

Uma Abordagem Bayesiana à Determinação de Modelos

Dissertação para a obtenção do Grau de Doutor
em Motricidade Humana na especialidade de Métodos Matemáticos



Universidade Técnica de Lisboa

Faculdade de Motricidade Humana

2005

Orientador: Doutor Luís Camilo do Canto de Loura

Co-orientador: Doutora Maria Antónia da Conceição Abrantes Amaral Turkman

Júri

Presidente: Reitor da Universidade Técnica de Lisboa

Vogais: Doutora Maria Antónia da Conceição Abrantes Amaral Turkman

Doutora Maria Manuela Costa Neves Figueiredo

Doutor Luís Camilo do Canto de Loura

Doutora Isabel Maria Simões Pereira

Doutor Paulo José de Jesus Soares



Uma Abordagem Bayesiana à Determinação de Modelos

Dissertação para a obtenção do Grau de Doutor
em Motricidade Humana na especialidade de Métodos Matemáticos



Universidade Técnica de Lisboa

Faculdade de Motricidade Humana

2005

Orientador: Doutor Luís Camilo do Canto de Loura

Co-orientador: Doutora Maria Antónia da Conceição Abrantes Amaral Turkman

Júri

Presidente: Reitor da Universidade Técnica de Lisboa

Vogais: Doutora Maria Antónia da Conceição Abrantes Amaral Turkman

Doutora Maria Manuela Costa Neves Figueiredo

Doutor Luís Camilo do Canto de Loura

Doutora Isabel Maria Simões Pereira

Doutor Paulo José de Jesus Soares



“Modelling in science remains, partly at least, an art. Some principles do exist, however, to guide the modeller. A first, though at first sight, not a very helpful principle, is that all models are wrong; some, though, are more useful than others and we should seek those. At the same time we must recognize that eternal truth is not within our grasp. A second principle is not to fall in love with one model to the exclusion of alternatives. Data will often point with almost equal emphasis at several possible models and it is important that the statisticians recognize and accept this. A third principle recommends thorough checks on the fit of a model to the data, for example by using residuals and other statistics derived from the fit to look for outlying observations and so on. Such diagnostic procedures are not yet fully formalized, and perhaps never will be. Some imagination or introspection is required here in order to determine the aspects of the model that are most important and most suspect.”

McCullagh e Nelder (1989)

Resumo

A determinação de modelos, que engloba a selecção e a validação, é um dos problemas importantes da estatística. A abordagem bayesiana deste problema é o assunto tratado nesta dissertação. Efectua-se uma revisão de alguns métodos de selecção e validação de modelos, propõe-se métodos de pesquisa do “melhor” subconjunto de covariáveis em modelos de regressão e, através de estudos de simulação, compara-se a performance de alguns critérios de selecção de modelos.

Durante muitos anos, o factor Bayes foi a ferramenta bayesiana preferida para a comparação de modelos, pelo facto de possibilitar a inclusão da informação *a priori* existente sobre o vector de parâmetros do modelo. Compara-se o factor Bayes com diversos critérios de verosimilhança penalizada, entre eles o critério de informação bayesiano, o critério de informação bayesiano estimado e o critério de informação da *deviance*. Como alternativa à selecção de modelos sugere-se, em certos casos, a utilização da média ponderada de modelos bayesianos.

Nos problemas de selecção de covariáveis em modelos de regressão, o número de modelos a comparar é, de um modo geral, elevado. Neste caso particular de selecção de modelos, não sendo viável a pesquisa exaustiva, propõe-se dois métodos bayesianos de pesquisa do “melhor” modelo: os métodos de selecção progressiva e de eliminação regressiva de covariáveis via medidas de discrepância. Dois exemplos, um de regressão linear normal e outro de regressão gama, são usados para ilustrar estes métodos de selecção de covariáveis.

Relativamente à validação de modelos, destaca-se a utilização do *p-value* preditivo *a posteriori* e, em particular, do *p-value* de discrepância, como medidas de adequabilidade do modelo.

Palavras-chave: estatística bayesiana, selecção de modelos, validação de modelos, selecção de covariáveis, estudos de simulação.

Abstract

Model determination, which refers to model selection and validation, is an important statistical problem. The Bayesian approach to this problem is the purpose of this thesis. A review of some methods of model selection and validation is made. Search methods for selection of the best subset of independent variables in regression models are proposed. Simulation studies are made to compare the performance of some model selection criteria.

For many years, Bayes factor was the favourite Bayesian tool to compare models, since it made possible the inclusion of existing prior information on the parameter vector of the model. Bayes factor is compared with several penalized likelihood criteria, namely, the Bayesian information criterion, the estimated Bayesian information criterion and the deviance information criterion. In some cases, instead of model selection, the use of Bayesian model averaging is suggested.

In variable selection problems the number of models to compare is large, even if the number of independent variables is moderate. In this particular case of model selection, since an exhaustive search is not feasible, two Bayesian methods of search of the best model are proposed: the forward variable selection and the backward variable elimination via discrepancy measures. Normal linear regression and gamma regression examples are used to illustrate these methods of variable selection.

As for model validation, the use of posterior predictive p-value and discrepancy p-value are highlighted as goodness of fit measures.

Keywords: Bayesian statistic, model selection, model validation, variable selection, simulation studies.

Agradecimentos

À Professora Doutora Maria Antónia Amaral Turkman, pela orientação, incentivo, paciência e amizade inextinguíveis, que tornaram possível a realização desta dissertação de doutoramento.

Ao Professor Doutor Luís Camilo do Canto de Loura, cujas críticas, correcções e comentários foram imprescindíveis para a elaboração final deste texto. O auxílio nas “questões burocráticas” do doutoramento foi também bastante importante.

À Ana Isabel Carita e à Paula Marta Bruno, por estarem sempre disponíveis para ajudar em tudo o que lhes era solicitado, tanto na revisão do texto, como na resolução dos pequenos grandes problemas que foram aparecendo no decurso deste trabalho.

Ao Carlos Ferreira, pela eficiência e rapidez na resolução de todos os contratemplos com computadores, impressoras, instalação de software, etc., e por algumas sugestões bastante úteis para a programação em Fortran.

À Luzia Gonçalves, a quem recorri sempre que tinha dúvidas de LaTeX, e que teve paciência para rever este texto. À Dulce Gomes, pela sua imprescindível ajuda com a programação em Fortran.

Ao Carlos Canto e Castro, que disponibilizou o editor CantoSol TeX e esteve sempre pronto para esclarecer as dúvidas e resolver os problemas que foram aparecendo.

Ao Carlos Andrade e à Ana Dinis, a preciosa colaboração na revisão do *Abstract*.

Um agradecimento muito especial aos meus pais, familiares e amigos pelo incentivo, paciência e carinho ao longo destes muitos meses de trabalho.

À Faculdade de Motricidade Humana pelas condições de trabalho proporcionadas, que possibilitaram a elaboração desta dissertação de doutoramento. Ao Fundo Social Europeu e ao Estado Português que co-financiou esta acção, no âmbito do III Quadro Comunitário de Apoio, através do programa PRODEP III.

Índice

Abreviaturas	xi
Lista de Figuras	xiii
Lista de Tabelas	xv
Nota Introdutória	xxiii
1 Fundamentos de Estatística Bayesiana	1
1.1 Introdução	1
1.2 Modelos hierárquicos bayesianos	4
1.3 Distribuições <i>a priori</i>	5
1.4 Inferência paramétrica	8
1.4.1 Estimação pontual	8
1.4.2 Estimação por regiões	10
1.4.3 Testes de hipóteses	10
1.5 Inferência preditiva	11
1.6 Decisão	12
1.7 Método de Monte Carlo	13
1.7.1 Monte Carlo directo	14
1.7.2 Monte Carlo com amostragem de importância	14
1.7.3 Monte Carlo via cadeias de Markov	15
2 Critérios de Selecção de Modelos	19
2.1 Introdução	19
2.2 Factor Bayes	21
2.3 Resíduos bayesianos	25

2.3.1	Ordenadas preditivas condicionais	27
2.4	Critérios de verosimilhança penalizada	28
2.4.1	Critério de informação de Akaike	28
2.4.2	Critério de informação bayesiano	29
2.4.3	Critério de informação bayesiano estimado	30
2.4.4	Critério de informação da <i>deviance</i>	31
2.4.5	Log-verosimilhança <i>a posteriori</i>	34
2.4.6	Relação entre DIC, \widehat{BIC} e LLP	34
2.5	Aplicação a dados de contagem	35
2.5.1	Modelo beta/binomial negativo	36
2.5.2	Modelo gama/Poisson	39
2.5.3	Critérios de selecção dos modelos	43
2.6	Média ponderada de modelos bayesianos	50
3	Seleccção de Covariáveis em Modelos de Regressão	57
3.1	Introdução	57
3.2	Métodos de selecção via medidas de discrepância	60
3.2.1	Seleccção informal por pesquisa manual	60
3.2.2	Seleccção progressiva de covariáveis	60
3.2.3	Eliminação regressiva de covariáveis	61
3.3	Métodos de selecção simultânea via amostrador Gibbs	65
3.3.1	Método de Carlin e Chib	66
3.3.2	Seleccção de covariáveis por pesquisa estocástica	68
3.3.3	Método de Kuo e Mallick	69
3.3.4	Método Gibbs de selecção de covariáveis	69
3.4	Comparação dos métodos de Carlin e Chib, SSVS, de Kuo e Mallick, e GVS	72
3.5	Exemplo de regressão linear normal	73
3.5.1	Introdução	73
3.5.2	Dados da massa gorda	74
3.5.3	Métodos de selecção via medidas de discrepância	78
3.5.4	Método Gibbs de selecção de covariáveis	87
3.5.5	Comparação de resultados	89
3.6	Exemplo de regressão gama	91
3.6.1	Introdução	91

3.6.2	Dados da homocisteinemia	92
3.6.3	Métodos de selecção via medidas de discrepância	93
3.6.4	Método Gibbs de selecção de covariáveis	101
3.6.5	Comparação de resultados	104
3.7	Comentários	105
4	Estudos de Simulação para Comparação de Critérios	107
4.1	Introdução	107
4.2	Dados de contagem	108
4.2.1	Planeamento da simulação	108
4.2.2	Performance dos critérios de selecção de modelos	110
4.2.3	Valores das medidas de selecção de modelos	113
4.3	Dados de regressão linear normal	120
4.3.1	Planeamento da simulação	120
4.3.2	Modelo de regressão linear normal	122
4.3.3	Performance dos critérios de selecção de modelos	124
4.3.4	Performance do método GVS	130
4.3.5	Valores das medidas de selecção de modelos	135
5	Validação de Modelos	143
5.1	Introdução	143
5.2	<i>p-values</i>	145
5.3	<i>p-value</i> preditivo <i>a posteriori</i>	151
5.3.1	<i>p-value</i> de discrepância	153
5.4	Aplicações	155
5.4.1	Modelos beta/binomial negativo e gama/Poisson — dados de contagem	155
5.4.2	Regressão linear normal — dados da massa gorda	158
5.4.3	Regressão gama — dados da homocisteinemia	162
6	Discussão e Trabalho Futuro	167
Anexos		
A	Distribuições de Probabilidade	171

B	Determinação de Modelos via Métodos MCMC	175
B.1	Estimação de distribuições preditivas	175
B.2	Estimação do valor médio preditivo e da variância preditiva	177
C	Conjuntos de Dados	179
D	Códigos do WinBUGS	187
D.1	Dados da massa gorda	187
D.1.1	Método SIMD	187
D.1.2	Método FVS	190
D.1.3	Método BVE	193
D.1.4	Método GVS	196
D.1.5	Método de Carlin e Chib	198
D.2	Dados da homocisteinemia	206
D.2.1	Método SIMD	206
D.2.2	Método FVS	208
D.2.3	Método BVE	209
D.2.4	Método GVS	211
D.3	Dados simulados de regressão linear normal — Método GVS	215
E	Resultados dos Estudos de Simulação	219
E.1	Dados de contagem	219
E.2	Dados de regressão linear normal	226
	Referências Bibliográficas	245

Abreviaturas

acr.	acrécimo
AIC	critério de informação de Akaike
BF	factor Bayes
BIC	critério de informação bayesiano
$\widehat{\text{BIC}}$	critério de informação bayesiano estimado
BMA	média ponderada de modelos bayesianos
BVE	eliminação regressiva de covariáveis
cov.	covariável
CPO	ordenada preditiva condicional
decr.	decrécimo
d.p.	desvio padrão
DIC	critério de informação da <i>deviance</i>
erro MC	erro de Monte Carlo
FVS	selecção progressiva de covariáveis
GVS	método Gibbs de selecção de covariáveis
HBP	homocisteinemia basal
HSP	homocisteinemia após sobrecarga com metionina oral
iid	independentes e identicamente distribuídas
it.	iteração
LLP	log-verosimilhança <i>a posteriori</i>
máx.	máximo
MCMC	Monte Carlo via cadeias de Markov
méd.	média
mín.	mínimo
p_{dis}	<i>p-value</i> de discrepância
p_{post}	<i>p-value</i> preditivo <i>a posteriori</i>
p_{prior}	<i>p-value</i> preditivo <i>a priori</i>
SC	critério de Schwarz
SIMD	selecção informal por pesquisa manual via medidas de discrepância
SlnCPO	soma dos logaritmos das ordenadas preditivas condicionais
SSVS	selecção de covariáveis por pesquisa estocástica
var.	variância

Lista de Figuras

2.1	Distribuição <i>a priori</i> , verosimilhanças e distribuições <i>a posteriori</i> do parâmetro θ_1 do modelo m_1 , obtidas com as amostras x_u e x_v	48
2.2	Distribuição <i>a priori</i> , verosimilhanças e distribuições <i>a posteriori</i> do parâmetro θ_2 do modelo m_2 , obtidas com as amostras x_u e x_v	49
3.1	Representação gráfica dos métodos de selecção de covariáveis SSVS, de Kuo e Mallick, e GVS	72
3.2	Caixas de bigodes das variáveis do exemplo da massa gorda	76
3.3	Diagramas de dispersão da percentagem de massa gorda <i>versus</i> cada uma das covariáveis	77
3.4	Gráficos das densidades marginais <i>a posteriori</i> dos coeficientes do modelo de regressão linear com 13 covariáveis, para o exemplo da massa gorda . . .	80
3.5	Gráficos das densidades marginais <i>a posteriori</i> dos coeficientes do modelo de regressão linear com 8 covariáveis, para o exemplo da massa gorda . . .	81
3.6	Gráficos das densidades marginais <i>a posteriori</i> dos coeficientes do modelo de regressão gama com 13 covariáveis, para o exemplo da homocisteinemia .	94
3.7	Gráficos das densidades marginais <i>a posteriori</i> dos coeficientes do modelo de regressão gama com 9 covariáveis, para o exemplo da homocisteinemia .	96
4.1	Percentagens de classificação correcta (no caso 1), das amostras geradas do modelo m_1 , para $\theta_1 = 0.534$ e para $n = 30$	111
4.2	Percentagens de classificação correcta (no caso 1), das amostras geradas do modelo m_2 , para $\theta_2 = 1.128$ e para $n = 30$	113
4.3	Médias/ $n \pm$ desvios padrão/ \sqrt{n} dos valores do DIC (no caso 1), referentes ao ajustamento dos modelos m_1 e m_2 a dados gerados de m_1 com $\theta_1 = 0.534$, <i>versus</i> n	115
4.4	Médias dos valores do DIC (no caso 1), obtidas com o ajustamento de m_1 e m_2 a dados gerados do modelo m_1 , <i>versus</i> θ_1 , para diferentes valores de n	116
4.5	Médias dos valores do DIC (no caso 1), obtidas com o ajustamento de m_1 e m_2 a dados gerados do modelo m_1 , <i>versus</i> n , para $\theta_1 = 0.073$ e para $\theta_1 = 0.421$	116

4.6	Médias/ $n \pm$ desvios padrão/ \sqrt{n} dos valores do DIC (no caso 1), referentes ao ajustamento dos modelos m_1 e m_2 a dados gerados de m_2 com $\theta_2 = 1.128$, <i>versus</i> n	117
4.7	Médias dos valores do DIC (no caso 1), obtidas com o ajustamento de m_1 e m_2 a dados gerados do modelo m_2 , <i>versus</i> θ_2 , para diferentes valores de n	118
4.8	Médias dos valores do DIC (no caso 1), obtidas com o ajustamento de m_1 e m_2 a dados gerados do modelo m_2 , <i>versus</i> n , para $\theta_2 = 0.632$ e para $\theta_2 = 2.741$	119
4.9	Caixas de bigodes dos valores do DIC, obtidos com os dados gerados do modelo m_8 , para os 16 modelos ajustados, nos casos 1 a 6	136
5.1	Valor observado de $t(X)$, comparado com os 10000 valores simulados da distribuição preditiva <i>a posteriori</i> de $t(X^{rep})$, para as amostras x_2 , x_5 e x_7	152
5.2	Esquema da simulação dos valores da distribuição preditiva <i>a posteriori</i> de $t(X^{rep}, \theta)$	155
5.3	Ordenadas preditivas condicionais <i>versus</i> números das observações, obtidas com as amostras x_u e x_v , para os modelos m_1 e m_2	156
5.4	Diagramas de dispersão de $t(x^{rep,j}, \theta^{rep,j})$ <i>versus</i> $t(x_{obs}, \theta^{rep,j})$, baseados em 10000 amostras simuladas da distribuição <i>a posteriori</i> conjunta de X^{rep} e θ , obtidas com as amostras x_u e x_v , para os modelos m_1 e m_2	157
5.5	Logaritmos das ordenadas preditivas condicionais <i>versus</i> números das observações, para os modelos m_1 , m_3 e m_4 do exemplo da massa gorda	158
5.6	Histogramas dos resíduos bayesianos, para os modelos m_1 , m_3 e m_4 do exemplo da massa gorda, com sobreposição da densidade normal ajustada	159
5.7	Resíduos bayesianos padronizados <i>versus</i> valores preditos das observações, para os modelos m_1 , m_3 e m_4 do exemplo da massa gorda	159
5.8	Valores simulados das quantidades de teste <i>versus</i> valores observados, para o modelo m_1 do exemplo da massa gorda	161
5.9	Logaritmos das ordenadas preditivas condicionais <i>versus</i> números das observações, para os modelos m_1 , m_3 e m_6 do exemplo da homocisteinemia	162
5.10	Resíduos bayesianos padronizados <i>versus</i> logaritmos dos valores preditos das observações, para os modelos m_1 , m_3 e m_6 do exemplo da homocisteinemia	162
5.11	Valores simulados das quantidades de teste <i>versus</i> valores observados, para o modelo m_1 do exemplo da homocisteinemia	165
5.12	Desvios padrão de Y_i <i>versus</i> resíduos bayesianos não padronizados, para o modelo m_1 do exemplo da homocisteinemia	165

Lista de Tabelas

1.1	Interpretação dos valores do factor Bayes	11
2.1	Valores dos critérios de selecção de modelos, obtidos com as amostras x_u e x_v , para os modelos m_1 e m_2	49
2.2	Exemplo ilustrativo da média ponderada de modelos bayesianos	54
3.1	Conjunto de dados, adaptado de Ryan (1997), usado para comparar a performance dos métodos FVS e BVE	63
3.2	Valores das medidas de discrepância, obtidos com os dados da tabela 3.1, para todos os modelos de regressão linear possíveis	64
3.3	Estatísticas sumárias das variáveis do exemplo da massa gorda	75
3.4	<i>Output</i> do SPSS com as estimativas dos coeficientes do modelo de regressão linear com 13 covariáveis, para o exemplo da massa gorda	78
3.5	<i>Output</i> do SPSS com as estimativas dos coeficientes do modelo de regressão linear com 4 covariáveis, para o exemplo da massa gorda	78
3.6	<i>Output</i> do WinBUGS com as estimativas dos parâmetros do modelo de regressão linear com 13 covariáveis, para o exemplo da massa gorda	79
3.7	Resultados do método SIMD, para o exemplo da massa gorda	81
3.8	<i>Output</i> do WinBUGS com as estimativas dos parâmetros do modelo de regressão linear com 8 covariáveis, para o exemplo da massa gorda	81
3.9	Resultados do método FVS via DIC, para o exemplo da massa gorda	82
3.10	Resultados do método FVS via \widehat{BIC} , para o exemplo da massa gorda	83
3.11	Resultados do método FVS via LLP, para o exemplo da massa gorda	84
3.12	Resultados do método FVS via SlnCPO, para o exemplo da massa gorda	84
3.13	Resultados do método BVE via DIC, para o exemplo da massa gorda	85
3.14	Resultados do método BVE via \widehat{BIC} , para o exemplo da massa gorda	86
3.15	Resultados do método BVE via LLP, para o exemplo da massa gorda	86
3.16	Resultados do método BVE via SlnCPO, para o exemplo da massa gorda	87

3.17	Estimativas das probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo, para o exemplo da massa gorda	88
3.18	Estimativas das probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo, para o exemplo da massa gorda, quando se considera um subconjunto de 8 covariáveis	89
3.19	Dez modelos com maior probabilidade <i>a posteriori</i> , para o exemplo da massa gorda, quando se considera um subconjunto de 8 covariáveis	89
3.20	Covariáveis incluídas nos modelos de regressão linear seleccionados por cada um dos métodos, no caso do exemplo da massa gorda	90
3.21	Valores das medidas de discrepância dos modelos de regressão linear m_1 a m_5 , para o exemplo da massa gorda	90
3.22	Estatísticas sumárias das variáveis do exemplo da homocisteinemia	92
3.23	<i>Output</i> do WinBUGS com as estimativas dos parâmetros do modelo de regressão gama com 13 covariáveis, para o exemplo da homocisteinemia . . .	94
3.24	Resultados do método SIMD, para o exemplo da homocisteinemia	95
3.25	<i>Output</i> do WinBUGS com as estimativas dos parâmetros do modelo de regressão gama com 9 covariáveis, para o exemplo da homocisteinemia . . .	95
3.26	Resultados do método FVS via DIC, para o exemplo da homocisteinemia . . .	97
3.27	Resultados do método FVS via \widehat{BIC} , para o exemplo da homocisteinemia . . .	97
3.28	Resultados do método FVS via LLP, para o exemplo da homocisteinemia . . .	98
3.29	Resultados do método FVS via SlnCPO, para o exemplo da homocisteinemia . . .	98
3.30	Resultados do método BVE via DIC, para o exemplo da homocisteinemia . . .	99
3.31	Resultados do método BVE via \widehat{BIC} , para o exemplo da homocisteinemia . . .	100
3.32	Resultados do método BVE via LLP, para o exemplo da homocisteinemia . . .	100
3.33	Resultados do método BVE via SlnCPO, para o exemplo da homocisteinemia . . .	101
3.34	Estimativas das probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo, para o exemplo da homocisteinemia	102
3.35	Estimativas das probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo, para o exemplo da homocisteinemia, quando se excluem algumas covariáveis	102
3.36	Dez modelos com maior probabilidade <i>a posteriori</i> obtidos pelo método GVS sem restrição no espaço dos modelos, para o exemplo da homocisteinemia, quando se considera um subconjunto de 8 covariáveis . . .	103
3.37	Dez modelos com maior probabilidade <i>a posteriori</i> obtidos pelo método GVS com restrição no espaço dos modelos, para o exemplo da homocisteinemia, quando se considera um subconjunto de 9 covariáveis . . .	103
3.38	Covariáveis incluídas nos modelos de regressão gama seleccionados por cada um dos métodos, no caso do exemplo da homocisteinemia	104

3.39	Valores das medidas de discrepância dos modelos de regressão gama m_1 a m_7 , para o exemplo da homocisteinemia	104
4.1	Percentagem de classificação correcta das amostras geradas de m_1 , para o caso 1	111
4.2	Percentagem de classificação correcta das amostras geradas de m_2 , para o caso 1	112
4.3	Médias e desvios padrão do DIC e do $\widehat{\text{BIC}}$, obtidos com o ajustamento dos modelos m_1 e m_2 , a dados gerados de m_1 com $\theta_1 = 0.534$, para o caso 1 . .	114
4.4	Médias/ n e desvios padrão/ \sqrt{n} do DIC e do $\widehat{\text{BIC}}$, obtidos com o ajustamento dos modelos m_1 e m_2 , a dados gerados de m_1 com $\theta_1 = 0.534$, para o caso 1	115
4.5	Médias e desvios padrão do DIC e do $\widehat{\text{BIC}}$, obtidos com o ajustamento dos modelos m_1 e m_2 , a dados gerados de m_2 com $\theta_2 = 1.128$, para o caso 1 . .	117
4.6	Médias/ n e desvios padrão/ \sqrt{n} do DIC e do $\widehat{\text{BIC}}$, obtidos com o ajustamento dos modelos m_1 e m_2 , a dados gerados de m_2 com $\theta_2 = 1.128$, para o caso 1	117
4.7	Modelos de regressão linear para um conjunto de 4 covariáveis	120
4.8	Número de vezes que cada modelo é seleccionado pelo DIC, por modelo gerador dos dados e modelo ajustado, para o caso 1	125
4.9	Número de vezes que cada modelo é seleccionado pelo $\widehat{\text{BIC}}$, por modelo gerador dos dados e modelo ajustado, para o caso 1	126
4.10	Número de vezes que cada modelo é seleccionado pelo LLP, por modelo gerador dos dados e modelo ajustado, para o caso 1	127
4.11	Frequências de classificação correcta dos dados gerados dos modelos m_1, \dots, m_{16} , nos casos 1, 5 e 6, para cada uma das medidas de selecção de modelos	129
4.12	Probabilidades <i>a posteriori</i> dos modelos e probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo, para o caso 1	131
4.13	Probabilidades <i>a posteriori</i> dos modelos e probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo, para o caso 2	132
4.14	Probabilidades <i>a posteriori</i> dos modelos e probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo, para o caso 3	133
4.15	Probabilidades <i>a posteriori</i> dos modelos e probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo, para o caso 4	133
4.16	Probabilidades <i>a posteriori</i> dos modelos e probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo, para o caso 5	134
4.17	Probabilidades <i>a posteriori</i> dos modelos e probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo, para o caso 6	135

4.18	Diferenças, por modelo gerador e modelo ajustado para o caso 1, entre as médias do DIC e a menor média do modelo gerador	139
4.19	Diferenças, por modelo gerador e modelo ajustado para o caso 1, entre as médias do $\widehat{\text{BIC}}$ e a menor média do modelo gerador	139
4.20	Diferenças, por modelo gerador e modelo ajustado para o caso 1, entre as médias do LLP e a menor média do modelo gerador	140
4.21	Diferenças, por modelo gerador e modelo ajustado para o caso 5, entre as médias do DIC e a menor média do modelo gerador	141
4.22	Diferenças, por modelo gerador e modelo ajustado para o caso 5, entre as médias do $\widehat{\text{BIC}}$ e a menor média do modelo gerador	141
4.23	Diferenças, por modelo gerador e modelo ajustado para o caso 5, entre as médias do LLP e a menor média do modelo gerador	142
5.1	Valores limite do p_{post} no modelo normal	149
5.2	Cálculo dos p -values preditivos <i>a posteriori</i> no modelo normal, para amostras geradas de várias distribuições	150
5.3	<i>Outputs</i> do WinBUGS com as estimativas dos p -values preditivos <i>a posteriori</i> para amostras geradas de várias distribuições	153
5.4	Ordenadas preditivas condicionais das amostras x_u e x_v , para os modelos m_1 e m_2	156
5.5	Estimativas dos p -values de discrepância, obtidas com as amostras x_u e x_v , para os modelos m_1 e m_2	158
5.6	Estimativas dos p -values de discrepância, obtidas com as quantidades de teste T_3, T_4 e T_6 , para os modelos de regressão linear m_1 a m_5 , do exemplo da massa gorda	160
5.7	<i>Outputs</i> do WinBUGS com as estimativas dos p -values de discrepância obtidas com as quantidades de teste T_3, T_4 e T_6 , para os modelos m_1, m_3 e m_4 , do exemplo da massa gorda	161
5.8	Estimativas dos p -values de discrepância, obtidas com as quantidades de teste T_1, T_2, T_5 e T_6 para os modelos de regressão gama m_1 a m_7 , do exemplo da homocisteinemia	164
5.9	<i>Outputs</i> do WinBUGS com as estimativas dos p -values de discrepância obtidas com as quantidades de teste T_1, T_2, T_5 e T_6 , para os modelos m_1, m_3 e m_6 , do exemplo da homocisteinemia	164
5.10	Estimativas dos p -values de discrepância, obtidas com as quantidades de teste T_1, T_2, T_5 e T_6 para os modelos de regressão gama m_1 a m_7 , do exemplo da homocisteinemia, quando se retiram as observações 35, 41 e 42	166
C.1	Dados da massa gorda	179
C.2	Dados da homocisteinemia	184

D.1	<i>Outputs</i> do WinBUGS com as estimativas dos parâmetros dos modelos de regressão linear m_1 e m_2 , para o exemplo da massa gorda	199
D.2	<i>Outputs</i> do WinBUGS com as estimativas dos parâmetros dos modelos de regressão linear m_3 , m_4 e m_5 , para o exemplo da massa gorda	203
E.1	Percentagem de classificação correcta das amostras geradas de m_1 , para o caso 2	219
E.2	Percentagem de classificação correcta das amostras geradas de m_2 , para o caso 2	220
E.3	Médias e variâncias das medidas de selecção de modelos, no caso 1, referentes às amostras geradas de m_1	221
E.4	Médias e variâncias das medidas de selecção de modelos, no caso 1, referentes às amostras geradas de m_2	222
E.5	Médias e variâncias das medidas de selecção de modelos, no caso 2, referentes às amostras geradas de m_1	223
E.6	Médias e variâncias das medidas de selecção de modelos, no caso 2, referentes às amostras geradas de m_2	225
E.7	Número de vezes que cada modelo é seleccionado pelo DIC, por modelo gerador dos dados e modelo ajustado, para o caso 2	226
E.8	Número de vezes que cada modelo é seleccionado pelo $\widehat{\text{BIC}}$, por modelo gerador dos dados e modelo ajustado, para o caso 2	227
E.9	Número de vezes que cada modelo é seleccionado pelo LLP, por modelo gerador dos dados e modelo ajustado, para o caso 2	227
E.10	Número de vezes que cada modelo é seleccionado pelo DIC, por modelo gerador dos dados e modelo ajustado, para o caso 3	228
E.11	Número de vezes que cada modelo é seleccionado pelo $\widehat{\text{BIC}}$, por modelo gerador dos dados e modelo ajustado, para o caso 3	228
E.12	Número de vezes que cada modelo é seleccionado pelo LLP, por modelo gerador dos dados e modelo ajustado, para o caso 3	229
E.13	Número de vezes que cada modelo é seleccionado pelo DIC, por modelo gerador dos dados e modelo ajustado, para o caso 4	229
E.14	Número de vezes que cada modelo é seleccionado pelo $\widehat{\text{BIC}}$, por modelo gerador dos dados e modelo ajustado, para o caso 4	230
E.15	Número de vezes que cada modelo é seleccionado pelo LLP, por modelo gerador dos dados e modelo ajustado, para o caso 4	230
E.16	Número de vezes que cada modelo é seleccionado pelo DIC, por modelo gerador dos dados e modelo ajustado, para o caso 5	231
E.17	Número de vezes que cada modelo é seleccionado pelo $\widehat{\text{BIC}}$, por modelo gerador dos dados e modelo ajustado, para o caso 5	231

E.18	Número de vezes que cada modelo é seleccionado pelo LLP, por modelo gerador dos dados e modelo ajustado, para o caso 5	232
E.19	Número de vezes que cada modelo é seleccionado pelo DIC, por modelo gerador dos dados e modelo ajustado, para o caso 6	232
E.20	Número de vezes que cada modelo é seleccionado pelo $\widehat{\text{BIC}}$, por modelo gerador dos dados e modelo ajustado, para o caso 6	233
E.21	Número de vezes que cada modelo é seleccionado pelo LLP, por modelo gerador dos dados e modelo ajustado, para o caso 6	233
E.22	Médias do DIC por modelo gerador dos dados e modelo ajustado, para o caso 1	234
E.23	Desvios padrão do DIC por modelo gerador dos dados e modelo ajustado, para o caso 1	234
E.24	Médias do $\widehat{\text{BIC}}$ por modelo gerador dos dados e modelo ajustado, para o caso 1	235
E.25	Médias do LLP por modelo gerador dos dados e modelo ajustado, para o caso 1	235
E.26	Médias do DIC por modelo gerador dos dados e modelo ajustado, para o caso 2	236
E.27	Desvios padrão do DIC por modelo gerador dos dados e modelo ajustado, para o caso 2	236
E.28	Médias do $\widehat{\text{BIC}}$ por modelo gerador dos dados e modelo ajustado, para o caso 2	237
E.29	Médias do LLP por modelo gerador dos dados e modelo ajustado, para o caso 2	237
E.30	Médias do DIC por modelo gerador dos dados e modelo ajustado, para o caso 3	238
E.31	Médias do $\widehat{\text{BIC}}$ por modelo gerador dos dados e modelo ajustado, para o caso 3	238
E.32	Médias do LLP por modelo gerador dos dados e modelo ajustado, para o caso 3	239
E.33	Médias do DIC por modelo gerador dos dados e modelo ajustado, para o caso 4	239
E.34	Médias do $\widehat{\text{BIC}}$ por modelo gerador dos dados e modelo ajustado, para o caso 4	240
E.35	Médias do LLP por modelo gerador dos dados e modelo ajustado, para o caso 4	240
E.36	Médias do DIC por modelo gerador dos dados e modelo ajustado, para o caso 5	241

E.37 Médias do $\widehat{\text{BIC}}$ por modelo gerador dos dados e modelo ajustado, para o caso 5	241
E.38 Médias do LLP por modelo gerador dos dados e modelo ajustado, para o caso 5	242
E.39 Médias do DIC por modelo gerador dos dados e modelo ajustado, para o caso 6	242
E.40 Médias do $\widehat{\text{BIC}}$ por modelo gerador dos dados e modelo ajustado, para o caso 6	243
E.41 Médias do LLP por modelo gerador dos dados e modelo ajustado, para o caso 6	243

Nota Introdutória

Um dos desafios mais importantes da Estatística é a construção de modelos que, aproximadamente, descrevam um determinado fenómeno aleatório presente na natureza. Na maior parte dos casos, dada uma amostra de uma experiência relacionada com esse fenómeno, diferentes propostas de modelos podem ser consideradas. A selecção ou comparação de modelos é o procedimento estatístico que determina qual a estrutura probabilística que deve ser escolhida, de um conjunto de modelos candidatos.

O modelo seleccionado deve incorporar toda a informação essencial, mas excluir as particularidades menos relevantes, de modo que os aspectos importantes não se diluam no excesso de pormenores. Por este motivo, qualquer procedimento de selecção de modelos deve ter em conta o equilíbrio entre o ajustamento — o modelo seleccionado deve descrever o conjunto de dados o melhor possível — e a parcimónia — deve evitar-se a desnecessária complicação do modelo.

De modo a impedir que maus modelos sejam aceites, meramente porque ajustam melhor os dados que outros ainda piores, deve também efectuar-se a validação do modelo, isto é, averiguar se o modelo seleccionado é adequado. A selecção e validação de modelos constitui uma etapa da análise estatística, por vezes, designada determinação de modelos.

A abordagem bayesiana ao problema de ajustamento de modelos passa, não só pela especificação da distribuição de probabilidade conjunta do vector de observações, condicional a um vector de parâmetros, bem como pela atribuição de uma distribuição de probabilidade ao próprio vector de parâmetros. Este é o modelo bayesiano padrão, definido através de uma hierarquia em dois níveis. Mas, em muitos casos, a especificação da própria distribuição *a priori* do vector de parâmetros é também feita à custa de diversos níveis de hierarquia. Esta estrutura hierárquica é uma das principais características da modelação bayesiana.

