1} \bar{\mathbf{x}} + K\boldsymbol{\mu}_0^T \Sigma^{-1} \boldsymbol{\mu}_0 + \text{Tr}(\Sigma^{-1} [A + \Psi]) \\
 &= [N + K] \left\{ \boldsymbol{\mu} - \frac{1}{N + K} [N\bar{\mathbf{x}} + K\boldsymbol{\mu}_0] \right\}^T \Sigma^{-1} \left\{ \boldsymbol{\mu} - \frac{1}{N + K} [N\bar{\mathbf{x}} + K\boldsymbol{\mu}_0] \right\} \\
 & \quad + \frac{NK}{N + K} [\bar{\mathbf{x}} - \boldsymbol{\mu}_0]^T \Sigma^{-1} [\bar{\mathbf{x}} - \boldsymbol{\mu}_0] + \text{Tr}([A + \Psi] \Sigma^{-1}) \tag{2.4.43}
 \end{aligned}$$

식 (2.4.42) 과 식 (2.4.43) 에서 알 수 있듯이, 다음 식들이 성립한다.

$$\begin{aligned}
 \pi(\Sigma, \bar{\mathbf{x}}, A) &= \int_{\mathbb{R}^p} \pi(\boldsymbol{\mu}, \Sigma, \bar{\mathbf{x}}, A) d\boldsymbol{\mu}, \\
 &= \frac{K^{p/2} N^{p/2} |\Psi|^{m/2}}{[N + K]^{p/2} \pi^{p/2} 2^{[N+m]p/2} \Gamma_p\left(\frac{m}{2}\right) \Gamma_p\left(\frac{N-1}{2}\right)} |\Sigma|^{-[N+m+p+1]/2} |A|^{[N-p-2]/2} \\
 & \quad \cdot \exp\left(-\frac{1}{2} \frac{NK}{N + K} [\bar{\mathbf{x}} - \boldsymbol{\mu}_0]^T \Sigma^{-1} [\bar{\mathbf{x}} - \boldsymbol{\mu}_0] + \text{Tr}([A + \Psi] \Sigma^{-1})\right) \tag{2.4.44}
 \end{aligned}$$

식 (2.4.44) 에서 알 수 있듯이, 다음 식들이 성립한다.

$$\begin{aligned}
 \pi(\bar{\mathbf{x}}, A) &= \int_{\Sigma \succ \mathcal{O}} \pi(\Sigma, \bar{\mathbf{x}}, A) d\Sigma, \\
 &= \frac{K^{p/2} N^{1/2} \Gamma_p\left(\frac{N+m}{2}\right)}{[N + K]^{p/2} \pi^{p/2} \Gamma_p\left(\frac{m}{2}\right) \Gamma_p\left(\frac{N-1}{2}\right)} \\
 & \quad \cdot |A|^{[N-p-2]/2} |\Psi|^{m/2} \left| \Psi + A + \frac{NK}{N + K} [\bar{\mathbf{x}} - \boldsymbol{\mu}_0][\bar{\mathbf{x}} - \boldsymbol{\mu}_0]^T \right|^{-[N+m]/2} \tag{2.4.45}
 \end{aligned}$$

식 (2.4.42) 과 식 (2.4.45) 에서 알 수 있듯이,  $\boldsymbol{\mu}$  와  $\Sigma$  가 주어진 조건 하에서  $\bar{\mathbf{x}}$  와  $A$  의 조건부 확률밀도함수는 다음과 같다.

$$\begin{aligned}
 \pi(\boldsymbol{\mu}, \Sigma | \bar{\mathbf{x}}, A) &= \frac{\pi(\boldsymbol{\mu}, \Sigma, \bar{\mathbf{x}}, A)}{\pi(\bar{\mathbf{x}}, A)} \\
 &= \frac{[N + K]^{p/2} |\Sigma|^{-[N+m+p+2]/2} \left| \Psi + A + \frac{NK}{N+K} [\bar{\mathbf{x}} - \boldsymbol{\mu}_0][\bar{\mathbf{x}} - \boldsymbol{\mu}_0]^T \right|^{[N+m]/2}}{\pi^{p/2} 2^{[N+m+1]p/2} \Gamma_p\left(\frac{N+m}{2}\right)} \\
 & \quad \cdot \exp\left(-\frac{N + K}{2} \left\{ \boldsymbol{\mu} - \frac{1}{N + K} [N\bar{\mathbf{x}} + K\boldsymbol{\mu}_0] \right\}^T \Sigma^{-1} \left\{ \boldsymbol{\mu} - \frac{1}{N + K} [N\bar{\mathbf{x}} + K\boldsymbol{\mu}_0] \right\}\right) \\
 & \quad \cdot \exp\left(-\frac{1}{2} \text{Tr}\left(\left\{ A + \Psi + \frac{NK}{N + K} [\bar{\mathbf{x}} - \boldsymbol{\mu}_0][\bar{\mathbf{x}} - \boldsymbol{\mu}_0]^T \right\} \Sigma^{-1}\right)\right) \tag{2.4.46}
 \end{aligned}$$

따라서 평균벡터  $\boldsymbol{\mu}$ 와 분산공분산행렬  $\Sigma$ 의 사후확률분포는 다음과 같다.

$$\boldsymbol{\mu} | (\Sigma, \bar{\boldsymbol{x}}, A) \stackrel{d}{\sim} \mathcal{N}_p \left( \frac{1}{N+K} [N\bar{\boldsymbol{x}} + K\boldsymbol{\mu}_0], \frac{1}{N+K} \Sigma \right) \quad (2.4.47)$$

$$\Sigma | (\bar{\boldsymbol{x}}, A) \stackrel{d}{\sim} \mathcal{IW}_p \left( \Sigma \mid \Psi + [N-1]S + \frac{NK}{N+K} [\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0][\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0]^T, N+m \right) \quad (2.4.48)$$

## 제2.5절 사전확률분포

고전통계학은 모수는 미지이지만 고정되어 (unknown but fixed) 있고 반복적 실험을 통해서 모수를 추정할 수 있다는 전제를 바탕으로 한다. 그러나, 고정된 모수라는 개념은 사회과학자나 행동과학자가 시행하는 연구방법과 모순되는 경우가 많다. 인문사회과학에서는 엄밀하게 동일한 조건 하에서 실험조사를 반복하거나, 사회에서 새로운 쇼크에 의해 발생하는 현상을 재현할 수 없다. 따라서, 모수가 고정되어있다는 가정 하에서 이러한 현상을 분석하는 것이 타당하지 않을 수도 있다. 반면에, 베이저안통계분석에서는 관찰값들 뿐만 아니라 모수도 확률분포를 따른다고 가정한다. 즉, 고전통계학과 베이저안통계학의 주된 차이는 기본적으로 확률에 대한 해석이 다른데서 출발한다. 고전통계학에서 확률은 단기적으로는 어떤 사상이 일어날 가능성을 나타내는 것으로 해석할 수 있으며, 대수법칙에서 알 수 있듯이 장기적으로는 이 사상이 발생하는 비율이 이에 해당하는 모집단의 비율로 수렴하는 것으로 해석할 수 있다. 반면에, 베이저안통계학에서는 확률을 ‘주관적으로’ 해석하는 사전확률분포에 이 사전확률분포로부터 ‘객관적으로’ 얻어지는 데이터가 포함하는 정보를 추가해서 갱신한 사후확률분포로 나타낸다. 특히, 연구자의 과거의 경험, 개인적인 직감, 그리고 전문가의 의견을 참고해서 사전확률분포를 채택한다. 물론, 연구자가 과거 경험을 명시적으로 반영한 사전확률분포와 연구자가 갖고 있는 일반적 지식을 나타내는 사전확률분포로 구분할 수도 있다. 그러나, 일반적으로 사전확률분포를 선택하는 가장 중요한 기준은 수리적으로 얼마나 편리하게 다룰 수 있느냐 여부이다.

### 2.5.1 공액사전확률분포

베이저안통계분석에서 발생하는 어려움 중 하나는 모수벡터의 차원이 높은 경우에 사후확률분포를 해석적으로 다루기가 어렵다는 것이다. 그러나, 앞에서 간단히 소개했듯이 공액사전확률분포 (conjugate prior probability distribution)를 이용하면, 사후확률분포를 간단히 계산할 수도 있다. 앞에서 설명했듯이, 공액이란 사전확률분포와 사후확률분포가 같은 확률분포족이

되는 것을 의미한다. 이러한 사전확률분포를 공액사전확률분포라 부른다. 공액을 좀 더 수학적으로 설명하면 다음과 같다. 우도함수  $f(\mathbf{x} | \boldsymbol{\theta})$ 의 모임 (family)  $\mathcal{F}$ 와 모수벡터  $\boldsymbol{\theta}$ 의 사전확률밀도함수  $\pi(\boldsymbol{\theta})$ 의 모임  $\mathcal{P}$ 에 대해서 다음 식이 성립하면, 모임  $\mathcal{P}$ 는 모임  $\mathcal{F}$ 에 대해서 공액 (conjugate)이라 한다.

$$f(\mathbf{x} | \boldsymbol{\theta}) \in \mathcal{F} \text{ and } \pi(\boldsymbol{\theta}) \in \mathcal{P} \Rightarrow \pi(\boldsymbol{\theta} | \mathbf{x}) \in \mathcal{P} \quad (2.5.1)$$

지금부터는 공액사전확률분포에 대한 예제들을 살펴보자.

**예제 2.5.1** 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 를 따르는 확률변수들  $x_1, x_2, \dots, x_n$ 이 서로 독립이라고 가정하자. 여기서 분산  $\sigma^2$ 는 알고 있다고 가정하고 평균  $\mu$ 를 추정하기로 하자. 모수  $\mu$ 의 사전확률분포가 다음과 같다고 하자.

$$\mu \stackrel{d}{\sim} \mathcal{N}(\mu_0, \sigma_0^2) \quad (1)$$

여기서  $\mu_0$ 와  $\sigma_0^2$ 는 알고있는 상수들이다. 관찰벡터  $\mathbf{x}_1 \doteq [x_1, x_2, \dots, x_{n_1}]^T$ 이 주어졌을 때  $\mu$ 의 사후확률분포는 다음과 같다.

$$\mu | \mathbf{x}_1 \stackrel{d}{\sim} \mathcal{N}(\mu_1, \sigma_1^2) \quad (2)$$

여기서  $\bar{x}_1, \mu_1$ 와  $\sigma_1^2$ 는 각각 다음과 같다.

$$\bar{x}_1 \doteq \frac{1}{n_1} \sum_{i=1}^{n_1} x_i \quad (3)$$

$$\mu_1 \doteq \frac{\bar{x}_1 n_1 / \sigma^2 + \mu_0 / \sigma_0^2}{n_1 / \sigma^2 + 1 / \sigma_0^2} \quad (4)$$

$$\sigma_1^2 \doteq \frac{1}{n_1 / \sigma^2 + 1 / \sigma_0^2} \quad (5)$$

사전확률분포와 사후확률분포가 모두 정규확률분포들이므로, 이들은 공액이다.

만약 같은 모집단에서 새로운 관찰값들  $x_{n_1+1}, x_{n_1+2}, \dots, x_{n_1+n_2}$ 가 얻어지면, 모수  $\mu$ 에 대한 사후확률분포가 새로운 사전확률분포가 되고, 모수  $\mu$ 에 대한 새로운 사후확률분포는 관찰벡터들  $\mathbf{x}_1$ 과  $\mathbf{x}_2 = [x_{n_1+1}, x_{n_1+2}, \dots, x_{n_1+n_2}]^T$ 가 주어진 조건 하에 모수  $\mu$ 에 대한 조건부확률분포이다. 식 (2)에서 알 수 있듯이, 새로운 사전확률분포도 정규확률분포이고, 새로

도출한 사후확률분포는 다음과 같다.

$$\mu \mid (\mathbf{x}_1, \mathbf{x}_2) \stackrel{d}{\sim} \mathcal{N}(\mu_2, \sigma_2^2) \quad (6)$$

여기서  $\mu_2$ 와  $\sigma_2^2$ 는 각각 다음과 같다.

$$\mu_2 \doteq \frac{\bar{x}_2 n_2 / \sigma^2 + \mu_1 / \sigma_1^2}{n_2 / \sigma^2 + 1 / \sigma_1^2} \quad (7)$$

$$\sigma_2^2 \doteq \frac{1}{n_2 / \sigma^2 + 1 / \sigma_1^2} \quad (8)$$

또한,  $\bar{x}_2$ 는 다음과 같다.

$$\bar{x}_2 \doteq \frac{1}{n_2} \sum_{i=1}^{n_2} x_{n_1+i} \quad (9)$$

이 공액사전확률분포들을 살펴보기 위해서, 다음 R프로그램 ConjugatePrior101.R을 실행해 보자.

```

1 # -----
2 # Filename: ConjugatePrior101.R
3 # Conjugate Prior: Normal Distributions
4 # Programmed by CBS
5 # -----
6 setEPS()
7 library(ggplot2)
8 set.seed(3)
9 mu <- seq(-5,15,length=301)
10 # Prior
11 ( mu0 <- 5); ( sigma0 <- 3 )
12 (sigma02 <- sigma0^2 )
13 prior <- dnorm(mu, mean = mu0, sd = sigma0, log = FALSE)
14 # likelihood 1
15 ( mux <- 7); ( sigmax <- 2 ); (nx1 <- 16)
16 ( sigmax2 <- sigmax^2 )
17 x1 <- mux + sigmax*rnorm(nx1,0,1)
18 likelihood1 <- dnorm(x1, mean = mux, sd = sigmax, log = FALSE)
19 # posterior 1
20 xbar1 = mean(x1)
21 ( sigma2Star1 <- 1/(nx1/sigmax2+1/sigma02) )
22 ( sigmaStar1 <- sqrt(sigma2Star1) )
23 ( muStar1 <- (xbar1*nx1/sigmax2 + mu0/sigma02)*sigma2Star1 )
24 posterior1 <- dnorm(mu, mean = muStar1, sd = sigmaStar1, log = FALSE)
25 # likelihood 2
26 (nx2 <- 16)
27 x2 <- mux + sigmax*rnorm(nx2,0,1)
28 likelihood2 <- dnorm(x2, mean = mux, sd = sigmax, log = FALSE)
29 # posterior 2
30 xbar2 = mean(x2)
31 ( sigma2Star2 <- 1/(nx2/sigmax2+1/sigma2Star1) )
32 ( sigmaStar2 <- sqrt(sigma2Star2) )

```

```

33 ( muStar2 <- (xbar2*nx2/sigmax2 + muStar1/sigma2Star1)*sigma2Star2 )
34 posterior2 <- dnorm(mu, mean = muStar2, sd = sigmaStar2, log = FALSE)
35 # Plotting
36 plot.new()
37 frame()
38 postscript('ConjugatePrior101.eps') # Start to save figure
39 plot(mu,prior, type="l", xlab="mean",ylab="PDF",lty=3,lwd=4,col=2,
40      ylim = c(0,1.2) )
41 lines(x1,likelihood1,type='p',lty=3,lwd=2,col =3)
42 lines(mu,posterior1,lty=1,lwd=4,col=4)
43 lines(x2,likelihood2,type='p',lty=3,lwd=2,col =5)
44 lines(mu,posterior2,lty=1,lwd=4,col=6)
45 legend(-2.5,0.5,c("Prior","Likelihood 1","Posterior 1",
46                  "Likelihood 2","Posterior 2"),
47        lty=c(3,2,1,2,1),lwd=c(4,2,4,2,4),col=c(2,3,4,5,6) )
48 dev.off() # End to save figure
49 # -----

```

이 R 프로그램 ConjugatePrior101.R에서 사전확률분포는  $\mathcal{N}(5, 3)$  이고, 첫 번째 관찰값들  $x_1, x_2, \dots, x_{16}$  을  $\mathcal{N}(7, 2)$  에서 발생시킨다.

이 R 프로그램 ConjugatePrior101.R을 실행하면, 그림 2.5.1이 출력된다. 그림 2.5.1에서 적색 점선은 사전확률밀도함수를, 녹색 점들은 첫 번째 우도함수를, 그리고 청색 실선은 첫 번째 사후확률밀도함수를 나타낸다. 두 번째 관찰값들  $x_{17}, x_{18}, \dots, x_{32}$  을  $\mathcal{N}(7, 2)$  에서 발생시킨다. 하늘색 점들은 두 번째 우도함수를, 그리고 분홍색 실선은 두 번째 사후확률밀도함수를 나타낸다. 이 경우에 사전확률밀도함수와 사후확률밀도함수들은 모두 정규확률분포를 따른다. 그림 2.5.1에서 알 수 있듯이, 사전확률밀도함수에 우도함수의 정보를 추가한 사후확률밀도함수는 사전확률밀도함수와 우도함수의 중간에 위치한다. ■

**예제 2.5.2** 확률변수들  $x_1, x_2, \dots, x_n$  이 서로 독립이고 모수가  $\theta$  인 Bernoulli 확률분포  $Bernoulli(\theta)$  를 따른다고 하자.

조건부확률변수  $x_i | \theta$  의 확률밀도함수  $f(x_i | \theta)$  는 다음과 같다.

$$f(x_i | \theta) = \theta^{x_i} [1 - \theta]^{1-x_i}, \quad (x_i = 0, 1) \quad (1)$$

따라서, 관찰벡터  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$  의 우도함수  $f(\mathbf{x} | \theta)$  는 다음과 같다.

$$f(\mathbf{x} | \theta) = \theta^{\sum_{i=1}^n x_i} [1 - \theta]^{n - \sum_{i=1}^n x_i} \quad (2)$$

이 우도함수는  $y = \sum_{i=1}^n x_i$  의 확률밀도함수에 비례한다. 조건부확률변수  $y | \theta$  의 확률밀도함수

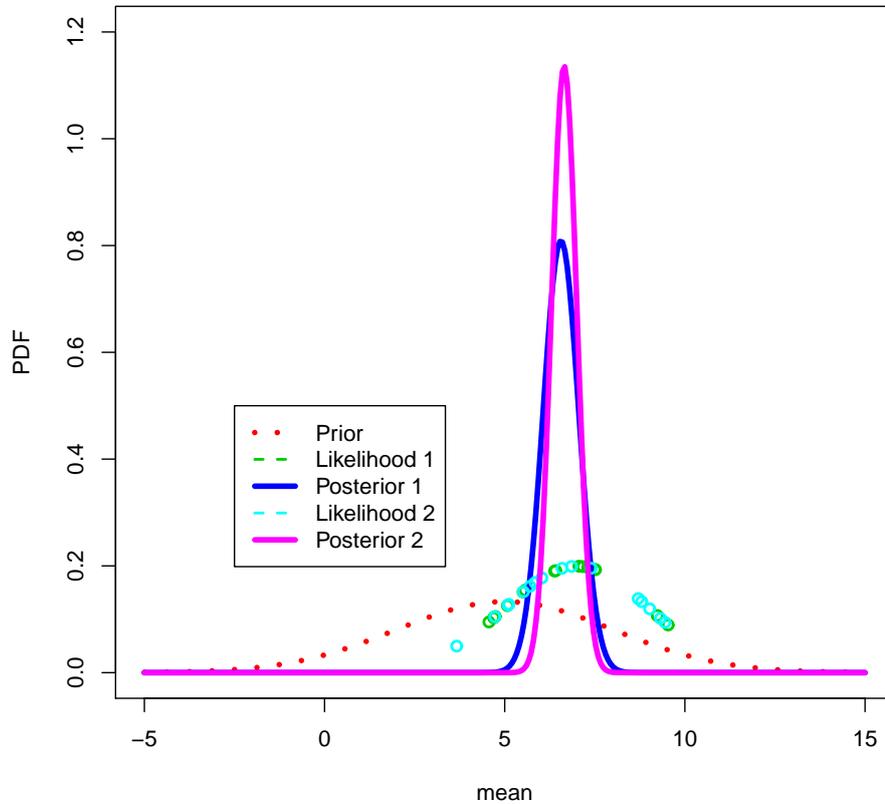


그림 2.5.1. 공액사전확률분포 1: 정규확률분포

$f(y | \theta)$ 는 다음과 같다.

$$f(y | \theta) = \binom{n}{y} \theta^y [1 - \theta]^{n-y} \quad (3)$$

모수  $\theta$ 의 사전확률분포가 모수들이  $\alpha$ 와  $\beta$ 인 베타확률분포  $Beta(\alpha, \beta)$  라고 하자. 여기서  $\alpha$ 와  $\beta$ 는 알고있는 상수들이다. 다음 식이 성립한다.

$$\pi(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} [1 - \theta]^{\beta-1}, \quad (0 \leq \theta \leq 1) \quad (4)$$

식 (3)과 식 (4)에서 알 수 있듯이,  $(y, \theta)$ 의 결합확률밀도함수는 다음과 같다.

$$h(y, \theta) = \frac{1}{B(\alpha, \beta)} \binom{n}{y} \theta^{\alpha+y-1} [1 - \theta]^{n-y+\beta-1}, \quad (0 \leq \theta \leq 1) \quad (5)$$

또한 확률변수  $y$ 의 주변확률밀도함수는 다음과 같다.

$$m(y) = \frac{1}{\text{Beta}(\alpha, \beta)} \binom{n}{y} B(y + \alpha, n - y + \beta), \quad (y = 0, 1, \dots, n) \quad (6)$$

따라서, 모수  $\theta$ 의 사후확률밀도함수는 다음과 같다.

$$\pi(\theta | \mathbf{x}) = \frac{1}{B(y + \alpha, n - y + \beta)} \theta^{y + \alpha - 1} [1 - \theta]^{n - y + \beta - 1} \quad (7)$$

즉, 모수  $\theta$ 의 사후확률분포를 다음과 같이 표기할 수 있다.

$$\theta | \mathbf{x} \stackrel{d}{\sim} \text{Beta}(y + \alpha, n - y + \beta), \quad (0 \leq \theta \leq 1) \quad (8)$$

■

**예제 2.5.3** 확률변수들  $x_1, x_2, \dots, x_n$ 이 서로 독립이고 비율모수 (rate parameter)가  $\theta$ 인 지수확률분포  $Exponential(\theta)$ 를 따른다고 하자. 즉, 이 지수확률분포의 크기모수 (scale parameter)는  $1/\theta$ 이다.

조건부확률변수  $x_i | \theta$ 의 확률밀도함수  $f(x_i | \theta)$ 는 다음과 같다.

$$f(x_i | \theta) = \theta \exp(-\theta x_i), \quad (x_i \geq 0) \quad (1)$$

따라서, 관찰벡터  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ 의 우도함수  $f(\mathbf{x} | \theta)$ 는 다음과 같다.

$$f(\mathbf{x} | \theta) = \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right) \quad (2)$$

모수  $\theta$ 의 사전확률분포가 형태모수 (shape parameter)가  $\alpha$ 이고 비율모수 (rate parameter)가  $\beta$ 인 감마확률분포  $Gamma(\alpha, \beta)$ 라고 하자. 여기서  $\alpha$ 와  $\beta$ 는 알려진 상수들이다. 따라서, 크기모수 (scale parameter)는  $1/\beta$ 이다. 다음 식이 성립한다.

$$\pi(\theta) = \frac{1}{\Gamma(\alpha)} \beta^\alpha \theta^{\alpha-1} \exp(-\beta\theta), \quad (\theta > 0) \quad (3)$$

식 (2)와 식 (3)에서 알 수 있듯이, 모수  $\theta$ 의 사후확률밀도함수는 다음과 같다.

$$\pi(\theta | \mathbf{x}) \propto f(\mathbf{x} | \theta)\pi(\theta) \propto \theta^{\alpha+n-1} \exp\left(-\theta \left[\sum_{i=1}^n x_i + \beta\right]\right) \quad (4)$$

즉, 모수  $\theta$ 의 사후확률분포를 다음과 같이 표기할 수 있다.

$$\theta | \mathbf{x} \stackrel{d}{\sim} \text{Gamma}\left(\alpha + n, \sum_{i=1}^n x_i + \beta\right) \quad (5)$$

사전확률분포와 사후확률분포가 모두 감마확률분포이므로, 이들은 공액이다. 이 공액사  
전확률분포들을 살펴보기 위해서, 다음 R프로그램 ConjugatePrior102.R을 실행해 보자.

```

1 # -----
2 # Filename: ConjugatePrior102.R
3 # Conjugate Prior: Gamma distribution
4 # Programmed by CBS
5 # -----
6 setEPS()
7 library(ggplot2)
8 set.seed(3)
9 theta <- seq(0,5,length=301)
10 # Prior
11 ( alpha0 <- 2.1); ( beta0 <- 1.9 )
12 prior <- dgamma(theta, shape=alpha0,rate=beta0,log=FALSE)
13 # likelihood 1
14 ( xmean = alpha0/beta0*1.05)
15 ( xtheta <- 1/xmean ); (nx1 <- 16)
16 ( x1 <- rexp(nx1,rate=xtheta) )
17 ( likelihood1 = dexp(x1,rate=xtheta) )
18 # posterior 1
19 ( sumx1 = sum(x1) )
20 ( alphaStar1 <- alpha0 + nx1 );
21 ( betaStar1 <- beta0 + sumx1 )
22 posterior1 <- dgamma(theta, shape=alphaStar1,rate=betaStar1,log=FALSE)
23 # likelihood 2
24 ( nx2 <- 16)
25 ( x2 <- rexp(nx2,rate=xtheta) )
26 ( likelihood2 = dexp(x2,rate=xtheta) )
27 # posterior 2
28 ( sumx2 = sum(x2) )
29 ( alphaStar2 <- alphaStar1 + nx2 );
30 ( betaStar2 <- betaStar1 + sumx2 )
31 posterior2 <- dgamma(theta, shape=alphaStar2,rate=betaStar2,log=FALSE)
32 # Plotting
33 plot.new()
34 frame()
35 postscript('ConjugatePrior102.eps') # Start to save figure
36 plot(theta,prior, type="l", xlab="theta",ylab="PDF",lty=3,lwd=4,col=2,
37       ylim = c(0,2.9) )
38 lines(x1,likelihood1,type='p',lty=3,lwd=2,col =3)
39 lines(theta,posterior1,lty=1,lwd=4,col=4)
40 lines(x2,likelihood2,type='p',lty=3,lwd=2,col =5)
41 lines(theta,posterior2,lty=1,lwd=4,col=6)
42 legend(3,1,c("Prior","Likelihood 1","Posterior 1",

```

```

43     "Likelihood 2","Posterior 2"),
44     lty=c(3,2,1,2,1),lwd=c(4,2,4,2,4),col=c(2,3,4,5,6) )
45 dev.off()           # End to save figure
46 # -----

```

이 R 프로그램 ConjugatePrior102.R에서 사전확률분포는  $Gamma(2.1, 1.9)$ 이다.

이 R 프로그램 ConjugatePrior102.R을 실행하면, 비율모수가 1.160526인 지수확률분포에서 첫 번째 관찰값들  $x_1, x_2, \dots, x_{16}$ 을 발생시키고, 그림 2.5.2를 출력한다. 그림 2.5.2에서 적색 점선은 사전확률밀도함수를, 녹색 점들은 첫 번째 우도함수를, 그리고 청색 실선은 첫 번째 사후확률밀도함수를 나타낸다. 동일한 지수확률분포에서 두 번째 관찰값들  $x_{17}, x_{18}, \dots, x_{32}$ 를 발생시킨다. 그림 2.5.2에서 하늘색 점들은 두 번째 우도함수를, 그리고 분홍색 실선은 두 번째 사후확률밀도함수를 나타낸다. 여기서 사전확률밀도함수와 사후확률밀도함수들은 모두 감마확률분포를 따른다. 그림 2.5.2에서 알 수 있듯이, 사전확률밀도함수에 우도함수의 정보를 추가한 사후확률밀도함수는 사전확률밀도함수보다 우측으로 쏠려있다. 또한, 두 번째 사후확률분포의 분산이 첫 번째 것보다 작다. ■

**예제 2.5.4** 확률변수들  $x_1, x_2, \dots, x_n$ 이 서로 독립이고 평균이  $\lambda$ 인 Poisson 확률분포  $Poisson(\lambda)$ 를 따른다고 하자. 조건부확률변수  $x_i | \lambda$ 의 확률밀도함수  $f(x_i | \lambda)$ 는 다음과 같다.

$$f(x_i | \lambda) = \exp(-\lambda) \frac{\lambda^{x_i}}{x_i!} \quad (1)$$

따라서, 관찰벡터  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ 의 우도함수  $f(\mathbf{x} | \lambda)$ 는 다음과 같다.

$$f(\mathbf{x} | \lambda) = \prod_{i=1}^n \frac{\exp(-\lambda) \lambda^{x_i}}{x_i!} = \frac{\exp(-n\lambda)}{\prod_{i=1}^n x_i!} \lambda^{\sum_{i=1}^n x_i} \quad (2)$$

모수  $\lambda$ 의 사전확률분포로서 모수들이  $\alpha$ 와  $\beta$ 인 감마확률분포  $Gamma(\alpha, \beta)$ 를 사용하자. 여기서  $\alpha$ 와  $\beta$ 는 알고있는 상수들이다. 다음 식이 성립한다.

$$\pi(\lambda) = \frac{1}{\Gamma(\alpha)} \beta^\alpha \lambda^{\alpha-1} \exp(-\beta\lambda), \quad (\lambda > 0) \quad (3)$$

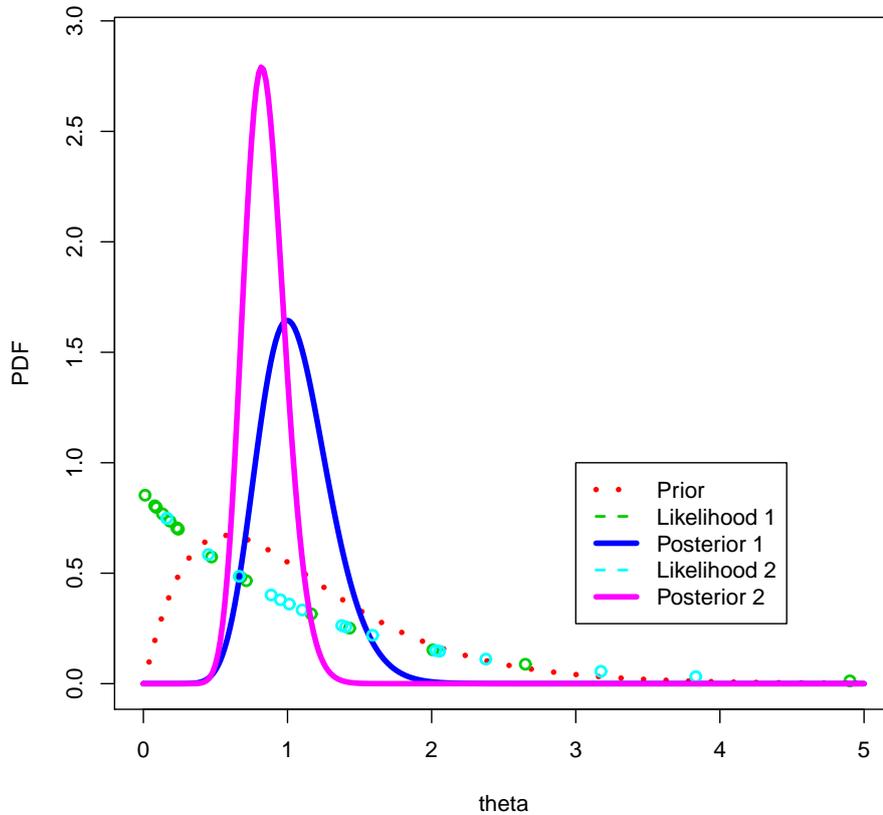


그림 2.5.2. 공액사전확률분포 2: 감마확률분포

식 (2)와 식 (3)에서 알 수 있듯이, 모수  $\lambda$ 의 사후확률밀도함수는 다음과 같다.

$$\begin{aligned} \pi(\lambda | \mathbf{x}) &\propto f(\mathbf{x} | \lambda)\pi(\lambda) \\ &= \exp(-n\lambda)\lambda^{\sum_{i=1}^n x_i} \lambda^{\alpha-1} \exp(-\beta\lambda) \propto \lambda^{\alpha+\sum_{i=1}^n x_i-1} \exp(-[\beta+n]\lambda) \end{aligned} \quad (4)$$

따라서, 모수  $\lambda$ 의 사후확률분포를 다음과 같이 표기할 수 있다.

$$\lambda | \mathbf{x} \stackrel{d}{\sim} \text{Gamma} \left( \alpha + \sum_{i=1}^n x_i, \beta + n \right) \quad (5)$$

사전확률분포와 사후확률분포가 모두 감마확률분포이므로, 이들은 공액이다. 이 확률분포들을 살펴보기 위해서, 다음 R프로그램 ConjugatePrior103.R을 실행해 보자.

```
1 # -----
2 # Filename: ConjugatePrior103.R
3 # Conjugate Prior: Gamma distribution
```

```

4 # Programmed by CBS
5 # -----
6 setEPS()
7 library(ggplot2)
8 set.seed(3)
9 lambda <- seq(0,5,length=301)
10 # Prior
11 ( alpha0 <- 2.1); ( beta0 <- 1.9 )
12 prior <- dgamma(lambda, shape=alpha0,rate=beta0,log=FALSE)
13 # likelihood 1
14 ( xmean = alpha0/beta0*1.05)
15 ( xlambda <- xmean ); (nx1 <- 16)
16 ( x1 <- rpois(nx1,xlambda) )
17 ( likelihood1 = dpois(x1,xlambda) )
18 # posterior 1
19 ( sumx1 = sum(x1) )
20 ( alphaStar1 <- alpha0 + sumx1 )
21 ( betaStar1 <- beta0 + nx1 )
22 posterior1 <- dgamma(lambda, shape=alphaStar1,rate=betaStar1,log=FALSE)
23 # likelihood 2
24 ( nx2 <- 16)
25 ( x2 <- rpois(nx2,xlambda) )
26 ( likelihood2 = dpois(x2,xlambda) )
27 # posterior 2
28 ( sumx2 = sum(x2) )
29 ( alphaStar2 <- alphaStar1 + sumx2 );
30 ( betaStar2 <- betaStar1 + nx2 )
31 posterior2 <- dgamma(lambda, shape=alphaStar2,rate=betaStar2,log=FALSE)
32 # Plotting
33 plot.new()
34 frame()
35 postscrip('ConjugatePrior103.eps') # Start to save figure
36 plot(lambda,prior, type="l", xlab="lambda",ylab="PDF",lty=3,lwd=4,col=2,
37       ylim = c(0,2.5) )
38 lines(x1,likelihood1,type='p',lty=3,lwd=2,col =3)
39 lines(lambda,posterior1,lty=1,lwd=4,col=4)
40 lines(x2,likelihood2,type='p',lty=3,lwd=2,col =5)
41 lines(lambda,posterior2,lty=1,lwd=4,col=6)
42 legend(3,1.3,c("Prior","Likelihood 1","Posterior 1",
43              "Likelihood 2","Posterior 2"),
44        lty=c(3,2,1,2,1),lwd=c(4,2,4,2,4),col=c(2,3,4,5,6) )
45 dev.off() # End to save figure
46 # -----

```

이 R프로그램 ConjugatePrior103.R에서 사전확률분포는  $Gamma(2.1, 1.9)$  이고, 비율모수가 1.160526인 Poisson 확률분포에서 첫 번째 관찰값들  $x_1, x_2, \dots, x_{16}$  을 발생시킨다.

이 R프로그램 ConjugatePrior103.R을 실행하면, 그림 2.5.3이 출력된다. 그림 2.5.3에서 적색 점선은 사전확률밀도함수를, 녹색 점들은 첫 번째 우도함수를, 그리고 청색 실선은 첫 번째 사후확률밀도함수를 나타낸다. 동일한 Poisson 확률분포에서 두 번째 관찰값들  $x_{17}, x_{18}, \dots, x_{32}$  를 발생시킨다. 하늘색 점들은 두 번째 우도함수를, 그리고 분홍색 실선은 두 번째 사후확률밀도함수를 나타낸다. 여기서 사전확률밀도함수와 사후확률밀도함수들은 모두 감마확률분포를 따른다. 그림 2.5.3에서 알 수 있듯이, 사전확률밀도함수에 우도함수의

정보를 추가한 사후확률밀도함수는 사전확률밀도함수보다 우측으로 쏠려있다. 또한, 두 번째 사후확률분포의 분산이 첫 번째 것보다 작다. ■

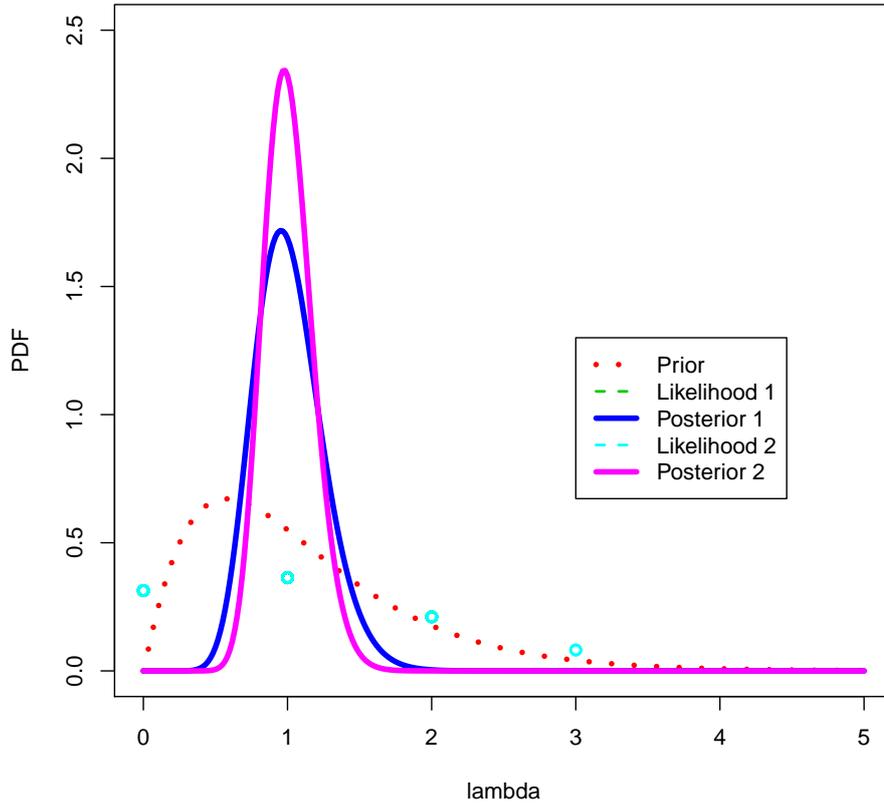


그림 2.5.3. 공액사전확률분포 3: 감마확률분포

**예제 2.5.5** 확률변수들  $x_1, x_2, \dots, x_n$  이 서로 독립이고 평균이  $\lambda$ 인 지수형 (exponential family) 확률분포를 따른다고 하자. Koopman-Darmois형이라고도 불리우는 지수형 확률분포의 확률밀도함수  $f(x_j | \theta)$ 는 다음과 같다.

$$f(x_j | \theta) = a(x_j) \exp(\varphi(\theta)t(x_j) + b(\theta)) \tag{1}$$

식 (1)에는 모수가 1개만 포함되어 있으므로,  $x_j$ 는 1모수 지수형 확률분포를 따른다고 한다. 대표적인 1모수 지수형 확률분포들은 다음과 같다.

첫째, 분산  $\sigma^2$ 가 알려진 정규확률분포  $\mathcal{N}(\theta, \sigma^2)$ 는 1모수 지수형 확률분포이다. 다음 식들이

성립한다.

$$a(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad t(x) = x, \quad \varphi(\theta) = \frac{\theta}{\sigma^2}, \quad b(\theta) = -\frac{\theta^2}{2\sigma^2} \quad (2)$$

둘째, 모수  $n$ 이 알려진 이항확률분포  $Binomial(n, \theta)$ 는 1모수 지수형 확률분포이다. 다음 식들이 성립한다.

$$a(x) = \binom{n}{x}, \quad t(x) = x, \quad \varphi(\theta) = \ln \frac{\theta}{1-\theta}, \quad b(\theta) = -n \ln(1-\theta) \quad (3)$$

셋째, 지수확률분포  $Exponential(\theta)$ 는 1모수 지수형 확률분포이다. 다음 식들이 성립한다.

$$a(x) = 1, \quad t(x) = x, \quad \varphi(\theta) = -\theta, \quad b(\theta) = \ln \theta \quad (4)$$

넷째, Poisson 확률분포  $Poisson(\theta)$ 는 1모수 지수형 확률분포이다. 다음 식들이 성립한다.

$$a(x) = \frac{1}{x!}, \quad t(x) = x, \quad \varphi(\theta) = \ln \theta, \quad b(\theta) = -\theta \quad (5)$$

그러나, 일양확률분포나  $t$  확률분포는 지수형 확률분포족에 속하지 않는다.

식 (1)을 만족하는 지수형 확률분포를 따르는 관찰벡터  $\mathbf{x} \doteq [x_1, x_2, \dots, x_n]^T$ 의 우도함수  $f(\mathbf{x} | \theta)$ 는 다음과 같다.

$$f(\mathbf{x} | \theta) = \left[ \prod_{i=1}^n a(x_i) \right] \exp\left(\varphi(\theta) \sum_{i=1}^n t(x_i) + nb(\theta)\right) \quad (6)$$

모수  $\theta$ 의 사전확률밀도함수가  $\pi(\theta)$ 이면, 이 모수의 사후확률밀도함수는 다음과 같다

$$\pi(\theta | \mathbf{x}) \propto \left[ \prod_{i=1}^n a(x_i) \right] \exp\left(\varphi(\theta) \sum_{i=1}^n t(x_i) + nb(\theta)\right) \pi(\theta) \quad (7)$$

따라서, 모수  $\theta$ 의 사전확률분포가 지수형 확률분포이면, 사후확률분포 역시 지수형 확률분포이다. 즉, 관찰값들이 지수형 확률분포족에서 발생된 경우에는 지수형 확률분포족은 공액성을 갖는다. ■

공액성이 없는 경우에 사후확률분포를 유도하기가 어렵기 때문에, 공액성은 오랫동안

베이저안통계분석에서 중요한 역할을 해왔다. 그러나, MCMC를 사용하면, 공액성이 없는 경우에도 사후확률분포를 수치적으로 쉽게 구할 수 있다. 그럼에도 불구하고, 오늘날에도 베이저안통계분석에서 공액성은 여전히 유용한 도구이다. 다양한 공액사전확률분포들이 표 2.5.1 ~ 표 2.5.4에 수록되어 있다. 이 표들은 다음과 같이 위키피디어 웹사이트에서 인용한 것이다.

[https://en.wikipedia.org/wiki/Conjugate\\_prior](https://en.wikipedia.org/wiki/Conjugate_prior)

Likelihood	Model parameters	Conjugate prior	Prior hyperparameters	Posterior hyperparameters
Bernoulli	$p$ (probability)	Beta	$\alpha, \beta$	$\alpha + \sum_{i=1}^n x_i, \beta + n - \sum_{i=1}^n x_i$
Binomial	$p$ (probability)	Beta	$\alpha, \beta$	$\alpha + \sum_{i=1}^n x_i, \beta + \sum_{i=1}^n N_i - \sum_{i=1}^n x_i$
Negative binomial with known failure number, $r$	$p$ (probability)	Beta	$\alpha, \beta$	$\alpha + \sum_{i=1}^n x_i, \beta + rn$
Poisson	$\lambda$ (rate)	Gamma	$k, \theta$	$k + \sum_{i=1}^n x_i, \frac{\theta}{n\theta + 1}$
Poisson	$\lambda$ (rate)	Gamma	$\alpha, \beta$	$\alpha + \sum_{i=1}^n x_i, \beta + n$
Categorical	$\mathbf{p}$ (probability vector), $k$ (number of categories; i.e., size of $\mathbf{p}$ )	Dirichlet	$\boldsymbol{\alpha}$	$\boldsymbol{\alpha} + (c_1, \dots, c_k)$ , where $c_i$ is the number of observations in category $i$
Multinomial	$\mathbf{p}$ (probability vector), $k$ (number of categories; i.e., size of $\mathbf{p}$ )	Dirichlet	$\boldsymbol{\alpha}$	$\boldsymbol{\alpha} + \sum_{i=1}^n \mathbf{x}_i$
Hypergeometric with known total population size, $N$	$M$ (number of target members)	Beta-binomial	$n = N$ $\alpha, \beta$	$\alpha + \sum_{i=1}^n x_i, \beta + \sum_{i=1}^n N_i - \sum_{i=1}^n x_i$
Geometric	$p_0$ (probability)	Beta	$\alpha, \beta$	$\alpha + n, \beta + \sum_{i=1}^n x_i - n$

표 2.5.1. Conjugate Discrete Distributions

Likelihood	Model parameters	Conjugate prior	Prior hyperparameters	Posterior hyperparameters
Normal with known variance $\sigma^2$	$\mu$ (mean)	Normal	$\mu_0, \sigma_0^2$	$\frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \left( \frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n x_i}{\sigma^2} \right), \left( \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)^{-1}$
Normal with known precision $\tau$	$\mu$ (mean)	Normal	$\mu_0, \tau_0$	$\frac{\tau_0 \mu_0 + \tau \sum_{i=1}^n x_i}{\tau_0 + n\tau}, \tau_0 + n\tau$
Normal with known mean $\mu$	$\sigma^2$ (variance)	Inverse gamma	$\alpha, \beta$	$\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2}$
Normal with known mean $\mu$	$\sigma^2$ (variance)	Scaled inverse chi-squared	$\nu, \sigma_0^2$	$\nu + n, \frac{\nu \sigma_0^2 + \sum_{i=1}^n (x_i - \mu)^2}{\nu + n}$
Normal with known mean $\mu$	$\tau$ (precision)	Gamma	$\alpha, \beta$	$\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2}$
Normal	$\mu$ and $\sigma^2$ Assuming exchangeability	Normal-inverse gamma	$\mu_0, \nu, \alpha, \beta$	$\frac{\nu \mu_0 + n\bar{x}}{\nu + n}, \nu + n, \alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{n\nu}{\nu + n} \frac{(\bar{x} - \mu_0)^2}{2}$
Normal	$\mu$ and $\tau$ Assuming exchangeability	Normal-gamma	$\mu_0, \nu, \alpha, \beta$	$\frac{\nu \mu_0 + n\bar{x}}{\nu + n}, \nu + n, \alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{n\nu}{\nu + n} \frac{(\bar{x} - \mu_0)^2}{2}$
Multivariate normal w/ known covariance matrix $\Sigma$	$\boldsymbol{\mu}$ (mean vector)	Multivariate normal	$\boldsymbol{\mu}_0, \Sigma_0$	$\left( \Sigma_0^{-1} + n\Sigma^{-1} \right)^{-1} \left( \Sigma_0^{-1} \boldsymbol{\mu}_0 + n\Sigma^{-1} \bar{\boldsymbol{x}} \right)$
Multivariate normal w/ known precision matrix $\Lambda$	$\boldsymbol{\mu}$ (mean vector)	Multivariate normal	$\boldsymbol{\mu}_0, \Lambda_0$	$(\Lambda_0 + n\Lambda)^{-1} (\Lambda_0 \boldsymbol{\mu}_0 + n\Lambda \bar{\boldsymbol{x}}), (\Lambda_0 + n\Lambda)$

표 2.5.2. Conjugate Continuous Distributions (1)

Likelihood	Model parameters	Conjugate prior	Prior hyperparameters	Posterior hyperparameters
Multivariate normal w/ known mean $\boldsymbol{\mu}$	$\Sigma$ (covariance matrix)	Inverse-Wishart	$\nu, \Psi$	$n + \nu, \Psi + \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T$
Multivariate normal w/ known mean $\boldsymbol{\mu}$	$\Lambda$ (precision matrix)	Wishart	$\nu, V$	$n + \nu, \left( V^{-1} + \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T \right)^{-1}$
Multivariate normal	$\boldsymbol{\mu}$ (mean vector), $\Sigma$ (covariance matrix)	normal-inverse-Wishart	$\boldsymbol{\mu}_0, \kappa_0, \nu_0, \Psi$	$\frac{\kappa_0 \boldsymbol{\mu}_0 + n \bar{\mathbf{x}}}{\kappa_0 + n}, \kappa_0 + n, \nu_0 + n,$ $\Psi + C + \frac{\kappa_0 n}{\kappa_0 + n} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T$ $\cdot C = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$
Multivariate normal	$\boldsymbol{\mu}$ (mean vector), $\Lambda$ (precision matrix)	normal-Wishart	$\boldsymbol{\mu}_0, \kappa_0, \nu_0, V$	$\frac{\kappa_0 \boldsymbol{\mu}_0 + n \bar{\mathbf{x}}}{\kappa_0 + n}, \kappa_0 + n, \nu_0 + n,$ $\left( V^{-1} + C + \frac{\kappa_0 n}{\kappa_0 + n} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \right)^{-1}$ $\cdot C = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$
Uniform	$U(0, \theta)$	Pareto	$x_m, k$	$\max \{x_1, \dots, x_n, x_m\}, k + n$
Pareto with known minimum $x_m$	$k$ (shape)	Gamma	$\alpha, \beta$	$\alpha + n, \beta + \sum_{i=1}^n \ln \frac{x_i}{x_m}$
Weibull with known shape $\beta$	$\theta$ (scale)	Inverse gamma	$a, b$	$a + n, b + \sum_{i=1}^n x_i^\beta$

표 2.5.3. Conjugate Continuous Distributions (2)

### 2.5.2 무정보사전확률분포

모수에 대한 정보가 거의 없는 경우에는 무정보사전확률분포 (noninformative prior probability distribution)를 사용한다. 무정보사전확률분포들에 대한 많은 논의가 있지만, 어떤 무정보사전확률분포가 가장 좋은 것이라는 일치된 의견은 없다. 가장 일반적으로 이용되는 것은 Jeffreys형 사전확률분포이다. Jeffreys형 사전확률분포는 확률밀도함수가 지지대인 유계구간 또는 무한구간에서 상수이다. 모수  $\theta$ 의 지지대가 유계구간  $[a, b]$ 인 경우에는 적분값  $\int_a^b d\theta$ 가 상수이므로, 사전확률분포에 대한 확률밀도함수를 정의할 수 있다. 이렇게 지지대에서

Likelihood	Model parameters	Conjugate prior	Prior hyperparameters	Posterior hyperparameters
Log-normal with known precision $\tau$	$\mu$ (mean)	Normal	$\mu_0, \tau_0$	$\left(\tau_0\mu_0 + \tau \sum_{i=1}^n \ln x_i\right) / (\tau_0 + n\tau), \tau_0 + n\tau$
Log-normal with known mean $\mu$	$\tau$ (precision)	Gamma	$\alpha, \beta$	$\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^n (\ln x_i - \mu)^2}{2}$
Exponential	$\lambda$ (rate)	Gamma	$\alpha, \beta$	$\alpha + n, \beta + \sum_{i=1}^n x_i$
Gamma with known shape $\alpha$	$\beta$ (rate)	Gamma	$\alpha_0, \beta_0$	$\alpha_0 + n\alpha, \beta_0 + \sum_{i=1}^n x_i$
Inverse Gamma with known shape $\alpha$	$\beta$ (inverse scale)	Gamma	$\alpha_0, \beta_0$	$\alpha_0 + n\alpha, \beta_0 + \sum_{i=1}^n \frac{1}{x_i}$
Gamma with known rate $\beta$	$\alpha$ (shape)	$\propto \frac{a^{\alpha-1}\beta^{\alpha c}}{\Gamma(\alpha)^b}$	$a, b, c$	$a \prod_{i=1}^n x_i, b + n, c + n$
Gamma	$\alpha$ (shape), $\beta$ (inverse scale)	$\propto \frac{p^{\alpha-1}e^{-\beta q}}{\Gamma(\alpha)^r\beta^{-\alpha s}}$	$p, q, r, s$	$p \prod_{i=1}^n x_i, q + \sum_{i=1}^n x_i, r + n, s + n$

표 2.5.4. Conjugate Continuous Distributions (3)

적분가능한 사전확률확률밀도함수를 정칙이라 (proper) 한다. 반면에, 지지대가 무한구간인  $(-\infty, \infty)$  이거나  $(0, \infty)$  인 경우에는 사전확률밀도함수를 지지대에서 적분하면, 적분값이  $\infty$  이다. 이러한 경우에는 사전확률밀도함수를 정의할 수는 없다. 그러나, Jeffreys는 평균  $\mu$  와 표준편차  $\sigma$  에 대한 무정보사전확률밀도함수들을 각각 다음과 같이 정의했다.

$$\pi(\mu) \propto \text{상수}, \quad (-\infty < \mu < \infty) \tag{2.5.2}$$

$$\pi(\sigma) \propto \frac{1}{\sigma}, \quad (0 < \sigma < \infty) \tag{2.5.3}$$

이 무정보사전확률밀도함수들을 지지대에서 적분한 값은 유한이 아니므로, 이에 해당하는 확률분포를 비정칙 (improper) 사전확률분포라 부른다.

**예제 2.5.6** 확률변수들  $x_1, x_2, \dots, x_n$  이 서로 독립이고 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$  를 따른다

고 하자. 평균  $\mu$ 를 알고 있다는 조건 하에 분산  $\sigma^2$ 를 추정하기로 하자.

분산  $\sigma^2$ 의 사전확률분포는 다음과 같은 역감마확률분포라고 가정하자.

$$\sigma^2 \stackrel{d}{\sim} \text{IG}\left(\frac{\nu_0}{2}, \frac{\nu_0 S_0}{2}\right) \quad (1)$$

앞에서 언급했듯이, 분산  $\sigma^2$ 의 사후확률분포는 다음과 같다.

$$\sigma^2 | \mathbf{x} \stackrel{d}{\sim} \text{IG}\left(\frac{\nu_0 + n}{2}, \frac{\nu_0 S_0 + n S_1}{2}\right) \quad (2)$$

즉, 사전확률분포와 사후확률분포가 모두 역감마확률분포를 따르므로, 역감마확률분포는 공액사전확률분포이다. 만약 사전확률분포의 자유도  $\nu_0$ 가 0에 가까이 가면, 점근적 사후확률분포는 다음과 같다.

$$\sigma^2 | \mathbf{x} \stackrel{d}{\sim} \text{IG}\left(\frac{n}{2}, \frac{n S_1}{2}\right) \quad (3)$$

분산  $\sigma^2$ 의 사전확률밀도함수를 다음과 같이 가정해도, 식 (3)의 사후확률분포를 얻는다.

$$\pi(\sigma^2) \propto \frac{1}{\sigma^2} \quad (4)$$

그러나, 이 사전확률밀도함수  $\pi(\sigma^2)$ 을 구간  $(0, \infty)$ 에서 적분하면, 그 값은 무한이 되고, 따라서  $\pi(\sigma^2)$ 는 (정칙인) 확률밀도함수가 아니다. ■

Jeffreys [56, p. 181]는 모수의 1대1 변환에 대해 불변인 사전확률분포를 제안하였다. 확률변수들  $x_1, x_2, \dots, x_n$ 이 서로 독립이고, 확률변수  $x_j$ 의 확률밀도함수가  $f(x_j | \theta)$ 라고 하자. 관찰벡터  $\mathbf{x} \doteq [x_1, x_2, \dots, x_n]^T$ 의 결합확률밀도함수를  $f(\mathbf{x} | \theta)$ 로 표기하면, 모수  $\theta$ 에 대한 Fisher 정보량  $I(\theta)$ 는 다음과 같다.

$$I(\theta) \doteq -E\left(\frac{d^2 \ln f(\mathbf{x} | \theta)}{d\theta^2} \middle| \theta\right) = E\left(\left[\frac{d \ln f(\mathbf{x} | \theta)}{d\theta}\right]^2 \middle| \theta\right) \quad (2.5.4)$$

모수  $\theta$ 에 대한 사전확률밀도함수  $\pi(\theta)$ 가 다음과 같다고 가정하자.

$$\pi(\theta) \propto |I(\theta)|^{1/2} \quad (2.5.5)$$

모수  $\eta = h(\theta)$ 에 대한 사전확률밀도함수  $\phi(\eta)$ 는 다음과 같다.

$$\phi(\eta) = \pi(\theta) \left| \frac{d\theta}{d\eta} \right| \propto |I(\theta)|^{1/2} \left| \frac{d\theta}{d\eta} \right| \quad (2.5.6)$$

Lehmann & Casella [66]의 식 (2.5.16)에서 알 수 있듯이, 모수  $\eta$ 에 대한 Fisher 정보량  $I(\eta)$ 는 다음과 같다.

$$I(\eta) = -E \left( \frac{d^2 \ln f(\mathbf{x} | h^{-1}(\eta))}{d\eta^2} \middle| \eta \right) = I(\theta) \left[ \frac{d\theta}{d\eta} \right]^2 \quad (2.5.7)$$

식 (2.5.6)와 식 (2.5.7)에서 알 수 있듯이, 다음 식이 성립한다.

$$\phi(\eta) \propto |I(\eta)|^{1/2} \quad (2.5.8)$$

식 (2.5.5)와 식 (2.5.8)에서 알 수 있듯이, 모수  $\theta$ 와 모수  $h(\theta)$ 의 사전확률밀도함수들은 같은 형태이다. 즉, 식 (2.5.5)에서 정의된 사전확률분포는 모수변환에 대해서 불변이다. 이러한 Jeffreys 원리를 모수벡터  $\boldsymbol{\theta}$ 로 확장할 수 있다. 그러나, 그 결과는 많은 논쟁요소를 포함하고 있다. 특히, 모수벡터  $\boldsymbol{\theta}$ 의 각 요소에 독립적으로 무정보사전확률분포를 적용해서 얻은 결과가 모수벡터  $\boldsymbol{\theta}$ 에 대해 Jeffreys 원리를 적용하여 얻은 결과와 다르다는 것에 유의하라.

**예제 2.5.7** 확률변수들  $x_1, x_2, \dots, x_n$ 이 서로 독립이고 모수가  $p$ 인 Bernoulli 확률분포를 따른다고 가정하자. 관찰점 개수  $n$ 을 알고 있다는 조건 하에서, Jeffreys의 사전확률분포를 구해보자.

조건부확률변수  $x_j | p$ 의 확률밀도함수  $f(x_j | p)$ 는 다음과 같다.

$$f(x_j | p) = p^{x_j} [1 - p]^{1-x_j} \quad (1)$$

관찰벡터  $\mathbf{x} \doteq [x_1, x_2, \dots, x_n]^T$ 의 우도함수  $f(\mathbf{x} | p)$ 는 다음과 같다.

$$f(\mathbf{x} | p) = p^x [1 - p]^{n-x} \quad (2)$$

여기서  $x = \sum_{i=1}^n x_i$ 이다. 즉, 대수우도함수는 다음과 같다.

$$\ln f(\mathbf{x} | p) = x \ln p + [n - x] \ln(1 - p) \quad (3)$$

따라서, 모수  $p$ 의 Fisher 정보량은 다음과 같다.

$$I(p) = -E \left( \frac{d^2 \ln f(x | p)}{dp^2} \middle| p \right) = \frac{n}{p[1-p]} \quad (4)$$

식 (2.5.5)와 식 (4)에서 알 수 있듯이, Jeffreys의 사전확률밀도함수는 다음과 같다.

$$\pi(p) \propto p^{-1/2}[1-p]^{-1/2} \quad (5)$$

이 사전확률밀도함수는 베타확률분포  $Beta(\frac{1}{2}, \frac{1}{2})$ 를 따른다. ■

**예제 2.5.8** 확률변수들  $x_1, x_2, \dots, x_n$ 이 서로 독립이고 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 를 따른다고 가정하자. 관찰점들 개수  $n$ 과 모수  $\mu$ 를 알고 있다는 조건 하에서, Jeffreys의 사전확률분포를 구해보자.

조건부확률변수  $x_j | \sigma$ 의 확률밀도함수  $f(x_j | \sigma)$ 는 다음과 같다.

$$f(x_j | \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{[x_j - \mu]^2}{2\sigma^2} \right) \quad (1)$$

관찰벡터  $\mathbf{x} \doteq [x_1, x_2, \dots, x_n]^T$ 의 우도함수  $f(\mathbf{x} | \sigma)$ 는 다음과 같다.

$$f(\mathbf{x} | \sigma) = \frac{1}{[\sqrt{2\pi}\sigma]^n} \exp \left( -\frac{1}{2\sigma^2} \sum_{j=1}^n [x_j - \mu]^2 \right) \quad (2)$$

따라서, 대수우도함수는 다음과 같다.

$$\ln f(\mathbf{x} | \sigma) = -n \ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{j=1}^n [x_j - \mu]^2 \quad (3)$$

다음 식이 성립한다.

$$\frac{d^2 \ln f(\mathbf{x} | \sigma)}{d\sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_{j=1}^n [x_j - \mu]^2 \quad (4)$$

따라서, 모수  $\sigma$ 의 Fisher정보량은 다음과 같다.

$$I(\sigma) = -E \left( \frac{d^2 \ln f(x | \sigma)}{d\sigma^2} \middle| \sigma \right) = \frac{2n}{\sigma^2} \quad (5)$$

식 (2.5.5)와 식 (5)에서 알 수 있듯이, Jeffreys의 사전확률밀도함수는 다음과 같다.

$$\pi(\sigma) \propto \frac{1}{\sigma} \quad (6)$$

■

**예제 2.5.9** 확률변수들  $x_1, x_2, \dots, x_n$ 이 서로 독립이고 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 를 따른다고 가정하자. 관찰점들 개수  $n$ 을 알고 있고 모수들  $\mu$ 와  $\sigma$ 를 모두 모른다는 조건 하에서, Jeffreys의 사전확률분포를 구해보자.

모수벡터를  $\boldsymbol{\theta} = [\mu, \sigma]^T$ 로 표기하자. 조건부확률변수  $x_j | \boldsymbol{\theta}$ 의 확률밀도함수  $f(x_j | \boldsymbol{\theta})$ 는 다음과 같다.

$$f(x_j | \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{[x_j - \mu]^2}{2\sigma^2}\right) \quad (1)$$

관찰벡터  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ 에 대한 모수벡터  $\boldsymbol{\theta}$ 의 우도함수  $L(\boldsymbol{\theta} | \mathbf{x})$ 는 다음과 같다.

$$f(\mathbf{x} | \boldsymbol{\theta}) = \frac{1}{[\sqrt{2\pi}\sigma]^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n [x_j - \mu]^2\right) \quad (2)$$

따라서, 대수우도함수는 다음과 같다.

$$\ln f(\mathbf{x} | \boldsymbol{\theta}) = -n \ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{j=1}^n [x_j - \mu]^2 \quad (3)$$

다음 식들이 성립한다.

$$\frac{\partial^2 \ln f(x | \boldsymbol{\theta})}{\partial \mu^2} = -\frac{n}{\sigma^2} \quad (4)$$

$$\frac{\partial^2 \ln f(x | \boldsymbol{\theta})}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_{j=1}^n [x_j - \mu]^2 \quad (5)$$

$$\frac{\partial^2 \ln f(x | \boldsymbol{\theta})}{\partial \mu \partial \sigma} = -\frac{2}{\sigma^3} \sum_{j=1}^n [x_j - \mu] \quad (6)$$

따라서, 모수벡터  $\boldsymbol{\theta}$ 의 Fisher 정보량은 다음과 같다.

$$I(\boldsymbol{\theta}) = \begin{bmatrix} n/\sigma^2 & 0 \\ 0 & 2n/\sigma^2 \end{bmatrix} \quad (8)$$

식 (2.5.5)와 식 (8)에서 알 수 있듯이, Jeffreys의 사전확률밀도함수는 다음과 같다.

$$\pi(\boldsymbol{\theta}) \propto |\det(I(\boldsymbol{\theta}))|^{1/2} \propto \frac{1}{\sigma^2} \quad (9)$$

이 사전확률밀도함수는  $\mu$ 에 의존하지 않음에 유의하라. ■

### 2.5.3 조건부공액

일반화선형모형과 같이 복잡한 모형이나 모수벡터  $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_m]^T$ 의 차원  $m$ 이 큰 경우에는, 모수벡터의 공액사전확률분포를 찾는 것이 쉽지 않다. 그러나, 이러한 경우에도 다른 모수들이 주어진 조건 하에서 모수  $\theta_j$ 의 공액인 조건부확률분포를 찾을 수 있다. 모수  $\theta_j$ 에 대해서, 다음과 같은 벡터를 정의하자.

$$\boldsymbol{\theta}_{-j} \doteq [\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_m] \quad (2.5.9)$$

사전조건부확률분포  $\pi(\theta_j | \boldsymbol{\theta}_{-j})$ 와 사후조건부확률분포  $\pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{y})$ 가 같은 확률분포족에 속할 때, 이 사전조건부확률분포를 조건부공액사전확률분포라고 한다. 즉, 우도함수  $f(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\phi})$ 의 모임  $\mathcal{F}$ 와 모수벡터  $\boldsymbol{\theta}$ 의 사전조건부확률밀도함수  $\pi(\boldsymbol{\theta} | \boldsymbol{\phi})$ 의 모임  $\mathcal{P}$ 에 대해서 다음 식이

성립하면, 모임  $\mathcal{P}$ 는 모임  $\mathcal{F}$ 에 대해서 조건부공액(conditional conjugate)이라 한다.

$$f(x | \boldsymbol{\theta}, \boldsymbol{\phi}) \in \mathcal{F} \text{ and } \pi(\boldsymbol{\theta} | \boldsymbol{\phi}) \in \mathcal{P} \Rightarrow \pi(\boldsymbol{\theta} | \mathbf{x}, \boldsymbol{\phi}) \in \mathcal{P} \quad (2.5.10)$$

조건부공액이라는 개념은 Gibbs샘플러를 구축하는 데 유용하다.

**예제 2.5.10** 모수들이 주어진 조건 하에서 서로 독립인 확률변수들  $y_1, y_2, \dots, y_n$ 이 다음 식들을 만족한다고 하자.

$$y_j \mid (\beta_1, \dots, \beta_K, \sigma^2) \stackrel{d}{\sim} \mathcal{N}\left(\sum_{k=1}^K \beta_k x_{jk}, \sigma^2\right), \quad (j = 1, 2, \dots, n) \quad (1)$$

따라서 우도함수는 다음과 같다.

$$f(\mathbf{y} | \boldsymbol{\beta}, \psi) \propto \psi^{n/2} \exp\left(-\frac{\psi}{2} \left\{ [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]^T X^T X [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}] + S_e \right\}\right) \quad (2)$$

여기서  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_K]^T$ 이고  $\psi = 1/\sigma^2$ 이며, 또한 다음 식들이 성립한다.

$$\hat{\boldsymbol{\beta}} = [X^T X]^{-1} X^T \mathbf{y}, \quad S_e = [\mathbf{y} - X\hat{\boldsymbol{\beta}}]^T [\mathbf{y} - X\hat{\boldsymbol{\beta}}] \quad (3)$$

모수들  $\boldsymbol{\beta}$ 와  $\psi$ 의 사전확률분포들이 각각 다음과 같다고 하자.

$$\boldsymbol{\beta} \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_0, B_0) \quad (4)$$

$$\psi \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \quad (5)$$

또한,  $\boldsymbol{\beta}$ 와  $\psi$ 는 독립이라고 가정하자. 따라서, 사전결합확률밀도함수는 다음과 같다.

$$\begin{aligned} \pi(\boldsymbol{\beta}, \psi) &= \pi(\boldsymbol{\beta})\pi(\psi) \\ &\propto \exp\left(-\frac{1}{2} [\boldsymbol{\beta} - \mathbf{b}_0]^T B_0^{-1} [\boldsymbol{\beta} - \mathbf{b}_0]\right) \psi^{n_0/2-1} \exp\left(-\frac{n_0 S_0 \psi}{2}\right) \end{aligned} \quad (6)$$

식 (4)와 독립성에서 알 수 있듯이, 조건부확률벡터  $\boldsymbol{\beta} | \psi$ 의 사전조건부확률분포는 다음과 같다.

$$\boldsymbol{\beta} | \psi \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_0, B_0) \quad (7)$$

식 (2)와 식 (6)에서 알 수 있듯이, 사후확률밀도함수는 다음과 같다.

$$\begin{aligned} \pi(\boldsymbol{\beta}, \psi | \mathbf{y}) &\propto f(\mathbf{y} | \boldsymbol{\beta}, \psi) \pi(\boldsymbol{\beta}) \pi(\psi) \\ &\propto \psi^{[n+n_0]/2-1} \exp\left(-\frac{\psi}{2} \{[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]^T X^T X [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}] + Se + n_0 S_0\}\right) \\ &\quad \cdot \exp\left(-\frac{1}{2} [\boldsymbol{\beta} - \mathbf{b}_0]^T B_0^{-1} [\boldsymbol{\beta} - \mathbf{b}_0]\right) \end{aligned} \quad (8)$$

식 (6)과 식 (8)에서 알 수 있듯이, 이 사전확률분포와 사후확률분포는 공액이 아니다.

조건부모수  $\boldsymbol{\beta} | \psi$ 의 사후조건부확률분포를 살펴보자. 식 (8)에서 알 수 있듯이, 조건부모수  $\boldsymbol{\beta} | \psi$ 의 사후조건부확률밀도함수는 다음과 같다.

$$\pi(\boldsymbol{\beta} | \psi, \mathbf{y}) \propto \exp\left(-\frac{1}{2} \left\{ \psi [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]^T X^T X [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}] + [\boldsymbol{\beta} - \mathbf{b}_0]^T B_0^{-1} [\boldsymbol{\beta} - \mathbf{b}_0] \right\}\right) \quad (9)$$

다음 벡터와 행렬을 정의하자.

$$\mathbf{b}_1 \doteq B_1 [B_0^{-1} \mathbf{b}_0 + \psi X^T \mathbf{y}], \quad B_1 \doteq [B_0^{-1} + \psi X^T X]^{-1} \quad (10)$$

식 (10)을 식 (9)에 대입하면,  $\boldsymbol{\beta} | (\psi, \mathbf{y})$ 의 확률밀도함수가 다음과 같음을 알 수 있다.

$$\pi(\boldsymbol{\beta} | \psi, \mathbf{y}) \propto \exp\left(-\frac{1}{2} [\boldsymbol{\beta} - \mathbf{b}_1]^T B_1^{-1} [\boldsymbol{\beta} - \mathbf{b}_1]\right) \quad (11)$$

즉, 다음 식이 성립한다.

$$\boldsymbol{\beta} | (\psi, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_1, B_1) \quad (12)$$

식 (7)과 식 (12)에서 알 수 있듯이,  $\boldsymbol{\beta}$ 의 사전조건부확률분포와 사후조건부확률분포는 같은 다변량정규확률분포족에 속한다. 따라서, 이들은 조건부공액이다.

조건부모수  $\psi | (\boldsymbol{\beta}, \mathbf{y})$ 의 사후조건부확률분포를 살펴보자. 식 (5)와 독립성에서 알 수 있듯이, 조건부확률변수  $\psi | \boldsymbol{\beta}$ 의 사전조건부확률분포는 다음과 같다.

$$\psi | \boldsymbol{\beta} \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \quad (13)$$

식 (8)에서 알 수 있듯이,  $\psi | (\boldsymbol{\beta}, \mathbf{y})$ 의 사후조건부확률밀도함수는 다음과 같다.

$$\pi(\psi | \boldsymbol{\beta}, \mathbf{y}) \propto \psi^{[n+n_0]/2-1} \exp\left(-\frac{\psi}{2} \left\{ [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]^T X^T X [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}] + S_e + n_0 S_0 \right\}\right) \quad (14)$$

다음 상수와 행렬을 정의하자.

$$n_1 \doteq n + n_0, \quad S_1 \doteq \frac{1}{n_1} \left\{ [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]^T X^T X [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}] + S_e + n_0 S_0 \right\} \quad (15)$$

식 (15)를 식 (14)에 대입하면,  $\psi | (\boldsymbol{\beta}, \mathbf{y})$ 의 사후조건부확률밀도함수가 다음과 같음을 알 수 있다.

$$\pi(\psi | \boldsymbol{\beta}, \mathbf{y}) \propto \psi^{n_1/2-1} \exp\left(-\frac{n_1 S_1}{2} \psi\right) \quad (16)$$

즉, 다음 식이 성립한다.

$$\psi | (\boldsymbol{\beta}, \mathbf{y}) \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_1}{2}, \frac{n_1 S_1}{2}\right) \quad (17)$$

식 (17)의 사후조건부확률분포는 식 (13)의 사전조건부확률분포와 같은 감마확률분포이므로, 이들은 조건부공액이다. ■

우도함수  $f(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\phi})$ 의 모임  $\mathcal{F}$ 와 모수벡터  $\boldsymbol{\phi}$ 가 주어진 조건 하에서 모수벡터  $\boldsymbol{\theta}$ 의 사전조건부확률밀도함수  $\pi(\boldsymbol{\theta} | \boldsymbol{\phi})$ 의 모임  $\mathcal{P}$ 에 대해서 다음 식이 성립하면, 모임  $\mathcal{P}$ 는 모임  $\mathcal{F}$ 에 대해서 세미공액 (semi-conjugate)이라 한다.

$$\begin{aligned} f(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\phi}) \in \mathcal{F} \text{ and } \pi(\boldsymbol{\theta}, \boldsymbol{\phi}) = \pi(\boldsymbol{\theta})\pi(\boldsymbol{\phi}) \text{ for } \pi(\boldsymbol{\theta}) \in \mathcal{P} \text{ and } \pi(\boldsymbol{\phi}) \notin \mathcal{P} \\ \Rightarrow \pi(\boldsymbol{\theta} | \mathbf{x}, \boldsymbol{\phi}) \in \mathcal{P} \end{aligned} \quad (2.5.11)$$

**예제 2.5.11** 제2.3.3소절에서 다룬 문제를 다시 생각해보자. 즉, 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 를 따르는 확률변수들  $x_1, x_2, \dots, x_n$ 이 서로 독립이라고 가정하고, 평균  $\mu$ 와 분산  $\sigma^2$ 를 모두 추정하기로 하자.

분산  $\sigma^2$ 의 사전확률분포는 역감마확률분포이고,  $\sigma^2$ 가 주어진 조건 하에서 평균  $\mu$ 의

조건부확률분포가 정규확률분포라고 하자. 즉, 다음 식들이 성립한다고 가정하자.

$$\sigma^2 \stackrel{d}{\sim} \text{IG} \left( \frac{\nu_0}{2}, \frac{\nu_0 S_0}{2} \right) \quad (1)$$

$$\mu \mid \sigma^2 \stackrel{d}{\sim} \mathcal{N} \left( \mu_0, \frac{\sigma^2}{\kappa_0} \right) \quad (2)$$

제2.3.3소절에서 알 수 있듯이, 평균  $\mu$ 와 분산  $\sigma^2$ 의 사후결합확률밀도함수는 다음과 같다.

$$\pi(\mu, \sigma^2 \mid \mathbf{x}) \propto \frac{1}{\sigma} [\sigma^2]^{-\nu_*/2-1} \exp \left( -\frac{1}{2\sigma^2} \left\{ \nu_* S_* + \kappa_* [\mu - \mu_*]^2 \right\} \right) \quad (3)$$

여기서 각 모수는 다음과 같다.

$$\mu_* \doteq \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{x} \quad (4)$$

$$\kappa_* \doteq \kappa_0 + n, \quad \nu_* \doteq \nu_0 + n \quad (5)$$

$$\nu_* S_* \doteq \nu_0 S_0 + [n - 1]s^2 + \frac{\kappa_0 n}{\kappa_0 + n} [\bar{x} - \mu_0]^2 \quad (6)$$

지금부터는 평균모수와 분산모수가 서로 독립이라고 가정하자. 즉, 다음 식들이 성립한다고 가정하자.

$$\mu \stackrel{d}{\sim} \mathcal{N}(\mu_0, \tau_0) \quad (7)$$

$$\sigma^2 \stackrel{d}{\sim} \text{IG} \left( \frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2} \right) \quad (8)$$

제2.3.3소절과 같은 방법을 적용하면, 다음 식들이 성립함을 알 수 있다.

$$\mu \mid (\sigma^2, \mathbf{x}) \stackrel{d}{\sim} \mathcal{N}(\mu_s, \tau_s) \quad (9)$$

$$\sigma^2 \mid (\mu, \mathbf{x}) \stackrel{d}{\sim} \text{IG} \left( \frac{\nu_0 + n}{2}, \frac{\nu_0 \sigma_0^2 + \sum_{i=1}^n [x_i - \mu]^2}{2} \right) \quad (10)$$

여기서 모수들은 다음과 같다.

$$\mu_s \doteq \frac{\mu_0/\tau_0^2 + n\bar{x}/\sigma^2}{1/\tau_0^2 + n/\sigma^2}, \quad \tau_s^2 \doteq \left[ \frac{1}{\tau_0^2} + \frac{n}{\sigma^2} \right]^{-1} \quad (11)$$

식 (9)와 식 (10)은 수리적으로 정교한 형태이지만, 이들로부터 사후결합확률밀도함수를

해석적으로 구하는 것은 쉽지 않은 일이다. 그러나, Gibbs샘플러를 사용해서 사후결합확률밀도함수로부터 난수벡터를 생성할 수 있다. 즉, 몬테카를로실험을 통해서 통계적 추론을 할 수 있다. ■

## 제2.6절 사후확률분포와 통계적 추론

### 2.6.1 사후주변확률분포

모수공간  $\Theta \subset \mathbb{R}^d$ 에서 정의되는 사후확률분포  $\pi(\boldsymbol{\theta} | \mathbf{y})$ 의 사후주변확률분포는 다음과 같다.

$$\pi(\theta_i | \mathbf{y}) = \int_{\Theta_{-i}} \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_{-i}, \quad (i = 1, 2, \dots, d) \quad (2.6.1)$$

여기서  $\boldsymbol{\theta}_{-i}$ 는 모수벡터  $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_d]^T$ 에서 제*i*번째 모수  $\theta_i$ 를 제거한 벡터이다. 즉,  $\boldsymbol{\theta}_{-i} = [\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d]^T$ 이다. 또한,  $\Theta_{-i}$ 는 모수벡터  $\boldsymbol{\theta}_{-i}$ 의 모수공간이다.

**예제 2.6.1** 제2.3.3소절에서 다룬 정규확률분포의 평균과 분산에 관한 문제에서 사후주변확률분포를 살펴보자.

정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 를 따르는 확률변수들  $x_1, x_2, \dots, x_n$ 이 서로 독립이라고 가정하고, 평균  $\mu$ 와 분산  $\sigma^2$ 를 모두 추정하기로 하자. 분산  $\sigma^2$ 의 사전확률분포는 역감마확률분포이고,  $\sigma^2$ 가 주어진 조건 하에서 평균  $\mu$ 의 조건부확률분포가 정규확률분포라고 하자. 즉, 다음 식들이 성립한다고 가정하자.

$$\sigma^2 \stackrel{d}{\sim} \text{IG} \left( \frac{\nu_0}{2}, \frac{\nu_0 S_0}{2} \right) \quad (1)$$

$$\mu | \sigma^2 \stackrel{d}{\sim} \mathcal{N} \left( \mu_0, \frac{\sigma^2}{\kappa_0} \right) \quad (2)$$

제2.3.3소절에서 알 수 있듯이, 평균  $\mu$ 와 분산  $\sigma^2$ 의 사후결합확률밀도함수는 다음과 같다.

$$\pi(\mu, \sigma^2 | \mathbf{x}) \propto \frac{1}{\sigma} [\sigma^2]^{-\nu_*/2-1} \exp \left( -\frac{1}{2\sigma^2} \left\{ \nu_* S_* + \kappa_* [\mu - \mu_*]^2 \right\} \right) \quad (3)$$

여기서 각 모수는 다음과 같다.

$$\mu_* \doteq \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{x} \quad (4)$$

$$\kappa_* \doteq \kappa_0 + n, \quad \nu_* \doteq \nu_0 + n \quad (5)$$

$$\nu_* S_* \doteq \nu_0 S_0 + [n - 1]s^2 + \frac{\kappa_0 n}{\kappa_0 + n} [\bar{x} - \mu_0]^2 \quad (6)$$

분산  $\sigma^2$ 가 주어진 조건 하에서  $\mu$ 의 사후조건부확률밀도함수  $\pi(\mu | \sigma^2, \mathbf{x})$ 는 분산  $\sigma^2$ 를 상수로 하는 사후결합확률밀도함수에 비례한다. 식 (3)에서 알 수 있듯이, 이에 해당하는 사후확률분포는 다음과 같다.

$$\mu | (\sigma^2, \mathbf{x}) \stackrel{d}{\sim} \mathcal{N}\left(\mu_*, \frac{\sigma^2}{\kappa_*}\right) \quad (7)$$

또한  $\sigma^2$ 의 사후주변확률분포  $\pi(\sigma^2 | \mathbf{x})$ 는 다음과 같이 자유도가  $\nu_*$ 인 척도화역카이제곱확률분포이다.

$$\sigma^2 | \mathbf{x} \stackrel{d}{\sim} Inv - \chi^2(\nu_*, S_*)$$

식 (3)의 사후결합확률밀도함수를  $\sigma^2$ 에 관해 적분하면,  $\mu$ 에 대한 사후주변확률밀도함수가 다음과 같음을 알 수 있다.

$$\pi(\mu | \mathbf{x}) \propto \left\{ 1 + \frac{1}{\nu_*} \left[ \frac{\mu - \mu_*}{\sqrt{S_*/\kappa_*}} \right]^2 \right\}^{-[\nu_*+1]/2} \quad (8)$$

즉,  $\frac{\mu - \mu_*}{\sqrt{S_*/\kappa_*}}$ 의 사후주변확률분포는 자유도가  $\nu_*$ 인  $t$  확률분포이다. ■

## 2.6.2 점추정

사후확률분포를 바탕으로 대표값들을 구하기로 하자. 첫째, 사후평균 (posterior mean)은 다음과 같다.

$$\hat{\boldsymbol{\theta}}^{mean} = \int_{\Theta} \boldsymbol{\theta} \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \quad (2.6.2)$$

사후확률분포의 모드를 사후모드 (posterior mode) 라고 한다. 즉, 사후모드는 다음과 같다.

$$\hat{\boldsymbol{\theta}}^{mode} = \arg \max_{\boldsymbol{\theta}} \pi(\boldsymbol{\theta} | \mathbf{y}) \quad (2.6.3)$$

따라서 사후모드를 최대사후확률추정량 (maximum a posterior estimator) 이라 부른다. 사후확률분포의 중앙값을 사후중앙값 (posterior median) 이라고 한다. 즉, 사후중앙값  $\hat{\boldsymbol{\theta}}^{median}$  의 제  $i$  번째 원소  $\hat{\theta}_i^{median}$  는 다음 식을 만족한다.

$$\int_{-\infty}^{\hat{\theta}_i^{median}} \pi(\theta_i | \mathbf{y}) d\theta_i = \frac{1}{2}, \quad (i = 1, 2, \dots, d) \quad (2.6.4)$$

**예제 2.6.2** 예제 2.6.1을 다시 살펴보자. 즉, 베이저안통계학을 이용하여 정규확률분포의 대표값을 추정하기로 하자.

예제 2.6.1에서 알 수 있듯이, 평균  $\mu$ 와 분산  $\sigma^2$ 의 사후결합확률밀도함수는 다음과 같다.

$$\pi(\mu, \sigma^2 | \mathbf{x}) \propto \frac{1}{\sigma} [\sigma^2]^{-\nu_*/2-1} \exp\left(-\frac{1}{2\sigma^2} \left\{ \nu_* S_* + \kappa_* [\mu - \mu_*]^2 \right\}\right) \quad (1)$$

여기서 각 모수는 다음과 같다.

$$\mu_* \doteq \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{x} \quad (2)$$

$$\kappa_* \doteq \kappa_0 + n, \quad \nu_* \doteq \nu_0 + n \quad (3)$$

$$\nu_* S_* \doteq \nu_0 S_0 + [n - 1]s^2 + \frac{\kappa_0 n}{\kappa_0 + n} [\bar{x} - \mu_0]^2 \quad (4)$$

또한,  $\mu$ 의 사후주변확률밀도함수는 다음과 같다.

$$\pi(\mu | \mathbf{x}) \propto \left\{ 1 + \frac{1}{\nu_*} \left[ \frac{\mu - \mu_*}{\sqrt{S_*/\kappa_*}} \right]^2 \right\}^{-[\nu_*+1]/2} \quad (5)$$

즉,  $\frac{\mu - \mu_*}{\sqrt{S_*/\kappa_*}}$ 의 사후주변확률분포는 자유도가  $\nu_*$ 인  $t$  확률분포이다. 따라서, 다음 식들이 성립한다.

$$\hat{\mu}^{mean} = \hat{\mu}^{mode} = \hat{\mu}^{median} = \mu_* \quad (6)$$

다음 식들이 성립한다.

$$\lim_{\kappa_0 \rightarrow \infty} \hat{\mu}^{mean} = \lim_{\kappa_0 \rightarrow \infty} \hat{\mu}^{mode} = \lim_{\kappa_0 \rightarrow \infty} \hat{\mu}^{median} = \bar{x} \quad (7)$$

즉, 식  $\kappa_0 \rightarrow \infty$ 이 성립하는 경우에는 베이저안통계분석에 의한 대표값들은 사전확률분포의 최우추정값인 표본평균에 수렴한다. ■

**예제 2.6.3** B당의 대통령후보 중 1명인 A는 본인의 IQ는 평균이  $\theta$ 이고 분산이 80인 정규확률분포를 따른다고 생각한다. 창당 당시 조사한 B당 당원들의 IQ  $\theta$ 는 평균이 110이고 분산이 120인 정규확률분포를 따른다고 알려져 있다. 자신이 뛰어났다고 생각한 A가 공개적으로 IQ테스트를 받은 결과 IQ가 98이라는 보고서를 받았다. 이 경우 관찰점이 한 개라는 것에 유의하라. 모수  $\theta$ 의 고전적 추정량은  $\hat{\theta}_F = 98$ 이고, 사후확률분포는  $\mathcal{N}(102.8, 48)$ 이다. 따라서, A씨 IQ의 Bayes추정값은  $\hat{\theta}_B = 102.8$ 이다. ■

**예제 2.6.4** 예제 2.5.4를 다시 생각해보자.

확률변수들  $x_1, x_2, \dots, x_n$ 이 서로 독립이고 평균이  $\lambda$ 인 Poisson 확률분포  $Poisson(\lambda)$ 를 따른다고 하면, 조건부확률변수  $x_i | \lambda$ 의 확률밀도함수  $f(x_i | \lambda)$ 는 다음과 같다.

$$f(x_i | \lambda) = \exp(-\lambda) \frac{\lambda^{x_i}}{x_i!} \quad (1)$$

모수  $\lambda$ 의 사전확률분포가 모수들이  $\alpha$ 와  $\beta$ 인 감마확률분포  $Gamma(\alpha, \beta)$ 라고 하자. 여기서  $\alpha$ 와  $\beta$ 는 알고있는 상수들이다. 모수  $\lambda$ 의 사전확률밀도함수는 다음과 같다.

$$\pi(\lambda) = \frac{1}{\Gamma(\alpha)} \beta^\alpha \lambda^{\alpha-1} \exp(-\beta\lambda), \quad (\lambda > 0) \quad (2)$$

예제 2.5.4에서 알 수 있듯이, 모수  $\lambda$ 의 사후확률분포를 다음과 같이 표기할 수 있다.

$$\lambda | \mathbf{x} \stackrel{d}{\sim} Gamma \left( \alpha + \sum_{i=1}^n x_i, \beta + n \right) \quad (3)$$

따라서, 다음 식들이 성립한다.

$$\hat{\lambda}^{mean} = \frac{\alpha + \sum_{i=1}^n x_i}{\beta + n}, \quad \hat{\lambda}^{mode} = \frac{\alpha - 1 + \sum_{i=1}^n x_i}{\beta + n}. \quad (4)$$

■

**예제 2.6.5** 확률변수들  $x_1, x_2, \dots, x_n$  이 서로 독립이고 모수들이  $\alpha$ 와  $\beta$ 인 베타확률분포  $Beta(\alpha, \beta)$ 를 따른다고 하자.

관찰벡터  $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$ 가 주어진 조건 하에서 우도함수는 다음과 같다.

$$L(\alpha, \beta | \mathbf{x}) = \prod_{j=1}^n \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x_j^{\alpha-1} [1 - x_j]^{\beta-1} \quad (1)$$

따라서 대수우도함수는 다음과 같다.

$$\begin{aligned} \ln L(\alpha, \beta | \mathbf{x}) &= n \ln \Gamma(\alpha + \beta) - n \ln \Gamma(\alpha) - n \ln \Gamma(\beta) \\ &\quad + [\alpha - 1] \sum_{j=1}^n \ln x_j + [\beta - 1] \sum_{j=1}^n \ln(1 - x_j) \end{aligned} \quad (2)$$

즉, 정규방정식은 다음과 같다.

$$\frac{\partial \ln L(\alpha, \beta | \mathbf{x})}{\partial \alpha} = \frac{n\Gamma'(\alpha + \beta)}{\Gamma(\alpha + \beta)} - \frac{n\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum_{j=1}^n \ln x_j \quad (3)$$

$$\frac{\partial \ln L(\alpha, \beta | \mathbf{x})}{\partial \beta} = \frac{n\Gamma'(\alpha + \beta)}{\Gamma(\alpha + \beta)} - \frac{n\Gamma'(\beta)}{\Gamma(\beta)} + \sum_{j=1}^n \ln(1 - x_j) \quad (4)$$

여기서 디감마함수(digamma function)  $\psi(a)$ 와 트라이감마함수(trigamma function)  $\psi'(a)$ 는 각각 다음과 같이 정의된다.

$$\psi(a) = \frac{\Gamma'(a)}{\Gamma(a)}, \quad \psi'(a) = \frac{\Gamma''(a)}{\Gamma(a)} - \frac{\Gamma'(a)^2}{\Gamma(a)^2} \quad (5)$$

이 정규방정식의 닫힌해(close solution)는 존재하지 않는다. 따라서, 최우추정값들을

구하기 위해서는 수치적 기법들을 적용해야 한다. 다음 함수들을 정의하자.

$$g_1(\alpha, \beta) \doteq \psi(\alpha) - \psi(\alpha + \beta) - \frac{1}{n} \sum_{j=1}^n \ln x_j \quad (6)$$

$$g_2(\alpha, \beta) \doteq \psi(\beta) - \psi(\alpha + \beta) - \frac{1}{n} \sum_{j=1}^n \ln(1 - x_j) \quad (7)$$

다음과 같이 벡터들과 행렬을 정의하자.

$$\boldsymbol{\eta} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}, \quad G = \begin{bmatrix} \frac{\partial g_1}{\partial \alpha} & \frac{\partial g_1}{\partial \beta} \\ \frac{\partial g_2}{\partial \alpha} & \frac{\partial g_2}{\partial \beta} \end{bmatrix} \quad (8)$$

다음 식들이 성립한다.

$$\frac{\partial g_1(\alpha, \beta)}{\partial \alpha} = \psi'(\alpha) - \psi'(\alpha + \beta) \quad (9)$$

$$\frac{\partial g_1(\alpha, \beta)}{\partial \beta} = -\psi'(\alpha + \beta) \quad (10)$$

$$\frac{\partial g_2(\alpha, \beta)}{\partial \alpha} = -\psi'(\alpha + \beta) \quad (11)$$

$$\frac{\partial g_2(\alpha, \beta)}{\partial \beta} = \psi'(\beta) - \psi'(\alpha + \beta) \quad (12)$$

예를 들어, Newton-Raphson 법을 적용하면 다음 알고리즘을 얻는다.

$$\hat{\boldsymbol{\eta}}_{i+1} = \hat{\boldsymbol{\eta}}_i - G_i^{-1} \mathbf{g}_i \quad (13)$$

여기서 아래첨자  $i$ 는 제  $i$ 번째 반복단계를 의미한다. 초기벡터  $\hat{\boldsymbol{\eta}}_0$ 를 적절히 선택하면  $\{\hat{\boldsymbol{\eta}}_i\}$ 는 2차수렴(quadratic convergence)한다.

적률추정법(method of moments)을 사용해서 모수들을 추정하기로 하자. 다음 식들이 성립한다.

$$E(X_1^r) = \int_0^1 x^r \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + r)}{\Gamma(\alpha)\Gamma(\alpha + \beta + r)} \quad (14)$$

따라서 다음 식들이 성립한다.

$$E(X_1) = \frac{\alpha}{\alpha + \beta} \quad (15)$$

$$E(X_1^2) = \frac{\alpha[\alpha + 1]}{[\alpha + \beta][\alpha + \beta + 1]} \quad (16)$$

$$Var(X_1) = \frac{\alpha\beta}{[\alpha + \beta]^2[\alpha + \beta + 1]} \quad (17)$$

식 (16)에서 알 수 있듯이, 다음 식이 성립한다.

$$\beta = \left[ \frac{1}{E(X_1)} - 1 \right] \alpha \quad (18)$$

식 (18)을 식 (17)에 대입하면, 다음 식을 얻는다.

$$\alpha = E(X_1) \left[ \frac{E(X_1) \{1 - E(X_1)\}}{Var(X_1)} - 1 \right] \quad (19)$$

$$\beta = \{1 - E(X_1)\} \left[ \frac{E(X_1) \{1 - E(X_1)\}}{Var(X_1)} - 1 \right] \quad (20)$$

예제 2.5.2에서 다룬 문제를 다시 생각해보자. 즉, 확률변수들  $x_1, x_2, \dots, x_n$  이 서로 독립이고 모수가  $\theta$ 인 지수확률분포  $Exponential(\theta)$ 를 따른다고 하자. 그리고, 모수  $\theta$ 의 사전확률분포는 모수들이  $\alpha$ 와  $\beta$ 인 베타확률분포  $Beta(\alpha, \beta)$ 라고 하자. 모수  $\theta$ 의 사전확률분포에서 난수들을 발생시켜 추정된 평균추정값과 분산추정값을 각각  $\hat{\mu}_0$ 와  $\hat{\sigma}_0^2$ 라고 하면, 식 (19)와 식 (20)에서 알 수 있듯이  $\alpha$ 와  $\beta$ 의 추정값들은 각각 다음과 같다.

$$\hat{\alpha}_0 = \hat{\mu}_0 \left\{ \frac{\hat{\mu}_0 [1 - \hat{\mu}_0]}{\hat{\sigma}_0^2} - 1 \right\} \quad (21)$$

$$\hat{\beta}_0 = [1 - \hat{\mu}_0] \left\{ \frac{\hat{\mu}_0 [1 - \hat{\mu}_0]}{\hat{\sigma}_0^2} - 1 \right\} \quad (22)$$

모수  $\theta$ 의 사후확률분포에서 난수들을 발생시켜 추정된 평균추정값과 분산추정값을 각각  $\hat{\mu}_1$ 와  $\hat{\sigma}_1^2$ 라고 하면, 식 (19)와 식 (20)에서 알 수 있듯이  $\alpha$ 와  $\beta$ 의 추정값들은 각각 다음과 같다.

$$\hat{\alpha}_1 = \hat{\mu}_1 \left\{ \frac{\hat{\mu}_1 [1 - \hat{\mu}_1]}{\hat{\sigma}_1^2} - 1 \right\} \quad (23)$$

$$\hat{\beta}_1 = [1 - \hat{\mu}_1] \left\{ \frac{\hat{\mu}_1 [1 - \hat{\mu}_1]}{\hat{\sigma}_1^2} - 1 \right\} \quad (24)$$



**예제 2.6.6** 결정이론(decision theory)을 바탕으로 베이지안 대표값들을 설명할 수 있다. 결정이론에서는 손실함수(loss function)의 사후기대값, 즉 사후위험함수를 최소화하는 추정량 또는 결정규칙(decision rule)을 베이지안추정량 또는 Bayes액션(Bayes action)이라 부른다.

첫째, 다음과 같은 모수벡터  $\theta$ 에 대한 이차손실함수(quadratic loss function)를 살펴보자.

$$L(\tilde{\theta}, \theta) = [\tilde{\theta} - \theta]^2 \quad (1)$$

이에 해당하는 위험함수는 다음과 같다.

$$R(\tilde{\theta}, \theta) = E(L(\tilde{\theta}, \theta)) = \int_{\Theta} [\tilde{\theta} - \theta]^2 \pi(\theta | \mathbf{y}) d\theta \quad (2)$$

따라서 다음 식이 성립한다.

$$\frac{dR(\tilde{\theta}, \theta)}{d\tilde{\theta}} = \frac{dE(L(\tilde{\theta}, \theta))}{d\tilde{\theta}} = 2 \int_{\Theta} [\tilde{\theta} - \theta] \pi(\theta | \mathbf{y}) d\theta \quad (3)$$

식 (3)에서 알 수 있듯이, 이차손실함수를 사용하는 위험함수를 최소화하는 추정량, 즉 베이지안추정량은 다음과 같은 사후평균  $\hat{\theta}$ 이다.

$$\tilde{\theta} = \int_{\Theta} \theta \pi(\theta | \mathbf{y}) d\theta = \hat{\theta} \quad (4)$$

둘째, 주어진 양수  $\epsilon$ 에 대해서 다음과 같은 제로원손실함수(zero-one loss function)를 살펴보자.

$$L_{\epsilon}(\tilde{\theta}, \theta) = 1_{\{\|\tilde{\theta} - \theta\| \geq \epsilon\}}(\tilde{\theta}) \quad (5)$$

임의의  $\delta > 0$ 에 대해서 다음 집합을 정의하자.

$$\Omega_{\delta} \doteq \left\{ \theta : \pi(\theta) > \pi(\tilde{\theta}^{mode}) \right\} \quad (6)$$

즉, 충분히 작은 양수  $\delta$ 에 대해서  $\Omega_{\delta}$ 는 모드  $\tilde{\theta}^{mode}$ 에서 열린 근방(neighborhood)이다. 다음

명제를 만족하는  $\epsilon^* > 0$ 가 존재한다.

$$\epsilon < \epsilon^* \Rightarrow N_\epsilon(\tilde{\theta}^{mode}) \quad (7)$$

여기서  $N_\epsilon(\tilde{\theta}^{mode})$ 는 중심이  $\tilde{\theta}^{mode}$ 이고 반경이  $\epsilon$ 인 열린(open)  $\epsilon$ -근방이다. 따라서 각  $\epsilon (< \epsilon^*)$ 에 대해서 다음 식이 성립한다.

$$N_\epsilon(\tilde{\theta}^{mode}(\epsilon)) \cup \Omega_\delta \neq \emptyset \quad (8)$$

여기서  $\tilde{\theta}^{mode}(\epsilon)$ 는 손실함수를  $L_\epsilon(\tilde{\theta}^{mode}, \theta)$ 로 하는 베이지안추정량이다. 집합  $\Omega_\delta$ 의  $\delta > 0$ 를 충분히 작게 할 수 있으므로, 다음 식이 성립한다.

$$\lim_{\delta \rightarrow 0} \tilde{\theta}^{mode}(\epsilon) = \tilde{\theta}^{mode} \quad (9)$$

셋째, 1차원 모수공간에서 정의되는 선형선형손실함수(linear-linear loss function)를 생각해보자. 즉, 다음 함수를 생각해보자.

$$L^{[q]}(\tilde{\theta}, \theta) = [1 - q][\tilde{\theta} - \theta]1_{(-\infty, \tilde{\theta}]}(\theta) + q[\theta - \tilde{\theta}]1_{(\tilde{\theta}, \infty)}(\theta) \quad (10)$$

이에 해당하는 위험함수는 다음과 같다.

$$R^{[q]}(\tilde{\theta}, \theta) = E(L^{[q]}(\tilde{\theta}, \theta)) = [1 - q] \int_{-\infty}^{\tilde{\theta}} [\tilde{\theta} - \theta] \pi(\theta) d\theta + q \int_{\tilde{\theta}}^{\infty} [\theta - \tilde{\theta}] \pi(\theta) d\theta \quad (11)$$

다음 식들이 성립한다.

$$\frac{dR^{[q]}(\tilde{\theta}, \theta)}{d\tilde{\theta}} = [1 - q]\Pr(\theta \leq \tilde{\theta}) - q\Pr(\theta > \tilde{\theta}) \quad (12)$$

$$\frac{d^2 R^{[q]}(\tilde{\theta}, \theta)}{d\tilde{\theta}^2} = \pi(\tilde{\theta}) \quad (13)$$

따라서 1차조건은 다음과 같다.

$$[1 - q]\Pr(\theta \leq \tilde{\theta}) = q\Pr(\theta > \tilde{\theta}) \quad (14)$$

식 (14)를 다음과 같이 쓸 수 있다.

$$\Pr(\theta \leq \tilde{\theta}) = q \quad (15)$$

다음과 같이  $q$  분위수 추정량을 정의하자.

$$\tilde{\theta}^{[q]} \doteq \left\{ \tilde{\theta} : \Pr(\theta \leq \tilde{\theta}) = q \right\} \quad (16)$$

만약 식  $\pi(\tilde{\theta}^{[q]}) > 0$ 가 성립하면, 식 (13)에서 알 수 있듯이  $\tilde{\theta}^{[q]}$ 는  $L^{[q]}(\tilde{\theta}, \theta)$ 를 손실함수로 하는 유일한 베이저안추정량이다. 특히,  $q = 0.5$ 일 때  $\tilde{\theta}^{[0.5]}$ 는 사후중앙값이다. 식 (11)에서 알 수 있듯이, 베이저안추정량  $\tilde{\theta}^{[q]}$ 의 위험함수는 다음과 같다.

$$R^{[q]}(\tilde{\theta}^{[q]}, \theta) = q[1 - q] \left\{ E\left(\tilde{\theta}^{[q]} - \theta \mid \theta \leq \tilde{\theta}^{[q]}\right) + E\left(\theta - \tilde{\theta}^{[q]} \mid \theta > \tilde{\theta}^{[q]}\right) \right\} \quad (17)$$

■

### 2.6.3 구간추정

고전통계학에서는 신뢰수준이  $1 - \alpha$ 인 신뢰구간  $[l, u]$ 가 다음과 같이 정의된다.

$$\int_l^u \pi(\theta | \mathbf{y}) d\theta = 1 - \alpha \quad (2.6.5)$$

베이저안통계학에서는 집합  $C \subset \Theta$ 가 다음 식을 만족하면, 신뢰수준이  $1 - \alpha$ 인 신용집합(credit set)이라 부른다.

$$\Pr(\theta \in C | \mathbf{y}) = E(1_C(\theta) | \mathbf{y}) = \int_C \pi(\theta | \mathbf{y}) d\mathbf{y} \geq 1 - \alpha \quad (2.6.6)$$

만약 사후확률분포가 이산형이면, 신뢰수준이  $1 - \alpha$ 인 신용집합은 다음과 같다.

$$\Pr(\theta \in C | \mathbf{y}) = E(1_C(\theta) | \mathbf{y}) = \sum_{\theta_i \in C} \pi(\theta_i | \mathbf{y}) \geq 1 - \alpha \quad (2.6.7)$$

어떤 사후확률분포에 대해서 신뢰수준이  $1 - \alpha$ 인 신용집합은 여러 개 존재한다.

신뢰수준  $1 - \alpha$ 가 미리 정해졌을 때, 구간길이를 최소화하는 신용집합은 다음과 같이 정의

되는 최고사후확률밀도구간 (highest posterior probability density interval: HPD interval)  $I_{HPD}$ 이다.

$$I_{HPD} \doteq \{\theta \mid \pi(\theta \mid \mathbf{y}) \geq k(\alpha)\} \quad (2.6.8)$$

여기서  $k(\alpha)$ 는 다음 식을 만족한다.

$$\int_{I_{HPD}} \pi(\theta \mid \mathbf{y}) d\theta \geq 1 - \alpha \quad (2.6.9)$$

**예제 2.6.7** 예제 2.6.3을 다시 생각해보자. B당 대통령후보 중 1명인 A는 본인의 IQ는 평균이  $\theta$ 이고 분산이 80인 정규확률분포를 따른다고 생각한다. 창당 당시 조사한 B당 당원들의 IQ  $\theta$ 는 평균이 110이고 분산이 120인 정규확률분포를 따른다고 알려져 있다. 자신이 뛰어났다고 생각한 A가 공개적으로 IQ테스트를 받은 결과 IQ가 98이라는 보고서를 받았다. 이 경우 관찰점이 한 개라는 점에 유의하라. 모수  $\theta$ 의 고전적 추정량은  $\hat{\theta}_F = 98$ 이고, 신뢰수준이 95%인 신뢰구간은  $[98 - 1.96\sqrt{80}, 98 + 1.96\sqrt{80}] = [80.4692, 115.5308]$ 이다. 따라서 이 신뢰구간의 길이는 35.06이다. 사후확률분포는  $\mathcal{N}(102.8, 48)$ 이다. 따라서, A씨 IQ의 베이즈추정값은  $\hat{\theta}_B = 102.8$ 이고 신뢰수준이 95%인 최고사후확률밀도구간  $I_{HPD}$ 는  $[102.8 - 1.96\sqrt{48}, 102.8 + 1.96\sqrt{48}] = [89.2207, 116.3793]$ 이고, 이 신뢰구간의 길이는 27.16이다. ■

**예제 2.6.8** 제2.3.1소절에서 다룬 문제를 다시 살펴보자. 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 를 따르는 확률변수들  $x_1, x_2, \dots, x_n$ 이 서로 독립이며 분산  $\sigma^2$ 는 알려졌다고 가정하고, 평균  $\mu$ 의 최고사후확률밀도구간을 추정하기로 하자.

평균  $\mu$ 의 사전확률분포가 다음과 같다고 가정하자.

$$\mu \stackrel{d}{\sim} \mathcal{N}(\mu_0, \sigma_0^2) \quad (1)$$

여기서  $\mu_0$ 와  $\sigma_0^2$ 는 알려진 상수들이다. 관찰벡터  $\mathbf{x} \doteq [x_1, x_2, \dots, x_n]^T$ 에 대한 모수  $\mu$ 의

우도함수  $L(\mu | \mathbf{x}) = f(\mathbf{x} | \mu)$ 는 다음과 같다.

$$L(\mu | \mathbf{x}) = \frac{1}{[\sqrt{2\pi}\sigma]^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n [x_j - \mu]^2\right) \quad (2)$$

따라서  $\mu$ 의 사후확률밀도함수가 다음과 같음을 알 수 있다.

$$\pi(\mu | \mathbf{x}) \propto \exp\left(-\frac{1}{2\sigma_*^2} [\mu - \mu_*]^2\right) \quad (3)$$

여기서  $\bar{x}$ ,  $\mu_*$ 와  $\sigma_*^2$ 는 각각 다음과 같다.

$$\bar{x} \doteq \frac{1}{n} \sum_{j=1}^n x_j \quad (4)$$

$$\mu_* \doteq \frac{\bar{x}n/\sigma^2 + \mu_0/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2} = \frac{n/\sigma^2}{n/\sigma^2 + 1/\sigma_0^2} \bar{x} + \frac{1/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2} \mu_0 \quad (5)$$

$$\sigma_*^2 \doteq \frac{1}{n/\sigma^2 + 1/\sigma_0^2} \quad (6)$$

즉, 평균  $\mu$ 의 사후확률분포는 다음과 같다.

$$\mu | \mathbf{x} \stackrel{d}{\sim} \mathcal{N}(\mu_*, \sigma_*^2) \quad (7)$$

따라서 신뢰수준이  $1 - \alpha$ 인 최고사후확률밀도구간  $I_{HPD}$ 는 다음과 같다.

$$I_{HPD} = [\mu_* - z_{\alpha/2}\sigma_*, \mu_* + z_{\alpha/2}\sigma_*] \quad (8)$$

여기서  $z_\alpha$ 는 확률  $\alpha$ 에 해당하는  $z$ 스코어( $z$ -score)로서 식  $N(z_\alpha) = \alpha$ 를 만족한다. ■

#### 2.6.4 가설검정

다음과 같은 가설검정문제를 살펴보자.

$$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1 \quad (2.6.10)$$

여기서  $\Theta_0 \cap \Theta_1 = \emptyset$ 이다. 사전오즈비 (prior odds ratio)와 관찰벡터  $\mathbf{x}$ 가 주어졌을 때 사후 오즈비 (posterior odds ratio)는 각각 다음과 같다.

$$O_{prior} = \frac{\Pr(H_0)}{\Pr(H_1)}, \quad O_{posterior} = \frac{\Pr(H_0|\mathbf{x})}{\Pr(H_1|\mathbf{x})} \quad (2.6.11)$$

가설  $H_i$  하에서 주변우도 (marginal likelihood)를 다음과 같이 표기하자.

$$m(\mathbf{x} | H_i) = \int_{\Theta_i} f(\mathbf{x} | \theta)\pi(\theta|H_i) d\theta, \quad (i = 0, 1) \quad (2.6.12)$$

Bayes정리에서 알 수 있듯이, 사후확률은 다음과 같다.

$$\Pr(H_i|\mathbf{x}) = \frac{m(\mathbf{x} | H_i)\Pr(H_i)}{m(\mathbf{x} | H_0)\Pr(H_0) + m(\mathbf{x} | H_1)\Pr(H_1)}, \quad (i = 1, 2) \quad (2.6.13)$$

만약 식  $\Pr(H_0|\mathbf{x}) < \Pr(H_1|\mathbf{x})$ 가 성립하면, 귀무가설  $H_0$ 를 기각한다.

다음과 같이 Bayes팩터 (Bayes factor)를 정의하자.

$$K_{01} = \frac{O_{posterior}}{O_{prior}} = \frac{\frac{\Pr(H_0|\mathbf{x})}{\Pr(H_1|\mathbf{x})}}{\frac{\Pr(H_0)}{\Pr(H_1)}} \quad (2.6.14)$$

식 (2.6.13)와 식 (2.6.14)에서 알 수 있듯이, 다음 식이 성립한다.

$$K_{01} = \frac{m(\mathbf{x} | H_0)}{m(\mathbf{x} | H_1)} \quad (2.6.15)$$

또한 다음 식이 성립한다.

$$\frac{\Pr(H_0|\mathbf{x})}{\Pr(H_1|\mathbf{x})} = \frac{\Pr(H_0)}{\Pr(H_1)} K_{01} \quad (2.6.16)$$

Jeffreys [56, p. 432]는 Bayes팩터를 사용해서 다음과 같이 결론을 내릴 것을 제안하였다.

(Grade 0)  $K > 10^0$ . Null hypothesis supported.

(Grade 1)  $10^0 > K > 10^{-1/2}$ . Evidence against  $H_0$ , but not worth more than a bare mention.

(Grade 2)  $10^{-1/2} > K > 10^{-1}$ . Evidence against  $H_0$  substantial.

(Grade 3)  $10^{-1} > K > 10^{-3/2}$ . Evidence against  $H_0$  strong.

(Grade 4)  $10^{-3/2} > K > 10^{-2}$ . Evidence against  $H_0$  very strong.

(Grade 5)  $10^{-2} > K$ . Evidence against  $H_0$  decisive.

**예제 2.6.9** 어떤 확률분포에서 관찰값들  $y_1 = 0.3, y_2 = 0.6, y_3 = 0.7, y_4 = 0.8, y_5 = 0.9$  이다. 귀무가설  $H_0$  하에서 확률밀도함수는 다음과 같다.

$$f(y_i | \boldsymbol{\theta}_0, H_0) = \begin{cases} 2\theta_{01}, & (0 \leq y_i < 0.5) \\ 2\theta_{02}, & (0.5 \leq y_i < 1) \\ 0, & (\text{otherwise}) \end{cases} \quad (1)$$

대립가설  $H_1$  하에서 확률밀도함수는 다음과 같다.

$$f(y_i | \boldsymbol{\theta}_1, H_1) = \begin{cases} 3\theta_{11}, & (0 \leq y_i < 1/3) \\ 3\theta_{12}, & (1/3 \leq y_i < 2/3) \\ 3\theta_{13}, & (2/3 \leq y_i < 1) \\ 0, & (\text{otherwise}) \end{cases} \quad (2)$$

여기서 모수벡터들  $\boldsymbol{\theta}_0 = [\theta_{01}, \theta_{02}]^T$  와  $\boldsymbol{\theta}_1 = [\theta_{11}, \theta_{12}, \theta_{13}]^T$  는 다음 식들을 만족한다.

$$\theta_{01} + \theta_{02} = 1, \quad \theta_{11} + \theta_{12} + \theta_{13} = 1 \quad (3)$$

또한, 각  $\theta_{ij}$  는 양수이다.

귀무가설  $H_0$  하에서 모수벡터의 사전확률밀도함수가 다음과 같다고 하자.

$$\pi(\boldsymbol{\theta}_0 | H_0) = \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)} \theta_{01}^{1-1} \theta_{02}^{1-1}. \quad (4)$$

대립가설  $H_1$  하에서 모수벡터의 사전확률밀도함수가 다음과 같다고 하자.

$$\pi(\boldsymbol{\theta}_1 | H_1) = \frac{\Gamma(1+1+1)}{\Gamma(1)\Gamma(1)\Gamma(1)} \theta_{11}^{1-1} \theta_{12}^{1-1} \theta_{13}^{1-1}. \quad (5)$$

귀무가설  $H_0$  하에서 모수벡터의 사후확률밀도함수는 다음과 같다.

$$\pi(\boldsymbol{\theta}_0 | \mathbf{y}, H_0) = \frac{\Gamma(2+5)}{\Gamma(2)\Gamma(5)} \theta_{01}^{2-1} \theta_{02}^{5-1}. \quad (6)$$

대립가설  $H_1$  하에서 모수벡터의 사후확률밀도함수는 다음과 같다.

$$\pi(\boldsymbol{\theta}_1 | \mathbf{y}, H_1) = \frac{\Gamma(2+2+4)}{\Gamma(2)\Gamma(2)\Gamma(4)} \theta_{11}^{2-1} \theta_{12}^{2-1} \theta_{13}^{4-1}. \quad (7)$$

따라서 해당 주변우도함수는 각각 다음과 같다.

$$m(\mathbf{y} | H_0) = 2^5 = \frac{\Gamma(2)\Gamma(5)}{\Gamma(7)} \frac{16}{15} \quad (8)$$

$$m(\mathbf{y} | H_1) = 2 \cdot 3^5 = \frac{\Gamma(2)\Gamma(2)\Gamma(4)}{\Gamma(8)} \frac{81}{140} \quad (9)$$

식 (2.6.15)에서 알 수 있듯이, Bayes팩터는 다음과 같다.

$$K_{01} = \frac{m(\mathbf{x} | H_0)}{m(\mathbf{x} | H_1)} = \frac{16/15}{81/140} = \frac{448}{243} = 1.8436 > 1 \quad (10)$$

따라서 Jeffreys의 기준에 따라 귀무가설을 채택한다. ■

### 2.6.5 예측

모수벡터가  $\boldsymbol{\theta}$ 인 확률밀도함수  $f(y|\boldsymbol{\theta})$ 를 따르는 확률변수  $y$ 를 살펴보자. 확률벡터  $\mathbf{x}$ 가 관찰되었을 때 확률변수  $y$ 의 사후예측확률밀도함수 (posterior predictive probability density function)는 다음과 같다.

$$f(y|\mathbf{x}) = \int_{\Theta} f(y|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta} \quad (2.6.17)$$

**예제 2.6.10** 모수  $\lambda$ 의 사전확률밀도함수가 다음과 같다고 가정하자.

$$\pi(\lambda) = 1_{[0.5, 1.5]}(\lambda). \quad (1)$$

관찰값들은 다음과 같은 Weibull 확률밀도함수에서 추출된 것이라고 하자.

$$f(x|\lambda) = 3.5[x\lambda]^{3.5-1} \exp(-[x\lambda]^{3.5}) 1_{[0,\infty)}(x) \quad (2)$$

관찰벡터  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$  의 결합확률밀도함수는 다음과 같다.

$$f(\mathbf{x}|\lambda) = \lambda^{3.5n} \prod_{j=1}^n x_j^{3.5-1} \exp\left(-\sum_{j=1}^n [x_j\lambda]^{3.5}\right) \quad (3)$$

관찰점들이 9이고 식  $\sum_{j=1}^9 x_j^{3.5} = 10.16$  를 만족한다고 하자. 사후확률밀도함수는 다음과 같다.

$$\pi(\lambda|x_1, x_2, \dots, x_9) = 1621.39\lambda^9 \exp(-10.16\lambda^{3.5}) 1_{[0.5,1.5]}(\lambda) \quad (4)$$

따라서 사후예측확률밀도함수는 다음과 같다.

$$f(y|\mathbf{x}) = \int_{0.5}^{1.5} 3.5\lambda^{2.5} y^{2.5} \exp(-[y\lambda]^{3.5}) \cdot 1621.39\lambda^9 \exp(-10.16\lambda^{3.5}) d\lambda \quad (5)$$

식 (5) 의 정적분은 그리 간단해 보이지 않는다. 아마도 각  $y$  에 대해서 수치적분을 해야할 것 같다. ■

## 제 3 장

# 베이지안통계모형들

### 제3.1절 선형모형의 베이지안통계분석

다음과 같은 선형회귀모형을 살펴보자.

$$y_j = \beta_1 x_{j1} + \beta_2 x_{j2} + \cdots + \beta_K x_{jK} + \epsilon_j, \quad (j = 1, 2, \cdots, n) \quad (3.1.1)$$

여기서  $y_j$ 는 종속변수  $y$ 의 제  $j$ 번째 관측값,  $x_{j1}, x_{j2}, \cdots, x_{jK}$ 는 설명변수들  $x_1, x_2, \cdots, x_K$ 의 제  $j$ 번째 관측값들 그리고  $\beta_1, \beta_2, \cdots, \beta_K$ 는 회귀계수들이다. 식  $n > K$ 가 성립하고, 다음 식이 성립한다고 가정하자.

$$\epsilon_j | \sigma^2 \stackrel{d}{\sim} \mathcal{N}(0, \sigma^2) \quad (3.1.2)$$

또한,  $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ 은 분산  $\sigma^2$ 가 주어진 조건 하에서 서로 독립인 오차항들이라고 가정하자. 다음 식들이 성립한다.

$$y_j | (\beta_1, \cdots, \beta_K, \sigma^2) \stackrel{d}{\sim} \mathcal{N}\left(\sum_{k=1}^K \beta_k x_{jk}, \sigma^2\right), \quad (j = 1, 2, \cdots, n) \quad (3.1.3)$$

식 (3.1.3)에서 알 수 있듯이, 모수들이 주어진 조건 하에서 종속변수열  $\{y_j | j = 1, 2, \cdots, n\}$ 는 서로 독립이다. 식 (3.1.3)을 다음과 같이 쓸 수 있다.

$$\mathbf{y} | (\boldsymbol{\beta}, \sigma^2) \stackrel{d}{\sim} \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \quad (3.1.4)$$

여기서  $I_n$ 은  $n$ 차원 단위행렬이고,  $\mathbf{y}$ ,  $X$ , 그리고  $\boldsymbol{\beta}$ 는 각각 다음과 같다.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nK} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} \quad (3.1.5)$$

**예제 3.1.1** 선형회귀모형을 그림으로 이해하기 위해서, 다음 MATLAB 프로그램 LinearRegressionModel101.m을 실행해 보자.

```

1 % -----
2 % Filename: LinearRegressionModel101.m
3 % -----
4 clear all, clf, close all
5 mu = @(x) -2.0 + 0.25*x;
6 x = 5:.1:13;
7 yhat = mu(x);
8 dy = -3.5:.1:3.5; sz = size(dy); k = (length(dy)+1)/2;
9 x1 = 7*ones(sz); y1 = mu(x1)+dy; z1 = normpdf(y1,mu(x1),1);
10 x2 = 8*ones(sz); y2 = mu(x2)+dy; z2 = normpdf(y2,mu(x2),1);
11 x3 = 9*ones(sz); y3 = mu(x3)+dy; z3 = normpdf(y3,mu(x3),1);
12 x4 = 10*ones(sz); y4 = mu(x4)+dy; z4 = normpdf(y4,mu(x4),1);
13 x5 = 11*ones(sz); y5 = mu(x5)+dy; z5 = normpdf(y5,mu(x5),1);
14 plot3(x,yhat,zeros(size(x)),'b-', ...
15       x1,y1,z1,'r-', x1([k k]),y1([k k]),[0 z1(k)],'k:', ...
16       x2,y2,z2,'r-', x2([k k]),y2([k k]),[0 z2(k)],'k:', ...
17       x3,y3,z3,'r-', x3([k k]),y3([k k]),[0 z3(k)],'k:', ...
18       x4,y4,z4,'r-', x4([k k]),y4([k k]),[0 z4(k)],'k:', ...
19       x5,y5,z5,'r-', x5([k k]),y5([k k]),[0 z5(k)],'k:', 'LineWidth',2);
20 set(gca, 'fontsize',11, 'fontweigh', 'bold')
21 xlim([5 13]); zlim([0 1]);
22 xlabel('X'); ylabel('Y'); zlabel('Probability density');
23 grid on; view([-45 45]);
24 saveas(gcf, 'LinearRegressionModel101', 'jpg')
25 % End of program
26 % -----

```

이 MATLAB 프로그램 LinearRegressionModel101.m을 실행한 결과가 그림 3.1.1에 그려져 있다. 이 그림은 다음 선형회귀모형을 묘사한 것이다.

$$y = -2.0 + 0.25x + \epsilon, \quad \epsilon \stackrel{d}{\sim} \mathcal{N}(0,1) \quad (1)$$

■

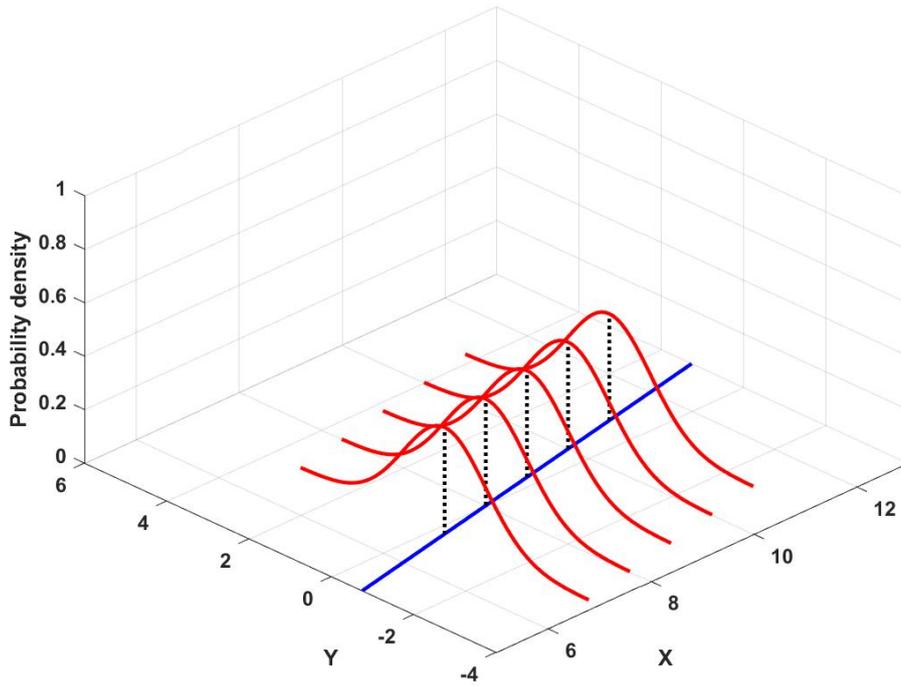


그림 3.1.1. 선형회귀모형

모수벡터  $(\beta, \sigma^2)$  의 사전확률분포가 다음과 같다고 가정하자.

$$\beta | \sigma^2 \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_0, \sigma^2 B_0) \tag{3.1.6}$$

$$\sigma^2 \stackrel{d}{\sim} \text{IG}\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \tag{3.1.7}$$

모수  $\sigma^2$  의 역수인 정밀도 (precision) 를  $\psi$  로 표기하자. 즉, 다음 식이 성립한다.

$$\psi = \frac{1}{\sigma^2} \tag{3.1.8}$$

식 (3.1.6)~식 (3.1.8) 에서 알 수 있듯이, 모수벡터  $(\beta, \psi)$  의 사전확률분포는 다음과 같다.

$$\beta | \psi \stackrel{d}{\sim} \mathcal{N}\left(\mathbf{b}_0, \frac{1}{\psi} B_0\right) \tag{3.1.9}$$

$$\psi \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \tag{3.1.10}$$

식 (3.1.9) 와 식 (3.1.10) 에서 정의한 모수벡터  $(\beta, \psi)$  의 사전확률분포는 정규감마확률분포

(normal-gamma distribution)이다. 이에 해당하는 사전확률밀도함수는 다음과 같다.

$$\begin{aligned}\pi(\boldsymbol{\beta}, \psi) &= \pi(\boldsymbol{\beta} \mid \psi)\pi(\psi) \\ &\propto \psi^{[n_0+K]/2-1} \exp\left(-\frac{1}{2}\psi \left\{n_0 S_0 + [\boldsymbol{\beta} - \mathbf{b}_0]^T B_0^{-1} [\boldsymbol{\beta} - \mathbf{b}_0]\right\}\right)\end{aligned}\quad (3.1.11)$$

정밀도  $\psi$ 를 사용해서 식 (3.1.4)를 표현하면, 다음과 같다.

$$\mathbf{y} \mid (\boldsymbol{\beta}, \psi) \stackrel{d}{\sim} \mathcal{N}\left(X\boldsymbol{\beta}, \frac{1}{\psi}I_n\right)\quad (3.1.12)$$

다음 확률벡터와 확률변수들을 정의하자.

$$\hat{\boldsymbol{\beta}} \doteq [X^T X]^{-1} X^T \mathbf{y}\quad (3.1.13)$$

$$S_e \doteq [\mathbf{y} - X\hat{\boldsymbol{\beta}}]^T [\mathbf{y} - X\hat{\boldsymbol{\beta}}]\quad (3.1.14)$$

$$Q(\boldsymbol{\beta}) \doteq [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]^T X^T X [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]\quad (3.1.15)$$

식 (3.1.12)~식 (3.1.15)에서 알 수 있듯이, 확률벡터  $\mathbf{y}$ 의 우도함수는 다음과 같다.

$$\begin{aligned}f(\mathbf{y} \mid \boldsymbol{\beta}, \psi) &\propto \psi^{n/2} \exp\left(-\frac{\psi}{2}[\mathbf{y} - X\boldsymbol{\beta}]^T [\mathbf{y} - X\boldsymbol{\beta}]\right) \\ &= \psi^{n/2} \exp\left(-\frac{\psi}{2}[Q(\boldsymbol{\beta}) + S_e]\right)\end{aligned}\quad (3.1.16)$$

다음 변수들을 정의하자.

$$n_* \doteq n_0 + n\quad (3.1.17)$$

$$B_* \doteq [B_0^{-1} + X^T X]^{-1}\quad (3.1.18)$$

$$\mathbf{b}_* \doteq B_* [B_0^{-1}\mathbf{b}_0 + X^T \mathbf{y}]\quad (3.1.19)$$

$$S_* \doteq \frac{1}{n_*} \left\{n_0 S_0 + [\mathbf{y} - X\mathbf{b}_*]^T \mathbf{y} + [\mathbf{b}_0 - \mathbf{b}_*]^T B_0^{-1} \mathbf{b}_0\right\}\quad (3.1.20)$$

다음 식이 성립한다.

$$\begin{aligned}n_0 S_0 + [\boldsymbol{\beta} - \mathbf{b}_0]^T B_0^{-1} [\boldsymbol{\beta} - \mathbf{b}_0] + Q(\boldsymbol{\beta}) + S_e \\ = n_* S_* + [\boldsymbol{\beta} - \mathbf{b}_*]^T B_*^{-1} [\boldsymbol{\beta} - \mathbf{b}_*]\end{aligned}\quad (3.1.21)$$

식 (3.1.11), 식 (3.1.16)과 식 (3.1.21)에서 알 수 있듯이, 모수벡터  $(\boldsymbol{\beta}, \psi)$ 의 사후확률밀도함수는 다음과 같다.

$$\begin{aligned} \pi(\boldsymbol{\beta}, \psi | \mathbf{y}) &\propto f(\mathbf{y} | \boldsymbol{\beta}, \psi) \pi(\boldsymbol{\beta}, \psi) \\ &\propto \psi^{K/2} \exp\left(-\frac{\psi}{2} [\boldsymbol{\beta} - \mathbf{b}_*]^T B_*^{-1} [\boldsymbol{\beta} - \mathbf{b}_*]\right) \psi^{n_*/2-1} \exp\left(-\frac{\psi}{2} n_* S_*\right) \end{aligned} \quad (3.1.22)$$

식 (3.1.22)에서 알 수 있듯이, 모수벡터  $(\boldsymbol{\beta}, \psi)$ 의 사후확률분포는 다음과 같다.

$$\boldsymbol{\beta} | (\psi, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}\left(\mathbf{b}_*, \frac{1}{\psi} B_*\right) \quad (3.1.23)$$

$$\psi | \mathbf{y} \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_*}{2}, \frac{n_* S_*}{2}\right) \quad (3.1.24)$$

식 (3.1.23)와 식 (3.1.24)에서 알 수 있듯이, 모수벡터  $(\boldsymbol{\beta}, \psi)$ 의 사후확률분포도 정규감마확률분포이다. 사전확률분포와 사후확률분포가 모두 정규감마확률분포이므로, 이들은 공액이다. 식 (3.1.23)과 식 (3.1.24)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\boldsymbol{\beta} | (\sigma^2, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_*, \sigma^2 B_*) \quad (3.1.25)$$

$$\sigma^2 | \mathbf{y} \stackrel{d}{\sim} \text{IG}\left(\frac{n_*}{2}, \frac{n_* S_*}{2}\right) \quad (3.1.26)$$

**예제 3.1.2** 식 (3.1.9)와 식 (3.1.10)에 기술한 사전확률분포와 식 (3.1.25)와 식 (3.1.26)의 사후확률분포를 그리기 위해서, 다음 R프로그램 Regression101.R을 실행해 보자.

```

1 # -----
2 # Filename: Regression101.R
3 # Linear Regression
4 # Programmed by CBS
5 # -----
6 setEPS()
7 # Making Data and Classical OLS
8 set.seed(3)
9 library(MASS)
10 x1 <- c(-10:10)
11 ( x2 <- sin(x1) )
12 ( sigmay <- 0.5 )
13 y <- pi*x1 + 0.123*x2 + sigmay*rnorm(21,0,1)
14 ny <- length(y)
15 lm(formula = y ~ x1 + x2 - 1) # -1 means no intercept
16 X <- cbind(x1, x2)
17 XtX <- t(X)%*%X
18 Xty <- t(X)%*%y
19
20 # Prior

```

```

21 ( b0 <- matrix(c(3.0,0.1),2,1 ) );
22 ( B0 <- diag(2) ); ( n0 <- 3 ); ( S0 <- 1.1 )
23 ( shape0 <- n0/2 );
24 ( rate0 <- n0*S0/2 );
25 Nbeta <- 20; Npsi <- 20
26 beta1.gd <- seq(-4,4,length=Nbeta)
27 beta2.gd <- seq(-4,4,length=Nbeta)
28 psi.gd <- seq(0.01,2.00,length=Npsi)
29
30 priorPDF <- array(0, dim=c(Nbeta,Nbeta,Npsi))
31 for (kk in 1:Npsi){
32   prior.psi <- 1/gamma(shape0)*(rate0)^(shape0)*
33     psi.gd[kk]^(-shape0-1)*
34     exp(-rate0/psi.gd[kk])
35   for (ii in 1:Nbeta){
36     for (jj in 1:Nbeta){
37       dumx = c( beta1.gd[ii]-b0[1],beta2.gd[jj]-b0[2])
38       prior.beta <- psi.gd[kk]/(2*pi*sqrt(det(B0)))*
39         exp(-1/2*psi.gd[kk]*t(dumx)%%ginv(B0)%%dumx)
40       priorPDF[ii,jj,kk] <- prior.psi*prior.beta
41     }
42   }
43 }
44
45 # Monte Carlo Simulations of Prior
46 plot.new()
47 postscript('Regression101.eps') # Start to save figure
48 par(mfrow=c(2,3))
49 NoSim <- 1024
50 psi0.Sim <- 1/rgamma(NoSim,shape=shape0,rate=rate0)
51 beta0.Sim <- matrix(0,NoSim,2)
52 for (kk in 1:NoSim){
53   beta0.Sim[kk,] <- mvrnorm(1,b0,B0/psi0.Sim[kk])
54 }
55 plot(beta0.Sim[,1],beta0.Sim[,2],xlab="beta_1",ylab="beta_2",col="red",
56       main="Prior 1")
57 plot(beta0.Sim[,1],psi0.Sim,xlab="beta_1",ylab="psi",col="green",
58       main="Prior 2")
59 plot(beta0.Sim[,2],psi0.Sim,xlab="beta_2",ylab="psi",col="blue",
60       main="Prior 3")
61
62 # Posterior
63 ( nStar <- n0 + ny )
64 ( BStar <- ginv(ginv(B0)+XtX) )
65 ( bStar <- BStar%*(ginv(B0)%*b0+Xty) )
66 ( SStar <- (n0*S0 + t(y-X%*bStar)%*y +
67   t(b0-bStar)%*ginv(B0)%*b0 )/nStar )
68 ( shapeStar <- nStar/2 )
69 ( rateStar <- nStar*SStar/2 )
70 Nbeta <- 20; Npsi <- 20
71 beta1.gd <- seq(-4,4,length=Nbeta)
72 beta2.gd <- seq(-4,4,length=Nbeta)
73 psi.gd <- seq(0.01,2.00,length=Npsi)
74 posteriorPDF <- array(0,dim=c(Nbeta,Nbeta,Npsi))
75 for (kk in 1:Npsi){
76   posterior.psi <- 1/gamma(shapeStar)*(rateStar)^(shapeStar)*
77     psi.gd[kk]^(-shapeStar-1)*
78     exp(-rateStar/psi.gd[kk])
79   for (ii in 1:Nbeta){
80     for (jj in 1:Nbeta){
81       dumx = c( beta1.gd[ii]-bStar[1],beta2.gd[jj]-bStar[2])

```

```

82     posterior.beta <- psi.gd[kk]/(2*pi*sqrt(det(BStar)))*
83         exp(-1/2*psi.gd[kk]*t(dumx)%*%ginv(BStar)%*%dumx)
84     posteriorPDF[ii,jj,kk] <- posterior.psi*posterior.beta
85   }
86 }
87 }
88
89 # Monte Carlo Simulations of Posterior
90 # postscript('Normal_MeanVariance1Star1.eps') # Start to save figure
91 NoSim <- 1024
92 psiStar.Sim <- 1/rgamma(NoSim,shape=shapeStar,rate=rateStar)
93 betaStar.Sim <- matrix(0,NoSim,2)
94 for (kk in 1:NoSim){
95     betaStar.Sim[kk,] <- mvrnorm(1,bStar,BStar/psiStar.Sim[kk])
96 }
97 plot(betaStar.Sim[,1],betaStar.Sim[,2],xlab="beta_1",
98     ylab="beta_2",col="red",main="Posterior 1")
99 plot(betaStar.Sim[,1],psiStar.Sim,xlab="beta_1",a
100     ylab="psi",col="green",main="Posterior 2")
101 plot(betaStar.Sim[,2],psiStar.Sim,xlab="beta_2",
102     ylab="psi",col="blue",main="Posterior 3")
103 dev.off() # End to save figure
104 # -----

```

이 R프로그램 Regression101.R에서 사전확률분포는 다음과 같은 정규감마확률분포이다.

$$\psi \stackrel{d}{\sim} \text{Gamma} \left( \frac{3}{2}, \frac{3 \times 1.1}{2} \right) \quad (1)$$

$$\boldsymbol{\beta} | \psi \stackrel{d}{\sim} \mathcal{N} \left( \begin{bmatrix} 3.0 \\ 0.1 \end{bmatrix}, \frac{1}{\psi} I_2 \right) \quad (2)$$

즉, 사전확률분포의 모수들은 다음과 같다.

$$n_0 = 3, \quad S_0 = 1.1, \quad \mathbf{b}_0 = \begin{bmatrix} 3.0 \\ 0.1 \end{bmatrix}, \quad B_0 = I_2 \quad (3)$$

다음 식들을 사용해서 관찰점들을 생성하고, 이들을 사용해서 우도함수를 계산한다.

$$x_{j,1} = j - 11, \quad x_{j,2} = \sin x_{j,1}, \quad (j = 1, 2, \dots, 21) \quad (4)$$

$$y_j = \pi x_{j,1} + 0.123 x_{j,2} + \epsilon_j, \quad (j = 1, 2, \dots, 21) \quad (5)$$

$$\epsilon_1, \epsilon_2, \dots, \epsilon_{21} \stackrel{d}{\sim} \mathcal{N}(0, 0.5^2) \quad (6)$$

사후확률분포의 모수들은 다음과 같다.

$$n_* = 24, \quad S_* = 0.2681365, \tag{7}$$

$$\mathbf{b}_* = \begin{bmatrix} 3.1460539 \\ 0.1260803 \end{bmatrix}, \quad B_* = \begin{bmatrix} 0.001308790 & -0.001039082 \\ -0.001039082 & 0.091710405 \end{bmatrix} \tag{8}$$

또한, 사후확률분포는 다음과 같은 정규감마확률분포이다.

$$\boldsymbol{\beta} \mid (\sigma^2, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_*, \sigma^2 B_*) \tag{9}$$

$$\sigma^2 \mid \mathbf{y} \stackrel{d}{\sim} \mathcal{IG}\left(\frac{n_*}{2}, \frac{n_* S_*}{2}\right) \tag{10}$$

이 R 프로그램 Regression101.R을 실행한 결과가 그림 3.1.2에 그려져 있다. ■

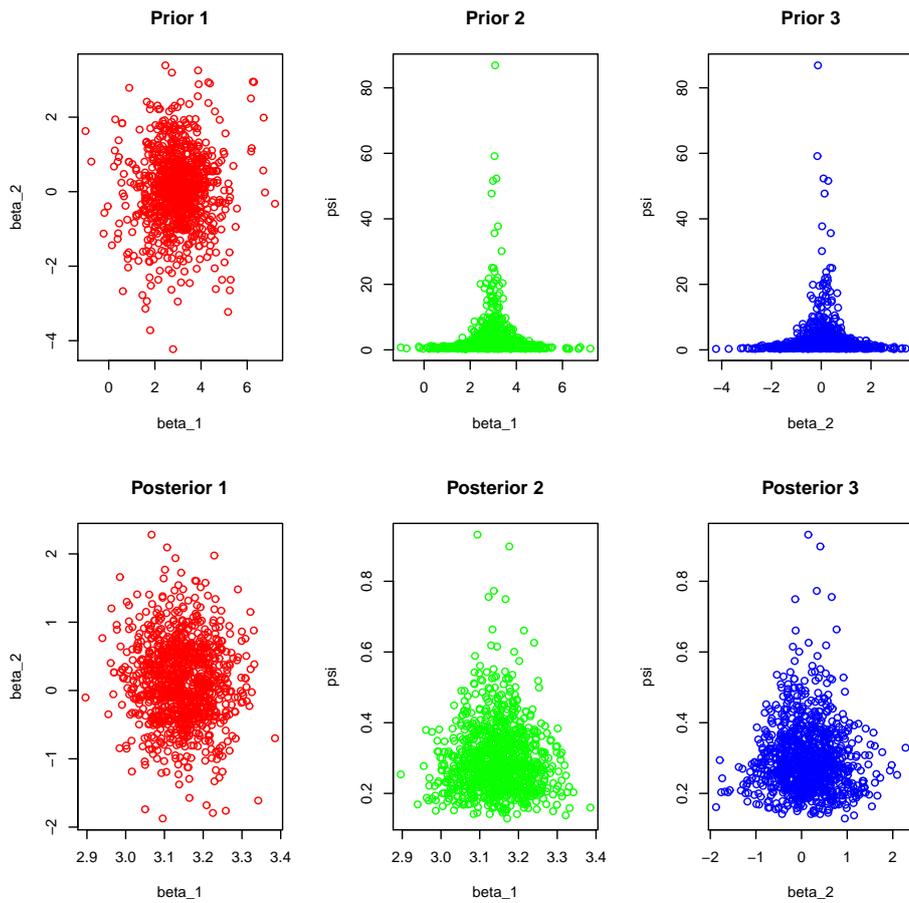


그림 3.1.2. 선형회귀모형의 베이زي안분석

식 (3.1.1)의 회귀모형을 지수형 확률분포족에 속하는 확률분포의 모형으로 확장하기로 하자. 확률변수  $y_j$ 의 확률밀도함수가 다음 식을 만족한다고 하자.

$$f(y_j | \theta_j) = a(y_j) \exp(y_j \theta_j + b(\theta_j)) \quad (3.1.27)$$

모수들이 주어진 조건 하에서  $\{y_j | j = 1, 2, \dots, n\}$ 은 서로 독립인 확률변수열이라고 하고, 다음 상수들을 정의하자.

$$\mu_j \doteq E(y_j | \theta_j) \quad (3.1.28)$$

$$\eta_j \doteq g(\mu_j) \doteq \sum_{k=1}^K \beta_k x_{jk} \quad (3.1.29)$$

식 (3.1.27) ~ 식 (3.1.29)로 구성된 모형을 일반화선형모형 (generalized linear model)이라 부른다. 여기서  $g$ 는 미분가능한 함수로서 링크함수(link function)라 불리운다. 일반화선형모형의 베이지안추정에서는 모수벡터  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_K]^T$ 에 대한 사전확률분포가 공액이 아닌 경우가 대부분이고, 해석적으로 베이지안추정을 할 수 없는 경우가 많다. 이러한 경우에는 MCMC를 사용해서 수치적으로 베이지안추정을 한다.

**예제 3.1.3** 모수들  $\lambda_1, \lambda_2, \dots, \lambda_n$ 이 주어진 조건 하에서 확률변수들  $y_1, y_2, \dots, y_n$ 이 서로 독립이고 Poisson 확률분포를 따른다고 하자. 즉, 다음 식들이 성립한다고 하자.

$$y_j | \lambda_j \stackrel{d}{\sim} \text{Poisson}(\lambda_j), \quad (j = 1, 2, \dots, n) \quad (1)$$

확률변수  $y_j$ 의 기대값  $\lambda_j$ 는 다음 식을 만족한다.

$$\lambda_j = E(y_j | \lambda_j) \quad (2)$$

기대값  $\lambda_j$ 가  $\eta_j \doteq \sum_{k=1}^K \beta_k x_{jk}$ 를 통해서만 설명변수들  $x_{j1}, x_{j2}, \dots, x_{jK}$ 의 영향을 받는다고 하자. 임의의 실수를 취할 수 있는  $\eta_j$ 와는 달리, 기대값  $\lambda_j$ 는 양수만을 취할 수 있다. 따라서,  $\lambda_j$ 가 다음 식들을 만족한다고 가정하자.

$$\ln \lambda_j = \sum_{k=1}^K \beta_k x_{jk} \quad (3)$$

이러한 모형을 Poisson 회귀모형이라 부른다. 이 Poisson 회귀모형이 일반화선형모형의 일종

임이 명백하고, 링크함수는  $g(\lambda) = \ln \lambda$ 이다.

Poisson 회귀모형을 그림으로 이해하기 위해서, 다음 MATLAB 프로그램 PoissonRegressionModel101.m을 실행해 보자.

```

1 % -----
2 % Filename: PoissonRegressionModel101.m
3 % -----
4 clear all, clf, close all
5 mu = @(x) exp(-2.0 + 0.25*x);
6 x = 5:.1:15;
7 yhat = mu(x);
8 x1 = 7*ones(1,5); y1 = 0:4; z1 = poisspdf(y1,mu(x1));
9 x2 = 10*ones(1,7); y2 = 0:6; z2 = poisspdf(y2,mu(x2));
10 x3 = 13*ones(1,9); y3 = 0:8; z3 = poisspdf(y3,mu(x3));
11 plot3(x,yhat,zeros(size(x)),'b-', ...
12       [x1; x1],[y1; y1],[z1; zeros(size(y1))],'k-', x1,y1,z1,'r.', ...
13       [x2; x2],[y2; y2],[z2; zeros(size(y2))],'k-', x2,y2,z2,'r.', ...
14       [x3; x3],[y3; y3],[z3; zeros(size(y3))],'k-', x3,y3,z3,'r.', ...
15       'LineWidth',2);
16 set(gca,'fontsize',11,'fontweigh','bold')
17 zlim([0 1]);
18 xlabel('X'); ylabel('Y'); zlabel('Probability');
19 grid on; view([-45 45]);
20 saveas(gcf,'PoissonRegressionModel101','jpg')
21 % End of program
22 % -----

```

이 MATLAB 프로그램 PoissonRegressionModel101.m을 실행한 결과가 그림 3.1.3에 그려져 있다. 이 그림은 다음과 같은 Poisson 회귀모형을 묘사한 것이다.

$$y|x \stackrel{d}{\sim} \text{Poisson}(\exp(-2.0 + 0.25x)) \quad (1)$$

■

**예제 3.1.4** Poisson 회귀분석을 하는 예로서 다음 MATLAB 프로그램 PoissonRegression101.m을 실행해 보자.

```

1 % -----
2 % Filename: PoissonRegression101.m
3 % -----
4 clear all, clf, close all
5 rng('default') % for reproducibility
6 X = randn(100,2);
7 mu = exp(-2+0.25*X(:,1)-0.3*X(:,2));
8 y = poissrnd(mu);
9 % Fit a generalized linear model using the Poisson distribution.

```

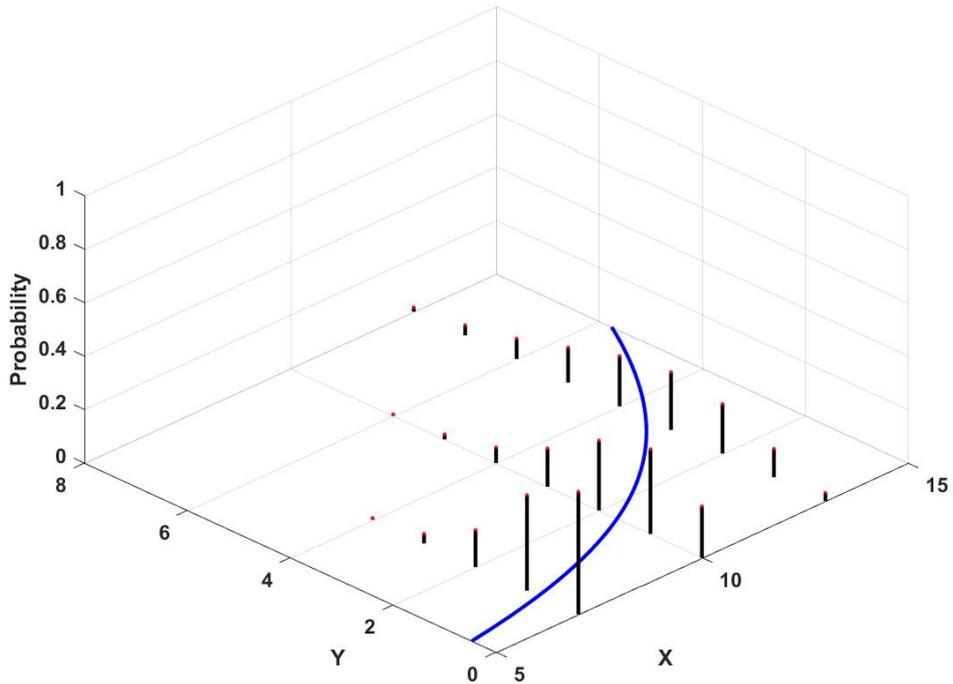


그림 3.1.3. Poisson 회귀모형

```

10 mdl = fitglm(X,y,'linear','Distribution','poisson')
11 gcf = plotSlice(mdl)
12 saveas(gcf,'PoissonRegression101','jpg')
13 % End of program
14 % -----
    
```

이 MATLAB 프로그램 PoissonRegression101.m을 실행한 결과가 그림 3.1.4에 그려져 있다. ■

**예제 3.1.5** 모수들이 주어진 조건 하에서 서로 독립인 확률변수들  $y_1, y_2, \dots, y_n$  이 다음 식들을 만족한다고 하자.

$$\Pr(y_j = 1 | p_j) = p_j = 1 - \Pr(y_j = 0 | p_j), \quad (j = 1, 2, \dots, n) \tag{1}$$

이 확률변수  $y_j$  의 기대값  $\mu_j = E(y_j | p_j)$ 는 다음 식을 만족한다.

$$\mu_j = p_j \tag{2}$$

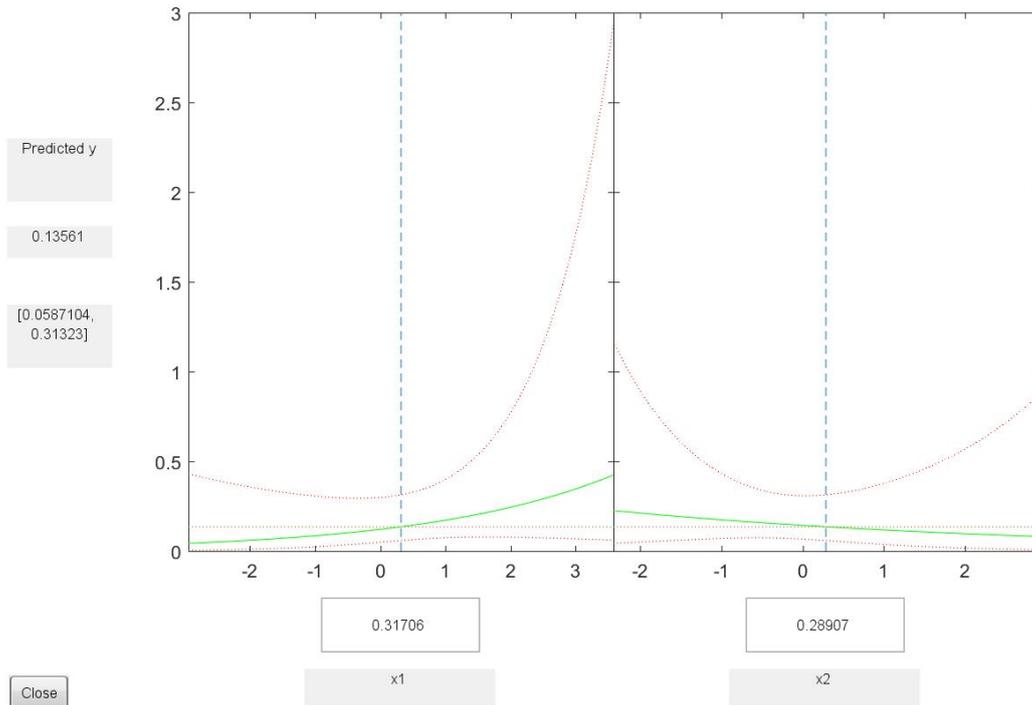


그림 3.1.4. Poisson 회귀분석

확률  $p_j$ 가  $\eta_j \doteq \sum_{k=1}^K \beta_k x_{jk}$ 를 통해서만  $x_{j1}, x_{j2}, \dots, x_{jK}$ 의 영향을 받는다고 하자. 임의의 실수를 취할 수 있는  $\eta_j$ 와는 달리, 확률  $p_j$ 는 구간  $[0, 1]$ 에 속하는 값만을 취할 수 있다. 따라서, 적당한 누적확률분포함수  $F$ 에 대해서  $p_j$ 가 다음 식들을 만족한다고 가정하자.

$$p_j = \mu_j, \quad F^{-1}(p_j) = \eta_j = \sum_{k=1}^K \beta_k x_{jk} \quad (3)$$

예를 들어, 누적확률분포함수  $F$ 로서 표준정규확률분포함수  $\Phi$ 를 사용하면, 식 (3)을 다음과 같이 쓸 수 있다.

$$p_j = \mu_j, \quad \Phi^{-1}(p_j) = \eta_j = \sum_{k=1}^K \beta_k x_{jk} \quad (4)$$

여기서  $g = \Phi^{-1}$ 는 링크함수(link function)이다. 이러한 일반화선형모형을 프로빗모형(probit model)이라 부른다.

다른 예로서, 누적확률분포함수  $F$ 로서 로지스틱확률분포함수를 사용하면, 링크함수는

다음과 같다.

$$g(p_j) = \ln \frac{p_j}{1 - p_j} \quad (5)$$

이러한 일반화선형모형을 로짓모형 (logit model) 이라 부른다.

누적확률분포함수  $F$ 로서 극값확률분포함수 (extreme distribution function) 를 사용하면, 링크함수는 다음과 같다.

$$g(p_j) = \ln(-\ln(1 - p_j)) \quad (6)$$

프로빗모형, 로짓모형, 극값확률분포함수를 링크함수로 이용하는 모형을 사용하는 예로서, 다음 MATLAB 프로그램 GLMfit102.m을 실행해 보자.

```

1 % -----
2 % Filename: GLMfit102.m
3 % -----
4 clear all, clf, close all
5 % Logit Model
6 x = [2100 2300 2500 2700 2900 3100 ...
7      3300 3500 3700 3900 4100 4300]';
8 n = [48 42 31 34 31 21 23 23 21 16 17 21]';
9 y = [1 2 0 3 8 8 14 17 19 15 17 21]';
10 bLogit = glmfit(x,[y n],'binomial','link','logit'); % Default
11 yfitLogit = glmval(bLogit,x,'logit','size',n);
12 % Probit
13 bProbit = glmfit(x,[y n],'binomial','link','probit');
14 yfitProbit = glmval(bProbit,x,'probit','size',n);
15 % Extreme Distribution Function
16 bExtreme = glmfit(x,[y n],'binomial','link','comploglog');
17 yfitExtreme = glmval(bExtreme,x,'comploglog','size',n);
18 % % Plotting
19 plot(x, y./n,'ko',x,yfitLogit./n,'r--',x,yfitProbit./n,'g-', ...
20      x,yfitExtreme./n,'b:','LineWidth',2)
21 legend('True Ratio','Logit','Probit','Extreme','location','NW')
22 set(gca,'fontsize',11,'fontweight','bold')
23 xlabel('\bf Regressor','FontSize',12);
24 ylabel('\bf Success Rate','FontSize',12);
25 axis([2000 4500 -0.05 1.05])
26 saveas(gcf,'GLMfit102','jpg')
27 % End of program
28 % -----

```

이 MATLAB 프로그램 GLMfit102.m을 실행한 결과가 그림 3.1.5에 그려져 있다. ■

### 예제 3.1.6

스위스도로관리청 (Swedish Road Administration) 의 웹사이트

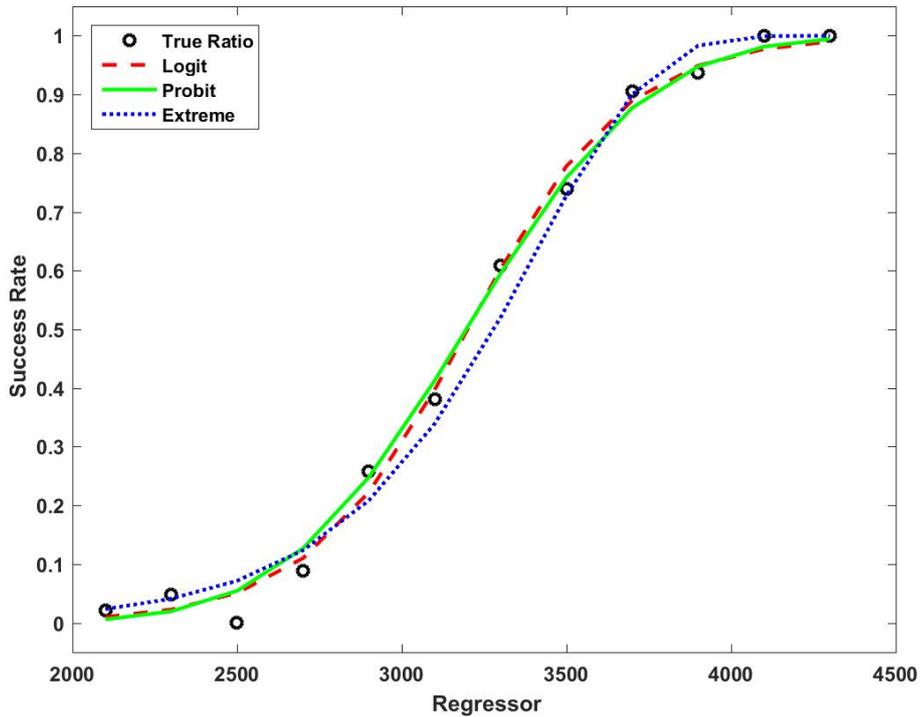


그림 3.1.5. 로짓모형, 프로빗모형, 극값링크함수모형

<http://www.trafikverket.se>에서 교통사고에 관한 다양한 통계를 얻을 수 있다. 이 웹사이트에서 1950년부터 2010년까지 교통사고 사망자수에 관한 데이터세트 1950-RoadData를 다운받을 수 있다.

제  $j$  번째 연도에 발생한 사망자의 수를 확률변수  $y_j$ 로 표기하고, 모수  $\mu_j$ 가 주어졌을 때 확률변수  $y_j | \mu_j$ 가 다음과 같이 Poisson 확률분포를 따른다고 하자.

$$y_j | \mu_j \stackrel{d}{\sim} \text{Poisson}(\mu_j), \quad (j = 1, 2, \dots, n) \tag{1}$$

또한,  $y_1 | \mu_1, y_2 | \mu_2, \dots, y_n | \mu_n$ 은 서로 독립이라고 가정하자. 확률변수  $y_j | \mu_j$ 의 기대값은  $E(y_j | \mu_j) = \mu_j$ 이다. 지금부터는 Poisson 회귀모형을 생각해보자. 즉, 기대값  $\mu_j$ 가 다음 식들을 만족한다고 가정하자.

$$\mu_j = E(y_j | \mu_j) = \exp \left( \sum_{k=1}^K \beta_k x_{jk} \right) \tag{2}$$

즉, 링크함수는  $g(x) = \ln x$ 이다.

다음 식들이 성립한다.

$$\Pr(y_j = l) = \frac{\exp\left(-\exp\left(\sum_{k=1}^K \beta_k x_{jk}\right)\right) \left[\exp\left(\sum_{k=1}^K \beta_k x_{jk}\right)\right]^l}{l!}, \quad (l = 0, 1, 2, \dots) \quad (3)$$

따라서 우도함수는 다음과 같다.

$$L(\boldsymbol{\beta}) = \prod_{j=1}^n \Pr(y_j = l_j) = \prod_{j=1}^n e^{-\mu_j} \frac{\mu_j^{l_j}}{l_j!} \quad (4)$$

여기서  $\mu_j = \mu_j(\boldsymbol{\beta})$ 는  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_K]^T$ 의 함수이다. 대수우도함수는 다음과 같다.

$$\ln L(\boldsymbol{\beta}) = -\sum_{j=1}^n \mu_j + \sum_{j=1}^n l_j \ln \mu_j - \sum_{j=1}^n \ln(l_j!) \quad (5)$$

이 대수우도함수를 최대화하기 위한 1차조건은 다음과 같다.

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{j=1}^n \left[ \frac{l_j}{\mu_j} - 1 \right] \frac{\partial \mu_j}{\partial \beta_k} = \sum_{j=1}^n [l_j - \mu_j] x_{jk} = 0, \quad (k = 1, 2, \dots, K) \quad (6)$$

이 1차조건은 비선형연립방정식을 이루므로 Newton-Raphson 법과 같은 수치적 기법을 적용해서 연립방정식 (6)을 풀어, 최우추정벡터  $\hat{\boldsymbol{\beta}}$ 을 구한다.

이 데이터세트 1950-RoadData를 분석하기 위해서, 다음 MATLAB 프로그램 RoadAccident1950\_2010.m을 실행하자.

```

1 % -----
2 % Filename: RoadAccident1950_2010.m
3 % http://www.math.chalmers.se/Stat/Grundutb/CTH
4 % /mve300/1112/files/lab4/lab4.pdf
5 % -----
6 clear all, clf, close all, clc
7
8 % (Step 0) Make a dataset
9 data = 10000* ...
10 [ 0.1950  0.0595  1.0583      NaN  0.0345      NaN  0.7042  0.0172  0.0008
11  0.1951  0.0708  1.1240      NaN  0.0410      NaN  0.7099  0.0173  0.0010
12  0.1952  0.0750  1.2846      NaN  0.0466      NaN  0.7151  0.0161  0.0010
13  0.1953  0.0921  1.5033      NaN  0.0542      NaN  0.7192  0.0170  0.0013
14  0.1954  0.0942  1.6963      NaN  0.0652      NaN  0.7235  0.0144  0.0013
15  0.1955  0.0902  1.8050      NaN  0.0755  0.1531  0.7290  0.0119  0.0012
16  0.1956  0.0889  1.9220  0.3275  0.0855  0.1657  0.7341  0.0104  0.0012
17  0.1957  0.0946  1.9934  0.3277  0.0987  0.1738  0.7393  0.0096  0.0013
18  0.1958  0.0941  2.0859  0.3104  0.1098  0.1893  0.7436  0.0086  0.0013
19  0.1959  0.1000  2.1475  0.3137  0.1216  0.2011  0.7471  0.0082  0.0013
20  0.1960  0.1036  2.1536  0.2983  0.1324  0.2127  0.7498  0.0078  0.0014
21  0.1961  0.1083  2.2989  0.3031  0.1439  0.2318  0.7542  0.0075  0.0014

```

```

22 | 0.1962 0.1123 2.2438 0.2942 0.1562 0.2394 0.7581 0.0072 0.0015
23 | 0.1963 0.1217 2.3400 0.3068 0.1697 0.2609 0.7628 0.0072 0.0016
24 | 0.1964 0.1308 2.4935 0.3370 0.1810 0.2857 0.7695 0.0072 0.0017
25 | 0.1965 0.1313 2.3618 0.3158 0.1934 0.3043 0.7773 0.0068 0.0017
26 | 0.1966 0.1313 2.1430 0.4700 0.2033 0.3165 0.7843 0.0065 0.0017
27 | 0.1967 0.1077 2.1001 0.5304 0.2126 0.3292 0.7893 0.0051 0.0014
28 | 0.1968 0.1262 2.3028 0.6111 0.2223 0.3451 0.7935 0.0057 0.0016
29 | 0.1969 0.1275 2.3199 0.6529 0.2350 0.3635 0.8004 0.0054 0.0016
30 | 0.1970 0.1307 2.2230 0.6614 0.2446 0.3781 0.8081 0.0053 0.0016
31 | 0.1971 0.1213 2.1872 0.7031 0.2513 0.3870 0.8115 0.0048 0.0015
32 | 0.1972 0.1194 2.1256 0.6657 0.2618 0.4025 0.8129 0.0046 0.0015
33 | 0.1973 0.1177 2.2551 0.7264 0.2667 0.4251 0.8144 0.0044 0.0014
34 | 0.1974 0.1197 2.0902 0.6982 0.2809 0.3921 0.8176 0.0043 0.0015
35 | 0.1975 0.1172 2.0809 0.6728 0.2931 0.4383 0.8208 0.0040 0.0014
36 | 0.1976 0.1168 2.1843 0.6679 0.3060 0.4624 0.8236 0.0038 0.0014
37 | 0.1977 0.1031 2.0916 0.6529 0.3039 0.4811 0.8267 0.0034 0.0012
38 | 0.1978 0.1034 2.0573 0.6431 0.3042 0.4945 0.8284 0.0034 0.0012
39 | 0.1979 0.0928 1.9552 0.6036 0.3059 0.4913 0.8303 0.0030 0.0011
40 | 0.1980 0.0848 1.9246 0.6064 0.3077 0.4751 0.8318 0.0028 0.0010
41 | 0.1981 0.0784 1.8554 0.5984 0.3093 0.4679 0.8323 0.0025 0.0009
42 | 0.1982 0.0758 1.9277 0.5950 0.3143 0.4712 0.8327 0.0024 0.0009
43 | 0.1983 0.0779 1.9803 0.6063 0.3222 0.4834 0.8331 0.0024 0.0009
44 | 0.1984 0.0801 2.0635 0.6068 0.3305 0.5028 0.8343 0.0024 0.0010
45 | 0.1985 0.0808 2.0671 0.5814 0.3383 0.5067 0.8358 0.0024 0.0010
46 | 0.1986 0.0844 2.1614 0.5804 0.3497 0.5328 0.8382 0.0024 0.0010
47 | 0.1987 0.0787 2.0468 0.5423 0.3626 0.5534 0.8414 0.0022 0.0009
48 | 0.1988 0.0813 2.2838 0.5869 0.3764 0.5739 0.8459 0.0022 0.0010
49 | 0.1989 0.0904 2.3531 0.5790 0.3887 0.5910 0.8527 0.0023 0.0011
50 | 0.1990 0.0772 2.2497 0.5501 0.3925 0.5630 0.8591 0.0020 0.0009
51 | 0.1991 0.0745 2.1057 0.4832 0.3945 0.5751 0.8644 0.0019 0.0009
52 | 0.1992 0.0759 2.0727 0.4705 0.3906 0.5879 0.8692 0.0019 0.0009
53 | 0.1993 0.0632 1.9741 0.4334 0.3882 0.5590 0.8745 0.0016 0.0007
54 | 0.1994 0.0589 2.1083 0.4221 0.3912 0.5655 0.8816 0.0015 0.0007
55 | 0.1995 0.0572 2.1173 0.3965 0.3953 0.5761 0.8837 0.0014 0.0006
56 | 0.1996 0.0537 2.0810 0.3837 0.3981 0.5683 0.8844 0.0013 0.0006
57 | 0.1997 0.0541 2.1280 0.3917 0.4053 0.5577 0.8848 0.0013 0.0006
58 | 0.1998 0.0531 2.1356 0.3883 0.4145 0.5427 0.8854 0.0013 0.0006
59 | 0.1999 0.0580 2.1964 0.4043 0.4259 0.5464 0.8861 0.0014 0.0007
60 | 0.2000 0.0591 2.2032 0.4104 0.4388 0.5361 0.8883 0.0014 0.0007
61 | 0.2001 0.0583 2.2330 0.4056 0.4428 0.5422 0.8909 0.0013 0.0006
62 | 0.2002 0.0560 2.4747 0.4592 0.4468 0.5515 0.8941 0.0012 0.0006
63 | 0.2003 0.0529 2.7103 0.4664 0.4511 0.5547 0.8976 0.0012 0.0006
64 | 0.2004 0.0480 2.6582 0.4022 0.4570 0.5546 0.9011 0.0011 0.0005
65 | 0.2005 0.0440 2.6459 0.3915 0.4633 0.5506 0.9048 0.0010 0.0005
66 | 0.2006 0.0445 2.6636 0.3959 0.4701 0.5363 0.9113 0.0010 0.0005
67 | 0.2007 0.0471 2.6749 0.3824 0.4782 0.5253 0.9183 0.0010 0.0005
68 | 0.2008 0.0397 2.6248 0.3657 0.4809 0.4930 0.9256 0.0008 0.0004
69 | 0.2009 0.0358 2.5281 0.3460 0.4837 0.4845 0.9341 0.0007 0.0004
70 | 0.2010 0.0266 2.3305 0.2888 0.4884 0.4550 0.9416 0.0005 0.0003 ];
71 traffic = struct('year',data(:,1),'killed',data(:,2), ...
72               'cars',data(:,5),'resident',data(:,7))
73
74 % (Step 1) Poisson regression w/ one regressor
75 X1 = traffic.year;
76 y = traffic.killed;
77 [beta1 dev1 stats1 ] = glmfit(X1,y,'poisson')
78 xx1 = [ ones(size(X1)), X1 ];
79 myFit1 = exp(xx1*beta1);
80 GLMfit1 = glmval(beta1,X1,'log');
81 dumA1 = max(abs(myFit1-GLMfit1))
82 resid1 = xx1'*(y-myFit1)

```

```

83| dumB1 = norm(resid1)
84|
85| % (Step 2) Poisson regression w/ two regressors
86| X2 = [ traffic.year, traffic.cars ];
87| [beta2 dev2 stats2 ] = glmfit(X2,y,'poisson')
88| xx2 = [ ones(size(X1)), X2 ];
89| myFit2 = exp(xx2*beta2);
90| GLMfit2 = glmval(beta2,X2,'log');
91| dumA2 = max(abs(myFit2-GLMfit2));
92| resid2 = xx2'*(y-myFit2);
93| dumB2 = norm(resid2)
94|
95| % (Step 3) Poisson regression w/ three regressors
96| X3 = [ traffic.year, traffic.cars, traffic.resident ];
97| [beta3 dev3 stats3 ] = glmfit(X3,y,'poisson')
98| xx3 = [ ones(size(X1)), X3 ];
99| myFit3 = exp(xx3*beta3);
100| GLMfit3 = glmval(beta3,X3,'log');
101| dumA3 = max(abs(myFit3-GLMfit3))
102| resid3 = xx3'*(y-myFit3);
103| dumB3 = norm(resid3)
104|
105| % (Step 4) Plotting
106| figure
107| plot(X1,y,'ro','LineWidth',2)
108| set(gca,'fontsize',11,'fontweigh','bold','xlim',[1948 2012])
109| hold on
110| plot(X1,myFit1,'b:','LineWidth',2)
111| plot(X1,myFit2,'g-','LineWidth',2)
112| plot(X1,myFit2,'k--','LineWidth',2)
113| hold off
114| legend('No of Accidents','One Regressor','Two Regressors', ...
115|       'Three Regressors','location','SW')
116| xlabel('Year');
117| ylabel('No of Killed Persons');
118| saveas(gcf,'RoadAccident1950_2010a','jpg')
119|
120| % (Step 5) Model selection-Deviance
121| diffDev21 = dev1 - dev2
122| prob21 = chi2cdf(diffDev21,1)
123| diffDev32 = dev2 - dev3
124| prob32 = chi2cdf(diffDev32,1)
125| diffDev31 = dev1 - dev3
126| prob31 = chi2cdf(diffDev31,2)
127| % End of program
128| % -----

```

유감스럽게도 웹사이트 <http://www.trafikverket.se>는 독일어로 작성되어 읽기가 쉽지 않다. 따라서, 이 데이터셋 1950-RoadData를 ASCII로 바꾸어 이 MATLAB프로그램에 포함시킨다. 이 데이터셋에서 첫 번째 변수, 두 번째 변수, 다섯 번째 변수, 그리고 여섯 번째 변수를 뽑아서 각각 traffic.year, traffic.killed, traffic.cars, 그리고 traffic.resident로 명명한다. 즉, MATLAB명령문 struct.m을 적용해서 traffic이라는 구조를 만들고 이 구조 안에 year, killed, cars, 그리고 resident라는 변수들을 포함시킨다.

이 MATLAB프로그램 RoadAccident1950\_2010a의 첫 번째 단계는 MATLAB함수 glm-

fit.m을 사용해서 변수 traffic.killed를 종속변수로 하고 변수 traffic.year를 설명변수로 하는 Poisson 회귀모형을 추정한 식은 다음과 같다.

$$\hat{\mu}_j = \exp(33.7538 - 0.0137 * \text{year}_j) \quad (7)$$

두 번째 단계는 MATLAB함수 glmfit.m을 사용해서 변수 traffic.killed를 종속변수로 하고 변수 traffic.year와 변수 traffic.cars를 설명변수들로 하는 Poisson 회귀모형을 추정한 식은 다음과 같다.

$$\hat{\mu}_j = \exp(184.9700 - 0.0915 * \text{year}_j + 0.0010 * \text{cars}_j) \quad (8)$$

세 번째 단계는 MATLAB함수 glmfit.m을 사용해서 변수 traffic.killed를 종속변수로 하고 변수 traffic.year, 변수 traffic.cars와 변수 traffic.resident를 설명변수들로 하는 Poisson 회귀모형을 추정한 식은 다음과 같다.

$$\hat{\mu}_j = \exp(186.4392 - 0.0926 * \text{year}_j + 0.0010 * \text{cars}_j - 0.0001 * \text{resident}_j) \quad (9)$$

식 (8)과 식 (9)에서 알 수 있듯이, 설명변수 traffic.resident는 종속변수 traffic.year를 추정하는데 추가적인 정보를 주지 못한다.

네 번째 단계로 관찰값과 적합값들  $\{\hat{\mu}_j\}$ 를 그린 그래프가 그림 3.1.6에 그려져 있다.

다섯 번째로 이탈통계량(deviance)을 사용해서 적합한 모형을 선택한다. 식  $q < p$ 를 만족하는  $(p, q)$ 에 대해서 다음과 같이 이탈통계량을 정의한다.

$$\text{DEV}(p, q) = 2 [\ln L(\beta_1, \dots, \beta_p) - \ln L(\beta_1, \dots, \beta_q)] \quad (10)$$

다음 가설검정문제를 생각해보자.

$$H_0 : \beta_{q+1} = \beta_{q+2} = \dots = \beta_p = 0 \text{ vs } H_1 : H_0 \text{ is not true.} \quad (11)$$

만약 적절한 정규성조건(regularity condtions)이 만족되고 또한 관찰점들의 개수  $n$ 이 충분히 크면, 귀무가설  $H_0$  하에서 다음 식이 성립한다.

$$\text{DEV}(p, q) \stackrel{d}{\sim} \chi_{p-q}^2 \quad (12)$$

MATLAB 함수 `glmfit.m`에서 출력되는 `dev`는 다음과 같다.

$$\text{dev}(p) = -2 \ln L(\beta_1, \dots, \beta_p) \quad (13)$$

따라서 다음 식이 성립한다.

$$\text{DEV}(p, q) = 2 [\ln L(\beta_1, \dots, \beta_p) - \ln L(\beta_1, \dots, \beta_q)] = \text{dev}(q) - \text{dev}(p) \quad (14)$$

이 MATLAB 프로그램 `RoadAccident1950_2010a`를 실행하면, 다음 식들을 얻는다.

$$\text{DEV}(2, 1) = \text{dev}(1) - \text{dev}(2) = 2972.4 - 481.8 = 2490.6 \quad (15)$$

$$\text{DEV}(3, 2) = \text{dev}(2) - \text{dev}(3) = 481.8714 - 480.0320 = 1.8394 \quad (16)$$

$$\text{DEV}(3, 1) = \text{dev}(1) - \text{dev}(3) = 2972.4 - 480.0 = 2492.4 \quad (17)$$

식 (15)에서 알 수 있듯이,  $p_{21} \doteq \Pr(\chi_{2-1}^2 < 2490.6) = 1 > 0.95$ 이다. 따라서, 첫 번째 설명변수 `traffic.year`만 사용한 Poisson 선형모형보다 두 번째 설명변수 `traffic.cars`를 추가한 Poisson 선형모형이 데이터세트 1950-RoadData에 더 적합하다. 식 (16)에서 알 수 있듯이,  $p_{32} \doteq \Pr(\chi_{3-2}^2 < 1.8394) = 0.8250 < 0.95$ 이다. 따라서, 첫 번째 설명변수 `traffic.year`와 두 번째 설명변수 `traffic.cars`를 사용한 Poisson 선형모형에 세 번째 설명변수 `traffic.resident`를 추가한 Poisson 선형모형이 데이터세트 1950-RoadData에 더 적합하다고 할 수 없다. 식 (17)에서 알 수 있듯이,  $p_{31} \doteq \Pr(\chi_{3-1}^2 < 2492.4) = 1 > 0.95$ 이다. 따라서, 첫 번째 설명변수 `traffic.year`만 사용한 Poisson 선형모형보다 두 번째 설명변수 `traffic.cars`와 세 번째 설명변수 `traffic.resident`를 추가한 Poisson 선형모형이 데이터세트 1950-RoadData에 더 적합하다. ■

**예제 3.1.7** 최소제곱추정법과 GLM추정법 (generalized linear model estimation: GLM estimation)을 비교하는 예로서, 다음 MATLAB 프로그램 `GLMfit103.m`을 실행해 보자.

```

1 % -----
2 % Filename: GLMfit103.m
3 % Generalized Linear Models encompasses techniques
4 % like Poisson regression and Logistic regression.
5 % -----
6
7 clear all, clf, close all, clc
8

```

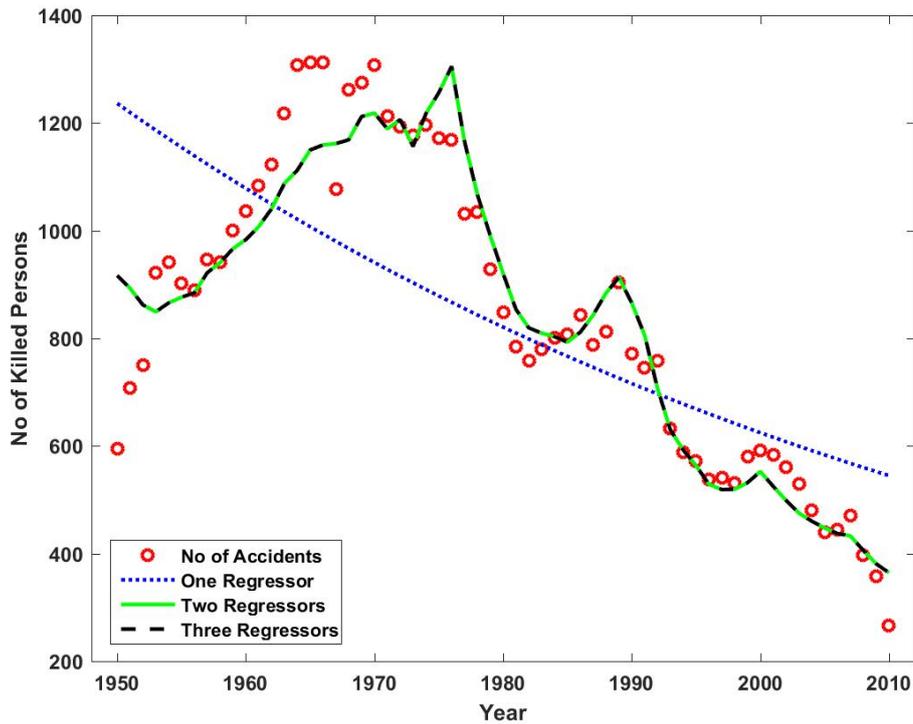


그림 3.1.6. 교통사고 사망자의 Poisson 회귀분석

```

9 % Create a dataset from Poisson regression model
10 Npoints = 16
11 xx = linspace(0,Npoints-1,Npoints)';
12 mu = @(x) exp(-1.0 + 1/4*x);
13 yy = mu(xx);
14 rng(1); % Seed number
15 NoisedY = poissrnd(yy,Npoints,1); % mean + noise
16
17 % Fitting the nonlinear regression model using OLS method
18 OLSfit = fit(xx, NoisedY, 'exp(B1 + B2*x)', 'StartPoint', [0.5 0.3])
19 SSE_OLS = sum((OLSfit(xx) - NoisedY).^2)
20
21 % Fitting the nonlinear regression model using GLM method
22 [b dev stats] = glmfit(xx,NoisedY, 'poisson')
23 GLMfit = exp(b(1) + b(2)*xx)
24 SSE_GLM = sum(stats.resid.^2)
25 DiffSSE = SSE_OLS - SSE_GLM
26
27 % OLS vs GLM plot
28 figure
29 plot(xx,yy,'g','LineWidth',2)
30 hold on
31 plot(xx,NoisedY,'b*');
32 plot(OLSfit)
33 plot(xx, GLMfit, 'k')
34 hold off
35 set(gca,'fontsize',11,'fontweigh','bold')
36 legend('Mean Curve','Poisson Observations','OLS','GLM','location','NW')
37 xlabel('\bf Time','FontSize',12');
38 ylabel('\bf Count','FontSize',12');

```

```

39 saveas(gcf,'GLMfit103a','jpg')
40
41 % Perform a Monte Carlo simulation
42 Nsimulation = 1024
43 parfor jj = 1:Nsimulation
44     NoisedY = poissrnd(yy,Npoints,1);
45     OLSfit = fit(xx, NoisedY,'exp(B1 + B2*x)','Startpoint',[0.5 0.3]);
46     B = coeffvalues(OLSfit);
47     OLS_Prediction(:,jj) = exp(B(1) + B(2)*xx);
48     b = glmfit(xx,NoisedY, 'poisson');
49     GLM_Prediction(:,jj) = exp(b(1) + b(2)*xx);
50 end
51
52 % Plot to show GLM is more robust than OLS.
53 figure
54 subplot(2,2,1)
55 plot(xx,yy,'g','LineWidth',2)
56 set(gca,'fontsize',11,'fontweigh','bold')
57 hold on
58 OLS_Mean = plot(xx, mean(OLS_Prediction, 2),'r-','LineWidth',2)
59 GLM_Mean = plot(xx, mean(GLM_Prediction, 2),'k--','LineWidth',2)
60 hold off
61 set(gca,'fontsize',11,'fontweigh','bold')
62 legend('Mean Curve','OLS','GLM','location','NW')
63 xlabel('\bf Time','FontSize',12');
64 ylabel('\bf Count','FontSize',12');
65 subplot(2,2,2)
66 OLS_std = std(OLS_Prediction, [], 2);
67 plot(xx, OLS_std,'r-','LineWidth',2)
68 set(gca,'fontsize',11,'fontweigh','bold')
69 legend('Std of OLS','location','NW')
70 xlabel('\bf Time','FontSize',12');
71 ylabel('\bf Std','FontSize',12');
72 subplot(2,2,3)
73 GLM_std = std(GLM_Prediction, [], 2);
74 plot(xx, GLM_std,'k--','LineWidth',2)
75 set(gca,'fontsize',11,'fontweigh','bold')
76 legend('Std of GLM','location','NW')
77 xlabel('\bf Time','FontSize',12');
78 ylabel('\bf Std','FontSize',12');
79 subplot(2,2,4)
80 plot(xx, OLS_std./GLM_std,'b:','LineWidth',2)
81 set(gca,'fontsize',11,'fontweigh','bold')
82 legend('Std Ratio','location','SW')
83 xlabel('\bf Time','FontSize',12');
84 ylabel('\bf Std Ratio','FontSize',12');
85 saveas(gcf,'GLMfit103b','jpg')
86 % End of program
87 % -----

```

이 MATLAB 프로그램 GLMfit103.m에서는 확률변수들  $y_1, y_2, \dots, y_{16}$  이 서로 독립이고 Poisson 확률분포를 따른다. 즉,  $y_1, y_2, \dots, y_{16}$  은 다음 Poisson 회귀모형을 따른다.

$$y_j | \mu_j \stackrel{d}{\sim} \text{Poisson}(\mu_j), \quad (j = 1, 2, \dots, 16) \quad (1)$$

확률변수  $y_j$ 의 기대값  $\mu_j$ 가 다음 식을 만족한다고 가정하자.

$$\ln \mu_j = \alpha + \beta x_j \quad (2)$$

여기서  $\alpha = -1.0$ ,  $\beta = 0.25$  그리고  $x_j = j - 1$ 이라고 하자. 즉, 평균곡선(mean curve)은 다음과 같다.

$$\mu_j = \exp(-1.0 + 0.25[j - 1]), \quad (j = 1, 2, \dots, 16) \quad (3)$$

MATLAB 프로그램 GLMfit103.m을 실행하면, 그림 3.1.7이 그려진다. 식 (2)의 평균곡선이 그림 3.1.7에 녹색 실선으로 그려져 있다. 식 (1)의 Poisson 회귀모형에서 발생된 관찰값들  $y_1, y_2, \dots, y_{16}$ 을 변수 NoiseY에 저장한다. 이 관찰값들이 그림 3.1.7에 청색 별표로 그려져 있다.

다음과 같은 비선형회귀모형을 살펴보자.

$$y_j = \exp(B_1 + B_2 x_j) + \epsilon_j, \quad (j = 1, 2, \dots, 16) \quad (4)$$

이 비선형회귀모형의 최소제곱추정식은 다음과 같다.

$$y_j = \exp(-0.5841 + 0.2014x_j) + \hat{\epsilon}_j, \quad (j = 1, 2, \dots, 16) \quad (5)$$

이 최소제곱추정식이 그림 3.1.7에 적색 실선으로 그려져 있다. 또한 최소제곱추정식에 의한 오차제곱합(SSE)은 45.0445이다.

일반화선형모형을 추정하는데는 주로 반복재가중최소제곱추정법(iteratively reweighted least squares method) 또는 최우추정법을 사용한다. 반복재가중최소제곱추정법을 사용해서 일반화선형모형 (1)을 추정한 결과는 다음과 같다.

$$\hat{\mu}_j = \exp(-0.8609 + 0.2240[j - 1]), \quad (j = 1, 2, \dots, 16) \quad (6)$$

이 일반화선형모형추정식이 그림 3.1.7에 적색 실선으로 그려져 있다. 또한 일반화선형모형추정식에 의한 오차제곱합은 46.3384이다.

그림 3.1.7에서 알 수 있듯이, 최소제곱추정식과 일반화선형모형추정식이 거의 비슷하다. 또한, 오차제곱합들도 아주 비슷하다. 그렇다면, 왜 최소제곱추정법 대신 일반화선형모형을

사용해야 할까? 그 답은 적합값들의 안정성에 있다. MATLAB 프로그램 GLMfit103.m을 실행하면, 그림 3.1.8이 그려진다. 그림 3.1.8의 좌측상단에는 각 추정법에 의해 추정된 평균곡선이 그려져 있다. 이 그래프에서 확인할 수 있듯이, 추정된 평균곡선들 사이에는 차이가 거의 없다. 그림 3.1.8의 우측상단에는 최소제곱추정법에 의한 추정값들의 표본편차들이 그려져 있고, 좌측하단에는 일반선형모형추정법에 의한 추정값들의 표본편차들이 그려져 있다. 이 그래프들에서 확인할 수 있듯이, 일반적으로 전자가 후자보다 크다. 또한, 그림 3.1.8의 우측하단에는 최소제곱추정법에 의한 추정값의 표본편차를 일반선형모형추정법에 의한 추정값의 표본편차로 나눈 값들의 그래프가 그려져 있다. 이 그래프에서 알 수 있듯이, 최소제곱추정법에 의한 추정값의 표본편차가 일반선형모형추정법에 의한 추정값의 표본편차보다 크다. ■

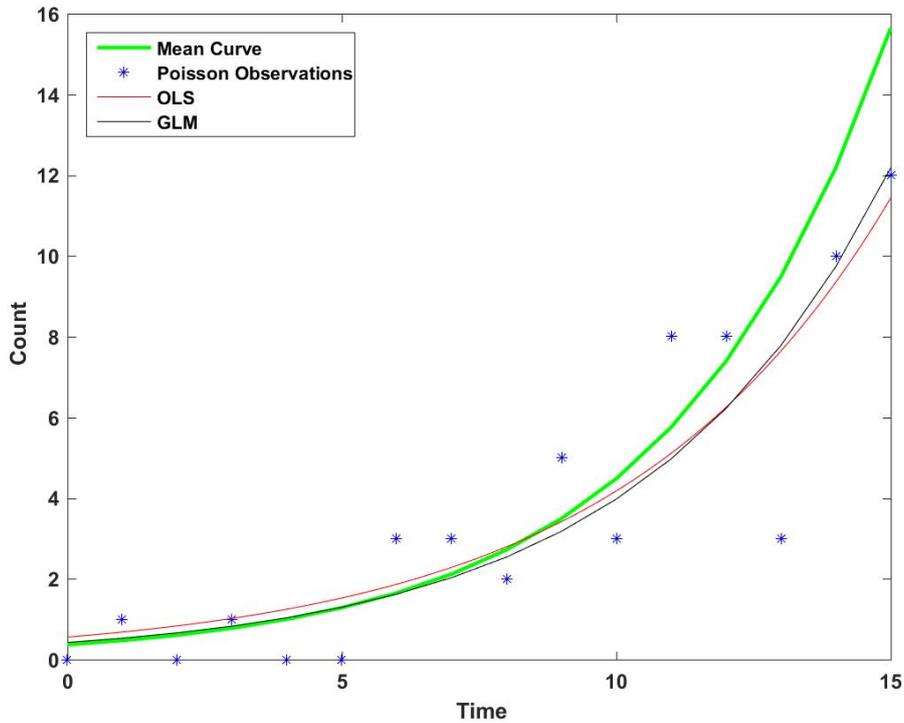


그림 3.1.7. 최소제곱추정법과 GLM

**예제 3.1.8** 로지스틱회귀모형을 베이지안추정하는 예로서, 다음 MATLAB 프로그램 PoissonBayesianRegress101.m을 실행해 보자.

```

1 % -----
2 % Filename: PoissonBayesianRegress101.m
3 % https://kr.mathworks.com/help/stats/examples
    
```

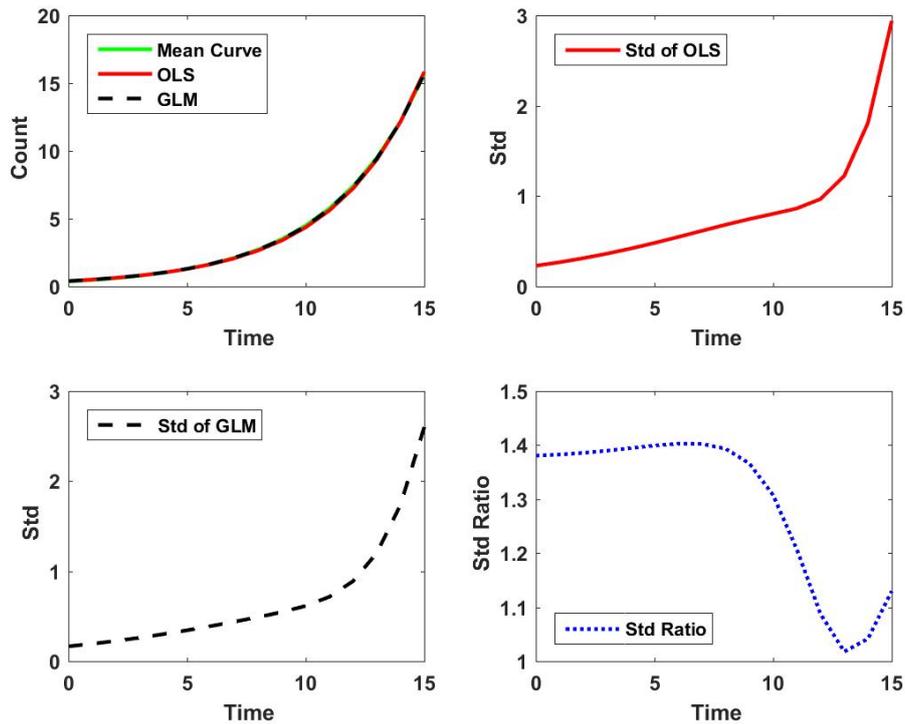


그림 3.1.8. GLM의 강건성

```

4 % /bayesian-analysis-for-a-logistic-regression-model.html
5 % -----
6 clear all, clf, close all
7 % Making a dataset
8 % % A set of car weights
9 weight = [2100 2300 2500 2700 2900 3100 3300 3500 3700 3900 4100 4300]';
10 weight = (weight-2800)/1000; % recenter and rescale
11 % % The number of cars tested at each weight
12 total = [48 42 31 34 31 21 23 23 21 16 17 21]';
13 % % The number of cars that have poor mpg performances at each weight
14 poor = [1 2 0 3 8 8 14 17 19 15 17 21]';
15
16 % Logistic Regression
17 % % Model
18 logitp = @(b,x) exp(b(1)+b(2).*x)./(1+exp(b(1)+b(2).*x))
19 prior1 = @(b1) normpdf(b1,0,20); % prior for intercept
20 prior2 = @(b2) normpdf(b2,0,20); % prior for slope
21 post = @(b) prod(binopdf(poor,total,logitp(b,weight))) ... % likelihood
22 *prior1(b(1))*prior2(b(2)); % priors
23 % % 3D-Plot of Posterior PDF w/o the constant term
24 b1 = linspace(-2.5, -1, 50);
25 b2 = linspace(3, 5.5, 50);
26 simpost = zeros(50,50);
27 for i = 1:length(b1)
28     for j = 1:length(b2)
29         simpost(i,j) = post([b1(i), b2(j)]);
30     end;
31 end;
32 mesh(b2,b1,simpost)
33 set(gca,'fontsize',11,'fontweigh','bold')

```

```

34 xlabel('Slope')
35 ylabel('Intercept')
36 zlabel('Posterior PDF')
37 saveas(gcf, 'PoissonBayesianRegress101M', 'jpg')
38 % End of program
39 % -----

```

이 MATLAB 프로그램 PoissonBayesianRegress101.m은 MathWorks에서 제공한 것을 약간 변형한 것이다. 이 예제에서 사용하는 데이터세트는 개스마일리지 (miles per gallon; mpg) 테스트에 관한 것으로, 차의 무게와 개스마일리지 사이의 관계에 대해서 살펴보자. 변수 *weight*는 차의 무게를 나타내고, 변수 *total*은 각 *weight* 별로 검사받는 차량의 대수를 나타내고, 변수 *poor*는 각 *weight* 별로 검사에 통과하지 못하는 차량의 대수를 나타낸다.

우리가 베이지안추정하고자 하는 로지스틱회귀모형은 다음과 같다.

$$\Pr(\text{failure}) = \frac{\exp(\beta_1 + \beta_2 * \text{weight})}{1 + \exp(\beta_1 + \beta_2 * \text{weight})} \quad (1)$$

사전확률밀도함수는 다음과 같다.

$$\pi_0(\beta_1, \beta_2) = \frac{1}{\sqrt{40\pi}} \exp\left(-\frac{1}{40}\beta_1^2\right) \frac{1}{\sqrt{40\pi}} \exp\left(-\frac{1}{40}\beta_2^2\right) \quad (2)$$

우도함수는 다음과 같다.

$$L(\beta_1, \beta_2) = \left[ \frac{\exp(\beta_1 + \beta_2 * \text{weight})}{1 + \exp(\beta_1 + \beta_2 * \text{weight})} \right]^{poor} \left[ \frac{1}{1 + \exp(\beta_1 + \beta_2 * \text{weight})} \right]^{total-poor} \quad (3)$$

따라서 사후확률밀도함수는 다음과 같다.

$$\pi_1(\beta_1, \beta_2 | \text{poor}, \text{total}) = C\pi_0(\beta_1, \beta_2)L(\beta_1, \beta_2 | \text{poor}, \text{total}) \quad (4)$$

여기서  $C$ 는 정규화상수 (normalizing constant)이다. 식 (4)의 우변을 적분해서 이 정규화상수를 구하는 것은 불가능해 보인다. 그러나 뒤에서 배우는 MCMC를 사용해서  $C$ 를 수치적으로 구할 수 있다. 이  $C$ 를 몰라도 사후확률밀도함수의 형태를 짐작할 수 있다.