Na literatura estatística encontra-se uma grande variedade de critérios para comparação de modelos. Os mais conhecidos e utilizados são os que se baseiam na minimização de uma quantidade, que pode ser expressa como menos duas vezes o máximo da log-verosimilhança mais uma função de penalização, que depende do número de parâmetros do modelo. Entre eles destaca-se o critério de informação de Akaike (Akaike, 1973) e o critério de informação bayesiano (Schwarz, 1978). A função de penalização do critério de informação bayesiano depende, não só do número de parâmetros do modelo, como também da dimensão da amostra. Apesar de algumas aproximações assintóticas que procuram justificar a utilização destes critérios na abordagem bayesiana (Kass e Raftery, 1995), eles não são intrinsecamente bayesianos, na medida em que não permitem incorporar

a informação *a priori* eventualmente existente.

O factor Bayes (Jeffreys, 1961) foi durante muito tempo considerado como o único modo correcto de efectuar a comparação bayesiana de modelos. Um dos problemas inerentes à utilização do factor Bayes na comparação de modelos é o facto de ele não se encontrar bem definido quando se utilizam distribuições *a priori* impróprias, o que acontece frequentemente no caso de distribuições *a priori* não informativas. Para contornar este problema, diversas modificações do factor Bayes têm sido propostas, tais como os factores Bayes *a posteriori* (Aitkin, 1991), parcial (O'Hagan, 1991), intrínseco (Berger e Pericchi, 1996), fraccionário (O'Hagan, 1995) e o factor pseudo-Bayes (Gelfand *et al.*, 1992).

Alternativamente a estas modificações do factor Bayes, Carlin e Louis (2000) propõem um critério de verosimilhança penalizada, de cariz bayesiano, para comparação de modelos. Este critério, designado critério de informação bayesiano estimado, consiste em escolher o modelo que minimiza menos duas vezes o valor esperado *a posteriori* da log-verosimilhança mais uma função de penalização, que depende da dimensão da amostra e do número de parâmetros do modelo.

A estrutura tipicamente hierárquica da modelação bayesiana dificulta, contudo, a determinação do número de parâmetros do modelo. Se algum dos modelos comparados tiver muitos níveis de hierarquia, a utilização do critério de informação bayesiano estimado pode estar inviabilizada. O critério de informação da *deviance* (Spiegelhalter *et al.*, 2002) surgiu como uma alternativa adequada para a comparação de modelos nesta situação. Este critério é constituído por duas componentes: uma medida da qualidade do ajustamento — o valor esperado *a posteriori* da *deviance* bayesiana — mais uma medida da complexidade do modelo — a diferença entre o valor esperado *a posteriori* da *deviance* bayesiana menos a *deviance* bayesiana calculada no valor esperado *a posteriori* do vector de parâmetros.

Note-se que, a *deviance* bayesiana, tal como é definida em Spiegelhalter *et al.* (2002), é menos duas vezes a log-verosimilhança mais duas vezes o logaritmo de uma função que depende apenas dos dados. Na comparação de modelos é usual considerar que esta função é igual a 1. Deste modo, a medida da qualidade de ajustamento do modelo do critério de informação bayesiano estimado e do critério de informação da *deviance* é a mesma, havendo apenas diferença na medida da complexidade do modelo.

O modelo seleccionado depende, obviamente, do critério de comparação de modelos utilizado; além disso, pode também depender do método de pesquisa utilizado na selecção de modelos. De facto, quando existem muitos modelos para comparar não é possível examinar exaustivamente todos eles e, deste modo, a arbitrariedade na escolha do melhor modelo é devida ao critério de selecção de modelos e ao método de pesquisa utilizados.

Por exemplo, no caso particular da selecção de covariáveis em modelos de regressão, o número de modelos a comparar é extremamente elevado, mesmo com um número moderado de possíveis covariáveis. Assim, a pesquisa exaustiva do melhor modelo é uma solução que não pode ser considerada na maior parte das aplicações práticas, sendo necessário recorrer a métodos alternativos.

Ao contrário do que acontece na abordagem clássica, onde o número de *packages* estatísticos para análise e modelação de dados é bastante elevado, a utilização da metodologia bayesiana nas aplicações práticas é um fenómeno que só recentemente foi despoletado, com o aparecimento do BUGS (Spiegelhalter *et al.*, 1996a) — de *Bayesian*

inference Using Gibbs Sampling. Este programa utiliza uma técnica de simulação conhecida por amostragem Gibbs, que possibilita a utilização das metodologias de inferência bayesiana nas aplicações estatísticas. O WinBUGS, isto é, a versão do BUGS para Windows, gratuitamente disponível em “<http://www.mrc-bsu.cam.ac.uk/bugs/>”, foi determinante para o rápido crescimento da utilização da inferência bayesiana nas aplicações práticas. Além dos manuais do próprio programa (Spiegelhalter *et al.*, 1996b, 1996c), outros exemplos do modo como a inferência bayesiana pode ser efectuada recorrendo ao WinBUGS podem ser vistos, por exemplo, em Congdon (2001).

A generalização da utilização dos métodos de Monte Carlo via cadeias de Markov possibilitou o aparecimento de diversos métodos bayesianos de selecção de covariáveis via amostrador Gibbs, que actuam simultaneamente no espaço dos modelos e no espaço dos parâmetros, entre os quais se destaca o método Gibbs de selecção de covariáveis (*e.g.*, Dellaportas *et al.*, 2000). Este método permite estimar as probabilidades *a posteriori* de todos os possíveis modelos, com base nas respectivas probabilidades *a priori* e nos dados.

Se nenhum dos modelos candidatos for claramente superior aos restantes, a média ponderada de modelos bayesianos (Hoeting *et al.*, 1999) pode ser a alternativa adequada à selecção de modelos. Neste caso, as inferências são efectuadas à custa da distribuição *a posteriori* da quantidade de interesse, que é uma mistura das distribuições *a posteriori* da quantidade de interesse para cada um dos modelos, ponderada pela probabilidade *a posteriori* do respectivo modelo.

Após a selecção de um modelo deve proceder-se à validação do mesmo. Mais precisamente, deve averiguar-se se o modelo é adequado, isto é, se possibilita uma boa descrição dos dados, permitindo efectuar as inferências necessárias para dar resposta às questões pretendidas. Destaca-se o *p-value* preditivo *a posteriori* (Rubin, 1984) e, em particular, o *p-value* de discrepância (Gelman *et al.*, 1995), como medidas de adequabilidade dos modelos.

Muitas vezes relegada para um segundo plano nas aplicações práticas, a selecção e validação de modelos é uma etapa fulcral para o sucesso de um processo de inferência, seja ele clássico ou bayesiano, paramétrico ou preditivo.

O objectivo desta dissertação é fazer uma revisão de alguns métodos de selecção e validação de modelos, propor métodos de pesquisa do melhor subconjunto de covariáveis em modelos de regressão e, através de estudos de simulação, comparar a performance de alguns dos critérios de selecção de modelos.

Esta tese está organizada em seis capítulos.

Com o capítulo 1, intitulado “Fundamentos de Estatística Bayesiana”, pretende-se despertar a curiosidade dos utilizadores de ferramentas estatísticas para a abordagem bayesiana, apresentando algumas noções básicas, indicando algumas referências importantes e enquadrando o problema da abordagem bayesiana à determinação de modelos, que é o objectivo desta dissertação.

No segundo capítulo são apresentados alguns critérios de selecção de modelos, com destaque para os diversos factores Bayes e para os critérios de verosimilhança penalizada, nomeadamente o critério de informação bayesiano estimado e o critério de informação da

deviance. Estes critérios são ilustrados através de um exemplo simples — a comparação dos modelos beta/binomial negativo e gama/Poisson para dados de contagem —, em que é possível obter analiticamente as suas expressões. Faz-se também referência à média ponderada de modelos bayesianos como uma alternativa à selecção de modelos.

A selecção de covariáveis em modelos de regressão é o assunto tratado no terceiro capítulo. A comparação dos subconjuntos de covariáveis pode ser efectuada à custa das medidas de selecção de modelos que, no contexto da selecção de covariáveis, são designadas medidas de discrepância. Dá-se destaque ao método de selecção informal por pesquisa manual via medidas de discrepância (Paulino *et al.*, 2003) e ao método Gibbs de selecção de covariáveis (Dellaportas *et al.*, 1997). Neste trabalho propõe-se, para problemas de selecção de covariáveis em modelos de regressão, dois métodos bayesianos de pesquisa do melhor modelo, inspirados nos seus homólogos clássicos — os métodos de selecção progressiva e de eliminação regressiva de covariáveis via medidas de discrepância. Recorre-se a dois conjuntos de dados da literatura, o da massa gorda (*e.g.*, Johnson, 1996) e o da homocisteinemia (*e.g.*, Teles, 1995), para exemplificar estes métodos de selecção de covariáveis, respectivamente, em modelos de regressão linear normal e modelos de regressão gama.

Os dois estudos de simulação efectuados são apresentados no capítulo 4. No primeiro estudo, simula-se dados de contagem dos modelos beta/binomial negativo e gama/Poisson e, para cada amostra, ajusta-se estes dois modelos, determinando as percentagens de classificação correcta obtidas com cada um dos critérios de selecção de modelos. No segundo estudo, simula-se dados de regressão linear normal de 16 modelos, cada um deles resultante da inclusão no modelo de uma das combinações possíveis de 4 covariáveis; para cada amostra simulada, ajusta-se os 16 modelos possíveis, sendo registados, de modo a avaliar a sua performance, os resultados obtidos com cada um dos critérios de selecção de covariáveis.

Os modelos seleccionados devem ser criteriosamente validados. No capítulo 5 destaca-se alguns métodos de validação de modelos, com especial destaque para o *p-value* e, em particular, para o *p-value* preditivo *a posteriori*. São também sugeridos alguns gráficos que permitem efectuar a validação informal de modelos. Ilustra-se estes métodos com modelos beta/binomial negativo e gama/Poisson para os dados de contagem, com modelos de regressão linear normal para os dados da massa gorda e com modelos de regressão gama para os dados da homocisteinemia.

Por fim, no capítulo 6, denominado “Discussão e Trabalho Futuro”, resume-se as conclusões efectuadas nos restantes capítulos e discute-se as conclusões gerais, perspectivando o muito trabalho que ainda há para fazer.

Em anexo são listadas as distribuições de probabilidade (anexo A) e os conjuntos de dados (anexo C) referidos nesta dissertação. Além disso, são também apresentados alguns aspectos da determinação de modelos via métodos MCMC (anexo B), os códigos do WinBUGS para implementação de alguns dos métodos de selecção e validação de modelos (anexo D), e algumas tabelas com resultados dos estudos de simulação (anexo E).

Capítulo 1

Fundamentos de Estatística Bayesiana

1.1 Introdução

A interpretação dos fenômenos aleatórios é usualmente feita com recurso à noção de vector aleatório, visto que a quantidade observada $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ numa ocorrência desse fenómeno pode ser encarada como uma observação de um vector aleatório $X = (X_1, X_2, \dots, X_n)$. Designa-se por espaço-amostra, e representa-se por \mathcal{X} , o contradomínio do vector aleatório X ; trata-se de um subconjunto de \mathbb{R}^n que corresponde à ideia intuitiva dos valores que uma amostra pode tomar.

Para a interpretação descrita ter interesse prático torna-se necessário estabelecer uma família \mathcal{F} de distribuições à qual pertença a distribuição F do vector aleatório X . Uma forma de isto ser feito é considerar que \mathcal{F} é uma família indexada num determinado conjunto de parâmetros $\Theta \subseteq \mathbb{R}^k$:

$$\mathcal{F} = \{F_\theta : \theta \in \Theta\}.$$

Diz-se que \mathcal{F} é um modelo estatístico para X .

Tendo em conta a natureza aleatória dos fenómenos a que se aplica o modelo \mathcal{F} , pode refinar-se a construção anterior e considerar que o próprio vector de parâmetros θ tem natureza aleatória. Sob este ponto de vista, θ pode ser encarado como um dos possíveis valores de um novo vector aleatório Θ_0 . Chama-se espaço-parâmetro de Θ_0 ao subconjunto de \mathbb{R}^k que é contradomínio do vector aleatório Θ_0 .

Por convenção utiliza-se letras latinas para indicar variáveis aleatórias e letras gregas para os parâmetros. Ainda por convenção, reserva-se as letras maiúsculas para variáveis aleatórias e as minúsculas para valores observados da variável aleatória. Assim, por coerência, o parâmetro enquanto variável aleatória deveria ser representado por letra maiúscula grega. Contudo, para não sobrecarregar a notação, não se fará distinção na representação do parâmetro e da variável aleatória que representa esse parâmetro, à semelhança do que acontece em trabalhos de índole bayesiana.

Se as variáveis aleatórias X_i , $i = 1, 2, \dots, n$, forem independentes, condicionalmente

a θ , então $f(x|\theta) = \prod_i^n f_i(x_i|\theta)$, onde $f_i(x_i|\theta)$ é a densidade da variável aleatória X_i condicional a θ . Se, além de independentes, as variáveis aleatórias X_i forem identicamente distribuídas, sendo $f_0(x_i|\theta)$ a respectiva densidade, então $f(x|\theta) = \prod_i^n f_0(x_i|\theta)$.

Considere-se que foi observada a amostra x . A densidade de probabilidade conjunta para X e θ pode ser escrita como o produto de duas densidades,

$$p(x, \theta) = f(x|\theta) h(\theta).$$

Nesta expressão $f(x|\theta)$ representa a especificação do modelo para o vector de observações x , dado o vector de parâmetros θ — denominada informação amostral — e $h(\theta)$ é a densidade de probabilidade da quantidade aleatória θ — muitas vezes referida por distribuição *a priori* de θ . A função de verosimilhança $L(\theta|x)$ é proporcional a $f(x|\theta)$ a menos de factores que não dependem de θ , isto é, $L(\theta|x) \propto f(x|\theta)$, e representa a plausibilidade ou verosimilhança de cada $\theta \in \Theta$, quando foi observada a amostra x .

Representa-se por $f(x)$ a distribuição do vector de dados X , desconhecido mas observável,

$$f(x) = \int_{\Theta} p(x, \theta) d\theta = \int_{\Theta} f(x|\theta) h(\theta) d\theta, \quad (1.1)$$

que se designa por distribuição marginal de X . Também é usual designar $f(x)$ por distribuição preditiva *a priori*, uma vez que sumaria a informação relativa a x antes de x ter sido observado.

Condicional aos dados observados x , a distribuição de probabilidade de θ pode ser actualizada usando propriedades básicas da Teoria de Probabilidades, nomeadamente o teorema de Bayes para densidades,

$$h(\theta|x) = \frac{p(x, \theta)}{f(x)} = \frac{f(x|\theta) h(\theta)}{\int_{\Theta} f(x|t) h(t) dt}, \quad \theta \in \Theta. \quad (1.2)$$

A distribuição $h(\theta|x)$, que se designa por distribuição *a posteriori* de θ , é proporcional ao produto da informação amostral $f(x|\theta)$ pela distribuição *a priori* $h(\theta)$, combinando estas duas fontes de informação. O teorema de Bayes para densidades pode ser entendido como uma forma de actualizar a informação sobre θ , através da multiplicação de $h(\theta)$ pela razão $f(x|\theta)/f(x)$. Quanto maior for a dimensão da amostra, mais forte é a informação amostral e, conseqüentemente, menor o impacto da distribuição *a priori* $h(\theta)$ na distribuição *a posteriori* $h(\theta|x)$ (e.g., Sorensen e Gianola, 2002, capítulo 7). A distribuição $h(\theta|x)$ tem um papel fundamental em todo o processo de inferência, seja ele paramétrico ou preditivo. Nas secções 2.5.1 e 2.5.2 exemplifica-se, para os modelos beta/binomial negativo e gama/Poisson, a determinação de distribuições *a posteriori* dos parâmetros.

Quando não for necessária a especificação completa da distribuição *a posteriori* utiliza-se a densidade *a posteriori* não normalizada,

$$h(\theta|x) \propto f(x|\theta)h(\theta),$$

evitando-se deste modo o cálculo do integral (1.1) que por vezes só é possível recorrendo a métodos de integração numérica ou de simulação estocástica.

Uma das vantagens da Estatística Bayesiana é a facilidade com que as inferências podem ser actualizadas, quando os dados são obtidos em várias etapas. Para ilustrar esta

situação, considere-se que os dados x foram obtidos em 2 etapas: numa primeira etapa foi observada a amostra $x^{(1)}$ e numa segunda etapa foi observada a amostra $x^{(2)}$. Por (1.2) a distribuição *a posteriori* de θ depois de observada a amostra $x^{(1)}$ na primeira etapa é dada por

$$h(\theta|x^{(1)}) = \frac{f(x^{(1)}|\theta) h(\theta)}{f(x^{(1)})}. \quad (1.3)$$

Na segunda etapa, quando se observa a amostra $x^{(2)}$, a informação *a priori* sobre o parâmetro θ já foi actualizada para $h(\theta|x^{(1)})$ e, por (1.2), conclui-se que

$$h(\theta|x^{(1)}, x^{(2)}) = \frac{f(x^{(2)}|x^{(1)}, \theta) h(\theta|x^{(1)})}{\int_{\Theta} f(x^{(2)}|x^{(1)}, t) h(t|x^{(1)}) dt}.$$

Mas, como $\int_{\Theta} f(x^{(2)}|x^{(1)}, t) h(t|x^{(1)}) dt = f(x^{(2)}|x^{(1)})$ logo

$$h(\theta|x^{(1)}, x^{(2)}) = \frac{f(x^{(2)}|x^{(1)}, \theta) h(\theta|x^{(1)})}{f(x^{(2)}|x^{(1)})}$$

e, substituindo $h(\theta|x^{(1)})$ pela expressão (1.3), obtém-se

$$h(\theta|x^{(1)}, x^{(2)}) = \frac{f(x^{(2)}|x^{(1)}, \theta) f(x^{(1)}|\theta) h(\theta)}{f(x^{(2)}|x^{(1)}) f(x^{(1)})}. \quad (1.4)$$

Uma vez que $f(x^{(2)}|x^{(1)}, \theta) f(x^{(1)}|\theta) = f(x^{(1)}, x^{(2)}|\theta)$ e $f(x^{(2)}|x^{(1)}) f(x^{(1)}) = f(x^{(1)}, x^{(2)})$ então

$$h(\theta|x^{(1)}, x^{(2)}) = \frac{f(x^{(1)}, x^{(2)}|\theta) h(\theta)}{f(x^{(1)}, x^{(2)})}.$$

Fica assim demonstrado que a distribuição *a posteriori* de θ , com base na amostra $x = (x^{(1)}, x^{(2)})$, tanto pode ser obtida determinando primeiro $h(\theta|x^{(1)})$ e usando a distribuição *a posteriori* de θ , observados os dados $x^{(1)}$, para distribuição *a priori* de θ , quando se observa os dados $x^{(2)}$, como considerando de uma só vez os dados $x = (x^{(1)}, x^{(2)})$. Este resultado é bastante útil quando os dados são obtidos sequencialmente no tempo, pois permite a actualização da distribuição *a posteriori* à medida que a informação amostral fica disponível. O resultado (1.4) pode ser generalizado para as amostras $x^{(1)}, x^{(2)}, \dots, x^{(k)}$ obtidas em k etapas.

Numa análise estatística tem-se, em geral, como objectivo não só a inferência sobre os parâmetros do modelo, mas também a predição de observações futuras. Depois do conjunto de dados x ter sido observado, pode-se querer predizer um conjunto de dados desconhecido mas observável, y , de um modelo que partilha do mesmo vector de parâmetros. A distribuição de Y , condicional a x , que se representa por $p(y|x)$, é dada por

$$p(y|x) = \int_{\Theta} p(y, \theta|x) d\theta = \int_{\Theta} f(y|x, \theta) h(\theta|x) d\theta,$$

e designa-se por distribuição preditiva *a posteriori*. A distribuição $p(y|x)$ sumaria a informação relativa à plausibilidade de observar uma amostra y , condicional à amostra observada x . Por vezes as observações a predizer são, condicionalmente ao parâmetro, independentes de x e, neste caso, $f(y|x, \theta) = f(y|\theta)$ e

$$p(y|x) = \int_{\Theta} f(y|\theta) h(\theta|x) d\theta.$$

Uma vez obtida a distribuição preditiva *a posteriori*, a predição de observações futuras é efectuada com base nesta distribuição, através dos métodos inferenciais mais adequados ao caso em análise.

Embora a inferência bayesiana seja bastante intuitiva, pois permite uma interpretação directa, em termos probabilísticos, dos testes de hipóteses e das regiões de confiança, ainda não é muito utilizada na maior parte das áreas científicas de aplicação, devido a dificuldades de índole prática na implementação desta metodologia. Nomeadamente, dificuldades na especificação da distribuição *a priori*, no cálculo do integral do denominador de (1.2) com vista à determinação da distribuição *a posteriori* e no cálculo de outros integrais necessários ao processo de inferência. O desenvolvimento de software para análise estatística de dados segundo a metodologia bayesiana está a ser um factor fulcral para uma maior utilização da Estatística Bayesiana por parte de investigadores de outras áreas.

1.2 Modelos hierárquicos bayesianos

A maioria das aplicações estatísticas envolve vários parâmetros que estão relacionados através da estrutura do próprio problema, o que se traduz no facto do modelo de probabilidade conjunta para esses parâmetros reflectir, de algum modo, a sua dependência. A modelação não hierárquica de problemas deste tipo pode dar origem a uma de duas situações: caso se considere poucos parâmetros, os modelos não ajustam convenientemente os dados; com muitos parâmetros, os modelos têm tendência a ajustar bem os dados conduzindo, no entanto, a más predições de observações futuras.

A modelação hierárquica simplifica a compreensão dos problemas com vários parâmetros e facilita o desenvolvimento de estratégias computacionais para a aproximação dos integrais envolvidos. Mas, mais importante ainda, com a modelação hierárquica é possível ter o número suficiente de parâmetros para ajustar adequadamente os dados, permitindo que a distribuição da população estruture a dependência entre parâmetros, por forma a evitar problemas de sobrestimação.

O modelo bayesiano padrão é definido por uma hierarquia de dois níveis, o primeiro constituído pela distribuição amostral dos dados X condicional ao vector de parâmetros θ , e o segundo referente à distribuição *a priori* de θ . Este modelo pode ser generalizado, através da especificação hierárquica em vários níveis da distribuição *a priori*.

Considere-se que a distribuição *a priori* de θ está indexada num vector de parâmetros η , que se designa vector de hiperparâmetros, sendo representada por $h(\theta|\eta)$. De modo a salientar que a distribuição *a posteriori* é condicional a η , a expressão (1.2) deve ser actualizada para

$$h(\theta|x, \eta) = \frac{p(x, \theta|\eta)}{f(x|\eta)} = \frac{f(x|\theta) h(\theta|\eta)}{\int_{\Theta} f(x|t) h(t|\eta) dt}, \quad \theta \in \Theta.$$

Por sua vez, caso exista incerteza relativamente ao valor a atribuir a η , deve expressar-se essa incerteza através da especificação, num terceiro nível, de outra distribuição *a priori* $g(\eta)$, que se designa por *hiperprior*. Neste caso a distribuição *a posteriori* é obtida

integrando também sobre η ,

$$h(\theta|x) = \frac{p(x, \theta)}{f(x)} = \frac{\int p(x, \theta, \eta) d\eta}{\iint p(x, t, \eta) d\eta dt} = \frac{\int f(x|\theta) h(\theta|\eta) g(\eta) d\eta}{\iint f(x|t) h(t|\eta) g(\eta) d\eta dt}, \quad \theta \in \Theta. \quad (1.5)$$

Também a distribuição *a priori* de η pode depender de um vector de parâmetros desconhecidos ξ , conduzindo à generalização de (1.5), através de duas distribuições *a priori*, $h(\theta|\eta)$ e $g(\eta|\xi)$, e de uma *hiperprior* $l(\xi)$, originando um modelo hierárquico com quatro níveis.

Uma informação mais detalhada acerca dos modelos hierárquicos bayesianos pode ser encontrada em Gelman *et al.* (1995, capítulo 5) e nas referências aí indicadas.

1.3 Distribuições *a priori*

Tal como na Estatística Clássica, o alicerce matemático da Estatística Bayesiana é a noção de espaço probabilizável (Ω, \mathcal{A}) . A partir de um espaço destes, constrói-se um espaço de probabilidade $(\Omega, \mathcal{A}, \text{Pr})$ atribuindo probabilidades aos elementos de \mathcal{A} . É na noção do espaço que é probabilizável e no conceito de probabilidade que reside a diferença fundamental entre a Estatística Clássica e a Estatística Bayesiana.

Na Estatística Clássica pressupõe-se que só são probabilizáveis os elementos de \mathcal{A} que representam situações físicas repetíveis e a atribuição de probabilidade é baseada em critérios frequentistas. Na Estatística Bayesiana os elementos de \mathcal{A} podem representar situações repetíveis ou não repetíveis, e a atribuição de probabilidade aos elementos não repetíveis baseia-se na quantificação das crenças sobre eles, devendo incluir não só a informação acumulada no passado como também a opinião de especialistas da área científica do estudo. Indivíduos distintos podem ter, e é natural que tenham, graus de credibilidade diferentes, uma vez que é pouco provável que estejam na posse do mesmo tipo de informação. Contudo, estas probabilidades subjectivas ou pessoais têm que obedecer a critérios de coerência e consistência.

Além da dificuldade inerente ao cálculo de integrais, já referida na secção 1.1, a atribuição de uma distribuição de probabilidade ao vector de parâmetros θ é outro factor que traz uma dificuldade acrescida a uma aplicação generalizada da metodologia bayesiana. Nesta secção faz-se um breve resumo sobre o modo de representação da informação *a priori*. Outros aspectos relativos à eliciação da distribuição *a priori* podem ser vistos em Paulino *et al.* (2003, capítulo 2) e nas referências aí mencionadas.

Na maior parte dos casos não existe dúvida relativamente à forma da distribuição *a priori*, bastando eliciar os hiperparâmetros para que o modelo fique completamente especificado. Considere-se que a distribuição *a priori* de θ pertence a uma família de distribuições representada por $h(\theta|\eta)$, em que η é escolhido de modo a representar as credibilidades do eliciador sobre o vector de parâmetros θ . Por exemplo, se η for um vector com duas componentes, o conhecimento de dois momentos ou de dois quantis é suficiente para a especificação completa de $h(\theta|\eta)$. Este modo de eliciar os hiperparâmetros faz parte dos métodos de eliciação estrutural. Existem também os métodos preditivos de eliciação de hiperparâmetros que, como o próprio nome indica, são baseados na distribuição preditiva

a priori; neste caso, são pedidas ao especialista informações referentes à distribuição preditiva *a priori*.

Para exemplificar estes dois métodos, considere-se o caso particular de uma verosimilhança binomial negativa em que o parâmetro θ , que se pretende eliciar, é a probabilidade de sucesso numa prova de Bernoulli. Neste caso o espaço-parâmetro é $\Theta = [0, 1]$ e, caso se opte pela beta para distribuição *a priori* de θ (que como se irá ver é, por vezes, conveniente), basta eliciar os dois parâmetros da beta de modo que a especificação da distribuição *a priori* esteja completa. Caso se pretenda utilizar o método de eliciação estrutural, basta fazer uso da relação existente entre os parâmetros da beta e o seu valor médio e variância (ver anexo A), e decidir quais os valores a atribuir a estes momentos da distribuição de modo a obter os valores dos hiperparâmetros. Para utilizar o método preditivo na eliciação dos hiperparâmetros da distribuição beta é necessário determinar a distribuição preditiva *a priori* (secção 2.5.1). Para encontrar os seus parâmetros, pede-se ao especialista que elicie duas ou mais (caso se pretenda testar a coerência da informação) probabilidades associadas à distribuição preditiva *a priori*.

A escolha de uma distribuição *a priori*, que represente a informação que o eliciador tem sobre θ , pode ser feita em função da conveniência de simplificação da derivação analítica da distribuição *a posteriori* e da distribuição preditiva *a posteriori* ou de diminuição do esforço computacional. Nesta perspectiva, a selecção da distribuição *a priori* adequada, que incorpore as crenças existentes sobre θ , deve ser feita na família conjugada natural¹ da distribuição $f(x|\theta)$, pois deste modo a distribuição *a posteriori* tem a mesma forma da distribuição *a priori*. Muitas das distribuições usadas para representar a informação amostral pertencem à família exponencial, para a qual existem distribuições *a priori* conjugadas naturais. Por exemplo, identifica-se facilmente a distribuição conjugada natural da binomial negativa se a forma da distribuição for analisada não como função de x , mas como função de θ ; nesta perspectiva a estrutura $\theta^a(1-\theta)^b$ revela que a distribuição conjugada natural da binomial negativa é a beta.

Quando não existe informação de natureza objectiva nem subjectiva sobre θ , ou quando essa informação é escassa, diz-se que se está, respectivamente, num estado de ignorância *a priori*, ou num estado de conhecimento vago. Nestes casos de ausência ou escassez de informação sobre θ , ou quando se pretende fazer inferência baseada apenas nos dados, deve utilizar-se distribuições *a priori* não informativas. A utilização de distribuições *a priori* não informativas foi inicialmente justificada com o Princípio da Razão Insuficiente de Bayes e Laplace: na ausência de informação *a priori* suficiente para privilegiar umas hipóteses em detrimento de outras, deve atribuir-se igual credibilidade a todos os valores de Θ . No caso de Θ ser finito, $\Theta = \{\theta_1, \dots, \theta_k\}$, a distribuição não informativa obtida por este argumento é a uniforme discreta, $h(\theta) = 1/k$, para $\theta \in \Theta$. No caso uniparamétrico, se o espaço-parâmetro for um intervalo limitado, digamos $\Theta = (a, b)$, com $-\infty < a < b < +\infty$, então a distribuição uniforme no intervalo (a, b)

$$h(\theta) = 1/(b - a), \quad a < \theta < b$$

pode ser encarada como uma distribuição não informativa para θ , pela generalização do

¹Seja $f(x|\theta)$ um elemento de $\mathcal{F} = \{f(x|\theta) : \theta \in \Theta\}$ e seja $\mathcal{H} = \{h(\theta|a) : a \in \mathcal{A}\}$ uma família de distribuições *a priori* indexada no conjunto \mathcal{A} . A família \mathcal{H} diz-se conjugada natural de \mathcal{F} se (i) $f(x|\theta)$, para cada x , é proporcional a um membro de \mathcal{H} e (ii) \mathcal{H} é fechada em relação a produtos, isto é, $\forall a_0, a_1 \in \mathcal{A}, \exists a_2 \in \mathcal{A} : h(\theta|a_0)h(\theta|a_1) \propto h(\theta|a_2)$.

Princípio da Razão Insuficiente de Bayes e Laplace, embora este argumento não seja consensual. Se o espaço-parâmetro for um intervalo não limitado, por exemplo $\Theta = (-\infty, +\infty)$, então esse princípio levaria à distribuição *a priori*

$$h(\theta) = c,$$

onde c é uma constante positiva. Mas como $\int_{\Theta} h(\theta)d\theta = \infty$ esta distribuição é imprópria podendo originar distribuições *a posteriori* impróprias e inviabilizar o processo de inferência.

A consideração de distribuições *a priori* uniformes, além de dar origem a distribuições *a priori* impróprias quando o espaço-parâmetro não é limitado, tem também a desvantagem de não ser invariante sobre reparametrização. Uma distribuição é invariante sobre reparametrização se a ignorância sobre θ implicar ignorância sobre $\psi(\theta)$, sendo ψ uma aplicação injectiva de θ , do espaço parâmetro Θ para o espaço-parâmetro $\psi(\Theta)$. O método de Jeffreys permite a obtenção de distribuições *a priori* invariantes perante reparametrização. No caso de θ ser uniparamétrico, a distribuição *a priori* de Jeffreys baseia-se na medida de informação de Fisher sobre θ ,

$$h(\theta) \propto [I(\theta)]^{1/2},$$

sendo $I(\theta)$ definida por

$$I(\theta) = E_{X|\theta} \left[\left(\frac{\partial \ln f(X|\theta)}{\partial \theta} \right)^2 \right].$$

No caso de θ ser multiparamétrico, a distribuição *a priori* de Jeffreys é baseada na matriz de informação de Fisher

$$h(\theta) \propto |I(\theta)|^{1/2}, \quad (1.6)$$

em que $|I(\theta)|$ é o determinante da matriz $I(\theta)$, de elementos

$$I_{ij}(\theta) = E_{X|\theta} \left[\frac{\partial \ln f(X|\theta)}{\partial \theta_i} \frac{\partial \ln f(X|\theta)}{\partial \theta_j} \right].$$

A utilização de (1.6) para obter distribuições *a priori* não informativas pode ser muito trabalhosa, principalmente quando o vector de parâmetros θ tem dimensão elevada. Atendendo ao argumento de que a ignorância *a priori* é coerente com independência, a dificuldade de aplicação do método de Jeffreys no caso multiparamétrico pode ser ultrapassada encontrando distribuições *a priori* não informativas para cada um dos parâmetros e considerando como distribuição *a priori* conjunta o seu produto.

Existem outros métodos para encontrar distribuições que representem situações de ignorância *a priori*, como por exemplo o método da entropia máxima devido a Jaynes e o método de Berger-Bernardo (*e.g.*, Paulino *et al.*, 2003, secções 2.4.5 e 2.4.6).

Veja-se três situações muito usuais em Inferência Bayesiana (Berger, 1985):

1. a densidade de X envolve μ apenas através do termo $x - \mu$, isto é, $f(x|\mu) = f_0(x - \mu)$. Neste caso μ designa-se por parâmetro de localização e uma distribuição *a priori* não informativa para μ , invariante sobre transformações de localização (isto

é, transformações do tipo $Y = X + c$), é a distribuição uniforme no intervalo de variação de μ . No caso $\Theta = \mathbb{R}$ pode utilizar-se

$$h(\mu) = 1, \mu \in \mathbb{R};$$

2. a densidade de X é da forma $f(x|\sigma) = \frac{1}{\sigma}f_0(\frac{x}{\sigma})$, com $\sigma > 0$. Neste caso σ designa-se por parâmetro de escala e uma distribuição *a priori* não informativa para σ , invariante sobre transformações de escala (isto é, transformações do tipo $Y = cX$, para $c > 0$), no caso $\Theta = \mathbb{R}^+$, é dada por

$$h(\sigma) = \frac{1}{\sigma}, \sigma > 0;$$

3. a densidade de X combina parâmetros de localização e de escala na forma $f(x|\mu, \sigma) = \frac{1}{\sigma}f_0(\frac{x-\mu}{\sigma})$. Neste caso, e atendendo ao argumento de que a ignorância *a priori* é coerente com independência, uma distribuição *a priori* (conjunta) não informativa para o vector de parâmetros (μ, σ) é o produto das distribuições *a priori* não informativas de μ e σ . No caso $\Theta = \mathbb{R} \times \mathbb{R}^+$ obtém-se

$$h(\mu, \sigma) = \frac{1}{\sigma}, \mu \in \mathbb{R}, \sigma > 0.$$

No caso das distribuições *a priori* serem não informativas, as inferências efectuadas a partir da distribuição *a posteriori* baseiam-se apenas nos dados, permitindo a comparação destas inferências com as da Estatística Clássica.

1.4 Inferência paramétrica

A inferência paramétrica é efectuada a partir da informação sobre θ , expressa em termos da distribuição *a posteriori*, e engloba estimativas pontuais, estimativas por regiões e testes de hipóteses.

No caso de θ ser uniparamétrico, a representação gráfica da densidade *a posteriori* $h(\theta|x)$ permite visualizar a forma da densidade e efectuar algumas inferências de natureza qualitativa. Por exemplo, se a densidade tiver duas modas distintas, pode deduzir-se que os valores mais prováveis para θ estão agrupados em dois intervalos separados. Ou, caso exista uma assimetria pronunciada, conclui-se que intervalos de valores de θ de igual amplitude e equidistantes da moda têm probabilidades associadas bastante distintas. Além disso, o esboço do gráfico de $h(\theta|x)$ possibilita a escolha das medidas de localização e de dispersão mais adequadas para sumariar a informação sobre θ contida na distribuição *a posteriori*.

1.4.1 Estimação pontual

Como, numa perspectiva bayesiana, a inferência é efectuada com base na distribuição *a posteriori*, toma-se como estimativa de localização central do vector de parâmetros $\theta = (\theta_1, \dots, \theta_k)$ uma medida resumo de $h(\theta|x)$ — a média, a mediana (se $k = 1$) ou

a moda — que se representa por $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$. Deve seleccionar-se a medida sumária da distribuição *a posteriori* mais adequada, tendo em conta a forma de $h(\theta|x)$ e o uso final da estimativa.

A moda *a posteriori* é dada por $\hat{\theta} = M_{\theta|x}[\theta]$, tal que

$$h(\hat{\theta}|x) = \max_{\theta \in \Theta} h(\theta|x).$$

Assim, para determinar a moda não é necessário calcular o integral (1.1), uma vez que

$$\max_{\theta \in \Theta} h(\theta|x) = \max_{\theta \in \Theta} \{f(x|\theta) h(\theta)\}.$$

Evitar o cálculo do referido integral pode ser uma vantagem da utilização da moda *a posteriori* como medida de localização. Note-se que a moda *a posteriori* (máximo da distribuição $h(\theta|x)$) não corresponde necessariamente ao vector das modas *a posteriori* marginais dos θ_i , obtidas pela maximização separada de $h(\theta_i|x) = \int \dots \int h(\theta|x) d\theta_1 \dots d\theta_{i-1} d\theta_{i+1} \dots d\theta_k$.

A média *a posteriori* é o valor médio de θ com respeito à distribuição *a posteriori*, isto é, $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$, com

$$\hat{\theta}_i = E_{\theta|x}[\theta_i] = \int_{\Theta} \theta_i h(\theta|x) d\theta = \frac{\int_{\Theta} \theta_i f(x|\theta) h(\theta) d\theta}{\int_{\Theta} f(x|\theta) h(\theta) d\theta}, \quad i = 1, 2, \dots, k;$$

estes integrais, na maior parte dos casos, não podem ser resolvidos analiticamente.

Pelo facto de não haver relação de ordem (compatível com as operações algébricas) em \mathbb{R}^k , não tem interesse definir a mediana *a posteriori* de vectores k -dimensionais. A mediana *a posteriori* é, no entanto, bastante utilizada nos subespaços univariados como estatística sumária de $h(\theta_i|x)$. Neste contexto, define-se $\hat{\theta}$ como o vector das medianas (de quantidades escalares) *a posteriori*, $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$, com $\hat{\theta}_i$ o menor valor que verifica

$$\Pr \left\{ \theta_i \leq \hat{\theta}_i | x \right\} \geq \frac{1}{2} \quad \text{e} \quad \Pr \left\{ \theta_i \geq \hat{\theta}_i | x \right\} \geq \frac{1}{2},$$

para $i = 1, \dots, k$. Do mesmo modo, define-se o vector dos quantis (de quantidades escalares) *a posteriori* de probabilidade α , $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$, com $\hat{\theta}_i$ o menor valor que verifica

$$\Pr \left\{ \theta_i \leq \hat{\theta}_i | x \right\} \geq \alpha \quad \text{e} \quad \Pr \left\{ \theta_i \geq \hat{\theta}_i | x \right\} \geq 1 - \alpha,$$

para $i = 1, \dots, k$.

A matriz de covariância *a posteriori* é dada por

$$V = E_{\theta|x} \left[(\theta - E_{\theta|x}[\theta]) (\theta - E_{\theta|x}[\theta])' \right]$$

e, no caso univariado, a variância *a posteriori* é

$$\text{Var}_{\theta|x}[\theta] = E_{\theta|x} \left[(\theta - E_{\theta|x}[\theta])^2 \right] = \int_{\Theta} (\theta - E_{\theta|x}[\theta])^2 h(\theta|x) d\theta.$$

A variância *a posteriori* é bastante utilizada nos subespaços univariados de $h(\theta|x)$ como medida de precisão de $E_{\theta|x}[\theta_i]$ enquanto estimativa de θ_i .

1.4.2 Estimação por regiões

De modo análogo ao que se passa na estatística clássica, também na estatística bayesiana tem interesse sumariar $h(\theta|x)$, não apenas através de estimativas pontuais, mas principalmente através de regiões de Θ que contenham a maior parte da massa probabilística *a posteriori*. Neste contexto, $R(x)$ é uma região de credibilidade γ , com $\gamma \in (0, 1)$, para θ se

$$\Pr[\theta \in R(x) | x] = \int_{R(x)} h(\theta|x) d\theta \geq \gamma.$$

Uma vez que existem infinitas regiões de credibilidade γ , interessa encontrar a região que inclua todos os valores de θ mais credíveis *a posteriori*, isto é, que verifique a condição

$$h(\theta_1|x) \geq h(\theta_2|x), \forall \theta_1 \in R(x), \theta_2 \notin R(x).$$

De acordo com esta condição, define-se região de credibilidade γ com densidade *a posteriori* máxima (ou região HPD, de *highest posterior density*) como sendo a região $R(x)$ tal que

$$R(x) = \{\theta \in \Theta : h(\theta|x) \geq c_\gamma\}$$

onde c_γ é a maior constante para a qual se verifica $\Pr[\theta \in R(x) | x] \geq \gamma$.

Ao contrário do que acontece com as regiões de confiança clássicas, a região de credibilidade tem uma interpretação probabilística directa.

1.4.3 Testes de hipóteses

De modo idêntico às regiões de credibilidade, também os testes de hipóteses bayesianos têm uma interpretação probabilística directa, consistindo em calcular as probabilidades *a posteriori* de cada uma das hipóteses e, caso seja esse o objectivo, optar por uma delas.

Considere-se o teste de hipóteses $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$, onde $\Theta_1 = \Theta - \Theta_0$. Representa-se por $O(H_0, H_1|x)$ a razão das vantagens *a posteriori* de H_0 sobre H_1 ,

$$O(H_0, H_1|x) = \frac{\Pr\{H_0|x\}}{\Pr\{H_1|x\}},$$

que é uma medida da plausibilidade relativa entre as hipóteses em confronto. A razão das vantagens *a priori* de H_0 sobre H_1 é o quociente entre as probabilidades *a priori* de H_0 e H_1 ,

$$O(H_0, H_1) = \frac{\Pr\{H_0\}}{\Pr\{H_1\}}.$$

Para avaliar a influência do vector de dados x na alteração da credibilidade relativa das duas hipóteses, utiliza-se o quociente entre a razão das vantagens *a posteriori* e a razão das vantagens *a priori*, que se designa por factor Bayes a favor de H_0 (ou contra H_1) e se representa por

$$\text{BF}(x) = \frac{O(H_0, H_1|x)}{O(H_0, H_1)}.$$

Aplicando logaritmos à expressão anterior obtém-se

$$\ln \text{BF}(x) = \ln O(H_0, H_1|x) - \ln O(H_0, H_1).$$

Pode interpretar-se $\ln \text{BF}(x)$ como o peso relativo da evidência contida nos dados a favor da hipótese H_0 , isto é, o peso da evidência *a posteriori* menos o peso da evidência *a priori*.

Um factor Bayes muito grande ou muito pequeno quando comparado com 1, constitui evidência a favor de H_0 ou H_1 , respectivamente. Os valores aproximados para a interpretação de $\text{BF}(x)$ e $2 \ln \text{BF}(x)$, sugeridos por Jeffreys (1961), são apresentados na tabela 1.1 (*e.g.*, Raftery, 1996).

Tabela 1.1: Interpretação dos valores do factor Bayes.

BF	$2 \ln \text{BF}$	Evidência a favor de H_0
inferior a 1	negativo	negativa (favorece H_1)
1 a 3	0 a 2	fraca
3 a 20	2 a 6	moderada
20 a 150	6 a 10	forte
superior a 150	superior a 10	muito forte

Além da interpretação probabilística directa, os testes de hipóteses bayesianos têm duas outras vantagens sobre os seus homólogos clássicos: não existe limite no número de hipóteses que se pode considerar simultaneamente e as hipóteses não necessitam de ser encaixadas umas nas outras. Deste modo, o conceito de hipótese H_0 e H_1 é muitas vezes substituído pelo conceito de modelo m_i , com $i = 1, 2, \dots, M$, como se irá ver no capítulo 2.

1.5 Inferência preditiva

Em muitas situações a realização de inferências sobre o vector de parâmetros θ não é o objectivo final da análise estatística, mas sim uma etapa do processo de predição de observações futuras.

Como já foi referido, a distribuição preditiva *a posteriori* $p(y|x)$, definida na secção 1.1, sumaria a informação respeitante à plausibilidade de observar y , dados o modelo amostral $f(\cdot|\theta)$, a distribuição *a priori* $h(\theta)$ e os dados x observados até esse momento. A inferência preditiva é efectuada de modo análogo à inferência paramétrica, tendo apenas em atenção que a densidade de interesse é agora $p(y|x)$ em vez de $h(\theta|x)$. A obtenção de medidas sumárias da distribuição preditiva *a posteriori* pode ser feita através de predições pontuais ou através de regiões de predição, como é o caso da região de predição com a mais alta densidade preditiva. Uma predição pontual para y pode ser obtida à custa do valor esperado de Y , calculado em relação à distribuição $p(y|x)$, isto é,

$$\hat{y} = E_{Y|x}[Y] = \int y p(y|x) dy = \int \int y f(y|\theta) h(\theta|x) dy d\theta.$$

1.6 Decisão

Embora a distinção entre inferência e decisão não seja uma questão consensual (Barnett, 1999), a decisão estatística pode ser encarada como uma extensão da inferência paramétrica na qual, além da informação inicial sobre $\theta \in \Theta$ representada pela distribuição *a priori* $h(\theta)$ e da informação amostral $f(x|\theta)$, é também incorporada a informação respeitante aos custos ou às consequências de adoptar uma determinada acção. Estas consequências devem ser quantificadas de um modo adequado e a informação daí resultante deve influenciar a decisão a tomar, isto é, deve influenciar a escolha da acção a adoptar.

Para formalizar esta abordagem estatística é necessário introduzir o conceito de espaço de acções \mathcal{A} , constituído por todas as possíveis acções a entre as quais o decisor tem que optar. No contexto da decisão estatística o espaço-parâmetro Θ é designado por espaço de estados da natureza, pois cada $\theta \in \Theta$ representa um estado da natureza ou “cenário” onde se desenrola a acção.

O decisor desconhece qual o estado da natureza que prevalece e tem que tomar uma decisão numa situação de incerteza, tendo em conta a informação *a priori*, a informação amostral e, além disso, a informação sobre a consequência de adoptar a acção a . A quantificação desta nova informação, referente à interacção da acção a com o estado da natureza θ , pode ser feita através da função perda $l(\theta, a)$. Esta função, que tem como domínio o produto cartesiano $\Theta \times \mathcal{A}$ e contradomínio \mathbb{R} , representa a perda ocorrida quando θ é o verdadeiro estado da natureza e foi tomada a acção a .

Exemplos de funções perda muito utilizadas são a função perda quadrática $l(\theta, a) = |\theta - a|^2$, a função perda absoluta $l(\theta, a) = |\theta - a|$ e a função perda 0 – 1 definida por

$$l(\theta, a) = \begin{cases} 0 & \text{se } \theta = a \\ 1 & \text{se } \theta \neq a \end{cases} .$$

A função ou regra de decisão d é uma função com domínio em \mathcal{X} e contradomínio em \mathcal{A} , que face aos dados observados x , induz uma determinada acção a . Designe-se por \mathcal{D} o conjunto de todas as possíveis regras de decisão; quando o decisor utiliza a regra de decisão $d \in \mathcal{D}$ e observa $x \in \mathcal{X}$ incorre numa perda $l(\theta, d(x))$.

Sob o ponto de vista bayesiano, e perante os dados observados x , a opinião do decisor estatístico face a um determinado estado da natureza θ é sumariada através da distribuição *a posteriori* $h(\theta|x)$. Deste modo, associado a cada acção a , tem interesse definir a perda esperada ou risco *a posteriori*,

$$r(a|x) = E_{\theta|x}[l(\theta, a)] = \int_{\Theta} l(\theta, a) h(\theta|x) d\theta .$$

Deve assim escolher-se, caso exista, a acção $\bar{a} \in \mathcal{A}$ que minimiza a perda esperada *a posteriori*, isto é, tal que

$$r(\bar{a}|x) = \inf_{a \in \mathcal{A}} r(a|x) .$$

Considerando funções perda específicas, a estimação pontual e os testes de hipóteses podem ser vistos como casos particulares de decisão estatística. Na estimação pontual o

espaço de acções é igual ao espaço-parâmetro, isto é, $\mathcal{A} = \Theta$, e uma acção $a \in \mathcal{A}$ consiste em propor a como estimativa de θ . Relativamente ao teste de hipóteses $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$, em que $\Theta_1 = \Theta - \Theta_0$, o espaço de acções é constituído por dois elementos, isto é, $\mathcal{A} = \{a_0, a_1\}$ — a acção a_0 representa a aceitação de H_0 e a acção a_1 representa a rejeição de H_0 .

1.7 Método de Monte Carlo

Conforme referido na secção 1.4, a realização de inferências sobre o vector de parâmetros θ é feita com base na distribuição *a posteriori*

$$h(\theta|x) = \frac{f(x|\theta) h(\theta)}{\int_{\Theta} f(x|t) h(t) dt} \quad (1.7)$$

e as inferências sumárias podem ser obtidas na forma de valores esperados *a posteriori*

$$E_{\theta|x}[g(\theta)] = \int_{\Theta} g(\theta) h(\theta|x) d\theta \quad (1.8)$$

para certas escolhas de $g(\cdot)$. Caso se pretenda determinar a média ou a variância *a posteriori* da i -ésima componente de θ , então considera-se $g(\theta) = \theta_i$ ou $g(\theta) = (\theta_i - E_{\theta_i|x}[\theta_i])^2$, respectivamente. Se y for uma observação futura de um modelo indexado no mesmo vector de parâmetros θ , as inferências preditivas de y são baseadas na distribuição preditiva *a posteriori*

$$p(y|x) = \int_{\Theta} f(y|x, \theta) h(\theta|x) d\theta \quad (1.9)$$

onde $f(y|x, \theta)$ é a distribuição de Y sob o modelo paramétrico considerado, o qual pode ser ou não estocasticamente dependente de X , conforme foi referido na secção 1.1. De modo análogo, as predições sumárias podem ser obtidas na forma de valores esperados preditivos

$$E_{Y|x}[g(Y)] = \int g(y) p(y|x) dy \quad (1.10)$$

para escolhas adequadas de $g(\cdot)$.

A operação de integração desempenha assim um papel fundamental na Estatística Bayesiana, como se pode ver pela necessidade de cálculo dos integrais das expressões (1.7) a (1.10). As dificuldades surgem pois, na maior parte das situações, não é possível obter expressões explícitas para os integrais envolvidos. Alguns métodos analíticos, como a aproximação da distribuição *a posteriori* à distribuição normal multivariada, ou o método de Laplace de aproximação de integrais, assim como métodos de integração numérica, como a quadratura iterativa, são algumas das soluções possíveis para aproximar estes integrais (*e.g.*, Paulino *et al.*, 2003, capítulo 5). A utilização de métodos de Monte Carlo é uma alternativa apropriada aos métodos numéricos para a aproximação de integrais, particularmente no caso multidimensional. Os métodos de Monte Carlo assentam em simulação estocástica, isto é, na reprodução de valores de distribuições de probabilidade. Nesta secção sumaria-se a ideia geral subjacente ao método de Monte Carlo directo, ao método de Monte Carlo com amostragem via função de importância e ao método de Monte Carlo via cadeias de Markov, em particular, ao algoritmo Gibbs.

1.7.1 Monte Carlo directo

Considere-se o problema de aproximar um integral da forma apresentada em (1.8). Caso se possa simular uma amostra aleatória $\theta_1, \theta_2, \dots, \theta_N$ da densidade *a posteriori* $h(\theta|x)$, o método de Monte Carlo directo aproxima este integral pela média empírica

$$\hat{E}_{\theta|x}[g(\theta)] = \frac{1}{N} \sum_{i=1}^N g(\theta_i),$$

a qual converge quase certamente para $E_{\theta|x}[g(\theta)]$, quando $N \rightarrow \infty$ (Lei Forte dos Grandes Números²). A precisão desta estimativa pode ser medida pelo erro padrão estimado de Monte Carlo, dado por

$$\frac{1}{\sqrt{N(N-1)}} \left\{ \sum_{i=1}^N \left[g(\theta_i) - \frac{1}{N} \sum_{i=1}^N g(\theta_i) \right]^2 \right\}^{1/2}.$$

A distribuição preditiva *a posteriori*, apresentada em (1.9), é o valor médio *a posteriori* de $f(Y|x, \theta)$, isto é, $p(y|x) = E_{\theta|x}[f(Y|x, \theta)]$, logo a aproximação de Monte Carlo para esta distribuição é dada por

$$\hat{p}(y|x) = \frac{1}{N} \sum_{i=1}^N f(y|x, \theta_i),$$

sendo $(\theta_1, \theta_2, \dots, \theta_N)$ uma amostra de valores simulados de $h(\theta|x)$. A obtenção de estimativas de Monte Carlo associadas com a distribuição preditiva $p(y|x)$ passa pela simulação de amostras desta distribuição. Para isso, começa-se por gerar uma amostra $(\theta_1, \theta_2, \dots, \theta_N)$ de $h(\theta|x)$; para cada i , $i = 1, \dots, N$, obtém-se y_i de $f(y|x, \theta_i)$. Com base na amostra (y_1, y_2, \dots, y_N) , pode calcular-se vários resumos da distribuição preditiva.

1.7.2 Monte Carlo com amostragem de importância

Quando não é possível simular uma amostra independente e identicamente distribuída directamente da distribuição *a posteriori* $h(\theta|x)$, uma das soluções possíveis é a amostragem via função de importância, que consiste em simular de uma distribuição semelhante à distribuição *a posteriori*. Para estimar o valor esperado *a posteriori* apresentado em (1.8) procede-se do seguinte modo. Seja $k(\theta)$ uma função densidade da qual é fácil simular valores e que constitui uma boa aproximação de $h(\theta|x)$. Escrevendo $w(\theta) = f(x|\theta) h(\theta)/k(\theta)$, tem-se

$$\int_{\Theta} g(\theta) h(\theta|x) d\theta = \frac{\int_{\Theta} g(\theta) f(x|\theta) h(\theta) d\theta}{\int_{\Theta} f(x|\theta) h(\theta) d\theta} = \frac{\int_{\Theta} g(\theta) w(\theta) k(\theta) d\theta}{\int_{\Theta} w(\theta) k(\theta) d\theta},$$

o que sugere aproximar $E_{\theta|x}[g(\theta)]$ por

$$\hat{E}_{\theta|x}[g(\theta)] = \frac{1}{\sum_{i=1}^N w_i} \sum_{i=1}^N w_i g(\theta_i),$$

²Seja $\{X_k\}_{k \geq 1}$ uma sucessão de variáveis aleatórias independentes e identicamente distribuídas com X e defina-se $S_N = \sum_{k=1}^N X_k$ e $\bar{X}_N = \frac{1}{N} S_N$. Então, com $N \rightarrow \infty$, \bar{X}_N converge quase certamente para μ finito se e só se $E(|X|) < \infty$, e tem-se $E(X) = \mu$.

onde $w_i = f(x|\theta_i)h(\theta_i)/k(\theta_i)$ e $\theta_1, \theta_2, \dots, \theta_N$ é uma amostra simulada de $k(\theta)$.

Este método, que se designa por método de Monte Carlo com amostragem de importância, atribui mais peso a regiões onde $k(\theta) < h(\theta|x)$ e menos peso a regiões onde $k(\theta) > h(\theta|x)$. A função de importância $k(\theta)$ deve constituir uma boa aproximação de $h(\theta|x)$, mas com caudas mais pesadas do que $h(\theta|x)$ e, além disso, deve ser escolhida de modo a que seja fácil a geração de amostras dessa distribuição.

1.7.3 Monte Carlo via cadeias de Markov

Como a simulação da distribuição *a posteriori* $h(\theta|x)$ nem sempre constitui tarefa fácil, a ideia básica subjacente ao método de Monte Carlo via cadeias de Markov, que se representa por MCMC, é a transformação do problema estático em consideração num problema de natureza dinâmica, através da construção de um processo estocástico temporal, que deve ser fácil de simular e que deve convergir para a distribuição original. Este processo temporal é, de um modo geral, uma cadeia de Markov homogénea cuja distribuição de equilíbrio é a distribuição da qual se pretende simular amostras. Das muitas referências existentes sobre o método MCMC, destaca-se Gamerman (1997) e Gilks *et al.* (1996).

Para compreender integralmente o método MCMC é necessário ter alguns conhecimentos sobre cadeias de Markov. No entanto, como este capítulo pretende apenas dar a conhecer algumas ferramentas da Estatística Bayesiana, a teoria das cadeias de Markov necessária para a compreensão integral dos métodos MCMC não é aqui abordada. Conceitos básicos de cadeias de Markov relacionados com os métodos MCMC podem ser vistos em Roberts (1996) e Paulino *et al.* (2003, secção 8.1).

O algoritmo de Metropolis-Hastings permite, precisamente, a construção de cadeias de Markov com distribuições de equilíbrio específicas. Os denominados métodos de Monte Carlo via cadeias de Markov não são mais do que a associação deste algoritmo para simulação de distribuições ao método de Monte Carlo para aproximação de integrais. O método de amostragem Gibbs, que a seguir se apresenta, pode ser interpretado como um caso particular do algoritmo de Metropolis-Hastings (Gelman *et al.*, 1995; Gamerman, 1997; Paulino *et al.*, 2003).

Algoritmo Gibbs

O método de amostragem Gibbs baseia-se no facto que, se a distribuição *a posteriori* $h(\theta|x)$, com $\theta = (\theta_1, \dots, \theta_k)$, for positiva em $\Theta_1 \times \dots \times \Theta_k$, com Θ_i suporte da distribuição de θ_i , para $i = 1, \dots, k$, então $h(\theta|x)$ é unicamente determinada pelas distribuições condicionais completas $h(\theta_i|x, \theta_{(-i)})$, onde $\theta_{(-i)} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$.

A geração de vectores de $h(\theta_1, \dots, \theta_k|x)$ através do algoritmo Gibbs pode ser resumida do seguinte modo:

- inicializa-se o vector θ atribuindo-lhe um valor inicial arbitrário $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)})$ pertencente a $\Theta_1 \times \Theta_2 \times \dots \times \Theta_k$;
- gera-se um valor $\theta_1^{(1)}$ de $h(\theta_1|x, \theta_2^{(0)}, \theta_3^{(0)}, \theta_4^{(0)}, \dots, \theta_k^{(0)})$;

- gera-se um valor $\theta_2^{(1)}$ de $h(\theta_2 | x, \theta_1^{(1)}, \theta_3^{(0)}, \theta_4^{(0)}, \dots, \theta_k^{(0)})$;
- gera-se um valor $\theta_3^{(1)}$ de $h(\theta_3 | x, \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)}, \dots, \theta_k^{(0)})$;
- ⋮
- gera-se um valor $\theta_k^{(1)}$ de $h(\theta_k | x, \theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}, \dots, \theta_{k-1}^{(1)})$.

Completa-se assim um ciclo deste algoritmo, transitando-se do estado $\theta^{(0)}$ para o estado $\theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_k^{(1)})$. Considerando como vector inicial o vector $\theta^{(1)}$, resultante do primeiro ciclo, é efectuado um novo ciclo que possibilita a transição do estado $\theta^{(1)}$ para o estado $\theta^{(2)} = (\theta_1^{(2)}, \theta_2^{(2)}, \dots, \theta_k^{(2)})$, e assim sucessivamente. Este ciclo é repetido T vezes, tomando sempre como valor inicial em cada ciclo o vector gerado no ciclo anterior.

A sucessão $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots, \theta^{(t)}, \dots$ é uma realização de uma cadeia de Markov com espaço de estados Θ e função de transição $p(\theta^{(t)}, \theta^{(t+1)})$ dada por (e.g., Smith e Roberts, 1993):

$$p(\theta^{(t)}, \theta^{(t+1)}) = \prod_{i=1}^k h(\theta_i^{(t+1)} | x, \theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, \dots, \theta_k^{(t)}).$$

O método de amostragem Gibbs permite, através de um processo iterativo, gerar realizações de uma cadeia de Markov de tal modo que, à medida que o número de iterações aumenta, a cadeia aproxima-se da sua condição de equilíbrio. Se num determinado instante t a cadeia já se encontrar no estado de equilíbrio, então o vector $\theta^{(t)}$ gerado nesse instante pode ser considerado uma realização da distribuição $h(\theta|x)$.

A teoria subjacente às cadeias de Markov permite concluir que, quando $t \rightarrow \infty$, $\theta^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_k^{(t)})$ tende em distribuição para um vector aleatório cuja função densidade de probabilidade conjunta é $h(\theta|x)$ e a média ergódica $\frac{1}{t} \sum_{l=1}^t g(\theta^{(l)})$ converge quase certamente para $E_{\theta|x}[g(\theta)]$, para qualquer função $g(\cdot)$, onde $E_{\theta|x}[g(\theta)]$ representa o valor esperado de $g(\theta)$ em relação à distribuição *a posteriori* $h(\theta|x)$. Este resultado é conhecido por Teorema Ergódico. Em particular, verifica-se que se $t \rightarrow \infty$, então $\theta_i^{(t)}$ tende em distribuição para uma quantidade aleatória cuja densidade é a marginal *a posteriori* de θ_i , $h(\theta_i|x)$.

O número total de iterações T a efectuar depende de três aspectos essenciais: (i) a velocidade de convergência da cadeia para o seu estado de equilíbrio, (ii) a correlação entre os sucessivos valores gerados, (iii) a precisão pretendida para as estimativas dos parâmetros.

Relativamente ao primeiro aspecto, é necessário efectuar um determinado número m de iterações que vão ser descartadas, isto é, cujos valores não vão ser considerados nas inferências a efectuar. Estas primeiras m iterações, correspondentes à simulação de valores que se assume não pertencerem à distribuição de equilíbrio, constituem o chamado período de aquecimento (*burn-in*). A natureza dinâmica do método de simulação faz com que observações consecutivas geradas pelo algoritmo Gibbs sejam obviamente correlacionadas. Para obter observações independentes a técnica usual consiste em considerar apenas as observações espaçadas de um valor h (espaçamento de amostragem), depois do período de aquecimento. Em relação ao terceiro aspecto, a precisão das estimativas dos parâmetros pode ser controlada pela dimensão da amostra final N .

Os três aspectos essenciais para determinar o número de iterações T dão origem a outras tantas questões: (i) quantas iterações m deve ter o período de aquecimento? (ii) qual deve ser o espaçamento h entre as iterações? (iii) qual deve ser a dimensão da amostra final N ? Embora a escolha adequada do número de iterações do período de aquecimento dependa do valor inicial atribuído a θ e da estrutura de correlação da cadeia, valores $m = 1000$ ou $m = 5000$ são usuais em muitas aplicações. Relativamente à segunda questão, alguns autores sugerem que valores de $h = 10$ ou $h = 20$ são suficientes na maioria das aplicações; no entanto, há casos de cadeias com estrutura de correlação elevada que obrigam a considerar $h = 40$. Valores de $N = 5000$ ou $N = 10000$ conduzem a estimativas com uma precisão aceitável em muitas aplicações.

O número total de iterações T do algoritmo Gibbs, necessárias para obter uma amostra final de dimensão N , é $T = m + h \times N$. No processo de inferência é apenas utilizado o subconjunto de dimensão N extraído desta cadeia, constituído pelos vectores $\theta^{(m+1)}, \theta^{(m+(h+1))}, \dots, \theta^{(m+[(N-1)h+1])}$.

A utilização da metodologia bayesiana nas aplicações práticas é um fenómeno recente, ao qual não é alheio o aparecimento do WinBUGS (Spiegelhalter *et al.*, 1996a). Este programa possibilita a utilização das metodologias de inferência bayesiana nas aplicações práticas, recorrendo, para isso, ao método de amostragem Gibbs.

Um risco inerente à utilização de métodos MCMC é o facto de se considerar que há convergência do método MCMC antes da distribuição de equilíbrio ter sido atingida. Existem diversos métodos de diagnóstico da convergência do método de amostragem Gibbs que, embora não provem que a cadeia convergiu, permitem detectar situações de não convergência (*e.g.*, Gill, 2002; Paulino *et al.*, 2003). Entre estes métodos destaca-se: (i) as representações gráficas das estimativas dos parâmetros, com a exclusão do período de aquecimento; (ii) as representações gráficas das autocorrelações dos parâmetros; (iii) a monitorização das médias ergódicas de certas quantidades escalares (por exemplo, do primeiro e do segundo momentos ou de alguns percentis); (iv) o diagnóstico de Geweke (Geweke, 1992); (v) o diagnóstico de Gelman e Rubin (Gelman e Rubin, 1992); (vi) o diagnóstico dos quantis de Raftery e Lewis (Raftery e Lewis, 1992). Os *softwares* CODA (Best *et al.*, 1996) e BOA (Smith, 2005) possibilitam a análise do *output* do método MCMC e, em particular, a análise da convergência deste método.

Capítulo 2

Critérios de Selecção de Modelos

2.1 Introdução

Perante uma amostra aleatória de uma dada população, um dos principais objectivos da Estatística é o ajustamento de modelos, com o objectivo de estimar parâmetros populacionais ou prever comportamentos futuros. Na abordagem bayesiana ao problema do ajustamento de modelos é necessário especificar a distribuição de probabilidade conjunta do vector de observações, condicional a um vector de parâmetros, assim como a distribuição *a priori* desse vector de parâmetros. A incerteza inerente à especificação, quer da distribuição do vector de observações, quer da distribuição do vector de parâmetros, proporciona o aparecimento de vários modelos candidatos ao ajustamento dos dados. A selecção de modelos constitui, portanto, um dos problemas mais importantes da Estatística e da investigação científica em geral, na medida em que possibilita a escolha, num conjunto de modelos em comparação, do modelo que melhor descreve o fenómeno aleatório que é objecto de estudo.

Num problema de selecção de modelos, o investigador deve ter sempre presente a necessidade de equilíbrio de dois conceitos distintos: a bondade ou qualidade do ajustamento e a parcimónia do modelo. O primeiro conceito implica a escolha do modelo que melhor descreve a situação que é objecto de investigação e o segundo conceito previne contra a desnecessária complicação do modelo. Um modelo demasiado complicado pode ajustar bem os dados mas ser totalmente inútil, por não permitir uma simplificação e uma melhor percepção do fenómeno em estudo, nem proporcionar, às vezes, boas predições.

Numa situação real que envolva incerteza, a informação disponível deve ser usada para seleccionar o melhor modelo — isto é, para escolher um modelo parcimonioso que possibilite o melhor ajustamento da realidade — no conjunto de modelos candidatos. Pode acontecer que nenhum dos modelos, do conjunto de modelos candidatos, seja um bom modelo e, neste caso, o melhor modelo não é sequer um bom modelo. De modo a evitar que maus modelos sejam aceites, meramente porque ajustam melhor os dados do que outros modelos ainda piores, deve também efectuar-se a validação do modelo, isto é, averiguar se o modelo seleccionado providencia um ajustamento adequado dos dados. A validação de modelos é o assunto tratado no capítulo 5. A selecção e validação de modelos constitui uma etapa da análise estatística muitas vezes designada por determinação de

modelos.

Os testes de hipóteses bayesianos têm a vantagem, sobre os seus homólogos clássicos, de não impor um limite ao número de hipóteses a serem consideradas simultaneamente, nem exigir que estas sejam encaixadas. Por este motivo, a designação “hipótese” é, por vezes, substituída por “modelo”. Não é portanto de estranhar que, sendo o factor Bayes a ferramenta utilizada nos testes de hipóteses, seja também um método eleito para a selecção de modelos. Existem, no entanto, diversas dificuldades na sua aplicação do ponto de vista prático, o que impulsionou o aparecimento de soluções alternativas para a selecção de modelos: outros factores Bayes, critérios definidos à custa dos resíduos, métodos baseados na verosimilhança penalizada, etc. Neste capítulo são descritos e exemplificados alguns destes critérios de selecção de modelos.

Alternativamente à selecção de modelos, os factores Bayes podem ser convertidos em pesos que, “anexados” aos vários modelos, possibilitam a obtenção de estimativas ou predições compostas, que têm em conta a incerteza na selecção do modelo (*e.g.*, Hoeting *et al.*, 1999). Esta metodologia, designada por média ponderada de modelos bayesianos, (abrevia-se BMA, de *Bayesian model averaging*), origina inferências mais precisas, na medida em que é contemplada a incerteza inerente ao processo de selecção de modelos.

Antes de se avançar na exposição dos diversos critérios de selecção de modelos, convém clarificar alguns conceitos e notações utilizados.

Um dos conceitos que pode suscitar alguma dúvida é o conceito de modelo. Entende-se por modelo, no caso mais simples, o modelo bayesiano padrão, definido através de uma hierarquia em dois níveis: a distribuição dos dados, condicional ao vector de parâmetros θ e a distribuição *a priori* de θ . Se a distribuição *a priori* de θ estiver indexada por outro vector de parâmetros η , pode ser necessário especificar essa incerteza através de uma distribuição *a priori* para η . Assim, no caso mais geral, o modelo engloba a distribuição dos dados, condicional ao vector de parâmetros, e os vários níveis de hierarquia referentes à distribuição do vector de parâmetros.

Na apresentação de alguns critérios de selecção de modelos de índole clássica, nomeadamente do critério de informação de Akaike e do critério de informação bayesiano, o conceito de modelo não é o mesmo. Neste caso, o conceito de modelo refere-se apenas à distribuição dos dados, condicional a um vector de parâmetros.

Neste capítulo, θ_i representa o vector de parâmetros do modelo m_i e não, como anteriormente se considerou, a i -ésima componente do vector de parâmetros θ .

A distribuição preditiva *a priori* ou distribuição marginal de x , relativa ao modelo m_i , é representada por $f(x|m_i)$. Esta notação pode suscitar alguma dúvida, uma vez que $f(x|m_i)$ indica apenas que o modelo m_i origina uma determinada forma para a distribuição preditiva *a priori*, o que poderia também ser representado por $f_i(x)$. O mesmo se passa com a distribuição dos dados, condicional ao vector de parâmetros θ_i , para o modelo m_i , que é representada por $f(x|\theta_i, m_i)$ em vez de $f_i(x|\theta_i)$; do mesmo modo, a verosimilhança do vector de parâmetros θ_i , para o modelo m_i , é representada por $L(\theta_i|x, m_i)$ em vez de $L_i(\theta_i|x)$.

2.2 Factor Bayes

Considere-se o conjunto $I = \{1, 2, \dots, M\}$, que supostamente contém o índice i^* , tal que m_{i^*} é o modelo “verdadeiro”, embora desconhecido. Representa-se por $\Pr\{m_i\}$ a probabilidade *a priori* de m_i ser o modelo “verdadeiro”. Uma vez observado o conjunto de dados x , o teorema de Bayes pode ser usado para obter a probabilidade *a posteriori* de m_i ser o modelo “verdadeiro”,

$$\Pr\{m_i|x\} = \frac{f(x|m_i) \Pr\{m_i\}}{f(x)}, \quad i \in I, \quad (2.1)$$

sendo $f(x)$ a distribuição preditiva global, definida por

$$f(x) = \sum_{j=1}^M f(x|m_j) \Pr\{m_j\}, \quad (2.2)$$

e $f(x|m_i)$, $i \in I$, a distribuição preditiva *a priori* de x ou distribuição marginal de x , para o modelo m_i , definida por

$$f(x|m_i) = \int f(x|\theta_i, m_i) h_i(\theta_i) d\theta_i, \quad i \in I. \quad (2.3)$$

Substituindo $f(x)$ em (2.1) pela expressão apresentada em (2.2) obtém-se, para a probabilidade *a posteriori* do modelo m_i , a seguinte expressão:

$$\Pr\{m_i|x\} = \frac{f(x|m_i) \Pr\{m_i\}}{\sum_{j=1}^M f(x|m_j) \Pr\{m_j\}}, \quad i \in I. \quad (2.4)$$

Observado o conjunto de dados x , e caso se pretenda seleccionar um entre vários modelos, o modelo que maximiza $\Pr\{m_i|x\}$ ou, de modo equivalente, o modelo que maximiza $f(x|m_i) \Pr\{m_i\}$, para $i \in I$, deve ser o escolhido. Se as probabilidades *a priori* dos modelos m_i forem iguais, isto é, se $\Pr\{m_i\} = \frac{1}{M}$, então

$$\Pr\{m_i|x\} = \frac{f(x|m_i)}{\sum_{j=1}^M f(x|m_j)};$$

neste caso, o modelo escolhido deve ser o modelo com maior valor de $f(x|m_i)$, sugerindo a utilização desta quantidade nos critérios de selecção de modelos.