MATLAB 프로그램 PoissonBayesianRegress101.m을 실행하면, 이 사후확률밀도함수를 상수  $C$ 로 나눈 3D그래프가 그림 3.1.9에 그려진다. 두 모수들, 즉 절편 (intercept)과 기울기 (slope)의 사전주변확률분포들은 서로 독립이라고 가정했지만, 그림 3.1.9의 3D그래프에서 알 수 있듯이 사후결합확률밀도함수는 모수공간에서 대각선을 따라 길쭉한 형태를 띠고

있다. 즉, 사후주변확률분포들은 서로 독립이 아니다. 우도함수를 계산할 때 관찰값들이 서로 독립이라고 가정한 것을 고려한다면, 이 결과는 의외이다. 사후확률분포가 내포하는 상관성은 사전확률밀도함수와 우도함수가 합쳐지는 과정에서 발생한 것이다. ■

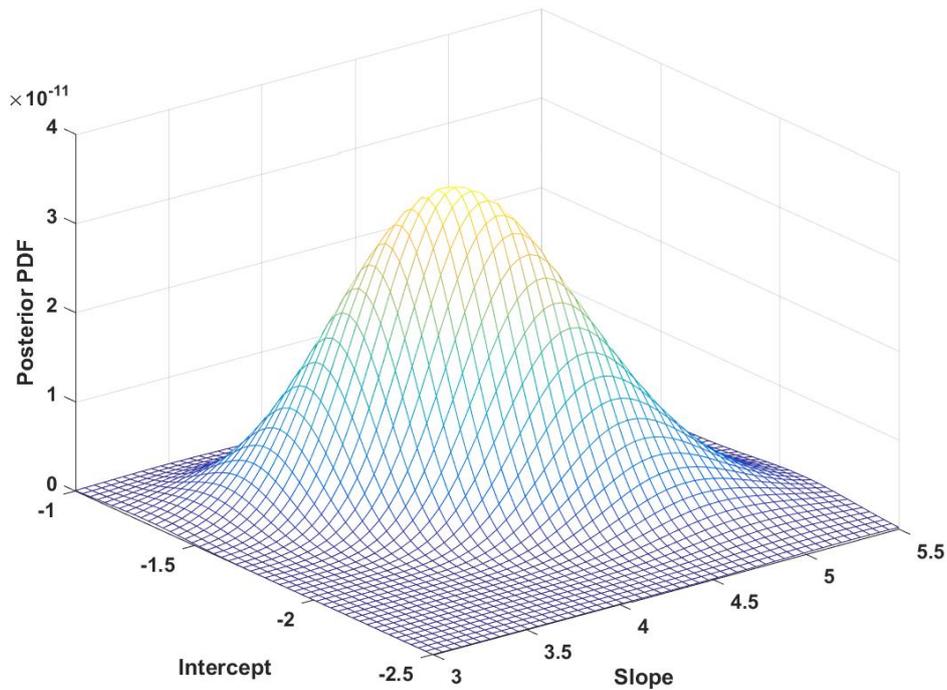


그림 3.1.9. 로지스틱회귀모형의 베이지안추정

**예제 3.1.9** 베이지안데이터분석에서는 사후확률분포를 요약하기 위해서 Monte Carlo 방법을 자주 사용한다. 예제 3.1.8의 결과를 분석하기 위해서, 다음 MATLAB 프로그램 PoissonBayesianRegress102.m을 실행해 보자.

```

1 % -----
2 % Filename: PoissonBayesianRegress102.m
3 % https://kr.mathworks.com/help/stats/examples
4 % /bayesian-analysis-for-a-logistic-regression-model.html
5 % -----
6 clear all, clf, close all
7 % Making a dataset
8 % % A set of car weights
9 weight = [2100 2300 2500 2700 2900 3100 3300 3500 3700 3900 4100 4300]';
10 weight = (weight-2800)/1000; % recenter and rescale
11 % % The number of cars tested at each weight
12 total = [48 42 31 34 31 21 23 23 21 16 17 21]';
13 % % The number of cars that have poor mpg performances at each weight
    
```

```

14 | poor = [1 2 0 3 8 8 14 17 19 15 17 21]';
15 |
16 | % (Pre-step) Logistic Regression Model
17 | logitp = @(b,x) exp(b(1)+b(2).*x)./(1+exp(b(1)+b(2).*x))
18 | prior1 = @(b1) normpdf(b1,0,20); % prior for intercept
19 | prior2 = @(b2) normpdf(b2,0,20); % prior for slope
20 | post = @(b) prod(binopdf(poor,total,logitp(b,weight))) ... % likelihood
21 |         *prior1(b(1))*prior2(b(2)); % priors
22 | b1 = linspace(-2.5, -1, 50);
23 | b2 = linspace(3, 5.5, 50);
24 |
25 | % (Step 1) Slice Sampler
26 | initial = [1 1];
27 | nsamples = 1000;
28 | trace = slicesample(initial,nsamples,'pdf',post,'width',[20 2]);
29 | figure
30 | subplot(2,2,1)
31 | plot(trace(:,1))
32 | set(gca,'fontsize',11,'fontweigh','bold')
33 | ylabel('Intercept');
34 | xlabel('Sample Number');
35 | subplot(2,2,2)
36 | plot(trace(:,2))
37 | set(gca,'fontsize',11,'fontweigh','bold')
38 | ylabel('Slope');
39 | xlabel('Sample Number');
40 | % % Smoothing
41 | movavg = filter((1/50)*ones(50,1), 1, trace);
42 | subplot(2,2,3)
43 | plot(movavg(:,1))
44 | set(gca,'fontsize',11,'fontweigh','bold')
45 | xlabel('Sample Number');
46 | ylabel('Means of the intercept');
47 | subplot(2,2,4)
48 | plot(movavg(:,2))
49 | set(gca,'fontsize',11,'fontweigh','bold')
50 | xlabel('Sample Number');
51 | ylabel('Means of the slope');
52 | saveas(gcf,'PoissonBayesianRegress102aM','jpg')
53 |
54 | % (Step 2) Eleminate Burn-in samples
55 | trace = slicesample(initial,nsamples,'pdf',post,'width',[20 2], ...
56 |                 'burnin',50);
57 | figure
58 | subplot(2,2,1)
59 | plot(trace(:,1))
60 | set(gca,'fontsize',11,'fontweigh','bold')
61 | ylabel('Intercept');
62 | xlabel('Sample Number');
63 | subplot(2,2,2)
64 | plot(trace(:,2))
65 | set(gca,'fontsize',11,'fontweigh','bold')
66 | ylabel('Slope');
67 | xlabel('Sample Number');
68 |
69 | % (Step 3) Thining-Out
70 | trace = slicesample(initial,nsamples,'pdf',post, ...
71 |                 'width',[20 2],'burnin',50,'thin',10);
72 | subplot(2,2,3)
73 | plot(trace(:,1))
74 | set(gca,'fontsize',11,'fontweigh','bold')

```

```

75 ylabel('Intercept');
76 xlabel('Sample Number');
77 subplot(2,2,4)
78 plot(trace(:,2))
79 set(gca,'fontsize',11,'fontweigh','bold')
80 ylabel('Slope');
81 xlabel('Sample Number');
82 saveas(gcf,'PoissonBayesianRegress102bM','jpg')
83
84 % (Step 4) Independence Test through Spectral Analysis
85 traceFFT = fft(detrend(trace,'constant'));
86 tracePeriodogram = traceFFT .* conj(traceFFT);
87 ACF = ifft(tracePeriodogram);
88 ACF = ACF(1:21,:); % Retain lags up to 20.
89 ACF = real([ACF(1:21,1) ./ ACF(1,1) ...
90           ACF(1:21,2) ./ ACF(1,2)]); % Normalize.
91 figure
92 bounds = sqrt(1/nsamples) * [2 ; -2]; % 95% CI for iid normal
93 labs = {'Sample ACF for intercept','Sample ACF for slope' };
94 for ii = 1:2
95     subplot(2,1,ii)
96     lineHandles = stem(0:20, ACF(:,ii) , 'filled' , 'r-o');
97     lineHandles.MarkerSize = 4;
98     grid('on')
99     xlabel('Lag')
100    ylabel(labs{ii})
101    hold on
102    plot([0.5 0.5 ; 20 20] , [bounds([1 1]) bounds([2 2])] , '-b');
103    plot([0 20] , [0 0] , '-k');
104    hold off
105    set(gca,'fontsize',11,'fontweigh','bold')
106    a = axis;
107    axis([a(1:3) 1]);
108 end
109 saveas(gcf,'PoissonBayesianRegress102cM','jpg')
110
111 % (Step 4) 3D Histogram
112 figure
113 subplot(1,1,1)
114 hist3(trace,[25,25]);
115 set(gca,'fontsize',11,'fontweigh','bold')
116 xlabel('Intercept')
117 ylabel('Slope')
118 zlabel('Posterior density')
119 view(-110,30)
120 saveas(gcf,'PoissonBayesianRegress102dM','jpg')
121
122 % (Step 6) Statistical Inference
123 figure
124 subplot(2,2,1)
125 hist(trace(:,1))
126 set(gca,'fontsize',11,'fontweigh','bold')
127 xlabel('Intercept');
128 subplot(2,2,2)
129 ksdensity(trace(:,2))
130 set(gca,'fontsize',11,'fontweigh','bold')
131 xlabel('Slope');
132 csum = cumsum(trace);
133 subplot(2,2,3)
134 plot(csum(:,1)'./(1:nsamples))
135 set(gca,'fontsize',11,'fontweigh','bold')

```

```

136 xlabel('Number of samples')
137 ylabel('Means of the intercept');
138 subplot(2,2,4)
139 plot(csum(:,2)'./(1:nsamples))
140 set(gca,'fontsize',11,'fontweigh','bold')
141 xlabel('Number of samples')
142 ylabel('Means of the slope');
143 saveas(gcf,'PoissonBayesianRegress102eM','jpg')
144 % % Estimation
145 bHat = mean(trace)
146 % End of program
147 % -----

```

이 MATLAB 프로그램 PoissonBayesianRegress102.m의 사전 단계는 예제 3.1.9의 로지스틱회귀분석을 다시 실행하는 것이다.

제1단계에서는 슬라이스샘플러(slice sampler)를 사용해서 사후확률밀도함수로부터 난수들을 발생시킨다. 예제 3.1.9에서 언급했듯이, 정규화상수(normalizing constant)  $C$ 를 해석적으로 구하는 것은 현재로서는 거의 불가능하다. 그러나 정규화상수를 모르는 경우에도 MCMC를 사용해서 사후확률분포로부터 난수들(random numbers)을 생성할 수 있다. 또한, 이 난수들을 사용해서 사후확률분포의 적률들을 추정할 수 있다. Gibbs샘플러와 마찬가지로 슬라이스샘플링은 정규화상수가 미지인 사후확률분포에서 난수들을 생성할 수 있는 MCMC 알고리즘이다. 다른 MCMC 알고리즘과는 달리 슬라이스샘플링은 목적확률분포(target probability distribution)나 주변확률분포(marginal probability distribution)를 사용하지 않고 오로지 사후확률밀도함수에 비례하는 확률밀도함수, 즉 스케일드(scaled) 사후확률밀도함수로부터 난수들을 생성한다. 이 MATLAB 프로그램 PoissonBayesianRegress102.m에서는 슬라이스샘플링을 적용해서 1000개 난수들을 생성하고 이를 변수 trace에 저장한다. 이 난수들이 그림 3.1.10의 상단에 그려져 있다. 좌측상단은 절편의 난수들의 그래프이고, 우측상단은 기울기의 난수들의 그래프이다. 이 그래프들에서 알 수 있듯이, 이 난수들은 초기값들에 영향을 많이 받아, 첫 50개 정도 난수들은 정상성(stationarity)을 보이지 않는다. 따라서 난수들의 정상성을 조사하기 위해서 이동평균윈도우(moving average window)를 적용해보기로 하자. MATLAB 함수 filter를 사용해서 이동평균을 실행할 수 있다. 다음 MATLAB 명령문을 실행해보자.

```
>> MoAv = filter(1/4*ones(4,1),1,1:12)
```

이 MATLAB 명령문을 실행한 결과는 다음과 같다.

```

MoAv =
    0.2500    0.7500    1.5000    2.5000    3.5000    4.5000
    5.5000    6.5000    7.5000    8.5000    9.5000   10.5000

```

즉, 이 MATLAB 필터는 입력된 수열  $\{x_i\}$ 를 이동평균한 수열을 출력한다. 이 이동평균값들의 시계열산점도들이 그림 3.1.10의 하단에 그려져 있다. 좌측하단은 절편의 이동평균값들의 그래프이고, 우측하단은 기울기의 이동평균값들의 그래프이다. 이 그래프들에서 알 수 있듯이, 첫 50개 또는 100개 정도 이후의 이동평균값들은 정상성을 보인다. 또한, 이 두 그래프들에서 절편과 기울기는 음의 상관관계를 보임을 알 수 있다.

제2단계에서는 초기값들의 영향력을 제거하기 위해서 첫 50개 관찰값들을 번인(burn-in)시키는 것이다. 즉, 첫 50개 관찰점들을 데이터세트에서 제거한다. 남은 관찰점들의 시계열산점도들이 그림 3.1.11의 상단에 그려져 있다.

제3단계는 생성된 난수들의 난수성을 조사하기 위한 것이다. 우리가 MCMC에서 계산하는 것은 표본평균이므로 설사 관찰점들 사이에 자기상관성 (autocorrelation)이 있다고 해도 그 목적을 달성할 수는 있다. 그러나 난수성의 결여는 계산된 표본평균의 통계적 추론을 어렵게 한다. 그림 3.1.11의 상단 그래프들에서 알 수 있듯이, 절편의 시계열산점도는 고빈도 (high frequency) 잡음처럼 보이고, 기울기의 산점도는 저빈도 (low frequency) 성분을 갖는 것처럼 보인다. 즉, 관찰점들 사이에 자기상관성이 있는 것처럼 보인다. 올바르게 생성된 난수들이라면 자기상관성이 없어야 한다. 자기상관성을 제거하는 직관적인 방법은 데이터를 씌닝 (thinning out)하는 것이다. 즉, 일정한 주기로 관찰점을 선택하고 나머지 관찰점들은 버리는 것이다. 이렇게 씌닝된 관찰점들의 자기상관성이 전체 관찰점들 자기상관성보다 작다면, 씌닝의 목적이 달성된 것이다. 본저자는 다양한 주기의 씌닝을 해서 좋은 것을 선택하는 방법이 다른 어떤 고급통계기법을 사용하는 것보다 효율적이라고 생각한다. 이 MATLAB 프로그램 PoissonBayesianRegress102.m에서는 주기 10인 씌닝을 했다. 이 주기는 임의로 선택한 것이다. 씌닝된 관찰점들의 시계열산점도들이 그림 3.1.11의 하단에 그려져 있다. 이 시계열산점도들에서 알 수 있듯이, 씌닝된 관찰점들의 난수성이 씌닝되지 않은 관찰점들의 난수성보다 크다.

제4단계는 발생된 난수들을 통계기법을 사용해서 난수성검정 (randomness test), 즉 독립성검정 (independence test)을 하는 것이다. 대표적인 시계열데이터의 독립성검정법으로는 런검정법 (run test), Durbin-Watson 검정법, 스펙트럼검정법 등이 있다. 시계열데이터의 독립성검정법에 대해서는 최병선 [1, 제6장]을 참조하라. 이 MATLAB 프로그램 PoissonBayesianRegress102.m에서는 스펙트럼검정법을 사용한다. 스펙트럼검정을 하기 위해서는 우선 시계열데이터 trace에 FFT (fast Fourier transform)을 적용해서 traceFFT를 계산한다. 이 traceFFT의 노름제곱이 시계열데이터 trace의 피어리어도그램 (periodogram)이다. 피어리어도그램은 스펙트럼 (spectrum)의 추정량이다. 정상적 시계열데이터  $\{x_t\}$ 에

대해서 시차 (lag) 가  $k$  인 자기공분산 (autocovariance) 을 다음과 같이 정의한다.

$$\sigma(k) \doteq Cov(x_t, x_{t+k}) \quad (1)$$

이 자기공분산의 Fourier 변환을 스펙트럴함수 (spectral function) 또는 스펙트럼이라 부른다. 즉, 스펙트럼은 다음과 같이 정의된다.

$$S(\lambda) \doteq \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \sigma(k) \exp(ik\lambda), \quad (-\pi \leq \lambda \leq \pi) \quad (2)$$

따라서, 자기공분산은 스펙트럼의 역함수이다. 즉, 다음 식이 성립한다.

$$\sigma(k) = \int_{-\pi}^{\pi} S(\lambda) \exp(-ik\lambda), \quad (k = \dots, -1, 0, 1, \dots) \quad (3)$$

피어리오도그램은 Schuster (1897, 1906) [91], [92], Knott (1897) [60] 등이 일본 지진의 주기, 밀 (wheat) 가격의 주기, 태양흑점 (sunspot) 의 주기 등을 연구하는 과정에 제시한 것이다. 최근 경제학에서 피어리오도그램이 다시 사용되고 있는데, 이에 대해서는 다음 웹사이트를 참조하라.

<https://lectures.quantecon.org/j1/estspec.html>

이렇게 추정된 자기상관함수들이 그림 3.1.12에 그려져 있다. 그림 3.1.12의 상단 그래프는 절편의 자기상관함수추정값들을 그린 것이다. 이 그래프에서 알 수 있듯이, 각 자기상관함수 추정값이 95% 신뢰구간에 포함되어 있다. 그림 3.1.12의 하단 그래프는 기울기의 자기상관함수추정값들을 그린 것이다. 이 그래프에서 알 수 있듯이, 시차 3 이후 자기상관함수추정값이 95% 신뢰구간에 포함되어 있다. 즉, 슬라이스샘플러에 의해서 생성되고, 변인되고 씨닝된 관찰값들은 난수성을 갖는 것처럼 판단된다.

제5단계는 발생된 난수들의 3D 히스토그램을 그리는 것이다. 이 3D 히스토그램이 그림 3.1.13에 그려져 있다. 이 히스토그램이 그림 3.1.9에 그려진 사후확률밀도함수를 잘 반영하고 있음을 알 수 있다.

제6단계는 발생된 난수들을 사용해서 통계적 추론을 하는 것이다. 우선 절편의 히스토그램을 그림 3.1.14의 좌측상단 그래프에 출력하고, 커널스무딩밀도함수 (kernel smoothing density; ks density) 를 그림 3.1.14의 우측상단 그래프에 출력한다. 또한, 관찰점들의 개수에 대한 절편의 표본평균을 그림 3.1.14의 좌측하단 그래프에 출력하고, 관찰점들의 개수에 대한 기울기의 표본평균을 그림 3.1.14의 우측하단 그래프에 출력한다. 이 그래프들에서 알 수

있듯이 관찰점들의 개수가 100보다 크면, 표본평균들이 매우 안정적이다. 즉, 이 표본평균들이 사후확률분포의 평균들을 잘 나타냄을 알 수 있다. 따라서, 이 예제에서 사용된 관찰점들의 개수 1000은 MCMC를 사용해서 분석하기에 충분하다. 이 관찰점들을 사용해서 구한 절편의 추정값은  $\hat{\beta}_1 = -1.7107$  이고, 기울기의 추정값은  $\hat{\beta}_2 = 4.2739$  이다. ■

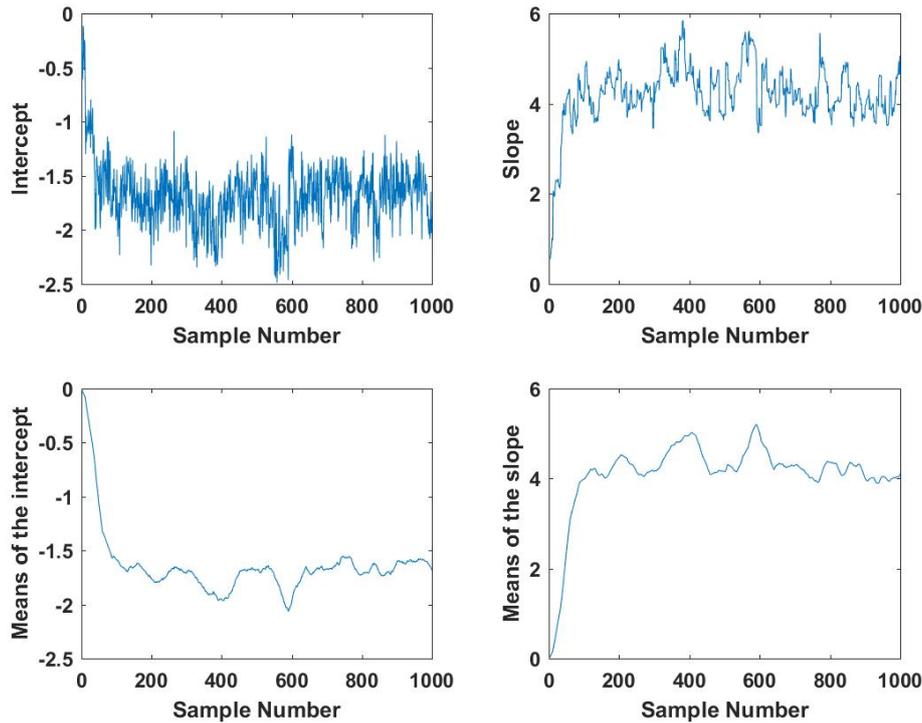


그림 3.1.10. 슬라이스샘플과 이동평균

### 제 3.2 절 계층모형과 베이저안통계분석

계층모형 (hierarchical model) 은 모수들에 계층적 구조를 부여하는 모형이다. 관측값들이 몇 개 그룹들로 나누어져 있으며 각 그룹 별로 설정한 모수가 있을 때, 이 모수들에 대해 공통적인 구조를 부여하는 모형이 대표적인 계층모형이다. 또한, 회귀모형에서 회귀계수벡터에 계층적 구조를 부여하는 계층모형을 생각할 수 있다.

계층모형에서는 모수벡터들  $\theta_1, \theta_2, \dots, \theta_p$  와 초모수벡터 (hyperparameter vector)  $\lambda$  가 주어졌을 때, 관찰값들  $y_1, y_2, \dots, y_p$  은 서로 독립이라고 가정한다. 따라서 이들의 우도함수는

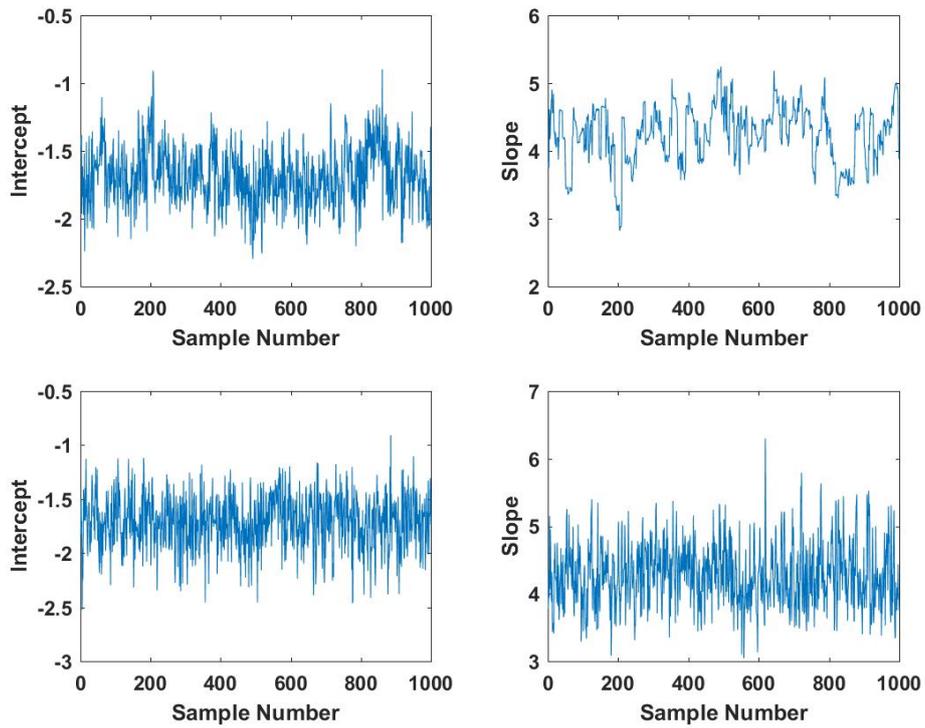


그림 3.1.11. 번인과 씨닝

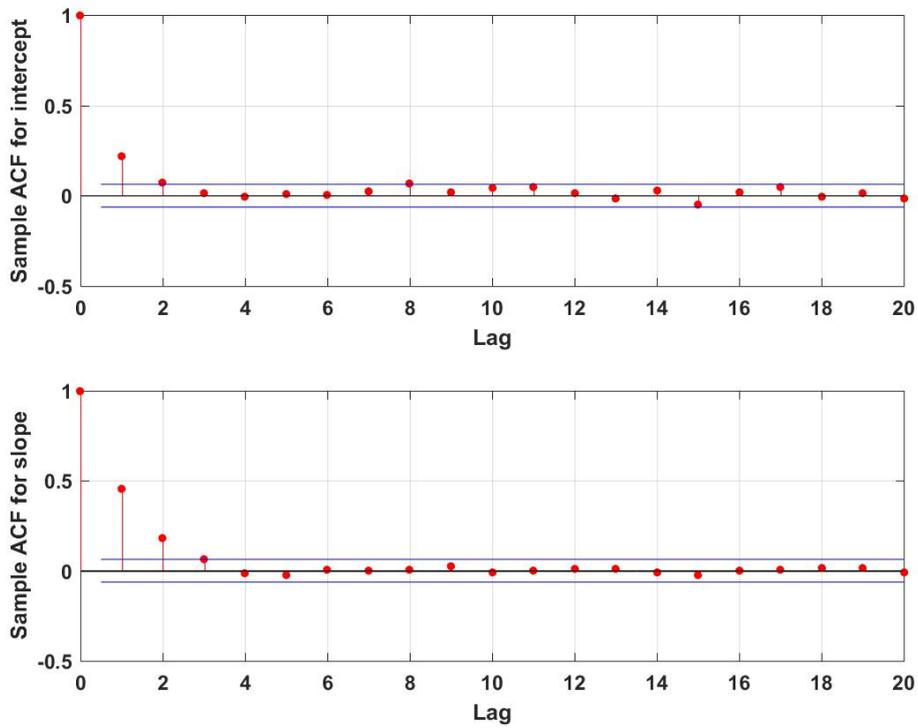


그림 3.1.12. 자기상관함수

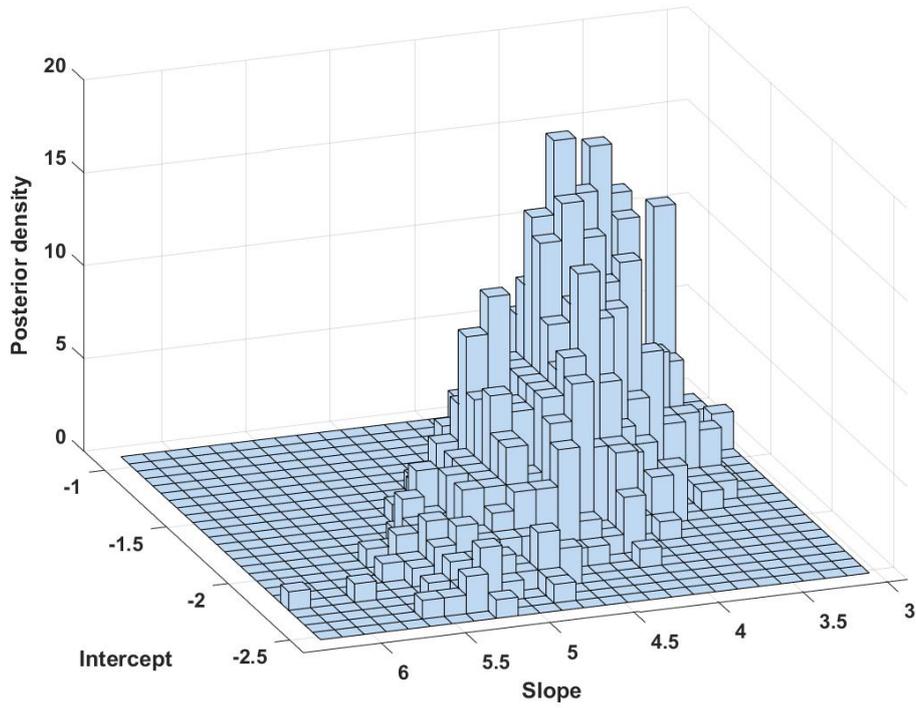


그림 3.1.13. 3D 히스토그램

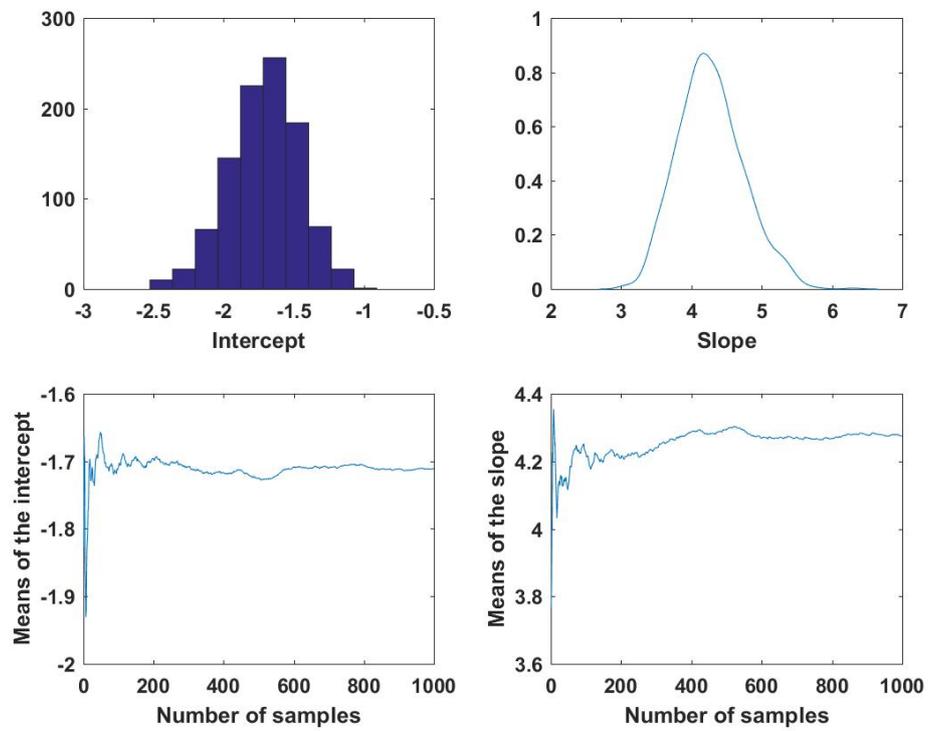


그림 3.1.14. 히스토그램과 표본평균

다음과 같다.

$$f(y_1, y_2, \dots, y_p | \theta_1, \theta_2, \dots, \theta_p, \lambda) = \prod_{i=1}^p f(y_i | \theta_i, \lambda) \quad (3.2.1)$$

또한, 초모수벡터  $\lambda$ 가 주어진 조건 하에서 모수벡터들  $\theta_1, \theta_2, \dots, \theta_p$ 의 사전확률밀도함수가 다음과 같다고 가정하자.

$$\pi(\theta_1, \theta_2, \dots, \theta_p | \lambda) = \prod_{i=1}^p \pi(\theta_i | \lambda) \quad (3.2.2)$$

즉, 초모수벡터  $\lambda$ 가 주어진 조건 하에서 모수벡터들  $\theta_1, \theta_2, \dots, \theta_p$ 는 서로 독립이다. 초모수벡터  $\lambda$ 의 사전확률분포가 다음과 같다고 하자.

$$\lambda \stackrel{d}{\sim} \int \pi(\lambda) d\lambda \quad (3.2.3)$$

즉, 초모수벡터  $\lambda$ 의 사전확률밀도함수는  $\pi(\lambda)$ 이다. 초모수벡터  $\lambda$ 의 사전확률분포를 초사전확률분포(hyper-prior probability distribution)라 부른다. 일반적으로 초사전확률분포로 무정보사전확률분포를 사용한다. 식 (3.2.1)~식 (3.2.3)으로 구성된 모형을 계층적 베이시안모형이라 부른다. 식 (3.2.2)와 식 (3.2.3)에서 알 수 있듯이, 다음 식이 성립한다.

$$\pi(\theta_1, \theta_2, \dots, \theta_p) = \int \prod_{i=1}^p \pi(\theta_i | \lambda) \pi(\lambda) d\lambda \quad (3.2.4)$$

식 (3.2.4)에서 알 수 있듯이, 초모수벡터  $\lambda$ 가 주어졌다는 조건이 없으면 일반적으로 모수벡터들  $\theta_1, \theta_2, \dots, \theta_p$ 는 서로 독립이 아니다.

초모수벡터  $\lambda$ 를 다루는 방법에 따라 계층모형분석을 두 가지로 나눌 수 있다. 첫째, 식 (3.2.3)에서처럼 초모수벡터  $\lambda$ 의 사전확률분포를 가정하고, 식 (3.2.4)를 이용해서 모수벡터들  $\theta_1, \theta_2, \dots, \theta_p$ 의 사후확률분포를 구하는 방법을 계층적 베이시안추정법이라 한다. 둘째, 관찰값들  $y_1, y_2, \dots, y_p$ 를 바탕으로 초모수벡터  $\lambda$ 를 추정된 후, 이를 사용해서 모수벡터들  $\theta_1, \theta_2, \dots, \theta_p$ 의 사후확률분포를 구하는 것을 경험적 베이시안추정법이라 한다. 점추정에 관한 한, 일반적으로 이 두 접근법들은 비슷한 결과를 보여준다. 그러나, 경험적 베이시안추정법에서는 초모수가 가지고 있는 불확실성을 무시하므로, 모수벡터 추정량의 표준오차를 과소추정(underestimate)하는 것으로 알려져 있다. 계층모형의 예제들을 살펴보자.

**예제 3.2.1** 다음 모형을 살펴보자.

$$y_{ij} = \mu_i + \epsilon_{ij}, \quad (i = 1, 2, \dots, p, j = 1, 2, \dots, J_i) \quad (1)$$

여기서  $\{\epsilon_{ij}\}$ 는 서로 독립이고 평균이 0이고 분산이  $\sigma^2$ 인 정규확률변수들이다. 이 모형에서 유의할 점은 제  $i$  번째 그룹에 속하는 관측값들은 동일한 모평균  $\mu_i$ 를 갖는다는 것이다.

만약 모평균들  $\{\mu_i | i = 1, 2, \dots, p\}$ 에 아무런 구조도 부과하지 않으면, 이 모형을 다음과 같이 Kronecker 델타함수  $\delta_{k,i}$ 를 더미변수로 사용한 선형회귀모형으로 표현할 수 있다.

$$y_{ij} = \sum_{k=1}^p \mu_k \delta_{k,i} + \epsilon_{ij}, \quad (i = 1, 2, \dots, p, j = 1, 2, \dots, J_i) \quad (2)$$

만약  $p$ 개 그룹들이 유사하다면,  $\{\mu_i | i = 1, 2, \dots, p\}$ 를 어떤 평균의 모집단에서 추출한  $p$ 개 관찰값들로 간주할 수 있다. 각 그룹의 평균이 다음과 같은 확률분포를 따른다고 가정하자.

$$\mu_i | (\mu, \tau^2) \stackrel{d}{\sim} \mathcal{N}(\mu, \tau^2), \quad (i = 1, 2, \dots, p) \quad (3)$$

또한,  $(\mu, \sigma^2, \tau^2)$ 에 대한 사전확률분포가 다음과 같다고 가정하자.

$$\mu \stackrel{d}{\sim} \mathcal{N}(\nu, \eta^2), \quad \sigma^2 \stackrel{d}{\sim} \pi(\sigma^2), \quad \tau^2 \stackrel{d}{\sim} \pi(\tau^2) \quad (4)$$

이 모형에서 사전확률분포를 두 계층으로 나눌 수 있다. 첫 번째 계층은 식 (3)에 의해 주어지고, 두 번째 계층은 식 (4)로 주어진다. 따라서, 모수들의 사전결합확률밀도함수는 다음과 같다.

$$\pi(\mu_1, \dots, \mu_p, \mu, \sigma^2, \tau^2) = \left[ \prod_{i=1}^p \pi(\mu_i | \mu, \tau^2) \right] \pi(\mu | \nu, \eta^2) \pi(\sigma^2) \pi(\tau^2) \quad (5)$$

■

**예제 3.2.2** 다음 조건들을 만족하는 계층모형을 사용해서, 정규확률분포의 평균, 분산, 그리고 공분산을 추정하기로 하자.

**첫째,**  $y_1 | (\mu_1, \sigma^2), y_2 | (\mu_2, \sigma^2), \dots, y_p | (\mu_p, \sigma^2)$ 는 서로 독립이고, 다음 식들을 만족한다

고 하자.

$$y_i | (\mu_i, \sigma^2) \stackrel{d}{\sim} \mathcal{N}(\mu_i, \sigma^2), \quad (i = 1, 2, \dots, p) \quad (1)$$

여기서  $\sigma^2 (> 0)$ 는 알려진 상수이다.

**둘째**,  $\mu_1 | \nu, \mu_2 | \nu, \dots, \mu_p | \nu$ 는 서로 독립이고, 다음 식들을 만족한다고 하자.

$$\mu_i | \nu \stackrel{d}{\sim} \mathcal{N}(\nu, \tau^2), \quad (i = 1, 2, \dots, p) \quad (2)$$

여기서  $\tau^2 (> 0)$ 는 알려진 상수이다.

**셋째**, 초모수  $\nu$ 는 다음 식을 만족한다고 하자.

$$\nu \stackrel{d}{\sim} \text{Uniform}(-\infty, \infty) \quad (3)$$

즉, 초모수  $\nu$ 는 무정보사전확률분포를 따른다.

우리의 목적은 관찰값들  $y_1, y_2, \dots, y_p$ 가 주어졌을 때, 모수들  $\mu_1, \mu_2, \dots, \mu_p$ 의 사후 확률분포를 구하는 것이다. 식 (3)에서 초모수  $\nu$ 의 사전확률분포가 주어졌으므로, 계층적 베이지안추정법을 적용하기로 하자. 다음 벡터들을 정의하자.

$$\mathbf{y} \doteq \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}, \quad \boldsymbol{\mu} \doteq \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}, \quad \mathbf{1}_p \doteq \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad (4)$$

모수벡터  $[\mu_1, \mu_2, \dots, \mu_p, \nu]$ 의 사후결합확률밀도함수는 다음과 같다.

$$\pi(\boldsymbol{\mu}, \nu | \mathbf{y}) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^p [y_i - \mu_i]^2\right) \exp\left(-\frac{1}{2\tau^2} \sum_{i=1}^p [\mu_i - \nu]^2\right) \quad (5)$$

다음 식이 성립한다.

$$\sum_{i=1}^p [\mu_i - \nu]^2 = \sum_{i=1}^p [\mu_i - \bar{\mu}]^2 + p[\nu - \bar{\mu}]^2 \quad (6)$$

여기서  $\bar{\mu} \doteq \frac{1}{p} \sum_{i=1}^p \mu_i$  이다. 식 (6)을 식 (5)에 대입한 다음  $\nu$ 에 대해서 적분하면, 다음 식을 얻는다.

$$\pi(\boldsymbol{\mu} | \mathbf{y}) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^p [y_i - \mu_i]^2\right) \exp\left(-\frac{1}{2\tau^2} \sum_{i=1}^p [\mu_i - \bar{\mu}]^2\right) \quad (7)$$

다음 식들이 성립한다.

$$\sum_{i=1}^p [y_i - \mu_i]^2 = \|\boldsymbol{\mu} - \mathbf{y}\|^2 = \boldsymbol{\mu}^T \boldsymbol{\mu} - 2\boldsymbol{\mu}^T \mathbf{y} + \mathbf{y}^T \mathbf{y} \quad (8)$$

$$\sum_{i=1}^p [\mu_i - \bar{\mu}]^2 = \|\boldsymbol{\mu} - \bar{\mu} \mathbf{1}_p\|^2 = \boldsymbol{\mu}^T \left[ I_p - \frac{1}{p} \mathbf{1}_p \mathbf{1}_p^T \right] \boldsymbol{\mu} \quad (9)$$

식 (8)과 식 (9)를 식 (7)에 대입하면, 다음 식을 얻는다.

$$\pi(\boldsymbol{\mu} | \mathbf{y}) \propto \exp\left(-\frac{1}{2} \left[ \boldsymbol{\mu}^T H \boldsymbol{\mu} - \frac{2}{\sigma^2} \boldsymbol{\mu}^T \mathbf{y} \right]\right) \quad (10)$$

여기서  $H$ 는 다음과 같다.

$$H \doteq \left[ \frac{1}{\sigma^2} + \frac{1}{\tau^2} \right] I_p - \frac{1}{\tau^2} \frac{1}{p} \mathbf{1}_p \mathbf{1}_p^T \quad (11)$$

식 (10)을 다음과 같이 쓸 수 있다.

$$\pi(\boldsymbol{\mu} | \mathbf{y}) \propto \exp\left(-\frac{1}{2} \left[ \boldsymbol{\mu} - \frac{1}{\sigma^2} H^{-1} \mathbf{y} \right]^T H \left[ \boldsymbol{\mu} - \frac{1}{\sigma^2} H^{-1} \mathbf{y} \right]\right) \quad (12)$$

따라서, 모수벡터  $\boldsymbol{\mu}$ 의 사후확률분포는 다음과 같다.

$$\boldsymbol{\mu} | \mathbf{y} \stackrel{d}{\sim} \mathcal{N}\left(\frac{1}{\sigma^2} H^{-1} \mathbf{y}, H^{-1}\right) \quad (13)$$

행렬  $H$ 를 다음과 같이 쓸 수 있다.

$$H = \left[ \frac{1}{\sigma^2} + \frac{1}{\tau^2} \right] \left[ I_p - \frac{\sigma^2}{\sigma^2 + \tau^2} \frac{1}{p} \mathbf{1}_p \mathbf{1}_p^T \right] = \frac{1}{\sigma^2[1 - \alpha]} \left[ I_p - \alpha \frac{1}{p} \mathbf{1}_p \mathbf{1}_p^T \right] \quad (14)$$

여기서  $\alpha$ 는 다음과 같다.

$$\alpha \doteq \frac{\sigma^2}{\sigma^2 + \tau^2} \quad (15)$$

다음 등식이 성립함을 쉽게 알 수 있다.

$$[B + \mathbf{u}\mathbf{v}^T]^{-1} = B^{-1} - \frac{1}{1 + \mathbf{v}^T B^{-1} \mathbf{u}} B^{-1} \mathbf{u}\mathbf{v}^T B^{-1} \quad (16)$$

식 (16)을 식 (14)에 적용하면, 행렬  $H$ 의 역행렬이 다음과 같음을 알 수 있다.

$$H^{-1} = \sigma^2 [1 - \alpha] \left\{ I_p - \mathbf{1}_p \left[ -\frac{p}{\alpha} + \mathbf{1}_p^T \mathbf{1}_p \right]^{-1} \mathbf{1}_p^T \right\} = \sigma^2 \left\{ [1 - \alpha] I_p + \frac{\alpha}{p} \mathbf{1}_p \mathbf{1}_p^T \right\} \quad (17)$$

식 (17)을 식 (13)에 대입하면, 다음 식을 얻는다.

$$\boldsymbol{\mu} | \mathbf{y} \stackrel{d}{\sim} \mathcal{N} \left( [1 - \alpha] \mathbf{y} + \alpha \bar{y} \mathbf{1}_p, \sigma^2 \left\{ [1 - \alpha] I_p + \frac{\alpha}{p} \mathbf{1}_p \mathbf{1}_p^T \right\} \right) \quad (18)$$

여기서  $\bar{y} \doteq \frac{1}{p} \sum_{i=1}^p y_i$ 이다.

식 (18)에서 알 수 있듯이, 계층적 베이esian추정량들은 다음과 같다.

$$\hat{\boldsymbol{\mu}}^{HB} = E(\boldsymbol{\mu} | \mathbf{y}) = [1 - \alpha] \mathbf{y} + \alpha \bar{y} \mathbf{1}_p \quad (19)$$

$$\hat{\mu}_i^{HB} = E(\mu_i | \mathbf{y}) = [1 - \alpha] y_i + \alpha \bar{y}, \quad (i = 1, 2, \dots, p) \quad (20)$$

$$Var^{HB}(\mu_i | \mathbf{y}) = \sigma^2 \left[ 1 - \alpha + \frac{\alpha}{p} \right], \quad (i = 1, 2, \dots, p) \quad (21)$$

$$Cov^{HB}(\mu_i, \mu_j | \mathbf{y}) = \frac{\alpha}{p} \sigma^2, \quad (i \neq j) \quad (22)$$

식 (18)에 기술한 사후확률분포를 그리기 위해서, 다음 R프로그램 Hierarchical101.R을 실행해 보자.

```

1 # -----
2 # Filename: Hierarchical101.R
3 # Hierarchical Bayes Estimate
4 # Programmed by CBS
5 # -----
6 setEPS()
7 # Prior parameters
8 ( p <- 2 )
9 ( sigma0 <- 3 )
10 ( tau0 <- 2 ); ( nu0 <- 3 ) # nu0 from Uniform(-inf,inf)

```

```

11 |
12 | # Likelihood
13 | y <- matrix(0,p,1)
14 | prior.mu <- matrix(0,p,1)
15 | set.seed(3)
16 | for (ii in 1:p){
17 |   prior.mu[ii] <- nu0 + tau0*rnorm(1,0,1)
18 |   y[ii] <- prior.mu[ii]+sigma0*rnorm(1,0,1)
19 | }
20 | y
21 | ( ybar <- mean(y) )
22 |
23 | # Posterior
24 | ( A = sigma0^2/(sigma0^2 + tau0^2) )
25 | onep = matrix(1,p)
26 | Ip = diag(p)
27 | ( H <- 1/(sigma0^2*(1-A))*(Ip - A/p*onep%*%t(onep)) )
28 | ( invH <- sigma0^2*((1-A)*Ip + A/p*onep%*%t(onep)) )
29 | ( posterior.mu <- 1/sigma0^2*invH%*%y )
30 |   # posterior.mu <- (1-A)*y+A*ybar*onep
31 | ( posterior.Tau <- invH )
32 |   # posterior.Tau <- sigma0^2*((1-A)*Ip+A/p*onep%*%t(onep))
33 |
34 | # Monte Carlo Simulations and Contour Plotting of Posterior
35 | library(MASS)
36 | library(ggplot2)
37 | # install.packages("gridExtra")
38 | library(gridExtra)
39 |
40 | NoSim <- 512
41 | mu.Sim <- mvrnorm(n=NoSim,posterior.mu,posterior.Tau)
42 | # Plotting
43 | plot.new()
44 | postscript('Hierarchical101.eps') # Start to save figure
45 | par(mfrow=c(2,2))
46 | plot1 <- qplot(mu.Sim[,1],mu.Sim[,2]) + geom_point()
47 | plot2 <- qplot(mu.Sim[,1]) +
48 |   geom_histogram(bins=30,colour="black",fill="white")
49 | plot3 <- qplot(mu.Sim[,2],bins=30,fill=I("blue"),col=I("red"))
50 | plot4 <- qplot(mu.Sim[,1],mu.Sim[,2]) + geom_point() +
51 |   stat_ellipse(linetype=1,col="red",level=0.99,lwd=1.5) +
52 |   stat_ellipse(linetype=2,col="green",level=0.95,lwd=1.5) +
53 |   stat_ellipse(linetype=3,col="blue",level=0.90,lwd=1.5)
54 | grid.arrange(plot1,plot2,plot3,plot4,ncol=2)
55 | dev.off() # End to save figure
56 | # -----

```

이 R프로그램 Hierarchical101.R에서는 차수가  $p = 2$ 인 경우를 다룬다. 가정에 의해서  $\nu$ 는 무정보사전확률분포를 따르므로, 이 확률분포에서 실제로  $\nu$ 값을 생성할 수는 없다. 따라서,  $\nu = 3$ 을 할당하였다. 또한,  $\mu_i$ 의 표준편차로  $\tau = 2$ 를, 그리고 관찰값의 표준편차로  $\sigma = 3$ 를

할당하였다. 즉, 사전확률분포와 관찰값의 확률분포는 각각 다음과 같다.

$$\mu_i \stackrel{d}{\sim} \mathcal{N}(3, 2^2), \quad (i = 1, 2) \tag{23}$$

$$y_i | \mu_i \stackrel{d}{\sim} \mathcal{N}(\mu_i, 3^2), \quad (i = 1, 2) \tag{24}$$

식 (19)에서 알 수 있듯이,  $\boldsymbol{\mu}$ 의 계층적 베이지안기법에 의한 사후확률분포는 다음과 같다.

$$\boldsymbol{\mu} | \mathbf{y} \stackrel{d}{\sim} \mathcal{N} \left( [1 - \alpha] \begin{bmatrix} 0.19855600 \\ 0.06118077 \end{bmatrix} + \alpha \bar{y} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, 3^2 \left\{ [1 - \alpha] I_2 + \frac{\alpha}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right\} \right) \tag{25}$$

여기서  $\alpha = 0.6923077$ 이고  $\bar{y} = 0.1298684$ 이다.

이 R프로그램 Hierarchical101.R을 실행하면, 그림 3.2.1이 그려진다. 그림 3.2.1은 식 (25)의 사후확률분포에서 512개 관찰점들을 생성하고, 이들을 바탕으로 그린 것이다. 그림 3.2.1의 좌측상단 그래프는 이 관찰점들의 산점도를 그린 것이고, 우측상단 그래프는  $\mu_1 | \mathbf{y}$ 의 히스토그램을 그린 것이고, 좌측하단 그래프는  $\mu_2 | \mathbf{y}$ 의 히스토그램을 그린 것이다. 또한, 우측하단 그래프는 관찰점들의 산점도에 신뢰영역을 나타내는 등가곡선을 추가한 것이다. 이 그래프에서 적색 실선으로 그려진 타원은 99% 신뢰영역을, 녹색 긴점선으로 그려진 타원은 95% 신뢰영역을 나타내며, 청색 점선으로 그려진 타원은 90% 신뢰영역을 나타낸다. ■

**예제 3.2.3** 경험적 베이지안추정법을 사용해서 예제 3.2.2의 계층모형을 추정하기로 하자. 즉, 다음과 같은 모형을 추정하기로 하자.

**첫째,** 조건부확률변수들  $y_1 | (\mu_1, \sigma^2), y_2 | (\mu_2, \sigma^2), \dots, y_p | (\mu_p, \sigma^2)$ 는 서로 독립이고, 다음 식들을 만족한다고 하자.

$$y_i | (\mu_i, \sigma^2) \stackrel{d}{\sim} \mathcal{N}(\mu_i, \sigma^2), \quad (i = 1, 2, \dots, p) \tag{1}$$

여기서  $\sigma^2 (> 0)$ 는 알려진 상수이다.

**둘째,** 조건부 모수들  $\mu_1 | \nu, \mu_2 | \nu, \dots, \mu_p | \nu$ 는 서로 독립이고, 다음 식들을 만족한다고 하자.

$$\mu_i | \nu \stackrel{d}{\sim} \mathcal{N}(\nu, \tau^2), \quad (i = 1, 2, \dots, p) \tag{2}$$

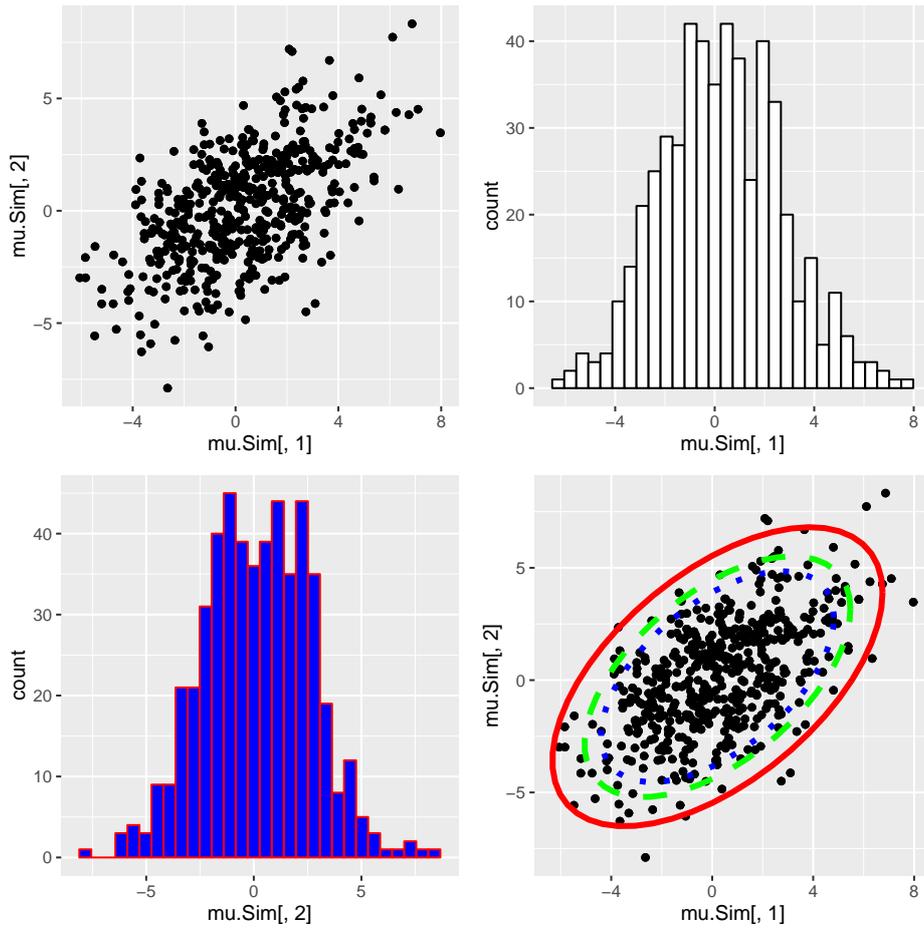


그림 3.2.1. 계층적 베이زي안분석

여기서  $\tau^2 (> 0)$ 는 알려진 상수이다.

우리의 목표는 관찰값들  $y_1, y_2, \dots, y_p$ 가 주어졌을 때, 모수들  $\mu_1, \mu_2, \dots, \mu_p$ 의 사후확률분포를 구하는 것이다. 이 모형에서는 초모수  $\nu$ 의 사전확률분포가 주어지지 않았으므로, 관찰값들  $y_1, y_2, \dots, y_p$ 로부터 초모수  $\nu$ 를 추정하기로 하자. 여기서도 예제 3.2.2의 표기법을 사용하기로 하자.

예제 3.2.2의 식 (18)을 유도하는 방법을 사용하면, 다음 식을 유도할 수 있다.

$$\boldsymbol{\mu} \mid \mathbf{y} \stackrel{d}{\sim} \mathcal{N}([1 - \alpha]\mathbf{y} + \alpha\nu\mathbf{1}_p, \sigma^2[1 - \alpha]I_p) \tag{3}$$

식 (3)에서 알 수 있듯이, 모수벡터  $\boldsymbol{\mu}$ 의 베이زي안추정량은 다음과 같다.

$$\hat{\boldsymbol{\mu}}^B = [1 - \alpha]\mathbf{y} + \alpha\nu\mathbf{1}_p \tag{4}$$

지금부터는 초모수  $\nu$ 를 추정하기로 하자. 모수들  $\tau$ 와  $\sigma^2$ 는 주어진 상수들이므로 조건부확률변수  $y_i | (\nu, \tau, \sigma^2)$ 를  $y_i | \nu$ 로 표기하자. 식 (1)과 식 (2)에서 알 수 있듯이,  $y_1 | \nu, y_2 | \nu, \dots, y_n | \nu$ 는 서로 독립이고, 다음 식들을 만족한다.

$$y_i | \nu \stackrel{d}{\sim} \mathcal{N}(\nu, \sigma^2 + \tau^2), \quad (i = 1, 2, \dots, p) \quad (5)$$

따라서,  $\bar{y}$ 를 초모수  $\nu$ 의 추정량으로 사용하자. 식 (3)에 이 초모수추정량을 대입하면, 다음 식을 얻는다.

$$\boldsymbol{\mu} | \mathbf{y} \stackrel{d}{\sim} \mathcal{N}([1 - \alpha]\mathbf{y} + \alpha\bar{y}\mathbf{1}_p, \sigma^2[1 - \alpha]I_p) \quad (6)$$

식 (6)에서 알 수 있듯이, 경험적 베이지안추정량들은 다음과 같다.

$$\hat{\boldsymbol{\mu}}^{EB} = E(\boldsymbol{\mu} | \mathbf{y}) = [1 - \alpha]\mathbf{y} + \alpha\bar{y}\mathbf{1}_p \quad (7)$$

$$\hat{\mu}_i^{EB} = E(\mu_i | \mathbf{y}) = [1 - \alpha]y_i + \alpha\bar{y}, \quad (i = 1, 2, \dots, p) \quad (8)$$

$$Var^{EB}(\mu_i | \mathbf{y}) = \sigma^2[1 - \alpha], \quad (i = 1, 2, \dots, p) \quad (9)$$

$$Cov^{EB}(\mu_i, \mu_j | \mathbf{y}) = 0, \quad (i \neq j) \quad (10)$$

식 (8)과 예제 3.2.2의 식 (20)에서 알 수 있듯이, 사후평균의 계층적 베이지안추정량과 경험적 베이지안추정량은 같다. 그러나, 사후분산의 경험적 베이지안추정량  $\sigma^2[1 - \alpha]$ 는 계층적 베이지안추정량  $\sigma^2[1 - \alpha + \alpha/p]$ 보다 작다. 즉, 사후분산의 경험적 베이지안추정량은 과소추정된(underestimated) 것이다. 이에 대해서 좀 더 자세히 살펴보자. 사후분산의 계층적 베이지안모형을 다음과 같이 쓸 수 있다.

$$Var^{HB}(\boldsymbol{\mu} | \mathbf{y}) = E_\nu(Var^{HB}(\boldsymbol{\mu} | \mathbf{y}, \nu)) + Var_\nu(E^{HB}(\boldsymbol{\mu} | \mathbf{y}, \nu)) \quad (11)$$

여기서  $E_\nu(\cdot)$ 와  $Var_\nu(\cdot)$ 는 각각  $\nu$ 의 확률분포에 대한 기대값과 분산을 의미한다. 식 (3)과 식 (4)에서 알 수 있듯이, 다음 식들이 성립한다.

$$E(Var^{HB}(\boldsymbol{\mu} | \mathbf{y}, \nu)) = E(\sigma^2[1 - \alpha]I_p | \mathbf{y}) = \sigma^2[1 - \alpha]I_p \quad (12)$$

$$Var_\nu(E^{HB}(\boldsymbol{\mu} | \mathbf{y}, \nu)) = Var([1 - \alpha]\mathbf{y} + \alpha\nu\mathbf{1}_p | \mathbf{y}) = \alpha^2 Var(\nu | \mathbf{y})\mathbf{1}_p\mathbf{1}_p^T \quad (13)$$

또한, 예제 3.2.2의 식 (2)와 식 (3)에서 알 수 있듯이, 계층적 베이저안모형에서 초모수  $\nu$ 의 사후확률분포는 다음과 같다.

$$\nu | \mathbf{y} \stackrel{d}{\sim} \mathcal{N} \left( \bar{y}, \frac{\sigma^2}{p\alpha} \right) \quad (14)$$

식 (14)를 유도하는 자세한 과정은 예제 3.2.4의 식 (11)을 유도하는 과정을 참조하라. 식 (12)~식 (14)를 식 (11)에 대입하면, 다음 식을 얻는다.

$$Var^{HB}(\boldsymbol{\mu} | \mathbf{y}) = \sigma^2 [1 - \alpha] I_p + \frac{\alpha \sigma^2}{p} \mathbf{1}_p \mathbf{1}_p^T \quad (15)$$

식 (6)과 식 (15)에서 알 수 있듯이, 다음 식이 성립한다.

$$Var^{HB}(\boldsymbol{\mu} | \mathbf{y}) = Var^{EB}(\boldsymbol{\mu} | \mathbf{y}) + \frac{\alpha \sigma^2}{p} \mathbf{1}_p \mathbf{1}_p^T \quad (16)$$

식 (16)의 우변의 두 번째 항은 초모수  $\nu$ 의 추정에서 발생하는 불확실성을 나타낸다. 즉, 계층적 베이저안추정법에서는 초모수  $\nu$ 의 사전확률분포를 가정하므로, 초모수  $\nu$ 의 추정에서 발생하는 불확실성을 모형에 반영한다.

식 (6)과 식 (14)에 기술한 사후확률분포를 그리기 위해서, 다음 R프로그램 Hierarchical102.R을 실행해 보자.

```

1 # -----
2 # Filename: Hierarchical102.R
3 # Empirical Bayes Estimate
4 # Programmed by CBS
5 # -----
6 setEPS()
7 # Prior parameters
8 ( p <- 2 )
9 ( sigma0 <- 3 )
10 ( tau0 <- 2 ); ( nu0 <- 3 ) # nu0 from Uniform(-inf,inf)
11
12 # Likelihood
13 y <- matrix(0,p,1)
14 prior.mu <- matrix(0,p,1)
15 set.seed(3)
16 for (ii in 1:p){
17     std0 <- sqrt(tau0^2 + sigma0^2)
18     y[ii] <- nu0+std0*rnorm(1,0,1)
19 }
20 y
21 ( ybar <- mean(y) )
22
23 # Posterior
24 ( A = sigma0^2/(sigma0^2 + tau0^2) )
25 onep = matrix(1,p)

```

```

26 Ip = diag(p)
27 ( H <- 1/(sigma0^2*(1-A))*(Ip - A/p*onep%*%t(onep)) )
28 ( invH <- sigma0^2*((1-A)*Ip + A/p*onep%*%t(onep)) )
29 ( posterior.mu <- 1/sigma0^2*invH%*%y )
30 # posterior.mu <- (1-A)*y+A*ybar*onep
31 ( posterior.Tau <- invH )
32 # posterior.Tau <- sigma0^2*((1-A)*Ip+A/p*onep%*%t(onep))
33
34 # Monte Carlo Simulations and Contour Plotting of Posterior
35 library(MASS)
36 library(ggplot2)
37 # install.packages("gridExtra")
38 library(gridExtra)
39
40 NoSim <- 512
41 mu.Sim <- mvrnorm(n=NoSim,posterior.mu,posterior.Tau)
42 # Plotting
43 plot.new()
44 postscript('Hierarchical102.eps') # Start to save figure
45 par(mfrow=c(2,2))
46 plot1 <- qplot(mu.Sim[,1],mu.Sim[,2]) + geom_point() +
47   stat_ellipse(linetype=1,col="red",level=0.99,lwd=1.5) +
48   stat_ellipse(linetype=2,col="green",level=0.95,lwd=1.5) +
49   stat_ellipse(linetype=3,col="blue",level=0.90,lwd=1.5)
50 plot2 <- qplot(mu.Sim[,1]) +
51   geom_histogram(bins=30,colour="black",fill="yellow")
52 plot3 <- qplot(mu.Sim[,2]) +
53   geom_histogram(bins=30,colour="black",fill="pink")
54 plot4 <- qplot(nu.Sim,bins=30,fill=I("blue"),col=I("red"))
55 grid.arrange(plot1,plot2,plot3,plot4,ncol=2)
56 dev.off() # End to save figure
57 # -----

```

이 R프로그램 Hierarchical102.R에서는 차수가  $p = 2$ 인 경우를 다룬다. 식 (5)에서 알 수 있듯이, 관찰값들을 발생시키기 위해서는  $\nu$ 값이 주어져야 한다. 이 R프로그램 Hierarchical102.R에서도 예제 3.2.2에서와 마찬가지로  $\nu = 3$ 을 할당하였다. 또한,  $\mu_i$ 의 표준편차로  $\tau = 2$ 를, 그리고 관찰값의 표준편차로  $\sigma = 3$ 를 할당하였다. 따라서, 사전확률분포와 관찰값의 확률분포는 각각 다음과 같다.

$$\mu_i \stackrel{d}{\sim} \mathcal{N}(3, 2^2), \quad (i = 1, 2) \quad (17)$$

$$y_i \stackrel{d}{\sim} \mathcal{N}(3, 3^2 + 2^2), \quad (i = 1, 2) \quad (18)$$

식 (6)에서 알 수 있듯이,  $\boldsymbol{\mu}$ 의 계층적 베이시안기법에 의한 사후확률분포는 다음과 같다.

$$\boldsymbol{\mu} | \mathbf{y} \stackrel{d}{\sim} \mathcal{N} \left( [1 - \alpha] \begin{bmatrix} 0.19855600 \\ 0.06118077 \end{bmatrix} + \alpha \bar{y} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, 3^2 [1 - \alpha] I_2 \right) \quad (19)$$

여기서  $\alpha = 0.6923077$ 이고  $\bar{y} = 0.1298684$ 이다. 식 (14)에서 알 수 있듯이,  $\nu$ 의 계층적

베이지안기법에 의한 사후확률분포는 다음과 같다.

$$\nu | \mathbf{y} \stackrel{d}{\sim} \mathcal{N} \left( 0.1298684, \frac{1}{2\alpha} 3^2 \right) \quad (20)$$

이 R프로그램 Hierarchical102.R을 실행하면, 그림 3.2.2가 그려진다. 이 그림은 식 (19)와 식 (20)의 사후확률분포들에서 512개의 관찰점들을 생성하고, 이들을 바탕으로 그린 것이다. 그림 3.2.2의 좌측상단 그래프는 이 관찰점들의 산점도와 신뢰영역을 나타내는 등가곡선을 그린 것이다. 이 그래프에서 적색 실선으로 그려진 타원은 99% 신뢰영역을, 녹색 긴점선으로 그려진 타원은 95% 신뢰영역을 나타내며, 청색 점선으로 그려진 타원은 90% 신뢰영역을 나타낸다. 우측상단 그래프는  $\mu_1 | \mathbf{y}$ 의 히스토그램을 그린 것이고, 좌측하단 그래프는  $\mu_2 | \mathbf{y}$ 의 히스토그램을 그린 것이다. 또한, 우측하단 그래프는  $\nu | \mathbf{y}$ 의 히스토그램을 그린 것이다. ■

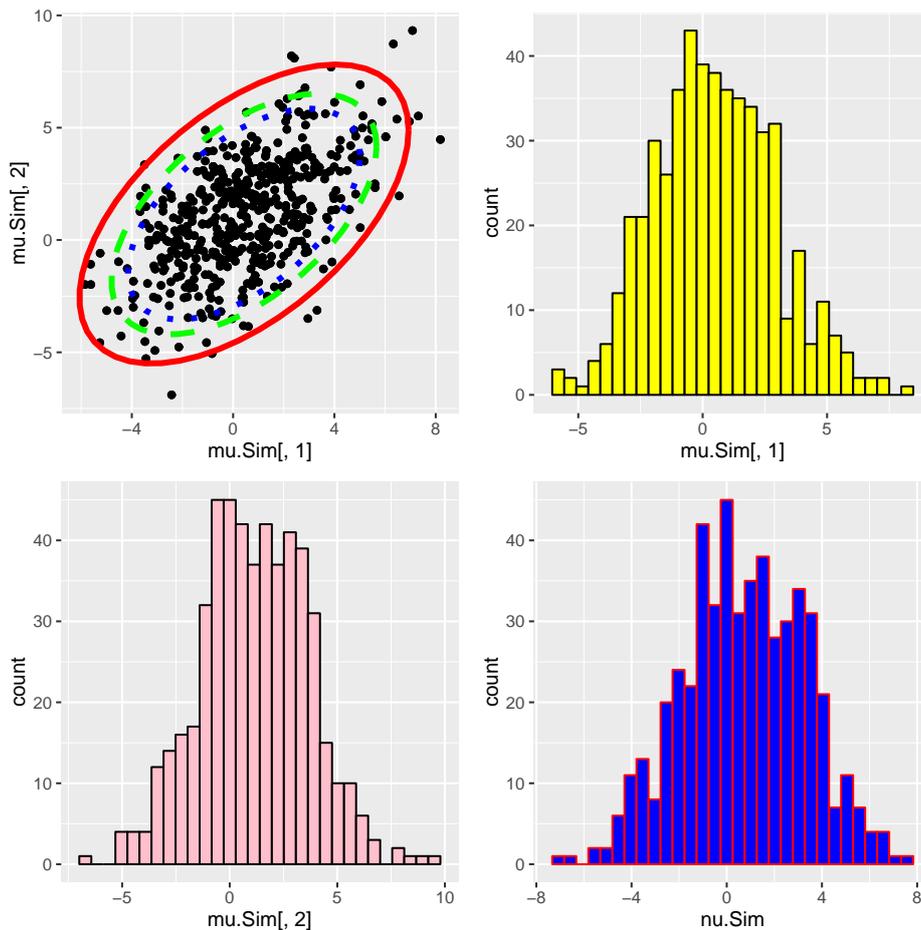


그림 3.2.2. 계층적 베이지안분석

**예제 3.2.4** 예제 3.2.2의 계층모형에서는 분산  $\tau$ 를 사전에 알고있는 상수라고 가정했다.

이 예제에서는  $\tau$ 역시 확률변수라고 가정하는 계층모형을 살펴보자. 이 계층모형은 다음과 같다.

**첫째,** 조건부확률변수들  $y_1 | (\mu_1, \nu, \tau^2), y_2 | (\mu_2, \nu, \tau^2), \dots, y_p | (\mu_p, \nu, \tau^2)$ 는 서로 독립이고, 다음 식들을 만족한다고 하자.

$$y_i | (\mu_i, \nu, \tau^2) \stackrel{d}{\sim} \mathcal{N}(\mu_i, \sigma^2), \quad (i = 1, 2, \dots, p) \quad (1)$$

여기서  $\sigma^2 (> 0)$ 는 알려진 상수이다.

**둘째,** 조건부 모수들  $\mu_1 | (\nu, \tau^2), \mu_2 | (\nu, \tau^2), \dots, \mu_p | (\nu, \tau^2)$ 는 서로 독립이고, 다음 식들을 만족한다고 하자.

$$\mu_i | (\nu, \tau^2) \stackrel{d}{\sim} \mathcal{N}(\nu, \tau^2), \quad (i = 1, 2, \dots, p) \quad (2)$$

**셋째,** 초모수  $\nu$ 와 분산  $\tau^2$ 의 사전결합확률밀도함수는 다음과 같다고 가정하자.

$$\pi(\nu, \tau^2) \propto 1 \quad (3)$$

이 예제에서는 예제 3.2.2의 표기법을 사용하기로 하자. 예제 3.2.2의 식 (18)을 구하는 방법을 적용하면,  $(\mathbf{y}, \nu, \tau^2)$ 가 주어졌다는 조건 하에 모수벡터  $\boldsymbol{\mu}$ 의 사후확률분포가 다음과 같음을 알 수 있다.

$$\boldsymbol{\mu} | (\mathbf{y}, \nu, \tau^2) \stackrel{d}{\sim} \mathcal{N}([1 - \alpha]\mathbf{y} + \alpha\nu\mathbf{1}_p, \sigma^2[1 - \alpha]I_p) \quad (4)$$

여기서  $\alpha = \sigma^2 / [\sigma^2 + \tau^2]$ 이다. 식 (1)과 식 (2)에서 알 수 있듯이, 다음 식이 성립한다.

$$\mathbf{y} | (\nu, \tau^2) \stackrel{d}{\sim} \mathcal{N}\left(\nu\mathbf{1}_p, \frac{\sigma^2}{\alpha}I_p\right) \quad (5)$$

분산  $\sigma^2$ 는 상수이므로  $\alpha = \sigma^2 / [\sigma^2 + \tau^2]$ 는  $\tau^2$ 의 함수이다. 이 성질과 식 (5)에서 알 수 있듯이,  $(\nu, \alpha)$ 의 사후결합확률밀도함수는 다음과 같다.

$$\pi(\nu, \alpha | \mathbf{y}) \propto \alpha^{p/2} \exp\left(-\frac{\alpha}{2\sigma^2} \|\mathbf{y} - \nu\mathbf{1}_p\|^2\right) \frac{1}{\alpha^2} \quad (6)$$

여기서 우변의  $1/\alpha^2$ 는 Jacobian이다. 식 (6)을 다음과 같이 쓸 수 있다.

$$\pi(\nu, \alpha | \mathbf{y}) \propto \alpha^{[p-4]/2} \exp\left(-\frac{\alpha}{2\sigma^2} \|\mathbf{y} - \nu \mathbf{1}_p\|^2\right) \quad (7)$$

다음 식들이 성립한다.

$$\begin{aligned} E(\boldsymbol{\mu} | \mathbf{y}) &= E(E(\boldsymbol{\mu} | (\mathbf{y}, \nu, \tau^2) | \mathbf{y})) = E([1 - \alpha]\mathbf{y} + \alpha\nu \mathbf{1}_p | \mathbf{y}) \\ &= [1 - E(\alpha | \mathbf{y})]\mathbf{y} + E(\alpha \mathbf{1}_p E(\nu | \alpha, \mathbf{y}) | \mathbf{y}) \end{aligned} \quad (8)$$

여기서 첫 번째와 세 번째 등호들은 반복조건부기대값법칙에 의해서, 그리고 두 번째 등호는 식 (4)에 의해서 성립한다. 다음 식이 성립한다.

$$\pi(\nu | \alpha, \mathbf{y}) \propto \pi(\nu, \alpha | \mathbf{y}) \quad (9)$$

식 (7)에서 알 수 있듯이, 다음 식이 성립한다.

$$\pi(\nu, \alpha | \mathbf{y}) \propto \alpha^{[p-4]/2} \exp\left(-\frac{\alpha}{2\sigma^2} \left\{ \sum_{i=1}^p [y_i - \bar{y}]^2 + p[\bar{y} - \nu]^2 \right\}\right) \quad (10)$$

식 (9)와 식 (10)에서 알 수 있듯이,  $(\alpha, \mathbf{y})$ 가 주어진 조건 하에서 초모수  $\nu$ 의 사후확률분포는 다음과 같다.

$$\nu | (\alpha, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}\left(\bar{y}, \frac{\sigma^2}{p\alpha}\right) \quad (11)$$

식 (11)을 식 (8)에 대입하면, 다음 식을 얻는다.

$$E(\boldsymbol{\mu} | \mathbf{y}) = [1 - E(\alpha | \mathbf{y})]\mathbf{y} + E(\alpha | \mathbf{y}) \bar{y} \mathbf{1}_p \quad (12)$$

다음 식이 성립한다.

$$Var(\boldsymbol{\mu} | \mathbf{y}) = E(Var(\boldsymbol{\mu} | \mathbf{y}, \alpha, \nu)) + Var_\nu(E(\boldsymbol{\mu} | \mathbf{y}, \alpha, \nu)) \quad (13)$$

식 (4)와 식 (5)에서 알 수 있듯이, 다음 식들이 성립한다.

$$E(\text{Var}(\boldsymbol{\mu} | \mathbf{y}, \alpha, \nu)) = E(\sigma^2 [1 - \alpha] I_p | \mathbf{y}) = \sigma^2 [1 - E(\alpha | \mathbf{y})] I_p \quad (14)$$

$$\text{Var}_\nu(E(\boldsymbol{\mu} | \mathbf{y}, \alpha, \nu)) = \text{Var}([1 - \alpha]\mathbf{y} + \alpha\nu\mathbf{1}_p | \mathbf{y}) = \text{Var}(\alpha[\mathbf{y} - \nu\mathbf{1}_p] | \mathbf{y}) \quad (15)$$

식 (14)와 식 (15)를 식 (13)에 대입하면, 다음 식을 얻는다.

$$\text{Var}(\boldsymbol{\mu} | \mathbf{y}) = \sigma^2 [1 - E(\alpha | \mathbf{y})] I_p + \text{Var}(\alpha[\mathbf{y} - \nu\mathbf{1}_p] | \mathbf{y}) \quad (16)$$

다음 식이 성립한다.

$$\begin{aligned} & \text{Var}(\alpha[\mathbf{y} - \nu\mathbf{1}_p] | \mathbf{y}) \\ &= E(\text{Var}(\alpha[\mathbf{y} - \nu\mathbf{1}_p] | (\alpha, \mathbf{y}) | \mathbf{y})) + \text{Var}(E(\alpha[\mathbf{y} - \nu\mathbf{1}_p] | (\alpha, \mathbf{y}) | \mathbf{y})) \end{aligned} \quad (17)$$

식 (11)에서 알 수 있듯이, 다음 식들이 성립한다.

$$E(\text{Var}(\alpha[\mathbf{y} - \nu\mathbf{1}_p] | (\alpha, \mathbf{y}) | \mathbf{y})) = E\left(\alpha^2 \frac{\sigma^2}{\alpha p} \mathbf{1}_p \mathbf{1}_p^T \middle| \mathbf{y}\right) = E(\alpha | \mathbf{y}) \frac{\sigma^2}{p} \mathbf{1}_p \mathbf{1}_p^T \quad (18)$$

$$\text{Var}(E(\alpha[\mathbf{y} - \nu\mathbf{1}_p] | (\alpha, \mathbf{y}) | \mathbf{y})) = \text{Var}(\alpha | \mathbf{y}) [\mathbf{y} - \bar{y}\mathbf{1}_p] [\mathbf{y} - \bar{y}\mathbf{1}_p]^T \quad (19)$$

식 (18)과 식 (19)를 식 (17)에 대입하면, 다음 식을 얻는다.

$$\text{Var}(\alpha[\mathbf{y} - \nu\mathbf{1}_p] | \mathbf{y}) = E(\alpha | \mathbf{y}) \sigma^2 \frac{1}{p} \mathbf{1}_p \mathbf{1}_p^T + \text{Var}(\alpha | \mathbf{y}) [\mathbf{y} - \bar{y}\mathbf{1}_p] [\mathbf{y} - \bar{y}\mathbf{1}_p]^T \quad (20)$$

식 (20)을 식 (16)에 대입하면, 다음 식을 얻는다.

$$\text{Var}(\boldsymbol{\mu} | \mathbf{y}) = [\sigma^2 - \sigma^2 E(\alpha | \mathbf{y})] \left[ I_p - \frac{1}{p} \mathbf{1}_p \mathbf{1}_p^T \right] + \text{Var}(\alpha | \mathbf{y}) [\mathbf{y} - \bar{y}\mathbf{1}_p] [\mathbf{y} - \bar{y}\mathbf{1}_p]^T \quad (21)$$

식 (4)에서 알 수 있듯이, 관찰벡터  $\mathbf{y}$ 가 주어진 조건 하에서 모수벡터  $\boldsymbol{\mu}$ 는  $p$ 변량 정규 확률분포를 따른다. 따라서, 조건부확률벡터  $\boldsymbol{\mu} | \mathbf{y}$ 의 확률분포를 알기 위해서는 이 조건부확률벡터의 평균벡터와 분산공분산행렬을 구하면 된다. 식 (12)와 식 (21)에서 알 수 있듯이, 조건부확률벡터  $\boldsymbol{\mu} | \mathbf{y}$ 의 평균벡터와 분산공분산행렬을 구하기 위해서는  $E(\alpha | \mathbf{y})$ 와  $\text{Var}(\alpha | \mathbf{y})$ 를 먼저 알아야 한다. 지금부터 이들을 구하기로 하자. 식 (10)의 양변을  $\nu$ 에 대해서 적분하면,

다음 식을 얻는다.

$$\pi(\alpha | \mathbf{y}) \propto \alpha^{[p-5]/2} \exp\left(-\frac{\alpha}{2\sigma^2} \sum_{i=1}^p [y_i - \bar{y}]^2\right) \quad (22)$$

다음 확률변수를 정의하자.

$$w \doteq \frac{1}{2\sigma^2} \sum_{i=1}^p [y_i - \bar{y}]^2 \alpha \quad (23)$$

식 (23)을 식 (22)에 대입하면, 다음 식이 성립함을 알 수 있다.

$$\pi(w | \mathbf{y}) = w^{[p-3]/2-1} \exp(-w), \quad (w > 0) \quad (24)$$

즉, 조건부확률변수  $w | \mathbf{y}$ 는 형상모수가  $\alpha = \frac{p-3}{2}$ 이고 비율이  $\beta = 1$ 인 감마확률분포를 따른다. 따라서, 다음 식들이 성립한다.

$$E(w | \mathbf{y}) = \frac{p-3}{2}, \quad Var(w | \mathbf{y}) = \frac{p-3}{2} \quad (25)$$

식 (23)을 식 (25)에 대입하면, 다음 식들을 얻는다.

$$E(\alpha | \mathbf{y}) = \frac{\sigma^2[p-3]}{\sum_{i=1}^p [y_i - \bar{y}]^2}, \quad Var(\alpha | \mathbf{y}) = \frac{2\sigma^4[p-3]}{\left\{\sum_{i=1}^p [y_i - \bar{y}]^2\right\}^2} \quad (26)$$

식 (26)을 다음과 같이 구할 수도 있다. 식 (22)에서 알 수 있듯이, 다음 식들이 성립한다.

$$E(\alpha | \mathbf{y}) = \frac{\int_0^1 \alpha^{[p-3]/2} \exp\left(-\frac{\alpha}{2\sigma^2} \sum_{i=1}^p [y_i - \bar{y}]^2\right) d\alpha}{\int_0^1 \alpha^{[p-5]/2} \exp\left(-\frac{\alpha}{2\sigma^2} \sum_{i=1}^p [y_i - \bar{y}]^2\right) d\alpha} \quad (27)$$

$$E(\alpha^2 | \mathbf{y}) = \frac{\int_0^1 \alpha^{[p-1]/2} \exp\left(-\frac{\alpha}{2\sigma^2} \sum_{i=1}^p [y_i - \bar{y}]^2\right) d\alpha}{\int_0^1 \alpha^{[p-5]/2} \exp\left(-\frac{\alpha}{2\sigma^2} \sum_{i=1}^p [y_i - \bar{y}]^2\right) d\alpha} \quad (28)$$

식 (27), 식 (28), 그리고 다음 식을 사용해서,  $Var(\alpha | \mathbf{y})$ 를 구할 수 있다.

$$Var(\alpha | \mathbf{y}) = E(\alpha^2 | \mathbf{y}) - [E(\alpha | \mathbf{y})]^2 \quad (29)$$

식 (4), 식 (11)과 식 (22)에 기술한 사후확률분포를 그리기 위해서, 다음 R프로그램 Hierarchical103.R을 실행해 보자.

```

1 # -----
2 # Filename: Hierarchical103.R
3 # Hierarchical Bayes Estimate when tau is a rv
4 # Programmed by CBS
5 # -----
6 setEPS()
7 # Prior parameters
8 ( p <- 2 )
9 ( sigma0 <- 3 )
10 ( tau0 <- 2 ); ( nu0 <- 0.5 ) # pi(nu0,tau^2) propotional to 1
11 ( A = sigma0^2/(sigma0^2 + tau0^2) )
12 onep = matrix(1,p)
13 Ip = diag(p)
14
15 # Likelihood
16 y <- matrix(0,p,1)
17 prior.nu <- matrix(0,p,1)
18 set.seed(3)
19 for (ii in 1:p){
20     std0 <- sqrt(sigma0^2/A)
21     y[iii] <- nu0+std0*rnorm(1,0,1)
22 }
23 y
24 ( ybar <- mean(y) )
25 ( yVar <- var(y) )
26
27 # Posterior
28 ( Nnu = 17 ); ( NumA <- 17 )
29 nuL <- -2; nuU <- 2
30 AL <- 0.01; AU <- 0.99
31 ( nu.gd <- seq(nuL,nuU,length=Nnu) )
32 ( A.gd <- seq(AL,AU,length=NumA) )
33 ( shapeStar <- (p-3)/2 );
34 ( rateStar <- 1 );
35 posteriorPDF <- matrix(0,Nnu,NumA)
36 for (jj in 1:NumA){
37     w <- A.gd[jj]/(2*sigma0^2)*(p-1)*yVar
38     posterior.w <- w^((p-5)/2)*exp(-w)
39     for (ii in 1:Nnu){
40         posterior.nu <- dnorm(nu.gd[ii],mean=ybar,
41                               sd=sigma0/(p*sqrt(A.gd[jj])),log = FALSE)
42         posteriorPDF[ii,jj] <- posterior.nu*posterior.w
43     }
44 }
45 # -----
46 # Plotting using plot3D
47 # install.packages("plot3D")
48 library(plot3D)
49 # install.packages("reshape")
50 library(reshape)
51 plot.new()
52 postscript('Hierarchical103.eps') # Start to save figure
53 par(mfrow = c(1,1), mar = c(3, 3, 3, 2))
54 contour(nu.gd,A.gd,posteriorPDF,ylim=c(0.01,0.3),33,
55         xlab="nu",ylab="A",col="red") # no of levels = 33
56 dev.off() # End to save figure

```

```

57 # -----
58 # Plotting using ggplot2
59 # http://www.cookbook-r.com/Graphs/Bar_and_line_graphs_(ggplot2)/
60 # install.packages("ggplot2")
61 library(ggplot2)
62 DataHier103 <- data.frame(nu.gd,A.gd,posteriorPDF)
63 PDF3d <- melt(posteriorPDF) # using melt
64 # nu3 <- PDF3d[,1]; A3 <- PDF3d[,2]
65 nu3 <- nuL+(nuU-nuL)/(Nnu-1)*(PDF3d[,1]-1)
66 A3 <- AL+(AU-AL)/(NumA-1)*(PDF3d[,2]-1)
67 PDF3 <- PDF3d[,3]
68 plot201 <- ggplot(PDF3d, aes(nu3,A3,z=PDF3))
69 plot201 + stat_contour(lwd=1.2,col="blue",bins=30)
70
71 # -----

```

이 R 프로그램 Hierarchical103.R에서는 차수가  $p = 2$ 인 경우를 다룬다. 식 (1)에서 알 수 있듯이, 관찰값들을 발생시키기 위해서는  $\nu$  값과  $\tau$  값이 주어져야 한다. 여기서는  $\nu = 0.5$  그리고  $\tau = 2$ 를 할당하였다. 또한, 관찰값의 표준편차로  $\sigma = 3$ 를 할당하였다.

식 (4)에서 알 수 있듯이,  $\boldsymbol{\mu}$ 의 계층적 베이저안기법에 의한 사후확률분포는 다음과 같다.

$$\boldsymbol{\mu} | (\mathbf{y}, 0.5, 2^2) \stackrel{d}{\sim} \mathcal{N}([1 - \alpha]\mathbf{y} + 0.5\alpha\mathbf{1}_2, \sigma^2[1 - \alpha]I_2) \quad (30)$$

여기서  $\alpha = 0.6923077$ 이다. 식 (11)에서 알 수 있듯이, 초모수  $\nu$ 의 확률분포는 다음과 같다.

$$\nu | (\alpha, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}\left(-1.761508, \frac{3^2}{2\alpha}\right) \quad (31)$$

여기서  $\mathbf{y} = [-2.9683003, -0.5547165]^T$ 이다. 식 (22)에서 알 수 있듯이,  $\alpha$ 의 사후확률밀도 함수는 다음과 같다.

$$\pi(\alpha | \mathbf{y}) \propto \alpha^{-3/2} \exp\left(-\frac{\alpha}{2 \cdot 3^2} \sum_{i=1}^2 [y_i + 1.761508]^2\right) \quad (32)$$

이 R 프로그램을 실행하면, 그림 3.2.3이 그려진다. 그림 3.2.3에서 알 수 있듯이,  $\alpha$ 가 작은 경우  $\nu$ 의 변화는 사후확률밀도함수에 미미한 영향을 끼친다. 그러나,  $\alpha$  값이 커지면  $\mu$  값이 작은 경우 사후확률밀도함수가 많이 작아진다. ■

**예제 3.2.5** 경험적 베이저안추정법을 적용해서 예제 3.2.4의 계층모형을 살펴보기로 하자.

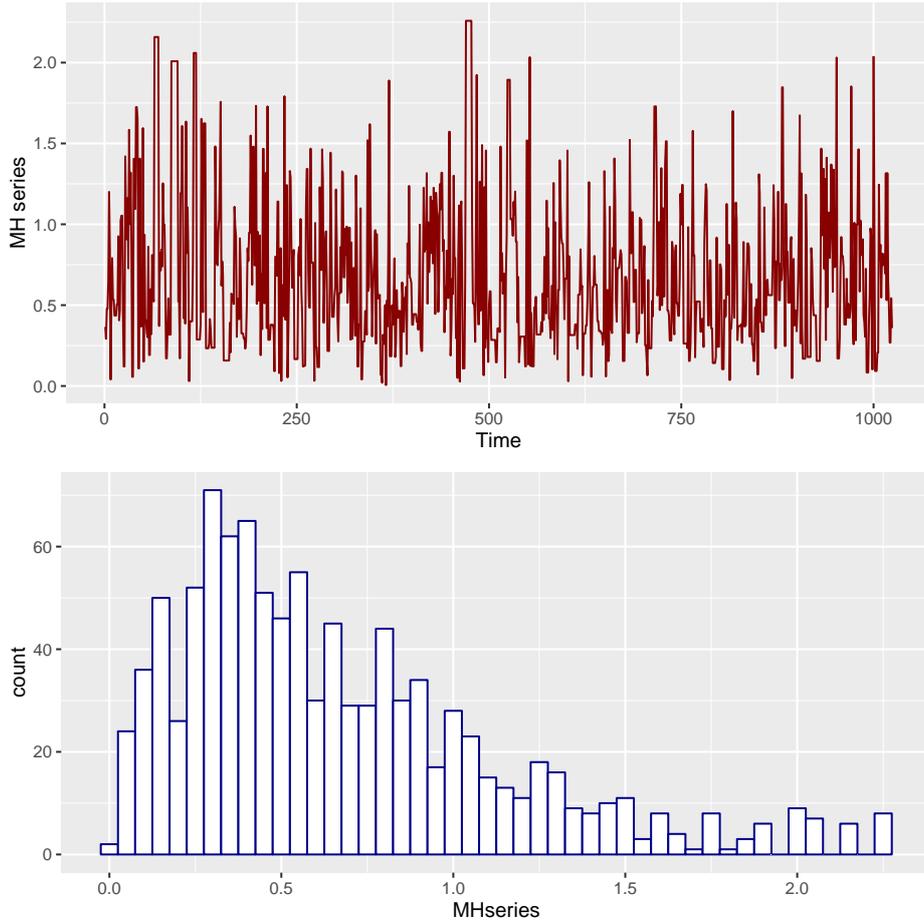


그림 3.2.3. 계층적 베이지안분석

예제 3.2.4의 식 (5)에서 알 수 있듯이, 다음 식이 성립한다.

$$\mathbf{y} \mid (\nu, \tau^2) \stackrel{d}{\sim} \mathcal{N} \left( \nu \mathbf{1}_p, \frac{\sigma^2}{\alpha} I_p \right) \tag{1}$$

따라서,  $(\nu, \tau^2)$ 의 충분통계량은  $(\bar{y}, S_\tau)$ 이다. 여기서  $S_\tau$ 는 다음과 같다.

$$S_\tau \doteq \sum_{i=1}^p [y_i - \bar{y}]^2 \tag{2}$$

다음 식이 성립한다.

$$S_\tau \stackrel{d}{\sim} \frac{\sigma^2}{\alpha} \chi_{p-1}^2 \tag{3}$$

여기서  $\alpha = \sigma^2 / [\sigma^2 + \tau^2]$ 이다. 식 (3)에서 알 수 있듯이, 각  $p(\geq 4)$ 에 대해서 다음 식들이

성립한다.

$$E\left(\frac{1}{S_\tau}\right) = \frac{\alpha}{\sigma^2} E\left(\frac{1}{\chi_{p-1}^2}\right) = \frac{\alpha}{\sigma^2[p-3]} \quad (4)$$

따라서,  $\alpha$ 를 다음과 같이 추정할 수 있다.

$$\hat{\alpha} = \frac{\sigma^2[p-3]}{S_\tau} \quad (5)$$

예제 3.2.4의 식 (12)에서 알 수 있듯이,  $\boldsymbol{\mu}$ 의 경험적 베이저안추정량은 다음과 같다.

$$\hat{\boldsymbol{\mu}}^{EB} = \left[1 - \frac{\sigma^2[p-3]}{S_\tau}\right] \mathbf{y} + \frac{\sigma^2[p-3]}{S_\tau} \bar{y} \mathbf{1}_p = \mathbf{y} - \frac{\sigma^2[p-3]}{S_\tau} [\mathbf{y} - \bar{y} \mathbf{1}_p] \quad (6)$$

■

**예제 3.2.6** 다음 선형회귀모형을 살펴보자.

$$\mathbf{y} \mid (\boldsymbol{\beta}_1, \sigma^2) \stackrel{d}{\sim} \mathcal{N}(X_1 \boldsymbol{\beta}_1, \sigma^2 I_n) \quad (1)$$

회귀계수벡터  $\boldsymbol{\beta}_1$ 을 피설명변수벡터로 하는 선형회귀모형이 다음과 같다고 하자.

$$\boldsymbol{\beta}_1 \mid \boldsymbol{\beta}_2 \stackrel{d}{\sim} \mathcal{N}(X_2 \boldsymbol{\beta}_2, C_2) \quad (2)$$

여기서 분산공분산행렬  $C_2$ 는 알고 있다고 가정하자. 선형회귀모형 (2)의 회귀계수벡터  $\boldsymbol{\beta}_2$ 와 정밀도  $\psi = 1/\sigma^2$ 의 사전확률분포들이 다음과 같다고 하자.

$$\boldsymbol{\beta}_2 \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_1, B_1), \quad \psi \stackrel{d}{\sim} \Gamma\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \quad (3)$$

여기서  $\mathbf{b}_1$ ,  $B_1$ ,  $n_0$ , 그리고  $S_0$ 는 이미 알고 있고,  $\boldsymbol{\beta}_2$ 와  $\psi$ 는 독립이다. 이 모형에서 사전확률 분포를 두 계층으로 나눌 수 있다. 첫 번째 계층은 식 (2)에 의해 주어지고, 두 번째 계층은 식 (3)으로 주어진다. 이 2단계 모형의 사후확률밀도함수는 다음과 같다.

$$\pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \psi \mid \mathbf{y}) \propto f(\mathbf{y} \mid \boldsymbol{\beta}_1, \psi) \pi(\boldsymbol{\beta}_1 \mid \boldsymbol{\beta}_2) \pi(\boldsymbol{\beta}_2) \pi(\psi) \quad (4)$$

여기서  $f(\mathbf{y} | \boldsymbol{\beta}_1, \psi)$ 는  $(\boldsymbol{\beta}_1, \psi)$ 가 주어진 조건 하에서 확률벡터  $\mathbf{y}$ 의 조건부확률밀도함수이다. 모수벡터  $\boldsymbol{\beta}_2$ 가 주어졌을 때 사후조건부확률밀도함수  $\pi(\boldsymbol{\beta}_1, \psi | \boldsymbol{\beta}_2, \mathbf{y})$ 는 다음과 같다.

$$\pi(\boldsymbol{\beta}_1, \psi | \boldsymbol{\beta}_2, \mathbf{y}) \propto f(\mathbf{y} | \boldsymbol{\beta}_1, \psi) \pi(\boldsymbol{\beta}_1 | \boldsymbol{\beta}_2) \pi(\psi) \quad (5)$$

식 (3.1.16)에서 알 수 있듯이, 우도함수는 다음과 같다.

$$f(\mathbf{y} | \boldsymbol{\beta}_1, \psi) \propto \psi^{n/2} \exp\left(-\frac{\psi}{2} \left\{ [\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1]^T X_1^T X_1 [\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1] + S_e \right\}\right) \quad (6)$$

여기서 다음 식들이 성립한다.

$$\hat{\boldsymbol{\beta}}_1 = [X_1^T X_1]^{-1} X_1^T \mathbf{y}, \quad S_e = [\mathbf{y} - X_1 \hat{\boldsymbol{\beta}}_1]^T [\mathbf{y} - X_1 \hat{\boldsymbol{\beta}}_1] \quad (7)$$

식 (2)에서 알 수 있듯이, 다음 식이 성립한다.

$$\pi(\boldsymbol{\beta}_1 | \boldsymbol{\beta}_2) \propto \exp\left(-\frac{1}{2} [\boldsymbol{\beta}_1 - X_2 \boldsymbol{\beta}_2]^T C_2^{-1} [\boldsymbol{\beta}_1 - X_2 \boldsymbol{\beta}_2]\right) \quad (8)$$

식 (3)에서 알 수 있듯이, 다음 식이 성립한다.

$$\pi(\psi) \propto \psi^{n_0/2-1} \exp\left(-\frac{n_0 S_0 \psi}{2}\right) \quad (9)$$

식 (7)~식 (9)에서 알 수 있듯이,  $(\boldsymbol{\beta}_1, \psi) | (\boldsymbol{\beta}_2, \mathbf{y})$ 의 사후조건부확률밀도함수는 다음과 같다.

$$\begin{aligned} & \pi(\boldsymbol{\beta}_1, \psi | \boldsymbol{\beta}_2, \mathbf{y}) \\ & \propto \exp\left(-\frac{1}{2} \left\{ \psi [\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1]^T X_1^T X_1 [\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1] + [\boldsymbol{\beta}_1 - X_2 \boldsymbol{\beta}_2]^T C_2^{-1} [\boldsymbol{\beta}_1 - X_2 \boldsymbol{\beta}_2] \right\}\right) \\ & \quad \cdot \psi^{n_0/2-1} \exp\left(-\frac{n_0 S_0 \psi}{2}\right) \end{aligned} \quad (10)$$

다음 행렬과 벡터를 정의하자.

$$B_{1*} \doteq [C_2^{-1} + \psi X_1^T X_1]^{-1}, \quad \mathbf{b}_{1*} \doteq B_{1*} [C_2^{-1} X_2 \boldsymbol{\beta}_2 + \psi X_1^T \mathbf{y}] \quad (11)$$

식 (11)을 식 (10)에 대입하면,  $(\boldsymbol{\beta}_1, \psi) | (\boldsymbol{\beta}_2, \mathbf{y})$ 의 사후조건부확률밀도함수가 다음과 같음을

알 수 있다.

$$\pi(\boldsymbol{\beta}_1, \psi \mid \boldsymbol{\beta}_2, \mathbf{y}) \propto \exp\left(-\frac{1}{2}[\boldsymbol{\beta}_1 - \mathbf{b}_{1*}]^T B_{1*}^{-1}[\boldsymbol{\beta}_1 - \mathbf{b}_{1*}]\right) \psi^{n_0/2-1} \exp\left(-\frac{n_0 S_0 \psi}{2}\right) \quad (12)$$

따라서, 다음 식이 성립함을 알 수 있다.

$$\pi(\boldsymbol{\beta}_1 \mid \psi, \boldsymbol{\beta}_2, \mathbf{y}) \propto \exp\left(-\frac{1}{2}[\boldsymbol{\beta}_1 - \mathbf{b}_{1*}]^T B_{1*}^{-1}[\boldsymbol{\beta}_1 - \mathbf{b}_{1*}]\right) \quad (12)$$

즉, 다음 식이 성립한다.

$$\boldsymbol{\beta}_1 \mid (\psi, \boldsymbol{\beta}_2, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_{1*}, B_{1*}) \quad (13)$$

식 (2)에서 알 수 있듯이, 다음 식이 성립한다.

$$\boldsymbol{\beta}_1 \mid (\psi, \boldsymbol{\beta}_2) \stackrel{d}{\sim} \mathcal{N}(X_2 \boldsymbol{\beta}_2, C_2) \quad (14)$$

식 (13)과 식 (14)에서 알 수 있듯이,  $(\psi, \boldsymbol{\beta}_2)$ 가 주어진 조건 하에서  $\boldsymbol{\beta}_1$ 의 사전조건부확률분포와 사후조건부확률분포는 정규확률분포족에 속한다. 따라서, 이들은 조건부공액이다.

식 (1)에서 알 수 있듯이, 다음 식이 성립한다.

$$\mathbf{y} \mid (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma^2) \stackrel{d}{\sim} \mathcal{N}(X_1 \boldsymbol{\beta}_1, \sigma^2 I_n) \quad (15)$$

즉,  $\mathbf{y}$ 와  $\boldsymbol{\beta}_2$ 는 독립이다. 따라서,  $\boldsymbol{\beta}_2$ 의 사후조건부확률분포는  $\mathbf{y}$ 에 의존하지 않는다. 또한 다음 식들이 성립한다.

$$\begin{aligned} \pi(\boldsymbol{\beta}_2 \mid \psi, \boldsymbol{\beta}_1) &\propto \pi(\boldsymbol{\beta}_1 \mid \boldsymbol{\beta}_2) \pi(\boldsymbol{\beta}_2) \\ &= \exp\left(-\frac{1}{2}[\boldsymbol{\beta}_1 - X_2 \boldsymbol{\beta}_2]^T C_2^{-1}[\boldsymbol{\beta}_1 - X_2 \boldsymbol{\beta}_2]\right) \exp\left(-\frac{1}{2}[\boldsymbol{\beta}_2 - \mathbf{b}_1]^T B_1^{-1}[\boldsymbol{\beta}_2 - \mathbf{b}_1]\right) \end{aligned} \quad (16)$$

식 (16)을 다음과 같이 쓸 수 있다.

$$\pi(\boldsymbol{\beta}_2 \mid \psi, \boldsymbol{\beta}_1) \propto \exp\left(-\frac{1}{2}[\boldsymbol{\beta}_2 - \mathbf{b}_{2*}]^T B_{2*}^{-1}[\boldsymbol{\beta}_2 - \mathbf{b}_{2*}]\right) \quad (17)$$

여기서  $B_{2*}$ 와  $\mathbf{b}_{2*}$ 는 각각 다음과 같다.

$$B_{2*} \doteq [B_1^{-1} + X_2^T C_2^{-1} X_2]^{-1}, \quad \mathbf{b}_{2*} \doteq B_{2*} [B_1^{-1} \boldsymbol{\beta}_1 + X_2^T C_2^{-1} \boldsymbol{\beta}_1] \quad (18)$$

식 (17)에서 알 수 있듯이  $(\psi, \boldsymbol{\beta}_1)$ 가 주어진 조건 하에서  $\boldsymbol{\beta}_2$ 의 사후조건부확률분포는 정규확률분포족에 속한다. 또한, 식 (2)와 식 (3)에서 알 수 있듯이, 같은 조건 하에서  $\boldsymbol{\beta}_2$ 의 사전조건부확률분포는 정규확률분포족에 속한다. 따라서, 이들은 조건부공액이다.

지금까지는 사전확률분포의 계층이 2개인 경우를 다루었다. 이를 확장해서, 다음과 같이 좀 더 많은 계층들로 구성된 모형을 생각할 수도 있다.

$$\boldsymbol{\beta}_j | \boldsymbol{\beta}_{j+1} \stackrel{d}{\sim} \mathcal{N}(X_{j+1} \boldsymbol{\beta}_{j+1}, C_{j+1}), \quad (j = 1, 2, \dots) \quad (4)$$

■

### 제3.3절 다변량선형회귀모형

#### 3.3.1 고전적 통계추론

다변량선형회귀모형에 대해서 간단히 살펴보자. 이에 대한 좀 더 자세한 내용은 최병선 [4, 제 12.4절]을 참조하라. 종속변수들  $y_1, y_2, \dots, y_p$ 에 대해서 종속벡터  $\mathbf{y}$ 를 다음과 같이 정의하자.

$$\mathbf{y} = [y_1, y_2, \dots, y_p]^T \quad (3.3.1)$$

이 종속벡터에 영향을 미치는 설명변수들을  $x_1, x_2, \dots, x_k$ 라 하고, 이들이 다음 다변량선형 회귀모형을 만족한다고 하자.

$$\mathbf{y} = x_1 \boldsymbol{\beta}_1 + x_2 \boldsymbol{\beta}_2 + \dots + x_k \boldsymbol{\beta}_k + \boldsymbol{\epsilon} \quad (3.3.2)$$

여기서  $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \dots, \epsilon_p]^T$ 는 평균벡터가  $\mathbf{0}$ 이고, 분산공분산행렬이  $\Sigma$ 인 오차벡터이며,  $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_k$ 는 회귀계수벡터들로서 다음과 같이 정의된다.

$$\boldsymbol{\beta}_i = [\beta_{1i}, \beta_{2i}, \dots, \beta_{ki}]^T, \quad (i = 1, 2, \dots, k) \quad (3.3.3)$$

다음과 같은  $n$  개 관찰점들이 관찰되었다고 하자.

$$\left[ \mathbf{y}_{(j)}^T, \mathbf{x}_{(j)} \right] = [y_{j1}, y_{j2}, \dots, y_{jp}, x_{j1}, x_{j2}, \dots, x_{jk}], \quad (j = 1, 2, \dots, n) \quad (3.3.4)$$

다음 행렬들을 정의하자.

$$Y \doteq \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix} \doteq \begin{bmatrix} \mathbf{y}_{(1)}^T \\ \mathbf{y}_{(2)}^T \\ \vdots \\ \mathbf{y}_{(n)}^T \end{bmatrix}, \quad X \doteq \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \doteq \begin{bmatrix} \mathbf{x}_{(1)} \\ \mathbf{x}_{(2)} \\ \vdots \\ \mathbf{x}_{(n)} \end{bmatrix} \quad (3.3.5)$$

$$\mathcal{E} \doteq \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{1p} \\ \epsilon_{21} & \epsilon_{22} & \cdots & \epsilon_{2p} \\ \vdots & \vdots & & \vdots \\ \epsilon_{n1} & \epsilon_{n2} & \cdots & \epsilon_{np} \end{bmatrix} \doteq \begin{bmatrix} \boldsymbol{\epsilon}_{(1)}^T \\ \boldsymbol{\epsilon}_{(2)}^T \\ \vdots \\ \boldsymbol{\epsilon}_{(n)}^T \end{bmatrix}, \quad B \doteq \begin{bmatrix} \beta_{11} & \beta_{21} & \cdots & \beta_{p1} \\ \beta_{12} & \beta_{22} & \cdots & \beta_{p2} \\ \vdots & \vdots & & \vdots \\ \beta_{1k} & \beta_{2k} & \cdots & \beta_{pk} \end{bmatrix} \doteq \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \boldsymbol{\beta}_2^T \\ \vdots \\ \boldsymbol{\beta}_k^T \end{bmatrix} \quad (3.3.6)$$

여기서  $\boldsymbol{\epsilon}_{(1)}, \boldsymbol{\epsilon}_{(2)}, \dots, \boldsymbol{\epsilon}_{(n)}$  은 서로 독립이며 평균벡터가  $\mathbf{0}$  이고 분산공분산행렬이  $\Sigma_p$  인 확률 벡들이다. 이 행렬들을 사용해서 다변량선형회귀모형 (3.3.2) 를 다음과 같이 표기할 수 있다.

$$Y = XB + \mathcal{E} \quad (3.3.7)$$

식 (3.3.7) 을  $vec$  연산자를 사용해서 다음과 같이 열벡터로 표현할 수 있다.

$$vec(Y) = (I_n \otimes X) vec(B) + vec(\mathcal{E}) \quad (3.3.8)$$

다음 식이 성립함을 쉽게 알 수 있다.

$$Var(vec(\mathcal{E}^T)) = I_n \otimes \Sigma \quad (3.3.9)$$

다음 오차제곱합  $S_\epsilon$  을 최소화하는 최소제곱추정량을 구해보자.

$$S_\epsilon \doteq vec(\mathcal{E}^T)^T (I_n \otimes \Sigma)^{-1} vec(\mathcal{E}^T) \quad (3.3.10)$$

다음 식이 성립한다.

$$S_\epsilon = \text{vec}(Y^T - B^T X^T)^T (I_n \otimes \Sigma)^{-1} \text{vec}(Y^T - B^T X^T) \quad (3.3.11)$$

여기서 유의할 점은  $S_\epsilon$ 은 스칼라이고  $B$ 는 행렬이라는 것이다. 따라서,  $S_\epsilon$ 을 최소로 하는  $B$ 를 구하는 것보다는,  $S_\epsilon$ 을 최소로 하는  $\text{vec}(B)$ 를 구하는 것이 더 쉬운 방법일 것이다. 이 점에 유의하여,  $S_\epsilon$ 을 다음과 같이 표현하자.

$$\begin{aligned} S_\epsilon &= \text{vec}(Y^T)^T (I_n \otimes \Sigma)^{-1} \text{vec}(Y^T) - 2\text{vec}(B^T X^T)^T (I_n \otimes \Sigma)^{-1} \text{vec}(Y^T) \\ &\quad + \text{vec}(B^T X^T)^T (I_n \otimes \Sigma)^{-1} \text{vec}(B^T X^T) \end{aligned} \quad (3.3.12)$$

다음 식들이 성립한다.

$$\begin{aligned} &\text{vec}(B^T X^T)^T (I_n \otimes \Sigma)^{-1} \text{vec}(Y^T) \\ &= \text{vec}(B^T)^T (X^T \otimes I_n) (I_n \otimes \Sigma^{-1}) \text{vec}(Y^T) \\ &= \text{vec}(B^T)^T (X^T \otimes \Sigma^{-1}) \text{vec}(Y^T) \end{aligned} \quad (3.3.13)$$

$$\begin{aligned} &\text{vec}(B^T X^T)^T (I_n \otimes \Sigma)^{-1} \text{vec}(B^T X^T) \\ &= \text{vec}(B^T)^T (X^T \otimes I_n) (I_n \otimes \Sigma^{-1}) (X \otimes I_n) \text{vec}(B^T) \\ &= \text{vec}(B^T)^T (X^T X) \otimes \Sigma^{-1} \text{vec}(B^T) \end{aligned} \quad (3.3.14)$$

식 (3.3.13)과 식 (3.3.14)를 식 (3.3.12)에 대입한 다음 미분하면, 다음 식을 얻는다.

$$\frac{\partial S_\epsilon}{\partial \text{vec}(B^T)} = -2 (X^T \otimes \Sigma^{-1}) \text{vec}(Y^T) + 2 (X^T X) \otimes \Sigma^{-1} \text{vec}(B^T) \quad (3.3.15)$$

따라서, 정규방정식은 다음과 같다.

$$(X^T X) \otimes \Sigma^{-1} \text{vec}(B^T) = (X^T \otimes \Sigma^{-1}) \text{vec}(Y^T) \quad (3.3.16)$$

오차제곱합의 Hessian 행렬은 다음과 같다.

$$\frac{\partial S_{\epsilon}}{\partial \text{vec}(B^T) \partial \text{vec}(B^T)^T} = 2(X^T X) \otimes \Sigma^{-1} \quad (3.3.17)$$

이 Hessian 행렬은 양반정치이다. 식 (3.3.16)에서 알 수 있듯이,  $\text{vec}(B)$ 의 최소제곱추정량은 다음과 같다.

$$\begin{aligned} \text{vec}(\hat{B}^T) &= [(X^T X) \otimes \Sigma^{-1}]^{-1} (X' \otimes \Sigma^{-1}) \text{vec}(Y^T) \\ &= [(X^T X)^{-1} \otimes \Sigma] (X' \otimes \Sigma^{-1}) \text{vec}(Y^T) \\ &= [(X^T X)^{-1} X^T] \otimes I_n \text{vec}(Y^T) \end{aligned} \quad (3.3.18)$$

즉, 다음 식이 성립한다.

$$\text{vec}(\hat{B}^T) = \text{vec}(Y^T X (X^T X)^{-1}) \quad (3.3.19)$$

따라서,  $B$ 의 최소제곱추정량은 다음과 같다.

$$\hat{B} = (X^T X)^{-1} X^T Y \quad (3.3.20)$$

행렬들  $Y$ ,  $X$ ,  $B$  그리고  $\mathcal{E}$ 을 각각 다음과 같이 분할하자.

$$Y = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_p \end{bmatrix}, \quad X = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k \end{bmatrix} \quad (3.3.21)$$

$$B = \begin{bmatrix} \beta_1^* & \beta_2^* & \cdots & \beta_p^* \end{bmatrix}, \quad \mathcal{E} = \begin{bmatrix} \epsilon_1 & \epsilon_2 & \cdots & \epsilon_p \end{bmatrix} \quad (3.3.22)$$

다음 식들이 성립한다.

$$E(\epsilon_i) = \mathbf{0}, \quad \text{Var}(\epsilon_i) = \sigma_{ii} I, \quad \text{Cov}(\epsilon_i, \epsilon_j) = \sigma_{ij} I \quad (3.3.23)$$

따라서, 각  $i(=1, 2, \dots, p)$ 에 대해서 다음 식이 성립한다.

$$\mathbf{y}_i = X \beta_i^* + \epsilon_i \quad (3.3.24)$$

식 (3.3.24)의 선형회귀모형에서 회귀계수벡터  $\beta_i^*$ 의 최소제곱추정량은 다음과 같다.

$$\hat{\beta}_i^* = (X^T X)^{-1} X^T \mathbf{y}_i \quad (3.3.25)$$

이 최소제곱추정량들에서 알 수 있듯이, 모수행렬  $B$ 의 다음과 같은 추정량을 얻을 수 있다.

$$\tilde{B} = [\hat{\beta}_1^*, \hat{\beta}_2^*, \dots, \hat{\beta}_p^*] = (X^T X)^{-1} X^T Y \quad (3.3.26)$$

어떤 책에서는 이 추정량  $\tilde{B}$ 를 회귀계수행렬  $B$ 의 최소제곱추정량이라고 하였다. 그러나, 일반적인 경우에 이러한 결론을 도출할 수는 없다. 이 경우에 추정량  $\tilde{B}$ 가 최소제곱추정량  $\hat{B}$ 과 동일한 이유는 디자인행렬이 동일하기 때문이다.

최우추정법을 적용해서 이 다변량선형회귀모형을 추정해보자. 만약 오차벡터  $\epsilon$ 이 평균벡터가  $\mathbf{0}$ 이고 분산공분산행렬이  $\Sigma$ 인 다변량정규확률분포  $\mathcal{N}_p(\mathbf{0}, \Sigma)$ 를 따르면, 이 다변량선형회귀모형의 우도함수  $L(B, \Sigma)$ 는 다음과 같다.

$$L(B, \Sigma) = \prod_{j=1}^n \left[ \frac{1}{[2\pi]^{p/2} |\Sigma|^{1/2}} \exp \left( -\frac{1}{2} [\mathbf{y}_{(j)}^T - \mathbf{x}_{(j)} B] \Sigma^{-1} [\mathbf{y}_{(j)}^T - \mathbf{x}_{(j)} B]^T \right) \right] \quad (3.3.27)$$

최소제곱추정량  $\hat{B}$ 의 정규방정식 (3.3.16)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\begin{aligned} & [Y - XB]^T [Y - XB] \\ &= \hat{\mathcal{E}}^T \hat{\mathcal{E}} + [\hat{B} - B]^T X^T X [\hat{B} - B] \\ &= \hat{S} + [\hat{B} - B]^T X^T X [\hat{B} - B] \end{aligned} \quad (3.3.28)$$

여기서  $\hat{\mathcal{E}}$ 와  $\hat{S}$ 는 각각 다음과 같다.

$$\hat{\mathcal{E}} \doteq Y - X\hat{B}, \quad \hat{S} \doteq \hat{\mathcal{E}}^T \hat{\mathcal{E}} \quad (3.3.29)$$

다음 식들이 성립한다.

$$\begin{aligned}
& \sum_{j=1}^n [\mathbf{y}_{(j)}^T - \mathbf{x}_{(j)}B] \Sigma^{-1} [\mathbf{y}_{(j)}^T - \mathbf{x}_{(j)}B]^T \\
&= \text{Tr} \left( \Sigma^{-1} \sum_{j=1}^n [\mathbf{y}_{(j)}^T - \mathbf{x}_{(j)}B]^T [\mathbf{y}_{(j)}^T - \mathbf{x}_{(j)}B] \right) \\
&= \text{Tr} \left( \Sigma^{-1} \sum_{j=1}^n \boldsymbol{\epsilon}_{(j)} \boldsymbol{\epsilon}_{(j)}^T \right) = \text{Tr} (\Sigma^{-1} \boldsymbol{\mathcal{E}}^T \boldsymbol{\mathcal{E}}) \tag{3.3.30}
\end{aligned}$$

또한, 다음 식들이 성립한다.

$$\begin{aligned}
\text{Tr} (\Sigma^{-1} \boldsymbol{\mathcal{E}}^T \boldsymbol{\mathcal{E}}) &= \text{Tr} (\Sigma^{-1} [Y - XB]^T [Y - XB]) \\
&= \text{Tr} \left( \Sigma^{-1} \left\{ \hat{S} + [\hat{B} - B]^T X^T X [\hat{B} - B] \right\} \right) \\
&= \text{Tr} (\Sigma^{-1} \hat{S}) + \text{Tr} \left( \Sigma^{-1} [\hat{B} - B]^T X^T X [\hat{B} - B] \right) \tag{3.3.31}
\end{aligned}$$

여기서 두 번째 등호는 식 (3.3.28)에 의해서 성립한다. 식 (3.3.30)과 식 (3.3.31)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\begin{aligned}
& \sum_{j=1}^n [\mathbf{y}_{(j)}^T - \mathbf{x}_{(j)}B] \Sigma^{-1} [\mathbf{y}_{(j)}^T - \mathbf{x}_{(j)}B]^T \\
&= \text{Tr} (\Sigma^{-1} \hat{S}) + \text{Tr} \left( X [\hat{B} - B] \Sigma^{-1} [\hat{B} - B]^T X^T \right) \\
&\leq \text{Tr} (\Sigma^{-1} \hat{S}) \tag{3.3.32}
\end{aligned}$$

식 (3.3.27)과 식 (3.3.32)에서 알 수 있듯이, 다음 식들이 성립한다.

$$L(B, \Sigma) \leq L(\hat{B}, \Sigma) = \frac{1}{[2\pi]^{pn/2} |\Sigma|^{n/2}} \exp \left( -\frac{1}{2} \text{Tr} (\Sigma^{-1} \hat{S}) \right) \tag{3.3.33}$$

즉, 최소제곱추정량  $\hat{B}$ 가 최우추정량이다.

다음 식이 성립한다.

$$\ln L(\hat{B}, \Sigma) = -\frac{pn}{2} \ln(2\pi) - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \text{Tr} (\Sigma^{-1} \hat{S}) \tag{3.3.34}$$

따라서, 다음 식이 성립한다.

$$\frac{\partial \ln L(\hat{B}, \Sigma)}{\partial \Sigma} = -\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}\hat{S}\Sigma^{-1} \quad (3.3.35)$$

즉,  $\Sigma$ 의 최우추정량은 다음과 같다.

$$\hat{\Sigma} = \frac{1}{n}\hat{S} = \frac{1}{n} [Y - \hat{Y}]^T [Y - \hat{Y}] \quad (3.3.36)$$

여기서  $\hat{Y} = X\hat{B}$ 이다.

다음 명제의 증명에 대해서는 Anderson [15, 제7장]을 참조하라.

**명제 3.3.1**

만약 오차벡터  $\epsilon$ 이 평균벡터가  $\mathbf{0}$ 이고 분산공분산행렬이  $\Sigma$ 인 다변량정규확률분포  $\mathcal{N}_p(\mathbf{0}, \Sigma)$ 를 따르면, 다음 성질들이 성립한다.

a)  $vec(\hat{B}^T) \stackrel{d}{\sim} \mathcal{N}(vec(B^T), (X^T X)^{-1} \otimes \Sigma)$

b) 오차벡터의 분산공분산행렬  $\Sigma$ 의 최우추정량  $\hat{\Sigma}$ 는 다음 식을 만족한다.

$$n\hat{\Sigma} \stackrel{d}{\sim} W_p(\Sigma, n - k)$$

c) 최우추정량들  $\hat{B}$ 과  $\hat{\Sigma}$ 는 서로 독립이다.

단변량인 확률변수의 표본분산  $s^2$ 는 관찰값들이 표본평균 주위에 얼마나 퍼져있는가를 나타내는 척도이다. 마찬가지로 다변량인 확률벡터의 표본분산공분산행렬  $S_\epsilon$ 도 관찰벡터들이 표본평균벡터 주위에 퍼져있는 정도를 나타낸다고 할 수 있다. 다변량인 경우에 이 퍼짐의 정도를 하나의 척도로 나타내기 위해서 일반화분산(generalized variance)  $|S_\epsilon|$ 이나 총변동량(total variation)  $\text{Tr}(S_\epsilon)$ 을 사용한다. 만약 이 통계량값들이 작으면, 표본평균벡터 주위에 관찰벡터들이 모여있음을 뜻한다. 총변동량  $\text{Tr}(S_\epsilon)$ 은 변수들 사이의 상관관계를 무시하고 표본분산들의 합을 전체 변동량으로 간주하는데 반해서, 일반화분산  $|S_\epsilon|$ 은 변수들 사이의

상관관계를 반영한다. 최소제곱추정량  $\hat{B}$ 는 다음 식들을 최소화한다.

$$\det(\mathcal{E}^T \mathcal{E}) = \det\left([Y - XB]^T [Y - XB]\right) \quad (3.3.37)$$

$$\text{Tr}(\mathcal{E}^T \mathcal{E}) = \text{Tr}\left([Y - XB]^T [Y - XB]\right) \quad (3.3.38)$$

즉, 최소제곱추정량  $\hat{B}$ 은 일반화분산과 총변동량을 최소화한다.

다변량선형회귀모형에 의한 총제곱교차곱합행렬 (total sum of squares and crossproducts)  $SST$ , 모형제곱교차곱합행렬 (sum of squares and crossproducts due to model)  $SSM$ , 그리고 잔차제곱교차곱합행렬 (sum of squares and crossproducts due to error)  $SSE$ 를 각각 다음과 같이 정의하자.

$$SST \doteq Y^T Y \quad (3.3.39)$$

$$SSM \doteq \hat{Y}^T \hat{Y} = \hat{B}^T X^T Y = Y^T \hat{Y} \quad (3.3.40)$$

$$SSE \doteq [Y - \hat{Y}]^T [Y - \hat{Y}] \quad (3.3.41)$$

이후 편의상 잔차제곱교차곱합행렬  $SSE$ 를  $\hat{S}$ 로 표기하자. 명제 3.3.1에서 알 수 있듯이, 분산공분산행렬  $\Sigma$ 의 불편추정량  $\tilde{\Sigma}$ 를 다음과 같이 정의할 수 있다.

$$\tilde{\Sigma} \doteq \frac{1}{n-k} \hat{S} = \frac{1}{n-k} [Y - \hat{Y}]^T [Y - \hat{Y}] \quad (3.3.42)$$

**예제 3.3.1** 다변량선형회귀모형을 적용하는 예로서 다음 MATLAB 프로그램 MultivariateLinearRegression101.m을 실행해 보자.

```

1 % -----
2 % Filename: MultivariateLinearRegression101.m
3 % https://kr.mathworks.com/help/stats/mvregress.html#btkqm0f-4
4 % -----
5 clear all, clf, close all
6
7 % Load the dataset
8 load('flu')
9 summary(flu)
10 Y = double(flu(:,2:end-1));
11 [n,d] = size(Y)
12 x = flu.WtdILI;
13 % % Figure of the Dataset
14 figure;
15 regions = flu.Properties.VarNames(2:end-1);
16 plot(x,Y,'x')
17 set(gca,'fontsize',11,'fontweigh','bold')

```

```

18 legend(regions, 'Location', 'NorthWest')
19 title('\bf Dataset flu')
20 xlabel('\bf Regressor WtdILLI', 'FontSize', 12);
21 saveas(gcf, 'MultivariateLinearRegression101a', 'jpg')
22
23 % Multivariate Regression A; Panel data w/ different Intercepts
24 X = cell(n,1);
25 for i = 1:n
26     X{i} = [eye(d) repmat(x(i),d,1)];
27 end
28 [betaA, SigmaA] = mvregress(X,Y);
29 B = [betaA(1:d)'; repmat(betaA(end),1,d)]
30 display(SigmaA)
31 CondNum = cond(SigmaA)
32 xx = linspace(.5,3.5)';
33 fits = [ones(size(xx)), xx]*B;
34 %% Plot of fitted values
35 figure;
36 h = plot(x,Y, 'x', xx, fits, '-');
37 for i = 1:d
38     set(h(d+i), 'color', get(h(i), 'color'));
39 end
40 regions = flu.Properties.VarNames(2:end-1);
41 legend(regions, 'Location', 'NorthWest');
42 set(gca, 'fontsize', 11, 'fontweight', 'bold')
43 title('\bf Panel Data w/ Different Intercepts')
44 xlabel('\bf Regressor WtdILLI', 'FontSize', 12);
45 saveas(gcf, 'MultivariateLinearRegression101b', 'jpg')
46
47 % Multivariate Regression 2; Panel data w/ different Slopes
48 X = cell(n,1);
49 for i = 1:n
50     X{i} = [eye(d) x(i)*eye(d)];
51 end
52 [betaB, SigmaB] = mvregress(X,Y, 'algorithm', 'cwls')
53 % Regression coefficient and Fitted values
54 B = [betaB(1:d)'; betaB(d+1:end)']
55 display(SigmaB)
56 CondNum = cond(SigmaB)
57 xx = linspace(.5,3.5)';
58 fits = [ones(size(xx)), xx]*B;
59 %% Plot of fitted values
60 figure;
61 h = plot(x,Y, 'x', xx, fits, '-');
62 for i = 1:d
63     set(h(d+i), 'color', get(h(i), 'color'));
64 end
65 regions = flu.Properties.VarNames(2:end-1);
66 legend(regions, 'Location', 'NorthWest');
67 set(gca, 'fontsize', 11, 'fontweight', 'bold')
68 title('\bf Panel Data w/ Different Slopes')
69 xlabel('\bf Regressor WtdILLI', 'FontSize', 12);
70 saveas(gcf, 'MultivariateLinearRegression101c', 'jpg')
71
72 % Multivariate Regression 3; With a Single Design matrix
73 X = cell(n,1);
74 X = [ones(size(x)), x];
75 [betaC, SigmaC, E, CovB, logL] = mvregress(X,Y)
76 B = betaC
77 display(SigmaC)
78 CondNum = cond(SigmaC)

```

```

79 xx = linspace(.5,3.5)';
80 fits = [ones(size(xx)),xx]*B;
81 % % Plot of fitted values
82 figure;
83 h = plot(x,Y,'x', xx,fits,'-');
84 for i = 1:d
85     set(h(d+i), 'color',get(h(i), 'color'));
86 end
87 regions = flu.Properties.VarNames(2:end-1);
88 legend(regions, 'Location', 'NorthWest');
89 set(gca, 'fontsize',11, 'fontweigh', 'bold')
90 title('\bf A Single Design Matrix')
91 xlabel('\bf Regressor WtdILI', 'FontSize',12');
92 saveas(gcf, 'MultivariateLinearRegression101d', 'jpg')
93
94 % End of program
95 % -----

```

이 예제에서 사용하는 데이터세트는 Google Flu Trends이다. 이 데이터세트는 MATLAB 데이터세트 'flu'에 저장되어 있다. 이 데이터세트는 미국 내 9개 지역의 ILI (influenza-like illness) 퍼센트들을 나타내는 변수들 NE, MidAtl, ENCentral, WNCentral, SATl, ESCentral, WSCentral, Mtn, Pac과 미국 CDC (Center for Disease Control)가 제공하는 가중ILI퍼센트 (weighted ILI)인 변수 WtdILI로 구성되어 있다.

이 MATLAB 프로그램 MultivariateLinearRegression101.m을 실행하면,  $X$  축에는 변수 WtdILI를, 그리고  $Y$  축에는 변수들 NE, MidAtl, ENCentral, WNCentral, SATl, ESCentral, WSCentral, Mtn, Pac을 표시한 산점도가 그림 3.3.1에 수록된다. 그림 3.3.1에서 변수 WtdILI 값이 커지면 변수들 NE, MidAtl, ENCentral, WNCentral, SATl, ESCentral, WSCentral, Mtn, Pac의 값들도 커진다. 다만 상승 기울기가 지역마다 차이가 있음을 확인할 수 있다.

첫 번째 모형은, 절편은 다르지만 기울기가 같은 다변량선형회귀모형이다. 즉, 다음 다변량선형회귀모형을 데이터세트 flu에 적합시키자.

$$\mathbf{y}_j = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_9 \end{bmatrix} + \beta x_j + \epsilon_j, \quad (j = 1, 2, \dots, n) \quad (1)$$

여기서  $x$ 는 설명변수 WtdILI이고  $\mathbf{y}$ 는 종속벡터이다. MATLAB 명령문 mvregress를 사용해

서 이 다변량선형회귀모형을 추정한 결과는 다음과 같다.

$$\boldsymbol{\alpha} = \begin{bmatrix} 0.3036 & 0.3138 & 0.3555 & 0.5405 & 0.3791 & 0.6428 & 0.3726 & 0.3503 & 0.1435 \end{bmatrix}^t \quad (2)$$

$$\hat{\beta} = 0.5870 \quad (3)$$

또한, 추정된 분산공분산행렬의 조건수(condition number)는 613.8684이다. 이렇게 추정된 선형회귀선들이 그림 3.3.2에 그려져 있다.

두 번째 모형은, 절편 뿐아니라 기울기도 동일하지 않다고 가정하는 다변량선형회귀모형이다. 즉, 다음 다변량선형회귀모형을 데이터세트 flu에 적합시키자.

$$\mathbf{y}_j = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_9 \end{bmatrix} + x_j \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_9 \end{bmatrix} + \boldsymbol{\epsilon}_j, \quad (j = 1, 2, \dots, n) \quad (4)$$

여기서  $x$ 는 설명변수 WtdILI이고  $\mathbf{y}$ 는 종속벡터이다. MATLAB명령문 mvregress에서는 옵션 algorithm에 cwls를 할당했다. 이 CWLS는 Covariance-Weighted Least Squares의 두음문자(acronym)로 분산공분산행렬을 가중행렬로 하는 일반화최소제곱추정법이다. 이 옵션을 적용해서, 다변량선형회귀모형 (4)를 추정한 결과는 다음과 같다.

$$\hat{\boldsymbol{\alpha}} = \begin{bmatrix} 0.1857 & 0.2425 & 0.2452 & 0.0681 & 0.1813 & 0.2542 & 0.1539 & -0.0161 & 0.1599 \end{bmatrix}^t \quad (5)$$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 0.6622 & 0.6325 & 0.6574 & 0.8884 & 0.7132 & 0.8350 & 0.7265 & 0.8208 & 0.5765 \end{bmatrix}^t \quad (6)$$

또한, 추정된 분산공분산행렬의 조건수가 477.2614로서 613.8684보다 작다. 따라서, 모든 종속변수들의 기울기들이 동일하다는 가정을 하는 다변량선형회귀모형보다는 그러한 제약을 부과하지 않는 다변량선형회귀모형을 적용하는 것이 더 타당하다는 것을 알 수 있다. 이렇게 추정된 선형회귀선들이 그림 3.3.3에 그려져 있다.

세 번째 모형은, 디자인행렬이 하나로 되어 있는 다변량선형회귀모형이다. 즉, 다음

다변량선형회귀모형을 데이터세트  $\text{flu}$ 에 적합시키자.

$$Y_{n \times p} = \begin{bmatrix} \boldsymbol{\alpha}^t + x_1 \boldsymbol{\beta}^t \\ \boldsymbol{\alpha}^t + x_2 \boldsymbol{\beta}^t \\ \vdots \\ \boldsymbol{\alpha}^t + x_n \boldsymbol{\beta}^t \end{bmatrix}_{n \times p} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}_{n \times 2} \begin{bmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \end{bmatrix}_{p \times 2}^T \quad (7)$$

여기서  $x$ 는 설명변수 WtdILI이고  $\mathbf{y}$ 는 종속벡터이다. MATLAB 명령문 `mvregress`를 사용해서 이 다변량선형회귀모형을 추정한 결과는 다음과 같다.

$$\boldsymbol{\alpha} = \begin{bmatrix} 0.1857 & 0.2425 & 0.2452 & 0.0681 & 0.1813 & 0.2542 & 0.1539 & -0.0161 & 0.1599 \end{bmatrix}^t \quad (8)$$

$$\boldsymbol{\beta} = \begin{bmatrix} 0.6622 & 0.6325 & 0.6574 & 0.8884 & 0.7132 & 0.8350 & 0.7265 & 0.8208 & 0.5765 \end{bmatrix}^t \quad (9)$$

식 (5)와 식 (6)의 추정벡터들은 각각 식 (8)과 식 (9)의 추정벡터들과 동일하다. 이 결과는 당연하다. 다변량선형회귀모형 (7)의 각 행이 식 (4)를 이루기 때문이다. 이렇게 추정된 선형회귀선들이 그림 3.3.4에 그려져 있다. 그림 3.3.3과 그림 3.3.4가 동일함을 확인할 수 있다. ■

**예제 3.3.2** 부분상관계수(partial correlation)는 두 변수들 사이의 상관관계를 분석하는데 중요한 통계량이다. 예를 들어, 데이터분석 관점에서 보면 혈압이 높은 사람이 연봉이 많은 것처럼 보일 것이다. 즉, 언뜻보아 혈압과 연봉은 양의 상관관계를 갖는 것 같다. 그렇다면 회사에 가서 핏대를 올리는 사람이 월급을 더 받는 경향이 있다고 결론을 내야할까? 나이를 고려한다면 나이가 많은 사람이 혈압도 높고 연봉도 높다. 그렇다면, 나이 효과를 제거해도 같은 결론을 내릴 수 있을까. 아마 그렇지 않을 것이다. 같은 나이에서 좀 더 냉정하게 사고할 수 있는 사람이 더 능력을 발휘하고 높은 연봉을 받거나 또는 혈압과 연봉은 상관성이 별로 없을 것이다. 이렇게 나이의 효과를 제거한 상관계수를 부분상관계수라고 부른다.

다변량선형회귀모형을 사용해서 부분상관계수를 정의하자. 종속벡터를  $\mathbf{y}$  그리고 설명변수들을  $x_1, x_2, \dots, x_k$ 로 하는 다변량선형회귀모형은 다음과 같다.

$$\mathbf{y} = x_1 \boldsymbol{\beta}_1 + x_2 \boldsymbol{\beta}_2 + \dots + x_k \boldsymbol{\beta}_k + \boldsymbol{\epsilon} \quad (1)$$

여기서  $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \dots, \epsilon_p]^T$ 는 평균벡터가  $\mathbf{0}$ 이고, 분산공분산행렬이  $\Sigma$ 인 오차벡터이며,

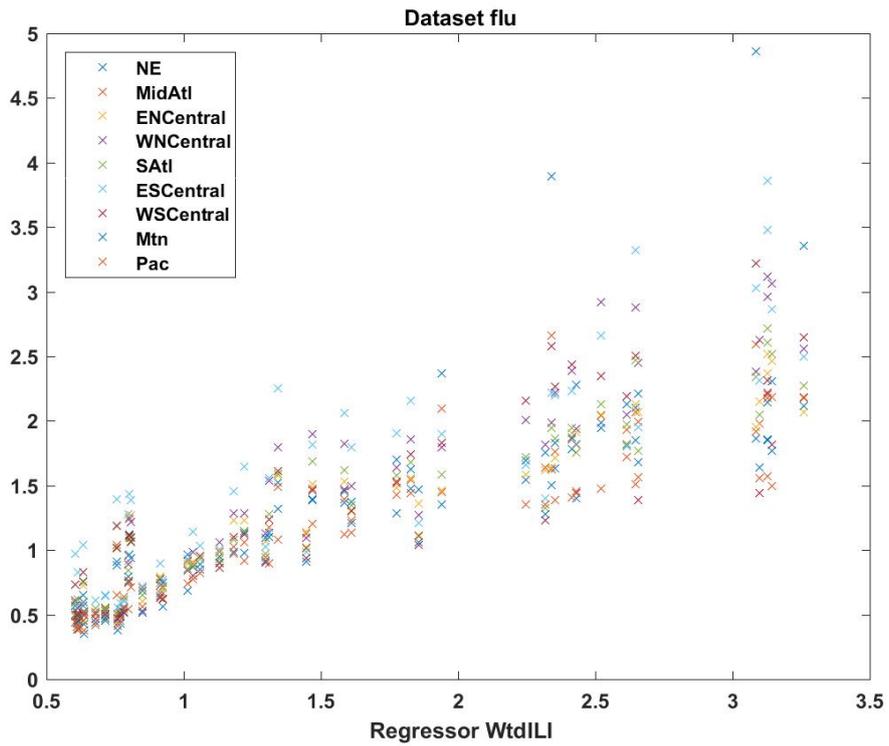


그림 3.3.1. 데이터세트 FLU

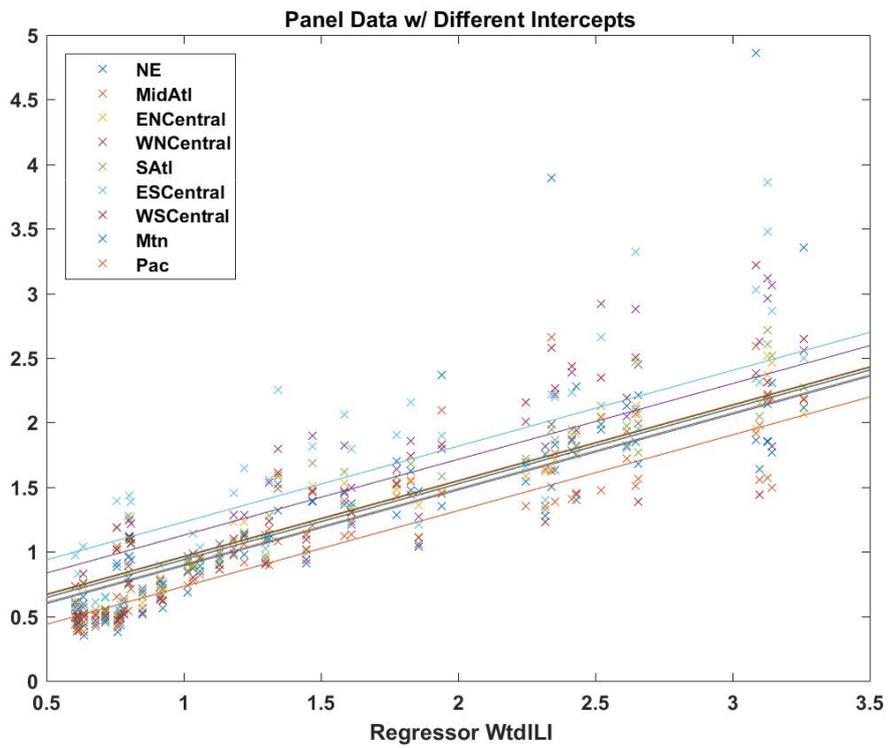


그림 3.3.2. 절편이 다른 패널데이터

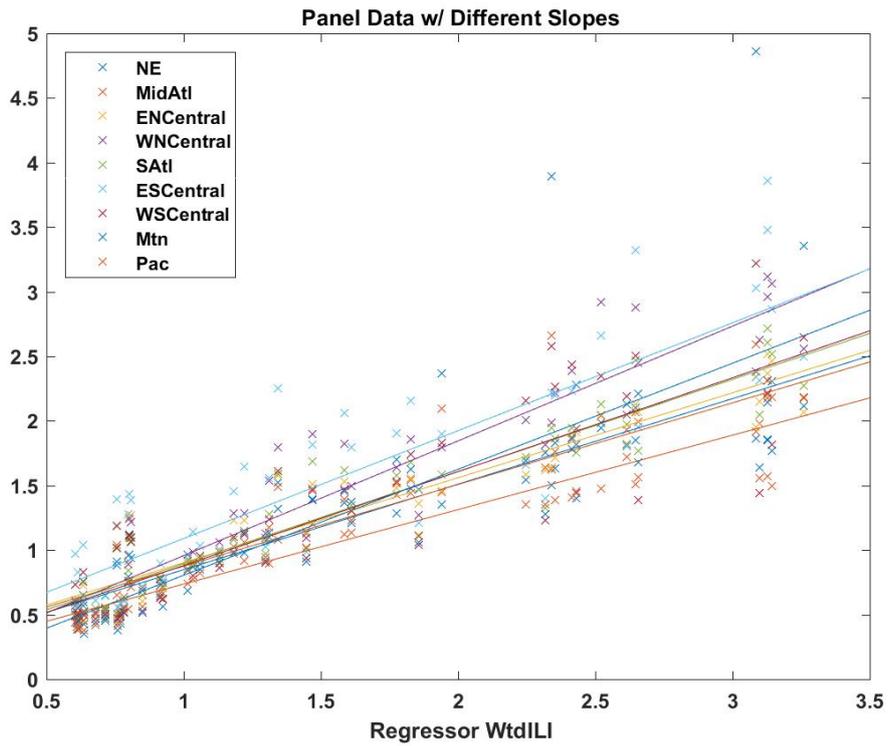


그림 3.3.3. 기울기가 다른 패널데이터

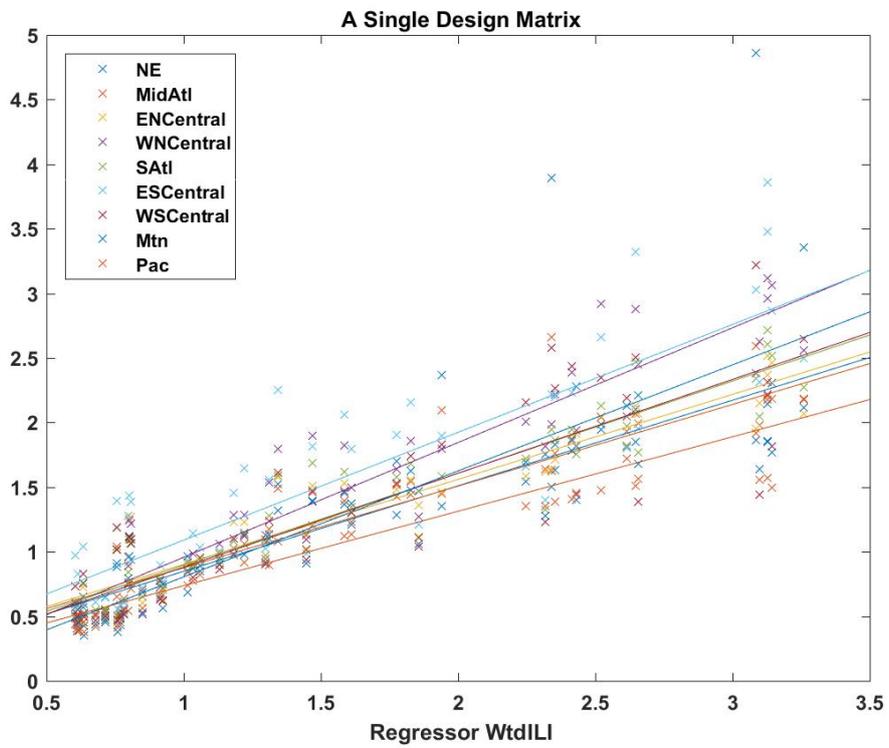


그림 3.3.4. 싱글디자인행렬 다변량선형회귀모형

$\beta_1, \beta_2, \dots, \beta_k$ 는 회귀계수벡터들이다. 이 다변량선형회귀모형의 추정식이 다음과 같다고 하자.

$$y_j = x_1\hat{\beta}_{j1} + x_2\hat{\beta}_{j2} + \dots + x_k\hat{\beta}_{jk} + \hat{\epsilon}_j, \quad (j = 1, 2, \dots, n) \quad (2)$$

잔차벡터들  $\{\hat{\epsilon}_j \mid j = 1, 2, \dots, n\}$ 의 상관계수행렬을  $\{y_j \mid j = 1, 2, \dots, n\}$ 의 부분상관계수행렬이라고 한다.

부분상관계수를 살펴보기 위해서, 다음 MATLAB 프로그램 PartialCorrelation101.m을 실행해 보자.

```

1 % -----
2 % Filename: PartialCorrelation101.m
3 % How to calculate the partial correlation using mvregress
4 % Dataset from https://kr.mathworks.com/matlabcentral/fileexchange
5 %           /21859-demo-files-for-mathworks-statistics
6 %           -webinar--oct-14--2008-?focused=6784250&tab=example
7 % Programmed by CBS
8 % -----
9 clear all, clf, close all
10
11 % Delete missing observations to Make a dataset
12 load carbig
13 X = [MPG,Acceleration,Displacement,Weight,Horsepower];
14 varNames = {'MPG'; 'Acceleration'; 'Displacement'; 'Weight'; 'Horsepower'};
15 xy = [Displacement,Horsepower,Weight,MPG,Acceleration];
16 xy(any(isnan(xy), 2),:)=[]; % Eliminate missing observations
17 Displacement = xy(:,1);
18 Horsepower = xy(:,2);
19 Weight = xy(:,3);
20 MPG = xy(:,4);
21 Acceleration = xy(:,5);
22
23 % Plotting
24 figure
25 gplotmatrix(X,[],Cylinders,['c' 'b' 'm' 'g' 'r'],[],[],false);
26 text([.08 .24 .43 .66 .83], repmat(-.1,1,5), varNames, 'FontSize',8);
27 text(repmat(-.12,1,5), [.86 .62 .41 .25 .02], varNames, 'FontSize',8, ...
28      'Rotation',90);
29 set(gca,'fontsize',11,'fontweigh','bold')
30 saveas(gcf,'PartialCorrelation101','jpg')
31
32 % Pearson correlation
33 [Rho,pvalRho] = corr([MPG,Acceleration,Displacement])
34
35 % Partial correlation
36 [PartialRho,pvalPartial] = partialcorri(MPG,[Acceleration,Displacement])
37
38 % Multivariate Linear Regression for dummy 'Displacement'
39 Y1 = [MPG,Acceleration];
40 [n,p] = size(Y1);
41 x1 = Displacement;
42 X1 = cell(n,1);
43 for ii = 1:n

```

```

44 X1{ii} = [eye(p) x1(ii)*eye(p)];
45 end
46 [beta1,Sigma1] = mvregress(X1,Y1,'algorithm','cwls');
47 PartialCorr1 = corrcov(Sigma1)
48
49 % Multivariate Linear Regression for dummy 'Accerlation'
50 Y2 = [MPG,Displacement];
51 [n,p] = size(Y2);
52 x2 = Acceleration;
53 X2 = cell(n,1);
54 for ii = 1:n
55     X2{ii} = [eye(p) x2(ii)*eye(p)];
56 end
57 [beta2,Sigma2] = mvregress(X2,Y2,'algorithm','cwls');
58 PartialCorr2 = corrcov(Sigma2)
59 save PartialCorrelation101.txt Rho pvalRho PartialRho ...
60     pvalPartial PartialCorr1 PartialCorr2 -ascii
61 % End of program
62 % -----

```

이 예제에서 사용하는 MATLAB 데이터세트 'carbig'에 저장되어 있다. 이 데이터세트는 변수들 MPG, Acceleration, Displacement, Weight, Horsepower로 구성되어 있다. 여기서 변수 MPG는 마일리지, 변수 Acceleration는 가속도, 변수 Displacement는 엔진 속 피스톤의 부피, 변수 Weight는 무게, 그리고 변수 Horsepower는 마력이다.

이 MATLAB 프로그램 PartialCorrelation101.m을 실행하면, 그림 3.3.5에 이 변수들의 산점도들이 수록된다. 그림 3.3.5에서 알 수 있듯이, 변수 MPG와 변수 Acceleration은 작은 양의 상관계수를 보이고, 변수 MPG와 변수 Displacement는 음의 상관관계를 보이며, 변수 Acceleration과 변수 Displacement는 음의 상관관계를 보인다. 이를 확인하기 위해서 MATLAB 함수 corr.m을 실행하면 다음과 같은 결과를 얻는다.

$$\text{Corr}(\text{MPG}, \text{Acceleration}) = 0.4631 \quad (3)$$

$$\text{Corr}(\text{Acceleration}, \text{Displacement}) = -0.6717 \quad (4)$$

$$\text{Corr}(\text{MPG}, \text{Displacement}) = -0.8048 \quad (5)$$

이 표본상관계수들의  $p$ 값들은 모두 0.0000이다. 즉, 이에 해당하는 각 상관계수는 유의적으로 0과 다르다.

MATLAB 함수 partialcorri.m을 사용해서 변수 MPG와 변수 Acceleration에서 변수 Displacement의 영향을 제거한 부분상관계수와 변수 MPG와 변수 Displacement에서 변수

Acceleration의 영향을 제거한 부분상관계수는 각각 다음과 같다.

$$\text{Corr}(\text{MPG}, \text{Acceleration} \mid \text{Displacement}) = -0.1766 \quad (6)$$

$$\text{Corr}(\text{MPG}, \text{Displacement} \mid \text{Acceleration}) = -0.7520 \quad (7)$$

이 표본상관계수들의  $p$ 값들은 각각 0.0922와 0.0000이다. 즉, 유의수준 0.05에서 첫 번째 부분상관계수는 0과 유의적으로 다르지 않고, 두 번째 부분상관계수는 0과 유의적으로 다르다. 따라서, 변수 MPG와 변수 Acceleration의 상관계수는 유의적으로 0과 다르지만, 변수 Displacement의 영향을 제거한 변수 MPG와 변수 Acceleration의 부분상관계수는 0과 다르지 않다. 즉, 변수 Displacement를 고려하지 않으면 변수 MPG와 변수 Acceleration은 양의 상관관계를 이루지만, 변수 Displacement를 고려하면 변수 MPG와 변수 Acceleration은 양의 상관관계를 이룬다고 할 수 없다.

MATLAB함수 mvregress.m을 사용해서, 변수 MPG와 변수 Acceleration에서 변수 Displacement의 영향을 제거한 부분상관계수추정값이 -0.1766임을 확인할 수 있다. 또한, 같은 방법으로 변수 MPG와 변수 Displacement에서 변수 Acceleration의 영향을 제거한 부분상관계수추정값이 -0.7520임을 확인할 수 있다. ■

**예제 3.3.3** SAS/STAT의 PROC REG를 사용해서 다변량선형모형을 분석하는 예로서 다음 SAS프로그램을 수행해보자.

```

1 * -----*
2 | Multivariate Multiple Regression
3 | Programmed by CBS
4 | Dataset is an artificail one made by CBS.
5 * -----*;
6 TITLE 'MULTIVARIATE MULTIPLE REGRESSION';
7 DATA MULTI2REG;
8     INPUT  ObsNo  X1-X3  Y1-Y3;
9     1      199    213.283  100.000    1239.39    -1474.32    537.692
10    2      198    216.462  100.000    1248.24    -1463.15    560.719
11    3      197    219.537  100.000    1251.94    -1454.58    494.979
12    4      196    222.504  100.000    1260.95    -1444.30    506.805
13    5      195    225.365  100.000    1265.69    -1435.29    532.535
14    6      194    228.120  100.000    1273.73    -1424.83    555.231
15    7      193    230.771  100.000    1279.60    -1415.23    570.780
16    8      192    233.319  100.000    1284.46    -1405.13    569.721
17    9      191    235.768  100.000    1291.53    -1396.90    560.003
18    10     190    238.122  100.000    1295.87    -1387.05    570.373
19    11     189    240.387  100.000    1301.58    -1376.77    638.838
20    12     188    242.565  100.000    1303.98    -1369.23    576.066
21    13     187    244.665  100.000    1309.82    -1358.52    633.023
22    14     186    246.691  100.000    1311.97    -1351.76    609.357

```

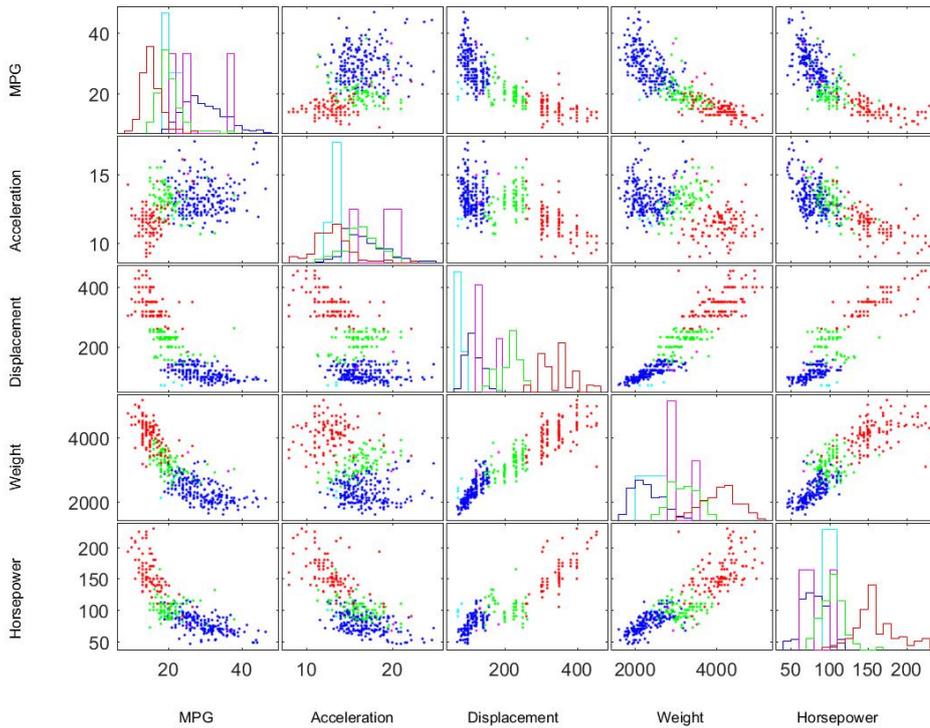


그림 3.3.5. 데이터세트 CARBIG의 산점도들

23	15	185	248.650	100.000	1316.95	-1341.47	643.933
24	16	184	250.549	100.000	1320.35	-1331.53	682.194
25	17	183	252.394	100.000	1325.57	-1321.93	685.426
26	18	182	254.192	100.000	1327.05	-1314.22	658.414
27	19	181	255.951	100.000	1329.16	-1306.98	668.311
28	20	180	257.676	100.000	1334.46	-1298.45	686.457
29	21	179	259.373	100.000	1338.84	-1287.67	709.615
30	22	178	261.050	100.000	1339.10	-1281.09	677.734
31	23	177	262.711	100.000	1343.38	-1271.10	742.429
32	24	176	264.362	100.000	1344.92	-1262.26	697.858
33	25	175	266.006	100.001	1348.64	-1253.59	756.798
34	26	174	267.649	100.001	1350.55	-1244.49	721.693
35	27	173	269.292	100.001	1355.82	-1237.24	731.112
36	28	172	270.939	100.001	1358.77	-1228.08	746.745
37	29	171	272.590	100.001	1360.49	-1218.26	791.422
38	30	170	274.247	100.001	1363.85	-1212.25	715.810
39	31	169	275.910	100.002	1367.55	-1202.31	764.490
40	32	168	277.577	100.002	1369.83	-1192.81	754.039
41	33	167	279.246	100.002	1372.69	-1184.86	751.371
42	34	166	280.915	100.003	1375.69	-1178.44	772.469
43	35	165	282.580	100.003	1379.92	-1166.73	820.098
44	36	164	284.236	100.004	1382.36	-1157.98	818.240
45	37	163	285.879	100.004	1386.40	-1149.79	834.116
46	38	162	287.501	100.005	1388.88	-1139.86	828.090
47	39	161	289.096	100.006	1389.89	-1132.71	837.483
48	40	160	290.657	100.007	1393.53	-1123.42	818.780
49	41	159	292.176	100.008	1394.75	-1116.66	789.688
50	42	158	293.642	100.010	1396.64	-1107.54	815.167
51	43	157	295.049	100.012	1401.17	-1098.34	879.008
52	44	156	296.385	100.014	1402.05	-1090.00	887.799

53	45	155	297.642	100.016	1403.96	-1081.91	884.821
54	46	154	298.808	100.019	1406.62	-1074.61	899.420
55	47	153	299.875	100.023	1407.20	-1066.04	905.144
56	48	152	300.832	100.027	1405.88	-1060.00	922.786
57	49	151	301.670	100.032	1406.70	-1051.64	934.244
58	50	150	302.377	100.038	1407.19	-1044.06	852.595
59	51	149	302.946	100.045	1407.89	-1035.63	854.448
60	52	148	303.367	100.053	1407.71	-1026.34	898.523
61	53	147	303.632	100.063	1406.49	-1019.73	826.725
62	54	146	303.732	100.074	1405.00	-1013.63	898.622
63	55	145	303.660	100.087	1400.85	-1006.52	850.664
64	56	144	303.411	100.103	1399.64	-999.69	909.132
65	57	143	302.977	100.122	1395.65	-992.51	859.386
66	58	142	302.355	100.144	1392.02	-987.39	922.789
67	59	141	301.541	100.170	1389.23	-981.19	897.198
68	60	140	300.531	100.201	1382.83	-974.77	905.877
69	61	139	299.325	100.237	1378.61	-968.37	886.986
70	62	138	297.921	100.280	1371.13	-964.80	831.672
71	63	137	296.320	100.331	1364.44	-957.39	852.421
72	64	136	294.522	100.391	1358.39	-951.87	903.908
73	65	135	292.532	100.462	1350.77	-948.39	917.692
74	66	134	290.352	100.546	1341.44	-943.49	842.546
75	67	133	287.987	100.645	1332.48	-940.66	923.574
76	68	132	285.444	100.762	1322.40	-936.02	887.563
77	69	131	282.728	100.900	1312.52	-931.96	865.029
78	70	130	279.848	101.063	1300.87	-928.59	805.137
79	71	129	276.812	101.256	1291.82	-925.54	852.043
80	72	128	273.630	101.484	1279.99	-921.91	801.422
81	73	127	270.312	101.753	1268.03	-920.58	789.958
82	74	126	266.869	102.071	1257.98	-914.33	798.386
83	75	125	263.314	102.447	1245.54	-915.66	735.397
84	76	124	259.657	102.891	1234.43	-912.62	781.367
85	77	123	255.911	103.415	1222.36	-909.23	743.478
86	78	122	252.090	104.034	1207.84	-909.35	743.532
87	79	121	248.206	104.766	1196.65	-909.28	714.882
88	80	120	244.272	105.630	1185.04	-906.23	723.324
89	81	119	240.303	106.651	1173.04	-907.87	707.861
90	82	118	236.310	107.858	1160.56	-910.00	701.413
91	83	117	232.307	109.283	1150.16	-908.89	728.898
92	84	116	228.306	110.966	1140.71	-909.99	696.968
93	85	115	224.320	112.955	1130.22	-912.95	707.907
94	86	114	220.360	115.305	1120.98	-918.02	661.052
95	87	113	216.438	118.080	1112.37	-924.05	628.065
96	88	112	212.565	121.359	1106.62	-928.90	647.528
97	89	111	208.749	125.233	1101.13	-937.88	671.320
98	90	110	205.001	129.810	1096.33	-949.00	597.758
99	91	109	201.328	135.216	1093.15	-963.81	602.509
100	92	108	197.737	141.603	1094.17	-978.60	612.605
101	93	107	194.236	149.148	1095.22	-996.85	582.083
102	94	106	190.829	158.061	1102.51	-1020.19	613.329
103	95	105	187.521	168.591	1112.41	-1047.62	652.241
104	96	104	184.316	181.031	1123.80	-1080.97	581.212
105	97	103	181.215	195.727	1140.94	-1122.83	547.012
106	98	102	178.221	213.088	1162.99	-1170.60	565.897
107	99	101	175.334	233.597	1197.62	-1225.88	585.277
108	100	100	172.552	257.827	1234.16	-1295.36	562.882
109							
110							
111							
112							
113							

```

110 RUN;
111 PROC CORR DATA = MULTI2REG;
112 VAR X1-X3 Y1-Y3;
113 RUN;

```

```

114 PROC REG DATA = MULTI2REG;
115 MODEL Y1 Y2 Y3 = X1 X2 X3;
116 MMR1: MTEST X1, X2 / DETAILS PRINT;
117 MMR2: MTEST Y1-Y2+Y3 / DETAILS PRINT;
118 MMR3: MTEST Y1-Y2, Y2-Y3, Y3, X1-X2, X2-X3 / DETAILS PRINT;
119 RUN;
120 *-----*;

```

이 예제는 최병선 [4, 제12.4절]에서 인용한 것이다.

첫 번째 DATA 스텝을 수행하면, 변수들 ObsNo, X1, X2, X3, Y1, Y2, Y3를 읽어서 SAS 데이터세트 MULTI2REG 만든다.

첫 번째 PROC CORR는 종속벡터와 설명변수들의 요약통계량들과 상관계수들을 출력하기 위한 것이다. 이 PROC CORR를 수행한 결과물은 아래와 같다.

The CORR \Pr ocedure

6 Variables: X1 X2 X3 Y1 Y2 Y3

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
X1	100	149.50000	29.01149	14950	100.00000	199.00000
X2	100	258.71395	36.76001	25871	172.55200	303.73200
X3	100	110.28058	27.94369	11028	100.00000	257.82700
Y1	100	1297	99.28150	129675	1093	1408
Y2	100	-1114	175.38236	-111351	-1474	-906.23000
Y3	100	738.79419	121.29394	73879	494.97900	934.24400

Pearson Correlation Coefficients,  $N = 100$

\Pr ob > |r| under H0: Rho=0

	X1	X2	X3	Y1	Y2	Y3
X1	1.00000	0.26361	-0.56344	0.56050	-0.83364	-0.14750
		0.0080	<.0001	<.0001	<.0001	0.1431

X2	0.26361 0.0080	1.00000	-0.69044 <.0001	0.87594 <.0001	0.23399 0.0191	0.88663 <.0001
X3	-0.56344 <.0001	-0.69044 <.0001	1.00000	-0.53367 <.0001	0.03014 0.7659	-0.42055 <.0001
Y1	0.56050 <.0001	0.87594 <.0001	-0.53367 <.0001	1.00000	-0.21067 0.0354	0.66233 <.0001
Y2	-0.83364 <.0001	0.23399 0.0191	0.03014 0.7659	-0.21067 0.0354	1.00000	0.55775 <.0001
Y3	-0.14750 0.1431	0.88663 <.0001	-0.42055 <.0001	0.66233 <.0001	0.55775 <.0001	1.00000

첫 번째 PROC REG는 설명변수들이 X1, X2, X3이고 종속변수들이 Y1, Y2, Y3인 다중선형회귀모형을 추정하기 위한 것이다.

첫 번째 MODEL스테이트먼트의 등호의 좌변에 종속변수들 Y1, Y2, Y3를 그리고 우변에 설명변수들 X1, X2, X3를 지정하였다. 이는 다음 다변량선형회귀모형을 추정하기 위한 것이다.

$$y_1 = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \beta_{13}x_3 + \epsilon_1 \quad (1)$$

$$y_2 = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}x_3 + \epsilon_2 \quad (2)$$

$$y_3 = \beta_{30} + \beta_{31}x_1 + \beta_{32}x_2 + \beta_{33}x_3 + \epsilon_3 \quad (3)$$

이 MODEL스테이트먼트를 수행한 결과물은 아래와 같다.

```

The REG \Pr ocedure
Model: MODEL1
Dependent Variable: Y1

Number of Observations Read      100
Number of Observations Used      100

Analysis of Variance
    
```

Source	DF	Sum of Squares	Mean Square	F Value	\Pr > F
Model	3	975735	325245	348578	<.0001
Error	96	89.57400	0.93306		
Corrected Total	99	975825			

Root MSE	0.96595	R-Square	0.9999
Dependent Mean	1296.74680	Adj R-Sq	0.9999
Coeff Var	0.07449		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	\Pr >  t
Intercept	1	1.95479	1.79904	1.09	0.2799
X1	1	2.00053	0.00414	482.90	<.0001
X2	1	2.99755	0.00373	802.78	<.0001
X3	1	1.99677	0.00574	348.15	<.0001

The REG Procedure

Model: MODEL1

Dependent Variable: Y2

Number of Observations Read 100  
 Number of Observations Used 100

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	\Pr > F
Model	3	3045053	1015018	1142992	<.0001
Error	96	85.25139	0.88804		
Corrected Total	99	3045138			

Root MSE                    0.94236    R-Square        1.0000  
 Dependent Mean        -1113.51310    Adj R-Sq       1.0000  
 Coeff Var                -0.08463

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	\Pr >  t
Intercept	1	5.33757	1.75509	3.04	0.0030
X1	1	-7.00062	0.00404	-1732.2	<.0001
X2	1	0.99893	0.00364	274.22	<.0001
X3	1	-2.99867	0.00560	-535.93	<.0001

The REG Procedure

Model: MODEL1

Dependent Variable: Y3

Number of Observations Read        100  
 Number of Observations Used       100

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	\Pr > F
Model	3	1377437	459146	557.44	<.0001
Error	96	79073	823.67289		
Corrected Total	99	1456510			

Root MSE                    28.69970    R-Square        0.9457  
 Dependent Mean        738.79419    Adj R-Sq       0.9440  
 Coeff Var                3.88467

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	\Pr >  t
----------	----	--------------------	----------------	---------	----------

Intercept	1	27.09353	53.45180	0.51	0.6134
X1	1	-1.54776	0.12309	-12.57	<.0001
X2	1	3.46704	0.11094	31.25	<.0001
X3	1	0.41819	0.17041	2.45	0.0159

이 결과물에서 알 수 있듯이, 추정모형은 다음과 같다.

$$y_1 = 1.9548 + 2.0005x_1 + 2.9976x_2 + 1.9968x_3 + \hat{\epsilon}_1 \quad (4)$$

$$y_2 = 5.3376 - 7.0006x_1 - 0.9989x_2 - 2.9987x_3 + \hat{\epsilon}_2 \quad (5)$$

$$y_3 = 27.0935 - 1.5478x_1 + 3.4670x_2 + 0.4182x_3 + \hat{\epsilon}_3 \quad (6)$$

종속변수가 Y1인 선형회귀모형의 모형제곱합과 잔차제곱합 그리고 이들의 자유도들은 각각 다음과 같다.

$$SSM = 975735, \quad d.f. = 3 \quad (7)$$

$$SSE = 89.574, \quad d.f. = 96 \quad (8)$$

또한, 검정통계량  $F$  값은 348,578 이고 이에 해당하는  $p$  값은 0.0001 미만이다. 따라서, 설명변수들 X1, X2, X3가 종속변수 Y1에 유의적인 영향을 미친다고 할 수 있다. 종속변수가 Y2인 선형회귀모형의 모형제곱합과 잔차제곱합 그리고 이들의 자유도들은 각각 다음과 같다.

$$SSM = 3045053, \quad d.f. = 3 \quad (9)$$

$$SSE = 85.251, \quad d.f. = 96 \quad (10)$$

또한, 검정통계량  $F$  값은 1,142,992 이고 이에 해당하는  $p$  값은 0.0001 미만이다. 따라서, 설명변수들 X1, X2, X3가 종속변수 Y2에 유의적인 영향을 미친다고 할 수 있다. 종속변수가 Y3인 선형회귀모형의 모형제곱합과 잔차제곱합 그리고 이들의 자유도들은 각각 다음과 같다.

$$SSM = 1377437, \quad d.f. = 3 \quad (11)$$

$$SSE = 79073, \quad d.f. = 96 \quad (12)$$

또한, 검정통계량  $F$  값은 557.44 이고 이에 해당하는  $p$  값은 0.0001 미만이다. 따라서, 설명변

수들 X1, X2, X3가 종속변수 Y3에 유의적인 영향을 미친다고 할 수 있다.

첫 번째 MTEST스테이트먼트 MMR1은 다음 가설들을 검정하기 위한 것이다.

$$H_0 : \beta_{11} = \beta_{21} = \beta_{31} = 0 \text{ and } \beta_{12} = \beta_{22} = \beta_{32} = 0 \quad \text{대} \quad H_1 : H_0 \text{가 틀림}$$

즉, 설명변수들 X1과 X2에 해당하는 회귀계수들이 모두 0이라는 귀무가설을 검정하기 위한 것이다. 이 가설들을 다음과 같이 쓸 수 있다.

$$H_0 : [LB - \mathbf{c}\mathbf{j}] M = O \quad \text{대} \quad H_1 : [LB - \mathbf{c}\mathbf{j}] M \neq O \quad (13)$$

여기서 각 행렬은 다음과 같다.

$$L = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} \beta_{10} & \beta_{20} & \beta_{30} \\ \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \\ \beta_{13} & \beta_{23} & \beta_{33} \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{j} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}, \quad M = I_3 \quad (14)$$

이 MTEST스테이트먼트를 실행한 결과물은 다음과 같다.

```

The REG \Pr ocedure
      Model: MODEL1
      Multivariate Test: MMR1

      L Ginv(X'X) L'      LB-cj

0.0000183933      3.4790993E-6      2.000534072      -7.000617567      -1.547764986
3.4790993E-6      0.0000149428      2.9975468424      0.9989296152      3.4670446059

      Inv(L Ginv(X'X) L')      Inv ()(LB-cj)

56872.291442      -13241.40706      74083.218854      -411368.3962      -133933.4903
-13241.40706      70004.600005      183352.18171      162627.69503      263203.65705

      Error Matrix (E)
    
```

89.574003578	43.515620435	1023.1849686
43.515620435	85.251391068	792.27954595
1023.1849686	792.27954595	79072.597536

Hypothesis Matrix (H)

697812.75683	-335472.359	521026.7804
-335472.359	3042286.4415	1200539.0727
521026.7804	1200539.0727	1119836.3861

Hypothesis + Error Matrix (T)

697902.33083	-335428.8434	522049.96537
-335428.8434	3042371.6929	1201331.3523
522049.96537	1201331.3523	1198908.9837

Eigenvectors

-0.000319	0.000519	-0.000000972
0.001204	0.000288	-0.000016777
-0.003577	-0.001822	0.003618

Eigenvalues

0.999982  
 0.999860  
 9.645995E-16

Multivariate Statistics and F Approximations

S=2 M=0 N=46

Statistic	Value	F Value	Num DF	Den DF	\Pr > F
Wilks' Lambda	0.000000	630672	6	188	<.0001
Pillai's Trace	1.999842	400655	6	190	<.0001
Hotelling-Lawley Trace	64040.000705	998101	6	123.57	<.0001
Roy's Greatest Root	56923.296598	1802571	3	95	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

이 결과물은 MTEST 스테이트먼트의 옵션 DETAILS와 옵션 PRINT에 의해서 출력된 것이다. 옵션 PRINT에 의해서 출력된 내용은 다음과 같다.

$$E = \begin{bmatrix} 89.574 & 43.516 & 1023.185 \\ 43.516 & 85.251 & 792.280 \\ 1023.185 & 792.280 & 79072.598 \end{bmatrix} \quad (15)$$

$$H = \begin{bmatrix} 697812.757 & -335472.359 & 521026.780 \\ -335472.359 & 3042286.442 & 1200539.073 \\ 521026.780 & 1200539.073 & 1119836.386 \end{bmatrix} \quad (16)$$

또한, 오차벡터의 분산공분산행렬의 추정행렬은 다음과 같다.

$$\hat{S} = \frac{1}{100 - 4} E \quad (17)$$

옵션 DETAILS에 의해서 출력된 내용은 다음과 같다.

$$L(X'X)^{-1}L^T = \begin{bmatrix} 0.184E - 4 & 3.479E - 6 \\ 3.479E - 6 & 0.149E - 4 \end{bmatrix} \quad (18)$$

$$L\hat{B} - \mathbf{c}j = \begin{bmatrix} 2.001 & -7.001 & -1.548 \\ 2.998 & 0.999 & 3.467 \end{bmatrix} \quad (19)$$

$$(L(X'X)^{-1}L^T)^{-1} = \begin{bmatrix} 56872.291 & -13241.407 \\ -13241.407 & 70004.600 \end{bmatrix} \quad (20)$$

$$(L(X^T X)^- L^T)^{-1}(L\hat{B} - \mathbf{c}j) = \begin{bmatrix} 74083.219 & -411368.396 & -133933.490 \\ 183352.182 & 162627.695 & 263203.657 \end{bmatrix} \quad (21)$$

$$T = \begin{bmatrix} 697902.331 & -335428.843 & 522049.965 \\ -335428.843 & 3042371.693 & 1201331.352 \\ 522049.965 & 1201331.352 & 1198908.984 \end{bmatrix} \quad (22)$$

행렬  $[E + H]^{-1}H$ 의 고유값들은 1.000, 1.000, 0.000이다. 또한, 행렬  $[E + H]^{-1}H$ 의 고유벡터들로 이루어진 고유행렬은 다음과 같다.

$$\begin{bmatrix} -0.000319 & 0.000519 & -0.000000972 \\ 0.001204 & 0.000288 & -0.000016777 \\ -0.003577 & -0.001822 & 0.003618 \end{bmatrix} \quad (23)$$

이 경우에는 행렬  $M$ 은 단위행렬이어서 출력을 하지 않았다. 만약 행렬  $M$ 이 단위행렬이 아니면, 옵션 DETAILS는 이 행렬을 출력하도록 한다. 각 검정통계값과 이에 해당하는  $F$ 값과  $p$ 값은 다음과 같다.

$$\text{Wilks } \Lambda : \quad \Lambda = 0.000, \quad F_W = 630672, \quad p < 0.0001 \quad (24)$$

$$\text{Pillai 트레이스} : \quad V = 2.000, \quad F_P = 400655, \quad p < 0.0001 \quad (25)$$

$$\text{Hotelling - Lawley 트레이스} : \quad U = 64040.001, \quad F_H = 998101, \quad p < 0.0001 \quad (26)$$

$$\text{Roy 최대고유값} : \quad \Theta = 56923.297, \quad F_R = 1802571, \quad p < 0.0001 \quad (27)$$

따라서, 각 검정통계량은 유의수준 0.01에서 귀무가설을 기각한다.

두 번째 MTEST스테이트먼트 MMR2는 다음 가설들을 검정하기 위한 것이다.

$$H_0 : \beta_{11} - \beta_{21} + \beta_{31} = 0 \text{ and } \beta_{12} - \beta_{22} + \beta_{32} = 0 \text{ and } \beta_{13} - \beta_{23} + \beta_{33} = 0$$

대  $H_1 : H_0$ 가 틀림

이 가설들을 다음과 같이 쓸 수 있다.

$$H_0 : [LB - \mathbf{c}j] M = O \quad \text{대} \quad H_1 : [LB - \mathbf{c}j] M \neq O \quad (28)$$

여기서 각 행렬은 다음과 같다.

$$L = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} \beta_{10} & \beta_{20} & \beta_{30} \\ \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \\ \beta_{13} & \beta_{23} & \beta_{33} \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad M = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \quad (29)$$

$$\mathbf{j} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \quad (30)$$

이 MTEST스테이트먼트를 실행한 결과물은 다음과 같다.

```

The REG \Pr ocedure
Model: MODEL1
Multivariate Test: MMR2

M, Across Dependents

          Y1          Y2          Y3
1.000000      -1.000000      1.000000

          L Ginv(X'X) L'   LB-cj

0.0000183933    3.4790993E-6    0.0000139195    2.000534072    -7.000617567
-1.547764986
3.4790993E-6    0.0000149428    0.0000156075    2.9975468424    0.9989296152
3.4670446059
0.0000139195    0.0000156075    0.0000352544    1.996770294    -2.998672746
0.4181943907

          Inv(L Ginv(X'X) L')   Inv ()(LB-cj)

83325          27831.6055          -45220.606          159825.88          -419922.845
    
```

```

-51385.1035
 27831.6055   133778.49889   -70213.85028   316484.46155   149345.23728
391376.20015
 -45220.606   -70213.85028   77304.11434   -146576.1092   14623.73347
-141115.5333
    
```

Error Matrix (E)

79622.202535

Hypothesis Matrix (H)

5354128.9679

Hypothesis + Error Matrix (T)

5433751.1704

Eigenvectors

0.000429

Eigenvalues

0.985347

Multivariate Statistics and Exact F Statistics

S=1 M=0.5 N=47

Statistic	Value	F Value	Num DF	Den DF	\Pr > F
Wilks' Lambda	0.01465327	2151.81	3	96	<.0001
Pillai's Trace	0.98534673	2151.81	3	96	<.0001
Hotelling-Lawley Trace	67.24417056	2151.81	3	96	<.0001
Roy's Greatest Root	67.24417056	2151.81	3	96	<.0001

세 번째 MTEST스테이트먼트 MMR3는 다음 가설들을 검정하기 위한 것이다.

$$H_0 : \beta_{11} - \beta_{12} = \beta_{21} - \beta_{22} = \beta_{31} - \beta_{32} = 0 \text{ and } \beta_{12} - \beta_{13} = \beta_{22} - \beta_{23} = \beta_{32} - \beta_{33} = 0$$

대  $H_1 : H_0$ 가 틀림

이 가설들을 다음과 같이 쓸 수 있다.

$$H_0 : [LB - \mathbf{c}\mathbf{j}]M = O \text{ 대 } H_1 : [LB - \mathbf{c}\mathbf{j}]M \neq O \tag{31}$$

여기서 각 행렬들은 다음과 같다.

$$L = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} \beta_{10} & \beta_{20} & \beta_{30} \\ \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \\ \beta_{13} & \beta_{23} & \beta_{33} \end{bmatrix}, \quad M = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{32}$$

$$\mathbf{j} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \tag{33}$$

이 MTEST스테이트먼트를 실행한 결과물은 다음과 같다.

```

The REG \Pr ocedure
Model: MODEL1
Multivariate Test: MMR3

M, Across Dependents

      Y1          Y2          Y3
-----
1.000000      -1.000000          0
      0          1.000000      -1.000000
      0          0          1.000000

L Ginv(X'X) L'   LB-cj

0.0000263779      -9.775799E-6      -0.99701277      -7.999547182      -5.014809592
    
```

-9.775799E-6      0.0000189823      1.0007765484      3.9976023611      3.0488502152

Inv(L Ginv(X'X) L')      Inv()(LB-cj)

46852.799435      24128.97943      -22565.12262      -278343.1145      -161392.2239  
 24128.97943      65106.969543      41100.627628      67250.865774      77499.160615

Error Matrix (E)

87.794153777      -272.6411933      230.90542266  
 -272.6411933      77573.289835      -78280.31799  
 230.90542266      -78280.31799      79072.597536

Hypothesis Matrix (H)

1869461.8988      -788240.2818      -1362406.086  
 -788240.2818      339344.62781      555240.9323  
 -1362406.086      555240.9323      1045634.6049

Hypothesis + Error Matrix (T)

1869549.693      -788512.9229      -1362175.181  
 -788512.9229      416917.91764      476960.61431  
 -1362175.181      476960.61431      1124707.2024

Eigenvectors

0.000215      -0.000450      -0.000446  
 0.007979      0.006948      0.006864  
 0.003912      0.005694      0.002073

Eigenvalues

0.999972  
 0.995598  
 8.324193E-15

Statistic	S=2	M=0	N=46	Num DF	Den DF	\Pr > F
	Value	F Value				
Wilks' Lambda	0.000000	88912.7		6	188	<.0001
Pillai's Trace	1.995570	14265.1		6	190	<.0001
Hotelling-Lawley Trace	35693.499385	556304		6	123.57	<.0001
Roy's Greatest Root	35467.314307	1123132		3	95	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.



### 3.3.2 베이저안 통계추론

이 소절에서는 베이저안추정법을 사용해서 다변량선형회귀모형을 추정하기로 하자.

다변량선형회귀모형의 오차벡터  $\epsilon$ 이 다변량정규확률분포  $\mathcal{N}_p(\mathbf{0}, \Sigma)$ 를 따른다고 하자. 식 (3.3.27)에서 알 수 있듯이, 우도함수  $L(B, \Sigma)$ 는 다음과 같다.

$$L(B, \Sigma) \propto |\Sigma|^{-n/2} \exp\left(-\frac{1}{2}\text{Tr}\left(\hat{S}\Sigma^{-1}\right)\right) \cdot \exp\left(-\frac{1}{2}\text{Tr}\left([\hat{B} - B]^T X^T X [\hat{B} - B] \Sigma^{-1}\right)\right) \quad (3.3.43)$$

다음 식들이 성립한다.

$$\text{Tr}(A^T B) = (\text{vec}(A))^T \text{vec}(B) \quad (3.3.44)$$

$$\text{vec}(ABC) = (C^T \otimes A) \text{vec}(B) \quad (3.3.45)$$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1} \quad (3.3.46)$$

따라서, 다음 식들이 성립한다.

$$\begin{aligned}
 & \text{Tr} \left( [\hat{B} - B]^T X^T X [\hat{B} - B] \Sigma^{-1} \right) \\
 &= \left( \text{vec}(\hat{B} - B) \right)^T \text{vec} \left( X^T X [\hat{B} - B] \Sigma^{-1} \right) \\
 &= \left( \text{vec}(\hat{B} - B) \right)^T (\Sigma^{-1} \otimes X^T X) \text{vec}(\hat{B} - B) \quad (3.3.47)
 \end{aligned}$$

여기서 첫 번째 등호는 식 (3.3.44)에 의해서, 그리고 두 번째 등호는 식 (3.3.45)에 의해서 성립한다. 다음 벡터들을 정의하자.

$$\boldsymbol{\beta} \doteq \text{vec}(B), \quad \hat{\boldsymbol{\beta}} \doteq \text{vec}(\hat{B}) \quad (3.3.48)$$

식 (3.3.48)을 식 (3.3.47)에 대입하면, 다음 식을 얻는다.

$$\text{Tr} \left( [\hat{B} - B]^T X^T X [\hat{B} - B] \Sigma^{-1} \right) = [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]^T (\Sigma^{-1} \otimes X^T X) [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}] \quad (3.3.49)$$

식 (3.3.49)를 식 (3.3.43)에 대입하면, 우도함수  $L(B, \Sigma)$ 가 다음과 같음을 알 수 있다.

$$\begin{aligned}
 L(B, \Sigma) &\propto |\Sigma|^{-n/2} \exp \left( -\frac{1}{2} \text{Tr}(\hat{S} \Sigma^{-1}) \right) \\
 &\quad \cdot \exp \left( -\frac{1}{2} [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]^T (\Sigma^{-1} \otimes X^T X) [\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}] \right) \quad (3.3.50)
 \end{aligned}$$

모수  $(B, \Sigma)$ 의 사전확률분포가 다음과 같다고 하자.

$$\pi(B, \Sigma) = \pi(B | \Sigma) \pi(\Sigma) \quad (3.3.51)$$

$$\Sigma \stackrel{d}{\sim} \mathcal{IW}_p(V_a, n_a), \quad (3.3.52)$$

$$\boldsymbol{\beta} | \Sigma \stackrel{d}{\sim} \mathcal{N}_{kp}(\boldsymbol{\beta}_a, \Sigma \otimes A_a^{-1}) \quad (3.3.53)$$

여기서  $\boldsymbol{\beta}_a \doteq \text{vec}(B_a)$ 이다. 식 (3.3.52)와 식 (3.3.53)을 식 (3.3.51)에 대입하면, 사전확률밀도함수가 다음과 같음을 알 수 있다.

$$\begin{aligned}
 \pi(B, \Sigma) &\propto \Sigma \otimes A_a^{-1/2} \exp \left( -\frac{1}{2} [\boldsymbol{\beta} - \boldsymbol{\beta}_a]^T (\Sigma^{-1} \otimes A_a) [\boldsymbol{\beta} - \boldsymbol{\beta}_a] \right) \\
 &\quad \cdot |\Sigma|^{-[n_a+p+1]/2} \exp \left( -\frac{1}{2} \text{Tr}(V_a \Sigma^{-1}) \right) \quad (3.3.54)
 \end{aligned}$$

식 (3.3.50)과 식 (3.3.54)에서 알 수 있듯이, 사후확률밀도함수는 다음과 같다.

$$\pi(B, \Sigma | Y, X) \propto \Sigma^{-[n_b+p+1]/2} \exp\left(-\frac{1}{2}\text{Tr}(V_b \Sigma^{-1})\right) \cdot \exp\left(-\frac{1}{2}[\beta - \beta_b]^T (\Sigma^{-1} \otimes A_b) [\beta - \beta_b]\right) \quad (3.3.55)$$

여기서 행렬들과 벡터들은 각각 다음과 같다.

$$A_b \doteq X^T X + A_a, \quad B_b \doteq A_b^{-1} [X^T X \hat{B} + A_a B_a]^{-1} \quad (3.3.56)$$

$$\beta_b \doteq \text{vec}(B_b), \quad n_b \doteq n_a + n \quad (3.3.57)$$

$$V_b \doteq V_a + [Y - X B_1]^T [Y - X B_b] + [B_b - B_a]^T A_a [B_b - B_a] \quad (3.3.58)$$

즉, 모수  $(B, \Sigma)$ 의 사후확률분포는 다음과 같다.

$$\Sigma | (X, Y) \stackrel{d}{\sim} \mathcal{IW}_p(V_b, n_b), \quad \beta | (\Sigma, X, Y) \stackrel{d}{\sim} \mathcal{N}_{kp}(\beta_b, \Sigma \otimes A_b^{-1}) \quad (3.3.59)$$

지금까지 내용을 바탕으로, 다변량선형회귀모형의 모수들  $(\beta, \Sigma)$ 의 사후확률분포에서 샘플링하는 Gibbs샘플러를 다음과 같이 요약할 수 있다.

**(1단계)** 초기모수  $(\beta^{(0)}, \Sigma^{(0)})$ 를 선택한다.

**(2단계)** 각  $j(= 1, 2, \dots)$ 에 대해서 주어진  $Y$ 에 대한 조건부샘플  $(\beta^{(j)}, \Sigma^{(j)})$ 를 발생시킨다.

우선, 식 (3.3.54)의 사후조건부확률분포를 사용해서  $\beta^{(j)} | (\Sigma^{(j-1)}, X, Y)$ 를 발생시킨다. 식 (3.3.59)의 사후조건부확률분포를 사용해서  $\Sigma^{(j)} | (\beta^{(j)}, X, Y)$ 를 발생시킨다.

**(3단계)** 충분히 큰  $N$ 에 대해서  $(\beta^{(j)}, \Sigma^{(j)})$ ,  $(j = N, N + 1, \dots)$ 를 기록한다.

**예제 3.3.4** Gibbs샘플러를 사용해서 다변량선형회귀모형들을 추정하는 예로서, 다음 MATLAB프로그램 BayesMultivariateRegression101.m을 실행해 보자.

```

1 % -----
2 %  Filename: BayesMultivariateRegression101.m
3 %  Gibbs Sampler for Multivariate Linear Regression
4 %  (Ref)  [1] (1997),  [2] (), f12 .4
5 % -----
6 clear, close all
7 % Make a dataset
8 XY = [ 1    199    213.283    100.000    1239.39    -1474.32    537.692
9        2    198    216.462    100.000    1248.24    -1463.15    560.719

```

10	3	197	219.537	100.000	1251.94	-1454.58	494.979
11	4	196	222.504	100.000	1260.95	-1444.30	506.805
12	5	195	225.365	100.000	1265.69	-1435.29	532.535
13	6	194	228.120	100.000	1273.73	-1424.83	555.231
14	7	193	230.771	100.000	1279.60	-1415.23	570.780
15	8	192	233.319	100.000	1284.46	-1405.13	569.721
16	9	191	235.768	100.000	1291.53	-1396.90	560.003
17	10	190	238.122	100.000	1295.87	-1387.05	570.373
18	11	189	240.387	100.000	1301.58	-1376.77	638.838
19	12	188	242.565	100.000	1303.98	-1369.23	576.066
20	13	187	244.665	100.000	1309.82	-1358.52	633.023
21	14	186	246.691	100.000	1311.97	-1351.76	609.357
22	15	185	248.650	100.000	1316.95	-1341.47	643.933
23	16	184	250.549	100.000	1320.35	-1331.53	682.194
24	17	183	252.394	100.000	1325.57	-1321.93	685.426
25	18	182	254.192	100.000	1327.05	-1314.22	658.414
26	19	181	255.951	100.000	1329.16	-1306.98	668.311
27	20	180	257.676	100.000	1334.46	-1298.45	686.457
28	21	179	259.373	100.000	1338.84	-1287.67	709.615
29	22	178	261.050	100.000	1339.10	-1281.09	677.734
30	23	177	262.711	100.000	1343.38	-1271.10	742.429
31	24	176	264.362	100.000	1344.92	-1262.26	697.858
32	25	175	266.006	100.001	1348.64	-1253.59	756.798
33	26	174	267.649	100.001	1350.55	-1244.49	721.693
34	27	173	269.292	100.001	1355.82	-1237.24	731.112
35	28	172	270.939	100.001	1358.77	-1228.08	746.745
36	29	171	272.590	100.001	1360.49	-1218.26	791.422
37	30	170	274.247	100.001	1363.85	-1212.25	715.810
38	31	169	275.910	100.002	1367.55	-1202.31	764.490
39	32	168	277.577	100.002	1369.83	-1192.81	754.039
40	33	167	279.246	100.002	1372.69	-1184.86	751.371
41	34	166	280.915	100.003	1375.69	-1178.44	772.469
42	35	165	282.580	100.003	1379.92	-1166.73	820.098
43	36	164	284.236	100.004	1382.36	-1157.98	818.240
44	37	163	285.879	100.004	1386.40	-1149.79	834.116
45	38	162	287.501	100.005	1388.88	-1139.86	828.090
46	39	161	289.096	100.006	1389.89	-1132.71	837.483
47	40	160	290.657	100.007	1393.53	-1123.42	818.780
48	41	159	292.176	100.008	1394.75	-1116.66	789.688
49	42	158	293.642	100.010	1396.64	-1107.54	815.167
50	43	157	295.049	100.012	1401.17	-1098.34	879.008
51	44	156	296.385	100.014	1402.05	-1090.00	887.799
52	45	155	297.642	100.016	1403.96	-1081.91	884.821
53	46	154	298.808	100.019	1406.62	-1074.61	899.420
54	47	153	299.875	100.023	1407.20	-1066.04	905.144
55	48	152	300.832	100.027	1405.88	-1060.00	922.786
56	49	151	301.670	100.032	1406.70	-1051.64	934.244
57	50	150	302.377	100.038	1407.19	-1044.06	852.595
58	51	149	302.946	100.045	1407.89	-1035.63	854.448
59	52	148	303.367	100.053	1407.71	-1026.34	898.523
60	53	147	303.632	100.063	1406.49	-1019.73	826.725
61	54	146	303.732	100.074	1405.00	-1013.63	898.622
62	55	145	303.660	100.087	1400.85	-1006.52	850.664
63	56	144	303.411	100.103	1399.64	-999.69	909.132
64	57	143	302.977	100.122	1395.65	-992.51	859.386
65	58	142	302.355	100.144	1392.02	-987.39	922.789
66	59	141	301.541	100.170	1389.23	-981.19	897.198
67	60	140	300.531	100.201	1382.83	-974.77	905.877
68	61	139	299.325	100.237	1378.61	-968.37	886.986
69	62	138	297.921	100.280	1371.13	-964.80	831.672
70	63	137	296.320	100.331	1364.44	-957.39	852.421

```

71 | 64    136    294.522    100.391    1358.39    -951.87    903.908
72 | 65    135    292.532    100.462    1350.77    -948.39    917.692
73 | 66    134    290.352    100.546    1341.44    -943.49    842.546
74 | 67    133    287.987    100.645    1332.48    -940.66    923.574
75 | 68    132    285.444    100.762    1322.40    -936.02    887.563
76 | 69    131    282.728    100.900    1312.52    -931.96    865.029
77 | 70    130    279.848    101.063    1300.87    -928.59    805.137
78 | 71    129    276.812    101.256    1291.82    -925.54    852.043
79 | 72    128    273.630    101.484    1279.99    -921.91    801.422
80 | 73    127    270.312    101.753    1268.03    -920.58    789.958
81 | 74    126    266.869    102.071    1257.98    -914.33    798.386
82 | 75    125    263.314    102.447    1245.54    -915.66    735.397
83 | 76    124    259.657    102.891    1234.43    -912.62    781.367
84 | 77    123    255.911    103.415    1222.36    -909.23    743.478
85 | 78    122    252.090    104.034    1207.84    -909.35    743.532
86 | 79    121    248.206    104.766    1196.65    -909.28    714.882
87 | 80    120    244.272    105.630    1185.04    -906.23    723.324
88 | 81    119    240.303    106.651    1173.04    -907.87    707.861
89 | 82    118    236.310    107.858    1160.56    -910.00    701.413
90 | 83    117    232.307    109.283    1150.16    -908.89    728.898
91 | 84    116    228.306    110.966    1140.71    -909.99    696.968
92 | 85    115    224.320    112.955    1130.22    -912.95    707.907
93 | 86    114    220.360    115.305    1120.98    -918.02    661.052
94 | 87    113    216.438    118.080    1112.37    -924.05    628.065
95 | 88    112    212.565    121.359    1106.62    -928.90    647.528
96 | 89    111    208.749    125.233    1101.13    -937.88    671.320
97 | 90    110    205.001    129.810    1096.33    -949.00    597.758
98 | 91    109    201.328    135.216    1093.15    -963.81    602.509
99 | 92    108    197.737    141.603    1094.17    -978.60    612.605
100 | 93    107    194.236    149.148    1095.22    -996.85    582.083
101 | 94    106    190.829    158.061    1102.51    -1020.19    613.329
102 | 95    105    187.521    168.591    1112.41    -1047.62    652.241
103 | 96    104    184.316    181.031    1123.80    -1080.97    581.212
104 | 97    103    181.215    195.727    1140.94    -1122.83    547.012
105 | 98    102    178.221    213.088    1162.99    -1170.60    565.897
106 | 99    101    175.334    233.597    1197.62    -1225.88    585.277
107 | 100   100    172.552    257.827    1234.16    -1295.36    562.882 ];
108 | %% Matrix form of the model
109 | [ nn dum11 ] = size(XY);
110 | kk = 3+1, pp = 3
111 | [ nn dum11 ] = size(XY);
112 | ONEs = ones(nn,1);
113 | X = [ ONEs, XY(:,2:kk) ];
114 | Y = XY(:,kk+1:kk+pp);
115 | XX = X'*X;
116 |
117 | % OLS
118 | BetaHat = inv(X'*X)*X'*Y
119 | ResidLSE = Y - X*BetaHat;
120 | SigmaLSE = ResidLSE'*ResidLSE/(nn-kk)
121 |
122 | % Gibbs Sampler for Multivariate Linear Regression
123 | randn('state',3)
124 | iter = 5000;
125 | burn = 1000;
126 | NoIter = iter-burn;
127 | %% Initial Parameters
128 | V_a = 0.01*eye(pp)
129 | n_a = 3
130 | Beta_a = BetaHat
131 | A_a = eye(kk)

```

```

132 Sigma_old = eye(pp)
133 Beta_old = zeros(kk,pp)
134 norm_old = 10000;
135 BetaAvg = zeros(kk,pp);
136 SigmaAvg = zeros(pp,pp);
137 for ii = 1:iter
138     A_b = XX+A_a;
139     Beta_b = inv(A_b)*(XX*BetaHat+A_a*Beta_a);
140     bbeta_b = reshape(Beta_b, kk*pp, 1);
141     n_b = n_a + nn;
142     dumE = Y - X*Beta_b; dumBmB = Beta_b - Beta_a;
143     V_b = V_a + dumE'*dumE + dumBmB'*A_a*dumBmB;
144     Sigma_new = inv(wishrnd(inv(V_b), n_b));
145     bbeta_new = mvnrnd(bbeta_b, kron(Sigma_new, inv(A_b)))';
146     Beta_new = reshape(bbeta_new, kk, pp);
147     norm_new = norm(Sigma_new);
148     if ii > burn
149         SSigma(ii-burn, :, :) = Sigma_new;
150         BBeta(ii-burn, :, :) = Beta_new;
151         BetaAvg = BetaAvg + Beta_new;
152         SigmaAvg = SigmaAvg + Sigma_new;
153         error(ii) = abs(norm_new - norm_old)/norm_old;
154         if abs(error(ii)) < 10^-6
155             NoIter = ii-burn; break
156         end
157     end
158     % For the next step
159     Sigma_old = Sigma_new;
160     Beta_old = Beta_new;
161     norm_old = norm_new;
162     A_a = A_b; Beta_a = Beta_b; n_a = n_b; V_a = V_b;
163 end
164 NoIter
165 BetaAvg = BetaAvg/NoIter
166 SigmaAvg = SigmaAvg/NoIter
167 ResidG = Y - X*BetaAvg;
168 SigmaG = ResidG'*ResidG/(nn-kk)
169
170 save('BayesMultivariateRegression101', 'BetaAvg', 'SigmaAvg', 'SigmaG')
171 % End of Program
172 % -----

```

이 MATLAB 프로그램 BayesMultivariateRegression101.m에서는 Gibbs 샘플러를 사용해서 다변량선형회귀모형을 추정한다. 데이터세트는 예제 3.3.3에서 사용한 데이터와 같다.