Quando os modelos são comparados aos pares, o que constitui uma situação muito usual, o factor Bayes emerge. Considere-se a comparação de dois modelos paramétricos, m_1 e m_2 , com vector de parâmetros θ_1 e θ_2 , respectivamente. Designe-se por $\Pr\{m_1\}$ e $\Pr\{m_2\} = 1 - \Pr\{m_1\}$ as probabilidades *a priori* dos modelos m_1 e m_2 , respectivamente. A observação do conjunto de dados x , que se assume obtido de um dos modelos m_1 ou m_2 , com função de probabilidade $f(x|\theta_1, m_1)$ ou $f(x|\theta_2, m_2)$, possibilita que a opinião *a priori* sobre a plausibilidade relativa dos modelos seja transformada numa opinião *a posteriori*, representada pelas probabilidades *a posteriori* de m_1 e m_2 , $\Pr\{m_1|x\}$ e $\Pr\{m_2|x\} = 1 - \Pr\{m_1|x\}$, respectivamente.

No caso em que se compara apenas dois modelos, a probabilidade *a posteriori* do modelo m_i , apresentada em (2.4), pode ser simplificada,

$$\Pr\{m_i|x\} = \frac{f(x|m_i) \Pr\{m_i\}}{f(x|m_1) \Pr\{m_1\} + f(x|m_2) \Pr\{m_2\}}, \quad i \in \{1, 2\}.$$

Deste modo, supondo que $\Pr\{m_2|x\} \neq 0$, tem-se

$$\frac{\Pr\{m_1|x\}}{\Pr\{m_2|x\}} = \frac{f(x|m_1) \Pr\{m_1\}}{f(x|m_2) \Pr\{m_2\}},$$

e o quociente das probabilidades *a posteriori* dos modelos em competição pode ser obtido através do produto do quociente das probabilidades *a priori* desses modelos pelo factor

$$\text{BF}(x) = \frac{f(x|m_1)}{f(x|m_2)}$$

que, como se viu na secção 1.4.3, é o factor Bayes a favor de m_1 e contra m_2 . Assim, como $\text{BF}(x)$ é o quociente entre as distribuições marginais de x para o modelo m_1 e m_2 , atendendo a (2.3), obtém-se

$$\text{BF}(x) = \frac{\int f(x|\theta_1, m_1) h_1(\theta_1) d\theta_1}{\int f(x|\theta_2, m_2) h_2(\theta_2) d\theta_2}. \quad (2.5)$$

O “melhor” modelo corresponde ao modelo que apresenta o maior valor da distribuição preditiva *a priori* para x . Assim, se $\text{BF}(x) > 1$, os dados x favorecem m_1 , e se $\text{BF}(x) < 1$, os dados x favorecem m_2 . Valores aproximados para a interpretação dos factores Bayes obtidos são apresentados na tabela 1.1.

Existe alguma similaridade entre a estatística razão de verosimilhanças e o factor Bayes, embora neste último caso os parâmetros sejam eliminados por integração em vez de serem por maximização. Já que em (2.5) x está fixo, $f(x|\theta_i, m_i)$ não é mais do que a verosimilhança de θ_i para o modelo m_i e, portanto, o factor Bayes pode ser interpretado como o quociente das médias das verosimilhanças dos parâmetros para os dois modelos, calculadas relativamente às respectivas distribuições *a priori* do vector de parâmetros.

A comparação de M modelos, cada um com vector de parâmetros θ_i , $i \in I$, pode ser feita aos pares. O modelo seleccionado corresponde ao modelo com maior valor da distribuição marginal de x . Considere-se, por exemplo, o caso $M = 3$, em que se pretende comparar os modelos m_1 , m_2 e m_3 ; pode começar-se por comparar os modelos m_1 e m_2 e depois compara-se o modelo seleccionado com o modelo m_3 .

De modo a identificar inequivocamente os modelos comparados, sempre que necessário, representa-se por $\text{BF}_{ij}(x)$ o factor Bayes a favor do modelo m_i e contra o modelo m_j .

Se o objectivo for o cálculo dos factores Bayes de todas as combinações possíveis dos M modelos, não é necessário calcular $\binom{M}{2}$ factores Bayes, pois verifica-se que:

$$\text{BF}_{ij}(x) = \text{BF}_{ik}(x) \times \text{BF}_{kj}(x),$$

para $i, j, k \in I$.

O factor Bayes não depende das probabilidades *a priori* dos modelos mas depende, salvo raras excepções, das distribuições *a priori* dos vectores de parâmetros dos modelos

em comparação. Assim, se a distribuição *a priori* do vector de parâmetros de um dos modelos for imprópria, o que é usual quando se considera distribuições *a priori* não informativas, a correspondente distribuição preditiva *a priori* também pode ser imprópria e, conseqüentemente, o factor Bayes pode não estar definido. Esta é uma das críticas apontadas ao factor Bayes (Gelfand e Dey, 1994).

Na ausência de informação *a priori* substancial sobre o vector de parâmetros, a especificação de uma distribuição *a priori* não informativa, para representar a ignorância *a priori*, parece ser a solução mais adequada. Esta solução pode, no entanto, no caso do espaço-parâmetro ser não limitado e da hipótese nula ser simples — isto é, do modelo estar completamente especificado —, conduzir ao paradoxo de Lindley (Lindley, 1957), como se ilustra no exemplo 2.1. Este paradoxo ocorre quando, à medida que $n \rightarrow \infty$ e independentemente dos dados, o factor Bayes favorece m_1 , quer ele seja ou não o modelo “correcto”. Isto deve-se ao uso inapropriado de algumas distribuições na representação da ignorância *a priori*. Neste caso, um factor Bayes a tender para infinito não quer dizer que exista uma forte evidência fornecida pelos dados a favor do modelo nulo m_1 , mas sim uma forte rejeição da distribuição *a priori* difusa do modelo m_2 (Aitkin, 1991).

Exemplo 2.1

Em Aitkin (1991) é apresentado o seguinte exemplo para ilustrar o paradoxo de Lindley. Considere-se, para a amostra $x = (x_1, x_2, \dots, x_n)$, dois modelos alternativos, $m_1 : X_i \stackrel{\text{iid}}{\sim} N(\mu_1, \sigma^2)$, $i = 1, 2, \dots, n$, com μ_1 e σ^2 conhecidos, e $m_2 : X_i | \mu_2 \stackrel{\text{iid}}{\sim} N(\mu_2, \sigma^2)$, $i = 1, 2, \dots, n$, com $h(\mu_2) = \frac{1}{2C}$, para $\mu_2 \in (-C, C)$, com C grande, e σ^2 conhecido. As distribuições preditivas *a priori* de cada um dos modelos são dadas por:

$$f(x|m_1) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_1)^2 \right\},$$

$$f(x|m_2) = \int_{-C}^C \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_2)^2 \right\} \frac{1}{2C} d\mu_2$$

e o factor Bayes a favor de m_1 e contra m_2 é:

$$\begin{aligned} \text{BF}(x) &= \frac{\exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_1)^2 \right\}}{\int_{-C}^C \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_2)^2 \right\} \frac{1}{2C} d\mu_2} \\ &= 2C \frac{\exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_1)^2 \right] \right\}}{\int_{-C}^C \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_2)^2 \right] \right\} d\mu_2} \\ &= 2C \frac{\exp \left\{ -\frac{n}{2\sigma^2} (\bar{x} - \mu_1)^2 \right\}}{\int_{-C}^C \exp \left\{ -\frac{n}{2\sigma^2} (\bar{x} - \mu_2)^2 \right\} d\mu_2} \\ &= 2C \frac{\frac{\sqrt{n}}{\sigma} \phi(z)}{\Phi \left[\frac{\sqrt{n}}{\sigma} (\bar{x} + C) \right] - \Phi \left[\frac{\sqrt{n}}{\sigma} (\bar{x} - C) \right]}, \end{aligned} \tag{2.6}$$

sendo $z = \frac{\sqrt{n}}{\sigma} (\bar{x} - \mu_1)$, e $\phi(\cdot)$ e $\Phi(\cdot)$ a função densidade de probabilidade e a função distribuição da normal padrão, respectivamente.

O denominador de (2.6) tende para 1 à medida que C aumenta. Assim, o valor do factor Bayes pode ser arbitrariamente grande à medida que $C \rightarrow \infty$ ou $n \rightarrow \infty$, independentemente dos dados (isto é, qualquer que seja o valor de z fixo). O paradoxo de Lindley ocorre porque é atribuído, à medida que $C \rightarrow \infty$ ou $n \rightarrow \infty$, através da distribuição *a priori* $h(\mu_2)$, peso não nulo a intervalos com verosimilhança insignificante do parâmetro μ_2 . Aitkin (1991) refere que a distribuição *a priori* não necessita de ser uniforme; qualquer distribuição *a priori* própria que se considere produz o mesmo resultado, à medida que $n \rightarrow \infty$. ■

Além da eventual ocorrência do paradoxo de Lindley, produzindo um factor Bayes que favorece o modelo m_1 quer ele seja o modelo “correcto” ou não, a utilização de distribuições *a priori* não informativas pode também originar situações em que o próprio factor Bayes não está bem definido. De modo a obviar estes problemas, Aitkin (1991) defende que é mais apropriado efectuar a selecção de modelos através do quociente das médias das verosimilhanças com respeito às distribuições *a posteriori* dos vectores parâmetros, observados os dados x , em vez de usar o quociente das médias das verosimilhanças com respeito às distribuições *a priori*, como acontece em (2.5). Surge deste modo o factor Bayes *a posteriori*, que é dado por

$$\text{BF}_{post}(x) = \frac{\int f(x|\theta_1, m_1) h_1(\theta_1|x) d\theta_1}{\int f(x|\theta_2, m_2) h_2(\theta_2|x) d\theta_2}.$$

O factor Bayes *a posteriori* é alvo de críticas devido ao facto dos dados serem utilizados duas vezes, uma na especificação da distribuição amostral e a outra na determinação da distribuição *a posteriori* (e.g., Cuzick, 1991 e O’Hagan, 1991).

O’Hagan (1991) propõe uma solução para este problema. Dividir o conjunto de dados x em duas partes $x = (x^{(1)}, x^{(2)})$: a amostra $x^{(1)}$ é designada por amostra de treino e é usada para determinar as distribuições *a posteriori* $h_i(\theta_i|x^{(1)})$, que conjuntamente com a informação amostral $f(x^{(2)}|\theta_i, m_i)$, para $i \in \{1, 2\}$, permite calcular o factor Bayes parcial

$$\text{BF}_{parc}(x^{(2)}|x^{(1)}) = \frac{p(x^{(2)}|x^{(1)}, m_1)}{p(x^{(2)}|x^{(1)}, m_2)},$$

onde

$$p(x^{(2)}|x^{(1)}, m_i) = \int f(x^{(2)}|\theta_i, m_i) h_i(\theta_i|x^{(1)}) d\theta_i, \quad i \in \{1, 2\},$$

é a distribuição preditiva *a posteriori* de $x^{(2)}$ condicional a $x^{(1)}$, para o modelo m_i . Embora deste modo se consiga contornar o problema das distribuições *a priori* impróprias, surge outro problema que é necessário solucionar: como dividir a amostra x em duas partes?

Berger e Pericchi (1996) propõem a utilização do factor Bayes intrínseco que, grosso modo, consiste em determinar a média aritmética ou geométrica dos factores Bayes parciais, obtidos usando todas as amostras de treino de dimensão mínima, isto é, de dimensão n_1 , tal que n_1 representa a menor dimensão da amostra que conduz a distribuições *a posteriori* $h_i(\theta_i|x^{(1)})$ próprias.

O’Hagan (1995) sugere a utilização do factor Bayes fraccionário definido por

$$\text{BF}_{frac}(x) = \frac{f(x, b|m_1)}{f(x, b|m_2)}$$

onde $b = n_1/n$, sendo n_1 a já mencionada dimensão de amostra mínima e

$$f(x, b|m_i) = \frac{\int f(x|\theta_i, m_i) h_i(\theta_i) d\theta_i}{\int f(x|\theta_i, m_i)^b h_i(\theta_i) d\theta_i}, \quad i \in \{1, 2\}.$$

De modo a reduzir a sensibilidade do factor Bayes à escolha da distribuição *a priori*, O'Hagan (1995,1997) aconselha a utilização de valores elevados para b , nomeadamente $b = \max(n_1, \ln n)/n$ ou $b = \max(n_1, \sqrt{n})/n$.

O assunto “factor Bayes” não fica de modo algum esgotado. Na secção 2.3.1 é também definido o factor pseudo-Bayes. Cavalcante (1998) faz uma revisão dos factores Bayes *a posteriori*, parcial, intrínseco e fraccionário. Han e Carlin (2001) comparam diversas técnicas, baseadas nos métodos MCMC, para calcular o factor Bayes. Destaca-se ainda, como uma referência obrigatória, o artigo de Kass e Raftery (1995), que apresenta com maior detalhe os assuntos mencionados nesta secção e outros aspectos não referidos, ilustrados com aplicações práticas.

2.3 Resíduos bayesianos

Considere-se que o vector $y = (y_1, y_2, \dots, y_{n^*})$ constitui uma amostra de validação, isto é, uma amostra independente de x (condicionalmente a θ), proveniente do mesmo modelo que gerou os dados x . O valor médio preditivo e a variância preditiva de cada componente de Y , são dados por:

$$E_{Y_i|x}[Y_i] = \int y_i p(y_i|x) dy_i \quad \text{e} \quad Var_{Y_i|x}[Y_i] = \int (y_i - E_{Y_i|x}[Y_i])^2 p(y_i|x) dy_i,$$

respectivamente, para $i = 1, 2, \dots, n^*$, sendo a distribuição preditiva *a posteriori* de Y_i definida por

$$p(y_i|x) = \int_{\Theta} f(y_i|\theta) h(\theta|x) d\theta.$$

A estimação, via métodos MCMC, do valor médio preditivo e da variância preditiva é apresentada no anexo B.2.

O resíduo bayesiano associado à i -ésima observação é a diferença entre o i -ésimo valor da amostra de validação, y_i , e o correspondente valor predito $\hat{y}_i = E_{Y_i|x}[Y_i]$, isto é,

$$r_i = y_i - E_{Y_i|x}[Y_i], \quad i = 1, 2, \dots, n^*. \quad (2.7)$$

De modo a eliminar a dependência dos resíduos da escala dos dados, é usual considerar o resíduo bayesiano padronizado:

$$d_i = \frac{y_i - E_{Y_i|x}[Y_i]}{\sqrt{Var_{Y_i|x}[Y_i]}}, \quad i = 1, 2, \dots, n^*. \quad (2.8)$$

A determinação dos resíduos bayesianos pressupõe a existência de duas amostras independentes, uma para efectuar o ajustamento e a outra para a validação do modelo. No caso de não ser possível obter duas amostras independentes, pode particionar-se o conjunto

de dados x de modo a obter duas amostras $x^{(1)}$ e $x^{(2)}$ que são usadas, respectivamente, como amostras de ajustamento e validação. Surge então outro problema: como particionar a amostra, isto é, quantas observações devem ser usadas para o ajustamento do modelo e quantas devem ser usadas para a validação do modelo? Carlin e Louis (2000, secção 2.4.1) sugerem que se reserve 20% a 30% da amostra para subsequente validação do modelo. No entanto, esta solução não é viável caso a amostra tenha dimensão reduzida.

Quando não é possível obter duas amostras independentes, Gelfand *et al.* (1992) sugerem a utilização de uma metodologia do tipo *Jackknife* (ou *leave one out*) na determinação dos resíduos, que dá origem ao resíduo bayesiano de eliminação. Neste caso, o resíduo bayesiano pode ser determinado à custa de uma única amostra x , pois o valor médio preditivo e a variância preditiva de X_i são calculados condicionalmente ao conjunto de dados x excluído x_i , isto é, condicionalmente a $x_{(-i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$, para $i = 1, 2, \dots, n$.

Seja $h(\theta|x_{(-i)})$ a distribuição *a posteriori* de θ observado o vector de dados $x_{(-i)}$, $f(x_i|x_{(-i)}, \theta)$ a distribuição de X_i condicional a $x_{(-i)}$, isto é,

$$f(x_i|x_{(-i)}, \theta) = \frac{f(x|\theta)}{f(x_{(-i)}|\theta)},$$

e $p(x_i|x_{(-i)})$ a distribuição preditiva de X_i condicional a $x_{(-i)}$, definida por:

$$p(x_i|x_{(-i)}) = \int_{\Theta} f(x_i|x_{(-i)}, \theta) h(\theta|x_{(-i)}) d\theta. \quad (2.9)$$

O valor médio preditivo e a variância preditiva de cada componente de X , condicionalmente a $x_{(-i)}$, são dados por:

$$E_{X_i|x_{(-i)}}[X_i] = \int x_i p(x_i|x_{(-i)}) dx_i,$$

$$Var_{X_i|x_{(-i)}}[X_i] = \int (x_i - E_{X_i|x_{(-i)}}[X_i])^2 p(x_i|x_{(-i)}) dx_i.$$

Com esta metodologia o resíduo r_i , definido em (2.7), é substituído pelo resíduo bayesiano de eliminação

$$r'_i = x_i - E_{X_i|x_{(-i)}}[X_i], \quad i = 1, 2, \dots, n,$$

e o resíduo d_i definido em (2.8) é substituído pelo resíduo bayesiano de eliminação padronizado

$$d'_i = \frac{x_i - E_{X_i|x_{(-i)}}[X_i]}{\sqrt{Var_{X_i|x_{(-i)}}[X_i]}}, \quad i = 1, 2, \dots, n.$$

As somas dos quadrados dos resíduos — $\sum_i^n d_i^2$ ou $\sum_i^n d_i'^2$ — ou as somas dos seus valores absolutos — $\sum_i^n |d_i|$ ou $\sum_i^n |d_i'|$ — podem também ser usadas na selecção de modelos, devendo optar-se pelo modelo que apresentar o menor valor para estas somas.

Se os modelos em competição não tiverem o mesmo número de parâmetros, deve representar-se graficamente no mesmo eixo de coordenadas (ou em gráficos com a mesma

escala para o eixo das abcissas e das ordenadas) os resíduos não padronizados $y_i - E_{Y_i|x}[Y_i]$ versus o desvio padrão $(Var_{Y_i|x}[Y_i])^{1/2}$. Um modelo com mais parâmetros que o necessário será detectado pela presença de variâncias maiores para os resíduos não padronizados do que um modelo com menos parâmetros mas também adequado (Paulino *et al.*, 2003, secção 8.4.2). O mesmo é válido para os resíduos bayesianos de eliminação não padronizados.

2.3.1 Ordenadas preditivas condicionais

O valor de $p(x_i|x_{(-i)})$, definido em (2.9), quando x_i é a i -ésima componente do vector observado x , designa-se por ordenada preditiva condicional da observação x_i e representa-se por $CPO(x_i)$ (de *conditional predictive ordinate*). O modo de estimar, via métodos MCMC, as ordenadas preditivas condicionais é apresentado no anexo B.1.

Gelfand *et al.* (1992), citando Besag (1975), afirmam que se a distribuição preditiva *a priori* $f(x)$ for própria e estritamente positiva no seu domínio, então $f(x)$ é equivalente ao conjunto $\{p(x_i|x_{(-i)}) : i = 1, \dots, n\}$ no sentido em que um determina unicamente o outro. Com base neste resultado, se X_i , para $i = 1, 2, \dots, n$, forem condicionalmente independentes dado θ , Geisser e Eddy (1979) sugerem que se use $\prod_{i=1}^n p(x_i|x_{(-i)})$ em vez da distribuição preditiva *a priori*, na comparação de modelos. Seguindo esta sugestão, Gelfand *et al.* (1992) propõem a substituição, no cálculo do factor Bayes, da distribuição preditiva *a priori* pelo produto das ordenadas preditivas condicionais, que designam por distribuição pseudo-preditiva, dando origem ao factor pseudo-Bayes:

$$BF_{pseudo}(x) = \frac{\prod_{i=1}^n CPO_{m_1}(x_i)}{\prod_{i=1}^n CPO_{m_2}(x_i)},$$

onde $CPO_{m_j}(x_i)$ é a ordenada preditiva condicional da observação x_i para o modelo m_j . Deste modo, o produto das ordenadas preditivas condicionais constitui uma medida de selecção de modelos. O factor pseudo-Bayes pode ser considerado uma versão “validação cruzada” do factor Bayes. A interpretação do valor do factor pseudo-Bayes é idêntica à apresentada para os outros factores Bayes definidos na secção 2.2.

Se $SlnCPO_{m_j}(x)$ representar a soma dos logaritmos das ordenadas preditivas condicionais do modelo m_j , isto é, se $SlnCPO_{m_j}(x) = \sum_{i=1}^n \ln CPO_{m_j}(x_i)$, então o logaritmo do factor pseudo-Bayes é dado por:

$$\ln BF_{pseudo}(x) = SlnCPO_{m_1}(x) - SlnCPO_{m_2}(x).$$

Como $\ln BF_{pseudo}(x)$ representa a alteração do valor do $SlnCPO$ do modelo m_1 para o modelo m_2 , é usual representar esta diferença por $\Delta SlnCPO(x)$. Se $\Delta SlnCPO > 0$, selecciona-se o modelo m_1 ; no caso contrário, selecciona-se m_2 .

Caso se pretenda utilizar o factor pseudo-Bayes com vários modelos em competição, deve seleccionar-se o modelo m_j que apresenta o maior valor de $\prod_{i=1}^n CPO_{m_j}(x_i)$ ou, equivalentemente, o que apresenta o maior valor do $SlnCPO_{m_j}(x)$.

Designa-se por $C(x_i)$ o quociente entre as ordenadas preditivas condicionais da i -ésima observação, dos modelos m_1 e m_2 , isto é, $C(x_i) = CPO_{m_1}(x_i)/CPO_{m_2}(x_i)$. Então, o factor pseudo-Bayes pode ser dado por

$$BF_{pseudo}(x) = \prod_{i=1}^n C(x_i),$$

sendo $C(x_i) > 1$ para as observações x_i que dão preferência ao modelo m_1 e $C(x_i) < 1$ para as observações x_i que dão preferência ao modelo m_2 . A representação gráfica de $\ln C(x_i)$ versus i ou x_i permite ter uma perspectiva global das observações que se encontram melhor ajustadas para cada um dos modelos em competição.

2.4 Critérios de verosimilhança penalizada

Akaike (1973) encontrou uma relação bastante simples entre a discrepância de Kullback-Liebler¹ e o máximo da log-verosimilhança (*e.g.*, Burnham e Anderson, 1998, capítulo 2), que conduziu ao aparecimento de uma metodologia bastante geral para seleccionar modelos parcimoniosos para os dados observados. A discrepância de Kullback-Liebler entre modelos constitui a base teórica, preconizada por Akaike, para a selecção de modelos. Surge, assim, o critério de informação de Akaike (Akaike, 1973), bem como outros critérios de selecção de modelos baseados na verosimilhança, com penalizações impostas ao incremento do número de parâmetros do modelo: o critério de informação bayesiano (Schwarz, 1978), o critério de informação bayesiano estimado (Carlin e Louis, 2000, secção 6.5.1) e o critério de informação da *deviance* (Spiegelhalter *et al.*, 2002).

2.4.1 Critério de informação de Akaike

O critério de informação de Akaike é o mais conhecido dos critérios de selecção de modelos baseados na verosimilhança penalizada. Nesta secção, a notação m_i , isto é, o conceito de modelo, refere-se apenas à distribuição de X , condicional ao vector de parâmetros θ_i , e assim a dimensão de θ_i coincide com p_i , o número de parâmetros do modelo m_i . Designando por $L(\theta_i|x, m_i)$ a verosimilhança do vector de parâmetros $\theta_i \in \Theta_i$ para o modelo m_i , o critério de informação de Akaike consiste em escolher o modelo que apresente o menor valor do

$$\text{AIC}_{m_i}(x) = -2 \ln [\sup_{\Theta_i} L(\theta_i|x, m_i)] + 2p_i,$$

onde AIC abrevia *Akaike information criterion*. Se $\hat{\theta}_i$ for o valor de θ_i que maximiza a verosimilhança, $\sup_{\Theta_i} L(\theta_i|x, m_i) = L(\hat{\theta}_i|x, m_i)$, então

$$\text{AIC}_{m_i}(x) = -2 \ln [L(\hat{\theta}_i|x, m_i)] + 2p_i.$$

A primeira parcela é uma medida do ajustamento do modelo e a segunda parcela é uma penalização ao incremento da complexidade do modelo, permitindo que modelos com diferente número de parâmetros possam ser comparados. Burnham e Anderson (1998, secção 2.7.1) argumentam que esta interpretação do AIC não faz justiça à derivação teórica que lhe deu origem, baseada na distância de Kullback-Liebler.

¹A discrepância (também designada distância) de Kullback-Liebler entre os modelos f e g é definida, para funções contínuas, como o integral (usualmente multidimensional), sempre que exista, $I(f, g) = \int f(x) \ln \left(\frac{f(x)}{g(x)} \right) dx$. É utilizada a notação $I(f, g)$ pois esta discrepância quantifica a “informação” perdida quando g é usada para aproximar f . Encontrar o modelo que origina a menor perda de informação é equivalente a encontrar a função g que minimiza $I(f, g)$.

Quando se pretende comparar os modelos m_1 e m_2 usa-se a diferença

$$\Delta\text{AIC}(x) = \text{AIC}_{m_1}(x) - \text{AIC}_{m_2}(x),$$

onde ΔAIC representa a alteração do valor do AIC do modelo m_1 para o modelo m_2 . Se $\Delta\text{AIC} < 0$, selecciona-se o modelo m_1 ; no caso contrário, selecciona-se m_2 . A quantidade ΔAIC aproxima assintoticamente $-2 \ln \text{BF}$, se a informação da distribuição *a priori* aumentar à mesma razão que a informação da verosimilhança, o que não é uma suposição realista do ponto de vista prático. Este critério tem tendência para manter demasiados parâmetros no modelo, devido ao facto da penalização utilizada para o incremento no número de parâmetros ser demasiado pequena (Carlin e Louis, 2000, secção 2.3.3).

Como os valores do AIC estão numa escala relativa, Burnham e Anderson (1998, secção 2.2) definem $\Delta_i = \text{AIC}_i - \min \text{AIC}$, onde AIC_i é o valor do AIC do i -ésimo modelo e $\min \text{AIC}$ é o menor valor do AIC dos modelos que estão a ser comparados, e recomendam que se use as diferenças Δ_i , em vez dos valores do AIC, dos modelos candidatos. Burnham e Anderson (1998, secção 4.2.1) apresentam uma regra empírica sobre o que constitui uma diferença importante no valor do AIC. Estudos de simulação e a ideia do percentil 0.95 de uma distribuição amostral aplicada a Δ_i levam estes autores a sugerir que modelos com $\Delta_i \geq 4$ não devem ser tidos em conta, uma vez que constituem modelos pouco plausíveis para os dados. Acrescentam que o valor considerado deve ser maior para amostras de pequena dimensão ou quando existem modelos para os quais a razão p/n é muito grande.

Para amostras de dimensão elevada e observações independentes, Burnham e Anderson (1998, secção 4.2.3) sugerem a seguinte interpretação para o valor de Δ_i :

- se $\Delta_i \leq 2$, não existe evidência de que o i -ésimo modelo não é o melhor modelo;
- se $2 < \Delta_i \leq 4$, existe uma evidência fraca de que o i -ésimo modelo não é o melhor modelo;
- se $4 < \Delta_i \leq 7$, existe alguma evidência de que o i -ésimo modelo não é o melhor modelo;
- se $7 < \Delta_i \leq 10$, existe uma evidência forte de que o i -ésimo modelo não é o melhor modelo;
- se $\Delta_i > 10$, existe uma evidência muito forte de que o i -ésimo modelo não é o melhor modelo.

2.4.2 Critério de informação bayesiano

Nesta secção, tal como na secção 2.4.1, a notação m_i , isto é, o conceito de modelo, refere-se apenas à distribuição de X , condicional ao vector de parâmetros θ_i .

Para dar resposta ao problema de selecção de modelos com vectores de parâmetros de diferentes dimensões, Schwarz (1978) defende a utilização do seguinte procedimento: escolher o modelo m_i que apresente o maior valor de

$$\text{SC}_{m_i}(x) = \sup_{\Theta_i} [\ln L(\theta_i|x, m_i)] - \frac{1}{2}p_i \ln n,$$

onde $L(\theta_i|x, m_i)$ representa a verosimilhança do vector de parâmetros θ_i para o modelo m_i , p_i o número de parâmetros do modelo m_i e n a dimensão da amostra. Mas, para cada um dos modelos m_i , o supremo de $L(\theta_i|x, m_i)$ não é mais do que $L(\hat{\theta}_i|x, m_i)$, em que $\hat{\theta}_i$ é o valor de θ_i que maximiza a verosimilhança. Assim, o procedimento apresentado, designado por critério de Schwarz, consiste em escolher o modelo m_i que apresente o maior valor de

$$SC_{m_i}(x) = \ln L(\hat{\theta}_i|x, m_i) - \frac{1}{2}p_i \ln n .$$

Quando se pretende comparar dois modelos, m_1 e m_2 , usa-se a diferença

$$\Delta SC(x) = [\ln L(\hat{\theta}_1|x, m_1) - \ln L(\hat{\theta}_2|x, m_2)] - \frac{1}{2}(p_1 - p_2) \ln n .$$

Se $\Delta SC > 0$, selecciona-se o modelo m_1 ; no caso contrário, selecciona-se m_2 . Para amostras de dimensão elevada, pode obter-se aproximações para $\ln BF$, independentes das distribuições *a priori* consideradas, fazendo uso de ΔSC (Kass e Raftery, 1995).

O critério de informação bayesiano consiste em escolher o modelo m_i que apresente o menor valor de $-2SC_{m_i}(x)$, isto é, o menor valor de $BIC_{m_i}(x)$, com

$$\begin{aligned} BIC_{m_i}(x) &= -2 \ln [\sup_{\Theta_i} L(\theta_i|x, m_i)] + p_i \ln n \\ &= -2 \ln L(\hat{\theta}_i|x, m_i) + p_i \ln n , \end{aligned} \quad (2.10)$$

onde BIC abrevia *Bayesian information criterion*. A primeira parcela de (2.10) é uma medida do ajustamento do modelo e a segunda parcela é uma penalização que permite corrigir uma eventual diferença no número de parâmetros dos modelos.

Para comparar os modelos m_1 e m_2 usa-se a diferença

$$\Delta BIC(x) = BIC_{m_1}(x) - BIC_{m_2}(x) ,$$

onde ΔBIC representa a alteração do valor do BIC do modelo m_1 para o modelo m_2 . Se $\Delta BIC < 0$, selecciona-se o modelo m_1 ; no caso contrário, selecciona-se m_2 .

Embora seja apenas utilizada informação relativa à verosimilhança da amostra, (2.10) é designado por critério de informação bayesiano, pois $\Delta BIC(x)$ constitui uma aproximação de $-2 \ln BF(x)$, independente das distribuições *a priori* consideradas, o que equivale a dizer que $\exp \left\{ -\frac{1}{2} \Delta BIC(x) \right\}$ aproxima o factor Bayes (*e.g.*, Kass e Raftery, 1995).

2.4.3 Critério de informação bayesiano estimado

Volta a considerar-se que m_i engloba a distribuição de X , condicional ao vector de parâmetros, e também os vários níveis de hierarquia referentes à distribuição do vector de parâmetros. Seja ℓ_{m_i} o logaritmo da verosimilhança do vector de parâmetros θ_i para o modelo m_i , isto é, $\ell_{m_i} = \ln L(\theta_i|x, m_i)$. De modo a evitar o cálculo do máximo da verosimilhança por simulação, Carlin e Louis (2000, secção 6.5.1) estimam ℓ_{m_i} através do seu valor esperado *a posteriori*,

$$\hat{\ell}_{m_i} = E_{\theta_i|x}[\ell_{m_i}] = E_{\theta_i|x}[\ln L(\theta_i|x, m_i)] , \quad (2.11)$$

e sugerem a utilização de $-2\hat{\ell}_{m_i}$ como medida de ajustamento relativo do modelo m_i . Para que modelos com vectores de parâmetros de diferentes dimensões possam ser comparados, estes autores sugerem que se use a mesma penalização imposta no BIC. Surge, deste modo, o critério de informação bayesiano estimado, que consiste em escolher o modelo m_i que apresente o menor valor de

$$\widehat{\text{BIC}}_{m_i}(x) = -2\hat{\ell}_{m_i} + p_i \ln n,$$

onde p_i representa o número de parâmetros do modelo m_i . Atendendo a (2.11), a expressão do critério de informação bayesiano estimado é dada por

$$\widehat{\text{BIC}}_{m_i}(x) = -2E_{\theta_i|x}[\ln L(\theta_i|x, m_i)] + p_i \ln n. \quad (2.12)$$

A comparação dos modelos m_1 e m_2 pode ser feita à custa de

$$\Delta\widehat{\text{BIC}}(x) = \widehat{\text{BIC}}_{m_1}(x) - \widehat{\text{BIC}}_{m_2}(x),$$

onde $\Delta\widehat{\text{BIC}}$ representa a alteração do valor do $\widehat{\text{BIC}}$ do modelo m_1 para o modelo m_2 . Se $\Delta\widehat{\text{BIC}} < 0$, selecciona-se o modelo m_1 ; no caso contrário, selecciona-se m_2 .

A comparação de modelos através do critério de informação bayesiano estimado requer o conhecimento do número de observações e também do número de parâmetros desses modelos. No caso dos modelos com vários níveis hierárquicos surge um problema: o que são exactamente p e n ?

Para motivar o leitor para esta questão, Carlin e Louis (2000, secção 6.5.1) apresentam, como exemplo, o caso de um estudo longitudinal, no qual existem s_i observações do paciente i , $i = 1, \dots, m$. Deve considerar-se $n = \sum_i s_i$ (o número total de observações) ou $n = m$ (o número de pacientes)? Se as observações de cada paciente forem independentes, $n = \sum_i s_i$ parece ser a escolha mais apropriada, enquanto se as observações de cada paciente forem perfeitamente correlacionadas deve escolher-se $n = m$. Mas certamente a modelação mais conveniente será tomar n algures entre estes dois valores. Caso se considere m efeitos aleatórios, um para cada paciente, que contribuição pode advir daí para p ? Se os efeitos aleatórios nada tiverem em comum (isto é, se forem essencialmente como efeitos fixos), então m deve ter uma contribuição elevada para p , e p estará próximo de m . Mas, se os dados e a distribuição *a priori* indicarem que eles são essencialmente idênticos, então a contribuição de m para p é pequena, pelo que p deve ser pouco superior a 1.

O problema da definição exacta de p e n pode ser ultrapassado através da utilização do critério de informação da *deviance*, proposto por Spiegelhalter *et al.* (2002).

2.4.4 Critério de informação da *deviance*

A abordagem clássica à validação de modelos pode ser feita à custa da estatística *deviance*, definida por “ $2 \times \log$ -verosimilhança (modelo saturado) $- 2 \times \log$ -verosimilhança (modelo ajustado)”². Dempster (1974) sugere que se utilize a distribuição *a posteriori* da *deviance* na selecção bayesiana de modelos.

²Um modelo saturado é um modelo com tantos parâmetros quantas as observações e que permite um ajustamento perfeito dos dados.