우선, 최소제곱추정법을 사용해서 추정한 회귀계수행렬은 다음과 같다.

$$\hat{B} = \begin{bmatrix} 1.9548 & 5.3376 & 27.0935 \\ 2.0005 & -7.0006 & -1.5478 \\ 2.9975 & 0.9989 & 3.4670 \\ 1.9968 & -2.9987 & 0.4182 \end{bmatrix} \quad (1)$$

즉, 최소제곱추정법으로 추정된 회귀식들은 다음과 같다.

$$y_1 = 1.9548 + 2.0005x_1 + 2.9975x_2 + 1.9968x_3 + \hat{\epsilon}_1 \quad (2)$$

$$y_2 = 5.3376 - 7.0006x_1 - 0.9989x_2 - 2.9987x_3 + \hat{\epsilon}_2 \quad (3)$$

$$y_3 = 27.0935 - 1.5478x_1 + 3.4670x_2 + 0.4182x_3 + \hat{\epsilon}_3 \quad (4)$$

이 추정식들은 예제 3.3.3의 결과와 일치한다. 오차벡터의 분산공분산행렬  $\Sigma$ 의 불편추정행렬은 다음과 같다.

$$\hat{S} = \begin{bmatrix} 0.9331 & 0.4533 & 10.6582 \\ 0.4533 & 0.8880 & 8.2529 \\ 10.6582 & 8.2529 & 823.6729 \end{bmatrix} \quad (5)$$

다음으로 위에서 기술한 Gibbs샘플러를 사용해서 다중선형회귀모형을 추정하기 위해서, 초기값들을 다음과 같이 선택했다.

$$V_a = 0.01I_p, \quad n_a = 3, \quad B_a = \hat{B}, \quad A_a = I_k \quad (6)$$

이 MATLAB프로그램에서 Gibbs샘플러를 사용해서 추정한 회귀계수행렬  $B_G$ 는 다음과 같다.

$$B_G = \begin{bmatrix} 1.9582 & 5.3384 & 27.1764 \\ 2.0005 & -7.0006 & -1.5478 \\ 2.9975 & 0.9989 & 3.4669 \\ 1.9968 & -2.9987 & 0.4180 \end{bmatrix} \quad (7)$$

식 (7)의 각 회귀계수추정값은 번인횟수를 1000회로 하고, 다음 멈춤기준을 적용해서 얻은 543개 Gibbs추정값들의 표본평균이다.

$$\frac{|norm(\Sigma^{(j)}) - norm(\Sigma^{(j-1)})|}{norm(\Sigma^{(j)})} < 10^{-6} \quad (8)$$

또한, 이 Gibbs샘플러에 의한 분산공분산행렬  $\Sigma$ 의 추정행렬  $\Sigma_G$ 는 다음과 같다.

$$\Sigma_G = \begin{bmatrix} 0.9331 & 0.4533 & 10.6582 \\ 0.4533 & 0.8880 & 8.2529 \\ 10.6582 & 8.2529 & 823.6729 \end{bmatrix} \quad (9)$$

식 (5)와 식 (9)에서 알 수 있듯이, 이 Gibbs샘플러에 의한 분산공분산행렬  $\Sigma$ 의 추정행렬  $\Sigma_G$ 는 불편추정량  $\hat{S}$ 와 거의 일치한다. ■

## 제 4 장

# Markov 체인

A drunk man will find his way home, but a drunk  
bird may get lost forever.

Shizuo Kakutani

### 제 4.1 절 베이지안분석과 Markov 체인

베이지안통계분석 대상인 모형에 모수들이 여러 개이면, 그 사후확률분포는 단변량확률분포가 아닌 다변량확률분포이다. 각 모수에 관한 통계적 추론을 시행할 때는 그 모수의 사후주변확률분포를 사용해야 하고, 사후주변확률밀도함수를 구하기 위해서는 다변량확률밀도함수에 대한 다중적분을 계산해야만 한다. 모수의 공액사전확률분포가 존재하는 경우에는 사후확률분포가 알려져 있어, 해당 모수의 사후주변확률분포를 해석적으로 구할 수 있는 경우가 많다. 그러나, 공액사전확률분포가 반드시 존재하는 것은 아니며, 겨우 공액사전조건부확률분포가 존재하는 경우가 일반적이다. 공액사전조건부확률분포가 존재하지 않는 경우에는 해석적인 방법으로 사후주변확률분포를 구하는 것은 어려우며, 이러한 경우에는 수치해석적 기법을 이용해서 다중적분을 계산해야 한다. 모수들의 개수가 적은 경우에는 수치해석기법을 적용해서 다중적분을 할 수 있지만, 모수들의 개수가 많은 경우에는 수치해석기법에 의한 적분은 정밀도가 떨어진다. 이러한 경우에 Markov 체인몬테카를로법(Markov Chain Monte Carlo method: MCMC method)을 사용해서 다중적분을 계산할 수 있다. MCMC는 수치해석기법에 비해 정밀도가 높은 결과를 제공하는 것으로 알려져 있다. Gelfand & Smith [42] 이후 베이지안분석에서 MCMC를 집중적으로 사용하게 되었다.

MCMC는 Markov 체인(Markov chain)을 이용하는 몬테카를로법(Monte Carlo method)이다. MCMC의 목적은 확률분포에서 좋은 샘플들을 발생시키는 것이므로, 샘플을 발생시키는

목표가 되는 확률분포라는 의미에서 이 확률분포를 목표분포(target probability distribution)라고 부른다. 보통 몬테카를로법에서는 목표분포에서 서로 독립적인 샘플들을 발생시킨다. 따라서, 샘플의 개수, 즉 샘플수를 크게 함으로써, 대수법칙(law of large numbers)에 입각해서 목표분포가 포함하는 모수를 추정할 수 있다. 그러나, 목표분포가 복잡한 다변량확률분포인 경우에는 그렇게 단순하게 문제를 해결할 수 없다. 그러한 복잡한 문제의 해결책들 중 하나가 MCMC이다. 즉, 불변분포(invariant distribution)가 목표분포가 되도록 Markov 체인을 구축하고, 이 Markov 체인에서 발생하는 샘플을 목표분포에서 발생한 샘플로 간주하는 것이 MCMC이다. 목표분포를 불변분포로 갖는 Markov 체인을 구성하는 대표적인 방법들이 Gibbs 샘플러와 Metropolis-Hastings 알고리즘이다. MCMC를 사용하는 최종목표는 발생한 샘플들로부터 정적분, 즉 기대값을 추정하는 것이다. 그러나 Markov 체인에서 발생하는 샘플들은 서로 독립은 아니므로, 기존의 대수법칙을 적용할 수 없다. 따라서, Markov 체인에서 표본평균이 모평균으로 수렴하는지 여부는 매우 중요한 문제이다. Birkhoff [17]가 제시한 에르고딕정리(ergodic theorem)에 의하면, 에르고딕성(ergodicity)을 갖는 Markov 체인에서는 표본평균이 모평균으로 수렴한다. MCMC에서는 이 에르고딕정리를 바탕으로 기대값을 추정하는 것이다. MCMC를 이해하기 위해서는 먼저 Markov 체인을 알고있어야 한다. Markov 체인이란 Markov 성(Markov property)을 만족시키는 이산시간형 Markov 확률과정이다. 이 장에서는 MCMC를 이해하는데 필요한 정도만큼 Markov 체인을 다루기로 하자. Markov 체인에 관한 자세한 내용은 Ross [85]와 최병선 [11]을 참조하고, Markov 체인과 몬테카를로법 사이의 관계는 Tierney [95]를 참조하라.

## 제 4.2 절 Markov 체인의 정의

이산시간형 확률변수열  $\{X_t\}$ 에서 정보  $\{X_s | s \leq n\}$ 가 주어졌을 때  $X_{n+1}$ 의 조건부확률밀도 함수  $f(X_{n+1} | X_n, X_{n-1}, \dots)$ 이  $f(X_{n+1} | X_n)$ 과 같으면, 이 확률과정  $\{X_t\}$ 는 (1차) Markov 성(Markov property)을 갖는다고 한다. 또한, 이  $f(\cdot | \cdot)$ 을 추이핵(transition kernel)이라 부른다. 서로 독립인 확률변수열  $\{X_t\}$ 는 Markov성을 가지므로, Markov성은 확률변수들의 독립성을 느슨하게 한 성질이라고 할 수 있다. 만약 조건부확률밀도함수  $f(X_{n+1} | X_n, X_{n-1}, \dots)$ 이  $f(X_{n+1} | X_n, X_{n-1})$ 과 같으면, 이 확률과정  $\{X_t\}$ 는 2차 Markov성을 갖는다고 한다. 확률변수  $X_t$ 의 지지대(support)  $S$ 를 상태공간(state space)이라 부른다.

Markov 체인은 유한 또는 가산인 상태공간  $S$ 에서 정의되며 Markov성을 갖는 이산시간형 확률과정이다. 이 장에서는 상태공간  $S$ 가 정수들의 집합  $\mathbb{Z}$ 의 부분집합인 이산형 상태공간모

형을 다룬다. 비음인 정수들의 집합을  $\mathbb{Z}_{\geq 0}$ 라 하자. 즉,  $\mathbb{Z}_{\geq 0} \doteq \{0, 1, 2, \dots\}$ 이다. 확률과정  $\{X_n | n \in \mathbb{Z}_{\geq 0}\}$ 에서  $X_n = i$ 는 시점  $n$ 에서 상태  $i$ 에 있음을 뜻한다. Markov체인의 특징은 현재시점에서 상태  $i$ 에 있을 때 다음 시점에서 상태  $j$ 로 이동할 확률이 과거와 독립적이라는 것이다. 즉, Markov 체인에서는 각  $n \in \mathbb{Z}_{\geq 0}$ 과 상태들  $i_0, i_1, \dots, i_{n-1}, i \in S$ 와  $j \in S$ 에 대해서, 다음 식이 성립한다.

$$\begin{aligned} & \Pr(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) \\ &= \Pr(X_{n+1} = j | X_n = i) \end{aligned} \tag{4.2.1}$$

이러한 확률과정  $\{X_n\}$ 을 Markov 체인 또는 이산시간 Markov 체인 (discrete time Markov chain: DTMC)라 부른다. 특히, 상태공간  $S$ 가 유한집합인 경우에는 확률과정  $\{X_n\}$ 을 유한 Markov 체인 (finite Markov Chain)이라고 한다.

식 (4.2.1) 우변의 확률  $\Pr(X_{n+1} = j | X_n = i)$ 가 시점  $n$ 에 의존하지 않는 경우, 즉 다음 식이 성립하는 경우에는  $\{X_n\}$ 을 시간동질적 (time-invariant, time-homogeneous, temporally homogeneous) 또는 정상적 (stationary) Markov 체인이라고 한다.

$$p(i, j) \doteq \Pr(X_{n+1} = j | X_n = i), \quad (i, j \in S) \tag{4.2.2}$$

여기서  $p(i, j)$ 는 시점  $n$ 의 상태  $i$ 에서 다음 시점  $n+1$ 의 상태  $j$ 로 이동하는 확률을 나타내고, 이를 추이확률 (transition probability)이라 부른다. 추이확률들은 다음 식들을 만족한다.

$$p(i, j) \geq 0, \quad (i, j \in S) \tag{4.2.3}$$

$$\sum_{j \in S} p(i, j) = 1, \quad (i \in S) \tag{4.2.4}$$

이 절에서는 시간동질적 Markov 체인만을 다루기로 하자.

만약 행렬  $M = [m(i, j)]$ 의 원소들이 식 (4.2.3)과 식 (4.2.4)를 만족하면, 행렬  $M$ 을 확률 행렬 (stochastic matrix)이라 부른다. 추이확률들 전체를 나열해서, 추이확률행렬 (transition probability matrix)  $\mathbb{P} \doteq [p(i, j)]$ 를 정의한다. 추이확률행렬은 식 (4.2.3)과 식 (4.2.4)를

만족하므로, 확률행렬이다. 만약 상태공간이  $S = \mathbb{Z}_{\geq 0}$  이면, 추이확률행렬은 다음과 같다.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & \cdots & j & \cdots \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ \vdots \\ i \\ \vdots \end{matrix} & \left[ \begin{array}{cccccc} p(0,0) & p(0,1) & p(0,2) & \cdots & p(0,j) & \cdots \\ p(1,0) & p(1,1) & p(1,2) & \cdots & p(1,j) & \cdots \\ p(2,0) & p(2,1) & p(2,2) & \cdots & p(2,j) & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \\ p(i,0) & p(i,1) & p(i,2) & \cdots & p(i,j) & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \end{array} \right. \end{matrix} \quad (4.2.5)$$

시간동질적 Markov 체인  $\{X_n\}$  은 각  $i_0, i_1, \dots, i_{n-1}, i_n (\in S)$  에 대해서 다음 식들을 만족한다.

$$\begin{aligned} & \Pr(X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) \\ &= \Pr(X_0 = i_0) \prod_{k=1}^n \Pr(X_k = i_k | X_{k-1} = i_{k-1}, X_{k-2} = i_{k-2}, \dots, X_0 = i_0) \\ &= \pi^{(0)}(i_0) p(i_0, i_1) p(i_1, i_2) \cdots p(i_{n-2}, i_{n-1}) p(i_{n-1}, i_n) \end{aligned} \quad (4.2.6)$$

여기서  $\pi^{(0)}(i_0) \doteq \Pr(X_0 = i_0)$  는 시점  $t = 0$  에서 상태가  $i_0$  인 확률, 즉 초기확률이다. 초기확률들의 집합  $\{\pi^{(0)}(i) | i \in S\}$  를 Markov 체인의 초기분포(initial distribution)라 하고, 벡터  $\boldsymbol{\pi}^{(0)} \doteq [\pi^{(0)}(0), \pi^{(0)}(1), \dots]$  를 초기상태벡터(initial state vector) 또는 초기벡터(initial vector)라 한다. 명확하게 말하면,  $\{\pi^{(0)}(i) | i \in S\}$  는 확률밀도함수이다. 그러나, 확률과정론에서는  $\{\pi^{(0)}(i) | i \in S\}$  를 관습적으로 초기분포라고 부름을 기억하라. 또한, 초기상태벡터의 상태벡터라는 단어도 혼란을 일으킨다. 따라서, 본서에서는 상태벡터보다는 상태확률벡터라는 용어를 사용할 것이다. 초기시점의 상태확률벡터와 추이확률행렬이 주어지면, 식 (4.2.6)에 의해서 각 유한차원 확률분포가 결정된다. 즉, Markov 체인이 일의적으로 결정된다. 따라서, Markov 체인이 주어졌다는 것은 Markov 체인의 확률법칙을 결정하는 추이확률행렬  $\mathbb{P}$ 와 초기시점에서 상태확률벡터  $\boldsymbol{\pi}^{(0)}$  의 쌍  $(\mathbb{P}, \boldsymbol{\pi}^{(0)})$  가 주어짐을 의미한다. 특히, 초기시점의 상태확률벡터  $\boldsymbol{\pi}^{(0)}$  가 문제되지 않는 경우에는, 추이확률행렬  $\mathbb{P}$ 가 주어지면 Markov 체인이 주어졌다고 한다.

식 (4.2.1)이 성립하면, 각  $n \in \mathbb{Z}_{\geq 0}$ ,  $m_a > m_{a-1} > \cdots > m_1 (> 0)$ 와  $k_b > k_{b-1} > \cdots >$

$k_1(\geq 0)$ 에 대해서 다음 식들이 성립한다.

$$\begin{aligned}
 & \Pr(X_{n+m_a} = i_{n+m_a}, \dots, X_{n+m_1} = i_{n+m_1} \mid X_{n-k_1} = i_{n-k_1}, \dots, X_{n-k_b} = i_{n-k_b}) \\
 &= \Pr(X_{n+m_1} = i_{n+m_1} \mid X_{n-k_1} = i_{n-k_1}, \dots, X_{n-k_b} = i_{n-k_b}) \\
 & \quad \cdot \prod_{j=1}^a \Pr(X_{n+m_j} = i_{n+m_j} \mid X_{n+m_{j-1}} = i_{n+m_{j-1}}, \dots, X_{n+m_1} = i_{n+m_1}, \\
 & \quad \quad \quad X_{n-k_1} = i_{n-k_1}, \dots, X_{n-k_b} = i_{n-k_b}) \\
 &= \Pr(X_{n+m_1} = i_{n+m_1} \mid X_{n-k_1} = i_{n-k_1}) \\
 & \quad \cdot \prod_{j=2}^a \Pr(X_{n+m_j} = i_{n+m_j} \mid X_{n+m_{j-1}} = i_{n+m_{j-1}}) \tag{4.2.7}
 \end{aligned}$$

여기서 두 번째 등호는 Markov성의 정의에 의해서 성립한다. 식 (4.2.7)의 우변은  $k_2, k_3, \dots, k_b$ 에 의존하지 않는다. 따라서, 다음 식들이 성립한다.

$$\begin{aligned}
 & \Pr(X_{n+m_a} = i_{n+m_a}, \dots, X_{n+m_1} = i_{n+m_1} \mid X_{n-k_1} = i_{n-k_1}, \dots, X_{n-k_b} = i_{n-k_b}) \\
 &= \Pr(X_{n+m_a} = i_{n+m_a}, \dots, X_{n+m_1} = i_{n+m_1} \mid X_{n-k_1} = i_{n-k_1}) \tag{4.2.8}
 \end{aligned}$$

식 (4.2.8)의 양변에  $\sum_{i_{n+m_1}} \dots \sum_{i_{n+m_a-1}}$ 을 취하면, 다음 식을 얻는다.

$$\begin{aligned}
 & \Pr(X_{n+m_a} = i_{n+m_a} \mid X_{n-k_1} = i_{n-k_1}, \dots, X_{n-k_b} = i_{n-k_b}) \\
 &= \Pr(X_{n+m_a} = i_{n+m_a} \mid X_{n-k_1} = i_{n-k_1}) \tag{4.2.9}
 \end{aligned}$$

식 (4.2.1)에 기술한 Markov체인의 정의는 식 (4.2.9)의 특별한 경우이다. 즉, 식 (4.2.1)과 식 (4.2.9)는 동치이다. 따라서, 각  $m, n(\in \mathbb{Z}_{\geq 0})$  그리고 각  $i, j, i_0, i_1, \dots, i_{n-1}(\in \mathbb{Z}_{\geq 0})$ 에 대해서 다음 식이 성립하면, 확률과정  $\{X_l \mid l \in \mathbb{Z}_{\geq 0}\}$ 은 Markov체인이다.

$$\begin{aligned}
 & \Pr(X_{n+m} = i \mid X_n = j, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) \\
 &= \Pr(X_{n+m} = i \mid X_n = j) \tag{4.2.10}
 \end{aligned}$$

또한, 각  $n(\in \mathbb{Z}_{\geq 0})$  그리고 각  $i, j, k_b > k_{b-1} > \dots > k_1(> 0)$ 에 대해서 다음 식이 성립하면,

확률과정  $\{X_l | l \in \mathbb{Z}_{\geq 0}\}$  은 Markov 체인이다.

$$\begin{aligned} & \Pr(X_{n+1} = i | X_n = j, X_{n-k_1} = i_{n-k_1}, \dots, X_{n-k_b} = i_{n-k_b}) \\ &= \Pr(X_{n+1} = i | X_n = j) \end{aligned} \quad (4.2.11)$$

## 제 4.3 절 통계모형과 Markov 성

### 4.3.1 ARMA 모형과 Markov 성

다음과 같은  $ARMA(p, q)$  모형을 살펴보자.

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + v_t + \theta_1 v_{t-1} + \dots + \theta_q v_{t-q}, \quad (t \in \mathbb{Z}) \quad (4.3.1)$$

여기서  $\{v_t | t \in \mathbb{Z}\}$  는 서로 독립이고 동일한 확률분포를 따르는 확률변수들이다. 이  $\{v_t\}$  를 오차항들 또는 쇄신항들 (innovations) 이라 부른다. 확률변수  $y_t$  가 확률변수  $y_{t-1}$  뿐 아니라 확률변수  $y_{t-2}$  에도 의존한다. 따라서 이  $ARMA(p, q)$  모형은 Markov 성을 갖지 않는 것 같다.

다음과 같이 벡터들과 행렬들을 정의하자.

$$\Phi = \begin{bmatrix} 1 & -\phi_1 & -\phi_2 & \dots & -\phi_{p-1} & -\phi_p \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}, \quad \mathbf{y}_t = \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p+1} \\ y_{t-p} \end{bmatrix} \quad (4.3.2)$$

$$\Theta = \begin{bmatrix} 1 & \theta_1 & \theta_2 & \cdots & \theta_{q-1} & \theta_q \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \quad \mathbf{v}_t = \begin{bmatrix} v_t \\ v_{t-1} \\ v_{t-2} \\ \vdots \\ v_{t-q+1} \\ v_{t-q} \end{bmatrix} \quad (4.3.3)$$

식 (4.3.1)을 다음과 같이 쓸 수 있다.

$$\mathbf{y}_{t+1} = \Phi \mathbf{y}_t + \Theta \mathbf{v}_t, \quad (t \in \mathbb{Z}) \quad (4.3.4)$$

식 (4.3.4)에서 알 수 있듯이, 확률벡터과정  $\{\mathbf{y}_t\}$ 는 Markov성을 갖는다.

### 4.3.2 Markov스위칭 ARMA모형과 Markov성

다음과 같은 Markov스위칭 ARMA모형 (MS-ARMA model)을 살펴보자.

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \Phi_{1t}\mathbf{y}_{t-1} + \cdots + \Phi_{pt}\mathbf{y}_{t-p} + \mathbf{v}_t + \Theta_{1t}\mathbf{v}_{t-1} + \cdots + \Theta_{qt}\mathbf{v}_{t-q}, \quad (t \in \mathbb{Z}) \quad (4.3.5)$$

여기서  $\mathbf{v}_t = \Sigma_t \boldsymbol{\epsilon}_t \in \mathbb{R}^d$ 이고,  $\{\boldsymbol{\epsilon}_t \in \mathbb{R}^d \mid t \in \mathbb{Z}\}$ 는 서로 독립이고 동일한 다변량확률분포를 따르는 확률벡터들이고,  $\Sigma_t$ 는 시점  $t$ 에 의존하는 양정치행렬이다. 또한,  $\Phi_{1t}, \dots, \Phi_{pt}, \Theta_{1t}, \dots, \Theta_{qt} \in \mathbb{R}^{d \times d}$ 는 시점  $t$ 에 의존하는 계수행렬들이고,  $\boldsymbol{\mu}_t \in \mathbb{R}^d$ 는 추세벡터 (drift vector)이다.

Stelzer (2009)는 이 MS-ARMA의 상태공간표현 (state space representation)을 제시하였다. 다음과 같은 행렬들과 벡터들을 정의하자.

$$\mathbf{Y}_t = [\mathbf{y}_t^T, \dots, \mathbf{y}_{t-p+1}^T, \mathbf{v}_t^T, \dots, \mathbf{y}_{t-q+1}^T]^T \in \mathbb{R}^{d[p+q]} \quad (4.3.6)$$

$$\Sigma_t = \left[ \Sigma_t^T, \underbrace{O^T, \dots, O^T, O^T}_{p-1}, \Sigma_t^T, \underbrace{O^T, \dots, O^T}_{q-1} \right]^T \in \mathbb{R}^{d[p+q] \times d} \quad (4.3.7)$$

$$\mathbf{m}_t = \left[ \boldsymbol{\mu}_t^T, \underbrace{O^T, \dots, O^T}_{p+q-1} \right]^T \in \mathbb{R}^{d[p+q]} \quad (4.3.8)$$

$$\mathbf{c}_t = \mathbf{m}_t + \Sigma_t \boldsymbol{\epsilon}_t \quad (4.3.9)$$

$$\Phi_t = \begin{bmatrix} \Phi_{1,t} & \Phi_{2,t} & \cdots & \Phi_{p-1,t} & \Phi_{p,t} \\ I_d & O & \cdots & O & O \\ O & I_d & \cdots & O & O \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & \cdots & I_d & O \end{bmatrix} \in \mathbb{R}^{dp \times dp} \quad (4.3.10)$$

$$\Theta_t = \begin{bmatrix} \Theta_{1,t} & \Theta_{2,t} & \cdots & \Theta_{q-1,t} & \Theta_{q,t} \\ O & O & \cdots & O & O \\ O & O & \cdots & O & O \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & \cdots & O & O \end{bmatrix} \in \mathbb{R}^{dp \times dq} \quad (4.3.11)$$

$$\mathbf{J} = \begin{bmatrix} O & O & \cdots & O & O \\ I_d & O & \cdots & O & O \\ O & I_d & \cdots & O & O \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & \cdots & I_d & O \end{bmatrix} \in \mathbb{R}^{dp \times dq} \quad (4.3.12)$$

$$\mathbf{A}_t = \begin{bmatrix} \Phi_t & \Theta_t \\ O & \mathbf{J} \end{bmatrix} \in \mathbb{R}^{d[p+q] \times d[p+q]} \quad (4.3.13)$$

이 행렬들과 벡터들을 사용해서 식 (4.3.5)의 MS-ARMA모형을 다음과 같이 쓸 수 있다.

$$\mathbf{Y}_t = \mathbf{A}_t \mathbf{y}_{t-1} + \mathbf{c}_t, \quad (t \in \mathbb{Z}) \quad (4.3.14)$$

식 (4.3.14)에서 알 수 있듯이, 확률과정  $\{\mathbf{Y}_t\}$ 는 Markov성을 갖는다.

### 4.3.3 상태공간모형과 Markov성

상태공간모형 (state space model) 은 복잡한 구조를 갖는 시계열데이터를 분석하는 데 자주 사용되는 모형으로서, 동적모형 (dynamic model) 이라고 불리우기도 한다. 상태공간모형은 관측방정식 (measurement equation) 과 상태방정식 (state equation) 에 의해 정의된다. 출력방정식 (output equation) 이라고도 불리우는 관측방정식은 다음과 같다.

$$y_t = \mathbf{F}_t^T \boldsymbol{\theta}_t + \epsilon_t, \quad (t = 1, 2, \dots, n) \tag{4.3.15}$$

여기서  $y_t \in \mathbb{R}$  는 관측값이고,  $\mathbf{F}_t \in \mathbb{R}^d$  는 시점  $t$  에서 알고있는 설명변수벡터이고,  $\boldsymbol{\theta}_t \in \mathbb{R}^d$  는 시점  $t$  에서 미지인 상태벡터 (state vector) 또는 회귀계수벡터 (regression coefficient vector) 이고,  $\epsilon_t$  는 평균이 0 이고 분산이  $\sigma_t^2$  인 정규확률분포를 따르는 오차항이다. 오차항들  $\{\epsilon_t\}$  는 서로 독립이라고 가정하자. 따라서, 조건부확률변수들  $\{y_t | (\boldsymbol{\theta}_t, \sigma_t^2)\}$  는 서로 독립이다. 상태방정식은 다음과 같다.

$$\boldsymbol{\theta}_t = G_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad (t = 2, 3, \dots, n) \tag{4.3.16}$$

$$\boldsymbol{\theta}_1 \stackrel{d}{\sim} \mathcal{N}(\mathbf{a}_1, R_1) \tag{4.3.17}$$

여기서  $G_t \in \mathbb{R}^{d \times d}$  는 시점  $t$  에서 알고있는 상태추이확률행렬 (state transition matrix) 이고, 쇄신오차벡터 (innovation error vector)  $\mathbf{w}_t$  는 평균벡터가  $\mathbf{0}$  이고 분산공분산행렬이  $W_t$  인  $d$  변량 정규확률분포를 따르며,  $\{\mathbf{w}_t\}$  는 서로 독립이다. 또한, 시점  $t$  에서  $\epsilon_t$  와  $\mathbf{w}_t$  는 서로 독립이다. 식 (4.3.15) 과 식 (4.3.16) 에서 알 수 있듯이, 다음 식이 성립한다.

$$\begin{bmatrix} y_t \\ \boldsymbol{\theta}_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{F}_t^T \\ 0 & G_{t+1} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ \boldsymbol{\theta}_t \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \mathbf{w}_{t+1} \end{bmatrix} \tag{4.3.18}$$

식 (4.3.18) 에서 알 수 있듯이, 확률과정  $\{[y_t, \boldsymbol{\theta}_{t+1}^T]^T\}$  는 Markov 성을 갖는다.

### 4.3.4 확률미분방정식과 Markov성

확률미분방정식이란 결정적 미분방정식이 우연변동하는 쇄신항에 의해 교란되는 방정식으로 생각할 수 있다. 확률미분방정식은 다음과 같다.

$$dS_t = \mu(S_t, t)dt + \sigma(S_t, t)dW_t, \quad (t \geq 0) \tag{4.3.19}$$

여기서  $\{W_t\}$ 는 표준Brown운동이다. 추세계수(drift coefficient)  $\mu(S_t, t)$ 와 확산계수(diffusion coefficient)  $\sigma(S_t, t)$ 는  $t$ 와  $S_t$ 에 의존한다. 만약  $\mu(S_t, t) = 0$ 이고  $\sigma(S_t, t) = 1$ 이면,  $\{S_t\}$ 는 표준Brown운동이다. 확률미분방정식 (4.3.19)에서 알 수 있듯이, 다음 식이 성립한다.

$$dS_t | S_t = [\mu(S_t, t) | S_t] dt + [\sigma(S_t, t) | S_t] dW_t = \mu(S_t, t)dt + \sigma(S_t, t)dW_t \quad (4.3.20)$$

여기서 첫 번째 등호는  $S_t$ 와  $dW_t$ 의 독립성에 의해서 성립한다. 식 (4.3.20)을 바탕으로 확률미분방정식 (4.3.19)를 만족하는 확률과정  $\{S_t | t \geq 0\}$ 는 Markov성을 갖는 것처럼 보인다. 그러나, 확률미분방정식 (4.3.19)는 다음 확률적분방정식을 미분방정식의 형태로 표현한 것에 불과하다.

$$S_t = S_0 + \int_0^t \mu(S_u, u)du + \int_0^t \sigma(S_u, u)dW_u \quad (4.3.21)$$

식 (4.3.21)의 우변의 첫 번째 적분은 Riemann적분이고, 두 번째는 Ito적분이다. 확률미분방정식 (4.3.19)의 추세계수와 확산계수는  $S_t$ 와  $t$ 에 의존하므로, 이 계수들 자체가 확률변수들이다. 시점  $t$ 에서 정보집합이 주어지면, 이 계수들이 관찰된다.

식 (4.3.21)에서 알 수 있듯이, 각  $s > 0$ 에 대해서 다음 식이 성립한다.

$$S_{t+s} = S_t + \int_t^{t+s} \mu(S_u, u)du + \int_t^{t+s} \sigma(S_u, u)dW_u \quad (4.3.22)$$

즉, 확률과정  $\{S_t | t \geq 0\}$ 는 Markov성을 갖는다.

## 제 4.4절 Markov 체인 예제

이 절에서는 Markov 체인 예제들을 살펴보자.

**예제 4.4.1** 제  $n$  번째 날 오전 8시에 자동차가 작동하면  $X_n = 1$ 이고, 작동하지 않으면  $X_n = 0$ 라 하자. 이 확률과정  $\{X_n\}$ 의 상태공간은  $S = \{0, 1\}$ 이다. 이 확률과정의 추이확률들이 다음과 같다고 하자.

$$p(0, 1) = p, \quad p(1, 0) = q, \quad (p, q \in [0, 1]) \quad (1)$$

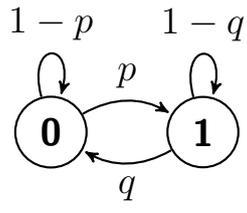
따라서,  $p(0, 1)$ 은 오전 8시에 고장난 상태인 차가 다음 날 오전 8시 전에 수리가 될 확률을

나타내고,  $p(0, 0)$ 는 오전 8시에 고장난 상태인 차가 다음 날 오전 8시까지 수리가 되지 않을 확률을 나타내고,  $p(1, 0)$ 는 오전 8시에 고장나지 않은 차가 다음 날 오전 8시까지 고장날 확률을 나타내고,  $p(1, 1)$ 은 오전 8시에 고장나지 않은 차가 다음 날 오전 8시까지 고장이 나지 않을 확률이다.

이 Markov 체인의 추이확률행렬은 다음과 같다.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} \end{matrix} \tag{2}$$

또한, 이 Markov 체인의 상태추이도 (transition diagram 또는 state diagram 또는 state transition diagram)는 다음과 같다.

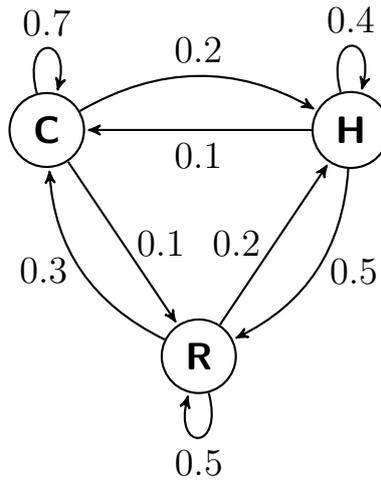


**예제 4.4.2** 서울의 7월 제  $n$  일의 날씨  $X_n$ 을 ‘맑다(Clear)’, ‘흐리다(Haze)’, ‘비가 온다(Rain)’로 분류하기로 하자. 실제로  $X_n$ 이  $X_{n-1}$  뿐 아니라  $X_{n-i}$ , ( $i = 2, 3, \dots$ )에도 의존한다고 가정하는 것이 타당할 것이다. 이 경우에  $\{X_n | n = 1, 2, \dots, 31\}$ 은 Markov 체인을 이루지 않는다. 그러나, 우리는 이 확률과정이 근사적으로 Markov 체인이라고 가정할 수도 있다.

이 Markov 체인의 추이확률행렬이 다음과 같다고 하자.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} C & H & R \end{matrix} \\ \begin{matrix} C \\ H \\ R \end{matrix} & \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.4 & 0.5 \\ 0.3 & 0.2 & 0.5 \end{bmatrix} \end{matrix} \tag{1}$$

이 Markov 체인의 상태추이도는 다음과 같다.



**예제 4.4.3** 다음과 같은 미로 (maze) 를 살펴보자.

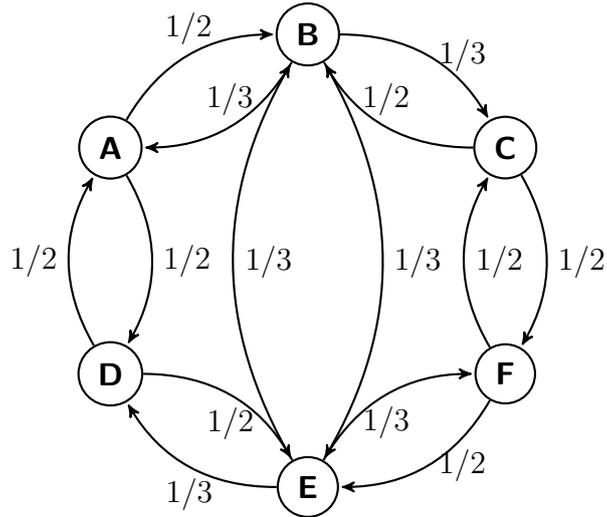
<b>A</b>	<b>B</b>	<b>C</b>
<b>D</b>	<b>E</b>	<b>F</b>

이 미로 속에 갇혀있는 쥐는 한 방에서 다른 방으로 이동할 때 일양확률분포를 따른다고 하고, 쥐가  $n$  번째 들어간 방을  $X_n$  이라 하자. 확률과정  $\{X_n | n = 0, 1, \dots\}$  은 상태공간이  $S = \{A, B, C, D, E, F\}$  인 Markov체인이다.

이 Markov체인의 추이확률행렬은 다음과 같다.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} A & B & C & D & E & F \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \end{matrix} & \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 \end{bmatrix} \end{matrix} \tag{2}$$

이 Markov 체인의 상태추이도는 다음과 같다.



**예제 4.4.4**    도박사의 파산

도박자금  $\$K$  를 지닌 도박사가 카지노에 가서 이기면  $\$1$  을 벌고 지면  $\$1$  을 잃는 게임을 반복해서 한다고 하자. 만약 도박사의 자금이  $\$N$  이 되면 도박사가 승리하고,  $\$0$  가 되면 도박사가 패배한다고 하자. 또한, 이 단위게임들은 서로 독립이며 각 단위게임에서 도박사가 이길 확률은  $p$  라고 하자. 시점  $n$  에서 도박자금을  $X_n$  이라 하면, 확률과정  $\{X_n | n = 0, 1, \dots\}$  은 추이확률들이 식  $p(0, 0) = 1$  과 식  $p(N, N) = 1$ , 그리고 각  $i (= 1, 2, \dots, N - 1)$  에 대해서 다음 식을 만족하는 Markov 체인이다.

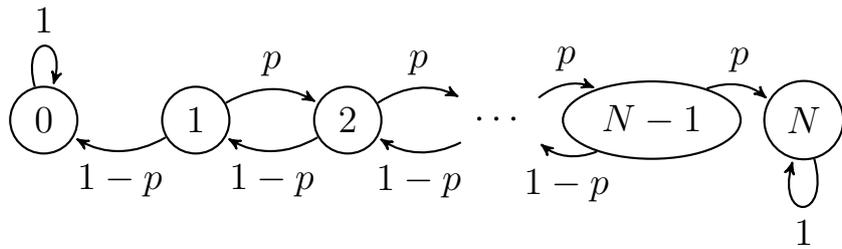
$$p(i, j) = \begin{cases} p, & (j = i + 1) \\ 1 - p, & (j = i - 1) \\ 0, & (\text{otherwise}) \end{cases} \quad (1)$$

이 Markov 체인은 도박사파산 (gambler's ruin) 을 설명하는데 사용되는 확률과정이다.

이 Markov 체인  $\{X_n\}$ 의 상태공간은  $S = \{0, 1, \dots, N\}$  이고 추이확률행렬은 다음과 같다.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & \cdots & N-2 & N-1 & N \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ \vdots \\ N-2 \\ N-1 \\ N \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1-p & 0 & p & \cdots & 0 & 0 & 0 \\ 0 & 1-p & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & p & 0 \\ 0 & 0 & 0 & \cdots & 1-p & 0 & p \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix} \end{matrix} \quad (2)$$

이 Markov 체인의 상태추이도는 다음과 같다.



**예제 4.4.5** 성공횟수

제  $k$  번째 전구 (light bulb)의 수명을  $Y_k$  라고 하고, 확률변수들  $\{Y_k\}$ 는 지지대 (support)가  $\mathbb{Z}_{\geq 0}$ 인 동일한 확률분포를 따르며 서로 독립이라고 가정하자. 다음과 같이 조건부확률을 정의하자.

$$r_m \doteq \Pr(Y_k > m \mid Y_k \geq m), \quad (m = 0, 1, 2, \dots) \quad (1)$$

여기서  $m$ 은 음이 아닌 정수이므로, 전구의 수명을 정수 단위, 예를 들어 분 (minute), 시 (hour) 또는 일 (day)로 나타내야 한다. 조건부확률  $r_m$ 은  $k$ 에 의존하지 않음에 유의하라. 시점  $n$ 에서 살아있는 전구의 나이 (age)를 확률변수  $X_n$ 으로 나타내자. 만약 시점  $n$ 에서 전구가 죽었으면,  $X_n = 0$ 라 하자. 이 확률과정  $\{X_n \mid n = 0, 1, \dots\}$ 은 Markov 체인이고 또한 추이확률들은

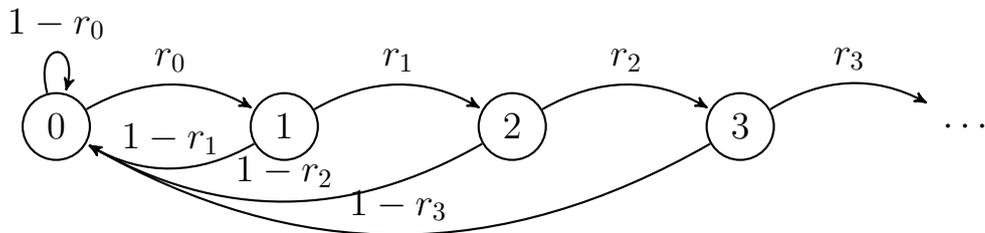
다음과 같다.

$$\Pr(X_{n+1} = j | X_n = i) = p(i, j) = \begin{cases} r_i, & (j = i + 1) \\ 1 - r_i, & (j = 0) \\ 0, & (\text{otherwise}) \end{cases} \quad (2)$$

이 Markov 체인  $\{X_n\}$ 은 성공횟수(success runs)를 설명하는데 사용되는 확률과정으로서 추이확률행렬은 다음과 같다.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & \dots \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{matrix} & \begin{bmatrix} 1 - r_0 & r_0 & 0 & 0 & 0 & \dots \\ 1 - r_1 & 0 & r_1 & 0 & 0 & \dots \\ 1 - r_2 & 0 & 0 & r_2 & 0 & \dots \\ 1 - r_3 & 0 & 0 & 0 & r_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \end{matrix} \quad (2)$$

이 Markov 체인의 상태추이도는 다음과 같다.



**예제 4.4.6** Ehrenfest 체인

항아리(urn) 두 개가 있고, 좌측 항아리와 우측 항아리에 들어있는 공들의 합이  $N$ 이라고 하자. 이  $N$  개 공들 중에서 하나를 무작위로(at random) 선택한 다음, 다른 항아리에 넣기로 하자. 즉, 두 항아리들 중 하나를 무작위로 선택하고 이 항아리에서 공 하나를 뽑아 다른 항아리에 넣는 실험을 반복한다고 하자. 제  $n$  번째 공을 뽑은 직후 좌측 항아리에 들어있는 공들의 개수를 확률변수  $X_n$ 이라 하자. 이 확률변수의 상태공간은  $S = \{0, 1, 2, \dots, N - 1, N\}$

이다. 각  $i, i_0, i_1, \dots, i_{n-1} (\in S)$  에 대해서 다음 식이 성립한다.

$$\Pr(X_{n+1} = i + 1 | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \frac{N - i}{N} \quad (1)$$

따라서, 확률과정  $\{X_n | n = 0, 1, \dots\}$  의 추이확률들은 다음 식들을 만족한다.

$$p(0, 1) = 1 \quad (2)$$

$$p(N, N - 1) = 1 \quad (3)$$

또한, 각  $i (= 1, 2, \dots, N - 1)$  에 대해서 다음 식이 성립한다.

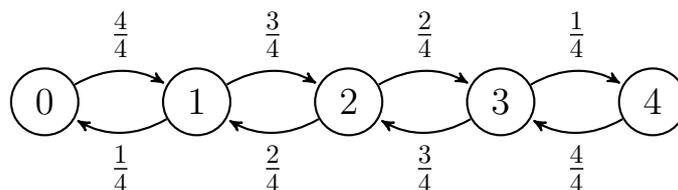
$$p(i, j) = \begin{cases} \frac{N-i}{N}, & (j = i + 1) \\ \frac{i}{N}, & (j = i - 1) \\ 0, & (\text{otherwise}) \end{cases} \quad (4)$$

이러한 Markov체인을 Ehrenfest 체인이라 부른다.

예를 들어,  $N = 4$  인 경우를 살펴보자. 이 Ehrenfest 체인의 추이확률행렬은 다음과 같다.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 0 & \frac{4}{4} & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 \\ 0 & \frac{2}{4} & 0 & \frac{2}{4} & 0 \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 & \frac{4}{4} & 0 \end{bmatrix} \end{matrix} \quad (5)$$

이 추이확률행렬에서는 제 1 대각행렬들의 패턴을 강조하기 위해서 1 대신  $\frac{4}{4}$  를 사용했음에 유의하라. 이 Markov 체인의 상태추이도는 다음과 같다.





서로 독립인 확률변수열  $\{Y_k | k = 0, 1, \dots\}$  과 이와 독립인 확률변수  $X_0$  에 대해서 다음 확률변수들을 정의하자.

$$X_{n+1} = f(X_n, Y_{n+1}), \quad (n = 0, 1, \dots) \quad (4.4.1)$$

식 (4.4.1) 에서 알 수 있듯이,  $X_1$  은  $X_0$  와  $Y_1$  의 함수이고,  $X_2$  는  $X_0$  와  $Y_1, Y_2$  의 함수이고,  $X_3$  는  $X_0$  와  $Y_1, Y_2, Y_3$  의 함수이다. 이와같은 과정을 반복하면,  $X_n$  은  $X_0$  와  $Y_1, Y_2, \dots, Y_n$  의 함수이다. 다음 식들이 성립한다.

$$\begin{aligned} & \Pr(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) \\ &= \Pr(f(X_n, Y_{n+1}) = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) \\ &= \Pr(f(i, Y_{n+1}) = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) \\ &= \Pr(f(i, Y_{n+1}) = j) \\ &= \Pr(f(X_n, Y_{n+1}) = j | X_n = i) \end{aligned} \quad (4.4.2)$$

여기서 세 번째 등호는  $Y_{n+1}$  과  $X_0, X_1, \dots, X_n$  이 독립이어서 성립한다. 식 (4.4.2) 에서 알 수 있듯이, 다음 식이 성립한다.

$$\begin{aligned} & \Pr(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) \\ &= \Pr(X_{n+1} = j | X_n = i) \end{aligned} \quad (4.4.3)$$

즉,  $\{X_n | n = 0, 1, \dots\}$  은 Markov 체인이다. 만약  $\{Y_k | k = 0, 1, \dots\}$  가 동일한 확률분포를 따르면, 다음 식들이 성립한다.

$$\Pr(X_{n+1} = j | X_n = i) = \Pr(f(i, Y_{n+1}) = j) = \Pr(f(i, Y_1) = j) \quad (4.4.4)$$

즉, 다음 식이 성립한다.

$$\Pr(X_{n+1} = j | X_n = i) = \Pr(X_1 = j | X_0 = i) \quad (4.4.5)$$

즉, Markov 체인  $\{X_n\}$  은 시간동질적이다.

**예제 4.4.7** 무작위공급

어떤 신문사에서 매일 신문용지 1톤을 사용한다고 하자. 신문용지 공급회사는 제  $n$  번째 달에 신문용지  $Y_n$  톤을 신문사에 규칙적으로 공급한다고 하자. 이를 정규공급(regular supply)이라고 하자. 여기서 확률변수들  $\{Y_k | k = 0, 1, 2, \dots\}$  는 서로 독립이고 다음 식들을 만족하는 동일한 확률분포를 따른다고 하자.

$$a_m \doteq \Pr(Y_k = m), \quad (m = 0, 1, 2, \dots) \quad (1)$$

또한, 아침에 신문사에 신문용지 재고가 0인 경우에는 정규공급이 되기 전에 그리고 인쇄를 시작하기 전에 신문용지 1톤을 비상공급(emergency supply)한다고 하자. 제  $n$  번째 달 아침 신문사가 가지고 있는 신문용지 양를  $X_n$  이라 하고, 정규공급량  $Y_n$  은 재고량  $X_n$  과 독립이라고 가정하자. 다음 식이 성립함을 쉽게 알 수 있다.

$$X_{n+1} = \begin{cases} X_n - 1 + Y_n, & (X_n > 0) \\ 1 - 1 + Y_n, & (X_n = 0) \end{cases} \quad (2)$$

만약  $X_0$  와  $\{Y_k | k = 0, 1, 2, \dots\}$  가 서로 독립적이면, 식 (4.4.5) 에서 알 수 있듯이, 확률과정  $\{X_n | n = 0, 1, \dots\}$  은 시간동질적 Markov 체인이다. 또한 추이확률행렬은 다음과 같다.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & \dots \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{matrix} & \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & a_0 & a_1 & a_2 & \dots \\ 0 & 0 & a_0 & a_1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{matrix} \quad (3)$$

■

**예제 4.4.8** 재고관리

상점 A에서 제  $n$  번째 달에 상품 B를  $D_n$  개 판다고 하자. 확률변수들  $\{D_n | n = 1, 2, \dots\}$

은 서로 독립이며 다음 식들을 만족하는 동일한 확률분포를 따른다고 하자.

$$\Pr(D_n = m) = a_m (\geq 0), \quad (m = 0, 1, \dots), \quad \sum_{m=0}^{\infty} a_m = 1 \quad (1)$$

초기시점에 이 상점이 가지고 있는 상품 B의 재고를  $X_0$  라고 하고, 제  $n$  번째 달 영업을 마친 후 상품 B의 재고를  $X_n$  이라고 하자. 이 상점에서는 상품 B의 재고  $X_n$  이  $s$  이하이면, 재고를  $S$  로 유지하기 위해서 상품 B를 주문한다고 하자. 여기서  $s$  와  $S$  는 음이 아닌 정수들이다.

각  $n \in \mathbb{Z}_{\geq 0}$  에 대해서 다음 식이 성립함을 쉽게 알 수 있다.

$$X_{n+1} = \begin{cases} [X_n - D_{n+1}]^+, & (X_n > s) \\ [S - D_{n+1}]^+, & (X_n \leq s) \end{cases} \quad (2)$$

여기서  $[A]^+ \doteq \max\{A, 0\}$  이다. 식 (2) 에서 알 수 있듯이,  $X_n$  이 주어진 조건 하에서  $X_{n+1}$  은  $D_{n+1}$  과  $X_n$  에만 의존한다. 만약  $X_0$  와  $\{D_n \mid n = 1, 2, \dots\}$  이 서로 독립적이면, 식 (4.4.5) 에서 알 수 있듯이, 확률과정  $\{X_n \mid n = 0, 1, \dots\}$  은 시간동질적 Markov체인이다. 또한, 추이확률은 다음과 같다.

$$p(i, j) = \begin{cases} \sum_{k=i}^{\infty} a_k, & (i \geq s \text{ and } j = 0) \\ a_{i-j}, & (i \geq s \text{ and } i \geq j) \\ \sum_{k=s}^{\infty} a_k, & (i < s \text{ and } j = 0) \\ a_{s-j}, & (i < s \text{ and } s \geq j > 0) \\ 0 & (\text{otherwise}) \end{cases} \quad (3)$$



**예제 4.4.9** 예제 4.4.8의 특별한 경우를 살펴보자. 월요일에서 토요일까지 영업을 하는 어떤 카메라가게에서 제  $n$  번째 주에 캐논 EOS-1D Mark IV를  $D_n$  대 판다고 하자. 확률변수열  $\{D_n \mid n = 1, 2, \dots\}$  은 서로 독립이고 동일한 확률분포를 따른다고 하자. 처음 시점에 이 카메라가게가 가지고 있는 Mark IV의 재고를  $X_0$  라고 하고, 제  $n$  번째 주 토요일 영업을 마친 후 Mark IV의 재고를  $X_n$  이라고 하자. 토요일 영업을 마친 후 캐논사에 Mark IV를 주문하면, 월요일 영업을 시작하기 전에 주문한 카메라가 배달된다고 하자. 이 카메라가게에서는 Mark IV의 재고  $X_n$  이 0이면, 재고를 3으로 유지하기 위해서 캐논사에 Mark IV를 주문한다고

하자. 각  $n \in \mathbb{Z}_{\geq 0}$  에 대해서 다음 식이 성립함을 쉽게 알 수 있다.

$$X_{n+1} = \begin{cases} [X_n - D_{n+1}]^+, & (X_n > 0) \\ [3 - D_{n+1}]^+, & (X_n = 0) \end{cases} \quad (1)$$

또한, 확률변수  $D_{n+1}$  의 확률밀도함수가 다음과 같다고 하자.

$$\Pr(D_{n+1} = m) = e^{-1} \frac{1}{m!}, \quad (m = 0, 1, 2, \dots) \quad (2)$$

즉, 확률변수  $D_{n+1}$  은 평균이 1인 Poisson 확률분포를 따른다. 즉, 다음 식들이 성립한다.

$$\Pr(D_{n+1} = m) = \begin{cases} \frac{1}{e} = 0.368, & (m = 0) \\ \frac{1}{e} = 0.368, & (m = 1) \\ \frac{1}{2e} = 0.184, & (m = 2) \\ 1 - \frac{5}{2e} = 0.080, & (m \geq 3) \end{cases} \quad (3)$$

식 (1)에서 알 수 있듯이,  $X_n = 0$  이면, 다음 식이 성립한다.

$$X_{n+1} = [3 - D_{n+1}]^+ \quad (4)$$

따라서, 다음 식들이 성립한다.

$$p(0, 3) = \Pr(X_{n+1} = 3 | X_n = 0) = \Pr(D_{n+1} = 0) = 0.368 \quad (5)$$

$$p(0, 2) = \Pr(X_{n+1} = 2 | X_n = 0) = \Pr(D_{n+1} = 1) = 0.368 \quad (6)$$

$$p(0, 1) = \Pr(X_{n+1} = 1 | X_n = 0) = \Pr(D_{n+1} = 2) = 0.184 \quad (7)$$

$$p(0, 0) = 1 - \sum_{j=1}^3 p(0, j) = 0.080 \quad (8)$$

만약  $X_n = 1$  이면, 다음 식이 성립한다.

$$X_{n+1} = [1 - D_{n+1}]^+ \quad (9)$$

따라서, 다음 식들이 성립한다.

$$p(1, 3) = 0 \quad (10)$$

$$p(1, 2) = 0 \quad (11)$$

$$p(1, 1) = \Pr(X_{n+1} = 1 | X_n = 1) = \Pr(D_{n+1} = 0) = 0.368 \quad (12)$$

$$p(1, 0) = 1 - \sum_{j=1}^3 p(1, j) = 0.632 \quad (13)$$

만약  $X_n = 2$ 이면, 다음 식이 성립한다.

$$X_{n+1} = [2 - D_{n+1}]^+ \quad (14)$$

따라서, 다음 식들이 성립한다.

$$p(2, 3) = 0 \quad (15)$$

$$p(2, 2) = \Pr(X_{n+1} = 2 | X_n = 2) = \Pr(D_{n+1} = 0) = 0.368 \quad (16)$$

$$p(2, 1) = \Pr(X_{n+1} = 1 | X_n = 2) = \Pr(D_{n+1} = 1) = 0.368 \quad (17)$$

$$p(2, 0) = 1 - \sum_{j=1}^3 p(2, j) = 0.264 \quad (18)$$

만약  $X_n = 3$ 이면, 다음 식이 성립한다.

$$X_{n+1} = [3 - D_{n+1}]^+ \quad (19)$$

식 (3) 과 식 (19) 에서 알 수 있듯이, 다음 식들이 성립한다.

$$p(3, 3) = \Pr(X_{n+1} = 3 | X_n = 3) = \Pr(D_{n+1} = 0) = 0.368 \quad (20)$$

$$p(3, 2) = \Pr(X_{n+1} = 2 | X_n = 3) = \Pr(D_{n+1} = 1) = 0.368 \quad (21)$$

$$p(3, 1) = \Pr(X_{n+1} = 1 | X_n = 3) = \Pr(D_{n+1} = 2) = 0.184 \quad (22)$$

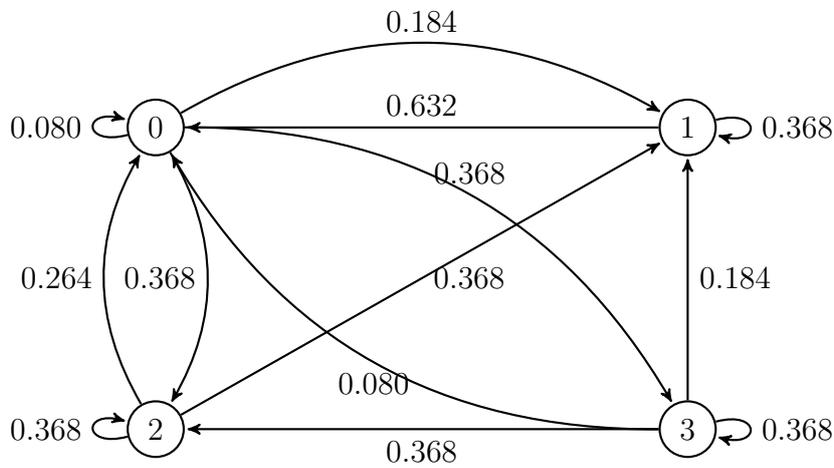
$$p(3, 0) = 1 - \sum_{j=1}^3 p(3, j) = 0.080 \quad (23)$$

지금까지 내용을 정리하면, 확률과정  $\{X_n | n = 0, 1, \dots\}$ 의 추이확률행렬이 다음과 같음을

알 수 있다.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{bmatrix} 0.080 & 0.184 & 0.368 & 0.368 \\ 0.632 & 0.368 & 0 & 0 \\ 0.264 & 0.368 & 0.368 & 0 \\ 0.080 & 0.184 & 0.368 & 0.368 \end{bmatrix} \end{matrix} \tag{24}$$

또한 이 Markov 체인의 상태추이도는 다음과 같다.



지금부터는 서로 독립이며 동일한 확률분포 (independent and identically distributed: iid) 를 따르는 확률변수들  $\{\xi_n | n = 1, 2, \dots\}$  을 바탕으로 하는 Markov 체인에 대해서 살펴보자.

**예제 4.4.10** 서로 독립인 확률변수들  $\{\xi_n | n = 1, 2, \dots\}$  이 다음 식들을 만족하는 동일한 확률분포를 따른다고 하자.

$$\Pr(\xi_n = m) = a_m (\geq 0), \quad (m = 0, 1, \dots), \quad \sum_{m=0}^{\infty} a_m = 1 \tag{1}$$

다음 확률변수들을 정의하자.

$$X_0 \doteq 0, \quad X_n \doteq \xi_n, \quad (n = 1, 2, \dots) \tag{2}$$

이렇게 정의되는 확률과정  $\{X_n | n = 0, 1, \dots\}$ 은 음이 아닌 정수들의 집합  $\mathbb{Z}_{\geq 0}$ 를 상태공간으로 하는 Markov체인이다. 이 Markov체인의 추이확률행렬은 다음과 같다.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & \dots \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{matrix} & \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{matrix} \quad (3)$$

■

**예제 4.4.11** 서로 독립인 확률변수들  $\{\xi_n | n = 1, 2, \dots\}$ 이 다음 식들을 만족하는 동일한 확률분포를 따른다고 하자.

$$\Pr(\xi_n = m) = a_m (\geq 0), \quad (m = 0, 1, \dots), \quad \sum_{m=0}^{\infty} a_m = 1 \quad (1)$$

다음 확률변수들을 정의하자.

$$X_0 \doteq 0, \quad X_n \doteq \sum_{k=0}^n \xi_k, \quad (n = 1, 2, \dots) \quad (2)$$

이렇게 정의되는 확률과정  $\{X_n | n = 0, 1, \dots\}$ 은 음이 아닌 정수들의 집합  $\mathbb{Z}_{\geq 0}$ 를 상태공간으로 하는 Markov체인이다. 이 Markov체인의 추이확률행렬은 다음과 같다.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & \dots \end{matrix} \\ \begin{matrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \end{matrix} & \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & a_0 + a_1 & a_2 & a_3 & \dots \\ 0 & 0 & a_0 + a_1 + a_2 & a_3 & \dots \\ 0 & 0 & 0 & a_0 + a_1 + a_2 + a_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{matrix} \quad (3)$$



**예제 4.4.12** 예제 4.4.11의 특별한 경우를 살펴보자. 서로 독립인 확률변수들  $\{\xi_n | n = 1, 2, \dots\}$ 이 다음 식들을 만족하는 동일한 확률분포를 따른다고 하자.

$$\Pr(\xi_n = 1) = 1 - \Pr(\xi_n = 0) = p \doteq 1 - q \tag{1}$$

다음 확률변수들을 정의하자.

$$X_0 \doteq 0, \quad X_n \doteq \sum_{k=1}^n \xi_k, \quad (n = 1, 2, \dots) \tag{2}$$

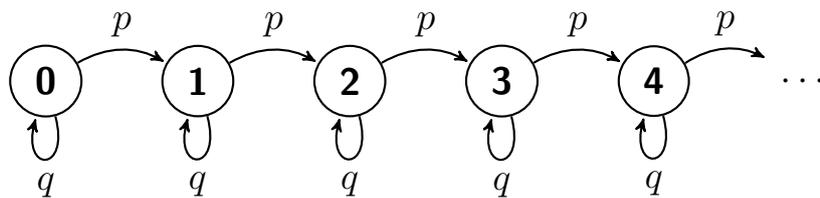
이렇게 정의되는 확률과정  $\{X_n | n = 0, 1, \dots\}$ 은 음이 아닌 정수들의 집합  $\mathbb{Z}_{\geq 0}$ 를 상태공간으로 하는 Markov 체인이다. 또한, 다음 식들이 성립한다.

$$X_0 \doteq 0, \quad X_n = X_{n-1} + \xi_n, \quad (n = 1, 2, \dots) \tag{3}$$

이 Markov 체인의 추이확률행렬은 다음과 같다.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 & \dots \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ \vdots \end{matrix} & \left[ \begin{array}{cccccc} q & p & 0 & 0 & 0 & \dots \\ 0 & q & p & 0 & 0 & \dots \\ 0 & 0 & q & p & 0 & \dots \\ 0 & 0 & 0 & q & p & \dots \\ 0 & 0 & 0 & 0 & q & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array} \right] \end{matrix} \tag{4}$$

이 Markov 체인의 상태추이도는 다음과 같다.





**예제 4.4.13** 단순확률보행

서로 독립인 확률변수들  $\{\xi_n | n = 1, 2, \dots\}$ 이 다음 식들을 만족하는 동일한 확률분포를 따른다고 하자.

$$\Pr(\xi_n = 1) = p, \quad \Pr(\xi_n = 0) = r, \quad \Pr(\xi_n = -1) = q \tag{1}$$

여기서  $p+q+r = 1$ 이다. 확률변수  $X_0$ 가  $\{\xi_n | n = 1, 2, \dots\}$ 과는 독립이며 정수값을 갖는다고 하고, 다음 확률변수들을 정의하자.

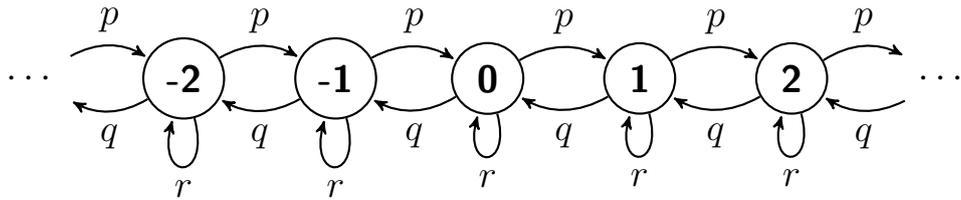
$$X_n \doteq X_0 + \sum_{k=1}^n \xi_k, \quad (n = 1, 2, \dots) \tag{2}$$

이렇게 정의되는 확률과정  $\{X_n | n = 0, 1, \dots\}$ 은 정수들의 집합  $Z$ 를 상태공간으로 하는 Markov체인이다. 이  $\{X_n\}$ 을 단순확률보행 (simple random walk)이라 부른다.

이 Markov체인의 추이확률행렬은 다음과 같다.

$$\mathbb{P} = \begin{matrix} & & & -2 & -1 & 0 & 1 & 2 & & & \\ & & & & & & & & & & \\ & & & & & & & & & & \\ \vdots & & & & & & & & & & \\ & & & & & & & & & & \\ -2 & \cdots & q & r & p & 0 & 0 & \cdots & & & \\ & & & & & & & & & & \\ -1 & \cdots & 0 & q & r & p & 0 & \cdots & & & \\ & & & & & & & & & & \\ 0 & \cdots & 0 & 0 & q & r & p & \cdots & & & \\ & & & & & & & & & & \\ 1 & \cdots & 0 & 0 & 0 & q & r & \cdots & & & \\ & & & & & & & & & & \\ 2 & \cdots & 0 & 0 & 0 & 0 & q & \cdots & & & \\ & & & & & & & & & & \\ \vdots & & & & & & & & & & \end{matrix} \tag{3}$$

이 Markov체인의 상태추이도는 다음과 같다.



**예제 4.4.14** 서로 독립인 확률변수들  $\{\xi_n | n = 1, 2, \dots\}$ 이 다음 식들을 만족하는 동일한 확률분포를 따른다고 하자.

$$\Pr(\xi_n = m) = a_m (\geq 0), \quad (m = 0, 1, \dots), \quad \sum_{m=0}^{\infty} a_m = 1 \quad (1)$$

다음 확률변수들을 정의하자.

$$X_0 \doteq 0, \quad X_n \doteq \max\{\xi_1, \xi_2, \dots, \xi_n\}, \quad (n = 1, 2, \dots) \quad (2)$$

각  $n \in \mathbb{Z}_{\geq 0}$ 에 대해서, 다음 식이 성립한다.

$$X_{n+1} = \max\{X_n, \xi_{n+1}\}, \quad (n = 1, 2, \dots) \quad (3)$$

따라서, 확률과정  $\{X_n | n = 0, 1, \dots\}$ 은 음이 아닌 정수들의 집합  $\mathbb{Z}_{\geq 0}$ 를 상태공간으로 하는 Markov 체인이다. 이 Markov 체인의 추이확률행렬은 다음과 같다.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{matrix} & \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ 0 & 0 & 0 & a_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{matrix} \quad (4)$$



## 제4.5절 Chapman-Kolmogorov 식

이후 특별한 언급이 없는 한 Markov 체인의 상태공간을  $\mathbb{Z}_{\geq 0} = \{0, 1, 2, \dots\}$  라고 하자. 각  $i (= 0, 1, \dots)$  에 대해서, 다음 식이 성립함은 명백하다.

$$\sum_{j=0}^{\infty} p(i, j) = 1 \quad (4.5.1)$$

식 (4.5.1) 을 다음과 같이 쓸 수 있다.

$$\mathbb{P}\mathbf{1} = \mathbf{1} \quad (4.5.2)$$

여기서  $\mathbf{1}$  은 모든 원소들이 1 인 열벡터이다. 식 (4.5.2) 는 열벡터  $\mathbf{1}$  이 추이확률행렬  $\mathbb{P}$  의 우고유벡터 (right eigenvector) 이고 이에 해당하는 고유값이 1 임을 의미한다. 사실, 1 은 추이확률행렬  $\mathbb{P}$  의 고유값들 중에서 절대값이 가장 큰 것이다.

Markov 체인이 시점  $m$  의 상태  $i$  에서 시점  $[m+n]$  의 상태  $j$  로 추이하는 확률을 제  $n$  단계 추이확률이라 하고, 이를  $p^{(n)}(i, j)$  로 표기하자. 각  $n (\in \mathbb{Z}_{\geq 0})$  에 대해서, 다음 식들이 성립한다.

$$p^{(n)}(i, j) \doteq \Pr(X_{n+m} = j \mid X_m = i), \quad (i, j \in \mathbb{Z}_{\geq 0}) \quad (4.5.3)$$

**예제 4.5.1** 독립동일확률변수들  $\{\xi_n \mid n = 1, 2, \dots\}$  의 확률질량함수가 다음과 같다고 하자.

$$\Pr(\xi = 1) = p, \quad \Pr(\xi = -1) = 1 - p \doteq q \quad (1)$$

다음 확률변수들을 정의하자.

$$X_0 = 0, \quad X_n = \sum_{k=1}^n \xi_k, \quad (n = 1, 2, \dots) \quad (2)$$

각 자연수  $n$  에 대해서  $X_n$  은 이항확률분포를 따르고,  $\{X_n \mid n = 0, 1, \dots\}$  은 확률보행과정이다. 따라서,  $\{X_n\}$  은 Markov 체인이다. 만약  $X_n = j - i$  이면,  $\{\xi_k \mid k = 1, 2, \dots, n\}$  중에서 값  $-1$  을 갖는 개수는  $[n + j - i]/2$  이고 값  $0$  을 갖는 개수는  $[n - j + i]/2$  이다. 따라서, 다음

식이 성립한다.

$$p^{(n)}(i, j) = \binom{n}{\frac{n+j-i}{2}} p^{\frac{n+j-i}{2}} q^{\frac{n-j+i}{2}} \quad (3)$$

■

자연수들  $n$  과  $m$  에 대해서, 다음 식들이 성립함을 알 수 있다.

$$\begin{aligned} p^{(n+m)}(i, j) &= \Pr(X_{n+m} = j \mid X_0 = i) \\ &= \sum_{k=0}^{\infty} \Pr(X_{n+m} = j, X_m = k \mid X_0 = i) \\ &= \sum_{k=0}^{\infty} \Pr(X_{n+m} = j \mid X_m = k, X_0 = i) \Pr(X_m = k \mid X_0 = i) \\ &= \sum_{k=0}^{\infty} \Pr(X_{n+m} = j \mid X_m = k) \Pr(X_m = k \mid X_0 = i) \end{aligned} \quad (4.5.4)$$

여기서 네 번째 등호는 Markov성에 의해서 성립한다. 식 (4.5.4)로부터 다음 정리가 성립함을 알 수 있다.

#### 정리 4.5.1: Chapman-Komogorov 식

Markov체인  $\{X_n \mid n = 0, 1, \dots\}$  의 제  $l$  단계 추이확률들  $\{p^{(l)}(i, j)\}$  는 각  $m, n \in \mathbb{Z}_{\geq 0}$  에 대해서 다음 식을 만족한다.

$$p^{(m+n)}(i, j) = \sum_{k=0}^{\infty} p^{(m)}(i, k) p^{(n)}(k, j), \quad (i, j \in \mathbb{Z}_{\geq 0})$$

정리 4.5.1에서 알 수 있듯이, 제  $m+n$  단계 추이확률행렬  $\mathbb{P}^{(m+n)} \doteq [p^{(m+n)}(i, j)]$  는 다음 식을 만족한다.

$$\mathbb{P}^{(m+n)} = \mathbb{P}^{(m)} \mathbb{P}^{(n)} \quad (4.5.5)$$

식  $\mathbb{P}^{(1)} = \mathbb{P}$ 이 성립함을 자명하다. 식 (4.5.5)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\mathbb{P}^{(n)} = \mathbb{P}^n, \quad (n = 1, 2, \dots) \quad (4.5.6)$$

즉, 제  $n$  단계 추이확률행렬  $\mathbb{P}^{(n)}$ 은 추이확률행렬  $\mathbb{P}$ 의  $n$ 제곱이다. 식 (4.5.5)와 식 (4.5.6)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\mathbb{P}^{(n+1)} = \mathbb{P}^{(n)}\mathbb{P}, \quad (n = 1, 2, \dots) \quad (4.5.7)$$

시점  $n$ 에서 확률변수  $X_n$ 이 상태  $x$ 에 존재할 확률, 즉 상태확률 (state probability)을 다음과 같이 정의하자.

$$\pi^{(n)}(i) \doteq \Pr(X_n = i), \quad (i = 0, 1, \dots) \quad (4.5.8)$$

또한, 행벡터인 확률벡터  $\boldsymbol{\pi}^{(n)}$ 을 다음과 같이 정의하자.

$$\boldsymbol{\pi}^{(n)} \doteq [\pi^{(n)}(0), \pi^{(n)}(1), \pi^{(n)}(2), \dots] \quad (4.5.9)$$

이  $\boldsymbol{\pi}^{(n)}$ 을 제  $n$  단계 상태확률벡터라 부른다.

각 시점  $n$ 의 각 상태  $j$ 에 대해 다음 식이 성립한다.

$$\Pr(X_n = j) = \sum_{i=0}^{\infty} \Pr(X_0 = i) \Pr(X_n = j | X_0 = i) \quad (4.5.10)$$

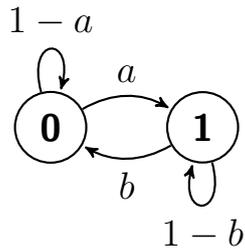
식 (4.5.6)과 식 (4.5.10)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(0)}\mathbb{P}^{(n)} = \boldsymbol{\pi}^{(0)}\mathbb{P}^n, \quad (n = 1, 2, \dots) \quad (4.5.11)$$

**예제 4.5.2** 상태공간이  $S = \{0, 1\}$  인 Markov체인의 추이확률행렬이 다음과 같다고 하자.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \end{matrix} \quad (1)$$

이 Markov체인의 상태추이도는 다음과 같다.



식 (4.5.7)에서 알 수 있듯이, 각  $n(=0, 1, \dots)$ 에 대해서 다음 식이 성립한다.

$$\begin{bmatrix} p^{(n+1)}(0,0) & p^{(n+1)}(0,1) \\ p^{(n+1)}(1,0) & p^{(n+1)}(1,1) \end{bmatrix} = \begin{bmatrix} p^{(n)}(0,0) & p^{(n)}(0,1) \\ p^{(n)}(1,0) & p^{(n)}(1,1) \end{bmatrix} \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \quad (2)$$

즉, 다음 식들이 성립한다.

$$p^{(n+1)}(0,0) = [1-a]p^{(n)}(0,0) + bp^{(n)}(0,1) \quad (3)$$

$$p^{(n+1)}(0,1) = ap^{(n)}(0,0) + [1-b]p^{(n)}(0,1) \quad (4)$$

$$p^{(n+1)}(1,0) = [1-a]p^{(n)}(1,0) + bp^{(n)}(1,1) \quad (5)$$

$$p^{(n+1)}(1,1) = ap^{(n)}(1,0) + [1-b]p^{(n)}(1,1) \quad (6)$$

식 (3)과 식 (4)에서 알 수 있듯이, 다음 식들이 성립한다.

$$p^{(n+1)}(0,0) + p^{(n+1)}(0,1) = p^{(n)}(0,0) + p^{(n)}(0,1) = \dots = p^{(1)}(0,0) + p^{(1)}(0,1) \quad (7)$$

따라서, 각  $n(=0, 1, \dots)$ 에 대해서 다음 식이 성립한다.

$$p^{(n)}(0,0) + p^{(n)}(0,1) = 1 \quad (8)$$

식 (8)을 식 (3)에 대입하면, 다음 식을 얻는다.

$$p^{(n+1)}(0, 0) = [1 - a]p^{(n)}(0, 0) + b [1 - p^{(n)}(0, 0)] \quad (9)$$

점화식 (9)의 고정값  $z$ 는 다음 식을 만족한다.

$$z = [1 - a]z + b[1 - z] \quad (10)$$

즉, 고정값은  $z = b/[a + b]$ 이다. 식 (9)의 양변에서 이 고정값을 빼면, 다음 식들이 성립함을 알 수 있다.

$$p^{(n+1)}(0, 0) - \frac{b}{a+b} = [1 - a - b] \left[ p^{(n)}(0, 0) - \frac{b}{a+b} \right], \quad (n = 0, 1, \dots) \quad (11)$$

식 (11)에서 알 수 있듯이, 다음 식들이 성립한다.

$$p^{(n)}(0, 0) - \frac{b}{a+b} = [1 - a - b]^n \left[ p^{(0)}(0, 0) - \frac{b}{a+b} \right], \quad (n = 0, 1, \dots) \quad (12)$$

정의에서 알 수 있듯이, 식  $p^{(0)}(0, 0) = 1$ 이 성립한다. 따라서, 식 (12)에서 알 수 있듯이 다음 식들이 성립한다.

$$p^{(n)}(0, 0) = \frac{b}{a+b} + \frac{a}{a+b}[1 - a - b]^n, \quad (n = 0, 1, \dots) \quad (13)$$

식 (8)과 식 (13)에서 알 수 있듯이, 다음 식들이 성립한다.

$$p^{(n)}(0, 1) = \frac{a}{a+b} - \frac{a}{a+b}[1 - a - b]^n, \quad (n = 0, 1, \dots) \quad (14)$$

식 (13)과 식 (14)에서  $a$ 와  $b$ 를 교환하면, 다음 식들을 얻는다.

$$p^{(n)}(1, 1) = \frac{a}{a+b} + \frac{b}{a+b}[1 - a - b]^n, \quad (n = 0, 1, \dots) \quad (15)$$

$$p^{(n)}(1, 0) = \frac{b}{a+b} - \frac{b}{a+b}[1 - a - b]^n, \quad (n = 0, 1, \dots) \quad (16)$$

식 (13)~식 (16)에서 알 수 있듯이, 제 $n$  단계 추이확률행렬은 다음과 같다.

$$\mathbb{P}^{(n)} = \frac{1}{a+b} \begin{bmatrix} b+a[1-a-b]^n & a-a[1-a-b]^n \\ b-b[1-a-b]^n & a+b[1-a-b]^n \end{bmatrix} \quad (17)$$

따라서, 만약 조건  $0 < a+b < 2$ 가 만족되면, 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} \mathbb{P}^{(n)} = \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{bmatrix} \quad (18)$$

식 (4.5.11)과 식 (17)에서 알 수 있듯이, 제 $n$  단계 상태확률들은 다음과 같다.

$$\pi^{(n)}(0) = \frac{b}{a+b} + \left[ \pi^{(0)}(0) - \frac{b}{a+b} \right] [1-a-b]^n \quad (19)$$

$$\pi^{(n)}(1) = \frac{a}{a+b} + \left[ \pi^{(0)}(1) - \frac{a}{a+b} \right] [1-a-b]^n \quad (20)$$

만약 조건  $0 < a+b < 2$ 가 만족되면, 다음 식들이 성립한다.

$$\lim_{n \rightarrow \infty} \pi^{(n)}(0) = \frac{b}{a+b}, \quad \lim_{n \rightarrow \infty} \pi^{(n)}(1) = \frac{a}{a+b} \quad (21)$$

식 (21)에서 알 수 있듯이, 극한상태확률들은 초기시점의 상태확률들  $\pi^{(0)}(0)$ 와  $\pi^{(0)}(1)$ 에 의존하지 않는다. ■

**예제 4.5.3** 예제 4.5.2를 약간 달리 생각해 보자. 서울대입구역에서 서울대 행정관까지 학교버스를 타고 온 학생이 다음에 시내버스를 탈 확률이  $\alpha$ 이고, 시내버스를 타고 온 학생이 다음에 학교버스를 탈 확률이  $\beta$ 라고 하자. 여기서  $\alpha$ 와  $\beta$ 는 다음 식들을 만족한다고 가정하자.

$$0 \leq \alpha \leq 1, \quad 0 \leq \beta \leq 1, \quad \alpha + \beta > 0 \quad (1)$$

시점  $n$ 에서 고객이 학교버스를 타면  $X_n = 0$ 라 하고, 시내버스를 타면  $X_n = 1$ 이라 하면,

$\{X_n\}$ 은 추이확률행렬이 다음과 같은 Markov 체인이다.

$$\mathbb{P} = \begin{array}{cc} & \begin{array}{cc} School & Public \end{array} \\ \begin{array}{c} School \\ Public \end{array} & \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix} \end{array} \quad (2)$$

다음 식이 성립한다.

$$\mathbb{P}^{(m+1)} = \mathbb{P} \mathbb{P}^{(m)} \quad (3)$$

따라서, 다음 식들이 성립한다.

$$p^{(m+1)}(0, 0) = [1 - \alpha]p^{(m)}(0, 0) + \alpha p^{(m)}(1, 0) \quad (4)$$

$$p^{(m+1)}(0, 1) = [1 - \alpha]p^{(m)}(0, 1) + \alpha p^{(m)}(1, 1) \quad (5)$$

$$p^{(m+1)}(1, 0) = \beta p^{(m)}(0, 0) + [1 - \beta]p^{(m)}(1, 0) \quad (6)$$

$$p^{(m+1)}(1, 1) = \beta p^{(m)}(0, 1) + [1 - \beta]p^{(m)}(1, 1) \quad (7)$$

식 (4)와 식 (6)에서 알 수 있듯이, 다음 식이 성립한다.

$$p^{(m+1)}(0, 0) - p^{(m+1)}(1, 0) = [1 - \alpha - \beta] [p^{(m)}(0, 0) - p^{(m)}(1, 0)] \quad (8)$$

즉, 다음 식들이 성립한다.

$$p^{(m)}(0, 0) - p^{(m)}(1, 0) = [p^{(1)}(0, 0) - p^{(1)}(1, 0)] [1 - \alpha - \beta]^{m-1} = [1 - \alpha - \beta]^m \quad (9)$$

같은 방법으로, 다음 식들이 성립함을 알 수 있다.

$$p^{(m)}(0, 1) - p^{(m)}(1, 1) = [p^{(1)}(0, 1) - p^{(1)}(1, 1)] [1 - \alpha - \beta]^{m-1} = -[1 - \alpha - \beta]^m \quad (10)$$

식 (4)를 다음과 같이 쓸 수 있다.

$$p^{(m+1)}(0, 0) - p^{(m)}(0, 0) = \alpha [p^{(m)}(1, 0) - p^{(m)}(0, 0)] \quad (11)$$

식 (9)를 식 (11)에 대입하면, 다음 식을 얻는다.

$$p^{(m+1)}(0, 0) = p^{(m)}(0, 0) - \alpha[1 - \alpha - \beta]^m \quad (12)$$

식 (12)에서 알 수 있듯이, 각 자연수  $n$ 에 대해서 다음 식들이 성립한다.

$$p^{(n)}(0, 0) = 1 - \alpha - \sum_{k=1}^{n-1} \alpha[1 - \alpha - \beta]^k = \frac{\beta}{\alpha + \beta} + \frac{\alpha}{\alpha + \beta}[1 - \alpha - \beta]^n \quad (13)$$

식 (13)을 식 (9)에 대입하면, 다음 식을 얻는다.

$$p^{(n)}(1, 0) = \frac{\beta}{\alpha + \beta} - \frac{\beta}{\alpha + \beta}[1 - \alpha - \beta]^n \quad (14)$$

다음 식들이 성립한다.

$$p^{(n)}(0, 1) = 1 - p^{(n)}(0, 0) = \frac{\alpha}{\alpha + \beta} - \frac{\alpha}{\alpha + \beta}[1 - \alpha - \beta]^n \quad (15)$$

여기서 두 번째 등호는 식 (13)에 의해서 성립한다. 식 (15)를 식 (10)에 대입하면, 다음 식을 얻는다.

$$p^{(n)}(1, 1) = \frac{\alpha}{\alpha + \beta} + \frac{\beta}{\alpha + \beta}[1 - \alpha - \beta]^n \quad (16)$$

지금까지 내용을 정리하면, 다음과 같다.

$$\mathbb{P}^{(n)} = \frac{1}{\alpha + \beta} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{[1 - \alpha - \beta]^n}{\alpha + \beta} \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} \quad (17)$$

행렬  $\mathbb{P}$ 의 고유방정식은 다음과 같다.

$$\det(\mathbb{P} - \lambda I) = \begin{vmatrix} 1 - \alpha - \lambda & \alpha \\ \beta & 1 - \beta - \lambda \end{vmatrix} = \{\lambda - [1 - \alpha - \beta]\} \{\lambda - 1\} \quad (18)$$

따라서, 행렬  $\mathbb{P}$ 의 고유값들은 1과  $1 - \alpha - \beta$ 이다. 또한, 이에 해당하는 고유벡터들은 각각

다음과 같다.

$$\mathbf{x}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} \frac{\alpha}{\sqrt{\alpha^2 + \beta^2}} \\ \frac{-\beta}{\sqrt{\alpha^2 + \beta^2}} \end{bmatrix} \quad (19)$$

다음 행렬들을 정의하자.

$$\Lambda \doteq \begin{bmatrix} 1 & 0 \\ 0 & 1 - \alpha - \beta \end{bmatrix}, \quad E \doteq \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{\alpha}{\sqrt{\alpha^2 + \beta^2}} \\ \frac{1}{\sqrt{2}} & \frac{-\beta}{\sqrt{\alpha^2 + \beta^2}} \end{bmatrix} \quad (20)$$

다음 식이 성립한다.

$$E^{-1} = \frac{1}{\alpha + \beta} \begin{bmatrix} \sqrt{2}\beta & \sqrt{2}\alpha \\ \sqrt{\alpha^2 + \beta^2} & -\sqrt{\alpha^2 + \beta^2} \end{bmatrix} \quad (21)$$

또한, 다음 식이 성립한다.

$$\mathbb{P} = E\Lambda E^{-1} \quad (22)$$

식 (22)에서 알 수 있듯이, 다음 식들이 성립한다.

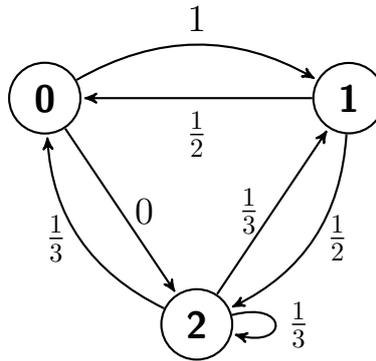
$$\mathbb{P}^n = E\Lambda^n E^{-1} = \frac{1}{\alpha + \beta} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{[1 - \alpha - \beta]^n}{\alpha + \beta} \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} \quad (23)$$

식 (17)과 식 (23)으로부터 식 (4.5.6)이 성립함을 확인할 수 있다. ■

**예제 4.5.4** 상태공간이  $S = \{0, 1, 2\}$ 인 Markov 체인의 추이확률행렬이 다음과 같다고 하자.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \end{matrix} \quad (1)$$

이 Markov 체인의 상태추이도는 다음과 같다.



이 Markov 체인의 초기시점 상태확률벡터가  $\pi^{(0)} = [0.1, 0.2, 0.7]$  이라 하자. 다음 식들이 성립한다.

$$\pi^{(1)} = \pi^{(0)}\mathbb{P} = [0.3333, 0.3333, 0.3333] \quad (2)$$

$$\pi^{(2)} = \pi^{(1)}\mathbb{P} = [0.2778, 0.4444, 0.2778] \quad (3)$$

$$\pi^{(3)} = \pi^{(2)}\mathbb{P} = [0.3148, 0.3704, 0.3148] \quad (4)$$

$$\pi^{(4)} = \pi^{(3)}\mathbb{P} = [0.3066, 0.3868, 0.3066] \quad (5)$$

또한, 다음 식들이 성립한다.

$$\pi^{(17)} = \pi^{(16)}\mathbb{P} = [0.3001, 0.3999, 0.3001] \quad (6)$$

$$\pi^{(18)} = \pi^{(17)}\mathbb{P} = [0.3000, 0.4001, 0.3000] \quad (7)$$

$$\pi^{(19)} = \pi^{(18)}\mathbb{P} = [0.3000, 0.4000, 0.3000] \quad (8)$$

$$\pi^{(20)} = \pi^{(19)}\mathbb{P} = [0.3000, 0.4000, 0.3000] \quad (9)$$

식 (2)~식 (9)에서 다음 식들이 성립함을 유추할 수 있다.

$$\lim_{n \rightarrow \infty} \pi^{(n)}(0) = 0.3, \quad \lim_{n \rightarrow \infty} \pi^{(n)}(1) = 0.4, \quad \lim_{n \rightarrow \infty} \pi^{(n)}(2) = 0.3 \quad (10)$$

식 (10)을 증명하기 위해서, 행렬  $\mathbb{P}$ 를 Schur분해하면 다음과 같다.

$$\mathbb{P} = Q\Lambda Q^{-1}, \quad \Lambda \doteq \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{2}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad Q \doteq \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{9}{\sqrt{118}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{-6}{\sqrt{118}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{118}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad (11)$$

따라서, 다음 식들이 성립한다.

$$\lim_{n \rightarrow \infty} \mathbb{P}^n = \lim_{n \rightarrow \infty} Q\Lambda^n Q^{-1} = Q \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} Q^{-1} = \begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} \quad (11)$$

즉, 다음 식들이 성립한다.

$$\lim_{n \rightarrow \infty} \pi^{(n)} = \lim_{n \rightarrow \infty} \pi^{(0)}\mathbb{P}^n = \pi^{(0)} \lim_{n \rightarrow \infty} \mathbb{P}^n = [0.3, 0.4, 0.3] \quad (12)$$

식 (12)에서 알 수 있듯이, 극한상태확률들은 초기시점의 상태확률들에 의존하지 않는다.

이 Markov체인의 상태확률벡터들을 구하기 위해서, 다음 MATLAB 프로그램 MarkovChain001.m을 실행하라.

```

1 % -----
2 % Filename: MarkovChain001.m
3 % Chapman-Kolmogorov Equation
4 % Programmed by CBS
5 % -----
6 clear all, close all
7 P = [ 0 1 0 ; 1/2 0 1/2 ; 1/3 1/3 1/3 ]
8 pii(1,:) = [0.1 0.2 0.7 ];
9 nMax = 20;
10 for kk = 1:nMax
11     pii(kk+1,:) = pii(kk,:)*P;
12 end
13 pii1 = pii(:,1); pii2 = pii(:,2); pii3 = pii(:,3);
14 pii12 = pii1+pii2;
15 nn = (0:nMax)';
16 disp('      n      \pi(1)      \pi(2)      \pi(3)')
17 disp([nn pii])
18 % Plotting
19 plot(nn,pii1,'g--',nn,pii2,'r-',nn,pii3,'k:', 'linewidth',2)
20 set(gca,'fontsize',11,'fontweigh','bold','ylim',[0,0.8])
21 legend('\pi^{(n)}(1)', '\pi^{(n)}(2)', '\pi^{(n)}(3)')
22 xlabel('n')
23 text(-3,0.45, '\bf \pi^{(n)}(i)')
24 saveas(gcf, 'MarkovChain001.jpg')
25 % Eigen-Analysis of the Matrix P
26 [ Q Lambda ] = eig(P)
27 invQ = inv(Q)

```

```

28 DC = Q*Lambda*invQ - P
29 % End of program
30 % -----

```

이 MATLAB 프로그램 MarkovChain001.m을 실행하면, 이 Markov체인의 상태확률벡터들을 그린 그림 4.5.1을 출력한다. 그림 4.5.1에서 식 (12)가 성립함을 확인할 수 있다.

다음 Python 프로그램 MarkovChain001.py를 실행하면, MATLAB 프로그램 MarkovChain001.m을 실행한 것과 동일한 결과를 얻는다.

```

1 # MarkovChain001.py
2 """
3 Filename : MarkovChain001.py
4 Chapman-Kolmogorov Equation
5 Programmed by Hwang Seol Woong based on the R program by CBS
6 """
7
8 from numpy import *
9 from numpy import linalg as LA
10 import pandas as pd
11 import matplotlib.pyplot as plt
12
13 P = matrix([[0.0, 1.0, 0.0],[0.5, 0.0, 0.5],[1/3,1/3,1/3]])
14 nMax = 20
15 pii = matrix(zeros((nMax,3)))
16 pii[0,:] = matrix([0.1, 0.2, 0.7])
17
18 for kk in range(nMax-1):
19     pii[kk+1,:] = pii[kk,:]*P
20
21
22
23 # Making DataFrame
24
25 pii1 = array(pii[:,0]).T[0]
26 pii2 = array(pii[:,1]).T[0]
27 pii3 = array(pii[:,2]).T[0]
28 pii12 = pii1 + pii2
29 nn = array([arange(0,nMax)])[0]
30
31 # create a dictionary
32 ex_dict = {
33     '$ \pi^{(n)}(1)$': pii1,
34     '$ \pi^{(n)}(2)$': pii2,
35     '$ \pi^{(n)}(3)$': pii3
36 }
37
38 # create a list of strings
39 columns = ['$ \pi^{(n)}(1)$', '$ \pi^{(n)}(2)$', '$ \pi^{(n)}(3)$' ]
40
41 index = nn
42
43 # Passing a dictionary
44 # key: column name
45 # value: series of values
46 df = pd.DataFrame(ex_dict, columns=columns, index=nn)
47 df.head()

```

```

48
49
50
51
52 #plotting
53 fig, ax = plt.subplots(figsize=(16,10))
54
55 ax.plot(pii1, label='$ \pi^{(n)}(1)$', linewidth=4.0, ls='--')
56 ax.plot(pii2, label='$ \pi^{(n)}(2)$', linewidth=2.0)
57 ax.plot(pii3, label='$ \pi^{(n)}(3)$', linewidth=2.0)
58
59 ax.tick_params(axis="y",direction = "in")           # axis ticks inside
60 ax.tick_params(axis="x",direction = "in")           # axis ticks inside
61 ax.legend(loc='upper right')
62 ax.set_xlabel('n')
63 ax.set_ylabel(r'$ \pi^{(n)}(i) $')
64
65
66 # End of program
67 # -----

```

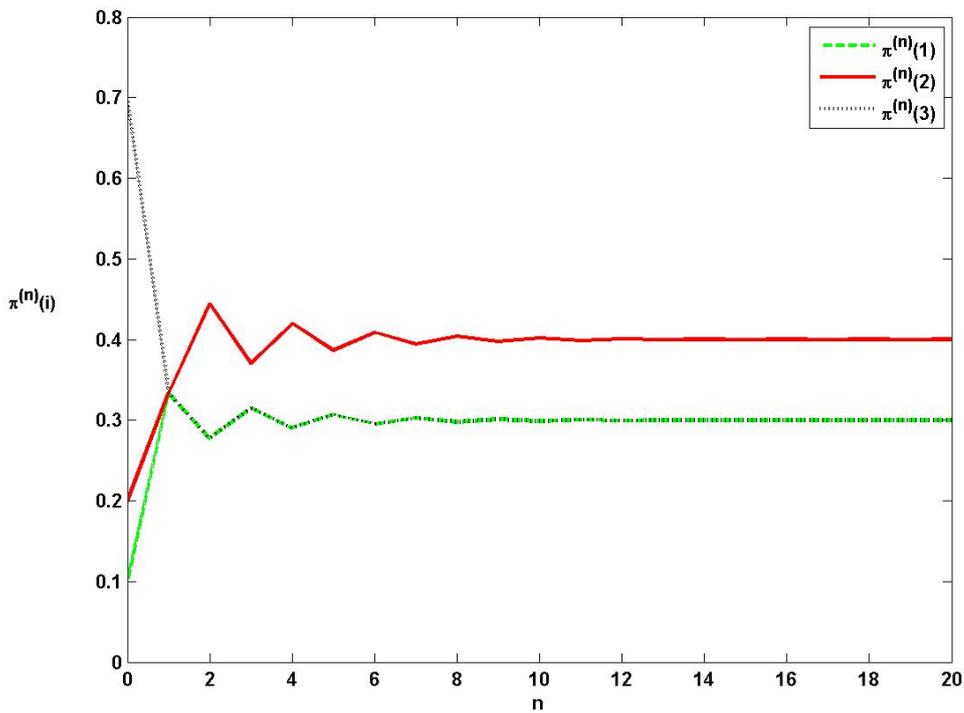


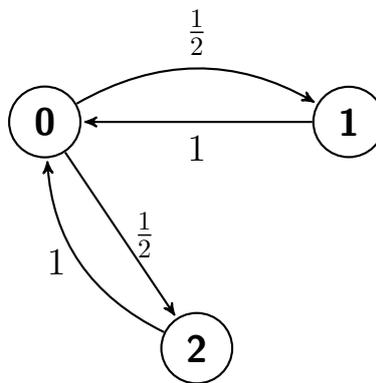
그림 4.5.1. 상태확률의 수렴

**예제 4.5.5** 상태공간이  $S = \{0, 1, 2\}$  인 Markov 체인의 추이확률행렬이 다음과 같다고

하자.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} & \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{matrix} \tag{1}$$

이 Markov 체인의 상태추이도는 다음과 같다.



식 (1)을 반복적용하면, 다음 식들이 성립함을 쉽게 알 수 있다.

$$\mathbb{P} = \mathbb{P}^3 = \mathbb{P}^5 = \dots = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \tag{2}$$

$$\mathbb{P}^2 = \mathbb{P}^4 = \mathbb{P}^6 = \dots = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{3}$$

식 (2)와 식 (3)을 증명하기 위해서, 행렬  $\mathbb{P}$ 를 Schur분해하면 다음과 같다.

$$\mathbb{P} = Q\Lambda Q^{-1}, \quad \Lambda \doteq \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad Q \doteq \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{4}$$

각 자연수  $m$ 에 대해서 다음 식들이 성립한다.

$$\mathbb{P}^{2m-1} = Q\Lambda^{2m-1}Q^{-1} = Q \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} Q^{-1} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad (5)$$

$$\mathbb{P}^{2m} = Q\Lambda^{2m}Q^{-1} = Q \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} Q^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad (6)$$

식 (5)와 식 (6)에서 알 수 있듯이, 추이확률행렬들은 주기적(periodic)이며 주기가 2이다.

이 Markov체인의 초기시점 상태확률벡터가  $\boldsymbol{\pi}^{(0)} = [0.1, 0.2, 0.7]$ 이라 하자. 다음 식들이 성립한다.

$$\boldsymbol{\pi}^{(1)} = \boldsymbol{\pi}^{(0)}\mathbb{P}^1 = [0.9000, 0.0500, 0.0500] \quad (7)$$

$$\boldsymbol{\pi}^{(2)} = \boldsymbol{\pi}^{(0)}\mathbb{P}^2 = [0.1000, 0.4500, 0.4500] \quad (8)$$

$$\boldsymbol{\pi}^{(3)} = \boldsymbol{\pi}^{(0)}\mathbb{P}^3 = [0.9000, 0.0500, 0.0500] \quad (9)$$

$$\boldsymbol{\pi}^{(4)} = \boldsymbol{\pi}^{(0)}\mathbb{P}^4 = [0.1000, 0.4500, 0.4500] \quad (10)$$

또한, 다음 식들이 성립한다.

$$\boldsymbol{\pi}^{(17)} = \boldsymbol{\pi}^{(0)}\mathbb{P}^{(17)} = [0.9000, 0.0500, 0.0500] \quad (11)$$

$$\boldsymbol{\pi}^{(18)} = \boldsymbol{\pi}^{(0)}\mathbb{P}^{(18)} = [0.1000, 0.4500, 0.4500] \quad (12)$$

$$\boldsymbol{\pi}^{(19)} = \boldsymbol{\pi}^{(0)}\mathbb{P}^{(19)} = [0.9000, 0.0500, 0.0500] \quad (13)$$

$$\boldsymbol{\pi}^{(20)} = \boldsymbol{\pi}^{(0)}\mathbb{P}^{(20)} = [0.1000, 0.4500, 0.4500] \quad (14)$$

따라서, 상태확률들이 진동한다고 짐작할 수 있다.

이 Markov체인의 상태확률벡터들을 구하기 위해서, 다음 MATLAB 프로그램 MarkovChain002.m 을 실행하라.

```

1 % -----
2 % Filename: MarkovChain002.m
3 % Chapman-Kolmogorov Equation
4 % Programmed by CBS
5 % -----
6 clear all, close all

```

```

7 % Eigen-Analysis of the Matrix P
8 P = [ 0 1/2 1/2 ; 1 0 0 ; 1 0 0 ]
9 P2 = P*P
10 P3 = P2*P
11 P4 = P3*P
12 [ Q Lambda ] = eig(P)
13 invQ = inv(Q)
14 DC = Q*Lambda*invQ - P
15 % State Probability Vector
16 pii(1,:) = [0.1 0.2 0.7 ];
17 nMax = 20;
18 for kk = 1:nMax
19     pii(kk+1,:) = pii(kk,:)*P;
20 end
21 pii1 = pii(:,1); pii2 = pii(:,2); pii3 = pii(:,3);
22 pii12 = pii1+pii2;
23 nn = (0:nMax)';
24 disp('      n      \pi(1)      \pi(2)      \pi(3)')
25 disp([nn pii])
26 % Plotting
27 plot(nn,pii1,'r--',nn,pii2,'g-',nn,pii3,'k:','linewidth',2)
28 set(gca,'fontsize',11,'fontweigh','bold','ylim',[0,1])
29 legend('\pi^{(n)}(1)', '\pi^{(n)}(2)', '\pi^{(n)}(3)',2)
30 xlabel('n')
31 text(-3,0.45, '\bf \pi^{(n)}(i)')
32 saveas(gcf, 'MarkovChain002.jpg')
33 % End of program
34 % -----

```

이 MATLAB 프로그램 MarkovChain002.m을 실행하면, 이 Markov체인의 상태확률벡터들을 그린 그림 4.5.2를 출력한다. 그림 4.5.2에서 상태확률들이 진동함을 알 수 있다. ■

## 제4.6절 상태의 분류

### 4.6.1 불변분포

Markov 체인에서 시간동질성, 즉 시불변성(time-invariance)이 성립하는지 여부는 중요한 문제이다. 비록 Markov 체인이 처음에는 시불변적이 아니라 하더라도, 시간이 아주 많이 흐른 다음에는 시불변적 Markov 체인으로 수렴하는지 여부는 중요한 문제이다. 이는 Markov 체인이 초기벡터  $\pi^{(0)}$ 에 얼마나 민감한가와 연결되는 문제이다.

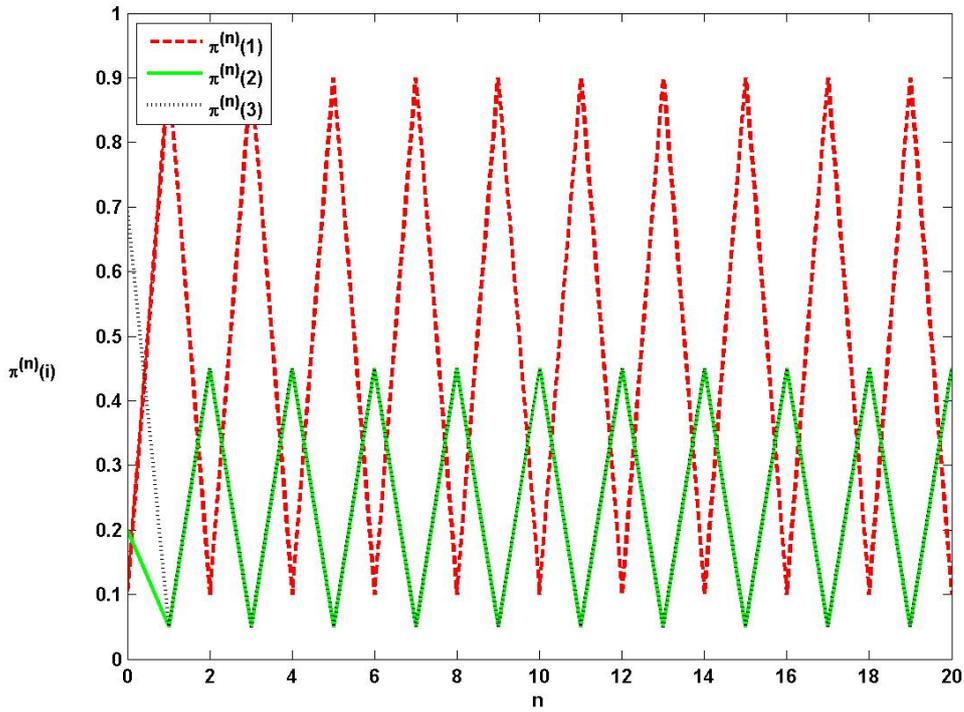


그림 4.5.2. 상태확률의 진동

각  $j \in S$ 에 대해서, 다음 식들이 성립함을 쉽게 알 수 있다.

$$\begin{aligned}
 \pi^{(n+1)}(j) &= \Pr(X_{n+1} = j) \\
 &= \sum_{i=0}^{\infty} \Pr(X_{n+1} = j | X_n = i) \Pr(X_n = i) \\
 &= \sum_{i=0}^{\infty} p(i, j) \pi^{(n)}(i)
 \end{aligned} \tag{4.6.1}$$

식 (4.6.1)을 다음과 같이 쓸 수 있다.

$$\boldsymbol{\pi}^{(n+1)} = \boldsymbol{\pi}^{(n)} \mathbb{P} \tag{4.6.2}$$

식 (4.6.2)에서 알 수 있듯이, 각 자연수  $m$ 에 대해서 다음 식들이 성립한다.

$$\boldsymbol{\pi}^{(n+m)} = \boldsymbol{\pi}^{(n)} \mathbb{P}^{(m)} = \boldsymbol{\pi}^{(n)} \mathbb{P}^m \tag{4.6.3}$$

다음 벡터를 정의하자.

$$\boldsymbol{\pi} \doteq \lim_{n \rightarrow \infty} \boldsymbol{\pi}^{(n)} \quad (4.6.4)$$

이  $\boldsymbol{\pi}$  를 극한안정상태분포 (limiting steady state distribution) 또는 정상분포 (stationary distribution) 라고 부른다.

만약  $\{\mathbb{P}^m\}$  이 행렬  $\mathbb{P}^\infty$  로 수렴하면, 이 Markov 체인은 안정성 (stability) 을 갖는다고 한다. 좀 더 명확하게 말하면, 점근적 안정성을 갖는다고 한다. 만약 Markov 체인이 안정성을 가지면, 식 (4.6.3) 과 식 (4.6.4) 에서 알 수 있듯이, 다음 식들이 성립한다.

$$\boldsymbol{\pi} = \lim_{n \rightarrow \infty} \boldsymbol{\pi}^{(0)} \mathbb{P}^n = \boldsymbol{\pi}^{(0)} \mathbb{P}^\infty \quad (4.6.5)$$

식 (4.6.2) 에서 알 수 있듯이, 이 극한안정상태분포는 다음 식을 만족한다.

$$\boldsymbol{\pi} = \boldsymbol{\pi} \mathbb{P} \quad (4.6.6)$$

식 (4.6.6) 에서 알 수 있듯이, 극한안정상태분포  $\boldsymbol{\pi}$  를 초기벡터로 하면, 각 시점  $t$  에서 상태벡터는  $\boldsymbol{\pi}$  이다. 따라서, 이  $\boldsymbol{\pi}$  를 불변분포 (invariant distribution) 또는 평형분포 (equilibrium distribution) 라 부르기도 한다.

식 (4.6.6) 에서 알 수 있듯이, 불변분포  $\boldsymbol{\pi}$  는 추이확률행렬  $\mathbb{P}$  의 고유값 1 에 대응하는 좌 고유벡터 (left eigenvector) 이다. 식 (4.6.5) 는 임의의 초기상태벡터  $\boldsymbol{\pi}^{(0)}$  에 대해서 성립한다. 따라서, 만약 제  $i$  번째 원소만 1 이고 나머지 원소들은 0 인  $\infty$  차원 열벡터  $\mathbf{e}_i^t$  를 초기상태벡터로 선택하면, 다음 식이 성립한다.

$$\mathbf{e}_i^t \mathbb{P}^\infty = \boldsymbol{\pi} \quad (4.6.7)$$

따라서, 다음 식들이 성립한다.

$$\mathbf{1}_\infty \boldsymbol{\pi} = \begin{bmatrix} \boldsymbol{\pi} \\ \boldsymbol{\pi} \\ \boldsymbol{\pi} \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1^t \\ \mathbf{e}_2^t \\ \mathbf{e}_3^t \\ \vdots \end{bmatrix} \mathbb{P}^\infty = I_\infty \mathbb{P}^\infty = \mathbb{P}^\infty \quad (4.6.8)$$

식 (4.6.8)과 수학적귀납법을 사용해서, 다음 식들이 성립함을 증명할 수가 있다.

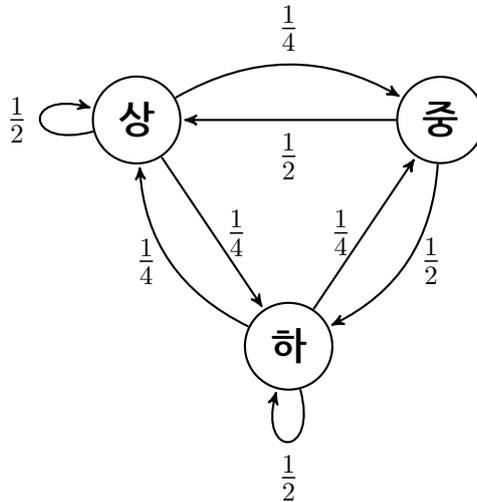
$$\det(\mathbb{P}^\infty - \lambda I_\infty) = \det(\mathbf{1}\boldsymbol{\pi} - \lambda I_\infty) = \lim_{n \rightarrow \infty} [-1]^n \lambda^n \left[ \lambda - \sum_{i=0}^n \pi(i) \right] \quad (4.6.9)$$

식 (4.6.9)에서 알 수 있듯이, 추이확률행렬  $\mathbb{P}^\infty$ 의 두 번째 이후 고유값들은 모두 0이다. 추이확률행렬  $\mathbb{P}$ 의 고유값을  $\lambda$ 라고 표기하면,  $\mathbb{P}^m$ 의 고유값은  $\lambda^m$ 이다. 따라서,  $\mathbb{P}^m$ 이  $\mathbb{P}^\infty$ 에 수렴하는 속도는  $\mathbb{P}$ 의 고유값들 중에서 절대값이 두 번째로 큰 고유값에 의해서 결정된다.

**예제 4.6.1** 어떤 과목에서 각 월별 학생성적을 ‘상’, ‘중’, 그리고 ‘하’로 분류한다고 하자. 다음 달 성적은 이번 달 성적에만 의존한다고 하고, 추이확률행렬이 다음과 같다고 하자.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} \text{상} & \text{중} & \text{하} \end{matrix} \\ \begin{matrix} \text{상} \\ \text{중} \\ \text{하} \end{matrix} & \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \end{matrix} \quad (1)$$

이 Markov 체인의 상태추이도는 다음과 같다.



첫 달에 성적이 ‘중’인 경우에는 초기분포가  $\boldsymbol{\pi}^{(0)} = [0, 1, 0]$ 로 표현되고, 다음 달 성적의 상태확률벡터는  $\boldsymbol{\pi}^{(1)} = [0.5, 0, 0.5]$ 이다. 행렬  $\mathbb{P}$ 의 고유값들은 다음과 같다.

$$\lambda_0 = 1, \quad \lambda_1 = 0.25, \quad \lambda_2 = -0.25 \quad (2)$$

다음 행렬들을 정의하자.

$$\Lambda \doteq \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & -\frac{1}{4} \end{bmatrix}, \quad E \doteq \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{3\sqrt{2}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-4}{3\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{3\sqrt{2}} \end{bmatrix} \quad (3)$$

다음 식이 성립한다.

$$E^{-1} = \begin{bmatrix} \frac{2\sqrt{3}}{5} & \frac{\sqrt{3}}{5} & \frac{2\sqrt{3}}{5} \\ -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ \frac{3\sqrt{2}}{10} & -\frac{3\sqrt{2}}{5} & \frac{3\sqrt{2}}{10} \end{bmatrix} \quad (4)$$

또한, 다음 식이 성립함을 확인할 수 있다.

$$\mathbb{P} = E\Lambda E^{-1} \quad (5)$$

식 (5)에서 알 수 있듯이, 추이확률행렬  $\mathbb{P}$ 의 고유값  $\lambda_0 = 1$ 에 대응하는 고유벡터는  $\left[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right]$ 이다. 따라서, 식 (4.5.2)가 성립함을 확인할 수 있다. 식 (5)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\begin{aligned} \mathbb{P}^n &= E\Lambda^n E^{-1} = E \begin{bmatrix} 1 & 0 & 0 \\ 0 & \left[\frac{1}{4}\right]^n & 0 \\ 0 & 0 & \left[\frac{-1}{4}\right]^n \end{bmatrix} E^{-1} \\ &= \begin{bmatrix} \frac{2}{5} + \frac{1}{2} \left[\frac{1}{4}\right]^n + \frac{1}{10} \left[\frac{-1}{4}\right]^n & \frac{1}{5} - \frac{1}{5} \left[\frac{-1}{4}\right]^n & \frac{2}{5} - \frac{1}{2} \left[\frac{1}{4}\right]^n + \frac{1}{10} \left[\frac{-1}{4}\right]^n \\ \frac{2}{5} - \frac{2}{5} \left[\frac{-1}{4}\right]^n & \frac{1}{5} + \frac{4}{5} \left[\frac{-1}{4}\right]^n & \frac{2}{5} - \frac{2}{5} \left[\frac{-1}{4}\right]^n \\ \frac{2}{5} - \frac{1}{2} \left[\frac{1}{4}\right]^n + \frac{1}{10} \left[\frac{-1}{4}\right]^n & \frac{1}{5} - \frac{1}{5} \left[\frac{-1}{4}\right]^n & \frac{2}{5} + \frac{1}{2} \left[\frac{1}{4}\right]^n + \frac{1}{10} \left[\frac{-1}{4}\right]^n \end{bmatrix} \quad (6) \end{aligned}$$

따라서 다음 식이 성립한다.

$$\mathbb{P}^\infty = E \left[ \lim_{n \rightarrow \infty} \Lambda^n \right] E^{-1} = \begin{bmatrix} \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \end{bmatrix} \quad (7)$$

여기서 두 번째 등호의 수렴속도는 두 번째 고유값  $\lambda_1 = 0.25$  또는 세 번째 고유값  $\lambda_2 = -0.25$

에 의해 결정된다. 행렬  $\mathbb{P}^\infty$ 의 고유값들은 1, 0, 0이고 이에 대응하는 좌고유벡터들은 다음과 같다.

$$\mathbf{y}_0 = \begin{bmatrix} \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \end{bmatrix} \quad (8)$$

$$\mathbf{y}_1 = \begin{bmatrix} -0.8018 & 0.2673 & 0.5345 \end{bmatrix} \quad (9)$$

$$\mathbf{y}_2 = \begin{bmatrix} 0.7160 & -0.6979 & -0.0180 \end{bmatrix} \quad (10)$$

불변분포  $\boldsymbol{\pi}$ 는 추이확률행렬  $\mathbb{P}$ 의 고유값 1에 대응하는 좌고유벡터이므로, 식  $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbb{P}$ 를 만족하는 불변분포는  $\boldsymbol{\pi} = \mathbf{y}_0$ 이다. 즉, 이 과목에서 ‘상’, ‘중’, 그리고 ‘하’의 비율은 대략 0.4대 0.2대 0.4이다. ■

**예제 4.6.2** 예제 4.5.3의 버스문제를 다시 살펴보자. 예제 4.5.3의 식 (17)을 다음과 같이 쓸 수 있다.

$$\mathbb{P}^n = \frac{1}{\alpha + \beta} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{[1 - \alpha - \beta]^n}{\alpha + \beta} \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} \quad (1)$$

만약 초기분포가  $\boldsymbol{\pi}^{(0)} = [\pi^{(0)}(0), \pi^{(0)}(1)]$ 이면, 시점  $n$ 에서 상태확률벡터는 다음과 같다.

$$\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(0)}\mathbb{P}^n = \frac{1}{\alpha + \beta} [\beta, \alpha] + [\pi^{(0)}(0)\alpha - \pi^{(0)}(0)\beta] \frac{[1 - \alpha - \beta]^n}{\alpha + \beta} \begin{bmatrix} 1 & 1 \end{bmatrix} \quad (2)$$

또한, 예제 4.5.3의 식 (18)에서 알 수 있듯이, 행렬  $\mathbb{P}$ 의 고유값들은 1과  $1 - \alpha - \beta$ 이다.

만약 식  $|1 - \alpha - \beta| < 1$ 이 성립하면, 다음 식이 성립한다.

$$\mathbb{P}^\infty = \frac{1}{\alpha + \beta} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} \quad (3)$$

즉, 이 Markov 체인은 안정적이다. 식 (3)을 식 (4.6.5)에 대입하면, 식  $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbb{P}$ 를 만족하는 불변분포가 다음과 같음을 알 수 있다.

$$\boldsymbol{\pi} = \lim_{m \rightarrow \infty} \boldsymbol{\pi}^{(0)}\mathbb{P}^m = \boldsymbol{\pi}^{(0)}\mathbb{P}^\infty = \frac{1}{\alpha + \beta} [\beta, \alpha] \quad (4)$$

식 (4)에서 확인할 수 있듯이, 불변분포는 초기분포에 의존하지 않는다.

만약  $\alpha = \beta = 0$ 이면, 추이확률행렬과 그 멱행렬들은 다음과 같다.

$$\mathbb{P}^m = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (m = 1, 2, \dots) \quad (5)$$

즉, 이 Markov 체인은 안정적이다. 또한, 불변분포는 다음과 같다.

$$\boldsymbol{\pi} = \lim_{m \rightarrow \infty} \boldsymbol{\pi}^{(0)} \mathbb{P}^m = \boldsymbol{\pi}^{(0)} \mathbb{P}^\infty = \boldsymbol{\pi}^{(0)} \quad (6)$$

만약  $\alpha = \beta = 1$ 이면, 추이확률행렬과 그 멱행렬은 다음과 같다.

$$\mathbb{P}^m = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad (m = 1, 3, \dots) \quad (7)$$

$$\mathbb{P}^m = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (m = 2, 4, \dots) \quad (8)$$

따라서, 이 Markov 체인은 안정적이 아니며, 또한 불변분포는 존재하지 않는다. ■

#### 4.6.2 최초방문시점

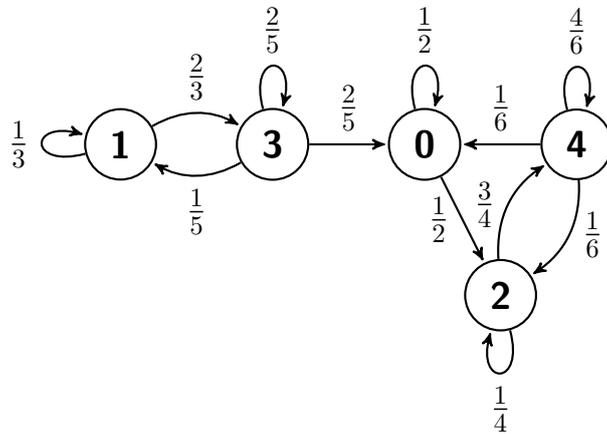
상태추이도에서 상태는 점에 해당하고 이 점들을 연결하는 방향선분에 할당된 확률이 추이확률이다. 이 추이확률에 따라 상태를 분류할 수 있을 것이다. 우선 다음 예제를 살펴보자.

**예제 4.6.3** 상태공간이  $S = \{0, 1, 2, 3, 4\}$  인 Markov 체인의 추이확률행렬이 다음과 같다

고 하자.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{2}{5} & \frac{1}{5} & 0 & \frac{2}{5} & 0 \\ \frac{1}{6} & 0 & \frac{1}{6} & 0 & \frac{4}{6} \end{bmatrix} \end{matrix} \tag{1}$$

이 Markov 체인의 상태추이도는 다음과 같다.



이 상태추이도에서 알 수 있듯이, 상태들의 부분집합  $\{0, 2, 4\}$ 에서는 서로 오갈 수 있고 부분집합  $\{1, 3\}$ 에서도 서로 오갈 수 있다. 그러나, 부분집합  $\{1, 3\}$ 에서 부분집합  $\{0, 2, 4\}$ 로 갈 수는 있으나 돌아갈 수는 없다. 따라서, 상태들의 순서를  $(0, 2, 4, 1, 3)$ 으로 바꾸면 좀 더 효율적으로 이 Markov 체인을 분석할 수 있을 것 같다. 이렇게 순서를 바꾼 경우에 추이확률

행렬은 다음과 같다.

$$\bar{\mathbb{P}} = \begin{matrix} & \begin{matrix} 0 & 2 & 4 & 1 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 2 \\ 4 \\ 1 \\ 3 \end{matrix} & \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{4}{6} & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{2}{5} & 0 & 0 & \frac{1}{5} & \frac{2}{5} \end{bmatrix} \end{matrix} \quad (2)$$

이 추이확률행렬  $\bar{\mathbb{P}}$ 를 정준형 (canonical) 추이확률행렬이라 부른다. 행렬  $\bar{\mathbb{P}}$ 를 Schur분해하면 다음과 같다.

$$\bar{\mathbb{P}} = \bar{Q}\bar{\Lambda}\bar{Q}^{-1} \quad (3)$$

여기서  $\bar{\Lambda}$ 와  $\bar{Q}$ 는  $\Lambda$ 와  $Q$ 의 공액복소수를 나타내는 것이 아니고 각각 다음과 같이 정의된다.

$$\bar{\Lambda} \doteq \begin{bmatrix} 0.2083 + 0.1998i & 0 & 0 & 0 & 0 \\ 0 & 0.2083 - 0.1998i & 0 & 0 & 0 \\ 0 & 0 & 0.2427 & 0 & 0 \\ 0 & 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 0 & 0.8239 \end{bmatrix} \quad (4)$$

$$\bar{Q} \doteq \begin{bmatrix} -0.1196 + 0.4360i & -0.1196 - 0.4360i & 0.0000 & 0.4472 & 0.0000 \\ -0.1045 - 0.3022i & -0.1045 + 0.3022i & -0.0000 & 0.4472 & 0.0000 \\ 0.0863 - 0.0110i & 0.0863 + 0.0110i & -0.0000 & 0.4472 & -0.0000 \\ -0.4211 - 0.1836i & -0.4211 + 0.1836i & 0.6181 & 0.4472 & 0.9044 \\ 0.6890 & 0.6890 & -0.7861 & 0.4472 & 0.4267 \end{bmatrix} \quad (5)$$

Chapman-Kolmogorov식을 적용하면, 다음 식들이 성립함을 알 수 있다.

$$\begin{aligned} \bar{\mathbb{P}}^{10} &= \bar{Q}\bar{\Lambda}^{10}\bar{Q}^{-1} \\ &= \begin{bmatrix} 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.2031 & 0.2147 & 0.3943 & 0.1052 & 0.0827 \\ 0.1949 & 0.2334 & 0.4832 & 0.0496 & 0.0390 \end{bmatrix} \end{aligned} \quad (6)$$

$$\begin{aligned} \bar{\mathbb{P}}^{20} &= \bar{Q}\bar{\Lambda}^{20}\bar{Q}^{-1} \\ &= \begin{bmatrix} 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.1897 & 0.2449 & 0.5383 & 0.0152 & 0.0119 \\ 0.1886 & 0.2476 & 0.5511 & 0.0072 & 0.0056 \end{bmatrix} \end{aligned} \quad (7)$$

$$\begin{aligned} \bar{\mathbb{P}}^{30} &= \bar{Q}\bar{\Lambda}^{30}\bar{Q}^{-1} \\ &= \begin{bmatrix} 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.1878 & 0.2493 & 0.5590 & 0.0022 & 0.0017 \\ 0.1877 & 0.2497 & 0.5609 & 0.0010 & 0.0008 \end{bmatrix} \end{aligned} \quad (8)$$

$$\begin{aligned} \bar{\mathbb{P}}^{40} &= \bar{Q}\bar{\Lambda}^{40}\bar{Q}^{-1} \\ &= \begin{bmatrix} 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.1878 & 0.2499 & 0.5620 & 0.0003 & 0.0002 \\ 0.1877 & 0.2500 & 0.5623 & 0.0001 & 0.0001 \end{bmatrix} \end{aligned} \quad (9)$$

$$\begin{aligned} \bar{\mathbb{P}}^{50} &= \bar{Q}\bar{\Lambda}^{50}\bar{Q}^{-1} \\ &= \begin{bmatrix} 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.1875 & 0.2500 & 0.5625 & -0.0000 & -0.0000 \\ 0.1875 & 0.2500 & 0.5624 & 0.0000 & 0.0000 \\ 0.1875 & 0.2500 & 0.5623 & 0.0000 & 0.0000 \end{bmatrix} \end{aligned} \quad (10)$$

식 (6)~식 (10)에서 추측할 수 있듯이, 다음 식들이 성립한다.

$$\begin{aligned} \lim_{n \rightarrow \infty} \bar{\mathbb{P}}^n &= \lim_{n \rightarrow \infty} \bar{Q}\bar{\Lambda}^n\bar{Q}^{-1} \\ &= \bar{Q} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \bar{Q}^{-1} = \begin{bmatrix} \frac{3}{16} & \frac{4}{16} & \frac{9}{16} & 0 & 0 \\ \frac{3}{16} & \frac{4}{16} & \frac{9}{16} & 0 & 0 \\ \frac{3}{16} & \frac{4}{16} & \frac{9}{16} & 0 & 0 \\ \frac{3}{16} & \frac{4}{16} & \frac{9}{16} & 0 & 0 \\ \frac{3}{16} & \frac{4}{16} & \frac{9}{16} & 0 & 0 \end{bmatrix} \end{aligned} \quad (11)$$

식 (11)에서 알 수 있듯이, 상태확률행렬  $\mathbb{P}$ 의 멱행렬들의 극한은 다음과 같다.

$$\mathbb{P}^\infty = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} \frac{3}{16} & 0 & \frac{4}{16} & 0 & \frac{9}{16} \\ \frac{3}{16} & 0 & \frac{4}{16} & 0 & \frac{9}{16} \\ \frac{3}{16} & 0 & \frac{4}{16} & 0 & \frac{9}{16} \\ \frac{3}{16} & 0 & \frac{4}{16} & 0 & \frac{9}{16} \\ \frac{3}{16} & 0 & \frac{4}{16} & 0 & \frac{9}{16} \end{bmatrix} \end{matrix} \quad (12)$$

식 (12)에서 알 수 있듯이, 초기시점의 어느 상태에서 시작하더라도 이 Markov 체인은 궁극적으로 상태 0, 상태 2 그리고 상태 4로 이동한다.

이 Markov 체인의 초기시점 상태확률벡터가  $\pi^{(0)} = [0.1, 0.2, 0.2, 0.2, 0.3]$  이라 하자. 정준 상태확률행렬에 해당하는 초기시점 상태확률벡터는 다음과 같다.

$$\bar{\pi}^{(0)} = [0.1, 0.2, 0.2, 0.3, 0.2] \quad (13)$$

다음 식들이 성립한다.

$$\bar{\pi}^{(1)} = \bar{\pi}^{(0)}\bar{\mathbb{P}} = [0.1633, 0.1333, 0.2833, 0.2400, 0.1800] \quad (14)$$

$$\bar{\pi}^{(2)} = \bar{\pi}^{(1)}\bar{\mathbb{P}} = [0.2009, 0.1622, 0.2889, 0.1960, 0.1520] \quad (15)$$

$$\bar{\pi}^{(3)} = \bar{\pi}^{(2)}\bar{\mathbb{P}} = [0.2094, 0.1891, 0.3143, 0.1611, 0.1261] \quad (16)$$

$$\bar{\pi}^{(4)} = \bar{\pi}^{(3)}\bar{\mathbb{P}} = [0.2075, 0.2044, 0.3514, 0.1326, 0.1041] \quad (17)$$

또한, 다음 식들이 성립한다.

$$\bar{\pi}^{(46)} = \bar{\pi}^{(45)}\bar{\mathbb{P}} = [0.1875, 0.2500, 0.5624, 0.0000, 0.0000] \quad (18)$$

$$\bar{\pi}^{(47)} = \bar{\pi}^{(46)}\bar{\mathbb{P}} = [0.1875, 0.2500, 0.5624, 0.0000, 0.0000] \quad (19)$$

$$\bar{\pi}^{(48)} = \bar{\pi}^{(47)}\bar{\mathbb{P}} = [0.1875, 0.2500, 0.5625, 0.0000, 0.0000] \quad (20)$$

$$\bar{\pi}^{(49)} = \bar{\pi}^{(48)}\bar{\mathbb{P}} = [0.1875, 0.2500, 0.5625, 0.0000, 0.0000] \quad (21)$$

식 (14)~식 (21)로부터 다음 식들이 성립함을 유추할 수 있다.

$$\lim_{n \rightarrow \infty} \bar{\pi}^{(n)}(0) = \frac{3}{16}, \quad \lim_{n \rightarrow \infty} \bar{\pi}^{(n)}(1) = \frac{4}{16} \quad (22)$$

$$\lim_{n \rightarrow \infty} \bar{\pi}^{(n)}(2) = \frac{9}{16}, \quad \lim_{n \rightarrow \infty} \bar{\pi}^{(n)}(3) = \frac{0}{16} \quad (23)$$

$$\lim_{n \rightarrow \infty} \bar{\pi}^{(n)}(4) = \frac{0}{16} \quad (24)$$

따라서, 극한상태확률들은 다음과 같다.

$$\lim_{n \rightarrow \infty} \pi^{(n)}(0) = \frac{3}{16}, \quad \lim_{n \rightarrow \infty} \pi^{(n)}(1) = \frac{9}{16} \quad (25)$$

$$\lim_{n \rightarrow \infty} \pi^{(n)}(2) = \frac{0}{16}, \quad \lim_{n \rightarrow \infty} \pi^{(n)}(3) = \frac{4}{16} \quad (26)$$

$$\lim_{n \rightarrow \infty} \pi^{(n)}(4) = \frac{0}{16} \quad (27)$$

또한, 식 (11)을 이용해서 다음과 같이 식 (25)~식 (27)을 유도할 수도 있다.

$$\lim_{n \rightarrow \infty} \pi^{(n)} = \lim_{n \rightarrow \infty} \pi^{(0)} \mathbb{P}^n = \pi^{(0)} \lim_{n \rightarrow \infty} \mathbb{P}^n = \left[ \frac{3}{16}, \frac{4}{16}, \frac{9}{16}, 0, 0 \right] \quad (28)$$

식 (28)에서 알 수 있듯이, 극한상태확률들은 초기시점의 상태확률들에 의존하지 않는다.

이 Markov 체인의 상태확률벡터들을 구하기 위해서, 다음 MATLAB 프로그램 StatesClassification001.m을 실행하라.

```

1 % -----
2 % Filename: StatesClassification001
3 % State Classification of Markov Chain
4 % Programmed by CBS
5 %-----
6 clear all, close all
7 % Eigen-Analysis of the canonical state transition Matrix
8 Pbar = [ 1/2 1/2 0 0 0
9          0 1/4 3/4 0 0
10         1/6 1/6 4/6 0 0
11         0 0 0 2/3 1/3
12         2/5 0 0 1/5 2/5 ]
13 [ Qbar Lambdabar ] = eig(Pbar)
14 invQbar = inv(Qbar)
15 DC = Qbar*Lambdabar*invQbar - Pbar
16 for nn=10:10:100
17     nn
18     LambdabarN = Lambdabar^nn;
19     PbarN = real(Qbar*LambdabarN*invQbar)
20 end
21 LambdaInf = Lambdabar^1000
22 PbarInf = real(Qbar*LambdaInf*invQbar)
23 % State Probability Vector

```

```

24 piibar(1,:) = [0.1 0.2 0.2 0.3 0.2 ];
25 nMax = 50;
26 for kk = 1:nMax
27     piibar(kk+1,:) = piibar(kk,:)*Pbar;
28 end
29 pii0 = piibar(:,1); pii2 = piibar(:,2); pii4 = piibar(:,3);
30 pii1 = piibar(:,4); pii3 = piibar(:,5);
31 nn = (0:nMax)';
32 disp([nn piibar])
33 % Plotting
34 plot(nn,pii0,'r--',nn,pii1,'g-',nn,pii2,'m--', ...
35      nn,pii3,'k:',nn,pii4,'b-.','linewidth',2)
36 set(gca,'fontsize',11,'fontweigh','bold','ylim',[-0.1,0.7])
37 legend('\pi^{(n)}(0)', '\pi^{(n)}(1)', '\pi^{(n)}(2)', '\pi^{(n)}(3)', ...
38        '\pi^{(n)}(4)', 'location', 'NW')
39 xlabel('n')
40 text(-7,0.45, '\bf \pi^{(n)}(i)')
41 saveas(gcf, 'StatesClassification001.jpg')
42 % End of program
43 % -----

```

이 MATLAB 프로그램 StatesClassification001.m을 실행하면, 이 Markov 체인의 상태 확률 벡터들을 그린 그림 4.6.1을 출력한다. 그림 4.6.1에서 상태 확률들이 극한 상태 확률들로 수렴함을 확인할 수 있다. ■

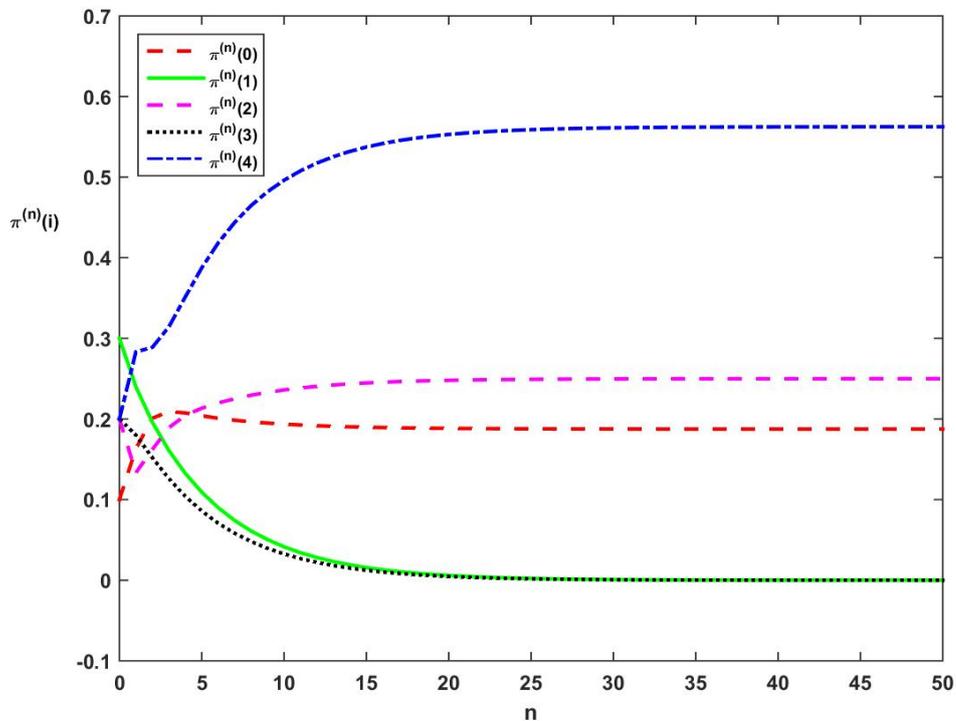
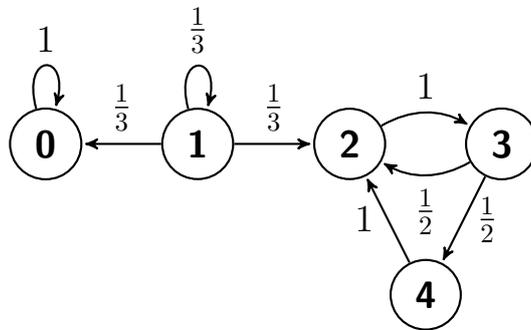


그림 4.6.1. Markov 체인의 극한 상태 확률

**예제 4.6.4** 상태공간이  $\{0, 1, 2, 3, 4\}$  이고 추이확률행렬이 다음과 같은 Markov 체인을 생각해 보자.

$$\mathbb{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{1}$$

이 Markov 체인의 상태추이도는 다음과 같다.



식 (1)의 추이확률행렬  $\mathbb{P}$ 에 예제 4.6.3에서 한 분석을 다시 적용하면, 다음과 같은 결론을 얻을 수 있다. 상태 0에 도달하면 그 상태에서 빠져나갈 수 없다. 시점 0에서 상태 1에 있지 않는 한 상태 1에 도달할 수 없다. 상태들의 부분집합  $\{2, 3, 4\}$ 는 하나의 상태와 비슷하다. 즉, 이 부분집합 내의 한 상태에서 이 부분집합 내의 다른 상태로 이동할 수 있다. 따라서, 시점 0에서 상태가 1이면 극한상태벡터는  $[1, 0, 0, 0, 0]$ 이고, 그렇지 않으면 극한추이확률행렬의 첫 번째 열과 두 번째 열은 영벡터들이다. ■

예제 4.6.3과 예제 4.6.4에서 짐작할 수 있듯이, Markov 체인의 극한행태를 분석하기 위해서는 최초방문시점 (first visit time 또는 first hitting time 또는 first return time 또는 time of the first return to) 을 정의할 필요가 있다.

**정의 4.6.1**

부분집합  $A \subset S$ 를 처음 방문하는 시점, 즉 최초방문시점  $T_A$ 는 다음과 같다.

$$T_A \doteq \min \{n \geq 1 \mid X_n \in A\}$$

만약 임의의  $n$ 에 대해서  $X_n \notin A$ 이면,  $T_A = \infty$ 라 한다.

최초방문시점  $T_A$ 는 지지대가  $\mathbb{Z}_{>0} \cup \{\infty\}$ 인 확률변수이고, 초기시점에서 상태  $X_0$ 에 관계없이 집합  $A$ 를 처음 방문하는 시점을 의미한다. 만약  $A = \{y\}$ 이면,  $T_A$ 를  $T_y$ 로 표기하자. 다음 식이 성립함은 자명하다.

$$\{T_y = n\} = \{X_1 \neq y, X_2 \neq y, \dots, X_{n-1} \neq y, X_n = y\} \quad (4.6.10)$$

즉,  $\{T_y = n\}$ 은 시점  $n$ 까지 확률변수값들  $\{X_k = x_k \mid k = 1, 2, \dots, n\}$ 에 의존한다. 시점 0에서 상태가  $x$ 라는 조건 하에 집합  $A$ 의 확률을  $P_x(A)$ 로 표기하자. 즉, 다음 식이 성립한다.

$$P_x(A) \doteq \Pr(A \mid X_0 = x) \quad (4.6.11)$$

우리집 멍멍이가 제  $n$ 번째 밤을 지내는 곳을  $X_n$ 이라 하고, 확률과정  $\{X_n \mid n = 0, 1, \dots\}$ 은 상태공간이  $\{H, G, B\}$ 인 Markov 체인을 이룬다고 하자. 여기서  $H$ 는 우리집,  $G$ 는 멍멍이의 여자친구가 사는 옆집, 그리고  $B$ 는 동네 푸줏간이다. 지금부터 오늘 아침 집을 나간 멍멍이가 제  $n$ 번째 밤을 여자친구와 지내는 확률  $p^{(n)}(H, G)$ 를 생각해 보자. 상태  $G$ 를  $y$ 로 표기하면, 다음 식이 성립한다.

$$\{X_n = y\} \subset \{T_y \leq n\} \quad (4.6.12)$$

다음 식들이 성립함은 자명하다.

$$\{X_n = y, T_y = m\} \cap \{X_n = y, T_y = k\} = \emptyset, \quad (1 \leq m < k \leq n) \quad (4.6.13)$$

따라서,  $\{X_n = y\}$ 를 다음과 같이 배반사상들(disjoint events)의 합집합으로 나타낼 수 있다.

$$\{X_n = y\} = \cup_{m=1}^n \{X_n = y, T_y = m\} \quad (4.6.14)$$

상태  $H$ 를  $x$ 로 표기하면, 다음 식들이 성립한다.

$$\begin{aligned}
 & P_x(X_n = y) \\
 &= \sum_{m=1}^n P_x(X_n = y, T_y = m) \\
 &= \sum_{m=1}^n P_x(X_n = y | T_y = m) P_x(T_y = m) \\
 &= \sum_{m=1}^n P_x(T_y = m) \Pr(X_n = y | X_0 = x, X_1 \neq y, \dots, X_{m-1} \neq y, X_m = y) \\
 &= \sum_{m=1}^n P_x(T_y = m) \Pr(X_n = y | X_m = y) \tag{4.6.15}
 \end{aligned}$$

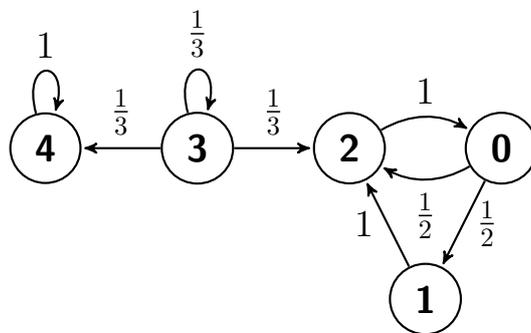
식 (4.6.15)와 Markov과정의 정상성에서 알 수 있듯이, 다음 식이 성립한다.

$$p^{(n)}(x, y) = \sum_{m=1}^n P_x(T_y = m) p^{(n-m)}(y, y) \tag{4.6.16}$$

식 (4.6.16)을 다음과 같이 해석할 수 있다. 시점 0에서 상태  $x$ 를 출발한 멍멍이(Markov체인)가 시점  $n$ 에서 상태  $y$ 에 도달할 확률  $p^{(n)}(x, y)$ 는 시점  $m(\leq n)$ 에서 상태  $y$ 를 처음 방문한 확률  $P_x(T_y = m)$ 에 시점 0에 상태  $y$ 를 출발해서 시점  $[n - m]$ 에 상태  $y$ 에 도달하는 확률  $p^{(n-m)}(y, y)$ 를 곱한 값  $P_x(T_y = m) p^{(n-m)}(y, y)$ 를  $m = 1, 2, \dots, n$ 에 대해서 더한 것이다. 식 (4.6.16)은 배리어옵션(barrier option)의 공정한 가치를 평가할 때 자주 사용된다.

### 4.6.3 과도상태, 재귀상태와 흡수상태

다음 Markov체인의 상태추이도를 살펴보자.



식  $p^{(1)}(0, 1) > 0$ 가 성립하므로 상태 0에서 상태 1로 단번에 갈 수 있고, 식  $p^{(2)}(0, 2) > 0$ 가 성립하므로 상태 0에서 상태 1을 거쳐 상태 2로 갈 수 있고, 식  $p^{(2)}(1, 1) > 0$ 가 성립하므로

상태 1에서 상태 2와 상태 0을 거쳐 상태 1로 갈 수 있다. 반면에, 상태 2에서 상태 3으로 갈 수 없고, 상태 4에서 상태 3으로 갈 수 없다. 즉, 어떤 상태에서 다른 상태로 갈 수 있는 경우와 그렇지 않은 경우가 있다. 예를 들어, 상태 3에서 출발해서 상태들 0, 1, 2, 3, 4에 도달할 수 있고, 상태들 0, 1, 2는 서로 오갈 수가 있다. 또한, 상태 4에 도달하면, 그 상태에서 빠져나올 수 없다.

식  $p^{(n)}(i, j) > 0$ 를 만족하는 정수  $n(\geq 0)$ 이 존재하면, 상태  $i$ 에서 유한 단계들을 거쳐 상태  $j$ 에 도달할 수 있으므로, 상태  $j$ 는 상태  $i$ 에서 도달가능(assessible)이라 하고 식  $i \rightarrow j$ 로 표기한다. 만약 식  $i \rightarrow j$ 와 식  $j \rightarrow i$ 가 성립하면, 즉 식  $p^{(n)}(i, j) > 0$ 를 만족하는 정수  $n(\geq 0)$ 과 식  $p^{(m)}(j, i) > 0$ 를 만족하는 정수  $m(\geq 0)$ 이 존재하면, 상태  $i$ 와 상태  $j$ 는 상호도달가능(communicative)이라 하고 식  $i \leftrightarrow j$ 로 표기한다. 도달가능과 상호도달가능에 대해서는 뒤에서 정형적(formally)으로 정의하기로 하자. 만약  $p(i, i) = 1$ 이면, 상태  $i$ 에서 빠져나오지 못한다. 따라서, 다음과 같은 정의를 하자.

**정의 4.6.2**

만약  $p(i, i) = 1$ 이면, 상태  $i$ 를 흡수상태라 한다.

만약  $y$ 가 흡수상태(absorbing state)이면, 다음 식들이 성립한다.

$$p^{(n)}(x, y) = \sum_{m=1}^n P_x(T_y = m) p^{(n-m)}(y, y) = \sum_{m=1}^n P_x(T_y = m) \tag{4.6.17}$$

여기서 첫 번째 등호는 식 (4.6.16)에 의해서 성립하고, 두 번째 등호는 흡수상태의 정의에 의해서 성립한다. 따라서, 다음 식이 성립한다.

$$P_x(T_y \leq n) = p^{(n)}(x, y) \tag{4.6.18}$$

식 (4.6.18)을 다음과 같이 해석할 수 있다. 시점 0에서 상태  $x$ 를 출발한 Markov체인이 시점  $n$  이전에 흡수상태  $y$ 에 도달하는 확률  $P_x(T_y \leq n)$ 은 시점 0의 상태  $x$ 에서 출발한 Markov체인이 시점  $n$ 에서 상태가  $y$ 인  $n$ 단계 추이확률  $p^{(n)}(x, y)$ 이다.

**예제 4.6.5** 예제 4.5.2의 Markov체인을 다시 생각해 보자. 이 Markov체인은 상태공간이

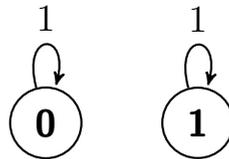
$S = \{0, 1\}$  이고 추이확률행렬은 다음과 같다.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \end{matrix} \tag{1}$$

만약  $a = b = 0$  이면, 다음 식들이 성립함은 자명하다.

$$\pi^{(n)}(0) = \pi^{(0)}(0) = 1, \quad \pi^{(n)}(1) = \pi^{(0)}(1) = 1 \tag{2}$$

상태 0와 상태 1은 흡수상태들이다. 다음 상태추이도에서 이를 확인할 수 있다.

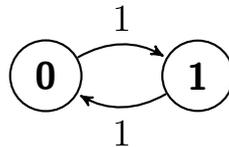


만약  $a = b = 1$  이면, 예제 4.5.2에서 알 수 있듯이 다음 식들이 성립한다.

$$\pi^{(n)}(0) = \frac{1}{2} + \left[ \pi^{(0)}(0) - \frac{1}{2} \right] [-1]^n \tag{3}$$

$$\pi^{(n)}(1) = \frac{1}{2} + \left[ \pi^{(0)}(1) - \frac{1}{2} \right] [-1]^n \tag{4}$$

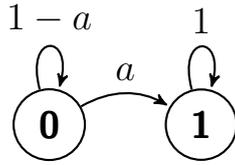
따라서,  $\lim_{n \rightarrow \infty} \pi^{(n)}(0)$  도  $\lim_{n \rightarrow \infty} \pi^{(n)}(1)$  도 존재하지 않는다. 이 Markov 체인은 상태 0와 상태 1 사이를 왕복한다. 즉, 상태 0과 상태 1은 상호도달가능이다. 다음 상태추이도에서 이를 확인할 수 있다.



만약 식  $a > 0$ 와 식  $b = 0$ 가 성립하면, 예제 4.5.2에서 알 수 있듯이 다음 식들이 성립한다.

$$\lim_{n \rightarrow \infty} \pi^{(n)}(0) = 0, \quad \lim_{n \rightarrow \infty} \pi^{(n)}(1) = 1 \tag{5}$$

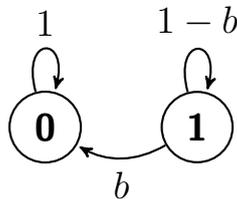
즉, 상태 1은 상태 0에서 도달가능이다. 또한, 상태 1은 흡수상태이다. 다음 상태추이도에서 이를 확인할 수 있다.



만약 식  $a = 0$ 와 식  $b > 0$ 가 성립하면, 예제 4.5.2에서 알 수 있듯이 다음 식들이 성립한다.

$$\lim_{n \rightarrow \infty} \pi^{(n)}(0) = 1, \quad \lim_{n \rightarrow \infty} \pi^{(n)}(1) = 0 \tag{6}$$

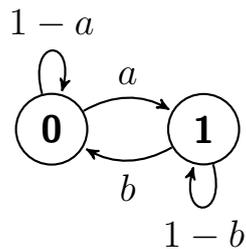
즉, 상태 0은 상태 1에서 도달가능이다. 또한, 상태 0은 흡수상태이다. 다음 상태추이도에서 이를 확인할 수 있다.



만약  $(a, b)$ 가  $ab \neq 0$ 이고  $ab \neq 1$ 이면, 예제 4.5.2에서 알 수 있듯이 다음 식들이 성립한다.

$$\lim_{n \rightarrow \infty} \pi^{(n)}(0) = \frac{b}{a+b}, \quad \lim_{n \rightarrow \infty} \pi^{(n)}(1) = \frac{a}{a+b} \tag{7}$$

즉, 상태 0과 상태 1은 상호도달가능이다. 다음 상태추이도에서 이를 확인할 수 있다.



■

#### 4.6.4 도달확률과 도달횟수

어떤 상태에 도달하는 확률과 횟수를 다음과 같이 정의하자.

**정의 4.6.3**

시점 0에서 상태  $x$ 를 출발하는 Markov체인  $\{X_n | n = 0, 1, \dots\}$ 이 단 한번이라도 상태  $y$ 에 도달하는 확률을  $f_{xy}$ 로 표기하자.

다음 식들이 성립한다.

$$f_{xy} \doteq \Pr(T_y < \infty | X_0 = x) = P_x(T_y < \infty) \quad (4.6.19)$$

**정의 4.6.4**

Markov체인  $\{X_n | n = 0, 1, \dots\}$ 이 양의 시점(positive time)에서 상태  $y$ 에 도달하는 횟수를  $N(y)$ 로 표기하자.

확률변수  $N(y)$ 의 지지대는  $\mathbb{Z}_{>0} \cup \{\infty\}$ 이다. 또한, 다음 식이 성립한다.

$$N(y) \doteq \sum_{n=1}^{\infty} 1(X_n = y) \quad (4.6.20)$$

**명제 4.6.1**

임의의 상태들  $x$ 와  $y$ 에 대해, 다음 식이 성립한다.

$$E_x(N(y)) = \sum_{n=1}^{\infty} p^{(n)}(x, y)$$

증명. 다음 식들이 성립함을 자명하다.

$$\begin{aligned} E_x(N(y)) &= E_x\left(\sum_{n=1}^{\infty} 1(X_n = y)\right) \\ &= \sum_{n=1}^{\infty} E_x(1(X_n = y)) = \sum_{n=1}^{\infty} P_x(X_n = y) \end{aligned} \quad (1)$$

따라서 다음 식이 성립한다.

$$E_x(N(y)) = \sum_{n=1}^{\infty} p^{(n)}(x, y) \quad (2)$$

■

다음 식들이 성립함은 자명하다.

$$\{N(y) \geq 1\} = \{T_y < \infty\} \quad (4.6.21)$$

$$\{N(y) = 0\} = \{T_y = \infty\} \quad (4.6.22)$$

따라서, 다음 식들이 성립한다.

$$\Pr(N(y) \geq 1 | X_0 = x) = \Pr(T_y < \infty | X_0 = x) = f_{xy} \quad (4.6.23)$$

다음 식들이 성립한다.

$$\begin{aligned} \Pr(N(y) \geq 2 | X_0 = x) &= P_x(N(y) \geq 2) \\ &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} P_x(T_y = m) P_y(T_y = n) \\ &= \left[ \sum_{m=1}^{\infty} P_x(T_y = m) \right] \left[ \sum_{n=1}^{\infty} P_y(T_y = n) \right] \\ &= P_x(T_y < \infty) P_y(T_y < \infty) = f_{xy} \cdot f_{yy} \end{aligned} \quad (4.6.24)$$

식 (4.6.23)과 식 (4.6.24)를 일반화하면, 다음과 같다. 자연수  $k(\geq 2)$ 에 대해서 다음 식들이 성립한다.

$$\begin{aligned} &\Pr(N(y) \geq k | X_0 = x) \\ &= P_x(N(y) \geq k) \\ &= \sum_{m=1}^{\infty} \sum_{n_1=1}^{\infty} \cdots \sum_{n_{k-1}=1}^{\infty} P_x(T_y = m) P_y(T_y = n_1) \cdots P_y(T_y = n_{k-1}) \\ &= \left[ \sum_{m=1}^{\infty} P_x(T_y = m) \right] \left[ \sum_{n_1=1}^{\infty} P_y(T_y = n_1) \right] \cdots \left[ \sum_{n_{k-1}=1}^{\infty} P_y(T_y = n_{k-1}) \right] \end{aligned} \quad (4.6.25)$$

다음 식들이 성립함을 자명하다.

$$\left[ \sum_{n_1=1}^{\infty} P_y(T_y = n_1) \right] = \cdots = \left[ \sum_{n_{k-1}=1}^{\infty} P_y(T_y = n_{k-1}) \right] = P_y(T_y < \infty) \quad (4.6.26)$$

식 (4.6.25) 과 식 (4.6.26) 에서 알 수 있듯이, 다음 식들이 성립한다.

$$\Pr(N(y) \geq k | X_0 = x) = P_x(T_y < \infty) [P_y(T_y < \infty)]^{k-1} = f_{xy} \cdot f_{yy}^{k-1} \quad (4.6.27)$$

다음 식들이 성립한다.

$$\begin{aligned} \Pr(N(y) = k | X_0 = x) &= P_x(N(y) \geq k) - P_x(N(y) \geq k+1) \\ &= f_{xy} \cdot f_{yy}^{k-1} - f_{xy} \cdot f_{yy}^k = f_{xy} [1 - f_{yy}] f_{yy}^{k-1} \end{aligned} \quad (4.6.28)$$

여기서 두 번째 등호는 식 (4.6.27) 에 의해서 성립한다. 다음 식이 성립함을 자명하다.

$$\Pr(N(y) = 0 | X_0 = x) = 1 - f_{xy} \quad (4.6.29)$$

식 (4.6.28) 을 사용해서, 식 (4.6.29) 를 증명할 수도 있다. 만약  $f_{yy} < 1$  이면, 식 (4.6.28) 에서 알 수 있듯이 다음 식들이 성립한다.

$$E(N(y) = k | X_0 = x) = \sum_{k=1}^{\infty} k f_{xy} [1 - f_{yy}] f_{yy}^{k-1} = \frac{f_{xy}}{1 - f_{yy}} \quad (4.6.30)$$

여기서 두 번째 등호는 다음 식에 의해서 성립한다.

$$\sum_{k=1}^{\infty} k a^{k-1} = \frac{1}{[1 - a]^2}, \quad (|a| < 1) \quad (4.6.31)$$

지금까지 다른 내용을 정리하면, 다음과 같다.

#### 명제 4.6.2

임의의 자연수  $k$  그리고 임의의 상태들  $x$ 와  $y$ 에 대해, 다음 식들이 성립한다.

$$P_x(N(y) = 0) = \Pr(N(y) = 0 | X_0 = x) = 1 - f_{xy}$$

$$P_x(N(y) \geq k) = \Pr(N(y) \geq k | X_0 = x) = f_{xy}f_{yy}^{k-1}$$

$$P_x(N(y) = k) = \Pr(N(y) = k | X_0 = x) = f_{xy}[1 - f_{yy}]f_{yy}^{k-1}$$

$$E_x(N(y)) = E(N(y) | X_0 = x) = \frac{f_{xy}}{1 - f_{yy}}, \quad (f_{yy} < 1)$$

**정의 4.6.5**

Markov체인  $\{X_n | n = 0, 1, \dots\}$ 에서 식  $f_{yy} < 1$ 을 만족하는 상태  $y$ 를 과도적(transient)이라 하고, 식  $f_{yy} = 1$ 을 만족하는 상태  $y$ 를 재귀적(recurrent)이라 한다.

만약 상태  $y$ 가 과도적이면, 상태  $y$ 로 다시 돌아오지 않을 확률이 0보다 크다. 또한 명제 4.6.2에서 알 수 있듯이, 다음 식들이 성립한다.

$$P_x(N(y) = \infty) = \lim_{k \rightarrow \infty} \Pr(N(y) \geq k | X_0 = x) = \lim_{k \rightarrow \infty} f_{xy}f_{yy}^{k-1} = 0 \quad (4.6.32)$$

즉, 상태  $x$ 를 출발한 Markov체인이 무한 횟수에 걸쳐 상태  $y$ 로 되돌아올 확률은 0이다. 만약 상태  $y$ 가 재귀적이면, 이 상태로 돌아올 확률이 1이다. 또한 명제 4.6.2에서 알 수 있듯이, 다음 식들이 성립한다.

$$P_y(N(y) = \infty) = \lim_{k \rightarrow \infty} \Pr(N(y) \geq k | X_0 = y) = \lim_{k \rightarrow \infty} f_{yy}f_{yy}^{k-1} = 1 \quad (4.6.33)$$

즉, 상태  $y$ 를 출발한 Markov체인이 무한 횟수에 걸쳐 상태  $y$ 로 돌아올 확률은 1이다. 명제 4.6.2에서 알 수 있듯이, 다음 명제가 성립한다.

$$E(N(y) | X_0 = y) < \infty \Leftrightarrow f_{yy} < 1 \quad (4.6.34)$$

명제 4.6.1과 식 (4.6.34)에서 알 수 있듯이, 다음 정리가 성립한다.

**정리 4.6.1**

상태  $y$ 가 재귀적(recurrent)이기 위한 필요충분조건은 다음과 같다.

$$E_x(N(y)) = \sum_{n=1}^{\infty} p^{(n)}(x, y) = \infty$$

시점 0에서 상태  $x$ 를 출발한 Markov 체인이 다시 상태  $x$ 에 도달하는 횟수를  $R_x(x)$ 라 하면, 확률변수  $R_x(x)$ 의 확률질량함수는 다음과 같다.

$$\Pr(R_x(x) = n) = f_{xx}^n [1 - f_{xx}], \quad (n = 0, 1, 2, \dots) \quad (4.6.35)$$

따라서, 다음 식들이 성립한다.

$$\Pr(R_x(x) < \infty) = \sum_{n=0}^{\infty} \Pr(R_x(x) = n) = \begin{cases} 1, & (f_{xx} < 1) \\ 0, & (f_{xx} = 1) \end{cases} \quad (4.6.36)$$

식 (4.6.36)에서 알 수 있듯이, 만약  $f_{xx} < 1$ 이면, 시점 0에서 상태  $x$ 를 출발한 Markov 체인이 다시 상태  $x$ 에 돌아오는 횟수는 유한하다. 이러한 상태  $x$ 는 과도적(transient)이다. 만약  $f_{xx} = 1$ 이면, 시점 0에서 상태  $x$ 를 출발한 Markov 체인이 다시 상태  $x$ 에 돌아오는 횟수는 무한이다. 이러한 상태  $x$ 는 재귀적(recurrent)이다. 각 상태는 과도적이거나 재귀적이다. 따라서, 상태  $x$ 가 과도적이기 위한 필요충분조건은 식  $\Pr(R_x(x) < \infty | X_0 = x) = 1$ 이 성립하는 것이다. 반면에 한 번 그 상태가 되면 거기서 헤어날 수 없는 상태는 흡수적(absorbing)이다. 상태  $x$ 가 흡수적이기 위한 필요조건은  $f_{xx} = 1$ 이다.

**제 4.7 절 상호도달가능과 동치류**

제 4.6.3 절에서 도달가능성과 상호도달가능성을 정의하였다. 도달확률을 사용해서 이들을 좀 더 정형적으로 정의하면 다음과 같다.

**정의 4.7.1**

만약 식  $f_{xy} > 0$ 가 성립하면, 상태  $y$ 는 상태  $x$ 에서 도달가능(assessible)하다. 만약 식

$f_{xy} > 0$ 와 식  $f_{yx} > 0$ 가 성립하면, 상태  $x$ 와 상태  $y$ 는 상호도달가능 (communicative) 하다.

#### 정리 4.7.1

상호도달가능은 동치관계 (equivalence relation)이다. 즉, 다음 세 성질들이 성립한다.

- a) 반사율 (reflexibility)  $x \leftrightarrow x$
- b) 대칭율 (symmetry)  $x \leftrightarrow y \Leftrightarrow y \leftrightarrow x$
- c) 추이율 (transitivity)  $x \leftrightarrow y$  and  $y \leftrightarrow z \Leftrightarrow x \leftrightarrow z$

증명. 반사율과 대칭율이 성립함은 명백하다. 만약 식  $x \rightarrow y$ 와 식  $y \rightarrow z$ 가 성립하면, 식  $p^{(m)}(x, y) > 0$ 와 식  $p^{(n)}(y, z) > 0$ 를 만족하는 정수들  $m(\geq 0)$ 과  $n(\geq 0)$ 이 존재한다. 따라서, Chapman-Kolmogorov식에 의해 다음 식들이 성립함을 알 수 있다.

$$p^{(m+n)}(x, z) = \sum_{w \in S} p^{(m)}(x, w)p^{(n)}(w, z) \geq p^{(m)}(x, y)p^{(n)}(y, z) > 0 \quad (1)$$

즉, 식  $x \rightarrow z$ 가 성립한다. 또한, 대칭율에 의해서 식  $z \rightarrow x$ 가 성립한다. 따라서, 추이율이 성립한다. ■

정리 4.7.1에서 알 수 있듯이, 상호도달가능은 동치관계이다. 따라서, 같은 동치류 (equivalent class)에 속하는 두 상태들이 상호도달가능하도록, 상태공간  $S$ 를 동치류들로 분할 (partition)할 수 있다. 즉, 서로 다른 동치류들은 서로소 (disjoint) 집합들이며 또한 모든 동치류들의 합집합은 상태공간  $S$ 이다. 상태  $x$ 가 속하는 동치류를  $[x]$ 로 표기하자. 상태공간  $S$ 가 단 하나의 동치류로 이루어진 Markov 체인을 불가약 (irreducible) 또는 기약 또는 축소불가능이라 한다. 즉, 불가약인 Markov 체인에서 모든 상태들은 상호도달가능하다. 만약 그렇지 않으면, Markov 체인이 가약 (reducible) 또는 불기약 또는 축소가능이라 한다. 불가약, 기약 그리고 축소불가능이라는 용어 중에서 축소불가능이 가장 친밀하게 느껴진다. 반면에 기약이라는 용어가 가장 널리 쓰인다. 본서에서는 불가약이라는 용어를 사용하기로 하자. 현실적인 문제에서는 불가약 Markov 체인을 다루는 것이 일반적이다. 이후 특별한 언급이 없는 한, 본서에서 다루는 Markov 체인은 불가약이라고 가정하자.

어떤 상태를 출발해서 정해진 시점에 다른 어떤 상태에 최초로 도달할 확률을 다음과 같이 정의하자.

**정의 4.7.2**

시점 0에서 상태  $x$  를 출발하는 Markov 체인  $\{X_k \mid k = 0, 1, \dots\}$  이 최초로 상태  $y$  에 도달하는 최초방문시점 (first visit time) 이  $n$  일 확률을  $f_{xy}^{(n)}$  로 표기하자.

다음 식들이 성립한다.

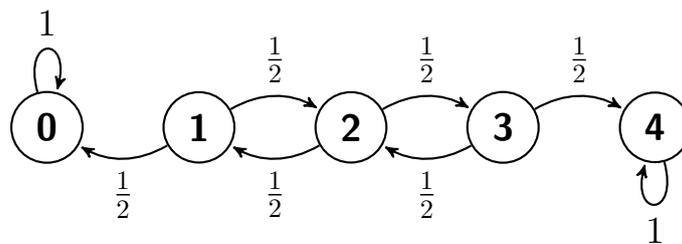
$$f_{xy} = \sum_{n=1}^{\infty} \Pr(X_1 \neq y, \dots, X_{n-1} \neq y, X_n = y \mid X_0 = x) = \sum_{n=1}^{\infty} f_{xy}^{(n)} \quad (4.7.1)$$

다음 예제들은 Markov 체인의 가약성에 관한 것들이다.

**예제 4.7.1** 상태공간이  $S = \{0, 1, 2, 3, 4\}$  인 Markov 체인  $\{X_t\}$  의 추이확률행렬이 다음과 같다고 하자.

$$\begin{array}{c}
 \\
 \\
 \\
 \\
 \mathbb{P} = \begin{array}{c}
 \begin{array}{ccccc}
 & 0 & 1 & 2 & 3 & 4 \\
 \begin{array}{c}
 0 \\
 1 \\
 2 \\
 3 \\
 4
 \end{array}
 & \begin{bmatrix}
 1 & 0 & 0 & 0 & 0 \\
 \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\
 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\
 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\
 0 & 0 & 0 & 0 & 1
 \end{bmatrix}
 & \\
 \end{array}
 \end{array}
 \end{array} \quad (1)$$

이 Markov 체인의 상태추이도는 다음과 같다.



행렬  $\mathbb{P}$ 를 Schur분해하거나 MATLAB을 사용해서, 다음 식이 성립함을 알 수 있다.

$$\mathbb{P}^\infty = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{3}{4} & 0 & 0 & 0 & \frac{1}{4} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{4} & 0 & 0 & 0 & \frac{3}{4} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix} \quad (2)$$

식 (2)에서 알 수 있듯이, 상태 0와 상태 4는 재귀적이며 흡수적이다. 또한, 상태 1, 상태 2, 그리고 상태 3은 과도적이다.

추이확률행렬  $\mathbb{P}$ 에서 상태들의 순서를 상태 0, 상태 4, 상태 1, 상태 2, 그리고 상태 3으로 바꾸면, 새로운 추이확률행렬  $\tilde{\mathbb{P}}$ 는 다음과 같다.

$$\tilde{\mathbb{P}} = \begin{matrix} & \begin{matrix} 0 & 4 & 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 4 \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix} \end{matrix} \quad (3)$$

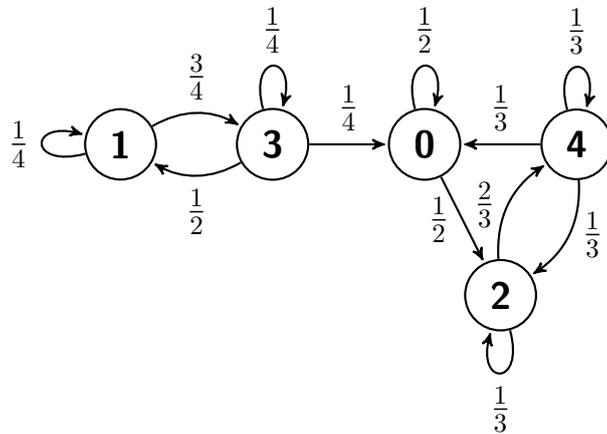
추이확률행렬  $\tilde{\mathbb{P}}$ 에서 알 수 있듯이, 이 Markov체인에서는 상태 1, 상태 2, 그리고 상태 3은 상호도달가능하고, 또한 상태 0과 상태 4는 흡수적이다. 즉, 이 Markov체인은 동치류들  $[1] = \{1, 2, 3\}$ ,  $[0] = \{0\}$ , 그리고  $[4] = \{4\}$ 로 이루어졌다. 따라서, 이 Markov체인은 가약이다. ■

**예제 4.7.2** 상태공간이  $S = \{0, 1, 2, 3, 4\}$ 인 Markov체인  $\{X_n\}$ 의 추이확률행렬이 다음과

같다고 하자.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \end{bmatrix} \end{matrix} \tag{1}$$

이 Markov 체인의 상태추이도는 다음과 같다.



행렬  $\mathbb{P}$ 를 Schur분해하거나 MATLAB을 사용하면, 다음 식이 성립함을 알 수 있다.

$$\mathbb{P}^\infty = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} \frac{1}{4} & 0 & \frac{3}{8} & 0 & \frac{3}{8} \\ \frac{1}{4} & 0 & \frac{3}{8} & 0 & \frac{3}{8} \\ \frac{1}{4} & 0 & \frac{3}{8} & 0 & \frac{3}{8} \\ \frac{1}{4} & 0 & \frac{3}{8} & 0 & \frac{3}{8} \\ \frac{1}{4} & 0 & \frac{3}{8} & 0 & \frac{3}{8} \end{bmatrix} \end{matrix} \tag{2}$$

임의의 초기분포  $\boldsymbol{\pi}^{(0)} = [\pi^{(0)}(0) \ \pi^{(0)}(1) \ \pi^{(0)}(2) \ \pi^{(0)}(3) \ \pi^{(0)}(4)]$ 에 대해서, 다음 식들이

성립한다.

$$\boldsymbol{\pi} = \lim_{m \rightarrow \infty} \boldsymbol{\pi}^{(0)} \mathbb{P}^m = \boldsymbol{\pi}^{(0)} \mathbb{P}^\infty = \left[ \frac{1}{4} \quad 0 \quad \frac{3}{8} \quad 0 \quad \frac{3}{8} \right] \quad (3)$$

즉, 초기분포에 상관없이 불변분포는  $\left[ \frac{1}{4} \quad 0 \quad \frac{3}{8} \quad 0 \quad \frac{3}{8} \right]$ 이다. 즉, 상태 1이나 상태 3에서 출발한 Markov체인은 궁극적으로 상태 1이나 상태 3으로 돌아오지 않고, 상태 0 또는 상태 2 또는 상태 4에 머물게 된다. 즉, 이 Markov체인은 가약이다.

추이확률행렬  $\mathbb{P}$ 에서 상태들의 순서를 상태 0, 상태 2, 상태 4, 상태 1, 그리고 상태 3로 바꾸면, 새로운 추이확률행렬  $\tilde{\mathbb{P}}$ 는 다음과 같다.

$$\tilde{\mathbb{P}} = \begin{matrix} & \begin{matrix} 0 & 2 & 4 & 1 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 2 \\ 4 \\ 1 \\ 3 \end{matrix} & \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix} \end{matrix} \quad (4)$$

추이확률행렬  $\tilde{\mathbb{P}}$ 에서 알 수 있듯이, 이 Markov체인에서는 상태 0, 상태 2, 그리고 상태 4가 상호도달가능하고, 또한 상태 1과 상태 3이 상호도달가능하다. 즉, 이 Markov체인은 동치류들  $[0] = \{0, 2, 4\}$ 와  $[1] = \{1, 3\}$ 으로 이루어졌다. 따라서, 이 Markov체인은 가약임을 확인할 수 있다. ■

### 정의 4.7.3

만약 자연수  $d$ 의 배수인  $n$ 에 대해서 식  $p^{(n)}(x, x) > 0$ 가 성립하고, 또한 만약  $d$ 가 이러한 성질을 만족하는 최대 정수라면,  $x$ 를 주기(period)가  $d$ 인 주기적(periodic) 상태라고 한다. 만약 각 자연수  $n$ 에 대해서  $p^{(n)}(x, x) = 0$ 이면, 상태  $x$ 의 주기를  $\infty$ 라고 한다. 또한, 주기가 1인 상태를 비주기적(aperiodic)이라 한다.

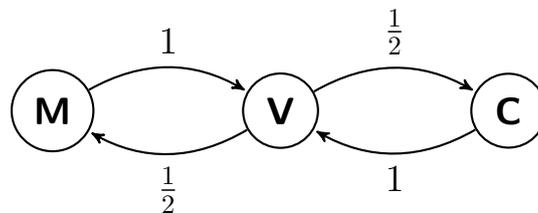
한 동치류에 속한 모든 상태들은 동일한 주기를 갖는다. 따라서, 불가약적 Markov체인의

모든 상태들은 동일한 주기를 갖는다.

**예제 4.7.3** 바닐라아이스크림 (V), 초코렛아이스크림 (C), 그리고 혼합아이스크림 (M) 을 파는 가게에서 어떤 아이스크림을 산 고객이 다음에 사는 아이스크림 종류를 나타내는 Markov 체인  $\{X_n\}$  의 추이확률행렬이 다음과 같다고 하자.

$$\begin{array}{c}
 V \quad C \quad M \\
 \mathbb{P} = \begin{array}{l}
 V \left[ \begin{array}{ccc} 0 & \frac{1}{2} & \frac{1}{2} \\
 C \left[ \begin{array}{ccc} 1 & 0 & 0 \\
 M \left[ \begin{array}{ccc} 1 & 0 & 0 \end{array} \right]
 \end{array} \right.
 \end{array} \quad (1)$$

이 Markov 체인의 상태추이도는 다음과 같다.



이 상태추이도에서 알 수 있듯이, 이 Markov 체인의 상태들은 상호도달가능하다. 즉, 이 Markov 체인은 불가약이다.

행렬  $\mathbb{P}$  의 고유값들로 이루어진 대각행렬  $\Lambda$  와 고유벡터들로 이루어진 행렬  $E$  는 각각 다음과 같다.

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad (2)$$

다음 식들이 성립한다.

$$E^{-1} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{4} & -\frac{\sqrt{3}}{4} \\ 0 & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}, \quad \Lambda^m = \begin{bmatrix} 1 & 0 & 0 \\ 0 & [-1]^m & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (3)$$

또한, 다음 식들이 성립한다.

$$\mathbb{P} = E\Lambda E^{-1} \quad (4)$$

식 (4)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\mathbb{P}^{2m-1} = E\Lambda^{2m-1}E^{-1} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad (5)$$

$$\mathbb{P}^{2m} = E\Lambda^{2m}E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad (6)$$

식 (5)와 식 (6)에서 알 수 있듯이, 각 상태는 주기가 2이다. 초기분포  $\boldsymbol{\pi}^{(0)}$ 에 대해 다음 식들이 성립함을 알 수 있다.

$$\boldsymbol{\pi}^{(2m-1)} = \left[ \pi^{(0)}(1) + \pi^{(0)}(2), \frac{1}{2}\pi^{(0)}(0), \frac{1}{2}\pi^{(0)}(0) \right] \quad (7)$$

$$\boldsymbol{\pi}^{(2m)} = \left[ \pi^{(0)}(0), \frac{\pi^{(0)}(1) + \pi^{(0)}(2)}{2}, \frac{\pi^{(0)}(1) + \pi^{(0)}(2)}{2} \right] \quad (8)$$

이 Markov 체인의 상태확률벡터가 주기 2로 진동함을 보이기 위해서, 다음 R프로그램 MCirreducibleR101.R을 실행해 보자.

```

1 # -----
2 # Filename: MCirreducibleR101.R
3 # Hierarchical Bayes Estimate when tau is a rv
4 # Programmed by CBS
5 # -----
6 setEPS()
7 # install.packages("markovchain")
8 library(markovchain)
9 stateNames <- c("Vanilla", "Chocolate", "Mixed")
10 Ice <- matrix(c(0,0.5,0.5,1,0,0,1,0,0),nrow=3, byrow=TRUE)
11 row.names(Ice) <- stateNames; colnames(Ice) <- stateNames
12 NewIce <- Ice
13 Nmax = 30
14 piMat <- matrix(0,Nmax,3)
15 piMat[1,] <- matrix(c(0.1,0.6,0.3),nrow=1, byrow=TRUE)
16 for (nn in 1:Nmax-1){
17   piMat[nn+1,] <- piMat[1,]%*%NewIce

```

```

18 OldIce <- NewIce
19 NewIce <- Ice%%OldIce
20 print(nn)
21 print(NewIce)
22 }
23 print(piMat)
24
25 # Plotting using ggplot2
26 plot.new()
27 postscript('MCirreducibleR101.eps') # Start to save figure
28 # install.packages("ggplot2")
29 library(ggplot2)
30 timee <- c(0:(Nmax-1))
31 IceData <- data.frame(timee, piMat)
32 ggplot(IceData, aes(timee)) +
33   geom_line(aes(y=piMat[,1], colour = "Vanilla"), lwd=1.5, lty=3) +
34   geom_line(aes(y=piMat[,2], colour = "Choco"), lwd=1.5, lty=1) +
35   geom_line(aes(y=piMat[,3], colour = "Mixed"), lwd=1.5, lty=2) +
36   xlab("Time") + ylab("States")
37 dev.off() # End to save figure
38 # -----

```

이 R 프로그램 MCirreducibleR101.R에서 초기분포는 다음과 같다.

$$\boldsymbol{\pi}^{(0)} = [\pi^{(0)}(0) = 0.1, \pi^{(0)}(1) = 0.6, \pi^{(0)}(2) = 0.3] \quad (9)$$

식 (7) 과 식 (8) 에서 알 수 있듯이, 각 자연수  $m$  에 대해서 상태확률벡터는 다음과 같다.

$$\boldsymbol{\pi}^{(2m-1)} = [0.9, 0.05, 0.05] \quad (10)$$

$$\boldsymbol{\pi}^{(2m)} = [0.1, 0.45, 0.45] \quad (11)$$

이 R 프로그램 MCirreducibleR101.R을 실행하면, 그림 4.7.1이 그려진다. 그림 4.7.1에서 적색 긴점선은 첫 번째 상태확률  $\pi^{(n)}(0)$ 를 나타내고, 녹색 실선은 두 번째 상태확률  $\pi^{(n)}(1)$ 을 나타내며, 청색 점선은 세 번째 상태확률  $\pi^{(n)}(2)$ 를 나타낸다. 이 그림은 식 (10)과 식 (11)에 부합된다.

다음 Python 프로그램 MCirreducible101.py을 실행하면, R 프로그램 MCirreducibleR101.R을 실행한 것과 동일한 결과를 얻는다.

```

1 # MCirreduciblePy101.py
2 """
3 Filename : MCirreduciblePy101.py
4 Hierarchical Bayes Estimate when tau is a r.v.
5 Programmed by Chang-Yong Han based on the R program by CBS
6 """
7
8 # Initialize

```

```

9 import numpy as np
10 import pandas as pd
11 import matplotlib.pyplot as plt
12
13 StateName = ['Vanilla', 'Chocolate', 'Mixed']
14 InitialState = [0.1, 0.6, 0.3]
15 TransMat = np.matrix([[0, 0.5, 0.5], [1, 0, 0], [1, 0, 0]])
16
17 # Update
18 N_Max = 30
19 piMat = pd.DataFrame(np.zeros([N_Max, 3]), columns = StateName)
20 for i in range(N_Max):
21     if i == 0:
22         piMat.iloc[i] = InitialState
23     else:
24         piMat.iloc[i] = piMat.iloc[i - 1] @ TransMat
25
26 # Plot
27 plt.xlabel('Time')
28 plt.ylabel('States')
29 plt.plot(piMat.index, piMat.Vanilla, label = 'Vanilla', ls = '--')
30 plt.plot(piMat.index, piMat.Chocolate, label = 'Chocolate', ls = '--')
31 plt.plot(piMat.index, piMat.Mixed, label = 'Mixed', ls = '--')
32 plt.legend(loc = 1, frameon = True, framealpha = 1, edgecolor = 'black')
33 plt.savefig('MCIrreduciblePy101.png')
34
35
36 # End of program
37 # -----

```



상태공간  $S$ 를 동치류로 분할할 수 있는 것과 상태의 재귀성에는 다음과 같은 밀접한 관계가 있다.

#### 정리 4.7.2

만약 상태  $x$ 가 재귀적이고 상태  $x$ 에서 상태  $y$ 로 도달가능하면, 상태  $y$ 에서 상태  $x$ 로 도달가능하고 또한 식  $f_{xy} = 1$ 과 식  $f_{yx} = 1$ 이 성립한다.

증명. 상태  $x$ 에서 상태  $y$ 로 도달가능하므로, 식  $p^{(n)}(x, y) > 0$ 를 만족하는 자연수  $n$ 이 존재한다. 이러한 자연수 중에서 최소값을  $m$ 이라 하자. 즉,  $m \doteq \min \{n \mid p^{(n)}(x, y) > 0\}$ 이다. 따라서, 다음 식들을 만족하는  $y_1, y_2, \dots, y_{m-1} \in S$ 가 존재한다.

$$p(x, y_1) p(y_1, y_2) \cdots p(y_{m-1}, y) > 0 \quad (1)$$

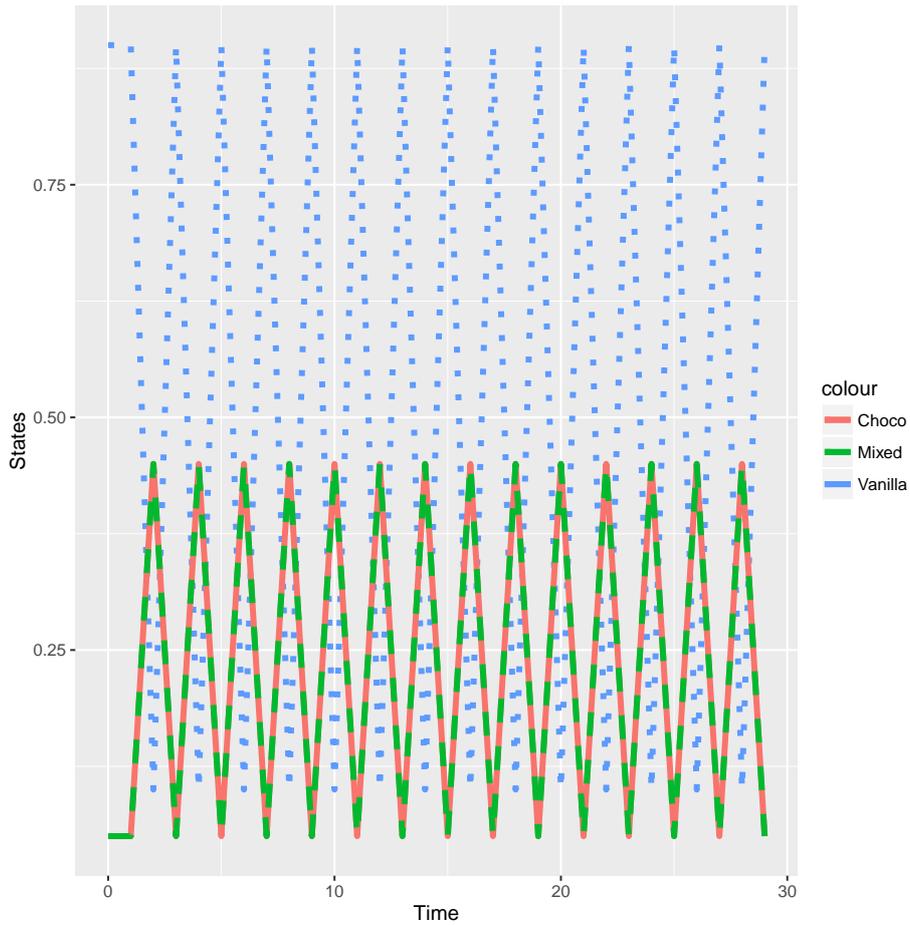


그림 4.7.1. 정규확률분포의 평균

부등식 (1)의 좌변을  $\alpha$ 로 표기하자. 최소값  $m$ 의 정의에서 알 수 있듯이, 각  $i(= 1, 2, \dots, m-1)$ 에 대해서  $y_i \neq x$ 이다. 따라서, 다음 식들이 성립한다.

$$\begin{aligned} \alpha [1 - f_{yx}] &\leq \Pr(\text{going from } x \text{ to } y, \text{ then never returning to } x) \\ &\leq P_x(T_x = \infty) = 0 \end{aligned} \tag{2}$$

여기서 마지막 등호는 상태  $x$ 의 재귀성에 의해서 성립한다. 식 (1)에서 알 수 있듯이,  $\alpha$ 는 양수이다. 식 (2)에서 알 수 있듯이, 식  $f_{yx} = 1$ 이 성립한다. 즉, 상태  $y$ 에서 상태  $x$ 로도달가능하다. 결과적으로, 식  $x \leftrightarrow y$ 가 성립한다.

임의의 자연수들  $j, k$ 와  $l$ 에 대해서, 다음 식이 성립한다.

$$p^{(j+k+l)}(y, y) \geq p^{(j)}(y, x)p^{(k)}(x, x)p^{(l)}(x, y) \tag{3}$$

식  $x \leftrightarrow y$ 에서 알 수 있듯이, 다음 식들을 만족하는 자연수들  $j$ 와  $l$ 이 존재한다.

$$p^{(j)}(y, x) > 0, \quad p^{(l)}(x, y) > 0 \quad (4)$$

상태  $x$ 가 재귀적이므로, 정리 4.6.1에서 알 수 있듯이 다음 식이 성립한다.

$$\sum_{k=1}^{\infty} p^{(k)}(x, x) = \infty \quad (5)$$

식 (3)~식 (5)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\sum_{n=1}^{\infty} p^{(n)}(y, y) \geq \sum_{k=1}^{\infty} p^{(j+k+l)}(y, y) \geq p^{(j)}(y, x) \left[ \sum_{k=1}^{\infty} p^{(k)}(x, x) \right] p^{(l)}(x, y) = \infty \quad (6)$$

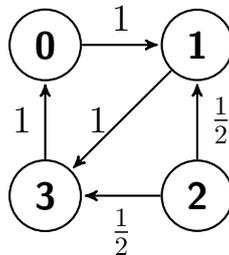
즉, 상태  $y$ 는 재귀적이다. 상태  $y$ 가 재귀적이고 상태  $y$ 에서 상태  $x$ 로 도달가능하므로, 식  $f_{xy} = 1$ 이 성립한다. ■

정리 4.7.2에서 알 수 있듯이, 상태  $x$ 가 재귀적이고 상태  $x$ 에서 상태  $y$ 로 도달가능하면, 상태  $x$ 와 상태  $y$ 는 같은 동치류에 속한다. 다음 따름정리가 성립한다.

#### 따름정리 4.7.1

만약 식  $x \rightarrow y$ 와 식  $y \nrightarrow x$ 가 성립하면, 상태  $x$ 는 과도적이다.

**예제 4.7.4** 상태공간이  $S = \{0, 1, 2, 3\}$ 이고 상태추이도가 다음과 같은 Markov 체인을 생각해 보자.



이 Markov 체인에서 상태들  $\{0, 1, 3\}$ 는 상호도달가능이다. 예를 들어, 상태 0은 재귀적이고 식  $0 \rightarrow 3$ 가 성립한다. 정리 4.7.2에서 알 수 있듯이, 식  $3 \rightarrow 0$ 가 성립하며 또한 식  $f_{03} = 1$ 과 식

$f_{30} = 1$ 이 성립한다. 즉, 상태 3는 재귀적이다. 그러나, 식  $2 \rightarrow 3$ 와 식  $3 \nrightarrow 2$ 가 성립하므로, 상태 2는 과도적이다. ■

앞에서도 언급했듯이, 재귀성과 과도성은 동치류의 성질이다. 따라서, 만약 상태  $x$ 가 재귀적이고 또한 만약 상태  $x$ 에서 상태  $y$ 로 도달할 수 있으면, 식  $f_{xy} = 1$ 과 식  $f_{yx} = 1$ 이 성립한다. 따라서 상태  $y$ 도 재귀적이다. 즉, 만약 상태  $x$ 에서 상태  $y$ 가 상호도달가능하면, 상태  $x$ 와 상태  $y$ 는 모두 재귀적이거나 또는 모두 과도적이다.

일반적으로  $f_{xx}$ 를 계산해서 상태  $x$ 가 재귀적인지 과도적인지를 판정하는 것은 쉬운 일이 아니다. 다음 식들이 성립한다.

$$\begin{aligned} E(R_x(x)) &= E\left(\sum_{n=1}^{\infty} 1(X_n = x) \mid X_0 = x\right) \\ &= \sum_{n=1}^{\infty} E(1(X_n = x) \mid X_0 = x) = \sum_{n=1}^{\infty} \Pr(X_n = x \mid X_0 = x) \\ &= \sum_{n=1}^{\infty} p^{(n)}(x, x) \end{aligned} \quad (4.7.2)$$

만약 상태  $x$ 가 과도적이면, 즉  $0 < f_{xx} < 1$ 이면, 다음 식들이 성립한다.

$$\begin{aligned} E(R_x(x)) &= \sum_{n=1}^{\infty} n \Pr(R_x(x) = n) \\ &= \sum_{n=1}^{\infty} n f_{xx}^n [1 - f_{xx}] = \frac{f_{xx}}{1 - f_{xx}} < \infty \end{aligned} \quad (4.7.3)$$

여기서 두 번째 등호는 식 (4.6.35)에 의해서, 그리고 세 번째 등호는 다음 식에 의해서 성립한다.

$$\sum_{n=1}^{\infty} n x^{n-1} = \frac{1}{[1-x]^2} \quad (4.7.4)$$

식 (4.7.2)와 식 (4.7.3)에서 알 수 있듯이, 다음 명제가 성립한다.

#### 명제 4.7.1

Markov체인에서 상태  $x$ 가 과도적이기 위한 필요충분조건은 식  $\sum_{n=1}^{\infty} p^{(n)}(x, x) < \infty$ 가

성립하는 것이다. 또한, 상태  $x$ 가 재귀적이기 위한 필요충분조건은 식  $\sum_{n=1}^{\infty} p^{(n)}(x, x) = \infty$ 가 성립하는 것이다.

**예제 4.7.5** 예제 4.5.3의 버스문제를 다시 살펴보자. 예제 4.5.3의 식 (17)은 다음과 같다.

$$\mathbb{P}^m = \frac{1}{\alpha + \beta} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{[1 - \alpha - \beta]^m}{\alpha + \beta} \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} \quad (1)$$

따라서, 각 자연수  $n$ 에 대해서 다음 식들이 성립한다.

$$p^{(n)}(0, 0) = \frac{\beta}{\alpha + \beta} + \frac{\alpha}{\alpha + \beta}[1 - \alpha - \beta]^n \quad (2)$$

$$p^{(n)}(1, 1) = \frac{\alpha}{\alpha + \beta} + \frac{\beta}{\alpha + \beta}[1 - \alpha - \beta]^n \quad (3)$$

만약  $\alpha = 0$ 이고  $\beta > 0$ 이면, 다음 식이 성립한다.

$$\sum_{n=1}^{\infty} p^{(n)}(0, 0) = \infty, \quad \sum_{n=1}^{\infty} p^{(n)}(1, 1) < \infty \quad (4)$$

따라서, 상태 0은 재귀적이고 상태 1은 과도적이다. 마찬가지로, 만약  $\alpha > 0$ 이고  $\beta = 0$ 이면, 상태 0은 과도적이고 상태 1은 재귀적이다. 만약  $\alpha > 0$ 이고  $\beta > 0$ 이면, 상태 0과 상태 1 모두 재귀적이다. 여기서 상태 0은 학교버스를 타는 것 그리고 상태 1은 공공버스를 타는 것임에 유의하라. ■

### 제4.8절 닫힌집합과 재귀성

Markov 체인에서 닫힌집합 (closed set)이라는 개념을 이용하면, 상태의 동치류를 이해하기 쉽다.

**정의 4.8.1**

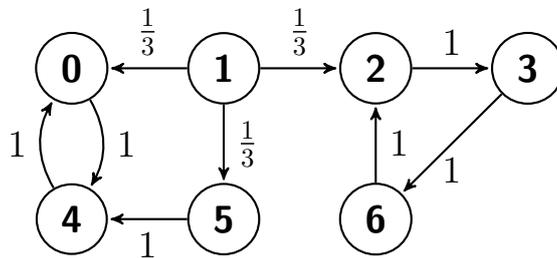
상태공간  $S$ 의 부분집합  $C$ 가 다음 명제를 만족하면,  $C$ 는 닫혀있다고 (closed) 한다.

$$\forall x \in C \text{ and } \forall y \notin C \Leftrightarrow x \not\rightarrow y$$

정의 4.8.1에서 알 수 있듯이, 닫힌집합  $C$ 에 속하는 어떤 상태도 집합  $C$ 에서 빠져나올 수 없다. 또한, 전체 상태공간  $S$ 는 항상 닫힌집합이다.

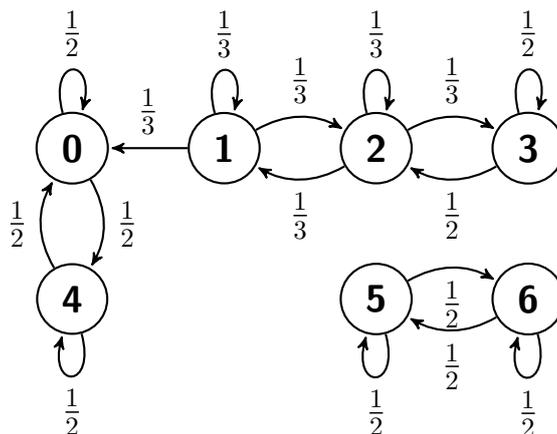
**예제 4.8.1** 상태들의 닫힌집합에 대한 예를 살펴보자.

첫 번째 예제로, 상태공간이  $S = \{0, 1, 2, 3, 4, 5, 6, 7\}$  이고 상태추이도가 다음과 같은 Markov 체인을 생각해 보자.



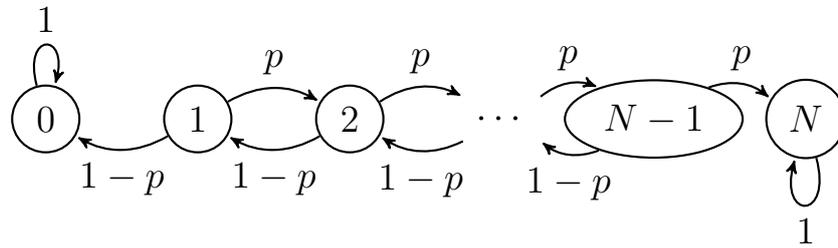
이 Markov 체인에서  $\{0, 4\}$ 와  $\{2, 3, 6\}$ 는 닫힌집합들이고 그들의 합집합인  $\{0, 2, 3, 4, 6\}$ 도 닫힌집합이다. 이 집합에 5를 추가한  $\{0, 2, 3, 4, 5, 6\}$ 도 닫힌집합이다. 전체 상태공간  $S$ 는 항상 닫힌집합이다. 또한, 불가약인 집합들은  $\{0, 4\}$ 와  $\{2, 3, 6\}$ 이다.

두 번째 예제로, 상태공간이  $S = \{0, 1, 2, 3, 4, 5, 6, 7\}$  이고 상태추이도가 다음과 같은 Markov 체인을 생각해 보자.



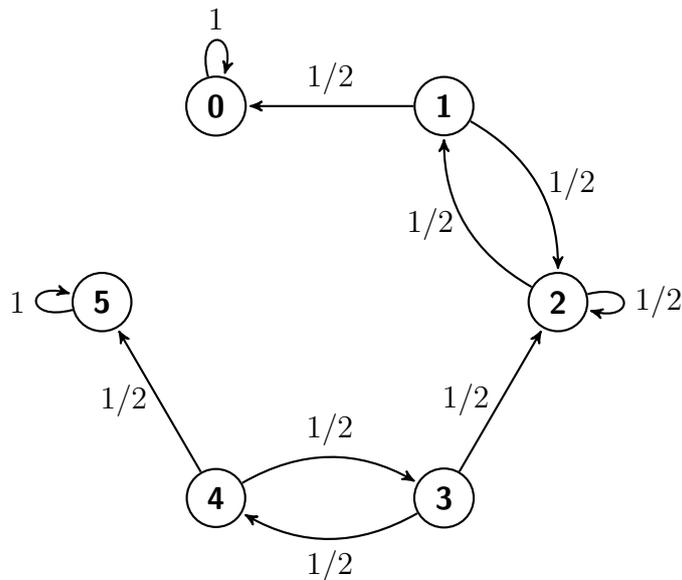
이 Markov 체인에서 닫힌집합들은  $\{0, 4\}$ ,  $\{5, 6\}$ ,  $\{0, 4, 5, 6\}$ 이다. 또한, 불가약인 집합들은  $\{0, 4\}$ ,  $\{1, 2, 3\}$ ,  $\{5, 6\}$ 이다.

세 번째 예제로, 상태공간이  $S = \{0, 1, \dots, N-1, N\}$  이고 상태추이도가 다음과 같은 Markov 체인을 생각해 보자.



여기서  $p \in (0, 1)$ 는 상수이다. 이 Markov 체인에서 닫힌집합들은  $\{0\}, \{N\}, \{0, N\}, S$ 이다. 또한, 불가약인 집합은  $\{0\}, \{N\}, \{1, 2, \dots, N-1\}$ 이다.

네 번째 예제로, 상태공간이  $S = \{0, 1, \dots, N-1, N\}$  이고 상태추이도가 다음과 같은 Markov 체인을 생각해 보자.



이 Markov 체인에서 닫힌집합들은  $\{0\}, \{0, 1, 2\}, \{5\}, S$ 이다. 또한, 불가약인 집합은  $\{0\}, \{1, 2\}, \{3, 4\}, \{5\}$ 이다. ■

**보조정리 4.8.1**

만약  $C \subset S$ 가 유한인 (finite) 닫힌집합이라면, 집합  $C$ 는 재귀적 상태를 적어도 하나 포함한다.

증명. 집합  $C$ 에 속한 각 상태가 과도적이라고 가정하고, 고정된 상태  $x \in C$ 를 생각해 보자. 만약  $X_0 = x$ 라 하면,  $C$ 가 닫힌집합이므로  $X_1 \in C, X_2 \in C, \dots$ 이다. 이 성질과 집합  $C$ 가 유한이라는 가정에 의해서, 식  $N(z) = \infty$ 를 만족하는  $z \in C$ 가 적어도 하나 존재한다. 따라서, 다음 식이 성립한다.

$$\sum_{y \in C} N(y) = \infty \quad (1)$$

다음 식이 성립한다.

$$E_x \left( \sum_{y \in C} N(y) \right) = \infty \quad (2)$$

집합  $C$ 에 속한 각 상태  $y$ 는 과도적이라고 가정했으므로, 명제 4.6.2에서 알 수 있듯이 다음 식들이 성립한다.

$$E_x(N(y)) = \frac{f_{xy}}{1 - f_{xy}} < \infty \quad (3)$$

집합  $C$ 가 유한이므로, 식 (3)에서 알 수 있듯이 다음 식들이 성립한다.

$$E_x \left( \sum_{y \in C} N(y) \right) = \sum_{y \in C} \frac{f_{xy}}{1 - f_{xy}} < \infty \quad (4)$$

식 (2)와 식 (4)는 모순이다. ■

정리 4.7.2와 보조정리 4.8.1에서 알 수 있듯이, 다음 정리가 성립한다.