Esta sugestão é seguida por Spiegelhalter *et al.* (2002) ao proporem o critério de informação da *deviance* para a selecção bayesiana de modelos. Estes autores definem a *deviance* bayesiana do modelo m_i , com vector de parâmetros θ_i , que representam por $D_{m_i}(\theta_i)$, como sendo

$$D_{m_i}(\theta_i) = -2 \ln f(x|\theta_i, m_i) + 2 \ln g_i(x), \quad (2.13)$$

onde $g_i(x)$ é uma função apenas dos dados x e $f(x|\theta_i, m_i)$ é a função densidade de probabilidade do modelo m_i com vector de parâmetros θ_i . Quando o objectivo de utilização da *deviance* bayesiana é a selecção de modelos, as funções $g_i(x)$ têm que ser idênticas, de modo a não influenciar na escolha do modelo. Nas aplicações efectuadas considera-se $g_i(x) = 1$.

Neste critério o ajustamento do modelo é sumariado através do valor esperado *a posteriori* da *deviance* bayesiana³,

$$\overline{D_{m_i}(\theta_i)} = E_{\theta_i|x}[D_{m_i}(\theta_i)]. \quad (2.14)$$

As estimativas para $\overline{D_{m_i}(\theta_i)}$ podem obter-se, no WinBUGS, através da monitorização da *deviance*. Quanto melhor for o ajustamento do modelo aos dados, maior é o valor da verosimilhança; deste modo, valores pequenos de $\overline{D_{m_i}(\theta_i)}$ identificam os modelos que ajustam melhor os dados.

Por sua vez a complexidade do modelo é expressa através do número efectivo de parâmetros do modelo $p_{D_{m_i}}$, que Spiegelhalter *et al.* (2002) propõem ser obtido como a diferença entre o valor esperado *a posteriori* da *deviance* bayesiana e o valor da *deviance* bayesiana calculado no valor esperado *a posteriori* de θ_i :

$$p_{D_{m_i}} = E_{\theta_i|x}[D_{m_i}(\theta_i)] - D_{m_i}(E_{\theta_i|x}[\theta_i]).$$

Escrevendo $\bar{\theta}_i = E_{\theta_i|x}[\theta_i]$, o número efectivo de parâmetros do modelo pode ser expresso por:

$$p_{D_{m_i}} = \overline{D_{m_i}(\theta_i)} - D_{m_i}(\bar{\theta}_i). \quad (2.15)$$

A estrutura hierárquica da modelação bayesiana induz uma dependência entre os parâmetros do modelo, originando uma redução — que pode depender da quantidade de dados disponíveis — na dimensionalidade efectiva do vector de parâmetros do modelo. Por este motivo, o valor de $p_{D_{m_i}}$ é muitas vezes inferior ao número total de parâmetros do modelo m_i .

A equação (2.15) pode ser reescrita na forma

$$\overline{D_{m_i}(\theta_i)} = D_{m_i}(\bar{\theta}_i) + p_{D_{m_i}} \quad (2.16)$$

e, assim, a medida de ajustamento do modelo pode ser interpretada como a soma de uma medida clássica de ajustamento por estimativa directa com uma medida da complexidade do modelo. Por este motivo, $\overline{D_{m_i}(\theta_i)}$ é também muitas vezes considerada uma medida de adequabilidade.

³Em Spiegelhalter *et al.* (2001), o valor esperado *a posteriori* de $D_{m_i}(\theta_i)$, quando se considera $g_i(x) = 1$, é simplesmente designado por *deviance*. Existe falta de coerência, nas referências de Estatística Bayesiana, no que se refere à definição de *deviance*, de *deviance* bayesiana e de valor esperado *a posteriori* da *deviance* bayesiana.

Por analogia com os resultados clássicos, Spiegelhalter *et al.* (2002) propõem a utilização do critério de informação da *deviance*, que consiste em escolher o modelo m_i que apresente o menor valor de

$$\text{DIC}_{m_i}(x) = \overline{D_{m_i}(\theta_i)} + p_{D_{m_i}} \quad (2.17)$$

$$= 2\overline{D_{m_i}(\theta_i)} - D_{m_i}(\bar{\theta}_i), \quad (2.18)$$

onde DIC abrevia *deviance information criterion*. O DIC do modelo m_i é definido à custa da medida de ajustamento do modelo $\overline{D_{m_i}(\theta_i)}$, penalizada pela complexidade do modelo $p_{D_{m_i}}$.

Quando se pretende comparar dois modelos, m_1 e m_2 , pode usar-se a diferença

$$\Delta\text{DIC}(x) = \text{DIC}_{m_1}(x) - \text{DIC}_{m_2}(x),$$

onde ΔDIC representa a alteração do valor do DIC do modelo m_1 para o modelo m_2 . Se $\Delta\text{DIC}(x) < 0$, selecciona-se o modelo m_1 ; no caso contrário, selecciona-se m_2 .

A utilização do DIC é particularmente útil quando a complexidade dos modelos comparados dificulta o conhecimento do número de parâmetros dos mesmos.

Atendendo a (2.16) e (2.18), o DIC pode ainda ser dado por

$$\text{DIC}_{m_i}(x) = D_{m_i}(\bar{\theta}_i) + 2p_{D_{m_i}}. \quad (2.19)$$

Spiegelhalter *et al.* (2002) apresentam uma justificação assintótica para o DIC, no caso em que o número de observações aumenta com o número de parâmetros p_i e a distribuição *a priori* é não hierárquica e está completamente especificada. Neste caso, $\text{AIC}_i(x) = D_{m_i}(\hat{\theta}_i) + 2p_i$, onde $\hat{\theta}_i$ representa a estimativa de máxima verosimilhança de θ_i . Esta fórmula é idêntica à expressão (2.19), mas com a média *a posteriori* $\bar{\theta}_i$ substituída pela estimativa de máxima verosimilhança $\hat{\theta}_i$, podendo o DIC ser considerado uma generalização do AIC. No caso em que a distribuição *a priori* de θ_i é dominada pela verosimilhança, devido ao incremento da dimensão da amostra ou pelo facto da distribuição *a priori* ser não informativa, o AIC e o DIC têm valores semelhantes, uma vez que a média *a posteriori* $\bar{\theta}_i$ está muito concentrada em torno da estimativa de máxima verosimilhança $\hat{\theta}_i$.

Tal como os valores dos outros critérios de verosimilhança penalizada, também os valores do DIC não têm significado intrínseco; apenas as diferenças dos valores do DIC entre modelos têm significado.

O cálculo do DIC via métodos MCMC é bastante fácil. Uma estimativa de $\overline{D_{m_i}(\theta_i)}$ pode ser facilmente obtida monitorizando $D_{m_i}(\theta_i)$ e considerando a média amostral dos valores simulados de $D_{m_i}(\theta_i)$. Depois, basta determinar $D_{m_i}(\bar{\theta}_i)$ por estimativa directa, sendo, para isso, necessário monitorizar θ_i por forma a calcular $\bar{\theta}_i$ e obter $D_{m_i}(\bar{\theta}_i)$.

O único inconveniente com o cálculo do DIC via métodos MCMC é o facto de não existir qualquer método eficiente para estimar a $\text{Var}(\text{DIC})$, de modo a possibilitar a determinação do erro de Monte Carlo associado a este critério. Zhu e Carlin (2000) sugerem que se utilize duas amostras independentes, de dimensões G_1 e G_2 , do método MCMC, para o cálculo de $\overline{D_{m_i}(\theta_i)}$ e $D_{m_i}(\bar{\theta}_i)$, respectivamente. Assim, estas duas estimativas são não correlacionadas

e, atendendo a (2.18), conclui-se que

$$\text{Var}[\text{DIC}] = \text{Var}[2\overline{D_{m_i}(\theta_i)} - D_{m_i}(\bar{\theta}_i)] = 4\text{Var}[\overline{D_{m_i}(\theta_i)}] + \text{Var}[D_{m_i}(\bar{\theta}_i)].$$

Zhu e Carlin (2000) estimam $4\text{Var}[\overline{D_{m_i}(\theta_i)}]$ directamente do *output* $\{D^{(g_1)}\}_{g_1=1}^{G_1}$ e recorrem ao método delta multivariado com o *output* $\{D^{(g_2)}\}_{g_2=1}^{G_2}$ para estimar $\text{Var}[D_{m_i}(\bar{\theta}_i)]$. Esta abordagem não deu grandes resultados e estes autores concluem que a abordagem do tipo “força bruta por replicação”, que consiste em replicar N vezes o cálculo do DIC e estimar $\text{Var}[\text{DIC}]$ através da sua variância amostral,

$$\widehat{\text{Var}}[\text{DIC}] = \frac{1}{N-1} \sum_{k=1}^N (\text{DIC}_k - \overline{\text{DIC}})^2,$$

deve ser preferida. Embora este modo de estimar a variância do DIC constitua uma abordagem bastante dispendiosa em termos de tempo computacional, parece ser, para já, o modo mais adequado de obter uma indicação da variabilidade inerente ao DIC.

Segundo referem Spiegelhalter *et al.* (2002), as regras apresentadas em Burnham e Anderson (1998) sobre o que constitui uma diferença importante no valor do AIC, parecem funcionar razoavelmente bem com o DIC. Acrescente-se, ainda, que em Spiegelhalter *et al.* (2001) é salientado “... *if the difference in DIC is, say, less than 5, and the models make very different inferences, then it could be misleading just to report the model with the lowest DIC*”.

2.4.5 Log-verosimilhança *a posteriori*

Quando se considera $g_i(x) = 1$ a *deviance bayesiana*, apresentada em (2.13), simplifica para

$$D_{m_i}(\theta_i) = -2 \ln f(x|\theta_i, m_i).$$

Atendendo a (2.16), o valor esperado *a posteriori* da *deviance bayesiana* pode ser considerado uma medida de adequabilidade do modelo, que se designa por log-verosimilhança *a posteriori* e se representa por LLP (*e.g.*, Paulino *et al.*, 2003, secção 9.4.4):

$$\text{LLP}(x) = \overline{D_{m_i}(\theta_i)} = -2 E_{\theta_i|x}[\ln f(x|\theta_i, m_i)]. \quad (2.20)$$

Os valores mais pequenos de LLP identificam os modelos mais adequados.

Quando se pretende comparar dois modelos, usa-se a diferença

$$\Delta\text{LLP}(x) = \text{LLP}_{m_1}(x) - \text{LLP}_{m_2}(x),$$

onde ΔLLP representa a alteração do valor do LLP do modelo m_1 para o modelo m_2 . Se $\Delta\text{LLP} < 0$, selecciona-se o modelo m_1 ; no caso contrário, selecciona-se m_2 .

2.4.6 Relação entre DIC, $\widehat{\text{BIC}}$ e LLP

Caso se considere $g_i(x) = 1$ na definição da *deviance bayesiana* apresentada em (2.13), tal como acontece nas aplicações efectuadas neste trabalho, é óbvia a relação existente entre DIC, $\widehat{\text{BIC}}$ e LLP.

Atendendo a (2.14), (2.15) e (2.17), tem-se

$$\text{DIC}_{m_i}(x) = -2 E_{\theta_i|x}[\ln L(\theta_i | x, m_i)] + p_{D_{m_i}}, \quad (2.21)$$

sendo

$$p_{D_{m_i}} = -2 E_{\theta_i|x}[\ln L(\theta_i | x, m_i)] + 2 \ln L(\bar{\theta}_i | x, m_i),$$

com $\bar{\theta}_i = E_{\theta_i|x}[\theta_i]$. Deste modo, comparando as expressões (2.21) e (2.12) verifica-se que a primeira parcela — referente à medida de ajustamento do modelo — do DIC e do $\widehat{\text{BIC}}$ coincidem. Essa parcela corresponde, precisamente, à log-verosimilhança *a posteriori*, definida em (2.20). Assim, o $\widehat{\text{BIC}}$, o DIC e o LLP encontram-se relacionados do seguinte modo:

$$\begin{aligned} \widehat{\text{BIC}}_{m_i}(x) &= \text{LLP}_{m_i}(x) + p_i \ln n, \\ \text{DIC}_{m_i}(x) &= \text{LLP}_{m_i}(x) + p_{D_{m_i}}. \end{aligned}$$

2.5 Aplicação a dados de contagem

A distribuição Poisson é usualmente utilizada para descrever fenómenos aleatórios que envolvam a contagem de acontecimentos raros que ocorrem em determinado período de tempo ou espaço e, de um modo geral, para dados de natureza discreta com suporte numerável. Uma das particularidades da distribuição Poisson reside no facto da variância ser igual à média. Esta característica dificulta a aplicação desta distribuição a dados de contagem que têm, muito frequentemente, variância superior à média, sendo mais adequado o ajustamento de uma distribuição binomial negativa. A distribuição binomial negativa pode, assim, ser vista como uma alternativa à distribuição Poisson, pois o suporte da distribuição é o mesmo, mas a existência de um parâmetro adicional permite a especificação independente do valor médio e da variância. Johnson e Kotz (1992) referem que a distribuição Poisson com parâmetro θ_2 é a forma limite da distribuição binomial negativa de parâmetros r e θ_1 , com $r \rightarrow \infty$, $\theta_1 \rightarrow 1$ e $r(1 - \theta_1) \rightarrow \theta_2$.

Para exemplificar alguns dos critérios de selecção de modelos apresentados neste capítulo, supõe-se a existência de uma amostra de dados de contagem e compara-se dois modelos alternativos: o modelo beta/binomial negativo (modelo m_1) e o modelo gama/Poisson (modelo m_2). O modelo beta/binomial negativo define-se para observações x_1, x_2, \dots, x_n independentes, condicionalmente a θ_1 , por

$$m_1 : X_i | \theta_1 \stackrel{\text{iid}}{\sim} \text{BiNeg}(r, \theta_1), \quad \theta_1 \sim \text{Be}(a, b), \quad i = 1, 2, \dots, n.$$

O modelo gama/Poisson define-se para observações x_1, x_2, \dots, x_n independentes, condicionalmente a θ_2 , por

$$m_2 : X_i | \theta_2 \stackrel{\text{iid}}{\sim} \text{Poi}(\theta_2), \quad \theta_2 \sim \text{Ga}(c, d), \quad i = 1, 2, \dots, n.$$

Nas secções 2.5.1 e 2.5.2 são apresentados alguns resultados referentes, respectivamente, aos modelos m_1 e m_2 . As expressões de alguns dos critérios de selecção de modelos, para comparação de m_1 e m_2 , são apresentadas na secção 2.5.3.

2.5.1 Modelo beta/binomial negativo

Considere-se uma sucessão de provas de Bernoulli e seja X_i a variável aleatória que representa o número de insucessos que precedem a ocorrência de r sucessos. A distribuição binomial negativa de parâmetros r e θ_1 , $\text{BiNeg}(r, \theta_1)$, com função de probabilidade definida por

$$f(x_i|\theta_1, m_1) = \binom{x_i + r - 1}{x_i} \theta_1^r (1 - \theta_1)^{x_i}, \quad x_i = 0, 1, 2, \dots, \quad 0 < \theta_1 < 1, \quad r \in \mathbb{N},$$

é adequada para modelar experiências em que se pretende estudar o número de insucessos até à ocorrência de r sucessos, numa sequência de provas de Bernoulli independentes em que a probabilidade de sucesso se mantém constante de prova para prova.

O parâmetro θ_1 é, de um modo geral, desconhecido. Num contexto bayesiano, para fazer inferência sobre o parâmetro θ_1 , é necessário incorporar, através da distribuição *a priori*, qualquer tipo de conhecimento ou juízo que o investigador possa ter sobre o parâmetro. No caso de não existir qualquer conhecimento sobre θ_1 , a informação *a priori* deve ser representada por uma distribuição *a priori* não informativa ou vaga.

Por exemplo, pode considerar-se que θ_1 representa a proporção de atletas esquerdinos na população dos praticantes de basquetebol em Portugal. Um problema que se coloca desde logo é a atribuição de uma distribuição *a priori* ao parâmetro θ_1 . No caso do investigador ser um treinador de basquetebol, a distribuição *a priori* reflectiria, certamente, a experiência por ele adquirida ao longo dos anos enquanto treinador dessa modalidade. Se o investigador fosse um leigo em basquetebol, mas um conhecedor das características da população portuguesa, então a distribuição *a priori* incorporaria essencialmente o conhecimento relativo à proporção de esquerdinos na população portuguesa. Não sendo um conhecedor de basquetebol nem das características da população portuguesa, o investigador pode sempre adoptar uma distribuição *a priori* vaga na família conjugada natural da binomial negativa, que simplifica substancialmente o cálculo da distribuição *a posteriori* e todo o processo inferencial.

Como já foi referido, a família conjugada natural da binomial negativa é a beta, como facilmente se verifica pela forma das respectivas funções de probabilidade. Considere-se, portanto, que θ_1 tem distribuição $\text{Be}(a, b)$, com função densidade de probabilidade

$$h_1(\theta_1) = \frac{1}{B(a, b)} \theta_1^{a-1} (1 - \theta_1)^{b-1}, \quad 0 < \theta_1 < 1, \quad a > 0, \quad b > 0,$$

onde $B(\cdot, \cdot)$ representa a função beta. No decurso desta secção são utilizadas algumas propriedades das distribuições binomial negativa e beta, que se encontram sumariadas no anexo A.

A média preditiva *a priori* de X_i para o modelo m_1 é dada por

$$\begin{aligned} E[X_i|m_1] &= \int_0^1 E[X_i|\theta_1, m_1] h_1(\theta_1) d\theta_1 \\ &= \int_0^1 \frac{r(1 - \theta_1)}{\theta_1} \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \theta_1^{a-1} (1 - \theta_1)^{b-1} d\theta_1 \\ &= \frac{rb}{a - 1} \int_0^1 \frac{\Gamma(a + b)}{\Gamma(a - 1)\Gamma(b + 1)} \theta_1^{(a-1)-1} (1 - \theta_1)^{(b+1)-1} d\theta_1. \end{aligned}$$

Como a função integranda do integral da expressão anterior é a função densidade de probabilidade de uma beta de parâmetros $a - 1$ e $b + 1$, então este integral é unitário e

$$E[X_i|m_1] = \frac{rb}{a-1}, \quad \text{para } a > 1. \quad (2.22)$$

Para determinar a variância preditiva *a priori* de X_i , $Var[X_i|m_1] = E[X_i^2|m_1] - (E[X_i|m_1])^2$, é necessário calcular

$$E[X_i^2|m_1] = \int_0^1 E[X_i^2|\theta_1, m_1] h_1(\theta_1) d\theta_1.$$

Como $E[X_i^2|\theta_1, m_1] = Var[X_i|\theta_1, m_1] + (E[X_i|\theta_1, m_1])^2$ logo

$$\begin{aligned} E[X_i^2|m_1] &= \int_0^1 \left[\frac{r(1-\theta_1)}{\theta_1^2} + \frac{r^2(1-\theta_1)^2}{\theta_1^2} \right] \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta_1^{a-1} (1-\theta_1)^{b-1} d\theta_1 \\ &= \int_0^1 \frac{r(1-\theta_1)}{\theta_1^2} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta_1^{a-1} (1-\theta_1)^{b-1} d\theta_1 + \\ &\quad + \int_0^1 \frac{r^2(1-\theta_1)^2}{\theta_1^2} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta_1^{a-1} (1-\theta_1)^{b-1} d\theta_1 \\ &= \frac{rb(a+b-1)}{(a-1)(a-2)} \int_0^1 \frac{\Gamma(a+b-1)}{\Gamma(a-2)\Gamma(b+1)} \theta_1^{(a-2)-1} (1-\theta_1)^{(b+1)-1} d\theta_1 + \\ &\quad + \frac{r^2b(b+1)}{(a-1)(a-2)} \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a-2)\Gamma(b+2)} \theta_1^{(a-2)-1} (1-\theta_1)^{(b+2)-1} d\theta_1. \end{aligned}$$

A função integranda do integral da primeira parcela é precisamente a função densidade de probabilidade de uma beta de parâmetros $a - 2$ e $b + 1$, donde este integral é unitário. Do mesmo modo, a função integranda do integral da segunda parcela é a função densidade de probabilidade de uma beta de parâmetros $a - 2$ e $b + 2$, donde este integral também é unitário. Portanto

$$\begin{aligned} E[X_i^2|m_1] &= \frac{rb(a+b-1)}{(a-1)(a-2)} + \frac{r^2b(b+1)}{(a-1)(a-2)} \\ &= \frac{rb}{(a-1)(a-2)} [a+b-1+r(b+1)] \end{aligned}$$

e assim obtém-se

$$Var[X_i|m_1] = \frac{rb}{a-1} \left[\frac{a+b+r-1}{a-2} + \frac{rb}{(a-1)(a-2)} \right], \quad \text{para } a > 2. \quad (2.23)$$

A covariância preditiva *a priori* entre X_i e X_j , supostos condicionalmente independentes e identicamente distribuídos, é

$$Cov[X_i, X_j|m_1] = E[X_i X_j|m_1] - E[X_i|m_1] E[X_j|m_1],$$

com

$$\begin{aligned}
 E[X_i X_j | m_1] &= \int_0^1 E[X_i X_j | \theta_1, m_1] h_1(\theta_1) d\theta_1 \\
 &= \int_0^1 E[X_i | \theta_1, m_1] E[X_j | \theta_1, m_1] h_1(\theta_1) d\theta_1 \\
 &= \int_0^1 \left[\frac{r(1-\theta_1)}{\theta_1} \right]^2 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta_1^{a-1} (1-\theta_1)^{b-1} d\theta_1 \\
 &= \frac{r^2 b(b+1)}{(a-1)(a-2)} \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a-2)\Gamma(b+2)} \theta_1^{(a-2)-1} (1-\theta_1)^{(b+2)-1} d\theta_1.
 \end{aligned}$$

Como a função integranda do integral da expressão anterior é a função densidade de probabilidade de uma beta de parâmetros $a-2$ e $b+2$, então este integral é unitário e

$$E[X_i X_j | m_1] = \frac{r^2 b(b+1)}{(a-1)(a-2)}.$$

Deste modo, atendendo a (2.22), obtém-se

$$\text{Cov}[X_i, X_j | m_1] = \frac{r^2 b(b+1)}{(a-1)(a-2)} - \left(\frac{rb}{a-1} \right)^2 = \frac{r^2 b(a+b-1)}{(a-1)^2(a-2)}, \quad \text{para } a > 2.$$

A distribuição preditiva *a priori* de X_i para o modelo m_1 é dada por

$$\begin{aligned}
 f(x_i | m_1) &= \int_0^1 f(x_i | \theta_1, m_1) h_1(\theta_1) d\theta_1 \\
 &= \int_0^1 \binom{x_i+r-1}{x_i} \theta_1^r (1-\theta_1)^{x_i} \frac{1}{B(a,b)} \theta_1^{a-1} (1-\theta_1)^{b-1} d\theta_1 \\
 &= \binom{x_i+r-1}{x_i} \frac{1}{B(a,b)} \int_0^1 \theta_1^{(a+r)-1} (1-\theta_1)^{(b+x_i)-1} d\theta_1
 \end{aligned}$$

Como o integral da expressão anterior é a função beta de parâmetros $a+r$ e $b+x_i$, logo

$$f(x_i | m_1) = \binom{x_i+r-1}{x_i} \frac{B(a+r, b+x_i)}{B(a,b)}, \quad x_i \in \mathbb{N}_0,$$

que é a função de probabilidade da distribuição binomial-negativa-beta de parâmetros $a > 0$, $b > 0$, $r \in \mathbb{N}$ (Bernardo e Smith, 2000, pág. 118).

Seja $x = (x_1, x_2, \dots, x_n)$ uma amostra com n observações, condicionalmente independentes e identicamente distribuídas, do modelo BiNeg(r, θ_1). Neste caso, a função de verossimilhança é dada por

$$L(\theta_1 | x, m_1) = \prod_i^n \binom{x_i+r-1}{x_i} \theta_1^r (1-\theta_1)^{x_i} = \left[\prod_i^n \binom{x_i+r-1}{x_i} \right] \theta_1^{nr} (1-\theta_1)^{n\bar{x}},$$

e o logaritmo da função de verossimilhança é dado por

$$\ln L(\theta_1 | x, m_1) = \sum_i^n \ln \binom{x_i+r-1}{x_i} + nr \ln \theta_1 + n\bar{x} \ln(1-\theta_1). \quad (2.24)$$

Aplicando o teorema de Bayes, a densidade *a posteriori* de θ_1 é dada por:

$$\begin{aligned} h_1(\theta_1|x) &= \frac{f(x|\theta_1, m_1) h_1(\theta_1)}{\int_0^1 f(x|u_1, m_1) h_1(u_1) du_1} \\ &= \frac{\left[\prod_i^n \binom{x_i+r-1}{x_i} \right] \theta_1^{nr} (1-\theta_1)^{n\bar{x}} \frac{1}{B(a,b)} \theta_1^{a-1} (1-\theta_1)^{b-1}}{\int_0^1 \left[\prod_i^n \binom{x_i+r-1}{x_i} \right] u_1^{nr} (1-u_1)^{n\bar{x}} \frac{1}{B(a,b)} u_1^{a-1} (1-u_1)^{b-1} du_1} \\ &= \frac{\theta_1^{(a+nr)-1} (1-\theta_1)^{(b+n\bar{x})-1}}{\int_0^1 u_1^{(a+nr)-1} (1-u_1)^{(b+n\bar{x})-1} du_1}. \end{aligned}$$

Como o integral do denominador da expressão anterior é a função beta de parâmetros $a + nr$ e $b + n\bar{x}$, logo

$$h_1(\theta_1|x) = \frac{1}{B(a + nr, b + n\bar{x})} \theta_1^{(a+nr)-1} (1-\theta_1)^{(b+n\bar{x})-1},$$

donde se conclui que a distribuição *a posteriori* de θ_1 é a distribuição beta de parâmetros $a + nr$ e $b + n\bar{x}$, isto é,

$$\theta_1|x \sim \text{Be}(a + nr, b + n\bar{x}). \quad (2.25)$$

A distribuição preditiva *a priori* ou distribuição marginal de X para o modelo m_1 é dada por:

$$\begin{aligned} f(x|m_1) &= \int_0^1 f(x|\theta_1, m_1) h_1(\theta_1) d\theta_1 \\ &= \int_0^1 \left[\prod_i^n \binom{x_i+r-1}{x_i} \right] \theta_1^{nr} (1-\theta_1)^{n\bar{x}} \frac{1}{B(a,b)} \theta_1^{a-1} (1-\theta_1)^{b-1} d\theta_1 \\ &= \left[\prod_i^n \binom{x_i+r-1}{x_i} \right] \frac{1}{B(a,b)} \int_0^1 \theta_1^{(a+nr)-1} (1-\theta_1)^{(b+n\bar{x})-1} d\theta_1. \end{aligned}$$

Como o integral da expressão anterior é a função beta de parâmetros $a + nr$ e $b + n\bar{x}$, logo

$$f(x|m_1) = \frac{B(a + nr, b + n\bar{x})}{B(a, b)} \prod_i^n \binom{x_i+r-1}{x_i}, \quad x \in \mathbb{N}_0^n. \quad (2.26)$$

O logaritmo de $f(x|m_1)$ é dado por

$$\begin{aligned} \ln f(x|m_1) &= \sum_i^n \ln(x_i + r - 1)! - \sum_i^n \ln x_i! - n \ln(r - 1)! + \ln \Gamma(a + nr) + \\ &\quad + \ln \Gamma(b + n\bar{x}) - \ln \Gamma(a + nr + b + n\bar{x}) - \ln \Gamma(a) - \ln \Gamma(b) + \ln \Gamma(a + b). \end{aligned}$$

2.5.2 Modelo gama/Poisson

A distribuição Poisson de parâmetro θ_2 , $\text{Poi}(\theta_2)$, tem função de probabilidade

$$f(x_i|\theta_2, m_2) = \frac{1}{x_i!} \theta_2^{x_i} e^{-\theta_2}, \quad x_i = 0, 1, 2, \dots, \quad \theta_2 > 0.$$

Como a família conjugada natural da Poisson é a gama, considera-se que θ_2 tem distribuição $\text{Ga}(c, d)$, com função densidade de probabilidade

$$h_2(\theta_2) = \frac{d^c}{\Gamma(c)} \theta_2^{c-1} e^{-d\theta_2}, \quad \theta_2 > 0, \quad c > 0, \quad d > 0.$$

No decurso desta secção são utilizadas algumas propriedades das distribuições Poisson e gama, que se encontram sumariadas no anexo A.

A média preditiva *a priori* de X_i para o modelo m_2 é dada por:

$$\begin{aligned} E[X_i|m_2] &= \int_0^\infty E[X_i|\theta_2, m_2] h_2(\theta_2) d\theta_2 \\ &= \int_0^\infty \theta_2 \frac{d^c}{\Gamma(c)} \theta_2^{c-1} e^{-d\theta_2} d\theta_2 \\ &= \frac{c}{d} \int_0^\infty \frac{d^{c+1}}{\Gamma(c+1)} \theta_2^{(c+1)-1} e^{-d\theta_2} d\theta_2 \end{aligned}$$

Como a função integranda do integral da expressão anterior é a função densidade de probabilidade de uma gama de parâmetros $c + 1$ e d , então este integral é unitário e

$$E[X_i|m_2] = \frac{c}{d}. \quad (2.27)$$

Para determinar a variância preditiva *a priori* de X_i , $\text{Var}[X_i|m_2] = E[X_i^2|m_2] - (E[X_i|m_2])^2$, é necessário calcular

$$E[X_i^2|m_2] = \int_0^\infty E[X_i^2|m_2, \theta_2] h(\theta_2) d\theta_2.$$

Como $E[X_i^2|\theta_2, m_2] = \text{Var}[X_i|\theta_2, m_2] + (E[X_i|\theta_2, m_2])^2$, logo

$$\begin{aligned} E[X_i^2|m_2] &= \int_0^\infty (\theta_2 + \theta_2^2) \frac{d^c}{\Gamma(c)} \theta_2^{c-1} e^{-d\theta_2} d\theta_2 \\ &= \int_0^\infty \theta_2 \frac{d^c}{\Gamma(c)} \theta_2^{c-1} e^{-d\theta_2} d\theta_2 + \int_0^\infty \theta_2^2 \frac{d^c}{\Gamma(c)} \theta_2^{c-1} e^{-d\theta_2} d\theta_2. \end{aligned}$$

A primeira e a segunda parcelas são $E[\theta_2]$ e $E[\theta_2^2]$, donde

$$E[X_i^2|m_2] = \frac{c}{d} + \frac{c(c+1)}{d^2}$$

e assim obtém-se

$$\text{Var}[X_i|m_2] = \frac{c(d+1)}{d^2}. \quad (2.28)$$

A covariância preditiva *a priori* entre X_i e X_j , supostos condicionalmente independentes e identicamente distribuídos, é

$$\text{Cov}[X_i, X_j|m_2] = E[X_i X_j|m_2] - E[X_i|m_2] E[X_j|m_2],$$

sendo

$$\begin{aligned}
 E[X_i X_j | m_2] &= \int_0^\infty E[X_i X_j | \theta_2, m_2] h_2(\theta_2) d\theta_2 \\
 &= \int_0^\infty E[X_i | \theta_2, m_2] E[X_j | \theta_2, m_2] h_2(\theta_2) d\theta_2 \\
 &= \int_0^\infty \theta_2^2 \frac{d^c}{\Gamma(c)} \theta_2^{c-1} e^{-d\theta_2} d\theta_2 \\
 &= \frac{(c+1)c}{d^2} \int_0^\infty \frac{d^{c+2}}{\Gamma(c+2)} \theta_2^{(c+2)-1} e^{-d\theta_2} d\theta_2.
 \end{aligned}$$

Como a função integranda do integral da expressão anterior é a função densidade de probabilidade de uma gama de parâmetros $c+2$ e d , então este integral é unitário e

$$E[X_i X_j | m_2] = \frac{(c+1)c}{d^2}.$$

Deste modo, atendendo a (2.27), obtém-se

$$Cov[X_i, X_j | m_2] = \frac{(c+1)c}{d^2} - \left(\frac{c}{d}\right)^2 = \frac{c}{d^2}.$$

A distribuição preditiva *a priori* de X_i para o modelo m_2 é dada por

$$\begin{aligned}
 f(x_i | m_2) &= \int_0^\infty f(x_i | \theta_2, m_2) h_2(\theta_2) d\theta_2 \\
 &= \int_0^\infty \frac{1}{x_i!} \theta_2^{x_i} e^{-\theta_2} \frac{d^c}{\Gamma(c)} \theta_2^{c-1} e^{-d\theta_2} d\theta_2 \\
 &= \frac{1}{x_i!} \frac{d^c}{\Gamma(c)} \frac{\Gamma(c+x_i)}{(d+1)^{c+x_i}} \int_0^\infty \frac{(d+1)^{c+x_i}}{\Gamma(c+x_i)} \theta_2^{(c+x_i)-1} e^{-(d+1)\theta_2} d\theta_2.
 \end{aligned}$$

Como a função integranda do integral da expressão anterior é a função densidade de probabilidade de uma gama de parâmetros $c+x_i$ e $d+1$, então este integral é unitário e

$$f(x_i | m_2) = \frac{1}{x_i!} \frac{d^c}{\Gamma(c)} \frac{\Gamma(c+x_i)}{(d+1)^{c+x_i}}, \quad x_i \in \mathbb{N}_0,$$

que é a função de probabilidade da distribuição Poisson-gama com parâmetros $c > 0$ e $d > 0$ (Bernardo e Smith, 2000, pág. 119)⁴.

Considere-se uma amostra aleatória de dimensão n , $x = (x_1, x_2, \dots, x_n)$, do modelo $\text{Poi}(\theta_2)$. Neste caso, a função de verossimilhança é dada por

$$L(\theta_2 | x, m_2) = \prod_i^n \frac{1}{x_i!} \theta_2^{x_i} e^{-\theta_2} = \frac{1}{\prod_i^n (x_i!)} \theta_2^{n\bar{x}} e^{-n\theta_2},$$

⁴Esta distribuição, tal como está definida em Bernardo e Smith (2000, pág. 119), tem um parâmetro adicional que, neste caso, é 1.

e o logaritmo da função de verosimilhança é dado por

$$\ln L(\theta_2|x, m_2) = - \sum_i^n \ln x_i! + n\bar{x} \ln \theta_2 - n\theta_2. \quad (2.29)$$

Pelo teorema de Bayes, obtém-se a densidade *a posteriori* de θ_2 para o modelo m_2 :

$$\begin{aligned} h_2(\theta_2|x) &= \frac{f(x|\theta_2, m_2) h_2(\theta_2)}{\int_0^\infty f(x|u_2, m_2) h_2(u_2) du_2} \\ &= \frac{\frac{1}{\prod_i^n (x_i!)} \theta_2^{n\bar{x}} e^{-n\theta_2} \frac{d^c}{\Gamma(c)} \theta_2^{c-1} e^{-d\theta_2}}{\int_0^\infty \frac{1}{\prod_i^n (x_i!)} u_2^{n\bar{x}} e^{-nu_2} \frac{d^c}{\Gamma(c)} u_2^{c-1} e^{-du_2} du_2} \\ &= \frac{\theta_2^{(c+n\bar{x})-1} e^{-(d+n)\theta_2}}{\int_0^\infty u_2^{(c+n\bar{x})-1} e^{-(d+n)u_2} du_2} \\ &= \frac{(d+n)^{c+n\bar{x}} \theta_2^{(c+n\bar{x})-1} e^{-(d+n)\theta_2}}{\int_0^\infty (d+n)^{c+n\bar{x}} u_2^{(c+n\bar{x})-1} e^{-(d+n)u_2} du_2}. \end{aligned}$$

Como o integral do denominador da expressão anterior é a função gama de parâmetro $c + n\bar{x}$, logo

$$h_2(\theta_2|x) = \frac{(d+n)^{c+n\bar{x}}}{\Gamma(c+n\bar{x})} \theta_2^{(c+n\bar{x})-1} e^{-(d+n)\theta_2},$$

donde se conclui que a distribuição *a posteriori* de θ_2 é a distribuição gama de parâmetros $c + n\bar{x}$ e $d + n$, isto é,

$$\theta_2|x \sim \text{Ga}(c + n\bar{x}, d + n). \quad (2.30)$$

A distribuição preditiva *a priori* ou distribuição marginal de X para o modelo m_2 , é dada por

$$\begin{aligned} f(x|m_2) &= \int_0^\infty f(x|\theta_2, m_2) h_2(\theta_2) d\theta_2 \\ &= \frac{1}{\prod_i^n (x_i!)} \frac{d^c}{\Gamma(c)} \int_0^\infty e^{-(d+n)\theta_2} \theta_2^{(c+n\bar{x})-1} d\theta_2 \\ &= \frac{1}{\prod_i^n (x_i!)} \frac{d^c}{\Gamma(c)} \frac{1}{(d+n)^{c+n\bar{x}}} \int_0^\infty e^{-(d+n)\theta_2} \theta_2^{(c+n\bar{x})-1} (d+n)^{c+n\bar{x}} d\theta_2. \end{aligned}$$

Como o integral da expressão anterior é a função gama de parâmetro $c + n\bar{x}$, logo

$$f(x|m_2) = \frac{1}{\prod_i^n (x_i!)} \frac{\Gamma(c+n\bar{x})}{\Gamma(c)} \frac{d^c}{(d+n)^{c+n\bar{x}}}, \quad x \in \mathbb{N}_0^n. \quad (2.31)$$

O logaritmo de $f(x|m_2)$ é dado por

$$\ln f(x|m_2) = - \sum_i^n \ln x_i! + \ln \Gamma(c+n\bar{x}) - \ln \Gamma(c) + c \ln d - (c+n\bar{x}) \ln(d+n).$$

2.5.3 Critérios de selecção dos modelos

Com os modelos m_1 e m_2 , descritos nas secções 2.5.1 e 2.5.2 respectivamente, exemplifica-se o cálculo das expressões de alguns dos critérios de selecção de modelos apresentados, nomeadamente do factor Bayes, do critério de informação bayesiano, do critério de informação bayesiano estimado, do critério de informação da *deviance* e da log-verosimilhança *a posteriori*. No caso do critério de informação bayesiano, que não tem em conta a informação *a priori*, os modelos comparados são, simplesmente, os modelos binomial negativo e Poisson.