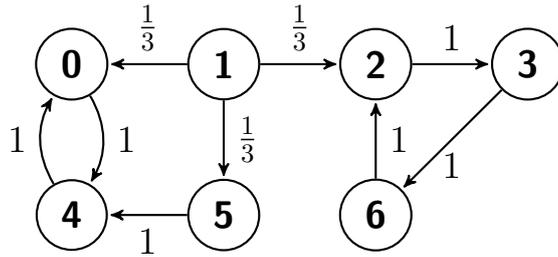
#### 정리 4.8.1

만약  $C(C \subset S)$ 가 유한이고 닫혔으며 불가약이라면, 집합  $C$ 의 각 상태는 재귀적이다.

따름정리 4.7.1과 정리 4.8.1에서 알 수 있듯이, 재귀성과 과도성은 동치류의 성질이다.

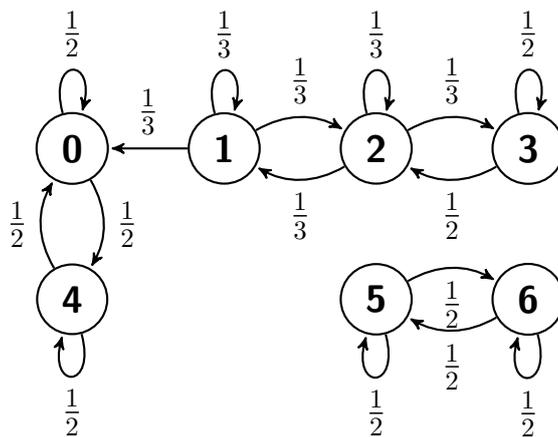
**예제 4.8.2** 예제 4.8.1을 다시 생각해 보자.

첫 번째 예는 상태공간이  $S = \{0, 1, 2, 3, 4, 5, 6, 7\}$ 이고 상태추이도가 다음과 같은 Markov 체인이다.



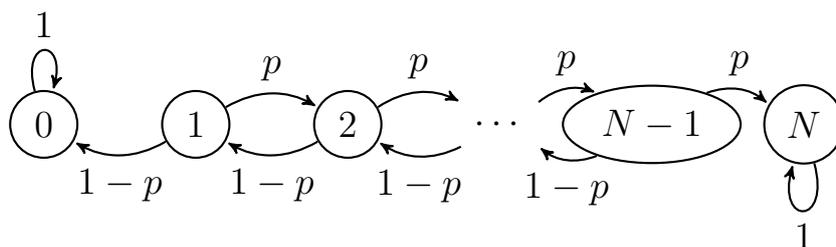
이 Markov 체인에서 집합  $\{0, 4\}$ 는 유한이고 닫혔으며 불가약이다. 따라서, 상태들 0와 4는 재귀적이다. 또한, 집합  $\{2, 3, 6\}$ 는 유한이고 닫혔으며 불가약이다. 따라서, 상태들 2, 3과 6은 재귀적이다. 식  $1 \rightarrow 0$ 와 식  $0 \not\rightarrow 1$ 이 성립한다. 따라서, 상태 1은 과도적이다. 식  $5 \rightarrow 4$ 와 식  $4 \not\rightarrow 5$ 가 성립한다. 따라서, 상태 5는 과도적이다.

두 번째 예는 상태공간이  $S = \{0, 1, 2, 3, 4, 5, 6, 7\}$ 이고 상태추이도가 다음과 같은 Markov 체인이다.



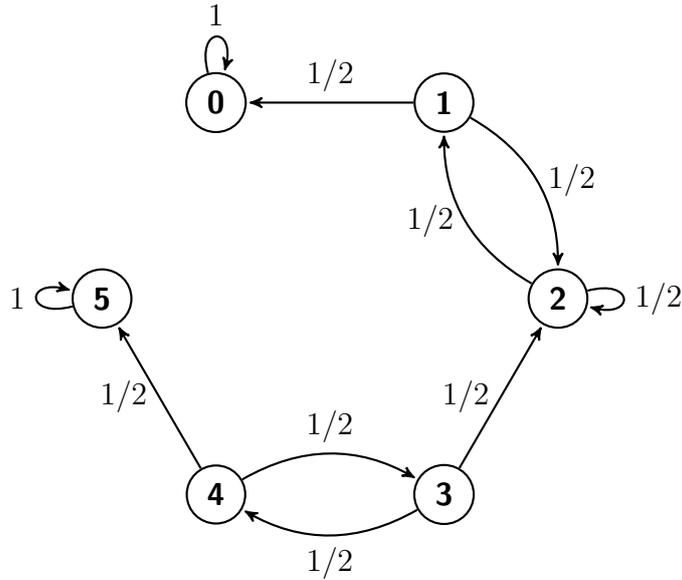
이 Markov 체인에서 집합  $\{0, 4\}$ 는 유한이고 닫혔으며 불가약이다. 따라서, 상태들 0과 4는 재귀적이다. 또한, 집합  $\{5, 6\}$ 는 유한이고 닫혔으며 불가약이다. 따라서, 상태들 5와 6은 재귀적이다. 집합  $\{1, 2, 3\}$ 에서 상태 1로 도달할 수 있어도 그 역은 성립하지 않는다. 따라서, 상태들 1, 2와 3은 과도적이다.

세 번째 예는 상태공간이  $S = \{0, 1, \dots, N-1, N\}$ 이고 상태추이도가 다음과 같은 Markov 체인이다.



여기서  $p \in (0, 1)$ 는 상수이다. 이 Markov체인에서 집합  $\{0\}$ 는 유한이고 닫혔으며 불가약이다. 따라서, 상태 0은 재귀적이다. 같은 이유로 상태  $N$ 도 재귀적이다. 그러나, 집합  $\{1, 2, \dots, N-1\}$ 에서 상태 0 또는 상태  $N$ 에 도달할 수 있어도 그 역은 성립하지 않는다. 따라서, 상태들 1과  $N$ 은 과도적이다.

네 번째 예는 상태공간이  $S = \{0, 1, \dots, N-1, N\}$ 이고 상태추이도가 다음과 같은 Markov 체인이다.



이 Markov체인에서 집합  $\{0\}$ 은 유한이고 닫혔으며 불가약이다. 따라서, 상태 0은 재귀적이다. 같은 이유로 상태 5도 재귀적이다. 그러나, 집합  $\{1, 2\}$ 에서 상태 0에 도달할 수 있어도 그 역은 성립하지 않는다. 따라서, 상태들 1과 2는 과도적이다. 집합  $\{3, 4\}$ 에서 상태 2 또는 상태 5에 도달할 수 있어도 그 역은 성립하지 않는다. 따라서, 상태들 3과 4는 과도적이다. ■

비록 집합  $C$ 가 불가약인 닫힌집합이라 하더라도 무한집합(infinite set)이면, 어떤 상태들은 과도적일 수 있다. 이 경우에는 모든 상태들이 과도적이거나 또는 모든 상태들이 재귀적이다.

**예제 4.8.3** 예제 4.4.13에서 다룬 단순확률보행을 살펴보자. 서로 독립인 확률변수들  $\{\xi_n | n = 1, 2, \dots\}$ 이 다음 식들을 만족하는 동일한 확률분포를 따른다고 하자.

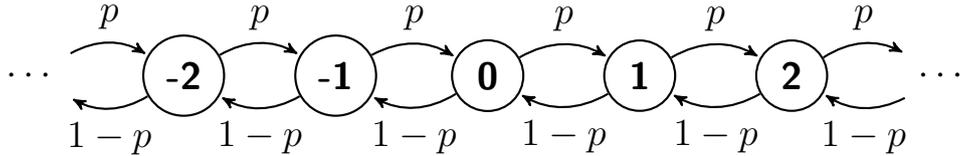
$$\Pr(\xi_n = 1) = p = 1 - \Pr(\xi_n = -1) \tag{1}$$

여기서  $p$ 는 개구간  $(0, 1)$ 에 속한다. 확률변수  $X_0$ 가  $\{\xi_n | n = 1, 2, \dots\}$ 과는 독립적이며 정수

값을 갖는다고 하자. 다음 확률변수들을 정의하자.

$$X_n \doteq X_0 + \sum_{k=1}^n \xi_k, \quad (n = 1, 2, \dots) \quad (2)$$

이렇게 정의되는 단순확률보행  $\{X_n | n = 0, 1, \dots\}$ 의 상태공간은  $S = \mathbb{Z}_{\geq 0}$ 이고 상태추이도는 다음과 같다.



상태집합  $S$ 는 불가약인 닫힌집합이다. 따라서, 모든 상태들이 과도적이거나 또는 모든 상태들이 재귀적이다. 즉, 상태 0이 과도적인지 재귀적인지 여부만 조사하면 나머지 상태들도 과도적인지 재귀적인지를 알 수 있다. 각 자연수  $n$ 에 대해서 다음 식이 성립함을 자명하다.

$$p^{(2n-1)}(0, 0) = 0 \quad (3)$$

이항확률모형에서 알 수 있듯이, 각 자연수  $n$ 에 대해서 다음 식이 성립한다.

$$p^{(2n)}(0, 0) = \binom{2n}{n} p^n [1-p]^n \quad (4)$$

식 (3)과 식 (4)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\begin{aligned} \sum_{n=1}^{\infty} p^{(n)}(0, 0) &= \sum_{n=1}^{\infty} p^{(2n-1)}(0, 0) + \sum_{n=1}^{\infty} p^{(2n)}(0, 0) \\ &= \sum_{n=1}^{\infty} \frac{(2n)!}{n!n!} p^n [1-p]^n \approx \sum_{n=1}^{\infty} \frac{\{4p[1-p]\}^n}{\sqrt{\pi n}} \end{aligned} \quad (5)$$

만약 식  $p \neq \frac{1}{2}$ 이 성립하면, 식  $4p[1-p] < 1$ 이 성립한다. 따라서, 다음 식들이 성립한다.

$$\sum_{n=1}^{\infty} p^{(n)}(0, 0) \leq \sum_{n=1}^{\infty} \frac{\{4p[1-p]\}^n}{\sqrt{\pi n}} < \infty \quad (6)$$

식 (6)과 명제 4.7.1에서 알 수 있듯이, 상태 0은 과도적이다. 따라서, 모든 상태들이 과도적이다.

만약 식  $p = \frac{1}{2}$ 이 성립하면, 식  $4p[1 - p] = 1$ 이 성립한다. 따라서, 다음 식들이 성립한다.

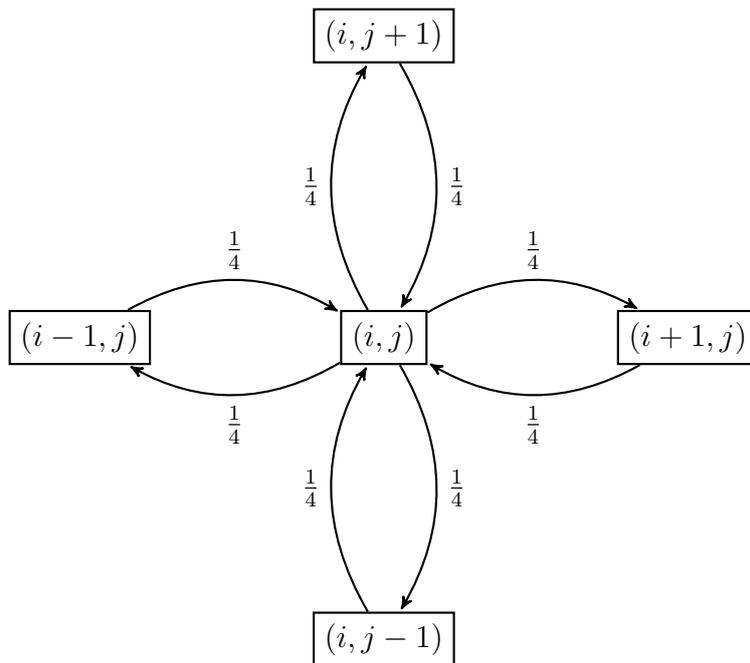
$$\sum_{n=1}^{\infty} p^{(n)}(0, 0) \approx \sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty \quad (7)$$

식 (7)과 명제 4.7.1에서 알 수 있듯이, 상태 0은 재귀적이다. 따라서, 모든 상태들이 재귀적이다. ■

**예제 4.8.4** 상태공간이 2차원인 대칭확률보행을 살펴보자. 시점  $n$ 에서 상태가  $X_n = (i, j)$ 인 경우 시점  $[n + 1]$ 에서 상태가  $X_{n+1} = (k, l)$ 이 되는 추이확률이 다음과 같다고 하자.

$$p((i, j), (k, l)) \doteq \Pr(X_{n+1} = (k, l) | X_n = (i, j)) = \begin{cases} \frac{1}{4}, & ((k, l) \in \{(i, j - 1), (i, j + 1), (i - 1, j), (i + 1, j)\}) \\ 0, & (\text{otherwise}) \end{cases} \quad (1)$$

이렇게 정의되는 단순확률보행  $\{X_n | n = 0, 1, \dots\}$ 의 상태공간은  $S = \mathbb{Z}_{\geq 0}^2$ 이고 상태추이도의 일부는 다음과 같다.



시점 0에서 원점  $O(0, 0)$ 를 출발한 대칭확률보행이 시점  $n$ 에서 다시 원점  $O$ 로 돌아올

확률  $p_{OO}^{(n)}$ 는 다음 식을 만족한다.

$$p_{OO}^{(n)} = \begin{cases} \sum_{q+r=m} \frac{(2m)!}{q!r!q!r!} \frac{1}{4^{2m}}, & (n = 2m) \\ 0, & (\text{otherwise}) \end{cases} \quad (2)$$

예제 4.8.3의 식 (5)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\begin{aligned} p_{OO}^{(2m)} &= \binom{2m}{m} \frac{1}{4^{2m}} \sum_{q=0}^m \binom{m}{q} \binom{m}{m-q} \\ &= \left\{ \binom{2m}{m} \frac{1}{4^m} \right\}^2 \approx \left[ \frac{1}{\sqrt{\pi m}} \right]^2 = \frac{1}{\pi m} \end{aligned} \quad (3)$$

따라서 다음 식들이 성립한다.

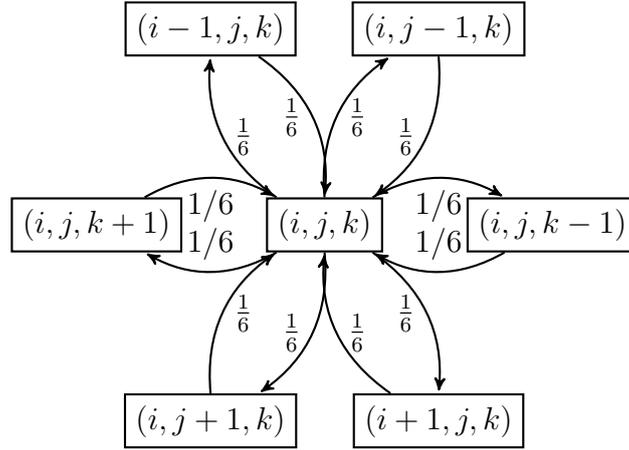
$$\sum_{m=1}^{\infty} p_{OO}^{(2m)} \approx \sum_{m=1}^{\infty} \frac{1}{\pi m} = \infty \quad (4)$$

식 (4)와 명제 4.7.1에서 알 수 있듯이, 상태  $O$ 는 재귀적이다. 상태집합  $S$ 는 불가약인 단한집합이다. 따라서, 모든 상태들이 재귀적이다. ■

**예제 4.8.5** 상태공간이 3차원인 대칭확률보행을 살펴보자. 시점  $n$ 에서 상태가  $X_n = (i, j, k)$ 인 경우 시점  $[n + 1]$ 에서 상태가  $X_{n+1} = (p, q, r)$ 이 되는 추이확률이 다음과 같다고 하자.

$$\begin{aligned} p((i, j), (p, q, r)) &\doteq \Pr(X_{n+1} = (k, l) \mid X_n = (i, j)) \\ &= \begin{cases} \frac{1}{6}, & ((p, q, r) \in \{(i, j, k - 1), (i, j, k + 1), (i, j - 1, k), \\ & \quad (i, j + 1, k), (i - 1, j, k), (i + 1, j, k)\}), \\ 0, & (\text{otherwise}) \end{cases} \end{aligned} \quad (1)$$

이렇게 정의되는 단순확률보행  $\{X_n \mid n = 0, 1, \dots\}$ 의 상태공간은  $S = \mathbb{Z}_{\geq 0}^3$ 이고 상태추이도의 일부는 다음과 같다.



시점 0에서 원점  $O(0, 0, 0)$ 를 출발한 대칭확률보행이 시점  $n$ 에서 다시 원점  $O$ 로 돌아올 확률  $p_{OO}^{(n)}$ 는 다음과 같다.

$$p_{OO}^{(n)} = \begin{cases} \sum_{a+b+c=m} \frac{(2m)!}{a!b!c!a!b!c!} \frac{1}{6^{2m}}, & (n = 2m) \\ 0, & (\text{otherwise}) \end{cases} \quad (2)$$

따라서, 다음 식들이 성립한다.

$$\begin{aligned} p_{OO}^{(2m)} &= \frac{1}{2^{2m}} \binom{2m}{m} \sum_{a+b+c=m} \left[ \frac{1}{3^m} \frac{m!}{a!b!c!} \right]^2 \\ &= \frac{1}{2^{2m}} \binom{2m}{m} \max_{a,b} \frac{1}{3^m} \frac{m!}{a!b!(m-a-b)!} \sum_{a+b+c=m} \frac{1}{3^m} \frac{m!}{a!b!c!} \\ &\leq \frac{1}{2^{2m}} \binom{2m}{m} \max_{a,b} \frac{1}{3^m} \frac{m!}{a!b!(m-a-b)!} \frac{1}{3^m} [1+1+1]^m \end{aligned} \quad (3)$$

식 (3)의 우변은 식  $a = b = c = \lfloor \frac{m}{3} \rfloor$  또는 식  $a = b = c = \lfloor \frac{m}{3} + 1 \rfloor$ 이 성립할 때 최대값을 갖는다. 여기서  $\lfloor x \rfloor$ 는  $x$ 를 넘지 않는 최대 정수이다. 따라서, 아주 큰  $m$ 에 대해서 다음 식이 성립한다.

$$p_{OO}^{(2m)} \leq \frac{1}{2^{2m}} \binom{2m}{m} \frac{1}{3^m} \left( \frac{m}{3} \right)! \quad (4)$$

식 (4)에 Stirling식을 적용하면, 다음 식을 얻는다.

$$p_{OO}^{(2m)} \leq \frac{1}{2} \left[ \frac{3}{\pi m} \right]^{3/2} \quad (5)$$

식 (2)와 식 (5)에서 알 수 있듯이, 다음 식이 성립한다.

$$\sum_{n=1}^{\infty} p_{OO}^{(2m)} < \infty \tag{6}$$

식 (6)과 명제 4.7.1에서 알 수 있듯이, 상태  $O$ 는 과도적이다. 상태집합  $S$ 는 불가약인 닫힌 집합이다. 따라서, 모든 상태들이 과도적이다. 이 결과를 바탕으로 고정점이론으로 유명한 Shizuo Kakutani는 다음과 같은 유머를 만들었다.

A drunk man will find his way home, but a drunk bird may get lost forever.



**정리 4.8.2**

어떤 Markov 체인의 상태집합  $S$ 가 가산이라고 (countable) 하자. 이 Markov 체인의 모든 과도적 상태들의 집합을  $T$ 라고 하면, 재귀적 상태들을 불가약한 (irreducible) 닫힌집합들  $C_1, C_2, \dots$ 로 분할할 수 있다. 이 성질을 다음 식들로 나타낼 수 있다.

$$S = T \cup C_1 \cup C_2 \cup \dots, \quad C_i \cap C_j = \emptyset, \quad (i \neq j)$$

증명. 주어진  $x \notin T$ 에 대해서 집합  $C_x$ 를 다음과 같이 정의하자.

$$C_x \doteq \{y \mid x \rightarrow y\} \tag{1}$$

상태  $x$ 가 재귀적이므로, 정의 4.6.5에서 알 수 있듯이 식  $f_{xx} = 1$ 이 성립한다. 따라서, 식  $x \in C_x$ 가 성립한다. 정리 4.7.2에서 알 수 있듯이, 각  $y \in C_x$ 에 대해서 식  $x \leftrightarrow y$ 가 성립한다. 따라서,  $y$ 는 재귀적이다. 다음 식이 성립한다.

$$T \cap C_x = \emptyset \tag{2}$$

지금부터는  $C_x$ 가 닫힌집합임을 증명하자. 만약 어떤  $y \in C_x$ 에서  $z$ 로 도달가능하면, Markov 체인의 추이율 (transitivity)에 의해서 식  $x \rightarrow z$ 가 성립한다. 따라서,  $z \in C_x$ 가 성립한다. 즉,  $C_x$ 는 닫힌집합이다.

지금부터는 집합  $C_x$ 가 불가약임을 증명하자. 상태들  $y \in C_x$ 와  $z \in C_x$ 에 대해서 다음 명제들이 성립한다.

$$x \rightarrow y \text{ and } x \rightarrow z \quad (3)$$

$$x \rightarrow y \text{ and } z \rightarrow x \quad (4)$$

$$z \rightarrow y \quad (5)$$

여기서 식 (4)는 정리 4.7.2에 의해서 성립하고, 식 (5)는 Markov 체인의 추이율에 의해서 성립한다. 따라서 집합  $C_x$ 는 불가약이다.

각  $x \in S \setminus T$ 에 대해서  $C_x$ 가 존재한다. 상태집합  $S$ 가 가산적이므로 이러한 집합들은 가산 개이다. 지금부터는 각  $x \in S \setminus T$ 와 각  $y \in S \setminus T$ 에 대해서 식  $C_x = C_y$  또는 식  $C_x \cap C_y = \emptyset$ 가 성립함을 증명하자. 만약 상태  $z \in C_x \cap C_y$ 가 존재하면, 집합  $C_x$ 가 불가약이므로 임의의  $w \in C_x$ 에 대해서 식  $z \rightarrow w$ 가 성립한다. 집합  $C_y$ 가 닫혀있고 식  $z \in C_y$ 가 성립하므로, 식  $z \rightarrow w$ 로부터 식  $w \in C_y$ 이 성립함을 알 수 있다. 즉, 식  $C_x \subset C_y$ 가 성립한다. 대칭성에 의해서, 식  $C_y \subset C_x$ 가 성립함을 알 수 있다. 즉, 식  $C_x = C_y$ 가 성립한다. 반면에, 만약 집합  $C_x \cap C_y$ 에 원소가 존재하지 않으면, 식  $C_x \cap C_y = \emptyset$ 가 성립한다. ■

## 제4.9절 에르고딕

상태  $x$ 로 재귀하기 위해 필요한 평균재귀시간(mean recurrent time)을  $\mu_{xx}$ 라 하자. 만약 상태  $x$ 가 과도적이면,  $\mu_{xx} = \infty$ 이다. 만약 상태  $x$ 가 재귀적이면, 다음 식이 성립한다.

$$\mu_{xx} = \sum_{n=1}^{\infty} n f_{xx}^{(n)} \quad (4.9.1)$$

### 정의 4.9.1

만약 재귀적 상태  $x$ 의 평균재귀시간  $\mu_{xx}$ 가 유한이면, 상태  $x$ 는 양재귀적(positive recurrent)이라 한다. 만약 재귀적 상태  $x$ 의 평균재귀시간이 무한이면, 즉  $\mu_{xx} = \infty$ 이면, 상태  $x$ 는 영재귀적(zero recurrent)이라 한다.

양재귀성과 영재귀성은 각각 동치류의 성질이다. 즉, 만약 상태  $x$ 와 상태  $y$ 가 상호도달가능이고, 또한 만약 상태  $x$ 가 양재귀적(영재귀적)이면, 상태  $y$ 도 양재귀적(영재귀적)이다. 다음 명제들에 대한 자세한 내용은 Ross [85]의 제4.3절을 참조하라.

#### 명제 4.9.1

Markov체인  $\{X_n \mid n = 0, 1, \dots\}$ 에서 상태  $x$ 와 상태  $y$ 가 상호도달가능하면, 다음 성질들이 성립한다.

a)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p^{(k)}(x, y) = \frac{1}{\mu_{yy}}$$

b) 만약 상태  $y$ 가 비주기적이면,  $\lim_{n \rightarrow \infty} p^{(n)}(x, y) = \frac{1}{\mu_{yy}}$ 이다.

c) 만약 상태  $y$ 가 주기  $d$ 를 갖는다면,  $\lim_{n \rightarrow \infty} p^{(nd)}(y, y) = \frac{d}{\mu_{yy}}$ 이다.

#### 명제 4.9.2

주기가  $d$ 인 재귀적 상태  $y$ 에 대해서 다음 상수를 정의하자.

$$\pi_y \doteq \lim_{n \rightarrow \infty} p^{(nd)}(y, y)$$

만약  $\pi_y > 0$ 이면, 재귀적 상태  $y$ 는 양재귀적이다. 만약  $\pi_y = 0$ 이면, 재귀적 상태  $y$ 는 영재귀적이다.

**예제 4.9.1** 예제 4.7.3의 아이스크림문제를 다시 살펴보자. 이 Markov체인  $\{X_n\}$ 의 추이 확률행렬이 다음과 같다. .

$$\mathbb{P} = \begin{array}{c} \\ V \\ C \\ M \end{array} \begin{array}{ccc} V & C & M \\ \left[ \begin{array}{ccc} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{array} \right] \end{array} \quad (1)$$

예제 4.7.3에서 알 수 있듯이, Markov 체인은 불가약이다. 또한, 다음 식들이 성립한다.

$$\mathbb{P}^{2n-1} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \mathbb{P}^{2n} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad (n = 1, 2, \dots) \quad (2)$$

식 (2)에서 알 수 있듯이, 각 상태는 주기가 2이다.

다음 식들이 성립한다.

$$\pi(0) = \lim_{n \rightarrow \infty} p^{(2n)}(0, 0) = 1 > 0 \quad (3)$$

$$\pi(1) = \lim_{n \rightarrow \infty} p^{(2n)}(1, 1) = \frac{1}{2} > 0 \quad (4)$$

$$\pi(2) = \lim_{n \rightarrow \infty} p^{(2n)}(2, 2) = \frac{1}{2} > 0 \quad (5)$$

명제 4.9.2에서 알 수 있듯이, 이 Markov 과정의 각 상태는 양재귀적이다. 명제 4.9.1에서 알 수 있듯이, 다음 식들이 성립한다.

$$\mu_{0,0} = \frac{2}{\lim_{n \rightarrow \infty} p^{(2n)}(0, 0)} = 2 \quad (6)$$

$$\mu_{1,1} = \frac{2}{\lim_{n \rightarrow \infty} p^{(2n)}(1, 1)} = 4 \quad (7)$$

$$\mu_{2,2} = \frac{2}{\lim_{n \rightarrow \infty} p^{(2n)}(2, 2)} = 4 \quad (8)$$

또한 다음 식들이 성립한다.

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p^{(k)}(i, 0) = \frac{1}{\mu_{0,0}} = \frac{1}{2} \quad (9)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p^{(k)}(i, 1) = \frac{1}{\mu_{1,1}} = \frac{1}{4} \quad (10)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p^{(k)}(i, 2) = \frac{1}{\mu_{2,2}} = \frac{1}{4} \quad (11)$$

■

**명제 4.9.3**

불가약인 Markov 체인의 각 상태는 다음 중 하나에 속한다.

- a) 각 상태는 과도적이든지 또는 영재귀적이다. 이 경우에 각  $(i, j) (\in \mathbb{Z}_{\geq 0}^2)$  에 대해서 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} p^{(n)}(i, j) = 0$$

즉, 불변분포는 존재하지 않는다.

- b) 각 상태는 양재귀적이다. 즉, 각  $(i, j) (\in \mathbb{Z}_{\geq 0}^2)$  에 대해서 다음 식이 성립한다.

$$\pi(j) = \lim_{n \rightarrow \infty} p^{(n)}(i, j) > 0$$

**정의 4.9.2**

불가약이고 비주기적인 Markov 체인  $\{X_t\}$  의 각 상태가 양재귀적이라고 하자. 즉, 각  $(i, j) (\in \mathbb{Z}_{\geq 0}^2)$  에 대해서 다음 식들이 성립한다고 하자.

$$\pi(j) = \lim_{n \rightarrow \infty} p^{(n)}(i, j) > 0$$

이 Markov 체인  $\{X_t\}$  는 에르고딕(ergodic)하다고 한다.

정의 4.9.2의 Markov 체인  $\{X_t\}$  는 안정적(stationary)이다. 즉, 에르고딕한 Markov 체인은 안정적이다. 또한, 극한분포  $\boldsymbol{\pi} \doteq [\pi(0), \pi(1), \dots]$  은 다음 식들을 만족하는 일의적 불변분포이다.

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbb{P}, \quad \boldsymbol{\pi}\mathbf{1} = 1, \quad \pi(j) \geq 0, \quad (j = 0, 1, 2, \dots) \quad (4.9.2)$$

다음 명제는 이 성질의 역에 해당하는 것이다.

**명제 4.9.4**

불가약적이고 비주기적인 Markov 체인에 대해 다음 식들을 만족하는 불변분포  $\boldsymbol{\pi} =$

$[\pi(0), \pi(1), \dots]$ 이 존재한다고 하자.

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbb{P}, \quad \boldsymbol{\pi}\mathbf{1} = 1, \quad \pi(j) \geq 0, \quad (j = 0, 1, 2, \dots)$$

이 Markov 체인은 다음 성질들을 만족한다.

- a) 각 상태는 재귀적이다.
- b) 각  $(i, j)$ 에 대해서, 극한  $\lim_{n \rightarrow \infty} p^{(n)}(i, j) = \pi(j)$ 가 성립한다.
- c) 각  $j$ 에 대해서,  $\pi(j)$ 는 양수이다.
- d) 불변분포  $\boldsymbol{\pi}$ 는 일의적으로 존재한다.

**명제 4.9.5**

불가약적이고 비주기적인 Markov 체인에 대해 다음 식들을 만족하는 불변분포  $\boldsymbol{\pi} = [\pi(0), \pi(1), \dots]$ 이 존재하지 않는다고 하자.

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbb{P}, \quad \boldsymbol{\pi}\mathbf{1} = 1, \quad \pi(j) \geq 0, \quad (j = 0, 1, 2, \dots)$$

이 Markov 체인에서는 각  $(x, y)$ 에 대해 극한  $\lim_{n \rightarrow \infty} p^{(n)}(x, y) = 0$ 가 성립한다. 즉, 각 상태는 과도적이든지 또는 영재귀적이다.

**예제 4.9.2** 상태공간이  $S = \{0, 1, 2, \dots\}$ 인 Markov 체인  $\{X_n\}$ 의 추이확률행렬이 다음과 같다고 하자.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & \dots \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{matrix} & \begin{bmatrix} q_0 & p_0 & 0 & 0 & \dots \\ q_1 & 0 & p_1 & 0 & \dots \\ 0 & q_2 & 0 & p_2 & \dots \\ 0 & 0 & q_3 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{matrix} \quad (1)$$

여기서 각  $j$ 에 대해서 식  $p_j + q_j = 1$ 이 성립함을 상기하라.

만약 각  $j$ 에 대해서  $p_j$ 와  $q_j$ 가 양수들이면, 이 Markov체인은 불가약이고 비주기적이다. 다음 방정식들을 살펴보자.

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbb{P}, \quad \boldsymbol{\pi}\mathbf{1} = 1, \quad \pi(j) \geq 0, (j = 0, 1, 2, \dots) \quad (2)$$

다음 식들이 성립한다.

$$\pi(0) = q_0\pi(0) + q_1\pi(1) \quad (3)$$

$$\pi(j) = p_{j-1}\pi(j-1) + q_{j+1}\pi(j+1), \quad (j = 1, 2, \dots) \quad (4)$$

식 (3)에서 알 수 있듯이, 다음 식이 성립한다.

$$\pi(1) = \frac{p_0}{q_1}\pi(0) \quad (5)$$

식 (4)에서 알 수 있듯이, 다음 식들이 성립한다.

$$q_{j+1}\pi(j+1) - p_j\pi(j) = q_j\pi(j) - p_{j-1}\pi(j-1), \quad (j = 1, 2, \dots) \quad (6)$$

식 (6)에서 알 수 있듯이, 다음 식들이 성립한다.

$$q_{j+1}\pi(j+1) - p_j\pi(j) = q_1\pi(1) - p_0\pi(0) = 0, \quad (j = 1, 2, \dots) \quad (7)$$

여기서 두 번째 등호는 식 (5)에 의해서 성립한다. 따라서, 다음 식들이 성립한다.

$$\pi(j+1) = \frac{p_j}{q_{j+1}}\pi(j), \quad (j = 0, 1, 2, \dots) \quad (8)$$

즉, 다음 식들이 성립한다.

$$\pi(j) = \frac{\prod_{k=0}^{j-1} p_k}{\prod_{k=1}^j q_k} \pi(0), \quad (j = 1, 2, \dots) \quad (9)$$

식 (9)를 식 (2)에 대입하면, 다음 식들이 성립함을 알 수 있다.

$$\sum_{j=0}^{\infty} \pi(j) = \left[ 1 + \sum_{j=1}^{\infty} \frac{\prod_{k=0}^{j-1} p_k}{\prod_{k=1}^j q_k} \right] \pi(0) = 1 \quad (10)$$

즉, 다음 식이 성립한다.

$$\pi(0) = \frac{1}{1 + \sum_{j=1}^{\infty} \frac{\prod_{k=0}^{j-1} p_k}{\prod_{k=1}^j q_k}} \quad (11)$$

만약 각  $j$ 에 대해서  $p_j = p$ 이고  $q_j = q$ 이면, 식 (11)을 다음과 같이 쓸 수 있다.

$$\pi(0) = \frac{1}{1 + \sum_{j=1}^{\infty} \left[ \frac{p}{q} \right]^j} \quad (12)$$

만약 조건  $p < 0.5$ 이면, 식 (12)에서 알 수 있듯이 다음 식이 성립한다.

$$\pi(0) = 1 - \frac{p}{q} \quad (13)$$

식 (9)와 식 (13)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\pi(j) = \left[ 1 - \frac{p}{q} \right] \frac{p^j}{q^j}, \quad (j = 0, 1, 2, \dots) \quad (14)$$

식 (14)에서 알 수 있듯이, 명제 4.9.4의 조건이 만족된다. 따라서, 각 상태는 재귀적이고, 또한 각  $(i, j)$ 에 대해서 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} p^{(n)}(i, j) = \pi(j) > 0 \quad (15)$$

즉, 각 상태는 양재귀적이다.

만약  $p = 0.5$ 이면, 식 (8)에서 알 수 있듯이 다음 식이 성립한다.

$$\pi(j+1) = \pi(j), \quad (j = 0, 1, 2, \dots) \quad (16)$$

즉,  $\pi_1 \neq 1$ 이다. 따라서, 명제 4.9.5에서 알 수 있듯이 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} p^{(n)}(i, j) = 0 \quad (17)$$

명제 4.9.3에서 알 수 있듯이, 불변분포는 존재하지 않는다. 즉, 각 상태는 과도적이거나 영재귀적이다. 이 Markov체인은 앞으로 나아가는 확률보다 뒤로 돌아오는 확률이 크다. 다음 식이 성립함을 증명할 수 있다.

$$\sum_{n=1}^{\infty} p^{(n)}(i, i) = \infty \quad (18)$$

따라서, 명제 4.7.1에서 알 수 있듯이 각 상태는 영재귀적이다.

만약  $p > 0.5$ 이면, 식 (12)에서 알 수 있듯이 다음 식이 성립한다.

$$\pi(0) = 0 \quad (19)$$

식 (9)와 식 (19)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\pi(j) = 0, \quad (j = 0, 1, 2, \dots) \quad (20)$$

식 (20)에서 알 수 있듯이, 명제 4.9.5의 조건이 만족된다. 따라서, 각  $(i, j)$ 에 대해서 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} p^{(n)}(i, j) = \pi(j) = 0 \quad (21)$$

명제 4.9.3에서 알 수 있듯이, 불변분포는 존재하지 않는다. 즉, 각 상태는 과도적이거나 영재귀적이다. 이 Markov체인은 앞으로 나아가는 확률보다 뒤로 돌아오는 확률이 크다. 다음 식이 성립함을 증명할 수 있다.

$$\sum_{n=1}^{\infty} p^{(n)}(i, i) < \infty \quad (22)$$

명제 4.7.1에서 알 수 있듯이, 각 상태는 과도적이다.

식 (18)과 식 (22)를 좀 더 자세하게 증명할 필요가 있으나, 본서의 목적에 비해 너무 수치적인 면에 흐르는 것 같아 증명하지 않았다. 관심있는 독자는 이 식들을 증명해보라. ■

MCMC에서는 기대값의 추정값으로 표본평균을 사용한다. 표본평균이 일치성 (consistency)을 갖기 위한 충분조건은 Markov 체인  $\{X_n\}$ 이 에르고딕한 것이다. Markov 체인이 에르고딕하다는 것을 간단히 설명하면 Markov 체인이 불가역 (irreducible), 즉 하나의 동치류로 이루어졌다는 것이다. 좀 더 명확하게 설명하면, Markov 체인이 비주기적이고 양재귀적이면 에르고딕하다고 한다. 즉, 불가역 Markov 체인의 각 상태가 비주기적이고 양재귀적이면, Markov 체인은 에르고딕하다고 한다. Markov 체인  $\{X_n\}$ 이 에르고딕하면, 다음 성질들이 성립한다. 이에 대한 자세한 내용은 Walters [96]를 참조하라.

**첫째,** 불변분포  $\pi$ 가 존재하고, 임의의 초기분포  $\pi^{(0)}$ 에 대해서 상태확률벡터  $\{\pi^{(n)}\}$ 은 이 불변분포  $\pi$ 에 수렴한다.

**둘째,** 강대수법칙 (strong law of large numbers: SLLN)이 성립한다. 즉, 평활한 (smooth) 함수  $g(\cdot)$ 에 대해서, 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n g(X_t) = E(g(x)) \text{ a.s.} \quad (4.9.3)$$

여기서  $x$ 는 불변분포  $\pi$ 에 따르는 확률변수이다.

**셋째,** 중심극한정리가 성립한다. 즉, 다음 식이 성립한다.

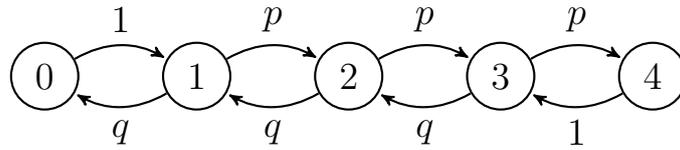
$$\sqrt{n} \left[ \frac{1}{n} \sum_{t=1}^n g(X_t) - E(g(x)) \right] \xrightarrow{d} \mathcal{N}(0, \sigma_g^2) \quad (4.9.4)$$

여기서  $\sigma_g^2$ 는 점근분산이다.

**예제 4.9.3** 상태공간이  $S = \{0, 1, 2, 3, 4\}$ 이고 추이확률행렬이 다음과 같은 확률과정을 살펴보자.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & q & 0 \end{bmatrix} \end{matrix} \quad (1)$$

이 Markov 체인의 상태추이도는 다음과 같다.



이 유한 Markov 체인의 상태공간을 분할하면, 다음과 같다.

$$S = \{0, 1, 2, 3, 4\} = [0] \tag{2}$$

이 유한 Markov 체인에서는 어떤 시점에서 상태 0이나 상태 4에 이르면, 다음 시점에서 다른 상태로 추이한다. 또한, 각 상태  $x$ 의 평균재귀시간  $\mu_{xx}$ 가 유한임은 자명하다. 즉, 상태  $x$ 는 양재귀적이다. 따라서, 이 유한 Markov 체인은 에르고딕 동치류  $S = [0]$ 로 이루어진다. 즉, 이 유한 Markov 체인은 에르고딕이다. ■

지금까지 내용을 정리하면, 상태들을 그림 4.9.1처럼 분류할 수 있다.

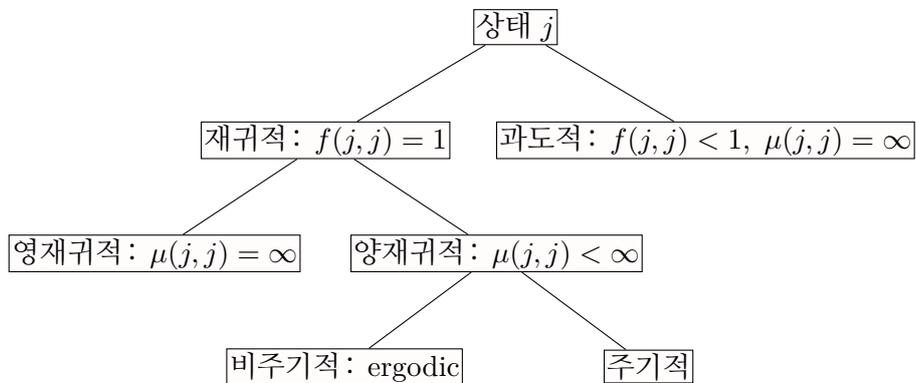


그림 4.9.1. 상태의 분류

### 제 4.10절 시간가역성

불가약이고 양재귀적 Markov 체인  $\{X_t\}$ 의 초기상태 (initial state)가 불변분포에서 선택되면, 이 Markov 체인은 정상적 (stationary)이다. 따라서, 에르고딕한 Markov 체인이 시점  $t = -\infty$ 에서 시작된다고 가정하면, 이 Markov 체인은 정상적이다. 이러한 Markov 체인은 안정상태

(steady state)에 있다고 한다. 지금까지는 Markov 체인에서 시간이 앞으로 흐르는 경우를 살펴보았다. 지금부터는 시간이 역행하는 경우를 살펴보자.

정상적이고 에르고딕한 Markov 체인  $\{X_t\}$ 는 다음 식들을 만족한다.

$$\begin{aligned} & \Pr(X_t | X_{t+1}, X_{t+2}, \dots) \\ &= \frac{\Pr(X_{t+2}, \dots | X_t, X_{t+1}) \Pr(X_t, X_{t+1})}{\Pr(X_{t+2}, \dots | X_{t+1}) \Pr(X_{t+1})} \\ &= \frac{\Pr(X_{t+2}, \dots | X_{t+1}) \Pr(X_t, X_{t+1})}{\Pr(X_{t+2}, \dots | X_{t+1}) \Pr(X_{t+1})} \\ &= \Pr(X_t | X_{t+1}) \end{aligned} \quad (4.10.1)$$

여기서 두 번째 등호는 Markov 체인의 정의에 의해서 성립한다. 식 (4.10.1)에서 알 수 있듯이, 후향확률과정  $\{X_t | t = \dots, n+1, n, n-1, \dots\}$ 도 Markov 체인이다. 이를 역 Markov 체인(reversed Markov chain)이라고 부른다. 식 (4.10.1)을 다음과 같이 설명할 수 있다. 시점  $t$ 에서  $X_t$ 가 주어진 조건 하에서 과거 확률변수들  $X_{t-1}, X_{t-2}, \dots$ 와 미래 확률변수들  $X_{t+1}, X_{t+2}, \dots$ 는 서로 독립이다. 따라서, 후향확률과정  $\{X_t | t = \dots, n+1, n, n-1, \dots\}$ 에서도 시점  $t$ 에서  $X_t$ 가 주어진 조건 하에서 과거 확률변수들  $X_{t+1}, X_{t+2}, \dots$ 와 미래 확률변수들  $X_{t-1}, X_{t-2}, \dots$ 는 서로 독립이다.

추이확률행렬이  $\mathbb{P} = [p(i, j)]$ 이고 불변분포가  $\boldsymbol{\pi} = \{\pi(i)\}$ 인 정상적 Markov 체인  $\{X_t\}$ 에 대해서, 다음 식들이 성립한다.

$$p(j, i) = \Pr(X_{t+1} = i | X_t = j) = \Pr(X_t = j | X_{t+1} = i) \frac{\Pr(X_{t+1} = i)}{\Pr(X_t = j)} \quad (4.10.2)$$

식 (4.10.2)에서 극한  $t \rightarrow \infty$ 를 취하면, 다음 식을 얻는다.

$$\lim_{t \rightarrow \infty} \Pr(X_t = j | X_{t+1} = i) = \frac{\pi(j)}{\pi(i)} p(j, i) \quad (4.10.3)$$

다음과 같은 추이확률을 정의하자.

$$r(i, j) \doteq \frac{\pi(j)}{\pi(i)} p(j, i) \quad (4.10.4)$$

식 (4.10.2)에서 알 수 있듯이,  $[r(i, j)]$ 는 정상적 후향확률과정  $\{X_t | t = \dots, n+1, n, n-1, \dots\}$ 의 추이확률행렬이다. 만약 각  $i \neq j$ 에 대해서 다음 식이 성립하면, 이 Markov 체인은 시간적

으로 가역(time-reversible) 또는 가역적(reversible)이라 한다.

$$r(i, j) = p(i, j) \tag{4.10.5}$$

식 (4.10.4)에서 알 수 있듯이, 식 (4.10.5)는 다음 식과 동치이다.

$$\pi(i)p(i, j) = \pi(j)p(j, i) \tag{4.10.6}$$

식 (4.10.6)을 다음과 같이 쓸 수 있다.

$$\Pr(X_t = i, X_{t+1} = j) = \Pr(X_t = j, X_{t+1} = i) \tag{4.10.7}$$

식 (4.10.7)을 다음과 같이 해석할 수 있다. 상태  $i$ 에서 상태  $j$ 로 가는 비율(rate)  $\pi(i)p(i, j)$ 는 상태  $j$ 에서 상태  $i$ 로 가는 비율  $\pi(j)p(j, i)$ 와 같다. 후향확률과정에서 상태  $i$ 에서 상태  $j$ 로 이동하는 것은 전향확률과정에서 상태  $j$ 에서 상태  $i$ 로 이동하는 것과 동치이다. 즉, 시간을 거슬러가며 보았을 때 상태  $i$ 에서 상태  $j$ 로 이동하는 것은 시간을 따라가며 보았을 때 상태  $j$ 에서 상태  $i$ 로 이동하는 것이다. 따라서, 식 (4.10.6)이 시간가역성의 필요조건임은 명백하다. 시간가역성에 대한 자세한 내용은 Kroese&Taimre&Botev[63, pp. 234-5]를 참조하라.

만약 식 (4.10.6)이 성립하면, 벡터  $\pi\mathbb{P}$ 의 제  $j$ 번째 원소는 다음 식들을 만족한다.

$$\sum_{i=1}^{\infty} \pi(i)p(i, j) = \sum_{i=1}^{\infty} \pi(j)p(j, i) = \pi(j) \sum_{i=1}^{\infty} p(j, i) = \pi(j) \tag{4.10.8}$$

식 (4.10.8)에서 알 수 있듯이, 식 (4.6.6)이 성립한다. 즉, Markov체인이 시간적으로 가역이면, 불변분포가 존재한다.

정상적이고 에르고딕한 Markov체인  $\{X_t\}$ 의 추이확률행렬이  $\mathbb{P} = [p(i, j)]$ 라고 하자. 수열  $\{\phi(0), \phi(1), \phi(2), \dots\}$ 의 합이  $\sum_{i=0}^{\infty} \phi(i) = 1$ 이며, 각  $(i, j)$ 에 대해서 다음 식을 만족한다고 하자.

$$\phi(i)p(i, j) = \phi(j)p(j, i) \tag{4.10.9}$$

다음 식들이 성립한다.

$$\sum_{i=1}^{\infty} \phi(i)p(j, i) = \sum_{i=1}^{\infty} \phi(j)p(j, i) = \phi(j) \sum_{i=1}^{\infty} p(j, i) = \phi(j) \quad (4.10.10)$$

식 (4.6.6)과 식 (4.10.10) 그리고 불변확률의 일의성에서 알 수 있듯이,  $\phi(i)$ 는 불변확률  $\pi(i)$ 이다. 따라서, 식 (4.10.9)는 식 (4.10.6)이 된다. 즉, 이 Markov체인은 시간적으로 가역이다. 이 성질은 매우 중요한 것으로, 이 성질을 이용할 때 유의할 점은 다음과 같다. 먼저 불변확률  $\pi(i)$ 를 구하는 것이 아니라 가역성조건인 식 (4.10.9)를 만족하는  $\{\phi(i)\}$ 를 구하면, 이  $\{\phi(i)\}$ 가 불변확률분포라는 것이다.

**예제 4.10.1** 상태공간이  $S = \{0, 1, 2, \dots, M\}$ 이고 추이확률들이 다음과 같은 확률과정  $\{X_n\}$ 을 살펴보자.

$$p(0, 1) = \alpha_0 = 1 - p(0, 0) \quad (1)$$

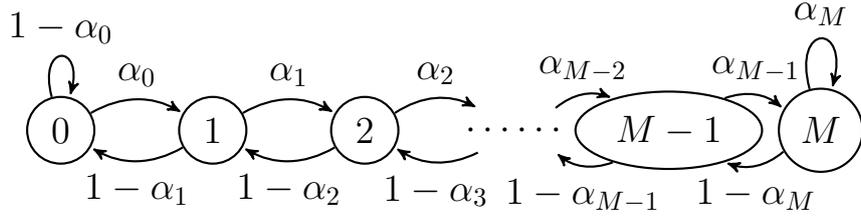
$$p(i, i + 1) = \alpha_i = 1 - p(i, i - 1), \quad (i = 1, 2, \dots, M) \quad (2)$$

$$p(M, M) = \alpha_M = 1 - p(M, M - 1) \quad (3)$$

이 Markov체인  $\{X_n\}$ 의 추이확률행렬은 다음과 같다.

$$\mathbb{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & \cdots & M-2 & M-1 & M \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ \vdots \\ M-2 \\ M-1 \\ M \end{matrix} & \left[ \begin{array}{ccccccc} 1 - \alpha_0 & \alpha_0 & 0 & \cdots & 0 & 0 & 0 \\ 1 - \alpha_1 & 0 & \alpha_1 & \cdots & 0 & 0 & 0 \\ 0 & 1 - \alpha_2 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \alpha_{M-2} & 0 \\ 0 & 0 & 0 & \cdots & 1 - \alpha_{M-1} & 0 & \alpha_{M-1} \\ 0 & 0 & 0 & \cdots & 0 & 1 - \alpha_M & \alpha_M \end{array} \right] \end{matrix} \quad (4)$$

이 Markov체인의 상태추이도는 다음과 같다.



이 상태추이도에서 알 수 있듯이, 다음 식들이 성립한다.

$$p(i, j) = 0 = p(j, i), \quad (|i - j| > 1) \tag{5}$$

다음 방정식들을 생각해 보자.

$$\phi(i)p(i, j) = \phi(j)p(j, i), \quad (i, j = 0, 1, 2, \dots, M) \tag{6}$$

식 (1)~식 (3)을 식 (6)에 대입하면, 다음 식들을 얻는다.

$$\phi(i)\alpha_i = \phi_{i+1}[1 - \alpha_{i+1}], \quad (i = 0, 1, 2, \dots, M - 1) \tag{7}$$

식 (7)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\phi(i) = \phi(0) \prod_{k=1}^i \frac{\alpha_{k-1}}{1 - \alpha_k}, \quad (i = 1, 2, \dots, M) \tag{8}$$

따라서, 조건  $\sum_{l=0}^M \phi(l) = 1$ 을 만족하는  $\phi(0)$ 는 다음과 같다.

$$\phi(0) = \frac{1}{1 + \sum_{l=1}^M \prod_{k=1}^l \frac{\alpha_{k-1}}{1 - \alpha_k}} \tag{9}$$

식 (8)과 식 (9)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\phi(i) = \frac{1}{1 + \sum_{l=1}^M \prod_{k=1}^l \frac{\alpha_{k-1}}{1 - \alpha_k}} \prod_{k=1}^i \frac{\alpha_{k-1}}{1 - \alpha_k}, \quad (i = 1, 2, \dots, M) \tag{10}$$

즉, 식 (10)의  $\phi(0), \phi(1), \dots, \phi(M)$ 이 연립방정식 (6)의 해이다. 따라서, 각  $i(= 0, 1, 2, \dots, M)$ 에 대해서, 확률값  $\pi(i) \doteq \phi(i)$ 를 정의하면  $\{\pi(i)\}$ 는 불변분포이다. 식 (4.10.10)에서 알 수

있듯이,  $\{X_n\}$ 은 시간적으로 가역인 Markov과정이다. ■

**예제 4.10.2** 상태들  $S = \{0, 1, 2, \dots, M\}$ 을 연결한 임의의 그래프를 살펴보자. 꼭지점  $x$ 와 꼭지점  $y$ 를 연결한 무게(weight)를  $w(i, j)$ 라 하자. 다음 식들이 성립함을 자명하다.

$$w(i, j) = w(j, i), \quad (i, j = 0, 1, 2, \dots, M) \quad (1)$$

추이확률이 다음 식을 만족하는 Markov체인  $\{X_n\}$ 을 구축하자.

$$p(i, j) = \frac{w_{ij}}{\sum_{k=0}^M w_{ik}} \quad (2)$$

다음과 같이  $\phi(i)$ 를 정의하자.

$$\phi(i) = \frac{\sum_{k=0}^M w_{ik}}{\sum_{l=0}^M \sum_{k=0}^M w_{lk}} \quad (3)$$

다음 식이 성립함을 자명하다.

$$\sum_{i=0}^M \phi(i) = 1 \quad (4)$$

또한, 다음 식들이 성립함을 쉽게 증명할 수 있다.

$$\phi(i)p(i, j) = \frac{w(i, j)}{\sum_{l=0}^M \sum_{k=0}^M w(l, k)} \quad (5)$$

$$\phi(j)p(j, i) = \frac{w(j, i)}{\sum_{l=0}^M \sum_{k=0}^M w(l, k)} \quad (6)$$

식 (1), 식 (5), 그리고 식 (6)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\phi(i)p(i, j) = \phi(j)p(j, i), \quad (i, j = 0, 1, 2, \dots, M) \quad (7)$$

따라서,  $\{X_n\}$ 은 가역적 Markov과정이고, 또한  $\{\phi(i)\}$ 는 불변분포이다. ■

**예제 4.10.3** 주어진 양수들  $a_1, a_2, \dots, a_M$ 의 합을  $S \doteq \sum_{i=1}^M a_i$ 로 표기하고,  $M$ 이 충분히 커서 합  $S$ 를 직접 계산하기 어려운 경우를 생각해 보자. 다음과 같이 확률을 정의하자.

$$\pi(1), \doteq \frac{a_1}{S}, \pi(2) \doteq \frac{a_2}{S}, \dots, \pi(M) \doteq \frac{a_M}{S} \quad (1)$$

우리의 목적이 이 확률밀도함수로부터 난수들을 생성하는 것이라고 하자.

상태공간이  $S = \{1, 2, \dots, M\}$ 인 불가역적 추이확률행렬  $\mathbb{Q} = [q(i, j)]$ 가 대칭이라고 하자. 즉, 다음 식들이 성립한다고 가정하자.

$$q(i, j) = q(j, i), \quad (i, j = 1, 2, \dots, M) \quad (2)$$

이 대칭 추이확률밀도함수를 이용해서, 다음과 같이 Markov 체인  $\{X_t \mid t = 0, 1, 2, \dots\}$ 를 정의하자.

(1단계) 적절한 초기값  $X_0$ 를 선택한다.

(2단계) 일양확률분포  $U(0, 1)$ 에서 일양난수  $u_t$ 를 생성한다.

(3단계) 만약  $X_t = i$ 이면, 추이확률행렬  $\mathbb{Q}$ 를 사용해서 후보값  $y (y \in \{1, 2, \dots, M\})$ 를 생성한다.

(4단계) 다음과 같이  $X_{t+1}$ 을 선택한다.

$$X_{t+1} = \begin{cases} y, & \left(u_t \leq \min \left\{1, \frac{a_y}{a_i}\right\}\right) \\ X_t, & (\text{otherwise}) \end{cases} \quad (3)$$

(5단계) 만약 원하는 개수만큼 실현값이 생성되면, 이 알고리즘을 멈춘다. 그렇지 않으면,  $t$ 를  $t + 1$ 으로 바꾼 다음 제2단계로 돌아간다.

이렇게 정의된 Markov 체인  $\{X_t \mid t = 0, 1, 2, \dots\}$ 의 추이확률행렬  $\mathbb{P} = [p(i, j)]$ 는 다음 식을 만족한다.

$$p(i, j) = \begin{cases} q(i, j) \min \left\{1, \frac{a_j}{a_i}\right\}, & (j \neq i) \\ q(i, i) + \sum_{j \neq i} \left[1 - \min \left\{1, \frac{a_j}{a_i}\right\}\right], & (j = i) \end{cases} \quad (4)$$

지금부터 확률밀도함수  $\{\pi(1), \pi(2), \dots, \pi(M)\}$ 이 Markov체인  $\{X_t\}$ 의 불변분포임을 증명하기로 하자. 다음 식들이 성립함을 자명하다.

$$\pi(i) \min \left\{ 1, \frac{\pi(j)}{\pi(i)} \right\} = \pi(j) \min \left\{ 1, \frac{\pi(i)}{\pi(j)} \right\}, \quad (i, j = 1, 2, \dots, M) \quad (5)$$

식 (1)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\frac{a_j}{a_i} = \frac{\pi(j)}{\pi(i)}, \quad (i, j = 1, 2, \dots, M) \quad (6)$$

식 (2), 식 (5)와 식 (6)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\pi(i)q(i, j) \min \left\{ 1, \frac{a_j}{a_i} \right\} = \pi(j)q(j, i) \min \left\{ 1, \frac{a_i}{a_j} \right\}, \quad (i \neq j) \quad (7)$$

식 (4)와 식 (7)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\pi(i)p(i, j) = \pi(j)p(j, i), \quad (i, j = 1, 2, \dots, M) \quad (8)$$

식 (8)에서 알 수 있듯이,  $\{X_t\}$ 는 시간적으로 가역인 Markov과정이다.

추이확률행렬  $\mathbb{Q}$ 가 불가약(irreducible)이므로, Markov체인  $\{X_t\}$ 도 불가약이다. 만약 식  $a_1 = a_2 = \dots = a_M$ 이 성립하지 않는다면, 식  $p(i, i) > 0$ 를 만족하는 상태  $i$ 가 반드시 존재한다. 즉,  $\{X_t\}$ 는 비주기적이다. 식 (4.10.9)에서 알 수 있듯이,  $\{X_t\}$ 는 시간적으로 가역인 Markov과정이고  $\{\pi(i)\}$ 는 불변분포이다.

우리는 확률변수열의 실현값들을 생성하기 쉬운 추이확률행렬  $\mathbb{Q}$ 에서 실현값을 생성한 다음, 앞에서 기술한 알고리즘을 적용해서 불변분포  $\{a_i/S\}$ 에서 실현값들을 생성할 수 있다. 이 과정에 합  $S = \sum_{i=1}^M a_i$ 를 계산할 필요가 없다는 점을 상기하라. ■

**예제 4.10.4** 만약 예제 4.10.3의 가정 (2)가 성립하지 않으면, 예제 4.10.3의 식 (3)을 다음과 같이 변형해야한다.

$$X_{t+1} = \begin{cases} y, & \left( u_t \leq \min \left\{ 1, \frac{a_y q(y,x)}{a_x q(x,y)} \right\} \right) \\ X_t, & (\text{otherwise}) \end{cases} \quad (1)$$

이렇게  $\{X_t\}$ 를 생성하는 알고리즘을 Metropolis-Hastings 샘플러라 부른다.

만약  $q(x, y)$ 가  $x$ 에 의존하지 않으면,  $q(x, y)$ 를  $q(y)$ 로 쓸 수 있다. 이 경우에는 식 (3)을 다음과 같이 쓸 수 있다.

$$X_{t+1} = \begin{cases} y, & \left( u_t \leq \min \left\{ 1, \frac{a_y q(x)}{a_x q(y)} \right\} \right) \\ X_t, & \text{(otherwise)} \end{cases} \quad (2)$$

이렇게  $\{X_t\}$ 를 생성하는 알고리즘을 독립샘플러라 부른다. 식 (1)과 식 (2)에 대한 자세한 내용은 제6장을 참조하라.

이 예제에서는 독립샘플러를 사용해서 다음과 같은 감마확률밀도함수에서 난수들을 생성하기로 하자.

$$\pi(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} \quad (3)$$

여기서 형태모수 (scale parameter)는  $\alpha = 2$ 이고 비율 (rate)은  $\beta = 3$ 이다. 즉, 척도모수 (scale parameter)는  $1/\beta = 1/3$ 이다. 이 감마확률분포의 평균과 분산은 각각 다음과 같다.

$$E(x) = \frac{2}{3}, \quad Var(x) = \frac{2}{3^2} = \frac{2}{9} \quad (4)$$

식 (3)의 확률밀도함수에 해당하는 확률분포를 목적분포 (target distribution)라 부른다. 제안분포 (proposal distribution 또는 candidate distribution 또는 instrumental distribution)로는 정규확률분포  $\mathcal{N}(\frac{2}{3}, \frac{2}{9})$ 를 사용하기로 하자. 즉,  $a_x$ 와  $q(x)$ 는 각각 다음과 같다.

$$a_x = \pi(x) = 9xe^{-3x}, \quad (x > 0) \quad (5)$$

$$q(x) = n \left( x \mid \frac{2}{3}, \frac{2}{9} \right) \quad (6)$$

이 감마난수들을 생성하기 위해서, 다음 R프로그램 MCMCexample001.R을 실행하라.

```

1 # -----
2 # Filename: MCMCexample001.R
3 # Metropolis-Hastings Independence Sampler
4 # Programmed by CBS
5 # -----
6 # Function for Independence Sampler
7 set.seed(1)
8 gammaa <- function(alpha, betaa, n)
9 {
10     mu <- alpha/betaa
11     sigma <- sqrt(alpha/betaa^2)
12     MH <- vector("numeric", n)

```

```

12     x <- alpha/betaa
13     MH[1] <- x
14     for (ii in 2:n) {
15         u <- runif(1)
16         y <- rnorm(1, mu, sigma) # candidate
17         ratioNum <- dgamma(y,alpha,betaa)*dnorm(x,mu,sigma)
18         ratioDen <- dgamma(x,alpha,betaa)*dnorm(y, mu, sigma)
19         alphaPr <- min(1,ratioNum/ratioDen) # Acceptance Prob
20         if (u < alphaPr)
21             x <- y
22         MH[ii] <- x
23     }
24     MH
25 }
26
27 # Generating Samples from Gamma distribution
28 NoSeries <- 1024
29 burnIn <- 5000
30 MHseriesExt <- gammaa(2,3,burnIn+NoSeries)
31 MHseries <- MHseriesExt[burnIn+(1:NoSeries)]
32
33 # Lag of MHseries
34 MHseries2 <- MHseries[2:NoSeries]
35 MHseriesLag <- c(NA, head(MHseries, -1))
36 MHseriesLag2 <- MHseriesLag[2:NoSeries]
37 MHmat <- data.frame(MHseries2,MHseriesLag2)
38
39 # Plotting
40 library(ggplot2)
41 # install.packages("grid")
42 library(grid)
43 setEPS()
44 plot.new()
45 postscript('MCMCexample001.eps') # Start to save figure
46 timee = c(1:NoSeries)
47 MHdata <- data.frame(timee,MHseries)
48 plot1 <- ggplot(MHdata, aes(timee,MHseries)) +
49     geom_line(colour="dark red") + xlab("Time") + ylab("MH series")
50 plot2 <- ggplot(MHdata, aes(x=MHseries)) +
51     geom_histogram(aes(y =..density..),
52         breaks=seq(0.0,2.5, by=0.05),
53         col="red", fill="green") +
54     geom_density(lwd=1.0)
55 plot3 <- ggplot(MHmat, aes(x=MHseries2, y=MHseriesLag2)) +
56     geom_point(colour="blue",lwd=1.5) + xlab("Lag(MH Series)") +
57     ylab("Lag(MH Series)")
58 pushViewport(viewport(layout = grid.layout(2,2)))
59 print(plot1, vp = viewport(layout.pos.row = 1, layout.pos.col = 1:2))
60 print(plot2, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
61 print(plot3, vp = viewport(layout.pos.row = 2, layout.pos.col = 2))
62 dev.off() # End to save figure
63 # -----

```

이 R 프로그램 MCMCexample001.R의 첫 번째 단계에서는 감마확률분포로부터 난수들을 발생시키는 독립샘플러함수 gammaa를 만든다. 이 함수 gammaa를 작동시켜서 6024개 난수들을 생성하고 이 중에서 첫 5000개를 버린 나머지 1024 난수들을 사용한다. 이 난수들의 표본평균은 0.6425157이고 분산은 0.1914459이다. 따라서,  $\alpha$ 의 적률추정량은 2.15636이고

$\beta$ 의 적률추정량은 3.356121이다. 이 R 프로그램 MCMCexample001.R을 실행하면, 이렇게 생성된 난수들이 그림 4.10.6에 출력된다. 그림 4.10.6의 상단 그래프는 이 난수들의 시계열 산점도 (time series plot)를 그린 것이고, 좌측하단 그래프는 이 난수들의 히스토그램을 그린 것이고, 우측하단 그래프는  $X_t$  대  $X_{t+1}$ 을 그린 산점도이다. 이 히스토그램은 감마확률밀도 함수처럼 보인다. 독자들은 적합도검정을 사용해서 이 추측이 옳은지를 검정해보라. 또한, 이 산점도에서 볼 수 있듯이 대각선에 난수들이 모여있다. 따라서,  $Corr(X_t, X_{t+1})$ 이 0이라고 하기에는 무리가 따른다. 즉, 이 난수들에는 계열상관(serial correlation)이 존재한다. ■

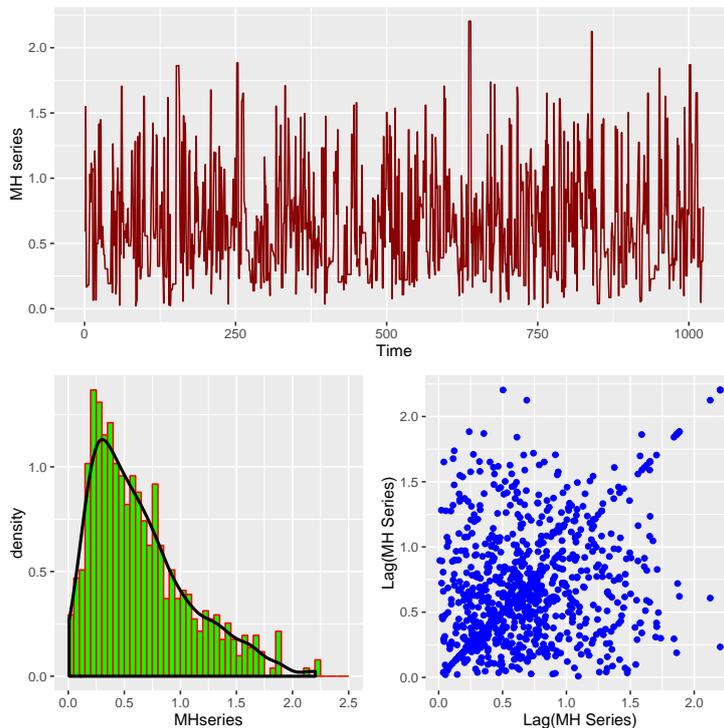


그림 4.10.1. Metropolis-Hastings알고리즘

**예제 4.10.5** 본저자는 프로페셔널프로그래머가 짠 프로그램을 보면서 해당 컴퓨터언어의 최신 기능을 배운다. 단, 프로그램을 보면서 과연 이 방법이 최선인가를 비판하면서 연구한다. 이러한 과정은 독학하는 사람에게 꼭 필요한 과정이라고 생각한다. 또한 너무나도 빨리 변하는 시대에 사는 우리들은 독학을 하지 않으면 해당 분야에 깊숙이 들어갈 수도 없다고 생각한다. 이러한 생각을 바탕으로 다음 웹사이트에서 제시하는 예제를 사용해서 MCMC를 구현하는 과정을 설명하고자 한다.

<https://theclevermachine.wordpress.com/2012/10/20/mcmc-the-metropolis-hastings-sampler/>

MCMC를 사용해서 사후확률분포를 생성하는 예로서 다음 MATLAB 프로그램 MCM-CexampleWeb001.m을 실행하라.

```

1  % -----
2  % This program is from the Web
3  % https://theclevermachine.wordpress.com/2012/10/20
4  %     /mcmc-the-metropolis-hastings-sampler/
5  % METROPOLIS-HASTINGS BAYESIAN POSTERIOR
6  clear all, close all
7  rand('seed',12345)
8  % PRIOR OVER SCALE PARAMETERS
9  B = 1;
10 % DEFINE LIKELIHOOD
11 likelihood = inline('(B.^A/gamma(A)).*y.^(A-1).*exp(-(B.*y))','y','A','B');
12 % CALCULATE AND VISUALIZE THE LIKELIHOOD SURFACE
13 dumNoy = 101; dumNoA = 100;
14 yy = linspace(0,10,dumNoy); % (Original) yy = linspace(0,10,dumNo);
15 AA = linspace(0.1,5,dumNoA); % (Original) AA = linspace(0.1,5,dumNo);
16 likeSurf = zeros(numel(yy),numel(AA)); % numel(A) = numbe of elements of A
17 for iA = 1:numel(AA); likeSurf(:,iA)=likelihood(yy(:),AA(iA),B); end;
18 figure;
19 surf(likeSurf); ylabel('p(y|A)'); xlabel('A'); colormap hot
20 % DISPLAY CONDITIONAL AT A = 2
21 hold on;
22 ly = plot3(ones(1,numel(AA))*40,1:dumNoy,likeSurf(:,40),'g','linewidth',3)
23 % Remind that 5*(40/dumNo) = 2
24 xlim([0 dumNoA]); ylim([0 dumNoy]); axis normal
25 set(gca,'XTick',[0,dumNoA]); set(gca,'XTickLabel',[0 5]);
26 set(gca,'YTick',[0,dumNoy]); set(gca,'YTickLabel',[0 10]);
27 view(65,25)
28 legend(ly,'p(y|A=2)','Location','Northeast');
29 hold off;
30 saveas(gcf,'MCMCexampleWeb001a','eps')
31 title('p(y|A)');
32 % DEFINE PRIOR OVER SHAPE PARAMETERS
33 prior = inline('sin(pi*A).^2','A');
34 % DEFINE THE POSTERIOR
35 p = inline('(B.^A/gamma(A)).*y.^(A-1).*exp(-(B.*y)).*sin(pi*A).^2', ...
36     'y','A','B');
37 % CALCULATE AND DISPLAY THE POSTERIOR SURFACE
38 postSurf = zeros(size(likeSurf));
39 for iA = 1:numel(AA); postSurf(:,iA)=p(yy(:),AA(iA),B); end;
40 figure
41 surf(postSurf); ylabel('y'); xlabel('A'); colormap hot
42 % DISPLAY THE PRIOR
43 hold on;
44 pA = plot3(1:dumNo,ones(1,numel(AA))*dumNoy,prior(AA),'b','linewidth',3)
45 % SAMPLE FROM p(A | y = 1.5)
46 y = 1.5;
47 target = postSurf(16,:);
48 % DISPLAY POSTERIOR
49 psA = plot3(1:dumNo,ones(1,numel(AA))*16,postSurf(16,:),'m','linewidth',3)
50 xlim([0 dumNoA]); ylim([0 dumNoy]); axis normal
51 set(gca,'XTick',[0,dumNoA]); set(gca,'XTickLabel',[0 5]);
52 set(gca,'YTick',[0,dumNo]); set(gca,'YTickLabel',[0 10]);
53 view(65,25)
54 legend([pA,psA],[p(A),'p(A|y = 1.5)'],'Location','Northeast');
55 hold off
56 title('p(A|y)');
57 saveas(gcf,'MCMCexampleWeb001b','eps')

```

```

58 % INITIALIZE THE METROPOLIS-HASTINGS SAMPLER
59 % DEFINE PROPOSAL DENSITY
60 q = inline('exp-pdf(x,mu)','x','mu');
61 % MEAN FOR PROPOSAL DENSITY
62 mu = 5;
63 % DISPLAY TARGET AND PROPOSAL
64 figure; hold on;
65 th = plot(AA,target,'m','Linewidth',2);
66 qh = plot(AA,q(AA,mu),'k','Linewidth',2)
67 legend([th,qh],{'Target, p(A)', 'Proposal, q(A)'});
68 xlabel('A');
69 hold off
70 saveas(gcf,'MCMCexampleWeb001c','eps')
71 % SOME CONSTANTS
72 nSamples = 5000;
73 burnIn = 500;
74 minn = 0.1; maxx = 5;
75 % INITIALIZE SAMPLER
76 x = zeros(1,nSamples);
77 x(1) = mu;
78 t = 1;
79 % RUN METROPOLIS-HASTINGS SAMPLER
80 while t < nSamples
81     t = t+1;
82     % SAMPLE FROM PROPOSAL
83     xStar = exprnd(mu);
84     % CORRECTION FACTOR
85     c = q(x(t-1),mu)/q(xStar,mu);
86     % CALCULATE THE (CORRECTED) ACCEPTANCE RATIO
87     alpha = min([1, p(y,xStar,B)/p(y,x(t-1),B)*c]);
88     % ACCEPT OR REJECT?
89     u = rand;
90     if u < alpha
91         x(t) = xStar;
92     else
93         x(t) = x(t-1);
94     end
95 end
96 % DISPLAY MARKOV CHAIN
97 figure;
98 subplot(211);
99 stairs(x(1:t),1:t, 'k');
100 hold on;
101 hb = plot([0 maxx/2],[burnIn burnIn],'g--','Linewidth',2)
102 ylabel('t'); xlabel('samples, A');
103 set(gca, 'YDir', 'reverse');
104 ylim([0 t])
105 axis tight;
106 xlim([0 maxx]);
107 title('Markov Chain Path');
108 legend(hb,'Burnin');
109 % DISPLAY SAMPLES
110 subplot(212);
111 nBins = dumNoA;
112 sampleBins = linspace(minn,maxx,nBins);
113 counts = hist(x(burnIn:end), sampleBins);
114 bar(sampleBins, counts/sum(counts), 'k');
115 xlabel('samples, A'); ylabel('p(A | y)');
116 title('Samples');
117 xlim([0 10])
118 % OVERLAY TARGET DISTRIBUTION

```

```

119 hold on;
120 plot(AA, target/sum(target), 'm-', 'LineWidth', 2);
121 legend('Sampled Distribution', sprintf('Target Posterior'))
122 axis tight
123 saveas(gcf, 'MCMCexampleWeb001d', 'eps')
124 % End of program
125 % -----

```

이 MATLAB 프로그램 MCMCexampleWeb001.m의 첫 번째 단계는 우도함수를 그리는 것이다. 이 MATLAB 프로그램에서는 관찰값들의 우도함수가 다음과 같은 감마확률분포에서 발생했다고 가정한다.

$$p(y|\boldsymbol{\theta}) = \frac{B^A}{\Gamma(A)} y^{A-1} e^{-By} \quad (1)$$

여기서 모수  $A$ 는 형상모수(shape parameter)이고  $B$ 는 비율(rate)로서 척도모수(scale parameter)의 역수이다. 즉, 척도모수는  $1/B$ 이다. 또한, 모수벡터는  $\boldsymbol{\theta} = [A, B]$ 이다. 이 예제에서는 관찰값과 모수들을 다음과 같이 선택했다.

$$A = \left[ 0.1 + 4.9 \times \frac{0}{99}, 0.1 + 4.9 \times \frac{1}{99}, \dots, 0.1 + 4.9 \times \frac{98}{99}, 0.1 + 4.9 \times \frac{99}{99} \right] \quad (2)$$

$$B = 1 \quad (3)$$

$$y = \left[ 10 \times \frac{0}{99}, 10 \times \frac{1}{99}, \dots, 10 \times \frac{98}{99}, 10 \times \frac{99}{99} \right] \quad (4)$$

우도함수곡면  $\{(y, A, p(y|\boldsymbol{\theta}))\}$ 를 그리기 위해서 MATLAB 명령문 “surf(likeSurf)”를 사용하였다. 또한, 이 곡면 위에  $A = 2$ 인 주변확률밀도함수를 그리기 위해서, 다음 명령문을 사용하였다.

```
ly = plot3(ones(1,numel(AA))*40,1:100,likeSurf(:,40),'g','linewidth',3)
```

그러나, 이 명령문을 실행하면,  $A$ 값은 2가 아니라 다음과 같다.

$$A = 0.1 + 4.9 \times \frac{40-1}{100-1} = 2.011 \quad (5)$$

만약  $A = 2$ 를 지정하려면, 변수 AA를 지정하는 MATLAB 함수 linspace를 정확하게 사용해야 한다. 이 MATLAB 프로그램을 실행하면, 이렇게 주변확률밀도함수가 중첩된 우도함수곡면이 그림 4.10.2에 출력한다.

이 MATLAB 프로그램 MCMCexampleWeb001.m의 두 번째 단계는 사전확률밀도함수와 높이가 조정되지 않은 (unnormalized) 사후확률밀도함수를 그리는 것이다. 이 MATLAB 프로그램에서는 사전확률밀도함수가 다음과 같다고 가정한다.

$$\Pr(B = 1) = 1, \quad \pi_0(A) = \sin^2(\pi A) \tag{6}$$

모수  $B$ 에 관한 사전확률분포는 비정칙 (improper) 이다. 모수  $B$ 가 상수이므로, 우리는 모수  $A$ 의 사후확률분포만 구하면 된다. 이 사후확률분포는 다음 식을 만족한다.

$$\pi_*(A|y) = \frac{p(y|A)\pi_0(A)}{\int p(y|A)\pi_0(A)dA} = \frac{\frac{1}{\Gamma(A)}y^{A-1}e^{-y}\sin^2(\pi A)}{\int \frac{1}{\Gamma(z)}y^{z-1}e^{-y}\sin^2(\pi z)dz} \tag{7}$$

그러나, 식 (7) 우변의 분모를 계산하는 과정이 쉽지 않다. 다음 비례식이 성립한다.

$$\pi_*(A|y) \propto \frac{1}{\Gamma(A)}y^{A-1}e^{-y}\sin^2(\pi A) \tag{8}$$

식 (8)의 우변은 높이가 조정되지 않은 사후확률밀도함수이다. 우리는 높이가 조정되지 않은 사후확률밀도함수를 MCMC에 적용해서 사후확률밀도함수를 구하기로 하자. 높이가 조정되지 않은 사후확률밀도함수  $\frac{1}{\Gamma(A)}y^{A-1}e^{-y}\sin^2(\pi A)$ 를 그리기 위해서 MATLAB 명령문 “surf(postSurf)”를 사용하였다. 이 그래프에 사전확률밀도함수를 추가하기 위해서 다음 명령문을 사용하였다.

```
pA = plot3(1:100,ones(1,numel(AA))*100,prior(AA),'b','linewidth',3)
```

이 그래프에  $y = 1.5$ 인 사후확률밀도함수를 추가하기 위해서 다음 명령문을 사용하였다.

```
psA = plot3(1:100, ones(1,numel(AA))*16,postSurf(16,:), 'm','linewidth',3)
```

그러나, 이 명령문을 실행하면,  $y$ 값은 1.5가 아니라 다음과 같다.

$$y = 0 + 10 * \times \frac{16 - 1}{100 - 1} = \frac{50}{33} = 1.51515 \tag{9}$$

이 MATLAB 프로그램을 실행하면, 이렇게 사전확률밀도함수와  $y = 50/33$ 인 사후확률밀도함수가 중첩된 높이가 조정되지 않은 사후확률밀도함수곡면을 그림 4.10.3에 출력한다.

이 MATLAB 프로그램 MCMCexampleWeb001.m의 세 번째 단계에서는 제안분포  $q(A)$ 를 지정한다. 사후확률분포의 지지대가  $(0, \infty)$ 이므로 정규확률분포와 마찬가지로 대칭이

아니고  $A > 0$ 에서 정의되는 비대칭확률분포를 제안분포로 사용해야 한다. 여기서는 확률밀도함수가 다음과 같은 지수확률분포를 제안분포로 사용한다.

$$q(A) = 5e^{-5A}, \quad (A > 0) \quad (10)$$

이 MATLAB 프로그램을 실행하면, 높이가 조정되지 않은 사후확률밀도함수  $\pi_*(A|y = 1.5) = p(A|y = 1.5)$ 와 제안분포  $q(A)$ 가 그려진 그림 4.10.4를 출력한다.

이 MATLAB 프로그램 MCMCexampleWeb001.m의 네 번째 단계는 Metropolis-Hastings 알고리즘을 사용해서 사전확률밀도함수와 높이가 조정된(normalized) 사후확률밀도함수  $\pi_*(A|y = 1.5)$ 를 구하는 것이다. 우선 Metropolis-Hastings 알고리즘을 사용해서 Markov 체인  $\{X_t\}$ 를 생성한다. 그 과정은 “while-end” 루프 사이에 있는 MATLAB 명령문들에 의해 실행된다. 이 MATLAB 프로그램을 실행하면, 이렇게 생성된 Markov 체인이 그림 4.10.5에 출력된다. 그림 4.10.5의 상단 그래프는 이 Markov 체인의 시계열산점도(time series plot)를 그린 것이고, 하단 그래프는 생성된 Markov 체인의 히스토그램과 높이가 조정된 사후확률밀도함수를 그린 것이다.

다음 Python 프로그램 MCMCexampleWeb002.py를 실행하면 MATLAB 프로그램 MCMCexampleWeb001.m을 실행한 것과 동일한 결과를 얻을 수 있다.

```

1 # MCMCexampleWeb002.py
2 """
3 Filename : GibbsMultiIntegration101.py
4 This program is from the Web
5     https://theclevermachine.wordpress.com/page/2/
6 METROPOLIS-HASTINGS BAYESIAN POSTERIOR
7 Programmed by Kyung-Bae Park based on the R program by CBS
8 """
9
10 import numpy as np
11 import math
12 import matplotlib.pyplot as plt
13 from matplotlib import cm
14 from mpl_toolkits.mplot3d import Axes3D
15 from scipy.stats import expon
16
17 np.random.seed(12345)
18
19 # PRIOR OVER SCALE PARAMETERS
20 B = 1
21
22 # DEFINE LIKELIHOOD
23 def likelihood(y,A,B):
24     return ((B**A/math.gamma(A)) * (y**(A-1)) * np.exp((-B*y)))
25
26
27 # CALCULATE AND VISUALIZE THE LIKELIHOOD SURFACE

```

```

28 yy = np.linspace(0,10,100)
29 AA = np.linspace(0.1,5,100)
30 likeSurf = np.zeros((len(yy), len(AA)))
31 for iA in range(len(AA)):
32     likeSurf[:,iA] = likelihood(yy,AA[iA],B)
33
34
35 # DISPLAY CONDITIONAL AT A = 2
36 fig_a = plt.figure()
37 ax_a = plt.axes(projection='3d')
38
39 X_a, Y_a = np.meshgrid(AA, yy)
40
41 ax_a.plot_surface(X_a, Y_a, likeSurf, rstride=1, cstride=1,
42                 cmap=cm.OrRd_r, linewidth=0)
43
44 ax_a.view_init(azim=330)
45 ax_a = fig_a.gca(projection='3d')
46
47 ax_a.plot(np.linspace(1,1,len(AA))*AA[39], yy, likeSurf[:,39], color = 'green')
48
49 plt.xlabel('A', fontsize = 10)
50 plt.ylabel('y', fontsize = 10)
51 plt.title('p(y|A)')
52 plt.legend(['p(y|A=2)'])
53 plt.show()
54 fig_a.savefig('MCMCexampleWeb002a.eps', format='eps')
55
56
57 # DEFINE PRIOR OVER SHAPE PARAMETERS
58 def prior(A):
59     return(np.sin(math.pi*A)**2)
60
61
62 # DEFINE THE POSTERIOR
63 def p(y,A,B):
64     return((B**A/math.gamma(A)) * (y**(A-1))
65           * np.exp((-B*y)) * math.sin(math.pi*A)**2)
66
67
68 # CALCULATE AND DISPLAY THE POSTERIOR SURFACE
69 postSurf = np.zeros((likeSurf.shape))
70 for iA in range(len(AA)):
71     postSurf[:,iA] = p(yy,AA[iA],B)
72
73
74 fig_b = plt.figure()
75 ax_b = plt.axes(projection='3d')
76
77 X_b, Y_b = np.meshgrid(AA, yy)
78
79 ax_b.plot_surface(X_b, Y_b, postSurf, rstride=1, cstride=1,
80                 cmap=cm.OrRd_r, linewidth=0)
81 ax_b.view_init(azim=330)
82
83 ax_b.plot(AA, np.linspace(1,1,len(yy))*yy[99], prior(AA), color = 'blue')
84
85 # SAMPLE FROM p(A | y = 1.5)
86 y = 1.5
87 target = postSurf[15,:]
88

```

```

89 # DISPLAY POSTERIOR
90 ax_b.plot(AA, np.linspace(1,1,len(yy))*yy[15],postSurf[15,:], color = 'yellow')
91
92 plt.xlabel('A', fontsize = 10)
93 plt.ylabel('y', fontsize = 10)
94 plt.title('p(A|y)')
95
96 plt.legend(['p(A)', 'p(A|y = 1.5)'])
97 plt.show()
98 fig_b.savefig('MCMCexampleWeb002b.eps', format='eps')
99
100
101 # INITIALIZE THE METROPOLIS-HASTINGS SAMPLER
102 # DEFINE PROPOSAL DENSITY
103 def q(x,mu):
104     return(expon.pdf(x,scale = mu))
105
106
107 # MEAN FOR PROPOSAL DENSITY
108 mu = 5
109
110
111 # DISPLAY TARGET AND PROPOSAL
112 fig_c = plt.figure()
113 plt.plot(AA, target, color = 'purple', linewidth = 2)
114 plt.plot(AA, q(AA,mu), color = 'black', linewidth = 2)
115 plt.legend(['Target, p(A)', 'Proposal, q(A)'])
116 plt.xlabel('A', fontsize = 10)
117 plt.show()
118 fig_c.savefig('MCMCexampleWeb002c.eps', format='eps')
119
120
121
122
123
124 # SOME CONSTANTS
125 nSamples = 5000
126 burnIn = 500
127 minn = 0.1
128 maxx = 5
129
130
131 # INTIIALZE SAMPLER
132 x = np.zeros((1, nSamples))
133 x[0,0] = mu
134 t = 1
135
136
137 # RUN METROPOLIS-HASTINGS SAMPLER
138 while t < nSamples:
139     t = t+1
140
141     # SAMPLE FROM PROPOSAL
142     xStar = np.random.exponential(mu)
143
144     # CORRECTION FACTOR
145     c = q(x[0,t-2],mu)/q(xStar,mu)
146
147     # CALCULATE THE (CORRECTED) ACCEPTANCE RATIO
148     alpha = min(1, p(y,xStar,B)/p(y,x[0,t-2],B)*c)
149

```

```

150     # ACCEPT OR REJECT?
151     u = np.random.uniform()
152
153     if u < alpha:
154         x[0,t-1] = xStar
155     else:
156         x[0,t-1] = x[0,t-2]
157
158
159 # DISPLAY MARKOV CHAIN
160 fig_d = plt.figure()
161 plt.subplot(211)
162 plt.step(x[0,0:t],np.linspace(1,t,t),color = 'black')
163 plt.plot(np.linspace(0,maxx/2,2), np.linspace(burnIn,burnIn,2),
164         color = 'green', ls = '--', linewidth = 2)
165 plt.xlabel('samples, A', fontsize = 10)
166 plt.ylabel('t', fontsize = 10)
167 plt.title('Markov Chain Path')
168 plt.xlim(0,maxx)
169 plt.ylim(0,t)
170 plt.legend(['time series plot', 'Burnin'])
171
172 # DISPLAY SAMPLES
173 plt.subplot(212)
174 nBins = 100
175 sampleBins = np.linspace(minn,maxx,nBins)
176 counts = np.asarray(np.histogram(x[0,(burnIn-1):], bins = sampleBins))[0]
177 plt.bar(np.linspace(minn + (maxx-minn)/(nBins-1),
178                 maxx - (maxx-minn)/(nBins-1),nBins-1),counts/np.sum(counts),
179         width=0.03, color = 'black')
180 plt.xlabel('samples, A', fontsize = 10)
181 plt.ylabel('p(A | y)', fontsize = 10)
182 plt.title('Samples')
183 plt.xlim(0,maxx)
184
185 plt.plot(AA, target/sum(target), color = 'purple', linewidth = 2, )
186 plt.legend(['Target Posterior', 'Sampled Distribution'])
187 plt.subplots_adjust(hspace=1)
188
189 plt.show()
190 fig_d.savefig('MCMCexampleWeb002d.eps', format='eps')
191
192
193 # End of program
194 # -----

```



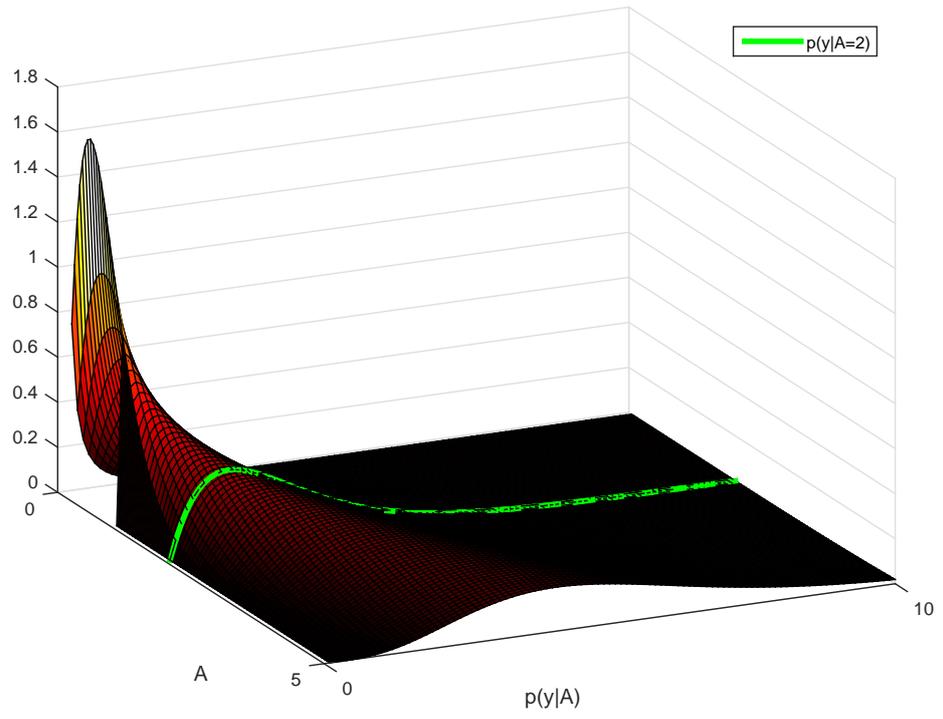


그림 4.10.2. 우도함수

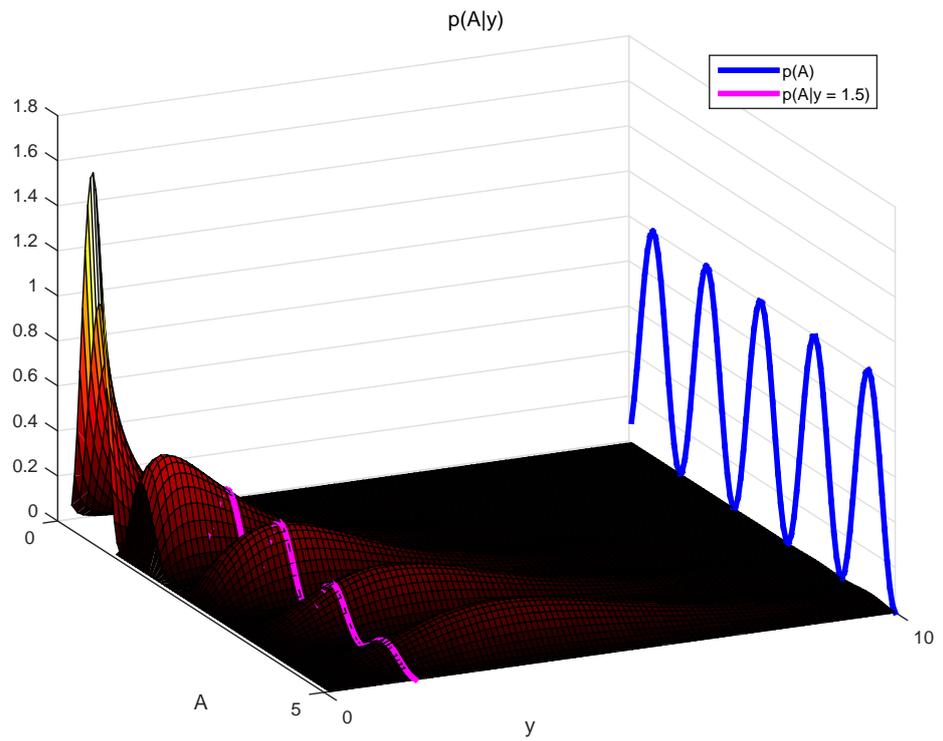


그림 4.10.3. 사전확률밀도함수

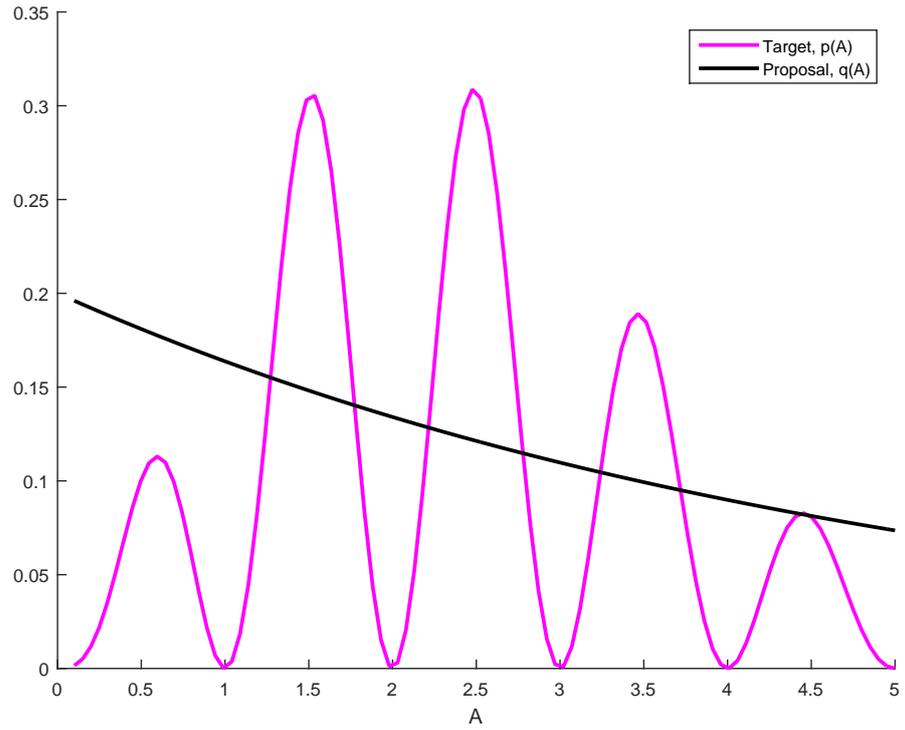


그림 4.10.4. 비정규화 사후확률밀도함수

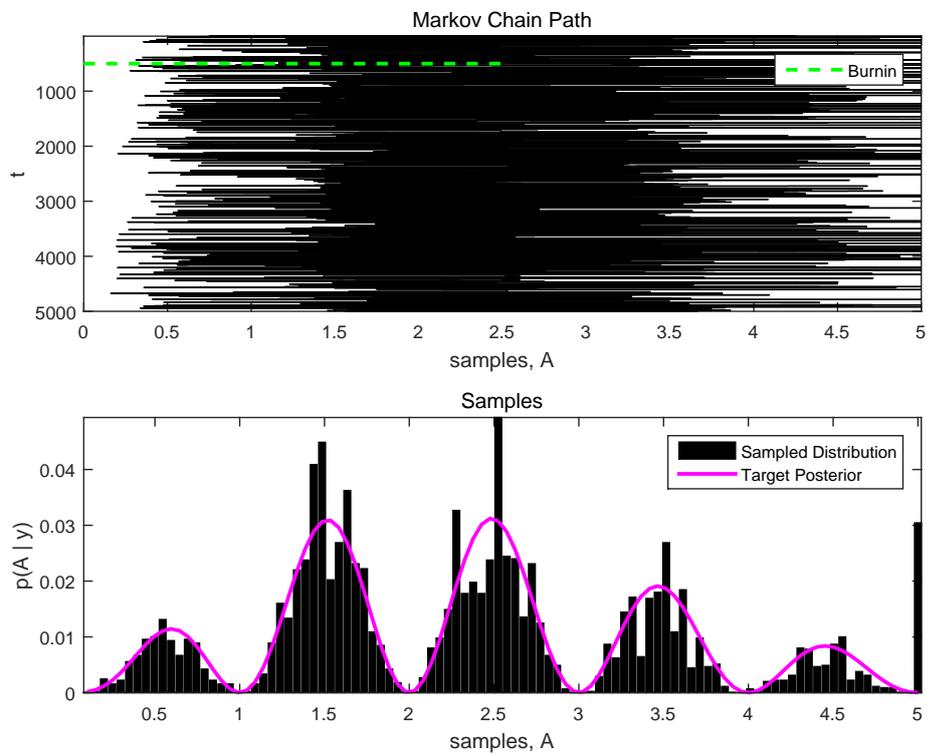


그림 4.10.5. 사후확률밀도함수

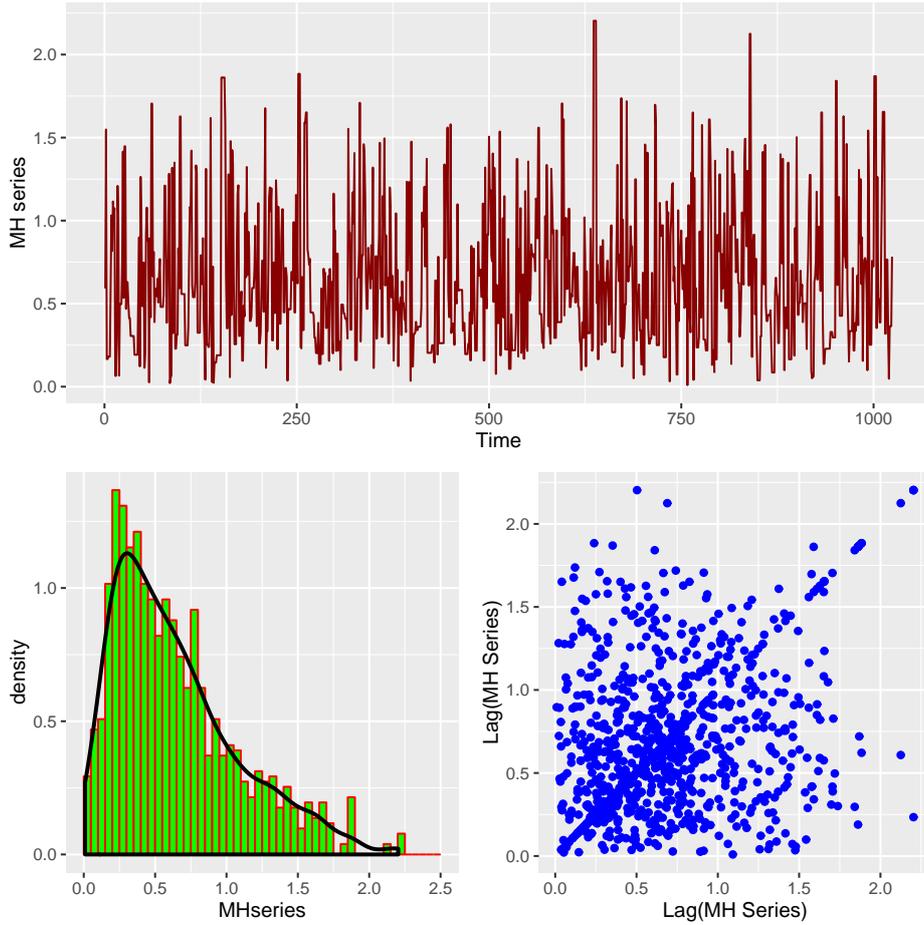


그림 4.10.6. Metropolis-Hastings 알고리즘

### 제 4.11 절 연속형 상태공간

상태공간이 연속인 경우를 살펴보자. 확률변수열  $\{X_n\}$ 이 각 가측집합  $A$ 에 대해서 다음 식을 만족하는 추이확률분포  $\mathbb{P}$ 를 가지면,  $\{X_n\}$ 은 Markov 체인이다.

$$\mathbb{P}(X_{n+1} \in A | X_0, X_1, \dots, X_n) = \mathbb{P}(X_{n+1} \in A | X_n) \tag{4.11.1}$$

초기상태  $X_0$ 가 주어진 조건 하에서 확률변수  $X_n$ 의 조건부확률분포를 다음과 같이 표기하자.

$$\mathbb{P}(X_n \in A | X_0) = \mathbb{P}^n(X_0, A) \tag{4.11.2}$$

여기서  $\mathbb{P}^n$ 은 추이확률분포  $\mathbb{P}$ 를  $n$ 번 반복적용한 것을 의미한다.

**정의 4.11.1**

집합  $S$ 와 Borel집합족  $\mathcal{B}(S)$ 에 대해  $S \times \mathcal{B}(S)$  상에서 정의되는 함수  $K$ 가 다음 조건들을 만족하면,  $K$ 를 추이핵(transition kernel)이라 부른다.

- a) 임의의  $x(\in S)$ 에 대해서  $K(x, \cdot)$ 는 확률측도이다.
- b) 임의의  $A(\in \mathcal{B}(S))$ 에 대해서  $K(\cdot, A)$ 는 측도이다.

만약  $S$ 가 이산집합이면, 추이핵  $K = [k_{ij}]$ 는 다음과 같은 원소를 갖는 추이확률행렬이다.

$$k_{ij} = \Pr(x^{(n)} = j \mid x^{(n)} = i), \quad (i, j \in S) \tag{4.11.3}$$

따라서, 식 (4.11.1)의 추이확률분포  $\mathbb{P}$ 를 추이핵(transition kernel)이라 부르기도 한다. 만약  $S$ 가 연속집합이면, 추이핵은 다음과 같이  $K(y, \cdot)$ 의 조건부확률밀도함수를 나타낸다.

$$\Pr(x \in S \mid y) = \int_A K(y, z) dz \tag{4.11.4}$$

각 가측집합  $A$ 에 대해서 다음 식을 만족하는  $\pi(\cdot)$ 를 이 Markov체인의 불변분포라 한다.

$$\pi(A) = \int \pi(dx) \mathbb{P}(x, A) \tag{4.11.5}$$

불가약인 Markov체인이 불변분포  $\pi(\cdot)$ 를 가지면,  $\pi(\cdot)$ 는 일의적인 불변분포이다. 더구나, 만약 이 Markov체인이 비주기적이면, 거의 모든  $x$ 에 대해서 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} \|\mathbb{P}^n(x, \cdot) - \pi(\cdot)\| = 0 \tag{4.11.6}$$

여기서  $\|\cdot\|$ 는 총변분거리(total variation distance)이다. 따라서, 목표분포  $\pi$ 를 불변분포로 갖는 Markov체인을 구성할 수 있다면, 반복횟수  $n$ 을 충분히 크게 함으로써, 목표분포에서 샘플들을 추출할 수 있다. 또한, 에르고딕성에 의해서, 목표분포에 관한 함수  $h$ 의 기대값  $E(h(x))$ 를 다음과 같이 추정할 수 있다.

$$E(h(x)) \approx \frac{1}{n} \sum_{t=b+1}^{b+n} h(X_t) \tag{4.11.7}$$

식 (4.11.7)에서  $b$ 는 번인기간(burn-in period), 즉 번인구간에 포함된 샘플수이다.

요구되는 정확성을 만족시키기 위해서, 샘플수  $n$ 을 얼마로 해야할까? 정의에서 알 수 있듯이, 샘플들  $h(X_{b+1}), h(X_{b+2}), \dots, h(X_{b+n})$ 은 독립이 아니다. 따라서, 추정량  $\frac{1}{n} \sum_{t=b+1}^{b+n} h(X_t)$ 의 분산을 추정하는 것은 쉽지 않다. 즉, 추정량의 분산을 바탕으로  $n$ 을 결정하는 것은 쉽지 않다. 시뮬레이션을 통해서  $n$ 을 결정하는 방법들 중 하나는 초기값  $X_0$ 를 달리하는 추정값들  $\frac{1}{n} \sum_{t=b+1}^{b+n} h(X_t)$ 을 계산한 다음, 이 추정값들 사이에 변동이 너무 크면  $n$ 을 증가시킨다.

## 제 5 장

# Gibbs 샘플러

### 제 5.1 절 Gibbs 샘플러의 정의

Gibbs 샘플러는 Geman & Geman [43]의 1984년 논문을 기점으로 통계학의 응용부분에 사용되기 시작했다. 이들은 Gibbs 샘플러를 이미지프로세싱과 격자모형에서 정의되는 Gibbs 분포를 분석하는데 사용하였다. 그러나, Gibbs 샘플러의 아이디어는 1970년대까지 거슬러 올라가서, 통계물리학의 열목욕알고리즘(heat bath algorithm)이나 공간통계학 등에서 독립적으로 사용되었다. Geman & Geman [43]은 상태공간이 이산적 확률분포인 Gibbs 분포를 추정하는데 사용한 반면에, Gelfand & Smith [42]는 상태공간이 연속적 확률분포를 추정하는데 사용하였다.

확률벡터를 구성하는 확률변수들의 조건부확률분포들로부터 샘플들을 축차적으로 생성하면, 적절한 조건 하에서 이들의 결합확률분포가 원래의 결합확률분포로 수렴한다. Gibbs 샘플러는 이 성질을 이용해서 결합확률분포로부터 샘플을 생성하는 기법이다. 즉, Gibbs 샘플러는 확률벡터의 결합확률분포로부터 직접 샘플을 생성하기 어려운 경우, 조건부확률분포들을 이용해서 결합확률분포로부터 샘플을 생성하는 기법이다. 확률벡터  $\mathbf{x} = [x_1, x_2, \dots, x_p]^t$ 의 결합확률밀도함수를  $f(\mathbf{x}) = f(x_1, x_2, \dots, x_p)$ 라 하자. 각  $i$ 에 대해서 확률변수  $x_i$ 를 제외한 나머지 확률변수들로 이루어진 확률벡터를  $\mathbf{x}_{-i}$ 라 하자. 즉,  $\mathbf{x}_{-i} \doteq [x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p]^t$ 라 하자. 결합확률밀도함수  $f(\mathbf{x})$ 에서 난수를 추출하는 것은 어렵지만, 조건부확률밀도함수  $f(x_i | \mathbf{x}_{-i})$ 에서 난수를 추출하는 것은 쉽다고 하자. 이러한 경우, 다음과 같은 Gibbs 샘플러를 사용한다.

**알고리즘 5.1.1: 체계적쓸기Gibbs샘플러**

(1단계) 초기벡터를  $\mathbf{x}^{(0)} = [x_1^{(0)}, x_2^{(0)}, \dots, x_p^{(0)}]^t$  라고 하자.

(2단계) 제  $j$  단계 관찰벡터  $\mathbf{x}^{(j)} = [x_1^{(j)}, x_2^{(j)}, \dots, x_p^{(j)}]^t$  가 주어진 경우, 제  $j + 1$  단계 관찰벡터  $\mathbf{x}^{(j+1)} = [x_1^{(j+1)}, x_2^{(j+1)}, \dots, x_p^{(j+1)}]^t$  의 각 성분을 다음과 같이 추출한다.

$$\begin{aligned} x_1^{(j+1)} &\stackrel{d}{\sim} f\left(x_1 \mid x_2^{(j)}, x_3^{(j)}, \dots, x_{p-1}^{(j)}, x_p^{(j)}\right) \\ x_2^{(j+1)} &\stackrel{d}{\sim} f\left(x_2 \mid x_1^{(j+1)}, x_3^{(j)}, \dots, x_{p-1}^{(j)}, x_p^{(j)}\right) \\ &\vdots \\ x_{p-1}^{(j+1)} &\stackrel{d}{\sim} f\left(x_{p-1} \mid x_1^{(j+1)}, x_2^{(j+1)}, \dots, x_{p-2}^{(j+1)}, x_p^{(j)}\right) \\ x_p^{(j+1)} &\stackrel{d}{\sim} f\left(x_p \mid x_1^{(j+1)}, x_2^{(j+1)}, \dots, x_{p-2}^{(j+1)}, x_{p-1}^{(j+1)}\right) \end{aligned}$$

(3단계) 만약  $j + 1$  이 사전에 정해진 개수이면, 그 결과를 출력하고 알고리즘 실행을 끝낸다. 그렇지 않으면  $j$  대신  $j + 1$  을 할당한 다음, 제2단계로 넘어간다.

Gibbs샘플러에서 사용되는 조건부확률분포들을 충만한(full) 조건부확률분포들이라 부르기도 한다. 알고리즘 5.1.1을 체계적쓸기Gibbs샘플러(systematic sweep Gibbs sampler)라 부른다. 이 체계적쓸기Gibbs샘플러에서 발생된 Markov체인은 시간가역적이(time reversible) 아니다. 그러나, 다음에 소개하는 무작위적쓸기Gibbs샘플러(random sweep Gibbs sampler)에서 발생된 Markov체인은 시간가역적이다. 제4장에서 설명했듯이, 시간가역성은 MCMC에서 매우 중요한 것이다.

**알고리즘 5.1.2: 무작위적쓸기Gibbs샘플러**

(1단계) 초기벡터를  $\mathbf{x}^{(0)} = [x_1^{(0)}, x_2^{(0)}, \dots, x_p^{(0)}]^t$  라고 하자.

(2단계) 제  $j$  단계 관찰벡터  $\mathbf{x}^{(j)} = [x_1^{(j)}, x_2^{(j)}, \dots, x_p^{(j)}]^t$  가 주어진 경우, 우선 지지대가  $\{1, 2, \dots, p\}$  인 이산형 확률분포로부터 난수  $k$  를 선택한다. 다음으로, 제  $j + 1$  단계

관찰벡터  $\mathbf{x}^{(j+1)} = [x_1^{(j+1)}, x_2^{(j+1)}, \dots, x_p^{(j+1)}]^t$ 의 각 성분을 다음과 같이 추출한다.

$$\begin{aligned} x_k^{(j+1)} &\stackrel{d}{\sim} f\left(x_k \mid x_1^{(j)}, \dots, x_{k-1}^{(j)}, x_{k+1}^{(j)}, \dots, x_p^{(j)}\right) \\ x_l^{(j+1)} &= x_l^{(j)}, \quad (l \neq k) \end{aligned}$$

**(3단계)** 만약  $j+1$ 이 사전에 정해진 개수이면, 그 결과를 출력하고 알고리즘 실행을 끝낸다. 그러나, 그렇지 않으면  $j$  대신  $j+1$ 을 대입한 다음, 제2단계로 넘어간다.

체계적쓰기Gibbs샘플러와 무작위적쓰기Gibbs샘플러를 간단히 Gibbs샘플러라 부르기로 하자. Gibbs샘플러를 적용하기 위해서는 먼저 초기벡터  $\mathbf{x}^{(0)}$ 를 적당히 선택한다. 각 초기벡터 별로 Gibbs샘플링을  $b+n$ 회 실행해서  $b+n$ 개 관찰벡터들을 발생시킨다. 여기서  $b$ 는 번인기간(burn-in period)을 의미한다. 확률변수들 사이에 상관성이 크면, 번인기간  $b$ 를 크게 선택할 필요가 있다. Gibbs샘플들  $\left\{ \mathbf{x}^{(j)} = [x_1^{(j)}, x_2^{(j)}, \dots, x_p^{(j)}]^t \mid j = b+1, b+2, \dots, b+n \right\}$ 은 정규조건(regularity conditions) 하에서 다음 성질들을 만족한다. 이에 대한 자세한 내용은 Zellner & Min [100]을 참조하라.

**첫째,** 확률벡터  $\mathbf{x}^{(j)} = [x_1^{(j)}, x_2^{(j)}, \dots, x_p^{(j)}]^t$ 의 결합확률분포는 극한  $j \rightarrow \infty$ 에서 결합확률밀도함수  $f(\mathbf{x}) = f(x_1, x_2, \dots, x_p)$ 에 해당하는 결합확률분포로 분포수렴한다.

**둘째,** 다음 식이 성립한다.

$$f(x_i) = \int f(x_i, \mathbf{x}_{-i}) d\mathbf{x}_{-i} = \int f(x \mid \mathbf{x}_{-i}) f(\mathbf{x}_{-i}) d\mathbf{x}_{-i} \quad (5.1.1)$$

따라서, 확률변수  $x_i$ 의 주변확률밀도함수  $f(x_i)$ 는 다음과 같다.

$$f(x_i) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=b+1}^{b+n} f\left(x_i \mid \mathbf{x}_{-i}^{(j)}\right) \quad (5.1.2)$$

**셋째,** 강대수법칙이 성립한다. 즉, 스무스한(smoothing) 함수  $g(\cdot)$ 에 대해서 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=b+1}^{b+n} g\left(\mathbf{x}^{(j)}\right) = E(g(\mathbf{x})) \quad a.s. \quad (5.1.3)$$

넷째, 중심극한정리가 성립한다. 즉, 스무스한 함수  $g(\cdot)$  에 대해서 다음 식이 성립한다.

$$\sqrt{n} \left[ \frac{1}{n} \sum_{j=b+1}^{b+n} g(\mathbf{x}^{(j)}) - E(g(\mathbf{x})) \right] \stackrel{d}{\sim} \mathcal{N}(0, \sigma_g^2) \quad (5.1.4)$$

**예제 5.1.1** 확률벡터  $[x, y]^t$  가 다음과 같은 2변량 정규확률분포를 따른다고 하자.

$$\begin{bmatrix} x \\ y \end{bmatrix} \stackrel{d}{\sim} \mathcal{N} \left( \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \quad (1)$$

다음 식들이 성립한다.

$$y | x \stackrel{d}{\sim} \mathcal{N}(\mu_y + \rho[x - \mu_x], 1 - \rho^2) \quad (2)$$

$$x | y \stackrel{d}{\sim} \mathcal{N}(\mu_x + \rho[y - \mu_y], 1 - \rho^2) \quad (3)$$

따라서, 다음 식들을 반복적용해서 Gibbs샘플들  $\{[x^{(j)}, y^{(j)}]^t\}$  를 생성할 수 있다.

$$y^{(j+1)} | x^{(j)} \stackrel{d}{\sim} \mathcal{N}(\mu_y + \rho[x^{(j)} - \mu_x], 1 - \rho^2) \quad (4)$$

$$x^{(j+1)} | y^{(j)} \stackrel{d}{\sim} \mathcal{N}(\mu_x + \rho[y^{(j)} - \mu_y], 1 - \rho^2) \quad (5)$$

식 (4)와 식 (5)를 사용해서 Gibbs샘플을 발생시키기 위해서, 다음 MATLAB프로그램 GibbsBivariateNormal101.m을 실행해 보자.

```

1 % -----
2 % Filename: GibbsBivariateNormal101.m
3 % Markov Chain Monte Carlo 1 for Bivariate Normal Distribution
4 % Programmed by CBS
5 % -----
6 clear, clf, close all
7 nn = 1050;
8 x = zeros(nn,1); y = zeros(nn,1);
9 rho = 0.9;
10 mux = 1; muy = -1;
11 sigma = sqrt(1-rho^2);
12 % Gibbs' Sampling
13 rng((1001^2+1)/2, 'twister')
14 x(1) = 10; y(1)=10;
15 for jj=2:nn
16     muxdum = mux + rho*(y(jj-1)-muy);
17     x(jj) = muxdum + sigma*randn(1);
18     muydum = muy + rho*(x(jj)-mux);
19     y(jj) = muydum + sigma*randn(1);

```

```

20 end
21 % Plotting
22 plot(x(1:50),y(1:50),'kd',x(1001:1050),y(1001:1050),'ro','linewidth',2)
23 hold on
24 set(gca,'fontsize',11,'fontweigh','bold')
25 legend('\bf Observations 1-50','\bf Observations 1001-1050', ...
26         'location','NW')
27 axis([-3 12 -4 11])
28 plot([-3 12],[-1 -1],'b-',[1 1],[-4 11],'b-')
29 hold off
30 saveas(gcf,'GibbsBivariateNormal101a','eps')
31 % Generating Bivariate Normal Random Numbers using SVD
32 zmat = zeros(2,nn);
33 mu = [ mux  muy ]';
34 Sigma = [ 1 rho ; rho 1 ];
35 [ P,D,Q ] = svd(Sigma)          % Schur decomposition
36 srLamb = sqrt(D)
37 for j=1:nn
38     z = mu + P*srLamb*randn(2,1);
39     zmat(:,j) = z;
40 end
41 % Plotting
42 figure
43 plot(zmat(1,1:50),zmat(2,1:50),'kd', ...
44      zmat(1,1001:1050),zmat(2,1001:1050),'ro','linewidth',2)
45 hold on
46 set(gca,'fontsize',11,'fontweigh','bold')
47 legend('\bf Observations 1-50','\bf Observations 1001-1050', ...
48         'location','NW')
49 axis([-3 12 -4 11])
50 plot([-3 12],[-1 -1],'b-',[1 1],[-4 11],'b-')
51 hold off
52 saveas(gcf,'GibbsBivariateNormal101b','eps')
53 save('GibbsBivariateNormal101','x','y','zmat')
54 % End of program
55 % -----

```

이 MATLAB 프로그램 GibbsBivariateNormal101.m을 실행하면, 평균벡터가  $[\mu_x, \mu_y]^t = [1, -1]^t$ 이고 또한 상관계수가  $\rho = 0.9$ 인 2변량 정규확률분포에서 Gibbs샘플들  $\{[x^{(j)}, y^{(j)}]^t \mid j = 1, 2, \dots, 1050\}$ 이 발생된다. 여기서 초기벡터는  $[10, 10]^t$ 이다. 이 Gibbs 샘플들의 일부가 그림 5.1.1에 그려져 있다. 이 그래프에서 흑색 다이아몬드점들은 첫 50개 Gibbs샘플들이고, 적색 원은 마지막 50개 Gibbs샘플들이다. 이 그래프에서 확인할 수 있듯이, 반복횟수가 커짐에 따라 Gibbs샘플들은 점점 목표확률분포인 2변량 정규확률분포에 가까운 확률분포로 가까이 가는 것처럼 보인다.

최병선 [7]의 제4.3절에는 Cholesky분해, Schur분해, 그리고 특이값분해(singular value decomposition: SVD)를 사용해서 다변량정규확률분포로부터 정규난벡터(normal random vector)를 생성하는 방법에 대해 다루었다. 이 중에서 특이값분해를 사용해서 식 (1)의 2변량 정규확률분포로부터 생성한 난벡터들의 일부가 그림 5.1.2에 그려져 있다. 이 그래프에서 흑색 다이아몬드점들은 첫 50개 샘플들이고, 적색 원은 마지막 50개 샘플들이다. 이 그래프에서 알

수 있듯이, 특이값분해를 사용해서 발생시킨 정규난벡터들은 반복횟수에 상관없이 식 (1)의 2변량 정규확률분포를 따른다고 할 수 있다. ■

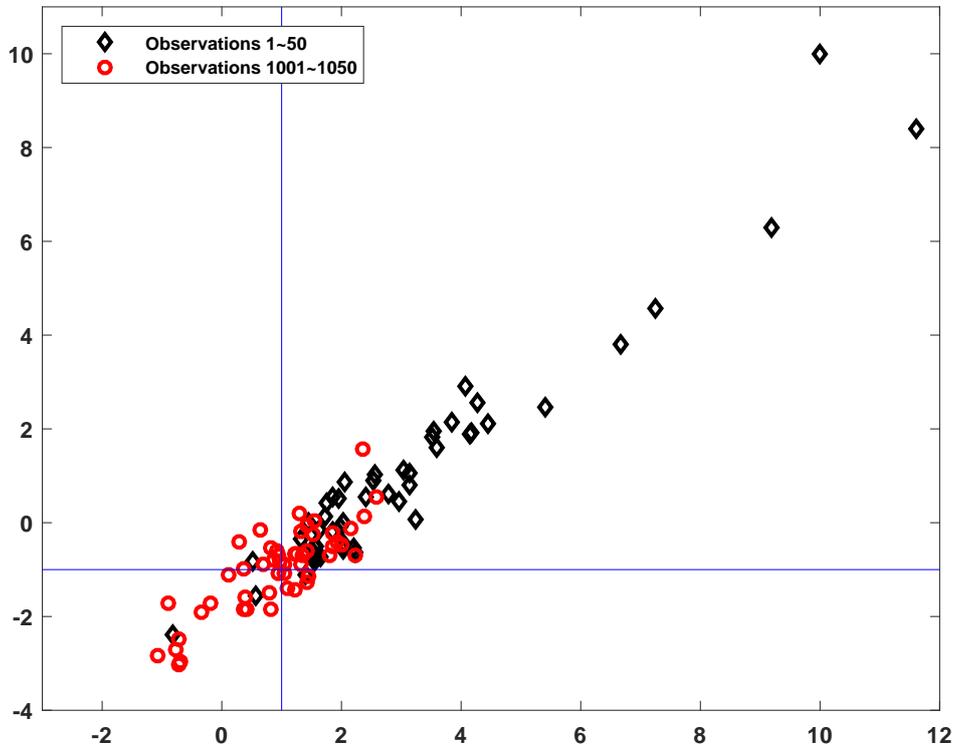


그림 5.1.1. Gibbs샘플러와 2변량 정규확률분포 1

**예제 5.1.2** 예제 5.1.1의 2변량 정규확률분포에서 Gibbs샘플을 발생시키기 위해서, 다음 R프로그램 GibbsBivariateNormal101R.R을 실행해 보자.

```

1 # -----
2 # Filename: GibbsBivariateNormal101R.R
3 # Gibbs Sampler for Bivariate Normal Distribution
4 # Programmed by CBS
5 # -----
6 # Gibbs Sampler Function
7 GibbsSampler <- function (Nobs,mux,muy,rho)
8 {
9     zmat <- matrix(ncol=2, nrow=Nobs)
10    set.seed(1)
11    x <- 10
12    y <- 10
13    zmat[1, ] <- c(x, y)
14    for (ii in 2:Nobs) {
15        y <- rnorm(1,muy+rho*(x-mux),sqrt(1 - rho^2))
16        x <- rnorm(1,mux+rho*(y-muy),sqrt(1 - rho^2))
17        zmat[ii, ] <- c(x, y)
18    }
19    zmat
20 }

```

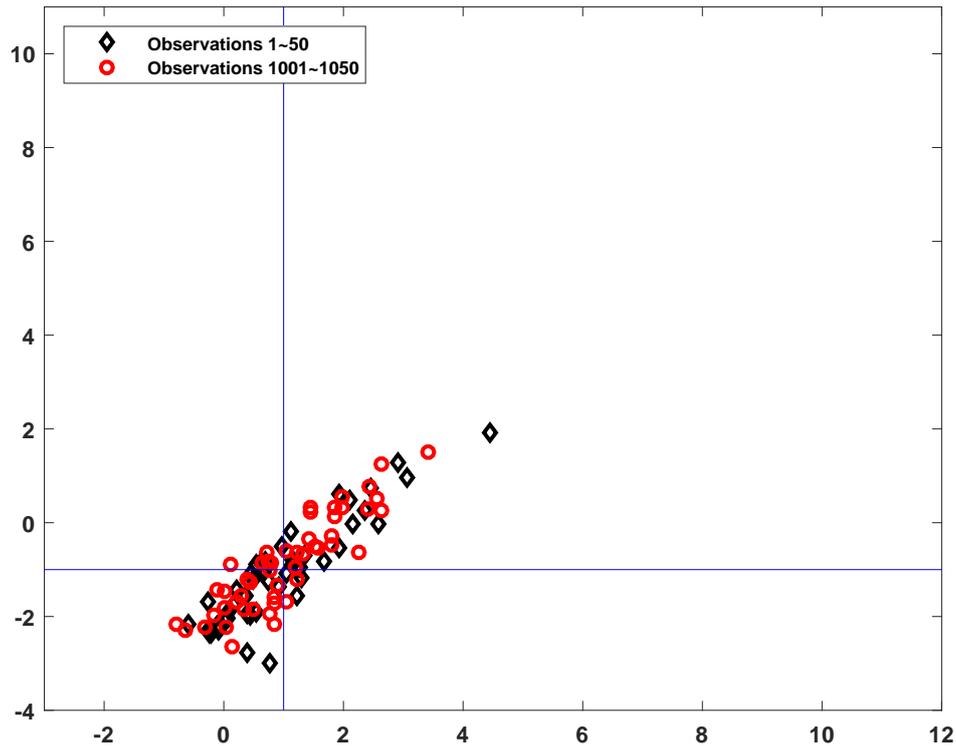


그림 5.1.2. 특이값분해와 2변량 정규확률분포

```

21
22 # Generating Gibbs Samples
23 Nobs = 1050
24 bivNormal <- GibbsSampler(Nobs,1,-1,0.9)
25
26 # Plotting
27 library(ggplot2)
28 # install.packages("grid")
29 library(grid)
30 setEPS()
31 plot.new()
32 postscript('GibbsBivariateNormal101R.eps') # Start to save figure
33 timee = c(1:Nobs)
34 GibbsData <- data.frame(timee,bivNormal)
35 plot1 <- ggplot(GibbsData, aes(timee,bivNormal[,1])) +
36   geom_line(colour="dark red") + xlab("Time") + ylab("x")
37 plot2 <- ggplot(GibbsData, aes(timee,bivNormal[,2])) +
38   geom_line(colour="dark blue") + xlab("Time") + ylab("y")
39 plot3 <- ggplot(GibbsData, aes(x=bivNormal[,1])) +
40   geom_histogram(aes(y =..density..), bins=30,
41     # breaks=seq(0.0,2.5, by=0.05),
42     col="red",fill="green") +
43   geom_density(lwd=1.0)
44 plot4 <- ggplot(GibbsData, aes(x=bivNormal[,2])) +
45   geom_histogram(aes(y =..density..), bins=30,
46     # breaks=seq(0.0,2.5, by=0.05),
47     col="red",fill="pink") +
48   geom_density(lwd=1.0)
49 pushViewport(viewport(layout = grid.layout(2,2)))
50 print(plot1, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
51 print(plot2, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
52 print(plot3, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))

```

```

53 print(plot4, vp = viewport(layout.pos.row = 2, layout.pos.col = 2))
54 dev.off() # End to save figure
55 # -----

```

이 R프로그램 GibbsBivariateNormal101R.R을 실행하면, 평균벡터가  $[\mu_x, \mu_y]^t = [1, -1]^t$ 이고 또한 상관계수가  $\rho = 0.9$ 인 2변량 정규확률분포에서 Gibbs샘플들  $\{[x^{(j)}, y^{(j)}]^t \mid j = 1, 2, \dots, 1050\}$ 이 발생된다. 여기서 초기벡터는  $[10, 10]^t$ 이다. 이 Gibbs샘플들이 그림 5.1.3에 그려져 있다. 그림 5.1.3의 좌측상단 그래프에는  $x$ 의 시계열산점도가 그려져 있고, 우측상단 그래프에는  $y$ 의 시계열산점도가 그려져 있고, 좌측하단 그래프에는  $x$ 의 히스토그램이 그려져 있고, 좌측하단 그래프에는  $y$ 의 히스토그램이 그려져 있다. 이 그래프들에서 알 수 있듯이, Gibbs샘플러는 초기값들에서 빨리 벗어나서 점근적 목표확률분포로부터 난수들을 생성한다. ■

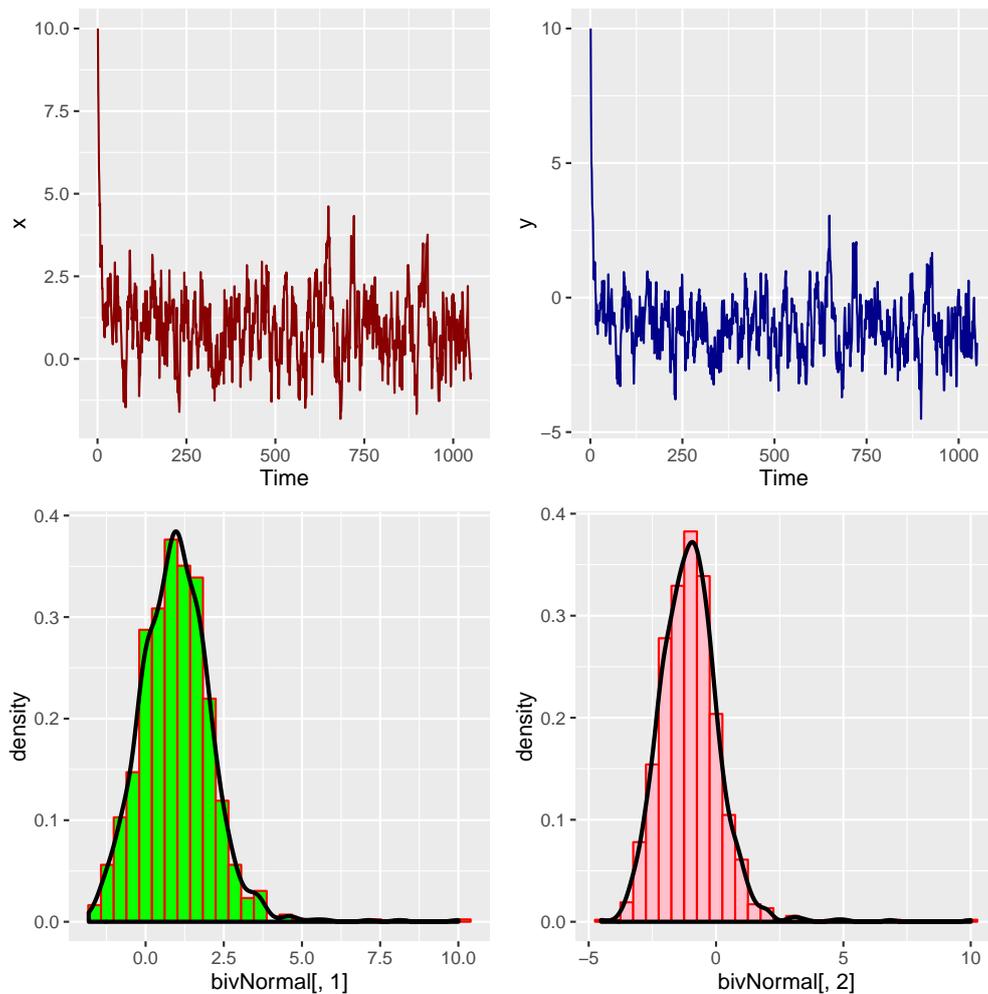


그림 5.1.3. Gibbs샘플러와 2변량 정규확률분포 2

**예제 5.1.3** 확률벡터  $[x, y]^t$ 가 다음과 같은 2변량 정규확률분포를 따른다고 하자.

$$\begin{bmatrix} x \\ y \end{bmatrix} \stackrel{d}{\sim} N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \quad (1)$$

예제 5.1.1에서 알 수 있듯이, 다음 식들이 성립한다.

$$y|x \stackrel{d}{\sim} \mathcal{N}(\rho x, 1 - \rho^2) \quad (2)$$

$$x|y \stackrel{d}{\sim} \mathcal{N}(\rho y, 1 - \rho^2) \quad (3)$$

따라서 다음 식들을 반복적용해서 Gibbs샘플들  $\{[x^{(j)}, y^{(j)}]^t\}$ 를 발생시킬 수 있다.

$$y^{(j+1)} \stackrel{d}{\sim} \mathcal{N}(\rho x^{(j)}, 1 - \rho^2) \quad (4)$$

$$x^{(j+1)} \stackrel{d}{\sim} \mathcal{N}(\rho y^{(j)}, 1 - \rho^2) \quad (5)$$

상관계수들이 각각  $\rho = 0.3$ 과  $0.9$ 인 경우, 초기벡터  $[1, 1]^t$ 을 출발하는 Gibbs샘플을 발생시키기 위해서, 다음 MATLAB 프로그램 GibbsBivariateNormal102.m을 실행해 보자.

```

1 % -----
2 % Filename: GibbsBivariateNormal102.m
3 % Markov Chain Monte Carlo 2 for Bivariate Normal Distribution
4 % Programmed by CBS
5 % -----
6 clear, clf, close all
7 nn = 100; % number of observations to generate
8 % rho = 0.3
9 rho = 0.3
10 x3 = zeros(1,nn);
11 y3 = zeros(1,nn);
12 % Gibb's sampling
13 rng((1001^2+1)/2, 'twister')
14 x3(1) = 1;
15 y3(1) = 1;
16 for ii=2:nn
17     x3(ii) = normrnd(rho*y3(ii-1), 1-rho^2);
18     y3(ii) = normrnd(rho*x3(ii), 1-rho^2);
19 end
20 % Plotting
21 subplot(2,2,1)
22 plot(x3(1:10), y3(1:10), 'g-', x3(1:10), y3(1:10), 'k.', 'linewidth', 2)
23 set(gca, 'fontsize', 11, 'fontweigh', 'bold')
24 dxy3 = 3*sqrt(1-rho^2);
25 axis([-dxy3, dxy3, -dxy3, dxy3])
26 axis equal
27 subplot(2,2,2)
28 plot(x3, y3, 'g-', x3, y3, 'k.', 'linewidth', 2)

```

```

29 set(gca, 'fontsize', 11, 'fontweigh', 'bold')
30 dxy3 = 3*sqrt(1-rho^2);
31 axis([-dxy3, dxy3, -dxy3, dxy3])
32 axis equal
33 %
34 % rho = 0.9
35 rho = 0.9
36 x9 = zeros(1, nn);
37 y9 = zeros(1, nn);
38 % Gibb's sampling
39 x9(1) = 1;
40 y9(1) = 1;
41 for ii=2:nn
42     x9(ii) = normrnd(rho*y9(ii-1), 1-rho^2);
43     y9(ii) = normrnd(rho*x9(ii), 1-rho^2);
44 end
45 % Ploting
46 subplot(2,2,3)
47 plot(x9(1:10), y9(1:10), 'g-', x9(1:10), y9(1:10), 'k.', 'linewidth', 2)
48 set(gca, 'fontsize', 11, 'fontweigh', 'bold')
49 dxy9 = 3*sqrt(1-rho^2);
50 axis([-dxy9, dxy9, -dxy9, dxy9])
51 axis equal
52 subplot(2,2,4)
53 plot(x9, y9, 'g-', x9, y9, 'k.', 'linewidth', 2)
54 set(gca, 'fontsize', 11, 'fontweigh', 'bold')
55 dxy9 = 3*sqrt(1-rho^2);
56 axis([-dxy9, dxy9, -dxy9, dxy9])
57 axis equal
58 saveas(gcf, 'GibbsBivariateNormal102', 'eps')
59 % End of program
60 % -----

```

이 MATLAB 프로그램 GibbsBivariateNormal102.m을 실행하면, 평균벡터가  $[\mu_x, \mu_y]^t = [1, -1]^t$ 인 2변량 정규확률분포에서 Gibbs샘플들  $\{[x^{(j)}, y^{(j)}]^t \mid j = 1, 2, \dots, 1050\}$ 가 발생된다. 여기서 초기벡터는  $[1, 1]^t$ 이다. 이 Gibbs샘플들의 일부가 그림 5.1.4에 그려져 있다. 그림 5.1.4의 좌측상단 그래프와 우측상단 그래프는 각각  $\rho = 0.3$ 인 첫 10개 샘플들과 첫 100개 샘플들의 산점도들이다. 그리고, 좌측하단 그래프와 우측하단 그래프는 각각  $\rho = 0.9$ 인 첫 10개 샘플들과 첫 100개 샘플들의 산점도들이다. 이 그래프들에서 알 수 있듯이, 상관계수  $\rho$ 가 크면 초기벡터에 의존해서 Gibbs샘플들의 움직임이 제한된다. ■

**예제 5.1.4** 다음 정적분을 계산해 보자.

$$V = \int_0^1 \int_0^1 \int_0^1 x_1 x_2 x_3 \sin(x_1 + x_2 + x_3) dx_1 dx_2 dx_3 \quad (1)$$

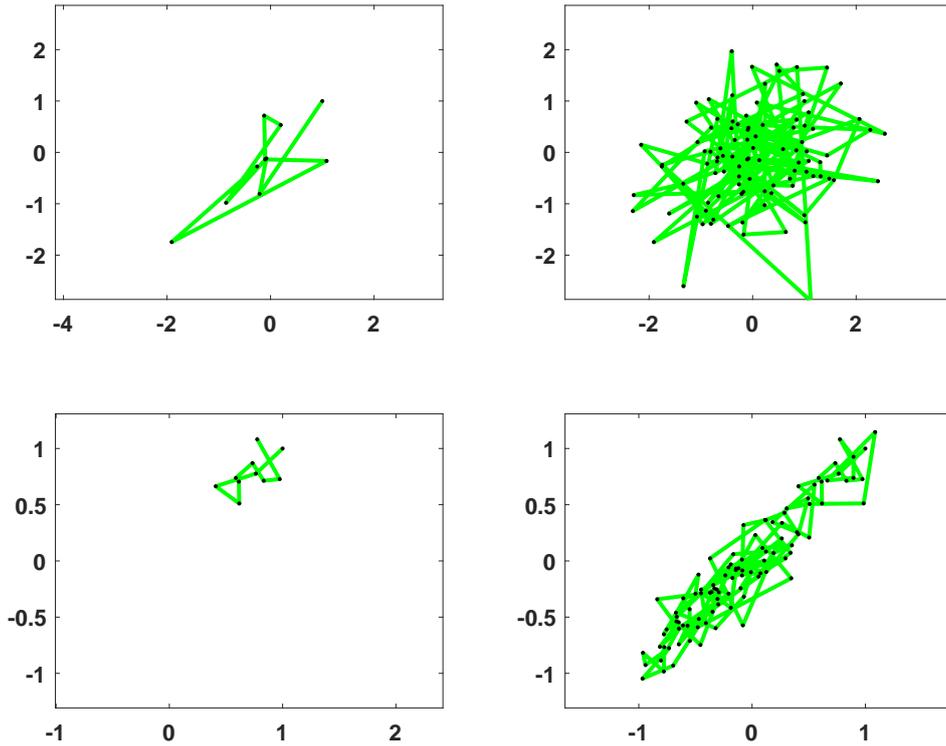


그림 5.1.4. Gibbs샘플러와 2변량 정규확률분포

이 정적분값을 해석적으로 구하면 다음과 같다.

$$V = 6 \cos 2 - 3 \cos 1 - 2 \cos 3 + 3 \sin 1 - 2 \sin 3 = 0.104369994617573 \quad (2)$$

첫째, 단순몬테카를로법을 사용해서 정적분  $V$ 를 계산하기 위해서, 확률변수들  $x_1, x_2$ , 그리고  $x_3$ 는 지지대가  $(0, 1)$ 인 일양확률분포를 따르며 서로 독립이라고 하자. 다음 식이 성립한다.

$$V = E_U (x_1 x_2 x_3 \sin (x_1 + x_2 + x_3)) \quad (3)$$

여기서  $E_U(\cdot)$ 는 이 일양결합확률분포 하에서 기대값연산자이다.

둘째, Gibbs샘플러를 사용해서 정적분  $V$ 를 계산하기 위해서, 확률벡터  $\mathbf{x} = [x_1, x_2, x_3]^t$ 의 확률밀도함수가 다음과 같다고 가정하자.

$$p(\mathbf{x}) = \frac{1}{C} \sin (x_1 + x_2 + x_3) \prod_{i=1}^3 1_{(0,1)}(x_i) \quad (4)$$

해석적 방법 또는 MATLAB의 Symbolic Toolbox의 함수 Int를 사용하면, 다음 식들이 성립

함을 알 수 있다.

$$C = \cos 3 - 3 \cos 2 + 3 \cos 1 - 1 = 0.879354930645401 \quad (5)$$

그러나, Gibbs샘플러에서 이  $C$  값을 사용하지는 않는다. 다음 식이 성립한다.

$$V = C E_T(x_1 x_2 x_3) \quad (6)$$

여기서  $E_T(\cdot)$ 는 식 (4)에서 정의한 확률밀도함수  $p(\mathbf{x})$  하에서 기대값연산자이다.

식 (3)과 식 (6)을 바탕으로 식 (1)의 정적분을 계산하기 위해서, 다음 MATLAB프로그램 GibbsMultiIntegration101.m을 실행해 보자.

```

1 % -----
2 % Filename: GibbsMultiIntegration101.m
3 % Gibbs Sampler for Multiple Integration 1
4 % Programmed by CBS
5 % -----
6 clear, close all, format long
7 % True Integration
8 syms xx yy zz
9 Inte = int(int(int(xx*yy*zz*sin(xx+yy+zz),xx,0,1),yy,0,1),zz,0,1)
10 integ = double(Inte)
11
12 % Simple Monte carlo Mehtod
13 rng((1001^2+1)/2, 'twister')
14 nn = 100000
15 stepW = nn/20; Nobs = stepW*(1:20);
16 xran = rand(nn,3);
17 MC = prod(xran,2).*sin(sum(xran,2));
18 for ii=1:20
19     nSample = stepW*ii;
20     ExpMC(ii) = mean(MC(1:nSample));
21     SDMC(ii) = std(MC(1:nSample));
22     LowerMC(ii) = ExpMC(ii)-2*SDMC(ii)/sqrt(nSample);
23     UpperMC(ii) = ExpMC(ii)+2*SDMC(ii)/sqrt(nSample);
24 end
25 ExpMC, SDMC, LowerMC, UpperMC
26 % Plotting
27 subplot(2,1,1)
28 plot(Nobs,integ*ones(1,20), Nobs,ExpMC,'k-',Nobs,LowerMC,'r--', ...
29     Nobs,UpperMC,'b--','linewidth',2)
30 set(gca,'fontsize',11,'fontweigh','bold','ylim',[0.10,0.11])
31 legend('\bf True value','\bf MC Integration','location','NE')
32
33 % Gibbs Sampling
34 C = int(int(int(sin(xx+yy+zz),xx,0,1),yy,0,1),zz,0,1)
35 C = double(C)
36 nk = 110000
37 x = 0.5*ones(3,1);
38 for jj=1:nk
39     ii = ceil(3*rand);
40     S_ii = sum(x) - x(ii);

```

```

41     x(ii) = acos( cos(S_ii) - rand*(cos(S_ii)-cos(S_ii +1)) ) - S_ii;
42     IntGibbs(jj) = C*prod(x);
43 end
44 bb = 1/11*nk; nn = nk-bb;
45 for ii=1:20
46     nSample = stepW*ii;
47     ExpGibbs(ii) = mean(IntGibbs(bb+1:(bb+ii*stepW)));
48     SDGibbs(ii) = std(IntGibbs(bb+1:(bb+ii*stepW)));
49     LowerGibbs(ii) = ExpGibbs(ii)-2*SDGibbs(ii)/sqrt(nSample);
50     UpperGibbs(ii) = ExpGibbs(ii)+2*SDGibbs(ii)/sqrt(nSample);
51 end
52 ExpGibbs, SDGibbs, LowerGibbs, UpperGibbs
53 % Plotting
54 subplot(2,1,2)
55 plot(Nobs,UpperGibbs,'r--',Nobs,LowerGibbs,'b--', ...
56      Nobs,integ*ones(1,20),'g-',Nobs,ExpGibbs,'k-', 'linewidth',2)
57 set(gca,'fontsize',11,'fontweigh','bold','ylim',[0.10,0.11])
58 legend('\bf True value','\bf Gibbs Integration','location','NE')
59 saveas(gcf,'GibbsMultiIntegration101','eps')
60 % End of program
61 % -----

```

이 MATLAB 프로그램 GibbsMultiIntegration101.m을 실행하면, 우선 단순몬테카를로 방법을 적용하기 위해 100,000개 샘플들을 발생시킨다. 이 샘플들의 표본평균과 표본표준편차는 각각 0.1042과 0.1016이고, 신뢰수준 95%의 신뢰구간은 (0.1036, 0.1044)이다. 또한 이 MATLAB 프로그램을 실행하면, 단순몬테카를로법에 의해서 생성된 샘플수가 500에서 500씩 증가해서 100,000이 될 때까지, 표본평균들과 95% 신뢰구간들이 그려진다. 이 그래프가 그림 5.1.5의 상단에 그려져 있다.

Gibbs샘플러를 적용하기 위해서, 다음 변수들을 정의하자.

$$S_{-1} \doteq x_2 + x_3, \quad S_{-2} \doteq x_1 + x_3, \quad S_{-3} \doteq x_1 + x_2 \quad (7)$$

조건부확률변수  $x_1 | (x_2, x_3)$ 의 누적확률분포함수  $F(x_1 | x_2, x_3)$ 를 구하면, 다음과 같다.

$$F(x_1 | x_2, x_3) = \frac{\int_0^{x_1} \sin(z + S_{-1}) dz}{\int_0^1 \sin(z + S_{-1}) dz} = \frac{\cos(S_{-1}) - \cos(S_{-1} + x_1)}{\cos(S_{-1}) - \cos(S_{-1} + 1)} \quad (8)$$

따라서, 조건부확률변수  $x_1 | (x_2, x_3)$ 의 확률분포는 다음과 같다.

$$x_1 | (x_2, x_3) \stackrel{d}{\sim} \cos^{-1}(\cos(S_{-1}) - U[\cos(S_{-1}) - \cos(S_{-1} + 1)]) - S_{-1} \quad (9)$$

여기서  $U$ 는 일양확률변수  $Uniform(0, 1)$ 을 따른다. 같은 방법을 사용해서, 조건부확률변수들

$x_2 | (x_1, x_3)$ 와  $x_3 | (x_1, x_2)$ 가 다음과 같음을 증명할 수 있다.

$$x_2 | (x_1, x_3) \stackrel{d}{\sim} \cos^{-1}(\cos(S_{-2}) - U[\cos(S_{-2}) - \cos(S_{-2} + 1)]) - S_{-2} \quad (10)$$

$$x_3 | (x_1, x_2) \stackrel{d}{\sim} \cos^{-1}(\cos(S_{-3}) - U[\cos(S_{-3}) - \cos(S_{-3} + 1)]) - S_{-3} \quad (11)$$

식 (9) ~ 식 (11)을 바탕으로한 Gibbs샘플러를 해서, 식 (6)의 기대값을 추정할 수 있다. 유의할 점은 식 (9) ~ 식 (11)에 상수  $C$ 가 포함되지 않는다는 것이다.

MATLAB 프로그램 GibbsMultiIntegration101.m을 실행하면, 번인기간이  $b = 10,000$ 이고 샘플수가  $n = 100,000$ 인 Gibbs샘플들이 생성된다. 이 Gibbs샘플들의 표본평균과 표본표준편차는 각각 0.1045과 0.1139이고, 신뢰수준 95%의 신뢰구간은 (0.1038, 0.1052)이다. 또한, 이 MATLAB 프로그램을 실행하면, Gibb샘플러에서 생성된 샘플수가 500에서 500씩 증가해서 100,000이 될 때까지, 표본평균들과 95% 신뢰구간들이 그려진다. 이 그래프가 그림 5.1.5의 하단에 그려져 있다.

이 예제에서는 Gibbs샘플러에 의한 신뢰구간이 단순몬테카를로법에 의한 신뢰구간과 크게 다르지 않다. 확률변수들  $x_1, x_2$ , 그리고  $x_3$ 가 서로 독립이기 때문에, 식 (9) ~ 식 (11)이 별로 유용하지 않나 생각한다.

다음 Python 프로그램 GibbsMultiIntegration101.py를 실행하면, 앞의 MATLAB 프로그램 GibbsMultiIntegration101.m과 동일한 결과를 얻는다.

```

1 # GibbsMultiIntegration101.py
2 """
3 Filename : GibbsMultiIntegration101.py
4 Gibbs Sampler for Multiple Integration 1
5 Programmed by Kyung-Bae Park based on the R program by CBS
6 """
7
8 import numpy as np
9 import matplotlib.pyplot as plt
10 import math
11 import cmath
12 from sympy import *
13
14 # True Integration
15 xx, yy, zz = symbols('xx yy zz')
16 Inte = integrate(integrate(integrate(xx*yy*zz*sin(xx+yy+zz),
17                               (xx,0,1)),(yy,0,1)),(zz,0,1))
18 integ = float(Inte)
19
20
21 # Simple Monte Carlo Method
22 np.random.seed(int((1001**2+1)/2))
23 nn = 100000
24 stepW = nn/20
25 Nobs = stepW*np.linspace(1,20,20)

```

```

26 xran = np.random.uniform(size=(3,nn)).T
27 MC = np.product(xran, axis=1)*np.sin(np.sum(xran, axis=1))
28 ExpMC = np.empty(20)
29 SDMC = np.empty(20)
30 LowerMC = np.empty(20)
31 UpperMC = np.empty(20)
32
33 for ii in range(20):
34     nSample = int(stepW*(ii+1))
35     ExpMC[ii] = np.mean(MC[range(nSample)])
36     SDMC[ii] = np.std(MC[range(nSample)])
37     LowerMC[ii] = ExpMC[ii]-2*SDMC[ii]/math.sqrt(nSample)
38     UpperMC[ii] = ExpMC[ii]+2*SDMC[ii]/math.sqrt(nSample)
39
40
41 # Gibbs Sampling
42 C = integrate(integrate(integrate(sin(xx+yy+zz),(xx,0,1)),(yy,0,1)),(zz,0,1))
43 C = float(C)
44 nk = 110000
45 x = 0.5*np.linspace(1,1,3)
46 IntGibbs = np.empty(nk)
47 for jj in range(nk):
48     ii = math.ceil(3*np.random.uniform())
49     S_ii = sum(x) - x[ii-1]
50     x[ii-1] = math.acos(math.cos(S_ii)-np.random.uniform()*
51                        (math.cos(S_ii)-math.cos(S_ii+1)))-S_ii
52     IntGibbs[jj] = C*np.product(x)
53 bb = int(1/11*nk)
54 nn = int(nk-bb)
55 ExpGibbs = np.empty(20)
56 SDGibbs = np.empty(20)
57 LowerGibbs = np.empty(20)
58 UpperGibbs = np.empty(20)
59 for ii in range(20):
60     nSample = int(stepW*(ii+1))
61     ExpGibbs[ii] = np.mean(IntGibbs[bb:(bb+nSample)])
62     SDGibbs[ii] = np.std(IntGibbs[bb:(bb+nSample)])
63     LowerGibbs[ii] = ExpGibbs[ii]-2*SDGibbs[ii]/math.sqrt(nSample)
64     UpperGibbs[ii] = ExpGibbs[ii]+2*SDGibbs[ii]/math.sqrt(nSample)
65
66
67 # Plotting
68 fig_a = plt.figure()
69 plt.subplot(211)
70 plt.plot(Nobs, integ*np.linspace(1,1,20), color = 'green', linewidth = 2)
71 plt.plot(Nobs, ExpMC, color = 'black', linewidth = 2)
72 plt.plot(Nobs, LowerMC, color = 'red', linewidth = 2, ls = '--')
73 plt.plot(Nobs, UpperMC, color = 'blue', linewidth = 2, ls = '--')
74 plt.ylim(0.10,0.11)
75 plt.title('MC Integration')
76 plt.legend(['True value', 'MC Integration'], loc=1)
77
78 plt.subplot(212)
79 plt.plot(Nobs, integ*np.linspace(1,1,20), color = 'green', linewidth = 2)
80 plt.plot(Nobs, ExpGibbs, color = 'black', linewidth = 2)
81 plt.plot(Nobs, LowerGibbs, color = 'red', linewidth = 2, ls = '--')
82 plt.plot(Nobs, UpperGibbs, color = 'blue', linewidth = 2, ls = '--')
83 plt.ylim(0.10,0.11)
84 plt.title('Gibbs Integration')
85 plt.legend(['True value', 'Gibbs Integration'], loc=1)
86 plt.subplots_adjust(hspace=1)

```

```

87 fig_a.savefig('GibbsMultiIntegration101.eps', format = 'eps')
88
89
90 # End of program
91 # -----
    
```

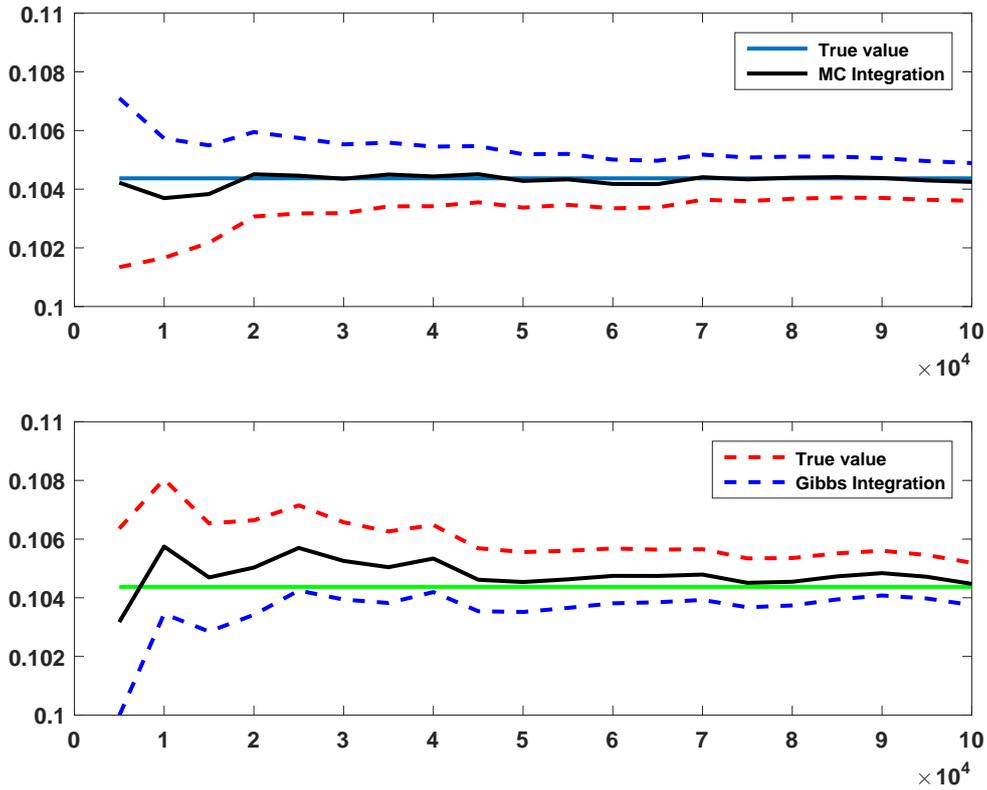


그림 5.1.5. Gibbs샘플러를 이용한 다중적분 1

**예제 5.1.5** 다음 정적분을 계산해 보자.

$$V = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 \exp\left(-\frac{1}{2} [x_1^2 x_2^2 + x_1^2 + x_2^2 - 2x_1 - 2x_2]\right) dx_1 dx_2 \quad (1)$$

첫째, 단순몬테카를로법을 사용해서 정적분  $V$ 를 계산하기 위해서,  $x_1$ 과  $x_2$ 가 서로 독립인 표준정규확률변수들이라고 하자. 다음 식이 성립한다.

$$V = E_N\left(2\pi x_1 x_2 \exp\left(-\frac{1}{2} [x_1^2 x_2^2 - 2x_1 - 2x_2]\right)\right) \quad (2)$$

여기서  $E_N(\cdot)$ 은 이 2변량 정규확률분포 하에서 기대값연산자이다.

둘째, Gibbs 샘플러를 사용해서 정적분  $V$ 를 계산하기로 하자. 다음과 같이 상수  $C$ 를 정의하자.

$$C \doteq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}[x_1^2 x_2^2 + x_1^2 + x_2^2 - 2x_1 - 2x_2]\right) dx_1 dx_2 \quad (3)$$

MATLAB의 Symbolic Toolbox의 함수 `int`를 사용하면, 다음 식이 성립함을 알 수 있다.

$$C = 9.788588777847213 \quad (4)$$

확률벡터  $\mathbf{x} = [x_1, x_2]^t$ 의 확률밀도함수를 다음과 같이 표기할 수 있다.

$$p(\mathbf{x}) = \frac{1}{C} \exp\left(-\frac{1}{2}[x_1^2 x_2^2 + x_1^2 + x_2^2 - 2x_1 - 2x_2]\right) \quad (5)$$

즉, 다음 식들이 성립한다.

$$V = C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2) dx_1 dx_2 = C E_p(x_1 x_2) \quad (6)$$

여기서  $E_p(\cdot)$ 는 확률밀도함수  $p(\mathbf{x})$  하에서 기대값연산자이다. 다음 식들이 성립한다.

$$x_1^2 x_2^2 + x_1^2 + x_2^2 - 2x_1 - 2x_2 = [x_2^2 + 1] \left[ x_1 - \frac{1}{x_2^2 + 1} \right]^2 + \left[ \frac{1}{x_2^2 + 1} + x_2^2 - 2x_2 \right] \quad (7)$$

$$x_2^2 x_1^2 + x_2^2 + x_1^2 - 2x_2 - 2x_1 = [x_1^2 + 1] \left[ x_2 - \frac{1}{x_1^2 + 1} \right]^2 + \left[ \frac{1}{x_1^2 + 1} + x_1^2 - 2x_1 \right] \quad (8)$$

조건부확률변수들  $x_1 | x_2$ 와  $x_2 | x_1$ 의 확률밀도함수들  $p(x_1 | x_2)$ 와  $p(x_2 | x_1)$ 은 각각 다음과 같다.

$$p(x_1 | x_2) = \frac{p(x_1, x_2)}{p(x_2)} = \frac{1}{2\pi\sqrt{x_2^2 + 1}} \exp\left(-\frac{1}{2}[x_2^2 + 1] \left[ x_1 - \frac{1}{x_2^2 + 1} \right]^2\right) \quad (9)$$

$$p(x_2 | x_1) = \frac{p(x_1, x_2)}{p(x_1)} = \frac{1}{2\pi\sqrt{x_1^2 + 1}} \exp\left(-\frac{1}{2}[x_1^2 + 1] \left[ x_2 - \frac{1}{x_1^2 + 1} \right]^2\right) \quad (10)$$

즉, 확률변수들  $x_1|x_2$ 와  $x_2|x_1$ 은 각각 다음과 같은 확률분포를 따른다.

$$x_1|x_2 \stackrel{d}{\sim} N\left(\frac{1}{x_2^2+1}, \frac{1}{x_2^2+1}\right) \quad (11)$$

$$x_2|x_1 \stackrel{d}{\sim} N\left(\frac{1}{x_1^2+1}, \frac{1}{x_1^2+1}\right) \quad (12)$$

비록 상수  $C$ 를 알지 못하더라도, 식 (11)과 식 (12)의 조건부확률밀도함수들을 사용해서 난수들을 생성할 수 있다.

식 (6)을 바탕으로 식 (1)의 정적분을 계산하기 위해서, 다음 MATLAB프로그램 Gibbs-MultiIntegration102.m을 실행해 보자.

```

1 % -----
2 % Filename: GibbsMultiIntegration102.m
3 % Gibbs Sampler for Multiple Integration 2
4 % Programmed by CBS
5 % -----
6 clear, close all, format long;
7 rng((1001^2+1)/2,'twister')
8
9 % Simple Monte carlo Mehtod
10 tic;
11 nn = 100000
12 x1 = randn(nn,1); x2 = randn(nn,1);
13 dum1 = x1.^2.*x2.^2 - 2*(x1 + x2);
14 MC = 2*pi*x1.*x2.*exp(-dum1/2);
15 ExpMC = mean(MC)
16 SDMC = std(MC)
17 CIEExpMC = [ ExpMC-2*SDMC/sqrt(nn) ExpMC+2*SDMC/sqrt(nn) ]
18 toc;
19 % Plotting
20 stepW = nn/20;
21 Nobs = stepW*(1:20);
22 for ii=1:20
23     nSample = stepW*ii;
24     ExpMC(ii) = mean(MC(1:nSample));
25     SDMC(ii) = std(MC(1:nSample));
26     LowerMC(ii) = ExpMC(ii)-2*SDMC(ii)/sqrt(nSample);
27     UpperMC(ii) = ExpMC(ii)+2*SDMC(ii)/sqrt(nSample);
28 end
29 ExpMC, SDMC, LowerMC, UpperMC
30 subplot(2,1,1)
31 plot(Nobs,ExpMC,'k-', ...
32     Nobs,LowerMC,'r--',Nobs,UpperMC,'b--','linewidth',2)
33 set(gca,'fontsize',11,'fontweigh','bold','ylim',[1.7,2.5])
34 legend('\bf MC Integration','\bf Lower Bound','\bf Upper Bound', ...
35     'location','SE')
36
37 % Gibbs Sampling
38 syms xx yy zz
39 ftn = exp(-1/2*(xx^2*yy^2 + xx^2 + yy^2 - 2*(xx + yy)))
40 % C = int(int(ftn,xx,-inf,inf),yy,-inf,inf)
41 % C = double(C)
42 C = 9.788588777847213
43 tic;

```

```

44 nk = 110000
45 bb = 1/11*nk; % nn = nk-bb;
46 x = zeros(2,1);
47 for jj=1:nk
48     ii = ceil(2*rand);
49     dum2 = 1/(x(3-ii)^2+1);
50     x(ii) = dum2 + sqrt(dum2)*randn;
51     IntGibbs(jj) = C*x(1)*x(2);
52 end
53 toc;
54 for ii=1:20
55     nSample = stepW*ii;
56     ExpGibbs(ii) = mean(IntGibbs(bb+1:(bb+ii*stepW)));
57     SDGibbs(ii) = std(IntGibbs(bb+1:(bb+ii*stepW)));
58     LowerGibbs(ii) = ExpGibbs(ii)-2*SDGibbs(ii)/sqrt(nSample);
59     UpperGibbs(ii) = ExpGibbs(ii)+2*SDGibbs(ii)/sqrt(nSample);
60 end
61 ExpGibbs, SDGibbs, LowerGibbs, UpperGibbs
62 % Plotting
63 subplot(2,1,2)
64 plot(Nobs,ExpGibbs,'k-', ...
65      Nobs,LowerGibbs,'r--',Nobs,UpperGibbs,'b--','linewidth',2)
66 set(gca,'fontsize',11,'fontweigh','bold','ylim',[1.7,2.5])
67 legend('\bf Gibbs Sampler','\bf Lower Bound','\bf Upper Bound', ...
68        'location','SE')
69 saveas(gcf,'GibbsMultiIntegration102','eps')
70 % End of program
71 % -----

```

이 MATLAB 프로그램 GibbsMultiIntegration102.m을 실행하면, 단순몬테카를로법을 적용하기 위해 2변량 정규확률분포로부터 샘플들  $\{[x_1^{(j)}, x_2^{(j)}] \mid j = 1, 2, \dots, 100000\}$ 을 발생시킨다. 각  $[x_1^{(j)}, x_2^{(j)}]$ 로부터, 다음과 같이  $V^{(j)}$ 를 계산한다.

$$V^{(j)} = 2\pi x_1^{(j)} x_2^{(j)} \exp\left(-\frac{1}{2} \left\{ [x_1^{(j)}]^2 [x_2^{(j)}]^2 - 2x_1^{(j)} - 2x_2^{(j)} \right\}\right) \quad (13)$$

이 샘플들  $\{V^{(j)} \mid j = 1, 2, \dots, 100000\}$ 의 표본평균과 표본표준편차는 각각 2.2578과 11.0107이다. 또한, 신뢰수준 95%의 신뢰구간은 (2.1882, 2.3275)이다. 이 MATLAB 프로그램을 실행하면, 단순몬테카를로법에 의해서 생성된 샘플수가 500에서 500씩 증가해서 100,000이 될 때까지, 표본평균들과 95% 신뢰구간들이 그려진다. 이 그래프가 그림 5.1.6의 상단에 그려져 있다.

식 (11)과 식 (12)의 조건부확률분포들을 사용해서 변인기간이  $b = 10,000$ 이고 샘플수가  $n = 100,000$ 인 Gibbs 샘플들을 발생시킨다. 이 샘플들의 표본평균과 표본표준편차는 각각 2.3120과 6.5642이다. 또한, 신뢰수준 95%의 신뢰구간은 (2.2705, 2.3535)이다. 이 MATLAB 프로그램 GibbsMultiIntegration102.m을 실행하면, Gibb 샘플러에서 생성된 샘플수가 500에서 500씩 증가해서 100,000이 될 때까지, 표본평균들과 95% 신뢰구간들을 그린다. 이

그래프가 그림 5.1.6의 하단에 그려져 있다.

그림 5.1.6에서 알 수 있듯이, Gibbs샘플러에 의한 표본표준편차가 단순몬테카를로법에 의한 표본표준편차보다 작다. ■

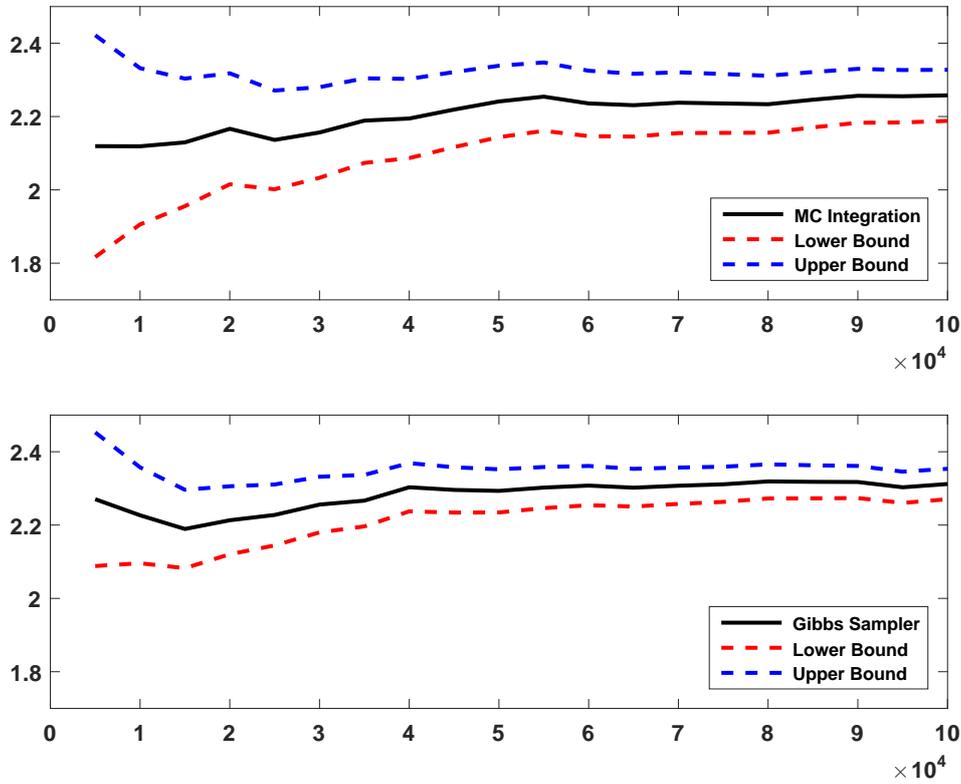


그림 5.1.6. Gibbs샘플러를 이용한 다중적분 2

**예제 5.1.6** 확률변수들  $x_1, x_2, \dots, x_p$ 가 서로 독립이며, 확률변수  $x_i$ 의 확률밀도함수가 다음과 같다고 하자.

$$f^{(i)}(x_i) = \lambda_i \exp(-\lambda_i x_i) 1(x_i > 0), \quad (i = 1, 2, \dots, p) \quad (1)$$

즉, 확률변수  $x_i$ 는 강도모수가  $\lambda_i (> 0)$ 인 지수확률분포를 따른다. 이 확률변수들의 합을 다음과 같이 표기하자.

$$S_p \doteq \sum_{i=1}^p x_i \quad (2)$$

지금부터는 조건부기대값  $E(S_p | S_p > c)$ 를 계산하는 문제를 생각해 보자. 만약  $c$ 가 큰 값이면, 조건  $S_p > c$ 를 만족하는 확률이 0에 가깝다. 따라서, 단순몬테카를로법을 적용하기

위해서는 불필요하게 많은 샘플들을 발생시켜야 한다. Gibbs샘플러를 사용하면, 이 난제를 피할 수 있다.

확률벡터  $\mathbf{x} \doteq [x_1, x_2, \dots, x_p]^t$ 의 결합확률밀도함수는 다음과 같다.

$$p(\mathbf{x}) = \left[ \prod_{i=1}^p \lambda_i \right] \exp \left( - \sum_{i=1}^p \lambda_i x_i \right) 1 \left( \min_{1 \leq i \leq p} x_i > 0 \right) \quad (3)$$

또한, 각  $i(= 1, 2, \dots, p)$ 에 대해서 확률변수  $x_i$ 는 다음 식을 만족한다.

$$\Pr(x_i > x) = \exp(-\lambda_i x) \quad (4)$$

각  $i(= 1, 2, \dots, p)$ 와 각  $j(= 1, 2, \dots)$ 에 대해서, 다음 확률변수들을 정의하자.

$$S_{-i} \doteq S_p - x_i, \quad S_{-i}^{(j)} \doteq \sum_{k=1}^p x_k^{(j)} - x_i^{(j)} \quad (5)$$

Gibbs샘플링을 하기 위해서는, 조건  $x_i^{(j)} + S_{-i}^{(j)} > c$ 가 만족되는 경우에 한해서 모수가  $\lambda_i$ 인 지수확률분포로부터 샘플  $x$ 를 추출한다. 다음 식들이 성립한다.

$$\begin{aligned} \Pr(x_i \leq x \mid x_i + S_{-i}^{(j)} > c) &= \frac{\Pr(x \geq x_i > c - S_{-i}^{(j)})}{\Pr(x_i > c - S_{-i}^{(j)})} \\ &= \frac{1}{\exp(-\lambda_i [c - S_{-i}^{(j)}])} \int_{c - S_{-i}^{(j)}}^x \lambda_i \exp(-\lambda_i x_i) dx_i \\ &= 1 - \frac{\exp(-\lambda_i x)}{\exp(-\lambda_i [c - S_{-i}^{(j)}])} \doteq U \end{aligned} \quad (6)$$

식 (6)에서 알 수 있듯이, 다음 식이 성립한다.

$$x = c - S_{-i}^{(j)} - \frac{1}{\lambda_i} \ln(1 - U) \quad (7)$$

만약  $U$ 가 구간  $(0, 1)$ 에서 일양확률분포를 따르면, 확률변수  $x_i$ 는 식 (6)의 조건부누적확률 분포  $\Pr(x_i \leq x \mid x_i + S_{-i}^{(j)} > c)$ 를 따른다. 식  $x_i > 0$ 를 만족하는 지지대만을 고려하므로, 식  $c > S_{-i}^{(j)}$ 가 성립해야한다. 따라서, 식 (7)의  $x$ 를 다음과 같이 변형한다.

$$x = \max \left\{ c - S_{-i}^{(j)}, 0 \right\} - \frac{1}{\lambda_i} \ln(1 - U) \quad (8)$$

식 (8)을 사용해서,  $\mathbf{x}_{-i}^{(j)} \doteq [x_1^{(j)}, \dots, x_{i-1}^{(j)}, x_{i+1}^{(j)}, \dots, x_p^{(j)}]^t$ 가 주어진 조건 하에서 Gibbs샘플  $x_i^{(j+1)}$ 를 추출한다.

식 (8)을 바탕으로 Gibbs샘플을 발생시키기 위해서, 다음 MATLAB프로그램 GibbsMultiExponential101.m을 실행해 보자.

```

1 % -----
2 % Filename: GibbsMultiExponential101.m
3 % Gibbs Sampler for Multi-Exponential Distribution
4 % Programmed by CBS
5 % -----
6 clear, close all, clf
7 rng((1001^2+1)/2, 'twister')
8 p = 10           % Number of Random Variables
9 c = 15;
10 % Sampling distribution of the summation S_p = sum_{i=1}^p x_i
11 %     where x_i is an exponential r.v. w/ lambda_i = i
12 nn = 5000       % Number of samples
13 XMC = - diag(1./[1:p])*log(rand(p,nn));
14 SMC = sum(XMC);
15 meanSMC = mean(SMC), sdSMC = std(SMC)
16 ProbPlau = mean(SMC > c)
17 % Histogram
18 xHistL = floor(min(SMC)); xHistR = ceil(max(SMC));
19 HistIncrement = ceil(max(SMC)-min(SMC))/20;
20 PMFind = (xHistL:HistIncrement:xHistR)';
21 [Nbar, xcenter] = hist(SMC,PMFind);
22 Nheight = Nbar/nn;
23 bar(xcenter, Nheight, 'w');
24 set(gca, 'fontsize', 11, 'fontweigh', 'bold')
25 ylabel('\bf Probability', 'fontsize', 12)
26 saveas(gcf, 'GibbsMultiExponential101', 'epsc')
27 % Monte carlo Simulation
28 nn = 1000000;
29 XMC = - diag(1./[1:p])*log(1-rand(p,nn));
30 SMC = sum(XMC);
31 meanSMC = mean(SMC), sdSMC = std(SMC)
32 ProbPlau = mean(SMC > c)
33 RSMC = SMC.*(SMC > c)/mean(SMC > c);
34 ExpMC = mean(RSMC)
35 CILMC = ExpMC - 2*std(RSMC)/sqrt(nn);
36 CIUMC = ExpMC + 2*std(RSMC)/sqrt(nn);
37 CIExpMC = [ CILMC CIUMC ]
38 % Gibbs Sampling
39 nk = 1300;
40 k = 300;           % burn-in number
41 nn = nk-k;        % number of observations to generate
42 np = zeros(1,p);
43 xGibbs = 1./(1:p)
44 for mm = 1:nk
45     ii = ceil(p*rand);
46     S_j_i = sum(xGibbs) - xGibbs(ii);
47     xGibbs(ii) = max(c-S_j_i, 0) - log(rand)/ii;
48     RSGibbs(mm) = S_j_i + xGibbs(ii);
49 end
50 ExpGibbs = mean(RSGibbs)
51 CILGibbs = ExpGibbs - 2*std(RSGibbs)/sqrt(nn);
52 CIUGibbs = ExpGibbs + 2*std(RSGibbs)/sqrt(nn);

```

```

53 CIExpGibbs = [ CILGibbs CIUGibbs ]
54 % End of program
55 % -----

```

이 MATLAB 프로그램 GibbsMultiExponential101.m은 단순몬테카를로법과 Gibbs샘플러를 사용해서, 확률변수  $x_k$ 는 모수가  $k$ 인 지수확률분포를 따르며 서로 독립인 확률변수들  $x_1, x_2, \dots, x_{10}$ 에 대한 조건부기대값  $E(S_{10} | S_{10} > 15)$ 을 계산한다.

확률변수  $S_{10}$ 의 히스토그램을 그리기 위해서, 확률벡터  $[x_1, x_2, \dots, x_{10}]^t$ 의 결합확률분포로부터 5,000개 샘플벡터를 발생시키고 이를 사용해서 확률변수  $S_{10} = \sum_{i=1}^{10} x_i$ 의 5,000개 샘플들을 만든다. 이들의 히스토그램이 그림 5.1.7에 그려져 있다. 이 샘플들의 표본평균과 표본표준편차는 각각 2.9265과 1.2464이다. 또한, 다음 식이 성립한다.

$$\Pr(S_{10} > 15) = 0 \quad (9)$$

식 (9)에서 알 수 있듯이, 단순몬테카를로법을 적용해서 조건부기대값  $E(S_{10} | S_{10} > 15)$ 을 계산하기 위해서는 아주 많은 샘플들을 발생시켜야 한다. 그렇지 않으면, 조건  $S_{10} > 15$ 를 만족하는 샘플이 하나도 없다. 따라서, 이 MATLAB 프로그램을 실행하면, 단순몬테카를로법을 적용하기 위해서  $10^6$ 개 샘플들을 발생시킨다. 이 샘플들의 표본평균과 표본표준편차는 각각 2.9296과 1.2442이다. 또한, 다음 식이 성립한다.

$$\Pr(S_{10} > 15) = 3 \cdot 10^{-6} \quad (10)$$

식 (10)에서 알 수 있듯이,  $10^6$ 개 샘플들 중에서 3개 정도만 조건  $S_{10} > 15$ 을 만족한다. 이 샘플들로부터 계산된 조건부기대값은  $E(S_{10} | S_{10} > 15) = 15.5370$ 이고 신뢰수준이 95%인 신뢰구간은  $(-2.4105, 33.4846)$ 이다.

Gibbs샘플러를 사용하면, 훨씬 적은 샘플들을 발생시켜 이 조건부기대값을 계산할 수 있다. 변인기간을  $b = 300$ 으로 하고, 샘플수가  $n = 1000$ 인 Gibbs샘플러에 의한 조건부기대값은  $E(S_{10} | S_{10} > 15) = 15.9573$ 이고 신뢰수준 95%의 신뢰구간은  $(15.8844, 16.0301)$ 이다.

단순몬테카를로법에서는 샘플수가  $n = 10^6$ 이고, Gibbs샘플러에서는 샘플수가  $n = 10^3$ 이다. 그럼에도 불구하고 Gibbs샘플러에 의한 신뢰구간이 단순몬테카를로법에 의한 신뢰구간보다 훨씬 짧다. ■

**예제 5.1.7** 이 예제는 Casella & George [24]에서 인용한 것으로, 베타이항확률분포에서

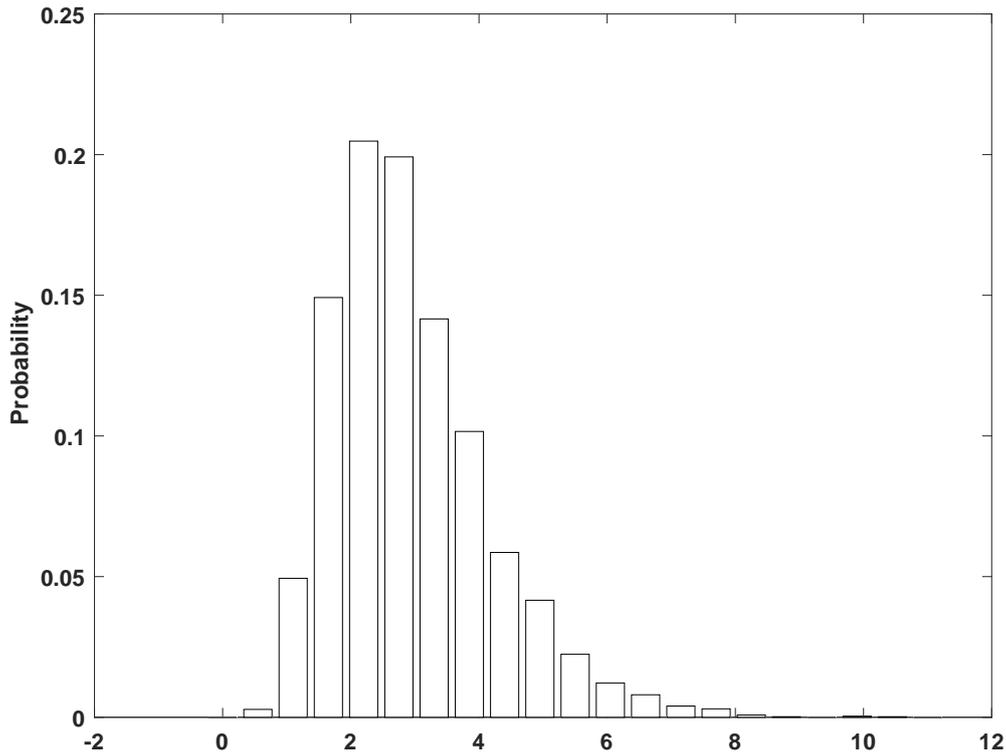


그림 5.1.7. Gibbs 샘플러와 조건부기대값

주변확률밀도함수를 추정하기 위한 것이다.

확률벡터  $[x, y]^t$  의 결합확률밀도함수  $f(x, y)$  가 다음과 같다고 하자.

$$f(x, y) \propto \binom{m}{x} y^{x+\alpha-1} [1-y]^{m-x+\beta-1}, \quad (x = 0, 1, \dots, m, 0 \leq y \leq 1) \quad (1)$$

다음 식들이 성립한다.

$$y|x \stackrel{d}{\sim} \text{Beta}(x + \alpha, m - x + \beta) \quad (2)$$

$$x|y \stackrel{d}{\sim} \text{Binomial}(m, y) \quad (3)$$

식 (2)와 식 (3)을 바탕으로 하는 다음 식들을 반복적용해서, 관찰벡터열  $\{[x^{(j)}, y^{(j)}]^t\}$  를 생성하기로 하자.

$$y^{(j+1)} \sim \text{Beta}(x^{(j)} + \alpha, m - x^{(j)} + \beta) \quad (4)$$

$$x^{(j+1)} \sim \text{Binomial}(m, y^{(j)}) \quad (5)$$

지금부터 확률변수  $x$  의 주변확률밀도함수  $f(x)$  를 구해보자. 식 (1)에서 알 수 있듯이,

다음 식이 성립한다.

$$f(x) \propto \binom{m}{x} \text{Beta}(x + \alpha, m - x + \beta), \quad (x = 0, 1, \dots, m) \quad (6)$$

다음 식이 성립함을 증명할 수 있다.

$$f(x) = \binom{m}{x} \frac{\Gamma(\alpha + \beta) \Gamma(x + \alpha) \Gamma(m - x + \beta)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(\alpha + \beta + m)}, \quad (x = 0, 1, \dots, m) \quad (7)$$

즉, 확률변수  $x$ 는 베타이항확률분포를 따른다.

Gibbs샘플러를 사용해서, 이 주변확률밀도함수  $f(x)$ 를 구하는 방법을 살펴보자. 다음 식들이 성립한다.

$$f(x) = \int f(x, y) dy = \int f(x|y) f(y) dy \approx \frac{1}{n} \sum_{j=1}^n f\left(x \mid y^{(j)}\right) \quad (8)$$

식 (6)에서  $n$ 은 큰 자연수이다.

식 (8)을 이용해서 주변확률밀도함수  $f(x)$ 를 유도하기 위해서, 다음 MATLAB프로그램 GibbsBetaBinomial101.m을 실행해 보자.

```

1 % -----
2 % Filename: GibbsBetaBinomial101.m
3 % Gibbs Sampler 1 for Beta-Binomial Distribution
4 % Casella & George (1992)
5 % Programmed by CBS
6 % -----
7 clear, close all, clf
8 % Problem set up
9 nk = 13000;
10 k = 3000; % burn-in number
11 nn = nk-k; % number of observations to generate
12 al = 3; % alpha
13 be = 4; % beta
14 x = zeros(1, nk);
15 y = zeros(1, nk);
16 m = 20;
17 % Gibb's sampling
18 x(1) = binornd(m, 0.5, 1, 1);
19 y(1) = betarnd(x(1)+al, m-x(1)+be, 1, 1);
20 for ii=2:nk
21     x(ii)=binornd(m,y(ii-1),1,1);
22     y(ii)=betarnd(x(ii)+al, m-x(ii)+be, 1, 1);
23 end
24 % Find the marginal pdf f(x) using Gibbs
25 fhat = zeros(1, m+1);
26 for ii=1:m+1
27     fhat(ii) = mean(binopdf(ii-1, m, y(k+1:nk)));
28 end
29 fhat

```

```

30 % Find the true pdf
31 for ii = 1:m+1
32     xi = ii-1;
33     dumf = log(gamma(m+1)) - log(gamma(xi+1)) - log(gamma(m-xi+1)) ...
34           + log(gamma(al+be)) - log(gamma(al)) - log(gamma(be)) ...
35           + log(gamma(xi+al)) + log(gamma(m-xi+be)) - log(gamma(al+be+m));
36     f(ii) = exp(dumf);
37 end % We can use the 'factorial' function
38 f
39 error = max(abs(f-fhat))
40 % Plotting
41 hold on
42 bar(0:1:m,fhat,0.2,'y')
43 set(gca,'fontsize',11,'fontweight','bold')
44 plot(0:1:m,f,'r*','linewidth',2)
45 legend('\bf Gibbs pdf','\bf True pdf','location','NE')
46 axis([-0.5 m+0.5 0 0.1 ])
47 plot(0:1:m,f,'r-','linewidth',2)
48 hold off
49 saveas(gcf,'GibbsBetaBinomial101','eps')
50 % End of program
51 % -----

```

MATLAB 프로그램 GibbsBetaBinomial101.m을 실행하면, 식 (4)와 식 (5)를 이용해서 Gibbs샘플들  $\{[x^{(j)}, y^{(j)}]^t \mid j = 1, 2, \dots, 10000\}$ 를 발생시킨다. 여기서 사용된 번인기간은  $b = 3000$ 이다. 식 (8)을 사용해서 구한 주변확률밀도함수의 추정값  $\hat{f}(x)$ 와 진짜주변확률밀도함수  $f(x)$ 는 다음 식을 만족한다.

$$\max_x |f(x) - \hat{f}(x)| = 0.0028 \quad (9)$$

즉, Gibbs샘플러에 의해 추정된 주변확률밀도함수  $\hat{f}(x)$ 는 진짜주변확률밀도함수  $f(x)$ 와 아주 가까움을 알 수 있다. MATLAB 프로그램 GibbsBetaBinomial101.m을 실행하면, 이 두 주변확률밀도함수들을 그린 그림 5.1.8을 출력한다. ■

**예제 5.1.8** 예제 5.1.7에서 다른 베타이항확률분포를 다시 생각해 보자. Gibbs샘플들  $\{[x^{(j)}, y^{(j)}]^t\}$ 로부터 도출한 확률변수  $x$ 의 히스토그램과 예제 5.1.7의 식 (7)에 주어진 확률변수  $x$ 의 주변확률밀도함수  $f(x)$ 에서 발생시킨 난수들의 히스토그램을 비교해보자.

이 히스토그램들을 그리기 위해서, 다음 MATLAB 프로그램 GibbsBetaBinomial102.m을 실행해 보자.

```

1 % -----
2 % Filename: GibbsBetaBinomial102.m
3 % Gibbs Sampler 2 for Beta-Binomial Distribution
4 % Programmed by CBS
5 % -----

```

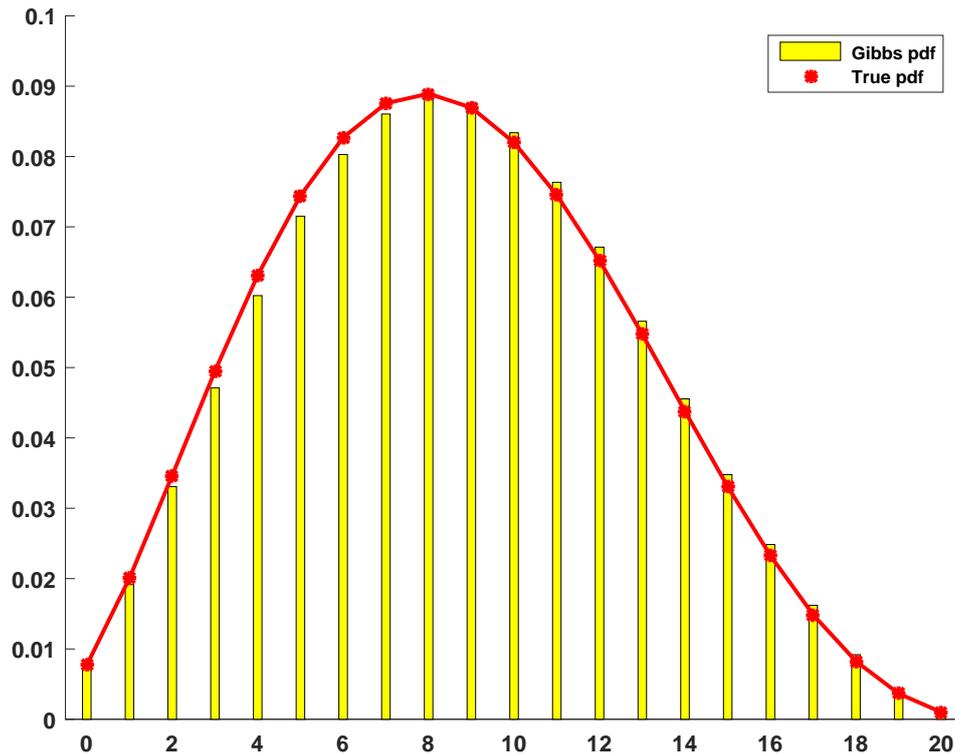


그림 5.1.8. Gibbs샘플러와 베타이항확률분포

```

6 function GibbsBetaBinomial102
7 clear, close all, clf
8 % Problem set up
9 nk = 13000;
10 k = 3000; % burn-in number
11 nn = nk-k; % number of observations to generate
12 al = 3; % alpha
13 be = 4; % beta
14 x = zeros(1,nk);
15 y = zeros(1,nk);
16 m = 20;
17 % Gibb's sampling
18 x(1) = binornd(m, 0.5, 1,1);
19 y(1) = betarnd(x(1)+al, m-x(1)+be, 1, 1);
20 for ii=2:nk
21     x(ii)=binornd(m,y(ii-1),1,1);
22     y(ii)=betarnd(x(ii)+al, m-x(ii)+be, 1, 1);
23 end
24 % Find the marginal pdf f(x) using MCMC
25 fhat = zeros(1,m+1);
26 for ii=1:m+1
27     fhat(ii) = mean(binopdf(ii-1,m,y(k+1:nk)));
28 end
29 fhat
30 % Random sampling from true pdf
31 for ii = 1:m+1
32     xi = ii-1;
33     dumf = log(gamma(m+1)) - log(gamma(xi+1)) - log(gamma(m-xi+1)) ...
34           + log(gamma(al+be)) - log(gamma(al)) - log(gamma(be)) ...
35           + log(gamma(xi+al)) + log(gamma(m-xi+be)) - log(gamma(al+be+m));
36     fpdf(ii) = exp(dumf);

```

```

37 end
38 TrueRS = GenerateDiscreteRS(0:1:m,fpdf,nn);
39 % Histogram of the random samples
40 [ Nbar,zcenter ] = hist(TrueRS,0:m);
41 fTrueRS = Nbar/nn
42 hold on
43 bar(0-0.1:1:m-0.1,fhat,0.3,'k')
44 set(gca,'fontsize',11,'fontweigh','bold')
45 bar(0.2:1:(m+0.2),fTrueRS,0.3,'w')
46 legend('\bf Gibbs histogram','\bf True histogram');
47 axis( [ -2 m+2 0 1.2*max(max(fhat),max(fTrueRS)) ] )
48 ylabel('Relative Frequency')
49 hold off
50 saveas(gcf,'GibbsBetaBinomial102','eps')
51 end
52 % End of program
53 % -----
54 % Filename: GenerateDiscreteRS.m
55 % Generate Discrete Random Samples
56 % -----
57 function RS = GenerateDiscreteRS(x,p,nrs)
58 % x = Support set (i.e., domain) of the distribution
59 % p = Probability density function
60 % nrs = Number of the generated random numbers
61 % (Ex) RS = GenerateDiscreteRS([1 3 5],[0.2 0.5 0.3],1200)
62 cumdf = cumsum(p);
63 RS = zeros(1,nrs);
64 for ii = 1:1:nrs
65     u = rand;
66     idum = 1;
67     while cumdf(idum) <= u
68         idum = idum+1;
69     end;
70     RS(ii) = x(idum);
71 end
72 end
73 % End of the function
74 % -----

```

MATLAB 프로그램 GibbsBetaBinomial102.m을 실행하면, Gibbs샘플들  $\{[x^{(j)}, y^{(j)}]^t \mid j = 1, 2, \dots, 10000\}$ 를 발생시킨다. 확률변수  $x$ 의 주변확률밀도함수  $f(x)$ 에서 난수들을 발생시키기 위해서, MATLAB 프로그램 GenerateDiscreteRS.m을 사용하였다.

MATLAB 프로그램 GibbsBetaBinomial102.m을 실행하면, 이 난수들의 히스토그램을 그린 그림 5.1.9이 그려진다. 그림 5.1.9에서 흑색막대는 Gibbs샘플러에 의한 히스토그램이고, 흰색막대는 진짜 주변확률밀도함수에서 발생된 샘플들의 히스토그램이다. 그림 5.1.9에서 볼 수 있듯이, 두 히스토그램들 사이에는 큰 차이가 없다. ■

예제 5.1.7에서는 조건부확률분포들로부터 결합확률분포를 유도하였다. 지금부터는 이러한 과정을 좀 더 자세히 살펴보자. 확률벡터  $[x, y]^t$ 의 결합확률밀도함수를  $f(x, y)$ , 조건

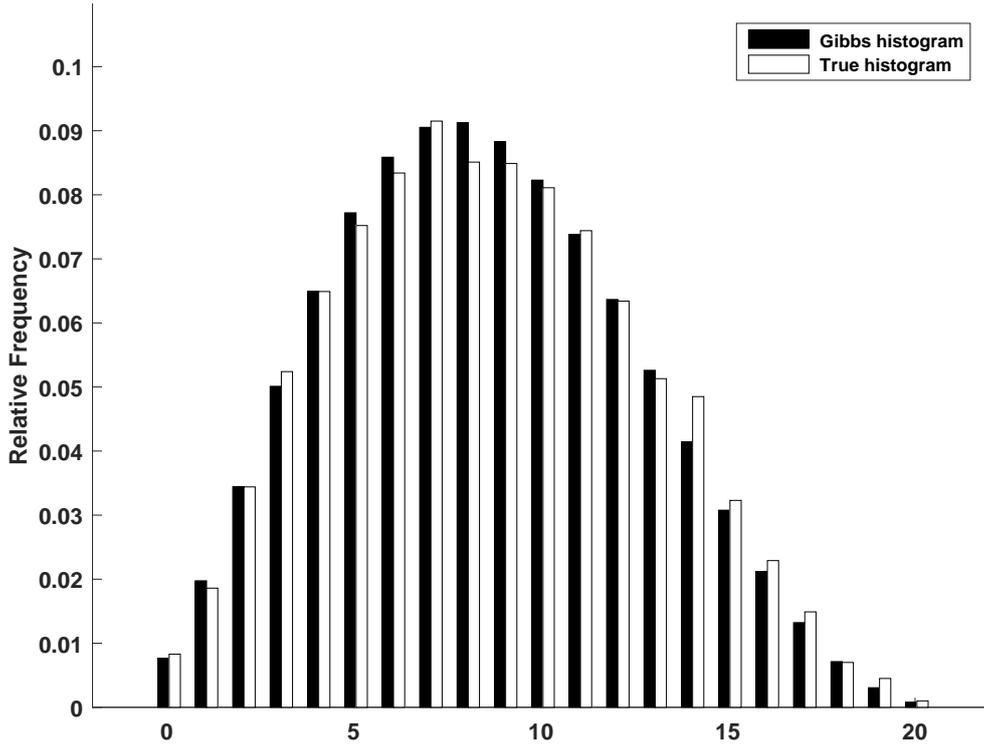


그림 5.1.9. 베타이항확률분포의 히스토그램

부확률밀도함수들을  $f_{X|Y}(x|y)$ 와  $f_{Y|X}(y|x)$ 라 하자. 다음 식들이 성립한다.

$$f_X(x) = \int f_{X|Y}(x|y)f_Y(y)dy = \int f_{X|Y}(x|y) \int f_{Y|X}(y|z)f_X(z)dzdy \quad (5.1.5)$$

여기서  $f_X(x)$ 와  $f_Y(y)$ 는 주변확률밀도함수들이다. 식 (5.1.5)를 다음과 같이 쓸 수 있다.

$$f_X(x) = \int K(x, z)f_X(z)dz \quad (5.1.6)$$

여기서  $K(x, z)$ 는 다음과 같다.

$$K(x, z) \doteq \int f_{X|Y}(x|y)f_{Y|X}(y|z)dy \quad (5.1.7)$$

즉,  $K(x, z)$ 은 추이핵이다. 식 (5.1.6)은 해가  $f_X(x)$ 인 고정점적분방정식 (fixed point integral equation)이다. Gibbs샘플들을 식 (5.1.5)에 적용하면, 다음 식을 얻는다.

$$\begin{aligned} & f_X(x^{(j+1)}) \\ &= \int \left[ \int f_{Y|X}(y^{(j+1)}|x^{(j)}) f_{X|Y}(x^{(j+1)}|y^{(j+1)}) dy^{(j+1)} \right] f_X(x^{(j)}) dx^{(j)} \end{aligned} \quad (5.1.8)$$

Gibbs샘플들을 식 (5.1.7)에 적용하면, 다음 식이 성립함을 알 수 있다.

$$K(x^{(j)}, x^{(j+1)}) = \int f_{Y|X}(y^{(j+1)} | x^{(j)}) f_{X|Y}(x^{(j+1)} | y^{(j+1)}) dy^{(j+1)} \quad (5.1.9)$$

식 (5.1.9)를 사용하면, 식 (5.1.8)을 다음과 같이 쓸 수 있다.

$$f_X(x^{(j+1)}) = \int K(x^{(j)}, x^{(j+1)}) f_X(x^{(j)}) dx^{(j)} \quad (5.1.10)$$

만약 Gibbs샘플러에 의해서 발생된 Gibbs열  $\{x^{(j)}\}$ 가 수렴한다면, 식 (5.1.10)에서 알 수 있듯이 이 Gibbs샘플열의 불변분포는 고정점적분방정식 (5.1.6)을 만족하는 확률밀도함수  $f_X(x)$ 이다.

**예제 5.1.9** 확률벡터  $[x, y]^t$ 가 다음과 같은 조건부확률밀도함수들을 갖는다고 하자.

$$f(x|y) = y \exp(-yx) 1_{(0, \infty)}(x) \quad (1)$$

$$f(y|x) = x \exp(-xy) 1_{(0, \infty)}(y) \quad (2)$$

식 (5.1.5)에서 알 수 있듯이, 주변확률밀도함수는 다음 식을 만족한다.

$$f_X(x) = \int_0^\infty \left[ \int_0^\infty y \exp(-yx) z \exp(-zy) dy \right] f_X(z) dz \quad (3)$$

즉, 다음 식이 성립한다.

$$f_X(x) = \int_0^\infty \frac{z}{[x+z]^2} f_X(z) dz \quad (4)$$

다음 식이 성립함을 쉽게 증명할 수 있다.

$$\frac{1}{x} = \int_0^\infty \frac{z}{[x+z]^2} \frac{1}{z} dz \quad (5)$$

즉,  $f_X(x) = x^{-1}$ 은 고정점적분방정식 (4)의 해이다. 그러나, 이  $f_X(x)$ 는 확률밀도함수가 아니다. 따라서, 식 (1)과 식 (2)의 조건부확률밀도함수들에 Gibbs샘플러를 적용해서 발생한 Gibbs샘플열은 수렴하지 않는다. 또한, 이 Gibbs샘플열의 히스토그램은 함수  $f_X(x) = x^{-1}$ 에 근사하지도 않는다.

지금까지 내용을 확인하기 위해서 다음 R프로그램 FixedPointIntegralEquation10R.R을

실행해 보자.

```

1 # -----
2 # Filename: FixedPointIntegralEquation101R.R
3 # Fixed Point Integral Equationhelp(dexp)
4 # Programmed by CBS
5 # -----
6 # Gibbs Sampler Function
7 GibbsSampler <- function (Nobs)
8 {
9     zmat <- matrix(ncol=2, nrow=Nobs)
10    set.seed(1)
11    x <- 1
12    y <- 1
13    zmat[1, ] <- c(x, y)
14    for (ii in 2:Nobs) {
15        x <- rexp(1,rate=y) # mean = 1/y
16        y <- rexp(1,rate=x) # mean = 1/x
17        zmat[ii, ] <- c(x, y)
18    }
19    zmat
20 }
21
22
23 # Generating Gibbs Samples
24 Nobs = 1024
25 burnin <- 9000
26 dum <- GibbsSampler(burnin+Nobs)
27 GibbsRN <- dum[burnin+1:burnin+Nobs,]
28 GibbsRN
29 # Plotting
30 library(ggplot2)
31 # install.packages("grid")
32 library(grid)
33 setEPS()
34 plot.new()
35 # postscript('FixedPointIntegralEquation101R.eps') # Start to save figure
36 timee = c(burnin+(1:Nobs))
37 GibbsData <- data.frame(timee,GibbsRN)
38 plot1 <- ggplot(GibbsData, aes(timee,GibbsRN[,1])) +
39     geom_line(colour="dark red") + xlab("Time") + ylab("x")
40 plot2 <- ggplot(GibbsData, aes(timee,GibbsRN[,2])) +
41     geom_line(colour="dark blue") + xlab("Time") + ylab("y")
42 pushViewport(viewport(layout = grid.layout(2,1)))
43 print(plot1, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
44 print(plot2, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
45 # dev.off() # End to save figure
46 # -----

```

R프로그램 FixedPointIntegralEquation10R.R을 실행하면, 이 Gibbs샘플러의 시계열산점도들이 그려진다. 그림 5.1.10에서 상단 그래프는 생성된  $\{x_t | t = 9001, 9002, \dots, 10024\}$ 의 시계열산점도이고, 하단 그래프는 생성된  $\{y_t | t = 9001, 9002, \dots, 10024\}$ 의 시계열산점도이다. 이 산점도들에서 알 수 있듯이, Gibbs샘플러들은 수렴하지 않는다.

다음 Python프로그램 FixedPointIntegralEquation101.py를 실행해도 R프로그램 FixedPointIntegralEquation10R.R과 동일한 결과를 얻는다.

```

1 # FixedPointIntegralEquation101Py.py
2 """
3 Filename : FixedPointIntegralEquation101Py.py
4 Fixed Point Integral Equationhelp(dexp)
5 Programmed by Chang-Yong Han based on the R program by CBS
6 """
7
8 import matplotlib.pyplot as plt
9
10 # Gibbs Sampler Function
11 def GibbsSampler(Nobs):
12     import numpy as np
13     np.random.seed(10)
14     zmat = np.zeros([Nobs, 2])
15     x, y = 1, 1
16     zmat[0] = [x, y]
17     for i in range(1, Nobs):
18         x, y = np.random.exponential(scale = (1/y)), np.random.exponential(scale
19 = (1/x))
20         zmat[i] = [x, y]
21     return(zmat)
22
23 # Generating Gibbs Samples
24 Nobs = 1024
25 burnin = 9000
26 dum = GibbsSampler(burnin + Nobs)
27 GibbsRN = dum[burnin:]
28
29 # Plot
30 time = list(range(burnin, burnin + Nobs))
31 f, ax = plt.subplots(2, sharex = True)
32 ax[0].yaxis.set_major_formatter(plt.FuncFormatter('{:.1e}'.format))
33 ax[0].plot(time, GibbsRN.T[0], color = 'firebrick')
34 ax[0].set(ylabel = 'x')
35 ax[1].yaxis.set_major_formatter(plt.FuncFormatter('{:.1e}'.format))
36 ax[1].plot(time, GibbsRN.T[1], color = 'navy')
37 ax[1].set(ylabel = 'y')
38 ax[1].set(xlabel = 'Time')
39 #plt.show()
40 f.savefig('FixedPointIntegralEquation101Py.png', bbox_inches = 'tight')
41
42 # End of program
43 # -----

```

■

**예제 5.1.10** 몬테카를로법을 써서 정적분을 구하는 것은 근본적으로 대수법칙을 이용하는 것이다. 따라서, 샘플수  $n$ 이 유한이면, 계산된 적분값에는 오차가 따르기 마련이다. 이러한 오차를 분산으로 표현할 수 있다. 분산을 작게 할수록 좋은 추정값이다.

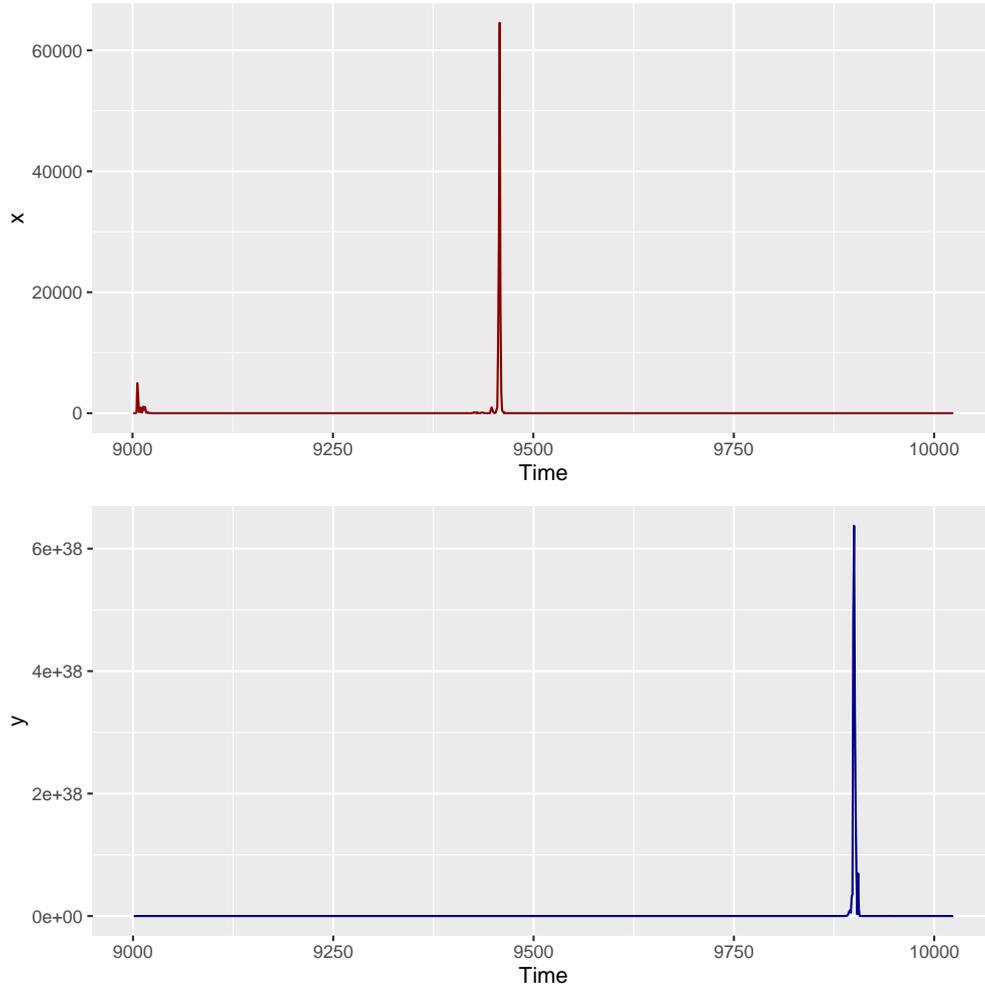


그림 5.1.10. 고정점적분방정식의 Gibbs샘플열

Rao-Blackwell정리에 의하면, 다음 식들이 성립한다.

$$E(g(x)) = E(E(g(x)|y)) \quad (1)$$

$$\text{Var}(E(g(x)|y)) \leq \text{Var}(g(x)) \quad (2)$$

즉, 기대  $E(g(x))$ 를 계산하기 위해서는 확률변수  $x$ 의 주변확률분포를 사용하는 것보다는 조건부확률변수  $x|y$ 의 조건부확률분포와 확률변수  $y$ 의 주변확률분포를 사용함으로써 분산을 줄일 수 있다.

예제 5.1.7에서 다른 확률벡터  $[x, y]^t$ 의 결합확률밀도함수  $f(x, y)$ 는 다음과 같다.

$$f(x, y) \propto \binom{m}{x} y^{x+\alpha-1} [1-y]^{m-x+\beta-1}, \quad (x = 0, 1, \dots, m, 0 \leq y \leq 1) \quad (3)$$

다음 식이 성립한다.

$$x|y \stackrel{d}{\sim} \text{Binomial}(m, y) \quad (4)$$

확률변수  $y$ 의 주변확률밀도함수  $f(y)$ 를 구해보자. 식 (3)에서 알 수 있듯이, 다음 식이 성립한다.

$$f(y) \propto y^{\alpha-1} [1-y]^{\beta-1} \sum_{x=0}^m \binom{m}{x} y^x [1-y]^{m-x}, \quad (0 \leq y \leq 1) \quad (5)$$

즉, 다음 식이 성립한다.

$$f(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} [1-y]^{\beta-1}, \quad (0 \leq y \leq 1) \quad (6)$$

따라서, 확률변수  $y$ 는 베타확률분포를 따른다.

Rao-Blackwell정리를 사용하면 몬테카를로법으로 계산한 기대값의 분산을 감소시킬 수 있음을 보이기 위해서, 다음 MATLAB 프로그램 RaoBlackwellization101.m을 실행해 보자.

```

1 % -----
2 % Filename: RaoBlackwellization101.m
3 % Reducing the Variance of a Monte Carlo Integration
4 % Programmed by CBS
5 % -----
6 function RaoBlackwellization101
7 clear, close all, format long
8 rng((1001^2+1)/2, 'twister')
9 % Problem set up
10 nk = 13000; k = 3000; nn = nk-k;
11 al = 3; be = 4;
12 x = zeros(1,nk); y = zeros(1,nk);
13 m = 20;
14 % Gibb's sampling
15 x(1) = binornd(m, 0.5, 1,1);
16 y(1) = betarnd(x(1)+al, m-x(1)+be, 1, 1);
17 for ii=2:nk
18     x(ii) = binornd(m,y(ii-1),1,1);
19     y(ii) = betarnd(x(ii)+al, m-x(ii)+be, 1, 1);
20 end
21 % Find the marginal pdf f(x)
22 fhat = zeros(1,m+1);
23 for ii=1:m+1
24     fhat(ii) = mean(binopdf(ii-1,m,y(k+1:nk)));
25 end
26 % Rao-Blackwellization
27 nx = 1024, ny = 512
28 for kk=1:1:100
29     % Mean and Variance by Single Expectation using Gibbs sampler
30     SEcondiRS = GenerateDiscreteRS(0:1:m,fhat,nx);
31     SEmean(kk) = mean(SEcondiRS);

```

```

32     SEvariance(kk) = var(SEcondiRS);
33 % Mean and Variance by Double Expectation
34     for iy = 1:ny
35         ydum = betarnd(al,be,1,1);
36         xdum = binornd(m,ydum,1,nx);
37         xmeanD(iy) = mean(xdum);
38     end
39     DEmean(kk) = mean(xmeanD);
40     DEvariance(kk) = var(xmeanD);
41 end
42 % Plotting
43 subplot(2,1,1)
44 plot(1:1:100,SEmean,'k-',1:1:100,DEmean,'r--','linewidth',2)
45 set(gca,'fontsize',11,'fontweigh','bold','ylim',[8 10])
46 legend('\bf Single Mean','\bf Double Mean','location','NE')
47 subplot(2,1,2)
48 plot(1:1:100,SEvariance,'k-',1:1:100,DEvariance,'r--','linewidth',2)
49 set(gca,'fontsize',11,'fontweigh','bold','ylim',[5 20])
50 legend('\bf Single Variance','\bf Double Variance','location','SE')
51 saveas(gcf,'RaoBlackwellization101','eps')
52 end % End of function RaoBlackwellization101
53 % End of program
54 % -----
55 % Filename: GenerateDiscreteRS.m
56 % Generate Discrete Random Samples
57 % -----
58 function RS = GenerateDiscreteRS(x,p,nrs)
59 % x = Support set (i.e., domain) of the distribution
60 % p = Probability density function
61 % nrs = Number of the generated random numbers
62 % (Ex) RS = GenerateDiscreteRS([1 3 5],[0.2 0.5 0.3],1200)
63 cumdf = cumsum(p);
64 RS = zeros(1,nrs);
65 for ii = 1:1:nrs
66     u = rand;
67     idum = 1;
68     while cumdf(idum) <= u
69         idum = idum+1;
70     end;
71     RS(ii) = x(idum);
72 end
73 end % End of function GenerateDiscreteRS
74 % End of the function
75 % -----

```

이 MATLAB 프로그램 RaoBlackwellization101.m을 실행하면, 그림 5.1.11에 시계열산점도들이 그려진다. 그림 5.1.11의 상단 그래프에서 흑색 실선은 Gibbs샘플러를 적용해서 추정된 주변확률밀도함수  $f(x)$ 로부터 계산한 100개 표본평균들을 그린 것이고, 적색 긴점선은 식 (4)의 조건부확률밀도함수  $f(x|y)$ 와 식 (6)의 주변확률밀도함수  $f(y)$ 로부터 계산한 100개 표본평균들을 그린 것이다. 이 상단 그래프에서 알 수 있듯이, 두 표본평균들은 큰 차이가 없다. 또한, 그림 5.1.11의 하단 그래프에는 이들에 해당하는 표본분산들을 그린 것이다. 즉, 흑색 실선은 주변확률밀도함수  $f(x)$ 로부터 계산한 100개 표본분산들을 그린 것이고, 적색 긴점선은 조건부확률밀도함수  $f(x|y)$ 와 주변확률밀도함수  $f(y)$ 로부터 계산한 100개 표본분산들을

그런 것이다. 이 하단 그래프에서 알 수 있듯이, 조건부기대값을 이용한 표본분산이 그렇지 않은 표본분산에 비해 작다. 즉, 식 (2)가 성립함을 확인할 수 있다. ■

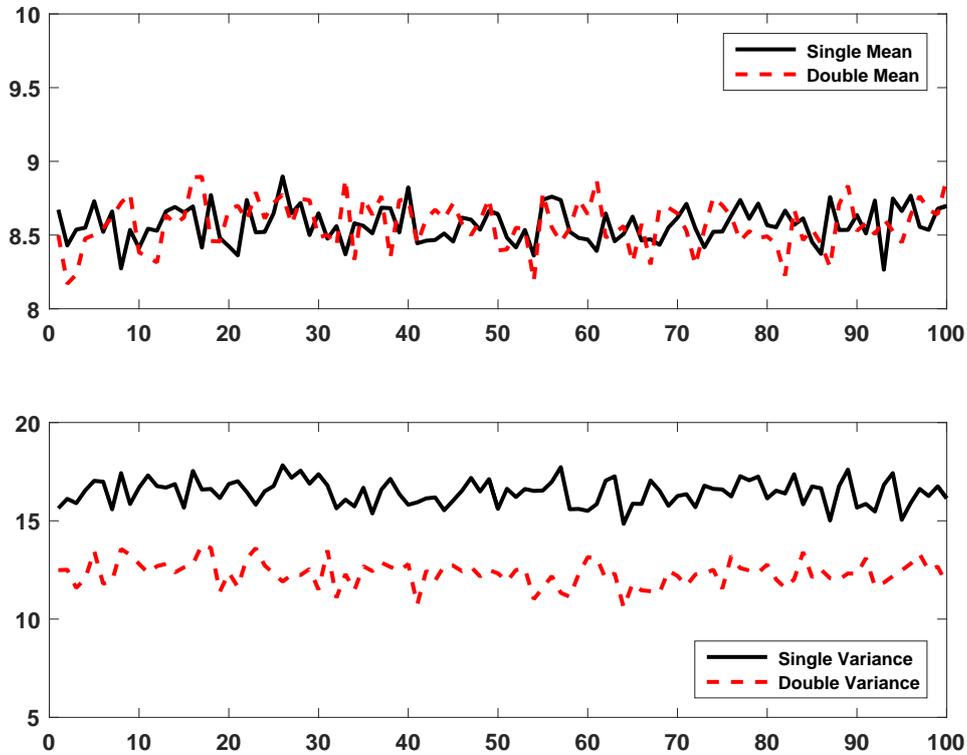


그림 5.1.11. Rao-Blackwellization

## 제 5.2절 Gibbs샘플러의 수렴성

알고리즘 5.1.1이나 알고리즘 5.1.2에 의해 발생된 Gibbs샘플들에 대해서 다음과 같은 의문이 생긴다.

**첫째,** Gibbs샘플러를 사용하면, 조건부확률분포들로부터 결합확률분포를 일의적으로 발생시킬 수 있는가?

**둘째,** Gibbs샘플들은 Markov연쇄를 이루는가?

**셋째,** Gibbs샘플들은 수렴하는가?

**넷째,** Gibbs샘플들은 원하는 불변분포로 수렴하는가?

이 절에서는 이러한 의문들에 대해서 생각해 보자.

**예제 5.2.1** Gibbs샘플러의 수렴성에 대해서 Casella & George [24]가 제시한 간단한 예제를 살펴보자.

다음과 같은 결합확률질량함수를 갖는 확률벡터  $[x, y]^t$  를 살펴보자.

$$f_{x,y}(0, 0) \doteq \Pr(x = 0, y = 0) = p_1 \quad (1)$$

$$f_{x,y}(1, 0) \doteq \Pr(x = 1, y = 0) = p_2 \quad (2)$$

$$f_{x,y}(0, 1) \doteq \Pr(x = 0, y = 1) = p_3 \quad (3)$$

$$f_{x,y}(1, 1) \doteq \Pr(x = 1, y = 1) = p_4 \quad (4)$$

여기서 다음 식들이 성립한다.

$$p_i \geq 0, \quad \sum_{i=1}^4 p_i = 1 \quad (5)$$

이 결합확률질량함수의 주변확률질량함수들은 다음과 같다.

$$\mathbf{f}_x \doteq [f_x(0) \quad f_x(1)] = [p_1 + p_3 \quad p_2 + p_4] \quad (6)$$

$$\mathbf{f}_y \doteq [f_y(0) \quad f_y(1)] = [p_1 + p_2 \quad p_3 + p_4] \quad (7)$$

이 주변확률질량함수들은 각각 성공확률이  $p_2 + p_4$ 와  $p_3 + p_4$ 인 Bernouille 확률분포를 나타낸다. 또한,  $x$ 가 주어진 조건 하에서 확률변수  $y$ 의 조건부확률질량함수를 다음과 같은 행렬로 나타낼 수 있다.

$$A_{y|x} \doteq \Pr(y|x) = \begin{bmatrix} \frac{p_1}{p_1+p_3} & \frac{p_3}{p_1+p_3} \\ \frac{p_2}{p_2+p_4} & \frac{p_4}{p_2+p_4} \end{bmatrix} \quad (8)$$

또한,  $y$ 가 주어진 조건 하에서 확률변수  $x$ 의 조건부확률질량함수를 다음과 같은 행렬로 나타낼 수 있다.

$$A_{x|y} \doteq \Pr(x|y) = \begin{bmatrix} \frac{p_1}{p_1+p_2} & \frac{p_2}{p_1+p_2} \\ \frac{p_3}{p_3+p_4} & \frac{p_4}{p_3+p_4} \end{bmatrix} \quad (9)$$

이  $A_{y|x}$ 는  $x$ 에서  $y$ 로 가는 추이확률행렬, 그리고  $A_{x|y}$ 는  $y$ 에서  $x$ 로 가는 추이확률행렬이다.

초기벡터를  $[x^{(0)} \quad y^{(0)}]^t$ 라 하고, 추이확률행렬들  $A_{y|x}$ 와  $A_{x|y}$ 를 알고리즘 5.1.1의 체계

적절한 Gibbs 샘플러에 적용해서, 다음과 같이 Gibbs 샘플들을 발생시킨다고 하자.

$$x^{(1)} \rightarrow y^{(1)} \rightarrow x^{(2)} \rightarrow y^{(2)} \rightarrow x^{(3)} \rightarrow y^{(3)} \rightarrow \dots \quad (10)$$

다음 식이 성립함을 자명하다.

$$\begin{aligned} & \Pr \left( x^{(1)} = a^{(1)} \mid x^{(0)} = a^{(0)} \right) \\ &= \sum_{y=0}^1 \Pr \left( x^{(1)} = a^{(1)} \mid y^{(1)} = y \right) \Pr \left( y^{(1)} = y \mid x^{(0)} = a^{(0)} \right) \end{aligned} \quad (11)$$

식 (11)에서 알 수 있듯이,  $x^{(0)}$ 에서  $x^{(1)}$ 으로 가는 추이확률행렬  $A_{x^{(1)}|x^{(0)}}$ 는 다음과 같다.

$$A_{x^{(1)}|x^{(0)}} = A_{y|x^{(0)}} A_{x^{(1)}|y} \quad (12)$$

식 (12)의 우변은 시점에 의존하지 않으므로,  $A_{x^{(1)}|x^{(0)}}$ 를  $A_{x|x}$ 로 표기할 수 있다. 식 (11)과 식 (12)에서 알 수 있듯이, 다음 식이 성립한다.

$$\Pr(x^{(k)} = j \mid x^{(0)} = i) = A_{x|x}^k(i, j) \quad (13)$$

즉, 시점 0에서 상태  $i$ 가 시점  $k$ 에서 상태  $j$ 로 추이하는 확률은 행렬  $A_{x|x}^k$ 의  $(i, j)$  원소이다.

다음과 같이  $x^{(k)}$ 의 주변확률질량함수를 정의하자.

$$\mathbf{f}_x^{(k)} \doteq \left[ f_x^{(k)}(0) \ f_x^{(k)}(1) \right] \quad (14)$$

다음 식들이 성립한다.

$$\mathbf{f}_x^{(k)} = \mathbf{f}_x A_{x|x}^k = \left[ \mathbf{f}_x A_{x|x}^{k-1} \right] A_{x|x} = \mathbf{f}_x^{(k-1)} A_{x|x} \quad (15)$$

만약 추이확률행렬  $A_{x|x}$ 의 각 원소가 양수이면, 임의의 초기벡터  $\mathbf{f}_x$ 에 대해서 다음 식들을 만족하는 확률질량함수  $\mathbf{f}_x^{(\infty)}$ 가 일의적으로 존재한다.

$$\lim_{k \rightarrow \infty} \mathbf{f}_x^{(k)} = \mathbf{f}_x^{(\infty)} \quad (16)$$

$$\mathbf{f}_x^{(\infty)} = \mathbf{f}_x^{(\infty)} A_{x|x} \quad (17)$$

만약 식 (10)의 Gibbs샘플들이 수렴하면, 식 (17)을 만족하는  $\mathbf{f}_x^{(\infty)}$ 는 확률변수  $x$ 의 주변확률질량함수이어야 한다. 다음 식들이 성립함을 명백하다.

$$\begin{aligned}
 \mathbf{f}_x A_{x|x} &= \mathbf{f}_x A_{y|x} A_{x|y} \\
 &= [p_1 + p_3, p_2 + p_4] \begin{bmatrix} \frac{p_1}{p_1+p_3} & \frac{p_3}{p_1+p_3} \\ \frac{p_2}{p_2+p_4} & \frac{p_4}{p_2+p_4} \end{bmatrix} A_{x|y} \\
 &= [p_1 + p_2, p_3 + p_4] \begin{bmatrix} \frac{p_1}{p_1+p_2} & \frac{p_2}{p_1+p_2} \\ \frac{p_3}{p_3+p_4} & \frac{p_4}{p_3+p_4} \end{bmatrix} \\
 &= [p_1 + p_3, p_2 + p_4] = \mathbf{f}_x \tag{18}
 \end{aligned}$$

여기서 첫 번째 등호는 식 (12)에 의해서, 두 번째 등호는 식 (6)과 식 (8)에 의해서, 세 번째 등호는 식 (9)에 의해서, 그리고 다섯 번째 등호는 식 (6)에 의해서 성립한다. 식 (18)과 확률질량함수  $\mathbf{f}_x^{(\infty)}$ 의 일의성에서 알 수 있듯이,  $\mathbf{f}_x^{(\infty)}$ 와  $\mathbf{f}_x$ 는 같다. 즉,  $k$ 가 커짐에 따라  $x^{(k)}$ 의 확률질량함수는  $\mathbf{f}_x$ 에 수렴한다. 따라서, 아주 큰  $k$ 에서 Gibbs샘플링을 멈추면,  $\mathbf{f}_x$ 는  $x^{(k)}$ 의 근사적 확률질량함수라고 할 수 있다. 특히  $k$ 가 크면 클수록,  $\mathbf{f}_x$ 는 좀 더 좋은 근사함수가 될 것이다. ■

좀 더 일반적인 경우를 살펴보기 위해서, 다음과 같이 양적조건(positivity condition)을 정의하자.

#### 정의 5.2.1: 양적조건

결합확률밀도함수  $f(x_1, x_2, \dots, x_p)$ 의 주변확률밀도함수를  $f_i(x_i)$ 라 하자. 다음 명제가 성립하면,  $f(x_1, x_2, \dots, x_p)$ 는 양적조건을 만족한다고 한다.

$$f_i(x_i) > 0, (i = 1, 2, \dots, p) \Rightarrow f(x_1, x_2, \dots, x_p) > 0$$

정의 5.2.1에서 알 수 있듯이, 양적조건을 만족하는 결합확률밀도함수  $f(x_1, x_2, \dots, x_p)$ 의 지지대(support)는 주변확률밀도함수들의 지지대들의 Cartesian 곱이다.

**명제 5.2.1: Hammersley-Clifford 정리**

확률벡터  $\mathbf{x} = [x_1, x_2, \dots, x_p]^t$  의 결합확률밀도함수  $f(x_1, x_2, \dots, x_p)$  가 양적조건을 만족하면, 이 결합확률밀도함수의 지지대에 속하는 임의의  $\boldsymbol{\xi} = [\xi_1, \xi_2, \dots, \xi_p]^t$  에 대해서 다음 식이 성립한다.

$$f(x_1, x_2, \dots, x_p) \propto \prod_{j=1}^p \frac{f(x_j | x_1, \dots, x_{j-1}, \xi_{j+1}, \dots, \xi_p)}{f(\xi_j | x_1, \dots, x_{j-1}, \xi_{j+1}, \dots, \xi_p)}$$

증명. 각  $i (= 1, 2, \dots, p)$  에 대해서, 다음 식들이 성립한다.

$$f(x_1, \dots, x_i, x_{i+1}, \dots, x_p) = f(x_{i+1}, \dots, x_p | x_1, \dots, x_i) f(x_1, \dots, x_i) \quad (1)$$

$$f(x_1, \dots, x_i, \xi_{i+1}, \dots, \xi_p) = f(\xi_{i+1}, \dots, \xi_p | x_1, \dots, x_i) f(x_1, \dots, x_i) \quad (2)$$

식 (1) 과 식 (2) 에서 알 수 있듯이, 각  $i (= 1, 2, \dots, p)$  에 대해서 다음 식이 성립한다.

$$\frac{f(x_1, \dots, x_i, x_{i+1}, \dots, x_p)}{f(x_1, \dots, x_i, \xi_{i+1}, \dots, \xi_p)} = \frac{f(x_{i+1}, \dots, x_p | x_1, \dots, x_i)}{f(\xi_{i+1}, \dots, \xi_p | x_1, \dots, x_i)} \quad (3)$$

식 (3) 을 반복적용하면, 다음 식을 얻는다.

$$f(x_1, x_2, \dots, x_p) = f(\xi_1, \xi_2, \dots, \xi_p) \prod_{j=1}^p \frac{f(x_j | x_1, \dots, x_{j-1}, \xi_{j+1}, \dots, \xi_p)}{f(\xi_j | x_1, \dots, x_{j-1}, \xi_{j+1}, \dots, \xi_p)} \quad (4)$$

■

다음 예제에서 알 수 있듯이, 명제 5.2.1의 Hammersley-Clifford 정리를 이용해서 모든 조건부확률밀도함수들로부터 결합확률밀도함수를 유도할 수 있는 것은 아니다. 특히, 베이즈 안통계분석에서 비정칙 (improper) 사전확률분포를 사용하는 경우에 자주 발생한다.

**예제 5.2.2** 확률변수들  $x_1$  과  $x_2$  가 다음과 같은 조건부확률분포들을 따른다고 가정하자.

$$x_1 | x_2 \stackrel{d}{\sim} \text{Exponential}(\lambda x_2) \quad (1)$$

$$x_2 | x_1 \stackrel{d}{\sim} \text{Exponential}(\lambda x_1) \quad (2)$$

여기서  $\lambda$ 는 양수이다. 다음 식들이 성립한다.

$$\begin{aligned} f(x_1, x_2) &\propto \frac{f(x_1 | \xi_2) f(x_2 | x_1)}{f(\xi_1 | \xi_2) f(\xi_2 | x_1)} \\ &= \frac{\lambda \xi_2 \exp(-\lambda \xi_2 x_1) \cdot \lambda x_1 \exp(-\lambda x_1 x_2)}{\lambda \xi_2 \exp(-\lambda \xi_2 \xi_1) \cdot \lambda x_1 \exp(-\lambda x_1 \xi_2)} \propto \exp(-\lambda x_1 x_2) \end{aligned} \quad (3)$$

다음 식들이 성립함을 쉽게 알 수 있다.

$$\begin{aligned} &\int_0^\infty \int_0^\infty \exp(-\lambda x_1 x_2) dx_1 dx_2 \\ &= \int_0^\infty \left[ \frac{-1}{\lambda x_2} \exp(-\lambda x_1 x_2) \right]_0^\infty dx_2 = \int_0^\infty \frac{1}{\lambda x_2} dx_2 = \infty \end{aligned} \quad (4)$$

식 (3) 과 식 (4) 에서 알 수 있듯이,  $f(x_1, x_2)$  를 결합확률밀도함수로 하는 결합확률분포는 존재하지 않는다. ■

지금부터는 체계적쓰기 Gibbs샘플러의 수렴성에 대해서만 다루기로 하자. 무작위적쓰기 Gibbs샘플러에 대해서도 비슷한 결과를 얻을 수 있다.

먼저 결합확률밀도함수  $f(x_1, x_2, \dots, x_p)$  가 Gibbs샘플러에 의해서 발생된 Markov연쇄의 불변분포인지를 살펴보자. Gibbs샘플러에서 제  $j$  단계 관찰벡터  $\mathbf{x}^{(j)} = [x_1^{(j)}, x_2^{(j)}, \dots, x_p^{(j)}]^t$  가 주어진 조건 하에서 제  $j+1$  단계 관찰벡터  $\mathbf{x}^{(j+1)} = [x_1^{(j+1)}, x_2^{(j+1)}, \dots, x_p^{(j+1)}]^t$  의 조건부 확률밀도함수, 즉 추이핵  $K(\mathbf{x}^{(j)}, \mathbf{x}^{(j+1)})$  은 다음과 같다.

$$K(\mathbf{x}^{(j)}, \mathbf{x}^{(j+1)}) = \prod_{k=1}^p f\left(x_k^{(j+1)} \mid x_1^{(j+1)}, \dots, x_{k-1}^{(j+1)}, x_{k+1}^{(j)}, \dots, x_p^{(j)}\right) \quad (5.2.1)$$

다음 벡터들을 정의하자.

$$\underline{\mathbf{x}}_k^{(j)} \doteq [x_1^{(j)}, \dots, x_{k-2}^{(j)}, x_{k-1}^{(j)}]^t \quad (5.2.2)$$

$$\bar{\mathbf{x}}_k^{(j)} \doteq [x_{k+1}^{(j)}, x_{k+2}^{(j)}, \dots, x_n^{(j)}]^t \quad (5.2.3)$$

식 (5.2.1) 을 다음과 같이 쓸 수 있다.

$$K(\mathbf{x}^{(j)}, \mathbf{x}^{(j+1)}) = \prod_{k=1}^p f\left(x_k^{(j+1)} \mid \underline{\mathbf{x}}_k^{(j+1)}, \bar{\mathbf{x}}_k^{(j)}\right) \quad (5.2.4)$$

다음 식들이 성립한다.

$$\begin{aligned}
& \int f(\mathbf{x}^{(j)}) K(\mathbf{x}^{(j)}, \mathbf{x}^{(j+1)}) d\mathbf{x}^{(j)} \\
&= \int f(\mathbf{x}^{(j)}) \prod_{k=1}^p f(x_k^{(j+1)} \mid \underline{\mathbf{x}}_k^{(j+1)}, \bar{\mathbf{x}}_k^{(j)}) d\mathbf{x}^{(j)} \\
&= \int \cdots \int \left[ \int f(\mathbf{x}^{(j)}) dx_1^{(j)} \right] \prod_{k=1}^p f(x_k^{(j+1)} \mid \underline{\mathbf{x}}_k^{(j+1)}, \bar{\mathbf{x}}_k^{(j)}) \prod_{k=2}^p dx_k^{(j)} \quad (5.2.5)
\end{aligned}$$

식 (5.2.5)의 우변을 다음과 같이 쓸 수 있다.

$$\begin{aligned}
& \int \cdots \int f(\bar{\mathbf{x}}_1^{(j)}) f(x_1^{(j+1)} \mid \bar{\mathbf{x}}_1^{(j)}) \prod_{k=2}^p f(x_k^{(j+1)} \mid \underline{\mathbf{x}}_k^{(j+1)}, \bar{\mathbf{x}}_k^{(j)}) \prod_{k=2}^p dx_k^{(j)} \\
&= \int \cdots \int f(x_1^{(j+1)}, \bar{\mathbf{x}}_1^{(j)}) \prod_{k=2}^p f(x_k^{(j+1)} \mid \underline{\mathbf{x}}_k^{(j+1)}, \bar{\mathbf{x}}_k^{(j)}) \prod_{k=2}^p dx_k^{(j)} \\
&= \int \cdots \int \left[ \int f(x_1^{(j+1)}, x_2^{(j)}, \bar{\mathbf{x}}_2^{(j)}) dx_2^{(j)} \right] f(x_2^{(j+1)} \mid \underline{\mathbf{x}}_2^{(j+1)}, \bar{\mathbf{x}}_2^{(j)}) \\
&\quad \cdot \prod_{k=3}^p f(x_k^{(j+1)} \mid \underline{\mathbf{x}}_k^{(j+1)}, \bar{\mathbf{x}}_k^{(j)}) \prod_{k=3}^p dx_k^{(j)} \quad (5.2.6)
\end{aligned}$$

식 (5.2.6)의 우변을 다음과 같이 쓸 수 있다.

$$\begin{aligned}
& \int \cdots \int f(x_1^{(j+1)}, \bar{\mathbf{x}}_2^{(j)}) f(x_2^{(j+1)} \mid x_1^{(j+1)}, \bar{\mathbf{x}}_2^{(j)}) \prod_{k=3}^p f(x_k^{(j+1)} \mid \underline{\mathbf{x}}_k^{(j+1)}, \bar{\mathbf{x}}_k^{(j)}) \prod_{k=3}^p dx_k^{(j)} \\
&= \int \cdots \int f(x_1^{(j+1)}, x_2^{(j+1)}, \bar{\mathbf{x}}_2^{(j)}) \prod_{k=3}^p f(x_k^{(j+1)} \mid \underline{\mathbf{x}}_k^{(j+1)}, \bar{\mathbf{x}}_k^{(j)}) \prod_{k=3}^p dx_k^{(j)} \\
&= \int \cdots \int f(\underline{\mathbf{x}}_3^{(j+1)}, \bar{\mathbf{x}}_2^{(j)}) \prod_{k=3}^p f(x_k^{(j+1)} \mid \underline{\mathbf{x}}_k^{(j+1)}, \bar{\mathbf{x}}_k^{(j)}) \prod_{k=3}^p dx_k^{(j)} \quad (5.2.7)
\end{aligned}$$

식 (5.2.7)의 우변을 다음과 같이 쓸 수 있다.

$$\begin{aligned}
& \int \cdots \int \left[ \int f(\underline{\mathbf{x}}_3^{(j+1)}, x_3^{(j)}, \bar{\mathbf{x}}_3^{(j)}) dx_3^{(j)} \right] \prod_{k=3}^p f(x_k^{(j+1)} \mid \underline{\mathbf{x}}_k^{(j+1)}, \bar{\mathbf{x}}_k^{(j)}) \prod_{k=4}^p dx_k^{(j)} \\
&= \int \cdots \int f(\underline{\mathbf{x}}_3^{(j+1)}, \bar{\mathbf{x}}_3^{(j)}) f(x_3^{(j+1)} \mid \underline{\mathbf{x}}_3^{(j+1)}, \bar{\mathbf{x}}_3^{(j)}) \prod_{k=4}^p f(x_k^{(j+1)} \mid \underline{\mathbf{x}}_k^{(j+1)}, \bar{\mathbf{x}}_k^{(j)}) \prod_{k=4}^p dx_k^{(j)} \\
&= \int \cdots \int f(\underline{\mathbf{x}}_3^{(j+1)}, x_3^{(j+1)}, \bar{\mathbf{x}}_3^{(j)}) \prod_{k=4}^p f(x_k^{(j+1)} \mid \underline{\mathbf{x}}_k^{(j+1)}, \bar{\mathbf{x}}_k^{(j)}) \prod_{k=4}^p dx_k^{(j)} \quad (5.2.8)
\end{aligned}$$

지금까지 과정을 반복적용하면, 식 (5.2.8)의 우변을 다음과 같이 쓸 수 있다.

$$\begin{aligned} & \int f(\underline{\mathbf{x}}_{p-1}^{(j+1)}, x_{p-1}^{(j+1)}, \bar{\mathbf{x}}_{p-1}^{(j)}) \prod_{k=p}^p f(x_k^{(j+1)} \mid \underline{\mathbf{x}}_k^{(j+1)}, \bar{\mathbf{x}}_k^{(j)}) \prod_{k=p}^p dx_k^{(j)} \\ &= \left[ \int f(\underline{\mathbf{x}}_{p-1}^{(j+1)}, x_{p-1}^{(j+1)}, x_p^{(j)}) dx_p^{(j)} \right] f(x_p^{(j+1)} \mid \underline{\mathbf{x}}_p^{(j+1)}) \\ &= f(\underline{\mathbf{x}}_{p-1}^{(j+1)}, x_{p-1}^{(j+1)}) f(x_p^{(j+1)} \mid \underline{\mathbf{x}}_p^{(j+1)}) \end{aligned} \quad (5.2.9)$$

식 (5.2.9)의 우변을 다음과 같이 쓸 수 있다.

$$f(\underline{\mathbf{x}}_p^{(j+1)}) f(x_p^{(j+1)} \mid \underline{\mathbf{x}}_p^{(j+1)}) = f(\underline{\mathbf{x}}_p^{(j+1)}, x_p^{(j+1)}) = f(\mathbf{x}^{(j+1)}) \quad (5.2.10)$$

식 (5.2.1)~식 (5.2.10)에서 알 수 있듯이, Gibbs샘플러에 의해서 발생된 Markov연쇄  $\{\mathbf{x}^{(j)}\}$ 의 불변분포는 결합확률밀도함수  $f(x_1, x_2, \dots, x_p)$ 이다.

지금부터는 Gibbs샘플러에 의해서 발생된 Markov연쇄  $\{\mathbf{x}^{(j)}\}$ 의 수렴성에 대해서 살펴보자. 우선 다음 예제를 살펴보자.

**예제 5.2.3** 다음과 같은 집합들을 정의하자.

$$A \doteq \{[x, y] \mid 0 < x < 1, 0 < y < 1\} \quad (1)$$

$$B \doteq \{[x, y] \mid -1 < x < 0, -1 < y < 0\} \quad (2)$$

여기서 집합들  $A$ 와  $B$ 는 서로소이다. 확률벡터  $[x, y]^t$ 의 결합확률밀도함수가 다음과 같다고 하자.

$$f(x, y) = \frac{1}{2} 1_{A \cup B}(x, y) \quad (3)$$

주변확률밀도함수들은 다음과 같다.

$$f(x) = \frac{1}{2} 1_{(-1,1)}(x) \quad (4)$$

$$f(y) = \frac{1}{2} 1_{(-1,1)}(y) \quad (5)$$

따라서, 조건부확률밀도함수들은 다음과 같다.

$$f(x|y) = 1_{AUB}(x, y) \quad (6)$$

$$f(y|x) = 1_{AUB}(x, y) \quad (7)$$

식 (6)과 식 (7)을 사용해서, Gibbs샘플들  $\{[x^{(j)}, y^{(j)}]^t\}$ 을 생성한다.

식 (6)과 식 (7)을 사용해서 Gibbs샘플을 발생시키기 위해서, 다음 MATLAB프로그램 GibbsBivariateUniform101.m을 실행해 보자.

```

1 % -----
2 % Filename: GibbsBivariateUnifrom101.m
3 % Gibbs Sampler for Bivariate Uniform Distribution 1
4 % Programmed by CBS
5 % -----
6 clear, clf, close all
7 rng((1001^2+1)/2, 'twister')
8 nn = 1000;
9 nk = round(1.1*1000);
10 x = zeros(nn,1); y = zeros(nn,1);
11 % Gibbs Sampler
12 x(1) = - 0.2; y(1) = - 0.7;
13 for jj=2:nk
14     x(jj) = -rand(1);
15     y(jj) = -rand(1);
16 end
17 % Plotting
18 hold on
19 plot( [0 1 1 0 0 ], [0 0 1 1 0 ], 'k-', 'linewidth', 3)
20 plot( [0 -1 -1 0 0 ], [0 0 -1 -1 0 ], 'k-', 'linewidth', 3)
21 set(gca, 'fontsize', 11, 'fontweigh', 'bold')
22 dumx = 0:0.01:1; dummy = ones(1,101);
23 A2 = area(-dumx, dummy)
24 A4 = area(dumx, -dummy)
25 set(A2, 'FaceColor', [1 1 0]);
26 set(A4, 'FaceColor', [0 1 0]);
27 for ii=1071:1099
28     plot([x(ii) x(ii+1)], [y(ii), y(ii)], 'r-', 'linewidth', 2)
29     plot([x(ii+1) x(ii+1)], [y(ii), y(ii+1)], 'r-', 'linewidth', 2)
30 end
31 axis square
32 xlabel('\bf x'), ylabel('\bf y', 'rotation', 0)
33 text(0.3, 0.5, '\bf A', 'fontsize', 50)
34 text(-0.7, -0.5, '\bf B', 'fontsize', 50)
35 hold off
36 saveas(gcf, 'GibbsBivariateUnifrom101', 'eps')
37 % End of program
38 % -----

```

이 MATLAB프로그램 GibbsBivariateUniform101.m을 실행하면, 식 (6)과 식 (7)에 기술한 조건부확률밀도함수들로부터 Gibbs샘플들  $\{[x^{(j)}, y^{(j)}]^t | j = 1, 2, \dots, 1100\}$ 를 발생시킨다. 단, 초기벡터는  $[x^{(1)}, y^{(1)}]^t = [-0.2, -0.7]^t$ 이다. 이 Gibbs샘플들 중 마지막 29개가 그림

5.2.1에 그려져 있다. 그림 5.2.1에서 알 수 있듯이, 초기벡터가 집합  $B$ 에 속하면, 발생된 Gibbs 샘플들도 집합  $B$ 에 속한다. 반대로, 초기벡터가 집합  $A$ 에 속하면, 발생된 Gibbs 샘플들도 집합  $A$ 에 속한다. 즉, 이 Gibbs샘플러에 의해서 발생된 Markov연쇄는 가약적(reducible)이다. ■

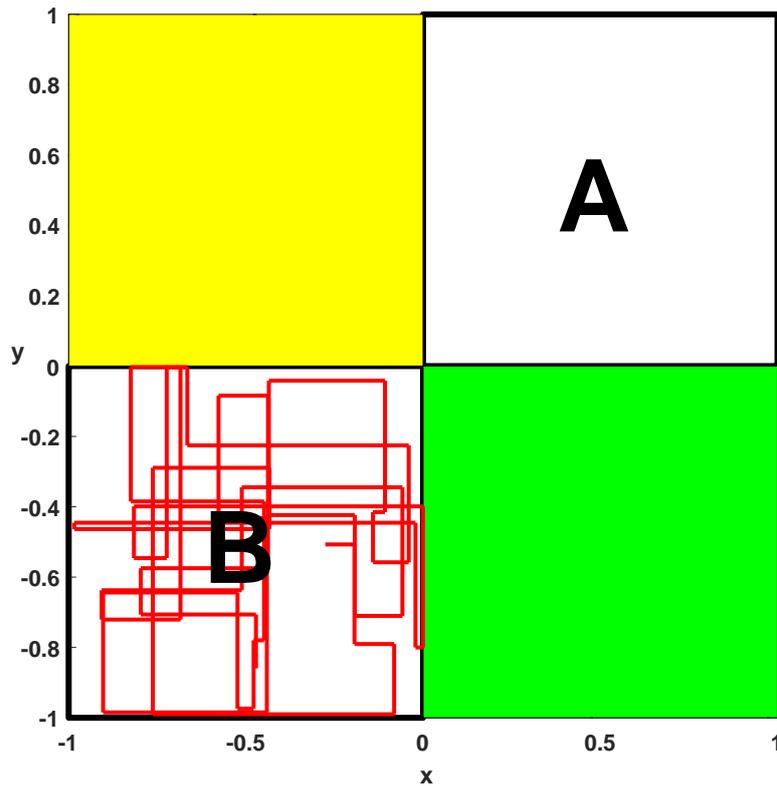


그림 5.2.1. Gibbs샘플러의 수렴성

예제 5.2.3에서 알 수 있듯이, Gibbs샘플러에 의해서 발생된 샘플열이 가약적 Markov체인을 이룰 수도 있다. 다음 명제는 Gibbs샘플러에 의해 발생한 샘플열이 불가약성(irreducibility)을 갖기 위한 충분조건을 제시한다.

#### 명제 5.2.2

만약 확률벡터  $\mathbf{x} = [x_1, x_2, \dots, x_p]^t$ 의 결합확률밀도함수  $f(x_1, x_2, \dots, x_p)$ 가 양적조건(positivity condition)을 만족하면, Gibbs샘플러는 불가약적이고 재귀적인 Markov연쇄를 발생시킨다.

증명. 다음 식을 만족하는 결합확률밀도함수  $f$ 의 지지대의 부분집합  $A$ 를 생각해 보자.

$$\int_A f(x_1^{(j)}, x_2^{(j)}, \dots, x_p^{(j)}) d\mathbf{x}^{(j)} > 0 \quad (1)$$

단,  $\mathbf{x}^{(j)} = [x_1^{(j)}, x_2^{(j)}, \dots, x_p^{(j)}]^t$ 이다. 식 (1)에서 알 수 있듯이, 각  $i$ 에 대해서 다음 식이 성립한다.

$$\int_{A_i} f(x_i^{(j)}) dx_i^{(j)} > 0 \quad (2)$$

여기서  $A_i$ 는 집합  $A$ 의 제  $i$ 번째 공간으로 사영이다. 따라서, 다음 명제를 만족하며 측도가 0이 아닌 집합  $B_i (\subset A_i)$ 가 존재한다.

$$x_i^{(j)} \in B_i \Rightarrow f(x_i^{(j)}) > 0 \quad (3)$$

양적조건에서 알 수 있듯이, 다음 명제가 성립한다.

$$[x_1^{(j)}, x_2^{(j)}, \dots, x_p^{(j)}] \in \prod_{i=1}^p B_i \Rightarrow f(x_1^{(j)}, x_2^{(j)}, \dots, x_p^{(j)}) > 0 \quad (4)$$

다음 식이 성립함을 기억하라.

$$\prod_{i=1}^p B_i \subset A \quad (5)$$

식 (5.2.1)에서 알 수 있듯이, 다음 식이 성립한다.

$$\int_A K(\mathbf{x}^{(j-1)}, \mathbf{x}^{(j)}) d\mathbf{x}^{(j)} = \int_A \prod_{k=1}^p f(x_k^{(j)} \mid x_1^{(j)}, \dots, x_{k-1}^{(j)}, x_{k+1}^{(j-1)}, \dots, x_p^{(j-1)}) d\mathbf{x}^{(j)} \quad (6)$$

식 (3)~식 (5)를 식 (6)에 대입하면, 다음 식을 얻는다.

$$\int_A K(\mathbf{x}^{(j-1)}, \mathbf{x}^{(j)}) d\mathbf{x}^{(j)} > 0 \quad (7)$$

따라서, Gibbs샘플러에 의해서 발생된 Markov체인  $\{\mathbf{x}^{(j)}\}$ 는 불가약적이다. 식 (5.2.4)에서 알 수 있듯이, 결합확률밀도함수  $f(x_1, x_2, \dots, x_p)$ 는 Markov연쇄  $\{\mathbf{x}^{(j)}\}$ 의 불변분포이다. 식 (7)과 불가약성에서 유추할 수 있듯이,  $f(x_1, x_2, \dots, x_p)$ 는 일의적 불변분포이고, 이 Markov

연쇄  $\{\mathbf{x}^{(j)}\}$ 는 재귀적이다. ■

명제 5.2.2에서 알 수 있듯이, Gibbs샘플러를 사용해서 양적조건을 만족하는 결합확률밀도함수로부터 불가역적이며 재귀적인 Markov연쇄를 생성할 수 있다. 만약 이 Gibbs샘플들이 비주기적이고 양재귀적이면, 에르고딕정리에서 알 수 있듯이 임의의 적분가능함수  $g(\cdot)$ 와 거의 모든 초기벡터  $\mathbf{x}^{(0)}$ 에 대해서 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n g(\mathbf{x}^{(j)}) = E(g(\mathbf{x})) \text{ a.s.} \quad (5.2.11)$$

여기서  $\mathbf{x}$ 는 불변분포  $f(x_1, x_2, \dots, x_p)$ 를 갖는 확률벡터이다. Gibbs샘플열  $\{\mathbf{x}^{(j)}\}$ 는 Markov연쇄를 이룬다. 따라서, 이들은 서로 독립이 아니고 상관되어 있다. 이러한 관계는 일반적으로 양의 상관이다. 이 상관계수의 절대값이 클수록, Gibbs샘플열은 천천히 변화한다. 이를 완만한 혼합(slowly mixing)이라 한다.

**예제 5.2.4** 확률벡터  $[x, y]^t$ 가 다음과 같은 2변량 정규확률분포를 따른다고 하자.

$$\begin{bmatrix} x \\ y \end{bmatrix} \stackrel{d}{\sim} \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \quad (1)$$

이 경우에 Gibbs샘플러는 다음과 같다.

$$y^{(j+1)} \stackrel{d}{\sim} \mathcal{N}(\rho x^{(j)}, 1 - \rho^2) \quad (2)$$

$$x^{(j+1)} \stackrel{d}{\sim} \mathcal{N}(\rho y^{(j)}, 1 - \rho^2) \quad (3)$$

식 (2)와 식 (3)에서 발생시킨 Gibbs샘플들을 사용해서 확률  $\Pr(x_1 \geq 0, x_2 \geq 0)$ 를 계산하기 위해, 다음 MATLAB프로그램 GibbsBivariateUniform103.m을 실행해 보자.

```

1 % -----
2 % Filename: GibbsBivariateNormal103.m
3 % Gibbs Sampler for Bivariate Normal Distribution 3
4 % Programmed by CBS
5 % -----
6 clear, clf, close all, format long
7 syms xx yy
8 rng((1001^2+1)/2, 'twister')
9 for nrho = [1 2]
10     rho = 0.3 + 0.6*(nrho-1)
11     TVnormal = 1/(2*pi*sqrt(1-rho^2)) ...
12         *exp(-1/(2*(1-rho^2))*(xx^2-2*rho*xx*yy+yy^2))

```

```

13 TrueProb(nrho) = double(int(int(TVnormal,xx,0,inf),yy,0,inf))
14 n1 = 200; n2 = 100;
15 nn = n1*n2;
16 xr = zeros(1,nn);
17 yr = zeros(1,nn);
18 % Gibb's sampling
19 xr(1) = 1; yr(1) = 1; gg(1) = 0;
20 for ndum = 1:n1
21     for ii=(ndum-1)*n2+2:ndum*n2+1
22         xr(ii) = normrnd(rho*yr(ii-1),1-rho^2);
23         yr(ii) = normrnd(rho*xr(ii), 1-rho^2);
24         gg(ii) = (xr(ii)>0 & yr(ii)>0 );
25     end
26     prob(ndum) = sum(gg(1:n2*ndum))/(n2*ndum);
27 end
28 xdum = n2*(1:n1);
29 % Ploting
30 subplot(2,1,nrho)
31 plot(xdum,prob,'k-',[1 nn],[TrueProb(nrho),TrueProb(nrho)],'r--', ...
32     'linewidth',2)
33 set(gca,'fontsize',11,'fontweigh','bold')
34 dumaxis(nrho) = round(10*TrueProb(nrho))/10
35 axis([1 nn dumaxis(nrho)-0.15 dumaxis(nrho)+0.15 ])
36 ylabel('\bf Probability')
37 xlabel('\bf Number of Gibbs Samples')
38 end
39 saveas(gcf,'GibbsBivariateNormal103','eps')
40 % End of program
41 % -----

```

이 MATLAB 프로그램 GibbsBivariateUniform103.m을 실행하면, 초기벡터가  $[x^{(1)}, y^{(1)}]^t = [1, 1]^t$  인 Gibbs 샘플들  $\{[x^{(j)}, y^{(j)}]^t \mid j = 1, 2, \dots, 20000\}$  이 발생된다. 이 Gibbs 샘플들로부터 계산된 확률  $\Pr(x_1 \geq 0, x_2 \geq 0)$  가 그림 5.2.2에 그려져 있다. 그림 5.2.2의 상단 그래프는 상관계수가  $\rho = 0.3$  인 경우이다. 이 그래프에서 알 수 있듯이, 상관계수의 절대값이 비교적 작은 경우에는 확률이 진짜확률 0.2985에 빠르게 수렴한다. 하단 그래프는 상관계수가  $\rho = 0.9$  인 경우이다. 이 그래프에서 알 수 있듯이, 상관계수의 절대값이 큰 경우에는 확률이 진짜확률 0.4282에 느리게 수렴한다. 즉, 이 경우에는 Gibbs 샘플열이 느리게 혼합한다.

다음 Python 프로그램 GibbsBivariateUniform103.py을 실행해도, MATLAB 프로그램 GibbsBivariateUniform103.m을 실행한 것과 동일한 결과를 출력한다.

```

1 # GibbsBivariateNormal103.py
2 """
3 Filename : GibbsBivariateNormal103.py
4 Gibbs Sampler for Bivariate Normal Distribution 3
5 Programmed by Kyo-Won Suh based on the R program by CBS
6 """
7
8 import numpy as np
9 import pandas as pd
10 import matplotlib.pyplot as plt

```

```

11 import scipy as sci
12 from scipy import integrate
13 import sympy as sym
14 import math as m
15
16
17
18 # xx, yy = sym.symbols('xx yy')
19 np.random.seed((1001**2+1)//2)
20 # nrho = np.array([1,2])
21 nrho = 1
22 rho = 0.3 + 0.6*(nrho-1)
23 # TVnormal = 1/(2*m.pi*np.sqrt(1-rho**2))*sym.exp(-1/(2*(1-rho**2))*(xx**2-2*rho*
    xx*yy+yy**2))
24 TVnormal = lambda xx, yy: 1/(2*m.pi*np.sqrt(1-rho**2))*sym.exp(-1/(2*(1-rho**2))
    *(xx**2-2*rho*xx*yy+yy**2))
25 #scipy.integrate(TVnormal, 0,1)
26 TrueProb = integrate.dblquad(TVnormal, 0, np.Inf, lambda yy: 0, lambda yy: np.Inf
    )
27 TrueProb = TrueProb[0]
28
29
30
31 n1 = 200; n2 = 100;
32 nn = n1*n2;
33
34
35 xr = np.zeros(nn+1);
36 yr = np.zeros(nn+1);
37 np.size(xr)
38
39
40
41 ## Gibb's sampling
42
43
44 xr[0] = 1;
45 yr[0] = 1;
46 gg = np.zeros(nn+1);
47 prob = np.zeros(n1);
48
49
50 for ndum in range (1, n1+1):
51     for ii in range ((ndum-1)*n2+1, ndum*n2+1):
52         xr[ii] = np.random.normal(rho*yr[ii-1],1-rho**2)
53         yr[ii] = np.random.normal(rho*xr[ii], 1-rho**2)
54         gg[ii] = 1 if xr[ii]>0 and yr[ii]>0 else 0
55     # print(ndum)
56     prob[ndum-1] = sum(gg[1:n2*ndum])/(n2*ndum)
57     # print(prob[ndum-1])
58
59 np.size(prob)
60
61
62
63
64 xdum = n2*np.arange(1,n1+1)
65 # print(xdum)
66 np.size(xdum)
67
68

```

```

69 fig = plt.figure(figsize=(10,5))
70 ax = fig.add_subplot(111)
71 ax.plot(xdum, prob, 'black')
72 ax.plot(range(1,nn+1), np.full((nn, 1), TrueProb), 'r--')
73 ax.set_xlabel('Number of Gibbs Samples ')
74 ax.set_ylabel('Probability')
75
76
77 np.random.seed((1001**2+1)//2)
78 # nrho = np.array([1,2])
79 nrho = 2
80 rho = 0.3 + 0.6*(nrho-1)
81 # TVnormal = 1/(2*m.pi*np.sqrt(1-rho**2))*sym.exp(-1/(2*(1-rho**2))*(xx**2-2*rho*
      xx*yy+yy**2))
82 TVnormal = lambda xx, yy: 1/(2*m.pi*np.sqrt(1-rho**2))*sym.exp(-1/(2*(1-rho**2))
      *(xx**2-2*rho*xx*yy+yy**2))
83 #scipy.integrate(TVnormal, 0,1)
84 TrueProb = integrate.dblquad(TVnormal, 0, np.Inf, lambda yy: 0, lambda yy: np.Inf
      )
85 TrueProb = TrueProb[0]
86 n1 = 200; n2 = 100;
87 nn = n1*n2;
88 xr = np.zeros(nn+1);
89 yr = np.zeros(nn+1);
90 np.size(xr)
91 ## Gibb's sampling
92 xr[0] = 1;
93 yr[0] = 1;
94 gg = np.zeros(nn+1);
95 prob = np.zeros(n1);
96 for ndum in range (1, n1+1):
97     for ii in range ((ndum-1)*n2+1, ndum*n2+1):
98         xr[ii] = np.random.normal(rho*yr[ii-1],1-rho**2)
99         yr[ii] = np.random.normal(rho*xr[ii], 1-rho**2)
100        gg[ii] = 1 if xr[ii]>0 and yr[ii]>0 else 0
101 #     print(ndum)
102     prob[ndum-1] = sum(gg[1:n2*ndum])/(n2*ndum)
103 #     print(prob[ndum-1])
104 xdum = n2*np.arange(1,n1+1)
105 # print(xdum)
106 np.size(xdum)
107 fig = plt.figure(figsize=(10,5))
108 ax = fig.add_subplot(111)
109 ax.plot(xdum, prob, 'black')
110 ax.plot(range(1,nn+1), np.full((nn, 1), TrueProb), 'r--')
111 ax.set_xlabel('Number of Gibbs Samples ')
112 ax.set_ylabel('Probability')
113
114
115 # End of program
116 # -----

```



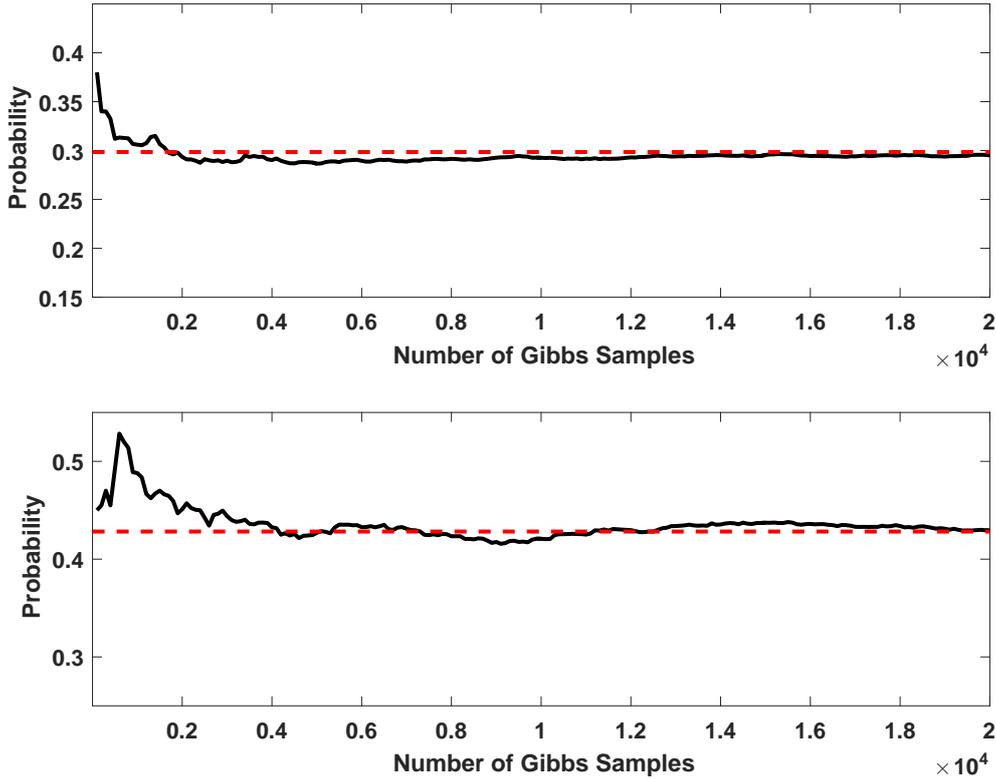


그림 5.2.2. Gibbs샘플러와 2변량 정규확률분포 3

### 제5.3절 베이지안통계분석과 Gibbs샘플러

모수벡터  $\theta$ 를  $p$ 개 벡터들로 나누기로 하자. 즉,  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ 로 나누기로 하자. 목표 분포인  $\theta$ 의 사후확률밀도함수를  $\pi(\theta | \mathbf{y})$ 로 두면, Gibbs샘플러에서는 각 그룹의 모수벡터  $\theta_i$ 를 다음과 같은 사후조건부확률밀도함수로부터 발생시킨다.

$$\pi(\theta_i | \theta_{-i}, \mathbf{y}) = \pi(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p, \mathbf{y}) \tag{5.3.1}$$

이러한 Gibbs샘플러를 시행하기 위해서는  $p$ 개 사후조건부확률밀도함수들  $\pi(\theta_1 | \theta_{-1}, \mathbf{y})$ ,  $\pi(\theta_2 | \theta_{-2}, \mathbf{y})$ ,  $\dots$ ,  $\pi(\theta_p | \theta_{-p}, \mathbf{y})$ 에서 샘플링이 가능해야 한다. 공액인 사후확률밀도함수  $\pi(\theta_i | \mathbf{y})$ 는 존재하지 않아도 공액인 사후조건부확률밀도함수  $\pi(\theta_i | \theta_{-i}, \mathbf{y})$ 가 존재하는 경우가 있다. 이러한 경우에 사후조건부확률밀도함수  $\pi(\theta_i | \theta_{-i}, \mathbf{y})$ 는 잘 알려진 함수 형태를 취하는 경우가 많으며, 또한 그 확률밀도함수로부터 샘플링이 쉬운 경우가 많다.

Gibbs샘플러를 사용해서 베이지안추정을 하는 절차는 다음과 같다. 먼저, 적당한 초기벡터  $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$ 를 선택한다. 이 초기벡터  $\theta^{(0)}$ 를 목표분포  $\pi(\theta | \mathbf{y})$ 에서 샘플링하는 것이 바람직하지만, 목표분포를 구하는 것이 목적이므로 실제로 이러한 샘플링은

가능하지 않다. 따라서, 편리하고 적당한 초기벡터를 선택하여 알고리즘을 진행한다. 다음 단계로,  $\pi(\theta_1 | \theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_p^{(0)}, \mathbf{y})$ 에서  $\theta_1^{(1)}$ 을 샘플링한다. 이렇게 얻어진  $\theta_1^{(1)}$ 을 이용해서,  $\pi(\theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_p^{(0)}, \mathbf{y})$ 에서  $\theta_2^{(1)}$ 를 샘플링한다. 같은 방법으로, 각  $i(= 3, 4, \dots, p-1)$ 에 대해서  $\pi(\theta_i | \theta_1^{(1)}, \dots, \theta_{i-1}^{(1)}, \theta_{i+1}^{(0)}, \dots, \theta_p^{(0)}, \mathbf{y})$ 로부터  $\theta_i^{(1)}$ 를 샘플링한다. 마지막으로,  $\pi(\theta_p | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{p-1}^{(1)}, \mathbf{y})$ 에서  $\theta_p^{(1)}$ 를 샘플링한다. 지금까지 결과를 취합해서, 제1단계 모수벡터  $\theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_p^{(1)})$ 를 얻는다. 동일한 절차를 통해서, 모수벡터들  $\theta^{(2)}, \theta^{(3)}, \dots$ 을 발생시킨다. 이렇게 발생된  $\{\theta^{(j)} | j = 1, 2, \dots\}$ 는 다음 식을 만족하는 Markov연쇄이다.

$$\begin{aligned} K(\theta^{(j-1)}, \theta^{(j)} | \mathbf{y}) &= \pi(\theta_1^{(j)} | \theta_2^{(j-1)}, \dots, \theta_p^{(j-1)}, \mathbf{y}) \\ &\cdot \left[ \prod_{i=2}^{p-1} \pi(\theta_i^{(j)} | \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)}, \theta_{i+1}^{(j-1)}, \dots, \theta_p^{(j-1)}, \mathbf{y}) \right] \pi(\theta_p^{(j)} | \theta_1^{(j)}, \dots, \theta_{p-1}^{(j)}, \mathbf{y}) \end{aligned} \quad (5.3.2)$$

여기서 추이핵  $K(\theta^{(j-1)}, \theta^{(j)} | \mathbf{y})$ 는 Gibbs샘플러에서 제  $j-1$ 단계 관찰벡터  $\theta^{(j-1)}$ 가 주어진 조건 하에서 제  $j$ 단계 관찰벡터  $\theta^{(j)}$ 의 조건부확률밀도함수이다. 이 Markov연쇄는  $\pi(\theta | \mathbf{y})$ 를 정상분포로 갖는다. 따라서, 충분히 큰  $k$ 에 대해  $\{\theta^{(j)} | j = k, k+1, \dots\}$ 는 목표분포  $\pi(\theta | \mathbf{y})$ 에서 발생된 샘플열이라 할 수 있다. 즉, 충분히 큰  $k$ 에 대해  $\{\theta^{(j)} | j \geq k\}$ 를 목표분포  $\pi(\theta | \mathbf{y})$ 로부터 추출된 샘플열로 간주한다.

모수벡터의 사후확률분포로부터 Gibbs샘플링하는 것은 알고리즘 5.1.1과 다르지 않다. 알고리즘 5.1.1과 다른 점은 관찰벡터  $\mathbf{y}$ 가 주어진 조건 하에서 조건부확률벡터로부터 샘플링을 한다는 것이다.

#### 알고리즘 5.3.1: 사후확률분포를 위한 Gibbs샘플러

(1단계) 초기벡터  $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$ 를 선택한다.

(2단계) 각  $j(= 1, 2, \dots)$ 에 대해서 다음 과정을 반복 실행한다.

모수벡터  $\theta_1^{(j)}$ 를  $\pi(\theta_1 | \theta_2^{(j-1)}, \theta_3^{(j-1)}, \dots, \theta_p^{(j-1)}, \mathbf{y})$ 로부터 발생시킨다.

모수벡터  $\theta_2^{(j)}$ 를  $\pi(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_p^{(j-1)}, \mathbf{y})$ 로부터 발생시킨다.

⋮

모수벡터  $\theta_i^{(j)}$ 를  $\pi(\theta_i | \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)}, \theta_{i+1}^{(j-1)}, \dots, \theta_p^{(j-1)}, \mathbf{y})$ 로부터 발생시킨다.

⋮

모수벡터  $\theta_p^{(j)}$  를  $\pi(\theta_p | \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_{p-1}^{(j)}, \mathbf{y})$  로부터 발생시킨다.

만약 충분히 큰  $k$ 에 대해 식  $j \geq k$ 이 성립하면, 모수벡터  $\theta^{(j)} = (\theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_p^{(j)})$  를 기록한다.

**(3단계)** 만약  $j$ 이 추출하고자 하는 샘플수에  $k$ 을 더한 값보다 크면, 이 알고리즘의 수행을 멈춘다. 그렇지 않으면,  $j$ 에 1을 더한 다음, 제2단계로 돌아간다.

알고리즘 5.3.1에서는 초기벡터를 적당히 선택하여 알고리즘을 시행한다고 가정했다. 그러나, 초기벡터가 적절하지 않을 가능성이 있으므로, 추출된 샘플들  $\{\theta^{(j)}\}$ 가 초기벡터에 상관없이 목표분포  $\pi(\theta | \mathbf{y})$ 에 수렴하고 있는지 여부를 유의해야 한다.

**5.3.1 정규확률분포와 Gibbs샘플러**

이 소절에서는 Gibbs샘플러를 사용해서 정규확률분포의 평균과 분산에 관해서 살펴보자. 서로 독립인 확률변수들  $y_1, y_2, \dots, y_n$ 이 동일한 확률분포  $N(\mu, \sigma^2)$ 를 따른다고 하자. 모수들  $\mu$ 와  $\sigma^2$ 의 사전확률분포가 다음과 같다고 가정하자.

$$\mu \stackrel{d}{\sim} N(\mu_0, \sigma_0^2) \tag{5.3.3}$$

$$\sigma^2 \stackrel{d}{\sim} IG\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \tag{5.3.4}$$

관찰벡터  $\mathbf{y} = [y_1, y_2, \dots, y_n]^t$ 의 우도함수는 다음과 같다.

$$f(\mathbf{y} | \mu, \sigma^2) = [2\pi\sigma^2]^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n [y_j - \mu]^2\right) \tag{5.3.5}$$

즉, 다음 식이 성립한다.

$$f(\mathbf{y} | \mu, \sigma^2) \propto [\sigma^2]^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \{[n-1]s^2 + n[\bar{y} - \mu]^2\}\right) \tag{5.3.6}$$

여기서  $\bar{y}$ 와  $s^2$ 는 각각 다음과 같다.

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j, \quad s^2 = \frac{1}{n-1} \sum_{j=1}^n [y_j - \bar{y}]^2 \tag{5.3.7}$$

따라서, 모수들  $\mu$ 와  $\sigma^2$ 의 사후조건부확률분포들은 각각 다음과 같다.

$$\pi(\mu | \sigma^2, \mathbf{y}) \propto \exp\left(-\frac{1}{2}\left\{n\sigma^{-2}[\bar{y} - \mu]^2 + \sigma_0^{-2}[\mu - \mu_0]^2\right\}\right) \quad (5.3.8)$$

$$\pi(\sigma^2 | \mu, \mathbf{y}) \propto [\sigma^2]^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n [y_j - \mu]^2\right) [\sigma^2]^{-n_0/2-1} \exp\left(-\frac{n_0 S_0}{2\sigma^2}\right) \quad (5.3.9)$$

다음 상수들과 변수들을 정의하자.

$$\mu_* \doteq \sigma_*^2 [\sigma_0^{-2} \mu_0 + n\sigma^{-2} \bar{y}] \quad (5.3.10)$$

$$\sigma_*^2 \doteq \frac{1}{\sigma_0^{-2} + n\sigma^{-2}} \quad (5.3.11)$$

$$n_* \doteq n_0 + n \quad (5.3.12)$$

$$S_* \doteq \frac{1}{n_*} \left\{ n_0 S_0 + \sum_{j=1}^n [y_j - \mu]^2 \right\} = \frac{1}{n_*} \left\{ n_0 S_0 + \sum_{j=1}^n [y_j - \bar{y}]^2 + n [\bar{y} - \mu]^2 \right\} \quad (5.3.13)$$

식 (5.3.8)~식 (5.3.13)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\pi(\mu | \sigma^2, \mathbf{y}) \propto \exp\left(-\frac{1}{2\sigma_*^2} [\mu - \mu_*]^2\right) \quad (5.3.14)$$

$$\pi(\sigma^2 | \mu, \mathbf{y}) \propto [\sigma^2]^{-n_*/2+1} \exp\left(-\frac{n_* S_*}{2\sigma^2}\right) \quad (5.3.15)$$

즉, 다음 식들이 성립한다.

$$\mu | (\sigma^2, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}(\mu_*, \sigma_*^2) \quad (5.3.16)$$

$$\sigma^2 | (\mu, \mathbf{y}) \stackrel{d}{\sim} IG\left(\frac{n_*}{2}, \frac{n_* S_*}{2}\right) \quad (5.3.17)$$

지금까지 내용을 바탕으로, 모수들  $(\mu, \sigma^2)$ 의 사후확률분포에서 샘플링하는 Gibbs샘플러를 다음과 같이 요약할 수 있다.

**(1단계)** 초기모수들  $(\mu^{(0)}, [\sigma^{(0)}]^2)$ 를 선택한다.

**(2단계)** 각  $j(= 1, 2, \dots)$ 에 대해서 주어진  $\mathbf{y}$ 에 대한 조건부샘플  $(\mu^{(j)}, [\sigma^{(j)}]^2)$ 를 발생시킨다. 즉, 식 (5.3.16)의 사후조건부확률분포에서  $\mu^{(j)} | ([\sigma^{(j-1)}]^2, \mathbf{y})$ 를 발생시키고, 식 (5.3.17)의 사후조건부확률분포에서  $[\sigma^{(j)}]^2 | (\mu^{(j)}, \mathbf{y})$ 를 발생시킨다.

(3단계) 충분히 큰  $b$ 에 대해서  $(\mu^{(j)}, [\sigma^{(j)}]^2)$ ,  $(j = b + 1, b + 2, \dots)$ 를 기록한다.

**예제 5.3.1** 이 Gibbs샘플러를 적용하는 예로서 다음 MATLAB 프로그램 GibbsSampler-Normal101R.R을 실행해 보자.

```

1 # -----
2 # Filename: GibbsSamplerNormal101R.R
3 # Gibbs Posterior for Normal Distribution
4 # Programmed by CBS
5 # -----
6 # Prior
7 mu0 <- 10; sigsig0 <- 25; n0 <- 1; S0 <-2
8 # Likelihood
9 y <- rnorm(16,10,2)
10 n <- length(y); ymean <- mean(y); yvar <- var(y); ysd <- sd(y)
11
12 # Posterior
13 sigsigSim <- 1/(1/sigsig0 + n/yvar)
14
15 ## Gibbs Sampler
16 set.seed(1)
17 Nsim <- 500; burnin <- 1000
18 ThetaDum <- matrix(nrow=(burnin+Nsim), ncol=2)
19 for (jj in 1:(burnin+Nsim)){
20     sigsigC <- 1/(1/sigsig0 + n/sigsigSim)
21     muC <- sigsigC*(mu0/sigsig0 + n*ymean/sigsigSim)
22     muSim <- rnorm(1,muC,sqrt(sigsigC))
23     nStar <- n0 + n;
24     SStar <- 1/nStar*(n0*S0+(n-1)*yvar+n*(ymean-muSim)^2)
25     sigsigSim <- 1/rgamma(1,nStar/2,nStar*SStar/2)
26     ThetaDum[jj,] <- c(muSim,sigsigSim)
27 }
28 Theta <- ThetaDum[(burnin+1):(burnin+Nsim),]
29 # PlottinTheta
30 library(ggplot2)
31 # install.packages("grid")
32 library(grid)
33 setEPS()
34 plot.new()
35 postscript('GibbsSamplerNormal101R.eps') # Start to save figure
36 timee = c(1:Nsim)
37 GibbsData <- data.frame(timee,Theta)
38 timeeS <- timee[1:10]; ThetaS <- Theta[1:10,]
39 GibbsDataS <- data.frame(timeeS,ThetaS)
40 timeeF <- timee[(Nsim-9):Nsim]; ThetaF <- Theta[(Nsim-9):Nsim,]
41 GibbsDataF <- data.frame(timeeF,ThetaF)
42 plot1 <- ggplot(GibbsDataS, aes(x=ThetaS[,1],y=ThetaS[,2])) +
43     geom_path(colour = "green",lwd=1.2) +
44     geom_point(colour="red",size=4) +
45     xlab(expression(mu)) + ylab(expression(sigma^2))
46 plot2 <- ggplot(GibbsDataF, aes(x=ThetaF[,1],y=ThetaF[,2])) +
47     geom_path(colour = "green",lwd=1.2) +
48     geom_point(colour="blue",size=4) +
49     xlab(expression(mu)) + ylab(expression(sigma^2))
50 plot3 <- ggplot(GibbsData, aes(x=Theta[,1])) +
51     geom_histogram(aes(y =..density..), bins=30,
52     # breaks=seq(0.0,2.5, by=0.05),
53     col="red",fill="green") +

```

```

54     geom_density(lwd=1.0) +
55     xlab(expression(mu))
56 plot4 <- ggplot(GibbsData, aes(x=Theta[,2])) +
57     geom_histogram(aes(y = ..density..), bins=30,
58                   # breaks=seq(0.0,2.5, by=0.05),
59                   col="red",fill="pink") +
60     geom_density(lwd=1.0) +
61     xlab(expression(sigma^2))
62 pushViewport(viewport(layout = grid.layout(2,2)))
63 print(plot1, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
64 print(plot2, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
65 print(plot3, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
66 print(plot4, vp = viewport(layout.pos.row = 2, layout.pos.col = 2))
67 dev.off()           # End to save figure
68 # -----

```

이 R 프로그램 GibbsSamplerNormal101R.R에서는 모수들이 다음과 같은 사전확률분포를 가정한다.

$$\mu_0 = 10, \quad \sigma_0^2 = 25, \quad n_0 = 1, \quad S_0 = 2 \quad (1)$$

또한, 정규확률분포  $\mathcal{N}(10, 2^2)$ 에서 추출된 16개 관찰값들이 우도함수를 구성한다. 앞에 기술한 Gibbs샘플러를 사용해서, 변인기간이 1000개인 Gibbs샘플들 500개를 발생시킨다.

이 R 프로그램 GibbsSamplerNormal101R.R을 실행하면, 그림 5.3.1에 그려진다. 그림 5.3.1의 상단좌측 그래프는 첫 10개 Gibbs샘플들의 산점도이고, 상단우측 그래프는 마지막 10개 Gibbs샘플들의 산점도이다. 이 산점도들에서 알 수 있듯이, 시간이 흐를수록 생성되는 관찰값들의 범위가 좁아진다. 또한, 하단좌측 그래프는 모수  $\mu$ 의 사후히스토그램이고, 하단우측 그래프는 모수  $\sigma^2$ 의 사후히스토그램이다.

다음 Python 프로그램 GibbsSamplerNormal101R.py를 실행하면, R 프로그램 GibbsSamplerNormal101R.R을 실행한 것과 동일한 결과를 얻는다.

```

1 # GibbsSamplerNormal101R.py
2 """
3 Filename : GibbsMultiIntegration101.py
4 Gibbs Posterior for Normal Distribution
5 Programmed by Kyung-Bae Park based on the R program by CBS
6 """
7
8 import numpy as np
9 import matplotlib.pyplot as plt
10 import math
11 import seaborn as sns
12
13
14 # Prior
15 mu0 = 10
16 sigsig0 = 25

```

```

17 n0 = 1
18 S0 = 2
19
20
21 # Likelihood
22 y = np.random.normal(10,2,16)
23 n = len(y)
24 ymean = np.mean(y)
25 yvar = np.var(y)
26 ysd = np.std(y)
27
28
29 # Posterior
30 sigsigSim = 1/(1/sigsig0 + n/yvar)
31
32
33 ## Gibbs Sampler
34 np.random.seed(12345)
35 Nsim = 500
36 burnin = 1000
37 ThetaDum = np.zeros(((burnin+Nsim),2))
38 for jj in range(burnin+Nsim):
39     sigsigC = 1/(1/sigsig0 + n/sigsigSim)
40     muC = sigsigC*(mu0/sigsig0 + n*ymean/sigsigSim)
41     muSim = np.random.normal(muC,math.sqrt(sigsigC),1)
42     nStar = n0 + n
43     SStar = 1/nStar*(n0*S0+(n-1)*yvar+n*(ymean-muSim)**2)
44     sigsigSim = 1/np.random.gamma(nStar/2, 1/(nStar*SStar/2))
45     ThetaDum[jj,:] = [muSim, sigsigSim]
46 Theta = ThetaDum[burnin:(burnin+Nsim),:]
47
48
49 # Plotting Theta
50 timee = (1+np.array(range(Nsim))).reshape((Nsim,1))
51 GibbsData = np.concatenate((timee,Theta), axis=1)
52 timeeS = timee[0:10]
53 ThetaS = Theta[0:10,:]
54 GibbsDataS = np.concatenate((timeeS, ThetaS), axis=1)
55 timeeF = timee[(Nsim-10):Nsim]
56 ThetaF = Theta[(Nsim-10):Nsim,:]
57 GibbsDataF = np.concatenate((timeeF, ThetaF), axis=1)
58
59 fig_a = plt.figure()
60 plt.subplot(221)
61 plt.plot(ThetaS[:,0], ThetaS[:,1], 'C3', zorder=1.2, lw=2, color='green')
62 plt.scatter(ThetaS[:,0], ThetaS[:,1], s=40, zorder=2, color='red')
63 plt.xlabel(r'$\mu$', fontsize = 10)
64 plt.ylabel(r'$\sigma^2$', fontsize = 10)
65
66 plt.subplot(222)
67 plt.plot(ThetaF[:,0], ThetaF[:,1], 'C3', zorder=1.2, lw=2, color='green')
68 plt.scatter(ThetaF[:,0], ThetaF[:,1], s=40, zorder=2, color='blue')
69 plt.xlabel(r'$\mu$', fontsize = 10)
70 plt.ylabel(r'$\sigma^2$', fontsize = 10)
71
72 plt.subplot(223)
73 sns.distplot(Theta[:,0], bins=30, color = 'black',
74             hist_kws={'edgecolor':'purple', 'color':'limegreen', 'alpha':0.75},
75             kde_kws={'linewidth': 2})
76 plt.xlabel(r'$\mu$', fontsize = 10)
77 plt.ylabel('density', fontsize = 10)

```

```

78
79 plt.subplot(224)
80 sns.distplot(Theta[:,1], bins=30, color = 'black',
81             hist_kws={'edgecolor': 'purple', 'color': 'lightpink', 'alpha': 0.75},
82             kde_kws={'linewidth': 2})
83 plt.xlabel(r'$\sigma^2$', fontsize = 10)
84 plt.ylabel('density', fontsize = 10)
85
86 plt.subplots_adjust(wspace=0.4, hspace=0.4)
87 plt.show()
88 fig_a.savefig('GibbsSamplerNormal101R.eps', format='eps')
89
90
91 # End of program
92 # -----

```

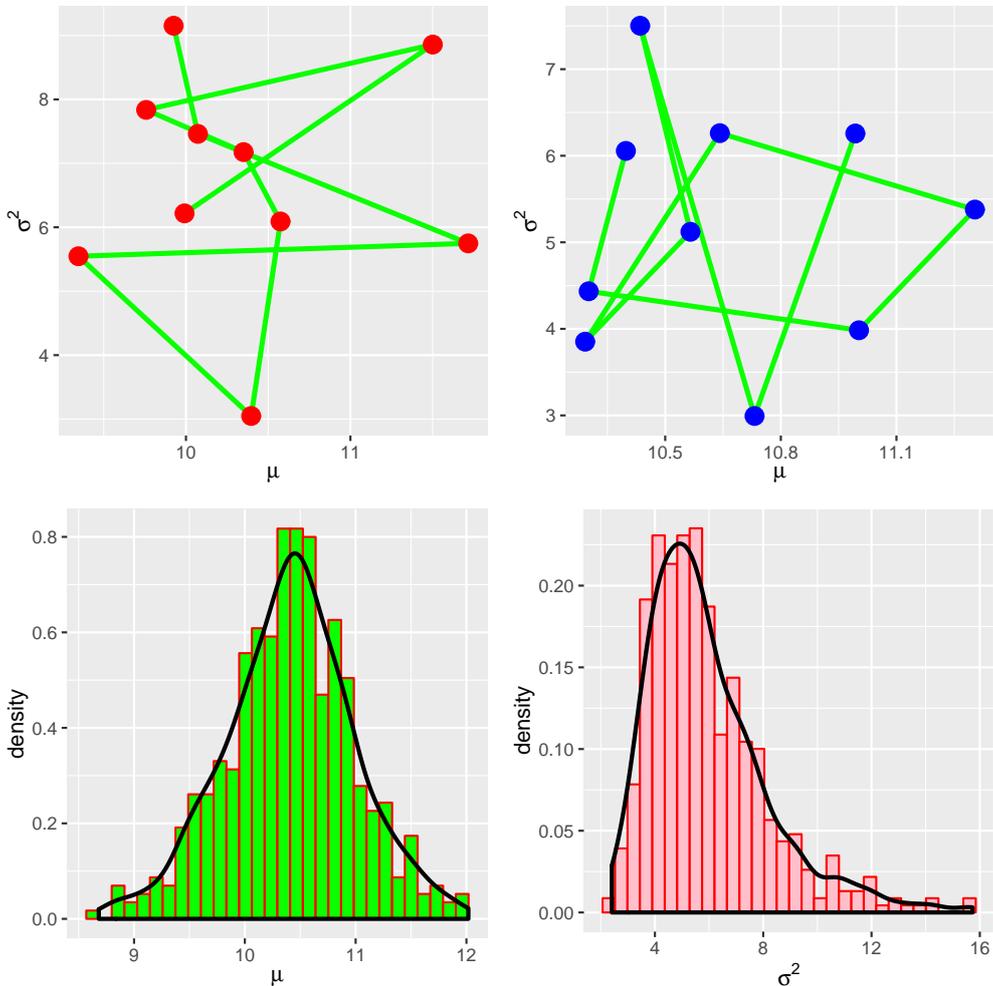


그림 5.3.1. Gibbs샘플러와 정규확률분포

### 5.3.2 시계열의 변환점과 Gibbs샘플러

베이지안추정법을 사용해서 시계열의 변환점(change-point)을 추정하는 문제를 살펴보자. 확률변수들  $y_1, y_2, \dots, y_n$ 이 서로 독립이라고 하자. 또한, 시점  $k$ 까지는 평균모수가  $\theta$ 인 Poisson 확률분포를 따르고, 시점  $k+1$ 부터는 평균모수가  $\lambda$ 인 Poisson 확률분포를 따른다고 가정하자. 즉, 다음 식들이 성립한다고 가정하자.

$$y_t \stackrel{d}{\sim} \text{Poisson}(\theta), \quad (t = 1, 2, \dots, k) \tag{5.3.18}$$

$$y_t \stackrel{d}{\sim} \text{Poisson}(\lambda), \quad (t = k + 1, k + 2, \dots, n) \tag{5.3.19}$$

모수들의 사전확률분포들이 다음과 같다고 하자.

$$\theta \stackrel{d}{\sim} \text{Gamma}(a_1, b_1) \tag{5.3.20}$$

$$\lambda \stackrel{d}{\sim} \text{Gamma}(a_2, b_2) \tag{5.3.21}$$

$$k \stackrel{d}{\sim} \text{Uniform}\{1, 2, \dots, n\} \tag{5.3.22}$$

또한,  $\theta, \lambda$ , 그리고  $k$ 는 서로 독립이고,  $a_1, b_1, a_2, b_2$ 는 알려진 상수들이다. 식 (5.3.22)에서 알 수 있듯이, 변환점  $k$ 는 지지대가  $\{1, 2, \dots, n\}$ 인 일양확률분포를 따른다. 관찰벡터를  $\mathbf{y} = [y_1, y_2, \dots, y_n]^t$ 이라 하면, 우도함수는 다음과 같다.

$$L(\mathbf{y} | k, \theta, \lambda) \propto \exp(-k\theta) \theta^{\sum_{t=1}^k y_t} \exp(-[n-k]\lambda) \lambda^{\sum_{t=k+1}^n y_t} \tag{5.3.23}$$

따라서, 사후결합확률분포는 다음과 같다.

$$\pi(\theta, \lambda, k | \mathbf{y}) \propto \theta^{a_1-1} e^{-b_1\theta} \lambda^{a_2-1} e^{-b_2\lambda} \prod_{t=1}^k \exp(-\theta)\theta^{y_t} \prod_{t=k+1}^n \exp(-\lambda)\lambda^{y_t} \tag{5.3.24}$$

식 (5.3.24)를 다음과 같이 쓸 수 있다.

$$\pi(\theta, \lambda, k | \mathbf{y}) \propto \theta^{a_1 + \sum_{t=1}^k y_t - 1} \exp(-[b_1 + k]\theta) \lambda^{a_2 + \sum_{t=k+1}^n y_t - 1} \exp(-[b_2 + n - k]\lambda) \tag{5.3.25}$$

따라서,  $\theta$ 와  $\lambda$ 의 사후조건부확률분포들은 각각 다음과 같다.

$$\theta | (k, \mathbf{y}) \stackrel{d}{\sim} \text{Gamma} \left( a_1 + \sum_{t=1}^k y_t, k + b_1 \right) \quad (5.3.26)$$

$$\lambda | (k, \mathbf{y}) \stackrel{d}{\sim} \text{Gamma} \left( a_2 + \sum_{t=k+1}^n y_t, n - k + b_2 \right) \quad (5.3.27)$$

또한,  $k$ 의 사후확률밀도함수는 다음과 같다.

$$\pi(k | \theta, \lambda, \mathbf{y}) = \frac{L(\mathbf{y} | k, \theta, \lambda)}{\sum_{t=1}^n L(\mathbf{y} | t, \theta, \lambda)} \quad (5.3.28)$$

즉, 다음 식이 성립한다.

$$\pi(k | \theta, \lambda, \mathbf{y}) \propto \theta^{\sum_{t=1}^k y_t} \lambda^{\sum_{t=k+1}^n y_t} \exp(k[\lambda - \theta]) \quad (5.3.29)$$

지금까지 내용을 바탕으로, 모수들  $(\theta, \lambda, k)$ 의 사후확률분포에서 샘플링하는 Gibbs샘플러를 다음과 같이 요약할 수 있다.

**(1단계)** 초기모수들  $(\theta^{(0)}, \lambda^{(0)}, k^{(0)})$ 를 선택한다.

**(2단계)** 각  $j(= 1, 2, \dots)$ 에 대해서 주어진  $\mathbf{y}$ 에 대한 조건부샘플  $(\theta^{(j)}, \lambda^{(j)}, k^{(j)})$ 를 발생시킨다. 즉, 식 (5.3.26)의 사후조건부확률분포에서  $\theta^{(j)} | (k^{(j-1)}, \mathbf{y})$ 를 발생시키고, 식 (5.3.27)의 사후조건부확률분포에서  $\lambda^{(j)} | (k^{(j-1)}, \mathbf{y})$ 를 발생시키고, 식 (5.3.29)의 사후조건부확률분포에서  $k^{(j)} | (\theta^{(j)}, \lambda^{(j)}, \mathbf{y})$ 를 발생시킨다.

**(3단계)** 충분히 큰  $b$ 에 대해서  $(\theta^{(j)}, \lambda^{(j)}, k^{(j)})$ ,  $(j = b + 1, b + 2, \dots)$ 를 기록한다.

**예제 5.3.2** Gibbs샘플러를 사용해서 변환점을 조사하기 위해서, 다음 MATLAB프로그램 GibbsSamplerPoissonGamma101.m을 실행해 보자.

```

1 % -----
2 % Filename: GibbsSamplerPoissonGamma101.m
3 % Gibbs Sampler for Poisson-Gamma Distribution 1
4 % Programmed by CBS
5 % -----
6 function GibbsSamplerPoissonGamma101
7 clear, close all, format long
8 % Creditcard Data
9 % from http://ecos.bok.or.kr (' : fÿ°) ->

```

```

10 %      ->   °£ -> 5.8.1 f i ½p
11 [ yi mmonthdumi ]= xlsread('BOK_Data_CreditCard_Number');
12 mmonthi = cell2mat(mmonthdumi(6:end-3,1));
13 yi = yi(1:end-3);
14 % Time series plots
15 ni = length(mmonthi)           % Number of data points
16 nni = 1:ni;
17 subplot(2,1,1)
18 plot(nni,yi,'g',nni,yi,'k.','linewidth',2);
19 set(gca,'fontsize',11,'fontweigh','bold','xlim',[1 ni])
20 set(gca,'xtick',[1 13 25 37 49 61 73 85 97 ])
21 set(gca,'xticklabel',{'03/12','04/12','05/12','06/12','07/12', ...
22                      '08/12','09/12','10/12','11/12'})
23 ylabel('\bf No of Creditcards')
24 % Net numbers of newly issued creditdards
25 n = ni-1, nn = 1:n;
26 y = diff(yi);
27 y = round((y+4000)/500);
28 meany = mean(y)
29 stdy = std(y);
30 para = (meany/stdy)^2;
31 subplot(2,1,2)
32 plot(nn,y,'g',nn,y,'k.',[1 ni],[meany meany],'r-','linewidth',2);
33 set(gca,'fontsize',11,'fontweigh','bold','xlim',[1 ni])
34 set(gca,'xtick',[1 13 25 37 49 61 73 85 96 ])
35 set(gca,'xticklabel',{'04/01','05/01','06/01','07/01', ...
36                      '08/01','09/01','10/01','11/01','11/12'})
37 ylabel('\bf Increaseament')
38 saveas(gcf,'GibbsSamplerPoissonGamma101a','eps')
39 % Gibbs sampling
40 nG = 200           % Number of Gibbs' samples
41 a1 = para; a2 = para; b1 = 3; b2 = 3;
42 theta = zeros(1,nG);
43 lambda = zeros(1,nG);
44 k = zeros(1,n);
45 L = zeros(1,n);           % Likelihood of k
46 % Starting points.
47 k(1) = unidrnd(n,1,1);
48 % Note that k denotes an index of a hypothesized change-point
49 theta(1) = meany;
50 lambda(1) = meany;
51 % Gibbs' sampler
52 for ii=2:nG
53     kdum = k(ii-1);
54     % Generating theta
55     dum1 = a1 + sum(y(1:kdum));
56     lam1 = kdum + b1;
57     theta(ii) = gamrnd(dum1,1/lam1,1,1);
58     % Generating lambda
59     dum2 = a2 + sum(y) - sum(y(1:kdum));
60     lam2 = n-kdum+b2;
61     lambda(ii) = gamrnd(dum2,1/lam2,1,1);
62     % Posterior probability of k
63     for jj=1:n
64         L(jj) = exp((lambda(ii)-theta(ii))*jj) ...
65                 *(theta(ii)/lambda(ii))^sum(y(1:jj));
66     end
67     pipi = L/sum(L);
68     % Now sample the variate for k
69     k(ii) = cssample(1:n,pipi,1);
70 end

```

```

71 % Convergence plots
72 figure
73 subplot(2,1,1)
74 plot(1:nG,theta,'k-.',1:nG,lambda,'r','linewidth',2)
75 set(gca,'fontsize',11,'fontweigh','bold')
76 legend('\bf \theta','\bf \lambda','location','SE')
77 subplot(2,1,2)
78 plot(k,'k','linewidth',2)
79 set(gca,'fontsize',11,'fontweigh','bold')
80 ylabel('Change Point')
81 saveas(gcf,'GibbsSamplerPoissonGamma101b','eps')
82 % Histogram of changing point
83 figure
84 hist(k,20)
85 h = findobj(gca,'Type','patch');
86 set(h,'FaceColor','y','EdgeColor','k')
87 set(gca,'fontsize',11,'fontweigh','bold')
88 xlabel('\bf k')
89 saveas(gcf,'GibbsSamplerPoissonGamma101c','eps')
90 end
91 % End of program
92 % -----
93 function rs = cssample(x,p,n)
94 % CSSAMPLE: Random sample from an arbitrary discrete/finite distribution.
95 % R = CSSAMPLE(X,P,N) This function will take an arbitrary discrete and
96 %   finite distribution and return a random sample from it.
97 %   X = Support set (i.e., domain) of the distribution
98 %   P = Probability density function
99 %   N = Number of the generated variates
100 %   Programmed by W. L. Martinez and A. R. Martinez on 9/15/01
101 %   for Computational Statistics Toolbox
102 % -----
103 if length(x) ~= length(p)
104     error('The sizes of the input vectors do not match.')
105     return
106 end
107 % Sort just in case they are not in order
108 [xs,ind1] = sort(x);
109 ps = p(ind1); % Sort these in the same order as x
110 % Get the CDF
111 F = cumsum(ps);
112 % Find all of the required variates
113 for ii=1:n
114     u = rand(1,1);
115     if u <= F(1)
116         rs(ii) = x(1);
117     elseif u > F(end-1)
118         rs(ii) = x(end);
119     else
120         ind2 = find(u <= F);
121         rs(ii) = xs(ind2(1));
122     end
123 end
124 end
125 % End of program
126 % -----

```

이 MATLAB 프로그램 GibbsSamplerPoissonGamma101.m에서는 한국은행의 경제통계 시스템인 ECOS(Economic Statistics System)에서 다운로드한 엑셀파일

‘BOK\_Data\_CreditCard\_Number’를 분석한다. 이 데이터세트에는 2003년 12월부터 2012년 3월까지 각 월별 신용카드 총발급장수를 천장 단위로 기록한 것이다. 이명박정권이 들어선 이후 한국의 가계부채가 심각한 상황에 이르렀고 특히 다시 한 번 카드대란이 올 수 있다는 가능성이 제기되면서, 2012년 초부터 신용카드 발급에 제약을 가하는 조치가 취해졌다. 이 MATLAB프로그램은 Gibbs샘플러를 사용해서 2003년 12월부터 2011년 12월까지 신용카드 발급장수에 어떤 변환점이 있는가를 조사하기 위한 것이다.

이 MATLAB프로그램 GibbsSamplerPoissonGamma101.m을 실행하면, 산점도들을 그린 그림 5.3.2가 그려진다. 이 기간동안 신용카드 총발급장수의 시계열산점도가 그림 5.3.2의 상단 그래프이다. 이 상단 그래프에서 알 수 있듯이, 신용카드 총발급장수가 계속 증가한다. 이 신용카드 총발급장수는 저장(stock)이므로, 시계열산점도에서 변환점을 찾기가 어렵다. 이 신용카드 총발급장수를 차분을 취한 월별 신용카드 순발급장수의 시계열산점도가 그림 5.3.2의 하단 그래프이다. 이 하단 그래프에서 적색 실선은 이 순발급장수의 평균이다. 월별 순발급장수는 유량(flow)으로서, 이 하단 그래프에서 알 수 있듯이 이 순발급장수의 시계열산점도는 총발급장수의 시계열산점도보다 좀 더 정상적이다. 그러나, 2005년도 중반에 순발급장수의 수준이 약간 증가함을 관찰할 수 있다. 즉, 2005년도부터 좀 더 많은 신용카드들이 발급되었다고 추측할 수 있다.

이 MATLAB프로그램 GibbsSamplerPoissonGamma101.m을 실행하면, Gibbs샘플들  $\{[\theta^{(j)}, \lambda^{(j)}, k^{(j)}] \mid j = 1, 2, \dots, 200\}$ 가 발생되고, 이  $\{\theta^{(j)}\}$ 와  $\{\lambda^{(j)}\}$ 의 시계열산점도들이 그림 5.3.3에 그려진다. 그림 5.3.3에서 알 수 있듯이, 변환점 이전의 평균모수  $\theta$ 가 변환점 후의 평균모수  $\lambda$ 보다 작다는 것을 알 수 있다. 즉, 신용카드의 각 월별 순발급장수가 증가했다는 결론을 얻을 수 있다. 또한, 이 MATLAB프로그램을 실행하면 변환점  $k$ 의 사후확률밀도함수  $\pi(k \mid \theta, \lambda, \mathbf{y})$ 에 해당하는 히스토그램을 출력한다. 이 히스토그램이 그림 5.3.4에 그려져 있다. 이 히스토그램에서 알 수 있듯이, 변환점의 사후확률분포의 최빈값이  $k = 14$ 이라는 것을 알 수 있다. 이 최빈값은 2005년 2월에 대응한다.

다음 Python프로그램 GibbsSamplerPoissonGamma101.py를 실행하면, MATLAB프로그램 GibbsSamplerPoissonGamma101.m을 실행한 것과 동일한 결과를 얻는다.

```

1 # GibbsSamplerPoissonGamma101.py
2 """
3 Filename : GibbsSamplerPoissonGamma101.py
4 Gibbs Sampler for Poisson-Gamma Distribution 1
5 Programmed by Hwang Seol Woong based on the R program by CBS
6 """
7
8

```

```

9 from numpy import *
10 import pandas as pd
11 import matplotlib.pyplot as plt
12
13
14 table = pd.read_excel('bokdata_credit_card.xlsx')
15
16
17
18 #def GibbsSamplerPoissonGamma10
19
20 yi = matrix(table)[: ,0]
21 ni = len(yi)
22 nni = arange(1, ni+1)
23 n = ni-1
24
25
26
27 ## number of card
28
29
30 fig, ax = plt.subplots( figsize=(16,6))
31
32 ax.plot(yi/1000, alpha = 0.3, linewidth = 4, color='orangered') #color='
    lightseagreen', alpha=1, ls='--'
33 ax.plot(yi/1000, color='orangered', ls=':', linewidth = 4)# linestyle='--')
34 ax.tick_params(axis="y",direction = "in") # axis ticks inside
35 ax.tick_params(axis="x",direction = "in") # axis ticks inside
36 ax.set_xlabel('Time')
37 ax.set_ylabel('Number of Creditcards(thousand)')
38
39 locs = arange(len(table.index),step=14) # locations
40 plt.xticks( locs, table.index[locs])
41
42
43 ## number of card newly issued
44
45 y = yi[1:] - yi[0:-1]
46 y = (y+4000)/500
47 meany = mean(y)
48 ymean = mean(y)*matrix(ones(len(y)))
49
50 fig, ax = plt.subplots( figsize=(16,6))
51
52 ax.plot(table.index[1:], array(y), linewidth = 3) #color='lightseagreen', alpha
    =1, ls='--'
53 ax.plot( array(ymean.T), linewidth = 3, ls='--')
54 ax.tick_params(axis="y",direction = "in") # axis ticks inside
55 ax.tick_params(axis="x",direction = "in") # axis ticks inside
56 ax.set_xlabel('Time')
57 ax.set_ylabel('Increasement')
58
59 locs = arange(len(table.index),step=14) # locations
60 plt.xticks( locs, table.index[locs])
61
62
63
64
65 def CSSAMPLE(x, p, n):
66     if len(x) != len(p):
67         print('The sizes of the input vectors do not match.')
```

```

68     F = cumsum(p)
69     rs = zeros(n)
70
71     for i in range(n):
72         u = random.rand(1)
73         if u <= F[0]:
74             rs[i] = x[1]
75         elif u > F[-2]:
76             rs[i] = x[-1]
77         else :
78             ind2 = len(F) - len(F.compress(u<F))-1
79             #ind2 = ind2.astype('int')
80             rs[i] = x[ind2]
81     return rs
82
83
84
85     ## Gibbs Sampling
86     random.seed(210)
87     nG = 200
88     n= 200
89
90     y = float64(y)
91     yst = y.std()
92
93     a1 = (mean(y)/yst)**2
94     a2 = (mean(y)/yst)**2
95     b1 = 3
96     b2 = 3
97     theta = zeros(nG)
98     lamb = zeros(nG)
99     k = zeros(n)
100    L = zeros(n)
101
102    k[0] = random.randint(n+1)
103
104    theta[0] = meany
105    lamb[0] = meany
106
107    for i in range(nG-1):
108        kdum = k[i]
109        kdum = kdum.astype('int')
110        dum1 = a1 + sum(y[0:kdum])
111        lam1 = kdum +b1
112        theta[i+1]= random.gamma(dum1, 1/lam1)
113
114        dum2 = a2 +sum(y) - sum(y[0:kdum])
115        lam2 = n -kdum +b2
116        lamb[i+1] = random.gamma(dum2, 1/lam2)
117
118        for j in range(n):
119            L[j] = exp( (lamb[i] - theta[i])*j) * (theta[i]/lamb[i] )**sum(y[0:j])
120
121        pipi = L/sum(L)
122        k[i+1] = CSSAMPLE(arange(1, n+1), pipi,1)
123
124
125    fig, ax = plt.subplots(figsize=(16,6))
126
127    ax.plot(theta, label='theta', color ='orangered', linewidth=3.0)
128    ax.plot(lamb, label='lambda', linewidth=3.0)

```

```

129 ax.legend()
130 ax.legend(loc='upper right')
131 ax.tick_params(axis="y",direction = "in") # axis ticks inside
132 ax.tick_params(axis="x",direction = "in") # axis ticks inside
133 ax.set_xlabel('Time')
134 ax.grid(linewidth=1,ls='--' )
135
136
137 fig, ax = plt.subplots(figsize=(16,6))
138
139 ax.plot(L, label='k', color = 'orangered', linewidth=3.0)
140 #ax.plot(lamb, label='lambda', linewidth=3.0)
141 ax.legend()
142 ax.legend(loc='upper right')
143 ax.tick_params(axis="y",direction = "in") # axis ticks inside
144 ax.tick_params(axis="x",direction = "in") # axis ticks inside
145 ax.set_xlabel('Time')
146 ax.grid(linewidth=1,ls='--' )
147
148
149 # End of program
150 # -----

```

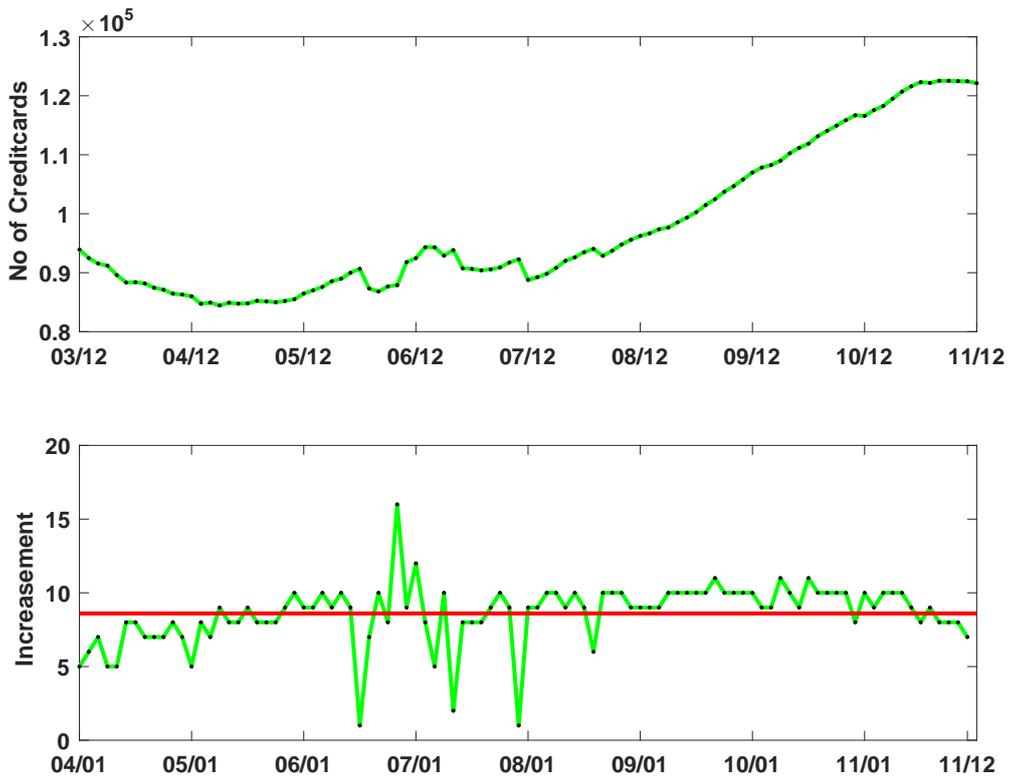


그림 5.3.2. 신용카드 총발급장수와 차분의 시계열산점도

**예제 5.3.3** 예제 5.3.2보다 약간 더 복잡한 모형을 사용해서 시계열의 변환점을 추정하는

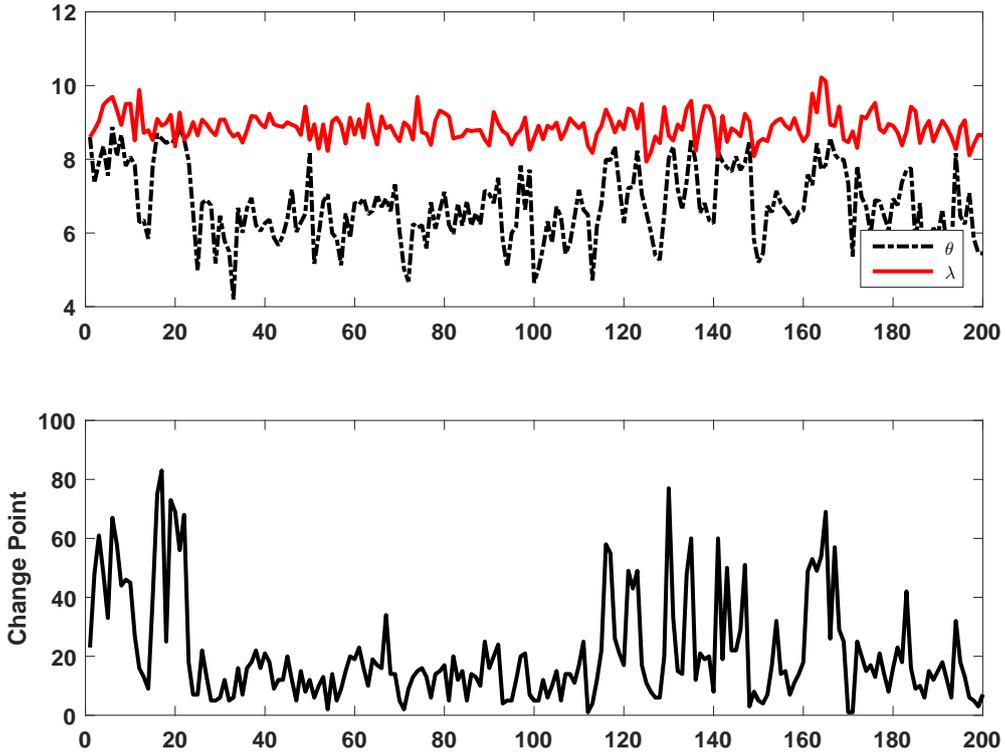


그림 5.3.3. Gibbs샘플들

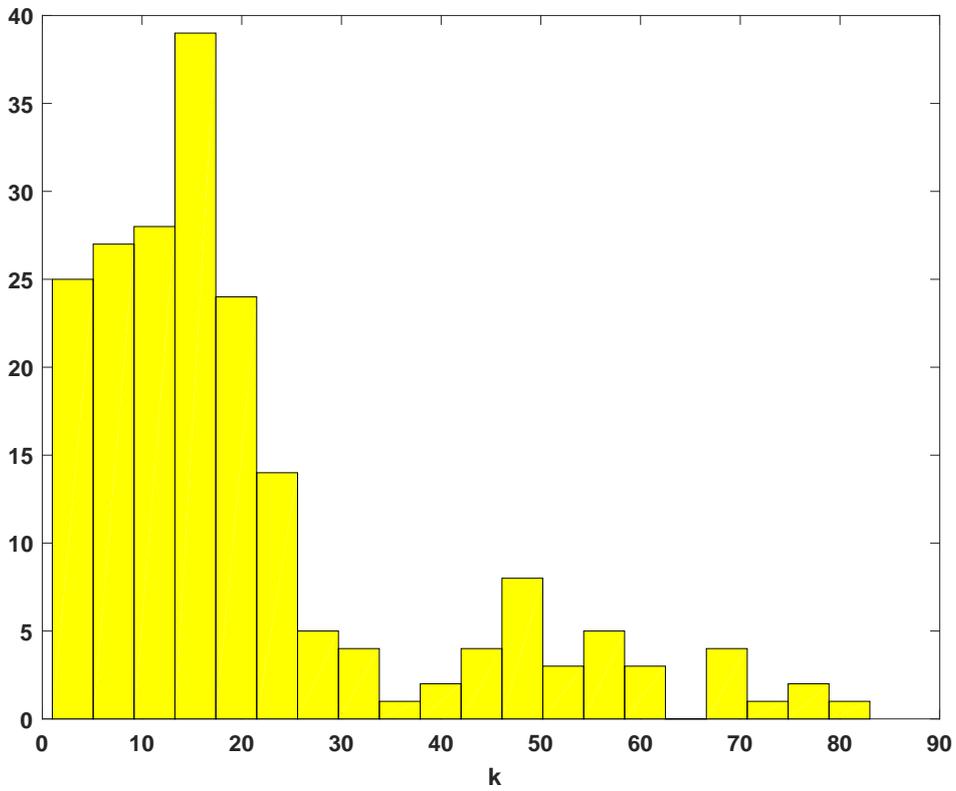


그림 5.3.4. 변환점의 히스토그램

문제를 살펴보자.

Raftery & Akman [81]과 Carlin & Gelfand & Smith [23]에서는 영국의 석탄광산에서 1851년부터 1962년까지 112년 동안 재해율이 감소되었다는 가설을 베이저안통계기법을 사용해서 검정하였다. 이 분석에서 사용된 데이터에서 알 수 있듯이, 전반에는 비교적 탄광 재해 건수가 많고, 후반에는 건수가 적다. 기술면에서 그리고 안전면에서 개선이 일어나서 탄광재해가 일어날 확률이 감소했다고 여겨지지만, 과연 그 변환점은 언제일까? 또한 그 전후에 평균 재해건수가 어떻게 변화했는가? 이러한 의문에 대해서 변환점이 있는 Poisson 확률모형을 적용해서 살펴보자.

Robert [83]가 제시한 모형을 사용해서 이 데이터세트를 분석하기로 하자. 이 모형에서는 제  $i$ 재 해의 재해횟수  $y_i$ 가 제  $k$ 번째 해까지 평균모수가  $\theta$ 인 Poisson 확률분포를 따르고, 제  $k+1$ 번째 해부터는 평균모수가  $\lambda$ 인 Poisson 확률분포를 따른다고 가정한다. 즉, 다음 식들이 성립한다고 가정한다.

$$y_i \stackrel{d}{\sim} \text{Poisson}(\theta), \quad (i = 1, 2, \dots, k) \quad (1)$$

$$y_i \stackrel{d}{\sim} \text{Poisson}(\lambda), \quad (i = k + 1, k + 2, \dots, n) \quad (2)$$

여기서  $n = 112$ 이다. 모수들의 사전확률분포는 다음과 같다.

$$\theta \stackrel{d}{\sim} \text{Gamma}(a_1, b_1) \quad (3)$$

$$\lambda \stackrel{d}{\sim} \text{Gamma}(a_2, b_2) \quad (4)$$

$$b_1 \stackrel{d}{\sim} \text{Gamma}(c_1, d_1) \quad (5)$$

$$b_2 \stackrel{d}{\sim} \text{Gamma}(c_2, d_2) \quad (6)$$

$$k \stackrel{d}{\sim} \text{Uniform}\{1, 2, \dots, n\} \quad (7)$$

또한,  $\theta$ ,  $\lambda$ , 그리고  $k$ 는 서로 독립이다. 식 (7)에서 알 수 있듯이,  $k$ 는 지지대가  $\{1, 2, \dots, n\}$ 인 일양확률분포를 따른다. 식 (3)~식 (6)에서 알 수 있듯이, 이 모형은 예제 5.3.2보다 한 단계가 더 많은 베이저안모형이다.

관찰벡터를  $\mathbf{y} = [y_1, y_2, \dots, y_n]^t$ 라 하면, 우도함수는 다음과 같다.

$$L(\mathbf{y} | k, \theta, \lambda) \propto \exp(-k\theta) \theta^{\sum_{t=1}^k y_t} \exp(-[n-k]\lambda) \lambda^{\sum_{t=k+1}^n y_t} \quad (8)$$

예제 5.3.2에서와 같은 방법을 사용해서, 다음과 같은 사후조건부확률분포들을 유도할 수 있다.

$$\theta | (\lambda, b_1, b_2, k, \mathbf{y}) \stackrel{d}{\sim} \text{Gamma} \left( a_1 + \sum_{i=1}^k y_i, k + b_1 \right) \quad (9)$$

$$\lambda | (\theta, b_1, b_2, k, \mathbf{y}) \stackrel{d}{\sim} \text{Gamma} \left( a_2 + \sum_{i=k+1}^n y_i, n - k + b_2 \right) \quad (10)$$

$$b_1 | (\theta, \lambda, b_2, k, \mathbf{y}) \stackrel{d}{\sim} \text{Gamma} (a_1 + c_1, \theta + d_1) \quad (11)$$

$$b_2 | (\theta, \lambda, b_1, k, \mathbf{y}) \stackrel{d}{\sim} \text{Gamma} (a_2 + c_2, \lambda + d_2) \quad (12)$$

또한, 변곡점  $k$ 의 사후확률밀도함수는 다음과 같다.

$$\pi(k | \theta, \lambda, b_1, b_2, \mathbf{y}) = \frac{L(\mathbf{y} | k, \theta, \lambda)}{\sum_{j=1}^n L(\mathbf{y} | j, \theta, \lambda)} \quad (13)$$

Gibbs샘플러를 사용해서 변환점을 조사하기 위해서, 다음 MATLAB 프로그램 GibbsSamplerPoissonGamma102.m을 실행해 보자.

```

1 % -----
2 % Filename: GibbsSamplerPoissonGamma102.m
3 % Gibbs Sampler for Poisson-Gamma Distribution 2
4 % Programmed by CBS
5 % -----
6 clear, close all
7 % Coal Disaster Data
8 % from http://orion.math.iastate.edu/burkardt/data/martinez/martinez.html
9 yyear = [1851:1962]';
10 y = [ 4 5 4 1 0 4 3 4 0 6 3 3 4 0 2 6 3 3 5 4 5 3 ...
11       1 4 4 1 5 5 3 4 2 5 2 2 3 4 2 1 3 2 2 1 1 1 ...
12       1 3 0 0 1 0 1 1 0 0 3 1 0 3 2 2 0 1 1 1 0 1 ...
13       0 1 0 0 0 2 1 0 0 0 1 1 0 2 3 3 1 1 2 1 1 1 ...
14       1 2 4 2 0 0 0 1 4 0 0 0 1 0 0 0 0 1 0 0 1 0 1 ]';
15 n = length(yyear) % Number of data points
16 plot(yyear,y,'g',yyear,y,'k*','linewidth',2);
17 set(gca,'fontsize',11,'fontweigh','bold')
18 axis([1850 1963 -0.5 6.5])
19 ylabel('\bf No of Disasters')
20 saveas(gcf,'GibbsSamplerPoissonGamma102a','eps')
21 % Parameter values by Roberts (2000)
22 rng((1001^2+1)/2,'twister')
23 nG = 200 % Number of Gibbs' samples
24 a1 = 0.5; a2 = 0.5; c1 = 0; c2 = 0; d1 = 1; d2 = 1;
25 theta = zeros(1,nG);
26 lambda = zeros(1,nG);
27 k = zeros(1,n);
28 L = zeros(1,n); % Likelihood of k
29 % Starting points.
30 k(1) = unidrnd(n,1,1);

```

```

31 % Note that k denotes an index of a hypothesized change-point
32 theta(1) = 1;
33 lambda(1) = 1;
34 b1 = 1;
35 b2 = 1;
36 % Gibbs' sampler
37 for ii=2:nG
38     kdum = k(ii-1);
39     % Generating theta
40     dum1 = a1 + sum(y(1:kdum));
41     lam1 = kdum + b1;
42     theta(ii) = gamrnd(dum1,1/lam1,1,1);
43     % Generating lambda
44     dum2 = a2 + sum(y) - sum(y(1:kdum));
45     lam2 = n-kdum+b2;
46     lambda(ii) = gamrnd(dum2,1/lam2,1,1);
47     % Generating b1 and b2
48     b1 = gamrnd(a1+c1,1/(theta(ii)+d1),1,1);
49     b2 = gamrnd(a2+c2,1/(lambda(ii)+d2),1,1);
50     % Posterior probability of k
51     for jj=1:n
52         L(jj) = exp((lambda(ii)-theta(ii))*jj) ...
53         *(theta(ii)/lambda(ii))^sum(y(1:jj));
54     end
55     pipi = L/sum(L);
56     % Now sample the variate for k
57     k(ii) = cssample(1:n,pipi,1);
58 end
59 % Convergence plots
60 figure
61 subplot(2,1,1)
62 plot(1:nG,theta,'k-.',1:nG,lambda,'r','linewidth',2)
63 set(gca,'fontsize',11,'fontweigh','bold')
64 legend('\bf \theta','\bf \lambda','location','SE')
65 axis([1 nG -1 4])
66 subplot(2,1,2)
67 plot(k,'k','linewidth',2)
68 set(gca,'fontsize',11,'fontweigh','bold')
69 ylabel('Change Point')
70 saveas(gcf,'GibbsSamplerPoissonGamma102b','eps')
71 % Histogram of changing point
72 figure
73 hist(k,20:60)
74 set(get(gca,'child'),'FaceColor','green','EdgeColor','r')
75 % set(gca,'fontsize',11,'fontweigh','bold')
76 xlabel('\bf k')
77 saveas(gcf,'GibbsSamplerPoissonGamma102c','eps')
78 % End of program
79 % -----

```

이 MATLAB 프로그램 GibbsPoissonGamma102.m을 실행하면, 재해횟수의 시계열산점도인 그림 5.3.5가 그려진다. 그림 5.3.5에서 석탄광산의 재해율이 감소하는 추세를 보인다는 것을 시각적으로 알 수 있다.

이 MATLAB 프로그램을 실행하면, Gibbs샘플들  $\left\{ [\theta^{(j)}, \lambda^{(j)}, k^{(j)}] \mid j = 1, 2, \dots, 200 \right\}$ 가 발생된다. 이들의 시계열산점도들이 그림 5.3.6에 그려져 있다. 그림 5.3.6에서 알 수 있듯이, 변환점 이전의 평균모수  $\theta$ 가 변환점 후의 평균모수  $\lambda$  보다 크다. 즉, 재해율이 감소했다는

결론을 얻을 수 있다.

이 MATLAB 프로그램을 실행하면, 변환점  $k$ 의 사후확률밀도함수  $\pi(k | \mathbf{y}, \theta, \lambda, b_1, b_2)$ 에 해당하는 히스토그램을 출력한다. 이 히스토그램이 그림 5.3.7에 그려져 있다. 이 히스토그램에서 알 수 있듯이, 변환점의 사후확률분포의 최빈값이  $k = 41$ 이라는 것을 알 수 있다. 이 최빈값은 1891년에 대응한다. ■

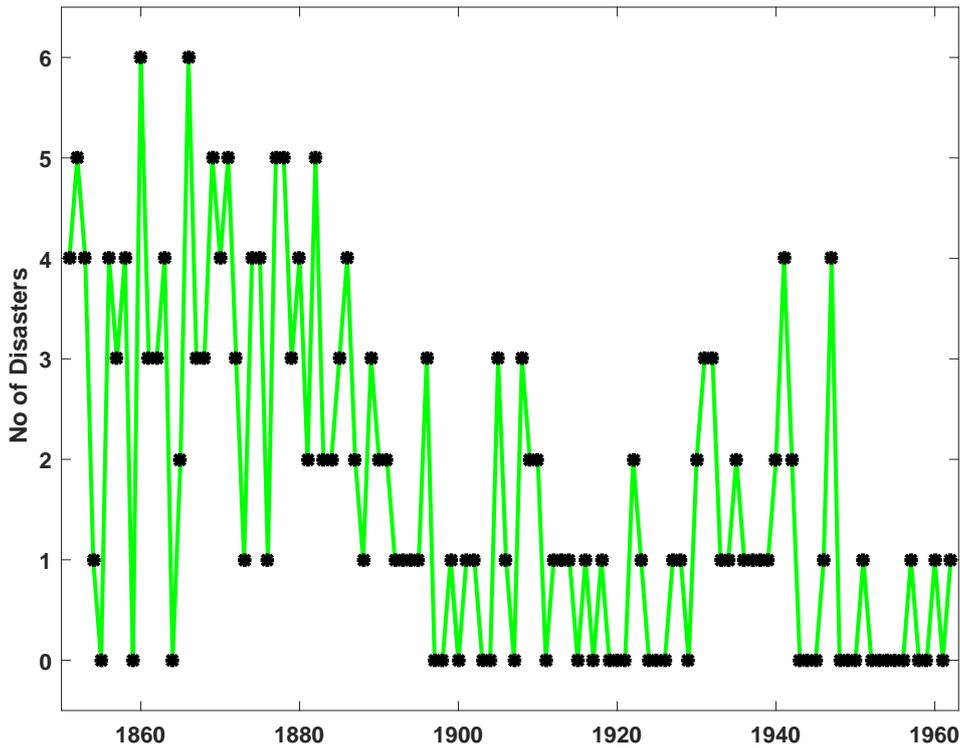


그림 5.3.5. 재해횟수의 시계열산점도

### 5.3.3 선형회귀모형과 Gibbs샘플러

다음 선형회귀모형을 살펴보자.

$$\mathbf{y} | (\boldsymbol{\beta}, \psi) \stackrel{d}{\sim} \mathcal{N}(X\boldsymbol{\beta}, \psi^{-1}I_n) \tag{5.3.30}$$

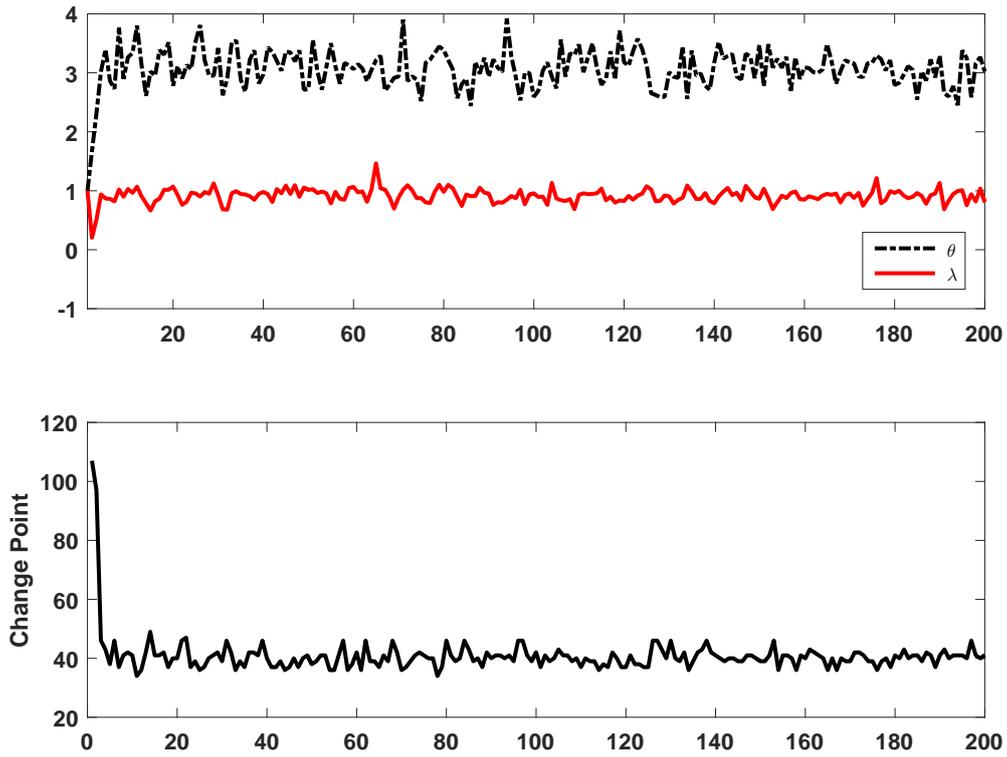


그림 5.3.6. Gibbs샘플들

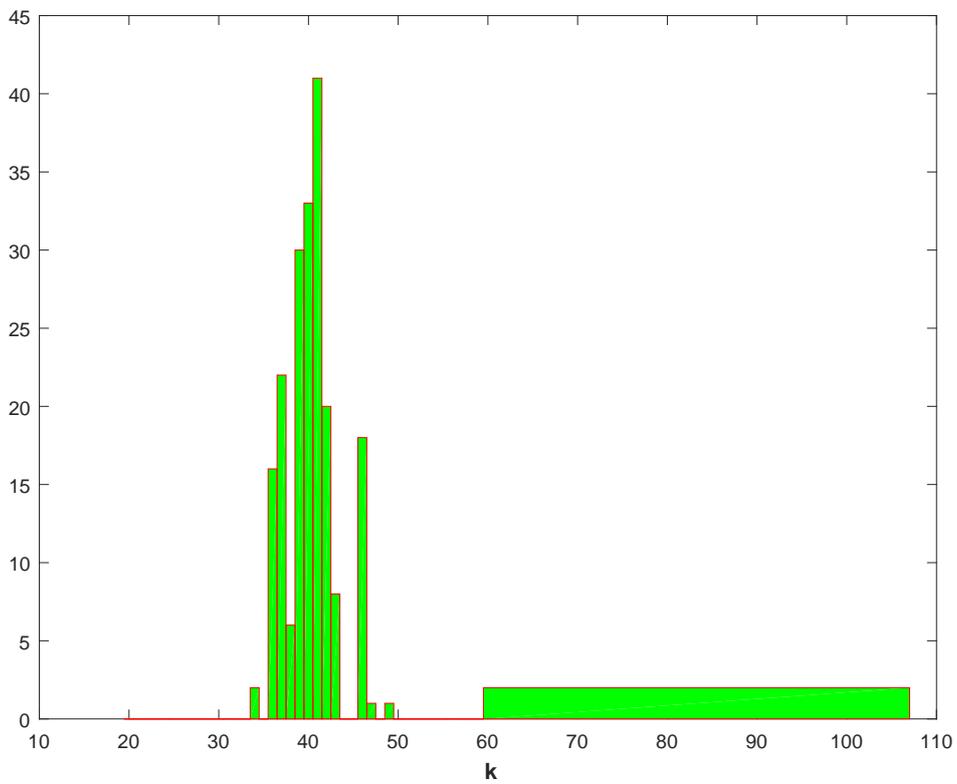


그림 5.3.7. 변환점의 히스토그램

여기서  $I_n$ 은  $n$ 차원 단위행렬이고,  $\mathbf{y}$ ,  $X$ , 그리고  $\boldsymbol{\beta}$ 는 각각 다음과 같다.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nK} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} \quad (5.3.31)$$

정밀도  $\psi$ 는 분산  $\sigma^2$ 의 역수이다. 식 (5.3.30)과 식 (5.3.31)에서 알 수 있듯이, 오차벡터  $\boldsymbol{\epsilon} \doteq [\epsilon_1, \epsilon_2, \dots, \epsilon_n]^T$ 의 원소들  $\{\epsilon_j\}$ 는 서로 독립이며 정규확률분포  $\mathcal{N}(0, \sigma^2)$ 를 따른다.

우선 Gibbs샘플러를 사용하지 않고, 모수벡터  $(\boldsymbol{\beta}, \psi)$ 의 사후조건부확률분포에서 샘플들을 얻는 방법을 살펴보자. 모수벡터  $(\boldsymbol{\beta}, \psi)$ 의 사전확률분포가 다음과 같다고 하자.

$$\boldsymbol{\beta} | \psi \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_0, \psi^{-1}B_0), \quad \psi \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \quad (5.3.32)$$

관찰벡터  $\mathbf{y} = [y_1, y_2, \dots, y_n]^t$ 의 우도함수는 다음과 같다.

$$f(\mathbf{y} | \boldsymbol{\beta}, \psi) \propto \psi^{n/2} \exp\left(-\frac{\psi}{2}[\mathbf{y} - X\boldsymbol{\beta}]^t[\mathbf{y} - X\boldsymbol{\beta}]\right) \quad (5.3.33)$$

따라서 모수벡터  $(\boldsymbol{\beta}, \psi)$ 의 사후확률분포는 다음과 같다.

$$\boldsymbol{\beta} | (\psi, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_*, \psi^{-1}B_*), \quad \psi | \mathbf{y} \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_*}{2}, \frac{n_* S_*}{2}\right) \quad (5.3.34)$$

여기서 상수들, 행렬들과 벡터는 다음과 같다.

$$n_* \doteq n_0 + n \quad (5.3.35)$$

$$S_* \doteq \frac{1}{n_*} \{n_0 S_0 + [\mathbf{y} - X\mathbf{b}_*]^t \mathbf{y} + [\mathbf{b}_0 - \mathbf{b}_*]^t B_0^{-1} \mathbf{b}_0\} \quad (5.3.36)$$

$$B_* \doteq [B_0^{-1} + X^t X]^{-1} \quad (5.3.37)$$

$$\mathbf{b}_* \doteq B_* [B_0^{-1} \mathbf{b}_0 + X^t \mathbf{y}] \quad (5.3.38)$$

식 (5.3.34)에서 알 수 있듯이,  $\psi$ 의 사후주변확률분포는 감마확률분포이다. 또한, 식 (5.3.34)을 이용해서  $\boldsymbol{\beta}$ 의 사후주변확률분포가 다변량  $t$ 확률분포임을 증명할 수 있다. 따라서, 이에 해당하는 사후확률밀도함수들을 해석적으로 구할 수 있다. 식 (5.3.34)의 사후확률분포에서 먼저  $\psi | \mathbf{y}$ 를 발생시킨다. 이렇게 얻어진  $\psi$ 를 이용해서  $\boldsymbol{\beta} | (\psi, \mathbf{y})$ 를 발생시키면, 모수  $(\boldsymbol{\beta}, \psi)$

의 독립된 샘플들을 얻는다. 따라서, 이 경우에는 보통 몬테카를로법에 의한 시뮬레이션이 가능하므로, Gibbs샘플러를 사용할 필요는 없다. 식 (5.3.32)의 사전확률분포와 식 (5.3.34)의 사후확률분포가 공액임에 유의하라.

이번에는 Gibbs샘플러를 사용해서 모수벡터  $(\boldsymbol{\beta}, \psi)$ 의 사후조건부확률분포에서 샘플들을 얻는 절차를 살펴보자. 사전확률분포가 다음과 같다고 하자.

$$\boldsymbol{\beta} | \psi \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_0, B_0) \quad (5.3.39)$$

$$\psi \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \quad (5.3.40)$$

이 경우 사후조건부확률분포는 다음과 같다.

$$\boldsymbol{\beta} | (\psi, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_+, B_+) \quad (5.3.41)$$

$$\psi | (\boldsymbol{\beta}, \mathbf{y}) \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_+}{2}, \frac{n_+ S_+}{2}\right) \quad (5.3.42)$$

여기서 상수, 행렬들과 벡터는 다음과 같다.

$$n_+ \doteq n_0 + n \quad (5.3.43)$$

$$S_+ \doteq \frac{1}{n_+} \{n_0 S_0 + [\mathbf{y} - X\boldsymbol{\beta}]^t [\mathbf{y} - X\boldsymbol{\beta}]\} \quad (5.3.44)$$

$$B_+ \doteq [B_0^{-1} + \psi X^t X]^{-1} \quad (5.3.45)$$

$$\mathbf{b}_+ \doteq B_+ [B_0^{-1} \mathbf{b}_0 + \psi X^t \mathbf{y}] \quad (5.3.46)$$

지금까지 내용을 바탕으로, 모수들  $(\boldsymbol{\beta}, \psi)$ 의 사후조건부확률분포들에서 샘플링하는 Gibbs 샘플러를 다음과 같이 요약할 수 있다.

**(1단계)** 초기모수들  $(\boldsymbol{\beta}^{(0)}, \psi^{(0)})$ 를 선택한다.

**(2단계)** 각  $j(= 1, 2, \dots)$ 에 대해서 주어진  $\mathbf{y}$ 에 대한 조건부샘플  $(\boldsymbol{\beta}^{(j)}, \psi^{(j)})$ 를 발생시킨다.

즉, 식 (5.3.41)의 사후조건부확률분포에서  $\boldsymbol{\beta}^{(j)} | (\psi^{(j-1)}, \mathbf{y})$ 를 발생시키고, 식 (5.3.42)의 사후조건부확률분포에서  $\psi^{(j)} | (\boldsymbol{\beta}^{(j)}, \mathbf{y})$ 를 발생시킨다.

**(3단계)** 충분히 큰  $b$ 에 대해서  $(\boldsymbol{\beta}^{(j)}, \psi^{(j)})$ ,  $(j = b + 1, b + 2, \dots)$ 를 기록한다.

**예제 5.3.4** Gibbs샘플러를 사용해서 선형회귀모형을 추정하는 예로서, 다음 R프로그램 GibbsSamplerLinearRegression101R.R을 실행해 보자.

```

1 # -----
2 # Filename: GibbsSamplerLinearRegression101R.R
3 # Gibbs Sampler for Linear Regression
4 # Programmed by CBS
5 # -----
6 # Making Data and Classical OLS (The same as Example 2.3.1)
7 set.seed(3)
8 library(MASS)
9 x1 <- c(-10:10)
10 ( x2 <- sin(x1) )
11 ( sigmay <- 0.5 )
12 y <- pi*x1 + 0.123*x2 + sigmay*rnorm(21,0,1)
13 ny <- length(y)
14 lm(formula = y ~ x1 + x2 - 1) # -1 means no intercept
15 X <- cbind(x1,x2)
16 XtX <- t(X)%*%X
17 Xty <- t(X)%*%y
18 yty <- t(y)%*%y
19
20 # Prior
21 ( b0 <- matrix(c(3.0,0.1),2,1 ) ); ( B0 <- diag(2) );
22 ( n0 <- 3 ); ( S0 <- 1.1 )
23 ( shape0 <- n0/2 ); ( rate0 <- n0*S0/2 );
24 Nbeta <- 20; Npsi <- 20
25 beta1.gd <- seq(-4,4,length=Nbeta)
26 beta2.gd <- seq(-4,4,length=Nbeta)
27 psi.gd <- seq(0.01,2.00,length=Npsi)
28 priorPDF <- array(0, dim=c(Nbeta,Nbeta,Npsi))
29 prior.psi <- 1/dgamma(psi.gd,shape0,rate0)
30 DUMBeta <- rbind(beta1.gd,beta2.gd) - b0%*%rep(1,Nbeta)
31 prior.beta <- 1/(2*pi*sqrt(det(B0)))*
32     exp(-1/2*t(DUMBeta)%*%ginv(B0)%*%DUMBeta)
33 for (kk in 1:Npsi){
34     for (ii in 1:Nbeta){
35         for (jj in 1:Nbeta){
36             priorPDF[ii,jj,kk] <- prior.psi[kk]*prior.beta[ii,jj]
37         }
38     }
39 }
40
41 # Posterior through Gibbs Sampler
42 set.seed(1)
43 Nsim <- 300; burnin <-1000
44 Nobs <- Nsim + burnin
45 dumB <- ginv(B0)%*%b0
46 nPost <- n0 + ny
47 psiPost <- 1/S0
48 theta <- matrix(ncol=3, nrow=Nobs)
49 for (ii in 1:Nobs){
50     BP <- ginv(ginv(B0)+psiPost*XtX)
51     bP <- BP%*%(dumB+psiPost*Xty)
52     betaPost <- mvrnorm(n = 1, bP, BP, tol = 1e-6,
53         empirical = FALSE, EISPACK = FALSE)
54     regDUM <- y-X%*%betaPost
55     Spost <- 1/nPost*(n0*S0 + t(regDUM)%*%regDUM)
56     psiPost <- rgamma(1,nPost/2,nPost*Spost/2)

```

```

57         theta[ii, ] <- c(betaPost, psiPost)
58     }
59
60 # Plotting
61 library(ggplot2)
62 # install.packages("grid")
63 library(grid)
64 setEPS()
65 plot.new()
66 postscript('GibbsSamplerLinearRegression101R.eps') # Start to save figure
67 timee = c(1:Nobs)
68 GibbsData <- data.frame(timee, theta)
69 plot1 <- ggplot(GibbsData, aes(theta[,1], theta[,2])) +
70     geom_point(colour="dark red", lwd=0.9) +
71     xlab(expression(beta_1)) +
72     ylab(expression(beta_2))
73 plot2 <- ggplot(GibbsData, aes(x=theta[,1])) +
74     xlab(expression(beta_1)) +
75     geom_histogram(aes(y = ..density..), bins=30,
76         # breaks=seq(0.0, 2.5, by=0.05),
77         col="red", fill="blue") +
78     geom_density(lwd=1.0)
79 plot3 <- ggplot(GibbsData, aes(x=theta[,2])) +
80     xlab(expression(beta_2)) +
81     geom_histogram(aes(y = ..density..), bins=30,
82         # breaks=seq(0.0, 2.5, by=0.05),
83         col="red", fill="green") +
84     geom_density(lwd=1.0)
85 plot4 <- ggplot(GibbsData, aes(x=theta[,3])) +
86     xlab(expression(psi)) +
87     geom_histogram(aes(y = ..density..), bins=30,
88         # breaks=seq(0.0, 2.5, by=0.05),
89         col="red", fill="pink") +
90     geom_density(lwd=1.0)
91 pushViewport(viewport(layout = grid.layout(2,2)))
92 print(plot1, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
93 print(plot2, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
94 print(plot3, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
95 print(plot4, vp = viewport(layout.pos.row = 2, layout.pos.col = 2))
96 dev.off() # End to save figure
97 # -----

```

이 R 프로그램 GibbsSamplerLinearRegression101R.R을 실행하면, 사후확률분포에 Gibbs 샘플러를 적용해서 발생시킨 샘플들을 그린 그림 5.3.8을 출력한다. 그림 5.3.8의 좌측상단 그래프에는  $[\beta_1, \beta_2]$ 의 산점도가 그려져 있고, 우측상단 그래프에는  $\beta_1$ 의 히스토그램이, 좌측 하단 그래프에는  $\beta_2$ 의 히스토그램이, 그리고 우측하단 그래프에는  $\psi$ 의 히스토그램이 그려져 있다. ■

베이지안추정법을 사용해서 오차항이  $t$  확률분포를 따르는 선형회귀모형을 추정하는 문제를 살펴보자. 즉, 다음과 같은 선형회귀모형을 살펴보자.

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (5.3.47)$$

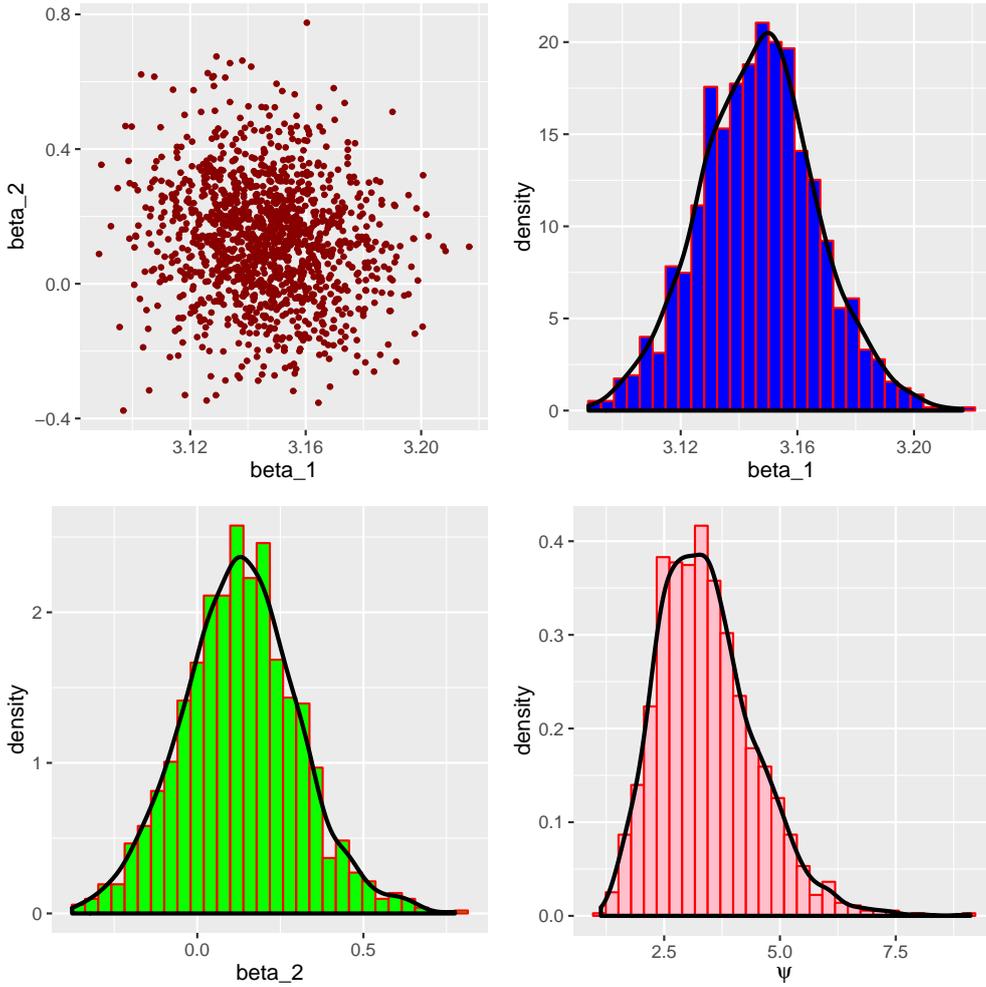


그림 5.3.8. 재해횟수의 시계열산점도

여기에서  $\mathbf{y}$ ,  $\boldsymbol{\epsilon}$ ,  $X$  그리고  $\boldsymbol{\beta}$ 는 각각 다음과 같다.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nK} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} \quad (5.3.48)$$

표준적 선형회귀모형과는 달리, 오차항  $\epsilon_i$ 가 다음 확률분포를 따르는 모형을 생각해 보자.

$$\epsilon_i | \eta_i \stackrel{d}{\sim} \mathcal{N}\left(0, \frac{1}{\eta_i} \sigma^2\right), \quad \eta_i \stackrel{d}{\sim} \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2}\right) \quad (5.3.49)$$

여기서  $\nu_0$ 는 알고 있는 상수이다. 식 (5.3.49)에서 알 수 있듯이, 오차항  $\epsilon_i$ 의 확률밀도함수는

다음과 같다.

$$\begin{aligned}
 f(\epsilon_i) &= \int f(\epsilon_i | \eta_i) f(\eta_i) d\eta_i \\
 &\propto \int_0^\infty \exp\left(-\frac{\eta_i}{\sigma^2} \epsilon_i\right) \epsilon_i^{\nu_0/2-1} \exp\left(-\frac{\nu_0}{2} \epsilon_i\right) d\epsilon_i \\
 &\propto \left[1 + \frac{1}{\nu_0} \frac{\epsilon_i^2}{\sigma^2}\right]^{-\frac{\nu_0+1}{2}} \quad (5.3.50)
 \end{aligned}$$

즉,  $\epsilon_i$ 는 자유도가  $\nu_0$ 인  $t$  확률분포인  $t_{\nu_0}(0, \sigma^2)$ 를 따른다. 여기서 자유도가  $\nu$ 인  $t$  확률분포  $t_\nu(\mu, \sigma^2)$ 의 확률밀도함수  $f(x)$ 는 다음과 같다.

$$f(x | \nu, \mu, \sigma^2) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}\right)} \left\{1 + \frac{1}{\nu} \frac{[x - \mu]^2}{\sigma^2}\right\}^{-\frac{\nu+1}{2}} \quad (5.3.51)$$

다음과 같이 대각행렬  $H$ 와 정밀도  $\psi$ 를 정의하자.

$$H \doteq \text{diag}(\eta_1, \eta_2, \dots, \eta_m), \quad \psi \doteq \frac{1}{\sigma^2} \quad (5.3.52)$$

식 (5.3.47)과 식 (5.3.49)에서 알 수 있듯이, 관찰벡터  $\mathbf{y} = [y_1, y_2, \dots, y_n]^t$ 의 우도함수는 다음과 같다.

$$\mathbf{y} | (\boldsymbol{\beta}, \psi, H) \stackrel{d}{\sim} \mathcal{N}\left(X\boldsymbol{\beta}, \frac{1}{\psi}H^{-1}\right) \quad (5.3.53)$$

사전확률분포가 다음과 같다고 하자.

$$\boldsymbol{\beta} \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_0, B_0), \quad \psi \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right), \quad \eta_i \stackrel{d}{\sim} \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0}{2}\right) \quad (5.3.54)$$

따라서, 사후결합확률밀도함수는 다음과 같다.

$$\begin{aligned}
 \pi(\boldsymbol{\beta}, \psi, H | \mathbf{y}) &\propto \psi^{[n+n_0-2]/2} \exp\left(-\frac{\psi}{2} \{\mathbf{y} - X\boldsymbol{\beta}\}^t H \{\mathbf{y} - X\boldsymbol{\beta}\} + n_0 S_0\right) \\
 &\cdot \exp\left(-\frac{1}{2} [\boldsymbol{\beta} - \mathbf{b}_0]^t B_0^{-1} [\boldsymbol{\beta} - \mathbf{b}_0]\right) \prod_{i=1}^n \left\{ \eta_i^{[\nu_0-2]/2} \exp\left(-\frac{\nu_0}{2} \eta_i\right) \right\} \quad (5.3.55)
 \end{aligned}$$

또한, 다음 식이 성립한다.

$$[\mathbf{y} - X\boldsymbol{\beta}]^t H [\mathbf{y} - X\boldsymbol{\beta}] = \sum_{i=1}^n \eta_i [y_i - \mathbf{x}_i^t \boldsymbol{\beta}]^2 \quad (5.3.56)$$

식 (5.3.55)와 식 (5.3.56)에서 알 수 있듯이, 사후조건부확률분포는 다음과 같다.

$$\boldsymbol{\beta} \mid (\psi, H, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_*, B_*) \quad (5.3.57)$$

$$\psi \mid (\boldsymbol{\beta}, H, \mathbf{y}) \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_*}{2}, \frac{n_* S_*}{2}\right) \quad (5.3.58)$$

$$\eta_i \mid (\psi, \boldsymbol{\beta}, \mathbf{y}) \stackrel{d}{\sim} \text{Gamma}\left(\frac{\nu_0 + 1}{2}, \frac{1}{2} [\nu_0 + \psi [y_i - \mathbf{x}_i^t \boldsymbol{\beta}]^2]\right) \quad (5.3.59)$$

여기서 상수들과 행렬들은 다음과 같다.

$$n_* \doteq n_0 + n \quad (5.3.60)$$

$$S_* \doteq \frac{1}{n_*} \{n_0 S_0 + [\mathbf{y} - X\boldsymbol{\beta}]^t H [\mathbf{y} - X\boldsymbol{\beta}]\} \quad (5.3.61)$$

$$B_* \doteq [B_0^{-1} + \psi X^t H X]^{-1} \quad (5.3.62)$$

$$\mathbf{b}_* \doteq B_* [B_0^{-1} \mathbf{b}_0 + \psi X^t H \mathbf{y}] \quad (5.3.63)$$

지금까지 내용을 바탕으로, 오차항이  $t$  확률분포인 선형회귀모형의 모수들  $(\boldsymbol{\beta}, \psi, H)$ 의 사후확률분포에서 샘플링하는 Gibbs샘플러를 다음과 같이 요약할 수 있다.

**(1단계)** 초기모수들  $(\boldsymbol{\beta}^{(0)}, \psi^{(0)}, H^{(0)})$ 를 선택한다.

**(2단계)** 각  $j(= 1, 2, \dots)$ 에 대해서 주어진  $\mathbf{y}$ 에 대한 조건부샘플  $(\boldsymbol{\beta}^{(j)}, \psi^{(j)}, H^{(j)})$ 를 발생시킨다. 우선, 식 (5.3.57)의 사후조건부확률분포를 사용해서  $\boldsymbol{\beta}^{(j)} \mid (\psi^{(j-1)}, H^{(j-1)}, \mathbf{y})$ 를 발생시킨다. 식 (5.3.58)의 사후조건부확률분포를 사용해서  $\psi^{(j)} \mid (\boldsymbol{\beta}^{(j)}, H^{(j-1)}, \mathbf{y})$ 를 발생시킨다. 식 (5.3.59)의 사후조건부확률분포를 사용해서, 각  $i$ 에 대해  $\eta_i^{(j)} \mid (\boldsymbol{\beta}^{(j)}, \psi^{(j)}, \mathbf{y})$ 를 발생시킨다.

**(3단계)** 충분히 큰  $b$ 에 대해서  $(\boldsymbol{\beta}^{(j)}, \psi^{(j)}, H^{(j)})$ ,  $(j = b + 1, b + 2, \dots)$ 를 기록한다.

## 5.3.4 구조변화가 있는 회귀모형과 Gibbs 샘플러

베이즈안추정법을 사용해서 다음과 같은 구조변화가 있는 회귀모형을 추정하는 문제를 살펴 보자.

$$y_t = \begin{cases} \alpha_1 x_{t1} + \alpha_2 x_{t2} + \cdots + \alpha_K x_{tK} + \epsilon_t, & (t = 1, 2, \dots, k) \\ \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + \eta_t, & (t = k+1, k+2, \dots, n) \end{cases} \quad (5.3.64)$$

여기서  $y_t$ 는 시점  $t$ 에서 종속변수  $\mathbf{x}_t = [x_{t1}, x_{t2}, \dots, x_{tK}]^t$ 는 설명변수벡터이고,  $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_K]^t$ 와  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_K]^t$ 는 모수벡터들이며,  $\{\epsilon_t\}$ 와  $\{\eta_t\}$ 는 서로 독립인 오차항들로서 다음 식들을 만족한다.

$$\epsilon_t | \sigma_1^2 \stackrel{d}{\sim} \mathcal{N}(0, \sigma_1^2), \quad (t = 1, 2, \dots, k) \quad (5.3.65)$$

$$\eta_t | \sigma_2^2 \stackrel{d}{\sim} \mathcal{N}(0, \sigma_2^2), \quad (t = k+1, k+2, \dots, n) \quad (5.3.66)$$

다음과 같이 정밀도들을 정의하자.

$$\psi_1 \doteq \frac{1}{\sigma_1^2}, \quad \psi_2 \doteq \frac{1}{\sigma_2^2} \quad (5.3.67)$$

모수들  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\beta}$ ,  $\psi_1$ ,  $\psi_2$ 와  $k$ 의 사전확률분포들이 다음과 같다고 하자.

$$\boldsymbol{\alpha} \stackrel{d}{\sim} \mathcal{N}(\mathbf{a}_0, A_0), \quad \boldsymbol{\beta} \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_0, B_0) \quad (5.3.68)$$

$$\psi_1 \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right), \quad \psi_2 \stackrel{d}{\sim} \text{Gamma}\left(\frac{m_0}{2}, \frac{m_0 T_0}{2}\right) \quad (5.3.69)$$

$$k \stackrel{d}{\sim} \text{Uniform}\{1, 2, \dots, n\} \quad (5.3.70)$$

관찰벡터  $\mathbf{y} = [y_1, y_2, \dots, y_n]^t$ 의 우도함수는 다음과 같다.

$$f(\mathbf{y} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \psi_1, \psi_2, k) \quad (5.3.71)$$

$$\propto \psi_1^{k/2} \psi_2^{[n-k]/2} \exp\left(-\frac{1}{2} \left\{ \psi_1 \sum_{t=1}^k [y_t - \mathbf{x}_t^t \boldsymbol{\alpha}]^2 + \psi_2 \sum_{t=k+1}^n [y_t - \mathbf{x}_t^t \boldsymbol{\beta}]^2 \right\}\right) \quad (5.3.72)$$

Bayes정리에서 알 수 있듯이, 모수들  $\alpha, \beta, \psi_1$ 와  $\psi_2$ 의 사후조건부확률분포들은 다음과 같다.

$$\alpha \mid (\psi_1, \psi_2, k, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}(\mathbf{a}_1, A_1) \tag{5.3.73}$$

$$\beta \mid (\psi_1, \psi_2, k, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_1, B_1) \tag{5.3.74}$$

$$\psi_1 \mid (\alpha, k, \mathbf{y}) \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_1}{2}, \frac{n_1 S_1}{2}\right) \tag{5.3.75}$$

$$\psi_2 \mid (\beta, k, \mathbf{y}) \stackrel{d}{\sim} \text{Gamma}\left(\frac{m_1}{2}, \frac{m_1 T_1}{2}\right) \tag{5.3.76}$$

여기서 행렬들, 벡터들 그리고 상수들은 다음과 같다.

$$A_1 \doteq \left[ A_0^{-1} + \psi_1 \sum_{t=1}^k \mathbf{x}_t \mathbf{x}_t^t \right]^{-1}, \quad B_1 \doteq \left[ B_0^{-1} + \psi_2 \sum_{t=k+1}^n \mathbf{x}_t \mathbf{x}_t^t \right]^{-1} \tag{5.3.77}$$

$$\mathbf{a}_1 \doteq A_1 \left[ A_0^{-1} \mathbf{a}_0 + \psi_1 \sum_{t=1}^k y_t \mathbf{x}_t \right], \quad \mathbf{b}_1 \doteq B_1 \left[ B_0^{-1} \mathbf{b}_0 + \psi_2 \sum_{t=k+1}^n y_t \mathbf{x}_t \right] \tag{5.3.78}$$

$$n_1 \doteq n_0 + k, \quad S_1 \doteq \frac{1}{n_1} \left\{ n_0 S_0 + \sum_{t=1}^k [y_t - \mathbf{x}_t^t \alpha]^2 \right\} \tag{5.3.79}$$

$$m_1 \doteq m_0 + n - k, \quad T_1 \doteq \frac{1}{m_1} \left\{ m_0 T_0 + \sum_{t=k+1}^n [y_t - \mathbf{x}_t^t \beta]^2 \right\} \tag{5.3.80}$$

다음 변수를 정의하자.

$$w_j \doteq \psi_1^{j/2} \psi_2^{[n-j]/2} \exp \left( -\frac{1}{2} \left\{ \psi_1 \sum_{t=1}^j [y_t - \mathbf{x}_t^t \alpha]^2 + \psi_2 \sum_{t=j+1}^n [y_t - \mathbf{x}_t^t \beta]^2 \right\} \right) \tag{5.3.81}$$

모수  $k$ 의 사후조건부확률분포는 다음과 같다.

$$\Pr(k = j) = \frac{w_j}{\sum_{t=1}^n w_t}, \quad (j = 1, 2, \dots, n) \tag{5.3.82}$$

지금까지 내용을 바탕으로, 구조변화가 있는 회귀모형의 모수들  $(\alpha, \beta, \psi_1, \psi_2)$ 의 사후확률분포에서 샘플링하는 Gibbs샘플러를 다음과 같이 요약할 수 있다.

(1단계) 초기모수들  $(\alpha^{(0)}, \beta^{(0)}, \psi_1^{(0)}, \psi_2^{(0)})$ 를 선택한다.

(2단계) 각  $j (= 1, 2, \dots)$ 에 대해서 주어진  $\mathbf{y}$ 에 대한 조건부샘플  $(\alpha^{(j)}, \beta^{(j)}, \psi_1^{(j)}, \psi_2^{(j)})$ 를 발생시킨다. 우선, 식 (5.3.73)의 사후조건부확률분포를 사용해서  $\alpha^{(j)} \mid (\psi_1^{(j-1)}, \psi_2^{(j-1)}, k^{(j-1)}, \mathbf{y})$ 를 발생시킨다. 식 (5.3.74)의 사후조건부확률분포를 사용해서  $\beta^{(j)} \mid (\psi_1^{(j-1)}, \psi_2^{(j-1)}, k^{(j-1)}, \mathbf{y})$

를 발생시킨다. 식 (5.3.75)의 사후조건부확률분포를 사용해서,  $\psi_1^{(j)} | (\boldsymbol{\alpha}^{(j)}, k^{(j-1)}, \mathbf{y})$ 를 발생시킨다. 식 (5.3.76)의 사후조건부확률분포를 사용해서,  $\psi_2^{(j)} | (\boldsymbol{\beta}^{(j)}, k^{(j-1)}, \mathbf{y})$ 를 발생시킨다. 식 (5.3.82)의 사후조건부확률분포를 사용해서, 확률분포를 사용해서,  $k^{(j)} | (\boldsymbol{\alpha}^{(j)}, \boldsymbol{\beta}^{(j)}, \psi_1^{(j)}, \psi_2^{(j)}, \mathbf{y})$ 를 발생시킨다.

(3단계) 충분히 큰  $b$ 에 대해서  $(\boldsymbol{\alpha}^{(j)}, \boldsymbol{\beta}^{(j)}, \psi_1^{(j)}, \psi_2^{(j)})$ ,  $(j = b + 1, b + 2, \dots)$ 를 기록한다.

### 5.3.5 프로빗모형, 로지스틱회귀모형, 토빗모형과 Gibbs 샘플러

회귀모형에서 종속변수  $y_i$ 가 0 또는 1이라는 이산값을 취하는 경우가 있다. 예를 들어, 제  $i$ 번째 고객에게 대출해주면  $y_i = 1$ 이라 하고, 대출을 해주지 않으면  $y_i = 0$ 으로 표기할 수 있다. 다음과 같은 모형을 사용해서, 종속변수  $y_i$ 를 설명변수벡터  $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ip}]^t$ 로 나타내기로 하자.

$$\Pr(y_i = 1) = F(\mathbf{x}_i^t \boldsymbol{\beta}) = 1 - \Pr(y_i = 0) \quad (5.3.83)$$

여기서  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_p]^t$ 은 모수벡터이다. 또한,  $F(\cdot)$ 는 누적확률분포함수이다. 함수  $F(\cdot)$ 로 다음과 같은 표준정규확률분포를 사용하면, 이에 해당하는 모형 (5.3.83)을 프로빗모형 (probit model)이라고 한다.

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz \quad (5.3.84)$$

함수  $F(\cdot)$ 로 다음과 같은 로지스틱함수 (logistic function)을 사용하면, 이에 해당하는 모형 (5.3.83)을 로지스틱회귀모형 (logistic model)이라고 한다.

$$F(x) = \frac{1}{1 + e^{-x}} \quad (5.3.85)$$

베이시안추정법을 사용해서 프로빗모형을 추정하는 문제를 살펴보자. 시점  $i$ 에서 종속변수를  $y_i$  그리고 설명변수벡터를  $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ip}]^t$ 라 하고, 프로빗모형의 모수벡터를  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_p]^t$ 라고 하면, 관찰벡터  $\mathbf{y} = [y_1, y_2, \dots, y_n]^t$ 의 우도함수는 다음과 같다.

$$f(\mathbf{y} | \boldsymbol{\beta}) = \prod_{i=1}^n F(\mathbf{x}_i^t \boldsymbol{\beta})^{y_i} [1 - F(\mathbf{x}_i^t \boldsymbol{\beta})]^{1-y_i} \quad (5.3.86)$$

식 (5.3.86)의 우도함수는 직접 다루기에 상당히 복잡한 형태이다. 그러나, 다음과 같이 잠재

변수  $y_i^*$ 를 도입함으로써, Gibbs샘플러를 이용해서 프로빗모형을 추정할 수 있다.

$$y_i^* = \mathbf{x}_i^t \boldsymbol{\beta} + \epsilon_i, \quad (i = 1, 2, \dots, n) \quad (5.3.87)$$

$$y_i = 1 (y_i^* > 0), \quad (i = 1, 2, \dots, n) \quad (5.3.88)$$

여기서  $\{\epsilon_t\}$ 는 서로 독립인 오차항들로서 동일한 표준정규확률분포를 따른다. 즉,  $\mathbf{y}^* = [y_1^*, y_2^*, \dots, y_n^*]^t$ 는 오차항의 분산공분산행렬이  $I_n$ 인 선형회귀모형이다. 모수벡터  $\boldsymbol{\beta}$ 의 사전 확률분포가 다음과 같다고 하자.

$$\boldsymbol{\beta} \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_0, B_0) \quad (5.3.89)$$

따라서,  $\mathbf{y}^*$ 가 주어진 경우 사후조건부확률분포는 다음과 같다.

$$\boldsymbol{\beta} | \mathbf{y}^* \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_1, B_1) \quad (5.3.90)$$

여기서 행렬과 벡터는 각각 다음과 같다.

$$B_1 \doteq \left[ B_0^{-1} + \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^t \right]^{-1}, \quad \mathbf{b}_1 \doteq B_1 \left[ B_0^{-1} \mathbf{b}_0 + \sum_{i=1}^n y_i^* \mathbf{x}_i \right] \quad (5.3.91)$$

잠재변수  $y_i^*$ 의 조건부확률분포는 다음과 같다.

$$y_i^* | (\boldsymbol{\beta}, y_i = 1) \stackrel{d}{\sim} \mathcal{TN}_{(0, \infty)}(\mathbf{x}_i^t \boldsymbol{\beta}, 1) \quad (5.3.92)$$

$$y_i^* | (\boldsymbol{\beta}, y_i = 0) \stackrel{d}{\sim} \mathcal{TN}_{(-\infty, 0]}(\mathbf{x}_i^t \boldsymbol{\beta}, 1) \quad (5.3.93)$$

여기서  $\mathcal{TN}_A(\mu, \sigma^2)$ 는 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 를 지지대  $A$ 로 한정된 절단정규확률분포 (truncated normal probability distribution)이다.

지금까지 내용을 바탕으로, 프로빗모형의 회귀계수벡터  $\boldsymbol{\beta}$ 의 사후확률분포에서 샘플링하는 Gibbs샘플러를 다음과 같이 요약할 수 있다.

**(1단계)** 초기모수  $\boldsymbol{\beta}^{(0)}$ 를 선택한다.

**(2단계)** 각  $j(= 1, 2, \dots)$ 에 대해서 샘플  $(y_i^{*(j)}, \boldsymbol{\beta}^{(j)})$ 를 발생시킨다. 우선, 식  $y_j = 1$ 을 만족하는  $j$ 에 대해 식 (5.3.92)의 사후조건부확률분포를 사용해서  $y_i^{*(j)} | \boldsymbol{\beta}^{(j-1)}$ 을, 그리고 식  $y_j = 0$ 을 만족하는  $j$ 에 대해 식 (5.3.93)의 사후조건부확률분포를 사용해서  $y_i^{*(j)} | \boldsymbol{\beta}^{(j-1)}$

을 발생시킨다. 다음으로 식 (5.3.90)의 사후조건부확률분포를 사용해서  $\beta^{(j)} | y^{*(j)}$ 를 생성한다.

(3단계) 충분히 큰  $b$ 에 대해서  $\beta^{(j)} | y^{*(j)}$ , ( $j = b + 1, b + 2, \dots$ )를 기록한다.

**예제 5.3.5** Gibbs샘플러를 사용해서 프로빗모형을 추정하기 위해서, 다음 R프로그램 GibbsSamplerProbit101R.R을 실행해 보자.

```

1 # -----
2 # Filename: GibbsSamplerProbit101R.R
3 # Gibbs Sampler for Probit Model
4 # Programmed by CBS
5 # -----
6 library(MASS)
7 # Obtaining Data
8 ## http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm
9 ## Program Effectiveness, 32 Cross Section Observations
10 ## Source: Spector and Mazzeo (1980)
11 mydata <- read.csv("SpectorMazzeo.csv",header=T)
12 nX <- nrow(mydata)
13 X <- cbind(matrix(1,nrow=nX,ncol=1),
14            mydata$GPA,mydata$TUCE,mydata$PSI)
15 y <- mydata$GRADE
16 XtX <- t(X)%*%X
17 Xty <- t(X)%*%y
18 yty <- t(y)%*%y
19 kX <- ncol(X)
20
21 # Prior
22 b0 <- c(-2,1,0,0)
23 B0 <- diag(2,4)
24
25 # Posterior
26 set.seed(1)
27 ## Truncated Normal Random Number
28 trunNormal <- function(mu,sigma,L,U){
29     uRN <- runif(1)
30     trunNormal <- qnorm(uRN*pnorm((U-mu)/sigma) +
31                        (1-uRN)*pnorm((L-mu)/sigma))*sigma+mu
32 }
33 ## Initials
34 Nsim <- 200
35 Nburnin <- 1000
36 Nobs <- Nburnin + Nsim
37 Bpost <- ginv(ginv(B0)+ XtX)
38 bpost <- Bpost%*%(ginv(B0)%*%b0+Xty)
39 ystar <- rep(0,nX)
40 betaa <- b0
41 betaAtotal <- matrix(0,Nobs,kX)
42 ## Gibbs Sampler
43 for (jj in 1:Nobs){
44     for (ii in 1:nX){
45         if (y[ii]==1){
46             ystar[ii] <- trunNormal(t(X[ii,])%*%betaa,1,0,Inf)
47         }
48         else{

```

```

49     ystar[ii] <- trunNormal(t(X[ii,])%*%betaa,1,-Inf,0)
50   }
51 }
52 bpost <- Bpost%*(ginv(B0)%*%b0+t(X)%*%ystar)
53 betaa <- mvrnorm(n=1,bpost,Bpost)
54 betaAtotal[jj,] <- betaa
55 }
56 betaA <- betaAtotal[(Nburnin+1):(Nobs),]
57
58 ## ACF
59 beta0acf <- acf(betaA[,1], plot = FALSE)
60 beta0acfdf <- with(beta0acf, data.frame(lag,acf))
61 beta1acf <- acf(betaA[,2], plot = FALSE)
62 beta1acfdf <- with(beta1acf, data.frame(lag,acf))
63 beta2acf <- acf(betaA[,3], plot = FALSE)
64 beta2acfdf <- with(beta2acf, data.frame(lag,acf))
65 beta3acf <- acf(betaA[,4], plot = FALSE)
66 beta3acfdf <- with(beta3acf, data.frame(lag,acf))
67
68 # Plotting
69 library(ggplot2)
70 library(grid)
71 setEPS()
72 ## Histogram
73 plot.new()
74 postscript('GibbsSamplerProbit101aR.eps') # Start to save figure
75 timee = c(1:Nsim)
76 GibbsData <- data.frame(timee,betaA)
77 plot11 <- ggplot(GibbsData, aes(x=betaA[,1])) +
78   xlab(expression(beta_0)) +
79   geom_histogram(aes(y =..density..), bins=30,
80     # breaks=seq(0.0,2.5, by=0.05),
81     col="green",fill="blue") +
82   geom_density(lwd=1.0)
83 plot12 <- ggplot(GibbsData, aes(x=betaA[,2])) +
84   xlab(expression(beta_1)) +
85   geom_histogram(aes(y =..density..), bins=30,
86     # breaks=seq(0.0,2.5, by=0.05),
87     col="green",fill="red") +
88   geom_density(lwd=1.0)
89 plot13 <- ggplot(GibbsData, aes(x=betaA[,3])) +
90   xlab(expression(beta_2)) +
91   geom_histogram(aes(y =..density..), bins=30,
92     # breaks=seq(0.0,2.5, by=0.05),
93     col="green",fill="dark blue") +
94   geom_density(lwd=1.0)
95 plot14 <- ggplot(GibbsData, aes(x=betaA[,4])) +
96   xlab(expression(beta_3)) +
97   geom_histogram(aes(y =..density..), bins=30,
98     # breaks=seq(0.0,2.5, by=0.05),
99     col="green",fill="dark red") +
100   geom_density(lwd=1.0)
101 pushViewport(viewport(layout = grid.layout(2,2)))
102 print(plot11, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
103 print(plot12, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
104 print(plot13, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
105 print(plot14, vp = viewport(layout.pos.row = 2, layout.pos.col = 2))
106 dev.off() # End to save figure
107
108 ## Autocorrelation Function
109 plot.new()

```

```

110 postscript('GibbsSamplerProbit101bR.eps') # Start to save figure
111 plot21 <- ggplot(data = beta0acfd, mapping = aes(x = lag, y = acf)) +
112     geom_hline(aes(yintercept = 0)) +
113     geom_segment(mapping = aes(xend = lag, yend = 0),
114                 lwd=2.0,color="blue")
115 plot22 <- ggplot(data = beta1acfd, mapping = aes(x = lag, y = acf)) +
116     geom_hline(aes(yintercept = 0)) +
117     geom_segment(mapping = aes(xend = lag, yend = 0),
118                 lwd=2.0,color="red")
119 plot23 <- ggplot(data = beta2acfd, mapping = aes(x = lag, y = acf)) +
120     geom_hline(aes(yintercept = 0)) +
121     geom_segment(mapping = aes(xend = lag, yend = 0),
122                 lwd=2.0,color="dark blue")
123 plot24 <- ggplot(data = beta3acfd, mapping = aes(x = lag, y = acf)) +
124     geom_hline(aes(yintercept = 0)) +
125     geom_segment(mapping = aes(xend = lag, yend = 0),
126                 lwd=2.0,color="dark red")
127 pushViewport(viewport(layout = grid.layout(2,2)))
128 print(plot21, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
129 print(plot22, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
130 print(plot23, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
131 print(plot24, vp = viewport(layout.pos.row = 2, layout.pos.col = 2))
132 dev.off() # End to save figure
133 # -----

```

Greene [51, pp. 711-713]에서 프로빗모형을 추정하는 예를 설명하기 위해서 Spector & Mazzeo (1980)가 제시한 데이터를 사용하였다. 이 예제에서도 같은 데이터를 사용해서 프로빗모형을 추정해보자. 이 데이터는 경제학을 가르치는 어떤 새로운 방법이 효율적인지를 조사하기 위한 것이다. 이 데이터세트에는 설명변수들 GPA, TUCE, PSI와 종속변수 GRADE가 있다. 이 중에서 설명변수 GPA는 학점평균(grade point average)을 나타내며 0과 4 사이 수를 취하며, 설명변수 TUCE는 학기 시작 전에 경제학에 대한 지식을 미리 테스트(pretest)한 결과를 나타내는 12와 29 사이 수를 취한다. 또한, 설명변수 PSI는 새로운 교육방법으로 배웠는지를 나타내는 2원변수(binary variable)이고, 종속변수 GRADE는 새로운 교육방법이 실시된 이후 학생의 성적이 올랐는가를 나타낸다. 만약 변수 GRADE가 값 1을 가지면 경제학과목 성적이 오른 것이고, 그렇지 않으면 변수 GRADE가 값 0을 갖는다. R함수 head를 사용해서 이 데이터세트의 앞 부분을 출력하면, 다음과 같다.

	OBS	GPA	TUCE	PSI	GRADE
	1	1 2.66	20	0	0
	2	2 2.89	22	0	0
	3	3 3.28	24	0	0
	4	4 2.92	12	0	0
	5	5 4.00	21	0	1
	6	6 2.86	17	0	0

이 R 프로그램에서는 이 데이터세트에 다음 프로빗모형을 적용한다.

$$y_j^* = \beta_0 + \beta_1 GPA_j + \beta_2 TUCE_j + \beta_3 PSE_j + \epsilon_j, \quad (j = 1, 2, \dots, 32) \quad (1)$$

$$GRADE_j = 1(y_j^* > 0), \quad (j = 1, 2, \dots, 32) \quad (2)$$

따라서,  $x_1 = GPA$ ,  $x_2 = TUCE$  그리고  $x_3 = PSE$ 이다.

식 (5.3.92)과 식 (5.3.93)에서 알 수 있듯이, 프로빗모형에서 Gibbs샘플을 생성하기 위해서는 절단정규확률변수에서 난수들을 생성할 필요가 있다. 확률변수  $y$ 는 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 를 지지대가  $[L, U]$ 가 되도록 절단한 절단정규확률분포를 따른다고 하자. 확률변수  $y$ 의 확률분포함수  $F(\cdot)$ 는 다음과 같다.

$$F(y) = \frac{N\left(\frac{y-\mu}{\sigma}\right) - N\left(\frac{L-\mu}{\sigma}\right)}{N\left(\frac{U-\mu}{\sigma}\right) - N\left(\frac{L-\mu}{\sigma}\right)} \quad (3)$$

따라서, 다음 식이 성립한다.

$$F(y) \stackrel{d}{\sim} Uniform(0, 1) \quad (4)$$

식 (4)에서 알 수 있듯이, 다음 식이 성립한다.

$$y \stackrel{d}{=} \mu + \sigma N^{-1}\left(w \cdot N\left(\frac{U-\mu}{\sigma}\right) + [1-w] \cdot N\left(\frac{L-\mu}{\sigma}\right)\right) \quad (5)$$

여기서  $x \stackrel{d}{=} z$ 는 확률변수들  $x$ 와  $z$ 가 동일한 확률분포를 따름을 의미하고,  $w \stackrel{d}{\sim} U(0, 1)$ 이다. 식 (5)를 이용해서 절단정규확률분포로부터 난수들을 생성할 수 있다.

이 R 프로그램 GibbsSamplerProbit101R.R에서는 베이저안통계기법을 사용해서 이 프로빗모형을 추정하기 위해서, 우선 사전확률분포의 모수들을 다음과 같이 선택한다.

$$\mathbf{b}_0 = \begin{bmatrix} -2 & 1 & 0 & 0 \end{bmatrix}^t, \quad B_0 = 2I_4 \quad (6)$$

프로빗모형의 회귀계수벡터  $\beta$ 의 사후확률분포에서 샘플링하는 Gibbs샘플러를 적용해서 Gibbs샘플들을  $Nsim = 200$ 개 발생시킨다. 이때 번인기간은  $Nburnin = 1000$ 이다. 이렇게

발생된 Gibbs샘플들의 표본평균은 다음과 같다.

$$\bar{\beta}_0 = -3.552067628, \quad \bar{\beta}_1 = 0.922784738 \quad (7)$$

$$\bar{\beta}_2 = -0.007735569, \quad \bar{\beta}_3 = 1.017461691 \quad (8)$$

또한, 분산공분산행렬은 다음과 같다.

$$Var(\beta) = \begin{bmatrix} 1.306921574 & -0.340458719 & -0.006459989 & -0.078255284 \\ -0.340458719 & 0.293872657 & -0.026454485 & -0.001152099 \\ -0.006459989 & -0.026454485 & 0.004121582 & -0.001751117 \\ -0.078255284 & -0.001152099 & -0.001751117 & 0.240304166 \end{bmatrix} \quad (9)$$

이 R프로그램 GibbsSamplerProbit101R.R 을 실행하면, 사후확률분포로부터 Gibbs샘플러를 사용해서 발생시킨 그래프들을 그린 그림 5.3.9을 출력한다. 그림 5.3.9의 좌측상단 그래프에는  $\beta_0$ 의 히스토그램이, 우측상단 그래프에는  $\beta_1$ 의 히스토그램이, 좌측하단 그래프에는  $\beta_2$ 의 히스토그램이, 그리고 우측하단 그래프에는  $\beta_3$ 의 히스토그램이 그려져 있다. 또한, 이 Gibbs샘플들의 추정된 자기상관함수들 (autocorrelation functions)이 그림 5.3.10에 그려져 있다. 그림 5.3.10의 좌측상단 그래프에는  $\beta_0$ 의 추정된 자기상관함수가, 우측상단 그래프에는  $\beta_1$ 의 추정된 자기상관함수가, 좌측하단 그래프에는  $\beta_2$ 의 추정된 자기상관함수가, 그리고 우측하단 그래프에는  $\beta_3$ 의 추정된 자기상관함수가 그려져 있다. ■

**예제 5.3.6** Gibbs샘플러를 사용해서 프로빗모형을 추정하는 다른 예로서, 다음 R프로그램 GibbsSamplerProbit102R.R을 실행해 보자.

```

1 # -----
2 # Filename: GibbsSamplerProbit102R.R
3 # Gibbs Sampler for Probit Model
4 # Programmed by CBS
5 # -----
6 library(MASS)
7 library(mvtnorm)
8 library(msm) # For rtnorm function
9
10 # Simulate data
11 set.seed(1)
12 n <- 256
13 x <- sample(0:4,n,replace=T) # Sampling w/ replacement
14 X <- matrix(c(rep(1,n),x), ncol=2)
15 betaTrue <- c(3.0,-1.0)
16 ystar <- X%*%betaTrue

```

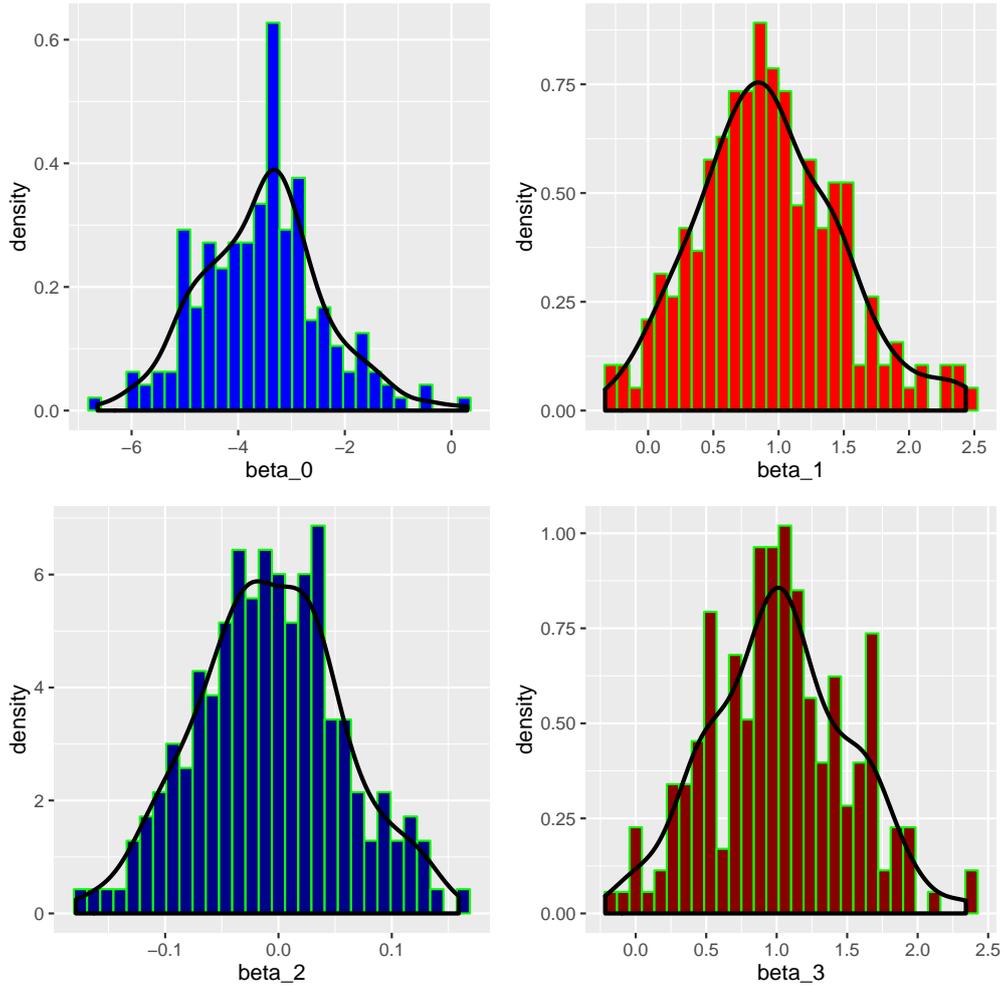


그림 5.3.9. 회귀계수벡터  $\beta$ 의 사후확률분포

```

17 y <- rbinom(n,1,pnorm(ystar))
18 p <- 2 # Dimension of beta
19 nOne <- sum(y)
20 nZero <- n-nOne
21 Xy <- cbind(X,y)
22 head(Xy)
23
24 # Priors
25 beta0 <- c(0,0)
26 B0 <- diag(pi,2)
27 Precision0 <- ginv(B0)
28
29 # Gibbs Sample
30 ## Initial Values
31 ( fitOutput <-glm(y~x,family=binomial(link=probit)) )
32 betaa <- rep(0,p)
33 nSim <- 1000
34 nBurnin <- 1000
35 z <- rep(0,n) # z = X^t*beta
36 betaMAT <- matrix(0,nrow=nBurnin+nSim,ncol=p)
37 ## Posterior
38 BPost <- ginv(Precision0 + t(X)%*%X)
39 for (ii in 1:(nBurnin+nSim)) {

```

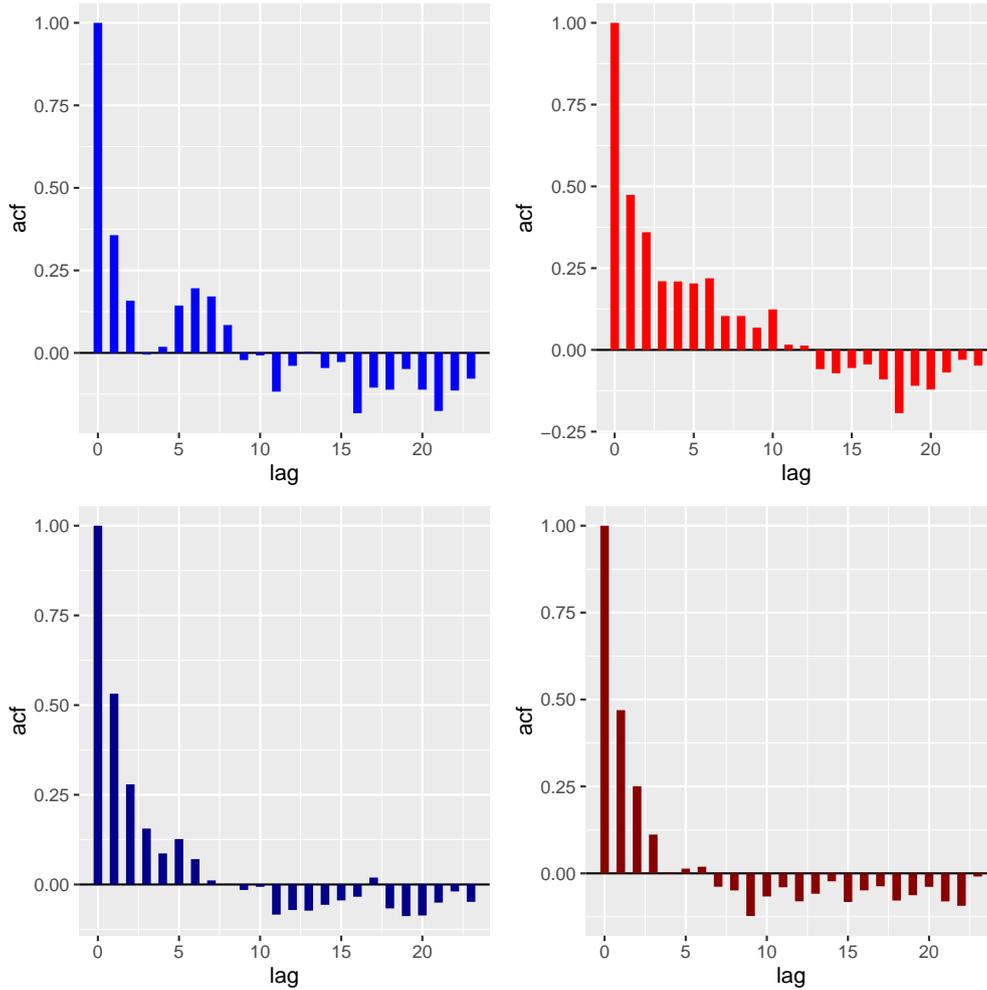


그림 5.3.10. 자기상관함수

```

40   muz <- X%%betaa
41   z[y==0] <- rtnorm(nZero,muz[y==0],1,-Inf,0)
42   z[y==1] <- rtnorm(nOne,muz[y==1],1,0,Inf)
43   # trom{msm}: Truncated norma;; dtnorm, ptnorm, qtnorm, rtnorm
44   bPost <- BPost%%(Precision%%beta0+t(X)%%z)
45   betaa <- c(rmvnorm(1,bPost,BPost))
46   betaMAT[ii,] <- betaa
47 }
48 betaM <- betaMAT[(nBurnin+1):(nBurnin+nSim),]
49 tail(betaM)
50 ( betaMean <- apply(betaM,2,mean) )
51 ( betaVar <- var(betaM) )
52
53 # Plotting
54 library(ggplot2)
55 library(grid)
56 setEPS()
57 plot.new()
58 postscript('GibbsSamplerProbit102R.eps') # Start to save figure
59 timee = c(1:nSim)
60 GibbsData <- data.frame(timee,betaM)
61 plot11 <- ggplot(GibbsData, aes(x=betaM[,1])) +
62   xlab(expression(beta_0)) +

```

```

63     geom_histogram(aes(y =..density..), bins=30,
64                   # breaks=seq(0.0,2.5, by=0.05),
65                   col="green",fill="blue") +
66     geom_density(lwd=1.0)
67 plot12 <- ggplot(GibbsData, aes(x=betaM[,2])) +
68     xlab(expression(beta_1)) +
69     geom_histogram(aes(y =..density..), bins=30,
70                   # breaks=seq(0.0,2.5, by=0.05),
71                   col="green",fill="red") +
72     geom_density(lwd=1.0)
73 plot13 <- ggplot(GibbsData, aes(x=betaM[,1])) +
74     xlab(expression(beta_0)) +
75     geom_line(stat="density",lwd=1.2,color="blue") +
76     geom_vline(xintercept=betaMean[1],lwd=1.1) +
77     geom_hline(yintercept=0,lwd=1.0)
78 plot14 <- ggplot(GibbsData, aes(x=betaM[,2])) +
79     xlab(expression(beta_1)) +
80     geom_line(stat="density",lwd=1.2,color="red") +
81     geom_vline(xintercept=betaMean[2],lwd=1.1)+
82     geom_hline(yintercept=0,lwd=1.0)
83 pushViewport(viewport(layout = grid.layout(2,2)))
84 print(plot11, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
85 print(plot12, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
86 print(plot13, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
87 print(plot14, vp = viewport(layout.pos.row = 2, layout.pos.col = 2))
88 dev.off()           # End to save figure
89 # -----

```

이 R프로그램 GibbsSamplerProbit102R.R에서는 지지대 {0, 1, 2, 3, 4}에서 무작위로 256개 관찰값들을 추출해서 설명변수  $x$ 에 저장한다. 또한, 다음 식들을 사용해서  $\{y_j\}$ 를 생성한다.

$$y_j^* = \beta_0 + \beta_1 x_j + \epsilon_j, \quad (j = 1, 2, \dots, 256) \quad (1)$$

$$y_j = 1 (y_j^* > 0), \quad (j = 1, 2, \dots, 256) \quad (2)$$

여기서  $\{\epsilon_j\}$ 는  $\mathcal{N}(0, 1)$ 를 따르고 서로 독립인 확률변수들이다. 이렇게 생성된 데이터세트의 앞 부분은 다음과 같다.

```

      y
[1,] 1 1 1
[2,] 1 1 1
[3,] 1 2 1
[4,] 1 4 1
[5,] 1 1 1
[6,] 1 4 0

```

이 R프로그램 GibbsSamplerProbit102R.R에서는 이 데이터세트에 다음 프로빗모형을

적용한다.

$$\Pr(y_j = 1) = \Pr(y_j^* > 0) = N(\beta_0 + \beta_1 x_j), \quad (j = 1, 2, \dots, 256) \quad (3)$$

이 R 프로그램 GibbsSamplerProbit102R.R에서는 베이지안통계기법을 사용해서 이 프로빗모형을 추정하기 위해서, 우선 사전확률분포의 모수들을 다음과 같이 선택한다.

$$\mathbf{b}_0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^t, \quad B_0 = \pi I_2 \quad (4)$$

프로빗모형에서 Gibbs 샘플을 생성하기 위해서는 절단정규확률분포에서 난수들을 생성할 필요가 있다. 이 R 프로그램에서는 라이브러리 msm에 포함된 함수 rtnorm을 사용해서 절단정규확률분포로부터 난수들을 생성했다. 프로빗모형의 회귀계수벡터  $\beta$ 의 사후확률분포에서 샘플링하는 Gibbs 샘플러를 적용해서 Gibbs 샘플들을  $nSim = 1000$  개 발생시켰다. 이때 번인기간은  $nBurnin = 1000$  이다. 이렇게 발생된 Gibbs 샘플들의 마지막 부분은 다음과 같다.