Para simplificar a apresentação dos critérios de selecção de modelos, usa-se a_1 e b_1 para representar, respectivamente, $a + nr$ e $b + n\bar{x}$ nas expressões obtidas na secção 2.5.1. De modo análogo, usa-se c_1 e d_1 para representar, respectivamente, $c + n\bar{x}$ e $d + n$ nas expressões obtidas na secção 2.5.2.

Mesmo quando os modelos comparados são bastante simples, como é o caso deste exemplo, a determinação analítica das expressões dos critérios de selecção de modelos pode não ser trivial.

Factor Bayes

Atendendo a que o factor Bayes, a favor do modelo m_1 e contra o modelo m_2 , é o quociente entre as distribuições preditivas *a priori* de X para os modelos m_1 e m_2 , de (2.26) e (2.31), conclui-se que:

$$\text{BF}(x) = \frac{\prod_i^n [(x_i + r - 1)!]}{(r - 1)!^n} \frac{B(a_1, b_1)}{B(a, b)} \frac{\Gamma(c)}{\Gamma(c_1)} \frac{d_1^{c_1}}{d^c}. \quad (2.32)$$

De modo a ultrapassar alguns problemas que surgem na implementação do cálculo do factor Bayes, deve logaritmicar-se $\text{BF}(x)$, obtendo-se:

$$\begin{aligned} \ln \text{BF}(x) &= \sum_i^n \ln(x_i + r - 1)! - n \ln(r - 1)! + \ln \Gamma(a_1) + \ln \Gamma(b_1) - \ln \Gamma(a_1 + b_1) - \\ &\quad - \ln \Gamma(a) - \ln \Gamma(b) + \ln \Gamma(a + b) + \ln \Gamma(c) - \ln \Gamma(c_1) + c_1 \ln d_1 - c \ln d. \end{aligned}$$

Critério de informação bayesiano

Por (2.24) e atendendo a que $\hat{\theta}_1 = \frac{r}{r + \bar{x}}$ e $p_1 = 1$, o critério de informação bayesiano do modelo binomial negativo é dado por

$$\begin{aligned} \text{BIC}_{m_1}(x) &= -2 \sum_i^n \ln \binom{x_i + r - 1}{x_i} - 2nr \ln \left(\frac{r}{r + \bar{x}} \right) - 2n\bar{x} \ln \left(\frac{\bar{x}}{r + \bar{x}} \right) + \ln n \\ &= -2 \sum_i^n \ln(x_i + r - 1)! + 2 \sum_i^n \ln x_i! + 2n \ln(r - 1)! - 2nr \ln \left(\frac{r}{r + \bar{x}} \right) - \\ &\quad - 2n\bar{x} \ln \left(\frac{\bar{x}}{r + \bar{x}} \right) + \ln n. \end{aligned}$$

Por (2.29) e atendendo a que $\hat{\theta}_2 = \bar{x}$ e $p_2 = 1$, o critério de informação bayesiano para o modelo Poisson é dado por

$$\text{BIC}_{m_2}(x) = 2n\bar{x} - 2n\bar{x} \ln \bar{x} + 2 \sum_i^n \ln x_i! + \ln n.$$

Assim, obtém-se:

$$\Delta \text{BIC}(x) = -2 \sum_i^n \ln(x_i + r - 1)! + 2n \ln(r - 1)! - 2nr \ln\left(\frac{r}{r + \bar{x}}\right) + 2n\bar{x} \ln(r + \bar{x}) - 2n\bar{x}.$$

Critério de informação bayesiano estimado

Atendendo a (2.24), o valor esperado *a posteriori* do logaritmo da função de verosimilhança para o modelo m_1 é:

$$\hat{\ell}_1 = \sum_i^n \ln \binom{x_i + r - 1}{x_i} + nr E_{\theta_1|x}[\ln \theta_1] + n\bar{x} E_{\theta_1|x}[\ln(1 - \theta_1)].$$

Como $\theta_1|x \sim \text{Be}(a_1, b_1)$, tendo em conta algumas propriedades da distribuição beta apresentadas no anexo A e sendo ψ a função digama, conclui-se que

$$\hat{\ell}_1 = \sum_i^n \ln \binom{x_i + r - 1}{x_i} + nr [\psi(a_1) - \psi(a_1 + b_1)] + n\bar{x} [\psi(b_1) - \psi(a_1 + b_1)],$$

donde o critério de informação bayesiano estimado do modelo m_1 é dado por

$$\begin{aligned} \widehat{\text{BIC}}_{m_1}(x) &= \\ &= -2 \sum_i^n \ln \binom{x_i + r - 1}{x_i} - 2nr [\psi(a_1) - \psi(a_1 + b_1)] - 2n\bar{x} [\psi(b_1) - \psi(a_1 + b_1)] + \ln n \\ &= -2 \sum_i^n \ln(x_i + r - 1)! + 2 \sum_i^n \ln x_i! + 2n \ln(r - 1)! - 2nr [\psi(a_1) - \psi(a_1 + b_1)] - \\ &\quad - 2n\bar{x} [\psi(b_1) - \psi(a_1 + b_1)] + \ln n. \end{aligned}$$

Atendendo a (2.29), o valor esperado *a posteriori* do logaritmo da função de verosimilhança para o modelo m_2 é:

$$\hat{\ell}_2 = -n E_{\theta_2|x}[\theta_2] + n\bar{x} E_{\theta_2|x}[\ln \theta_2] - \sum_i^n \ln x_i!.$$

Como $\theta_2|x \sim \text{Ga}(c_1, d_1)$, tendo em conta algumas propriedades da distribuição gama apresentadas no anexo A, conclui-se que

$$\hat{\ell}_2 = -n \frac{c_1}{d_1} + n\bar{x} [\psi(c_1) - \ln d_1] - \sum_i^n \ln x_i!,$$

donde o critério de informação bayesiano estimado para o modelo m_2 é

$$\widehat{\text{BIC}}_{m_2}(x) = 2n \frac{c_1}{d_1} - 2n\bar{x}[\psi(c_1) - \ln d_1] + 2 \sum_i^n \ln x_i! + \ln n.$$

Deste modo, obtém-se:

$$\begin{aligned} \Delta \widehat{\text{BIC}}(x) &= -2 \sum_i^n \ln(x_i + r - 1)! + 2n \ln(r - 1)! - 2nr[\psi(a_1) - \psi(a_1 + b_1)] - \\ &\quad - 2n\bar{x}[\psi(b_1) - \psi(a_1 + b_1)] - 2n \frac{c_1}{d_1} + 2n\bar{x}[\psi(c_1) - \ln d_1]. \end{aligned}$$

Critério de informação da *deviance*

Por (2.24) e considerando $g_1(x) = 1$, a *deviance* bayesiana do modelo m_1 é dada por

$$D_{m_1}(\theta_1) = -2 \sum_i^n \ln \binom{x_i + r - 1}{x_i} - 2nr \ln \theta_1 - 2n\bar{x} \ln(1 - \theta_1).$$

Assim, o valor médio *a posteriori* da *deviance* bayesiana é

$$\overline{D_{m_1}(\theta_1)} = -2 \sum_i^n \ln \binom{x_i + r - 1}{x_i} - 2nr E_{\theta_1|x}[\ln \theta_1] - 2n\bar{x} E_{\theta_1|x}[\ln(1 - \theta_1)]$$

e, atendendo a algumas propriedades da distribuição beta apresentadas no anexo A, obtém-se

$$\overline{D_{m_1}(\theta_1)} = -2 \sum_i^n \ln \binom{x_i + r - 1}{x_i} - 2nr[\psi(a_1) - \psi(a_1 + b_1)] - 2n\bar{x}[\psi(b_1) - \psi(a_1 + b_1)]. \quad (2.33)$$

Como $\bar{\theta}_1 = E_{\theta_1|x}[\theta_1]$ e $\theta_1|x \sim \text{Be}(a_1, b_1)$, logo $\bar{\theta}_1 = \frac{a_1}{a_1 + b_1}$, donde se obtém

$$D_{m_1}(\bar{\theta}_1) = -2 \sum_i^n \ln \binom{x_i + r - 1}{x_i} - 2nr \ln \left(\frac{a_1}{a_1 + b_1} \right) - 2n\bar{x} \ln \left(\frac{b_1}{a_1 + b_1} \right).$$

Deste modo, o número efectivo de parâmetros do modelo m_1 é

$$\begin{aligned} p_{D_{m_1}} &= -2nr \left[\psi(a_1) - \psi(a_1 + b_1) - \ln \left(\frac{a_1}{a_1 + b_1} \right) \right] - \\ &\quad - 2n\bar{x} \left[\psi(b_1) - \psi(a_1 + b_1) - \ln \left(\frac{b_1}{a_1 + b_1} \right) \right], \end{aligned}$$

e o critério de informação da *deviance* é dado por

$$\begin{aligned} \text{DIC}_{m_1}(x) &= \\ &= -2 \sum_i^n \ln \binom{x_i + r - 1}{x_i} - 4nr[\psi(a_1) - \psi(a_1 + b_1)] - 4n\bar{x}[\psi(b_1) - \psi(a_1 + b_1)] + \\ &\quad + 2nr \ln \left(\frac{a_1}{a_1 + b_1} \right) + 2n\bar{x} \ln \left(\frac{b_1}{a_1 + b_1} \right). \end{aligned}$$

Por (2.29) e considerando $g_2(x) = 1$, a *deviance* bayesiana do modelo m_2 é dada por

$$D_{m_2}(\theta_2) = 2n\theta_2 - 2n\bar{x} \ln \theta_2 + 2 \sum_i^n \ln x_i!.$$

Segue-se que o valor médio *a posteriori* da *deviance* bayesiana é

$$\overline{D_{m_2}(\theta_2)} = 2n E_{\theta_2|x}[\theta_2] - 2n\bar{x} E_{\theta_2|x}[\ln \theta_2] + 2 \sum_i^n \ln x_i!$$

e, atendendo a algumas propriedades da distribuição gama apresentadas no anexo A, obtém-se

$$\overline{D_{m_2}(\theta_2)} = 2n \frac{c_1}{d_1} - 2n\bar{x}[\psi(c_1) - \ln d_1] + 2 \sum_i^n \ln x_i!. \quad (2.34)$$

Como $\bar{\theta}_2 = E_{\theta_2|x}[\theta_2]$ e $\theta_2|x \sim \text{Ga}(c_1, d_1)$, logo $\bar{\theta}_2 = \frac{c_1}{d_1}$, donde

$$D_{m_2}(\bar{\theta}_2) = 2n \frac{c_1}{d_1} - 2n\bar{x} \ln \left(\frac{c_1}{d_1} \right) + 2 \sum_i^n \ln x_i!.$$

Assim, o número efectivo de parâmetros do modelo m_2 é dado por

$$p_{D_{m_2}} = 2n\bar{x}[\ln c_1 - \psi(c_1)]$$

e o critério de informação da *deviance* para esse modelo é

$$\text{DIC}_{m_2}(x) = 2n \frac{c_1}{d_1} + 2n\bar{x}[\ln c_1 - 2\psi(c_1) + \ln d_1] + 2 \sum_i^n \ln x_i!.$$

É então imediato que:

$$\begin{aligned} \Delta \text{DIC}(x) &= -2 \sum_i^n \ln(x_i + r - 1)! + 2n \ln(r - 1)! - 4nr[\psi(a_1) - \psi(a_1 + b_1)] - \\ &\quad - 4n\bar{x}[\psi(b_1) - \psi(a_1 + b_1)] + 2nr \ln \left(\frac{a_1}{a_1 + b_1} \right) + 2n\bar{x} \ln \left(\frac{b_1}{a_1 + b_1} \right) - \\ &\quad - 2n \frac{c_1}{d_1} - 2n\bar{x}[\ln c_1 - 2\psi(c_1) + \ln d_1]. \end{aligned}$$

Log-verosimilhança *a posteriori*

Como $\text{LLP}_{m_1}(x) = \overline{D_{m_1}(\theta_1)}$ e $\text{LLP}_{m_2}(x) = \overline{D_{m_2}(\theta_2)}$, as expressões (2.33) e (2.34) são, respectivamente, as log-verosimilhanças *a posteriori* para os modelos m_1 e m_2 . É também imediato que:

$$\begin{aligned} \Delta \text{LLP}(x) &= -2 \sum_i^n \ln(x_i + r - 1)! + 2n \ln(r - 1)! - 2nr[\psi(a_1) - \psi(a_1 + b_1)] - \\ &\quad - 2n\bar{x}[\psi(b_1) - \psi(a_1 + b_1) - \psi(c_1) + \ln d_1] - 2n \frac{c_1}{d_1}. \end{aligned}$$

Exemplo 2.2

Para ilustrar a aplicação dos critérios de selecção de modelos apresentados nesta secção, compara-se os modelos beta/binomial negativo e gama/Poisson, para as observações x_1, x_2, \dots, x_{20} , definidos respectivamente por:

$$m_1 : X_i | \theta_1 \stackrel{\text{iid}}{\sim} \text{BiNeg}(2, \theta_1), \quad \theta_1 \sim \text{Be}(3, 3), \quad i = 1, 2, \dots, 20 \quad \text{e}$$

$$m_2 : X_i | \theta_2 \stackrel{\text{iid}}{\sim} \text{Poi}(\theta_2), \quad \theta_2 \sim \text{Ga}(0.3, 0.1), \quad i = 1, 2, \dots, 20.$$

Tal como sugerem Bernardo e Smith (2000, exemplo 6.5), compara-se modelos para os quais as médias preditivas de X_i são iguais: $E[X_i | m_1] = E[X_i | m_2] = 3$. Opta-se por fixar os hiperparâmetros de modo que as variâncias preditivas de X_i dos dois modelos não sejam muito diferentes: $\text{Var}[X_i | m_1] = 30$ e $\text{Var}[X_i | m_2] = 33$.

Da distribuição $\text{Be}(3, 3)$ gera-se dois valores para o parâmetro θ_1 , 0.399 e 0.839. O parâmetro r da binomial negativa que representa o número de sucessos fixa-se em $r = 2$. Simula-se duas amostras de dimensão 20,

$$x_u = (2, 1, 2, 3, 2, 17, 1, 1, 4, 6, 4, 2, 6, 6, 5, 2, 1, 6, 0, 9) \quad \text{e}$$

$$x_v = (0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 2, 1, 0, 1, 0, 1),$$

constituídas por observações independentes e identicamente distribuídas $\text{BiNeg}(2, 0.399)$ e $\text{BiNeg}(2, 0.839)$, respectivamente. De notar que a distribuição $\text{BiNeg}(2, 0.399)$ se distingue, em princípio, da Poisson, enquanto a distribuição $\text{BiNeg}(2, 0.839)$ pode, eventualmente, confundir-se com uma Poisson.

Para a amostra x_u obtém-se $\bar{x}_u = 4$ e $s_u^2 = 14.947$, donde

- a verosimilhança é $L(\theta_1 | x, m_1) \propto \theta_1^{40} (1 - \theta_1)^{80}$, $\theta_1 \in (0, 1)$, e a distribuição *a posteriori* é $\theta_1 | x_u \sim \text{Be}(43, 83)$, no caso do modelo m_1 ;
- a verosimilhança é $L(\theta_2 | x, m_2) \propto \theta_2^{80} e^{-20\theta_2}$, $\theta_2 > 0$, e a distribuição *a posteriori* é $\theta_2 | x_u \sim \text{Ga}(80.3, 20.1)$, no caso do modelo m_2 .

Para a amostra x_v obtém-se $\bar{x}_v = 0.5$ e $s_v^2 = 0.368$, donde

- a verosimilhança é $L(\theta_1 | x, m_1) \propto \theta_1^{40} (1 - \theta_1)^{10}$, $\theta_1 \in (0, 1)$, e a distribuição *a posteriori* é $\theta_1 | x_v \sim \text{Be}(43, 13)$, no caso do modelo m_1 ;
- a verosimilhança é $L(\theta_2 | x, m_2) \propto \theta_2^{10} e^{-20\theta_2}$, $\theta_2 > 0$, e a distribuição *a posteriori* é $\theta_2 | x_v \sim \text{Ga}(10.3, 20.1)$, no caso do modelo m_2 .

Nas figuras 2.1 e 2.2 ilustra-se, para os modelos m_1 e m_2 , a influência da informação amostral na alteração das credibilidades *a priori* sobre os parâmetros θ_1 e θ_2 , respectivamente. Como a verosimilhança pode ser multiplicada por uma constante arbitrária, para a representar graficamente escolhe-se a constante conveniente, de modo a obter área unitária; neste caso, diz-se que a verosimilhança está normalizada. Para ajudar na interpretação destas figuras veja-se Barnett (1999, pág. 226) citando Lindley (1965, pág. 21):

“... if the prior [distribution of θ] ... is sensibly constant over that range of θ for which the likelihood function is appreciable, and not too large over that range of θ for which the likelihood function is small, then the posterior [distribution] ... is approximately equal to the [normalized] likelihood function ...”

Analisa-se, em primeiro lugar, o que se passa no caso do modelo m_1 (figura 2.1). Se x_u for a amostra observada, então a distribuição *a priori* $\text{Be}(3, 3)$ é praticamente constante no intervalo de θ_1 com verosimilhança relevante mas, além disso, verifica-se que grande parte da distribuição *a priori* está concentrada num intervalo de θ_1 com verosimilhança insignificante; por este motivo, a moda da distribuição *a posteriori* de θ_1 é ligeiramente superior à moda da verosimilhança normalizada. Se x_v for a amostra observada, então a distribuição *a priori* $\text{Be}(3, 3)$ não é constante no intervalo de θ_1 com verosimilhança relevante e, além disso, verifica-se que parte da distribuição *a priori* está concentrada num intervalo de θ_1 com verosimilhança insignificante, registando-se uma diferença mais apreciável (do que a verificada no caso da amostra x_u) entre a verosimilhança normalizada e a distribuição *a posteriori* de θ_1 . As modas das distribuições *a posteriori* de θ_1 obtidas com cada uma das amostras apresentam-se “deslocadas”, comparativamente às modas das respectivas verosimilhanças, na direcção da moda da distribuição *a priori* considerada; mas, mesmo assim, a influência da distribuição *a priori* é quase negligível.

Considera-se, agora, o modelo m_2 (figura 2.2). Se x_u for a amostra observada, então a distribuição *a priori* $\text{Ga}(0.3, 0.1)$ é praticamente constante no intervalo de θ_2 com verosimilhança relevante mas, além disso, verifica-se que a distribuição *a priori* está essencialmente concentrada num intervalo de θ_2 com verosimilhança insignificante, sendo a moda da distribuição *a posteriori* de θ_2 ligeiramente inferior à moda da verosimilhança normalizada. Se x_v for a amostra observada, então a distribuição *a priori* $\text{Ga}(0.3, 0.1)$ não é constante no intervalo de θ_2 com verosimilhança relevante e, além disso, verifica-se que parte da distribuição *a priori* está concentrada num intervalo de θ_2 com verosimilhança insignificante, sendo a moda da distribuição *a posteriori* de θ_2 ligeiramente inferior à moda da verosimilhança normalizada.

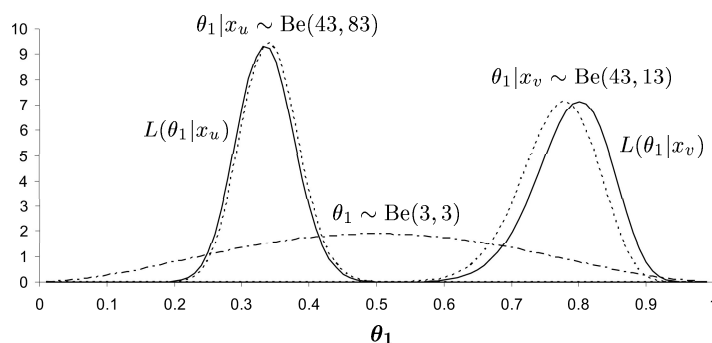


Figura 2.1: Distribuição *a priori*, verosimilhanças (a contínuo) e distribuições *a posteriori* (a tracejado) do parâmetro θ_1 do modelo m_1 , obtidas com as amostras x_u e x_v .

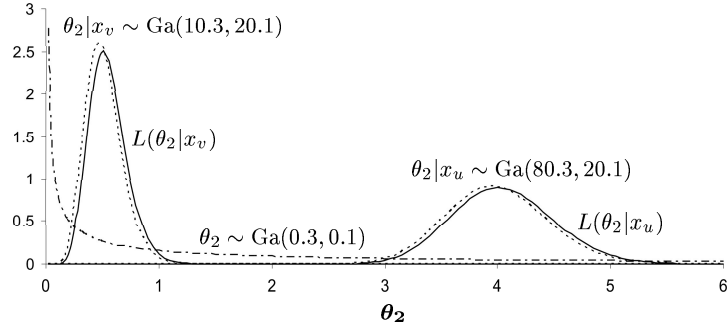


Figura 2.2: Distribuição *a priori*, verossimilhanças (a contínuo) e distribuições *a posteriori* (a tracejado) do parâmetro θ_2 do modelo m_2 , obtidas com as amostras x_u e x_v .

Atendendo a que $r = 2$, $a = 3$, $b = 3$, $c = 0.3$, $d = 0.1$ e $n = 20$, é fácil obter os valores dos critérios de selecção de modelos para as duas amostras, que são apresentados na tabela 2.1. As somas dos logaritmos das ordenadas preditivas condicionais dos modelos m_1 e m_2 obtêm-se por execução do programa em Fortran CPO_bbn_gp, apresentado no ficheiro com o mesmo nome no CD anexo a este documento.

Tabela 2.1: Valores dos critérios de selecção de modelos, obtidos com as amostras x_u e x_v , para os modelos m_1 e m_2 ¹.

amostra	critério	m_1	m_2	Δ critério	
x_u	$\ln f(x \cdot)$	-50.586	-61.165	$\ln \text{BF}(x)$	10.579
	$\text{BIC}(x)$	100.481	119.239	$\Delta \text{BIC}(x)$	-18.758
	$\widehat{\text{BIC}}(x)$	101.460	120.237	$\Delta \widehat{\text{BIC}}(x)$	-18.777
	$\text{DIC}(x)$	99.411	118.240	$\Delta \text{DIC}(x)$	-18.829
	$\text{LLP}(x)$	98.465	117.291	$\Delta \text{LLP}(x)$	-18.826
	$\text{SlnCPO}(x)$	-49.902	-60.575	$\Delta \text{SlnCPO}(x)$	10.673
x_v	$\ln f(x \cdot)$	-20.523	-19.905	$\ln \text{BF}(x)$	-0.618
	$\text{BIC}(x)$	39.748	38.245	$\Delta \text{BIC}(x)$	1.503
	$\widehat{\text{BIC}}(x)$	40.866	39.238	$\Delta \widehat{\text{BIC}}(x)$	1.628
	$\text{DIC}(x)$	38.687	37.228	$\Delta \text{DIC}(x)$	1.459
	$\text{LLP}(x)$	37.870	36.751	$\Delta \text{LLP}(x)$	1.119
	$\text{SlnCPO}(x)$	-19.180	-18.516	$\Delta \text{SlnCPO}(x)$	-0.663

¹No caso do critério de informação bayesiano os modelos comparados são $m_1 : X_i | \theta_1 \stackrel{\text{iid}}{\sim} \text{BiNeg}(2, \theta_1)$, para $i = 1, 2, \dots, 20$, e $m_2 : X_i | \theta_2 \stackrel{\text{iid}}{\sim} \text{Poi}(\theta_2)$, para $i = 1, 2, \dots, 20$.

No caso da amostra x_u , qualquer um dos critérios favorece o modelo beta/binomial negativo com uma evidência bastante forte; como a distribuição $\text{BiNeg}(2, 0.399)$ se distingue, em princípio, da Poisson não é de estranhar esta opção dos critérios de selecção de modelos por m_1 . No caso da amostra x_v , todos os critérios favorecem o modelo gama/Poisson, embora com uma evidência bastante fraca; esta situação ocorre porque a distribuição $\text{BiNeg}(2, 0.839)$ pode, eventualmente, confundir-se com uma Poisson. ■

2.6 Média ponderada de modelos bayesianos

Na análise estatística de alguns problemas pode haver motivos científicos subjacentes que justifiquem a opção por um só modelo. Noutros casos, o processo de selecção de modelos pode não conduzir à escolha de um único modelo mas sim de vários. A solução ideal para esta situação consiste em efectuar as inferências à custa da média ponderada desses modelos, em vez de as obter simplesmente à custa de um único modelo. De facto, em situações deste tipo escolher um único modelo pode não ser a solução óptima, pois as inferências, condicionais ao modelo escolhido, ignoram a incerteza inerente ao processo de selecção de modelos. Isto conduz à subestimação da incerteza sobre o vector de parâmetros de interesse, isto é, as estimativas obtidas para o vector de parâmetros de interesse são estimadas com demasiada confiança.

Considere-se o conjunto de modelos candidatos $\mathcal{M} = \{m_1, m_2, \dots, m_M\}$, sendo as probabilidades *a priori* e *a posteriori* do modelo m_i dadas, respectivamente, por $\Pr\{m_i\}$ e $\Pr\{m_i|x\}$, para $i = 1, 2, \dots, M$. Seja Δ uma quantidade de interesse (vector de parâmetros ou observações futuras) que está bem definida para todos os modelos do conjunto \mathcal{M} .

Cada um dos modelos m_i , para $i \neq 1$, pode ser comparado com o modelo m_1 , dando origem aos factores Bayes $\text{BF}_{21}(x)$, $\text{BF}_{31}(x)$, \dots , $\text{BF}_{M1}(x)$. Então, atendendo a (2.4), a probabilidade *a posteriori* do modelo m_i pode ser reescrita por

$$\Pr\{m_i|x\} = \frac{\frac{f(x|m_i)}{f(x|m_1)} \frac{\Pr\{m_i\}}{\Pr\{m_1\}}}{\sum_{j=1}^M \frac{f(x|m_j)}{f(x|m_1)} \frac{\Pr\{m_j\}}{\Pr\{m_1\}}}.$$

Como $\frac{\Pr\{m_i\}}{\Pr\{m_1\}}$ é a razão das vantagens *a priori* de m_i relativamente a m_1 e $\frac{f(x|m_i)}{f(x|m_1)}$ é o factor Bayes a favor de m_i e contra m_1 , então a igualdade anterior pode ser expressa por

$$\Pr\{m_i|x\} = \frac{O(m_i, m_1) \text{BF}_{i1}(x)}{\sum_{j=1}^M O(m_j, m_1) \text{BF}_{j1}(x)}, \quad (2.35)$$

para $i = 1, 2, \dots, M$; obviamente $\text{BF}_{11}(x) = 1$ e $O(m_1, m_1) = 1$. Na ausência de informação *a priori* sobre a plausibilidade relativa dos modelos em competição, a escolha natural consiste em considerar que as probabilidades *a priori* de todos os modelos são iguais, isto é, que $O(m_i, m_1) = 1$; neste caso,

$$\Pr\{m_i|x\} = \frac{\text{BF}_{i1}(x)}{\sum_{j=1}^M \text{BF}_{j1}(x)},$$

para $i = 1, 2, \dots, M$. Outros valores para $O(m_i, m_1)$ podem ser usados de modo a incorporar o conhecimento *a priori* sobre a plausibilidade relativa dos modelos. As probabilidades *a posteriori* dos modelos, dadas por (2.35), podem ser utilizadas como ponderações, que associadas às distribuições *a posteriori* da quantidade de interesse Δ , sob a validade de cada um dos modelos, permitem a realização de inferências que têm em conta a incerteza na escolha do modelo.

A distribuição *a posteriori* da quantidade de interesse Δ , sob a validade do modelo m_i , é

$$h(\Delta|x, m_i) = \int h(\Delta|x, \theta_i, m_i) h(\theta_i|x, m_i) d\theta_i.$$

A distribuição $h(\Delta|x, m_i)$ pode ser usada para fazer inferências sobre o parâmetro Δ condicionalmente ao modelo m_i . Mas este não é o procedimento adoptado no caso da média ponderada de modelos bayesianos. No caso da metodologia BMA, as inferências são efectuadas à custa da distribuição *a posteriori* de Δ sem condicionar no modelo m_i , $h(\Delta|x)$, que não é mais do que uma mistura das distribuições $h(\Delta|x, m_i)$, ponderada pela probabilidade *a posteriori* de cada um dos modelos m_i :

$$h(\Delta|x) = \sum_{i=1}^M h(\Delta|x, m_i) \Pr\{m_i|x\}. \quad (2.36)$$

Como já foi referido, as inferências realizadas à custa de $h(\Delta|x)$ têm em conta a incerteza no processo de selecção do modelo.

A média e a variância *a posteriori* de Δ são dadas (*e.g.*, Kass e Raftery, 1995), respectivamente, por:

$$E[\Delta|x] = \sum_{i=1}^M E[\Delta|x, m_i] \Pr\{m_i|x\} \quad (2.37)$$

e

$$Var[\Delta|x] = \sum_{i=1}^M \left\{ Var[\Delta|x, m_i] + (E[\Delta|x, m_i])^2 \right\} \Pr\{m_i|x\} - (E[\Delta|x])^2.$$

Como $E[\Delta^2|x, m_i] = Var[\Delta|x, m_i] + (E[\Delta|x, m_i])^2$, logo a $Var[\Delta|x]$ é dada por:

$$Var[\Delta|x] = \sum_{i=1}^M E[\Delta^2|x, m_i] \Pr\{m_i|x\} - (E[\Delta|x])^2. \quad (2.38)$$

Na maior parte das aplicações estatísticas, quando existem vários modelos em competição, a abordagem usual consiste em escolher um modelo, que supõe-se ser o “melhor” modelo. A equação (2.36) mostra que escolher um único modelo e efectuar as inferências condicionalmente ao modelo seleccionado só é razoável se a probabilidade *a posteriori* de um dos modelos for próxima de 1 ou se a soma em (2.36) for dominada por modelos m_i , cujas distribuições $h(\Delta|x, m_i)$ sejam idênticas (Kass e Raftery, 1995).

No exemplo 2.3 ilustra-se a aplicação da metodologia BMA numa situação simples, em que é possível determinar analiticamente o valor médio e a variância *a posteriori* da quantidade de interesse via média ponderada de modelos bayesianos. Exemplos da aplicação desta metodologia a modelos de regressão linear normal podem ser encontrados em Raftery *et al.* (1997) e Hoeting *et al.* (1999).

Exemplo 2.3

Utiliza-se os modelos beta/binomial negativo e gama/Poisson definidos, respectivamente, nas secções 2.5.1 e 2.5.2, para exemplificar a média ponderada de modelos bayesianos. Considere-se uma amostra de dados de contagem de dimensão n , $x = (x_1, \dots, x_n)$, e os modelos $m_1 : X_i|\theta_1 \stackrel{\text{iid}}{\sim} \text{BiNeg}(r, \theta_1)$, para $i = 1, 2, \dots, n$, $\theta_1 \sim \text{Be}(a, b)$, e $m_2 : X_i|\theta_2 \stackrel{\text{iid}}{\sim} \text{Poi}(\theta_2)$, para $i = 1, 2, \dots, n$, $\theta_2 \sim \text{Ga}(c, d)$. Supõe-se que as probabilidades *a priori* dos modelos m_1 e m_2 são iguais. Para ilustrar a metodologia

BMA, considera-se a quantidade de interesse $\Delta = E[X]$ e determina-se a média *a posteriori* e a variância *a posteriori* da quantidade de interesse Δ , com base na média ponderada dos modelos m_1 e m_2 .

Como se está a considerar apenas dois modelos, a expressão (2.37), referente ao valor médio *a posteriori* de Δ , simplifica para

$$\begin{aligned} E[\Delta|x] &= E[\Delta|x, m_1] \Pr\{m_1|x\} + E[\Delta|x, m_2] \Pr\{m_2|x\} \\ &= E[\Delta|x, m_1] (1 - \Pr\{m_2|x\}) + E[\Delta|x, m_2] \Pr\{m_2|x\} \\ &= E[\Delta|x, m_1] + (E[\Delta|x, m_2] - E[\Delta|x, m_1]) \Pr\{m_2|x\}, \end{aligned} \quad (2.39)$$

sendo a probabilidade *a posteriori* do modelo m_2 dada por

$$\Pr\{m_2|x\} = \frac{\text{BF}_{21}(x)}{1 + \text{BF}_{21}(x)}$$

e, atendendo a (2.32), o factor Bayes a favor de m_2 e contra m_1 dado por

$$\text{BF}_{21}(x) = \frac{(r-1)!^n}{\prod_i^n (x_i + r - 1)!} \frac{B(a, b)}{B(a_1, b_1)} \frac{\Gamma(c_1)}{\Gamma(c)} \frac{d^c}{d_1^{c_1}}.$$

Assim, para determinar $E[\Delta|x]$, é necessário calcular $E[\Delta|x, m_1]$ e $E[\Delta|x, m_2]$.

No caso do modelo m_1 , a quantidade de interesse Δ é dada por $\Delta = \frac{r(1-\theta_1)}{\theta_1}$, logo

$$E[\Delta|x, m_1] = E_{\theta_1|x} \left[\frac{r(1-\theta_1)}{\theta_1} \right] = \int_0^1 \frac{r(1-\theta_1)}{\theta_1} h(\theta_1|x) d\theta_1.$$

Escrevendo $a_1 = a + nr$ e $b_1 = b + n\bar{x}$ tem-se, por (2.25), que $\theta_1|x \sim \text{Be}(a_1, b_1)$. Assim,

$$\begin{aligned} E[\Delta|x, m_1] &= \int_0^1 \frac{r(1-\theta_1)}{\theta_1} \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1) \Gamma(b_1)} \theta_1^{a_1-1} (1-\theta_1)^{b_1-1} d\theta_1 \\ &= r \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1) \Gamma(b_1)} \int_0^1 \theta_1^{(a_1-1)-1} (1-\theta_1)^{(b_1+1)-1} d\theta_1 \\ &= r \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1) \Gamma(b_1)} \frac{\Gamma(a_1 - 1) \Gamma(b_1 + 1)}{\Gamma(a_1 + b_1)} \\ &= \frac{rb_1}{a_1 - 1}. \end{aligned}$$

No caso do modelo m_2 , a quantidade de interesse Δ é dada por $\Delta = \theta_2$. Pondo $c_1 = c + n\bar{x}$ e $d_1 = d + n$ tem-se, por (2.30), que $\theta_2|x \sim \text{Ga}(c_1, d_1)$. Deste modo,

$$E[\Delta|x, m_2] = E_{\theta_2|x}[\theta_2] = \frac{c_1}{d_1}.$$

Assim, substituindo as expressões de $E[\Delta|x, m_1]$ e $E[\Delta|x, m_2]$ em (2.39), obtém-se

$$E[\Delta|x] = \frac{rb_1}{a_1 - 1} + \left(\frac{c_1}{d_1} - \frac{rb_1}{a_1 - 1} \right) \Pr\{m_2|x\}.$$

Como se está a considerar apenas dois modelos, a expressão (2.38), referente à variância *a posteriori* de Δ , simplifica para

$$\begin{aligned}
 Var[\Delta|x] &= \\
 &= E[\Delta^2|x, m_1] \Pr\{m_1|x\} + E[\Delta^2|x, m_2] \Pr\{m_2|x\} - (E[\Delta|x])^2 \\
 &= E[\Delta^2|x, m_1] (1 - \Pr\{m_2|x\}) + E[\Delta^2|x, m_2] \Pr\{m_2|x\} - (E[\Delta|x])^2 \\
 &= E[\Delta^2|x, m_1] - (E[\Delta|x])^2 + (E[\Delta^2|x, m_2] - E[\Delta^2|x, m_1]) \Pr\{m_2|x\}. \quad (2.40)
 \end{aligned}$$

Para determinar $Var[\Delta|x]$ é, portanto, necessário calcular $E[\Delta^2|x, m_1]$ e $E[\Delta^2|x, m_2]$.