	[,1]	[,2]
[995,]	2.650992	-0.9064977
[996,]	2.728602	-0.9601093
[997,]	2.804436	-0.9648010
[998,]	2.455025	-0.8419398
[999,]	2.188603	-0.7357317
[1000,]	1.947153	-0.6608517

또한, 이 Gibbs 샘플들의 표본평균들은 다음과 같다.

$$\bar{\beta}_0 = 2.8410321, \quad \bar{\beta}_1 = -0.9624568 \quad (5)$$

또한, 분산공분산행렬은 다음과 같다.

$$\text{Var}(\beta) = \begin{bmatrix} 0.09307834 & -0.03052765 \\ -0.03052765 & 0.01136371 \end{bmatrix} \quad (6)$$

이 R 프로그램 GibbsSamplerProbit102R.R을 실행하면, 사후확률분포로부터 Gibbs 샘플러를 사용해서 발생시킨 그래프들을 그린 그림 5.3.11을 출력한다. 그림 5.3.11의 좌측상단

그래프에는  $\beta_0$  의 히스토그램이, 우측상단 그래프에는  $\beta_1$  의 히스토그램이, 좌측하단 그래프에는  $\beta_0$  의 추정된 확률밀도함수와 표본평균이, 그리고 우측하단 그래프에는  $\beta_1$  의 추정된 확률밀도함수와 표본평균이 그려져 있다. ■

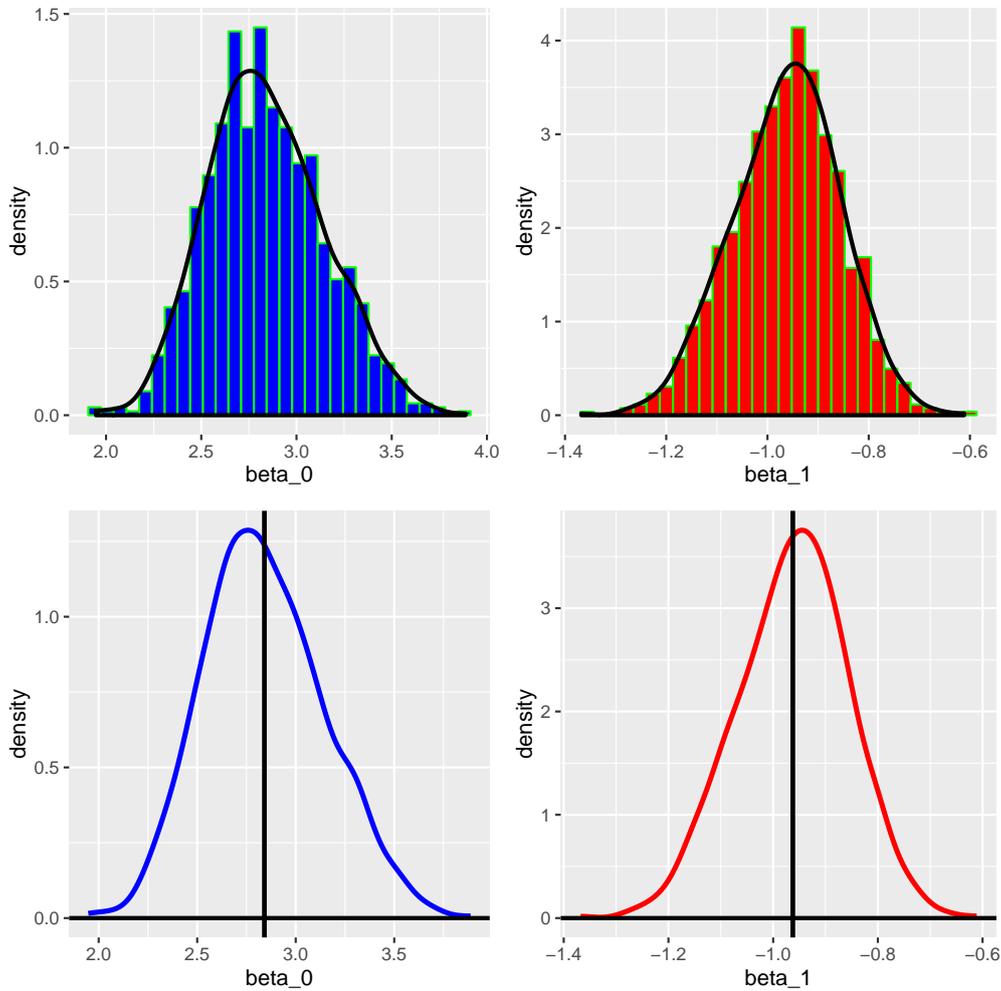


그림 5.3.11. 사후확률분포

**예제 5.3.7** 로지스틱회귀모형은 일반선형모형 (generalized linear model: GLM) 의 특수한 경우이다. 따라서 고전적 통계학에서는 최우추정법을 사용해서 로지스틱회귀모형을 추정하는 것이 일반적이다. 최우추정법을 사용해서 로지스틱회귀모형을 추정하는 예로서 다음 R프로그램 GibbsSamplerLogit101R.R을 실행해 보자.

```

1 # -----
2 # Filename: GibbsSamplerLogit101R.R
3 # Generalized Linear Model Approach for Logit Model
4 # Programmed by CBS
5 # -----
    
```

```

6 library(ggplot2)
7 library(grid)
8 library(MASS)
9 # install.package("aod")
10 library(aod)
11 # install.packages("Rcpp")
12 library(Rcpp)
13
14 # Obtaining Data
15 ## http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm
16 ## Program Effectiveness, 32 Cross Section Observations
17 ## Source: Spector and Mazzeo (1980)
18 mydata <- read.csv("SpectorMazzeo.csv",header=T)
19 sink("GibbsSamplerLogit101R.txt", append=TRUE, split=TRUE)
20 head(mydata)
21
22 # Descriptive Statistics
23 summary(mydata[,2:5])
24 sapply(mydata[,2:5], sd)
25 xtabs(~ GRADE+PSI, data=mydata) # Test if there exists 0 cells.
26
27 # GLM Analysis of Logit Model
28 ( mydata$PSIr <- factor(mydata$PSI) )
29 LogitModel <- glm(GRADE ~ GPA + TUCE + PSIr, data = mydata,
30                  family="binomial")
31 summary(LogitModel)
32 ## Confidence Intervals, Wald-test, and Odds Ratios
33 confint(LogitModel)
34 wald.test(b=coef(LogitModel), Sigma=vcov(LogitModel), Terms=4)
35 exp(cbind(OR = coef(LogitModel), confint(LogitModel)))
36 # -----

```

이 R프로그램을 실행하면, Spector-Mazzeo 데이터세트의 앞 부분을 출력한다. 그 내용은 다음과 같다.

	OBS	GPA	TUCE	PSI	GRADE
1	1	2.66	20	0	0
2	2	2.89	22	0	0
3	3	3.28	24	0	0
4	4	2.92	12	0	0
5	5	4.00	21	0	1
6	6	2.86	17	0	0

R함수 `summary`를 사용해서, 이 데이터세트를 요약한 결과는 다음과 같다.

	GPA	TUCE	PSI	GRADE
Min.	:2.060	Min. :12.00	Min. :0.0000	Min. :0.0000
1st Qu.:	2.812	1st Qu.:19.75	1st Qu.:0.0000	1st Qu.:0.0000
Median	:3.065	Median :22.50	Median :0.0000	Median :0.0000
Mean	:3.117	Mean :21.94	Mean :0.4375	Mean :0.3438

```
3rd Qu.:3.515 3rd Qu.:25.00 3rd Qu.:1.0000 3rd Qu.:1.0000
Max. :4.000 Max. :29.00 Max. :1.0000 Max. :1.0000
```

R함수 `sapply`를 사용해서 구한 각 변수의 표준편차는 다음과 같다.

```
GPA TUCE PSI GRADE
0.4667128 3.9015092 0.5040161 0.4825587
```

각 칸(cell)에 원소가 한 개 이상이어야 GLM을 적용할 수 있다. 이를 확인하기 위해서, 이 2원테이블을 출력해보아야 한다. 설명변수인 PSI와 종속변수인 GRADE의 2원테이블은 다음과 같다.

```
PSI
GRADE 0 1
0 15 6
1 3 8
```

R함수 `glm`을 사용해서 로지스틱회귀모형을 최우추정한 결과는 다음과 같다.

```
Call:
glm(formula = GRADE ~ GPA + TUCE + PSIr, family = "binomial",
    data = mydata)
```

Deviance Residuals:

```
Min 1Q Median 3Q Max
-1.9551 -0.6453 -0.2570 0.5888 2.0966
```

Coefficients:

```
Estimate Std. Error z value Pr(>|z|)
(Intercept) -13.02135 4.93127 -2.641 0.00828 **
GPA 2.82611 1.26293 2.238 0.02524 *
TUCE 0.09516 0.14155 0.672 0.50143
PSIr1 2.37869 1.06456 2.234 0.02545 *
```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 41.183 on 31 degrees of freedom
Residual deviance: 25.779 on 28 degrees of freedom
AIC: 33.779
```

```
Number of Fisher Scoring iterations: 5
```

이 결과물에서 알 수 있듯이, 최우추정법에 의한 추정식은 다음과 같다.

$$\hat{y}_i = 1(\hat{y}_i^* > 0) \tag{1}$$

$$\hat{y}_i^* = -13.02135 + 2.82611GPA + 0.09516TUCE + 2.37869PSI \tag{2}$$

R함수 confint를 사용해서 구한 각 회귀계수의 신뢰구간은 다음과 같다.

```
                2.5 %    97.5 %
(Intercept) -25.1681649 -4.9023529
GPA           0.6393666  5.7572531
TUCE          -0.1702054  0.4050443
PSIr1         0.4786090  4.8102604
```

R함수 wald.test를 사용해서 설명변수 PSI에 관한 귀무가설  $H_0, |\beta_4| = 0$ 를 검정한 결과는 다음과 같다.

```
Chi-squared test:
X2 = 5.0, df = 1, P(> X2) = 0.025
```

이  $p$ 값이 0.05보다 작으므로 유의수준 5퍼센트에서 귀무가설을 기각한다.

각 회귀계수의 오즈비(odds ratio)와 신뢰구간은 다음과 같다.

```
                OR        2.5 %    97.5 %
(Intercept) 2.212590e-06 1.173829e-11 7.429082e-03
GPA          1.687971e+01 1.895280e+00 3.164778e+02
TUCE         1.099832e+00 8.434915e-01 1.499369e+00
PSIr1        1.079073e+01 1.613828e+00 1.227636e+02
```



베이지안추정법을 사용해서 로지스틱회귀모형을 추정하는 과정은 예제 5.3.6에 기술한 프로빗모형을 추정하는 과정과 거의 동일하다. 다만, 정규확률분포함수  $\mu + \sigma N(x)$ 를 로지스틱확률분포  $\mu + \sigma \text{Logi}(x)$ 로 바꾸면 된다. 이  $\text{Logi}(x)$ 는 다음과 같다.

$$\text{Logi}(x) = \frac{1}{1 + e^{-x}} \tag{5.3.94}$$

따라서, 잠재변수의 조건부확률분포는 다음과 같다.

$$y_i^* | (\beta, y_i = 1) \stackrel{d}{\sim} \mathcal{TL}_{(0, \infty)}(\mathbf{x}_i^t \beta, 1) \tag{5.3.95}$$

$$y_i^* | (\beta, y_i = 0) \stackrel{d}{\sim} \mathcal{TL}_{(-\infty, 0]}(\mathbf{x}_i^t \beta, 1) \tag{5.3.96}$$

여기서  $\mathcal{TL}_A(\mu, \sigma^2)$ 는 로지스틱확률분포함수  $\mu + \sigma \text{Logi}(x)$ 를 지지대  $A$ 로 한정된 절단로지스틱확률분포(truncated logistic distribution)이다. 따라서, 로지스틱회귀모형의 회귀계수벡터  $\beta$ 의 사후확률분포에서 샘플링하는 Gibbs샘플러를 다음과 같이 요약할 수 있다.

(1단계) 초기모수  $\beta^{(0)}$ 를 선택한다.

(2단계) 각  $j(= 1, 2, \dots)$ 에 대해서 샘플  $(y_i^{*(j)}, \beta^{(j)})$ 를 발생시킨다. 우선, 식  $y_j = 1$ 을 만족하는  $j$ 에 대해 식 (5.3.95)의 사후조건부확률분포를 사용해서  $y_i^{*(j)} | \beta^{(j-1)}$ 을, 그리고 식  $y_j = 0$ 을 만족하는  $j$ 에 대해 식 (5.3.96)의 사후조건부확률분포를 사용해서  $y_i^{*(j)} | \beta^{(j-1)}$ 을 발생시킨다. 다음으로 이들을 사용해서  $\beta^{(j)} | y^{*(j)}$ 를 발생시킨다.

(3단계) 충분히 큰  $b$ 에 대해서  $\beta^{(j)} | y^{*(j)}$ ,  $(j = b + 1, b + 2, \dots)$ 를 기록한다.

**예제 5.3.8** Gibbs샘플러를 사용해서 로지스틱회귀모형을 추정하기 위해서, 다음 R프로그램 GibbsSamplerLogit102R.R을 실행해 보자.

```

1 # -----
2 # Filename: GibbsSa102plerLogit102R.R
3 # Gibbs Sampler for Logit Model
4 # Programmed by CBS
5 # -----
6 library(MASS)
7 # Obtaining Data
8 ## http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm
9 ## Program Effectiveness, 32 Cross Section Observations
10 ## Source: Spector and Mazzeo (1980)
11 mydata <- read.csv("SpectorMazzeo.csv", header=T)
12 nX <- nrow(mydata)
13 X <- cbind(matrix(1, nrow=nX, ncol=1),

```

```

14 |         mydata$GPA, mydata$TUCE, mydata$PSI)
15 | y <- mydata$GRADE
16 | XtX <- t(X)%*%X
17 | Xty <- t(X)%*%y
18 | yty <- t(y)%*%y
19 | kX <- ncol(X)
20 |
21 | # Prior
22 | b0 <- c(-2,1,0,0)
23 | B0 <- diag(2,4)
24 |
25 | # Posterior
26 | set.seed(1)
27 | ## Truncated Logistic Random Number
28 | trunLogistic <- function(mu, sigma, L, U){
29 |     uRN <- runif(1)
30 |     dumTL <- uRN/(1+exp(-(U-mu)/sigma)) +
31 |             (1-uRN)/(1+exp(-(L-mu)/sigma))
32 |     trunLogistic <- log(dumTL/(1-dumTL))
33 | }
34 | ## Initials
35 | Nsim <- 200
36 | Nburnin <- 1000
37 | Nobs <- Nburnin + Nsim
38 | Bpost <- ginv(ginv(B0)+ XtX)
39 | bpost <- Bpost%*%(ginv(B0)%*%b0+Xty)
40 | ystar <- rep(0, nX)
41 | betaa <- b0
42 | betaAtotal <- matrix(0, Nobs, kX)
43 | ## Gibbs Sampler
44 | for (jj in 1:Nobs){
45 |     for (ii in 1:nX){
46 |         if (y[ii]==1){
47 |             ystar[ii] <- trunLogistic(t(X[ii,])%*%betaa, 1, 0, Inf)
48 |         }
49 |         else{
50 |             ystar[ii] <- trunLogistic(t(X[ii,])%*%betaa, 1, -Inf, 0)
51 |         }
52 |     }
53 |     bpost <- Bpost%*%(ginv(B0)%*%b0+t(X)%*%ystar)
54 |     betaa <- mvrnorm(n=1, bpost, Bpost)
55 |     betaAtotal[jj,] <- betaa
56 | }
57 | betaA <- betaAtotal[(Nburnin+1):(Nobs),]
58 | ( colMeans(betaA) )
59 | ( var(betaA) )
60 |
61 | ## ACF
62 | beta0acf <- acf(betaA[,1], plot = FALSE)
63 | beta0acfdf <- with(beta0acf, data.frame(lag, acf))
64 | beta1acf <- acf(betaA[,2], plot = FALSE)
65 | beta1acfdf <- with(beta1acf, data.frame(lag, acf))
66 | beta2acf <- acf(betaA[,3], plot = FALSE)
67 | beta2acfdf <- with(beta2acf, data.frame(lag, acf))
68 | beta3acf <- acf(betaA[,4], plot = FALSE)
69 | beta3acfdf <- with(beta3acf, data.frame(lag, acf))
70 |
71 | # Plotting
72 | library(ggplot2)
73 | library(grid)
74 | setEPS()

```

```

75 ## Histogram
76 plot.new()
77 postscript('GibbsSamplerLogit102aR.eps') # Start to save figure
78 timee = c(1:Nsim)
79 GibbsData <- data.frame(timee,betaA)
80 plot11 <- ggplot(GibbsData, aes(x=betaA[,1])) +
81   xlab(expression(beta_0)) +
82   geom_histogram(aes(y =..density..), bins=30,
83     # breaks=seq(0.0,2.5, by=0.05),
84     col="green",fill="blue") +
85   geom_density(lwd=1.0)
86 plot12 <- ggplot(GibbsData, aes(x=betaA[,2])) +
87   xlab(expression(beta_1)) +
88   geom_histogram(aes(y =..density..), bins=30,
89     # breaks=seq(0.0,2.5, by=0.05),
90     col="green",fill="red") +
91   geom_density(lwd=1.0)
92 plot13 <- ggplot(GibbsData, aes(x=betaA[,3])) +
93   xlab(expression(beta_2)) +
94   geom_histogram(aes(y =..density..), bins=30,
95     # breaks=seq(0.0,2.5, by=0.05),
96     col="green",fill="dark blue") +
97   geom_density(lwd=1.0)
98 plot14 <- ggplot(GibbsData, aes(x=betaA[,4])) +
99   xlab(expression(beta_3)) +
100   geom_histogram(aes(y =..density..), bins=30,
101     # breaks=seq(0.0,2.5, by=0.05),
102     col="green",fill="dark red") +
103   geom_density(lwd=1.0)
104 pushViewport(viewport(layout = grid.layout(2,2)))
105 print(plot11, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
106 print(plot12, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
107 print(plot13, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
108 print(plot14, vp = viewport(layout.pos.row = 2, layout.pos.col = 2))
109 dev.off() # End to save figure
110
111 ## Autocorrelation Function
112 plot.new()
113 postscript('GibbsSamplerLogit102bR.eps') # Start to save figure
114 plot21 <- ggplot(data = beta0acfd, mapping = aes(x = lag, y = acf)) +
115   geom_hline(aes(yintercept = 0)) +
116   geom_segment(mapping = aes(xend = lag, yend = 0),
117     lwd=2.0,color="blue")
118 plot22 <- ggplot(data = beta1acfd, mapping = aes(x = lag, y = acf)) +
119   geom_hline(aes(yintercept = 0)) +
120   geom_segment(mapping = aes(xend = lag, yend = 0),
121     lwd=2.0,color="red")
122 plot23 <- ggplot(data = beta2acfd, mapping = aes(x = lag, y = acf)) +
123   geom_hline(aes(yintercept = 0)) +
124   geom_segment(mapping = aes(xend = lag, yend = 0),
125     lwd=2.0,color="dark blue")
126 plot24 <- ggplot(data = beta3acfd, mapping = aes(x = lag, y = acf)) +
127   geom_hline(aes(yintercept = 0)) +
128   geom_segment(mapping = aes(xend = lag, yend = 0),
129     lwd=2.0,color="dark red")
130 pushViewport(viewport(layout = grid.layout(2,2)))
131 print(plot21, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
132 print(plot22, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
133 print(plot23, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
134 print(plot24, vp = viewport(layout.pos.row = 2, layout.pos.col = 2))
135 dev.off() # End to save figure

```

예제 5.3.5에서 Spector & Mazzeo (1980)가 제시한 데이터를 사용해서 프로빗모형을 추정하였다. 이 예제에서는 같은 데이터를 사용해서 로지스틱회귀모형을 추정해보자.

이 R프로그램 GibbsSamplerLogit102R.R에서는 이 데이터세트에 다음 로지스틱회귀모형을 적용한다.

$$y_j^* = \beta_0 + \beta_1 GPA_j + \beta_2 TUCE_j + \beta_3 PSE_j + \epsilon_j, \quad (j = 1, 2, \dots, 32) \quad (1)$$

$$GRADE_j = 1(y_j^* > 0), \quad (j = 1, 2, \dots, 32) \quad (2)$$

따라서,  $x_1 = GPA$ ,  $x_2 = TUCE$  그리고  $x_3 = PSE$ 이다.

로지스틱회귀모형에서 Gibbs샘플을 생성하기 위해서는 절단로지스틱확률변수에서 난수들을 생성할 필요가 있다. 확률변수  $y$ 가 로지스틱확률분포함수  $\mu + \sigma \text{Logit}(x)$ 에서 지지대를  $[L, U]$ 인 절단로지스틱확률분포를 따른다고 하자. 확률변수  $y$ 의 확률분포함수  $F(\cdot)$ 는 다음과 같다.

$$F(y) = \frac{[1 + \exp(-\frac{y-\mu}{\sigma})]^{-1} - [1 + \exp(-\frac{L-\mu}{\sigma})]^{-1}}{[1 + \exp(-\frac{U-\mu}{\sigma})]^{-1} - [1 + \exp(-\frac{L-\mu}{\sigma})]^{-1}} \quad (3)$$

따라서 다음 식들이 성립한다.

$$F(y) \stackrel{d}{\sim} \text{Uniform}(0, 1) \quad (4)$$

$$F^{-1}(x) = \ln \frac{x}{1-x} \quad (5)$$

식 (3)에서 알 수 있듯이, 다음 식이 성립한다.

$$y \stackrel{d}{\equiv} \mu + \sigma F^{-1} \left( w \left[ 1 + \exp\left(-\frac{U-\mu}{\sigma}\right) \right]^{-1} + [1-w] \left[ 1 + \exp\left(-\frac{L-\mu}{\sigma}\right) \right]^{-1} \right) \quad (6)$$

여기서  $w \stackrel{d}{\sim} U(0, 1)$ 이다. 식 (6)을 이용해서 절단정규확률분포로부터 난수들을 생성할 수 있다.

이 R프로그램 GibbsSamplerLogit102R.R에서는 이 로지스틱회귀모형을 베이지안통계

기법으로 추정하기 위해서, 우선 사전확률분포의 모수들을 다음과 같이 선택한다.

$$\mathbf{b}_0 = \begin{bmatrix} -2 & 1 & 0 & 0 \end{bmatrix}^t, \quad B_0 = 2I_4 \quad (7)$$

로지스틱회귀모형의 회귀계수벡터  $\beta$ 의 사후확률분포에서 샘플링하는 Gibbs샘플러를 적용해서 Gibbs샘플들을  $Nsim = 200$  개 발생시켰다. 이때 번인기간은  $Nburnin = 1000$  이다. 이렇게 발생된 Gibbs샘플들의 표본평균은 다음과 같다.

$$\bar{\beta}_0 = -2.95578589, \quad \bar{\beta}_1 = 0.72332798 \quad (8)$$

$$\bar{\beta}_2 = 0.00884681, \quad \bar{\beta}_3 = 0.61462643 \quad (9)$$

또한, 분산공분산행렬은 다음과 같다.

$$Var(\beta) = \begin{bmatrix} 1.781483813 & -0.5973366662 & 0.006763879 & -0.1207609166 \\ -0.597336666 & 0.6102233673 & -0.056568527 & 0.0008136868 \\ 0.006763879 & -0.0565685272 & 0.007642873 & -0.0045655534 \\ -0.120760917 & 0.0008136868 & -0.004565553 & 0.4339825580 \end{bmatrix} \quad (10)$$

이 R프로그램 GibbsSamplerLogit102R.R을 실행하면, 사후확률분포로부터 Gibbs샘플러를 사용해서 발생시킨 그래프들을 그린 그림 5.3.12를 출력한다. 그림 5.3.12의 좌측상단 그래프에는  $\beta_0$ 의 히스토그램이, 우측상단 그래프에는  $\beta_1$ 의 히스토그램이, 좌측하단 그래프에는  $\beta_2$ 의 히스토그램이, 그리고 우측하단 그래프에는  $\beta_3$ 의 히스토그램이 그려져 있다. 또한, 이 Gibbs샘플들의 추정된 자기상관함수들이 그림 5.3.13에 그려져 있다. 그림 5.3.13의 좌측상단 그래프에는  $\beta_0$ 의 추정된 자기상관함수가, 우측상단 그래프에는  $\beta_1$ 의 추정된 자기상관함수가, 좌측하단 그래프에는  $\beta_2$ 의 추정된 자기상관함수가, 그리고 우측하단 그래프에는  $\beta_3$ 의 추정된 자기상관함수가 그려져 있다. ■

토빗모형 (tobit model)은 절삭이 있는 회귀모형 (censored regression model)으로, 종속변수  $y_i$ 가 어떤 경계값을 넘었을 때만 관측되는 회귀모형이다. 이 경계값을 0으로 하는

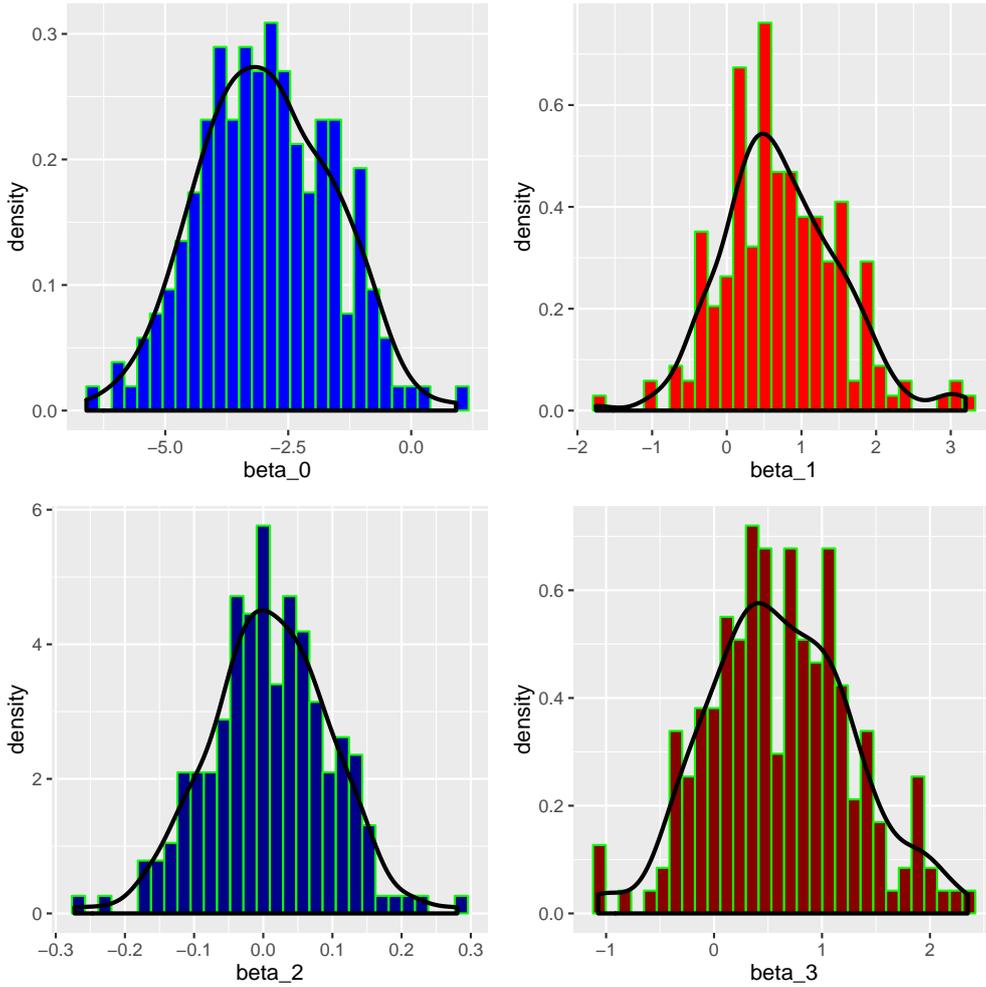


그림 5.3.12. 사후확률분포

토빗모형은 다음과 같다.

$$y_i = \max \{0, y_i^*\}, \quad (i = 1, 2, \dots, n) \tag{5.3.97}$$

$$y_i^* = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i, \quad (i = 1, 2, \dots, n) \tag{5.3.98}$$

여기에서  $y_i$ 는 시점  $i$ 에서 종속변수이고,  $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ip}]^t$ 는 설명변수벡터이고,  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_p]^t$ 은 모수벡터이며,  $\{\epsilon_i\}$ 는 서로 독립인 오차항들로서 다음 식들을 만족한다.

$$\epsilon_i | \sigma^2 \stackrel{d}{\sim} \mathcal{N}(0, \sigma^2), \quad (i = 1, 2, \dots, n) \tag{5.3.99}$$

식 (5.3.98)의  $y_i^*$ 는 종속변수이지만, 관측되는 것은 식 (5.3.97)의  $y_i$  뿐이다. 따라서, 식  $y_i^* \leq 0$

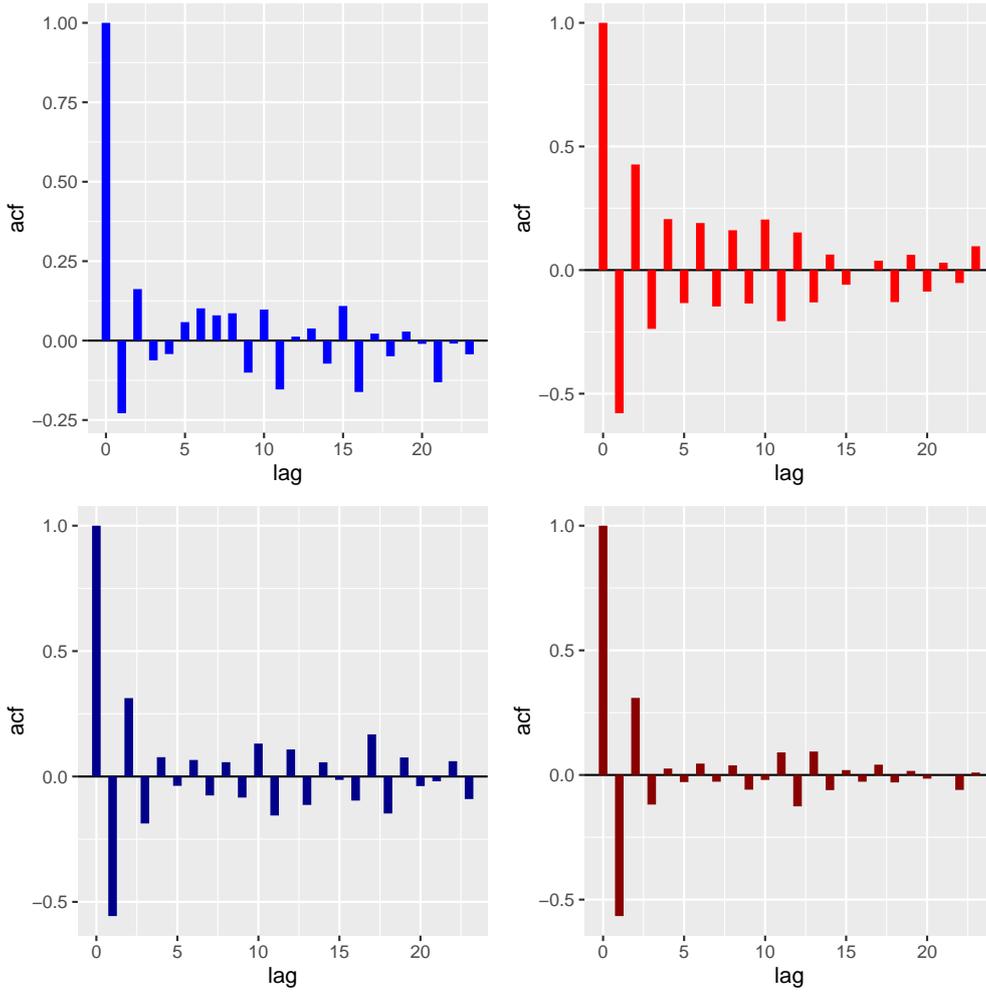


그림 5.3.13. 자기상관함수

가 성립하는 경우에는  $y_i^*$  는 관측되지 않는 잠재변수(latent variable)이다.

**예제 5.3.9** Fair [36], [37]는 *Psychology Today* 1970년 7월호에서 다룬 혼외정사에 관한 자료에서 발췌한 데이터를 분석하였다. 이 데이터세트 Affairs는 다음 확률변수들을 포함하고 있다.

$y$  = 지난해 혼외정사 횟수; 없음 ( $y = 0$ ), 1회 ( $y = 1$ ), 2회 ( $y = 2$ ), 3회 ( $y = 3$ ), 4회에서 10회 ( $y = 7$ ), 월별, 주별, 일별로 자주 ( $y = 12$ ). 표본평균은 1.46이고 빈도수는 (451, 34, 17, 19, , 42, 38)이다.

$z_1$  = 성별; 여자 ( $z_1 = 0$ ), 남자 ( $z_1 = 1$ ). 표본평균은 0.476이다.

$z_2$  = 나이. 표본평균은 32.5이다.

$z_3$  = 결혼기간. 표본평균은 32.5이다.

$z_4$  = 자녀 유무; 무 ( $z_4 = 0$ ), 유 ( $z_4 = 1$ ). 표본평균은 0.715이다.

$z_5$  = 종교적 성향; 종교에 대해 적대적 (anti)이면  $z_5 = 1$ , 아주 종교적이면  $z_5 = 5$ . 그 사이를 신심의 정도에 따라, 2, 3, 4를 선택한다. 표본평균은 3.12이다.

$z_6$  = 교육기간; 예를 들어 고등학교를 졸업했으면,  $z_6 = 12$ , 박사를 받았거나, 그 외 추가적 교육을 받았으면  $z_6 = 20$ . 표본평균은 16.2이다.

$z_7$  = 직업; Hollingshead 척도에 따라 1에서 7로 나타낸다. 표본평균은 4.19이다.

$z_8$  = 본인이 생각하는 결혼 행복도; 아주 나쁘면  $z_8 = 1$ , 아주 행복하면  $z_8 = 5$ . 그 사이를 행복도에 따라, 2, 3, 4를 선택한다. 표본평균은 3.93이다.

Greene [50]의 제22.3.6소절에서는 이 데이터세트 Affairs를 토빗모형으로 분석하였고, Greene [51]의 제18.4.6소절에서는 이 데이터세트 Affairs를 센서드Poisson 회귀분석(censored Poisson regression)과 센서음이항회귀분석(censored negative binomial regression)을 하였다.

최우추정법을 사용해서 데이터세트 Affairs에 적합한 토빗모형을 추정하는 예로서 다음 R 프로그램을 살펴보자.

```

1 # -----
2 # Filename: GibbsSamplerTobit101R.R
3 # Gibbs Sampler for Tobit Model
4 # https://cran.r-project.org/web/packages/AER/AER.pdf
5 # (Ref) Greene, W.H. (2012) Econometric Analysis (7th Ed), pp. 854-5.
6 # (Data Source) http://fairmodel.econ.yale.edu/raufair/work.ss.htm
7 # http://statsmodels.sourceforge.net/devel/datasets/generated/fair.html
8 # -----
9 #install.packages("AER")
10 library(AER)
11 remove(list=ls(all=TRUE));
12 data("Affairs")
13 sink("GibbsSamplerTobit101R.txt", append=TRUE, split=TRUE)
14 head(Affairs)
15 fm.tobit <- tobit(affairs ~ age + yearsmarried + religiousness +
16                   occupation + rating, data = Affairs)
17 summary(fm.tobit)
18 # -----

```

이 R프로그램은 본저자가 짠 것이 아니라 Kleiber & Zeileis가 제시한 R패키지 AER(Applied Econometrics with R)을 이용한 것이다. 이 R패키지는 Kleiber & Zeileis [59]을 바탕으로 한다.

이 R프로그램을 실행하면, 데이터세트 Affairs의 앞 부분을 출력한다. 그 내용은 다음과 같다.

	affairs	gender	age	yearsmarried	children	religiousness	education
4	0	male	37	10.00	no	3	18
5	0	female	27	4.00	no	4	14
11	0	female	32	15.00	yes	1	12
16	0	male	57	15.00	yes	5	18
23	0	male	22	0.75	no	2	17
29	0	female	32	1.50	no	2	17

	occupation	rating
4	7	4
5	6	4
11	1	4
16	6	5
23	6	3
29	5	5

또한, AER 패키지의 함수 `tobit`을 실행한 결과는 다음과 같다.

Call:

```
tobit(formula = affairs ~ age + yearsmarried + religiousness +
      occupation + rating, data = Affairs)
```

Observations:

Total	Left-censored	Uncensored	Right-censored
601	451	150	0

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	8.17420	2.74145	2.982	0.00287 **
age	-0.17933	0.07909	-2.267	0.02337 *
yearsmarried	0.55414	0.13452	4.119	3.80e-05 ***
religiousness	-1.68622	0.40375	-4.176	2.96e-05 ***
occupation	0.32605	0.25442	1.282	0.20001
rating	-2.28497	0.40783	-5.603	2.11e-08 ***
Log(scale)	2.10986	0.06710	31.444	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Scale: 8.247

```

Gaussian distribution
Number of Newton-Raphson Iterations: 4
Log-likelihood: -705.6 on 7 Df
Wald-statistic: 67.71 on 5 Df, p-value: 3.0718e-13

```

이 결과물에서 알 수 있듯이, 최우추정법에 의한 추정식은 다음과 같다.

$$\hat{y}_i = \max \{0, y_i^*\} \quad (1)$$

$$\hat{y}_i^* = 8.17420 - 0.17933z_2 + 0.55414z_3 - 1.68622z_5 + 0.32605z_7 - 2.28497z_8 \quad (2)$$

■

베이지안추정법을 사용해서 토빗모형을 추정하는 문제를 살펴보자. 정밀도를  $\psi = 1/\sigma^2$  라 하고, 모수  $(\boldsymbol{\beta}, \psi)$  의 사전확률분포가 다음과 같다고 하자.

$$\boldsymbol{\beta} \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_0, B_0), \quad \psi \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \quad (5.3.100)$$

벡터  $\mathbf{y}^* = [y_1^*, y_2^*, \dots, y_n^*]^t$  의 결합확률밀도함수는 다음과 같다.

$$f(\mathbf{y}^* | \boldsymbol{\beta}, \psi) \propto \psi^{n/2} \exp\left(-\frac{\psi}{2} \sum_{i=1}^n [y_i^* - \mathbf{x}_i^t \boldsymbol{\beta}]^2\right) \quad (5.3.101)$$

따라서,  $\mathbf{y}^*$  가 주어진 조건 하에서 사후조건부확률분포들은 다음과 같다.

$$\psi | (\boldsymbol{\beta}, \mathbf{y}^*) \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_{po}}{2}, \frac{n_{po} S_{po}}{2}\right) \quad (5.3.102)$$

여기서 행렬, 벡터 그리고 상수들은 다음과 같다.

$$B_{po} \doteq \left[ B_0^{-1} + \psi \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^t \right]^{-1}, \quad \mathbf{b}_{po} \doteq B_{po} \left[ B_0^{-1} \mathbf{b}_0 + \psi \sum_{i=1}^n y_i^* \mathbf{x}_i \right] \quad (5.3.103)$$

$$n_{po} \doteq n_0 + n, \quad S_{po} \doteq \frac{1}{n_{po}} \left\{ n_0 S_0 + \sum_{i=1}^k [y_i^* - \mathbf{x}_i^t \boldsymbol{\beta}]^2 \right\} \quad (5.3.104)$$

만약  $y_i > 0$  이면, 식  $y_i^* = y_i$  이 성립한다. 만약  $y_i = 0$  이면, 식  $y_i^* < 0$  이 성립한다. 이  $y_i^*$  를 샘플링하는 과정을 살펴보자. 조건  $y_i = 0$  하에서 잠재변수  $y_i^*$  의 조건부확률밀도함수는

다음과 같다.

$$\pi(y_i^* | \boldsymbol{\beta}, \psi, \mathbf{y}) \propto \exp\left(-\frac{\psi}{2} [y_i^* - \mathbf{x}_i^t \boldsymbol{\beta}]^2\right) 1(y_i^* \leq 0) \quad (5.3.105)$$

식 (5.3.105)에서 알 수 있듯이,  $y_i^*$ 는 구간  $(-\infty, 0]$ 에서 절단되고 평균이  $\mathbf{x}_i^t \boldsymbol{\beta}$ 이며 분산이  $\psi^{-1}$ 인 절단정규확률분포  $\mathcal{TN}_{(-\infty, 0]}(\mathbf{x}_i^t \boldsymbol{\beta}, \psi^{-1})$ 이다. 따라서, 이 절단정규확률분포로부터 잠재변수  $y_i^*(\leq 0)$ 를 샘플링한다. 절단정규확률분포  $\mathcal{TN}_{(\alpha, \beta]}(\mu, \sigma^2)$ 에서 난수를 발생시키기 위해서는 먼저  $\mathcal{N}(\mu, \sigma^2)$ 에서 정규난수를 발생시키고, 그 정규난수가 구간  $(\alpha, \beta]$ 에 속하면 절단정규난수로 선택하고, 그렇지 않으면 구간  $(\alpha, \beta]$ 에 속하는 정규난수가 발생할 때까지 반복해서 다음 절단정규난수를 발생시킨다. 절단정규확률분포에서 효율적으로 샘플링하는 방법에 관해서는 Geweke [44]를 참조하라.

지금까지 내용을 바탕으로, 토빗모형의 모수들  $(\boldsymbol{\beta}, \psi)$ 의 사후확률분포에서 샘플링하는 Gibbs샘플러를 다음과 같이 요약할 수 있다.

**(1단계)** 초기모수들  $(\boldsymbol{\beta}^{(0)}, \psi^{(0)})$ 를 선택한다.

**(2단계)** 각  $j(= 1, 2, \dots)$ 에 대해서 샘플  $(y_i^{*(j)}, \boldsymbol{\beta}^{(j)}, \psi^{(j)})$ 를 발생시킨다. 즉, 식  $y_j = 0$ 을 만족하는  $j$ 에 대해 다음 사후조건부확률분포로부터 샘플을 발생시킨다.

$$y_i^{*(j)} | (\boldsymbol{\beta}^{(j-1)}, \psi^{(j-1)}) \stackrel{d}{\sim} \mathcal{TN}_{(-\infty, 0]}(\mathbf{x}_i^t \boldsymbol{\beta}^{(j-1)}, 1/\psi^{(j-1)}) \quad (5.3.106)$$

$$\boldsymbol{\beta}^{(j)} | (\psi^{(j-1)}, \mathbf{y}^{*(j)}) \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_1^{(j)}, B_1^{(j)}) \quad (5.3.107)$$

$$\psi | (\boldsymbol{\beta}^{(j)}, \mathbf{y}^{*(j)}) \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_1}{2}, \frac{n_1 S_1^{(j)}}{2}\right) \quad (5.3.108)$$

여기서 행렬, 벡터와 상수는 다음과 같다.

$$B_1^{(j)} \doteq \left[ B_0^{-1} + \psi^{(j-1)} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^t \right]^{-1} \quad (5.3.109)$$

$$\mathbf{b}_1^{(j)} \doteq B_1^{(j)} \left[ B_0^{-1} \mathbf{b}_0 + \psi^{(j-1)} \sum_{i=1}^n y_i^{*(j)} \mathbf{x}_i \right] \quad (5.3.110)$$

$$S_1^{(j)} \doteq \frac{1}{n_1} \left\{ n_0 S_0 + \sum_{i=1}^k [y_i^{*(j)} - \mathbf{x}_i^t \boldsymbol{\beta}^{(j)}]^2 \right\} \quad (5.3.111)$$

**(3단계)** 충분히 큰  $b$ 에 대해서  $(\boldsymbol{\beta}^{(j)}, \psi^{(j)})$ ,  $(j = b + 1, b + 2, \dots)$ 를 기록한다.

**예제 5.3.10** Gibbs샘플러를 사용해서 데이터세트 Affairs에 적합한 토빗모형을 추정하기 위해서 다음 MATLAB 프로그램 GibbsSamplerTobit102.m을 실행하자.

```

1 % -----
2 % Filename: GibbsSamplerTobit102.m
3 % Gibbs Sampler for Tobit Model
4 % Programmed by CBS
5 % -----
6 function GibbsSamplerTobit102
7 clear all, close all, clc, format short
8 rng((1001^2+1)/2)
9
10 % Read Dataset
11 fileID = fopen('FairAffairs.txt');
12 sizeFA = [601,15];
13 FA = fscanf(fileID,'%d %d %f %d %f %f %d %d %f %f %d %d %f %f %f',sizeFA);
14 fclose(fileID);
15 Y = FA(:,13); % No of Affairs
16 n = length(Y);
17 z1 = FA(:,4); % Sex
18 z2 = FA(:,5); % Age
19 z3 = FA(:,6); % Married years
20 z4 = FA(:,7); % Children
21 z5 = FA(:,8); % Religiousness
22 z6 = FA(:,9); % Education
23 z7 = FA(:,11); % Occupation
24 z8 = FA(:,12); % Self-rating of marriage
25 X = [ones(n,1), z2, z3, z5, z7, z8 ];
26 [n,k] = size(X);
27
28 % Prior
29 b0 = 100*ones(6,1);
30 B0 = (1000^2)*eye(k);
31 n0 = 100;
32 S0 = 200;
33 psii0 = 1/S0;
34 invB0 = inv(B0);
35
36 % Posterior
37 n0 = 16;
38 Nburnin = 2000;
39 Nsim = 2000;
40 betaMat = zeros(Nburnin + Nsim, k);
41 psiiMat = zeros(Nburnin + Nsim, 1);
42 isY0 = (Y == 0); % Indicator function
43 betaStar = b0;
44 psiiStar = psii0;
45
46 % % Gibbs Sampling
47 Lower = -inf;
48 Upper = 0;
49 for ii = 1:(Nburnin + Nsim)
50     % % yStar
51     TNormalDum = normcdf(Lower)*ones(n,1) ...
52                 +(normcdf(Upper)-normcdf(Lower))*rand(n,1);
53     TNormal = X*betaStar + norminv(TNormalDum);
54     yStar = isY0.*TNormal + (1 - isY0).*Y;
55     % % Parameters
56     nStar = n0 + n;

```

```

57     BStar = inv( invB0 + psiiStar*X'*X );
58     bStar = BStar*( invB0*b0 + psiiStar*X'*yStar );
59     eDum = yStar - X*betaStar;
60     SStar = 1/nStar*(n0*S0 + eDum'*eDum);
61     betaStar = mvnrnd(bStar,BStar)';
62     psiiStar = gamrnd(nStar/2,nStar*SStar/2,1);
63     betaMat(ii,:) = betaStar';
64     psiiMat(ii) = psiiStar;
65 end
66 betaMat = betaMat(Nburnin+1:Nburnin+Nsim,:);
67 psiiMat = psiiMat(Nburnin+1:Nburnin+Nsim,:);
68 sigsigMat = 1./psiiMat;
69 %% Statistics
70 meanbeta = mean(betaMat,1)
71 stdbeta = std(betaMat,1)
72 covbeta = cov(betaMat)
73 meansigsig = mean(sigsigMat)
74
75 % Plotting
76 figure
77 subplot(2,2,1)
78 h0 = histfit(betaMat(:,1),24)
79 h0(1).FaceColor = [.8 .8 1];
80 h0(2).Color = [.3 .3 .3];
81 set(gca,'fontsize',11,'fontweigh','bold')
82 ylabel('\bf \beta_{0}','rotation',0)
83 subplot(2,2,2)
84 h1 = histfit(betaMat(:,2),24)
85 h1(1).FaceColor = [.8 .8 .9];
86 h1(2).Color = [1.0 .0 .0];
87 set(gca,'fontsize',11,'fontweigh','bold')
88 ylabel('\bf \beta_{1}','rotation',0)
89 subplot(2,2,3)
90 h2 = histfit(betaMat(:,3),24)
91 h2(1).FaceColor = [.8 .9 .8];
92 h2(2).Color = [0 1.0 .0];
93 set(gca,'fontsize',11,'fontweigh','bold')
94 ylabel('\bf \beta_{2}','rotation',0)
95 subplot(2,2,4)
96 h3 = histfit(betaMat(:,4),24)
97 h3(1).FaceColor = [.9 .8 .8];
98 h3(2).Color = [0 0 1.0];
99 set(gca,'fontsize',11,'fontweigh','bold')
100 ylabel('\bf \beta_{3}','rotation',0)
101 saveas(gcf,'GibbsSamplerTobit102a','eps')
102
103 figure
104 subplot(2,2,1)
105 [hist4Freq, hist4Out] = hist(betaMat(:,5),24);
106 binWidth = hist4Out(2)-hist4Out(1);
107 bar(hist4Out,hist4Freq/binWidth/sum(hist4Freq),'r');
108 hold on
109 ProbDensity = fitdist(betaMat(:,5),'normal');
110 plot(hist4Out, pdf(ProbDensity,hist4Out),'k','linewidth',1.8);
111 set(gca,'fontsize',11,'fontweigh','bold')
112 ylabel('\bf \beta_{4}','rotation',0)
113 hold off
114 subplot(2,2,2)
115 [hist5Freq, hist5Out] = hist(betaMat(:,6),24);
116 binWidth = hist5Out(2)-hist5Out(1);
117 ProbDensity = normpdf(hist5Out,meanbeta(6),stdbeta(6));

```

```

118 hold on
119 bar(hist5Out,hist5Freq/trapz(hist5Out,hist5Freq),'g');
120 plot(hist5Out,ProbDensity,'k','linewidth',1.8);
121 set(gca,'fontsize',11,'fontweigh','bold')
122 ylabel('\bf \beta_{5}','rotation',0)
123 hold off
124 subplot(2,2,3)
125 [histSFreq, histSOut] = hist(sigsigMat,24);
126 binWidth = histSOut(2)-histSOut(1);
127 bar(histSOut,histSFreq/binWidth/sum(histSFreq),'b');
128 hold on
129 ProbDensity = fitdist(sigsigMat,'normal');
130 plot(histSOut, pdf(ProbDensity,histSOut),'k','linewidth',1.8);
131 set(gca,'fontsize',11,'fontweigh','bold')
132 ylabel('\bf \sigma^{2}','rotation',0)
133 hold off
134 subplot(2,2,4)
135 [histPFreq, histPOut] = hist(psiMat,24);
136 binWidth = histPOut(2)-histPOut(1);
137 bar(histPOut,histPFreq/binWidth/sum(histPFreq),'w');
138 hold on
139 ProbDensity = fitdist(psiMat,'Gamma');
140 plot(histPOut, pdf(ProbDensity,histPOut),'k','linewidth',1.8);
141 set(gca,'fontsize',11,'fontweigh','bold')
142 ylabel('\bf \psi','rotation',0)
143 hold off
144 saveas(gcf,'GibbsSamplerTobit102b','epsc')
145 end
146 % End of program
147 % -----

```

예제 5.3.9에서 다룬 Fair의 데이터세트 Affairs를 Fair의 웹사이트에서 다운받아 ASCII 파일 FairAffairs.txt에 저장하였다. 이 MATLAB프로그램 GibbsSamplerTobit102.m에서는 베이지안통계기법을 사용해서 다음 토빗모형을 추정한다.

$$y = \max(0, y_i^*) \quad (1)$$

$$y_i^* = \beta_0 + \beta_1 z_2 + \beta_2 z_3 + \beta_3 z_5 + \beta_4 z_7 + \beta_5 z_8 + \epsilon \quad (2)$$

사전확률분포의 모수들을 다음과 같이 선택한다.

$$\mathbf{b}_0 = \left[ 100 \ 100 \ 100 \ 100 \ 100 \ 100 \right]^t, \quad B_0 = 1000^2 I_6 \quad (3)$$

$$n_0 = 100, \quad S_0 = 200, \quad \psi_0 = \frac{1}{S_0} \quad (4)$$

토빗모형의 회귀계수벡터  $\beta$ 의 사후확률분포에서 샘플링하는 Gibbs샘플러를 적용해서 Gibbs샘플들을  $Nsim = 2000$ 개 발생시켰다. 이때 번인기간은  $Nburnin = 2000$ 이다. 이렇게

발생된 Gibbs샘플들의 표본평균들은 다음과 같다.

$$\bar{\beta}_0 = 75.5006, \quad \bar{\beta}_1 = -0.0719, \quad \bar{\beta}_2 = -0.0676 \quad (5)$$

$$\bar{\beta}_3 = -0.0515, \quad \bar{\beta}_4 = -0.0408, \quad \bar{\beta}_5 = -0.0109 \quad (6)$$

또한, 분산공분산행렬은 다음과 같다.

$$Var(\beta) = 10^{-3} \begin{bmatrix} 0.1333 & -0.0002 & -0.0001 & -0.0001 & -0.0001 & -0.0000 \\ -0.0002 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ -0.0001 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ -0.0001 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ -0.0001 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ -0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix} \quad (7)$$

이 MATLAB프로그램 GibbsSamplerTobit102.m 을 실행하면, 사후확률분포로부터 Gibbs 샘플러를 사용해서 발생시킨 그래프들을 그린 그림 5.3.14와 그림 5.3.15를 출력한다. 그림 5.3.14의 좌측상단 그래프에는  $\beta_0$ 의 히스토그램이, 우측상단 그래프에는  $\beta_1$ 의 히스토그램이, 좌측하단 그래프에는  $\beta_2$ 의 히스토그램이, 그리고 우측하단 그래프에는  $\beta_3$ 의 히스토그램이 그려져 있다. 그림 5.3.15의 좌측상단 그래프에는  $\beta_4$ 의 히스토그램이, 우측상단 그래프에는  $\beta_5$ 의 히스토그램이, 좌측하단 그래프에는  $\sigma^2$ 의 히스토그램이, 그리고 우측하단 그래프에는  $\psi$ 의 히스토그램이 그려져 있다.

다음 Python프로그램 GibbsSamplerTobit102.py를 실행해도, MATLAB프로그램 GibbsSamplerTobit102.m를 실행한 것과 동일한 결과를 얻는다.

```

1 # GibbsSamplerTobit102.py
2 """
3 Filename : GibbsSamplerTobit102.py
4 Gibbs Sampler for Tobit Model
5 Programmed by Tae-Kon Yang based on the R program by CBS
6 """
7
8 import numpy as np
9 import pandas as pd
10 from scipy.stats import norm, multivariate_normal, gamma
11 from scipy.special import ndtri
12 from sympy import oo
13 import matplotlib.pyplot as plt
14 import matplotlib.mlab as mlab
15
16 #read dataset
17 FA = np.loadtxt('FairAffair.txt')

```

```

18 np.random.seed(12345)
19
20 Y = FA[:,12].reshape((-1,1))
21 n = len(Y)
22 z1 = FA[:,3]#Sex
23 z2 = FA[:,4]#Age
24 z3 = FA[:,5]#Married Years
25 z4 = FA[:,6]#Children
26 z5 = FA[:,7]#Religiousness
27 z6 = FA[:,8]#Education
28 z7 = FA[:,10]#Occupation
29 z8 = FA[:,11]#Self-rating of marriage
30 X = np.c_[np.ones((n,1)), z2, z3, z5, z7, z8]
31 n, k = X.shape
32
33
34 #Prior
35 b0 = 100*np.ones((6,1))
36 B0 = (1000**2)*np.eye((6))#n,k=shape(X)
37 n0 = 100
38 S0 = 200
39 psii0 = 1/S0
40 invB0 = np.linalg.inv(B0)
41
42 #Posterior
43 n02 =16 # assigned n0 twice in the original code
44 Nburnin = 2000
45 Nsim = 2000
46 betaMat = np.zeros((Nburnin + Nsim, k))
47 psiiMat = np.zeros((Nburnin+Nsim, 1))
48
49 isY0 = (Y==0)
50 betaStar = b0
51 psiiStar = psii0
52
53 #Gibbs Sampling
54 Lower = -np.inf
55 Upper = 0
56 TNormalDum = np.zeros((n,1))
57 TNormal = np.zeros((n,1))
58 for ii in range(Nburnin + Nsim):
59     TNormalDum = norm.cdf(Lower)*np.ones((n,1))+(norm.cdf(Upper)-norm.cdf(Lower))
60     *np.random.rand(n,1)
61     TNormal = np.dot(X,betaStar) + ndtri(TNormalDum)
62     #TNormal = TNormal.reshape(n,)
63     yStar = isY0*TNormal +(1-isY0)*Y
64
65 #parameters
66     nStar = n02 +n
67
68     BStar = np.linalg.inv(invB0+np.dot(psiiStar,np.dot(X.T,X)))
69
70     bStar = np.dot(BStar,np.dot(invB0,b0)+psiiStar*np.dot(X.T,yStar))#7|yStar
71     601,1
72     eDum = yStar - np.dot(X, betaStar)
73     SSStar = 1/nStar*(n0*S0 +np.dot(eDum.T,eDum))
74     betaStar = multivariate_normal.rvs(bStar[:,0], BStar).reshape((-1,1))#mean
75     vector must be 1-d array
76     psiiStar = gamma.rvs(nStar/2, nStar*SSStar/2, 1)
77     betaMat[ii,:] = betaStar.T
78     psiiMat[ii] = psiiStar

```

```

76
77 betaMat = betaMat[Nburnin : Nburnin+Nsim, :]
78 psiiMat = psiiMat[Nburnin : Nburnin+Nsim, :]#차원이 하나 부족함
79 sigsigMat = 1/psiiMat
80
81 #statistics
82 meanbeta = np.mean(betaMat,axis=0)
83 stdbeta = np.std(betaMat, axis=0)
84 covbeta = np.cov(betaMat.T)
85 meansigsig = np.mean(sigsigMat)
86
87
88
89
90 #plotting
91 plt.rc('text', usetex=True)
92 num_bins = 24
93
94 fig = plt.figure()
95 ax1 = plt.subplot(2, 2, 1)
96 n1, bins1, patches1= ax1.hist(betaMat[:,0], num_bins, color = 'm')#count,
    lefthand edge
97 mu1 = meanbeta[0]
98 sigma1 = stdbeta[0]
99 # add a 'best fit' line
100 binscenters1= 0.5*(bins1[1:]+bins1[:-1])
101 y = mlab.normpdf(binscenters1, mu1, sigma1)
102 ax1.plot(binscenters1, y, 'r-')
103 ax1.set_ylabel(r"\beta_{0}",rotation=1)
104
105 # add a 'best fit' line
106
107 ax2 = plt.subplot(2, 2, 2)
108 n2, bins2, patches2 = ax2.hist(betaMat[:,1], num_bins, color='b')
109 mu2 = meanbeta[1]
110 sigma2 = stdbeta[1]
111 binscenters2= 0.5*(bins2[1:]+bins2[:-1])
112 y2 = mlab.normpdf(binscenters2, mu2, sigma2)
113 ax2.plot(binscenters2, y2, 'r-')
114 ax2.set_ylabel(r"\beta_{1}",rotation=1)
115
116 ax3 = plt.subplot(2, 2, 3)
117 n3, bins3, patches3 = ax3.hist(betaMat[:,2], num_bins, color = 'g')
118 mu3 = meanbeta[2]
119 sigma3 = stdbeta[2]
120 binscenters3= 0.5*(bins3[1:]+bins3[:-1])
121 y3 = mlab.normpdf(binscenters3, mu3, sigma3)
122 ax3.plot(binscenters3, y3, 'r-')
123 ax3.set_ylabel(r"\beta_{2}",rotation=1)
124
125 ax4 = plt.subplot(2, 2, 4)
126 n4, bins4, patches4 = ax4.hist(betaMat[:,3], num_bins, histtype='bar', color='k')
127 mu4 = meanbeta[3]
128 sigma4 = stdbeta[3]
129 binscenters4= 0.5*(bins4[1:]+bins4[:-1])
130 y4 = mlab.normpdf(binscenters4, mu4, sigma4)
131 ax4.plot(binscenters4, y4, 'r-')
132 ax4.set_ylabel(r"\beta_{3}",rotation=1)
133
134 plt.show()
135 fig.savefig('GibbsSamplerTobit102a')

```

```

136
137 #group b
138 fig2 = plt.figure()
139 ax5 = plt.subplot(2, 2, 1)
140 n5, bins5 = np.histogram(betaMat[:,4], num_bins)#n5: frequency , bins5 : left
    edges
141 mu5, sigma5 = norm.fit(betaMat[:,4])
142 binscenters5 = 0.5*(bins5[1:]+bins5[:-1])
143 binwidth = n5[1] -n5[0]
144 n_bar = np.asarray([i/binwidth/sum(n5) for i in n5])
145 ax5.bar(binscenters5,n_bar,color='r')
146 ax5.plot(binscenters5, norm.pdf(binscenters5, mu5, sigma5),'k',linewidth = 1.8)
147 plt.ylabel(r"\beta_{4}",rotation=1)
148
149 ax6 = plt.subplot(2, 2, 2)
150 hist5freq, hist5out = np.histogram(betaMat[:,5], bins=num_bins)
151 mu6, sigma6 = norm.fit(betaMat[:,5])
152 center2 = (hist5out[:-1]+hist5out[1:])/2
153 ax6.bar(center2, hist5freq/np.trapz(center2,hist5freq), color='g')
154 ax6.plot(center2, norm.pdf(center2, mu6, sigma6),'r--')
155 plt.ylabel(r"\beta_{5}",rotation=0)
156
157 ax7 = plt.subplot(2, 2, 3)
158 hist6freq, hist6out = np.histogram(sigsigMat, bins=num_bins)
159 binwidth3 = hist6out[1] - hist6out[0]
160 hist6freq_bar = [i/binwidth3/sum(hist6freq) for i in hist6freq]
161 hist6freq_bar = np.asarray(hist6freq_bar)
162 center3 = (hist6out[:-1]+hist6out[1:])/2
163 ax7.bar(center3, hist6freq_bar,color='r')
164 mu7, sigma7 = norm.fit(sigsigMat)
165 ax7.plot(center3, norm.pdf(center3, mu7, sigma7),'g-')
166 plt.ylabel(r"\beta_{6}",rotation=0)
167
168 ax8 = plt.subplot(2, 2, 4)
169 hist7freq, hist7out = np.histogram(psiMat, bins=num_bins)
170 binwidth4 = hist7out[1] - hist7out[0]
171 hist7freq_bar = [i/binwidth4/sum(hist7freq) for i in hist7freq]
172 hist7freq_bar = np.asarray(hist7freq_bar)
173 center4 = (hist7out[:-1]+hist7out[1:])/2
174 ax8.bar(center4, hist7freq_bar,color='r')
175 alpha_g, loc_g, beta_g = gamma.fit(psiMat)
176 ax8.plot(center4, gamma.pdf(center4, alpha_g, loc_g, beta_g))
177 plt.ylabel(r"\beta_{7}",rotation=0)
178
179
180
181
182 plt.show()
183 fig2.savefig('GibssTobitb')
184
185
186 # End of program
187 # -----

```



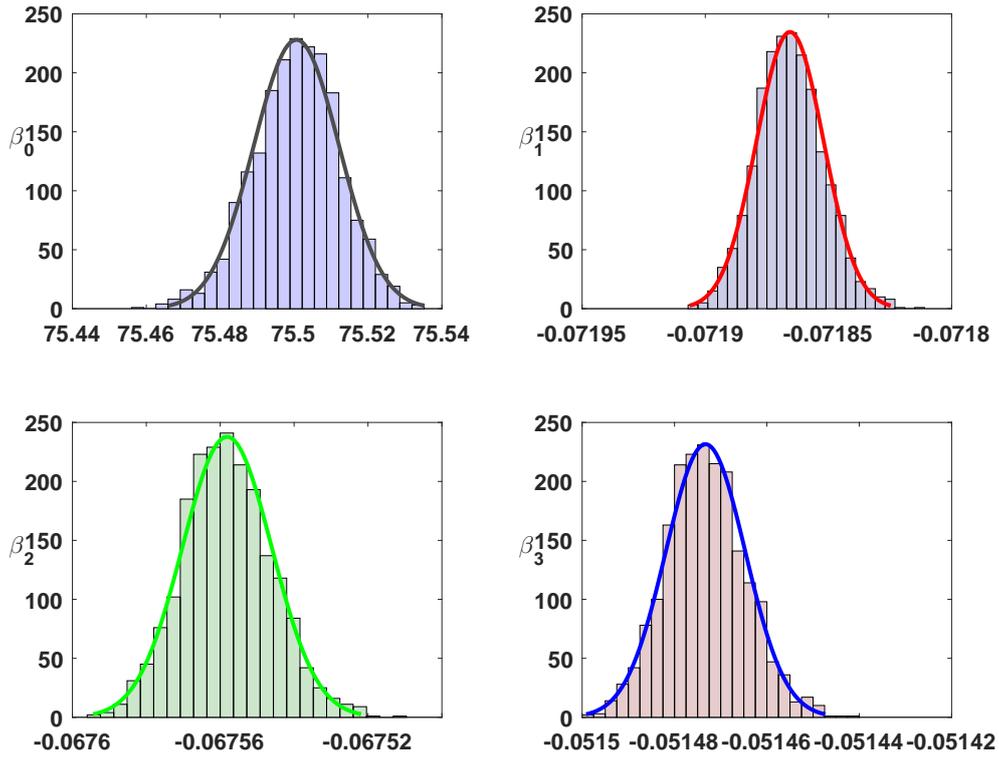


그림 5.3.14. 사후확률분포 1

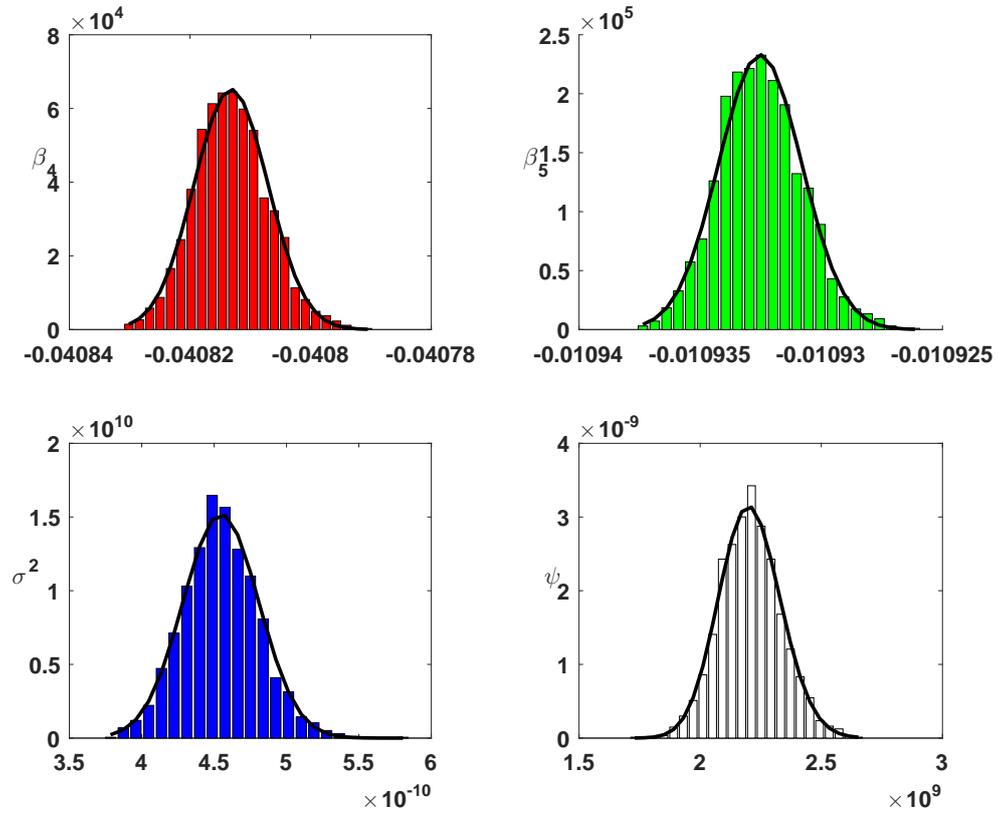


그림 5.3.15. 사후확률분포 2

### 5.3.6 SUR 모형과 Gibbs 샘플러

이 소절에서는 ‘보기에는 무관한 회귀모형 (seemingly unrelated regression model: SUR model)’의 추정에 대해서 살펴보자. 베이지안 추정법을 사용해서 ‘보기에는 무관한 회귀모형’을 추정하는 과정을 살펴보기 전에, 먼저 보기에는 무관한 회귀모형 자체에 대해서 살펴보자.

만약 두 선형회귀모형들이 외형적으로 어떤 관계를 가지고 있지 않으나 같은 시점에서 한 선형회귀모형의 오차항이 다른 선형회귀모형의 오차항과 관계를 가지면, 이 관계를 선형회귀모형들을 추정하는 데 반영하므로써 좀 더 효율적인 추정을 할 수 있다. 다음과 같은 선형회귀모형들을 살펴보자.

$$y_{1j} = \beta_{10} + \beta_{11}x_{1j1} + \beta_{12}x_{1j2} + \cdots + \beta_{1K}x_{1jK} + \epsilon_{1j} \quad (5.3.112)$$

$$y_{2j} = \beta_{20} + \beta_{21}x_{2j1} + \beta_{22}x_{2j2} + \cdots + \beta_{2K}x_{2jK} + \epsilon_{2j} \quad (5.3.113)$$

식 (5.3.112)와 식 (5.3.113)의 두 선형회귀모형들을 다음과 같은 선형회귀모형 한 개로 나타낼 수 있다.

$$\mathbf{y}_j = X_j \boldsymbol{\beta} + \boldsymbol{\epsilon}_j \quad (5.3.114)$$

여기서  $\{\boldsymbol{\epsilon}_j = [\epsilon_{1j}, \epsilon_{2j}]^t \mid j = 1, 2, \dots, T\}$ 는 서로 독립인 확률벡터들로서,  $\boldsymbol{\epsilon}_j$ 는 평균벡터가  $\mathbf{0}$ 이며 다음과 같은 분산공분산행렬을 갖는 확률벡터이다.

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \quad (5.3.115)$$

다음과 같은 벡터들과 행렬을 정의하자.

$$\mathbf{y} \doteq \begin{bmatrix} y_{11} \\ y_{21} \\ y_{12} \\ y_{22} \\ \vdots \\ y_{1T} \\ y_{2T} \end{bmatrix}, \quad \boldsymbol{\epsilon} \doteq \begin{bmatrix} \epsilon_{11} \\ \epsilon_{21} \\ \epsilon_{12} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{1T} \\ \epsilon_{2T} \end{bmatrix}, \quad X \doteq \begin{bmatrix} 1 & x_{111} & \cdots & x_{11K} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & x_{211} & \cdots & x_{21K} \\ 1 & x_{121} & \cdots & x_{12K} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & x_{221} & \cdots & x_{22K} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1T1} & \cdots & x_{1TK} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & x_{2T1} & \cdots & x_{2TK} \end{bmatrix} \quad (5.3.116)$$

$$\boldsymbol{\beta} \doteq [\beta_{10}, \beta_{11}, \dots, \beta_{1K}, \beta_{20}, \beta_{21}, \dots, \beta_{2K}]^t \quad (5.3.117)$$

여기서 오차항벡터  $\boldsymbol{\epsilon}$ 은 다음 식들을 만족한다.

$$E(\boldsymbol{\epsilon}) = \mathbf{0}, \quad Var(\boldsymbol{\epsilon}) = \begin{bmatrix} \Sigma & O & O & \cdots & O \\ O & \Sigma & O & \cdots & O \\ O & O & \Sigma & \cdots & O \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ O & O & O & \cdots & \Sigma \end{bmatrix} \quad (5.3.118)$$

만약  $\sigma_{12}$ 가 0이 아니라면, 선형회귀모형 (5.3.114)은 Gauss-Markov조건을 만족하지 못한다. 따라서, 최소제곱추정량  $\hat{\boldsymbol{\beta}}$ 은 효율적이지 못하고 일반화최소제곱추정량  $\tilde{\boldsymbol{\beta}}$ 가 최량선형불편추정량(best linear unbiased estimator: BLUE)이다. 좀 더 일반적으로, 만약 두 개 이상 선형회귀모형들이 외형적으로 어떤 관계를 가지고 있지 않으나 같은 시점에서 한 선형회귀모형의 오차항과 다른 선형회귀모형의 오차항의 공분산이 0이 아니면, 최소제곱추정량보다는 일반화최소제곱추정량이 더 효율적이다. Zellner [98]는 이러한 선형회귀모형을 ‘보기에는 무관한 회귀모형’이라 명명하고, 이 문제에 대한 일반화최소제곱추정법을 제시했다. 이를 결합일반화최소제곱추정법(joint generalized least squares method)이라고 한다.

결합일반화최소제곱추정법을 설명하기 위해서, 다음과 같은  $M$ 개 선형회귀모형들을 살펴보자.

$$\mathbf{y}_i = X_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \quad (i = 1, 2, \dots, M) \quad (5.3.119)$$

여기서  $\mathbf{y}_i$ 는  $T \times 1$  행렬,  $X_i$ 는  $T \times K_i$  행렬,  $\boldsymbol{\beta}_i$ 는  $K_i \times 1$  행렬 그리고  $\boldsymbol{\epsilon}_i$ 는  $T \times 1$  행렬이다.

또한 오차항벡터  $\boldsymbol{\epsilon}_i$ 는 다음 성질들을 만족한다.

$$E(\boldsymbol{\epsilon}_i) = \mathbf{0}, \quad \text{Var}(\boldsymbol{\epsilon}_i) = \sigma_{ii}I_T, \quad \text{Cov}(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) = \sigma_{ij}I_T, \quad (1 \leq i < j \leq M) \quad (5.3.120)$$

다음 상수와 행렬을 정의하자.

$$K \doteq \sum_{i=1}^M K_i \quad (5.3.121)$$

$$\Sigma \doteq \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2M} \\ \vdots & \vdots & & \vdots \\ \sigma_{1M} & \sigma_{2M} & \cdots & \sigma_{MM} \end{bmatrix} \quad (5.3.122)$$

또한, 다음과 같은 벡터들과 행렬을 정의하자.

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & O & O & \cdots & O \\ O & X_2 & O & \cdots & O \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ O & O & O & \cdots & X_M \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_M \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{bmatrix} \quad (5.3.123)$$

여기서  $\mathbf{y}$ 는  $TM \times 1$  행렬,  $X$ 는  $TM \times K$  행렬,  $\boldsymbol{\epsilon}$ 은  $TM \times 1$  행렬, 그리고  $\boldsymbol{\beta}$ 는  $K \times 1$  행렬이다. 식 (5.3.119)의  $M$ 개 선형회귀모형들을 다음과 같이 하나의 선형회귀모형으로 나타낼 수 있다.

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad E(\boldsymbol{\epsilon}) = \mathbf{0} \quad (5.3.124)$$

다음 식들이 성립한다.

$$\text{Var}(\boldsymbol{\epsilon}) = \begin{bmatrix} \sigma_{11}I_T & \sigma_{12}I_T & \cdots & \sigma_{1M}I_T \\ \sigma_{12}I_T & \sigma_{22}I_T & \cdots & \sigma_{2M}I_T \\ \vdots & \vdots & & \vdots \\ \sigma_{1M}I_T & \sigma_{2M}I_T & \cdots & \sigma_{MM}I_T \end{bmatrix} = \Sigma \otimes I_T \quad (5.3.125)$$

이 선형회귀모형의 일반화최소제곱추정량은 다음과 같다.

$$\tilde{\beta} = \left( X^t [\Sigma \otimes I_T]^{-1} X \right)^{-1} X^t [\Sigma \otimes I_T]^{-1} \mathbf{y} \tag{5.3.126}$$

이 추정량을 결합일반화최소제곱추정량이라 한다. 식  $(E \otimes F)^{-1} = E^{-1} \otimes F^{-1}$ 를 식 (5.3.126)에 적용하면, 다음 식을 얻는다.

$$\tilde{\beta} = \left( X^t [\Sigma^{-1} \otimes I_T] X \right)^{-1} X^t [\Sigma^{-1} \otimes I_T] \mathbf{y} \tag{5.3.127}$$

최소제곱추정량에 비해서 이 결합일반화최소제곱추정량은 효율적이다. 그러나, 다음과 같이 두 가지 특수한 경우에는 결합일반화최소제곱추정량과 최소제곱추정량은 동일하다. 첫째, 분산공분산행렬  $\Sigma$ 가 대각행렬이다. 둘째, 모든 선형회귀모형들의 디자인행렬들이 동일하다. 즉, 다음 식이 성립한다.

$$X_1 = X_2 = \dots = X_M \tag{5.3.128}$$

이에 대한 증명은 최병선 [4]의 제9.7절을 참조하라. 이로부터 다음과 같은 성질을 유추할 수 있다. 같은 시점에서 서로 다른 선형회귀모형들의 오차항들의 상관관계가 높다면, 일반적으로 결합일반화최소제곱추정량의 효율성과 최소제곱추정량의 효율성의 차이가 크다. 또한, 서로 다른 선형회귀모형들의 설명변수들의 상관관계가 높다면, 일반적으로 결합일반화최소제곱추정량의 효율성과 최소제곱추정량의 효율성의 차

**예제 5.3.11** SAS/ETS에서 보기에는 무관한 회귀를 실행하기 위해서는 PROC SYSLIN을 사용하거나 PROC MODEL을 사용한다. PROC MODEL을 사용하는 방법에 대해서는 최병선 [4, pp. 325-328]을 참조하라. 여기서는 PROC SYSLIN을 사용해서 보기에는 무관한 회귀를 수행하는 예제로서 다음 SAS프로그램을 수행해 보자.

```

1  * -----*
2  | Example 27.2 Grunfeld's Model Estimated with SUR of
3  |           SAS/ETS(R) 9.22 User's Guide
4  | http://support.sas.com/documentation/cdl/en/etsug/63348/HTML
5  |           /default/viewer.htm#etsug_syslin_sect045.htm
6  *----- Zellner's Seemingly Unrelated Technique -----*
7  | A. Zellner, "An Efficient Method of Estimating Seemingly Unrelated Regressions
8  |           and Tests for Aggregation Bias," JASA 57(1962) pp.348-364
9  | J.C.G. Boot, "Investment Demand: an Empirical Contribution to the Aggregation
10 |           Problem," IER 1(1960) pp.3-30.
11 | Y. Grunfeld, "The Determinants of Corporate Investment," Unpublished thesis,
12 |           Chicago, 1958

```

```

13 | %± (1997), ±@ (), f9 .7
14 *-----*
15 DATA Grunfeld;
16     INFILE 'C:\Documents and Settings\user\My Documents\My SASFiles\9.1
17           \GibbsSamplingSAS\GrunfeldData.txt';
18     INPUT YEAR GEI GEF GEC WI WF WC;
19     LABEL GEI = 'Gross Investment, GE'
20           GEC = 'Capital Stock Lagged, GE'
21           GEF = 'Value of Outstanding Shares Lagged, GE'
22           WI = 'Gross Investment, WH'
23           WC = 'Capital Stock Lagged, WH'
24           WF = 'Value of Outstanding Shares Lagged, WH';
25 RUN;
26
27 PROC SYSLIN DATA=Grunfeld SUR;
28     GE:      MODEL GEI = GEF GEC;
29     WESTING: MODEL WI = WF WC;
30 RUN;
31 *-----*
    ;

```

첫 번째 DATA스텝을 수행하면, SAS데이터세트 GRUNFELD가 만들어진다. 이 6변량 시계열데이터 GRUNFELD는 Grunfeld [52]가 그의 박사논문에서 투자함수의 추정에 관한 이론을 예증하기 위해서 사용한 데이터세트이다. Grunfeld는 두 독립된 투자함수모형들을 추정할 때 각각의 투자모형을 개별적으로 추정하는 것보다는 두 모형들이 보유하고 있는 정보를 이용해서 두 모형들을 동시에 추정하는 것이 효율적임을 보였다. 이 이론을 예증하기 위해서 사용한 시계열데이터를 Grunfeld데이터라고 부른다. 이 6변량시계열데이터는 다음과 같은 변수들을 포함하고 있다.

- GEI : GE의 총투자액
- GEC : GE의 자본금
- GEF : GE의 자산가치
- WI : Westinghouse의 총투자액
- WC : Westinghouse의 자본금
- WF : Westinghouse의 자산가치

SAS/ETS의 PROC SYSLIN은 선형연립방정식 (linear simultaneous equations)을 추정하는 데 사용한다. ‘보기에는 무관한 회귀모형’을 추정하기 위해서는 PROC SYSLIN스태이트먼트의 옵션 SUR을 사용해야 한다. MODEL스태이트먼트에는 추정하고자하는 선형모형을 지정한다. 이 MODEL스태이트먼트의 사용양식은 PROC REG의 MODEL스태이트먼트의 사용양식과 동일하다. 이 SAS프로그램에서는 MODEL스태이트먼트들에 레이블들 (labels) GE와 WESTING을 추가했다.

이 SAS 프로그램을 수행하면, 우선 각 선형회귀모형 별로 최소제곱추정법을 적용한다. 그 결과물은 아래와 같다.

The SYSLIN Procedure

Ordinary Least Squares Estimation

Model GE  
 Dependent Variable GEI  
 Label Gross Investment, GE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	\Pr > F
Model	2	31632.03	15816.02	20.34	<.0001
Error	17	13216.59	777.4463		
Corrected Total	19	44848.62			

Root MSE	27.88272	R-Square	0.70531
Dependent Mean	102.29000	Adj R-Sq	0.67064
Coeff Var	27.25850		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	\Pr >  t	Variable Label
Intercept	1	-9.95631	31.37425	-0.32	0.7548	Intercept
GEF	1	0.026551	0.015566	1.71	0.1063	Value of Outstanding Shares Lagged, GE
GEC	1	0.151694	0.025704	5.90	<.0001	Capital Stock Lagged, GE

The SYSLIN Procedure

Ordinary Least Squares Estimation

Model WESTING  
 Dependent Variable WI  
 Label Gross Investment, WH

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	\Pr > F
Model	2	5165.553	2582.776	24.76	<.0001
Error	17	1773.234	104.3079		
Corrected Total	19	6938.787			

Root MSE 10.21312 R-Square 0.74445  
 Dependent Mean 42.89150 Adj R-Sq 0.71438  
 Coeff Var 23.81153

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	\Pr >  t	Variable Label
Intercept	1	-0.50939	8.015289	-0.06	0.9501	Intercept
WF	1	0.052894	0.015707	3.37	0.0037	Value of Outstanding Shares Lagged, WH
WC	1	0.092406	0.056099	1.65	0.1179	Capital Stock Lagged, WH

이 결과물에서 알 수 있듯이, 최소제곱추정법에 의한 추정식들은 다음과 같다.

$$GEI = -9.9563 + 0.0266GEF + 0.1517GEC + \hat{\epsilon}_G \quad (1)$$

$$WI = -0.5094 + 0.0529WF + 0.0924WC + \hat{\epsilon}_W \quad (2)$$

이 최소제곱추정식들에 의한 잔차들의 분산공분산행렬과 상관계수행렬은 아래와 같다. 여기서 두 최소제곱추정식들에 의한 잔차들의 상관계수가 0.7290이다. 따라서, 이 선형연립방정식을 추정하기 위해서는 ‘보기에는 무관한 회귀모형’을 적용하는 것이 타당하다는 것을 알 수 있다.

The SYSLIN Procedure

Seemingly Unrelated Regression Estimation

Cross Model Covariance

	GE	WESTING
GE	777.446	207.587
WESTING	207.587	104.308

Cross Model Correlation

	GE	WESTING
GE	1.00000	0.72896
WESTING	0.72896	1.00000

Cross Model Inverse Correlation

	GE	WESTING
GE	2.13397	-1.55559
WESTING	-1.55559	2.13397

Cross Model Inverse Covariance

	GE	WESTING
GE	0.002745	-.005463
WESTING	-.005463	0.020458

이 SAS 프로그램을 수행해서 얻은 ‘보기에 무관한 회귀모형’의 추정 결과는 아래와 같다.



이 결과물에서 알 수 있듯이, ‘보기에 무관한 회귀모형’의 추정식은 다음과 같다.

$$GEI = -27.7193 + 0.0383GEF + 0.1390GEC + \tilde{\epsilon}_G \quad (3)$$

$$WI = -1.2520 + 0.0576WF + 0.0640WC + \tilde{\epsilon}_W \quad (4)$$

■

Gibbs샘플러를 사용해서 ‘보기에는 무관한 회귀모형’을 추정하는 문제를 다시 살펴보자. 이에 대한 자세한 내용은 Rossi & Allenby & McCulloch [86, pp. 65-67]를 참조하라. 다음 선형회귀모형을 다시 살펴보자.

$$\mathbf{y}_i = X_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, (i = 1, 2, \dots, M) \quad (5.3.129)$$

여기서  $\mathbf{y}_i$ 는  $T \times 1$  행렬,  $X_i$ 는  $T \times K_i$  행렬,  $\boldsymbol{\beta}_i$ 는  $K_i \times 1$  행렬 그리고  $\boldsymbol{\epsilon}_i$ 는  $T \times 1$  행렬이다. 또한 오차항벡터  $\boldsymbol{\epsilon}_i$ 는 다음 성질들을 만족한다.

$$E(\boldsymbol{\epsilon}_i) = \mathbf{0}, \quad Var(\boldsymbol{\epsilon}_i) = \sigma_{ii}I, \quad Cov(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) = \sigma_{ij}I_T, \quad (1 \leq i < j \leq M) \quad (5.3.130)$$

다음 식들이 성립한다.

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad E(\boldsymbol{\epsilon}) = \mathbf{0}, \quad Var(\boldsymbol{\epsilon}) = \Sigma \otimes I_T \quad (5.3.131)$$

여기서  $\mathbf{y}$ 는  $TM \times 1$  행렬,  $X$ 는  $TM \times K$  행렬,  $\boldsymbol{\epsilon}$ 은  $TM \times 1$  행렬, 그리고  $\boldsymbol{\beta}$ 는  $K \times 1$  행렬이다. 오차벡터  $\boldsymbol{\epsilon}$ 이 다변량정규확률분포를 따른다면, 다음 식들이 성립한다.

$$\mathbf{y} \stackrel{d}{\sim} \mathcal{N}_{TM}(X\boldsymbol{\beta}, \Sigma \otimes I_T), \quad E(\boldsymbol{\epsilon}) = \mathbf{0}, \quad Var(\boldsymbol{\epsilon}) = \Sigma \otimes I_T \quad (5.3.132)$$

따라서, 확률벡터  $\mathbf{y}$ 의 결합확률밀도함수는 다음과 같다.

$$f(\mathbf{y} | \boldsymbol{\beta}, \Sigma) = [2\pi]^{-MT/2} |\Sigma|^{-T/2} \exp\left(-\frac{1}{2} [\mathbf{y} - X\boldsymbol{\beta}]^t [\Sigma^{-1} \otimes I_T] [\mathbf{y} - X\boldsymbol{\beta}]\right) \quad (5.3.133)$$

행렬  $\Sigma$ 를 다음과 같이 Cholesky분해하자.

$$\Sigma = U^t U \quad (5.3.134)$$

다음과 같은 벡터들과 행렬을 정의하자.

$$\tilde{\mathbf{y}} \doteq ((U^{-1})^t \otimes I_T) \mathbf{y}, \quad \tilde{X} \doteq ((U^{-1})^t \otimes I_T) X, \quad \tilde{\boldsymbol{\epsilon}} \doteq ((U^{-1})^t \otimes I_T) \boldsymbol{\epsilon} \quad (5.3.135)$$

다음 식들이 성립한다.

$$\begin{aligned} \text{Var}(\tilde{\boldsymbol{\epsilon}}) &= \text{Var}(((U^{-1})^t \otimes I_T) \boldsymbol{\epsilon}) \\ &= ((U^{-1})^t \otimes I_T) (\Sigma \otimes I_T) ((U^{-1})^t \otimes I_T)^t \\ &= ((U^{-1})^t \otimes I_T) (\Sigma \otimes I_T) (U^{-1} \otimes I_T) \\ &= ((U^{-1})^t \Sigma (U^{-1})) \otimes (I_T I_T I_T) = I_M \otimes I_T \end{aligned} \quad (5.3.136)$$

식 (5.3.131), 식 (5.3.135), 그리고 식 (5.3.136)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\tilde{\mathbf{y}} = \tilde{X} \boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}}, \quad E(\tilde{\boldsymbol{\epsilon}}) = \mathbf{0}, \quad \text{Var}(\tilde{\boldsymbol{\epsilon}}) = I_{MT} \quad (5.3.137)$$

식 (5.3.137)의 모형은 Gauss-Markov조건을 만족하는 표준선형회귀모형이다. 오차벡터  $\tilde{\boldsymbol{\epsilon}}$ 가 다변량정규확률분포를 따르면, 다음 식이 성립한다.

$$\tilde{\mathbf{y}} \stackrel{d}{\sim} \mathcal{N}_{TM}(\tilde{X} \boldsymbol{\beta}, I_{TM}) \quad (5.3.138)$$

다음과 같이 정밀도행렬을 정의하자.

$$\Psi \doteq \Sigma^{-1} \quad (5.3.139)$$

모수  $(\beta, \Psi)$  의 사전확률밀도함수가 다음과 같다고 하자.

$$\pi(\beta, \Psi) = \pi(\beta)\pi(\Psi) \tag{5.3.140}$$

$$\beta \stackrel{d}{\sim} \mathcal{N}(\beta_a, A_a^{-1}) \tag{5.3.141}$$

$$\Psi \stackrel{d}{\sim} W(R_a, n_a) \tag{5.3.142}$$

Koop [61, pp. 104-141]에서 알 수 있듯이, 사후확률분포는 다음과 같다.

$$\beta | (\Psi, \mathbf{y}, X) \stackrel{d}{\sim} \mathcal{N}(\beta_z, A_z^{-1}) \tag{5.3.143}$$

$$\Psi | (\beta, \mathbf{y}, X) \stackrel{d}{\sim} W(R_z, n_z) \tag{5.3.144}$$

여기서 각 모수는 다음과 같다.

$$A_z \doteq \tilde{X}^t \tilde{X} + A_a \tag{5.3.145}$$

$$\beta_z \doteq A_z^{-1} [\tilde{X}^t \tilde{\mathbf{y}} + A_a \beta_a] \tag{5.3.146}$$

$$n_z \doteq n_a + T \tag{5.3.147}$$

$$R_z \doteq R_a + S \tag{5.3.148}$$

특히  $S$ 는  $(i, j)$  원소가  $\epsilon_i^t \epsilon_j$ 인  $M \times M$  행렬이다. 다음 식들이 성립한다.

$$\begin{aligned} ((U^{-1})^t \otimes I_T)^t ((U^{-1})^t \otimes I_T) &= (U^{-1} \otimes I_T) ((U^{-1})^t \otimes I_T) \\ &= (U^{-1}(U^{-1})^t) \otimes (I_T I_T) = \Sigma^{-1} \otimes I_T = \Psi \otimes I_T \end{aligned} \tag{5.3.149}$$

식 (5.3.135)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\tilde{X}^t \tilde{X} = X^t (\Psi \otimes I_T) X \tag{5.3.150}$$

$$\tilde{X}^t \tilde{\mathbf{y}} = X^t (\Psi \otimes I_T) \mathbf{y} \tag{5.3.151}$$

즉, 다음 식들이 성립한다.

$$A_z \doteq X^t (\Psi \otimes I_T) X + A_a \quad (5.3.152)$$

$$\beta_z \doteq A_z^{-1} [X^t (\Psi \otimes I_T) \mathbf{y} + A_a \beta_a] \quad (5.3.153)$$

지금까지 내용을 바탕으로, ‘보기에는 무관한 회귀모형’의 모수들  $(\beta, \Psi)$ 의 사후확률분포에서 샘플링하는 Gibbs샘플러를 다음과 같이 요약할 수 있다.

(1단계) 초기모수들  $(\beta^{(0)}, \Psi^{(0)})$ 를 선택한다.

(2단계) 각  $j(= 1, 2, \dots)$ 에 대해서 주어진  $\mathbf{y}$ 에 대한 조건부샘플  $(\beta^{(j)}, \Psi^{(j)})$ 를 발생시킨다.

우선, 식 (5.3.143)의 사후조건부확률분포를 사용해서  $\beta^{(j)} | (\Psi^{(j-1)}, \mathbf{y}, X)$ 를 발생시킨다.

다. 식 (5.3.144)의 사후조건부확률분포를 사용해서  $\Psi^{(j)} | (\beta^{(j)}, \mathbf{y}, X)$ 를 발생시킨다.

(3단계) 충분히 큰  $b$ 에 대해서  $(\beta^{(j)}, \Psi^{(j)})$ ,  $(j = b + 1, b + 2, \dots)$ 를 기록한다.

**예제 5.3.12** Gibbs샘플러를 사용해서 ‘보기에는 무관한 회귀모형’을 추정하는 예로서, 다음 MATLAB프로그램 Gibbs\_SUR\_Grunfeld\_RAM.m을 실행해 보자.

```

1 % -----
2 % Filename: GibbsSamplerSUR102.m
3 % Gibbs Sampler for SUR
4 % Dataset: Grunfeld (1958)
5 % (Ref)  (1997),  (), f9 .7
6 % (Ref) Rossi, Allenby & McCulloch (2005), pp. 65-67
7 % -----
8 clear all, close all, clc, format short
9 XY = [ 1935      33.1      1170.6      97.8      12.93      191.5      1.8
10        1936      45.0      2015.8      104.4      25.90      516.0      .8
11        1937      77.2      2803.3      118.0      35.05      729.0      7.4
12        1938      44.6      2039.7      156.2      22.89      560.4      18.1
13        1939      48.1      2256.2      172.6      18.84      519.9      23.5
14        1940      74.4      2132.2      186.6      28.57      628.5      26.5
15        1941     113.0      1834.1      220.9      48.51      537.1      36.2
16        1942      91.9      1588.0      287.8      43.34      561.2      60.8
17        1943      61.3      1749.4      319.9      37.02      617.2      84.4
18        1944      56.8      1687.2      321.3      37.81      626.7      91.2
19        1945      93.6      2007.7      319.6      39.27      737.2      92.4
20        1946     159.9      2208.3      346.0      53.46      760.5      86.0
21        1947     147.2      1656.7      456.4      55.56      581.4     111.1
22        1948     146.3      1604.4      543.4      49.56      662.3     130.6
23        1949      98.3      1431.8      618.3      32.04      583.8     141.8
24        1950      93.5      1610.5      647.4      32.24      635.2     136.7
25        1951     135.2      1819.4      671.3      54.38      723.8     129.7
26        1952     157.3      2079.7      726.1      71.78      864.1     145.5
27        1953     179.5      2371.6      800.3      90.08     1193.5     174.8
28        1954     189.6      2759.9      888.9      68.60     1188.9     213.5 ];

```

```

29 Year = XY(:,1); GEI = XY(:,2); GEF = XY(:,3); GEC= XY(:,4);
30     WI = XY(:,5); WF = XY(:,6); WC = XY(:,7);
31 M = 2 % Number of linear models
32 T = length(Year) % Number of data points
33 k(1) = 3; k(2) = 3; % Number of regressors
34 K = sum(k) % Total Number of regressors
35 %% OLS
36 y1 = GEI; X1 = [ ones(T,1) GEF GEC ];
37 y2 = WI; X2 = [ ones(T,1) WF WC ];
38 [ beta1 bint1 Res1 ResInt1 stats1 ]= regress(y1,X1);
39 beta1, stats1
40 [ beta2 bint2 Res2 ResInt2 stats2 ]= regress(y2,X2);
41 beta2, stats2
42 Resid = [ Res1 Res2 ];
43 CovResid = cov(Resid)
44 PsiResid = inv(CovResid)
45 CorrResid = CovResid(1,2)/sqrt(CovResid(1,1)*CovResid(2,2))
46 betaOLS1 = [ beta1 ; beta2 ]
47 %%% SUR Model Estimation by Gibbs Sampler
48 % Make a matrix-form regression
49 rng('default')
50 TM = T*M
51 y(:,1) = y1;
52 y(:,2) = y2;
53 X(:, :,1) = X1;
54 X(:, :,2) = X2;
55 yy = y(:,1);
56 XX = X(:, :,1);
57 for ii=2:M
58     yy = [yy ; y(:,ii)];
59     XX = [XX zeros((ii-1)*T,k(ii)) ; zeros(T,sum(k(1:ii-1))) X(:, :,ii)];
60 end
61 % OLS strikes back!
62 U = chol(inv(eye(M)));
63 Tilda = kron(inv(U)',eye(T));
64 tilyy = Tilda*yy;
65 tilXX = Tilda*XX;
66 betaOLS2 = (tilXX'*tilXX)^-1*tilXX'*tilyy;
67 diffbeta = betaOLS1-betaOLS2
68 % Run the Gibbs sampler
69 iter = 1200;
70 burn = 1000;
71 kmax = iter;
72 % Initial Parameters
73 bbeta_a = [ -28 0.04 0.14 -1.25 0.06 0.06 ]';
74 A_a = 10^-4*eye(K);
75 n_a = 0;
76 R_a = eye(M);
77 Psi_a = PsiResid; % = inv(Sigma_a)
78 PsiAvg = zeros(M,M);
79 for kk = 1:iter
80     U = chol(inv(Psi_a));
81     Tilda = kron(inv(U)',eye(T));
82     tilyy = Tilda*yy;
83     tilXX = Tilda*XX;
84     A_z = tilXX'*tilXX + A_a;
85     bbeta_z = inv(A_z)*(tilXX'*tilyy + A_a*bbeta_a);
86     bbetaNew = mvnrnd(bbeta_z,inv(A_z))';
87     epsilon(:,1) = y(:,1) - X(:, :,1)*bbetaNew(1:k(1));
88     E = epsilon(:,1);
89     for mm=2:M

```

```

90     dumbeta = (sum(k(1:mm-1))+1:sum(k(1:mm)))';
91     epsilon(:,mm) = y(:,mm) - X(:, :, mm)*bbetaNew(dumbeta);
92     E = [ E epsilon(:,mm) ];
93     end
94     S = E'*E;
95     R_z = R_a + S;
96     n_z = n_a + T;
97     PsiNew = wishrnd(inv(R_z),n_z);
98     bbetaM(:,kk) = bbetaNew;
99     if kk > burn
100         PsiAvg = PsiAvg + PsiNew;
101         error(kk) = abs(max(bbetaM(:,kk)-bbetaM(:,kk-1)));
102         if abs(error(kk)) < 10^-8
103             kmax = kk, break
104         end
105     end
106     % For the next step
107     A_a = A_z; n_a = n_z; R_a = R_z; Psi_a = PsiNew;
108 end
109 bbetaGibb = mean(bbetaM(:,burn+1:kmax),2);
110 betaG1 = bbetaGibb(1:k(1))
111 betaG2 = bbetaGibb(k(1)+1:k(1)+k(2))
112 invPsiAvg = inv(PsiAvg/(kmax-burn))
113 CorrPsiAvg= invPsiAvg(1,2)/sqrt(invPsiAvg(1,1)*invPsiAvg(2,2))
114 ResG1 = y1 - X1*betaG1;
115 ResG2 = y2 - X2*betaG2;
116 ResG = [ ResG1 ResG2 ];
117 CovResG = cov(ResG)
118 CorrResG = CovResG(1,2)/sqrt(CovResG(1,1)*CovResG(2,2))
119 % Scatter plot of residuals
120 plot(Res1,CovResid(1,2)/CovResid(1,1)*Res1,'g-', ...
121      ResG1,CovResG(1,2)/CovResG(1,1)*ResG1,'k:','linewidth',1.5)
122 set(gca,'fontsize',11,'fontweigh','bold')
123 hold on
124 plot(Res1,Res2,'bd',ResG1,ResG2,'r*','linewidth',1)
125 legend('\bf OLS Residual','\bf Gibbs Residual', ...
126        '\bf OLS Residual','\bf Gibbs Residual','location','SE')
127 xlabel('Res_{GE}'),ylabel('Res_W')
128 axis square
129 hold off
130 saveas(gcf,'GibbsSamplerSUR102','epsc')
131 % End of Program
132 % -----

```

이 MATLAB 프로그램 GibbsSamplerSUR102.m에서는 데이터세트 GrunfeldData를 불러들인 후, Gibbs샘플러를 사용해서 ‘보기에는 무관한 회귀모형’을 추정한다.

최소제곱추정법을 사용해서 추정한 회귀식들은 다음과 같다.

$$GEI = -9.9563 + 0.0266GEF + 0.1517GEC + \hat{\epsilon}_G \quad (1)$$

$$WI = -0.5094 + 0.0529WF + 0.0924WC + \hat{\epsilon}_W \quad (2)$$

또한, 잔차들의 분산공분산행렬  $\hat{\Sigma}$ 은 다음과 같고, 이들의 상관계수는 0.7290이다.

$$\hat{\Sigma} = \begin{bmatrix} 695.6099 & 185.7359 \\ 185.7359 & 93.3281 \end{bmatrix} \quad (3)$$

따라서, 이 데이터세트에는 독립적인 두 선형회귀모형들보다는 ‘보기에는 무관한 회귀모형’을 적용하는 것이 더 타당할 것이다.

Gibbs샘플러를 사용해서 ‘보기에는 무관한 회귀모형’을 추정하기 위해서, 초기값들을 다음과 같이 선택하자.

$$\beta_a = \begin{bmatrix} -28 & 0.04 & 0.14 & -1.25 & 0.06 & 0.06 \end{bmatrix}^t \quad (4)$$

$$A_a = 10^{-4}I_6, \quad n_a = 0, \quad R_a = I_2 \quad (5)$$

$$\Psi_a = \begin{bmatrix} 695.6099 & 185.7359 \\ 185.7359 & 93.3281 \end{bmatrix}^{-1} \quad (6)$$

이 MATLAB프로그램 GibbsSamplerSUR102.m을 수행해서 얻은 ‘보기에 무관한 회귀모형’의 추정식들은 다음과 같다.

$$GEI = -28.0686 + 0.0400GEF + 0.1400GEC + \tilde{\epsilon}_G \quad (7)$$

$$WI = -1.2621 + 0.0600WF + 0.0601WC + \tilde{\epsilon}_W \quad (8)$$

이 추정식들은 식 (1)과 식 (2)의 최소제곱추정식들과 크게 달라 보인다. 이 ‘보기에는 무관한 회귀모형’에 의한 잔차들의 분산공분산행렬  $\tilde{\Sigma}$ 은 다음과 같고, 이들의 상관계수는 0.7680이다.

$$\tilde{\Sigma} = \begin{bmatrix} 731.3963 & 202.6066 \\ 202.6066 & 95.1544 \end{bmatrix} \quad (9)$$

이 분산공분산행렬  $\tilde{\Sigma}$ 는 식 (3)에 기술한 최소제곱추정법에 의한 분산공분산행렬  $\hat{\Sigma}$ 와 크게 달라 보이지 않는다.

독자들이 유의할 점은 초기값에 따라 Gibbs샘플러를 사용한 ‘보기에는 무관한 회귀모형’의 추정식들이 크게 달라진다는 점이다. 본저자는 출력된 결과를 보고, 이 MATLAB프로그램이 잘못되지 않았나 하는 의구심에 며칠에 걸쳐 디버깅해보고 전혀 다른 방법으로 프로그램을 짜보았다. 그러나, 다양한 접근을 해도 계속 엉뚱한 결과를 출력하는 것 같았다. 그래서,

어쩌면 Gibbs샘플러 자체가 다양한 점들로 수렴할지도 모른다는 생각에, 여러 초기값들을 대입해서 실험한 결과 이 Gibbs샘플러는 초기값에 크게 의존한다는 결론을 얻었다. 행렬  $\tilde{X}^t \tilde{X}$ 의 대각원소들이 아주 커서,  $\beta$ 의 분산공분산행렬인  $A_z^{-1}$ 의 대각원소들이 0에 가깝기 때문인 것 같다.

적어도 이 예제에서는 최소제곱추정법과 ‘보기에는 무관한 회귀모형’에 적용한 일반화최소제곱추정법에 의해 추정된 회귀식들이 달라 보이지만, 잔차들의 구조에는 큰 변함이 없는 것 같다. 이를 확인해보기 위해서, 두 모형들에 의한 잔차들과 이 잔차들의 선형회귀선들을 그린 그래프가 그림 5.3.16에 그려져 있다. 그림 5.3.16에서 확인할 수 있듯이, 두 모형들에는 큰 차이가 없는 듯하다. 즉, 이 데이터세트 GrunfeldData에서 오차제곱합은 다양한 모수들에서 극소값들을 갖는 것 같다. ■

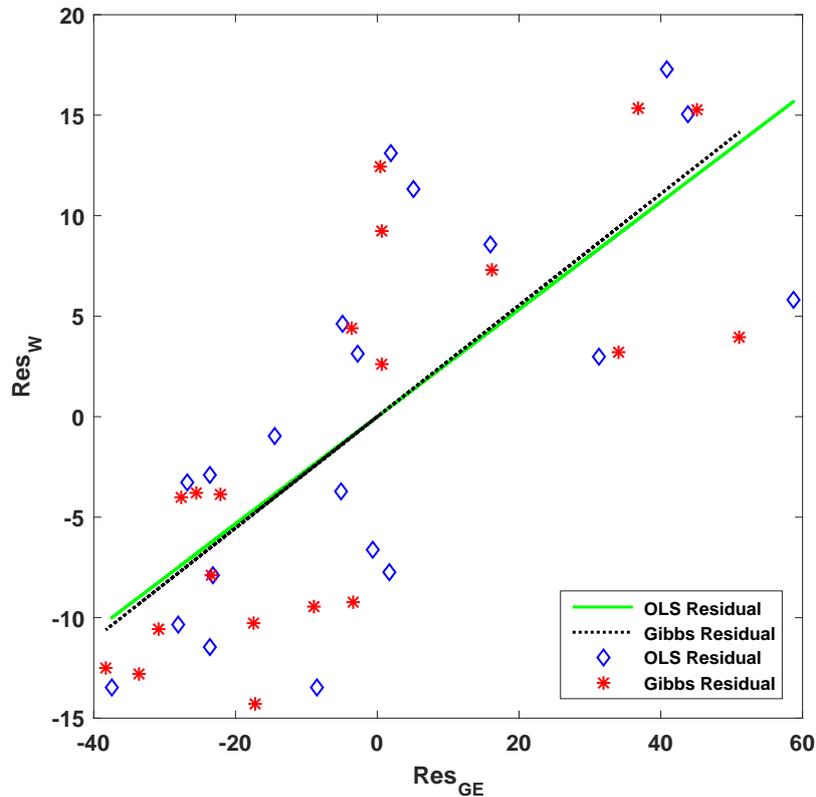


그림 5.3.16. 최소제곱추정식과 보기에는 무관한 회귀식

‘보기에는 무관한 회귀모형’을 추정하기 위해서, Koop가 제시한 MATLAB프로그램을 사용하고자 한다. 다음 웹사이트에서 이 프로그램을 다운받을 수 있다.

<http://www.wiley.com/legacy/wileychi/koopbayesian/supp/chapter6d.m>

동일한 분석을 시도한 다른 MATLAB프로그램을 다음 웹사이트에서 다운받을 수 있다.

[http://www.tcd.ie/Economics/staff/frainj/main/2005\\\_06\\\_MSc/Introduction\%20to\%20MATLAB/session\%208/Baseball.m](http://www.tcd.ie/Economics/staff/frainj/main/2005\_06\_MSc/Introduction\%20to\%20MATLAB/session\%208/Baseball.m)

Koop [61, pp. 137-143]에서는 결합일반화최소제곱추정법을 설명하기 위해서, ‘보기에는 무관한 회귀모형’을 다음과 같이 표현한다.

$$y_{mj} = \mathbf{x}_{mj}^t \boldsymbol{\beta}_m + \epsilon_{mj}, \quad (m = 1, 2, \dots, M, j = 1, 2, \dots, T) \quad (5.3.154)$$

여기서  $y_{mj}$ 는 제  $m$  번째 선형모형의 종속변수의 제  $j$  번째 관찰값이고,  $K_m \times 1$  행렬  $\mathbf{x}_{mj}$ 는 제  $m$  번째 선형모형의 설명변수들의 제  $j$  번째 관찰벡터이고,  $K_m \times 1$  행렬  $\boldsymbol{\beta}_m$ 은 제  $m$  번째 선형모형의 회귀계수벡터이다. 앞에서와 마찬가지로, 상수  $K$ 를  $\sum_{m=1}^M K_m$ 이라 하고, 다음과 같은 벡터들과 행렬을 정의하자.

$$\mathbf{y}_j = \begin{bmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{Mj} \end{bmatrix}, \quad \boldsymbol{\epsilon}_j = \begin{bmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ \vdots \\ \epsilon_{Mj} \end{bmatrix}, \quad X_j = \begin{bmatrix} \mathbf{x}_{1j}^t & O & O & \cdots & O \\ O & \mathbf{x}_{2j}^t & O & \cdots & O \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ O & O & O & \cdots & \mathbf{x}_{Mj}^t \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{bmatrix} \quad (5.3.155)$$

여기서  $\mathbf{y}_j$ 와  $\boldsymbol{\epsilon}_j$ 는  $M \times 1$  행렬들,  $X_j$ 는  $M \times K$  행렬, 그리고  $\boldsymbol{\beta}$ 는  $K \times 1$  행렬이다. 식 (5.3.154)를 다음과 같이 쓸 수 있다.

$$\mathbf{y}_j = X_j \boldsymbol{\beta} + \boldsymbol{\epsilon}_j, \quad (j = 1, 2, \dots, T) \quad (5.3.156)$$

오차항들  $\{\epsilon_j | j = 1, 2, \dots, T\}$ 는 서로 독립이며, 다음 식을 만족한다고 하자.

$$\boldsymbol{\epsilon}_j \stackrel{d}{\sim} \mathcal{N}(\mathbf{0}, H^{-1}), \quad (j = 1, 2, \dots, T) \quad (5.3.157)$$

또한, 다음과 같은 벡터들과 행렬을 정의하자.

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_T \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_T \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_T \end{bmatrix} \quad (5.3.158)$$

여기서  $\mathbf{y}$ 와  $\boldsymbol{\epsilon}$ 는  $TM \times 1$  행렬, 그리고  $X$ 는  $TM \times K$  행렬이다. 따라서, 식 (5.3.156)을 다음과 같이 나타낼 수 있다.

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (5.3.159)$$

$$\boldsymbol{\epsilon} \stackrel{d}{\sim} \mathcal{N}(\mathbf{0}, \Omega) \quad (5.3.160)$$

여기서  $\Omega$ 는 다음과 같다.

$$\Omega \doteq \begin{bmatrix} H^{-1} & O & \dots & O & O \\ O & H^{-1} & \dots & O & O \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & \dots & H^{-1} & O \\ O & O & \dots & O & H^{-1} \end{bmatrix} = H^{-1} \otimes I_T \quad (5.3.161)$$

이 선형회귀모형의 일반화최소제곱추정량은 다음과 같다.

$$\tilde{\boldsymbol{\beta}} = (X^t \Omega^{-1} X)^{-1} X^t \Omega^{-1} \mathbf{y} = \left[ \sum_{j=1}^T X_j^t H X_j \right]^{-1} \left[ \sum_{j=1}^T X_j^t H \mathbf{y}_j \right] \quad (5.3.162)$$

모수  $(\boldsymbol{\beta}, H)$ 의 사전확률밀도함수가 다음과 같다고 하자.

$$\pi(\boldsymbol{\beta}, H) = \pi(\boldsymbol{\beta})\pi(H) \quad (5.3.163)$$

$$\boldsymbol{\beta} \stackrel{d}{\sim} \mathcal{N}_K(\underline{\boldsymbol{\beta}}, \underline{V}) \quad (5.3.164)$$

$$H \stackrel{d}{\sim} \mathcal{W}_K(\underline{H}, \underline{\nu}) \quad (5.3.165)$$

Koop [61, pp. 104-141]에서 알 수 있듯이, 사후확률분포는 다음과 같다.

$$\boldsymbol{\beta} | (H, \mathbf{y}, X) \stackrel{d}{\sim} \mathcal{N}_K(\bar{\boldsymbol{\beta}}, \bar{V}) \quad (5.3.166)$$

$$H | (\boldsymbol{\beta}, \mathbf{y}, X) \stackrel{d}{\sim} \mathcal{W}_K(\bar{H}, \bar{\nu}) \quad (5.3.167)$$

여기서 모수들은 다음과 같다.

$$\bar{V} \doteq \left[ \underline{V}^{-1} + \sum_{j=1}^T X_j^t H X_j \right]^{-1} \quad (5.3.168)$$

$$\bar{\beta} \doteq \bar{V} \left[ \underline{V}^{-1} \underline{\beta} + \sum_{j=1}^T X_j^t H \mathbf{y}_j \right] \quad (5.3.169)$$

$$\bar{\nu} \doteq \underline{\nu} + T \quad (5.3.170)$$

$$\bar{H} \doteq \left( \underline{H}^{-1} + \sum_{j=1}^T [\mathbf{y}_j - X_j \beta] [\mathbf{y}_j - X_j \beta]^t \right)^{-1} \quad (5.3.171)$$

지금까지 내용을 바탕으로, ‘보기에는 무관한 회귀모형’의 모수들  $(\beta, H)$ 의 사후확률분포에서 샘플링하는 Koop의 Gibbs샘플러를 다음과 같이 요약할 수 있다.

**(1단계)** 초기모수들  $(\beta^{(0)}, H^{(0)})$ 를 선택한다.

**(2단계)** 각  $j (= 1, 2, \dots)$ 에 대해서 주어진  $\mathbf{y}$ 에 대한 조건부샘플  $(\beta^{(j)}, H^{(j)})$ 를 발생시킨다.

우선, 식 (5.3.166)의 사후조건부확률분포를 사용해서  $\beta^{(j)} | (H^{(j-1)}, \mathbf{y}, X)$ 를 발생시킨다. 식 (5.3.167)의 사후조건부확률분포를 사용해서  $H^{(j)} | (\beta^{(j)}, \mathbf{y}, X)$ 를 발생시킨다.

**(3단계)** 충분히 큰  $b$ 에 대해서  $(\beta^{(j)}, H^{(j)})$ ,  $(j = b + 1, b + 2, \dots)$ 를 기록한다.

**예제 5.3.13** Gibbs샘플러를 사용해서 ‘보기에는 무관한 회귀모형’을 추정하는 다른 예를 살펴보자. 이 예제에 대한 자세한 내용은 Koop [61, pp. 137-143]와 Koop & Poirier & Tobias [62, pp. 314-317]를 참조하라. 이 예제에서 사용할 데이터는 미국 프로야구팀들인 Yankees와 Red Sox의 승률에 관한 것이다. 데이터세트들 yankees.txt와 redsox.txt에는 1903년부터 1999년까지 연도를 나타내는 Year, 각 팀의 승률(team winning percentage)을 나타내는 PCT, 득점(runs)을 나타내는 R, 타율(batting average)을 나타내는 AVG, 출루율(on-base percentage)을 나타내는 OBP, 장타율(slugging percentage)를 나타내는 장타율, 도루(stolen bases)를 나타내는 SB, 그리고 방어율(earned run average)을 나타내는 ERA가 수록되어 있다.

우리가 다루고자 하는 선형모형의 종속변수는 PCT이고 설명변수들은 OBP, SLG, 그리고 ERA이다. 이들로 구성된 ‘보기에는 무관한 회귀모형’을 추정하기 위해서, 앞에서 설명한 Koop가 제시한 MATLAB프로그램을 사용하고자 한다. 이 MATLAB프로그램의 이름을

GibbsSamplerSUR103.m으로 바꾸었다. 또한, 데이터셋 yankees.txt와 redsox.txt를 이 MATLAB프로그램 안으로 불러 들였다.

위에서 기술한 Gibbs샘플러를 사용해서 ‘보기에는 무관한 회귀모형’을 Yankee팀과 Red Sox팀의 승률에 관한 데이터에 적합시키고자 하는 모형은 다음과 같다.

$$YPCT = \beta_1 + \beta_2YOBP + \beta_3YSLG + \beta_4YERA + \epsilon_Y \quad (1)$$

$$RPCT = \beta_5 + \beta_6ROBP + \beta_7RSLG + \beta_8RERA + \epsilon_R \quad (2)$$

초기값들을 다음과 같이 선택하자.

$$\underline{\beta} = \mathbf{0}, \quad \underline{V} = 0.1^2 I_8 \quad (3)$$

$$\underline{\nu} = 0, \quad \underline{H} = I_2 \quad (4)$$

반복횟수를 11,000으로 하고 번인개수를 1,000으로 하는 Gibbs샘플러를 적용해서 구한 ‘보기에 무관한 회귀모형’의 추정식들은 다음과 같다.

$$YPCT = 0.3257 + 0.2991YOBP + 0.5356YSLG - 0.0211YERA + \hat{\epsilon}_Y \quad (5)$$

$$RPCT = 0.4316 + 0.3309ROBP + 0.4817RSLG - 0.0576RERA + \hat{\epsilon}_R \quad (6)$$

또한, 이 ‘보기에는 무관한 회귀모형’에 의한 잔차들의 분산공분산행렬  $\tilde{\Sigma}$ 는 다음과 같고, 이들의 상관계수는  $-0.0010$ 이다.

$$\tilde{\Sigma} = \begin{bmatrix} 277.3609 & -0.0303 \\ -0.0303 & 280.2692 \end{bmatrix} \quad (7)$$

식 (7)에서 알 수 있듯이, 이 데이터세트에 ‘보기에는 무관한 회귀모형’을 적용하면, 최소제곱 추정법을 적용할 때 발생하는 잔차들의 상관성이 제거된다. ■

다음과 같은 ‘보기에는 무관한 회귀모형’을 다시 살펴보자.

$$\mathbf{y}_i = X_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \quad (i = 1, 2, \dots, M) \quad (5.3.172)$$

여기서  $\mathbf{y}_i$ 는  $T \times 1$  행렬,  $X_i$ 는  $T \times K_i$  행렬,  $\boldsymbol{\beta}_i$ 는  $K_i \times 1$  행렬 그리고  $\boldsymbol{\epsilon}_i$ 는  $T \times 1$  행렬이다.

또한 오차항벡터  $\boldsymbol{\epsilon}_i$ 는 다음 성질들을 만족한다.

$$E(\boldsymbol{\epsilon}_i) = \mathbf{0}, \quad Var(\boldsymbol{\epsilon}_i) = \sigma_{ii}I_T, \quad Cov(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) = \sigma_{ij}I_T, \quad (1 \leq i < j \leq M) \quad (5.3.173)$$

따라서, 확률벡터  $\mathbf{y}$ 의 결합확률밀도함수는 다음과 같다.

$$f(\mathbf{y} | \boldsymbol{\beta}, \Sigma) = [2\pi]^{-MT/2} |\Sigma|^{-T/2} \exp\left(-\frac{1}{2} [\mathbf{y} - X\boldsymbol{\beta}]^t [\Sigma^{-1} \otimes I_T] [\mathbf{y} - X\boldsymbol{\beta}]\right) \quad (5.3.174)$$

트레이스의 성질을 이용해서 다음 식을 유도할 수 있다.

$$[\mathbf{y} - X\boldsymbol{\beta}]^t [\Sigma^{-1} \otimes I_T] [\mathbf{y} - X\boldsymbol{\beta}] = \text{Tr}([Y - X^*B]^t [Y - X^*B] \Sigma^{-1}) \quad (5.3.175)$$

여기서  $K \doteq \sum_{i=1}^M K_i$ 이고,  $T \times M$  행렬  $Y$ ,  $T \times K$  행렬  $X^*$ , 그리고  $K \times M$  행렬  $B$ 는 각각 다음과 같다.

$$Y \doteq \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_M \end{bmatrix} \quad (5.3.176)$$

$$X^* \doteq \begin{bmatrix} X_1 & X_2 & \cdots & X_M \end{bmatrix} \quad (5.3.177)$$

$$B = \begin{bmatrix} \beta_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \beta_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \beta_M \end{bmatrix} \quad (5.3.178)$$

식 (5.3.174)와 식 (5.3.175)에서 알 수 있듯이, 확률벡터  $\mathbf{y}$ 의 결합확률밀도함수를 다음과 같이 쓸 수 있다.

$$f(\mathbf{y} | \boldsymbol{\beta}, \Sigma) = [2\pi]^{-MT/2} |\Sigma|^{-T/2} \exp\left(-\frac{1}{2} \text{Tr}(A\Sigma^{-1})\right) \quad (5.3.179)$$

여기서  $M \times M$  행렬  $A$ 는 다음과 같다.

$$A = [Y - X^*B]^t [Y - X^*B] \quad (5.3.180)$$

즉, 행렬  $A$ 의  $(i, j)$  원소  $a_{i,j}$ 는 다음과 같다.

$$a_{i,j} \doteq [\mathbf{y}_i - X_i \boldsymbol{\beta}_i]^t [\mathbf{y}_j - X_j \boldsymbol{\beta}_j] \quad (5.3.181)$$

다음과 같이 정밀도행렬을 정의하자.

$$\Psi \doteq \Sigma^{-1} \quad (5.3.182)$$

모수들을  $(R, n)$ 으로 하는 Wishart 확률변수  $W$ 의 확률밀도함수는 다음과 같다.

$$f(W | R, n) \propto |W|^{[n-p-1]/2} \exp\left(-\frac{1}{2} \text{Tr}(R^{-1}W)\right) \quad (5.3.183)$$

여기서  $W$ 는  $p \times p$  차원 행렬로서 거의 모든 점에서 양정치행렬이다.

Zellner [99, 제8장]에서 사용한 모수  $(\boldsymbol{\beta}, \Psi)$ 의 사전확률밀도함수는 다음과 같다.

$$\pi(\boldsymbol{\beta}, \Psi) = \pi(\boldsymbol{\beta})\pi(\Psi) \propto |\Psi|^{[M+1]/2} \quad (5.3.184)$$

여기서  $\boldsymbol{\beta}$ 와  $\Psi$ 는 서로 독립이다. 식 (5.3.174)에서 알 수 있듯이, 관찰벡터  $\mathbf{y}$ 의 우도함수는 다음과 같다.

$$f(\mathbf{y} | \boldsymbol{\beta}, \Psi) \propto |\Psi|^{T/2} \exp\left(-\frac{1}{2} [\mathbf{y} - X\boldsymbol{\beta}]^t [\Psi \otimes I_T] [\mathbf{y} - X\boldsymbol{\beta}]\right) \quad (5.3.185)$$

식 (5.3.184)와 식 (5.3.185)에서 알 수 있듯이, 모수  $(\boldsymbol{\beta}, \Psi)$ 의 사후확률밀도함수는 다음과 같다.

$$\pi(\boldsymbol{\beta}, \Psi | \mathbf{y}) \propto |\Psi|^{[T+M+1]/2} \exp\left(-\frac{1}{2} [\mathbf{y} - X\boldsymbol{\beta}]^t [\Psi \otimes I_T] [\mathbf{y} - X\boldsymbol{\beta}]\right) \quad (5.3.186)$$

다음 식이 성립한다.

$$\begin{aligned} & [\mathbf{y} - X\boldsymbol{\beta}]^t [\Psi \otimes I_T] [\mathbf{y} - X\boldsymbol{\beta}] \\ &= [\mathbf{y} - X\tilde{\boldsymbol{\beta}}]^t [\Psi \otimes I_T] [\mathbf{y} - X\tilde{\boldsymbol{\beta}}] + [\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}]^t X^t [\Psi \otimes I_T] X [\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}] \end{aligned} \quad (5.3.187)$$

여기서  $\tilde{\boldsymbol{\beta}} = (X^t [\Psi \otimes I_T] X)^{-1} X^t [\Psi \otimes I_T] \mathbf{y}$ 는 예제 5.3.13에 정의된 일반화최소제곱추정량이다. 식 (5.3.186)과 식 (5.3.187)에서 알 수 있듯이, 모수  $\boldsymbol{\beta}$ 의 사후조건부확률밀도함수는

다음과 같다.

$$\pi(\boldsymbol{\beta} | \Psi, \mathbf{y}) \propto \exp\left(-\frac{1}{2} [\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}]^t X^t [\Psi \otimes I_T] X [\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}]\right) \quad (5.3.188)$$

식 (5.3.188) 에서 알 수 있듯이, 모수  $\boldsymbol{\beta}$  의 사후평균은 일반화최소제곱추정량과 같다. 즉, 다음 식이 성립한다.

$$E(\boldsymbol{\beta} | \Psi, \mathbf{y}) = (X^t [\Psi \otimes I_T] X)^{-1} X^t [\Psi \otimes I_T] \mathbf{y} = \tilde{\boldsymbol{\beta}} \quad (5.3.189)$$

또한, 사후분산공분산행렬은 다음과 같다.

$$Var(\boldsymbol{\beta} | \Psi, \mathbf{y}) = (X^t [\Psi \otimes I_T] X)^{-1} \quad (5.3.190)$$

식 (5.3.186) 에서 알 수 있듯이, 모수  $(\boldsymbol{\beta}, \Psi)$  의 사후확률밀도함수는 다음과 같다.

$$\pi(\boldsymbol{\beta}, \Psi | \mathbf{y}) \propto |\Psi|^{[T+M+1]/2} \exp\left(-\frac{1}{2} \text{Tr}(A\Psi)\right) \quad (5.3.191)$$

따라서, 모수  $\Psi$  의 사후조건부확률밀도함수는 다음과 같다.

$$\pi(\Psi | \boldsymbol{\beta}, \mathbf{y}) \propto |\Psi|^{[T+M+1]/2} \exp\left(-\frac{1}{2} \text{Tr}(A\Psi)\right) \quad (5.3.192)$$

즉, 조건부확률행렬  $\Psi | (\boldsymbol{\beta}, \mathbf{y})$  은 모수들을  $(A, T + 2M + 2)$  으로 하는 Wishart 확률분포를 따른다. 식 (5.3.188) 과 식 (5.3.192) 에서 알 수 있듯이, 다음 식들이 성립한다.

$$\boldsymbol{\beta} | (\Psi, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}\left(\tilde{\boldsymbol{\beta}}, (X^t [\Psi \otimes I_T] X)^{-1}\right) \quad (5.3.193)$$

$$\Psi | (\boldsymbol{\beta}, \mathbf{y}) \stackrel{d}{\sim} \mathcal{W}(A, T + 2M + 2) \quad (5.3.194)$$

지금까지 내용을 바탕으로, ‘보기에는 무관한 회귀모형’ 의 모수들  $(\boldsymbol{\beta}, \Psi)$  의 사후확률분포에서 샘플링하는 Gibbs샘플러를 다음과 같이 요약할 수 있다.

**(1단계)** 초기모수들  $(\boldsymbol{\beta}^{(0)}, \Psi^{(0)})$  를 선택한다.

**(2단계)** 각  $j(= 1, 2, \dots)$  에 대해서 주어진  $\mathbf{y}$  에 대한 조건부샘플  $(\boldsymbol{\beta}^{(j)}, \Psi^{(j)})$  를 발생시킨다.

우선, 식 (5.3.193) 의 사후조건부확률분포를 사용해서  $\boldsymbol{\beta}^{(j)} | (\Psi^{(j-1)}, \mathbf{y})$  를 발생시킨다.

식 (5.3.194)의 사후조건부확률분포를 사용해서  $\Psi^{(j)} | (\beta^{(j)}, \mathbf{y})$ 를 발생시킨다.

(3단계) 충분히 큰  $b$ 에 대해서  $(\beta^{(j)}, \Psi^{(j)})$ ,  $(j = b + 1, b + 2, \dots)$ 를 기록한다.

**예제 5.3.14** Gibbs샘플러를 사용해서 ‘보기에는 무관한 회귀모형’을 추정하는 다른 예제를 Koop & Poirier & Tobias [62, pp. 136-138]에서 찾을 수 있다. 이 예제를 살펴보기 위해서 다음 MATLAB프로그램 gibbs\_sur\_000\_Tobias.m을 실행하자.

```

1 % -----
2 % Filename: gibbs_sur_000_Tobias.m
3 % http://www2.econ.iastate.edu/faculty/tobias/chapter_11.html
4 % -----
5 clear, close all, clc, format long;
6 rng((1001^2+1)/2,'twister')
7 % randn('seed',sum(100*clock));
8 % rand('seed',sum(100*clock));
9 % This m-file runs a gibbs sampling routine on a two-equation SUR model.
10 % Set up design of experiment.
11 nobs = 2500;
12 beta11 = 1; beta12 = .3;
13 beta21 = 2; beta22 = -.7;
14 corr_true = -.45;
15 Sigma_true = [.5 corr_true*sqrt(.5)*sqrt(.3);
16               corr_true*sqrt(.5)*sqrt(.3) .3];
17 H_true = chol(Sigma_true);
18
19 %Generate the data;
20 x1 = zeros(nobs,1);
21 x2 = x1;
22 y1 = x1;
23 y2 = x1;
24
25 for i = 1:nobs;
26     error_vec = H_true'*randn(2,1);
27     x1(i,1) = randn(1,1);
28     x2(i,1) = randn(1,1);
29     y1(i,1) = beta11 + beta12*x1(i,1) + error_vec(1);
30     y2(i,1) = beta21 + beta22*x2(i,1) + error_vec(2);
31 end;
32
33 %Define some terms that will be useful later when running the Gibbs
34 %sampler.
35 bigy = [y1' y2'];
36 bigX1 = [ones(nobs,1) x1];
37 bigX2 = [ones(nobs,1) x2];
38 X1X1 = bigX1'*bigX1;    X1X2 = bigX1'*bigX2;
39 X2X1 = bigX2'*bigX1;    X2X2 = bigX2'*bigX2;
40 X1Y1 = bigX1'*y1;      X1Y2 = bigX1'*y2;
41 X2Y1 = bigX2'*y1;      X2Y2 = bigX2'*y2;
42 ytilde = zeros(2,nobs);
43 xtilde = zeros(2,4,nobs);
44 for i = 1:nobs;
45     ytilde(:,i) = [y1(i,1) y2(i,1)]';
46     xtilde(:, :, i) = [1 x1(i,1) 0 0;
47                       0 0 1 x2(i,1)];
48 end;

```

```

49 |
50 | %Define the prior hyperparameters;
51 | mu_beta = zeros(4,1);
52 | V_beta = 1000*eye(4);
53 | Omega = eye(2);
54 | nu = 4;
55 |
56 | %Run the Gibbs sampler
57 | iter = 1000;
58 | burn = 200;
59 | sig1_final = zeros(iter-burn,1);
60 | sig2_final = sig1_final;
61 | corr_final =sig1_final;
62 | betas_final = zeros(4,iter-burn);
63 |
64 | sig1 = 1; sig2 = 1; sig12 = 0;
65 | for i = 1:iter;
66 |     %Sample from Conditional for Beta
67 |     matrix1 = [sig1*X1X1 sig12*X1X2;
68 |               sig12*X2X1 sig2*X2X2];
69 |     matrix2 = [ sig1*X1Y1 + sig12*X1Y2 ;
70 |               sig2*X2Y2 + sig12*X2Y1];
71 |     D_beta = inv(matrix1 + inv(V_beta));
72 |     d_beta = matrix2 + inv(V_beta)*mu_beta;
73 |     H_beta = chol(D_beta);
74 |     betas = D_beta*d_beta + H_beta'*randn(4,1);
75 |
76 |     %Sample from Wishart conditional for Sigmainv
77 |     temp = zeros(2,2);
78 |     for j = 1:nobs;
79 |         temp_term = (ytilde(:,j) - xtilde(:, :,j)*betas)* ...
80 |                     (ytilde(:,j) - xtilde(:, :,j)*betas)';
81 |         temp = temp_term + temp;
82 |     end;
83 |
84 |     Sigmainv = wishrnd(inv(temp + inv(Omega)),nobs+nu); % Modified
85 |     % Sigmainv = wish_rnd(inv(temp + inv(Omega)),nobs+nu);
86 |     Sigma = inv(Sigmainv);
87 |     sig1 = Sigma(1,1);
88 |     sig2 = Sigma(2,2);
89 |     sig12 = Sigma(1,2);
90 |
91 |     s1 = Sigma(1,1);
92 |     s2 = Sigma(2,2);
93 |     s12 = Sigma(1,2);
94 |
95 |     if i > burn;
96 |         sig1_final(i-burn,1) = s1;
97 |         sig2_final(i-burn,1) = s2;
98 |         corr_final(i-burn,1) = s12/(sqrt(s1)*sqrt(s2));
99 |         betas_final(:,i-burn,1) = betas;
100 |     end;
101 | end;
102 |
103 | disp('Posterior means and true values - error variances and correlations');
104 | [mean(sig1_final) .5;
105 |  mean(sig2_final) .3;
106 |  mean(corr_final) corr_true]
107 |
108 | disp('Posterior means and true values - regression parameters');
109 | [mean(betas_final)' [beta11 beta12 beta21 beta22]']

```

```

110 | % End of program
111 | % -----

```

이 MATLAB 프로그램 gibbs\_sur\_000\_Tobias.m을 실행한 결과는 다음과 같다.

Posterior means and true values - error variances and correlations

ans =

```

0.501545175992469    0.500000000000000
0.308025717831232    0.300000000000000
-0.470341724035113  -0.450000000000000

```

Posterior means and true values - regression parameters

ans =

```

0.995948606541718    1.000000000000000
0.300567312745961    0.300000000000000
1.996147372726544    2.000000000000000
-0.699301188427463  -0.700000000000000

```

■

### 5.3.7 상태공간모형과 Gibbs 샘플러

베이저안추정법을 사용해서 상태공간모형을 추정하는 문제를 살펴보자. 여기서는 다음과 같이 관측방정식과 상태방정식에서 오차항의 분산이 일정한 모형을 살펴보자.

$$y_t = F_t^t \boldsymbol{\theta}_t + \epsilon_t, \quad (t = 1, 2, \dots, n) \quad (5.3.195)$$

$$\boldsymbol{\theta}_t = G_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad (t = 2, 3, \dots, n) \quad (5.3.196)$$

$$\boldsymbol{\theta}_1 \stackrel{d}{\sim} \mathcal{N}(\mathbf{a}_1, R_1) \quad (5.3.197)$$

여기서  $\{\epsilon_t\}$ 는 서로 독립이고 평균이 0이고 분산은 시간에 의존하지 않고 일정한  $\sigma^2$ 이며 정규 확률분포를 따르는 오차항들이고,  $\{\boldsymbol{\omega}_t\}$ 는 서로 독립이고 평균벡터가  $\mathbf{0}$ 이고 분산공분산행렬이 시간에 의존하지 않고 일정한  $\Sigma$ 인  $d$ 변량 정규확률분포를 따르는 쇠신오차벡터열이다. 또한,

시점  $t$ 에서  $\epsilon_t$ 와  $\omega_t$ 는 서로 독립이다. 다음과 같이 정밀도와 정밀도행렬을 정의하자.

$$\psi \doteq \frac{1}{\sigma^2}, \quad \Psi \doteq \Sigma^{-1} \tag{5.3.198}$$

모수들  $(\psi, \Psi)$ 의 사전확률분포가 다음과 같다고 하자.

$$\psi \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right), \quad \Psi \stackrel{d}{\sim} \mathcal{W}_d(n_0, \Psi_0) \tag{5.3.199}$$

시점  $t$ 까지 관측벡터를  $\mathbf{y}_t \doteq [y_1 \ y_2 \ \cdots \ y_t]^t$ 로 표기하면,  $\mathbf{y}_t$ 가 주어진 조건 하에서  $\boldsymbol{\theta}_t$ 의 사후확률분포가 다음과 같음을 알 수 있다.

$$\boldsymbol{\theta}_t | \mathbf{y}_t \stackrel{d}{\sim} \mathcal{N}(\mathbf{m}_t, C_t) \tag{5.3.200}$$

여기서 평균벡터  $\mathbf{m}_t$ 와 분산공분산행렬  $C_t$ 는 각각 다음과 같다.

$$\mathbf{m}_t = G_t \mathbf{m}_{t-1} + \frac{1}{Q_t} R_t \mathbf{F}_t^t [y_t - \eta_t] \tag{5.3.201}$$

$$C_t = R_t - \frac{1}{Q_t} R_t \mathbf{F}_t^t \mathbf{F}_t^t R_t \tag{5.3.202}$$

식 (5.3.200)에서 알 수 있듯이, 상태변수  $\boldsymbol{\theta}$ 의 사후조건부확률분포는 다변량정규확률분포이다. 식 (5.3.199)와 식 (5.3.200)에서 알 수 있듯이, 시점  $n$ 까지 관측벡터  $\mathbf{y}_n$ 과 모수벡터  $\boldsymbol{\theta} \doteq [\boldsymbol{\theta}_1^t, \boldsymbol{\theta}_2^t, \dots, \boldsymbol{\theta}_n^t]^t$ 가 주어졌을 때, 모수들  $(\psi, \Psi)$ 의 사후조건부확률분포는 다음과 같다.

$$\psi | (\boldsymbol{\theta}, \mathbf{y}_n) \stackrel{d}{\sim} \text{Gamma}\left(\frac{n_*}{2}, \frac{n_* S_*}{2}\right) \tag{5.3.203}$$

$$\Psi | (\boldsymbol{\theta}, \mathbf{y}_n) \stackrel{d}{\sim} \mathcal{W}_d(n_* - 1, \Psi_*) \tag{5.3.204}$$

여기서 상수와 행렬은 다음과 같다.

$$S_* \doteq \frac{1}{n_*} \left\{ n_0 S_0 + \sum_{t=1}^n [y_t - F_t^t \boldsymbol{\theta}_t]^2 \right\} \tag{5.3.205}$$

$$\Psi_*^{-1} \doteq \Psi_0^{-1} + \sum_{t=2}^n [\boldsymbol{\theta}_t - G_t \boldsymbol{\theta}_{t-1}] [\boldsymbol{\theta}_t - G_t \boldsymbol{\theta}_{t-1}]^t \tag{5.3.206}$$

식 (5.3.195)~식 (5.3.197)으로 이루어진 상태공간모형에 Gibbs샘플러를 적용하기 위해서는, 먼저 Kalman필터를 적용해서 각  $t(= 1, 2, \dots, n)$ 에 대해서  $\boldsymbol{\theta}_t, \mathbf{m}_t, C_t, G_t$ 와  $R_t$ 를

계산한다. 식 (5.3.200)에서 알 수 있듯이, 다음 식이 성립한다.

$$\boldsymbol{\theta}_n | (\boldsymbol{\psi}, \boldsymbol{\Psi}, \mathbf{y}_n) \stackrel{d}{\sim} \mathcal{N}(\mathbf{m}_n, C_n) \quad (5.3.207)$$

각  $t (= n-1, n-2, \dots, 1)$ 에 대해서 다음과 같은 평활식이 성립한다.

$$\boldsymbol{\theta}_t | (\boldsymbol{\psi}, \boldsymbol{\Psi}, \boldsymbol{\theta}_{t+1}, \dots, \boldsymbol{\theta}_n, \mathbf{y}_n) \stackrel{d}{\sim} \mathcal{N}(\mathbf{h}_t^{(n)}, H_t^{(n)}) \quad (5.3.208)$$

여기서 평균벡터와 분산공분산행렬은 각각 다음과 같다.

$$\mathbf{h}_t^{(n)} \doteq \mathbf{m}_t + C_t G_{t+1}^t R_{t+1}^{-1} [\boldsymbol{\theta}_{t+1} - \mathbf{a}_{t+1}] \quad (5.3.209)$$

$$H_t^{(n)} \doteq C_t - C_t G_{t+1}^t R_{t+1}^{-1} G_{t+1}^t C_t \quad (5.3.210)$$

만약  $\boldsymbol{\theta}_t$ 의 차원이 높거나, 만약 사후조건부확률분포가 퇴화하면, Durbin & Koopman [34]이 제시한대로 상태방정식의 오차항  $\boldsymbol{\omega}_t$ 를 사후확률분포에서 샘플링하는 과정을 추가해서  $\boldsymbol{\theta}_t$ 를 샘플링하는 방법이 효율적이다.

지금까지 내용을 바탕으로, 상태공간모형의 모수들  $(\boldsymbol{\theta}, \boldsymbol{\psi}, \boldsymbol{\Psi})$ 의 사후확률분포에서 샘플링하는 Gibbs 샘플러를 다음과 같이 요약할 수 있다.

**(1단계)** 초기모수들  $(\boldsymbol{\theta}^{(0)}, \boldsymbol{\psi}^{(0)}, \boldsymbol{\Psi}^{(0)})$ 를 선택한다.

**(2단계)** 각  $j (= 1, 2, \dots)$ 에 대해서 주어진  $\mathbf{y}$ 에 대한 조건부샘플  $(\boldsymbol{\theta}^{(j)}, \boldsymbol{\psi}^{(j)}, \boldsymbol{\Psi}^{(j)})$ 를 발생시킨다. 우선, 식 (5.3.207)의 사후조건부확률분포에서  $\boldsymbol{\theta}_n^{(j)} | (\boldsymbol{\psi}^{(j-1)}, \boldsymbol{\Psi}^{(j-1)}, \mathbf{y}_n)$ 을 발생시킨다. 각  $t (= n-1, n-2, \dots, 1)$ 에 대해서, 식 (5.3.208)의 사후조건부확률분포에서  $\boldsymbol{\theta}_t^{(j)} | (\boldsymbol{\psi}^{(j-1)}, \boldsymbol{\Psi}^{(j-1)}, \boldsymbol{\theta}_{t+1}^{(j)}, \dots, \boldsymbol{\theta}_n^{(j)}, \mathbf{y}_n)$ 을 발생시킨다. 다음으로, 식 (5.3.203)의 사후조건부확률분포에서  $\boldsymbol{\psi}^{(j)} | (\boldsymbol{\theta}^{(j)}, \mathbf{y}_n)$ 을 발생시키고, 식 (5.3.204)의 사후조건부확률분포에서  $\boldsymbol{\Psi}^{(j)} | (\boldsymbol{\theta}^{(j)}, \mathbf{y}_n)$ 을 발생시킨다.

**(3단계)** 충분히 큰  $b$ 에 대해서  $(\boldsymbol{\theta}^{(j)}, \boldsymbol{\psi}^{(j)}, \boldsymbol{\Psi}^{(j)})$ ,  $(j = b+1, b+2, \dots)$ 를 기록한다.

## 제 6 장

# Metropolis-Hastings 샘플러

Gibbs 샘플러를 사용해서 베이즈안통계분석을 하기 위해서는 사후조건부확률분포들  $\pi(\theta_1 | \theta_{-1}, \mathbf{y})$ ,  $\pi(\theta_2 | \theta_{-2}, \mathbf{y})$ ,  $\dots$ ,  $\pi(\theta_p | \theta_{-p}, \mathbf{y})$  에서 샘플링할 수 있어야 한다. 그러나, 실제 문제에서는 모든 사후조건부확률분포들에서 샘플링을 쉽게 할 수 있는 것은 아니다. Metropolis-Hastings 샘플러, 즉 MH 샘플러는 그러한 경우에도 사용할 수 있는 방법이다. MH 샘플러는 Gibbs 샘플링보다 일반적인 샘플링법이다. MH 샘플러는 통계물리학 분야에서 시스템의 균형상태를 시뮬레이션하기 위해 Metropolis 등 [76]에 의해 도입된 알고리즘을 Hastings [53]가 일반화한 것이다. MH 샘플러는 Gibbs 샘플러와 마찬가지로 사후조건부확률분포  $\pi(\theta_j | \theta_{-j}, \mathbf{y})$  에서 샘플링을 하는데 사용할 수 있을 뿐만 아니라, 사후결합확률분포  $\pi(\theta | \mathbf{y})$  에서 샘플링을 하는데도 사용할 수 있다.

### 제 6.1 절 MH 샘플러

#### 6.1.1 MH 샘플러

MH 샘플러는 목표확률분포, 좀 더 정확하게 말하면 목표확률밀도함수  $\pi(\cdot)$  에서 샘플링을 하는 것이다. MH 샘플러는 제안분포(proposal distribution), 좀 더 명확하게 말하면 제안분포의 조건부확률밀도함수  $q_{x_t, \cdot}$  로부터 후보점(candidate point)  $y$  를 샘플링하고, 이로부터 시점  $t+1$  에서 상태  $x_{t+1}$  구한다. 따라서,  $x_{t+1}$  은 오직 시점  $t$  에서 상태  $x_t$  에만 의존한다. 엄격히 말하면,  $q_{x_t, \cdot}$  는 확률분포가 아니라 추이확률함수, 즉 조건부확률밀도함수이다. 그러나, 관습상  $q_{x_t, \cdot}$  를 제안분포라 부른다. 제안분포  $q_{x_t, \cdot}$  는 불가약성(irreducibility)과 비주기성(aperiodicity)을 만족해야 한다. 만약 목표확률밀도함수  $\pi(\cdot)$  의 지지대가 유한이면, 이 조건들이 만족된다. 또한, 만약 제안분포  $q_{x_t, \cdot}$  가 목표확률밀도함수  $\pi(\cdot)$  의 지지대의 각 점에서 양의 확률밀도를

가지면, 이 조건들이 만족된다.

MH 샘플러에서는 다음과 같은 확률로 후보점  $y$ 가 시점  $t+1$ 에서 상태  $x_{t+1}$ 이 채택된다.

$$\alpha(x_t, y) \doteq \min \left\{ \frac{\pi(y)q_{y,x_t}}{\pi(x_t)q_{x_t,y}}, 1 \right\} \quad (6.1.1)$$

이  $\alpha(x_t, y)$ 를 MH채택확률(Metropolis-Hastings acceptance probability)이라 부른다. 만약 후보점  $y$ 가 채택되지 않으면, Markov 체인은 움직이지 않는다. 즉,  $x_{t+1} = x_t$ 이다. MH 샘플러를 다음과 같이 요약할 수 있다.

#### 알고리즘 6.1.1: Metropolis-Hastings 샘플러

(1단계) 적절한 초기값  $x_0$ 를 선택한다.

(2단계) 제안분포의 조건부확률밀도함수  $q_{x_t, \cdot}$ 에서 후보점  $y$ 를 생성한다.

(3단계) 일양확률분포  $Uniform(0, 1)$ 에서 일양난수  $u_t$ 를 생성한다.

(4단계) 만약 식  $u_t \leq \alpha(x_t, y)$ 가 성립하면,  $x_{t+1} = y$ 라 하고, 그렇지 않으면  $x_{t+1} = x_t$ 라 한다. 여기서 MH채택확률  $\alpha(x_t, y)$ 는 다음과 같다.

$$\alpha(x_t, y) = \min \left\{ \frac{\pi(y)q_{y,x_t}}{\pi(x_t)q_{x_t,y}}, 1 \right\}$$

(5단계) 주어진 번인기간  $[1, b]$ 와 샘플수  $n$ 에 대해서, 만약  $t+1 > b+n$ 이면, 이 알고리즘을 멈춘다. 그렇지 않으면,  $t$ 를  $t+1$ 으로 바꾼 다음 제2단계로 돌아간다.

알고리즘 6.1.1에서 알 수 있듯이, 제안분포의 조건부확률밀도함수  $q_{x_t, \cdot}$ 로부터 샘플링이 쉬운 MH 샘플러가 유용하다.

**예제 6.1.1** MH 샘플러를 사용해서, 다음과 같은 Cauchy 확률밀도함수로부터 난수들을 생성하기로 하자. 즉, 목표확률밀도함수는 Cauchy 확률분포를 따른다.

$$\pi(x) = \frac{1}{\pi[1+x^2]}, \quad (-\infty < x < \infty) \quad (1)$$

제안분포로 분산이  $\sigma^2 = 1$ 인 정규확률분포를 사용하자. 따라서 제안분포의 조건부확률밀도

함수는 다음과 같다.

$$q_{x_t,y} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}[y-x_t]^2\right) \quad (2)$$

이 MH샘플러를 적용하기 위해서, 다음 MATLAB 프로그램 MHsampler101.m을 실행해 보자.

```

1 % -----
2 % Filename: MHsampler101.m
3 % Cauchy Random Numbers generated by MH sampler
4 % Programmed by CBS
5 % -----
6 clear all, close all
7 rand('seed', (101^2+1)/2)
8 % Target pdf w/o the denominator constant
9 cauchy1 = '1./(1+x.^2)';
10 target = inline(cauchy1, 'x');
11 % Proposal pdf w/o the denominator constant
12 normal1 = '1/sig*exp(-1/2*((x-mu)/sig).^2)';
13 proposal = inline(normal1, 'x', 'mu', 'sig');
14 % Generate n=10000 samples from the Markov chain with burn-in number 1000.
15 n = 10000; b = 1000;
16 sig = 1;
17 x = zeros(1, n+b);
18 x(1) = randn(1);
19 for ii=2:n+b
20     y = x(ii-1) + sig*randn(1);
21     u = rand(1);
22     dumalpha = target(y)*proposal(x(ii-1), y, sig) ...
23               /(target(x(ii-1))*proposal(y, x(ii-1), sig));
24     alpha = min(1, dumalpha);
25     if u <= alpha
26         x(ii) = y;
27     else
28         x(ii) = x(ii-1);
29     end
30 end
31 % Histogram
32 xHistL = -40;
33 xHistR = 40;
34 HistIncrement = 1;
35 PDFind = ( (xHistL+(HistIncrement/2)):HistIncrement:xHistR );
36 [Nbar, xcenter] = hist(x(b+1:b+n), PDFind);
37 Nheight = Nbar/length(x(b+1:b+n))/(xcenter(2)-xcenter(1));
38 % True Normal PDF
39 CAUCHYpdf = target(PDFind)/pi;
40 % Plotting
41 hold on
42 bar(xcenter, Nheight, 0.95, 'g'); % Histogram
43 plot(PDFind, CAUCHYpdf, 'r--', 'LineWidth', [2.5]); % Normal PDF
44 set(gca, 'fontsize', 11, 'fontweigh', 'bold')
45 legend('Histogram', 'Cauchy PDF', 'location', 'NE');
46 xlabel('\bf Cauchy', 'fontsize', 12)
47 ylabel('\bf PDF', 'fontsize', 12)
48 hold off;
49 saveas(gcf, 'MHsampler101a', 'eps')
50 save('MHsampler101.txt', 'x', '-ascii')
51
52 % Display history of our samples

```

```

53 % subplot( 3,1,2:3 );
54 figure
55 FirstObs = 1:700;
56 stairs(x(FirstObs),FirstObs,'k-' );
57 set( gca , 'YDir' , 'reverse' );
58 set(gca,'fontsize',11,'fontweigh','bold')
59 xlim([xHistL,xHistR])
60 ylabel( 'time' ); xlabel( 'x' );
61 saveas(gcf,'MHsampler101b','eps')
62 % End of Program
63 % -----

```

이 MATLAB 프로그램 MHsampler101.m을 실행하면, 목표확률밀도함수인 Cauchy 확률 밀도함수 (1)에서 샘플들  $\{x_j | j = 1, 2, \dots, 11000\}$ 이 발생된다. 여기서 초기값  $x_1$ 은 표준정규난수이다. 이 중에서 1,000개 번인샘플들을 버린 나머지 10,000개 샘플들의 히스토그램을 그림 6.1.1에 그려져 있다. 그림 6.1.1에서 적색 긴점선은 목표확률밀도함수인 Cauchy 확률 밀도함수를 그린 것이다. 그림 6.1.1에서 히스토그램은 진짜확률밀도함수와 아주 가까움을 확인할 수 있다. 또한, 그림 6.1.2에는 첫 700개 MH샘플들의 시계열산점도가 그려져 있다. 이 시계열산점도에서 번인샘플들을 나타내는 첫 100개는 초기값인 정규난수의 영향을 받아서 정규확률분포에서 생성된 것처럼 0 주변에 몰려있으나, 나머지 MH샘플들은 번인샘플들보다 널리 퍼져있음을 관찰할 수 있다. ■

**예제 6.1.2** Gibbs샘플링은 샘플링에서 항상 추이가 일어나는 경우에 해당한다. 여기서는 2변량 확률벡터  $[x, y]$ 에서 Gibbs샘플링하는 경우를 살펴보자.

조건  $x = i$  하에서  $y = j$ 가 되는 조건부확률은  $\Pr(y = j | x = i) = q_{ij}$  이고,  $y = j$ 인 확률은  $\Pr(y = j) = \pi(j)$ 이다. 또한, 조건  $y = j$ 하에서  $x = i$ 가 되는 조건부확률은  $\Pr(x = i | y = j) = q_{ji}$  이고,  $x = i$ 인 확률은  $\Pr(x = i) = \pi(i)$ 이다. 따라서, 다음 식들이 성립한다.

$$\frac{\pi(j)q_{ji}}{\pi(i)q_{ij}} = \frac{\Pr(y = j)\Pr(x = i | y = j)}{\Pr(x = i)\Pr(y = j | x = i)} = \frac{\Pr(x = i, y = j)}{\Pr(y = j, x = i)} = 1 \quad (1)$$

즉, 다음 식들이 성립한다.

$$\alpha_{ij} = \min \left( \frac{\pi(j)q_{ji}}{\pi(i)q_{ij}}, 1 \right) = 1 \quad (2)$$

따라서, 항상 추이가 일어난다. ■

**예제 6.1.3** MH샘플러를 사용해서, 다음 목표확률밀도함수에서 난수들을 생성하기로

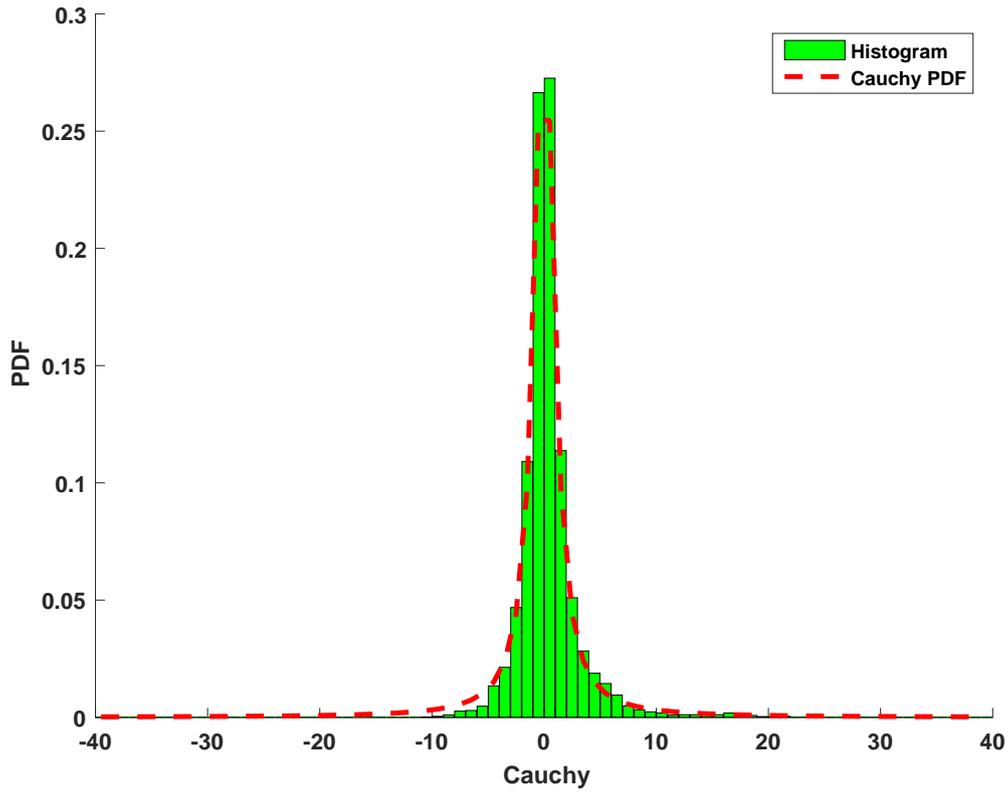


그림 6.1.1. MH샘플러와 Cauchy분포

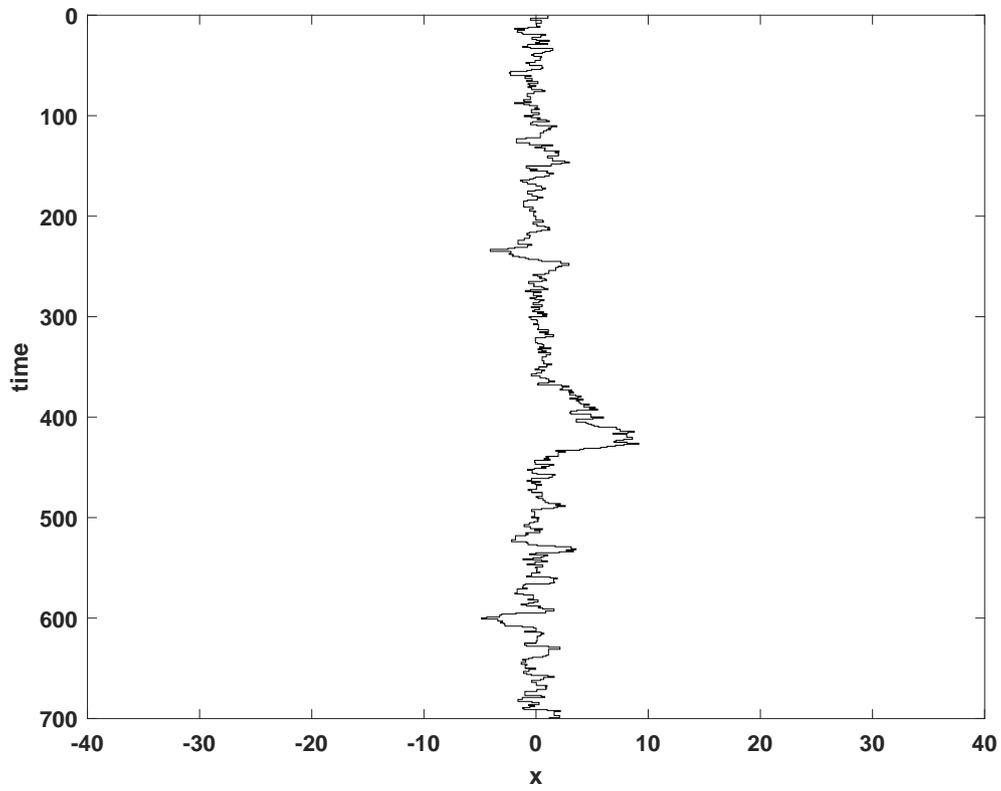


그림 6.1.2. MH샘플러와 Cauchy분포

하자.

$$\pi(x) = 0.7 \cdot \frac{1}{0.5\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left[\frac{x-7}{0.5}\right]^2\right) + 0.3 \cdot \frac{1}{0.5\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left[\frac{x-10}{0.5}\right]^2\right) \quad (1)$$

이 목표확률밀도함수는 혼합정규확률분포(mixed normal probability distribution)를 따른다. 다음과 같은 조건부확률밀도함수를 갖는 제안분포를 사용하기로 하자.

$$q_{x_t,y} = \frac{1}{0.1\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left[\frac{y-x_t}{0.1}\right]^2\right) \quad (2)$$

다음 식이 성립함을 자명하다.

$$q_{x_t,y} = q_{y,x_t} \quad (3)$$

따라서, MH채택확률은 다음과 같다.

$$\alpha(x_t, y) = \min\left\{\frac{\pi(y)}{\pi(x_t)}, 1\right\} \quad (4)$$

MH샘플러를 사용해서 목표확률밀도함수 (1)로부터 난수를 생성하는 예로서, 다음 R 프로그램 MHsampler103R.R을 실행해 보자.

```

1 # -----
2 # Filename:LogitMH101R.R
3 # Generic Metropolis-Hastings for Logit Model
4 # Robert and Casella (2004)
5 # About tThe space shuttle Challenger explosion in 1986
6 # Programmed by CBS
7 # -----
8 # ipak function: install and load multiple R packages.
9 # check to see if packages are installed.
10 # Install them if they are not, then load them into the R session.
11 ipak <- function(pkg){
12   new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
13   if (length(new.pkg))
14     install.packages(new.pkg, dependencies = TRUE)
15   sapply(pkg, require, character.only = TRUE)
16 }
17 # usage
18 packages <- c("ggplot2", "grid", "Rcpp")
19 # -----
20
21 MH <- function(x0, f, dprop, rprop, N, B) {
22   x <- matrix(NA, N + B, length(x0)); fx <- rep(NA, N + B)
23   x[1,] <- x0; fx[1] <- f(x0); ct <- 0
24   for(i in 2:(N + B)) {
25     u <- rprop(x[i-1,])
26     fu <- f(u)
27     r <- log(fu) + log(dprop(x[i-1,], u))
28     r <- r - log(fx[i-1]) - log(dprop(u, x[i-1,]))

```

```

29     R <- min(exp(r), 1)
30     if(runif(1) <= R) {
31         ct <- ct + 1
32         x[i,] <- u
33         fx[i] <- fu
34     } else {
35         x[i,] <- x[i-1,]
36         fx[i] <- fx[i-1]
37     }
38 }
39 return(list(x=x[-(1:B)], fx=fx[-(1:B)], rate=ct/(N + B)))
40 }
41
42 # Data
43 y <- c(1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0)
44 x <- c(53, 57, 58, 63, 66, 67, 67, 67, 68, 69, 70, 70, 70, 70, 72, 73, 75, 75,
45       76, 76, 78, 79, 81)
46
47 # Preliminary output from ML estimation
48 logreg.out <- glm(y ~ x, family=binomial(logit))
49 summary(logreg.out)
50 a.mle <- as.numeric(logreg.out$coefficients[1])
51 b.mle <- as.numeric(logreg.out$coefficients[2])
52 var.a.mle <- summary(logreg.out)$cov.scaled[1, 1]
53 var.b.mle <- summary(logreg.out)$cov.scaled[2, 2]
54 b.mme <- exp(a.mle + 0.577216)
55
56 # Posterior distribution
57 dpost <- function(theta) {
58     a <- theta[1]
59     b <- theta[2]
60     p <- 1 - 1/(1 + exp(a + b * x))
61     lik <- exp(sum(dbinom(y, size=1, prob=p, log=TRUE)))
62     dprior <- exp(a) * exp(-exp(a) / b.mme) / b.mme
63     return(lik * dprior)
64 }
65
66 # Proposal distribution (independent proposal, so "theta0" is not used)
67 dprop <- function(theta, theta0) {
68     a <- theta[1]
69     b <- theta[2]
70     pr1 <- exp(a) * exp(-exp(a) / b.mme) / b.mme
71     pr2 <- dnorm(b, b.mle, sqrt(var.b.mle))
72     return(pr1 * pr2)
73 }
74
75 rprop <- function(theta0) {
76     a <- log(rexp(1, 1 / b.mme))
77     b <- rnorm(1, b.mle, sqrt(var.b.mle))
78     return(c(a, b))
79 }
80
81 # Run Metropolis-Hastings
82 N <- 10000
83 B <- 1000
84 MH.out <- MH(c(a.mle, b.mle), dpost, dprop, rprop, N, B)
85 alpha.MH <- MH.out$x[,1]
86 beta.MH <- MH.out$x[,2]
87
88 p65 <- 1 - 1/(1 + exp(alpha.MH + beta.MH * 65))
89 p45 <- 1 - 1/(1 + exp(alpha.MH + beta.MH * 45))

```



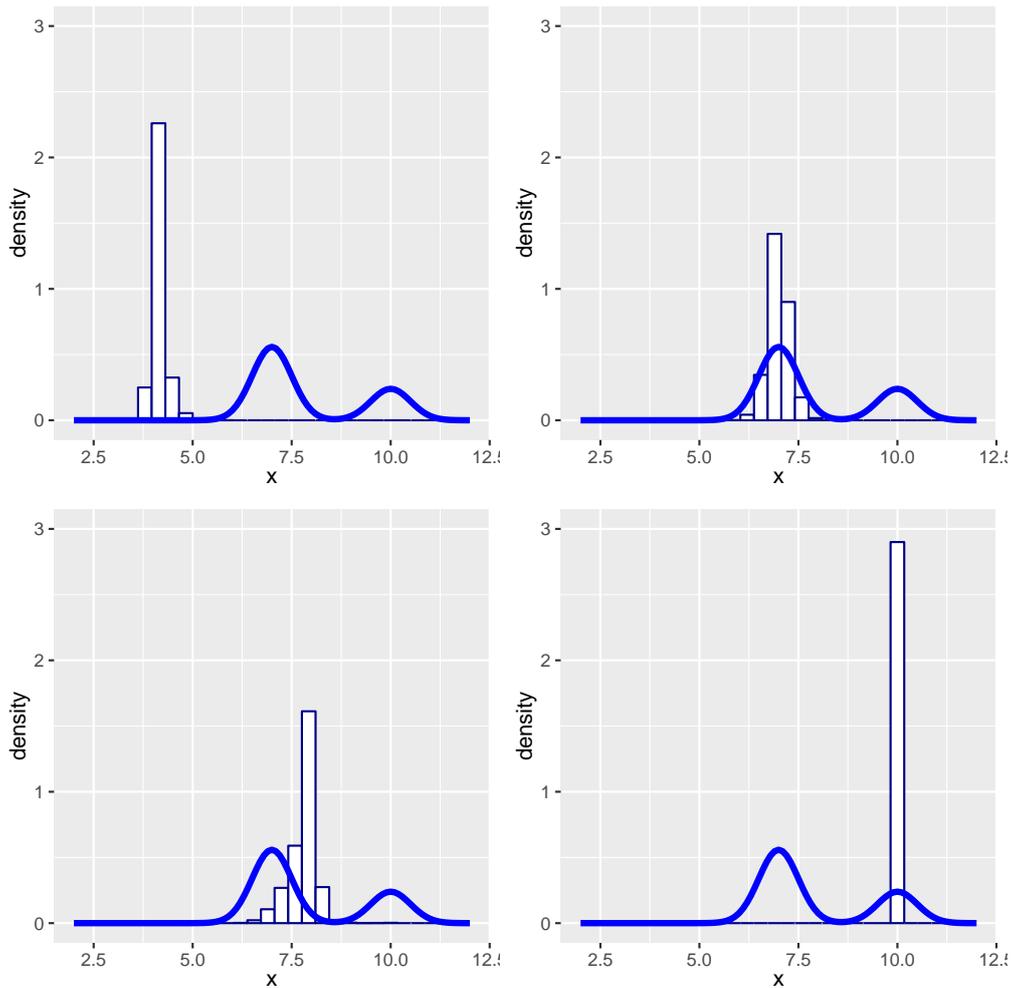


그림 6.1.3. MH샘플러와 혼합정규확률분포

**예제 6.1.4** MH샘플러를 사용해서, 다음 식을 만족하는 목표확률밀도함수에서 난수들을 생성하기로 하자.

$$\pi(x) \propto \sin^2(x) \sin^2(2x)\phi(x) \quad (1)$$

여기서  $\phi(x)$ 는 표준정규확률밀도함수이다. 제안분포는 일양확률분포  $Uniform(x-A, x+A)$ 이다. 즉, 제안분포의 조건부확률밀도함수는 다음과 같다.

$$q_{x_t, y} = \frac{1}{2A} 1_{(x_t-A, x_t+A)}(y) \quad (2)$$

MH샘플러를 사용해서 목표확률밀도함수 (1)로부터 난수를 생성하기 위해서, 다음 MATLAB 프로그램 MHsampler104R.R을 실행해 보자.

1 # -----

```

2 # Filename:MHsampler104R.R
3 # Metropolis-Hastings Sample
4 # From https://arxiv.org/pdf/1504.01896.pdf
5 # Modified by CBS
6 # -----
7 # Making Functions
8 set.seed(1)
9 target = function(x){
10     sin(x)^2*sin(2*x)^2*dnorm(x)}
11 metropolis = function(INx,A=1){
12     OUTx = runif(1,INx-A,INx+A)
13     if (runif(1) > target(OUTx)/target(INx)) OUTx = INx
14     return(OUTx)}
15
16 # MH Sampler
17 nBurnin <- 5000
18 nSim <- 10000
19 nTotal <- nBurnin + nSim
20 x = rep(0,nTotal)
21 for (t in 2:nTotal) x[t] <- metropolis(x[t-1])
22 x <- x[(nBurnin+1):(nBurnin+nSim)]
23
24 ## ACF
25 x_acf <- acf(x, plot = FALSE)
26 x_acfdf <- with(x_acf, data.frame(lag,acf))
27
28 # Plotting
29 # intall.packages("Rcpp")
30 # intall.packages("ggplot2")
31 library(ggplot2)
32 # install.packages("grid")
33 library(grid)
34 setEPS()
35 plot.new()
36 postscript('MHsampler104R.eps') # Start to save figure
37 timee = c(1:nSim)
38 MHdata <- data.frame(timee,x)
39 plot1 <- ggplot(MHdata, aes(x=x)) +
40     xlab(expression(x)) +
41     geom_histogram(aes(y = ..density..), bins=30,
42         # breaks=seq(0.0,2.5, by=0.05),
43         col="dark blue",fill="white") +
44     geom_density(lwd=2.0,col="red")
45 plot2 <- ggplot(data = x_acfdf, mapping = aes(x = lag, y = acf)) +
46     geom_hline(aes(yintercept = 0)) +
47     geom_segment(mapping = aes(xend = lag, yend = 0),lwd=2.0,color="blue")
48 pushViewport(viewport(layout = grid.layout(2,1)))
49 print(plot1, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
50 print(plot2, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
51 dev.off() # End to save figure
52 # -----

```

식 (2)에서 알 수 있듯이, 다음 식들이 성립한다.

$$q_{x_t,y} = \frac{1}{2A} 1(|y - x_t| < A) = q_{y,x_t} \quad (3)$$

따라서, MH채택확률  $\alpha(x_t, y)$ 는 다음과 같다.

$$\alpha(x_t, y) = \min \left\{ \frac{\pi(y)}{\pi(x_t)}, 1 \right\} \quad (4)$$

이 R프로그램 MHsampler104R.R을 실행하면, 목표확률밀도함수 (1)에서 샘플들  $\{x_j | j = 1, 2, \dots, 15000\}$ 이 발생된다. 이 중에서 5,000개 번인샘플들을 버린 나머지 10,000개 샘플들의 산점도들이 그림 6.1.4에 그려져 있다. 그림 6.1.4의 상단 그래프는 MH샘플들의 히스토그램을 그린 것이다. 이 그래프에서 적색 실선은 추정된 목표확률밀도함수를 그린 것이다. 또한, 하단 그래프는 이 MH샘플들의 자기상관함수를 그린 것이다. 이 그래프에서 알 수 있듯이, 이웃하는 MH샘플들의 자기상관성이 높다. ■

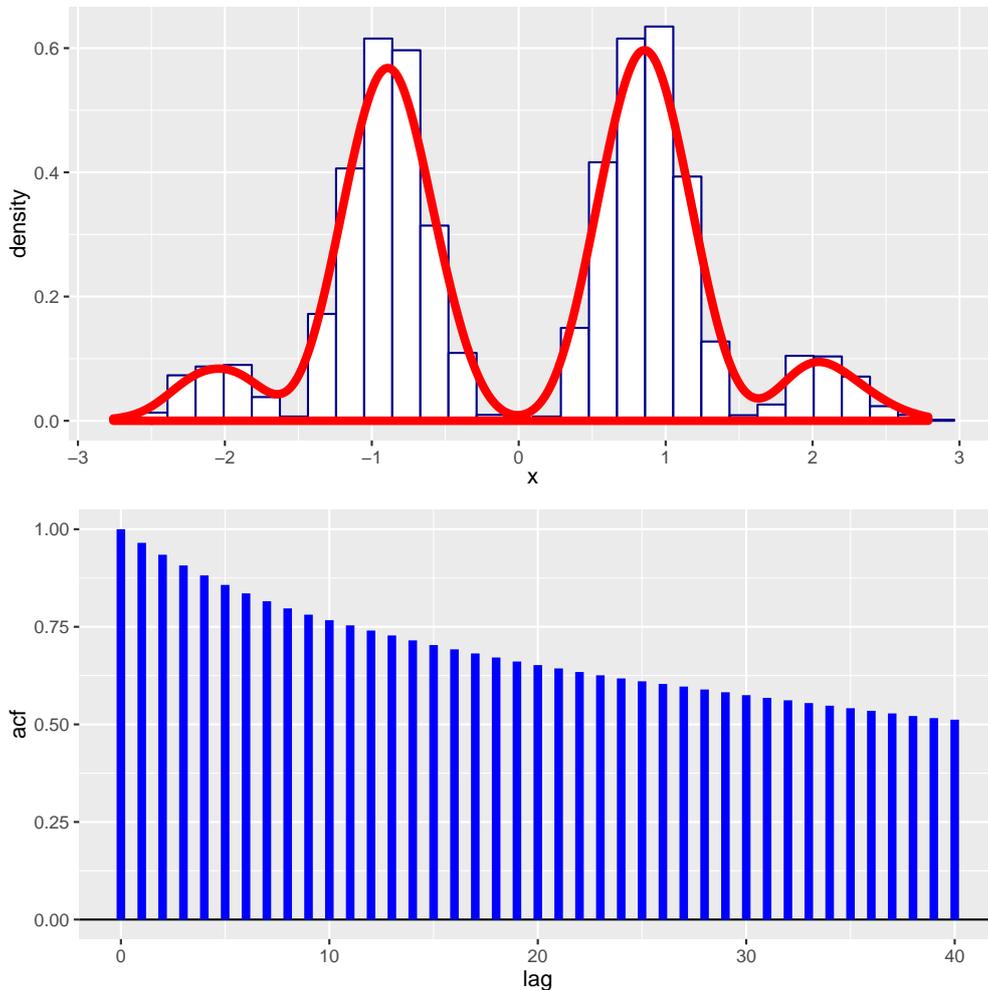


그림 6.1.4. MH샘플러와  $\sin^2(x) \sin^2(2x)\phi(x)$

**예제 6.1.5** MH샘플러를 사용해서 베타확률분포  $Beta(3, 7)$ 에서 난수들을 생성하기 위해,

다음 R프로그램 MHsampler105R.R을 실행해 보자.

```

1 # -----
2 # Filename:MHsampler105R.R
3 # Metropolis-Hastings Samples from Beta probability distribution
4 # Modified by CBS
5 # -----
6 # Making Functions
7 set.seed(1)
8 # MH Sampler
9 nBurnin <- 5000
10 nSim <- 10000
11 nTotal <- nBurnin + nSim
12 x <- rep(runif(1),nTotal) # Initial samples
13 for (ii in 2:nTotal) {
14   y <- runif(1)
15   alpha <- dbeta(y,3,7)/dbeta(x[ii-1],3,7)
16   x[ii] <- x[ii-1] + (y-x[ii-1])* ( runif(1)<alpha )
17 }
18 x <- x[(nBurnin+1):(nBurnin+nSim)]
19
20 # Plotting
21 # intall.packages("Rcpp")
22 # intall.packages("ggplot2")
23 library(ggplot2)
24 # install.packages("grid")
25 library(grid)
26 setEPS()
27 plot.new()
28 postscript('MHsampler105R.eps') # Start to save figure
29 timee = c(1:nSim)
30 MHdata <- data.frame(timee,x)
31 plot1 <- ggplot(MHdata, aes(x=x)) +
32   xlab(expression(x)) +
33   geom_histogram(aes(y =..density..), bins=30,
34     # breaks=seq(0.0,2.5, by=0.05),
35     col="dark blue",fill="green")
36 plot2 <- ggplot(MHdata, aes(timee,x)) +
37   geom_line(col="blue",lwd=1.2) +
38   xlim(9801,10000) + xlab("time")
39 pushViewport(viewport(layout = grid.layout(2,1)))
40 print(plot1, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
41 print(plot2, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
42 dev.off() # End to save figure
43 # -----

```

이 R프로그램 MHsampler105R.R에서 사용한 제안분포는 일양확률분포  $Uniform(0, 1)$ 이다. MH채택확률  $\alpha(x_t, y)$ 가 다음 식을 만족함은 자명하다.

$$\alpha(x_t, y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x_t)} \right\} \quad (1)$$

이 R프로그램 MHsampler105R.R을 실행하면, 베타확률분포  $Beta(3, 7)$ 에서 MH샘플들  $\{x_j | j = 1, 2, \dots, 15000\}$ 가 발생된다. 이 중에서 5,000개 번인샘플들을 버린 나머지 10,000

개 MH샘플들의 그래프들이 그림 6.1.5에 그려져 있다. 그림 6.1.5의 상단 그래프는 MH 샘플들의 히스토그램을 그린 것이고, 하단 그래프는 이 MH샘플들의 시계열산점도를 그린 것이다. ■

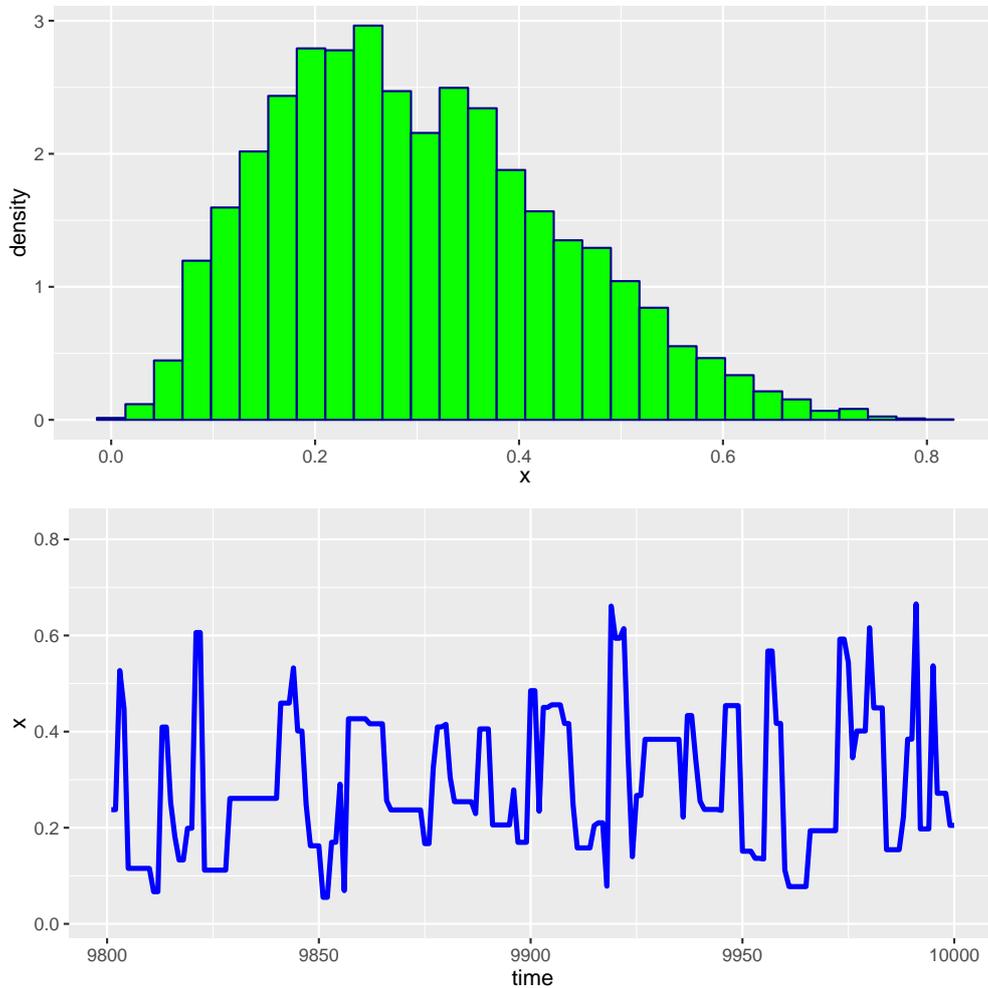


그림 6.1.5. MH샘플러와 베타확률분포

MH샘플러를 사용하는 경우에 제안분포의 조건부확률밀도함수  $q_{\cdot}$ 를 어떻게 선택하느냐는 중요한 문제이다. 특히 베이지안통계분석을 위해서 MH샘플러를 사용하는 경우, 제안분포를 어떻게 선택하느냐에 MH샘플러의 성공 여부가 달려있다. 만약 사후확률분포에 근사하는 제안분포를 찾는다면, 제안된 후보점  $\theta$ 를 수용하는 MH채택확률  $\alpha(\theta^{(j-1)}, \theta | \mathbf{y})$ 가 커져서 후보점의 채택률(acceptance rate)이 높아진다. 반면에, 제안분포가 사후확률분포의 좋은 근사함수가 아니면, 새로운 점으로 이동할 확률이 작아서 사후확률분포로부터 샘플링이 진행되지 않는다. 채택률이 높더라도  $\{\theta^{(j)} | j = 1, 2, \dots\}$ 의 자기상관성이 크다면, 이는 목표

확률분포의 상태공간중 일부분에서 샘플링을 반복하는 것이다. 따라서, 상태공간 전체에서 골고루 샘플링할 때까지는 긴 시간을 필요로 한다. 결론적으로, MH 샘플러에 의한 효율적 샘플링은 제안분포를 어떻게 선택하는냐에 달려있다고해도 과언이 아니다.

### 6.1.2 다변량 MH 샘플러

지금까지 소개한 단변량 MH 샘플러를 다변량 MH 샘플러로 쉽게 확장할 수 있다. 다변량 MH 샘플링을 블록단위 샘플링 (blockwise MH sampling) 과 성분단위 샘플링 (componentwise MH sampling) 으로 나눌 수 있다. 목표확률밀도 함수가 결합확률밀도 함수  $\pi(\mathbf{x})$  이고 제안분포의 조건부 결합확률밀도 함수가  $q_{\mathbf{x}_t, \mathbf{y}}$  라고 하자.

블록단위 샘플링은  $\mathbf{x}_t$  로부터  $\mathbf{x}_{t+1}$  을 한 번에 생성하는 방법이다. 지금까지 본서에서 난벡터 (random vector) 들을 생성하는데 사용한 샘플러는 블록단위 MH 샘플러이다. 다음 알고리즘을 사용해서 블록단위 샘플링을 실현할 수 있다.

#### 알고리즘 6.1.2: 블록단위 MH 샘플러

- (1단계) 적절한 초기벡터  $\mathbf{x}_0$  를 선택한다.
- (2단계) 제안분포의 조건부확률밀도 함수  $q_{\mathbf{x}_t, \mathbf{y}}$  에서 후보점  $\mathbf{y}$  를 생성한다.
- (3단계) 일양확률분포  $Uniform(0, 1)$  에서 일양난수  $u_t$  를 생성한다.
- (4단계) 만약 식  $u_t \leq \alpha(\mathbf{x}_t, \mathbf{y})$  가 성립하면,  $\mathbf{x}_{t+1} = \mathbf{y}$  라 하고, 그렇지 않으면  $\mathbf{x}_{t+1} = \mathbf{x}_t$  라 한다. 여기서 MH 채택확률  $\alpha(\mathbf{x}_t, \mathbf{y})$  는 다음과 같다.

$$\alpha(\mathbf{x}_t, \mathbf{y}) = \min \left\{ \frac{\pi(\mathbf{y})q_{\mathbf{y}, \mathbf{x}_t}}{\pi(\mathbf{x}_t)q_{\mathbf{x}_t, \mathbf{y}}}, 1 \right\}$$

- (5단계) 주어진 번인기간  $[1, b]$  와 샘플수  $n$  에 대해서, 만약  $t + 1 > b + n$  이면, 이 알고리즘을 멈춘다. 그렇지 않으면,  $t$  를  $t + 1$  으로 바꾼 다음 제2단계로 돌아간다.

**예제 6.1.6** 블록단위 MH 샘플러를 사용해서, 다음과 같은 2변량 정규확률분포에서 난벡터들을 생성하기로 하자.

$$\pi(\mathbf{x}) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp \left( -\frac{1}{2} \mathbf{x}^t \Sigma^{-1} \mathbf{x} \right) \quad (1)$$

여기서  $\Sigma$ 는 다음과 같다.

$$\Sigma = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix} \quad (2)$$

또한, 제안분포의 조건부결합확률밀도함수가 다음과 같다고 하자.

$$q_{\mathbf{x}_t, \mathbf{y}} = \frac{1}{2\pi} \exp\left(-\frac{1}{2} [\mathbf{x} - \mathbf{y}]^t [\mathbf{x} - \mathbf{y}]\right) \quad (3)$$

이 제안분포를 사용해서 목표확률분포로부터 블록단위MH샘플들을 생성하기 위해서 다음 MATLAB 프로그램 BlockwiseMHsampler101.m을 실행해 보자.

```

1 % -----
2 % Filename: BlockwiseMHsampler101.m
3 % Blockwise MH Sampler (BIVARIATE NORMAL)
4 % using MATLAB function mhsample
5 % Programmed by MathWorks
6 % -----
7 clear all, close all
8 rand('seed', (101^2+1)/2)
9 Dimen = 2; % Dimension of random vector
10 % Target distribution
11 pii = inline('mvnpdf(x,[0 0],[1 0.7 ; 0.7 1])','x');
12 % Proposal distribution
13 q = inline('mvnpdf(x,mu)','x','mu')
14
15 % Blockwise Sampler
16 nSim = 2000;
17 nBurnin = 5000;
18 nTot = nSim + nBurnin;
19 timee = 1;
20 x = zeros(nSim,2);
21 x(1,:) = randn(1,Dimen);
22 while timee < nTot
23     timee = timee + 1;
24     y = mvnrnd(x(timee-1,:),eye(Dimen));
25     alpha = min( [ 1, pii(y)*q(x(timee-1,:),y) ...
26                 /pii(x(timee-1,))/q(y,x(timee-1,)) ] );
27     if rand(1) < alpha
28         x(timee,:) = y;
29     else
30         x(timee,:) = x(timee-1,:);
31     end
32 end
33 xOut = x(nBurnin+1:nTot,:);
34 % Plotting
35 plot(xOut(:,1),xOut(:,2),'r','LineWidth',1.5)
36 set(gca,'fontsize',11,'fontweigh','bold')
37 hold all
38 for k = 0:4
39     theta = linspace(0,2*pi,201);
40     r = sqrt(51*k./10./(10-7*sin(2*theta)));
41     xx1 = r.*cos(theta);
42     xx2 = r.*sin(theta);
43     plot(xx1,xx2,'k-','LineWidth',1.5)

```

```

44 end
45 axis([-3.5 3.5 -3.5 3.5])
46 axis square
47 hold off
48 saveas(gcf, 'BlockwiseMHsampler101', 'eps')
49 save('BlockwiseMHsampler101.txt', 'xOut', '-ascii')
50 % End of program
51 % -----

```

이 MATLAB 프로그램 BlockwiseMHsampler101.m을 실행하면, 블록단위MH샘플러를 사용하여 2,000개 정규난벡터들을 발생시킨다. 이 블록단위MH샘플들의 산점도가 그림 6.1.6에 그려져 있다. 그림 6.1.6에서 적색 점들은 MH샘플들이고, 흑색 실선들은 각  $k(= 1, 2, 3, 4)$ 에 대해서 식  $\mathbf{x}^t \Sigma^{-1} \mathbf{x} = k$ 를 만족하는 등가곡선(isoquant curve)이다. ■

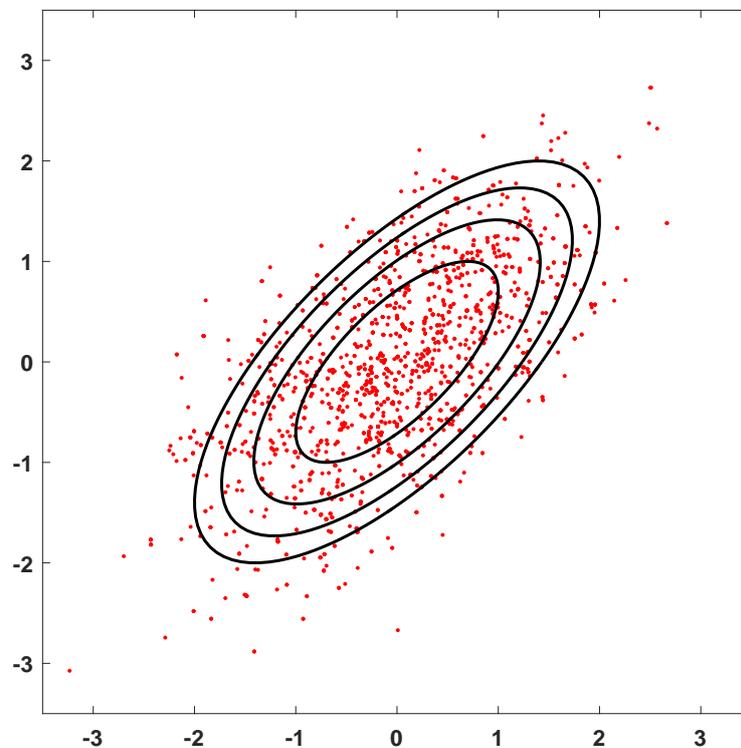


그림 6.1.6. 블록단위MH샘플러와 2변량 정규확률분포 1

**예제 6.1.7** 예제 6.1.6를 다시 살펴보자. 블록단위MH샘플들을 생성하기 위해서 다음 MATLAB 프로그램 BlockwiseMHsampler501.m을 실행해 보자.

```

1 % -----
2 % Filename: BlockwiseMHsampler501.m
3 % Blockwise MH Sampler (BIVARIATE NORMAL)
4 % https://theclevermachine.wordpress.com/2012/11/04/

```

```

5 %   mcmc-multivariate-distributions-block-wise-component-wise-updates/
6 % -----
7 clear all, close all
8 rand('seed',(101^2+1)/2)
9
10 D = 2; % # VARIABLES
11 nBurnIn = 5000;
12
13 % TARGET DISTRIBUTION IS A 2D NORMAL WITH STRONG COVARIANCE
14 p = inline('mvnpdf(x,[0 0],[1 0.7 ; 0.7 1])','x');
15
16 % PROPOSAL DISTRIBUTION STANDARD 2D GUASSIAN
17 q = inline('mvnpdf(x,mu)','x','mu')
18
19 nSamples = 2000;
20 minn = [-3 -3];
21 maxx = [3 3];
22
23 % INITIALIZE BLOCK-WISE SAMPLER
24 t = 1;
25 x = zeros(nSamples,2);
26 x(1,:) = randn(1,D);
27
28 % RUN SAMPLER
29 while t < nSamples
30     t = t + 1;
31
32     % SAMPLE FROM PROPOSAL
33     xStar = mvnrnd(x(t-1,:),eye(D));
34
35     % CORRECTION FACTOR (SHOULD EQUAL 1)
36     c = q(x(t-1,:),xStar)/q(xStar,x(t-1,:));
37
38     % CALCULATE THE M-H ACCEPTANCE PROBABILITY
39     alpha = min([1, p(xStar)/p(x(t-1,:))]);
40
41     % ACCEPT OR REJECT?
42     u = rand;
43     if u < alpha
44         x(t,:) = xStar;
45     else
46         x(t,:) = x(t-1,:);
47     end
48 end
49
50 % DISPLAY
51 nBins = 20;
52 bins1 = linspace(minn(1), maxx(1), nBins);
53 bins2 = linspace(minn(2), maxx(2), nBins);
54
55 % DISPLAY SAMPLED DISTRIBUTION
56 ax = subplot(121);
57 bins1 = linspace(minn(1), maxx(1), nBins);
58 bins2 = linspace(minn(2), maxx(2), nBins);
59 sampX = hist3(x, 'Edges', {bins1, bins2});
60 hist3(x, 'Edges', {bins1, bins2});
61 view(-15,40)
62
63 % COLOR HISTOGRAM BARS ACCORDING TO HEIGHT
64 colormap autumn
65 set(gcf,'renderer','opengl');

```

```

66 set(gca, 'fontsize', 11, 'fontweigh', 'bold')
67 set(get(gca, 'child'), 'FaceColor', 'interp', 'CDataMode', 'auto');
68 xlabel('x_1'); ylabel('x_2'); zlabel('Frequency');
69 axis square
70 set(ax, 'xTick', [minn(1), 0, maxx(1)]);
71 set(ax, 'yTick', [minn(2), 0, maxx(2)]);
72 title('Sampled Distribution');
73
74 % DISPLAY ANALYTIC DENSITY
75 ax = subplot(122);
76 [x1, x2] = meshgrid(bins1, bins2);
77 probX = p([x1(:), x2(:)]);
78 probX = reshape(probX, nBins, nBins);
79 surf(probX); axis xy
80 view(-15, 40)
81 xlabel('x_1'); ylabel('x_2'); zlabel('p(\{bfx\})');
82 colormap autumn
83 axis square
84 set(ax, 'xTick', [1, round(nBins/2), nBins]);
85 set(ax, 'xTickLabel', [minn(1), 0, maxx(1)]);
86 set(ax, 'yTick', [1, round(nBins/2), nBins]);
87 set(ax, 'yTickLabel', [minn(2), 0, maxx(2)]);
88 title('Analytic Distribution')
89
90 saveas(gcf, 'BlockwiseMHsampler501', 'eps')
91 % End of program
92 % -----

```

이 MATLAB 프로그램 BlockwiseMHsampler501.m을 실행하면, 블록단위 MH샘플러를 사용해서 2,000개 정규난벡터들을 발생시킨다. 이 블록단위 MH샘플들의 히스토그램을 그림 6.1.7의 좌측 그래프에 그려져 있고, 결합확률밀도함수가 우측 그래프에 그려져 있다. ■

**예제 6.1.8** 2변량 지수확률밀도함수 (bivariate exponential probability density) 에서 블록단위 MH샘플들을 생성하기 위해서 다음 MATLAB 프로그램 BlockwiseMHsampler601.m 을 실행해 보자.