No caso do modelo m_1 o valor esperado *a posteriori* de Δ^2 é

$$E[\Delta^2|x, m_1] = E_{\theta_1|x} \left[\left(\frac{r(1-\theta_1)}{\theta_1} \right)^2 \right] = \int_0^1 \left(\frac{r(1-\theta_1)}{\theta_1} \right)^2 h(\theta_1|x) d\theta_1$$

e, como $\theta_1|x \sim \text{Be}(a_1, b_1)$, então

$$\begin{aligned}
 E[\Delta^2|x, m_1] &= \int_0^1 \frac{r^2(1-\theta_1)^2}{\theta_1^2} \frac{\Gamma(a_1+b_1)}{\Gamma(a_1)\Gamma(b_1)} \theta_1^{a_1-1} (1-\theta_1)^{b_1-1} d\theta_1 \\
 &= r^2 \frac{\Gamma(a_1+b_1)}{\Gamma(a_1)\Gamma(b_1)} \int_0^1 \theta_1^{(a_1-2)-1} (1-\theta_1)^{(b_1+2)-1} d\theta_1 \\
 &= r^2 \frac{\Gamma(a_1+b_1)}{\Gamma(a_1)\Gamma(b_1)} \frac{\Gamma(a_1-2)\Gamma(b_1+2)}{\Gamma(a_1+b_1)} \\
 &= \frac{r^2 b_1 (b_1 + 1)}{(a_1 - 1)(a_1 - 2)}.
 \end{aligned}$$

No caso do modelo m_2 , como $\theta_2|x \sim \text{Ga}(c_1, d_1)$, o valor esperado *a posteriori* de Δ^2 é:

$$E[\Delta^2|x, m_2] = E_{\theta_2|x}[\theta_2^2] = Var_{\theta_2|x}[\theta_2] + (E_{\theta_2|x}[\theta_2])^2 = \frac{c_1}{d_1^2} + \left(\frac{c_1}{d_1} \right)^2 = \frac{c_1 + c_1^2}{d_1^2}.$$

Assim, substituindo as expressões de $E[\Delta^2|x, m_1]$ e $E[\Delta^2|x, m_2]$ em (2.40), obtém-se

$$\begin{aligned}
 Var[\Delta|x] &= \frac{r^2 b_1 (b_1 + 1)}{(a_1 - 1)(a_1 - 2)} - \left[\frac{r b_1}{a_1 - 1} + \left(\frac{c_1}{d_1} - \frac{r b_1}{a_1 - 1} \right) \Pr\{m_2|x\} \right]^2 + \\
 &+ \left[\frac{c_1 + c_1^2}{d_1^2} - \frac{r^2 b_1 (b_1 + 1)}{(a_1 - 1)(a_1 - 2)} \right] \Pr\{m_2|x\}.
 \end{aligned}$$

Para concretizar este exemplo, considere-se uma amostra de dados de contagem, de dimensão $n = 20$, gerada da distribuição binomial negativa com $r = 2$ e $\theta_1 = 0.4$,

$$x = (3, 9, 5, 3, 3, 6, 3, 4, 2, 8, 2, 2, 4, 4, 5, 2, 7, 5, 4, 9).$$

As estimativas para o valor médio e variância *a posteriori* de Δ , obtidas através da média ponderada dos modelos m_1 e m_2 , em três casos distintos de especificação dos hiperparâmetros, são apresentadas na tabela 2.2.

Tabela 2.2: Exemplo ilustrativo da média ponderada de modelos bayesianos¹.

	caso 1	caso 2	caso 3
$BF_{21}(x)$	1.047	0.040	0.238
$\Pr\{m_1 x\}$	0.489	0.962	0.807
$\Pr\{m_2 x\}$	0.511	0.038	0.193
$E[\Delta x, m_1]$	4.279	4.333	4.476
$E[\Delta x, m_2]$	4.136	3.957	4.136
$E[\Delta x]$	4.206	4.319	4.411
$Var[\Delta x, m_1]$	0.640	0.669	0.707
$Var[\Delta x, m_2]$	0.188	0.172	0.188
$Var[\Delta x]$	0.414	0.656	0.625

¹No caso 1, tem-se $a = 4$, $b = 2$, $c = 1$ e $d = 2$; no caso 2, tem-se $a = 3$, $b = 1$, $c = 1$ e $d = 3$; no caso 3, tem-se $a = 3$, $b = 4$, $c = 1$ e $d = 2$.

No caso 1, as probabilidades *a posteriori* dos dois modelos são idênticas; assim, o valor médio *a posteriori* de Δ é aproximadamente igual à média de $E[\Delta|x, m_1]$ e $E[\Delta|x, m_2]$, e a variância *a posteriori* de Δ é aproximadamente igual à média de $Var[\Delta|x, m_1]$ e $Var[\Delta|x, m_2]$. No caso 2, a probabilidade *a posteriori* do modelo m_1 está próxima de 1 e, assim, as inferências efectuadas com a metodologia BMA são muito idênticas às obtidas com o modelo m_1 . No caso 3, a probabilidade *a posteriori* do modelo m_1 é bastante superior à do modelo m_2 ; assim, neste caso, o valor médio *a posteriori* de Δ e a variância *a posteriori* de Δ têm valores mais próximos, respectivamente, de $E[\Delta|x, m_1]$ e $Var[\Delta|x, m_1]$, do que de $E[\Delta|x, m_2]$ e $Var[\Delta|x, m_2]$. ■

Hoeting *et al.* (1999) apontam quatro factores responsáveis pela pouca utilização da metodologia BMA nas aplicações práticas:

1. o número de modelos candidatos ser, de um modo geral, bastante elevado, o que dificulta o cálculo directo da soma em (2.36), devido ao número elevado de parcelas envolvidas;
2. o cálculo dos integrais intervenientes em (2.36) poder exigir a utilização de métodos MCMC;
3. a especificação da distribuição *a priori* de cada um dos possíveis modelos poder constituir uma tarefa árdua;
4. a escolha do conjunto de modelos \mathcal{M} sobre o qual deve ser efectuada a média ponderada de modelos bayesianos nem sempre ser óbvia.

Para contornar a primeira dificuldade de utilização desta metodologia, Madigan e Raftery (1994) apresentam um método, baseado na escolha de um menor número de parcelas para a soma em (2.36), mas que permite obter um resultado idêntico. Estes

autores sugerem que se limite o conjunto de modelos ao conjunto \mathcal{A}' de modelos que têm uma probabilidade *a posteriori* mínima (*Occam's Window*), isto é, ao conjunto

$$\mathcal{A}' = \left\{ m_i : \frac{\max_{1 \leq j \leq M} \Pr\{m_j|x\}}{\Pr\{m_i|x\}} \leq C \right\},$$

com $C = 20$, por exemplo. Posteriormente é eliminado o conjunto

$$\mathcal{B} = \left\{ m_i : \exists m_j \in \mathcal{A}', m_j \subset m_i, \frac{\Pr\{m_j|x\}}{\Pr\{m_i|x\}} > 1 \right\},$$

constituído pelos modelos que têm menor probabilidade *a posteriori* que os seus próprios submodelos (*Occam's Razor*). Esta estratégia reduz substancialmente o número de parcelas envolvidas em (2.36), pois a soma é efectuada em relação a um conjunto de índices $I_{\mathcal{A}}$, com $I_{\mathcal{A}} \subseteq I$, que contém apenas os índices correspondentes aos modelos pertencentes ao conjunto $\mathcal{A} = \mathcal{A}' \setminus \mathcal{B}$.

A descrição detalhada deste e doutros métodos de identificação dos modelos a considerar na soma em (2.36), ou doutro tipo de abordagens que permitem aproximar directamente (2.36), pode ser encontrada em Kass e Raftery (1995) e Hoeting *et al.* (1999).

Relativamente ao segundo ponto, para algumas classes de modelos os integrais envolvidos em (2.36) podem ser resolvidos analiticamente e, caso isso não seja possível, outros modos de aproximar estes integrais são sugeridos em Hoeting *et al.* (1999). O terceiro motivo pode ser contornado através da utilização de uma regra automática, por exemplo, $\Pr\{m_i\} = 1/M$, para todo $i \in I$. Esta suposição pode ser considerada válida no caso de existir pouca informação sobre a plausibilidade relativa dos modelos. No entanto, quando existe alguma informação *a priori* sobre a plausibilidade dos modelos a utilização desta regra reduz a performance preditiva e soluções alternativas devem ser utilizadas. Algumas sugestões para resolver ou contornar este problema, bem como o problema da escolha do conjunto inicial de modelos, podem ser encontradas em Hoeting *et al.* (1999).

Capítulo 3

Seleccção de Covariáveis em Modelos de Regressão

3.1 Introdução

A selecção de covariáveis em modelos de regressão consiste na identificação das covariáveis que possibilitam uma melhor compreensão da variabilidade inerente ao fenómeno aleatório em estudo, que se representa por Y (variável resposta). Deve, contudo, evitar-se a desnecessária complicação do modelo, originada pela inclusão de covariáveis que pouco acrescentam na explicação da variabilidade de Y . Assim, qualquer procedimento utilizado na selecção de covariáveis deve ter em conta o equilíbrio entre a qualidade do ajustamento e a parcimónia do modelo.

Se o número de possíveis covariáveis for reduzido, a escolha parcimoniosa do melhor subconjunto de covariáveis pode ser efectuada por pesquisa exaustiva, isto é, analisando todas as regressões possíveis. Admitindo que o termo constante está incluído no modelo e não considerando a eventual existência de interacções, o número total de modelos a comparar, para um conjunto de p possíveis covariáveis e numa situação de pesquisa exaustiva, é 2^p . Obviamente, este número é ainda maior quando se considera a possibilidade de existência de interacções, ou se admite que o termo constante pode ou não estar incluído no modelo. Deste modo, a pesquisa exaustiva não é uma solução aceitável, mesmo para um número moderado de covariáveis.

Na abordagem clássica a problemas de regressão existem vários métodos de selecção de covariáveis amplamente utilizados. Quando o número de covariáveis é reduzido, a execução de regressões com todos os possíveis subconjuntos de covariáveis é um método ainda muito utilizado. Quando o número de covariáveis é mais elevado são utilizados outros métodos de selecção de covariáveis: a selecção progressiva, a eliminação regressiva e o método passo a passo que combina os dois primeiros métodos. A selecção progressiva de covariáveis começa com o modelo só com a constante e, através de testes F parciais, as covariáveis mais significativas vão entrando, uma a uma, no modelo. A eliminação regressiva de covariáveis começa com o modelo com todas as covariáveis e, novamente através de testes F parciais, são excluídas do modelo, uma a uma, as covariáveis menos significativas. O método passo a passo começa do mesmo modo que a selecção progressiva mas, sempre

que uma covariável é incluída no modelo, testa se alguma das outras covariáveis que já estava no modelo deve ser excluída. A descrição destes procedimentos pode ser encontrada num dos muitos livros de regressão, dos quais destaca-se Weisberg (1985), Montgomery e Peck (1992) e Ryan (1997). Mais pormenores sobre a abordagem clássica aos métodos de selecção de covariáveis podem ser encontrados em Hocking (1976) e Miller (1984,1990).

Tanto na selecção progressiva de covariáveis como na eliminação regressiva de covariáveis, para um conjunto de p possíveis covariáveis e contando com o modelo só com a constante, o número máximo de modelos examinados é $p + 1$. Quando p é grande, a percentagem de modelos examinados, do conjunto de todos os modelos possíveis, é bastante pequena; esta é uma das críticas apontadas a estes dois métodos (Ryan, 1997).

Como já foi referido, os métodos de selecção progressiva, de eliminação regressiva e passo a passo baseiam-se na comparação sequencial de modelos encaixados, por aplicação de testes de significância baseados na distribuição F . O não conhecimento do nível de significância global e a impossibilidade de comparação de modelos não encaixados são outras críticas usuais a estes métodos de selecção de covariáveis. Estas fragilidades impulsionaram a aplicação de outros critérios, nomeadamente os que se baseiam no R^2 ajustado, no erro médio quadrático ou na estatística C_p de Mallows (Jobson, 1991, secção 4.2.2). Outros critérios também bastante utilizados são os que se baseiam na minimização de “ $-2 \times \log$ -verosimilhança + penalização”, onde a penalização pode depender do número de covariáveis presentes no modelo, como no caso do AIC, ou do número de covariáveis presentes no modelo e da dimensão da amostra, como no caso do BIC.

Na sua génese, a selecção bayesiana de modelos e, em particular, a selecção bayesiana de covariáveis, baseia-se no cálculo das probabilidades *a posteriori* dos modelos envolvidos. Considera-se, de um modo geral, que o melhor modelo é o que apresenta maior probabilidade *a posteriori*. Assim, caso exista um modelo com probabilidade *a posteriori* próxima de 1 deve optar-se, inequivocamente, por esse modelo; a escolha do modelo torna-se mais complicada se existirem dois ou mais modelos com probabilidades *a posteriori* relevantes.

O desenvolvimento da teoria dos modelos lineares generalizados (Nelder e Wedderburn, 1972) veio possibilitar a unificação dos métodos estatísticos subjacentes a diversos tipos de modelos, tais como modelos de regressão linear, análise de variância, regressão logística, entre muitos outros. Um modelo linear generalizado é definido à custa de três componentes: a componente aleatória, através da especificação da distribuição da variável resposta (que pertence à família exponencial); a componente sistemática, através da determinação das covariáveis a serem incluídas no preditor linear; a função de ligação (função monótona e diferenciável), que define a relação existente entre o preditor linear e o valor médio da variável resposta (McCullagh e Nelder, 1989, secção 2.2). A selecção de modelos pode referir-se à selecção de qualquer uma das componentes do modelo linear generalizado. No entanto, neste capítulo, considera-se que a incerteza na especificação do modelo é apenas relativa à escolha das covariáveis a serem incluídas no preditor linear.

Recentemente tem havido bastante investigação sobre a abordagem bayesiana à selecção de covariáveis em modelos lineares generalizados. Dellaportas *et al.* (2000) efectua uma revisão crítica de alguns dos métodos bayesianos de selecção do melhor subconjunto de covariáveis via amostrador Gibbs, entre eles o método de Carlin e Chib, a selecção de covariáveis via pesquisa estocástica, o método de Kuo e Mallick e o método

Gibbs de selecção de covariáveis. Com excepção do método de Carlin e Chib, que só é efectivamente razoável utilizar quando o número de modelos em comparação é pequeno, os outros métodos baseiam-se no cálculo das probabilidades *a posteriori* de todos os possíveis modelos. Estes métodos MCMC de selecção de covariáveis possibilitam a geração de amostras do espaço dos modelos e dos parâmetros, permitindo estimar não só a distribuição *a posteriori* dos parâmetros, mas também as probabilidades *a posteriori* dos modelos. Os métodos referidos fazem uso das denominadas distribuições *pseudoprior*, para acautelar o problema de os modelos comparados terem vectores de parâmetros com dimensões distintas. Green (1995) propõe que se utilize o método dos saltos reversíveis (de *reversible jump*) de modo a contornar o referido problema. Este método é, basicamente, um algoritmo MCMC que permite a geração de uma cadeia de Markov que salta entre modelos com vectores de parâmetros de dimensões distintas, mantendo as condições necessárias para a convergência do algoritmo.

Barbieri e Berger (2004) definem “modelo de probabilidade *a posteriori* mediana” como sendo o modelo que contém todas as covariáveis com probabilidade *a posteriori* de inclusão no modelo superior ou igual a 0.5. Estes autores acrescentam que, no caso da regressão linear normal, se o modelo com probabilidade *a posteriori* máxima tiver uma probabilidade bastante superior à dos outros modelos então, de um modo geral, este modelo coincide com o modelo de probabilidade *a posteriori* mediana. No caso da selecção de covariáveis ser efectuada com recurso aos métodos MCMC, é mais fácil determinar o modelo de probabilidade *a posteriori* mediana do que o de probabilidade *a posteriori* máxima, pois basta registar a proporção de vezes que cada covariável está presente nos modelos “visitados”. Obviamente, averiguar se p proporções são superiores ou inferiores a 0.5 é bastante mais fácil que estimar a proporção de vezes que 2^p modelos são “visitados”. Assim, com um número elevado de covariáveis, pode optar-se por determinar o modelo de probabilidade *a posteriori* mediana, uma vez que a determinação do modelo com probabilidade *a posteriori* máxima pode ser computacionalmente difícil ou até mesmo impossível.

Na abordagem clássica, o processo de selecção de covariáveis em modelos de regressão implica a escolha de métodos de pesquisa e de medidas para comparação dos modelos associados a cada um dos subconjuntos de covariáveis. Neste trabalho, sugere-se a utilização de dois métodos de pesquisa do melhor subconjunto de covariáveis, inspirados nos métodos de pesquisa da abordagem clássica, em conjunção com a utilização de algumas das medidas bayesianas de selecção de modelos apresentadas no capítulo 2. Nomeadamente, propõe-se a utilização da selecção progressiva e da eliminação regressiva de covariáveis, elegendo-se para comparação dos subconjuntos de covariáveis as medidas de selecção de modelos — também designadas por medidas de discrepância — DIC, \widehat{BIC} , LLP e SlnCPO. Entre subconjuntos de covariáveis em comparação, escolhe-se aquele cujo valor da medida de discrepância o indica como mais favorável. No caso da medida de discrepância usada ser a soma dos logaritmos das ordenadas preditivas condicionais, escolhe-se o subconjunto de covariáveis que maximiza SlnCPO. No caso das outras medidas de discrepância, escolhe-se o subconjunto de covariáveis que minimiza o valor de cada uma dessas medidas. Estes métodos podem ser utilizados na selecção do melhor subconjunto de covariáveis em qualquer modelo linear generalizado.

Os dois métodos de selecção de covariáveis via medidas de discrepância que se propõe são comparados com o método de selecção informal por pesquisa manual via medidas

de discrepância (Paulino *et al.*, 2003), e com o método Gibbs de selecção de covariáveis (Dellaportas *et al.*, 2000). Recorre-se a dois conjuntos de dados da literatura para ilustrar estes métodos de selecção de covariáveis no caso de modelos de regressão linear normal e de regressão gama.

3.2 Métodos de selecção via medidas de discrepância

Propõe-se dois métodos, inspirados em métodos da Estatística Clássica, para a selecção bayesiana de covariáveis em modelos de regressão: a selecção progressiva de covariáveis e a eliminação regressiva de covariáveis via medidas de discrepância. Faz-se também referência ao método de selecção informal por pesquisa manual via medidas de discrepância.

3.2.1 Selecção informal por pesquisa manual

O método de selecção informal por pesquisa manual via medidas de discrepância (SIMD) começa por ajustar o modelo de regressão com todas as covariáveis em estudo, sendo calculado o valor de uma ou mais medidas de discrepância. Através da análise das distribuições marginais *a posteriori* dos coeficientes de regressão, elimina-se do modelo, uma a uma, as covariáveis cujas distribuições marginais *a posteriori* dos coeficientes de regressão estão mais concentradas em torno de zero. Um a um, os modelos sem as respectivas covariáveis são novamente ajustados e as medidas de discrepância calculadas. O melhor modelo é aquele que, entre os modelos estimados, minimiza a medida de discrepância seleccionada (Paulino *et al.*, 2003, secção 9.4.4).

Este método tem como vantagem o facto do investigador poder decidir, passo a passo, qual a próxima covariável a sair do modelo ou se o processo de exclusão de covariáveis deve parar. Esta vantagem pode também constituir uma desvantagem pois impossibilita a implementação computacional deste método, uma vez que é requerida a intervenção do investigador ao longo de todo o processo. Refira-se também o cunho subjectivo do método SIMD, pelo facto do investigador ter que avaliar se as distribuições marginais *a posteriori* dos coeficientes de regressão estão ou não suficientemente concentradas em torno de zero.

3.2.2 Selecção progressiva de covariáveis

O método de selecção progressiva de covariáveis (FVS) via medidas de discrepância começa com o modelo só com a constante e as covariáveis mais significativas, de acordo com a medida de discrepância escolhida, vão entrando uma a uma no modelo, até já não existirem mais covariáveis para entrar no modelo, ou até que a inclusão de qualquer outra covariável produza valores menos favoráveis da medida de discrepância. No caso da medida considerada ser o DIC, o \widehat{BIC} ou o LLP, a covariável mais significativa é a que, ao ser incluída no modelo, produz a maior redução no valor da medida de discrepância, enquanto no caso do SlnCPO, a covariável mais significativa é a que, ao ser incluída no modelo, produz o maior aumento no valor do SlnCPO.

O método FVS para selecção do melhor subconjunto de covariáveis via DIC, num conjunto de p possíveis covariáveis, está descrito no algoritmo 3.1.

Algoritmo 3.1

Iteração 1. Ajusta-se um modelo apenas com a constante e calcula-se o DIC.

Iteração 2. Ajusta-se p modelos de regressão, cada um deles com uma das covariáveis x_k , com $k = 1, \dots, p$, e determina-se os valores do DIC de cada um destes modelos. Se o mínimo do DIC for obtido com a covariável x_{k_1} e se este valor for inferior ao valor do DIC da iteração anterior, escolhe-se x_{k_1} para incluir no modelo; no caso contrário, o processo de selecção de covariáveis pára e o modelo seleccionado é constituído apenas pela constante.

Iteração 3. Ajusta-se $p-1$ modelos de regressão, cada um dos quais com duas covariáveis, x_{k_1} e x_k , com $k = 1, \dots, p$ e $k \neq k_1$, e determina-se o DIC de cada um destes modelos. Se o mínimo do DIC for obtido com x_{k_1} e x_{k_2} , e se este valor for inferior ao valor do DIC da iteração anterior, então x_{k_2} é a segunda covariável a ser incluída no modelo; no caso contrário, o processo de selecção de covariáveis pára e o modelo seleccionado é constituído apenas por x_{k_1} .

Procede-se deste modo até à t -ésima iteração, com $t \leq p + 1$:

Iteração t . Ajusta-se $p - t + 2$ modelos de regressão, cada um dos quais com $t - 1$ covariáveis, $x_{k_1}, x_{k_2}, \dots, x_{k_{t-2}}, x_k$, com $k = 1, \dots, p$ e $k \notin \{k_1, k_2, \dots, k_{t-2}\}$, e determina-se o DIC de cada um destes modelos. Se o mínimo do DIC for obtido com as covariáveis $x_{k_1}, x_{k_2}, \dots, x_{k_{t-2}}, x_{k_{t-1}}$, e se este valor for inferior ao valor do DIC da iteração anterior, então $x_{k_{t-1}}$ é a $(t - 1)$ -ésima covariável a ser incluída no modelo; no caso contrário, o processo de selecção de covariáveis pára e o modelo seleccionado é constituído pelas covariáveis $x_{k_1}, x_{k_2}, \dots, x_{k_{t-2}}$. ■

O método FVS pode ser aplicado com outras medidas de discrepância, como por exemplo, o $\widehat{\text{BIC}}$, o LLP ou o SlnCPO. No caso de se utilizar a soma dos logaritmos das ordenadas preditivas condicionais é necessário modificar o algoritmo 3.1 pois, em cada iteração, o melhor modelo é identificado pelo máximo de SlnCPO, e a inclusão de uma covariável no modelo é apenas efectuada se originar um acréscimo no valor de SlnCPO, relativamente ao valor da iteração anterior; pode também optar-se por não modificar o algoritmo e considerar o simétrico da soma dos logaritmos das ordenadas preditivas condicionais, isto é, $-\text{SlnCPO}$.

Se existirem covariáveis conjuntamente, mas não individualmente, significativas, o método FVS não é o mais indicado, conforme se irá ilustrar no exemplo 3.1.

3.2.3 Eliminação regressiva de covariáveis

O método de eliminação regressiva de covariáveis (BVE) via medidas de discrepância começa com o modelo completo, isto é, o modelo com todas as covariáveis. As covariáveis menos significativas, de acordo com a medida de discrepância considerada, são excluídas uma a uma do modelo, até já não existirem mais covariáveis para sair do modelo, ou até que a exclusão de qualquer outra covariável origine valores menos favoráveis da medida de discrepância. No caso da medida considerada ser o DIC, o $\widehat{\text{BIC}}$ ou o LLP, a covariável

menos significativa é a que, ao ser excluída do modelo, produz o maior decréscimo no valor da medida de discrepância, enquanto no caso do SlnCPO, a covariável menos significativa é a que, ao ser excluída do modelo, produz o maior aumento no valor do SlnCPO.

O método BVE para selecção do melhor subconjunto de covariáveis via DIC, num conjunto de p possíveis covariáveis, está descrito no algoritmo 3.2.

Algoritmo 3.2

Iteração 1. Ajusta-se um modelo de regressão com todas as covariáveis e calcula-se o DIC.

Iteração 2. Ajusta-se p modelos de regressão, cada um deles correspondente à exclusão de uma das covariáveis x_k , com $k = 1, \dots, p$, e calcula-se os valores do DIC destes modelos. Se o mínimo do DIC for obtido com a exclusão de x_{k_1} e se este valor for menor ou igual do que o valor do DIC da iteração anterior, escolhe-se x_{k_1} para excluir do modelo; no caso contrário, o processo de exclusão de covariáveis pára e o modelo seleccionado é o modelo com todas as covariáveis.

Iteração 3. Ajusta-se $p-1$ modelos de regressão, cada um deles resultante da exclusão de duas covariáveis, x_{k_1} e x_k , com $k = 1, \dots, p$ e $k \neq k_1$, e determina-se o valor do DIC de cada um destes modelos. Se o mínimo do DIC for obtido com a exclusão de x_{k_1} e x_{k_2} , e se este valor for menor ou igual do que o valor do DIC da iteração anterior, então x_{k_2} é a segunda covariável a ser excluída do modelo; no caso contrário, o processo de exclusão de covariáveis pára e o modelo seleccionado é o que exclui x_{k_1} .

Procede-se deste modo até à t -ésima iteração, com $t \leq p+1$:

Iteração t . Ajusta-se $p-t+2$ modelos de regressão, cada um deles resultante da exclusão de $t-1$ covariáveis, $x_{k_1}, x_{k_2}, \dots, x_{k_{t-2}}, x_k$, com $k = 1, \dots, p$ e $k \notin \{k_1, k_2, \dots, k_{t-2}\}$, e determina-se o valor da DIC de cada um destes modelos. Se o mínimo do DIC for obtido com a exclusão de $x_{k_1}, x_{k_2}, \dots, x_{k_{t-2}}, x_{k_{t-1}}$, e se este valor for menor ou igual do que o valor do DIC da iteração anterior, então $x_{k_{t-1}}$ é a $(t-1)$ -ésima covariável a ser excluída do modelo; no caso contrário, o processo de eliminação de covariáveis pára e o modelo seleccionado é o que exclui $x_{k_1}, x_{k_2}, \dots, x_{k_{t-2}}$. ■

O método BVE pode também ser implementado com as medidas de discrepância $\widehat{\text{BIC}}$, LLP ou SlnCPO, tal como acontece com o método FVS.

Montgomery e Peck (1992, pág. 297 e 293) salientam que, “*The backward elimination algorithm is often less adversely affected by the correlative structure of the regressors than is forward selection*”, e acrescentam que a eliminação regressiva é, frequentemente, um bom procedimento para a selecção de covariáveis, “*It is particularly favored by analysts who like to see the effect of including all the candidate regressors, just that nothing “obvious” will be missed*”.

O mesmo acontece com os métodos FVS e BVE via medidas de discrepância propostos para a selecção bayesiana de covariáveis em modelos de regressão. De facto, quando existe multicolinearidade entre as covariáveis ou quando há covariáveis conjuntamente

significativas, mas não individualmente significativas, o método BVE deve ser preferido ao método FVS, uma vez que, nestes casos, a garantia de sucesso do método progressivo na pesquisa do melhor modelo é menor que a do método regressivo, conforme se ilustra no exemplo 3.1.

Considere-se, por exemplo, o caso de duas covariáveis x_{k_1} e x_{k_2} ($k_1, k_2 \in \{1, \dots, p\}$) conjuntamente, mas não individualmente, significativas. É natural que o método FVS via medidas de discrepância possa parar antes de seleccionar qualquer uma destas covariáveis, uma vez que, em cada iteração, testa qual das covariáveis não incluídas no modelo deve ser seleccionada, e não é de esperar que x_{k_1} ou x_{k_2} sejam escolhidas pois não são individualmente significativas. Ao contrário, como o método BVE via medidas de discrepância começa com todas as covariáveis incluídas no modelo, é plausível que nem x_{k_1} nem x_{k_2} sejam excluídas pois, como são conjuntamente significativas, quando se testa a exclusão de cada uma delas do modelo é natural obter valores mais desfavoráveis para a medida de discrepância.

Exemplo 3.1

Sejam Y uma variável resposta e x_1, x_2 e x_3 o conjunto de possíveis covariáveis. As observações destas variáveis em 20 indivíduos são apresentadas na tabela 3.1. Pretende seleccionar-se o subconjunto de covariáveis que permite obter o melhor modelo de regressão linear normal para a variável resposta Y .

Tabela 3.1: Conjunto de dados, adaptado de Ryan (1997, tabela 5.1), usado para comparar a performance dos métodos FVS e BVE.

y	x_1	x_2	x_3	y	x_1	x_2	x_3
11.3141	1.35	6.2112	1.71	18.3680	1.99	9.5602	8.47
12.0623	2.34	9.8328	3.31	8.0786	2.25	8.8968	1.13
10.2527	2.85	11.3507	0.58	15.5961	3.04	12.8535	5.93
10.3557	1.55	6.7705	1.62	20.0016	2.59	11.9332	6.46
7.5074	2.50	9.6899	-1.70	14.5630	3.56	14.5349	2.97
10.6675	1.48	6.5703	-0.45	13.4005	3.87	15.4560	4.46
7.1506	1.79	7.1276	-1.33	14.3493	2.05	9.1575	4.16
4.5246	1.84	6.9013	-7.02	5.2692	1.68	6.4475	-0.98
3.4678	3.50	12.6103	-6.81	16.6721	1.45	7.3855	6.40
-1.5040	1.67	5.3706	-12.81	4.4787	2.12	7.8802	-5.62

Na tabela 3.2 apresenta-se os valores das medidas de discrepância DIC, \widehat{BIC} , LLP e SlnCPO, para os modelos de regressão linear normal, resultantes de todas as combinações possíveis de covariáveis. Através da análise desta tabela facilmente se identifica as sucessivas iterações dos métodos FVS e BVE.

A aplicação do método FVS começa com o ajustamento do modelo só com a constante e na segunda iteração x_3 é incluída no modelo, pois é a covariável que origina o maior decréscimo no valor do DIC, do \widehat{BIC} e do LLP e o maior aumento no valor de SlnCPO. No prosseguimento da aplicação deste método verifica-se que a inclusão de qualquer uma das outras covariáveis não origina valores mais favoráveis das medidas de discrepância e, deste modo, o processo de selecção de covariáveis pára e o modelo escolhido é constituído

Tabela 3.2: Valores das medidas de discrepância, obtidos com os dados da tabela 3.1, para todos os modelos de regressão linear possíveis.

covariáveis	DIC	$\widehat{\text{BIC}}$	LLP	SlnCPO
nenhuma	126.6	130.6	124.6	-61.6
x_1	129.8	135.6	126.6	-63.4
x_2	125.9	131.7	122.7	-61.7
x_3	81.8	87.6	78.6	-39.6
x_1, x_2	-94.7	-87.1	-99.0	49.1
x_1, x_3	83.7	91.5	79.5	-40.5
x_2, x_3	83.1	90.8	78.8	-40.3
x_1, x_2, x_3	-92.6	-82.9	-97.9	48.2

apenas pela covariável x_3 .

Por sua vez, a aplicação do método BVE a este conjunto de dados começa com o ajustamento do modelo com todas as covariáveis. Na 2ª iteração x_3 sai do modelo, pois é a covariável cuja exclusão origina o maior decréscimo no valor do DIC, do $\widehat{\text{BIC}}$ e do LLP e o maior aumento no valor de SlnCPO. Na iteração seguinte verifica-se que a exclusão de qualquer uma das outras covariáveis não produz nenhum decréscimo no valor do DIC, do $\widehat{\text{BIC}}$ e do LLP, nem nenhum aumento no valor de SlnCPO e, conseqüentemente, o método BVE pára e o modelo escolhido é constituído por x_1 e x_2 .

Pela análise da tabela 3.2 constata-se que o melhor modelo é o modelo constituído pelas covariáveis x_1 e x_2 , pois é o que apresenta valores mais pequenos do DIC, do $\widehat{\text{BIC}}$ e do LLP e valor mais elevado do SlnCPO; este seria o modelo seleccionado numa situação de pesquisa exaustiva. Dos dois métodos de pesquisa apresentados, apenas o BVE escolhe este modelo, enquanto o método FVS opta pelo modelo constituído pela covariável x_3 , que apresenta valores das medidas de discrepância bastante mais desfavoráveis.

A incapacidade do método FVS na identificação do melhor modelo pode ser devida à conjunção de dois factores: a correlação linear entre y e x_3 ($r_{y,x_3} = 0.955$) ser bastante maior que a correlação linear entre y e x_1 ($r_{y,x_1} = 0.178$) ou entre y e x_2 ($r_{y,x_2} = 0.450$), e a contribuição conjunta de x_1 e x_2 na explicação da variabilidade de y ser superior à contribuição de x_3 . ■

Para um conjunto de p possíveis covariáveis, o número máximo de modelos a ajustar e comparar com os métodos de selecção progressiva e de eliminação regressiva via medidas de discrepância é $1 + p(p + 1)/2$, número bastante inferior ao da pesquisa exaustiva, mesmo para um número moderado de covariáveis. No caso dos exemplos das secções 3.5 e 3.6, utilizados para ilustrar estes métodos, em que existem 13 possíveis covariáveis, na pesquisa exaustiva teriam que ser ajustados e comparados 8192 modelos, enquanto com os métodos FVS ou BVE são ajustados e comparados, no máximo, 92 modelos. Obviamente, quanto maior for o valor de p , menor é a percentagem dos possíveis subconjuntos de covariáveis examinados quando se utiliza um destes métodos, sendo também menor a garantia que o subconjunto de covariáveis escolhido satisfaça um determinado critério de optimalidade. Refira-se ainda que o número de modelos ajustados com os métodos de selecção progressiva

e de eliminação regressiva via medidas de discrepância é bastante superior ao número de modelos ajustados pelos correspondentes métodos da abordagem clássica, nos quais o teste F parcial permite identificar a covariável que vai entrar ou sair do modelo, em cada passo, e só então é que o novo modelo é ajustado.