```

1 % -----
2 % Filename: BlockwiseMHsampler601.m
3 % Metropolis procedure to sample from Bivariate Exponential
4 % Blockwise updating. Use a uniform proposal distribution
5 % From http://www.cidlab.com/205c/205C\_v4.pdf, p. 26.
6 % Problem adopted from Gill (2008)
7 % -----
8 function BlockwiseMHsampler
9 format short
10
11 % Initialize the Metropolis sampler
12 T = 5000; % Set the maximum number of iterations
13 thetamin = [ 0 0 ]; % define minimum for theta1 and theta2
14 thetamax = [ 8 8 ]; % define maximum for theta1 and theta2

```

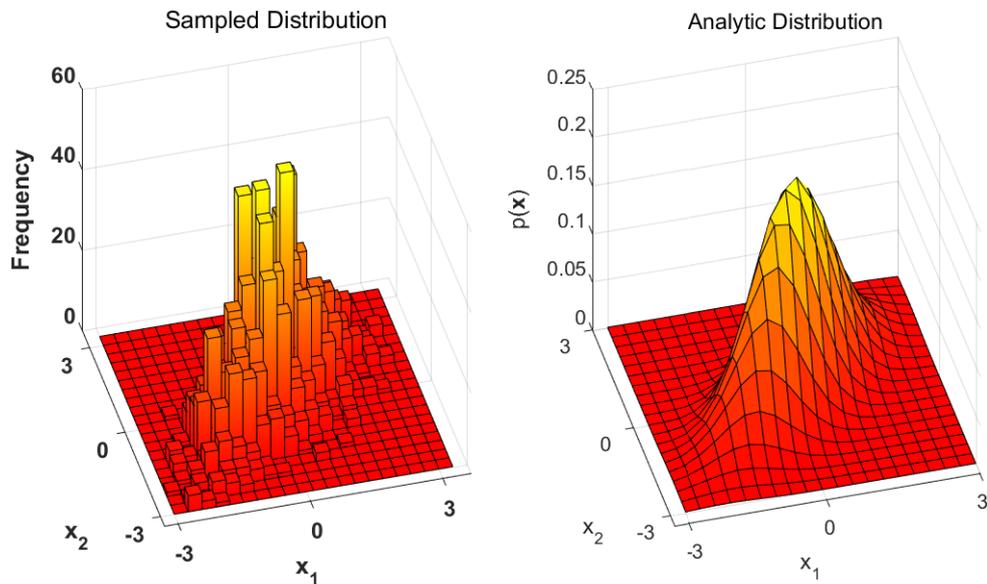


그림 6.1.7. 블록단위MH샘플러와 2변량 정규확률분포 2

```

15 seed = 1; % set the random seed
16 rand( 'state' , seed ); randn('state',seed );
17 theta = zeros( 2 , T ); % Init storage space for our samples
18 theta(1,1) = unifrnd(thetamin(1),thetamax(1)); % Start value for theta1
19 theta(2,1) = unifrnd(thetamin(2),thetamax(2)); % Start value for theta2
20
21 % Start sampling
22 t = 1;
23 while t < T % Iterate until we have T samples
24     t = t + 1;
25     % Propose a new value for theta
26     theta_star = unifrnd( thetamin , thetamax );
27     pratio = bivexp(theta_star(1), theta_star(2) ) / ...
28             bivexp( theta(1,t-1) , theta(2,t-1) );
29     alpha = min([1 pratio]); % Calculate the acceptance ratio
30     u = rand; % Draw a uniform deviate from [ 0 1 ]
31     if u < alpha % Do we accept this proposal?
32         theta(:,t) = theta_star; % proposal becomes new value for theta
33     else
34         theta(:,t) = theta(:,t-1); % copy old value of theta
35     end
36 end
37
38 % Plotting
39 figure( 1 ); clf;
40 % Display histogram of our samples
41

```

```

42 subplot(1,2,1);
43 nbins = 10;
44 thetabins1 = linspace( thetamin(1) , thetamax(1) , nbins );
45 thetabins2 = linspace( thetamin(2) , thetamax(2) , nbins );
46 hist3( theta' , 'Edges' , {thetabins1 thetabins2} );
47 xlabel( '\theta 1' ); ylabel( '\theta 2' ); zlabel( 'counts' );
48 set(gca, 'fontSize',11, 'fontweigh', 'bold')
49 az = 61; el = 30;
50 view(az, el);
51
52 % Plot the theoretical density
53 subplot(1,2,2);
54 nbins = 20;
55 thetabins1 = linspace( thetamin(1) , thetamax(1) , nbins );
56 thetabins2 = linspace( thetamin(2) , thetamax(2) , nbins );
57 [theta1grid,theta2grid] = meshgrid(thetabins1,thetabins2);
58 ygrid = bivexp( theta1grid , theta2grid );
59 mesh( theta1grid , theta2grid , ygrid );
60 set(gca, 'fontSize',11, 'fontweigh', 'bold')
61 xlabel( '\theta 1' ); ylabel( '\theta 2' );
62 zlabel( 'f(\theta 1,\theta 2)' );
63 view(az, el);
64
65 saveas(gcf, 'BlockwiseMHsampler601', 'eps')
66 % End of program
67 % -----
68 function y = bivexp(theta1,theta2)
69 % Returns the density of a bivariate exponential function
70 lambda1 = 0.5; % Set up some constants
71 lambda2 = 0.1;
72 lambda = 0.01;
73 maxval = 8;
74 y = exp( -(lambda1+lambda)*theta1 -(lambda2+lambda)*theta2 ...
75         -lambda*maxval );
76 % End of the function bivexp
77 % -----

```

이 MATLAB 프로그램 BlockwiseMHsampler601.m은 다음과 같은 2변량 지수확률밀도 함수에서 MH샘플들을 발생시키기 위한 것이다.

$$\pi(\theta_1, \theta_2) = \exp(-(\lambda_1 + \lambda)\theta_1 - (\lambda_2 + \lambda)\theta_2 - \lambda(\theta_1, \theta_2)) \quad (1)$$

여기서  $(\theta_1, \theta_2)$ 의 지지대는  $[0, 8] \times [0, 8]$ 이다. 또한, 다음과 같이 모수들을 정의하자.

$$\lambda_1 = 0.5, \quad \lambda_2 = 0.1, \quad \lambda = 0.01, \quad \max\{\theta_1, \theta_2\} = 8 \quad (2)$$

또한 제안분포는 다음과 같이 서로 독립인 일양확률분포들의 결합확률분포이다.

$$(\theta_1, \theta_2) \stackrel{d}{\sim} \text{Uniform}[0, 8] \times \text{Uniform}[0, 8] \quad (3)$$

이 MATLAB 프로그램 BlockwiseMHsampler601.m에서 사용한 샘플러는 앞 상태 (the previous state of the sampler)에 의존하지 않는다. 뒤에서 언급하게 되겠지만 이러한 샘플러를 독립샘플러 (independence sampler)라 부른다. 이 제안분포가 비록 좋은 것은 아니지만 실장 (implementation)하기가 쉽다.

이 MATLAB 프로그램 BlockwiseMHsampler601.m을 실행하면, 그림 6.1.8이 그려진다. 그림 6.1.8의 우변 그래프는 이 2변량 지수확률밀도함수를 그린 것이고, 좌변 그래프는 발생된 5000개 MH샘플들의 히스토그램을 그린 것이다. ■

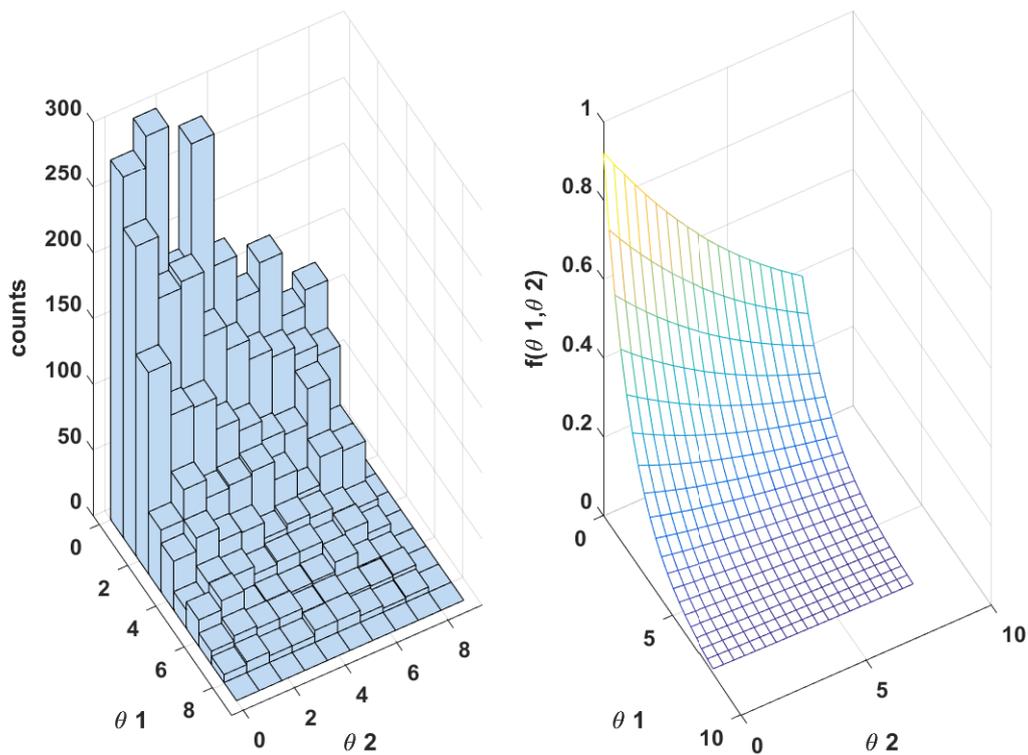


그림 6.1.8. 블록단위MH샘플러와 2변량 지수확률분포 2

**예제 6.1.9** 이 예제는 Steyvers and Vandekerckhove의 2015년 강의노트에서 인용한 것으로, Steyvers등이 쓴 논문 ‘The wisdom of crowds in the recollection of order information’에서 다룬 내용을 요약한 것이다 [94].

미국의 첫 다섯 대통령들의 순열(permutation)을 다음과 같이 표기하자.

$$\omega = (1, 2, 3, 4, 5) = (\textit{Washington}, \textit{Adams}, \textit{Jefferson}, \textit{Madison}, \textit{Monroe}) \quad (1)$$

어떤 개인이 기억하는 첫 다섯 대통령들의 순서를  $\theta$ 라고 하고,  $\theta$ 의 확률이 다음과 같다고 가정하자.

$$p(\theta | \omega, \lambda) \propto \exp(-d(\theta, \omega)\lambda) \quad (2)$$

여기서  $d(\theta, \omega)$ 는 두 순열들  $\theta$ 와  $\omega$ 의 Kendall의 타우거리(tau distance)이다. 이 Kendall의 타우거리는 두 순열들이 일치하기 위해서 이웃하는 원소들을 맞바꾸어야(adjacent pairwise swaps) 하는 횟수를 의미한다. 예를 들어, 다음 식들이 성립한다.

$$d(\theta, (\textit{Adams}, \textit{Washington}, \textit{Jefferson}, \textit{Madison}, \textit{Monroe})) = 1 \quad (3)$$

$$d(\theta, (\textit{Adams}, \textit{Jefferson}, \textit{Washington}, \textit{Madison}, \textit{Monroe})) = 2 \quad (4)$$

유의할 점은 Kendall의 타우거리는 이웃하는 원소들만을 맞바꾸는 것을 의미한다. 또한, 식 (2)의 모수  $\lambda$ 는 기억된 순열(remembered ordering)  $\theta$ 의 확률분포가 얼마나 뾰족한가를 나타낸다. 따라서,  $\lambda$ 가 증가하면, 확률밀도함수  $p(\theta | \omega, \lambda)$ 의 매스(mass)가 올바른 순열 주변에 더 많이 모인다.

첫 번째 순열  $\theta^{(1)}$ 를 임의로 선택하고, 제안분포에서는 Cayley거리  $S(\theta^*, \theta)$ 를 사용하기로 하자. Cayley거리는 두 순열들  $\theta^*$ 와  $\theta$ 가 일치하기 위해서 원소들을 맞바꾸어야(pairwise swaps) 하는 횟수를 의미한다. 이 Cayley거리를 계산할 때 맞바꾸어야 할 원소들이 바로 이웃할 필요는 없다. 즉, 제  $(t-1)$ 번째 MH샘플  $\theta^{(t-1)}$ 가 주어졌을 때 다음과 같은 제안분포의 조건부확률밀도함수에서 제안점  $\theta^*$ 를 뽑는다.

$$q\left(\theta = \theta^* \mid \theta^{(t-1)}\right) = \begin{cases} \frac{1}{\binom{N}{2}}, & (S(\theta^*, \theta^{(t-1)}) = 1) \\ 0, & (\textit{otherwise}) \end{cases} \quad (5)$$

여기서  $N$ 은 순열을 이루는 원소들의 개수이다. 즉, 이 문제에서는  $N = 5$ 이다. 따라서 MH

채택확률은 다음과 같다.

$$\alpha(\theta^{(t-1)}, \theta^*) = \min \left\{ \frac{p(\theta^* | \omega, \lambda)}{p(\theta^{(t-1)} | \omega, \lambda)}, 1 \right\} = \min \left\{ \frac{\exp(-d(\theta^*, \omega)\lambda)}{\exp(-d(\theta^{(t-1)}, \omega)\lambda)}, 1 \right\} \quad (6)$$

목표확률분포로부터 블록단위MH샘플들을 생성하기 위해서 다음 MATLAB 프로그램 BlockwiseMHsampler602.m을 실행해 보자.

```

1 % -----
2 % Filename: BlockwiseMHsampler602.m
3 % Blockwise MH sampler for Mallows model
4 % Samples orderings from a distribution over orderings
5 % From http://www.cidlab.com/205c/205C_v4.pdf, pp. 29-31.
6 % -----
7 function MetropolisSampler
8 clear all, close all, format long
9
10 % Initialize model parameters
11 lambda = 0.1; % scaling parameter
12 labels = { 'Washington' , 'Adams' , 'Jefferson' , 'Madison' , 'Monroe' };
13 omega = [ 1 2 3 4 5 ]; % correct ordering
14 L = length( omega ); % number of items in ordering
15
16 % Initialize the Metropolis sampler
17 T = 500; % Set the maximum number of iterations
18 rand('state',1); randn('state',1); % set the random seed
19 theta = zeros( L , T ); % Init storage space for our samples
20 theta(:,1) = randperm( L ); % Random ordering to start with
21
22 % Start sampling
23 t = 1;
24 while t < T % Iterate until we have T samples
25     t = t + 1;
26     % Our proposal is the last ordering but with two items switched
27     lasttheta = theta(:,t-1); % Get the last theta
28     % Propose two items to switch
29     whswap = randperm( L ); whswap = whswap(1:2);
30     thetastar = lasttheta;
31     thetastar( whswap(1) ) = lasttheta( whswap(2) );
32     thetastar( whswap(2) ) = lasttheta( whswap(1) );
33
34     % calculate Kendall tau distances
35     dist1 = kendalltau( thetastar , omega );
36     dist2 = kendalltau( lasttheta , omega );
37
38     % Calculate the acceptance ratio
39     pratio = exp( -dist1*lambda ) / exp( -dist2*lambda );
40     alpha = min( [ 1 pratio ] );
41     u = rand; % Draw a uniform deviate from [ 0 1 ]
42     if u < alpha % Do we accept this proposal?
43         theta(:,t) = thetastar; % proposal becomes new theta
44     else
45         theta(:,t) = lasttheta; % copy old theta
46     end
47

```

```

48 % Occasionally print the current state
49 if mod( t,50 ) == 0
50     fprintf( 't=%3d\t' , t );
51     for j=1:L
52         fprintf( '%15s' , labels{ theta(j,t)} )
53     end
54     fprintf( '\n' );
55 end
56 end
57
58 % End of program
59 % -----
60 function tau = kendalltau( order1 , order2 )
61 % Returns the Kendall tau distance between two orderings
62 % Note: this is not the most efficient implementation
63 [ dummy , ranking1 ] = sort( order1(:)' , 2 , 'ascend' );
64 [ dummy , ranking2 ] = sort( order2(:)' , 2 , 'ascend' );
65 N = length( ranking1 );
66 [ ii , jj ] = meshgrid( 1:N , 1:N );
67 ok = find( jj(:) > ii(:) );
68 ii = ii( ok );
69 jj = jj( ok );
70 nok = length( ok );
71 sign1 = ranking1( jj ) > ranking1( ii );
72 sign2 = ranking2( jj ) > ranking2( ii );
73 tau = sum( sign1 ~= sign2 );
74 % End of the function kendalltau
75 % -----

```

이 MATLAB 프로그램 BlockwiseMHsampler602.m을 실행하면, 블록단위MH샘플러를 사용해서 500개 순열들을 발생시킨다. 이 중에서 50개마다 출력한 결과는 다음과 같다.

t= 50	Madison	Washington	Adams	Monroe	Jefferson
t=100	Adams	Jefferson	Monroe	Madison	Washington
t=150	Jefferson	Adams	Washington	Madison	Monroe
t=200	Jefferson	Adams	Madison	Monroe	Washington
t=250	Jefferson	Monroe	Washington	Madison	Adams
t=300	Adams	Monroe	Washington	Madison	Jefferson
t=350	Monroe	Washington	Madison	Adams	Jefferson
t=400	Jefferson	Madison	Adams	Monroe	Washington
t=450	Jefferson	Washington	Adams	Madison	Monroe
t=500	Monroe	Adams	Madison	Jefferson	Washington

■

블록단위MH샘플러를 적용하는 데 발생하는 문제점은 난벡터의 차수  $d$ 가 클 때 적절한 제안분포를 찾는 것이 어렵다는 것이다. 즉, MH채택확률이 높은 제안분포를 찾는 것이 쉽지 않다. 이러한 문제점을 해결하는 한 가지 방법은 난벡터 자체를 한꺼번에 생성하는 것이 아니라

난벡터의 성분들을 차례로 생성하는 것이다. 시점  $t$ 에서 벡터들을 다음과 같이 표기하자.

$$\mathbf{x}_t = [x_{t,1}, x_{t,2}, \dots, x_{t,d}]^t \quad (6.1.2)$$

$$\mathbf{x}_{t,-i} = [x_{t,1}, \dots, x_{t,i-1}, x_{t,i+1}, \dots, x_{t,d}]^t \quad (6.1.3)$$

성분단위MH샘플러에서는 1차원 제안분포의 조건부확률밀도함수  $q_{x_{t-1,i}, \cdot}$ 를 이용한다. 다음 알고리즘을 사용해서 성분단위MH샘플러를 실현할 수 있다.

### 알고리즘 6.1.3: 성분단위MH샘플러

(1단계) 적절한 초기벡터  $\mathbf{x}_0$ 를 선택한다.

(2단계) 각  $i (= 1, 2, \dots, d)$ 에서 다음과 같이  $x_{t-1,i}$ 를 생성한다.

(2-1) 제안분포의 조건부확률밀도함수  $q_{x_{t-1,i}, \cdot}$ 에서 후보점  $y_i$ 를 생성한다.

(2-2) 일양확률분포  $Uniform(0, 1)$ 에서 일양난수  $u_{t,i}$ 를 생성한다.

(2-3) 만약  $u_{t,i} \leq \alpha(x_{t,i}, y_i)$ 가 성립하면,  $x_{t,i} = y_i$ 라 하고, 그렇지 않으면  $x_{t,i} = x_{t-1,i}$ 라 한다. 여기서 MH채택확률  $\alpha(x_{t,i}, y_i)$ 는 다음과 같다.

$$\alpha(x_{t,i}, y_i) = \min \left\{ \frac{\pi(y_i, \mathbf{x}_{t-1,-i}) q_{y_i, x_{t-1,i}}}{\pi(\mathbf{x}_{t-1}) q_{x_{t-1,i}, y_i}}, 1 \right\}$$

(3단계) 주어진 번인기간  $[1, b]$ 와 샘플수  $n$ 에 대해서, 만약  $t + 1 > b + n$ 이면, 이 알고리즘을 멈춘다. 그렇지 않으면,  $t$ 를  $t + 1$ 으로 바꾼 다음 제2단계로 돌아간다.

**예제 6.1.10** 성분단위MH샘플러를 사용해서, 다음과 같은 2변량 정규확률밀도함수에서 난벡터들을 생성하기로 하자.

$$\pi(\mathbf{x}) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^t \Sigma^{-1} \mathbf{x}\right) \quad (1)$$

여기서  $\Sigma$ 는 다음과 같다.

$$\Sigma = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix} \quad (2)$$

또한, 제안분포의 조건부결합확률밀도함수는 다음과 같다.

$$q_{\mathbf{x}_t, \mathbf{y}} = \frac{1}{2\pi} \exp\left(-\frac{1}{2} [\mathbf{x} - \mathbf{y}]^t [\mathbf{x} - \mathbf{y}]\right) \quad (3)$$

따라서 주변확률밀도함수는 다음과 같다.

$$q_{x_{t,i}, y} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} [x_{t,i} - y]^2\right) \quad (4)$$

이 제안분포를 사용해서 목표확률밀도함수로부터 성분단위MH샘플들을 생성하기 위해서 다음 MATLAB 프로그램 ComponentwiseMHsampler101.m을 실행해 보자.

```

1 % -----
2 % Filename: ComponentwiseMHsampler101.m
3 % Componentwise MH Sampler (BIVARIATE NORMAL)
4 % using MATLAB function mhsample
5 % Programmed by MathWorks
6 % -----
7 clear all, close all
8 rand('seed', (101^2+1)/2)
9 Dimen = 2; % Dimension of random vector
10 % TARGET DISTRIBUTION
11 pii = inline('mvnpdf(x,[0 0],[1 0.7 ; 0.7 1])','x');
12 % Componentwise Sampler
13 ProposalSD = 1;
14 nSim = 2000;
15 nBurnin = 5000;
16 nTot = nSim + nBurnin;
17 % INITIALIZE COMPONENT-WISE SAMPLER
18 x = zeros(nTot, Dimen);
19 xNew = randn(1, Dimen)
20 timee = 1;
21 x(timee, :) = xNew;
22 % Componentwise MH sampler
23 while timee < nTot
24     timee = timee + 1;
25     for dd = 1:Dimen % dd = 1:dimension
26         y = normrnd(xNew(:, dd), ProposalSD);
27         piiRatio = pii([y xNew(3-dd)]) / pii([xNew(1) xNew(2)]);
28         alpha = min([1, piiRatio]);
29         if rand < alpha
30             xNew(dd) = y;
31         end
32     end
33 x(timee, :) = xNew;
34 end
35 xOut = x(nBurnin+1:nTot, :);
36 % Plotting
37 plot(xOut(:, 1), xOut(:, 2), 'r.', 'LineWidth', 1.5)
38 set(gca, 'fontsize', 11, 'fontweigh', 'bold')
39 hold all
40 for k = 0:4
41     theta = linspace(0, 2*pi, 201);
42     r = sqrt(51*k./10./(10-7*sin(2*theta)));

```

```

43     xx1 = r.*cos(theta);
44     xx2 = r.*sin(theta);
45     plot(xx1,xx2,'k-','LineWidth',1.5)
46 end
47 axis([-3.5 3.5 -3.5 3.5])
48 axis square
49 hold off
50 saveas(gcf,'ComponentwiseMHsampler101','eps')
51 % End of program
52 % -----

```

이 MATLAB 프로그램 ComponentwiseMHsampler101.m을 실행하면, 성분단위MH샘플러를 사용해서 2,000개 정규난벡터들을 발생시킨다. 이 성분단위MH샘플들의 산점도가 그림 6.1.9에 그려져 있다. 그림 6.1.9에서 적색 점들은 MH샘플들이고, 흑색 실선들은 각  $k(=1,2,3,4)$ 에 대해서 식  $\mathbf{x}'\Sigma^{-1}\mathbf{x} = k$ 를 만족하는 등가곡선이다. ■

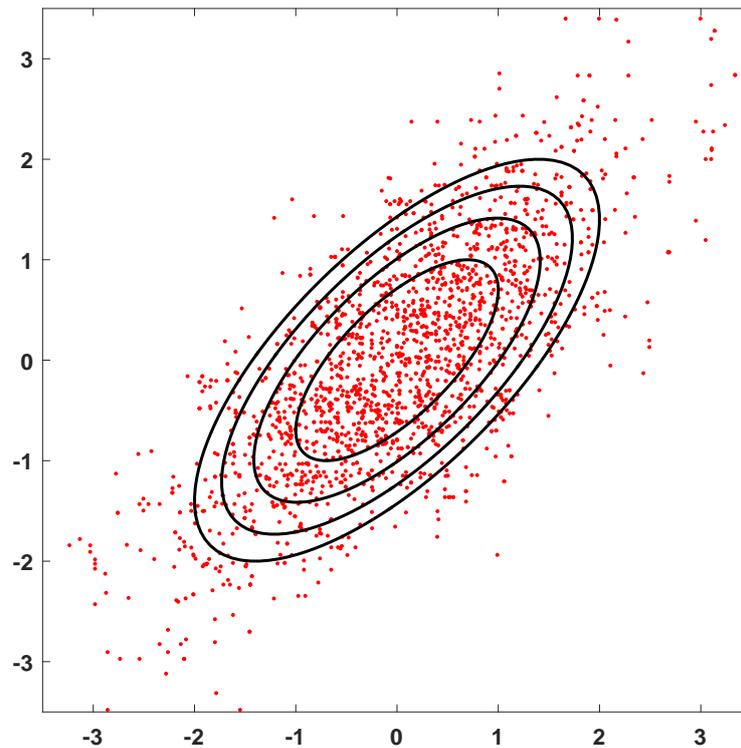


그림 6.1.9. 성분단위MH샘플러와 2변량 정규확률분포 1

**예제 6.1.11** 예제 6.1.10을 다시 살펴보자. 목표확률분포로부터 성분단위MH샘플들을 생성하기 위해서 다음 MATLAB 프로그램 ComponentMHsampler501.m을 실행해 보자.

```

1 % -----
2 % Filename: ComponentwiseMHsampler101.m

```

```

3 % Blockwise MH Sampler (BIVARIATE NORMAL)
4 % using MATLAB function mhsample
5 % Programmed by MathWorks
6 % -----
7 clear all, close all
8 rand('seed', (101^2+1)/2) % rand('seed', 12345);
9
10 % TARGET DISTRIBUTION
11 p = inline('mvnpdf(x,[0 0],[1 0.7 ; 0.7 1])','x');
12
13 nSim = 2000
14 nBurnin = 5000
15 nSamples = nSim + nBurnin;
16 propSigma = 1; % PROPOSAL VARIANCE
17 minn = [-3 -3];
18 maxx = [3 3];
19
20 % INITIALIZE COMPONENT-WISE SAMPLER
21 x = zeros(nSamples,2);
22 xCurrent(1) = randn;
23 xCurrent(2) = randn;
24 dims = 1:2; % INDICES INTO EACH DIMENSION
25 t = 1;
26 x(t,1) = xCurrent(1);
27 x(t,2) = xCurrent(2);
28
29 % RUN SAMPLER
30 while t < nSamples
31     t = t + 1;
32     for iD = 1:2 % LOOP OVER DIMENSIONS
33
34         % SAMPLE PROPOSAL
35         xStar = normrnd(xCurrent(:,iD), propSigma);
36
37         % NOTE: CORRECTION FACTOR c=1 BECAUSE
38         % N(mu,1) IS SYMMETRIC, NO NEED TO CALCULATE
39
40         % CALCULATE THE ACCEPTANCE PROBABILITY
41         pratio = p([xStar xCurrent(dims~=iD)]) / ...
42         p([xCurrent(1) xCurrent(2)]);
43         alpha = min([1, pratio]);
44
45         % ACCEPT OR REJECT?
46         u = rand;
47         if u < alpha
48             xCurrent(iD) = xStar;
49         end
50     end
51
52     % UPDATE SAMPLES
53     x(t,:) = xCurrent;
54 end
55
56 % Burn-In
57 x([1:nBurnin],:) = [];
58 sizex = size(x);
59 disp('size of x = '), sizex
60
61 % DISPLAY
62 nBins = 20;
63 bins1 = linspace(minn(1), maxx(1), nBins);

```

```

64 bins2 = linspace(minn(2), maxx(2), nBins);
65
66 % DISPLAY SAMPLED DISTRIBUTION
67 figure;
68 ax = subplot(121);
69 bins1 = linspace(minn(1), maxx(1), nBins);
70 bins2 = linspace(minn(2), maxx(2), nBins);
71 sampX = hist3(x, 'Edges', {bins1, bins2});
72 hist3(x, 'Edges', {bins1, bins2});
73 view(-15,40)
74
75 % COLOR HISTOGRAM BARS ACCORDING TO HEIGHT
76 colormap autumn
77 set(gca,'fontsize',11,'fontweigh','bold')
78 set(gcf,'renderer','opengl');
79 set(get(gca,'child'),'FaceColor','interp','CDataMode','auto');
80 xlabel('x_1'); ylabel('x_2'); zlabel('Frequency');
81 axis square
82 set(ax,'xTick',[minn(1),0,maxx(1)]);
83 set(ax,'yTick',[minn(2),0,maxx(2)]);
84 title('Sampled Distribution');
85
86 % DISPLAY ANALYTIC DENSITY
87 ax = subplot(122);
88 [x1 ,x2] = meshgrid(bins1,bins2);
89 probX = p([x1(:), x2(:)]);
90 probX = reshape(probX ,nBins, nBins);
91 surf(probX); axis xy
92 view(-15,40)
93 xlabel('x_1'); ylabel('x_2'); zlabel('p({\bf x})');
94 colormap autumn
95 axis square
96 set(gca,'fontsize',11,'fontweigh','bold')
97 set(ax,'xTick',[1,round(nBins/2),nBins]);
98 set(ax,'xTickLabel',[minn(1),0,maxx(1)]);
99 set(ax,'yTick',[1,round(nBins/2),nBins]);
100 set(ax,'yTickLabel',[minn(2),0,maxx(2)]);
101 title('Analytic Distribution')
102
103 saveas(gcf,'ComponentwiseMHsampler501','eps')
104 % End of program
105 % -----

```

이 MATLAB 프로그램 ComponentwiseMHsampler501.m을 실행하면, 성분단위MH샘플러를 사용해서 2,000개 정규난벡터들을 발생시킨다. 이 성분단위MH샘플들의 히스토그램을 그림 6.1.10의 좌측 그래프에 그려져 있고, 결합확률밀도함수가 우측 그래프에 그려져 있다. 좌측의 히스토그램이 우측의 결합확률밀도함수의 그래프를 잘 반영하는 것처럼 보인다. ■

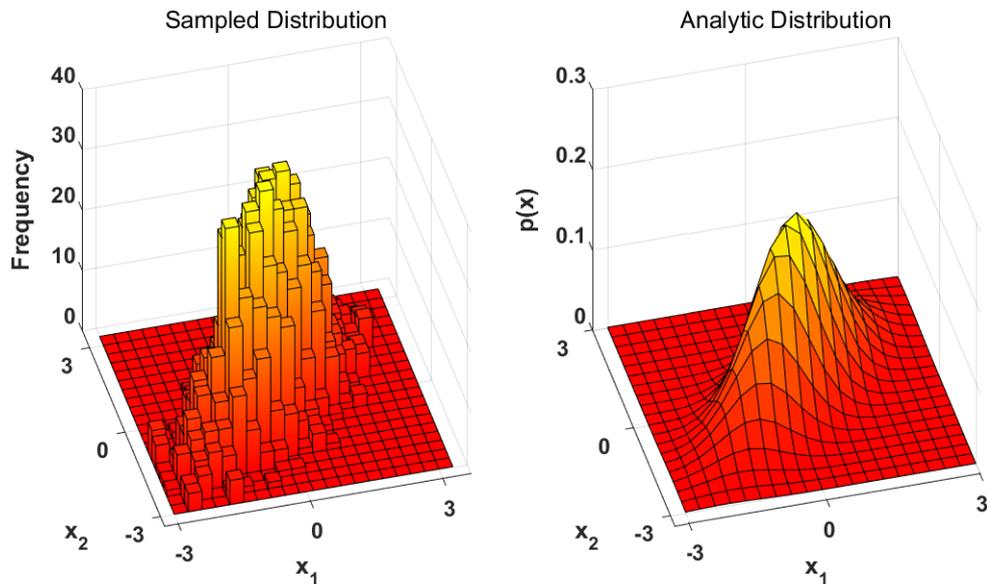


그림 6.1.10. 성분단위MH샘플러와 2변량 정규확률분포 2

### 6.1.3 Metropolis 샘플러

Metropolis 샘플러는 Metropolis 등 [76]이 제시한 방법이다. 이 방법에서는 오직 대칭확률분포만을 제안분포로 사용한다. 즉, 각  $[x, y]$ 에 대해서 다음 식을 만족하는  $q$ 를 사용한다.

$$q_{y,x} = q_{x,y} \quad (6.1.4)$$

식 (6.1.4)를 식 (6.1.1)에 적용하면, 다음 식을 얻는다.

$$\alpha(x_t, y) = \min \left\{ \frac{\pi(y)}{\pi(x_t)}, 1 \right\} \quad (6.1.5)$$

즉, MH채택확률은 제안분포에 의존하지 않는다. 식 (6.1.5)의 확률  $\alpha(x_t, y)$ 로 후보점  $y$ 가 시점  $t+1$ 에서 상태  $x_{t+1}$ 로 채택된다. 만약 후보점  $y$ 가 채택되지 않으면, Markov체인은 움직이지 않는다. 즉,  $x_{t+1} = x_t$ 이다. Metropolis 샘플러를 다음과 같이 요약할 수 있다.

**알고리즘 6.1.4: Metropolis샘플러**

(1단계) 적절한 초기값  $x_0$  를 선택한다.

(2단계) 제안분포의 조건부확률밀도함수  $q_{x_t, \cdot}$  에서 후보점  $y$  를 생성한다.

(3단계) 일양확률분포  $Uniform(0, 1)$  에서 일양난수  $u_t$  를 생성한다.

(4단계) 만약 식  $u_t \leq \alpha(x_t, y)$  가 성립하면  $x_{t+1} = y$  라 하고, 그렇지 않으면  $x_{t+1} = x_t$  라 한다. 여기서 MH채택확률  $\alpha(x_t, y)$  는 다음과 같다.

$$\alpha(x_t, y) = \min \left\{ \frac{\pi(y)}{\pi(x_t)}, 1 \right\}$$

(5단계) 주어진 번인기간  $[1, b]$  와 샘플수  $n$  에 대해서, 만약  $t + 1 > b + n$  이면, 이 알고리즘을 멈춘다. 그렇지 않으면,  $t$  를  $t + 1$  으로 바꾼 다음 제2단계로 돌아간다.

만약 식  $\pi(y) \geq \pi(x_t)$  이 성립하면,  $\alpha(x_t, y)$  는 1이 되고 Markov 체인은  $y$  로 움직인다. 이것은 목표확률밀도함수가 증가하는 움직임은 항상 채택된다는 것을 의미한다. 반대로, 감소하는 움직임은 확률  $\alpha(x_t, y) = \pi(y)/\pi(x_t)$  로 채택된다. 이러한 개념은 시뮬레이티드어닐링(simulated annealing) 과 같은 최적화 기법의 토대가 된다.

**예제 6.1.12** Metropolis샘플러를 사용해서, 표준정규확률분포에서 Metropolis샘플들을 생성하기로 하자. 즉, 목표확률분포는 표준정규확률분포이다. 다음과 같은 제안분포의 조건부확률밀도함수를 사용하자.

$$q_{x_t, y} = \frac{1}{2A} 1_{(-A, A)}(y - x_t) \quad (1)$$

이 제안분포는 조건 (6.1.4) 를 만족한다. 따라서, MH샘플러는 Metropolis샘플러이다.

Metropolis샘플러를 적용하기 위해서, 다음 R프로그램 MetropolisSampler101R.R을 실행해 보자.

```

1 # -----
2 # Filename: MetropolisSampler101R.R
3 # Metropolis Sampler
4 # Programmed by CBS
5 # -----
6 # Metropolis Sampler
7 MetropolisNormal <- function(nSim, A){

```

```

8         nTot <- 2*nSim
9         xvec <- vector("numeric", nSim)
10        x <- 0
11        xvec[1] <- x
12        for (ii in 2:nTot){
13            z <- runif(1,-A,A)
14            candidate <- x + z
15            alphaa <- min(1,dnorm(candidate)/dnorm(x))
16            if ( runif(1)<alphaa ) {x <- candidate}
17            xvec[ii] <- x
18        }
19        xMetropolis <- xvec[(nSim+1):(2*nSim)]
20    }
21
22    # Generate Metroipois Sampler
23
24
25    set.seed(1)
26    nSim <- 10000
27    IntervallLength = 2
28    MetNormal <- MetropolisNormal(nSim,IntervallLength/2)
29    head(MetNormal)
30
31    # Plotting
32    # install.packages("ggplot2")
33    library(ggplot2)
34    # install.packages("grid")
35    library(grid)
36    setEPS()
37    plot.new()
38    postscript('MetropolisSampler101R.eps') # Start to save figure
39    timee = c(1:nSim)
40    MetropolisData <- data.frame(timee,MetNormal)
41    MetropolisData2 <- MetropolisData[(nSim-50):nSim,]
42    plot1 <- ggplot(MetropolisData2, aes(timee,MetNormal)) +
43        geom_line(col="red",lwd=1.2) + xlab("time") +
44        geom_point(lwd=2)
45    plot2 <- ggplot(MetropolisData, aes(x=MetNormal)) +
46        geom_histogram(aes(y=..density..),
47            col="dark blue",fill="white",lwd=1.2) +
48            stat_function(fun=dnorm, colour="red",lwd=1.2)
49    pushViewport(viewport(layout=grid.layout(2,1)))
50    print(plot1, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
51    print(plot2, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
52    dev.off() # End to save figure
53    # -----

```

이 R 프로그램 MetropolisSampler101R.R을 실행하면, Metropolis샘플들 10,000개가 생성된다. 이 Metropolis샘플들의 그래프들이 그림 6.1.11에 수록되어 있다. 그림 6.1.11의 상단 그래프는 마지막 50개 Metropolis샘플들의 시계열산포도이고, 하단 그래프는 10,000개 Metropolis샘플들의 히스토그램이다. ■

상태들의 집합이 이산시간형인 경우를 사용해서, MH샘플러를 생각해 보자. 음이 아닌 상수들의 집합  $\{c_1, c_2, \dots\}$ 를 사용해서 다음과 같이 확률밀도함수  $\{\pi(1), \pi(2), \dots\}$ 를 정의

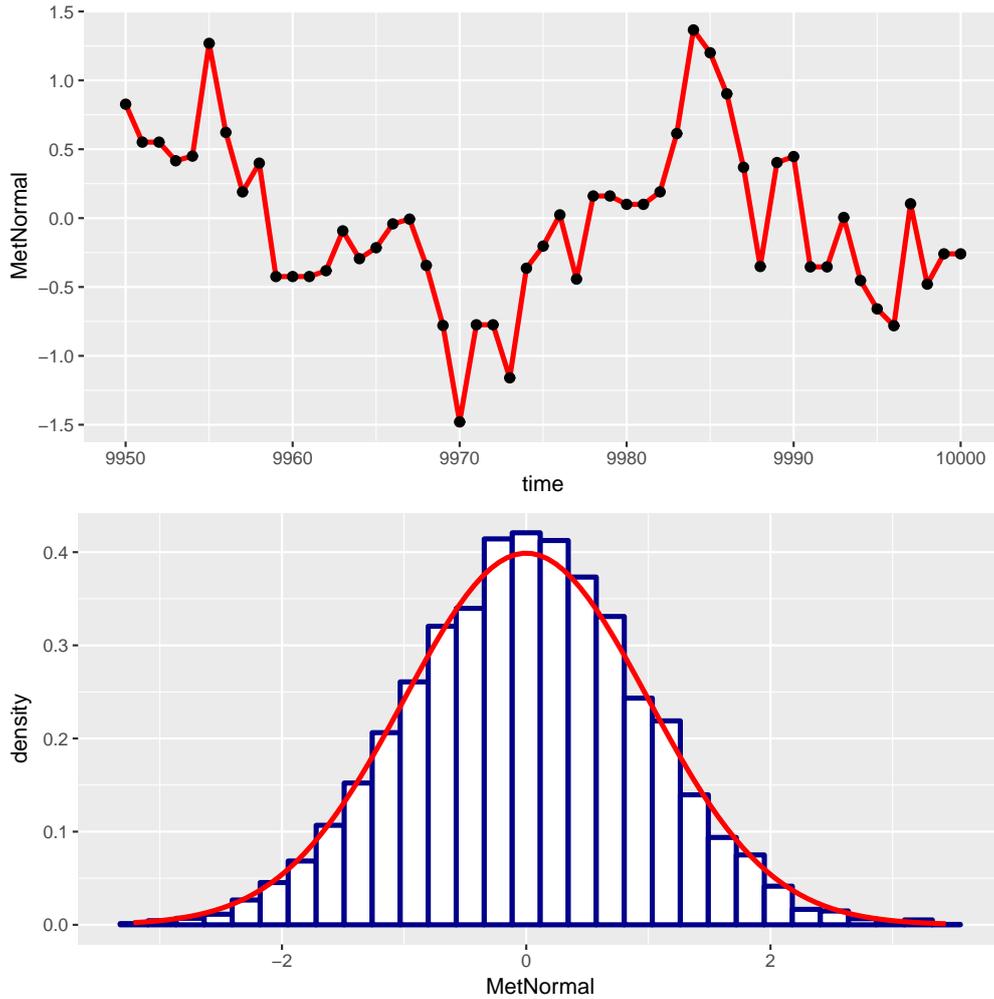


그림 6.1.11. Metropolis샘플러와 정규확률분포

하자.

$$\pi(i) \doteq \frac{c_i}{\sum_{j=1}^{\infty} c_j}, \quad (i = 1, 2, \dots) \quad (6.1.6)$$

여기서  $C \doteq \sum_{j=1}^{\infty} c_j$ 를 간단하게 구할 수는 없다고 가정하자. 이것은 사후확률분포를 표준화하는 적분이 일반적으로 쉽지 않은 것과 유사하다. 이러한 경우에는, MH샘플러를 사용해서  $\{c_1, c_2, \dots\}$ 로부터 쉽게 샘플을 발생시킬 수 있다.

**예제 6.1.13** Metropolis샘플러를 사용해서, 다음과 같은 Poisson 확률밀도함수에서 Metropolis샘플들을 생성하기로 하자.

$$\pi(k) = \frac{1}{k!} \lambda^k \exp(-\lambda), \quad (k = 0, 1, \dots) \quad (1)$$

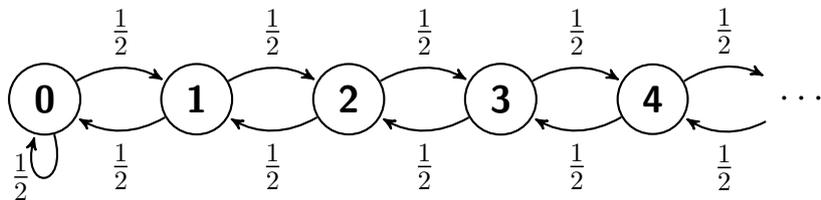
제안분포의 조건부확률밀도함수는 다음과 같다.

$$q_{x_t,y} = \begin{cases} \frac{1}{2}1_{\{x_t\}}(y) + \frac{1}{2}1_{\{x_t+1\}}(y), & (x_t = 0) \\ \frac{1}{2}1_{\{x_t-1\}}(y) + \frac{1}{2}1_{\{x_t+1\}}(y), & (x_t > 0) \end{cases} \quad (2)$$

이 제안분포에 해당하는 추이확률행렬은 다음과 같다.

$$Q = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & \dots \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ \vdots \end{matrix} & \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \dots \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & \dots \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \dots \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & \dots \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} & \end{matrix} \quad (3)$$

식 (3)의 추이확률행렬은 대칭이다. 따라서, 제안분포는 조건 (6.1.4)를 만족한다. 즉, 이 MH 샘플러는 Metropolis 샘플러이다. 이 Markov 체인의 상태추이도는 다음과 같다.



이 Metropolis 샘플러를 사용해서 Metropolis 샘플들을 생성하기 위한 알고리즘은 다음과 같다.

(1단계) 적절한 초기값  $x_0$ 를 선택한다.

(2단계) 일양확률분포  $Uniform(0, 1)$ 에서 일양난수  $u_1$ 을 발생시킨다. 추이확률행렬  $Q$ 를

이용해서, 다음과 같이 후보점  $y$ 를 선택한다.

$$y = \begin{cases} 0, & (x_t = 0 \text{ and } u_1 < \frac{1}{2}) \\ 1, & (x_t = 0 \text{ and } u_1 \geq \frac{1}{2}) \\ x_t - 1, & (x_t \neq 0 \text{ and } u_1 < \frac{1}{2}) \\ x_t + 1, & (x_t \neq 0 \text{ and } u_1 \geq \frac{1}{2}) \end{cases}$$

(3단계) MH채택확률  $\alpha(x_t, y)$ 는 다음과 같다.

$$\alpha(x_t, y) = \begin{cases} 1, & (x_t = 0 \text{ and } y = 0) \\ \frac{x_t}{\lambda}, & (y = x_t - 1 \text{ and } x_t > 0) \\ \frac{\lambda}{y}, & (y = x_t + 1) \end{cases}$$

(4단계) 일양확률분포  $Uniform(0, 1)$ 에서 일양난수  $u_2$ 를 발생시킨다. 새로운 Metropolis 샘플  $x_{t+1}$ 은 다음과 같다.

$$x_{t+1} = \begin{cases} y, & (\alpha(x_t, y) \geq 1) \\ y, & (u_2 < \alpha(x_t, y) < 1) \\ x_t, & (\alpha(x_t, y) < 1 \text{ and } \alpha(x_t, y) \leq u_2) \end{cases}$$

(5단계) 만약  $t$ 가 원하는 샘플개수를 초과하면, 이 알고리즘을 멈춘다. 그렇지 않으면,  $t$ 를  $t + 1$ 으로 증가시킨 뒤, 제2단계로 돌아간다.

이 알고리즘을 사용해서 Metropolis샘플들을 발생시키기 위해, 다음 R프로그램 Metropolis-Sampler102R.R을 실행해 보자.

```

1 # -----
2 # Filename: MetropolisSampler102R.R
3 # Metropolis Sampler
4 # Programmed by CBS
5 # -----
6 # Target function
7 pii = function(x, lambda){1/gamma(xx+1)*lambda^xx*exp(-lambda)}
8
9 # Metropois Algorithm
10 MetropolisPoisson = function(lambdaa, x0, nSim){
11     nBurnin <- nSim

```

```

12     nTot <- nBurnin + nSim
13     set.seed(5102, "Mersenne-Twister")
14     xOut = seq(1, nTot, 1)
15     x <- x0
16     for(kk in 1:nTot){
17         # Generate y using Q
18         if ( x==0 ){
19             if ( runif(1)<.5 ){ y <- 0 }
20             else { y <- 1 }
21         }
22         else{ if ( runif(1)<.5 ){ y <- x-1 }
23             else { y <- x+1 }
24         }
25         # Generate alpha
26         if ( x==0 && y==0 ) {alphaa <- 1}
27         if ( y-x+1==0 ) { alphaa <- x/lambdaa }
28         if ( y-x-1==0 ) { alphaa <- lambdaa/y }
29         # New x
30         if ( alphaa >= 1 ){ xNew <- y }
31         else { if( runif(1)<alphaa ) { xNew <- y }
32             else { xNew <- x }
33         }
34         x = xNew
35         xOut[kk] = x
36     }
37     xOutput <- xOut[(nBurnin+1):nTot]
38     return(xOutput)
39 }
40
41 # Practice Metropolis algorithm
42 nSim <- 30000
43 x0 <- 2
44 MS1 <- MetropolisPoisson(1,x0,nSim)
45 MS2 <- MetropolisPoisson(2,x0,nSim)
46 MS3 <- MetropolisPoisson(3,x0,nSim)
47 MS6 <- MetropolisPoisson(6,x0,nSim)
48
49 # Plotting
50 # intall.packages("Rcpp")
51 # intall.packages("ggplot2")
52 library(ggplot2)
53 # install.packages("grid")
54 library(grid)
55 setEPS()
56 plot.new()
57 postscript('MetropolisSampler102bR.eps') # Start to save figure
58 timee = c(1:nSim)
59 MetropolisData <- data.frame(timee, MS1, MS2, MS3, MS6)
60 head(MetropolisData)
61 tail(MetropolisData)
62 plot1 <- ggplot(MetropolisData, aes(x=MS1)) +
63     geom_histogram(aes(y = ..density..), bins=30,
64                   col="dark blue", fill="red") +
65     xlab(expression(x)) +
66     xlim(0, 12)
67 plot2 <- ggplot(MetropolisData, aes(x=MS2)) +
68     xlab(expression(x)) +
69     geom_histogram(aes(y = ..density..), bins=30,
70                   col="dark blue", fill="green") +
71     xlim(0, 12)
72 plot3 <- ggplot(MetropolisData, aes(x=MS3)) +

```

```

73     xlab(expression(x)) +
74     geom_histogram(aes(y = ..density..), bins=30,
75                   col="dark blue",fill="blue") +
76     xlim(0,12)
77 plot6 <- ggplot(MetropolisData, aes(x=MS6)) +
78     xlab(expression(x)) +
79     geom_histogram(aes(y = ..density..), bins=30,
80                   col="dark blue",fill="yellow") +
81     xlim(0,12)
82 pushViewport(viewport(layout = grid.layout(2,2)))
83 print(plot1, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
84 print(plot2, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
85 print(plot3, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
86 print(plot6, vp = viewport(layout.pos.row = 2, layout.pos.col = 2))
87 dev.off()           # End to save figure
88
89 plot.new()
90 postscript('MetropolisSampler102aR.eps') # Start to save figure
91 MetropolisData2 <- MetropolisData[(nSim-29):nSim,]
92 ggplot(MetropolisData2, aes(x=timee, y=MS1)) +
93     geom_line(lwd=1.1, color="red", lty="solid") +
94     geom_line(aes(x=timee, y=MS2), lwd=1.1, color="green", lty="dashed") +
95     geom_line(aes(x=timee, y=MS3), lwd=1.1, color="blue", lty="longdash") +
96     geom_line(aes(x=timee, y=MS6), lwd=1.1, color="dark red", lty="twodash")
97 dev.off()           # End to save figure
98 # -----

```

이 R 프로그램 MetropolisSampler102R.R을 실행하면, 강도모수 (intensity)  $\lambda$ 가 각각 1, 2, 3, 6인 Poisson 확률분포에서 Metropolis 샘플들을 30,000개씩 발생시킨다. 이렇게 발생된 마지막 30개 샘플들의 시계열산점도가 그림 6.1.12에 그려져 있다. 그림 6.1.12에서 적색 실선은 강도모수가  $\lambda = 1$ 인 Poisson 확률과정의 시계열산점도를, 녹색 대시선 (dashed line)은 강도모수가  $\lambda = 2$ 인 Poisson 확률과정의 시계열산점도를, 청색 긴점선 (long dash line)은 강도모수가  $\lambda = 3$ 인 Poisson 확률과정의 시계열산점도를, 암적색 이배점선 (two dash line)은 강도모수가  $\lambda = 6$ 인 Poisson 확률과정의 시계열산점도를 나타낸다.

이 R 프로그램 MetropolisSampler102R.R을 실행하면, 그림 6.1.13의 히스토그램들이 그려진다. 그림 6.1.13에서 적색 히스토그램은 강도모수가  $\lambda = 1$ 인 Poisson 확률과정의 Metropolis 샘플들을 그린 것이고, 녹색 히스토그램은 강도모수가  $\lambda = 2$ 인 Poisson 확률과정의 Metropolis 샘플들을 그린 것이고, 청색 히스토그램은 강도모수가  $\lambda = 3$ 인 Poisson 확률과정의 Metropolis 샘플들을 그린 것이고, 황색 히스토그램은 강도모수가  $\lambda = 6$ 인 Poisson 확률과정의 Metropolis 샘플들을 각각 그린 것이다. ■

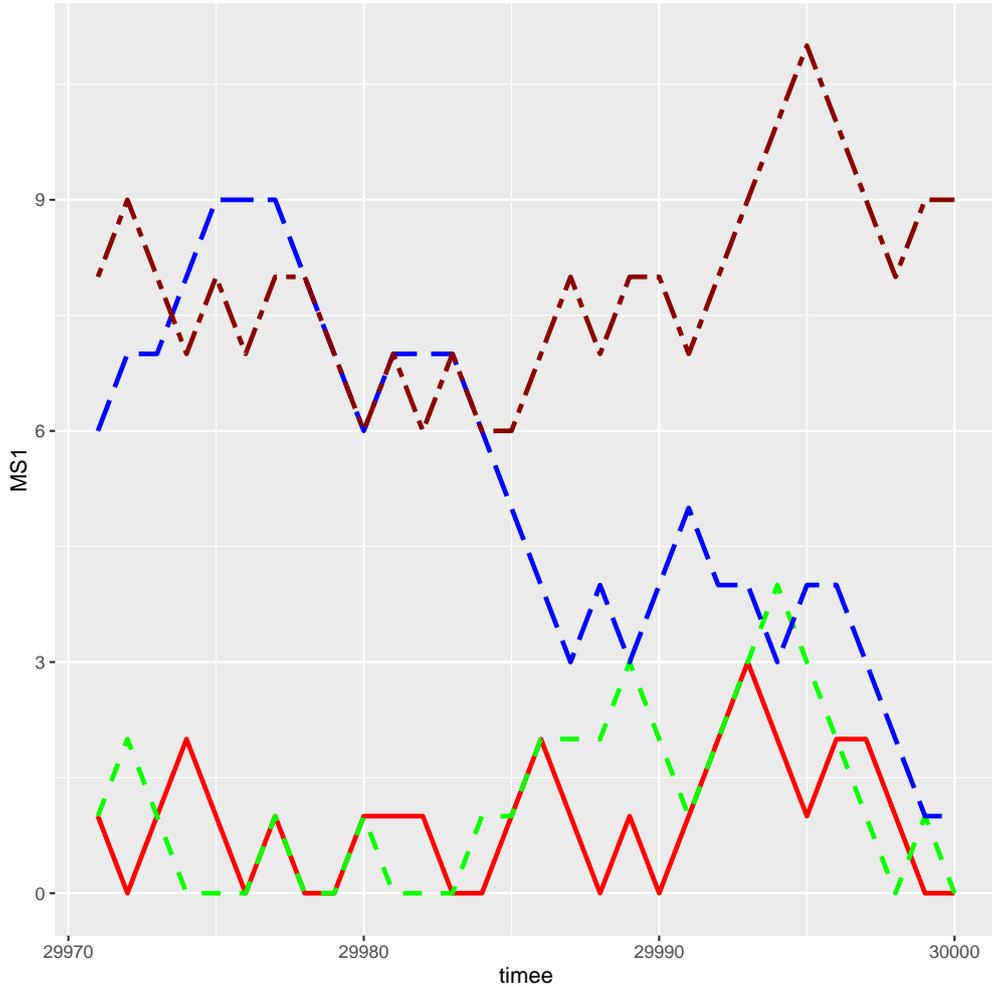


그림 6.1.12. Metropolis샘플러와 Poisson샘플들

### 6.1.4 확률보행Metropolis샘플러

MH샘플러를 사용할 때 가장 간단한 형태의 제안분포는 확률보행에서 유도되는 것이다. 만약 제안분포의 조건부확률밀도함수가 다음과 같다고 하자.

$$q_{x,y} = f(|x - y|) \tag{6.1.7}$$

조건 (6.1.7)이 만족되면,  $q_{x,y}$ 는 대칭이다. 따라서 이 샘플러는 Metropolis샘플러이다. 이러한 Metropolis샘플러를 확률보행Metropolis샘플러 (random-walk Metropolis sampler)라 부른다. 식 (6.1.7)에서 알 수 있듯이,  $z \doteq y - x_t$ 는 제안분포를 따르는 확률변수이다. 보통  $z$ 는  $Uniform(-A, A)$ 를 따른다고 가정하고,  $A$ 로는 작은 양수를 사용한다.

**예제 6.1.14** 확률보행Metropolis샘플러를 사용해서, 다음과 같은 표준정규확률밀도함수

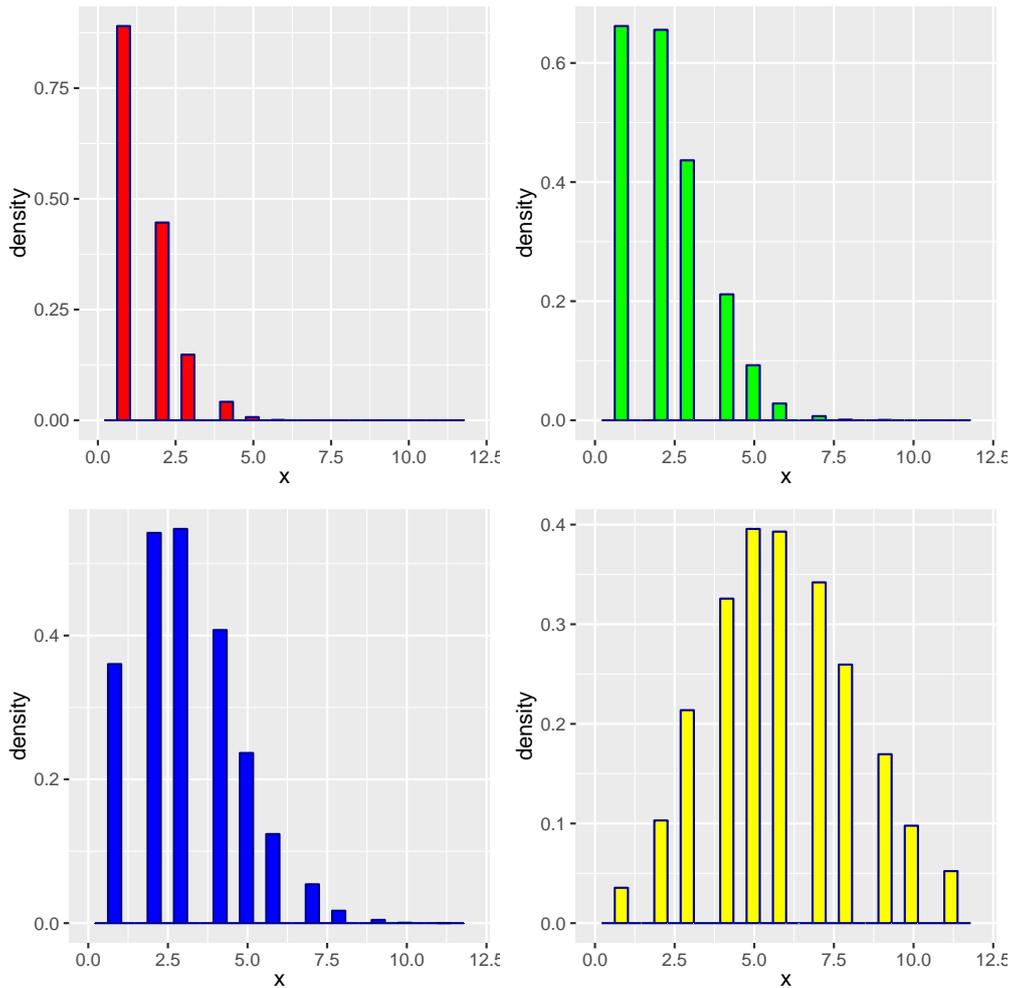


그림 6.1.13. Metropolis샘플러와 Poisson 확률밀도함수

로부터 난수들을 생성하기로 하자.

$$\pi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), \quad (-\infty < x < \infty) \quad (1)$$

이 예제에서는 올바른 제안분포를 선택하는 것이 중요함을 강조하고자 한다.

제안분포로 표준편차가 각각  $\sigma = 0.5$ ,  $\sigma = 0.1$ , 그리고  $\sigma = 10.0$ 인 정규확률분포를 사용하자. 각 제안분포의 조건부확률밀도함수는 다음과 같다.

$$q_{x_t, y} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{[y - x_t]^2}{\sigma^2}\right) \quad (2)$$

식 (2)에 해당하는 제안분포는 조건 (6.1.7)을 만족한다. 즉, 이 MH샘플러는 확률보행 Metropolis샘플러이다.

이 확률보행Metropolis샘플러를 적용하기 위해서, 다음 MATLAB프로그램 RWmetropolis101.m을 실행해 보자.

```

1 % -----
2 % Filename: RWmetropolis101.m
3 % Normal Random Numbers generated by Random Walk Metropolis Sampler
4 % Programmed by CBS
5 % -----
6 clear all, close all
7 % Variances for Proposal Distributions
8 sigmaa(1) = 0.5; sigmaa(2) = 0.1; sigmaa(3) = 10.0;
9 % Generate 500 samples from MC.
10 n = 1000;
11 x = zeros(n,3);
12 % Starting values
13 x(1,:) = [ -10 0 0 ];
14 rand('state',0); % Use twister or state for seed.
15 % Run the MC
16 for runn = 1:3;
17     for ii = 2:n;
18         y = randn(1)*sigmaa(runn) + x(ii-1,runn);
19         u = rand(1);
20         alpha = normpdf(y,0,1)/normpdf(x(ii-1,runn),0,1);
21         if u <= alpha
22             x(ii,runn) = y;
23         else
24             x(ii,runn) = x(ii-1,runn);
25         end
26     end
27     % Plotting
28     subplot(3,1,runn)
29     plot(1:n,x(:,runn),'k',[1 n],[-2.6 -2.6], 'r:', ...
30          [1 n],[2.6 2.6], 'r:', 'LineWidth',2);
31     set(gca,'fontsize',11,'fontweigh','bold')
32 end
33 saveas(gcf,'RWmetropolis101','eps')
34 save('RWmetropolis101.txt','x','-ascii')
35 % End of Program
36 % -----

```

이 MATLAB프로그램 RWmetropolis101.m을 실행하면, 각 제안분포에 대한 확률보행 Metropolis샘플러를 적용해서 Metropolis샘플링을 한다. 발생된 Metropolis샘플들의 시계열산점도가 그림 6.1.14에 그려져 있다. 그림 6.1.14의 상단 그래프는 제안분포가  $\sigma = 0.5$ 인 정규확률분포인 경우, 가운데 그래프는 제안분포가  $\sigma = 0.1$ 인 정규확률분포인 경우, 그리고 하단 그래프는 제안분포의 표준편차가  $\sigma = 10.0$ 인 정규확률분포인 경우이다. 각 그래프에는 수평선들  $\pm 2.6$ 이 그려져 있다. 만약 Metropolis샘플들이 잘 생성되었다면, Metropolis샘플들의 99%가 이 신뢰구간 안에 포함되어야 한다.

첫 번째 Metropolis샘플열은 표준편차가  $\sigma = 0.5$ 인 제안분포로부터 발생된 것이다. 이 Metropolis샘플열은 초기값이 중심에서 아주 멀리 떨어져 있음에도 불구하고 아주 빨리 신뢰구간 안으로 흡수되고, 또한 난수성을 보인다. 두 번째 Metropolis샘플열은 표준편차가

$\sigma = 0.1$ 인 제안분포로부터 발생된 것이다. 이 제안분포는 중심부분에 확률질량이 모여있다. 따라서, 이 Metropolis샘플열은 천천히 변화하고, 제 450번째 Metropolis샘플까지도 목표확률밀도함수로 수렴하는 것처럼 보이지는 않는다. 세 번째 Metropolis샘플열은 표준편차가  $\sigma = 10.0$ 인 제안분포로부터 발생된 것이다. 즉,  $z = y - x_t$ 의 절대값이 큰 경우가 많다. 이 경우에, 작은 MH채택확률  $\alpha(x_t, y)$ 로 후보점  $y$ 가 시점  $t+1$ 에서 상태  $x_{t+1}$ 로 채택된다. 즉, 이 Metropolis샘플열에는 변화가 심한 부분이 있다. 반면에,  $z = y - x_t$ 의 절대값이 작은 경우에는 높은 MH채택확률  $\alpha(x_t, y)$ 로 후보점  $y$ 가 시점  $t+1$ 에서  $x_{t+1}$ 로 채택된다. 그러나,  $|y - x_t|$ 가 작으므로,  $x_{t+1}$ 과  $x_t$ 가 비슷하다. 즉, Metropolis샘플열에 변화가 나타나지 않는 부분도 많다. 본저자는 이렇게 변화가 없는 부분은 변화가 심한 부분과 비교되어 더 변화가 없는 것처럼 보이는 것이 아닌가 생각한다. ■

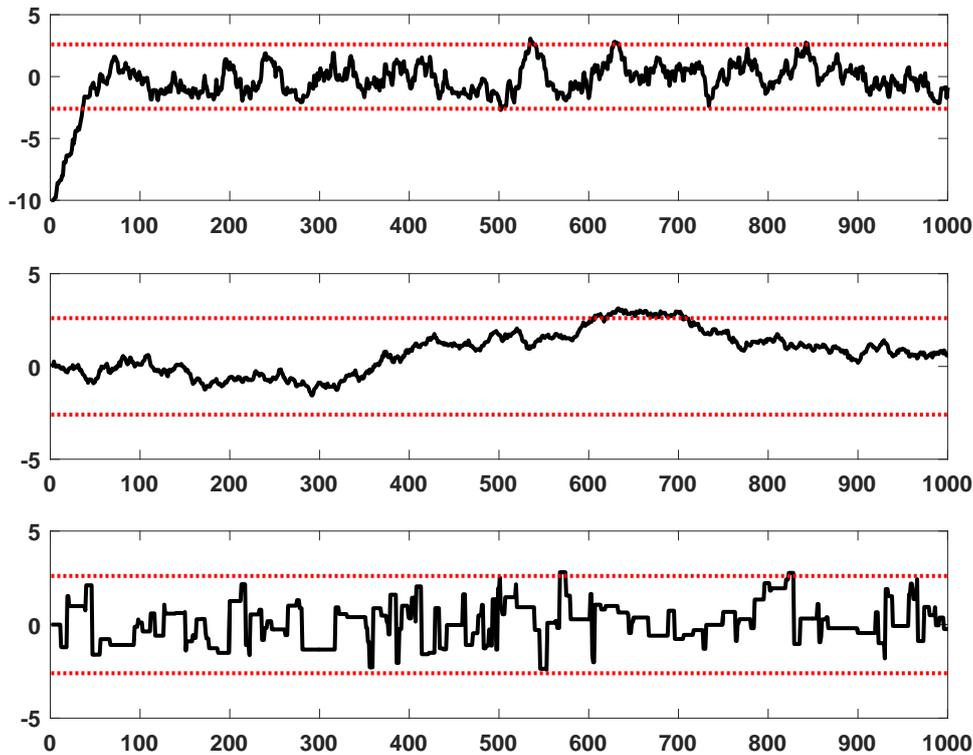


그림 6.1.14. 확률보행Metropolis샘플러와 정규확률분포

**예제 6.1.15** 확률보행Metropolis샘플러를 사용해서, 다음과 같은 표준정규확률밀도함수로부터 난수들을 생성하기로 하자.

$$\pi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), \quad (-\infty < x < \infty) \quad (1)$$

제안분포로는 지지대가  $(-A, A)$  인 일양확률분포를 사용하자. 즉, 이 제안분포의 조건부확률 밀도함수는 다음과 같다.

$$q_{x_t, y} = \frac{1}{2A} 1_{(-A, A)}(|y - x_t|) \quad (2)$$

이 제안분포는 조건 (6.1.7) 을 만족한다.

MATLAB 함수 `mhsample.m` 을 사용해서 확률보행 Metropolis 샘플들을 생성하기로 하고, 다음 MATLAB 프로그램 `RWmetropolis102.m` 을 실행해 보자.

```

1 % -----
2 % Filename: RWmetropolis102.m
3 % Normal Random Numbers generated by Random Walk Metropolis Sampler
4 % using MATLAB function mhsample
5 % Programmed by MathWorks
6 % -----
7 clear all, close all
8 rng default % For reproducibility
9 A = 1;
10 pdfx = @(x) normpdf(x);
11 proppdf = @(x,y) unifpdf(y-x,-A,A);
12 proprnd = @(x) x + rand*2*A - A;
13 nsamples = 15000;
14 x = mhsample(1, nsamples, 'pdf', pdfx, 'proprnd', proprnd, 'symmetric', 1);
15 % Plot the sample data.
16 figure;
17 h = histfit(x, 50);
18 set(gca, 'fontsize', 11, 'fontweigh', 'bold');
19 h(1).FaceColor = [.8 .8 .8];
20 saveas(gcf, 'RWmetropolis102', 'eps');
21 save('RWmetropolis102.txt', 'x', '-ascii');
22 % End of Program
23 % -----

```

MATLAB 함수 `mhsample.m` 의 기초적인 사용법은 다음과 같다.

```

smp1 = mhsample(start, nsamples, 'pdf', pdf, 'proppdf', proppdf, ...
               'proprnd', proprnd, 'symmetric', sym)

```

첫 번째 입력모수 `start` 는 MH 샘플의 초기값  $x_0$  이고, 두 번째 입력모수 `nsamples` 에는 발생시킬 MH 샘플의 개수를 지정하고, 세 번째 입력모수 `pdf` 에는 목표확률밀도함수 `pdf` 를 지정하고, 네 번째 입력모수 `proppdf` 에는 제안분포 `proppdf` 를 지정하고, 다섯 번째 입력모수 `proprnd` 에는 제안분포에서 생성된 난수를 지정하고, 여섯 번째 입력모수 `symmetric` 에는 제안분포의 대칭성을 나타내는 `sym` 을 할당한다. 만약 `sym` 이 1 이면, 제안분포는 대칭이고, 0 이면 비대칭이며, 디폴트는 비대칭이다.

이 MATLAB 프로그램 `RWmetropolis102.m` 을 실행하면, 다음과 같은 제안분포의 조건부

확률밀도함수를 사용해서 15,000개 표준정규난수들 생성한다.

$$q_{x_t, y} = \frac{1}{2} 1_{(-1, 1)}(|y - x_t|) \quad (3)$$

이 확률보행Metropolis샘플들의 히스토그램을 그림 6.1.15에 그려져 있다. ■

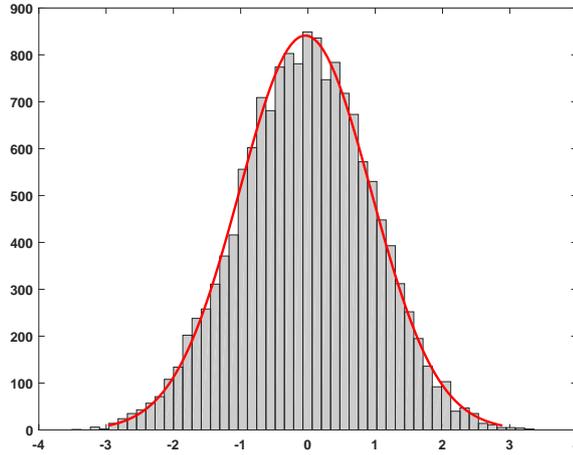


그림 6.1.15. 확률보행Metropolis샘플러와 정규확률분포

단변량 난수 경우와 마찬가지로 다변량 난벡터 후보점  $\mathbf{y}$ 를 찾는 가장 간단한 방법은 확률보행과정을 이용하는 것이다. 즉, 다음과 같이 후보점  $\mathbf{y}$ 를 발생시킨다.

$$\mathbf{y} = \mathbf{x}^{(j-1)} + \mathbf{z}, \quad \mathbf{z} \stackrel{d}{\sim} f(\mathbf{z}) \quad (6.1.8)$$

이 경우에 제안분포의 조건부결합확률밀도함수는 다음과 같다.

$$q(\mathbf{x}^{(j-1)}, \mathbf{y}) = f(\mathbf{y} - \mathbf{x}^{(j-1)}) \quad (6.1.9)$$

확률벡터  $\mathbf{z}$ 의 확률분포로는 평균벡터가  $\mathbf{0}$ 인 다변량정규확률분포나 다변량  $t$ 확률분포가 자주 이용된다. 이 경우에는 분산을 어떻게 선택하느냐가 문제이다. 만약 분산들을 작게 잡으면, 상태공간의 일부에서만 후보점을 선택한다. 즉, 상태공간 전체에서 샘플링이 진행되지 않는다. 반면에, 분산들을 크게 잡으면, 목표확률분포의 밀도가 낮은 점이 후보점  $\mathbf{y}$ 로 제안되기 쉬워서, 채택률이 낮아진다. 이 두 경우 모두 MH샘플열  $\{\mathbf{x}^{(j)}\}$ 의 자기상관이 커지고 따라서 비효율적이다.

### 6.1.5 독립샘플러

독립샘플러 (independence sampler)는 Tierney [95]에 의해 제안되었다. 독립샘플러에서 제안분포의 조건부확률밀도함수  $q_{x,y}$ 는  $x$ 에 의존하지 않는 비조건부확률밀도함수  $q_y$ 이다. 따라서, 독립샘플러에 의해 생성된 현재시점의 샘플은 과거의 샘플들로부터 독립적이다. 식 (6.1.1)에서 알 수 있듯이, MH채택확률은 다음과 같다.

$$\alpha(x_t, y) = \min \left\{ \frac{\pi(y)q_{x_t}}{\pi(x_t)q_y}, 1 \right\} \quad (6.1.10)$$

만약  $w(x) = \pi(x)/q_x$ 로 표기하면, 식 (6.1.10)을 다음과 같이 쓸 수 있다.

$$\alpha(x_t, y) = \min \left\{ \frac{w(y)}{w(x_t)}, 1 \right\} \quad (6.1.11)$$

여기서  $w$ 는 목표확률밀도함수  $\pi$ 와 제안분포의 확률밀도함수  $q$ 의 비(ratio)이므로, 제안분포의 확률밀도함수에 비해 목표확률밀도함수가 상대적으로 큰 점이 선택되기 쉽다. 독립샘플러를 요약하면, 다음과 같다.

#### 알고리즘 6.1.5: 독립샘플러

- (1단계) 적절한 초기값  $x_0$ 를 선택한다.
- (2단계) 제안분포의 확률밀도함수  $q_{x_t}$ 에서 후보점  $y$ 를 생성한다.
- (3단계) 일양확률분포  $Uniform(0, 1)$ 에서 일양난수  $u_t$ 를 생성한다.
- (4단계) 만약 식  $u_t \leq \alpha(x_t, y)$ 가 성립하면,  $x_{t+1} = y$ 라 하고, 그렇지 않으면  $x_{t+1} = x_t$ 라 한다. 여기서  $\alpha(x_t, y)$ 는 다음과 같다.

$$\alpha(x_t, y) = \min \left\{ \frac{\pi(y)q_{x_t}}{\pi(x_t)q_y}, 1 \right\}$$

- (5단계) 주어진 변인기간  $[1, b]$ 와 샘플수  $n$ 에 대해서, 만약  $t + 1 > b + n$ 이면, 이 알고리즘을 멈춘다. 그렇지 않으면,  $t$ 를  $t + 1$ 으로 바꾼 다음 제2단계로 돌아간다.

일반적으로 독립샘플러는 제안분포의 조건부확률밀도함수  $q$ 가 목표확률밀도함수  $\pi(\cdot)$ 와 매우 유사한 경우에만 사용한다. Gilks 등 [48]은 제안분포의 꼬리가 목표확률분포의 꼬리보다

더 두꺼운 경우에 독립샘플러를 적용할 것을 제안하였다. 독립샘플러에서 이전 값들에 대해 독립적으로 후보점  $y$ 을 생성하지만, 생성된 샘플들은 서로 독립이 아니라는 점에 유의해야 한다. 그 이유는 각 시점  $t$ 에서 다음 시점  $t+1$ 로 이동할 때  $y$ 를 채택하는 채택확률  $\alpha(x_t, y)$ 가  $x_t$ 에 의존하기 때문이다.

**예제 6.1.16** 독립샘플러를 사용해서, 다음 감마확률밀도함수로부터 난수들을 생성하기로 하자.

$$f(x) = \frac{1}{\Gamma(2.43)} x^{2.43-1} \exp(-x) 1_{[0,\infty)}(x) \quad (1)$$

제안분포로 지지대가  $(-A, A)$ 인 일양확률분포를 사용하자. 즉, 이 제안분포의 확률밀도함수는 다음과 같다.

$$q_y = \frac{1}{2A} 1_{(-A,A)}(y) \quad (2)$$

식 (2)의 제안분포의 확률밀도함수는 조건 (6.1.7)을 만족한다. 즉, 이 독립샘플러는 확률보행 Metropolis샘플러이다.

MATLAB함수 `mhsample.m`을 사용해서 독립Metropolis샘플들을 생성하기로 하고, 다음 MATLAB프로그램 `IndependenceSampler101.m`을 실행해 보자.

```

1 % -----
2 % Filename: IndependenceSampler101.m
3 % Normal Random Numbers generated by Independence Metropolis Sampler
4 % using MATLAB function mhsample
5 % Programmed by MathWorks
6 % -----
7 clear all, close all
8 rng default; % For reproducibility
9 alpha = 2.43;
10 beta = 1;
11 pdf = @(x)gampdf(x,alpha,beta); % Target distribution
12 proppdf = @(x,y) gampdf(x,floor(alpha),floor(alpha)/alpha);
13 proprnd = @(x) sum(exprnd(floor(alpha)/alpha,floor(alpha),1));
14 nsamples = 5000;
15 smpl = mhsample(1,nsamples,'pdf',pdf,'proprnd',proprnd,...
16                 'proppdf',proppdf);
17 % Plot the results.
18 xxhat = cumsum(smpl.^2)./(1:nsamples)';
19 figure;
20 subplot(2,1,1)
21 plot(1:nsamples,xxhat,'r-','linewidth',2)
22 set(gca,'fontsize',11,'fontweigh','bold')
23 subplot(2,1,2)
24 hist(xxhat,30)
25 set(gca,'fontsize',11,'fontweigh','bold')
26 % h(1).FaceColor = [.8 .8 .8 ];
27 saveas(gcf,'IndependenceSampler101','eps')
28 save('IndependenceSampler101.txt','xxhat','-ascii')
29 % End of Program
30 % -----

```

MATLAB 함수 `mhsample.m`의 사용법은 다음과 같다.

```
smp1 = mhsample(start, nsamples, 'pdf', pdf, 'proppdf', proppdf, ...
                'proprnd', proprnd, 'symmetric', sym)
```

첫 번째 입력모수 `start`는 MH 샘플의 사전값  $x_0$  이고, 두 번째 입력모수 `'nsamples'`는 발생시킬 MH 샘플의 개수이고, 세 번째 입력모수 `'pdf'`에는 목표확률밀도함수 `pdf`를 지정하고, 네 번째 입력모수 `'proppdf'`에는 제안분포의 조건부확률밀도함수 `proppdf`를 지정하고, 다섯 번째 입력모수 `'proprnd'`에는 제안분포에서 생성된 난수를 지정하고, 여섯 번째 입력모수 `'symmetric'`에는 제안분포의 대칭성을 나타내는 `sym`을 할당한다. 만약 `sym`이 1이면 제안분포는 대칭이고, 0이면 비대칭이다. 또한, 디폴트는 비대칭이다.

이 MATLAB 프로그램 `IndependenceSampler101.m`을 실행하면, 다음과 같은 제안분포의 확률밀도함수를 사용해서 15,000개 표준정규난수들을 생성한다.

$$q_{x_t, y} = \frac{1}{2} 1_{(-1, 1)}(|y - x_t|) \quad (3)$$

즉, 식 (2)에서  $A = 1$ 인 경우이다. 생성된 독립 Metropolis 샘플들의 시계열산점도와 히스토그램을 그림 6.1.16에 그려져 있다. ■

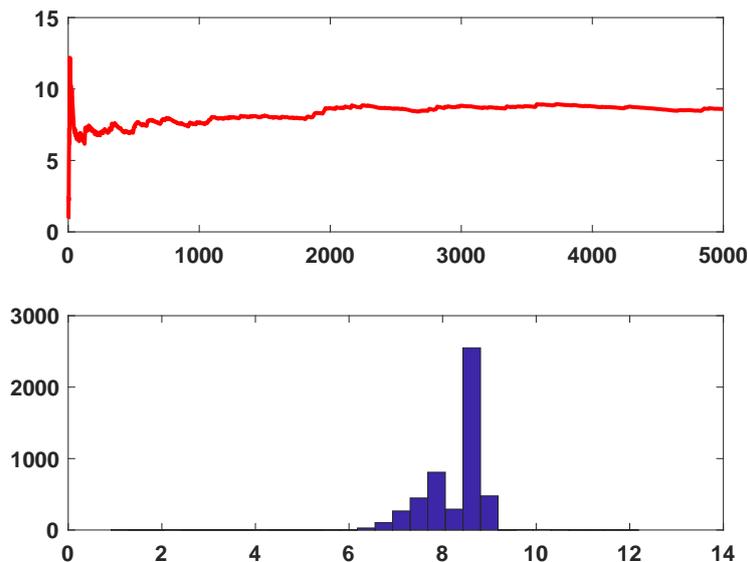


그림 6.1.16. 독립샘플러와 감마확률분포

### 6.1.6 AR-MH샘플러

AR-MH샘플러 (Acceptance-Rejection Metropolis-Hastings sampler)는 독립샘플러를 바탕으로 하는 것으로, 목표확률밀도함수  $\pi(\mathbf{x})$ 의 근사식으로 정규확률분포  $\mathcal{N}(\mathbf{m}, V)$ 의 확률밀도함수  $\phi(\mathbf{x}; \mathbf{m}, V)$ 를 사용하기도 한다. 즉, 다음과 같은 제안분포의 조건부확률밀도함수를 사용한다.

$$q(\mathbf{x}) \propto \min \{ \pi(\mathbf{x}), c \cdot \phi(\mathbf{x}; \mathbf{m}, V) \} \quad (6.1.12)$$

여기서  $c$ 는 미리 알고 있는 상수이다.

이 방법을 적용하기 위해서는 먼저 다음 식을 사용해서 후보점  $\mathbf{x}$ 를 발생시킨다.

$$\mathbf{x} \stackrel{d}{\sim} \phi(\mathbf{x}; \mathbf{m}, V) \quad (6.1.13)$$

이렇게 발생된  $\mathbf{x}$ 를 확률  $\pi(\mathbf{x}) / [c \cdot \phi(\mathbf{x}; \mathbf{m}, V)]$ 로 수용하고,  $\mathbf{x}$ 가 기각되는 경우에는 수용될 때까지 반복하여 후보점을 발생시킨다. 이러한 단계를 채택기각단계 (Acceptance-Rejection step)라 한다. 새로운 후보점  $\mathbf{x}$ 가 제안되면, 나머지 과정은 일반적인 MH샘플러를 따른다. 우선, 다음 집합을 정의하자.

$$C \doteq \{ \mathbf{x} \mid \pi(\mathbf{x}) \leq c \cdot \phi(\mathbf{x}; \mathbf{m}, V) \} \quad (6.1.14)$$

다음 식을 사용해서 MH채택확률을 계산하자.

$$\alpha(\mathbf{x}^{(j-1)}, \mathbf{x} \mid \mathbf{y}) = \begin{cases} 1, & (\mathbf{x}^{(j-1)} \in C) \\ c \cdot \phi(\mathbf{x}^{(j-1)}) / \pi(\mathbf{x}^{(j-1)}), & (\mathbf{x}^{(j-1)} \notin C, \mathbf{x} \in C) \\ \min \left\{ \frac{\pi(\mathbf{x}) \phi(\mathbf{x}^{(j-1)})}{\pi(\mathbf{x}^{(j-1)}) \phi(\mathbf{x})}, 1 \right\}, & (\mathbf{x}^{(j-1)} \notin C, \mathbf{x} \notin C) \end{cases} \quad (6.1.15)$$

이러한 샘플러는 채택과 기각을 포함하는 샘플러이므로 AR-MH샘플러라 불린다. 이 AR-MH 샘플러를 사용함으로써 단순히  $\phi(\mathbf{x}; \mathbf{m}, V)$ 를 제안분포의 조건부확률밀도함수로 사용하는 독립샘플러보다 채택률을 높일 수 있다.

## 6.1.7 자기회귀샘플러

Tierney [95]와 Chib & Greenberg [27]는 자기회귀생성밀도함수 (autoregressive generating density) 를 바탕으로 한 자기회귀샘플러를 제안하였다. 이 샘플러에서는 후보점  $\mathbf{y}$ 가 다음 모형에서 발생된다.

$$\mathbf{y} = \mathbf{a} + B[\mathbf{x}_t - \mathbf{a}] + \mathbf{z} \quad (6.1.16)$$

여기서  $\mathbf{a}$ 는 벡터,  $B$ 는 행렬, 그리고  $\mathbf{1} \doteq [1, 1, \dots, 1]^t$ 이다. 확률벡터  $\mathbf{z}$ 는 제안분포의 조건부 결합확률밀도함수  $q$ 를 갖는다. 즉,  $q_{\mathbf{x}_t, \cdot}$ 는 다음과 같다.

$$q_{\mathbf{x}_t, \mathbf{y}} = q(\mathbf{y} - \mathbf{a} - B[\mathbf{x}_t - \mathbf{a}]) \quad (6.1.17)$$

**예제 6.1.17** 자기회귀생성밀도함수를 사용해서 2변량 정규확률분포로부터 난벡터들을 생성하는 예로서, 다음 MATLAB 프로그램 ARMHsampler101.m을 실행해 보자.

```

1 % -----
2 % Filename: ARMHsampler101.m
3 % MCMC by Autoregressive Generating Density 1: Bivariate Normal Density
4 % Programmed by CBS
5 % -----
6 clear, clf, close all
7 rand('seed', (101^2+1)/2)
8 nSim = 2000;
9 nBurnin = 20000;
10 nTot = nSim+nBurnin;
11 xmat = zeros(2, nTot);
12 mux = [0 ; 0];
13 covx = [1 0.7 ; 0.7 1 ];
14 myftn='exp( -1/2*(x-mux)''*inv(covx)*(x-mux) )';
15 target = inline(myftn, 'x', 'mux', 'covx');
16 % Generating Random Vectors
17 xmat(:,1) = randn(2,1);
18 a = mux;
19 B = [ -1 0 ; 0 -1 ];
20 for ii=2:nTot
21     z = 2*rand(2,1)-1;
22     y = a + B*(xmat(:,ii-1)-a)+10*z;
23     u = rand(1);
24     dumalpha = target(y,mux,covx)/target(xmat(:,ii-1),mux,covx);
25     alpha = min(1,dumalpha);
26     if u < alpha
27         xmat(:,ii) = y;
28     else
29         xmat(:,ii) = xmat(:,ii-1);
30     end
31 end
32 xOut = xmat(:,(nBurnin+1):nTot);

```

```

33 % Summary statistics
34 muhat = mean(xmat,2)
35 covhat = cov(xmat')
36 corrhhat = corr(xmat')
37 % Plotting
38 plot(xOut(1,:),xOut(2,:), 'r.', 'LineWidth',1.5)
39 set(gca, 'fontsize',11, 'fontweigh', 'bold')
40 hold on
41 for k = 0:4
42     theta = linspace(0,2*pi,201);
43     r = sqrt(51*k./10./(10-7*sin(2*theta)));
44     xx1 = r.*cos(theta);
45     xx2 = r.*sin(theta);
46     plot(xx1,xx2, 'k-', 'LineWidth',1.5)
47 end
48 axis([-3.5 3.5 -3.5 3.5])
49 axis square
50 hold off
51 saveas(gcf, 'ARMHsampler101', 'eps')
52 save('ARMHsampler101.txt', 'xOut', '-ascii')
53 % End of program
54 % -----

```

이 MATLAB 프로그램 ARMHsampler101.m을 실행하면, 다음과 같은 2변량 정규확률분포에서 정규난벡터들을 발생한다.

$$\mathbf{x} \stackrel{d}{\sim} \mathcal{N} \left( \begin{bmatrix} 20 \\ 10 \end{bmatrix}, \begin{bmatrix} 10 & 5 \\ 5 & 10 \end{bmatrix} \right) \quad (1)$$

이 예제에서는 자기회귀생성밀도함수를 사용해서, 변인개수는  $b = 20,00$ 이고 샘플수는  $n = 2,000$ 인 난벡터들을 생성한다. 이 난벡터들의 표본평균벡터, 표본분산공분산행렬, 그리고 표본상관계수행렬은 각각 다음과 같다.

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} -0.0059 \\ -0.0577 \end{bmatrix}, \quad \hat{\boldsymbol{\Sigma}} = \begin{bmatrix} 0.8457 & 0.5571 \\ 0.5571 & 0.8296 \end{bmatrix}, \quad P = \begin{bmatrix} 1.0000 & 0.6652 \\ 0.6652 & 1.0000 \end{bmatrix} \quad (2)$$

이렇게 발생된 난벡터열의 산점도가 그림 6.1.17에 수록되어 있다. 원래 그림 6.1.17의 산점도에는 2,000개 점들이 나타나야 하나, 그보다 훨씬 적은 점들이 나타났다. 이는 비슷한 점들이 겹쳐 찍혀서 나타나는 현상 같다. ■

**예제 6.1.18** 동물들의 유전자의 연결성을 살펴보기 위해서 자기회귀샘플러를 사용하기로 하자. 이 연구에서는 197종류 동물들을 네 종류로 분류하고, 각 종류의 확률이 각각  $[2 + \theta]/4, [1 - \theta]/4, [1 - \theta]/4$ , 그리고  $\theta/4$ 라고 가정한다. 실제로 각 종류에 속하는 동물들의

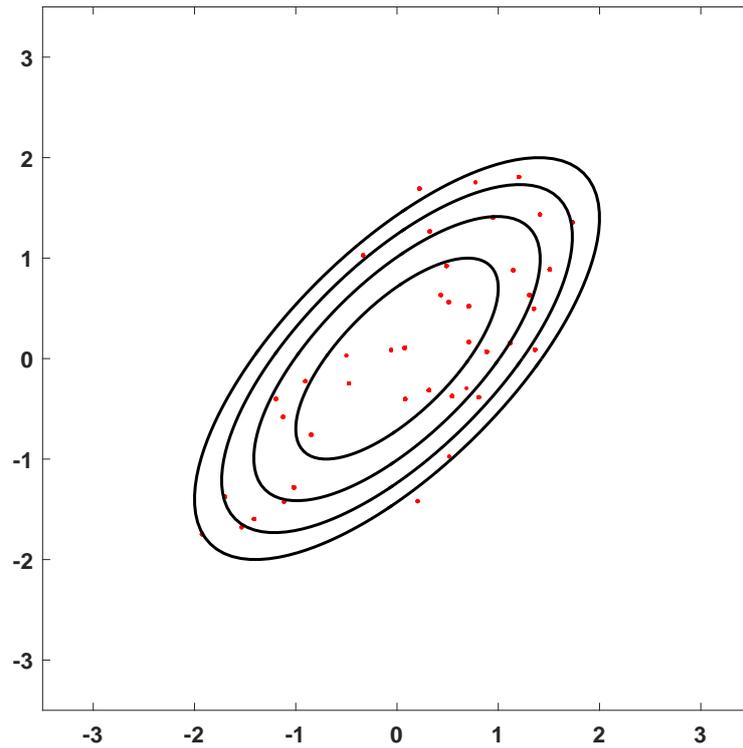


그림 6.1.17. 자기회귀생성밀도함수에 의한 2변량 정규난벡터

수가 다음과 같다고 하자.

$$z_1 = 125, \quad z_2 = 18, \quad z_3 = 20, \quad z_4 = 34 \quad (1)$$

관찰벡터  $\mathbf{z} \doteq [z_1, z_2, z_3, z_4]^t = [125, 18, 20, 34]^t$ 의 우도함수는 다음과 같다.

$$P(\mathbf{z} | \theta) \propto [2 + \theta]^{z_1} [1 - \theta]^{z_2 + z_3} \theta^{z_4} = [2 + \theta]^{125} [1 - \theta]^{38} \theta^{34} \quad (2)$$

모수  $\theta$ 의 사전확률밀도함수가 다음과 같다고 가정하자.

$$\pi(\theta) = \frac{1}{2a} I_{(-a,a)}(\theta) \quad (3)$$

Bayes 정리에서 알 수 있듯이, 모수  $\theta$ 의 사후확률밀도함수는 다음과 같다.

$$\pi(\theta | \mathbf{z}) \propto [2 + \theta]^{z_1} [1 - \theta]^{z_2 + z_3} \theta^{z_4} = [2 + \theta]^{125} [1 - \theta]^{38} \theta^{34} \quad (4)$$

자기회귀생성밀도함수를 사용해서 이 사후확률밀도함수로부터  $\theta$ 의 난수열을 발생시키기 위해서, 다음 MATLAB 프로그램 ARMHsampler102.m을 실행해 보자.

```

1 % -----
2 % Filename: ARMHsampler102.m
3 % MCMC by Autoregressive Generating Density 2: Discrete Distribution
4 % Robert (2000)
5 % Programmed by CBS
6 % -----
7 clear, clf, close all
8 rand('seed',(101^2+1)/2)
9 z1 = 125, z2 = 18, z3 = 20, z4 = 34
10 nSim = 1000, NBurnin = 100
11 nTot = nSim+NBurnin
12 % Step size for proposal distribution
13 a = 0.1;
14 % Set up the space to store values.
15 theta = zeros(1,nTot);
16 myftn = '(2+tt).^z1.*((1-tt).^(z2+z3)).*(tt.^z4)';
17 posterior = inline(myftn,'tt','z1','z2','z3','z4');
18 % Use random-walk Metropolis-Hastings
19 % w/ y = theta(i-1)+z and z ~ U(-a,a)
20 theta(1) = rand(1);
21 for ii = 2:nTot
22     y = theta(ii-1)+a*(2*rand(1)-1);
23     u = rand(1);
24     dumalpha = posterior(y,z1,z2,z3,z4)/posterior(theta(ii-1),z1,z2,z3,z4);
25     alpha = min(dumalpha,1);
26     if u<= alpha
27         theta(ii) = y;
28     else
29         theta(ii) = theta(ii-1);
30     end
31 end
32 % Histogram
33 xHistL = min(theta), xHistR = max(theta)
34 HistIncrement = 0.02
35 PMFind = (xHistL:HistIncrement:xHistR)';
36 [Nbar, xcenter] = hist(theta(NBurnin+1:NBurnin+nSim), PMFind);
37 Nheight = Nbar/nSim;
38 bar(xcenter, Nheight, 'g');
39 set(gca, 'fontsize', 11, 'fontweigh', 'bold')
40 ylabel('\bf Posterior', 'fontsize', 12)
41 saveas(gcf, 'ARMHsampler102', 'eps')
42 save('AR_MHsampler102.txt', 'theta', '-ascii')
43 % End of program
44 % -----

```

이 MATLAB 프로그램 ARMHsampler102.m을 실행하면, 변인개수  $b = 100$  다음에 사후 관찰값들  $\{\theta^{(j)} \mid j = 1, 2, \dots, 1000\}$ 이 생성된다. 이 중에서 최소값은 0.3485이고 최대값은 0.7613이다. 이렇게 발생된 사후관찰값들의 히스토그램을 그림 6.1.18에 그려져 있다. 이 히스토그램은  $\theta$ 의 사후확률밀도함수의 근사이다. ■

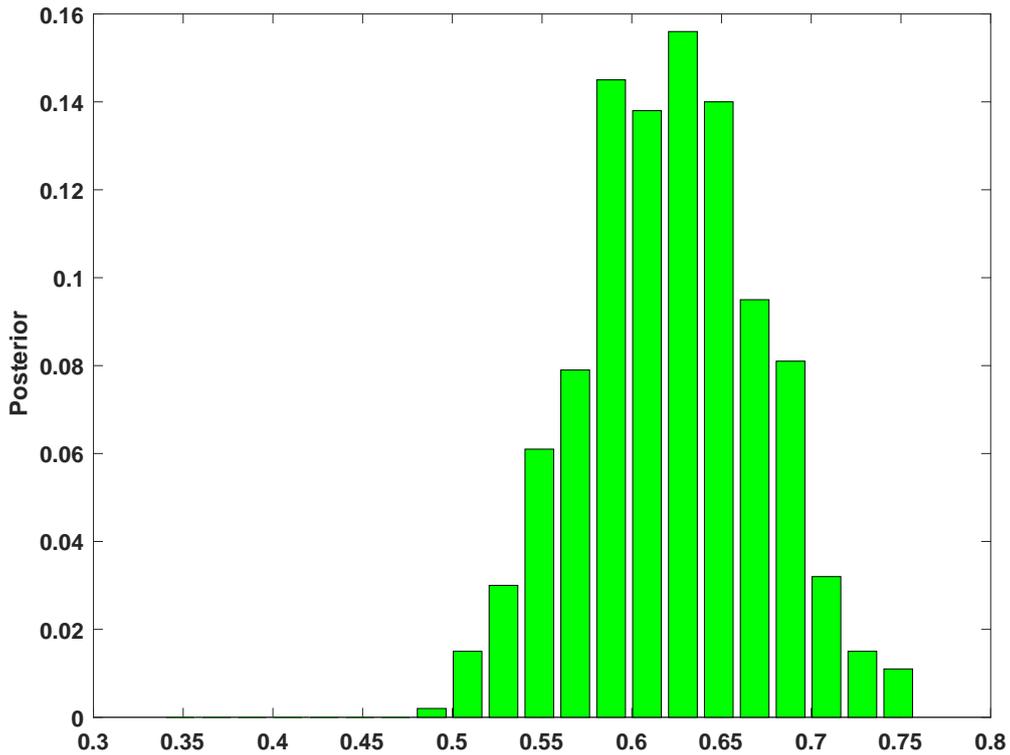


그림 6.1.18. 자기회귀생성밀도함수에 의한 이산형 확률변수

## 제 6.2 절 불변분포의 존재성

이 절에서는 MH 샘플러에서 발생된 MH 샘플들이 목표확률밀도함수에 수렴함을 증명하기로 하자.

이산형 Markov 연쇄에 대해서 살펴보자. 상태공간을  $S = \{0, 1, \dots\}$ 로 하는 어떤 불가약적 Markov 연쇄의 주어진 추이행렬  $Q = [q_{ij}]$ 를 추이행렬로 하는 새로운 Markov 연쇄  $\{x_t\}$ 를 다음과 같이 발생시키자. 첫째, 시점  $t$ 에서 상태가  $x_t = i$ 인 경우, 확률질량함수  $\{q_{i0}, q_{i1}, q_{i2}, \dots\}$ 로부터 실현값  $y$ 를 생성한다. 둘째, 시점  $t+1$ 에서  $x_{t+1}$ 을 구하기 위해서, 다음과 같이 MH 채택확률  $\alpha_{iy} \doteq \alpha(x_t, y)$ 를 정의하자.

$$\alpha_{iy} \doteq \min \left\{ \frac{\pi(y)q_{yi}}{\pi(i)q_{iy}}, 1 \right\} \quad (6.2.1)$$

셋째, 확률  $\alpha_{iy}$ 로  $x_{t+1} = y$ 라 하고, 확률  $1 - \alpha_{iy}$ 로  $x_{t+1} = i$ 라 하자. 즉, 다음과 같이 시점

$t + 1$ 에서 상태를 할당한다.

$$x_{t+1} = \begin{cases} y & \text{with probability } \alpha_{iy} \\ x_t & \text{with probability } 1 - \alpha_{iy} \end{cases} \quad (6.2.2)$$

이 새로운 Markov연쇄  $\{x_t\}$ 에서, 시점  $t$ 의 상태  $i$ 에서 시점  $t + 1$ 의 상태  $j$ 로 가는 추이확률  $p_{ij}$ 는 다음과 같다.

$$p_{ij} = \begin{cases} q_{ij}\alpha_{ij}, & (j \neq i) \\ q_{ii} + \sum_{k \neq i} q_{ik} [1 - \alpha_{ik}], & (j = i) \end{cases} \quad (6.2.3)$$

다음 식들이 성립한다.

$$\begin{aligned} \pi(i)q_{ij}\alpha_{ij} &= \pi(i)q_{ij} \min \left\{ \frac{\pi(j)q_{ji}}{\pi(i)q_{ij}}, 1 \right\} = \min \{ \pi(j)q_{ji}, \pi(i)q_{ij} \} \\ &= \pi(j)q_{ji} \min \left\{ \frac{\pi(i)q_{ij}}{\pi(j)q_{ji}}, 1 \right\} = \pi(j)q_{ji}\alpha_{ji} \end{aligned} \quad (6.2.4)$$

여기서 첫 번째와 네 번째 등식들은 식 (6.2.1)에 의해서 성립한다. 식 (6.2.4)에서 알 수 있듯이, 다음 식이 성립한다.

$$\frac{\pi(i)}{\pi(j)} = \frac{q_{ji}\alpha_{ji}}{q_{ij}\alpha_{ij}} \quad (6.2.5)$$

이 Markov연쇄  $\{x_t\}$ 가 시간적으로 가역이기 위한 필요충분조건은 다음과 같다.

$$\pi(i)p_{ij} = \pi(j)p_{ji} \quad (6.2.6)$$

식 (6.2.3)에서 알 수 있듯이, 다음 식은 식 (6.2.6)과 동치이다.

$$\pi(i)q_{ij}\alpha_{ij} = \pi(j)q_{ji}\alpha_{ji} \quad (6.2.7)$$

따라서, 식 (6.2.1)과 식 (6.2.2)에 의해서 생성된 Markov연쇄  $\{x_t\}$ 는 시간적으로 가역이다. Markov연쇄가 시간적으로 가역이면 불변분포가 존재하고,  $\{\pi_0, \pi_1, \dots\}$ 가 이 불변분포이다. 제4.9절에서 알 수 있듯이, 불변분포가 존재하면 추이확률  $p_{ij}^{(n)}$ 는 목표확률  $\pi(j)$ 에 수렴한다.

연속형 Markov연쇄의 MH샘플러에 의해 생성되는  $\{\theta^{(j)} \mid j = 1, 2, \dots\}$ 가 불변분포가

$\pi(\theta)$  인 Markov 연쇄를 구성함을 증명하기로 하자. 다음 함수들을 정의하자.

$$p(\theta^{(j-1)}, \theta^{(j)}) \doteq \begin{cases} q(\theta^{(j-1)}, \theta^{(j)}) \alpha(\theta^{(j-1)}, \theta^{(j)}), & (\theta^{(j-1)} \neq \theta^{(j)}) \\ 0, & (\theta^{(j-1)} = \theta^{(j)}) \end{cases} \quad (6.2.8)$$

$$r(\theta^{(j-1)}) \doteq 1 - \int p(\theta^{(j-1)}, \theta^{(j)}) \mu(d\theta^{(j)}) \quad (6.2.9)$$

$$\delta_{\theta^{(j-1)}}(d\theta^{(j)}) \doteq \begin{cases} 1, & (\theta^{(j-1)} \in d\theta^{(j)}) \\ 0, & (\theta^{(j-1)} \notin d\theta^{(j)}) \end{cases} \quad (6.2.10)$$

여기서  $\mu(\cdot)$  는 확률측도이다. 이 Markov 연쇄의 추이핵  $P$  는 다음과 같다.

$$P(\theta^{(j-1)}, \theta^{(j)}) = p(\theta^{(j-1)}, \theta^{(j)}) \mu(d\theta^{(j)}) + r(\theta^{(j-1)}) \delta_{\theta^{(j-1)}}(d\theta^{(j)}) \quad (6.2.11)$$

식 (6.2.11) 의 추이핵  $P$  는 추이확률  $p(\theta^{(j-1)}, \theta^{(j)})$  로서  $\theta^{(j-1)}$  에서  $\theta^{(j)}$  로 추이하고, 확률  $r(\theta^{(j-1)})$  로서  $\theta^{(j-1)}$  에 머무는 것을 나타낸다. 이 Markov 연쇄에서 조건부확률밀도함수  $p(\theta^{(j-1)}, \theta^{(j)})$  와 확률밀도함수  $\pi(\theta)$  는 다음 식을 만족한다.

$$\pi(\theta^{(j-1)}) p(\theta^{(j-1)}, \theta^{(j)}) = \pi(\theta^{(j)}) p(\theta^{(j)}, \theta^{(j-1)}) \quad (6.2.12)$$

즉, 이 Markov 연쇄는 시간적으로 가역이다. 식 (6.2.12) 를 상세균형방정식(detailed balance equation)이라 부르기도 한다. 앞에서도 언급했듯이, 가역성조건 (6.2.12) 는 이 Markov 연쇄의 불변분포가  $\pi(\theta)$  이기 위한 충분조건이다. 따라서, MH 샘플러에 의해 생성되는 Markov 연쇄는  $\pi(\theta)$  를 불변분포로 갖는다. 제 4.9 절에서 알 수 있듯이, 불변분포가 존재하면 추이확률  $p(\theta^{(j-1)}, \theta^{(j)})$  는 확률  $\pi(\theta^{(j)})$  에 수렴한다.

## 제 6.3 절 베이저안통계분석과 MH 샘플러

### 6.3.1 베이저안통계분석을 위한 MH 샘플러

사후확률밀도함수  $\pi(\theta | \mathbf{y})$  에서 샘플링이 어려운 경우에, MH 샘플러에서는 이 사후확률밀도함수와 근사한 확률밀도함수에서 샘플링을 한 다음, 목표확률밀도함수와 근사 확률밀도함수의 차이를 수정해서, 목표확률밀도함수에서 샘플열을 발생시킨다. 앞에서도 언급했듯이, 이 근사 확률밀도함수를 관습적으로 제안분포(proposal distribution) 또는 후보생성분포(candidate

generating distribution)라 부른다.

Gibbs샘플러에서와 마찬가지로, 적당한 초기벡터  $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$ 를 선택한다. 다음 단계로, 각  $j (= 1, 2, \dots)$ 에 대해서 다음과 같은 절차를 거쳐 주어진  $\theta^{(j-1)}$ 로부터  $\theta^{(j)}$ 를 발생시킨다. 우선, 관찰벡터  $\mathbf{y}$ 가 주어진 조건 하에서 조건부제한확률밀도함수  $q(\theta^{(j-1)}, \theta | \mathbf{y})$ 에서  $\theta$ 을 발생시킨다. 이 조건부제한확률밀도함수는 조건부 추이확률을 나타냄을 상기하라. 이렇게 구해진  $\theta$ 는 조건부목표확률밀도함수  $\pi(\theta | \mathbf{y})$ 에서 추출된 샘플이 아니다. 따라서, 조건부제한확률밀도함수와 조건부목표확률밀도함수사이에서 발생하는 괴리를 수정하기 위해서, 다음과 같이 정의되는 MH채택확률을 사용한다.

$$\alpha(\theta^{(j-1)}, \theta | \mathbf{y}) = \min \left\{ \frac{\pi(\theta | \mathbf{y})q(\theta, \theta^{(j-1)} | \mathbf{y})}{\pi(\theta^{(j-1)} | \mathbf{y})q(\theta^{(j-1)}, \theta | \mathbf{y})}, 1 \right\} \tag{6.3.1}$$

식 (6.3.1)은 조건  $\pi(\theta^{(j-1)} | \mathbf{y})q(\theta^{(j-1)}, \theta | \mathbf{y}) > 0$ 가 성립할 때만 정의된다. 만약 조건  $\pi(\theta^{(j-1)} | \mathbf{y})q(\theta^{(j-1)}, \theta | \mathbf{y}) = 0$ 가 성립하면, MH채택확률을  $\alpha(\theta^{(j-1)}, \theta | \mathbf{y}) = 1$ 이라 놓는다. 다음 단계로, 지지대가  $[0, 1]$ 인 일양확률분포에서 일양난수  $u$ 를 발생시킨다. 만약 식  $u \leq \alpha(\theta^{(j-1)}, \theta | \mathbf{y})$ 가 성립하면,  $\theta^{(j)} = \theta$ 로 한다. 그렇지 않으면,  $\theta^{(j)} = \theta^{(j-1)}$ 로 한다. 지금까지 설명한 베이저안통계분석을 위한 MH샘플러를 요약하면 다음과 같다.

**알고리즘 6.3.1: 베이저안통계분석을 위한 MH샘플러**

(1단계) 초기벡터  $\theta^{(0)}$ 를 선택한다.

(2단계) 각  $j (= 1, 2, \dots)$ 에 대해서 다음 과정을 실행한다.

(2-1) 다음 조건부제한확률밀도함수를 사용해서, 새로운 후보점  $\theta$ 를 추출한다.

$$\theta \stackrel{d}{\sim} q(\theta^{(j-1)}, \theta | \mathbf{y})$$

(2-2) 만약 조건  $\pi(\theta^{(j-1)} | \mathbf{y})q(\theta^{(j-1)}, \theta | \mathbf{y}) > 0$ 가 성립하면, 다음 MH채택확률을 계산한다.

$$\alpha(\theta^{(j-1)}, \theta | \mathbf{y}) = \min \left\{ \frac{\pi(\theta | \mathbf{y})q(\theta, \theta^{(j-1)} | \mathbf{y})}{\pi(\theta^{(j-1)} | \mathbf{y})q(\theta^{(j-1)}, \theta | \mathbf{y})}, 1 \right\}$$

만약 조건  $\pi(\theta^{(j-1)} | \mathbf{y})q(\theta^{(j-1)}, \theta | \mathbf{y}) = 0$ 가 성립하면, 확률  $\alpha(\theta^{(j-1)}, \theta | \mathbf{y})$ 를 1로 둔다.

(2-3) 다음과 같이 일양난수  $u$ 를 발생시킨다.

$$u \stackrel{d}{\sim} \text{Uniform}(0, 1)$$

(2-4) 다음과 같이  $\theta^{(j)}$ 를 정한다.

$$\theta^{(j)} = \begin{cases} \theta, & (u \leq \alpha(\theta^{(j-1)}, \theta | \mathbf{y})) \\ \theta^{(j-1)}, & (u > \alpha(\theta^{(j-1)}, \theta | \mathbf{y})) \end{cases}$$

(3단계) 충분히 큰  $b$ 에 대해  $\theta^{(j)}$ , ( $j = b + 1, b + 2, \dots$ )를 기록한다.

식 (6.3.1)에서는  $\pi$ 와  $q$ 가 분자와 분모 모두에 나타나므로, 이들의 정규화상수(normalizing constant)들이 상쇄된다. 따라서  $\pi$ 의 정규화상수나  $q$ 의 정규화상수를 구하지 못하는 경우에도, 식 (6.3.1)을 이용해서 MH샘플러를 실행할 수 있다.

Metropolis샘플러에서는 조건부제안확률밀도함수  $q$ 가  $(\theta^{(j-1)}, \theta)$ 에 관해 대칭이다. 즉, 다음 식이 성립한다.

$$q(\theta^{(j-1)}, \theta | \mathbf{y}) = q(\theta, \theta^{(j-1)} | \mathbf{y}) \quad (6.3.2)$$

조건 (6.3.2)가 성립하는 경우에는 식 (6.3.1)을 다음과 같이 쓸 수 있다.

$$\alpha(\theta^{(j-1)}, \theta | \mathbf{y}) = \min \left\{ \frac{\pi(\theta | \mathbf{y})}{\pi(\theta^{(j-1)} | \mathbf{y})}, 1 \right\} \quad (6.3.3)$$

따라서, 만약 식  $\pi(\theta | \mathbf{y}) \geq \pi(\theta^{(j-1)} | \mathbf{y})$ 가 성립하면  $\theta^{(j-1)}$ 에서 반드시  $\theta$ 로 이동하고, 그렇지 않으면 확률  $\pi(\theta | \mathbf{y})/\pi(\theta^{(j-1)} | \mathbf{y})$ 로  $\theta^{(j-1)}$ 에서  $\theta$ 로 이동한다. 즉, 사후확률밀도함수가 높은 후보점  $\theta$ 가 제안되면 반드시 그 점으로 이동하고, 그렇지 않으면 목표확률밀도함수값들의 비를 확률로 해서  $\theta$ 로 이동한다. 즉, 사후확률밀도함수가 낮은 후보점도 1보다 작은 확률이긴 하지만  $\theta$ 로 이동할 수 있다.

Gibbs샘플러에서 모수벡터  $\theta$ 가  $p$ 개 블록(block)으로 나누어진  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ 이면, 각 모수벡터  $\theta_j$ 를 사후조건부확률밀도함수  $\pi(\theta_j | \theta_{-j}, \mathbf{y})$ 에서 발생시킨다. 그러나, 실제 응용에서는 모든 사후조건부확률분포들에서 쉽게 샘플링할 수 있는 것은 아니다. 이러한 경우에는 다음과 같은 다중블록MH샘플러(multi-block algorithm)를 사용할 수도 있다.

**알고리즘 6.3.2: 다중블록MH샘플러**

(1단계) 초기벡터  $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$  를 결정한다.

(2단계) 각  $j(= 1, 2, \dots)$  에 대해,  $j$  를 고정하고  $k(= 1, 2, \dots, p)$  에 대해 축차적으로 다음 과정을 시행하여  $\theta^{(j)} = (\theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_p^{(j)})$  를 구한다.

(2-1) 조건부확률밀도함수  $q_j(\theta_k^{(j-1)}, \theta_k \mid \theta_{-k}^{(j)}, \mathbf{y})$  를 이용해서,  $\theta_{-k}^{(j)} = (\theta_1^{(j)}, \dots, \theta_{k-1}^{(j)}, \theta_{k+1}^{(j)}, \dots, \theta_p^{(j)})$  가 주어진 조건 하에서 새로운 후보점  $\theta_k$  를 제안한다.

(2-2) 만약  $\pi(\theta_k^{(j-1)} \mid \theta_{-k}^{(j)}, \mathbf{y}) q(\theta_k^{(j-1)}, \theta_k \mid \theta_{-k}^{(j)}, \mathbf{y}) = 0$  이면, 다음 확률을 정의한다.

$$\alpha_k(\theta_k^{(j-1)}, \theta_k \mid \theta_{-k}^{(j)}, \mathbf{y}) = 1$$

만약  $\pi(\theta_k^{(j-1)} \mid \theta_{-k}^{(j)}, \mathbf{y}) q(\theta_k^{(j-1)}, \theta_k \mid \theta_{-k}^{(j)}, \mathbf{y}) \neq 0$  이면, 다음 MH채택확률을 계산한다.

$$\alpha_k(\theta_k^{(j-1)}, \theta_k \mid \theta_{-k}^{(j)}, \mathbf{y}) = \min \left\{ \frac{\pi(\theta_k \mid \theta_{-k}^{(j)}, \mathbf{y}) q_k(\theta_k, \theta_k^{(j-1)} \mid \theta_{-k}^{(j)}, \mathbf{y})}{\pi(\theta_k^{(j-1)} \mid \theta_{-k}^{(j)}, \mathbf{y}) q_k(\theta_k^{(j-1)}, \theta_k \mid \theta_{-k}^{(j)}, \mathbf{y})}, 1 \right\}$$

(2-3) 일양난수  $u$  를 발생시킨다.

(2-4) 다음과 같이  $\theta_k^{(j)}$  를 결정한다.

$$\theta_k^{(j)} = \begin{cases} \theta_k, & (u \leq \alpha_k(\theta_k^{(j-1)}, \theta_k \mid \theta_{-k}^{(j)}, \mathbf{y})) \\ \theta_k^{(j-1)}, & (u > \alpha_k(\theta_k^{(j-1)}, \theta_k \mid \theta_{-k}^{(j)}, \mathbf{y})) \end{cases}$$

(3단계) 충분히 큰  $b$  에 대해  $\theta^{(j)}, (j = b + 1, b + 2, \dots)$  를 기록한다.

알고리즘 6.3.2에서 다음 식들이 성립한다고 가정하자.

$$q_k(\theta_k^{(j-1)}, \theta_k \mid \theta_{-k}^{(j)}, \mathbf{y}) = \pi(\theta_k \mid \theta_{-k}^{(j)}, \mathbf{y}) \tag{6.3.4}$$

$$q_k(\theta_k, \theta_k^{(j-1)} \mid \theta_{-k}^{(j)}, \mathbf{y}) = \pi(\theta_k^{(j-1)} \mid \theta_{-k}^{(j)}, \mathbf{y}) \tag{6.3.5}$$

다음 식이 성립한다.

$$\alpha_k \left( \boldsymbol{\theta}_k^{(j-1)}, \boldsymbol{\theta}_k \mid \boldsymbol{\theta}_{-k}^{(j)}, \mathbf{y} \right) = 1 \quad (6.3.6)$$

따라서, 알고리즘 6.3.2는 Gibbs샘플러가 된다. 즉, Gibbs샘플러는 다중블록MH샘플러에서 조건부제한확률밀도함수가 사후조건부확률밀도와 일치하며 MH채택확률이 1이 되는 특별한 경우이다.

### 6.3.2 MH샘플의 수렴성

사후확률분포에서 샘플링을 시행할 때는 단일체인(single chain)을 이용하는 방법과 다중체인(multiple chain)을 이용하는 방법이 있다. 첫째, 단일체인은 하나의 긴 MH샘플열을 발생시키고, 그 MH샘플들을 사용해서 추론을 시행하는 방법이다. 둘째, 다중체인은 비교적 짧은 복수의 MH샘플열들을 발생시키고, 모든 MH샘플열들의 MH샘플들을 같이 사용해서 추론을 시행하는 방법이다. 다중체인에서는 서로 독립인 MH샘플들을 얻을 수 있지만, 초기값을 잘 선택하지 않으면 MH샘플들이 사후확률분포에 수렴하지 않을 가능성이 높다. 물론, 단일체인에서도 MH샘플들이 사후확률분포에 수렴하지 않을 가능성이 있지만, 긴 수열이 상태공간 전체에서 샘플링될 가능성이 높다. 따라서, 오늘날에는 다중체인보다도 단일체인이 많이 사용되고 있다.

MCMC에서는 불변분포인 사후확률분포에서 초기값  $\boldsymbol{\theta}^{(0)}$ 를 샘플하지 않는 한, MH샘플  $\boldsymbol{\theta}^{(j)}$ 가 수렴할 때까지는 어느 정도 시간이 걸린다. 따라서, 충분히 큰  $b$ 에 대해서  $\{\boldsymbol{\theta}^{(j)} \mid j > b\}$ 의 경험확률분포(empirical probability distribution)가 사후확률분포에 수렴하고 있다고 간주하고, 이 MH샘플들을 이용해서 불변분포인 사후확률분포에 관한 통계적 추론을 실행한다. 이 경우에,  $\{\boldsymbol{\theta}^{(j)} \mid j \leq b\}$ 는 MH샘플들이 불변분포에 수렴할 때까지 초기값에 의존하는 번인샘플(burn-in sample)들로서, 통계적 추론에 사용하지 않고 버린다. 적당한  $b$ 를 선택하는 기준을 구하려는 다양한 시도가 이루어지고 있지만, 아직은 보편적으로 사용되는 기준은 없다. 번인샘플들을 버리고 남은 MH샘플들  $\{\boldsymbol{\theta}^{(j)} \mid j > b\}$ 가 사후확률분포에 수렴하고 있는지 여부를 판정하는 것은 중요한 문제이다. 많은 경우, 실제로 얻어진 MH샘플열을 이용해서 수렴여부를 판정하고 있다. 수렴여부를 판정하는 가장 간단한 방법은 MH샘플열의 시계열산점도, MH샘플평균의 시계열산점도, 그리고 MH샘플열의 자기상관함수 등 다양한 그림을 그려보는 것이다. 이 시계열산점도들과 자기상관함수를 바탕으로 발생된 MH샘플열의 변동이 초기값에 의존하는 여부와 안정성을 조사해서, MH샘플열의 수렴여부를 판정한다. 사후확률분포의

상태공간 전체에서 효율적으로 샘플링이 되었다면, MH샘플열의 자기상관함수의 절대값은 시차 (lag)가 커짐에 따라 감소한다. 따라서, MH샘플열의 자기상관계수의 절대값이 크면, MH샘플열이 상태공간의 일부에서 MH샘플들이 발생된 것이라고 생각할 수 있다. 이러한 경우에, 사후확률분포에서 좋은 MH샘플열을 얻기 위해서는 보다 많은 샘플링을 시행하거나 또는 제안분포를 바꾸는 등 샘플링법을 개선해야 한다. 특히, MH샘플러에서 후보점의 채택률 (acceptance rate)이 낮은 경우에는, 같은 값이 반복해서 샘플링된다. 이러한 경우에도 MH샘플들의 자기상관계수가 커지므로 보다 많은 샘플링을 시행하거나 또는 제안분포를 바꾸는 등 샘플링법을 개선해야 한다.

Geweke[45]는 얻어진 MH샘플열의 앞 10% 샘플들을 이용해서 추정한 모수함수  $g(\theta)$ 의 기대값과 후반 50% 샘플들을 이용해서 추정한 기대값이 동일한가를 검정함으로써, MH샘플열의 수렴성을 조사한다. MH샘플열이  $n$ 개 MH샘플들로 이루어졌을 때, 첫  $n_1$  샘플들과 마지막  $n_2$  샘플들을 사용해서 다음과 같이 모수함수  $g(\theta)$ 를 추정하자.

$$\bar{g}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} g(\theta^{(j)}), \quad \bar{g}_2 = \frac{1}{n_2} \sum_{j=n-n_2+1}^n g(\theta^{(j)}) \tag{6.3.7}$$

시계열분석에서 표본평균의 분산을 추정하는 대표적인 방법은 스펙트럴밀도함수를 사용하는 것이다. 이 방법을 적용해서,  $\bar{g}_1$ 과  $\bar{g}_2$ 의 분산들을 각각 다음과 같이 추정하자.

$$s_1^2 \doteq \frac{2\pi\hat{g}_1(0)}{n_1}, \quad s_2^2 \doteq \frac{2\pi\hat{g}_2(0)}{n_2} \tag{6.3.8}$$

여기서  $\hat{g}_1(\omega)$ 와  $\hat{g}_2(\omega)$ 는 첫  $n_1$  샘플들과 마지막  $n_2$  샘플들을 사용해서 추정한 스펙트럴밀도함수의 추정값들이다. 만약 MH샘플열이 사후확률분포에 분포수렴하면, 다음 점근관계 (asymptotic relation)가 성립한다.

$$z_g \doteq \frac{\bar{g}_1 - \bar{g}_2}{\sqrt{s_1^2 + s_2^2}} \stackrel{d}{\sim} \mathcal{N}(0, 1) \tag{6.3.9}$$

따라서, 만약  $|z_g|$ 가 아주 크면, MH샘플열이 불변분포로 수렴하지 않는다고 판단한다.

Geweke법 이외에도, 다른 초기값들을 갖는 여러 MH샘플열들을 샘플링하고 이 MH샘플열들의 안정성에 입각해서 수렴판정을 하는 방법, MH샘플 백분위수의 수렴을 이용하는 방법, 확률과정의 재생시간을 이용하는 방법, Markov연쇄의 일양 에르고딕성을 이용하는 방법 등 MH샘플열의 수렴성을 조사하는 다양한 방법들이 제시되었다. Cowles & Carlin [28]은 13종류 방법들을 비교검토하였다.

모수벡터  $\theta$ 의 차원이 큰 경우에 성분단위MH샘플러로  $\theta$ 의 성분을 하나씩 발생시키면, MH 샘플들의 자기상관이 상당히 커져서 수렴이 늦어지는 경우가 있다. 이러한 경우에 수렴을 가속시키기 위해서, 모수벡터  $\theta$ 를 블록화(blocking)해서 블록단위MH샘플러를 적용하거나 모수  $\theta$ 를 변수변환(reparameterization)을 하기도 한다. 첫째, 앞에서 설명했듯이 블록단위MH샘플러는 모수벡터  $\theta$ 의 성분들을 몇 개 블록들로 나누고, 각 블록에서 MH샘플을 발생시키는 방법이다. 예를 들어, 모수벡터  $\theta = (\theta_1, \theta_1, \dots, \theta_{36})$ 의 샘플링에서 모수  $\theta_j$ , ( $j = 1, 2, \dots, 36$ )를 한번에 하나씩 발생시키면 MH샘플들의 자기상관이 높다고 하자. 이렇게 한번에 하나씩 발생시키는 것을 단순이동(single-move)이라 한다. 이 경우에  $\theta$ 의 모수들을 3개 블록들  $\theta_1 = (\theta_1, \theta_2, \dots, \theta_{11})$ ,  $\theta_2 = (\theta_{12}, \theta_{13}, \dots, \theta_{23})$ , 그리고  $\theta_3 = (\theta_{24}, \theta_{25}, \dots, \theta_{36})$ 로 나누고,  $\theta = (\theta_1, \theta_2, \theta_3)$ 에 블록단위MH샘플러를 시행하면, 많은 경우에 MH샘플들의 자기상관계수의 절대값을 작게 할 수 있다. 이렇게 블록으로 발생시키는 것을 다중이동(multi-move)이라 한다. 특히 강한 상관이 있는  $\theta_j$ 들에 관해서는 같은 블록으로 모아서 샘플링을 시행하면 된다. 수렴을 가속하기 위해서는 적은 블록수로 샘플링하는 것이 바람직하다. 그러나, 블록수를 줄이면, 즉 블록 내의 모수들의 개수가 많아지면, MH샘플러에서 목표확률밀도함수에 대한 제안분포의 조건부확률밀도함수의 근사도가 떨어져서 채택률이 낮아진다. 따라서, 적당한 개수의 블록들을 사용하는 것이 중요하다. 시행오차를 거쳐 이 적당한 개수를 찾아야 한다.

### 6.3.3 사후확률분포를 이용한 추론과 모형선택

MCMC에 의해 얻어진 샘플들을 이용해서 모수벡터  $\theta$ 에 관한 베이지안추론을 할 수 있다. 즉, 모수들의 사후확률밀도함수들에서 얻어진 샘플들을 이용해서 사후확률밀도함수를 추정할 수 있다. 또한, 샘플들의 평균과 표준편차로 사후확률분포의 평균이나 표준오차를 추정할 수 있다. 샘플들의 순서통계량(order statistics)을 이용해서, 신용구간(credible interval)도 추정할 수 있다. 모수  $\theta$ 의 함수  $h(\theta)$ 에 관한 추론을 시행하기 위해서는 사후샘플  $\theta^{(j)} | \mathbf{y}$ 의 함수  $h(\theta | \mathbf{y})$ 를 이용한다. 사후평균의 표준오차를 구하기 위해서는, 얻어진 샘플들이 시계열데이터라는 점을 유의해야 한다. 만약  $n$ 개 샘플들을 이용해서  $h(\theta)$ 의 스펙트럴밀도함수  $f(h(\theta))$ 를  $\hat{f}(h(\theta))$ 로 추정하면,  $h(\theta)$ 의 분산추정값은  $2\pi\hat{f}(0)/n$ 이다. 또한, 다음과 같은 일괄평균(batch mean)를 사용할 수도 있다.

샘플들  $\theta^{(j)}$ , ( $j = 1, 2, \dots, n$ )에서  $\theta^{(j)}$ 와  $\theta^{(j+m)}$ 이 근사적으로 독립이 되도록 충분히 큰  $m$ 을 선택하자. 자연수  $k = n/m$ 에 대해서,  $n$ 개 샘플들을  $k$ 개 그룹들(batches)으로 나누고,

다음 통계량들을 정의하자.

$$\bar{h} \doteq \frac{1}{n} \sum_{j=1}^n h(\boldsymbol{\theta}^{(j)}), \quad \bar{h}_i \doteq \frac{1}{m} \sum_{i=[i-1]m+1}^{im} h(\boldsymbol{\theta}^{(i)}), \quad (i = 1, 2, \dots, k) \quad (6.3.10)$$

이 통계량들을 이용해서,  $h(\boldsymbol{\theta})$ 의 분산을 다음과 같이 추정할 수 있다.

$$\hat{\sigma}_h^2 \doteq \frac{m}{k-1} \sum_{i=1}^k [\bar{h}_i - \bar{h}]^2 \quad (6.3.11)$$

또한,  $\bar{h}$ 의 분산추정값은  $\hat{\sigma}_h^2/n$ 이다.

여러 모형들 중에서 어떤 모형이 주어진 데이터세트에 가장 적합한가를 선택하는 기준으로서 주변우도(marginal likelihood)나 주변우도들의 Bayes팩터(Bayes factor)를 사용하기도 한다. 지금부터는 사후확률분포가 정칙이라고 가정하자. 후보가 되는 모형들을  $M_1, M_2, \dots, M_K$ 라고 하고, 그에 해당하는 모수들, 사전확률분포들, 그리고 관찰벡터  $\mathbf{y}$ 의 확률밀도함수를  $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_K, \pi_1(\boldsymbol{\theta}_1), \pi_2(\boldsymbol{\theta}_2), \dots, \pi_K(\boldsymbol{\theta}_K)$ , 그리고  $f_1(\mathbf{y}|\boldsymbol{\theta}_1), f_2(\mathbf{y}|\boldsymbol{\theta}_2), \dots, f_K(\mathbf{y}|\boldsymbol{\theta}_K)$ 로 표기하자. 모형  $M_i$ 의 주변우도  $L_i(\mathbf{y})$ 는 다음과 같다.

$$L_i(\mathbf{y}) = \int f_i(\mathbf{y}|\boldsymbol{\theta}_i) \pi_i(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i \quad (6.3.12)$$

제*i*번째 주변우도가 크다는 것은 제*i*번째 모형에서 관찰벡터  $\mathbf{y}$ 가 발생하기 쉽다는 것이며, 또한 제*i*번째 모형의 적합도가 좋은 것을 의미한다. 모형  $M_i$ 의 주변우도를 모형  $M_j$ 의 주변우도로 나눈 것을 Bayes팩터  $B_{ij}$ 라고 한다. 즉, 다음 식들이 성립한다.

$$B_{ij} \doteq \frac{L_i(\mathbf{y})}{L_j(\mathbf{y})} = \frac{\pi_i(\boldsymbol{\theta}_i|\mathbf{y})/\pi_i(\boldsymbol{\theta}_i)}{\pi_j(\boldsymbol{\theta}_j|\mathbf{y})/\pi_j(\boldsymbol{\theta}_j)} \quad (6.3.13)$$

Jeffreys [56]는 모형  $M_i$ 가 모형  $M_j$ 보다 주어진 데이터세트에 더 타당한지 여부를 결정하기 위해서, 다음과 같은 판단기준을 사용할 것을 제안하였다. 만약  $\log_{10} B_{ij}$ 가 구간  $(2, \infty)$ 에 속하면, 결정적으로(decisively) 증거가 있다. 만약  $\log_{10} B_{ij}$ 가 구간  $(1, 2)$ 에 속하면, 증거가 강하게(strongly) 있다. 만약  $\log_{10} B_{ij}$ 가 구간  $(0.5, 1)$ 에 속하면, 증거가 충분히(substantially) 있다. 만약  $\log_{10} B_{ij}$ 가 구간  $(0, 0.5)$ 에 속하면, 증거가 충분하다고 할 수 없다.

## 제 6.4 절 MH 샘플러와 베이저안통계모형

### 6.4.1 MH 샘플러를 사용한 이항확률모형분석

시행횟수가  $n$  이고 성공확률  $\theta$  인 이항확률분포에서  $\theta$  의 지지대가  $S = \{0.1, 0.3, 0.5, 0.7, 0.9\}$  라고 하자. 모수  $\theta$  의 사전확률밀도함수가 다음과 같다고 하자.

$$\pi(\theta) = \frac{1}{5}, \quad (\theta = 0.1, 0.3, 0.5, 0.7, 0.9) \quad (6.4.1)$$

만약 시행횟수가  $n = 10$  이고 성공횟수가  $x = 7$  이면, 사후확률밀도함수는 다음과 같다.

$$\pi(\theta|x) \propto \theta^7 [1 - \theta]^3, \quad (\theta = 0.1, 0.3, 0.5, 0.7, 0.9) \quad (6.4.2)$$

이 사후확률밀도함수를 구하기 위해서는 다음 값을 계산해야 한다.

$$C = \sum_{k=1}^5 [0.2k - 0.1]^7 [0.9 - 0.2k]^3 \quad (6.4.3)$$

다음 식이 성립함을 알 수 있다.

$$C = \frac{187}{49820} \approx 0.00375 \quad (6.4.4)$$

여기서는 식 (6.4.4) 의  $C$  값을 사용하지 않고, Metropolis 샘플러를 사용해서 사후확률밀도함수를 구해보자.

다음 제안행렬을 생각해보자.

$$\mathbb{Q}_1 = \begin{matrix} & \begin{matrix} 0.1 & 0.3 & 0.5 & 0.7 & 0.9 \end{matrix} \\ \begin{matrix} 0.1 \\ 0.3 \\ 0.5 \\ 0.7 \\ 0.9 \end{matrix} & \begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \end{bmatrix} \end{matrix} \quad (6.4.5)$$

제안행렬  $\mathbb{Q}_1$  에서는 각 추이확률이  $q_{ij} = 0.2$  이다. 따라서, 행렬  $\mathbb{Q}_1$  이 Markov 행렬임이

자명하다. 또한, 다음과 같은 제안행렬을 생각해보자.

$$Q_2 = \begin{matrix} & \begin{matrix} 0.1 & 0.3 & 0.5 & 0.7 & 0.9 \end{matrix} \\ \begin{matrix} 0.1 \\ 0.3 \\ 0.5 \\ 0.7 \\ 0.9 \end{matrix} & \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \end{bmatrix} \end{matrix} \quad (6.4.6)$$

제안행렬  $Q_2$ 에서는 상태가 0.1 또는 0.9인 경우를 제외하고 반드시 이웃하는 상태로만 이동한다. 또한, 상태가 0.1 또는 0.9인 경우에는 자신과 이웃하는 상태로만 이동한다. 따라서, 행렬  $Q_2$ 도 Markov 행렬이다. 행렬  $Q_1$ 이나 행렬  $Q_2$  모두 대칭이므로, Metropolis샘플러를 이용할 수 있다.

**예제 6.4.1** Metropolis샘플러를 적용해서 식 (6.4.2)의 사후확률밀도함수를 구하기 위해서, 다음 MATLAB 프로그램 MetropolisSampler104.m을 실행해 보자.

```

1 % -----
2 % Filename: MetropolisSampler104.m
3 % Posterior Distribution of Discrete Uniform RV by Metropolis Sampler
4 % Programmed by CBS
5 % -----
6 function MetropolisSampler104
7 clear all, close all
8 % Prior Distribution
9 theta = 0.1:0.2:0.9
10 prior = 0.2*ones(1,5)
11 % MCMC
12 n = 1000; b = 100;
13 Q1 = 0.2*ones(5,5)
14 Q2 = zeros(5,5);
15 Q2(1,1) = 0.5; Q2(5,5) = 0.5;
16 for kk=1:4
17     Q2(kk,kk+1) = 0.5;
18     Q2(kk+1,kk) = 0.5;
19 end
20 Q2
21 % Initial Value
22 x = GenerateDiscreteRS(theta,prior,2)';
23 tt = 1;
24 % Iteration
25 for tt=2:n+b
26     ii = round((x(:,tt-1)-0.1)/0.2+1);
27     y(1) = GenerateDiscreteRS(theta,Q1(ii(1),:),1);
28     y(2) = GenerateDiscreteRS(theta,Q2(ii(2),:),1);
29     jj = round((y-0.1)/0.2+1);
30     alpha = min(1,prior(jj)./prior(ii));

```

```

31     u = rand(2,1);
32     for kk=1:2
33         if u(kk) <= alpha(kk)
34             x(kk,tt) = y(kk);
35         else
36             x(kk,tt) = x(kk,tt-1);
37         end
38     end
39 end
40 % Time Series Plot
41 subplot(2,1,1)
42 plot(101:150,x(1,101:150),'g-',101:150,x(1,101:150),'k*','linewidth',1.5)
43 set(gca,'fontsize',11,'fontweigh','bold')
44 axis([101 150 0 1 ])
45 xlabel('\bf Time','fontsize',12)
46 ylabel('\bf State by Q_{1}','fontsize',12)
47 subplot(2,1,2)
48 plot(101:150,x(2,101:150),'g-',101:150,x(2,101:150),'k*','linewidth',1.5)
49 set(gca,'fontsize',11,'fontweigh','bold')
50 axis([101 150 0 1 ])
51 xlabel('\bf Time','fontsize',12)
52 ylabel('\bf State by Q_{2}','fontsize',12)
53 saveas(gcf,'MetropolisSampler104a','eps')
54 % Histogram
55 xHistL = min(theta);
56 xHistR = max(theta);
57 HistIncrement = 0.2;
58 PMFind = (xHistL:HistIncrement:xHistR)';
59 [Nbar1, xcenter1] = hist(x(1,b+1:b+n), PMFind);
60 [Nbar2, xcenter2] = hist(x(2,b+1:b+n), PMFind);
61 Nheight1 = Nbar1/n;
62 Nheight2 = Nbar2/n;
63 figure
64 subplot(1,2,1)
65 hold on
66 bar(xcenter1, Nheight1, 0.2, 'g');
67 plot([0 1],[0.2,0.2],'r--','linewidth',2)
68 set(gca,'fontsize',11,'fontweigh','bold','xtick',0.1:0.2:0.9)
69 legend('Histogram','Uniform PMF');
70 ylabel('\bf PMF by Q_{1}','fontsize',12)
71 axis([0 1 0 0.3])
72 hold off;
73 subplot(1,2,2)
74 hold on
75 bar(xcenter2, Nheight2, 0.2, 'g');
76 plot([0 1],[0.2,0.2],'r--','linewidth',2)
77 set(gca,'fontsize',11,'fontweigh','bold','xtick',0.1:0.2:0.9)
78 legend('Histogram','Uniform PMF');
79 ylabel('\bf PMF by Q_{2}','fontsize',12)
80 axis([0 1 0 0.3])
81 saveas(gcf,'MetropolisSampler104b','eps')
82 save('MetropolisSampler104.txt','x','-ascii')
83 hold off;
84 end
85 % -----
86 % Filename: GenerateDiscreteRS.m
87 % Generate Discrete Random Samples
88 % -----
89 function RS = GenerateDiscreteRS(x,p,nrs)
90 % x = Support set (i.e., domain) of the distribution
91 % p = Probability density function

```

```

92 % nrs = Number of the generated random numbers
93 % (Ex) RS = GenerateDiscreteRS([1 3 5],[0.2 0.5 0.3],1200)
94 cumdf = cumsum(p);
95 RS = zeros(1,nrs);
96 for ii = 1:1:nrs
97     u = rand;
98     idum = 1;
99     while cumdf(idum) <= u
100         idum = idum+1;
101     end;
102     RS(ii) = x(idum);
103 end
104 end
105 % End of the function
106 % -----

```

이 MATLAB 프로그램 MetropolisSampler104.m을 실행하면, 식 (6.4.5)와 식 (6.4.6)의 제안행렬들을 사용한 Metropolis샘플러들을 적용해서 MH샘플들을 발생시킨다. 이때 번인 개수는  $b = 100$ 이고 MH샘플 개수는  $n = 1000$ 이다.

각 제안행렬에 대한 시계열산점도가 그림 6.4.1에 그려져 있다. 그림 6.4.1에서 상단 그래프는 제안행렬  $Q_1$ 에 의해 발생시킨 시간구간  $[101, 150]$ 에서 MH샘플들의 시계열산점도이고, 그리고 하단 그래프는 제안행렬  $Q_2$ 에 의해 발생시킨 시간구간  $[101, 150]$ 에서 MH샘플들의 시계열산점도이다. 또한, 이렇게 발생된 사후확률질량함수들이 그림 6.4.2에 그려져 있다. 그림 6.4.2에서 좌측 그래프는 제안행렬  $Q_1$ 에 의해 발생시킨 사후확률질량함수이고, 우측 그래프는 제안행렬  $Q_2$ 에 의해 발생시킨 사후확률질량함수이다. 그림 6.4.1과 그림 6.4.2에서 알 수 있듯이, 제안행렬  $Q_2$ 를 사용한 Metropolis샘플러에서는 좀 더 큰 번인기간을 사용했어야 할 것 같다. ■

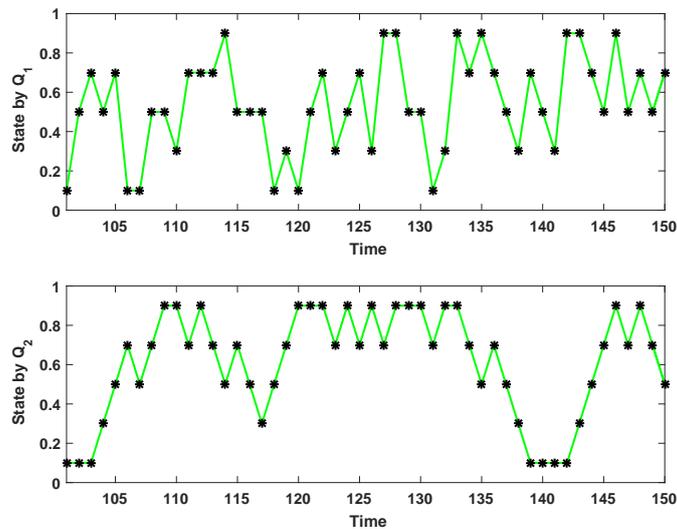


그림 6.4.1. Metropolis샘플들의 시계열산포도

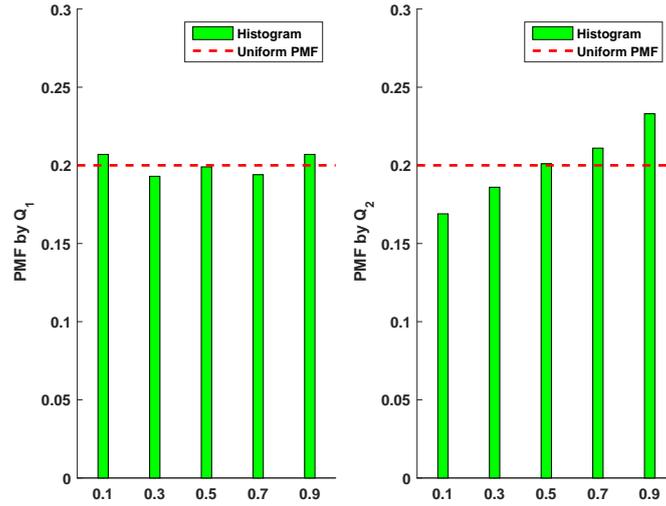


그림 6.4.2. Metropolis 샘플들의 히스토그램

### 6.4.2 MH 샘플러를 사용한 로지스틱 회귀 모형 분석

시점  $i$ 에서 종속변수를  $y_i$  그리고 설명변수 벡터를  $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ip}]^t$  라 하고, 로지스틱 회귀 모형의 모수 벡터를  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_p]^t$  라고 하면, 로지스틱 회귀 모형은 다음과 같다.

$$\Pr(y_i = 1) = \frac{\exp(\mathbf{x}_i^t \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^t \boldsymbol{\beta})} = 1 - \Pr(y_i = 0), \quad (i = 1, 2, \dots, n) \quad (6.4.7)$$

모수 벡터  $\boldsymbol{\beta}$ 의 사전 확률 분포가 다음과 같다고 가정하자.

$$\boldsymbol{\beta} \stackrel{d}{\sim} \mathcal{N}_p(\mathbf{b}_0, B_0) \quad (6.4.8)$$

관찰 벡터  $\mathbf{y} = [y_1, y_2, \dots, y_n]^t$ 가 주어졌을 때 모수 벡터  $\boldsymbol{\beta}$ 의 우도 함수는 다음과 같다.

$$f(\mathbf{y} | \boldsymbol{\beta}) = \prod_{i=1}^n \left[ \frac{\exp(\mathbf{x}_i^t \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^t \boldsymbol{\beta})} \right]^{y_i} \left[ \frac{1}{1 + \exp(\mathbf{x}_i^t \boldsymbol{\beta})} \right]^{1-y_i} \quad (6.4.9)$$

따라서,  $\boldsymbol{\beta}$ 의 사후 확률 밀도 함수는 다음과 같다.

$$\pi(\boldsymbol{\beta} | \mathbf{y}) \propto \frac{\prod_{i=1}^n [\exp(\mathbf{x}_i^t \boldsymbol{\beta})]^{y_i}}{\prod_{i=1}^n [1 + \exp(\mathbf{x}_i^t \boldsymbol{\beta})]} \exp\left(-\frac{1}{2} [\boldsymbol{\beta} - \mathbf{b}_0]^t B_0^{-1} [\boldsymbol{\beta} - \mathbf{b}_0]\right) \quad (6.4.10)$$

MH 샘플러를 사용해서 이 사후 확률 밀도 함수에서 샘플링하기 위해서, 먼저 제안 분포를 구축하기로 하자. 사후 확률 밀도 함수에서 표준화 상수를 제외한 나머지의 대수를  $g(\boldsymbol{\beta})$ 라 하자.

다음 식이 성립한다.

$$g(\boldsymbol{\beta}) = \sum_{i=1}^n y_i \mathbf{x}_i^t \boldsymbol{\beta} - \sum_{i=1}^n \ln(1 + \exp(\mathbf{x}_i^t \boldsymbol{\beta})) - \frac{1}{2} [\boldsymbol{\beta} - \mathbf{b}_0]^t B_0^{-1} [\boldsymbol{\beta} - \mathbf{b}_0] \quad (6.4.11)$$

이  $g(\boldsymbol{\beta})$ 를  $\boldsymbol{\beta}_*$ 의 주변에서 Taylor근사식을 구하면 다음과 같다.

$$g(\boldsymbol{\beta}) \approx g(\boldsymbol{\beta}_*) + [\boldsymbol{\beta} - \boldsymbol{\beta}_*]^t \frac{\partial g(\boldsymbol{\beta}_*)}{\partial \boldsymbol{\beta}} + \frac{1}{2} [\boldsymbol{\beta} - \boldsymbol{\beta}_*]^t \frac{\partial^2 g(\boldsymbol{\beta}_*)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t} [\boldsymbol{\beta} - \boldsymbol{\beta}_*] \quad (6.4.12)$$

여기서  $\boldsymbol{\beta}_*$ 로는 사후확률분포에서 모드(mode) 또는 그 근사값으로 한다. 근사식 (6.4.12)의 우변을  $h(\boldsymbol{\beta})$ 로 표기하자. 이  $h(\boldsymbol{\beta})$ 를 다음과 같이 쓸 수 있다.

$$h(\boldsymbol{\beta}) = a_0 - \frac{1}{2} [\boldsymbol{\beta} - \mathbf{m}]^t V^{-1} [\boldsymbol{\beta} - \mathbf{m}] \quad (6.4.13)$$

여기서  $\mathbf{m}$ 과  $V$ 는 각각 다음과 같다.

$$\mathbf{m} \doteq \boldsymbol{\beta}_* + V \frac{\partial g(\boldsymbol{\beta}_*)}{\partial \boldsymbol{\beta}}, \quad V \doteq - \left[ \frac{\partial^2 g(\boldsymbol{\beta}_*)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t} \right]^{-1} \quad (6.4.14)$$

다음과 같은 제안분포를 사용하자.

$$\boldsymbol{\beta} \stackrel{d}{\sim} \mathcal{N}_p(\mathbf{m}, V) \quad (6.4.15)$$

즉, 제안분포의 조건부확률밀도함수는  $q(\boldsymbol{\beta} | \mathbf{y}) = n(\boldsymbol{\beta} | \mathbf{m}, V)$ 이다. 다음 식이 성립한다.

$$\frac{\pi(\boldsymbol{\beta} | \mathbf{y})}{c \cdot n(\boldsymbol{\beta} | \mathbf{m}, V)} = \exp(g(\boldsymbol{\beta}) - h(\boldsymbol{\beta})) \quad (6.4.16)$$

여기서  $c$ 는 상수이다. 따라서, 이 독립샘플러에서는 제  $j-1$ 번째 단계에서 모수벡터를  $\boldsymbol{\beta}^{(j-1)}$ 라고 하자. 식 (6.4.15)의 제안분포를 사용해서 샘플후보  $\boldsymbol{\beta}$ 를 발생시키고, 다음 MH채택확률  $\alpha(\boldsymbol{\beta}^{(j-1)}, \boldsymbol{\beta} | \mathbf{y})$ 로 샘플후보를 채택한다.

$$\begin{aligned} \alpha(\boldsymbol{\beta}^{(j-1)}, \boldsymbol{\beta} | \mathbf{y}) &= \min \left\{ \frac{\pi(\boldsymbol{\beta} | \mathbf{y}) / n(\boldsymbol{\beta} | \mathbf{m}, V)}{\pi(\boldsymbol{\beta}^{(j-1)} | \mathbf{y}) / n(\boldsymbol{\beta}^{(j-1)} | \mathbf{m}, V)}, 1 \right\} \\ &= \min \left\{ \exp(g(\boldsymbol{\beta}) - h(\boldsymbol{\beta}) - g(\boldsymbol{\beta}^{(j-1)}) + h(\boldsymbol{\beta}^{(j-1)})), 1 \right\} \end{aligned} \quad (6.4.17)$$

지금까지 내용을 바탕으로, 회귀계수벡터  $\boldsymbol{\beta}$ 의 사후확률분포에서 샘플링하는 MH샘플러를

다음과 같이 요약할 수 있다.

#### 알고리즘 6.4.1: 로지스틱회귀모형의 MH샘플러

- (1단계) 초기모수  $\beta^{(0)}$  를 선택한다.
- (2단계) 사후확률분포의 모드  $\beta_*$  를 구한다.
- (3단계) 각  $j (= 1, 2, \dots)$  에 대해서 다음과 같이  $\beta^{(j)}$  를 발생시킨다.
- (3-1) 식 (6.4.15) 를 사용해서, 회귀계수의 후보  $\beta$  를 발생시킨다.
- (3-2) 식 (6.4.17) 을 사용해서,  $\alpha(\beta^{(j-1)}, \beta | \mathbf{y})$  를 계산한다.
- (4단계) 독립적으로 일양난수  $u$  를 발생시킨다. 만약 식  $\alpha(\beta^{(j-1)}, \beta | \mathbf{y}) \geq u$  가 성립하면, 회귀계수벡터의 후보  $\beta$  를  $\beta^{(j)}$  로 채택한다. 만약 그렇지 않으면,  $\beta^{(j-1)}$  을  $\beta^{(j)}$  로 채택한다.
- (5단계) 충분히 큰  $b$  에 대해서  $\beta^{(j)}$ , ( $j = b + 1, b + 2, \dots$ ) 를 기록한다.

**예제 6.4.2** 이 예제는 Robert & Casella [pp. 15-6][84]의 예제 1.13을 다시 다루는 것이다.

1986년 스페이스셔틀 챌린저호가 이륙하는 과정에서 폭발해서, 우주인 일곱 명이 사망했다. 이 폭발의 원인은 우주선의 부품들을 봉합하는 O링(O-ring)이 잘 작동하지 않았기때문으로 알려졌다. 그 날은 기온이 화씨 31도로서 기록적으로 낮았고, 또한 기온이 하강할수록 O링이 제대로 작동하지 않을 확률이 높아진다고 알려져있다. 이러한 가설을 검정하기 위해서, 다음 R프로그램 LogisticRegressionMH101R.R을 실행해 보자.

```

1 # -----
2 # Filename:LogisticRegressionMH101R.R
3 # Generic Metropolis-Hastings for Logit Model
4 # Robert and Casella (2004)
5 # About tThe space shuttle Challenger explosion in 1986
6 # -----
7 # ipak function: install and load multiple R packages.
8 # check to see if packages are installed.
9 # Install them if they are not, then load them into the R session.
10 ipak <- function(pkg){
11     new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
12     if (length(new.pkg))
13         install.packages(new.pkg, dependencies = TRUE)
14     sapply(pkg, require, character.only = TRUE)
15 }
16 # usagpropR

```

```

17 packages <- c("ggplot2", "grid", "Rcpp")
18 # -----
19
20 MH <- function(x0, fun, propD, propR, N, Burn) {
21   xx <- matrix(NA, Nsmpl+Burn, length(x0));
22   ff <- rep(NA, Nsmpl+Burn)
23   xx[1,] <- x0;
24   ff[1] <- fun(x0);
25   ct <- 0
26   for(ii in 2:(Nsmpl+Burn)) {
27     u <- propR(xx[ii-1,])
28     fu <- fun(u)
29     r <- log(fu) + log(propD(xx[ii-1,],u))
30     r <- r-log(ff[ii-1])-log(propD(u,xx[ii-1,]))
31     dumR <- min(exp(r), 1)
32     if(runif(1) <= dumR) {
33       ct <- ct+1
34       xx[ii,] <- u
35       ff[ii] <- fu
36     } else {
37       xx[ii,] <- xx[ii-1,]
38       ff[ii] <- ff[ii-1]
39     }
40   }
41   return(list(xx=xx[-(1:Burn),], ff=ff[-(1:Burn)], rate=ct/(Nsmpl+Burn)))
42 }
43
44 # Data
45 y <- c(1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0)
46 x <- c(53, 57, 58, 63, 66, 67, 67, 67, 68, 69, 70, 70, 70, 70, 72, 73, 75,
47       75, 76, 76, 78, 79, 81)
48
49 # Preliminary output from ML estimation
50 logregrOut <- glm(y ~ x, family=binomial(logit))
51 summary(logregrOut)
52 a.mle <- as.numeric(logregrOut$coefficients[1])
53 b.mle <- as.numeric(logregrOut$coefficients[2])
54 var.a.mle <- summary(logregrOut)$cov.scaled[1, 1]
55 var.b.mle <- summary(logregrOut)$cov.scaled[2, 2]
56 b.mme <- exp(a.mle + 0.577216)
57
58 # Posterior distribution
59 postD <- function(theta) {
60   a <- theta[1]
61   b <- theta[2]
62   p <- 1-1/(1 + exp(a + b * x))
63   likeli <- exp(sum(dbinom(y, size=1, prob=p, log=TRUE)))
64   priorD <- exp(a)*exp(-exp(a)/b.mme)/b.mme
65   return(likeli*priorD)
66 }
67
68 # Proposal distribution (independent proposal, so "theta0" is not used)
69 propD <- function(theta, theta0) {
70   a <- theta[1]
71   b <- theta[2]
72   pr1 <- exp(a)/b.mme*exp(-exp(a)/b.mme)
73   pr2 <- dnorm(b, b.mle, sqrt(var.b.mle))
74   return(pr1*pr2)
75 }
76
77 propR <- function(theta0) {

```

```

78     a <- log(rexp(1, 1/b.mme))
79     b <- rnorm(1, b.mle, sqrt(var.b.mle))
80     return(c(a, b))
81 }
82
83 # Run Metropolis-Hastings
84 Nsmpl <- 30000
85 Burn <- round(0.1*Nsmpl)
86 MH.out <- MH(c(a.mle, b.mle), postD, propD, propR, Nsmpl, Burn)
87 alpha.MH <- MH.out$xx[,1]
88 beta.MH <- MH.out$xx[,2]
89 p65 <- 1-1/(1+exp(alpha.MH+beta.MH*65))
90 p45 <- 1-1/(1+exp(alpha.MH+beta.MH*45))
91
92 # Plotting
93 setEPS()
94 ## Plot 1
95 plot.new()
96 # postscript('LogisticRegressionMH101aR.eps') # Start to save figure
97 par(mfrow=c(2,2))
98 hist(alpha.MH, freq=FALSE, col="gray", border="white",
99       xlab=expression(alpha))
100 plot(alpha.MH, type="l", col="gray", xlab="Iteration",
101       ylab=expression(alpha))
102 lines(1:Nsmpl, cumsum(alpha.MH)/(1:Nsmpl), col="red", lwd=2)
103 hist(beta.MH, freq=FALSE, col="gray", border="white",
104       xlab=expression(beta))
105 plot(beta.MH, type="l", col="gray", xlab="Iteration",
106       ylab=expression(beta))
107 lines(1:Nsmpl, cumsum(beta.MH)/(1:Nsmpl), col="red", lwd=2)
108 # dev.off() # End to save figure
109 ## Plot 2
110 plot.new()
111 # postscript('LogisticRegressionMH101bR.eps') # Start to save figure
112 par(mfrow=c(1,2))
113 hist(p65, freq=FALSE, col="gray", border="white", xlab="p(65)", main="")
114 hist(p45, freq=FALSE, col="gray", border="white", xlab="p(45)", main="")
115 # dev.off() # End to save figure
116
117 # -----

```

주어진 온도  $x$ 에 대해서 성공확률  $p(x) \doteq \Pr(Y = 1 | X = x)$ 가 다음 로지스틱회귀모형을 만족한다고 하자.

$$\ln \frac{p(x)}{1-p(x)} = \alpha + \beta x \quad (1)$$

이 R 프로그램 LogisticRegressionMH101R.R을 실행하면, 회귀계수들  $\alpha$ 와  $\beta$ 의 MH샘플들의 히스토그램들과 시계열산점도들을 그린 그림 6.4.3을 그린다. 그림 6.4.3의 좌측상단 그래프는 회귀계수  $\alpha$ 의 MH샘플들의 히스토그램을 그린 것이고, 우측상단 그래프는 MH샘플들의 시계열산점도를 그린 것이고, 좌측하단 그래프는 회귀계수  $\beta$ 의 MH샘플들의 히스토그램을 그린 것이고, 우측하단 그래프는 MH샘플들의 시계열산점도를 그린 것이다.

다음 식들이 성립한다.

$$p(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} = 1 - \frac{1}{1 + \exp(\alpha + \beta x)} \quad (2)$$

이 R프로그램 LogisticRegressionMH101R.R을 실행하면, MH샘플러를 사용해서 구한 확률  $p(0.65)$ 와 확률  $p(0.45)$ 의 히스토그램들이 그림 6.4.4에 출력된다. 그림 6.4.4의 상단 그래프는 확률  $p(0.65)$ 의 히스토그램을, 그리고 하단 그래프는 확률  $p(0.45)$ 의 히스토그램을 그린 것이다. ■

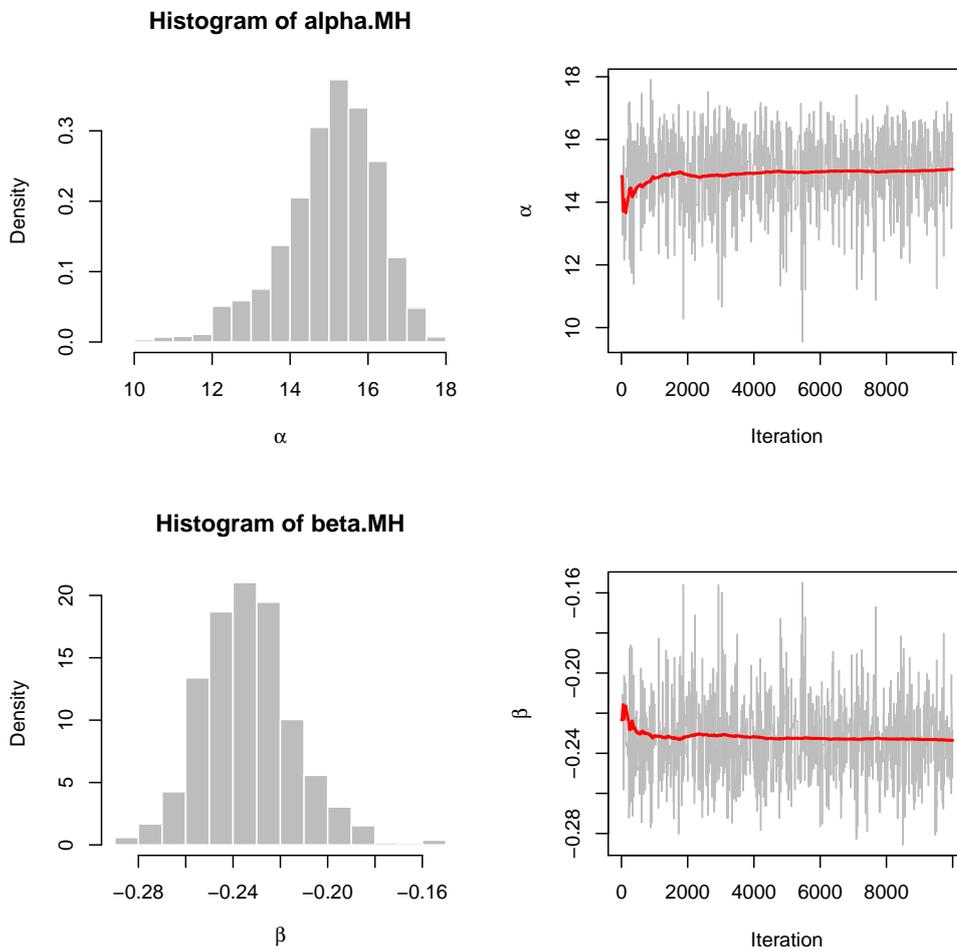


그림 6.4.3. 로지스틱회귀모형 I

식 (6.4.15)에서는  $\mathcal{N}_p(\mathbf{m}, V)$ 를 제안분포로 사용한다. 여기서는 제안분포의 조건부확률

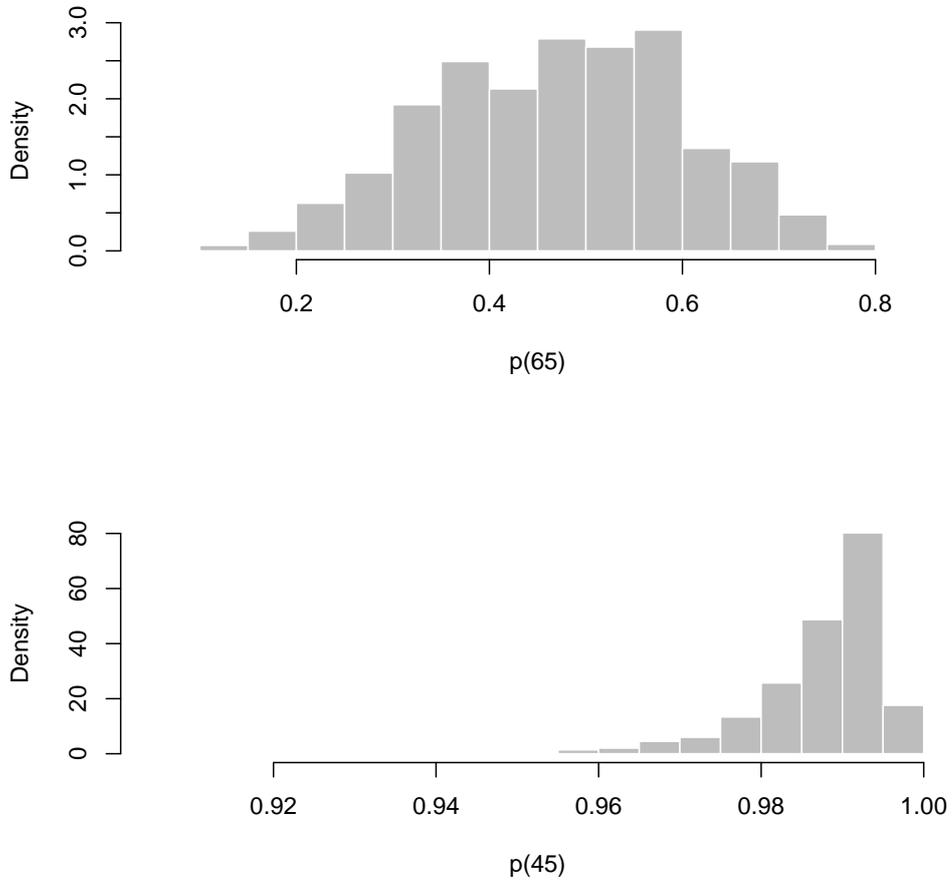


그림 6.4.4. 로지스틱회귀모형 II

밀도함수로 다음 함수를 사용하는 AR-MH 샘플러를 살펴보자.

$$q(\boldsymbol{\beta} | \mathbf{y}) \propto \min \{ \pi(\boldsymbol{\beta} | \mathbf{y}), c_1 n(\boldsymbol{\beta} | \mathbf{m}, V) \} \tag{6.4.18}$$

여기서  $c_1$ 은 상수이다. 식 (6.4.18)에서 알 수 있듯이, 다음 식이 성립한다.

$$\min \{ \pi(\boldsymbol{\beta} | \mathbf{y}), c_1 n(\boldsymbol{\beta} | \mathbf{m}, V) \} \leq c_1 n(\boldsymbol{\beta} | \mathbf{m}, V) \tag{6.4.19}$$

즉, 다음 식을 만족하는 상수  $c_2$ 가 존재한다.

$$q(\boldsymbol{\beta} | \mathbf{y}) \leq c_2 n(\boldsymbol{\beta} | \mathbf{m}, V) \tag{6.4.20}$$

첫째, 채택기각법을 적용하는 단계, 즉 AR 단계 (AR-step)를 살펴보자. 식 (6.4.18)의 제안

분포의 조건부확률밀도함수에서 샘플링을 하기 위해서, 우선 다음 확률분포를 따르는  $\beta$ 를 생성한다.

$$\beta \stackrel{d}{\sim} \mathcal{N}(\mathbf{m}, V) \quad (6.4.21)$$

다음 확률로 후보점  $\beta$ 를 채택한다.

$$\min \left\{ \frac{\pi(\beta | \mathbf{y})}{c_2 \cdot n(\beta | \mathbf{m}, V)}, 1 \right\} = \min \{ \exp(g(\beta) - h(\beta)), 1 \} \quad (6.4.22)$$

만약  $\beta$ 가 기각되면, 식 (6.4.21)의 확률분포에서  $\beta$ 가 채택될 때까지 새로운 후보점을 생성한다. 둘째, MH샘플러를 적용해보자. AR단계에서 발생된  $\beta$ 는 다음 식들을 만족한다.

$$\frac{\pi(\beta | \mathbf{y})}{q(\beta | \mathbf{y})} = \frac{c_3 \exp(g(\beta))}{\min \{ \exp(g(\beta)), \exp(h(\beta)) \}} = \frac{c_3}{\min \{ \exp(h(\beta) - g(\beta)), 1 \}} \quad (6.4.23)$$

여기서  $c_3$ 는 상수이다. 따라서, 제  $j - 1$  번째 단계에서 모수벡터를  $\beta^{(j-1)}$ 라고 하면, AR 단계에서 발생시킨 후보점  $\beta$ 를 다음과 같은 확률  $\alpha(\beta^{(j-1)}, \beta | \mathbf{y})$ 로 채택한다.

$$\begin{aligned} \alpha \left( \beta^{(j-1)}, \beta \mid \mathbf{y} \right) &= \min \left\{ \frac{\pi(\beta | \mathbf{y})/q(\beta | \mathbf{y})}{\pi(\beta^{(j-1)} | \mathbf{y})/q(\beta^{(j-1)} | \mathbf{y})}, 1 \right\} \\ &= \min \left\{ \frac{\min \{ \exp(h(\beta^{(j-1)}, 1) - g(\beta^{(j-1)})) \}}{\min \{ 1, \exp(h(\beta) - g(\beta)) \}}, 1 \right\} \end{aligned} \quad (6.4.24)$$

이러한 과정을 MH단계 (MR-step)라 부른다.

지금까지 내용을 바탕으로, 로지스틱회귀모형의 회귀계수벡터  $\beta$ 의 사후확률분포에서 샘플링하는 AR-MH샘플러를 다음과 같이 요약할 수 있다.

#### 알고리즘 6.4.2: 로지스틱회귀모형의 AR-MH샘플러

- (1단계) 초기모수  $\beta^{(0)}$ 를 선택한다.
- (2단계) 사후확률분포의 모드  $\beta_*$ 를 구한다.
- (3단계) 각  $j (= 1, 2, \dots)$ 에 대해서 다음과 같이  $\beta^{(j)}$ 를 발생시킨다.
  - (3-1) 식 (6.4.21)를 사용해서 후보점  $\beta$ 를 발생시킨다. 이와는 독립적으로 일양난수  $u_1$ 을 발생시킨다. 만약 식  $u_1 \leq \exp(g(\beta) - h(\beta))$ 가 성립하면, 후보점  $\beta$ 를 채택하고

제3-2단계로 넘어간다. 그렇지 않으면, 식 (6.4.21)를 사용해서 후보점  $\beta$ 를 다시 발생시킨다.

(3-2) 식 (6.4.24)를 사용해서, 확률  $\alpha(\beta^{(j-1)}, \beta | \mathbf{y})$ 를 계산한다. 이와는 독립적으로 일양난수  $u_2$ 를 발생시킨다. 만약 식  $\alpha(\beta^{(j-1)}, \beta | \mathbf{y}) \geq u_2$ 가 성립하면, 회귀계수벡터의 후보  $\beta$ 를  $\beta^{(j)}$ 로 채택한다. 만약 그렇지 않으면,  $\beta^{(j-1)}$ 을  $\beta^{(j)}$ 로 채택한다.

(4단계) 충분히 큰  $b$ 에 대해서  $\beta^{(j)}$ , ( $j = N, N + 1, \dots$ )를 기록한다.

**예제 6.4.3** R 패키지 mcmc에 포함된 R함수 metrop은 연속형 확률벡터의 Metropolis 샘플러를 구현한다. R함수 metrop의 간단한 사용법은 다음과 같다.

```
metrop(obj, initial, nbatch, scale)
```

여기서 ‘obj’는 목표확률밀도함수이고, ‘initial’은 사전확률밀도함수이고, ‘nbatch’는 생성할 Metropolis샘플수, 그리고 ‘scale’은 제안스텝크기(proposal step size)를 조절하는 모수이다. 만약  $x$ 가 현재상태이면, 제안스텝크기는  $x + scale \cdot z$ 이다. 여기서  $z$ 는 표준정규확률벡터이다.

Metropolis샘플러를 적용하기 위해서, 다음 R프로그램 MetropolisSampler103R.R을 실행해 보자.

```
1 # -----
2 # Filename: Metropolis103R.R
3 # Metropolis Sampler through mcmc package
4 # http://www.stat.umn.edu/geyer/mcmc/library/mcmc/doc/demo.pdf
5 # Modified by CBS
6 # -----
7 # install.packages("mcmc")
8 library(mcmc)
9
10 # Call dataset and GLM
11 data(logit)
12 head(logit)
13 ( nR <- nrow(logit) )
14 ( nC <- ncol(logit) )
15 outp <- glm(y ~ x1 + x2 + x3 + x4, data = logit, family = binomial())
16 summary(outp)
17
18 # Modify dataset
19 x <- logit
20 x$y <- NULL
21 head(x$y)
22 x <- as.matrix(x)
23 head(x)
24 x <- cbind(1, x)
25 head(x)
```

```

26 dimnames(x) <- NULL
27 head(x)
28 y <- logit$y
29
30 # log unnormalized posterior = log likelihood plus log prior
31 lupost <- function(beta, x, y) {
32   eta <- x %*% beta
33   p <- 1/(1 + exp(-eta))
34   logl <- sum(log(p[y == 1])) + sum(log(1-p[y == 0]))
35   return( logl + sum(dnorm(beta, 0, 2, log = TRUE)) )
36 }
37
38 # Metropolis algorithm to simulate the posterior.
39 set.seed(42)
40 ( beta.init <- as.numeric(coefficients(out)) )
41 nSim <- 1000
42 out <- metrop(lupost, beta.init, nSim, x = x, y = y)
43 names(out)
44 out$accept
45 out$time
46
47 # Plotting
48 # install.packages("ggplot2")
49 library(ggplot2)
50 # install.packages("grid")
51 library(grid)
52 setEPS()
53 plot.new()
54 postscript('MetropolisSampler103R.eps') # Start to save figure
55 timee = c(1:nSim)
56 MHdata <- data.frame(timee, out$batch[ ,1])
57 plot1 <- ggplot(MHdata, aes(timee, out$batch[ ,1])) +
58   geom_line(col="dark red", lwd=1.2) + xlab("time")
59 plot2 <- ggplot(MHdata, aes(timee, out$batch[ ,2])) +
60   geom_line(col="dark blue", lwd=1.2) + xlab("time")
61 plot3 <- ggplot(MHdata, aes(timee, out$batch[ ,3])) +
62   geom_line(col="dark red", lwd=1.2) + xlab("time")
63 plot4 <- ggplot(MHdata, aes(timee, out$batch[ ,4])) +
64   geom_line(col="dark blue", lwd=1.2) + xlab("time")
65 plot5 <- ggplot(MHdata, aes(timee, out$batch[ ,5])) +
66   geom_line(col="dark red", lwd=1.2) + xlab("time")
67 plot6 <- ggplot(MHdata, aes(out$batch[ ,4], out$batch[ ,5])) +
68   geom_point(colour="red", size=2)
69 pushViewport(viewport(layout = grid.layout(2,3)))
70 print(plot1, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
71 print(plot2, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))
72 print(plot3, vp = viewport(layout.pos.row = 1, layout.pos.col = 3))
73 print(plot4, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))
74 print(plot5, vp = viewport(layout.pos.row = 2, layout.pos.col = 2))
75 print(plot6, vp = viewport(layout.pos.row = 2, layout.pos.col = 3))
76 dev.off() # End to save figure
77 # -----

```

이 R 프로그램 MetropolisSampler103R.R을 실행하면, R 패키지 mcmc에 포함된 데이터 세트 logit을 읽어 들인다. 이 데이터세트의 앞 부분은 다음과 같다.

```

y      x1      x2      x3      x4
1 0 -0.162 -0.348 -0.524 -0.312

```

```

2 1  0.187 -0.410  0.362 -0.366
3 1  0.160  1.649 -0.664  0.051
4 1  1.536  0.084  0.403  1.732
5 1 -0.162 -0.061  0.192  0.256
6 0 -2.855 -3.341 -2.549 -1.461

```

이 데이터세트 logit에는 종속변수  $y$ 와 설명변수들  $x_1, x_2, x_3, x_4$ 로 구성되어 있다. 또한 관찰점들은 100개이다.

다음 단계로, 데이터세트 logit을 다음과 같은 이항로지스틱회귀모형(binary logistic regression model, binary logit regression model)에 적합시킨다.

$$\Pr(y = 1 | \boldsymbol{\beta}) \doteq \pi(\boldsymbol{\beta}) \doteq \frac{\exp(\mathbf{x}^t \boldsymbol{\beta})}{1 + \exp(\mathbf{x}^t \boldsymbol{\beta})} \quad (1)$$

여기서  $\mathbf{x} = [1, x_1, x_2, x_3, x_4]^t$ 이고  $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \beta_3, \beta_4]^t$ 이다. R함수 glm을 사용해서 데이터세트 logit을 이항로지스틱회귀모형에 적합시킨 결과는 다음과 같다.

```

Call:
glm(formula = y ~ x1 + x2 + x3 + x4, family = binomial(), data = logit)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7461  -0.6907   0.1540   0.7041   2.1943

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.6328     0.3007   2.104  0.03536 *
x1             0.7390     0.3616   2.043  0.04100 *
x2            1.1137     0.3627   3.071  0.00213 **
x3             0.4781     0.3538   1.351  0.17663
x4             0.6944     0.3989   1.741  0.08172 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 137.628  on 99  degrees of freedom
Residual deviance:  87.668  on 95  degrees of freedom

```

AIC: 97.668

Number of Fisher Scoring iterations: 6

따라서, 추정식은 다음과 같다.

$$\ln \frac{\pi(\boldsymbol{\beta})}{1 - \pi(\boldsymbol{\beta})} = 0.6328 + 0.7390x_1 + 1.1137x_2 + 0.4781x_3 + 0.6944x_4 + \hat{\epsilon} \quad (2)$$

베이지안통계분석을 하기 위해서, 다음과 같이 모수벡터  $\boldsymbol{\beta}$ 의 사전확률분포를 가정하자.

$$\boldsymbol{\beta} \stackrel{d}{\sim} \mathcal{N}(\mathbf{0}, 2I_5) \quad (3)$$

이 사전확률분포에 주어진 관찰값들의 정보를 담은 우도함수를 반영해서 사후확률분포를 구할 수 있다. 이 사후확률분포의 상수항을 구하는 과정을 피해가기 위해서, Metropolis샘플러를 적용해서 사후확률분포로부터 샘플링을 한다. 발생된 Metropolis샘플들의 앞 부분은 다음과 같다.

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	0.6327889	0.7390045	1.113686	0.4780903	0.6943901
[2,]	0.6327889	0.7390045	1.113686	0.4780903	0.6943901
[3,]	0.6327889	0.7390045	1.113686	0.4780903	0.6943901
[4,]	0.6327889	0.7390045	1.113686	0.4780903	0.6943901
[5,]	0.6327889	0.7390045	1.113686	0.4780903	0.6943901
[6,]	0.6327889	0.7390045	1.113686	0.4780903	0.6943901

발생된 샘플들의 시계열산점도가 그림 6.4.5에 그려져 있다. 그림 6.4.5의 상단좌측 그래프는  $\beta_0$ 의 사후확률샘플의 시계열산점도를 그린 것이고, 상단중간 그래프는  $\beta_1$ 의 사후확률샘플의 시계열산점도를 그린 것이고, 상단우측 그래프는  $\beta_2$ 의 사후확률샘플의 시계열산점도를 그린 것이고, 하단좌측 그래프는  $\beta_3$ 의 사후확률샘플의 시계열산점도를 그린 것이고, 하단중간 그래프는  $\beta_4$ 의 사후확률샘플의 시계열산점도를 그린 것이고, 하단우측 그래프는  $(\beta_3, \beta_4)$ 의 사후확률샘플의 산점도를 그린 것이다. ■

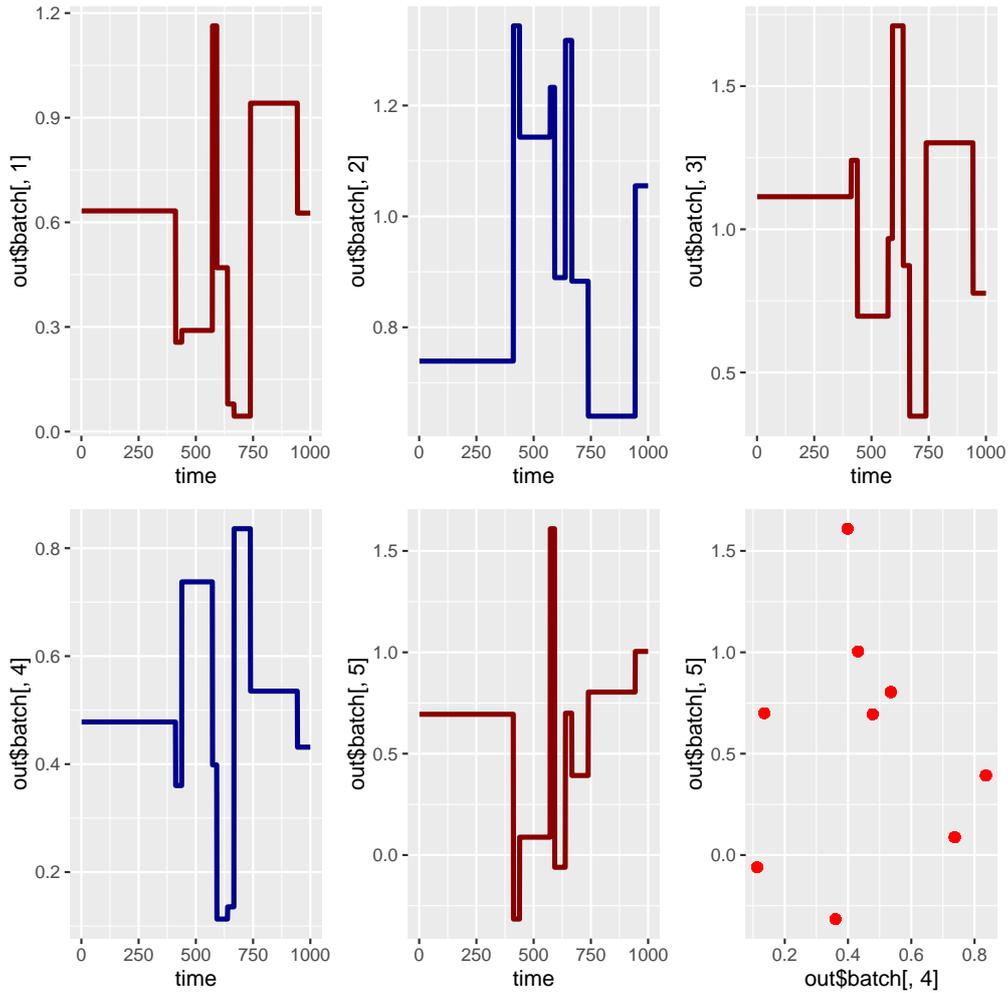


그림 6.4.5. Metropolis샘플러와 정규확률분포

### 6.4.3 MH샘플러를 사용한 AR모형분석과 ARX모형분석

다음과 같은 차수가 1인 자기회귀과정, 즉  $AR(1)$  과정  $\{y_t | t = 0, 1, \dots, n\}$ 을 살펴보자.

$$y_t = \phi y_{t-1} + \epsilon_t, \quad (t = 1, 2, \dots, n) \quad (6.4.25)$$

$$y_0 \stackrel{d}{\sim} \mathcal{N}\left(0, \frac{\sigma^2}{1 - \phi^2}\right) \quad (6.4.26)$$

여기서  $\{\epsilon_t | t = 1, 2, \dots, n\}$ 는 서로 독립이고 평균이 0이고 분산  $\sigma^2$ 를 따르는 정규확률변수열이다. 식  $|\phi| < 1$ 이 성립한다고 가정하자. 즉, 이 자기회귀과정은 안정성조건 (stability condition)을 만족한다. 관찰벡터를  $\mathbf{y} = [y_1, y_2, \dots, y_n]^t$ 로 표기하자.

**예제 6.4.4** 고전적 시계열분석을 적용해서 AR모형을 추정하는 예를 살펴보기 위해서, 다음 MATLAB 프로그램 ARmodelEstimation101.m을 실행해 보자.

```

1 % -----
2 % Filename: ARmodelEstimation101.m
3 % AR model Estimation
4 % https://kr.mathworks.com/help/ident/ug/estimating-ar-and-arma-models.html
5 % -----
6 clear, close all, clf
7 % Generatin time series of model:  $y_t = 0.9 y_{t-1} + b(L) e_t$ 
8 a = [1 -0.7]; % AR coeffs
9 b = [1 0.3 0.2]; % MA coeffs
10 Nobs = 256;
11 rng((1001^2+1)/2, 'twister')
12 e = randn(Nobs,1); % generate gaussian white noise
13 y = filter(b,a,e); % generate y
14
15 % Time Sereis Plot
16 plot(y, 'k-', 'LineWidth', 1.5)
17 set(gca, 'fontsize', 11, 'fontweigh', 'bold')
18 xlim([1 256])
19 xlabel('time'); ylabel('y');
20 saveas(gcf, 'ARmodelEstimation101', 'eps')
21
22 % Estimation
23 mdl1 = arima(1,0,0) % 1 AR(1) model
24 EstMdl1 = estimate(mdl1,y)
25 mdl2 = arima(1,0,2) % 1 ARMA(1,2) model
26 EstMdl2 = estimate(mdl2,y)
27 % End of program
28 % -----

```

이 MATLAB 프로그램 ARmodelEstimation101.m을 실행하면, 다음과 같은 ARMA모형에서 관찰값들을 256개 생성한다.

$$y_t - 0.7y_{t-1} = \epsilon_t + 0.3\epsilon_{t-1} + 0.2\epsilon_{t-2} \quad (1)$$

여기서  $\{\epsilon_t\}$ 는 서로 독립이고 동일한 확률분포  $\mathcal{N}(0, 1)$ 에서 생성된 난수열이다. 식 (6.4.25)와 식 (1)에서 유의할 점은  $\phi = 0.7$ 이라는 것이다. 즉, 계수의 부호가 반대이다. 이렇게 발생된  $\{y_t\}$ 의 시계열산점도가 그림 6.4.6에 그려져 있다. 그림 6.4.6의 시계열산점도는 정상성을 보이는 것 같다. 따라서, 이 시계열데이터에 ARMA모형을 적합시킬 수 있다.

적합시킨 첫 번째 모형은  $ARMA(1, 0)$  모형, 즉  $AR(1)$  모형이다. 이 모형을 추정된 결과는 다음과 같다.

$$y_t = 0.0949 + 0.8059y_{t-1} + \hat{\epsilon}_t \quad (2)$$

또한  $\hat{\epsilon}_t$ 의 표본분산은 0.9740이다. 적합시킨 두 번째 모형은  $ARMA(1, 2)$  모형이다. 이 모형을

추정한 결과는 다음과 같다.

$$y_t = 0.1767 + 0.6523y_{t-1} + \hat{\epsilon}_t + 0.2905\hat{\epsilon}_{t-1} + 0.1742\hat{\epsilon}_{t-2} \quad (3)$$

또한  $\hat{\epsilon}_t$ 의 표본분산은 0.9381이다. 이 경우 두 번째 모형의 표본분산이 첫 번째 모형의 표본분산보다 작다. 그렇다고 해서, 두 번째 모형이 주어진 시계열데이터에 더 적합하다는 것은 아니다. 올바른 ARMA모형을 선택하는 것은 그리 쉽지 않은 문제이다. 이에 대한 자세한 내용은 Choi [25]를 참조하라. ■

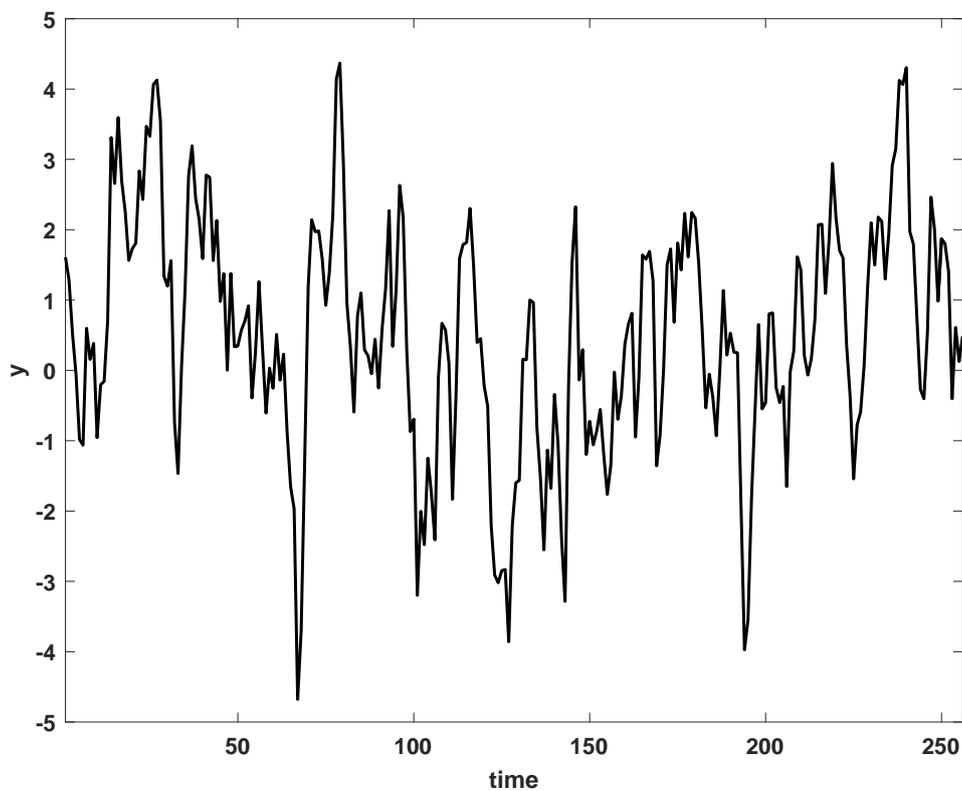


그림 6.4.6. ARMA(1,2) 모형에서 생성된 시계열데이터

식 (6.4.25)와 식 (6.4.26)로 구성된 AR(1)모형을 베이지안통계분석을 하기 위해서, 모수들  $(\phi, \sigma^2)$ 에 대한 사전확률분포를 다음과 같이 정의하자.

$$\phi \stackrel{d}{\sim} \text{Uniform}(-1, 1), \quad \sigma^2 \stackrel{d}{\sim} \text{IG}\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \quad (6.4.27)$$

따라서,  $\sigma^2$ 의 사후조건부확률분포는 다음과 같다.

$$\sigma^2 | (\phi, \mathbf{y}) \stackrel{d}{\sim} \mathcal{IG} \left( \frac{n_1}{2}, \frac{n_1 S_1}{2} \right) \quad (6.4.28)$$

여기서  $n_1$ 과  $S_1$ 은 각각 다음과 같다.

$$n_1 \doteq n_0 + n + 1, \quad S_1 \doteq \frac{1}{n_1} \left\{ n_0 S_0 + [1 - \phi^2] y_0^2 + \sum_{t=1}^n [y_t - \phi y_{t-1}]^2 \right\} \quad (6.4.29)$$

또한, 자기회귀계수  $\phi$ 의 사후조건부확률밀도함수는 다음과 같다.

$$\pi(\phi | \sigma^2, \mathbf{y}) \propto [1 - \phi^2]^{1/2} \exp \left( -\frac{[\phi - \hat{\phi}]^2}{2\hat{\sigma}_\phi^2} \right) \quad (6.4.30)$$

여기서  $\hat{\phi}$ 과  $\hat{\sigma}_\phi^2$ 는 각각 다음과 같다.

$$\hat{\phi} \doteq \frac{\sum_{t=1}^n y_t y_{t-1}}{\sum_{t=1}^{n-1} y_t^2}, \quad \hat{\sigma}_\phi^2 \doteq \frac{\sigma^2}{\sum_{t=1}^{n-1} y_t^2} \quad (6.4.31)$$

식 (6.4.30)의 사후조건부확률밀도함수는 우리에게 익숙한 것이 아니므로, 독립샘플러를 적용해서 샘플링하기로 하자. 즉,  $\phi^{(j-1)}$ 가 주어졌다고 가정하고,  $\phi^{(j)}$ 를 발생시키기로 하자. 자기회귀계수  $\phi$ 의 후보의 제안분포로 다음과 같은  $(-1, 1)$ 상에서 절단정규확률분포를 사용하자.

$$\phi | (\sigma^2, \mathbf{y}) \stackrel{d}{\sim} \mathcal{TN}_{(-1,1)}(\hat{\phi}, \hat{\sigma}_\phi^2) \quad (6.4.32)$$

다음 식들이 성립한다.

$$\frac{w(\phi | \sigma^2, \mathbf{y})}{w(\phi^{(j-1)} | \sigma^2, \mathbf{y})} \quad (6.4.33)$$

$$= \frac{\pi(\phi | \sigma^2, \mathbf{y})/q(\phi | \sigma^2, \mathbf{y})}{\pi(\phi^{(j-1)} | \sigma^2, \mathbf{y})/q(\phi^{(j-1)} | \sigma^2, \mathbf{y})} \quad (6.4.34)$$

$$= \frac{[1 - \phi^2]^{1/2} \exp\left(-\frac{[\phi - \hat{\phi}]^2}{2\hat{\sigma}_{\phi^2}^2}\right)}{\left\{1 - [\phi^{(j-1)}]^2\right\}^{1/2} \exp\left(-\frac{[\phi^{(j-1)} - \hat{\phi}]^2}{2\hat{\sigma}_{\phi^2}^2}\right)} \frac{\exp\left(-\frac{[\phi^{(j-1)} - \hat{\phi}]^2}{2\hat{\sigma}_{\phi^2}^2}\right)}{\exp\left(-\frac{[\phi - \hat{\phi}]^2}{2\hat{\sigma}_{\phi^2}^2}\right)} \quad (6.4.35)$$

$$= \frac{[1 - \phi^2]^{1/2}}{\left\{1 - [\phi^{(j-1)}]^2\right\}^{1/2}} \quad (6.4.36)$$

따라서, 절단정규확률분포  $\mathcal{TN}_{(-1,1)}(\hat{\phi}, \hat{\sigma}_{\phi^2}^2)$ 에서 발생된 자기회귀계수  $\phi$ 의 후보를 다음과 같은 확률로 채택한다.

$$\alpha\left(\phi^{(j-1)}, \phi \mid \sigma^2, \mathbf{y}\right) = \min \left\{ \frac{[1 - \phi^2]^{1/2}}{\left\{1 - [\phi^{(j-1)}]^2\right\}^{1/2}}, 1 \right\} \quad (6.4.37)$$

지금까지 내용을 바탕으로,  $AR(1)$  모형의 모수들  $(\phi, \sigma^2)$ 의 사후확률분포에서 샘플링하는 독립샘플러를 다음과 같이 요약할 수 있다.

#### 알고리즘 6.4.3: AR모형의 독립샘플러

- (1단계) 초기모수들  $(\phi^{(0)}, \sigma^{2(0)})$ 를 선택한다.
- (2단계) 각  $j(= 1, 2, \dots)$ 에 대해서 다음과 같이 샘플  $(\phi^{(j)}, \sigma^{2(j)})$ 를 발생시킨다.
- (2-1) 식 (6.4.28)의 사후조건부확률분포를 사용해서  $\sigma^{2(j)} \mid (\phi^{(j-1)}, \mathbf{y})$ 을 발생시킨다.
- (2-2) 식 (6.4.32)을 사용해서 자기회귀계수의 후보  $\phi \mid (\sigma^2, \mathbf{y})$ 를 발생시킨다.
- (2-3) 식 (6.4.37)를 사용해서, 확률  $\alpha(\phi^{(j-1)}, \phi \mid \sigma^2, \mathbf{y})$ 를 계산한다.
- (3단계) 독립적으로 일양난수  $u$ 를 발생시킨다. 만약 식  $\alpha(\phi^{(j-1)}, \phi \mid \sigma^2, \mathbf{y}) \geq u$ 가 성립하면, 자기회귀계수의 후보  $\phi \mid (\sigma^2, \mathbf{y})$ 를  $\phi^{(j)}$ 로 채택한다. 만약 그렇지 않으면,

$\phi^{(j-1)}$ 을  $\phi^{(j)}$ 로 채택한다.

(4단계) 충분히 큰  $b$ 에 대해서  $(\phi^{(j)}, \sigma^{2(j)})$ ,  $(j = b + 1, b + 2, \dots)$ 를 기록한다.

다음과 같은 ARX과정  $\{y_t | t = 0, 1, \dots, n\}$ 을 살펴보자.

$$y_t + a_1 y_{t-1} + \dots + a_p y_{t-p} = b_1 x_{t-1} + \dots + b_r x_{t-r} + \epsilon_t, \quad (t = 1, 2, \dots, n) \quad (6.4.38)$$

여기서  $\{\epsilon_t\}$ 는 서로 독립이고 평균이 0이고 분산  $\sigma^2$ 를 따르는 정규확률변수열이고,  $\{x_t\}$ 는 외생변수열이다. 이 ARX과정이 안정성조건을 만족한다고 가정하자.

**예제 6.4.5** 고전적 통계분석을 적용해서 ARX모형을 추정하는 예를 살펴보기 위해서, 다음 MATLAB프로그램 ARXmodelEstimation101.m을 실행해 보자.

```

1 % -----
2 % Filename: ARXmodelEstimation101.m
3 % ARX model Estimation
4 % Programmed by CBS
5 % -----
6 clear, close all
7 % Model
8 A = [1 -0.7]; % AR coeffs
9 B = [0 0.3 1.2]; % MA coeffs
10 mdl1 = idpoly(A,B)
11 Nobs = 256;
12 rng((1001^2+1)/2, 'twister') % Seed
13 x = 1.234*idinput(Nobs, 'rbs'); % rbs = random binary signal {-1,1}
14 e = randn(Nobs,1); % Gaussian random numbers
15 X = iddata([], x)
16 E = iddata([], e)
17 Y = sim(mdl1, [X E])
18 Z = [Y, X]
19 yNew = Z.Y;
20
21 % Time Series Plot
22 plot(1:Nobs, yNew, 'k-', 'LineWidth', 1.5)
23 set(gca, 'fontSize', 11, 'fontweight', 'bold')
24 xlim([1 Nobs])
25 xlabel('time'); ylabel('y', 'rotation', 0);
26 saveas(gcf, 'ARXmodelEstimation101', 'eps')
27
28 % ARX Estimation
29 mdl121 = arx(Z, [1 2 1])
30 mdl221 = arx(Z, [2 2 1])
31 % End of program
32 % -----

```

이 MATLAB프로그램 ARmodelEstimation101.m을 실행하면, 명령문 ‘mdl1 = id-

poly(A,B)'에서 지정했듯이 다음과 같은 ARX모형에서 관찰값들을 256개 생성한다.

$$y_t - 0.7y_{t-1} = 0 \cdot x_t + 0.3x_{t-1} + 1.2x_{t-2} + \epsilon_t \quad (1)$$

여기서  $\{x_t\}$ 는 지지대가  $\{-1, 1\}$ 인 일양난수들이고,  $\{\epsilon_t\}$ 는 서로 독립이고 동일한 확률분포  $\mathcal{N}(0, 1)$ 에서 생성된 난수열이다. 이렇게 발생된  $\{y_t\}$ 의 시계열산점도가 그림 6.4.7에 그려져 있다. 이 시계열데이터에 ARX모형을 적합시키자.

적합시킨 첫 번째 모형은  $ARX(1, 2)$  모형이다. 이 모형을 추정한 결과는 다음과 같다.

$$y_t - 0.7017y_{t-1} = 0.2782x_{t-1} + 1.188x_{t-2} + \hat{\epsilon}_t \quad (3)$$

추정된 모형의  $R^2$  값은 0.6243이고 FPE값은 1.105이다. 두 번째 모형은  $ARX(2, 2)$  모형이다. 이 모형을 추정한 결과는 다음과 같다.

$$y_t - 0.7053y_{t-1} + 0.0046y_{t-2} = 0.2784x_{t-1} + 1.186x_{t-2} + \hat{\epsilon}_t \quad (3)$$

추정된 모형의  $R^2$  값은 0.6243이고 FPE값은 1.114이다. 첫 번째 모형의 FPE값이 더 작으므로 첫 번째 모형인  $ARX(1, 2)$  모형이 이 데이터세트에 더 적합하다고 할 수는 있으나 FPE가 절대적인 기준은 아니라는 것을 기억하라. ■

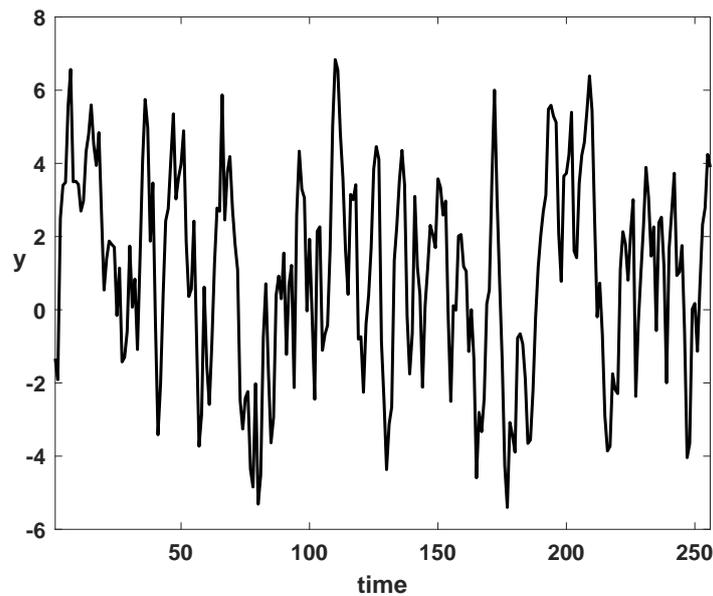


그림 6.4.7. ARX모형에서 생성된 시계열데이터

**예제 6.4.6** Enrique M. Quilis는 ARX모형을 추정하는 MATLAB프로그램을 제시하였다. 이 MATLAB프로그램을 다음 웹사이트에서 다운받을 수 있다.

<https://kr.mathworks.com/matlabcentral/profile/authors/1008800-enrique-m-quilis>

Quilis가 제시한 고전적 통계기법으로 ARX모형을 추정하는 예를 살펴보기 위해서, 다음 MATLAB프로그램 BayARX\_arx\_d.m을 실행해 보자.

```

1 % -----
2 % USAGE: BayARX_arx_d
3 % PURPOSE: demo of ARX
4 % Unrestricted ARX model
5 % https://kr.mathworks.com/matlabcentral/fileexchange
6 % /26955-bayesian-autoregressive-modeling?s_tid=prof_contriblnk
7 % -----
8 function BayARX_arx_d
9 close all; clc; clear all;
10
11 % Yield on Letras 12m: 1987.06 - 2007.03
12 % z = load('BayARX_y1.prn');
13 z = [ 15.2430  15.3080  15.0660  14.9600  14.1940  12.9810  12.3210
14       11.7440  11.6710  10.8050  9.6160  9.7910  10.0160  10.0030
15       10.0200  10.4320  11.5370  11.7520  12.6980  13.0360  13.4090
16       13.8610  14.1720  14.5120  14.4620  14.3340  14.0170  13.5670
17       13.0250  13.0750  13.0820  13.1030  13.4130  14.0440  14.3620
18       14.5170  14.5060  14.5120  14.5100  14.5150  14.4880  14.4970
19       14.5240  14.5170  14.3800  13.1750  12.5270  11.9920  11.6350
20       11.8620  12.3210  11.9960  11.8420  11.8450  11.9860  11.8740
21       11.7750  11.6880  11.8850  11.9110  11.9370  12.2660  12.8180
22       13.1880  13.6460  13.5280  13.6940  13.2820  12.4770  12.5360
23       12.5160  11.3010  10.6190  10.0440  9.6370  9.1850  8.9340
24       8.5650  8.0160  7.7300  7.7240  7.7280  7.7300  7.6960
25       7.7190  8.1290  8.0970  8.5860  8.7470  8.7510  8.8880
26       10.1500  9.7670  10.2310  10.2430  10.2290  10.2850  10.1680
27       10.0250  9.4640  9.5050  9.5040  9.0230  8.4830  8.3890
28       8.3130  7.7000  7.1910  7.1370  7.0640  7.1290  7.0230
29       6.6050  6.1530  5.8300  5.2810  5.2060  5.5650  5.3770
30       5.0170  4.9950  4.9730  4.9730  4.9310  4.7570  4.6980
31       4.4610  4.2450  4.1070  3.9670  4.0260  4.0250  4.0490
32       4.0060  3.9070  3.6590  3.3980  3.3190  2.9520  2.8450
33       2.8760  2.9290  2.5860  2.5590  2.7170  2.9350  3.0530
34       3.1510  3.3460  3.5170  3.6970  3.8470  3.9340  4.1140
35       4.3190  4.6970  4.7540  4.9710  5.1020  5.0670  5.0800
36       5.0230  4.7170  4.3850  4.3330  4.2550  4.3040  4.4240
37       4.1650  4.2240  3.9700  3.5350  3.2190  3.0070  3.1070
38       3.1620  3.5010  3.6960  3.7090  3.7750  3.7830  3.5690
39       3.2830  3.2120  2.8910  2.9710  2.8640  2.5550  2.2200
40       2.3440  2.3350  2.0530  1.8480  1.9860  2.1320  2.1450
41       2.1220  2.2910  2.3710  2.1310  1.9630  1.9430  2.0770
42       2.1760  2.2770  2.2120  2.1470  2.2470  2.1460  2.1980
43       2.1550  2.1630  2.1670  2.1710  2.1030  2.0600  1.9920
44       2.0620  2.0740  2.0880  2.2810  2.5230  2.6670  2.6180
45       2.7220  2.9190  3.0360  3.1210  3.2720  3.3600  3.4750
46       3.5350  3.5810  3.6600  3.7720  3.8360  3.8360  3.9480 ];
47 zsize = size(z); zlength = zsize(1)*zsize(2);
48 z = reshape(z,zlength,1);
49 z = [ z; 4.0770 ];
50

```

```

51 % Exogenous variables: unit step 1999.01 - 2007.12 (EMU)
52 % x = load('step9901.prn');
53 % rex.X = x(1:length(z));
54 rex.X = [];
55
56 % Intercept: yes (1) or not (0)
57 rex.opC = 1;
58
59 % Basic variable
60 rex.z = z;
61
62 % Order of AR model
63 rex.p = 6;
64
65 % Calling estimation function
66 res = BayARX_arx(rex);
67
68 % Printing output
69 file_sal = 'BayARX_arx.out';
70 arx_print(res,file_sal);
71 %arx_print(res);
72 edit(file_sal);
73 % End of BayARX_arx_d
74 % % % -----
75 function []=arx_print(res,file_sal);
76 % -----
77 % PURPOSE: Prints the output of the estimation of an ARX-BARX model
78 % SYNTAX: [] = arx_print(res,file_sal);
79 % -----
80 % OUTPUT: an ASCII file (file_sal) containing information about:
81 %         res.meth      = label
82 %         res.parm      = number of: observations, effective number
83 %                       of observations, series, order of AR model
84 %         res.mu        = 1x1 intercept
85 %         res.mu_dt     = 1x1 intercept: standard error
86 %         res.mu_t      = 1x1 intercept: t-ratios
87 %         res.b         = mx1 intercept
88 %         res.b_dt     = mx1 intercept: standard error
89 %         res.b_t       = mx1 intercept: t-ratios
90 %         res.phi       = p vector of phi(h) parameters
91 %         res.phi_dt    = phi(h) parameters: standard errors
92 %         res.phi_t     = phi(h) parameters: t-ratios
93 %         res.sigma     = variance of residuals
94 %         res.perform   = log-likelihood,statistics AIC and BIC
95 % -----
96 % INPUT: res a structure generated by arx/barx function
97 %         file_sal: name of the ASCII file for output (optional)
98 % -----
99 % NOTE: Prior specification: Litterman, Raynauld-Simonato.
100 % -----
101 % written by:
102 % Enrique M. Quilis
103 % Macroeconomic Research Department
104 % Ministry of Economy and Finance
105 % Paseo de la Castellana, 162. Office 2.5-1.
106 % 28046 - Madrid (SPAIN)
107 % <enrique.quilis@meh.es>
108 % September 2007
109 % Version 1.1
110 % -----
111 % Optional redirection to screen

```

```

112 if (nargin == 1)
113     fid = 1;
114 else
115     fid = fopen(file_sal, 'w');
116 end
117
118 % -----
119 % Auxiliary calculations
120 k=res.parm(3);
121 p=res.parm(4);
122 mu_L = [res.mu res.mu_dt res.mu_t]';
123 b_L = [res.b res.b_dt res.b_t]';
124 h=1:p; h=h';
125 phi_L = [h res.phi res.phi_dt res.phi_t]';
126
127 % -----
128 % Printing output in ASCII file
129 sep1='-----';
130 sep2='*****';
131 fprintf(fid, '\n ');
132 fprintf(fid, '%s \n ', sep2 );
133
134 switch res.meth
135     case {'ARX MODEL: UNRESTRICTED ESTIMATION'}
136         fprintf(fid, '%s\n ', res.meth );
137     case {'ARX MODEL: BAYESIAN ESTIMATION'}
138         fprintf(fid, '%s\n ', res.meth );
139         switch res.prior.meth
140             case {'LITTERMAN PRIOR'}
141                 fprintf(fid, '%s \n ', sep2 );
142                 fprintf(fid, '%s\n ', res.prior.meth );
143                 fprintf(fid, ' Hyperparameters: tightness \n ');
144                 fprintf(fid, ' Intercept %12.4f \n', res.prior.rex.th_mu);
145                 fprintf(fid, ' Exogenous %12.4f \n', res.prior.rex.th_b);
146                 switch res.prior.rex.type
147                     case 1
148                         fprintf(fid, ' Decay function: harmonic \n');
149                     case 2
150                         fprintf(fid, ' Decay function: geometric \n');
151                 end
152                 fprintf(fid, ' Global %12.4f \n', res.prior.rex.th(1));
153                 fprintf(fid, ' Decay %12.4f \n', res.prior.rex.th(2));
154             case {'RAYNAULD-SIMONATO PRIOR'}
155                 fprintf(fid, '%s \n ', sep2 );
156                 fprintf(fid, '%s\n ', res.prior.meth );
157                 fprintf(fid, ' Hyperparameters \n ');
158                 switch res.prior.rex.type
159                     case 1
160                         fprintf(fid, ' Mean: centered around (1-B)(1-B^s) \n');
161                     case 2
162                         fprintf(fid, ' Mean: centered around (1-B^s) \n');
163                 end
164                 fprintf(fid, 'Hyperparameters: tightness \n ');
165                 fprintf(fid, ' Intercept %12.4f \n', res.prior.rex.th_mu);
166                 fprintf(fid, ' Exogenous %12.4f \n', res.prior.rex.th_b);
167                 fprintf(fid, ' Global %12.4f \n', res.prior.rex.th(1));
168                 fprintf(fid, ' Decay: regular part %12.4f \n', ...
169                     res.prior.rex.th(2));
170                 fprintf(fid, ' Decay: seasonal part %12.4f \n', ...
171                     res.prior.rex.th(3));
172

```

```

173         case {'CANOVA PRIOR'}
174             fprintf(fid,'%s\n ', res.prior.meth );
175             fprintf(fid,' Under construction!! \n ');
176             % TO BE INCLUDED: INFO ABOUT HYPERPARAMETERS
177         end
178     end
179     fprintf(fid,'%s \n ', sep2 );
180     fprintf(fid,'Number of observations           : %4d\n ',res.parm(1));
181     fprintf(fid,'Number of effective observations : %4d\n ',res.parm(2));
182     fprintf(fid,'Number of series                : %4d\n ',res.parm(3) );
183     fprintf(fid,'Number of lags                  : %4d\n ',res.parm(4) );
184     fprintf(fid,'\n ');
185     fprintf(fid,'%s \n ', sep2 );
186     fprintf(fid,'mu: Estimate, s.d. and t-ratio (read columnwise) \n ');
187     fprintf(fid,'%s \n ', sep1 );
188     fprintf(fid,'%12.4f      %12.4f      %8.4f \n ',mu_L);
189     fprintf(fid,'%s \n ', sep1 );
190     fprintf(fid,'b: Estimate, s.d. and t-ratio (read columnwise) \n ');
191     fprintf(fid,'%s \n ', sep1 );
192     fprintf(fid,'%12.4f      %12.4f      %8.4f \n ',b_L);
193     fprintf(fid,'%s \n ', sep1 );
194     fprintf(fid,'phi: Lag, estimate, s.d. and t-ratio (read columnwise) \n ');
195     fprintf(fid,'%s \n ', sep1 );
196     fprintf(fid,'%4d %12.4f      %12.4f      %8.4f \n ',phi_L);
197     fprintf(fid,'%s \n ', sep1 );
198     fprintf(fid,'RESIDUALS: s.d. ');
199     fprintf(fid,'\n');
200     fprintf(fid,' %12.4f ', sqrt(res.sigma));
201     fprintf(fid,'\n');
202     fprintf(fid,'%s \n ', sep1 );
203     fprintf(fid,'Model performance (read columnwise): \n');
204     fprintf(fid,' - log-likelihood\n');
205     fprintf(fid,' - AIC and BIC \n');
206     fprintf(fid,' - BIC \n');
207     fprintf(fid,'%s \n ', sep1 );
208     fprintf(fid,'%12.4f      %8.4f      %8.4f \n ',res.perform);
209     fprintf(fid,'%s \n ', sep2 );
210     fprintf(fid,'\n');
211
212     if (nargin == 1)
213         % Do nothing
214     else
215         fclose(fid);
216     end
217 % End of function arx_print
218 %-----

```

이 MATLAB 프로그램 BayARX\_arx\_d.m을 실행하기 위해서는 같은 디렉토리에 다음 MATLAB 프로그램 BayARX\_arx.m가 있어야 한다.

```

1 % -----
2 % USAGE: BayARX_arx
3 % https://kr.mathworks.com/matlabcentral/fileexchange
4 %       /26955-bayesian-autoregressive-modeling?s_tid=prof_contriblnk
5 % % -----
6
7 function res = arx(rex);
8 % PURPOSE: Unrestricted estimation of an univariate ARX model
9 % -----

```

```

10 % SYNTAX: res = arx(rex)
11 % -----
12 % OUTPUT: a structure ...
13 %     res.meth      = 'ARX MODEL: UNRESTRICTED ESTIMATION';
14 %     res.parm     = [n,nobs,k,p];
15 %     res.mu       = [mu]; <===== Intercept
16 %     res.mu_dt    = [mu_dt];
17 %     res.mu_t     = [mu_t];
18 %     res.b        = [b]; <===== Exogenous variables parameters
19 %     res.b_dt     = [b_dt];
20 %     res.b_t      = [b_t];
21 %     res.phi      = [phi]; <===== Lagged variables parameters
22 %     res.phi_dt   = [phi_dt];
23 %     res.phi_t    = [phi_t];
24 %     res.u        = [u]; <===== Residuals
25 %     res.sigma    = [sigma];
26 %     res.perform  = [log_like,aic,bic];
27 %     res.G        = VCV matrix of estimated parameters
28 % -----
29 % INPUT: a structure ...
30 %     X      : nxm matrix of exogenous series, columnwise
31 %     z      : nx1 vector of original series, columnwise
32 %     opC    : 1x1 indicator to include (1) an intercept or not (0)
33 %     p      : 1x1 order of the model
34 % -----
35 % LIBRARY: lag, uisolator
36 % -----
37 % SEE ALSO: barx, arx_print
38 % -----
39 % written by:
40 % Enrique M. Quilis
41 % Macroeconomic Research Department
42 % Ministry of Economy and Finance
43 % Paseo de la Castellana, 162. Office 2.5-1.
44 % 28046 - Madrid (SPAIN)
45 % <enrique.quilis@meh.es>
46 % January 2009
47 % Version 1.2
48 % -----
49 % -----
50 % Loading the structure
51
52 % Exogenous variables
53 Xx = rex.X;
54
55 % Intercept: yes (1) or not (0)
56 opC = rex.opC;
57
58 % Basic variable
59 z = rex.z;
60
61 % Order of AR model
62 p = rex.p;
63
64 % -----
65 % Computing basic parameters
66 [n,k] = size(z);
67 if (k ~= 1)
68     error('*** ONLY UNIVARIATE MODELING, PLEASE ***');
69 end
70 [aux,m] = size(Xx);

```

```

71 nobs = n-p; % Number of effective observations
72
73 % -----
74 %           PREPARING THE MODEL IN LINEAR MODEL FORMAT: Y=XB+U
75 % -----
76 Xz = lag(z,p);
77 Xz = Xz(:,2:end); % Dropping vector of ones (no intercept is assumed)
78 z = z(p+1:end);
79 if (isempty(Xx) == 0) % Exogenous variables are included
80     Xx = Xx(p+1:end,:);
81 end
82
83 % Intercept
84 switch opC
85 case 0 % No intercept
86     X = [Xx Xz];
87 case 1 % Intercept
88     X = [ones(nobs,1) Xx Xz];
89 end
90
91 % -----
92 %           ESTIMATION OF THE MODEL: OLS
93 beta = (X'*X) \ (X'*z);
94
95 % Residuals
96 u = z - X * beta ;
97
98 % variance of residuals
99 sigma = var(u,1);
100
101 % VCV and t-ratios of beta parameters
102 G      = sigma * inv(X'*X);
103 beta_dt = sqrt(diag(G));
104 beta_t  = beta ./ beta_dt;
105
106 % Performance measures
107 log_like = -0.5*nobs*(1*(1+log(2*pi))+log(sigma));
108 npar = opC + m + p;
109 aic = -(2*log_like/nobs) + (2*npar)/nobs;
110 bic = -(2*log_like/nobs) + npar*(log(nobs)/nobs);
111
112 % -----
113 %           ISOLATING THE MATRICES
114 % -----
115 [mu,b,phi] = uisolator(beta,opC,m,p);
116 [mu_dt,b_dt,phi_dt] = uisolator(beta_dt,opC,m,p);
117 [mu_t,b_t,phi_t] = uisolator(beta_t,opC,m,p);
118
119 % -----
120 %           LOADING THE STRUCTURE
121 res.meth      = 'ARX MODEL: UNRESTRICTED ESTIMATION';
122 res.parm      = [n,nobs,k,p];
123 res.mu        = [mu];
124 res.mu_dt     = [mu_dt];
125 res.mu_t      = [mu_t];
126 res.b         = [b];
127 res.b_dt     = [b_dt];
128 res.b_t       = [b_t];
129 res.phi       = [phi];
130 res.phi_dt    = [phi_dt];
131 res.phi_t     = [phi_t];

```

```

132 res.u          = [u];
133 res.sigma      = [sigma];
134 res.perform    = [log_like,aic,bic];
135 res.G          = [G];
136 % End of function arx
137 % % -----
138
139 function X = lag(Z,p);
140 % PURPOSE: Create a matrix with the lags of Z up to p
141 % -----
142 % SYNTAX: X = lag(Z,p)
143 % -----
144 % OUTPUT: X = matrix of Z(1)..Z(p), including a vector of 1s,
145 %         in the common range of obs.: p+1 .. n
146 %         X : n-p x k*p+1
147 % -----
148 % INPUT: Z = an nxk matrix of original series, columnwise
149 %         p = the maximum lag, p>=1
150
151 % -----
152 % written by:
153 % Enrique M. Quilis
154 % Macroeconomic Research Department
155 % Ministry of Economy and Finance
156 % Paseo de la Castellana, 162. Office 2.5-1.
157 % 28046 - Madrid (SPAIN)
158 % <enrique.quilis@meh.es>
159
160 % Cheks
161 if (p < 1)
162     error(' *** The number of lags should be greater than one *** ');
163 end
164 [n,k] = size(Z);
165 if (n < p)
166     error(' *** The # of obs. should be greater than the # of lags *** ');
167 end
168
169 % Initial matrix
170 XX = [Z
171       zeros(p,k)];
172
173 % Recurrence
174 j = 1;
175 while (j <= p)
176     Aux = [zeros(j,k)
177           Z
178           zeros(p-j,k)];
179     XX = [ XX Aux ];
180     j = j + 1;
181 end;
182
183 % Final matrix
184 Xf = [zeros(p,k)
185       Z];
186
187 % Adding the last lag
188 XX = [XX Xf];
189
190 % Selecting the common range
191 X = XX(p+1:n,k+1:k*(p+1));
192

```

```

193 % Preparing X to include a constant
194 [nX,mX] = size(X);
195 X = [ ones(nX,1) X];
196 % End of function lag
197 % % -----
198
199 function [mu,b,phi] = uisolator(beta,opC,m,p);
200 % PURPOSE: Creates a sequence of vectors of dimension 1 (mu)
201 %           m (beta for exog.) and p (lagged endogenous)
202 % -----
203 % SYNTAX: [mu,b,phi] = uisolator(beta,opC,m,p);
204 % -----
205 % OUTPUT: Intercept mu : 1x1, b : mx1, phi : px1
206 %         Note thar mu and/or b may be empty
207 % -----
208 % INPUT: beta: vector to be segmented
209 %        opC : 1x1 indicator to include (1) an intercept or not (0)
210 %        m   : number of exogenous variables
211 %        p   : number of lags
212 % -----
213 % written by:
214 % Enrique M. Quilis
215 % Macroeconomic Research Department
216 % Ministry of Economy and Finance
217 % Paseo de la Castellana, 162. Office 2.5-1.
218 % 28046 - Madrid (SPAIN)
219 % <enrique.quilis@meh.es>
220 %
221 % September 2007
222 % Version 1.1
223 % -----
224 % Isolating intercept
225 switch opC
226 case 0 % No intercept
227     mu = [];
228 case 1
229     mu = beta(1,1);
230 end
231
232 % -----
233 % Isolating beta for exogenous variables
234 switch m
235 case 0 % No exogenous variables
236     b = [];
237 otherwise % Exogenous variables
238     b = beta(2:m+1,1);
239 end
240
241 % -----
242 % Isolating beta for lagged endogenous variable
243 phi = beta(m+opC+1:end,1);
244 % End of uisolator
245 %-----

```

이 MATLAB 프로그램 BayARX\_arx\_d.m을 실행하면, 다음과 같은 결과가 출력된다.

```

*****
ARX MODEL: UNRESTRICTED ESTIMATION
*****

```

```

Number of observations      : 239
Number of effective observations : 233
Number of series          : 1
Number of lags            : 6
    
```

\*\*\*\*\*

mu: Estimate, s.d. and t-ratio (read columnwise)

```

-----
          1.4434          0.2877          5.0175
-----
    
```

b: Estimate, s.d. and t-ratio (read columnwise)

```

-----
-----
    
```

phi: Lag, estimate, s.d. and t-ratio (read columnwise)

```

-----
 1      0.7160          0.0633          11.3063
 2      0.1334          0.0787           1.6942
 3      0.1860          0.0790           2.3541
 4     -0.0405          0.0783          -0.5175
 5      0.0334          0.0779           0.4289
 6     -0.2348          0.0623          -3.7672
-----
    
```

RESIDUALS: s.d.

```

          2.0118
-----
    
```

Model performance (read columnwise):

```

- log-likelihood
- AIC and BIC
- BIC
-----
-493.4905          4.2961          4.3997
    
```

\*\*\*\*\*

이 결과물에서 알 수 있듯이, 이 예제에서  $AR(6)$  모형을 추정한 결과는 다음과 같다.

$$\begin{aligned}
 \hat{y}_t = & 1.4434 + 0.7160\hat{y}_{t-1} + 0.1334\hat{y}_{t-2} + 0.1860\hat{y}_{t-3} \\
 & - 0.0405\hat{y}_{t-4} + 0.0334\hat{y}_{t-5} - 0.2348\hat{y}_{t-6}
 \end{aligned}
 \tag{1}$$

또한 잔차들의 표준편차는 다음과 같다.

$$\hat{\sigma} \doteq \sqrt{\frac{1}{233} \sum_{t=7}^{239} [\hat{y}_t - y_t]^2} = 2.0118 \quad (2)$$

■

**예제 6.4.7** Enrique M. Quilis는 ARX모형을 베이지안기법으로 추정하는 MATLAB 프로그램을 제시하였다. 이 MATLAB 프로그램도 다음 웹사이트에서 다운받을 수 있다.

<https://kr.mathworks.com/matlabcentral/profile/authors/1008800-enrique-m-quilis>

Quilis가 제시한 추정법을 적용하는 예로서, 다음 MATLAB 프로그램 BayARX\_barx\_d.m 을 실행해 보자.

```

1 % -----
2 % USAGE: BayARX_barx_L_d
3 % PURPOSE: demo of BARX
4 %         Litterman BARX model
5 % https://kr.mathworks.com/matlabcentral/fileexchange
6 %         /26955-bayesian-autoregressive-modeling?s_tid=prof_contriblnk
7 % -----
8 function BayARX_barx_L_d
9 close all; clc; clear all;
10
11 % Yield on Letras 12m: 1987.06 - 2007.03
12 % z = load('BayARX_y1.prn');
13 z = [ 15.2430    15.3080    15.0660    14.9600    14.1940    12.9810    12.3210
14        11.7440    11.6710    10.8050    9.6160    9.7910    10.0160    10.0030
15        10.0200    10.4320    11.5370    11.7520    12.6980    13.0360    13.4090
16        13.8610    14.1720    14.5120    14.4620    14.3340    14.0170    13.5670
17        13.0250    13.0750    13.0820    13.1030    13.4130    14.0440    14.3620
18        14.5170    14.5060    14.5120    14.5100    14.5150    14.4880    14.4970
19        14.5240    14.5170    14.3800    13.1750    12.5270    11.9920    11.6350
20        11.8620    12.3210    11.9960    11.8420    11.8450    11.9860    11.8740
21        11.7750    11.6880    11.8850    11.9110    11.9370    12.2660    12.8180
22        13.1880    13.6460    13.5280    13.6940    13.2820    12.4770    12.5360
23        12.5160    11.3010    10.6190    10.0440    9.6370    9.1850    8.9340
24        8.5650    8.0160    7.7300    7.7240    7.7280    7.7300    7.6960
25        7.7190    8.1290    8.0970    8.5860    8.7470    8.7510    8.8880
26        10.1500    9.7670    10.2310    10.2430    10.2290    10.2850    10.1680
27        10.0250    9.4640    9.5050    9.5040    9.0230    8.4830    8.3890
28        8.3130    7.7000    7.1910    7.1370    7.0640    7.1290    7.0230
29        6.6050    6.1530    5.8300    5.2810    5.2060    5.5650    5.3770
30        5.0170    4.9950    4.9730    4.9730    4.9310    4.7570    4.6980
31        4.4610    4.2450    4.1070    3.9670    4.0260    4.0250    4.0490
32        4.0060    3.9070    3.6590    3.3980    3.3190    2.9520    2.8450
33        2.8760    2.9290    2.5860    2.5590    2.7170    2.9350    3.0530
34        3.1510    3.3460    3.5170    3.6970    3.8470    3.9340    4.1140
35        4.3190    4.6970    4.7540    4.9710    5.1020    5.0670    5.0800
36        5.0230    4.7170    4.3850    4.3330    4.2550    4.3040    4.4240
37        4.1650    4.2240    3.9700    3.5350    3.2190    3.0070    3.1070

```

```

38      3.1620    3.5010    3.6960    3.7090    3.7750    3.7830    3.5690
39      3.2830    3.2120    2.8910    2.9710    2.8640    2.5550    2.2200
40      2.3440    2.3350    2.0530    1.8480    1.9860    2.1320    2.1450
41      2.1220    2.2910    2.3710    2.1310    1.9630    1.9430    2.0770
42      2.1760    2.2770    2.2120    2.1470    2.2470    2.1460    2.1980
43      2.1550    2.1630    2.1670    2.1710    2.1030    2.0600    1.9920
44      2.0620    2.0740    2.0880    2.2810    2.5230    2.6670    2.6180
45      2.7220    2.9190    3.0360    3.1210    3.2720    3.3600    3.4750
46      3.5350    3.5810    3.6600    3.7720    3.8360    3.8360    3.9480 ];
47  zsize = size(z); zlength = zsize(1)*zsize(2);
48  z = reshape(z,zlength,1);
49  z = [ z; 4.0770 ];
50
51  % Exogenous variables: unit step 1999.01 - 2007.12 (EMU)
52  % x = load('step9901.prn');
53  % rex.X = x(1:length(z));
54  rex.X = [];
55
56  % Intercept: yes (1) or not (0)
57  rex.opC = 0;
58
59  % Basic variable
60  rex.z = z;
61
62  % Order of AR model
63  rex.p = 6;
64
65  %-----
66  % Generating Litterman prior
67  % Type of decay function: harmonic (1) or geometric (2)
68  rex.type = 1;
69  rex.th_mu = 0.0000000001;
70  rex.th_b = 1000;
71  rex.th = [0.10 0.05];
72  rex.plt = 1; %Plots or no plots
73
74  prior = uprior_L(rex);
75
76  %-----
77  % Calling estimation function
78  res = barx(rex,prior);
79
80  % Printing output
81  file_sal = 'BayARX_barx_L.out';
82  arx_print(res,file_sal);
83  edit(file_sal);
84  % End of function BayARX_barx_L_d
85  % %-----
86
87  function res = uprior_L(rex);
88  % PURPOSE: Generates a Litterman prior for ARX models
89  % -----
90  % SYNTAX: res = prior_L(rex);
91  % -----
92  % OUTPUT: res: a structure with ...
93  %     prior_beta      : prior on the mean, (opC+m+p) x 1
94  %     prior_beta_var  : prior on the variance, (opC+m+p)x(opC+m+p)
95  %     g                : temporal decay function, px1
96  %     vu               : variance for scaling, 1x1
97  %     rex              : inherited from input, see below
98  % -----

```

```

99 % INPUT: rex: a structure with ...
100 %     z      : nx1 time series to be modeled
101 %     p      : order of AR model
102 %     th_mu  : hyperparameter to control the variance of the intercept mu
103 %     th_b   : hyperparameter to control the variance of the exog.param. b
104 %     th     : a 1x2 vector of hyperparameters to control
105 %                the prior variance of phi
106 %             th(1) = global tightness
107 %             th(2) = tightness on lag decay
108 %     type  : decay function g: harmonic (=1) or geometric (=2)
109 %     plt   : plot g (1) or not (0)
110 % -----
111 % LIBRARY: arx
112 % -----
113 % SEE ALSO: uprior_RS
114 % -----
115 % REFERENCE: Litterman, R.B. (1984) "Specifying vector autoregressions
116 % for macroeconomic forecasting", Federal Reserve Bank of Minneapolis,
117 % Staff Report n. 92.
118 % -----
119 % written by:
120 % Enrique M. Quilis
121 % Macroeconomic Research Department
122 % Ministry of Economy and Finance
123 % Paseo de la Castellana, 162. Office 2.5-1.
124 % 28046 - Madrid (SPAIN)
125 % <enrique.quilis@meh.es>
126 % September 2007
127 % Version 1.1
128
129 % -----
130 % Loading the structure
131
132 % Intercept: yes (1) or not (0)
133 opC = rex.opC;
134
135 % Number of exogenous variables
136 aux = size(rex.X);
137 m = aux(1,2); clear aux;
138
139 % Order of AR model
140 p = rex.p;
141
142 % Hyperparameters
143 th_mu = rex.th_mu;
144 th_b = rex.th_b;
145 th = rex.th;
146
147 % -----
148 % Computing the variance for scaling by means of ARX(p) model
149 res1 = arx(rex);
150 vu = res1.sigma;
151
152 % -----
153 % PRIOR ON THE INTERCEPT: mu 1x1
154 % -----
155 switch opC
156 case 1 % Intercept is included
157     % Mean
158     mup = 0;
159     % Variance

```

```

160     vmup = th_mu * vu;
161 otherwise
162     mup = [];
163     vmup = [];
164 end
165
166 % -----
167 % PRIOR ON THE EXOGENOUS VARIABLES PARAMETERS: b mx1
168 % -----
169 switch m
170 case 0 % No exogenous variables are included
171     % Mean
172     bp = [];
173     % Variance
174     vbp = [];
175 otherwise %
176     % Mean
177     bp = zeros(m,1);
178     % Variance
179     vbp = th_b * ones(m,1);
180 end
181
182 % -----
183 % PRIOR ON THE LAGGED TERMS: phi px1
184 % -----
185 % Mean
186 phip = zeros(p,1);
187 phip(1) = 1;
188
189 % Variance
190 g = NaN*ones(p,1);
191 g(1)=1;
192 h=2:p;
193 switch rex.type
194 case 1 % Harmonic decay: th(2) >= 0
195     g(h) = h.^(-th(2));
196     case 2 % Geometric decay: 0 <= th(2) <= 1
197         g(h) = th(2).^h;
198 end % of switch
199 vphp = (th(1) * g) .^2;
200
201 % -----
202 % GLOBAL PRIOR ON THE MEAN
203 % -----
204 prior_beta = [mup; bp; phip ];
205
206 % -----
207 % GLOBAL PRIOR ON THE VARIANCE
208 % -----
209 aux = [vmup; vbp; vphp ];
210 prior_beta_var = diag(aux); clear aux;
211
212 % -----
213 % LOADING THE STRUCTURE
214 % -----
215 res.meth          = 'LITTERMAN PRIOR';
216 res.prior_beta    = prior_beta;
217 res.prior_beta_var = prior_beta_var;
218 res.g             = g;
219 res.vu            = vu; %Variance for scaling
220 res.rex           = rex;

```

```

221 |
222 | % -----
223 | % OPTIONAL PLOTS
224 | switch rex.plt
225 | case 0
226 |     %Just do nothing
227 | case 1
228 |     h=0:p+1; h=h';
229 |     plot(h,[NaN; g; NaN],'-bd');
230 |     axis([0 p+1 0 1.05]);
231 |     grid on; xlabel('Lag');
232 |     set(gca,'fontsize',11,'fontweigh','bold')
233 |     saveas(gcf,'BayARX_barx_L_d1','eps')
234 |     switch rex.type
235 |         case 1
236 |             title('Decay function: harmonic');
237 |         case 2
238 |             title('Decay function: geometric');
239 |     end % of switch
240 |     figure;
241 |     plot(h,[NaN; phip; NaN],'-ro',h,[NaN; phip+sqrt(vphip); NaN],':b', ...
242 |          h,[NaN; phip-sqrt(vphip); NaN],':b');
243 |     grid on; title('Prior on phi parameters'); xlabel('Lag');
244 |     set(gca,'fontsize',11,'fontweigh','bold')
245 |     saveas(gcf,'BayARX_barx_L_d2','eps')
246 | end
247 | % End of function uprior_L
248 | % -----
249 |
250 | function res = arx(rex);
251 | % PURPOSE: Unrestricted estimation of an univariate ARX model
252 | % -----
253 | % SYNTAX: res = arx(rex)
254 | % -----
255 | % OUTPUT: a structure ...
256 | %     res.meth      = 'ARX MODEL: UNRESTRICTED ESTIMATION';
257 | %     res.parm      = [n,nobs,k,p];
258 | %     res.mu        = [mu]; <===== Intercept
259 | %     res.mu_dt     = [mu_dt];
260 | %     res.mu_t      = [mu_t];
261 | %     res.b         = [b]; <===== Exogenous variables parameters
262 | %     res.b_dt      = [b_dt];
263 | %     res.b_t       = [b_t];
264 | %     res.phi       = [phi]; <===== Lagged variables parameters
265 | %     res.phi_dt    = [phi_dt];
266 | %     res.phi_t     = [phi_t];
267 | %     res.u         = [u]; <===== Residuals
268 | %     res.sigma     = [sigma];
269 | %     res.perform   = [log_like,aic,bic];
270 | %     res.G         = VCV matrix of estimated parameters
271 | % -----
272 | % INPUT: a structure ...
273 | %     X      : nxm matrix of exogenous series, columnwise
274 | %     z      : nx1 vector of original series, columnwise
275 | %     opC    : 1x1 indicator to include (1) an intercept or not (0)
276 | %     p      : 1x1 order of the model
277 | % -----
278 | % LIBRARY: lag, uisolator
279 | % -----
280 | % SEE ALSO: arx_print
281 | % -----

```

```

282 % written by:
283 % Enrique M. Quilis
284 % Macroeconomic Research Department
285 % Ministry of Economy and Finance
286 % Paseo de la Castellana, 162. Office 2.5-1.
287 % 28046 - Madrid (SPAIN)
288 % <enrique.quilis@meh.es>
289 % January 2009
290 % Version 1.2
291
292 % -----
293 % Loading the structure
294
295 % Exogenous variables
296 Xx = rex.X;
297
298 % Intercept: yes (1) or not (0)
299 opC = rex.opC;
300
301 % Basic variable
302 z = rex.z;
303
304 % Order of AR model
305 p = rex.p;
306
307 % -----
308 % Computing basic parameters
309 [n,k] = size(z);
310 if (k ~= 1)
311     error('*** ONLY UNIVARIATE MODELING, PLEASE ***');
312 end
313 [aux,m] = size(Xx);
314 nobs = n-p; % Number of effective observations
315
316 % -----
317 %     PREPARING THE MODEL IN LINEAR MODEL FORMAT: Y=XB+U
318 % -----
319 Xz = lag(z,p);
320 Xz = Xz(:,2:end); % Dropping vector of ones (no intercept is assumed)
321 z = z(p+1:end);
322 if (isempty(Xx) == 0) % Exogenous variables are included
323     Xx = Xx(p+1:end,:);
324 end
325
326 % Intercept
327 switch opC
328 case 0 % No intercept
329     X = [Xx Xz];
330 case 1 % Intercept
331     X = [ones(nobs,1) Xx Xz];
332 end
333
334 % -----
335 %     ESTIMATION OF THE MODEL: OLS
336 beta = (X'*X) \ (X'*z);
337
338 % Residuals
339 u = z - X * beta ;
340
341 % variance of residuals
342 sigma = var(u,1);

```

```

343
344 % VCV and t-ratios of beta parameters
345 G          = sigma * inv(X'*X);
346 beta_dt   = sqrt(diag(G));
347 beta_t    = beta ./ beta_dt;
348
349 % Performance measures
350 log_like = -0.5*nobs*(1*(1+log(2*pi))+log(sigma));
351 npar = opC + m + p;
352 aic = -(2*log_like/nobs) + (2*npar)/nobs;
353 bic = -(2*log_like/nobs) + npar*(log(nobs)/nobs);
354
355 % -----
356 %           ISOLATING THE MATRICES
357 % -----
358 [mu,b,phi] = uisolator(beta,opC,m,p);
359 [mu_dt,b_dt,phi_dt] = uisolator(beta_dt,opC,m,p);
360 [mu_t,b_t,phi_t] = uisolator(beta_t,opC,m,p);
361
362 % -----
363 %           LOADING THE STRUCTURE
364 res.meth    = 'ARX MODEL: UNRESTRICTED ESTIMATION';
365 res.parm    = [n,nobs,k,p];
366 res.mu      = [mu];
367 res.mu_dt   = [mu_dt];
368 res.mu_t    = [mu_t];
369 res.b       = [b];
370 res.b_dt    = [b_dt];
371 res.b_t     = [b_t];
372 res.phi     = [phi];
373 res.phi_dt  = [phi_dt];
374 res.phi_t   = [phi_t];
375 res.u       = [u];
376 res.sigma   = [sigma];
377 res.perform = [log_like,aic,bic];
378 res.G       = [G];
379 % End of function arx
380 % %-----
381
382 function X = lag(Z,p);
383 % PURPOSE: Create a matrix with the lags of Z up to p
384 % -----
385 % SYNTAX: X = lag(Z,p)
386 % -----
387 % OUTPUT: X = matrix of Z(1)..Z(p), including a vector of 1s,
388 %          in the common range of obs.: p+1 .. n
389 %          X : n-p x k*p+1
390 % -----
391 % INPUT: Z = an nxk matrix of original series, columnwise
392 %          p = the maximum lag, p>=1
393 % -----
394 % written by:
395 % Enrique M. Quilis
396 % Macroeconomic Research Department
397 % Ministry of Economy and Finance
398 % Paseo de la Castellana, 162. Office 2.5-1.
399 % 28046 - Madrid (SPAIN)
400 % <enrique.quilis@meh.es>
401
402 % Cheks
403 if (p < 1)

```

```

404     error ( ' *** The number of lags should be greater than one *** ' );
405 end
406 [n,k] = size(Z);
407 if (n < p)
408     error ( ' *** The # of obs. should be greater than the # of lags *** ' );
409 end
410
411 % Initial matrix
412 XX = [Z; zeros(p,k)];
413
414 % Recurrence
415 j = 1;
416 while (j <= p)
417     Aux = [zeros(j,k)
418           Z
419           zeros(p-j,k)];
420     XX = [ XX Aux ];
421     j = j + 1;
422 end;
423
424 % Final matrix
425 Xf = [zeros(p,k); Z];
426
427 % Adding the last lag
428 XX = [XX Xf];
429
430 % Selecting the common range
431 X = XX(p+1:n,k+1:k*(p+1));
432
433 % Preparing X to include a constant
434 [nX,mX] = size(X);
435 X = [ ones(nX,1) X];
436 % End of function lag
437 % %-----
438
439 function [mu,b,phi] = uisolator(beta,opC,m,p);
440 % PURPOSE: Creates a sequence of vectors of dimension 1 (mu)
441 %           m (beta for exog.) and p (lagged endogenous)
442 % -----
443 % SYNTAX: [mu,b,phi] = uisolator(beta,opC,m,p);
444 % -----
445 % OUTPUT: Intercept mu : 1x1, b : mx1, phi : px1
446 %         Note thar mu and/or b may be empty
447 % -----
448 % INPUT: beta: vector to be segmented
449 %         opC : 1x1 indicator to include (1) an intercept or not (0)
450 %         m   : number of exogenous variables
451 %         p   : number of lags
452 % -----
453 % written by:
454 % Enrique M. Quilis
455 % Macroeconomic Research Department
456 % Ministry of Economy and Finance
457 % Paseo de la Castellana, 162. Office 2.5-1.
458 % 28046 - Madrid (SPAIN)
459 % <enrique.quilis@meh.es>
460 % September 2007
461 % Version 1.1
462
463 % -----
464 % Isolating intercept

```

```

465 switch opC
466 case 0 % No intercept
467     mu = [];
468 case 1
469     mu = beta(1,1);
470 end
471
472 % -----
473 % Isolating beta for exogenous variables
474 switch m
475 case 0 % No exogenous variables
476     b = [];
477 otherwise % Exogenous variables
478     b = beta(2:m+1,1);
479 end
480
481 % -----
482 % Isolating beta for lagged endogenous variable
483 phi = beta(m+opC+1:end,1);
484 % End of function uisolator
485 % %-----
486
487 function res = barx(rex,prior);
488 % PURPOSE: Bayesian estimation of an univariate BARX model
489 % -----
490 % SYNTAX: res = barx(rex,prior)
491 % -----
492 % OUTPUT: a structure ...
493 %     res.meth      = 'ARX MODEL: UNRESTRICTED ESTIMATION';
494 %     res.parm      = [n,nobs,k,p];
495 %     res.mu        = [mu]; <===== Intercept
496 %     res.mu_dt     = [mu_dt];
497 %     res.mu_t      = [mu_t];
498 %     res.b         = [b]; <===== Exogenous variables parameters
499 %     res.b_dt     = [b_dt];
500 %     res.b_t       = [b_t];
501 %     res.phi       = [phi]; <===== Lagged variables parameters
502 %     res.phi_dt    = [phi_dt];
503 %     res.phi_t     = [phi_t];
504 %     res.u         = [u]; <===== Residuals
505 %     res.sigma     = [sigma];
506 %     res.perform   = [log_like,aic,bic];
507 % -----
508 % INPUT: rex --> a structure ...
509 %     X      : nxm matrix of exogenous series, columnwise
510 %     z      : nx1 vector of original series, columnwise
511 %     opC    : 1x1 indicator to include (1) an intercept or not (0)
512 %     p      : 1x1 order of the model
513 %     prior  --> a structure that depends on the selected prior, see the
514 %     specific prior generating function
515 % -----
516 % LIBRARY: lag, uisolator
517 % -----
518 % SEE ALSO: arx, uprior_L, uprior_RS, uprior_C
519 % -----
520 % written by:
521 % Enrique M. Quilis
522 % Macroeconomic Research Department
523 % Ministry of Economy and Finance
524 % Paseo de la Castellana, 162. Office 2.5-1.
525 % 28046 - Madrid (SPAIN)

```

```

526 % <enrique.quilis@meh.es>
527 % September 2007
528 % Version 1.1
529
530 % -----
531 % Loading the structure
532
533 % Exogenous variables
534 Xx = rex.X;
535
536 % Intercept: yes (1) or not (0)
537 opC = rex.opC;
538
539 % Basic variable
540 z = rex.z;
541
542 % Order of AR model
543 p = rex.p;
544
545 % -----
546 % Computing basic parameters
547 [n,k] = size(z);
548 if (k ~= 1)
549     error('*** ONLY UNIVARIATE MODELING, PLEASE ***');
550 end
551 [aux,m] = size(Xx);
552 nobs = n-p; % Number of effective observations
553
554 % -----
555 %           PREPARING THE MODEL IN LINEAR MODEL FORMAT: Y=XB+U
556 % -----
557 Xz = lag(z,p);
558 Xz = Xz(:,2:end); % Dropping vector of ones (no intercept is assumed)
559 z = z(p+1:end);
560 if (isempty(Xx) == 0) % Exogenous variables are included
561     Xx = Xx(p+1:end,:);
562 end
563
564 % Intercept
565 switch opC
566 case 0 % No intercept
567     X = [Xx Xz];
568 case 1 % Intercept
569     X = [ones(nobs,1) Xx Xz];
570 end
571
572 % -----
573 %           ESTIMATION OF THE MODEL: MIXED THEIL-GOLDBERGER
574 sigmai = 1/prior.vu;
575 beta = (inv(prior.prior_beta_var) + (sigmai * X'*X)) \ ...
576     (inv(prior.prior_beta_var) * prior.prior_beta + ((sigmai * X'*z)));
577 beta_dt = sqrt(diag(inv((inv(prior.prior_beta_var) + (sigmai * X'*X))));
578
579 % Residuals
580 u = z - X * beta ;
581
582 % variance of residuals
583 sigma = var(u,1);
584
585 % VCV and t-ratios of beta parameters
586 beta_t = beta ./ beta_dt;

```

```

587
588 % Performance measures
589 log_like = -0.5*nobs*(1*(1+log(2*pi))+log(sigma));
590 npar = opC + m + p;
591 aic = -(2*log_like/nobs) + (2*npar)/nobs;
592 bic = -(2*log_like/nobs) + npar*(log(nobs)/nobs);
593
594 % -----
595 %           ISOLATING THE MATRICES
596 % -----
597 [mu,b,phi] = uisolator(beta,opC,m,p);
598 [mu_dt,b_dt,phi_dt] = uisolator(beta_dt,opC,m,p);
599 [mu_t,b_t,phi_t] = uisolator(beta_t,opC,m,p);
600
601 % -----
602 %           LOADING THE STRUCTURE
603 res.meth      = 'ARX MODEL: BAYESIAN ESTIMATION';
604 res.parm      = [n,nobs,k,p];
605 res.mu        = [mu];
606 res.mu_dt     = [mu_dt];
607 res.mu_t      = [mu_t];
608 res.b         = [b];
609 res.b_dt      = [b_dt];
610 res.b_t       = [b_t];
611 res.phi       = [phi];
612 res.phi_dt    = [phi_dt];
613 res.phi_t     = [phi_t];
614 res.u         = [u];
615 res.sigma     = [sigma];
616 res.perform   = [log_like,aic,bic];
617 res.prior     = prior;
618 % End of function barx
619 % %-----
620
621 function []=arx_print(res,file_sal);
622 % PURPOSE: Prints the output of the estimation of an ARX-BARX model
623 % -----
624 % SYNTAX:  [] = arx_print(res,file_sal);
625 % -----
626 % OUTPUT: an ASCII file (file_sal) containing information about:
627 %         res.meth      = label
628 %         res.parm      = number of: observations, effective number
629 %                       of observations, series, order of AR model
630 %         res.mu        = 1x1 intercept
631 %         res.mu_dt     = 1x1 intercept: standard error
632 %         res.mu_t      = 1x1 intercept: t-ratios
633 %         res.b         = mx1 intercept
634 %         res.b_dt      = mx1 intercept: standard error
635 %         res.b_t       = mx1 intercept: t-ratios
636 %         res.phi       = p vector of phi(h) parameters
637 %         res.phi_dt    = phi(h) parameters: standard errors
638 %         res.phi_t     = phi(h) parameters: t-ratios
639 %         res.sigma     = variance of residuals
640 %         res.perform   = log-likelihood,statistics AIC and BIC
641 % -----
642 % INPUT: res a structure generated by arx/barx function
643 %        file_sal: name of the ASCII file for output (optional)
644 % -----
645 % SEE ALSO: arx, barx
646 % -----
647 % NOTE: Prior specification: Litterman, Raynauld-Simonato.

```



```

709         switch res.prior.rex.type
710             case 1
711                 fprintf(fid,' Mean: centered around (1-B)(1-B^s) \n');
712             case 2
713                 fprintf(fid,' Mean: centered around (1-B^s) \n');
714             end
715             fprintf(fid,' Hyperparameters: tightness \n ');
716             fprintf(fid,' Intercept %12.4f \n', res.prior.rex.th_mu);
717             fprintf(fid,' Exogenous %12.4f \n', res.prior.rex.th_b);
718             fprintf(fid,' Global %12.4f \n', res.prior.rex.th(1));
719             fprintf(fid,' Decay: regular part %12.4f \n', ...
720                 res.prior.rex.th(2));
721             fprintf(fid,' Decay: seasonal part %12.4f \n', ...
722                 res.prior.rex.th(3));
723         case {'CANOVA PRIOR'}
724             fprintf(fid,'%s\n ', res.prior.meth );
725             fprintf(fid,' Under construction!! \n ');
726             % TO BE INCLUDED: INFO ABOUT HYPERPARAMETERS
727         end
728     end
729     fprintf(fid,'%s \n ', sep2 );
730     fprintf(fid,'Number of observations      : %4d\n ',res.parm(1));
731     fprintf(fid,'Number of effective observations : %4d\n ',res.parm(2));
732     fprintf(fid,'Number of series              : %4d\n ',res.parm(3) );
733     fprintf(fid,'Number of lags                : %4d\n ',res.parm(4) );
734     fprintf(fid,'\n ');
735     fprintf(fid,'%s \n ', sep2 );
736     fprintf(fid, ...
737         'mu: Estimate, s.d. and t-ratio (read columnwise) \n ');
738     fprintf(fid,'%s \n ', sep1 );
739     fprintf(fid,'%12.4f      %12.4f      %8.4f \n ',mu_L);
740     fprintf(fid,'%s \n ', sep1 );
741     fprintf(fid,'b: Estimate, s.d. and t-ratio (read columnwise) \n ');
742     fprintf(fid,'%s \n ', sep1 );
743     fprintf(fid,'%12.4f      %12.4f      %8.4f \n ',b_L);
744     fprintf(fid,'%s \n ', sep1 );
745     fprintf(fid,'phi: Lag, estimate, s.d. and t-ratio (read columnwise) \n ');
746     fprintf(fid,'%s \n ', sep1 );
747     fprintf(fid,'%4d %12.4f      %12.4f      %8.4f \n ',phi_L);
748     fprintf(fid,'%s \n ', sep1 );
749     fprintf(fid,'RESIDUALS: s.d. ');
750     fprintf(fid,'\n');
751     fprintf(fid,' %12.4f ', sqrt(res.sigma));
752     fprintf(fid,'\n');
753     fprintf(fid,'%s \n ', sep1 );
754     fprintf(fid,'Model performance (read columnwise): \n');
755     fprintf(fid,' - log-likelihood\n');
756     fprintf(fid,' - AIC and BIC \n');
757     fprintf(fid,' - BIC \n');
758     fprintf(fid,'%s \n ', sep1 );
759     fprintf(fid,'%12.4f      %8.4f      %8.4f \n ',res.perform);
760     fprintf(fid,'%s \n ', sep2 );
761     fprintf(fid,'\n');
762     if (nargin == 1)
763         % Do nothing
764     else
765         fclose(fid);
766     end
767     % End of function arx_print
768     % %-----

```

이 MATLAB 프로그램 BayARX\_barx\_L\_d.m에서는 Litterman [68], [69], [70]이 제시한 사전확률분포를 사용하였다.

이 MATLAB 프로그램 BayARX\_barx\_L\_d.m을 실행하면, 다음과 같은 결과가 출력된다.

```
*****
ARX MODEL: BAYESIAN ESTIMATION
*****
LITTERMAN PRIOR
Hyperparameters: tightness
Intercept      0.0000
Exogenous     1000.0000
Decay function: harmonic
Global        0.1000
Decay         0.0500
*****
Number of observations      : 239
Number of effective observations : 233
Number of series           : 1
Number of lags             : 6
*****
mu: Estimate, s.d. and t-ratio (read columnwise)
-----
-----
b: Estimate, s.d. and t-ratio (read columnwise)
-----
-----
phi: Lag, estimate, s.d. and t-ratio (read columnwise)
-----
1      0.8818      0.0495      17.7994
2      0.0802      0.0582      1.3787
3      0.1008      0.0587      1.7163
4     -0.0166      0.0580     -0.2853
5      0.0080      0.0565      0.1408
6     -0.0983      0.0476     -2.0658
-----
RESIDUALS: s.d.
```

```

2.1134
-----
Model performance (read columnwise):
- log-likelihood
- AIC and BIC
- BIC
-----
-504.9620      4.3859      4.4748
*****

```

이 결과물에서 알 수 있듯이, 이 예제에서  $AR(6)$  모형을 베이지안추정한 결과는 다음과 같다.

$$\begin{aligned} \tilde{y}_t = & 0.8818\tilde{y}_{t-1} + 0.0802\tilde{y}_{t-2} + 0.1008\tilde{y}_{t-3} \\ & - 0.0166\tilde{y}_{t-4} + 0.0080\tilde{y}_{t-5} - 0.0983\tilde{y}_{t-6} \end{aligned} \tag{1}$$

또한 잔차들의 표준편차는 다음과 같다.

$$\tilde{\sigma} \doteq \sqrt{\frac{1}{233} \sum_{t=7}^{239} [\tilde{y}_t - y_t]^2} = 2.1134 \tag{2}$$



#### 6.4.4 MH 샘플러를 사용한 자기회귀오차모형분석

베이지안추정법을 사용해서 오차항들이 자기회귀과정 (autoregression process) 을 이루는 선형회귀모형을 추정하는 문제를 살펴보자. 즉, 다음과 같은 모형을 살펴보자.

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_K x_{tK} + \epsilon_t, \quad (t = 1, 2, \dots, n) \tag{6.4.39}$$

여기서  $y_t$  는 피설명변수,  $\mathbf{x}_t \doteq [x_{t1}, x_{t2}, \dots, x_{tK}]^t$  는 설명변수들의 벡터,  $\beta_1, \beta_2, \dots, \beta_K$  는 모수들이고,  $\epsilon_t$  는 오차항이다. 오차항들  $\{\epsilon_t\}$  는 다음과 같은  $p$  차 자기회귀모형, 즉  $AR(p)$

모형을 따른다고 가정하자.

$$\epsilon_t = \phi_1\epsilon_{t-1} + \phi_2\epsilon_{t-2} + \cdots + \phi_p\epsilon_{t-p} + u_t, \quad (t = p+1, p+2, \dots, n) \quad (6.4.40)$$

여기서  $\phi_1, \phi_2, \dots, \phi_p$ 는 모수들이고, 오차항들  $\{u_t\}$ 는 서로 독립이며 정규확률분포  $\mathcal{N}(0, \sigma^2)$ 를 따른다고 하자. 다음 함수를 정의하자.

$$\phi(z) \doteq 1 - \phi_1z - \phi_2z^2 - \cdots - \phi_pz^p \quad (6.4.41)$$

식 (6.4.40)를 다음과 같이 쓸 수 있다.

$$\phi(L)\epsilon_t = u_t \quad (6.4.42)$$

여기서  $L$ 은 후진연산자(backshift operator 또는 lag operator)이다. 방정식  $\phi(z) = 0$ 의 근이 모두 단위원 밖에 있다고 가정하자. 다음과 같이  $\mathbf{y}$ ,  $X_n$ ,  $\boldsymbol{\beta}$  그리고  $\boldsymbol{\phi}$ 를 정의하자.

$$\mathbf{y} \doteq \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X_n \doteq \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nK} \end{bmatrix}, \quad \boldsymbol{\beta} \doteq \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}, \quad \boldsymbol{\phi} \doteq \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad (6.4.43)$$

또한, 정밀도  $\psi$ 는 분산  $\sigma^2$ 의 역수이다.

모수들  $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi})$ 의 사전확률분포가 다음과 같다고 가정하자.

$$\boldsymbol{\beta} \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_0, B_0), \quad \sigma^2 \stackrel{d}{\sim} \text{IG}\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right), \quad \boldsymbol{\phi} \stackrel{d}{\sim} \mathcal{TN}_S(\mathbf{q}_0, Q_0) \quad (6.4.44)$$

여기서  $\mathcal{TN}_S(\mathbf{q}_0, Q_0)$ 는 다변량정규확률분포  $\mathcal{N}(\mathbf{q}_0, Q_0)$ 를 영역  $S$ 에서 절단한 절단다변량정규확률분포를 나타낸다.

사후확률분포를 구하는 첫 번째 단계로  $\boldsymbol{\beta} \mid (\boldsymbol{\phi}, \sigma^2)$ 의 사후조건부확률분포를 구해보자. 각  $t(\geq p+1)$ 에 대해서 다음 변수들을 정의하자.

$$y_t^* \doteq \phi(L)y_t, \quad x_{tj}^* \doteq \phi(L)x_{tj}, \quad (j = 1, 2, \dots, p) \quad (6.4.45)$$

식 (6.4.39), 식 (6.4.40), 그리고 식 (6.4.45)에서 알 수 있듯이, 다음 식들이 성립한다.

$$y_t^* = \beta_1 x_{t1}^* + \beta_2 x_{t2}^* + \cdots + \beta_p x_{tp}^* + u_t, \quad (t = p+1, p+2, \dots, n) \quad (6.4.46)$$

여기서  $\{u_t | \sigma^2\}$  들은 서로 독립이며 평균이 0이고 분산이  $\sigma^2$  인 정규확률분포를 따른다. 따라서, 식 (6.4.46)는 표준적 선형회귀모형을 나타낸다. 확률벡터  $\mathbf{y}_p \doteq [y_1, y_2, \dots, y_p]^t$  는  $\boldsymbol{\epsilon}_p \doteq [\epsilon_1, \epsilon_2, \dots, \epsilon_p]^t$  의 함수이고,  $\boldsymbol{\epsilon}_p$  는  $u_p, u_{p-1}, \dots, u_1$  의 함수이다. 다음 식이 성립한다.

$$\boldsymbol{\epsilon}_p \stackrel{d}{\sim} \mathcal{N}(0, \sigma^2 \Sigma_p) \quad (6.4.47)$$

여기서  $\Sigma_p$ 는 분산공분산행렬로서 다음과 같이 구할 수 있다. 식 (6.4.40)에서 알 수 있듯이, 각  $j (= 1, 2, \dots, p)$ 에 대해서 다음 식들이 성립한다.

$$\text{Cov}(\epsilon_t, \epsilon_{t-j}) = \phi_1 \text{Cov}(\epsilon_{t-1}, \epsilon_{t-j}) + \cdots + \phi_p \text{Cov}(\epsilon_{t-p}, \epsilon_{t-j}) + \text{Cov}(u_t, \epsilon_{t-j}) \quad (6.4.48)$$

$$\text{Cov}(u_t, \epsilon_{t-j}) = \sigma^2 1_{\{0\}}(j) \quad (6.4.49)$$

식 (6.4.40)에서 알 수 있듯이, 다음 식이 성립한다.

$$\boldsymbol{\epsilon}_p = \Phi \boldsymbol{\epsilon}_{p-1} + u_p \mathbf{e}_1 \quad (6.4.50)$$

여기서  $\boldsymbol{\epsilon}_{p-1} \doteq [\epsilon_0, \epsilon_1, \dots, \epsilon_{p-1}]^t$  이고, 또한  $\Phi$ 와  $\mathbf{e}_1$ 은 각각 다음과 같다.

$$\Phi = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad \mathbf{e}_1 \doteq \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (6.4.51)$$

방정식  $\phi(z) = 0$ 의 근이 모두 단위원 밖에 있으므로,  $\boldsymbol{\epsilon}_p$ 와  $\boldsymbol{\epsilon}_{p-1}$ 은 동일하게 평균벡터  $\mathbf{0}$ 과 분산공분산행렬  $\sigma^2 \Sigma_p$ 를 갖는다. 따라서, 다음 식들이 성립한다.

$$\sigma^2 \Sigma_p = \text{Var}(\boldsymbol{\epsilon}_p) = \Phi \text{Var}(\boldsymbol{\epsilon}_{p-1}) \Phi^t + \sigma^2 \mathbf{e}_1 \mathbf{e}_1^t = \sigma^2 \Phi \Sigma_p \Phi^t + \sigma^2 \mathbf{e}_1 \mathbf{e}_1^t \quad (6.4.52)$$

즉, 다음 식이 성립한다.

$$\Sigma_p = \Phi \Sigma_p \Phi^t + \mathbf{e}_1 \mathbf{e}_1^t \quad (6.4.53)$$

식 (6.4.47)에서 알 수 있듯이, 다음 식이 성립한다.

$$\mathbf{y}_p \stackrel{d}{\sim} \mathcal{N}(X_p \boldsymbol{\beta}, \sigma^2 \Sigma_p) \quad (6.4.54)$$

행렬  $\Sigma_p$ 에 Cholesky 분해를 적용하면, 다음 식을 만족하는 하삼각행렬  $T$ 를 얻는다.

$$\Sigma_p = T T^t \quad (6.4.55)$$

다음 확률벡터들을 정의하자.

$$\mathbf{y}_p^* \doteq T^{-1} \mathbf{y}_p, \quad X_p^* \doteq T^{-1} X_p \quad (6.4.56)$$

식 (6.4.54)에서 알 수 있듯이, 다음 식이 성립한다.

$$\mathbf{y}_p^* \stackrel{d}{\sim} \mathcal{N}(X_p^* \boldsymbol{\beta}, \sigma^2 I_p) \quad (6.4.57)$$

이로부터  $y_1^*, y_2^*, \dots, y_n^*$ 의 우도함수가 다음과 같음을 알 수 있다.

$$f(y_1^*, y_2^*, \dots, y_n^* | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}) \propto [\sigma^2]^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{t=1}^n [y_t^* - [\mathbf{x}_t^*]^t \boldsymbol{\beta}]^2\right) \quad (6.4.58)$$

여기서  $\mathbf{x}_t^*$ 를  $\mathbf{x}_t^* \doteq [x_{t1}^*, x_{t2}^*, \dots, x_{tp}^*]^t$ 로 표기하자. 식 (6.4.44)과 식 (6.4.58)에서 알 수 있듯이,  $\boldsymbol{\beta} | (\boldsymbol{\phi}, \sigma^2)$ 의 사후조건부확률분포는 다음과 같다.

$$\boldsymbol{\beta} | (\sigma^2, \boldsymbol{\phi}, \mathbf{y}) \stackrel{d}{\sim} \mathcal{N}(\mathbf{b}_1, B_1) \quad (6.4.59)$$

여기서  $\mathbf{b}_1$  과  $B_1$  은 각각 다음과 같다.

$$\mathbf{b}_1 \doteq B_1 \left[ B_0^{-1} \mathbf{b}_0 + \frac{1}{\sigma} \sum_{t=1}^n [\mathbf{x}_t^*]^t y_t^* \right] \quad (6.4.60)$$

$$B_1 \doteq \left[ B_0^{-1} + \sigma^{-2} \sum_{t=1}^n \mathbf{x}_t^* [\mathbf{x}_t^*]^t \right]^{-1} \quad (6.4.61)$$

두 번째 단계로, 같은 방법을 사용해서  $\sigma^2 \mid (\boldsymbol{\beta}, \boldsymbol{\phi})$  의 사후조건부확률분포가 다음과 같음을 증명할 수 있다.

$$\sigma^2 \mid (\boldsymbol{\beta}, \boldsymbol{\phi}, \mathbf{y}) \stackrel{d}{\sim} \text{IG} \left( \frac{n_1}{2}, \frac{n_1 S_1}{2} \right) \quad (6.4.62)$$

여기서  $n_1$  과  $S_1$  은 각각 다음과 같다.

$$n_1 \doteq n_0 + n, \quad S_1 \doteq \frac{1}{n_1} \left\{ n_0 S_0 + \sum_{t=1}^n [y_t^* - [\mathbf{x}_t^*]^t \boldsymbol{\beta}]^2 \right\} \quad (6.4.63)$$

세 번째 단계로,  $\boldsymbol{\phi} \mid (\boldsymbol{\beta}, \sigma^2)$  의 사후조건부확률분포를 구해보자. 각  $t (\geq p+1)$  에 대해서  $\mathbf{z}_t \doteq [\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-p}]^t$  라고 하면, 다음 식이 성립한다.

$$\epsilon_t \mid \mathbf{z}_t \stackrel{d}{\sim} \mathcal{N}(\mathbf{z}_t^t \boldsymbol{\phi}, \sigma^2) \quad (6.4.64)$$

식 (6.4.47) 에서 알 수 있듯이  $\boldsymbol{\epsilon}_p$  는  $\mathcal{N}(0, \sigma^2 \Sigma_p)$  를 따르고,  $\Sigma_p$  는  $\boldsymbol{\phi}$  의 함수이다. 따라서,  $\boldsymbol{\phi} \mid (\boldsymbol{\beta}, \sigma^2)$  의 사후조건부확률밀도함수는 다음과 같다.

$$\begin{aligned} \pi(\boldsymbol{\phi} \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}) &\propto \Sigma_p^{-1/2} \exp \left( -\frac{1}{2\sigma^2} \left\{ \boldsymbol{\epsilon}_p^t \Sigma_p^{-1} \boldsymbol{\epsilon}_p + \sum_{t=p+1}^n [\epsilon_t - \mathbf{z}_t^t \boldsymbol{\phi}]^2 \right\} \right) \\ &\cdot \exp \left( -\frac{1}{2} [\boldsymbol{\phi} - \mathbf{q}_0]^t Q_0^{-1} [\boldsymbol{\phi} - \mathbf{q}_0] \right) 1(\boldsymbol{\phi} \in S) \\ &\cdot A(\boldsymbol{\phi}) \exp \left( -\frac{1}{2} [\boldsymbol{\phi} - \mathbf{q}_1]^t Q_1^{-1} [\boldsymbol{\phi} - \mathbf{q}_1] \right) 1(\boldsymbol{\phi} \in S) \end{aligned} \quad (6.4.65)$$

여기서 상수, 변수, 벡터들, 그리고 행렬은 다음과 같다.

$$A(\boldsymbol{\phi}) \doteq |\Sigma_p|^{-1/2} \exp\left(-\frac{1}{2\sigma^2} \boldsymbol{\epsilon}_p^t \Sigma_p^{-1} \boldsymbol{\epsilon}_p\right) \quad (6.4.66)$$

$$\mathbf{q}_1 \doteq Q_1 \left[ Q_0^{-1} \mathbf{q}_0 + \sigma^2 \sum_{t=p+1}^n \mathbf{z}_t \epsilon_t \right] \quad (6.4.67)$$

$$Q_1 \doteq \left[ Q_0^{-1} + \sum_{t=p+1}^n \mathbf{z}_t \mathbf{z}_t^t \right]^{-1} \quad (6.4.68)$$

$$\epsilon_t = y_t - \mathbf{x}_t^t \boldsymbol{\beta} \quad (6.4.69)$$

$$\boldsymbol{\epsilon}_p = \mathbf{y}_p - X_p \boldsymbol{\beta} \quad (6.4.70)$$

사후조건부확률분포 (6.4.65)에서  $\boldsymbol{\phi}$ 를 샘플링하기 위해서는 다음과 같은 제안분포에서  $\boldsymbol{\phi}$ 를 발생시킨다.

$$\boldsymbol{\phi} \stackrel{d}{\sim} \mathcal{TN}_S(\mathbf{q}_1, Q_1) \quad (6.4.71)$$

여기서  $\boldsymbol{\phi}^{(j-1)}$ 는 제  $j-1$ 번째 단계에서 모수벡터이다. 이 후보점  $\boldsymbol{\phi}$ 를 다음 확률로 채택한다.

$$\alpha\left(\boldsymbol{\phi}^{(j-1)}, \boldsymbol{\phi} \mid \boldsymbol{\beta}^{(j)}, \sigma^{2(j)}, \mathbf{y}\right) = \min\left\{\frac{A(\boldsymbol{\phi})}{A(\boldsymbol{\phi}^{(j-1)})}, 1\right\} \quad (6.4.72)$$

지금까지 내용을 바탕으로, 오차항이 자기회귀과정에 따르는 회귀모형의 모수들  $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi})$ 의 사후확률분포에서 샘플링하는 MH샘플러를 다음과 같이 요약할 수 있다.

#### 알고리즘 6.4.4: 자기회귀오차모형의 MH샘플러

(1단계) 초기모수들  $(\boldsymbol{\beta}^{(0)}, \sigma^{2(0)}, \boldsymbol{\phi}^{(0)})$ 를 선택한다.

(2단계) 각  $j(= 1, 2, \dots)$ 에 대해서 주어진  $\mathbf{y}$ 에 대한 조건부샘플  $(\boldsymbol{\beta}^{(j)}, \sigma^{2(j)}, \boldsymbol{\phi}^{(j)})$ 를 발생시킨다.

(2-1) 식 (6.4.59)의 사후조건부확률분포에서  $\boldsymbol{\beta}^{(j)} \mid (\sigma^{2(j-1)}, \boldsymbol{\phi}^{(j-1)}, \mathbf{y})$ 를 발생시킨다.

(2-2) 식 (6.4.62)의 사후조건부확률분포에서  $\sigma^{2(j)} \mid (\boldsymbol{\beta}^{(j)}, \boldsymbol{\phi}^{(j-1)}, \mathbf{y})$ 를 발생시킨다.

(2-3) 식 (6.4.71)에서 후보점  $\boldsymbol{\phi}$ 를 발생시키고, MH채택확률  $\alpha(\boldsymbol{\phi}^{(j-1)}, \boldsymbol{\phi} \mid \boldsymbol{\beta}^{(j)}, \sigma^{2(j)}, \mathbf{y})$

를 계산한다. 일양난수  $u$ 를 이용해서, 다음과 같이  $\phi^{(j)}$ 를 결정한다.

$$\phi^{(j)} = \begin{cases} \phi, & (u \leq \alpha(\phi^{(j-1)}, \phi | \beta^{(j)}, \sigma^{2(j)}, \mathbf{y})) \\ \phi^{(j-1)}, & (u > \alpha(\phi^{(j-1)}, \phi | \beta^{(j)}, \sigma^{2(j)}, \mathbf{y})) \end{cases}$$

(제3단계) 충분히 큰  $b$ 에 대해  $(\beta^{(j)}, \sigma^{2(j)}, \phi^{(j)})$ ,  $(j = b + 1, b + 2, \dots)$ 를 기록한다.

# 부록 A

We may conclude that characteristic function methods are most powerful techniques that the empirically interested finance researcher should know.

Jondeau, Poon, and Rockinger [57, p. 486]

## A.1 특성함수

중심극한정리를 증명할 때, 적률모함수(moment generating function) 또는 특성함수(characteristic function)를 사용한다.

특성함수(characteristic function)를 다음과 같이 정의한다.

### 정의 A.1

확률변수  $x$ 의 확률분포함수가  $F_x(\cdot)$ 라고 하면, 확률변수  $x$  또는 확률분포함수  $F_x(\cdot)$ 의 특성함수를 다음과 같이 정의한다.

$$\phi_x(t) \doteq E(e^{itx}) = \int_{-\infty}^{\infty} e^{itx} dF_x(x), \quad (t \in \mathbb{R})$$

만약 확률변수  $x$ 의 확률밀도함수  $f_x(\cdot)$ 가 존재하면, 특성함수를 다음과 같이 쓸 수 있다.

$$\phi_x(t) = \int_{-\infty}^{\infty} e^{itx} f_x(x) dx = \mathcal{F}[f_x](-t) = \hat{f}_x(-t) \quad (\text{A1.1})$$

여기서  $\mathcal{F}[f](t)$ 와  $\hat{f}(t)$ 는 함수  $f(x)$ 의 Fourier 변환이다.

Epps [35]는 특성함수를 다음과 같이 기하학적으로 설명했다. Euler식에서 알 수 있듯이, 확률변수  $\exp(itx)$ 는 복소평면의 단위원 위에 존재한다. 따라서,  $t$ 에서 계산되는 특성함수값  $\phi_x(t) = E(e^{itx})$ 는 단위원을 둘러싸은(wrapped around) 확률변수  $e^{itx}$ 의 질량중심(center of mass)으로 해석할 수 있다. 따라서, 특성함수는 무한히 긴 실린더 내부와 위에서 움직이는 곡선이라고 할 수 있다.

**예제 A.1** 표준정규확률분포  $\mathcal{N}(0, 1)$ 를 따르는 표준정규확률변수  $x$ 의 특성함수를 다음과 같이 구할 수 있다.

$$\begin{aligned}\phi_x(t) &= \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx \\ &= \int_0^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx + \int_{-\infty}^0 e^{itx} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx \\ &= \int_0^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx + \int_0^{\infty} e^{-itx} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx\end{aligned}\quad (1)$$

따라서 다음 식이 성립한다.

$$\phi_x(t) = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2}x^2} \cos(tx) dx \quad (2)$$

식 (2)에서 알 수 있듯이, 다음 식이 성립한다.

$$\begin{aligned}\phi'_x(t) &= -\frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2}x^2} x \sin(tx) dx \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \sin(tx) d\left(e^{-\frac{1}{2}x^2}\right) \\ &= \frac{2}{\sqrt{2\pi}} \left[ \sin(tx) e^{-\frac{1}{2}x^2} \right]_{x=0}^{x=\infty} - \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2}x^2} t \cos(tx) dx\end{aligned}\quad (3)$$

식 (2)와 식 (3)에서 알 수 있듯이, 다음 식이 성립한다.

$$\phi'_x(t) = -t \phi_x(t) \quad (4)$$

정규확률밀도함수의 성질에서 알 수 있듯이, 다음 식이 성립한다.

$$\phi_x(0) = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2}x^2} \cos(0 \cdot x) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1. \quad (5)$$

초기조건 (5)를 만족하는 미분방정식 (4)를 풀면, 표준정규확률분포의 특성함수가 다음과

같음을 알 수 있다.

$$\phi_x(t) = e^{-\frac{1}{2}t^2} \tag{6}$$

복소해석학(complex analysis)을 적용해서 특성함수  $\phi_x(t)$ 를 구하기로 하자. 복소해석학을 모르는 독자는 그냥 넘어가기를 권한다. 다음 식들이 성립한다.

$$\phi_x(t) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = e^{-\frac{1}{2}t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[x-it]^2} dx \tag{7}$$

따라서 우리의 목적은 다음 식을 증명하는 것이다.

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[x-it]^2} dx = 1 \tag{8}$$

다음 함수를 정의하자.

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \tag{9}$$

그림 A.1의 적분경로를  $C$ 로 표기하자. Cauchy정리에서 알 수 있듯이, 다음 식들이 성립한다.

$$\oint_C f(z) dz = \oint_C \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 0 \tag{10}$$

따라서 다음 식이 성립한다.

$$I_1 + I_2 + I_3 + I_4 = 0 \tag{11}$$

여기서  $I_k$ 는 다음과 같다.

$$I_k \doteq \int_{C_k} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz, \quad (k = 1, 2, 3, 4) \tag{12}$$

다음 식들이 성립한다.

$$\begin{aligned} \lim_{M \rightarrow \infty} I_3 &= \lim_{M \rightarrow \infty} \int_{M-ti}^{-M-ti} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\ &= \lim_{M \rightarrow \infty} \int_M^{-M} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[u-ti]^2} du = - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[x-it]^2} dx \end{aligned} \tag{13}$$

다음 식들이 성립한다.

$$\lim_{M \rightarrow \infty} I_1 = \lim_{M \rightarrow \infty} \int_{-M}^M \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1 \quad (14)$$

또한 다음 식들이 성립한다.

$$\begin{aligned} \lim_{M \rightarrow \infty} |I_2| &= \lim_{M \rightarrow \infty} \left| \int_0^{-t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[M+yi]^2} dy \right| \leq \lim_{M \rightarrow \infty} \int_{-t}^0 \left| \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[M+yi]^2} \right| dy \\ &= \lim_{M \rightarrow \infty} e^{-\frac{1}{2}M^2} \int_{-t}^0 \left| \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}y^2} e^{-Myi} \right| dy = \lim_{M \rightarrow \infty} e^{-\frac{1}{2}M^2} \int_{-t}^0 \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}y^2} dy \end{aligned} \quad (15)$$

즉, 다음 식이 성립한다.

$$\lim_{M \rightarrow \infty} I_2 = 0 \quad (16)$$

같은 방법으로 다음 식이 성립함을 증명할 수 있다.

$$\lim_{M \rightarrow \infty} I_4 = 0 \quad (17)$$

식 (13)~식 (17)을 식 (11)에 대입하면, 식 (8)이 성립함을 알 수 있다. ■

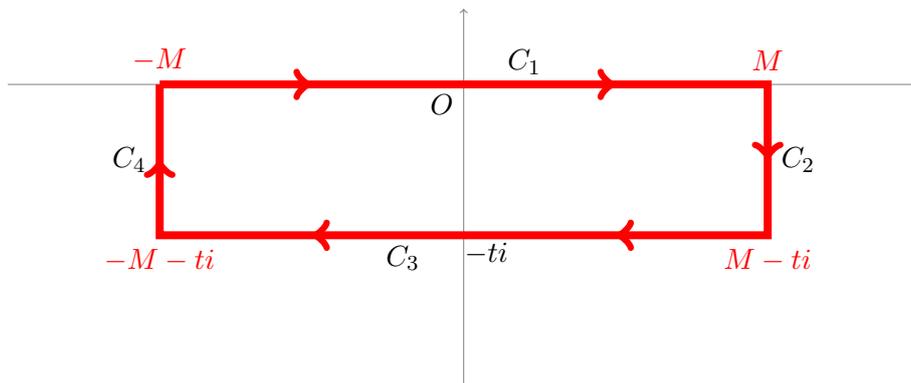


그림 A.1. 적분경로

**예제 A.2** 정규확률분포  $\mathcal{N}(\mu, \sigma^2)$ 를 따르는 정규확률변수  $x$ 의 특성함수를 다음과 같이

구할 수 있다.

$$\begin{aligned} \phi_x(t) &= \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2} \left[\frac{x-\mu}{\sigma}\right]^2\right) dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2} [x^2 - 2\mu x + \mu^2 - 2i\sigma^2 tx]\right) dx \\ &= \exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2} \left[\frac{x-\mu-i\sigma^2 t}{\sigma}\right]^2\right) dx \end{aligned} \quad (1)$$

다음 식이 성립한다.

$$\begin{aligned} &\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2} \left[\frac{x-\mu-i\sigma^2 t}{\sigma}\right]^2\right) dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} [u-i\sigma t]^2\right) du = 1 \end{aligned} \quad (2)$$

여기서 첫 번째 등호는 변수변환  $u = [x - \mu - i\sigma^2 t]/\sigma$ 에 의해서 성립하고, 두 번째 등호는 예제 A.1의 식 (8)에 의해서 성립한다. 식 (2)를 식 (1)에 대입하면, 다음 식을 얻는다.

$$\phi_x(t) = \exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right) \quad (3)$$

■

**예제 A.3** 다음과 같은 Cauchy 확률밀도함수의 특성함수  $\phi_x(t)$ 를 구하기로 하자.

$$f_x(x) = \frac{1}{\pi[1+x^2]} \quad (1)$$

다음 식이 성립한다.

$$\phi_x(t) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\pi[1+x^2]} dx \quad (2)$$

만약  $t > 0$ 이면, 그림 A.2의 좌측 적분경로를  $C_\alpha$ 로 표기하고 다음 적분을 고려해보자.

$$I = \int_{C_\alpha} \frac{e^{itz}}{1+z^2} dz \quad (3)$$

Cauchy정리에서 알 수 있듯이 함수  $f(z) = e^{itz}/[1+z^2]$ 에 대해서 다음 식들이 성립한다.

$$\oint_{C_\alpha} f(z)dz = 2\pi i \operatorname{Res}(f, i) = 2\pi i \lim_{z \rightarrow i} [z-i] \frac{e^{itz}}{1+z^2} = 2\pi i \frac{e^{-t}}{2i} = \pi e^{-t} \quad (4)$$

따라서 다음 식이 성립한다.

$$\int_{C_1} \frac{e^{itz}}{1+z^2} dz + \int_{C_R} \frac{e^{itz}}{1+z^2} dz = \pi e^{-t} \quad (5)$$

다음 식들이 성립한다.

$$\left| \int_{C_R} \frac{e^{itz}}{1+z^2} dz \right| \leq \int_{C_R} \left| \frac{e^{itz}}{1+z^2} \right| dz \leq \pi R \frac{\exp(-tR \sin(\arg(z)))}{R^2-1} \quad (6)$$

적분구간  $C_R$ 에서 부등식  $\sin(\arg(z)) > 0$ 가 성립하므로, 식 (6)에서 알 수 있듯이 다음 식이 성립한다.

$$\lim_{R \rightarrow \infty} \int_{C_R} \frac{e^{itz}}{1+z^2} dz = 0 \quad (7)$$

식 (6)과 식 (7)에서 알 수 있듯이, 다음 식이 성립한다.

$$\int_{-\infty}^{\infty} \frac{e^{itx}}{1+x^2} dx = \lim_{R \rightarrow \infty} \int_{C_1} \frac{e^{itz}}{1+z^2} dz = \pi e^{-t}, \quad (t > 0) \quad (8)$$

만약  $t < 0$ 이면, 그림 A.2의 우측 적분경로를  $C_\beta$ 로 표기하고, 다음 적분을 고려해보자.

$$I = \int_{C_\beta} \frac{e^{itz}}{1+z^2} dz \quad (9)$$

Cauchy정리에서 알 수 있듯이 함수  $f(z) = e^{itz}/[1+z^2]$ 에 대해서 다음 식들이 성립한다.

$$\oint_{C_\beta} f(z)dz = 2\pi i \operatorname{Res}(f, -i) = 2\pi i \lim_{z \rightarrow -i} [z+i] \frac{e^{itz}}{1+z^2} = -2\pi i \frac{e^t}{2i} = -\pi e^t \quad (10)$$

따라서 다음 식이 성립한다.

$$\int_{C_1} \frac{e^{itz}}{1+z^2} dz + \int_{C_R} \frac{e^{itz}}{1+z^2} dz = -\pi e^t \quad (11)$$

다음 식들이 성립한다.

$$\left| \int_{C_R} \frac{e^{itz}}{1+z^2} dz \right| \leq \int_{C_R} \left| \frac{e^{itz}}{1+z^2} \right| dz \leq \pi R \frac{\exp(-tR \sin(\arg(z)))}{R^2-1} \quad (12)$$

적분구간  $C_R$ 에서 부등식  $\sin(\arg(z)) < 0$ 가 성립하고 가정에서 식  $t < 0$ 이 성립하므로, 식 (12)에서 알 수 있듯이 다음 식이 성립한다.

$$\lim_{R \rightarrow \infty} \int_{C_R} \frac{e^{itz}}{1+z^2} dz = 0 \quad (13)$$

식 (11)과 식 (13)에서 알 수 있듯이, 다음 식이 성립한다.

$$\int_{-\infty}^{\infty} \frac{e^{itx}}{1+x^2} dx = - \lim_{R \rightarrow \infty} \int_{C_1} \frac{e^{itz}}{1+z^2} dz = \pi e^t, \quad (t < 0) \quad (14)$$

식 (2), 식 (8)과 식 (14)에서 알 수 있듯이, 다음 식이 성립한다.

$$\phi_x(t) = e^{-|t|} \quad (14)$$

■

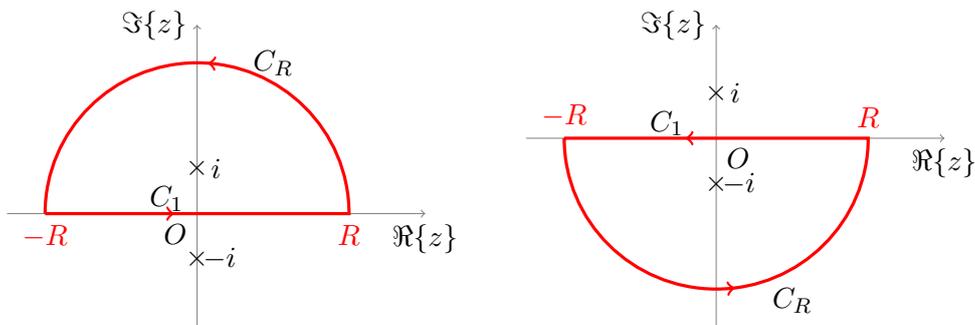


그림 A.2. 적분경로

**명제 A.1**

확률변수  $x$ 의 특성함수  $\phi_x(t)$ 는 다음 성질들을 갖는다.

(a)  $\phi_x(0) = 1$

$$(b) |\phi_x(t)| \leq 1, \quad (t \in \mathbb{R})$$

$$(c) \overline{\phi_x(t)} = \phi_x(-t)$$

(d) 각  $t \in \mathbb{R}$  에 대해서  $\phi_x(t)$  는 연속이다.

$$(e) \phi_{ax+b}(t) = e^{ibt} \phi_x(at)$$

증명. 다음 식들이 성립한다.

$$\phi_x(0) = \int_{-\infty}^{\infty} e^{i \cdot 0 \cdot x} dF_x(x) = \int_{-\infty}^{\infty} dF_x(x) = 1 \quad (1)$$

즉, 성질 (a) 가 성립한다. 다음 식들이 성립한다.

$$|\phi_x(t)| = \left| \int_{-\infty}^{\infty} e^{itx} dF_x(x) \right| \leq \int_{-\infty}^{\infty} |e^{itx}| dF_x(x) \leq \int_{-\infty}^{\infty} dF_x(x) = 1 \quad (2)$$

즉, 성질 (b) 가 성립한다. 다음 식들이 성립한다.

$$\overline{\phi_x(t)} = \int_{-\infty}^{\infty} \overline{e^{itx}} dF_x(x) = \int_{-\infty}^{\infty} e^{i[-t]x} dF_x(x) = \phi_x(-t) \quad (3)$$

즉, 성질 (c) 가 성립한다. 각 점  $x$  에서 지수함수가 연속이므로, 성질 (d) 가 성립한다. 다음 식들이 성립한다.

$$\begin{aligned} \phi_{ax+b}(t) &= E\left(e^{it[ax+b]}\right) = \int_{-\infty}^{\infty} e^{i[at]x} e^{itb} dF_x(x) \\ &= e^{itb} \int_{-\infty}^{\infty} e^{i[at]x} dF_x(x) = e^{itb} \phi_x(at) \end{aligned} \quad (4)$$

■

명제 A.1의 성질 (b)의 증명에서 알 수 있듯이, 특성함수는 항상 존재한다. 다음 명제가 성립함을 쉽게 알 수 있다.

### 명제 A.2

특성함수  $\phi_x(t)$ 가 점  $t = 0$ 에서  $n$ 번 미분가능하면, 다음 식이 성립한다.

$$\left. \frac{d^n \phi_x(t)}{dt^n} \right|_{t=0} = i^n E(x^n)$$

**명제 A.3**

만약 확률분포함수  $F_x(\cdot)$ 에 해당하는 확률밀도함수  $f_x(\cdot)$ 가 존재하면, 특성함수  $\phi_x(t)$ 는 다음 식을 만족한다.

$$\lim_{|t| \rightarrow \infty} \phi_x(t) = 0$$

만약 확률밀도함수  $f_x(\cdot)$ 가  $n$ 번 미분가능하면, 특성함수  $\phi_x(t)$ 는 다음 식을 만족한다.

$$\phi_x(t) = o(t^n) \text{ as } |t| \rightarrow \infty$$

증명. 첫 번째 성질의 증명은 Feller [39, Lemma 3, p. 513]를 참조하라.

부분적분을 사용해서 다음 식을 유도할 수 있다.

$$\phi_x(t) = \frac{1}{it} \int_{-\infty}^{\infty} e^{itx} f'(x) dx \tag{1}$$

따라서, 다음 식이 성립한다.

$$\lim_{t \rightarrow \infty} |\phi_x(t)| = \lim_{t \rightarrow \infty} \frac{1}{t} \left| \int_{-\infty}^{\infty} e^{itx} f'(x) dx \right| = 0 \tag{2}$$

즉, 다음 식이 성립한다.

$$\phi_x(t) = o(t) \text{ as } |t| \rightarrow \infty \tag{3}$$

같은 방법으로 자연수  $n$ 에 대해서 두 번째 성질을 증명할 수 있다. ■

다음 성질은 특성함수의 합성곱(convolution)에 관한 것이다.

**명제 A.4: 합성곱정리**

서로 독립이고 동일한 확률분포를 따르는 확률변수들  $x_1, x_2, \dots, x_n$ 의 공통적 특성함

수를  $\phi_x(t)$ 라고 하면, 실수들  $c_1, c_2, \dots, c_n$ 에 대해서 다음 식이 성립한다.

$$\phi_{\sum_{k=1}^n c_k x_k}(t) = \prod_{k=1}^n \phi_x(c_k t)$$

증명. 다음 식들이 성립한다.

$$\begin{aligned} \phi_{\sum_{k=1}^n c_k x_k}(t) &= E \left( \exp \left( i \left[ \sum_{k=1}^n c_k x_k \right] t \right) \right) \\ &= \prod_{k=1}^n E \left( \exp (i [c_k x_k] t) \right) = \prod_{k=1}^n \phi_x(c_k t) \end{aligned} \quad (1)$$

여기서 마지막 등호는 명제 A.1의 성질 (e)에 의해서 성립한다. ■

#### 따름정리 A.1

서로 독립인 확률변수들  $x$ 와  $y$ 의 확률분포함수들을  $F_x(\cdot)$ 와  $F_y(\cdot)$ 라고 하고 특성함수들을  $\phi_x(\cdot)$ 와  $\phi_y(\cdot)$ 라고 하면, 확률변수  $z = x + y$ 의 확률분포함수  $F_z(\cdot)$ 와 특성함수  $\phi_z(\cdot)$ 는 각각 다음과 같다.

$$F_z(z) = \int_{-\infty}^{\infty} F_x(z - y) dF_y(y), \quad \phi_z(t) = \phi_x(t) \phi_y(t)$$

**예제 A.4** 금융공학에서 확률변수의 무한분해가능 (infinitely divisible)은 중요한 역할을 한다. 확률변수  $x$ 의 특성함수  $\phi_x(\cdot)$ 가 각 자연수  $n$ 에 대해서 식  $\phi_x(t) = \phi_n(t)^n$ 을 만족하는  $\phi_n(\cdot)$ 이 존재하면, 이 확률변수는 무한분해가능하다고 한다. 무한분해가능성의 의미를 따름 정리 A.1에서 찾을 수 있다.

무한분해가능한 확률분포의 특성함수를 다음과 같이 표현할 수 있다.

$$\phi_x(t) = \exp \left( iat - \frac{1}{2} \sigma^2 t^2 + \int_{\mathbb{R} \setminus \{0\}} [e^{itx} - 1 - itx 1_{\{|x| < 1\}}] G(dx) \right) \quad (1)$$

여기서  $a \in \mathbb{R}$ 와  $\sigma \geq 0$ 는 상수들이고,  $G$ 는 다음 식을 만족하는  $\sigma$ -유한측도 (sigma-finite measure)이다.

$$\int_{\mathbb{R} \setminus \{0\}} \min \{x^2, 1\} G(dx) < \infty \quad (2)$$

이러한  $G$ 를 Lévy 측도라고 부른다. 또한, 식 (1)을 Lévy-Khintchine 식이라 부른다.

본저자는 식 (1)을 구하는 것은 그리 어려운 문제가 아니라고 생각한다. 다음 식을 이용해보라.

$$\lim_{h \rightarrow 0} \frac{\phi_{x+h}(t) - \phi_x(t)}{h} = \phi_x(t) \lim_{h \rightarrow 0} \frac{[\phi_h(t) - 1]}{h} \tag{3}$$



다음 명제는 특성함수의 일의성 (uniqueness)에 관한 것이다. 이 명제의 자세한 증명은 Dudley [33, pp. 303-304]를 참조하라.

**명제 A.5: 일의성정리**

확률변수들  $x$ 와  $y$ 의 확률분포함수들을  $F_x(\cdot)$ 과  $F_y(\cdot)$ 라 하고 특성함수들을  $\phi_x(\cdot)$ 과  $\phi_y(\cdot)$ 라 하자. 확률분포함수들  $F_x(\cdot)$ 과  $F_y(\cdot)$ 가 동일하기 위한 필요충분조건은 특성함수들  $\phi_x(\cdot)$ 과  $\phi_y(\cdot)$ 가 동일한 것이다. 즉, 다음 명제가 성립한다.

$$x \stackrel{d}{\sim} y \Leftrightarrow \phi_x(t) = \phi_y(t), (t \in \mathbb{R})$$

다음 명제를 Lévy의 연속성정리 (continuity theorem) 또는 수렴성정리 (convergence theorem)라 한다. 명제에 대한 증명은 Williams [97, pp. 185-187]를 참조하라.

**명제 A.6: Lévy의 연속성정리**

확률변수  $x_n$ 의 확률분포함수를  $F_n(\cdot)$ 이라하고 특성함수를  $\phi_n(\cdot)$ 라 하자. 만약 각 점  $t(\in \mathbb{R})$ 에서 극한함수  $\phi(t) \doteq \lim_{n \rightarrow \infty} \phi_n(t)$ 가 정의되고 함수  $\phi(t)$ 는 점  $t = 0$ 에서 연속이라고 하면, 함수  $\phi(t)$ 는 어떤 확률분포함수  $F(\cdot)$ 의 특성함수이며 또한 다음 식이 성립한다.

$$F_n \xrightarrow{d} F$$

적률모함수를 이용하는 연속성정리를 유도하기 위해서, 우선 다음과 같은 복소해석 (complex analysis)의 중요한 정리를 먼저 살펴보자.

**보조정리 A.1: Vitali-Porter 정리**

주어진 집합  $\Omega$ 의 부분집합  $E$ 를 살펴보자. 부분집합  $E$ 는 집합  $\Omega$ 에 속하는 집적점 (accumulation point)을 포함하고 있다고 하자. 정의역  $\Omega$ 에서 국소적으로 (locally) 유계인 (bounded) 해석적 (analytic) 함수열  $\{f_n\}$ 을 생각해보자. 만약 각 점  $z \in E$ 에서  $\lim_{n \rightarrow \infty} f_n(z)$ 이 존재하면, 집합  $\Omega$ 의 임의의 콤팩트한 (compact) 부분집합 상에서 함수열  $\{f_n\}$ 이 어떤 해석적 함수로 일양수렴한다.

보조정리 A.1을 명제 A.6에 적용하면, 다음 명제가 성립함을 알 수 있다.

**명제 A.7: 적률모함수의 연속성정리**

확률변수  $x_n$ 의 확률분포함수를  $F_n(\cdot)$ 이라하고 적률모함수를  $M_n(\cdot)$ 라 하자. 만약 어떤 양수  $\epsilon$ 에 대해서 각 점  $t \in (-\epsilon, \epsilon)$ 에서 극한함수  $M(t) \doteq \lim_{n \rightarrow \infty} M_n(t)$ 가 정의되고 함수  $M(t)$ 는 점  $t = 0$ 에서 연속이라고 하면, 함수  $M(t)$ 는 어떤 확률분포함수  $F(\cdot)$ 의 특성함수이며 또한 다음 식이 성립한다.

$$F_n \xrightarrow{d} F$$

정의 A.1에서 확률분포함수  $F_x(\cdot)$ 로부터 특성함수  $\phi_x(\cdot)$ 를 정의했다. 이 소절에서는 주어진 특성함수  $\phi_x(\cdot)$ 로부터 확률분포함수  $F_x(\cdot)$ 를 유도하기로 하자.

**예제 A.5** 확률변수  $x$ 의 확률분포함수를  $F_x(\cdot)$  그리고 특성함수  $\phi_x(\cdot)$ 라 하자. 또한, 표준 정규확률변수  $y$ 가 확률변수  $x$ 와 독립이라고 가정하자. 명제 A.4에서 알 수 있듯이, 양수  $\alpha$ 에 대해서 확률변수  $z_\alpha \doteq x + \alpha y$ 의 특성함수는 다음과 같다.

$$\phi_{z_\alpha}(t) = \phi_{x+\alpha y}(t) = \phi_x(t) \exp\left(-\frac{\alpha^2 t^2}{2}\right) \quad (1)$$

확률변수  $x$  대신 확률변수  $z_\alpha$ 를 사용하는 이유는 임의의  $\alpha$ 에 대해서  $z_\alpha$ 가 매우 평활한 확률변수이기 때문이다. 또한,  $\alpha$ 가 작은 양수이면 확률변수  $z_\alpha$ 의 확률분포는 확률변수  $x$ 의 확률분포와 매우 비슷하다. ■

**예제 A.6** 연속형 확률변수  $x$ 의 확률밀도함수를  $f_x(\cdot)$  그리고 특성함수  $\phi_x(\cdot)$ 라 하자. 또한, 표준정규확률변수  $w$ 가 확률변수  $x$ 와 독립이라고 가정하자. 양수  $\alpha$ 에 대해서 확률변수  $z_\alpha \doteq x + \alpha w$ 의 확률밀도함수  $f_{z_\alpha}(\cdot)$ 를 구해보자.

임의의 확률변수  $y$ 에 대해서 다음 식들이 성립한다.

$$\begin{aligned} E(\phi_y(x - \xi)) &= \int_{-\infty}^{\infty} \phi_y(x - \xi) f_x(x) dx \\ &= \int_{-\infty}^{\infty} E(\exp(i[x - \xi]y)) f_x(x) dx \\ &= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} \exp(i[x - \xi]y) f_y(y) dy \right] f_x(x) dx \\ &= \int_{-\infty}^{\infty} \exp(-i\xi y) \left[ \int_{-\infty}^{\infty} \exp(ixy) f_x(x) dx \right] f_y(y) dy \end{aligned} \tag{1}$$

즉, 다음 식들이 성립한다.

$$E(\phi_y(x - \xi)) = \int_{-\infty}^{\infty} [\exp(-i\xi y) \phi_x(y)] f_y(y) dy = E(e^{-i\xi y} \phi_x(y)) \tag{2}$$

다음 확률변수  $y$ 의 확률분포가  $\mathcal{N}(0, \alpha^{-2})$ 라고 하면, 확률밀도함수  $f_y(y)$ 와 특성함수  $\phi_y(t)$ 는 각각 다음과 같다.

$$f_y(y) = \frac{\alpha}{\sqrt{2\pi}} \exp\left(-\frac{\alpha^2 y^2}{2}\right), \quad \phi_y(t) = \exp\left(-\frac{t^2}{2\alpha^2}\right) \tag{3}$$

식 (3)을 식 (2)에 대입하면, 다음 식이 성립함을 알 수 있다.

$$\int_{-\infty}^{\infty} e^{-i\xi y} \phi_x(y) \frac{\alpha}{\sqrt{2\pi}} \exp\left(-\frac{\alpha^2 y^2}{2}\right) dy = \int_{-\infty}^{\infty} \exp\left(-\frac{[x - \xi]^2}{2\alpha^2}\right) f_x(x) dx \tag{4}$$

즉, 다음 식이 성립한다.

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\xi y} \phi_x(y) \exp\left(-\frac{\alpha^2 y^2}{2}\right) dy = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\alpha} \exp\left(-\frac{[\xi - x]^2}{2\alpha^2}\right) f_x(x) dx \tag{5}$$

식 (5)의 우변은 확률변수  $\alpha w$ 의 확률밀도함수와 확률변수  $x$ 의 확률밀도함수  $f_x(x)$ 의 합성곱이다. 즉, 확률변수  $z_\alpha = x + \alpha w$ 의 확률밀도함수  $f_{z_\alpha}(\cdot)$ 이다. 따라서, 다음 식이 성립한다.

$$f_{z_\alpha}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-izt} \phi_x(t) \exp\left(-\frac{\alpha^2 t^2}{2}\right) dt \tag{6}$$



**예제 A.7** 이 예제에서는 복소해석학을 적용한다. 따라서 복소해석학을 모르는 독자는 그냥 넘어가기를 권한다.

이 예제에서는 다음 식을 증명하기로 하자.

$$\int_0^{\infty} \frac{\sin x}{x} dx = \frac{\pi}{2} \quad (1)$$

다음 함수를 정의하자.

$$f(z) = \frac{e^{iz}}{z} \quad (2)$$

그림 A.3의 적분경로를  $C$ 로 표기하자. Cauchy 정리에서 알 수 있듯이, 다음 식들이 성립한다.

$$\oint_C f(z) dz = \oint_C \frac{e^{iz}}{z} dz = 0 \quad (3)$$

따라서 다음 식이 성립한다.

$$\int_{C_1} \frac{e^{iz}}{z} dz + \int_{C_R} \frac{e^{iz}}{z} dz + \int_{C_2} \frac{e^{iz}}{z} dz + \int_{C_\epsilon} \frac{e^{iz}}{z} dz = 0 \quad (4)$$

다음 식들이 성립한다.

$$\begin{aligned} \int_{C_1} \frac{e^{iz}}{z} dz + \int_{C_2} \frac{e^{iz}}{z} dz &= \int_{\epsilon}^R \frac{e^{ix}}{x} dx + \int_{-R}^{-\epsilon} \frac{e^{ix}}{x} dx \\ &= \int_{\epsilon}^R \frac{e^{ix}}{x} dx - \int_{\epsilon}^R \frac{e^{-iy}}{y} dy = 2i \int_{\epsilon}^R \frac{\sin x}{x} dx \end{aligned} \quad (5)$$

다음 식들이 성립한다.

$$\begin{aligned} \left| \int_{C_R} \frac{e^{iz}}{z} dz \right| &= \left| \int_0^{\pi} \frac{e^{iRe^{i\theta}}}{Re^{i\theta}} Re^{i\theta} i d\theta \right| \leq \int_0^{\pi} \left| e^{iRe^{i\theta}} \right| d\theta \\ &= \int_0^{\pi} \left| e^{-R \sin \theta} e^{iR \cos \theta} \right| d\theta = \int_0^{\pi} e^{-R \sin \theta} d\theta \leq \int_0^{\pi} \exp\left(-R \frac{2\theta}{\pi}\right) d\theta \end{aligned} \quad (6)$$

다음 식들이 성립한다.

$$\begin{aligned} \int_0^\pi e^{-R\sin\theta} d\theta &= \int_0^{\pi/2} e^{-R\sin\theta} d\theta + \int_{\pi/2}^\pi e^{-R\sin\theta} d\theta \\ &\leq \int_0^{\pi/2} \exp\left(-2R\frac{\theta}{\pi}\right) d\theta + \int_{\pi/2}^\pi \exp\left(-2R\left[1 - \frac{\theta}{\pi}\right]\right) d\theta \end{aligned} \quad (7)$$

여기서 마지막 부등호는 다음 식들에 의해서 성립한다.

$$\sin\theta \geq \frac{2\theta}{\pi}, \quad \left(0 \leq \theta \leq \frac{\pi}{2}\right) \quad \text{and} \quad \sin\theta \geq 2\left[1 - \frac{\theta}{\pi}\right], \quad \left(\frac{\pi}{2} \leq \theta \leq \pi\right) \quad (8)$$

식 (7)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\left| \int_{C_R} \frac{e^{iz}}{z} dz \right| \leq \frac{\pi}{2R} [1 - e^{-R}] + \frac{\pi}{2R} [1 - e^{-R}] < \frac{\pi}{R} \quad (9)$$

식 (9)에서 알 수 있듯이, 다음 식이 성립한다.

$$\lim_{R \rightarrow \infty} \int_{C_R} \frac{e^{iz}}{z} dz = 0 \quad (10)$$

Taylor전개를 적용하면, 각  $\epsilon (\in (0, 0.5))$ 에 대해서 다음 식이 성립함을 알 수 있다.

$$\left| e^{i\epsilon e^{i\theta}} - 1 \right| < 2\epsilon \quad (11)$$

다음 식들이 성립한다.

$$\int_{C_\epsilon} \frac{e^{iz}}{z} dz = \int_0^\pi \frac{e^{i\epsilon e^{i\theta}}}{\epsilon e^{i\theta}} \epsilon e^{i\theta} i d\theta = i \int_0^\pi e^{i\epsilon e^{i\theta}} d\theta = i \int_0^\pi [e^{i\epsilon e^{i\theta}} - 1] d\theta + \pi i \quad (12)$$

식 (11)과 식 (12)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\left| \int_{C_\epsilon} \frac{e^{iz}}{z} dz - \pi i \right| \leq \left| i \int_0^\pi [e^{i\epsilon e^{i\theta}} - 1] d\theta \right| \leq \int_0^\pi 2\epsilon d\theta = 2\pi\epsilon \quad (13)$$

즉, 다음 식이 성립한다.

$$\lim_{\epsilon \rightarrow +0} \int_{C_\epsilon} \frac{e^{iz}}{z} dz = \pi i \quad (14)$$

식 (5), 식 (10) 그리고 식 (14)를 식 (4)에 대입하면, 다음 식이 성립함을 알 수 있다.

$$\lim_{\epsilon \rightarrow +0} \lim_{R \rightarrow \infty} \int_{\epsilon}^R \frac{\sin x}{x} dx = \frac{\pi}{2} \quad (15)$$

즉, 식 (1)이 성립한다. ■

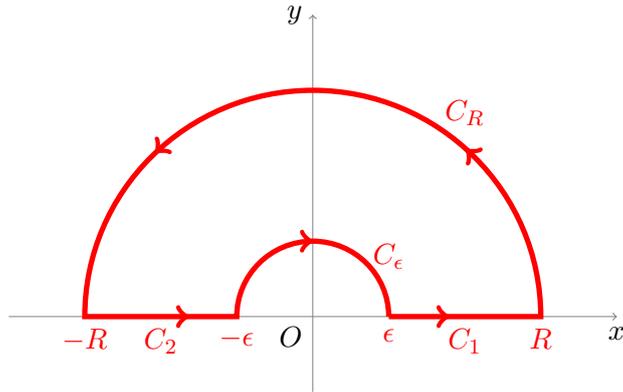


그림 A.3. 적분경로

예제 A.7에서 알 수 있듯이, 다음 보조정리가 성립한다.

**보조정리 A.2**

실수  $\alpha$ 에 대해서 다음 식이 성립한다.

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin(\alpha x)}{x} dx = \text{sgn}(\alpha)$$

여기서  $\text{sgn}(\alpha)$ 는 다음과 같다.

$$\text{sgn}(\alpha) \doteq \begin{cases} 1, & (\alpha > 0) \\ 0, & (\alpha = 0) \\ -1, & (\alpha < 0) \end{cases}$$

**명제 A.8: 특성함수의 역함수**

확률분포함수  $F_x(\cdot)$ 의 특성함수를  $\phi_x(\cdot)$ 로 표기하자. 만약  $F_x(\cdot)$ 가 점  $x$ 에서 연속이면, 다음 식이 성립한다.

$$F_x(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\Im \{e^{-itx} \phi_x(t)\}}{t} dt$$

여기서  $\Im \{z\}$ 는  $z$ 의 허수부이다.

증명. 다음 식이 성립한다.

$$\Im(z) = \frac{z - \bar{z}}{2i} \tag{1}$$

식 (1)에서 알 수 있듯이, 다음 식이 성립한다.

$$\Im \{e^{-itx} \phi_x(t)\} = \frac{e^{-itx} \phi_x(t) - \overline{e^{-itx} \phi_x(t)}}{2i} \tag{2}$$

명제 A.1의 성질 (c)를 식 (2)에 적용하면, 다음 식을 얻는다.

$$\Im \{e^{-itx} \phi_x(t)\} = \frac{e^{-itx} \phi_x(t) - e^{itx} \phi_x(-t)}{2i} \tag{3}$$

다음 식들이 성립한다.

$$e^{-itx} \phi_x(t) = e^{-itx} \int_{-\infty}^\infty e^{itu} dF_x(u) = \int_{-\infty}^\infty e^{it[u-x]} dF_x(u) \tag{4}$$

$$e^{itx} \phi_x(-t) = e^{itx} \int_{-\infty}^\infty e^{-itu} dF_x(u) = \int_{-\infty}^\infty e^{-it[u-x]} dF_x(u) \tag{5}$$

식 (4)와 식 (5)를 식 (3)에 대입하면, 다음 식이 성립한다.

$$\Im \{e^{-itx} \phi_x(t)\} = \int_{-\infty}^\infty \sin(t[u-x]) dF_x(u) \tag{6}$$

식 (6)에서 알 수 있듯이, 다음 식이 성립한다.

$$\int_0^\infty \frac{\Im \{e^{-itx} \phi_x(t)\}}{t} dt = \int_0^\infty \int_{-\infty}^\infty \frac{\sin(t[u-x])}{t} dF_x(u) dt \tag{7}$$

다음 식들이 성립한다.

$$\begin{aligned} \int_{-\infty}^{\infty} \left[ \int_0^{\infty} \frac{\sin(t[u-x])}{t} dt \right] dF_x(u) &= \frac{1}{2} \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} \frac{\sin(t[u-x])}{t} dt \right] dF_x(u) \\ &= \frac{\pi}{2} \int_{-\infty}^{\infty} \operatorname{sgn}(u-x) dF_x(u) = \frac{\pi}{2} \left[ \int_{-\infty}^x [-1] dF_x(u) + \int_x^{\infty} dF_x(u) \right] \end{aligned} \quad (8)$$

식 (7) 과 식 (8) 에서 알 수 있듯이, 다음 식이 성립한다.

$$\int_0^{\infty} \frac{\Im \{ e^{-itx} \phi_x(t) \}}{t} dt = \frac{\pi}{2} [1 - 2F_x(x)] \quad (9)$$

식 (9) 에서 알 수 있듯이, 다음 식이 성립한다.

$$F_x(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{\Im \{ e^{-itx} \phi_x(t) \}}{t} dt \quad (10)$$

■

명제 A.8의 증명에서 알 수 있듯이, 확률분포함수  $F_x(\cdot)$ 와 특성함수  $\phi_x(\cdot)$ 는 다음 식들을 만족한다.

$$F_x(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{e^{-itx} \phi_x(t) - e^{itx} \phi_x(-t)}{2it} dt = \frac{1}{2} - \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} \phi_x(t)}{it} dt \quad (A1.2)$$

식 (A1.2)에서 알 수 있듯이, 다음 식이 성립한다.

$$F_x(0) = \frac{1}{2} - \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\phi_x(t)}{it} dt \quad (A1.3)$$

식 (A1.2)와 식 (A1.3)에서 알 수 있듯이, 다음 식이 성립한다.

$$F_x(x) = F_x(0) - \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - 1}{it} \phi_x(t) dt \quad (A1.4)$$

식 (A1.4)에서 알 수 있듯이, 다음 식이 성립한다.

$$F_x(b) - F_x(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \phi_x(t) dt \quad (A1.5)$$

점  $x$ 에서 확률분포함수  $F(\cdot)$ 가 미분가능이면, 식 (A1.5)에서 알 수 있듯이 확률밀도함수

$f_x(x)$ 는 다음 식들을 만족한다.

$$f_x(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_x(t) dt = \frac{1}{\pi} \int_0^{\infty} \Re \{ e^{-itx} \phi_x(t) \} dt \quad (\text{A1.6})$$

여기서  $\Re(z)$ 는  $z$ 의 실수부이다. 지금까지 내용을 정리하면 다음과 같다.

**따름정리 A.2**

확률분포함수  $F_x(\cdot)$ 와 그 특성함수  $\phi_x(\cdot)$ 에 대해서, 다음 식들이 성립한다.

(a) 점  $x$ 에서 확률분포함수  $F_x(\cdot)$ 가 연속이면, 다음 식이 성립한다.

$$F_x(x) = F_x(0) - \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - 1}{it} \phi_x(t) dt$$

(b) 점  $a$ 와 점  $b$ 에서 확률분포함수  $F_x(\cdot)$ 가 연속이면, 다음 식이 성립한다.

$$F_x(b) - F_x(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \phi_x(t) dt$$

(c) 점  $x$ 에서 확률분포함수  $F(\cdot)$ 가 미분가능이면, 확률밀도함수  $f_x(\cdot)$ 는 다음과 같다.

$$f_x(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_x(t) dt = \frac{1}{\pi} \int_0^{\infty} \Re \{ e^{-itx} \phi_x(t) \} dt$$

**예제 A.8** 표준정규확률변수의 확률밀도함수  $f_N(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ 의 특성함수  $\phi_N(\cdot)$ 는 다음과 같다.

$$\begin{aligned} \phi_N(t) &= \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \exp\left(-\frac{t^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{[x-it]^2}{2}\right) dx = \exp\left(-\frac{t^2}{2}\right) \end{aligned} \quad (1)$$

여기서 세 번째 등호는 예제 A.1의 식 (8)에 의해서 성립한다. 식 (1)에서 알 수 있듯이, 표준정규확률분포의 특성함수는  $\phi_N(t) = \exp\left(-\frac{t^2}{2}\right)$ 이다.

특성함수  $\phi_N(t) = \exp\left(-\frac{t^2}{2}\right)$ 가 주어졌을 때, 따름정리 A.2의 성질 (c)에서 알 수 있듯이

확률밀도함수  $f_N(\cdot)$ 는 다음과 같다.

$$\begin{aligned} f_N(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \exp\left(-\frac{t^2}{2}\right) dt \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{[x+it]^2}{2}\right) dx = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \end{aligned} \quad (2)$$

여기서 세 번째 등호는 예제 A.1의 식 (8)에 의해서 성립한다. 식 (2)에서 알 수 있듯이, 함수  $\phi_N(t) = \exp\left(-\frac{t^2}{2}\right)$ 가 표준정규확률밀도함수의 특성함수이다. ■

**예제 A.9** 다음 지수확률밀도함수를 살펴보자.

$$f_x(x) = \lambda e^{-\lambda x} 1_{[0, \infty)}(x) \quad (1)$$

여기서  $\lambda(> 0)$ 는 양수이다. 이 지수확률밀도함수의 특성함수  $\phi_x(\cdot)$ 는 다음과 같다.

$$\begin{aligned} \phi_x(t) &= \int_0^{\infty} e^{itx} \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^{\infty} e^{-\lambda x} \cos(tx) dx + i\lambda \int_0^{\infty} e^{-\lambda x} \sin(tx) dx \\ &= \lambda \int_0^{\infty} e^{-\lambda x} \cos(tx) dx + i\lambda \int_0^{\infty} e^{-\lambda x} \sin(tx) dx \end{aligned} \quad (2)$$

다음 식들이 성립한다.

$$\int_0^{\infty} e^{-\lambda x} \cos(tx) dx = \left[ \frac{e^{-\lambda x} [-\lambda \cos(tx) + t \sin(tx)]}{\lambda^2 + t^2} \right]_{x=0}^{x=\infty} = \frac{\lambda}{\lambda^2 + t^2} \quad (3)$$

$$\int_0^{\infty} e^{-\lambda x} \sin(tx) dx = \left[ \frac{e^{-\lambda x} [-\lambda \sin(tx) - t \cos(tx)]}{\lambda^2 + t^2} \right]_{x=0}^{x=\infty} = \frac{t}{\lambda^2 + t^2} \quad (4)$$

식 (3)과 식 (4)를 식 (2)에 대입하면, 다음 식을 얻는다.

$$\phi_x(t) = \frac{\lambda[\lambda + it]}{\lambda^2 + t^2} = \frac{-\lambda}{it - \lambda} \quad (5)$$

특성함수  $\phi_x(t) = \frac{-\lambda}{it - \lambda}$ 가 주어졌을 때, 따름정리 A.2의 성질 (c)에서 알 수 있듯이 확률밀도함수  $f_x(\cdot)$ 는 다음과 같다.

$$f_x(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{-\lambda}{it - \lambda} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{g(t)}{t + i\lambda} dt \quad (6)$$

여기서  $g(t) = i\lambda \exp(-itx)$ 이다. 그림 A.4의 적분경로를  $C$ 로 표기하자. Cauchy 정리에서 알 수 있듯이, 다음 식들이 성립한다.

$$\frac{1}{2\pi} \oint_C \frac{g(z)}{z + i\lambda} dz = \frac{1}{2\pi} 2\pi i g(-i\lambda) = -\lambda e^{-\lambda x} \quad (7)$$

따라서 다음 식이 성립한다.

$$\frac{1}{2\pi} \int_{C_R} \frac{g(z)}{z + i\lambda} dz + \frac{1}{2\pi} \int_R^{-R} \frac{g(z)}{z + i\lambda} dz = -\lambda e^{-\lambda x} \quad (8)$$

지금부터 다음 식을 증명하기로 하자.

$$\lim_{R \rightarrow \infty} \int_{C_R} \frac{g(z)}{z + i\lambda} dz = 0 \quad (9)$$

만약 식  $z = Re^{i\theta} \in C_R$ 이 성립하면, 다음 식들이 성립한다.

$$\left| i\lambda \exp(-ixRe^{i\theta}) \right| = \left| i\lambda \exp(xR \sin \theta) \exp(-ixR \cos \theta) \right| = \lambda \exp(xR \sin \theta) \quad (10)$$

만약 식  $R \geq 2\lambda$ 가 성립하면, 다음 식들이 성립한다.

$$\left| Re^{i\theta} + i\lambda \right| \geq \left| Re^{i\theta} \right| - |i\lambda| = R - \lambda \geq R - \frac{R}{2} = \frac{R}{2} \quad (11)$$

식 (11)에서 알 수 있듯이, 식  $R \geq 2\lambda$ 가 성립하면 다음 식들이 성립한다.

$$\left| \frac{Re^{i\theta}}{Re^{i\theta} + i\lambda} \right| \leq \frac{R}{R/2} = 2 \quad (12)$$

다음 식들이 성립한다.

$$\begin{aligned} \left| \int_{C_R} \frac{g(z)}{z + i\lambda} dz \right| &= \left| \int_{\pi \leq \theta \leq 2\pi} \frac{i\lambda \exp(-ixRe^{i\theta})}{Re^{i\theta} + i\lambda} iRe^{i\theta} d\theta \right| \\ &\leq \int_{\pi}^{2\pi} \left| i\lambda \exp(-ixRe^{i\theta}) \right| \left| \frac{iRe^{i\theta}}{Re^{i\theta} + i\lambda} \right| d\theta \\ &\leq 2\lambda \int_{\pi}^{3\pi/2} \exp(xR \sin \theta) d\theta + 2\lambda \int_{3\pi/2}^{2\pi} \exp(xR \sin \theta) d\theta \end{aligned} \quad (13)$$

여기서 두 번째 부등호는 식 (10)과 식 (12)에 의해서 성립한다. 다음 부등식이 성립함을 쉽게

알 수 있다.

$$\sin \theta \leq \frac{2}{\pi} \left| \theta - \frac{3\pi}{2} \right| - 1, \quad (\pi \leq \theta \leq 2\pi) \quad (14)$$

식 (14)에서 알 수 있듯이, 다음 부등식들이 성립한다.

$$\int_{\pi}^{3\pi/2} \exp(xR \sin \theta) d\theta \leq \int_{\pi}^{3\pi/2} \exp\left(xR \left[2 - \frac{2}{\pi}\theta\right]\right) d\theta \quad (15)$$

$$\int_{3\pi/2}^{2\pi} \exp(xR \sin \theta) d\theta \leq \int_{3\pi/2}^{2\pi} \exp\left(xR \left[\frac{2}{\pi}\theta - 4\right]\right) d\theta \quad (16)$$

다음 식들이 성립한다.

$$\int_{\pi}^{3\pi/2} \exp\left(xR \left[2 - \frac{2}{\pi}\theta\right]\right) d\theta = \frac{\pi}{-2xR} \exp\left(xR \left[2 - \frac{2}{\pi}\theta\right]\right) = \frac{\pi}{2xR} [1 - e^{-xR}] \quad (17)$$

$$\int_{3\pi/2}^{2\pi} \exp\left(xR \left[\frac{2}{\pi}\theta - 4\right]\right) d\theta = \frac{\pi}{2xR} \exp\left(xR \left[\frac{2}{\pi}\theta - 4\right]\right) = \frac{\pi}{2xR} [1 - e^{-xR}] \quad (18)$$

식 (15)~식 (18)에서 알 수 있듯이, 다음 식이 성립한다.

$$\lim_{R \rightarrow \infty} \left[ 2\lambda \int_{\pi}^{3\pi/2} \exp(xR \sin \theta) d\theta + 2\lambda \int_{3\pi/2}^{2\pi} \exp(xR \sin \theta) d\theta \right] = 0 \quad (19)$$

식 (13)과 식 (19)에서 알 수 있듯이, 식 (9)가 성립한다.

식 (8)과 식 (9)를 식 (6)에 대입하면, 다음 식들이 성립함을 알 수 있다.

$$f_x(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{g(z)}{z + i\lambda} dz = \lambda e^{-\lambda x} \quad (20)$$

식 (20)에서 알 수 있듯이, 특성함수  $\phi_x(t) = \frac{-\lambda}{it - \lambda}$  는 모수가  $\lambda$ 인 지수확률변수이다. ■

**예제 A.10** 형태모수 (shape parameter) 가  $\alpha (> 0)$  이고 비율모수 (rate parameter) 가  $\beta (> 0)$  인 감마확률분포의 확률밀도함수는 다음과 같다.

$$f_x(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} 1_{[0, \infty)}(x) \quad (1)$$

이 예제에서는 형태모수가  $n$  이고 비율모수가 1인 감마확률분포를 살펴보자. 이 감마확률분

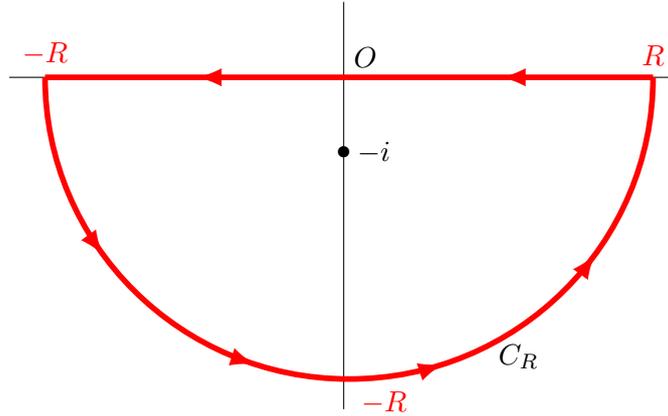


그림 A.4. 적분경로 C

포의 확률밀도함수는 다음과 같다.

$$f_x(x) = \frac{1}{\Gamma(n)} x^{n-1} e^{-x} 1_{[0, \infty)}(x) \quad (2)$$

이 지수확률밀도함수의 특성함수  $\phi_x(\cdot)$ 는 다음과 같다.

$$\begin{aligned} \phi_x(t) &= \int_0^\infty e^{itx} \frac{1}{\Gamma(n)} x^{n-1} e^{-x} dx = \int_0^\infty e^{itx} \sum_{k=0}^\infty \frac{1}{k!} [itx]^k \frac{1}{\Gamma(n)} x^{n-1} e^{-x} dx \\ &= \frac{1}{\Gamma(n)} \sum_{k=0}^\infty \frac{1}{k!} [it]^k \int_0^\infty x^{k+n-1} e^{-x} dx = \sum_{k=0}^\infty \frac{\Gamma(k+n)}{k! \Gamma(n)} [it]^k \\ &= \sum_{k=0}^\infty \binom{k+n-1}{k} [it]^k = \frac{1}{[1-it]^n} \end{aligned} \quad (3)$$

여기서 다섯 번째 등호는 다음 식에 의해서 성립한다.

$$\frac{1}{[1-x]^r} = \sum_{k=0}^\infty \binom{k+r-1}{k} x^k, \quad (|x| < 1) \quad (4)$$

여기서  $r$ 은 양수이다. 주의할 점은 식 (3)에서 여섯 번째 등호는  $t$ 가 구간  $(-1, 1)$ 에서 성립한다는 것이다. 그러나, 부분적분을 사용해서 이 여섯 번째 등호가 모든  $t (\in \mathbb{R})$ 에 대해서 성립함을 보일 수 있다. 식 (3)에서 알 수 있듯이, 감마확률변수의 특성함수는 다음과 같다.

$$\phi_x(t) = \frac{i^n}{[t+i]^n} \quad (5)$$

특성함수  $\phi_x(t) = i^n [t+i]^{-n}$ 가 주어졌을 때, 따름정리 A.2의 성질 (c)에서 알 수 있듯이

확률밀도함수  $f_x(\cdot)$ 는 다음과 같다.

$$f_x(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{i^n}{[t+i]^n} dt \quad (6)$$

그림 A.4의 적분경로  $C$ 에서, 다음 식이 성립한다.

$$\frac{1}{2\pi} \oint_C e^{-izx} \frac{i^n}{[z+i]^n} dz = \frac{i^n}{2\pi} \int_R^{-R} \frac{e^{-izx}}{[z+i]^n} dz + \frac{i^n}{2\pi} \oint_{C_R} \frac{e^{-izx}}{[z+i]^n} dz \quad (7)$$

일반화 Cauchy 적분 공식 (generalized Cauchy integral formula)에서 알 수 있듯이, 다음 식들이 성립한다.

$$\frac{i^n}{2\pi} \oint_C \frac{e^{-izx}}{[z+i]^n} = \frac{i^n}{2\pi} \frac{2\pi i}{\Gamma(n)} \frac{d^{n-1} e^{-izx}}{d^{n-1} z} \Big|_{z=-i} = \frac{i^{n+1}}{\Gamma(n)} [-ix]^{n-1} e^{-x} = -\frac{1}{\Gamma(n)} x^{n-1} e^{-x} \quad (8)$$

식 (7) 과 식 (8)에서 알 수 있듯이, 다음 식이 성립한다.

$$\frac{1}{2\pi} \oint_{C_R} \frac{e^{-izx}}{[z+i]^n} dz + \frac{1}{2\pi} \int_R^{-R} \frac{e^{-izx}}{[z+i]^n} dz = -\frac{1}{\Gamma(n)} x^{n-1} e^{-x} \quad (9)$$

즉, 다음 식이 성립한다.

$$\frac{1}{2\pi} \lim_{R \rightarrow \infty} \oint_{C_R} \frac{e^{-izx}}{[z+i]^n} dz + \frac{1}{2\pi} \int_{\infty}^{-\infty} \frac{e^{-izx}}{[z+i]^n} dz = -\frac{1}{\Gamma(n)} x^{n-1} e^{-x} \quad (10)$$

예제 A.9에서 사용한 방법을 사용해서, 다음 식이 성립함을 증명할 수 있다.

$$\lim_{R \rightarrow \infty} \oint_{C_R} \frac{e^{-izx}}{[z+i]^n} dz = 0 \quad (11)$$

식 (7), 식 (10) 과 식 (11)을 식 (6)에 대입하면, 다음 식을 얻는다.

$$f_x(x) = \frac{1}{\Gamma(n)} x^{n-1} e^{-x} \quad (12)$$

■

**예제 A.11** 다음 이중지수확률밀도함수 (double exponential probability density function)

를 살펴보자.

$$f_x(x) = \frac{1}{2}e^{-|x|}, \quad (x \in \mathbb{R}) \quad (1)$$

이 지수확률밀도함수의 특성함수  $\phi_x(\cdot)$ 는 다음과 같다.

$$\begin{aligned} \phi_x(t) &= \int_{-\infty}^{\infty} e^{itx} \frac{1}{2}e^{-|x|} dx = \frac{1}{2} \int_0^{\infty} e^{itx} \frac{1}{2}e^{-x} dx + \frac{1}{2} \int_{-\infty}^0 e^{itx} \frac{1}{2}e^x dx \\ &= \frac{1}{2} \int_0^{\infty} e^{[it-1]x} dx + \frac{1}{2} \int_{-\infty}^0 e^{[it+1]x} dx = \frac{1}{2} \left[ \frac{1}{1-it} + \frac{1}{1+it} \right] \end{aligned} \quad (2)$$

식 (2)에서 알 수 있듯이, 다음 식이 성립한다.

$$\phi_x(t) = \frac{1}{1+t^2} \quad (3)$$

특성함수  $\phi_x(t) = [1+t^2]^{-1}$ 이 주어졌을 때, 따름정리 A.2의 성질 (c)에서 알 수 있듯이 확률밀도함수  $f_x(\cdot)$ 는 다음과 같다.

$$f_x(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{1}{1+t^2} dt = \frac{1}{4\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{1}{1-it} dt + \frac{1}{4\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{1}{1+it} dt \quad (4)$$

예제 A.9와 같은 방법을 사용해서, 다음 식을 증명할 수 있다.

$$\frac{1}{4\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{1}{1-it} dt = \frac{1}{2} e^{-x} 1_{[0,\infty)}(x) \quad (5)$$

그림 A.5의 적분경로를  $C$ 로 표기하자. Cauchy정리에서 알 수 있듯이, 다음 식들이 성립한다.

$$\begin{aligned} \frac{1}{4\pi} \oint_C e^{-izx} \frac{1}{1+iz} dz &= \frac{1}{4\pi i} \oint_C e^{-izx} \frac{1}{z-i} dz \\ &= \frac{1}{4\pi i} 2\pi i e^{-i^2 x} 1_{(-\infty,0)}(x) = \frac{1}{2} e^x 1_{(-\infty,0)}(x) \end{aligned} \quad (6)$$

식 (6)을 다음과 같이 쓸 수 있다.

$$\frac{1}{4\pi} \int_{C_R} e^{-izx} \frac{1}{1+iz} dz + \frac{1}{4\pi} \int_{-R}^R e^{-izx} \frac{1}{1+iz} dz = \frac{1}{2} e^x 1_{(-\infty,0)}(x) \quad (7)$$

예제 A.9에서와 동일한 방법을 사용해서, 다음 식을 증명할 수 있다.

$$\lim_{R \rightarrow \infty} \int_{C_R} e^{-izx} \frac{1}{1+iz} dz = 0 \quad (8)$$

식 (8)을 식 (7)에 대입하면, 다음 식을 얻는다.

$$\frac{1}{4\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{1}{1+it} dt = \frac{1}{2} e^x 1_{(-\infty, 0)}(x) \quad (9)$$

식 (5)와 식 (9)를 식 (4)에 대입하면, 다음 식을 얻는다.

$$f_x(x) = \frac{1}{2} e^{-|x|} \quad (10)$$

식 (10)에서 알 수 있듯이, 특성함수  $\phi_x(t) = [1+t^2]^{-1}$ 는 이중지수확률변수이다. ■

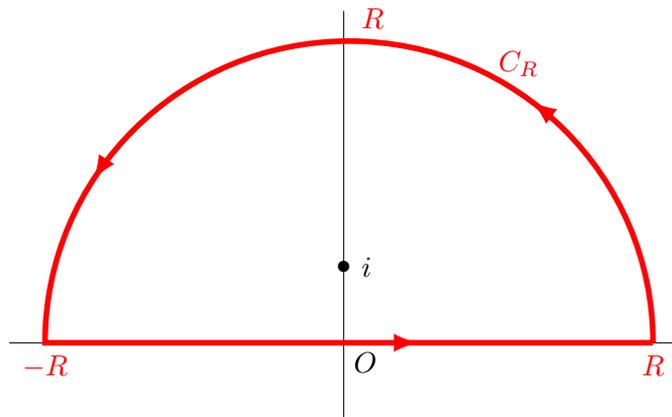


그림 A.5. 적분경로 C

**예제 A.12** 연속인 확률분포함수에 대해 명제 A.5를 증명하기로 하자.

연속형 확률변수들  $x$ 와  $y$ 의 확률분포함수들을  $F_x(\cdot)$ 과  $F_y(\cdot)$ 라 하고 특성함수들을  $\phi_x(\cdot)$ 과  $\phi_y(\cdot)$ 라 하자. 만약 항등식  $\phi_x(t) \equiv \phi_y(t)$ 이 성립하면, 따름정리 A.2의 성질 (b)에서 알 수 있듯이 다음 식이 성립한다.

$$F_x(b) - F_x(a) = F_y(b) - F_y(a) \quad (1)$$

식 (1)에 식  $a = -\infty$ 를 대입하면, 다음 식을 얻는다.

$$F_x(b) = F_y(b) \tag{2}$$

임의의  $b(\in \mathbb{R})$ 에 대해서 식 (2)가 성립하므로, 식  $F_x \equiv F_y$ 가 성립한다. ■

**예제 A.13** 연속인 확률분포함수에 대해 명제 A.6에서 언급한 Lévy의 연속성정리를 증명하기로 하자.

연속형 확률변수열  $\{x_n\}$ 과 연속형 확률변수  $x$ 를 살펴보자. 확률변수들  $x_n$ 과  $x$ 의 확률분포함수들을  $F_{x_n}(\cdot)$ 과  $F_x(\cdot)$ 라 하고 특성함수들을  $\phi_{x_n}(\cdot)$ 과  $\phi_x(\cdot)$ 라 하자. 또한, 표준정규확률변수  $y$ 가 확률변수열  $\{x_n\}$ 과 확률변수  $x$ 와 독립이라고 가정하자. 양수  $\alpha$ 에 대해서 확률변수들  $Z_n \doteq x_n + \alpha y$ 와  $Z \doteq x + \alpha y$ 의 특성함수들은 각각 다음과 같다.

$$\phi_{Z_n}(t) = \phi_{x_n + \alpha y}(t) = \phi_{x_n}(t) \exp\left(-\frac{\alpha^2 t^2}{2}\right) \tag{1}$$

$$\phi_Z(t) = \phi_{x + \alpha y}(t) = \phi_x(t) \exp\left(-\frac{\alpha^2 t^2}{2}\right) \tag{2}$$

식 (1), 식 (2)와 따름정리 A.2의 성질 (b)에서 알 수 있듯이, 임의의 실수들  $a$ 와  $b(> a)$ 에 대해서 다음 식들이 성립한다.

$$\Pr(a \leq x_n + \alpha y \leq b) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \phi_{x_n}(t) \exp\left(-\frac{\alpha^2 t^2}{2}\right) dt \tag{3}$$

$$\Pr(a \leq x + \alpha y \leq b) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \phi_x(t) \exp\left(-\frac{\alpha^2 t^2}{2}\right) dt \tag{4}$$

다음 식이 성립한다고 가정하자.

$$\lim_{n \rightarrow \infty} \phi_{x_n}(t) = \phi_x(t), \quad (t \in \mathbb{R}) \tag{5}$$

식 (3)과 식 (4)를 식 (5)에 대입하면, 다음 식을 얻는다.

$$\lim_{n \rightarrow \infty} \Pr(a \leq x_n + \alpha y \leq b) = \Pr(a \leq x + \alpha y \leq b) \tag{6}$$

식 (6)에 식  $b = \infty$ 를 대입하면, 다음 식을 얻는다.

$$\lim_{n \rightarrow \infty} \Pr(a \leq x_n + \alpha y) = \Pr(a \leq x + \alpha y) \quad (7)$$

식 (7)에 극한  $\alpha \rightarrow 0$ 을 적용하면, 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} F_{x_n}(a) = F_x(a) \quad (8)$$

■

**예제 A.14** 이 예제는 Lebesgue 측도를 사용해서 명제 A.8을 좀 더 자세하게 증명하고자 한다. 즉, 다음 명제를 증명하기로 하자.

확률변수  $x$ 의 특성함수를  $\phi_x(\cdot)$ 라고 하고, 다음 함수를 정의하자.

$$f_x(x) \doteq \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_x(t) dt \quad (1)$$

만약 특성함수  $\phi_x(\cdot)$ 가 실수집합  $\mathbb{R}$ 에서 적분가능하면, 함수  $f_x(x)$ 가 확률변수  $x$ 의 확률밀도함수이며 실수집합  $\mathbb{R}$ 에서 유계이고 연속이다.

첫째, 함수  $f_x(\cdot)$ 가 존재함을 증명하자. 다음 식들이 성립한다.

$$\left| \int_{-\infty}^{\infty} e^{-itx} \phi_x(t) dt \right| \leq \int_{-\infty}^{\infty} |e^{-itx} \phi_x(t)| dt = \int_{-\infty}^{\infty} |\phi_x(t)| dt < \infty \quad (2)$$

따라서  $f_x(\cdot)$ 가 존재한다.

둘째, 함수  $f_x(\cdot)$ 가 유계임을 증명하자. 다음 식들이 성립한다.

$$2\pi |f_x(x)| = \left| \int_{-\infty}^{\infty} e^{-itx} \phi_x(t) dt \right| \leq \int_{-\infty}^{\infty} |\phi_x(t)| dt < \infty \quad (3)$$

따라서 함수  $f_x(\cdot)$ 가 유계이다.

셋째, 식  $\lim_{x_n \rightarrow x} f_x(x_n) = f_x(x)$ 를 증명하자. 식  $\lim_{n \rightarrow \infty} x_n = x$ 가 성립한다고 가정하면, 다음

식이 성립한다.

$$\lim_{n \rightarrow \infty} \exp(-itx_n) \phi_x(t) = \exp(-itx) \phi_x(t) \quad (4)$$

각  $t \in \mathbb{R}$ 와 각  $n \in \mathbb{N}$ 에 대해서, 다음 식이 성립한다.

$$|\exp(-itx_n) \phi_x(t)| \leq |\phi_x(t)| \quad (5)$$

가정에 의해서 특성함수  $\phi_x(\cdot)$ 가 실수집합  $\mathbb{R}$ 에서 적분가능이므로, 다음 식이 성립한다.

$$\int_{-\infty}^{\infty} |\phi_x(t)| dt < \infty \quad (6)$$

식 (4)~식 (6)에서 알 수 있듯이, Lebesgue 지배수렴정리에 의해서 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx_n} \phi_x(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_x(t) dt \quad (7)$$

넷째, 함수  $f_x(\cdot)$ 가 비음임을 증명하자. 예제 A.6에서 알 수 있듯이, 표준정규확률변수  $y$ 와 양수  $\alpha$ 에 대해서 확률변수  $z_\alpha \doteq x + \alpha y$ 의 확률밀도함수  $f_{z_\alpha}(\cdot)$ 는 다음과 같다.

$$f_{z_\alpha}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-izt} \phi_x(t) \exp\left(-\frac{\alpha^2 t^2}{2}\right) dt \quad (8)$$

다음 식들이 성립한다.

$$\begin{aligned} & \sup_{x \in \mathbb{R}} |f_x(x) - f_{z_\alpha}(x)| \\ &= \sup_{x \in \mathbb{R}} \left| \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \phi_x(t) \left[1 - \exp\left(-\frac{\alpha^2 t^2}{2}\right)\right] dt \right| \\ &\leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |\phi_x(t)| \left[1 - \exp\left(-\frac{\alpha^2 t^2}{2}\right)\right] dt \end{aligned} \quad (9)$$

각  $t \in \mathbb{R}$ 에 대해서 다음 식이 성립한다.

$$\lim_{\alpha \rightarrow 0} |\phi_x(t)| \left[1 - \exp\left(-\frac{\alpha^2 t^2}{2}\right)\right] = 0 \quad (10)$$

각  $t (\in \mathbb{R})$  와 각  $n (\in \mathbb{Z})$  에 대해서, 다음 식이 성립한다.

$$\left| \phi_x(t) \left[ 1 - \exp\left(-\frac{\alpha^2 t^2}{2}\right) \right] \right| \leq |\phi_x(t)| \quad (11)$$

가정에 의해서 특성함수  $\phi_x(\cdot)$  가 실수집합  $\mathbb{R}$  에서 적분가능하므로, 다음 식이 성립한다.

$$\int_{-\infty}^{\infty} |\phi_x(t)| dt < \infty \quad (12)$$

식 (10)~식 (12) 에서 알 수 있듯이, Lebesgue 지배수렴정리에 의해서 다음 식이 성립한다.

$$\limsup_{\alpha \rightarrow 0} \sup_{x \in \mathbb{R}} |f_x(x) - f_{z_\alpha}(x)| = 0 \quad (13)$$

식 (13) 에서 알 수 있듯이,  $\alpha$  가 0 에 수렴하면  $f_{z_\alpha}(\cdot)$  가  $f_x(\cdot)$  로 일양수렴한다. 또한  $f_{z_\alpha}(\cdot)$  가 실수함수이고 비음이다. 따라서  $f_x(\cdot)$  도 실수함수이고 비음이다.

다섯째, 임의의 실수들  $a$  와  $b$  에 대해서 다음 식이 성립함을 증명하자.

$$\Pr(a < x \leq b) = \int_a^b f_x(x) dx, \quad (-\infty < a < b < \infty) \quad (14)$$

다음과 같이 함수  $g_n : \mathbb{R} \rightarrow \mathbb{R}$  을 정의하자.

$$g_n(x) = \begin{cases} 0, & (-\infty < x \leq a) \\ n[x - a], & (a < x \leq a + \frac{1}{n}) \\ 1, & (a + \frac{1}{n} < x \leq b) \\ n[b + \frac{1}{n} - x], & (b < x \leq b + \frac{1}{n}) \\ 0, & (b + \frac{1}{n} < x < \infty) \end{cases} \quad (15)$$

다음 식이 성립함은 자명하다.

$$0 \leq g_n(x) \leq 1, \quad (x \in \mathbb{R}) \quad (16)$$

다음 식이 성립한다.

$$E(g_n(z_\alpha)) = \int_{-\infty}^{\infty} g_n(x) f_{z_\alpha}(x) dx \quad (17)$$

식 (16)과 Lebesgue지배수렴정리에서 알 수 있듯이, 다음 식이 성립한다.

$$\lim_{\alpha \rightarrow 0} E(g_n(z_\alpha)) = E(g_n(x)) \tag{18}$$

다음 식들이 성립한다.

$$\lim_{\alpha \rightarrow 0} g_n(x)f_{z_\alpha}(x) = g_n(x)f_x(x), \quad (x \in \mathbb{R}) \tag{19}$$

$$g_n(x)f_x(x) \leq \int_{-\infty}^{\infty} |\phi_x(t)| dt 1_{(a,b+1]}(x) \tag{20}$$

$$\int_{-\infty}^{\infty} 1_{(a,b+1]}(x) dx < \infty \tag{21}$$

식 (19)~식 (21)과 Lebesgue지배수렴정리에서 알 수 있듯이, 다음 식이 성립한다.

$$\lim_{\alpha \rightarrow 0} \int_{-\infty}^{\infty} g_n(x)f_{z_\alpha}(x) dx = \int_{-\infty}^{\infty} g_n(x)f_x(x) dx \tag{22}$$

식 (18)과 식 (22)에서 알 수 있듯이, 다음 식이 성립한다.

$$E(g_n(x)) = \int_{-\infty}^{\infty} g_n(x)f_x(x) dx \tag{23}$$

다음 식들이 성립한다.

$$\lim_{n \rightarrow \infty} g_n(x) = 1_{(a,b]}(x), \quad (x \in \mathbb{R}) \tag{24}$$

$$g_n(x) \leq 1_{(a,b]}(x) \tag{25}$$

$$\int_{-\infty}^{\infty} 1_{(a,b]}(x) dx < \infty \tag{26}$$

식 (24)~식 (26)과 Lebesgue지배수렴정리에서 알 수 있듯이, 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} g_n(x)f_x(x) dx = \int_{-\infty}^{\infty} 1_{(a,b]}(x)f_x(x) dx \tag{27}$$

식 (23)과 식 (27)에서 알 수 있듯이, 다음 식이 성립한다.

$$\lim_{n \rightarrow \infty} E(g_n(x)) = \int_{-\infty}^{\infty} 1_{(a,b]}(x)f_x(x) dx \tag{28}$$

다음 식들이 성립한다.

$$\lim_{n \rightarrow \infty} g_n(x) = 1_{(a,b]}(x), \quad (x \in \mathbb{R}) \quad (29)$$

$$g_n(x) \leq 1_{(a,b]}(x) \quad (30)$$

$$\int_{-\infty}^{\infty} 1_{(a,b]}(x) dx < \infty \quad (31)$$

식 (29)~식 (31) 과 Lebesgue 지배수렴정리에서 알 수 있듯이, 다음 식들이 성립한다.

$$\lim_{n \rightarrow \infty} E(g_n(x)) = E\left(\lim_{n \rightarrow \infty} g_n(x)\right) = E(1_{(a,b]}(x)) \quad (32)$$

식 (28) 과 식 (32) 에서 알 수 있듯이, 다음 식이 성립한다.

$$E(1_{(a,b]}(x)) = \int_{-\infty}^{\infty} 1_{(a,b]}(x) f_x(x) dx \quad (33)$$

식 (33) 과 식 (14) 는 동치이다. ■

**예제 A.15** Lebesgue 측도를 이용해서 특성함수의 일의성정리에 대해서 살펴보자. 이에 대한 자세한 내용은 Chung [26]의 제6장을 참조하라.

확률변수  $x$ 의 확률측도를  $\mu(\cdot)$ 라 하면, 특성함수는 다음과 같이 정의된다.

$$\phi_x(t) = \int_{-\infty}^{\infty} e^{itx} \mu(dx) \quad (1)$$

만약 식  $a < b$ 가 성립하면, 다음 식이 성립한다.

$$\mu((x_1, x_2)) + \frac{1}{2}\mu(\{x_1\}) + \frac{1}{2}\mu(\{x_2\}) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} f_x(t) dt \quad (2)$$

여기서 피적분함수의 점  $t = 0$ 에서 값은 연속성에 의해서 정의된다. ■

## A.2 중심극한정리

이제 단변량 중심극한정리를 기술하고 증명하기로 하자.

**정리 A.1: 단변량 중심극한정리**

확률변수들  $x_1, x_2, \dots$  이 서로 독립이고 평균이  $\mu$  이고 분산이  $\sigma^2$  인 동일한 확률분포를 따른다고 하자. 만약  $E(|x_1|^3) < \infty$  이면, 다음 식이 성립한다.

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n [x_k - \mu] \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

증명. 다음 식이 성립한다.

$$\phi_{x_k}(t) = 1 - \frac{\sigma^2}{2} t^2 + O(t^3), \quad (k = 1, 2, \dots) \tag{1}$$

다음 식들이 성립한다.

$$\phi_{\frac{1}{\sqrt{n}} \sum_{k=1}^n x_k}(t) = \prod_{k=1}^n \phi_{x_k} \left( \frac{t}{\sqrt{n}} \right) = \left\{ 1 - \frac{\sigma^2}{2} \left[ \frac{t}{\sqrt{n}} \right]^2 + O \left( \left[ \frac{t}{\sqrt{n}} \right]^3 \right) \right\}^n \tag{2}$$

여기서 첫 번째 등호는 명제 A.1의 성질 (e)와 명제 A.4에 의해서 성립하고, 두 번째 등호는 식 (1)에 의해서 성립한다. 식 (2)에서 알 수 있듯이, 각  $t \in \mathbb{R}$  에 대해서 다음 식들이 성립한다.

$$\lim_{n \rightarrow \infty} \phi_{\frac{1}{\sqrt{n}} \sum_{k=1}^n x_k}(t) = \lim_{n \rightarrow \infty} \left\{ 1 - \frac{\sigma^2}{2} \left[ \frac{t}{\sqrt{n}} \right]^2 + O \left( \left[ \frac{t}{\sqrt{n}} \right]^3 \right) \right\}^n = \exp \left( -\frac{1}{2} \sigma^2 t^2 \right). \tag{3}$$

명제 A.6의 연속성정리와 예제 A.2에서 알 수 있듯이, 다음 식이 성립한다.

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n x_k \xrightarrow{d} \mathcal{N}(0, \sigma^2) \tag{4}$$

■

지금부터는 단변량 중심극한정리를 이용해서 다변량 중심극한정리를 유도하자. 서로 독립이고 동일한 확률분포를 따르는 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  의 평균벡터가  $\boldsymbol{\mu} = E(\mathbf{x}_i)$  이고 분산공분산행렬이  $\Sigma = E([\mathbf{x}_i - \boldsymbol{\mu}][\mathbf{x}_i - \boldsymbol{\mu}]^T)$  라고 하자. 실수  $u \in \mathbb{R}$  와 실벡터  $\mathbf{t} \in \mathbb{R}^p$  에 대해서 다음 함수를 정의하자.

$$m_N(\mathbf{t}, u) = E \left( \exp \left( u \mathbf{t}^T \frac{1}{\sqrt{N}} \sum_{i=1}^N [\mathbf{x}_i - \boldsymbol{\mu}] \right) \right) \tag{A2.1}$$

고정된  $\mathbf{t}$ 에 대해서  $m_N(\mathbf{t}, u)$ 는 확률변수  $\frac{1}{\sqrt{N}} \sum_{i=1}^N [\mathbf{t}^T \mathbf{x}_i - \mathbf{t}^T \boldsymbol{\mu}]$ 의 적률모함수이다. 다변량 중심극한정리에서 알 수 있듯이, 이 확률변수의 극한확률분포가  $\mathcal{N}(0, \mathbf{t}^T \Sigma \mathbf{t})$ 이다. 즉, 임의의  $u \in \mathbb{R}$ 와  $\mathbf{t} \in \mathbb{R}^p$ 에 대해서 다음 식이 성립한다.

$$\lim_{N \rightarrow \infty} m_N(\mathbf{t}, u) = \exp\left(\frac{1}{2} u^2 \mathbf{t}^T \Sigma \mathbf{t}\right) \quad (\text{A2.2})$$

식 (A2.2)에 식  $u = 1$ 을 대입하면, 임의의  $\mathbf{t} \in \mathbb{R}^p$ 에 대해 다음 식이 성립함을 알 수 있다.

$$\lim_{N \rightarrow \infty} E\left(\exp\left(\mathbf{t}^T \frac{1}{\sqrt{N}} \sum_{i=1}^N [\mathbf{x}_i - \boldsymbol{\mu}]\right)\right) = \exp\left(\frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t}\right) \quad (\text{A2.3})$$

함수  $\exp(\frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t})$ 가 점  $\mathbf{t} = \mathbf{0}$ 에서 연속이므로, 식 (A2.3)이 점  $\mathbf{t} = \mathbf{0}$ 의 이웃집합(neighborhood)에서 수렴한다. 따라서, 다음 정리가 성립한다.

#### 정리 A.2: 다변량 중심극한정리

서로 독립이고 동일한 확률분포를 따르는 확률벡터들  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ 의 평균벡터가  $E(\mathbf{x}_i) = \boldsymbol{\mu}$ 이고 분산공분산행렬이  $E([\mathbf{x}_i - \boldsymbol{\mu}][\mathbf{x}_i - \boldsymbol{\mu}]^T) = \Sigma$ 이면, 다음 식이 성립한다.

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N [\mathbf{x}_i - \boldsymbol{\mu}] \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma)$$

## 참고 문헌

- [1] 최병선 (1992) 『단변량시계열분석』, 세경사.  
<http://s-space.snu.ac.kr/handle/10371/94457>
- [2] 최병선 (1995) 『다변량시계열분석』, 세경사.  
<http://s-space.snu.ac.kr/handle/10371/94434>
- [3] 최병선 (1997a) 『회귀분석(상)』, 세경사.  
<http://s-space.snu.ac.kr/handle/10371/94433>
- [4] 최병선 (1997b) 『회귀분석(하)』, 세경사.  
<http://s-space.snu.ac.kr/handle/10371/94456>
- [5] 최병선 (2002) 『Fourier 해석입문』, 세경사.  
<http://s-space.snu.ac.kr/handle/10371/94413>
- [6] 최병선 (2004) 『이산형 재무모형의 수리적 배경』, 세경사.  
<http://s-space.snu.ac.kr/handle/10371/94455>
- [7] 최병선 (2007) 『계산재무론』, 세경사.
- [8] 최병선 (2009) 『금융공학 I: Elements of Financial Engineering』, 세경사.
- [9] 최병선 (2013) 『금융공학 II: Black-Scholes Formula』, 서울대학교 출판문화원.
- [10] 최병선 (2015) 『금융공학 III: Introduction to Financial Engineering』, 세경사.  
<http://s-space.snu.ac.kr/handle/10371/142795>
- [11] 최병선 (2016) 『금융공학 IV: Monte Carlo Methods for Finance and Economics』,  
김구재단.  
<http://s-space.snu.ac.kr/handle/10371/98466>

- [12] 최병선 (2018) 『금융공학 VII: Scientific Computing For Finance and Economics』, 김구재단.  
<http://s-space.snu.ac.kr/handle/10371/139655>
- [13] 최병선·이성백 (2001) 『행렬의 대각화와 Jordan표준형』, 세경사.  
<http://s-space.snu.ac.kr/handle/10371/94394>
- [14] Alaoglu, L. and Birkhoff, G. (1940) General ergodic theorems. *The Annals of Mathematics* **41**, 293-309.
- [15] Anderson, T.W. (2003) *An Introduction to Multivariate Statistical Analysis (Third Edition)*, Wiley.
- [16] Bauwens, L., Lubrano, M. and Richard, J.F. (1999) *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press.
- [17] Birkhoff, G.D. (1931) Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences of the United States of America* **17**, 656.
- [18] Blackwell, D. and Kendall, D. (1964) The Martin boundary for Pólya's urn scheme, and an application to stochastic population growth. *Journal of Applied Probability*, **1**, 284-296.
- [19] Bolstad, W.M. (2010) *Understanding Computational Bayesian Statistics*, Wiley.
- [20] Box, G.E.P. (1976) Science and statistics. *Journal of the American Statistical Association* **71**, 791-799.
- [21] Box, G.E.P. and Draper, N.R. (1987) *Empirical Model-Building and Response Surfaces*, Wiley.
- [22] Brandimarte, P. (2014) *Handbook in Monte Carlo Simulation: Applications in Financial Engineering, Risk Management, and Economics*, Wiley.
- [23] Carlin, B.P., Gelfand, A.E., and Smith, A.F.M. (1992) Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics* **41**, 389-405.
- [24] Casella, G. and George, E. (1992) An introduction to gibbs sampling. *The American Statistician* **46**, 167-174.

- [25] Choi, B. (1992) *ARMA Model Identification*, Springer.
- [26] Chung, K.L. (2001) *A Course in Probability Theory (Third Edition)*, Academic Press.
- [27] Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings algorithm. *The American Statistician* **49**, 327-335.
- [28] Cowles, M.K. and Carlin, B.P. (1996) Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association* **91**, 883-904.
- [29] de Finetti, B. (1931) Funzione caratteristica di un fenomeno aleatorio. *Atti della R. Accademia Nazionale dei Lincei, Serie 6. Memorie, Classe di Scienze Fisiche, Matematiche e Naturale*, **4**, 251-299.
- [30] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1-38.
- [31] Diaconis, P. and Freedman, D. (1980) Finite exchangeable sequences. *Annals of Probability* **8**, 745-764.
- [32] Diaconis, P. and Freedman, D. (1987) A dozen De Finetti-style results in search of a theory. *Ann. Inst. Henri Poincaré Suppl. au N.2* **23**, 397-423.
- [33] Dudley, R.M. (2002) *Real Analysis and Probability (Second Edition)*, Cambridge University Press.
- [34] Durbin, J. and Koopman, S.J. (2002) A simple and efficient simulation smoother for state space time series analysis. *Biometrika* **89**, 603-616.
- [35] Epps, T.W. (1993) Characteristic functions and their empirical counterparts: geometrical interpretations and applications to statistical inference. *The American Statistician* **47**, 33-38.
- [36] Fair, R. (1977) A note on the computation of the tobit estimator. *Econometrica* **45**, 1723-1727.

- [37] Fair, R. (1978) A theory of extramarital affairs. *Journal of Political Economy* **86**, 45-61.
- [38] Feller, W. (1968) *An Introduction to Probability Theory and Its Applications, Vol. I (Third Edition)*, Wiley.
- [39] Feller, W. (1971) *An Introduction to Probability Theory and Its Applications, Vol. II (Second Edition)*, Wiley.
- [40] Ferguson, T. (1967) *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press.
- [41] Flury, B. and Zoppe, A. (2000) Exercises in EM. *The American Statistician* **54**, 207-209.
- [42] Gelfand, A., and Smith, A.F.M. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398-409.
- [43] Geman, S., and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**, 721-740.
- [44] Geweke, J. (1991) Efficient simulation from the multivariate normal and student- $t$  distributions subject to linear constraints, In *Computing Science and Statistics: Proceedings of the TwentyThird Symposium on the Interface*, E. M. Keramidas (ed). 571-578. Fairfax: Interface Foundation of North America, Inc.
- [45] Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics 4*, Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. (eds.), 169-193. Oxford: Oxford University.
- [46] Geweke, J. (2005) *Contemporary Bayesian Econometrics and Statistics*, Wiley.
- [47] Geweke, J., Koop, G. and Van Dijk, H. (2011) *The Oxford Handbook of Bayesian Econometrics*. Oxford University Press.
- [48] Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1996) *Markov Chain Monte Carlo in Practice*, Chapman & Hall.

- [49] Greenberg E. (2008) *Introduction to Bayesian Econometrics*. Cambridge University Press.
- [50] Greene, W.H. (2003) *Econometric Analysis (Fifth Edition)*, Prentice Hall.
- [51] Greene, W.H. (2012) *Econometric Analysis (Seventh Edition)*, Prentice Hall.
- [52] Grunfeld, Y. (1958) The Determinants of Corporate Investment, Unpublished Ph.D. Thesis, Chicago.
- [53] Hastings, W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97-109.
- [54] Hewitt, E. and Savage, L.J. (1955) Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, **80**, 470–501.
- [55] Hoff, P.D. (2009) *A First Course in Bayesian Statistical Methods*, Springer.
- [56] Jeffreys, H. (1961) *Theory of Probability (Third Edition)*, Oxford University Press.
- [57] Jondeau, E., Poon, S., and Rockinger, M. (2007) *Financial Modeling Under Non-Gaussian Distributions*, Springer.
- [58] Keynes, J.M. (1921) *Treatise on Probability*, Macmillan & Co.
- [59] Kleiber, C. and Zeileis, A. (2008) *Applied Econometrics with R*, Springer.
- [60] Knott, C. (1897) On lunar periodicities in earthquake frequency. *Proceedings of the Royal Society of London* **60**, 457-466.
- [61] Koop, G. (2003) *Bayesian Econometrics*, Wiley.
- [62] Koop, G., Poirier, D.J., and Tobias, J.L. (2007) *Bayesian Econometric Methods*, Cambridge University Press.
- [63] Kroese, D.P., Taimre, T., and Botev, Z.I. (2011) *Handbook of Monte Carlo Methods*, Wiley.
- [64] Lancaster, T. (2004) *An Introduction to Modern Bayesian Econometrics*, Blackwell.

- [65] Lauritzen, S.L. (1988) *Extremal Families and Systems of Sufficient Statistics*, Springer-Verlag.
- [66] Lehmann, E.L. and Casella, G. (2003) *Theory of Point Estimation (Second Edition)*, Springer.
- [67] Leonard, T. and Hsu, J.S.J. (1999) *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers*, Cambridge University Press.
- [68] Litterman, R.B. (1984a) Specifying vector autoregressions for macroeconomic forecasting. Federal Reserve Bank of Minneapolis, Staff Report no. 92.
- [69] Litterman, R.B. (1984b) Forecasting and policy analysis with bayesian vector autoregression models. *Federal Reserve Bank of Minneapolis, Quarterly Review* **8(4)**, 30-41.
- [70] Litterman, R.B. (1986) Forecasting with bayesian vector autoregressions - five years of experience. *Journal of Business and Economic Statistics* **4(1)**, 25-38.
- [71] Mcgrayne, S.B. (2012) *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*, Yale University Press.
- [72] McLachlan, G.J. and Krishnan, T. (2008) *The EM Algorithm and Extensions (Second Edition)*, Wiley.
- [73] McLachlan, G.J. and Peel, D. (2000) *Finite Mixture Models*, Wiley.
- [74] McNeil, A.J., Frey, R. and Embrechts, P. (2015) *Quantitative Risk Management: Concepts, Techniques and Tools (Revised Edition)*, Princeton University Press.
- [75] Meng, X.L. and Rubin, D.B. (1991) Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association* **86**, 899-909.
- [76] Metropolis, N.A., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087-1092.

- [77] Phillips, P.C.B. (1982) The true characteristic function of the  $F$  distribution. *Biometrika* **69**, 261–264.
- [78] Pliny the Elder (1944) *Natural History*, *Loeb Classical Library*, Translated by H. Rackham, Harvard University Press, 1944.
- [79] Poirier, D. (1995) *Intermediate Statistics and Econometrics: A Comparative Approach*, MIT Press.
- [80] Rachev, S.T., Hsu, J., Bagasheva, B.S., and Fabozzi, F.J. (2008) *Bayesian Methods in Finance*, Wiley.
- [81] Raftery, A.E. and Akman, V.E. (1986) Bayesian analysis of a Poisson process with a change-point. *Biometrika*, **73**, 85–89.
- [82] Ramsey, F.P. (1931) Truth and Probability. In *The Foundations of Mathematics and Other Logical Essays*, ed. R. B. Braithwaite, 156–98. Kegan, Paul, Trench, Trubner and Co.
- [83] Robert, C.P. (2001) *The Bayesian Choice – A Decision-Theoretic Motivation (Second Edition)*, Springer.
- [84] Robert, C.P. and Casella, G. (2004) *Monte Carlo Statistical Methods (Second Edition)*, Springer.
- [85] Ross, S. (1996) *Stochastic Processes (Second Edition)*, Wiley.
- [86] Rossi, P.E., Allenby, G. and McCulloch, R. (2005) *Bayesian Statistics and Marketing*, Wiley.
- [87] Rossi, P.E. and McCulloch, R. (2008) *Bayesm: Bayesian Inference for Marketing/Microeconometrics. R package version 2.2.2*. Available on <http://faculty.chicogogsb.edu/peter.rossi/research/bsm.html>.
- [88] Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*, Wiley.
- [89] Ruiz, I. (2015) *XVA Desks - A New Era for Risk Management; Understanding, Building and Managing Counterparty, Funding and Capital Risk*, Palgrave.

- [90] Savage, L.J. (1954) *The Foundations of Statistics*, Wiley.
- [91] Schuster, A. (1897) On lunar and solar periodicities of earthquakes. *Proceedings of the Royal Society of London* **61**, 455–465.
- [92] Schuster, A. (1906) On the periodicities of sunspots. *Philosophical Transactions of the Royal Society of London. Series A* **206**, 69-100.
- [93] Silver, N. (2012) *The Signal and the Noise: Why So Many Predictions Fail-But Some Don't*, Penguin Press.
- [94] Steyvers, M., Lee, M.D., Miller, B., and Hemmer, P. (2009) The wisdom of crowds in the recollection of order information. In J. Lafferty and C. Williams (Eds.), *Advances in neural information processing systems*, **22**, NIPS.
- [95] Tierney, L. (1994) Markov chains for exploring posterior distributions. *Annals of Statistics* **22**, 1701-1762.
- [96] Walters, P. (1982) *An Introduction to Ergodic Theory*, Springer.
- [97] Williams, D. (1991) *Probability with Martingales*, Cambridge University Press.
- [98] Zellner, A. (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* **57**, 348-368.
- [99] Zellner, A. (1971) *An Introduction to Bayesian Inference in Econometrics*, Wiley.
- [100] Zellner, A. and Min, C. (2012) Gibbs sampler convergence criteria. *Journal of the American Statistical Association* **90**, 921-927.

# 찾아보기

- F* probability distribution, 33
- F* 확률분포, 33
- $\beta$ -CVaR, 94
- t*-probability distribution, 32
- t* 확률분포, 32, 486
- z*-score, 188
- z* 스코어, 188
- 2변량 지수확률밀도함수, 572
- 2차수렴, 182
- 2차원 대칭확률보행, 374
- 3차원 대칭확률보행, 375
- t* 확률분포, 489
- Metropolis-Hastings 샘플러, 395
- absorbing, 354
- absorbing state, 346, 347
- acceptance rate, 568, 613
- Acceptance-Rejection
  - Metropolis-Hastings sampler, 601
- Acceptance-Rejection step, 601
- admissible, 71
- All models are wrong, but some are
  - useful, 2
- aperiodic, 359
- aperiodicity, 556
- AR-MH 샘플러, 601, 626
- AR-step, 626
- area element, 135
- ARMA 모형, 294, 633, 634, 638
- ARX 모형, 632, 637, 639, 648
- AR 단계, 626
- AR 모형, 632
- assessible, 347, 355
- asymptotic relation, 613
- autocorrelation, 222
- autocorrelation function, 498
- autocovariance, 223
- average loss, 71
- backshift operator, 663
- barrier option, 346
- batch, 615
- Bayes factor, 189, 615
- Bayesian statistics, 1
- Bayes 결정과정, 71, 74
- Bayes 액션, 184
- Bayes 위험, 71

- Bayes 정리, 3, 5, 113  
 Bayes 팩터, 189, 615  
 Bernouille 확률분포, 447  
 Bernoulli 확률분포, 6, 10, 22, 88, 155, 169  
 best linear unbiased estimator, 527  
 beta probability distribution, 33  
 beta-binomial distribution, 6  
 binary logit regression model, 630  
 BIS 규제, 92  
 bivariate exponential probability density, 572  
 blockwise MH sampling, 568  
 BLUE, 527  
 Borel 집합족, 409  
 burn-in period, 410, 413  
 burn-in sample, 612  
  
 candidate distribution, 395  
 candidate generating distribution, 609  
 candidate point, 555  
 canonical, 338  
 Cauchy probability distribution, 40  
 Cauchy 확률밀도함수, 556  
 Cauchy 확률분포, 40  
 CDS, 109, 120  
 CDS 프리미엄, 109, 120  
 CE, 93  
 censored negative binomial regression, 514  
 censored Poisson regression, 514  
 censored regression model, 511  
 CESF, 93  
  
 Chapman-Kolmogorov 식, 315, 316, 355  
 characteristic function, 29  
 chi-squared probability distribution, 31  
 Cholesky 분해, 140, 416, 536, 665  
 closed set, 367  
 communicative, 347, 355  
 complete, 71  
 componentwise MH sampling, 568  
 concave function, 141  
 condition number, 259  
 conditional independence, 86  
 conditional probability density, 53  
 conditional value at risk, 91, 93  
 conjugate, 123  
 conjugate prior distribution, 123  
 conjugate prior probability distribution, 153  
 consistency, 386  
 contingency table, 21  
 convex term, 140  
 countale, 377  
 counterparty, 91  
 country risk, 107  
 credible interval, 614  
 credit default swap, 109  
 credit expected short fall, 93  
 credit risk, 91, 105  
 credit set, 186  
 current exposure, 93  
 CVaR, 91, 93–95, 97, 99, 103  
  
 data precision, 123, 144

- de Finetti 정리, 84, 87
- decesion rule, 184
- decision procedure, 71
- decision theory, 184
- default odds ratio, 108
- degenerate random variable, 44
- detailed balance equation, 608
- deviance, 210
- diffusion coefficient, 298
- digamma function, 181
- Dirac 델타함수, 44
- disjoint, 355
- drift coefficient, 298
- drift vector, 295
- Durbin-Watson 검정, 222
- dynamic model, 297
- earned run average, 545
- Economic Statistics System, 473
- ECOS, 473
- Ehrenfest 체인, 303
- empirical probability distribution, 612
- ENE, 93
- EPE, 93
- equilibrium distribution, 332
- equivalence relation, 355
- equivalent class, 355
- ergodic, 378, 381
- Erlang 확률분포, 29
- exchangeable, 84
- expected negative exposure, 93
- expected positive exposure, 93
- exponential family, 162
- exponential probability distribution, 28
- extreme distribution function, 205
- fast Fourier transform, 222
- FFT, 222
- first hitting time, 344
- first return time, 344
- first visit time, 344, 356
- Fisher 정보량, 168, 170
- fixed point integral equation, 439
- flow, 473
- Fourier 변환, 223
- Fubini 정리, 76
- gambler's ruin, 301
- gamma probability distribution, 30
- Gauss-Markov 조건, 527, 536
- generalized exponential probability distribution, 31
- generalized extreme value probability distribution, 36
- generalized least squares estimator, 114
- generalized linear model, 201, 202, 211, 503
- generalized Pareto distribution, 35
- generalized variance, 256
- Gibbs 샘플러, 173, 177, 221, 283, 288, 290, 411, 555, 609
- Gibbs 샘플링, 558
- GLM, 211, 503
- GLS 추정량, 114

- Google Flu Trends, 258  
 gradient, 141  
 Gram-Schmidt 정리, 65  
 Grunfeld 데이터, 530  
  
 Hammersley-Clifford 정리, 450  
 heat bath algorithm, 411  
 Hessian 행렬, 141, 252  
 Hewitt-Savage 정리, 88  
 hierarchical model, 224  
 high frequency, 222  
 highest posterior probability density  
     interval: HPD interval, 187  
 Hotelling  $T^2$  통계량, 78  
 hyper-ball, 134  
 hyper-parameter vector, 100  
 hyper-prior probability distribution, 227  
 hypergeometric random variable, 26  
 hyperparameter vector, 224  
  
 iid, 310  
 improper, 167, 401, 450  
 inadmissible, 71  
 independence sampler, 598  
 independence test, 222  
 independent and identically distributed,  
     310  
 infinite set, 372  
 initial distribution, 292  
 initial state, 388  
 innovation, 294  
 innovation error vector, 297  
  
 instrumental distribution, 395  
 intensity, 591  
 invariance, 135  
 invariant distribution, 290, 332  
 inverse gamma distribution, 41  
 inverse Gaussian probability  
     distribution, 40  
 inverse Wishart probability distribution,  
     61  
 irreducibility, 455, 556  
 irreducible, 355, 377, 386  
 isoquant curve, 570  
 iteratively reweighted least squares  
     method, 214  
 Ito 적분, 298  
  
 Jacobian, 135, 240  
 James-Stein 추정벡터, 77  
 Jeffreys 원리, 169  
 Jeffreys의 사전확률분포, 168, 169, 171  
 Jeffreys형 사전확률분포, 166  
 Jensen 부등식, 116  
 joint generalized least squares method,  
     527  
  
 Kakutani, 377  
 Kalman 필터, 554  
 Kendall의 타우거리, 576  
 kernel smoothing density; ks density, 223  
 Koopman-Darmois 형, 162  
 Kronecker 델타함수, 64, 228  
  
 lag, 613

- lag operator, 663
- latent variable, 513
- left eigenvector, 332
- level curve, 105
- limiting steady state distribution, 332
- linear-linear loss function, 185
- link function, 201, 202
- logistic model, 492
- logistic probability distribution, 39
- logistic regression model, 119
- logit model, 205
- loss function, 71, 92, 148, 184
- low frequency, 222
- lower incomplete gamma function, 30
- marginal likelihood, 615
- marginal probability distribution, 221
- market risk, 91
- Markov chain, 289
- Markov Chain Monte Carlo method, 4, 289
- Markov property, 290
- Markov 성, 290, 294
- Markov 스위칭 ARMA 모형, 295
- Markov 체인, 289, 298
- mass, 576
- maximum a posterior estimator, 179
- MCMC, 4, 6, 120, 164, 201, 221, 224, 289, 386, 397, 412
- MCMC method, 289
- mean curve, 214
- mean recurrent time, 378
- measurement equation, 297
- Metropolis-Hastings acceptance probability, 556
- Metropolis-Hastings 샘플러, 555
- Metropolis-Hastings 알고리즘, 290
- Metropolis 샘플러, 584, 592, 610, 616, 628
- Metropolis 알고리즘, 393
- mgf, 42
- MH 단계, 627
- MH 샘플러, 555
- MH 샘플 열, 612
- MH 샘플 평균, 612
- MH 채택 확률, 556
- minimal complete, 71
- minimax decision procedure, 74
- mixed normal probability distribution, 560
- mode, 621
- moment generating function, 29, 42
- Monte Carlo method, 289
- Monte Carlo 법, 218
- moving average window, 221
- MR-step, 627
- MS-ARMA model, 295
- multi-block algorithm, 610
- multi-move, 614
- multinomial probability distribution, 26
- multiple chain, 612
- natural conjugate distribution family, 123

- natural conjugate prior distribution, 101
- NBER-NSF 세미나, 3
- negative binomial distribution, 25
- Newton-Raphson 법, 182, 207
- non-negative definite matrix, 44
- noncentral  $F$  probability distribution, 39
- noncentral  $t$ -probability distribution, 38
- noncentral chi-square probability distribution, 37
- noncentrality, 37–39
- noninformative prior probability distribution, 166
- normal random vector, 416
- normal-gamma distribution, 196
- normalizing constant, 5, 217, 610
- odds ratio, 108, 506
- operational risk, 91
- order statistics, 614
- orthonormal matrix, 51
- output equation, 297
- partial correlation, 260
- partition, 355
- Pascal 확률분포, 24
- period, 359
- periodic, 359
- periodogram, 222
- permutation, 84, 576
- PFE, 93, 95, 99
- Poisson 확률모형, 478
- Poisson 확률밀도함수, 587
- Poisson 확률분포, 26, 159, 161, 163, 180, 201, 206, 308, 469, 478, 591
- Poisson 회귀모형, 202, 206, 213
- positive definite matrix, 45
- positive recurrent, 378
- positivity condition, 449, 455
- posterior mean, 178
- posterior median, 179
- posterior mode, 179
- posterior odds ratio, 189
- posterior precisio, 123
- posterior precision, 144
- posterior predictive probability density function, 191
- potential future exposure, 93
- precision, 195
- predictive distribution, 99
- prior assumption, 123, 144
- prior odds ratio, 189
- prior precision, 123, 144
- probit model, 204, 492
- PROC MODEL, 529
- PROC SYSLIN, 529
- proper, 167
- proposal distribution, 395, 555, 609
- proposal step size, 628
- quadratic convergence, 182
- quadratic loss function, 184
- queueing model, 29
- random number, 221

- random sweep Gibbs sampler, 412
- random vector, 568
- random-walk Metropolis sampler, 592
- randomness test, 222
- Rao-Blackwellization, 446
- Rao-Blackwell정리, 443
- rate, 29
- Rayleigh probability distribution, 35
- Rayleigh 확률분포, 35
- recurrent, 353, 354
- Red Sox, 545
- reducible, 355, 455
- reflexibility, 355
- regression coefficient vector, 297
- regularity conditions, 413
- regularity condtions, 210
- remembered ordering, 576
- reparameterization, 614
- reversed Markov chain, 388
- reversible, 389
- Riemann 적분, 298
- right eigenvector, 315
- risk, 91
- risk event, 91
- risk function, 71
- risk measure, 91
- risk metric, 91
- R 함수 metrop, 628
- SAS/ETS, 529
- SAS/STAT, 265
- scale parameter, 29
- scaled inverse- $\chi^2$  distribution, 129
- scaled inverse-chi-squared distribution, 42
- Schur complement, 137
- Schur보완, 137
- Schur분해, 63, 68, 325, 338, 357, 358, 416
- seemingly unrelated regression model, 526
- semi-conjugate, 175
- serial correlation, 397
- shape parameter, 29
- simple random walk, 313
- simulated annealing, 585
- single chain, 612
- single-move, 614
- singular, 44
- singular value decomposition, 416
- slice sampler, 221
- SLLN, 386
- slowly mixing, 457
- sovereign risk, 107
- spectral function, 223
- spectrum, 222
- stability, 332
- stability condition, 632
- state diagram, 299
- state equation, 297
- state probability, 317
- state space, 290
- state space model, 297

- state space representation, 295
- state transition diagram, 299
- state transition matrix, 297
- state vector, 297
- stationarity, 221
- stationary, 291, 381, 388
- stationary distribution, 332
- steady state, 388
- stochastics matrix, 291
- stock, 473
- strong law of large numbers, 386
- success runs, 303
- sum of squares and crossproducts due to error, 256
- sum of squares and crossproducts due to model, 256
- sunspot, 223
- support, 290
- SUR model, 526
- SVD, 416
- Swedish Road Administration, 206
- symmetry, 355
- systematic sweep Gibbs sampler, 412
- target probability distribution, 221, 290
- tau distance, 576
- Taylor 근사식, 621
- thining out, 222
- time of the first return to, 344
- time-invariance, 330
- time-invariant, time-homogeneous, temporally homogeneous, 291
- time-reversible, 389
- tobit model, 511
- Toeplitz 행렬, 85
- total sum of squares and crossproducts, 256
- total variation, 256
- total variation distance, 409
- transient, 353, 354
- transition diagram, 299
- transition kernel, 290, 409
- transition probability, 291
- transition probability matrix, 292
- transitivity, 355, 377
- trigamma function, 181
- truncated logistic distribution, 507
- truncated normal distribution, 95
- truncated normal probability distribution, 493
- truncated probability density function, 94
- unimodal, 121
- upper incomplete gamma function, 30
- autoregressive generating density, 602
- value at risk, 93
- VaR, 93, 95
- volume element, 135
- Weibull probability distribution, 34
- Weibull 확률밀도함수, 192
- Weibull 확률분포, 34
- Wishart probability distribution, 59

- Wishart 확률변수, 548
- Wishart 확률분포, 59, 146, 549
- Yankees, 545
- zero recurrent, 378
- zero-one loss function, 184
- 가산, 377
- 가설검정, 188
- 가약, 355
- 가약성, 356
- 가약적, 455
- 가역성조건, 390, 608
- 가역적, 389
- 가중평균, 99
- 감마확률밀도함수, 395, 397
- 감마확률분포, 30, 31, 41, 125, 157, 159, 175, 180, 397, 400, 483
- 강건성, 215
- 강대수법칙, 386, 414
- 강도모수, 591
- 거래상대방, 91
- 결정과정, 71
- 결정규칙, 184
- 결정이론, 184
- 결정적 미분방정식, 297
- 결측데이터, 4
- 결합일반화최소제곱추정법, 527, 529
- 결합확률밀도함수, 4
- 경험적 베이시안추정법, 227, 233, 235, 244, 246
- 경험확률분포, 612
- 계열상관, 397
- 계층모형, 224, 233, 244
- 계층모형분석, 227
- 계층적 베이시안모형, 227
- 계층적 베이시안추정량, 235
- 계층적 베이시안추정법, 227, 229, 233, 237
- 고빈도, 222
- 고전통계학, 1
- 고정점적분방정식, 439
- 공액, 123, 197
- 공액사전조건부확률분포, 289
- 공액사전확률분포, 153, 164, 289
- 공액사전확률분포족, 123, 144
- 과도상태, 346, 349
- 과도성, 366
- 과도적, 353, 354, 381
- 과소추정, 227
- 관측방정식, 297, 552
- 교환가능, 84, 87
- 구조변화, 490, 491
- 그레디언트, 141
- 극값확률분포함수, 205
- 극한상태확률, 320, 342
- 극한안정상태분포, 332
- 극한추이확률행렬, 344
- 기대디폴트오즈비, 111, 115
- 기대파산오즈비, 116
- 기대회수액, 107
- 기약, 355
- 기억된 순열, 576

- 기하확률분포, 24
- 난벡터, 568
- 난수, 221
- 난수성검정, 222
- 다변량  $t$  확률분포, 483
- 다변량 MH 샘플링, 568
- 다변량선형모형, 265
- 다변량선형회귀모형, 249, 253, 256, 281, 286
- 다변량정규확률벡터, 45
- 다변량정규확률분포, 49, 66, 69, 130, 133, 146, 253, 281
- 다변량확률분포, 295
- 다중MH샘플러, 614
- 다중블록MH샘플러, 610, 612
- 다중선형회귀모형, 287
- 다중이동, 614
- 다중적분, 3, 289
- 다중체인, 612
- 다항확률분포, 26
- 단변량정규확률분포, 45
- 단봉, 121
- 단순이동, 614
- 단순확률보행, 313, 372
- 단일체인, 612
- 단힌집합, 367, 368, 377
- 대기행렬모형, 29
- 대수디폴트오즈비, 110, 112
- 대수디폴트오즈비의, 120
- 대수정규확률분포, 95
- 대수파산오즈비, 116
- 대칭율, 355
- 대칭확률분포, 584
- 데이터정밀도, 123, 144
- 도달가능, 347, 355, 363
- 도달가능성, 354
- 도달확률, 349
- 도달횟수, 349
- 도박사파산, 301
- 도수주의자, 1, 3
- 독립동일확률변수, 315
- 독립동일확률분포, 310
- 독립샘플러, 395, 598, 601
- 독립성검정, 222
- 동적모형, 297
- 동치관계, 355
- 동치류, 355, 360, 363, 366
- 등가곡선, 105, 133, 570
- 디감마함수, 181
- 디자인행렬, 529
- 디폴트, 92, 107, 108
- 디폴트오즈비, 108, 109, 111, 112, 115, 120
- 디폴트오즈비모형, 110
- 디폴트확률, 107, 108, 119
- 런검정, 222
- 로지스틱확률분포, 39, 507, 510
- 로지스틱확률분포함수, 205
- 로지스틱회귀모형, 119, 120, 215, 492, 503, 505, 510, 620, 627
- 로짓모형, 205
- 리스크, 91

- 리스크관리, 2, 92
- 리스크모형, 92
- 리스크사상, 91
- 리스크지표, 91, 93, 97
- 링크함수, 201, 202
- 마아코프체인몬테카를로법, 289
- 매스, 576
- 면적요소, 135
- 모드, 621
- 모형제공교차곱합행렬, 256
- 모형제공합, 273
- 목적확률분포, 221
- 목표분포, 290
- 몬테카를로법, 19, 289
- 무작위공급, 306
- 무작위적쓰기Gibbs샘플러, 412, 451
- 무정보사전확률분포, 166, 169, 229
- 무한집합, 372
- 미니맥스결정과정, 74
- 반복재가중최소제곱추정법, 214
- 반복조건부기대값법칙, 240
- 반사율, 355
- 방어율, 545
- 배리어옵션, 346
- 백분위수, 613
- 변인, 223, 287
- 변인기간, 410, 413, 568
- 변인샘플, 567, 612
- 베이지안, 1
- 베이지안분석, 289
- 베이지안중심극한정리, 123, 144
- 베이지안추정량, 148, 184
- 베이지안통계학, 1
- 베타이항확률분포, 6, 434, 436
- 베타함수, 6
- 베타확률분포, 6, 7, 10, 18, 20, 33, 156, 181, 567
- 변수변환, 614
- 변환점, 469, 479, 481, 485
- 보기에는 무관한 회귀모형, 526, 529, 546
- 블록행, 140
- 부분상관계수, 260
- 부피요소, 135
- 분산공분산행렬, 45, 138, 256, 288
- 분할, 355
- 분할표, 21
- 분할행렬, 48
- 불가약, 355, 370, 377, 378, 381, 386, 388
- 불가약성, 455, 556
- 불가약적, 360
- 불가역적, 457
- 불기약, 355
- 불변, 135
- 불변분포, 290, 332, 336, 381, 388, 451, 608
- 불변확률, 390
- 불변확률분포, 390
- 불편추정량, 68, 287, 288
- 블록단위MH샘플러, 614
- 블록단위샘플링, 568
- 블록화, 614
- 비선형연립방정식, 207

- 비율, 29
- 비음정치행렬, 44
- 비정칙, 167, 401, 450
- 비주기성, 556
- 비주기적, 359, 382, 386, 457
- 비중심 $F$ 확률분포, 39
- 비중심 $t$ 확률분포, 38
- 비중심모수, 37-39, 69
- 비중심카이제곱확률분포, 37, 68
  
- 사전CVaR, 101
- 사전가정, 123, 144
- 사전오즈비, 189
- 사전자유도, 131
- 사전정밀도, 123, 144
- 사전평균, 123
- 사전확률밀도함수, 4
- 사전확률분포, 3, 4
- 사후CVaR, 102
- 사후기대값, 5, 72
- 사후디폴트확률, 118
- 사후모드, 179
- 사후분산공분산행렬, 73
- 사후예측확률밀도함수, 191
- 사후오즈비, 189
- 사후자유도, 131
- 사후적률, 3
- 사후정밀도, 123, 144
- 사후증양값, 179
- 사후평균, 11, 178
- 사후평균벡터, 72
- 사후확률밀도함수, 5
- 사후확률분포, 3, 5, 462
- 상불완비감마함수, 30
- 상세균형방정식, 608
- 상태공간, 290
- 상태공간모형, 297, 552, 614
- 상태공간표현, 295
- 상태방정식, 297, 552
- 상태벡터, 292, 297
- 상태변수, 614
- 상태추이도, 299
- 상태추이확률행렬, 297
- 상태확률, 317, 320
- 상태확률벡터, 292, 317
- 상호도달가능, 347, 355, 359
- 상호도달가능성, 354
- 샘플수, 410
- 서로소, 355
- 서브프라임모기지론, 107
- 선택가능, 71
- 선택불가능, 71
- 선형선형손실함수, 185
- 선형회귀모형, 111, 112, 193, 228, 246, 481, 486, 489, 527, 664
- 성공횟수, 303
- 성분단위MH샘플러, 579, 614
- 성분단위샘플링, 568
- 세미공액, 175
- 센서드Poisson회귀분석, 514
- 센서음이향회귀분석, 514
- 소버린리스크, 107
- 손실함수, 71, 92, 94, 148, 184

- 쇄신오차벡터, 297
- 쇄신항, 294, 297
- 수렴성, 451, 613
- 순서통계량, 614
- 순열, 84, 576
- 스위스도로관리청, 206
- 스페이스셔틀 챌린저, 622
- 스펙트럴밀도함수, 613, 615
- 스펙트럴함수, 223
- 스펙트럼, 222, 223
- 스펙트럼검정, 222
- 슬라이스샘플러, 221, 223
- 시가감손, 103
- 시간가역성, 389, 412
- 시간동질성, 330
- 시간동질적, 291
- 시간적으로 가역, 389, 394, 607
- 시계열데이터, 614
- 시계열산점도, 612, 619, 624, 631, 633
- 시뮬레이션터어널링, 585
- 시불변성, 330
- 시장리스크, 91
- 시차, 613
- 신뢰수준, 186
- 신용구간, 614
- 신용리스크, 91, 105
- 신용리스크분석, 107
- 신용집합, 186
- 신용카드, 473
- 씨닝, 222, 223
- 안전이자율, 109
- 안정상태, 388
- 안정성, 332, 613
- 안정성조건, 632, 637
- 안정적, 335, 336, 381
- 양반정치, 252
- 양재귀적, 378, 381, 386, 388, 457
- 양적조건, 449, 455, 457
- 양정치행렬, 45, 67, 295
- 에르고딕, 378, 381, 386-388, 409
- 에르고딕성, 613
- 에르고딕정리, 457
- 역Markov 체인, 388
- 역Wishart 확률분포, 146, 147
- 역Wishart 확률행렬, 61
- 역감마확률분포, 41, 125, 128, 168
- 역정규확률분포, 40
- 연속형 상태공간, 408
- 연속형 일양확률분포, 27
- 열목욕알고리즘, 411
- 영재귀적, 378, 381
- 예측분포, 3
- 예측확률, 16, 18-20
- 오목함수, 141
- 오즈비, 108, 506
- 오퍼레이셔널리스크, 91
- 완만한 혼합, 457
- 완전하다, 71
- 외생변수, 637
- 우고유벡터, 315
- 우도함수, 3, 4, 66, 665
- 우도함수곡면, 400

- 우연변동, 297
- 위키피디아, 164
- 위험함수, 71
- 유량, 473
- 유전자의 연결성, 603
- 유한Markov체인, 387
- 음이항확률분포, 25
- 이동평균원도우, 221
- 이산시간형 Markov 확률과정, 290
- 이산형 일양확률분포, 21
- 이자율스프레드, 108, 112
- 이차손실함수, 184
- 이탈통계량, 210
- 이항로지스틱회귀모형, 630
- 이항정리, 26
- 이항확률분포, 6, 20, 163, 315, 616
- 일반선형모형, 503
- 일반지수확률분포, 31
- 일반화Pareto확률분포, 35
- 일반화극한값확률분포, 36
- 일반화분산, 256
- 일반화선형모형, 201, 202, 204, 205, 214
- 일반화최소제곱추정량, 114, 549
- 일반화최소제곱추정법, 527
- 일양 에르고딕성, 613
- 일양확률분포, 469, 478, 563, 566, 574
- 일치성, 386
- 자기공분산, 223
- 자기상관계수, 614
- 자기상관성, 222, 568, 614
- 자기상관함수, 498, 612
- 자기회귀과정, 632, 667
- 자기회귀모형, 663
- 자기회귀샘플러, 603
- 자기회귀생성밀도함수, 602
- 자기회귀오차모형, 662, 667
- 자연공역사전확률분포, 101
- 자연공역사전확률분포족, 123
- 자유도, 273
- 잔차제곱교차곱합행렬, 256
- 잔차제곱합, 273
- 잠재변수, 513, 517
- 재고관리, 306
- 재귀상태, 346, 349
- 재귀성, 363, 366, 367
- 재귀적, 353, 354, 363, 370, 377, 457
- 저량, 473
- 저빈도, 222
- 적률생성함수, 29, 42
- 적률추정법, 182
- 적합도검정, 397
- 절단다변량정규확률분포, 663
- 절단로지스틱확률분포, 507, 510
- 절단정규확률분포, 95, 493, 502, 517, 635
- 절단확률밀도함수, 94
- 절삭이 있는 회귀모형, 511
- 점근관계, 613
- 점근분산, 386
- 정규감마확률분포, 196, 199
- 정규난벡터, 416
- 정규난수, 517
- 정규방정식, 253

- 정규성조건, 210
- 정규화상수, 5, 217, 610
- 정규확률벡터, 44, 65
- 정규확률분포, 28, 121, 125, 128
- 정규확률분포족, 249
- 정밀도, 195, 483
- 정밀도행렬, 536
- 정보집합, 298
- 정상분포, 332
- 정상성, 221
- 정상적, 223, 291, 388
- 정준형, 338
- 정칙, 167
- 제로원손실함수, 184
- 제안분포, 395, 555, 609
- 제안분포의 조건부확률밀도함수, 614
- 제안스텝크기, 628
- 제안행렬, 616
- 조건부 추이확률, 609
- 조건부공액, 172
- 조건부공액사전확률분포, 172
- 조건부독립, 86
- 조건부리스크가치, 91
- 조건부목표확률밀도함수, 609
- 조건부제안확률밀도함수, 609, 610
- 조건부확률밀도함수, 53
- 조건수, 259
- 조달이자율, 108
- 좌고유벡터, 332
- 주기, 359
- 주기적, 359
- 주변우도, 615
- 주변확률분포, 49, 221
- 중심극한정리, 121, 386, 414
- 중심카이제곱확률분포, 68
- 지수형 확률분포, 162
- 지수확률분포, 28, 157, 159, 163, 430
- 지지대, 290
- 직교성, 65
- 직교행렬, 51, 65
- 진짜디폴트확률, 118
- 채택기각단계, 601
- 채택기각법, 626
- 채택률, 568, 614
- 척도화역카이제곱확률분포, 42, 129, 131, 178
- 체계적쓰기Gibbs샘플러, 412, 448, 451
- 초구, 134
- 초구의 부피, 136
- 초기분포, 292
- 초기상태, 388
- 초기상태벡터, 332
- 초기하확률변수, 26
- 초모수, 229, 235
- 초모수벡터, 224, 227
- 초사전확률분포, 227
- 총변동량, 256
- 총변분거리, 409
- 총제곱교차곱합행렬, 256
- 최고사후확률밀도구간, 187
- 최대사후확률추정량, 179
- 최대엔트로피확률분포, 121

- 최량선형불편추정량, 527  
 최빈값, 473, 481  
 최소스프레드, 108  
 최소완전하다, 71  
 최소제공추정량, 254  
 최소제공추정법, 211, 214, 252, 253, 286, 527, 529  
 최우추정량, 68, 138, 180, 254  
 최우추정법, 207, 214, 503, 506, 514  
 최초방문시점, 336, 344, 345, 356  
 추세계수, 298  
 추세벡터, 295  
 추이율, 355, 377  
 추이핵, 290, 409  
 추이확률, 291  
 추이확률행렬, 292  
 축소가능, 355  
 축소불가능, 355  
 출력방정식, 297  
 충분통계량, 7, 122, 245  
 카이제곱확률분포, 31, 59  
 커널스므딩밀도함수, 223  
 컨트리리스크, 107  
  
 크기모수, 29  
 타우거리, 576  
 탄광재해, 478  
 태양흑점, 223  
 토빗모형, 492, 511, 516, 520  
 퇴화확률변수, 44  
 트라이감마함수, 181  
 특성함수, 29  
 특이, 44  
 특이값분해, 416  
 특이확률변수, 44  
 파산확률, 116  
 편의추정량, 68  
 평균곡선, 214  
 평균벡터, 45  
 평균손실, 71  
 평균위험, 75  
 평균재귀시간, 378  
 평형분포, 332  
 표준Brown운동, 298  
 표준정규확률분포, 28  
 프로빗모형, 204, 492, 502  
 피어리어도그램, 222  
 하불완비감마함수, 30  
 하이퍼모수벡터, 100  
 한국은행의 경제통계시스템, 473  
 형태모수, 29  
 혼외정사, 513  
 혼합정규확률분포, 560  
 확률미분방정식, 297  
 확률보행, 592  
 확률보행Metropolis샘플러, 592  
 확률보행과정, 315  
 확률행렬, 291  
 확산계수, 298  
 회귀계수벡터, 297  
 회귀모형, 107, 112  
 회사채스프레드, 109, 120  
 회사채이자율, 109

회수율, 108, 109

후보생성분포, 609

후보점, 555, 602

후보점의 채택률, 613

후진연산자, 663

후향확률과정, 388

흡수상태, 346, 347, 349

흡수적, 354

히스토그램, 397



## 저자 소개

---

최 병 선 (崔秉善)

서울대학교 수학과 졸업 (이학사)

미국 Stanford대학교 대학원 졸업 (경제학 석사, 통계학 석사,

Ph. D. in Statistics <major> and Economics <minor>)

연세대학교 상경대학 교수 역임

현재 서울대학교 경제학부 교수 (재무경제학 담당)

E-mail bschoi12@snu.ac.kr

## 금융공학 VIII: Bayesian Methods for Finance and Economics

---

2018년 11월 20일 제1판 1쇄 인쇄

2018년 11월 30일 제1판 1쇄 발행

저 자 최 병 선

발행처 서울대학교 금융경제연구원

서울특별시 관악구 관악로 1 우편번호 08826

전 화 02-880-5424

팩 스 02-888-4231

E-mail sirfe@snu.ac.kr

<http://www.sirfe.com>

---

ISBN 979-11-952542-5-5 93320

비매품

이 도서의 국립중앙도서관 출판예정도서목록(CIP)은 서지정보유통지원시스템 홈페이지(<http://seoji.nl.go.kr>)와 국가자료종합목록시스템(<http://www.nl.go.kr/kolisnet>)에서 이용하실 수 있습니다. (CIP제어번호 : CIP2018034655)



## 최병선 교수 Creative Commons Book List (<http://s-space.snu.ac.kr/handle/10371/314>) <sup>1</sup>

1. (제2판) SAS를 이용한 현대통계학 (이성백교수와 공저)  
<http://s-space.snu.ac.kr/handle/10371/94393>
2. 행렬의 대각화와 Jordan표준형 (이성백교수와 공저)  
<http://s-space.snu.ac.kr/handle/10371/94394>
3. Fourier 해석 입문  
<http://s-space.snu.ac.kr/handle/10371/94413>
4. Lebesgue 적분 입문  
<http://s-space.snu.ac.kr/handle/10371/94414>
5. Wavelet 해석  
<http://s-space.snu.ac.kr/handle/10371/94415>
6. 회귀분석(상)  
<http://s-space.snu.ac.kr/handle/10371/94433>
7. 다변량시계열분석  
<http://s-space.snu.ac.kr/handle/10371/94434>
8. 이산형 재무모형의 수리적 배경  
<http://s-space.snu.ac.kr/handle/10371/94455>
9. 회귀분석(하)  
<http://s-space.snu.ac.kr/handle/10371/94456>
10. 단변량시계열분석  
<http://s-space.snu.ac.kr/handle/10371/94457>
11. 금융공학 IV: Monte Carlo Methods for Finance and Economics  
<http://s-space.snu.ac.kr/handle/10371/98466>
12. 금융공학 V: Introduction to Financial Engineering with R  
<http://s-space.snu.ac.kr/handle/10371/99003>
13. SAS/IML 입문  
<http://s-space.snu.ac.kr/handle/10371/116955>
14. 금융공학 VII: Scientific Computing For Finance and Economics  
<http://s-space.snu.ac.kr/handle/10371/139655>

---

<sup>1</sup>현대문화사(hun1249@hanmail.net)에서 이 책들의 POD판을 구하실 수 있습니다.

15. 금융공학 III: Introduction to Financial Engineering

<http://s-space.snu.ac.kr/handle/10371/142795>