3.3 Métodos de selecção simultânea via amostrador Gibbs

Seja m um modelo, \mathcal{M} um conjunto finito de todos os modelos em consideração e $\Pr\{m\}$ a probabilidade *a priori* do modelo m . A contribuição dos dados para as probabilidades *a posteriori* dos modelos é expressa através da distribuição marginal, $f(y|m)$, definida por

$$f(y|m) = \int f(y|\theta_m, m) h(\theta_m|m) d\theta_m,$$

onde θ_m é o vector de parâmetros do modelo m , $h(\theta_m|m)$ é a distribuição *a priori* de θ_m para o modelo m e $f(y|\theta_m, m)$ é a distribuição de Y , condicional ao vector de parâmetros θ_m , para o modelo m . A probabilidade *a posteriori* de cada modelo m é dada por

$$\Pr\{m|y\} = \frac{\Pr\{m\} f(y|m)}{\sum_{m' \in \mathcal{M}} \Pr\{m'\} f(y|m')}, \quad m \in \mathcal{M}. \quad (3.1)$$

Em particular, a probabilidade *a posteriori* relativa de dois modelos em competição, m_1 e m_2 , reduz-se a

$$\frac{\Pr\{m_1|y\}}{\Pr\{m_2|y\}} = \frac{f(y|m_1)}{f(y|m_2)} \times \frac{\Pr\{m_1\}}{\Pr\{m_2\}},$$

que relaciona, através do factor Bayes, a razão das vantagens *a posteriori* com a razão das vantagens *a priori* dos dois modelos. Quando não é possível determinar analiticamente as densidades $f(y|m_1)$ e $f(y|m_2)$, a determinação do factor Bayes entre dois modelos em competição passa pela estimação da densidade marginal de y para cada modelo e, posteriormente, pelo cálculo do factor Bayes através do quociente $\hat{f}(y|m_1)/\hat{f}(y|m_2)$.

Uma abordagem alternativa consiste em incluir um identificador γ do modelo m como mais um vector de parâmetros do algoritmo de amostragem. Embora este procedimento complique a fase inicial do processo de amostragem, tem como vantagem o facto das iterações $\{\gamma^{(i)}\}_{i=1}^N$, obtidas após a convergência, constituírem uma amostra da distribuição *a posteriori* do identificador do modelo m , e permitirem estimar a $\Pr\{m|y\}$ através do quociente

$$\widehat{\Pr}\{m|y\} = \frac{\#\{i : \gamma^{(i)} = \gamma\}}{N}, \quad m \in \mathcal{M}.$$

Deste modo é possível obter estimativas das probabilidades *a posteriori* de cada modelo $m \in \mathcal{M}$, isto é, estimar (3.1). Estas estimativas podem ser usadas para estimar o factor Bayes entre quaisquer dois modelos em competição, m_1 e m_2 , através de

$$\widehat{\text{BF}}_{12}(y) = \frac{\widehat{\Pr}\{m_1|y\} / \widehat{\Pr}\{m_2|y\}}{\Pr\{m_1\} / \Pr\{m_2\}}.$$

Esta estratégia de inclusão do identificador do modelo como mais um parâmetro do algoritmo de amostragem deu origem a vários métodos de selecção simultânea de modelos

— no espaço dos modelos e no espaço dos parâmetros — e, em particular, de selecção de covariáveis, via amostrador Gibbs. Entre estes métodos de selecção de covariáveis, nomeia-se o método de Carlin e Chib (Carlin e Chib, 1995), o método de selecção de covariáveis por pesquisa estocástica (George e McCulloch, 1993), o método de Kuo e Mallick (Kuo e Mallick, 1998) e o método Gibbs de selecção de covariáveis (Dellaportas *et al.*, 2000). O facto destes métodos serem baseados na utilização do amostrador Gibbs tem a vantagem de poderem ser implementados de um modo bastante simples no WinBUGS (Spiegelhalter *et al.*, 1996a).

3.3.1 Método de Carlin e Chib

A selecção simultânea de modelos, no espaço dos modelos e no espaço dos parâmetros, exige que os modelos em competição tenham o mesmo número de parâmetros, o que raramente acontece, nomeadamente em problemas de selecção de covariáveis. Carlin e Chib (1995) propõem um método de selecção de modelos, baseado na utilização do amostrador Gibbs, que possibilita a comparação simultânea de modelos que não têm o mesmo número de parâmetros.

Considere-se que existem k modelos em competição e, para cada um deles, um vector de parâmetros θ_{m_j} de dimensão n_j , $j = 1, \dots, k$. Seja θ o vector $(\theta_{m_1}, \dots, \theta_{m_k})$ constituído pelos vectores de parâmetros do conjunto de modelos em competição, que se representa por $\mathcal{M} = \{m_1, \dots, m_k\}$. Designa-se genericamente por m um elemento do conjunto \mathcal{M} . De um modo geral, nos problemas de selecção de modelos, pretende determinar-se as probabilidades *a posteriori* $\Pr\{m|y\}$ de cada modelo m e, eventualmente, a distribuição *a posteriori* do vector de parâmetros θ_m para o modelo m , $h(\theta_m|y, m)$.

Supondo que $\theta_m \in \mathbb{R}^{n_m}$, a solução proposta por Carlin e Chib (1995) consiste em amostrar sobre o identificador do modelo m e o espaço produto $\mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_k}$, o que requer que a distribuição *a priori* do vector de parâmetros θ_m de cada modelo m seja vista como fazendo parte de um modelo geral.

Uma vez que o objectivo primordial desta abordagem é o cálculo de factores Bayes, assume-se que as distribuições *a priori* $h(\theta_m|m)$, $m \in \mathcal{M}$, são próprias, embora possivelmente difusas. Assume-se também que m apenas fornece indicação sobre que vector de parâmetros é relevante para os dados; assim, Y é independente de $\theta_{m'}$, dado m , para $m' \neq m$, isto é,

$$f(y|\theta, m) = f(y|\theta_m, m). \quad (3.2)$$

Para simplificar, assume-se também a existência de independência entre os vários vectores de parâmetros θ_{m_j} , dado o modelo m , isto é,

$$h(\theta|m) = \prod_{j=1}^k h(\theta_{m_j}|m). \quad (3.3)$$

Completa-se a especificação do modelo através da escolha de distribuições $h(\theta_{m'}|m \neq m')$ próprias, que são designadas distribuições *pseudoprior*.

As suposições de independência condicional (3.2) e (3.3) implicam que a distribuição

marginal de Y para o modelo m é dada por

$$\begin{aligned} f(y|m) &= \int f(y|\theta, m) h(\theta|m) d\theta \\ &= \int f(y|\theta_m, m) \prod_{j=1}^k h(\theta_{m_j}|m) d\theta \\ &= \int f(y|\theta_m, m) h(\theta_m|m) d\theta_m, \end{aligned}$$

e assim a forma de $h(\theta_{m'}|m \neq m')$ é irrelevante para a distribuição marginal $f(y|m)$. Deste modo, tal como o nome sugere, uma distribuição *pseudoprior* não é, na realidade, uma distribuição *a priori*, mas apenas uma densidade de ligação convenientemente escolhida, necessária para definir a distribuição conjunta do vector (m, θ) .

Atendendo a (3.2) e (3.3) a distribuição conjunta de Y , θ e m é

$$\begin{aligned} p(y, \theta, m) &= f(y|\theta, m) h(\theta|m) \Pr\{m\} \\ &= f(y|\theta_m, m) \left[\prod_{j=1}^k h(\theta_{m_j}|m) \right] \Pr\{m\}, \end{aligned}$$

onde $\Pr\{m\}$ é a probabilidade *a priori* do modelo m , verificando $\sum_{m \in \mathcal{M}} \Pr\{m\} = 1$. Assim, com o pressuposto de independência *a priori* dos parâmetros, Carlin e Chib (1995) sugerem que a especificação da distribuição *a priori* conjunta de m e θ seja feita através do produto da probabilidade *a priori* $\Pr\{m\}$, pela densidade *a priori* $h(\theta_m|m)$ e as distribuições *pseudoprior* ou densidades de ligação $h(\theta_{m'}|m \neq m')$.

Para a implementação do método de amostragem Gibbs são necessárias as distribuições condicionais completas. A distribuição *a posteriori* do vector de parâmetros $\theta_{m'}$, para $m' \in \mathcal{M}$, é

$$h(\theta_{m'}|y, \theta_{(-m')}, m) \propto \begin{cases} f(y|\theta_m, m) h(\theta_m|m) & , \text{ se } m' = m \\ h(\theta_{m'}|m) & , \text{ se } m' \neq m \end{cases},$$

onde $\theta_{(-m')}$ representa o conjunto de vectores de parâmetros θ com excepção do vector $\theta_{m'}$. Assim, quando $m' = m$, gera-se da distribuição condicional completa usual do modelo m , e quando $m' \neq m$, gera-se da correspondente distribuição *pseudoprior* $h(\theta_{m'}|m)$. Por sua vez, a distribuição condicional completa de m é dada por

$$h(m|\theta, y) = \frac{f(y|\theta_m, m) \left[\prod_{j=1}^k h(\theta_{m_j}|m) \right] \Pr\{m\}}{\sum_{m' \in \mathcal{M}} f(y|\theta_{m'}, m') \left[\prod_{j=1}^k h(\theta_{m'_j}|m') \right] \Pr\{m'\}}$$

Para assegurar a rápida convergência deste método, a especificação das distribuições *pseudoprior* tem que ser feita com muito cuidado, o que pode ser uma tarefa morosa se o número de modelos em comparação for elevado, tal como acontece em problemas de selecção de covariáveis mesmo que o número de potenciais covariáveis seja moderado.

Na secção 3.5 são efectuadas aplicações do método de Carlin e Chib. Outros pormenores sobre a implementação deste método podem ser vistos em Carlin e Chib (1995) e Spiegelhalter *et al.* (1996c).

3.3.2 Selecção de covariáveis por pesquisa estocástica

O método de selecção de covariáveis por pesquisa estocástica, que se representa por SSVS (de *Stochastic Search Variable Selection*), foi introduzido por George e McCulloch (1993) para modelos de regressão linear tendo posteriormente sido adaptado para modelos mais complexos (referências em Dellaportas *et al.*, 2000).

Considere-se um problema de selecção de covariáveis em modelos de regressão, em que existem p covariáveis, x_1, x_2, \dots, x_p , candidatas a ser incluídas no preditor linear. O preditor linear de um modelo linear generalizado pode ser representado por

$$\eta_i = \sum_{j=1}^p b_j x_{ij}, \quad i = 1, \dots, n.$$

Cada modelo $m \in \mathcal{M}$ pode ser representado por um vector γ de dimensão p , $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_p)$, onde γ_j é um indicador da presença da covariável x_j no modelo ($\gamma_j = 1$ indica que x_j está presente no modelo e $\gamma_j = 0$ indica que x_j está ausente do modelo).

O método SSVS considera que o vector de parâmetros b tem dimensão p para todos os modelos, independentemente do número de covariáveis presentes no modelo. Isto pode ser conseguido considerando um modelo bayesiano hierárquico em que a distribuição de b_j , condicional a γ_j , é definida por uma mistura de normais. Assim, o indicador γ_j é incluído no modelo através da distribuição *a priori* de b_j condicional a γ_j ,

$$b_j | \gamma_j \sim \gamma_j N(0, u_j^2 v_j^2) + (1 - \gamma_j) N(0, v_j^2)$$

com $\gamma_j \sim \text{Bern}(p_j)$ e para valores de u_j e v_j convenientemente escolhidos.

Uma distribuição *a priori* não informativa para o vector de parâmetros γ consiste em considerar que γ_j assume o valor 1 ou 0 com probabilidade 1/2, independentemente do valor de $\gamma_{j'}$, para $j' \neq j$. Esta distribuição *a priori* para γ pode ser considerada não informativa na medida em que atribui igual peso a todos os possíveis modelos.

Se $\gamma_j = 0$, a distribuição *a priori* de b_j deve estar concentrada em torno de zero, de modo a assegurar que x_j está “ausente” do preditor linear. Por sua vez, quando $\gamma_j = 1$, assume-se que há pouca informação *a priori* sobre o parâmetro b_j , sendo esta situação representada por uma distribuição *a priori* difusa para b_j . A escolha de u_j e v_j deve ser feita com bastante cuidado pois as probabilidades *a posteriori* dos modelos são bastante mais influenciadas pelos valores destes parâmetros do que pela distribuição *a priori* de γ_j .

Caso se escolha em primeiro lugar $u_j^2 v_j^2$ de modo a representar uma distribuição *a priori* difusa para b_j quando $\gamma_j = 1$, então u_j^2 deve ser escolhido considerando o valor de $|b_j|$ para o qual as duas densidades componentes da distribuição *a priori* assumem valores iguais. Este valor pode ser considerado o menor valor de $|b_j|$, que se designa por δ_j , a partir do qual o termo é considerado significativo do ponto de vista prático. Devido à influência que tem nas distribuições *a posteriori*, a escolha de δ_j deve ser feita com bastante cuidado (*e.g.*, Paulino *et al.*, 2003).

As distribuições condicionais completas *a posteriori* de b_j e γ_j são dadas, respectivamente, por:

$$h(b_j | y, b_{(-j)}, \gamma) \propto f(y | b, \gamma) h(b_j | \gamma_j)$$

e

$$\frac{h(\gamma_j = 1 | y, b, \gamma_{(-j)})}{h(\gamma_j = 0 | y, b, \gamma_{(-j)})} = \frac{h(b | \gamma_j = 1, \gamma_{(-j)})}{h(b | \gamma_j = 0, \gamma_{(-j)})} \frac{h(\gamma_j = 1, \gamma_{(-j)})}{h(\gamma_j = 0, \gamma_{(-j)})} \quad (3.4)$$

onde $\gamma_{(-j)}$ representa todos as componentes do vector γ excepto γ_j . Aplicações deste método a exemplos reais podem ser encontradas em George e McCulloch (1993, 1996).

3.3.3 Método de Kuo e Mallick

O método de Kuo e Mallick (Kuo e Mallick, 1998), também conhecido por método das distribuições *a priori* não condicionais, contrariamente ao método proposto na secção 3.3.2, inclui o parâmetro γ_j no preditor linear,

$$\eta_i = \sum_{j=1}^p \gamma_j b_j x_{ij}, \quad i = 1, \dots, n, \quad (3.5)$$

onde b_j é a j -ésima componente do vector de parâmetros b de dimensão p e x_{ij} é a observação da j -ésima covariável para o i -ésimo indivíduo. Este método considera que a distribuição *a priori* de b é independente de γ e, portanto, do modelo m ; assim, verifica-se que:

$$h(b_j | b_{(-j)}, \gamma) = h(b_j | b_{(-j)}).$$

As distribuições condicionais completas *a posteriori* de b_j e γ_j são dadas, respectivamente, por:

$$h(b_j | y, \gamma, b_{(-j)}) \propto \begin{cases} f(y | b, \gamma) h(b_j | b_{(-j)}) & , \text{ se } \gamma_j = 1 \\ h(b_j | b_{(-j)}) & , \text{ se } \gamma_j = 0 \end{cases}$$

e

$$\frac{h(\gamma_j = 1 | y, \gamma_{(-j)}, b)}{h(\gamma_j = 0 | y, \gamma_{(-j)}, b)} = \frac{f(y | \gamma_j = 1, \gamma_{(-j)}, b)}{f(y | \gamma_j = 0, \gamma_{(-j)}, b)} \frac{h(\gamma_j = 1, \gamma_{(-j)})}{h(\gamma_j = 0, \gamma_{(-j)})}. \quad (3.6)$$

Este método é muito fácil de utilizar, pois é apenas necessário especificar a distribuição *a priori* usual de b para o modelo completo e as distribuições *a priori* condicionais $h(b_j | b_{(-j)})$ substituem as distribuições *pseudoprior* do método de Carlin e Chib. A principal desvantagem deste método é a falta de flexibilidade que impossibilita a execução de alterações para melhorar a sua eficiência. Aplicações deste método a exemplos reais podem ser encontrados em Kuo e Mallick (1998).

3.3.4 Método Gibbs de selecção de covariáveis

Dellaportas *et al.* (1997) defendem a utilização de um método híbrido, que consiste em especificar o preditor linear tal como foi feito pelo método das distribuições *a priori* não condicionais de Kuo e Mallick e em considerar, tal como o método SSVS, que a distribuição *a priori* de b não é independente de γ .

Considere-se p covariáveis candidatas a serem incluídas no modelo, isto é, a fazerem parte do preditor linear. Cada modelo $m \in \mathcal{M}$ é identificado por um vector $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_p)$, de dimensão p , onde γ_j é um indicador da presença de x_j no modelo.

O preditor linear de um modelo de regressão (ou, mais genericamente, de um modelo linear generalizado) determinado por γ pode ser escrito como apresentado em (3.5).

A informação amostral de cada modelo é representada por $f(y|b, \gamma)$ e a distribuição *a priori* de (b, γ) representada por $h(b, \gamma) = h(b|\gamma) h(\gamma)$, onde $h(b|\gamma)$ é a distribuição *a priori* do vector de parâmetros b , condicional a γ , e $h(\gamma)$ é a probabilidade *a priori* do modelo identificado por γ . Se $m \in \mathcal{M}$ for o modelo identificado por γ então $h(\gamma) = \Pr\{m\}$.

O vector b é particionado em dois subvectores b_γ e $b_{(-\gamma)}$, sendo b_γ o vector com as componentes de b que estão incluídas no modelo identificado por γ e $b_{(-\gamma)}$ o vector com as componentes de b que estão excluídas do modelo identificado por γ . De acordo com esta partição de b , a distribuição *a priori* conjunta de (b, γ) pode ser reescrita:

$$h(b, \gamma) = h(b_\gamma|\gamma) h(b_{(-\gamma)}|b_\gamma, \gamma) h(\gamma).$$

Atendendo à partição de b e à forma do preditor linear, a informação amostral pode ser simplificada

$$f(y|b, \gamma) = f(y|b_\gamma, \gamma)$$

e, portanto, a distribuição *a posteriori* de (b, γ) é

$$\begin{aligned} h(b, \gamma|y) &\propto f(y|b, \gamma) h(b, \gamma) \\ &\propto f(y|b_\gamma, \gamma) h(b_\gamma|\gamma) h(b_{(-\gamma)}|b_\gamma, \gamma) h(\gamma). \end{aligned}$$

Como o vector de parâmetros $b_{(-\gamma)}$ não é influenciado pela informação amostral do modelo identificado por γ , então a distribuição *a posteriori* de b para cada modelo identificado por γ é dada por

$$h(b|y, \gamma) = h(b_\gamma|y, \gamma) h(b_{(-\gamma)}|b_\gamma, \gamma),$$

onde $h(b_\gamma|y, \gamma)$ é a distribuição *a posteriori* dos parâmetros incluídos no modelo identificado por γ e $h(b_{(-\gamma)}|b_\gamma, \gamma)$ é a distribuição *a priori* condicional dos parâmetros não incluídos no modelo identificado por γ .

A distribuição $h(b_\gamma|\gamma)$ é a efectiva distribuição *a priori* do vector de parâmetros b_γ do modelo identificado por γ , enquanto a distribuição $h(b_{(-\gamma)}|b_\gamma, \gamma)$ recebe a designação de distribuição *pseudoprior* pelo facto da distribuição do vector de parâmetros $b_{(-\gamma)}$ não ser influenciado pela informação amostral do modelo identificado por γ nem influenciar a distribuição *a posteriori* dos parâmetros do modelo identificado por γ , $h(b_\gamma|y, \gamma)$. As distribuições *pseudoprior* devem, no entanto, ser especificadas com algum cuidado pois têm influência na performance do algoritmo MCMC.

As distribuições condicionais completas *a posteriori* de b_γ , $b_{(-\gamma)}$ e γ_j são dadas, respectivamente, por:

$$\begin{aligned} h(b_\gamma|b_{(-\gamma)}, \gamma, y) &\propto f(y|b, \gamma) h(b_\gamma|\gamma) h(b_{(-\gamma)}|b_\gamma, \gamma), \\ h(b_{(-\gamma)}|b_\gamma, \gamma, y) &\propto h(b_{(-\gamma)}|b_\gamma, \gamma) \end{aligned}$$

e

$$\frac{h(\gamma_j = 1|\gamma_{(-j)}, b, y)}{h(\gamma_j = 0|\gamma_{(-j)}, b, y)} = \frac{f(y|b, \gamma_j = 1, \gamma_{(-j)}) h(b|\gamma_j = 1, \gamma_{(-j)}) h(\gamma_j = 1, \gamma_{(-j)})}{f(y|b, \gamma_j = 0, \gamma_{(-j)}) h(b|\gamma_j = 0, \gamma_{(-j)}) h(\gamma_j = 0, \gamma_{(-j)})}. \quad (3.7)$$

A escolha das distribuições *a priori* e das distribuições *pseudoprior* constitui um aspecto muito importante do método GVS. A solução mais simples consiste em considerar que $h(b|\gamma)$ é dada pelo produto de distribuições *a priori* e distribuições *pseudoprior* independentes, isto é, $h(b|\gamma) = \prod_{j=1}^p h(b_j|\gamma_j)$. Neste caso, a escolha mais simples para $h(b_j|\gamma_j)$ é:

$$h(b_j|\gamma_j) = \gamma_j h(b_j|\gamma_j = 1) + (1 - \gamma_j) h(b_j|\gamma_j = 0), \quad (3.8)$$

que dá origem à distribuição *a priori* $h(b_\gamma|\gamma) = \prod_{j=1}^p h(b_j|\gamma_j)$ e à distribuição *pseudoprior* $h(b_{(-\gamma)}|b_\gamma, \gamma) = \prod_{j=0} h(b_j|\gamma_j)$.

A distribuição *a priori* simplificada apresentada em (3.8) e a formulação do modelo (3.5), dão origem à distribuição condicional completa *a posteriori*

$$\begin{aligned} h(b_j|b_{(-j)}, \gamma, y) &\propto f(y|b_\gamma, \gamma) \prod_{k=1}^p h(b_k|\gamma_k) \\ &\propto \begin{cases} f(y|b, \gamma) h(b_j|\gamma_j = 1) & , \text{ se } \gamma_j = 1 \\ h(b_j|\gamma_j = 0) & , \text{ se } \gamma_j = 0 \end{cases} \end{aligned}$$

evidenciando que a distribuição *pseudoprior* $h(b_j|\gamma_j = 0)$ não afecta a distribuição *a posteriori* de cada um dos coeficientes do modelo.

De modo similar ao efectuado em George e McCulloch (1993), Dellaportas *et al.* (2000) consideram, para distribuição *a priori* de $b_j|\gamma_j$, uma mistura de distribuições normais:

$$b_j|\gamma_j \sim \gamma_j N(0, \Sigma_j) + (1 - \gamma_j) N(\bar{\mu}_j, S_j), \quad j = 1, 2, \dots, p.$$

Os valores dos hiperparâmetros Σ_j , $\bar{\mu}_j$ e S_j podem ser obtidos através das estimativas de máxima verosimilhança ou das estimativas obtidas via métodos MCMC para o modelo completo (*e.g.*, Dellaportas e Forster, 1999). Contudo, em Ntzoufras (1999), a escolha automática de $\bar{\mu}_j = 0$ e $S_j = \Sigma_j/k^2$, com $k = 10$, mostrou ser uma opção adequada para os valores destes hiperparâmetros; o parâmetro k é apenas um parâmetro da distribuição *pseudoprior* e não afecta a distribuição *a posteriori*.

No caso de não existir restrição em relação ao espaço dos modelos, é usual considerar uma distribuição *a priori* não informativa para o vector γ , isto é, considerar γ_j , para $j = 1, \dots, p$, independentes e identicamente distribuídas $\text{Bern}(1/2)$.

Menciona-se um pormenor sobre a implementação do método GVS no WinBUGS: é conveniente utilizar o código $1 + \sum_{j=1}^p \gamma_j 2^{j-1}$ que permite transformar o vector $\gamma = (\gamma_1, \dots, \gamma_p)$, que identifica o modelo, num índice. Por exemplo, no caso de 3 covariáveis — x_1 , x_2 e x_3 — o código do modelo é dado por $1 + \gamma_1 + 2\gamma_2 + 4\gamma_3$; neste caso 1, 2, 3, 4, 5, 6, 7 e 8 são os índices dos modelos identificados pelos vectores (0, 0, 0), (1, 0, 0), (0, 1, 0), (1, 1, 0), (0, 0, 1), (1, 0, 1), (0, 1, 1) e (1, 1, 1), respectivamente. Mais pormenores sobre a implementação do método GVS no WinBUGS podem ser encontrados em Dellaportas *et al.* (2000) e Ntzoufras (2002).

A implementação deste método no WinBUGS não constitui uma tarefa difícil. No entanto, surgem algumas dificuldades na sua aplicação pois, mesmo com um número moderado de covariáveis, o número de modelos em comparação pode ser extremamente elevado.

3.4 Comparação dos métodos de Carlin e Chib, SSVS, de Kuo e Mallick, e GVS

As representações gráficas do método SSVS, do método Kuo e Mallick e do método GVS, apresentadas na figura 3.1, permitem uma melhor compreensão das semelhanças e das diferenças entre estes três métodos de selecção de covariáveis.

No método SSVS, a distribuição *a priori* de b_j é condicional a γ_j e o indicador γ_j não está presente no preditor linear. No método de Kuo e Mallick, a distribuição *a priori* de b_j é independente de γ , sendo o indicador do modelo incluído no preditor linear. O método GVS é um método híbrido do método SSVS e do método de Kuo e Mallick pois, tal como no método SSVS, a distribuição *a priori* de b_j é condicional a γ_j e, além disso, tal como no método Kuo e Mallick, o indicador γ_j é incluído no preditor linear.

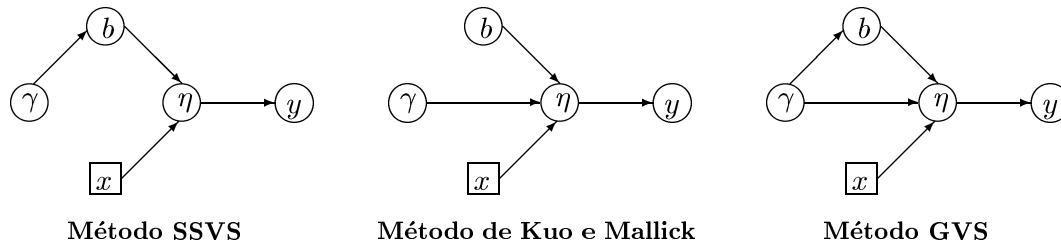


Figura 3.1: Representação gráfica dos métodos de selecção de covariáveis SSVS, de Kuo e Mallick, e GVS (Dellaportas *et al.*, 2000).

As semelhanças e as diferenças entre estes métodos podem também ser facilmente entendidas pela observação das respectivas distribuições condicionais completas (Dellaportas *et al.*, 2000). No método SSVS, $f(y|b, \gamma)$ é independente de γ e assim o primeiro quociente do lado direito de (3.6) e (3.7) não está presente em (3.4). No método de Kuo e Mallick, b e γ são *a priori* independentes, logo o segundo quociente do lado direito de (3.7), que coincide com o primeiro quociente do lado direito de (3.4), não está presente em (3.6).

As diferenças entre os métodos de Carlin e Chib, SSVS, de Kuo e Mallick e GVS dizem essencialmente respeito às distribuições *a priori* a às distribuições *pseudoprior*. No método de Carlin e Chib e no método GVS as distribuições *pseudoprior* apenas têm influência na eficiência do amostrador, funcionando unicamente como densidades de ligação. No método SSVS os parâmetros das distribuições *a priori* têm bastante impacto nas distribuições *a posteriori* e as distribuições *pseudoprior* não podem ser apenas consideradas como densidades de ligação. O método mais simples, mas também o menos flexível, é o de Kuo e Mallick que apenas requer a especificação das usuais distribuições *a priori* dos parâmetros do modelo.

Salienta-se, ainda, que o método de Carlin e Chib é um método de comparação de modelos, enquanto os outros métodos foram especificamente desenvolvidos para problemas de selecção de covariáveis em modelos de regressão. Conforme se ilustra na secção 3.5, para um número moderado de covariáveis, o método GVS permite estimar as probabilidades *a posteriori* dos modelos resultantes de todas as combinações possíveis de covariáveis; o mesmo não se passa com o método de Carlin e Chib, que só é viável utilizar quando o número de modelos em comparação for reduzido.

3.5 Exemplo de regressão linear normal

3.5.1 Introdução

Considere-se o modelo de regressão linear

$$Y_i = b_0 + b_1 x_{i1} + \cdots + b_p x_{ip} + \epsilon_i, \quad i = 1, \dots, n,$$

onde Y_i é uma variável resposta, b_0, b_1, \dots, b_p são parâmetros desconhecidos, x_{i1}, \dots, x_{ip} são observações de p covariáveis e ϵ_i é o erro aleatório, que tem distribuição normal com valor médio nulo e variância σ^2 , $\epsilon_i \sim N(0, \sigma^2)$. Escrevendo $b = (b_0, b_1, \dots, b_p)$ e $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$, tem-se

$$Y_i | b, \sigma^2, x_i \sim N(\mu_i, \sigma^2), \quad i = 1, \dots, n,$$

cuja função densidade de probabilidade é dada por

$$f(y_i | b, \sigma^2, x_i) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{y_i - \mu_i}{\sigma} \right)^2 \right\}, \quad y_i \in \mathbb{R} \text{ e } \sigma > 0,$$

com valor médio $\mu_i = b_0 + b_1 x_{i1} + \cdots + b_p x_{ip}$ e variância σ^2 (independente de i). Caso se pretenda representar por θ o vector de parâmetros deste modelo, tem-se $\theta = (b_0, b_1, \dots, b_p, \sigma^2)$. As variáveis aleatórias Y_i , para $i = 1, 2, \dots, n$, são condicionalmente independentes dado θ . O modelo de regressão linear é também designado por modelo de regressão normal, pelo facto das componentes do erro aleatório terem distribuição normal.

Um dos objectivos da regressão consiste na estimação dos parâmetros desconhecidos b_0, b_1, \dots, b_p e σ^2 . Perante um conjunto de dados, a informação *a priori* sobre estes parâmetros desconhecidos é actualizada pelas respectivas distribuições *a posteriori*, permitindo avaliar quais as covariáveis que têm um efeito relevante na variável resposta.

Outro objectivo da regressão é a predição de observações da variável resposta referentes a novos indivíduos, com base nas observações das p covariáveis. Dado um novo indivíduo com vector de observações $(x_{n+1,1}, \dots, x_{n+1,p})$, pode obter-se uma estimativa de uma observação futura de Y , através do valor esperado de Y_{n+1} , calculado relativamente à distribuição preditiva *a posteriori*, $p(y_{n+1} | \mathcal{D})$; isto é,

$$\hat{y}_{n+1} = E_{Y_{n+1} | \mathcal{D}}[Y_{n+1}] = \int y_{n+1} p(y_{n+1} | \mathcal{D}) dy_{n+1} = \iint y_{n+1} f(y_{n+1} | \theta) h(\theta | \mathcal{D}) dy_{n+1} d\theta,$$

onde \mathcal{D} representa a informação amostral.

Na secção 4.3.2 descreve-se com maior detalhe a análise bayesiana do modelo de regressão linear normal. Usando a distribuição *a priori* não informativa para o vector de parâmetros do modelo,

$$h(b, \sigma^2) = (\sigma^2)^{-1}, \quad \text{para } b \in \mathbb{R}^{p+1} \text{ e } \sigma^2 > 0, \quad (3.9)$$

apresenta-se a distribuição *a posteriori* do vector de parâmetros (b, σ^2) , assim como a distribuição *a posteriori* de σ^2 e a distribuição *a posteriori* de b condicional a σ^2 .

Na abordagem bayesiana aos modelos de regressão linear é usual modelar o parâmetro $\tau = 1/\sigma^2$, que é designado por precisão, em vez do parâmetro σ^2 . Esta opção deve-se ao facto de ser mais prático considerar uma distribuição *a priori* gama para o parâmetro τ do que uma distribuição gama invertida para σ^2 .

O WinBUGS só permite a utilização de distribuições *a priori* próprias, não sendo portanto possível considerar a distribuição não informativa (3.9). Na maior parte dos casos considera-se uma distribuição *a priori* para os coeficientes de regressão com variância elevada (Spiegelhalter *et al.*, 1996a, secção 9.2). Nas aplicações dos métodos de selecção de covariáveis, apresentadas nas secções 3.5.3 e 3.5.4, considera-se que a distribuição *a priori* de cada um dos coeficientes de regressão é $N(0, 100^2)$, o que se traduz numa distribuição *a priori* praticamente uniforme no intervalo com verosimilhança significativa do parâmetro. Especifica-se, também, uma distribuição *a priori* própria difusa para a precisão, $\tau \sim \text{Ga}(\epsilon, \epsilon)$, com valor médio 1 e variância $1/\epsilon$. Considera-se ϵ pequeno de modo que a variância *a priori* de τ seja elevada; nas aplicações já referidas usa-se $\epsilon = 0.001$.

Existem muitas referências na área dos modelos de regressão: a abordagem clássica a este tipo de modelos pode ser vista em Weisberg (1985), Jobson (1991) e Ryan (1997) e a abordagem bayesiana em Gelman *et al.* (1995), Paulino *et al.* (2003) e Rowe (2003).

3.5.2 Dados da massa gorda

Dois indivíduos do mesmo sexo, idade, altura e peso podem apresentar diferenças relativamente à sua composição corporal, isto é, diferenças nas proporções relativas dos principais componentes corporais — a água, a massa gorda e a massa livre de gordura (*e.g.*, Fragoso e Vieira, 2000). A percentagem de massa gorda é consensualmente aceite como um indicador da aptidão física e de potenciais problemas de saúde, sendo portanto uma característica muito estudada em populações humanas.

Existem diversos métodos para avaliação da composição corporal e, conseqüentemente, para determinação da percentagem de massa gorda. Os métodos directos, que se baseiam na dissecação de cadáveres, são os únicos que são precisos mas, obviamente, raramente utilizados. Os métodos indirectos baseiam-se em medidas quantitativas dos diferentes componentes corporais obtidas, por exemplo, através da hidrodensitometria (também conhecida como pesagem hidrostática), da excreção de creatinina ou da ressonância magnética. Estes métodos, apesar de terem uma precisão elevada, são pouco utilizados pois requerem equipamentos laboratoriais sofisticados e implicam exames morosos e caros, sendo essencialmente utilizados na validação dos métodos duplamente indirectos. Os métodos duplamente indirectos são os mais utilizados, pois são económicos e de fácil aplicação. Entre eles destaca-se a impedância bioeléctrica, isto é, a medida da resistência total do corpo à passagem de uma corrente eléctrica mínima, e as medidas antropométricas, como é o caso das circunferências corporais ou das espessuras de pregas adiposas.

A medição de circunferências corporais, em conjugação com o peso e a altura, permite estimar a percentagem de massa gorda através de equações preditivas, que podem ser determinadas por métodos de regressão linear múltipla. Selecciona-se vários indivíduos de uma determinada população e por aplicação de um método indirecto, por exemplo a hidrodensitometria, são calculadas estimativas relativamente precisas da percentagem de

massa gorda¹. Além disso, são também registadas diversas medidas de circunferências corporais de cada um dos indivíduos, assim como a sua idade, peso e altura. Recorrendo à regressão linear múltipla pode obter-se, com base nestas medidas, uma equação preditiva para a percentagem de massa gorda. Esta equação pode ser usada para estimar a percentagem de massa gorda em indivíduos provenientes da mesma população dos indivíduos que deram origem à equação preditiva.

O conjunto de dados utilizado para exemplificação da selecção bayesiana de covariáveis em modelos de regressão linear normal, originalmente referido por Penrose *et al.* (1985) e apresentado em Johnson (1996), é constituído por observações de 14 variáveis, nomeadamente, a percentagem de massa gorda², a idade, o peso, a altura e 10 medidas de circunferências corporais, de 251 homens, com idades compreendidas entre os 22 e os 81 anos (tabela C.1 do anexo C). A variável resposta no modelo de regressão linear normal é a percentagem de massa gorda (variável que é mais difícil de medir), sendo as restantes, as variáveis explicativas ou covariáveis, de fácil obtenção. A definição das variáveis e algumas medidas de estatística descritiva, nomeadamente a média, o desvio padrão, o mínimo e o máximo, do conjunto de dados da massa gorda são apresentados na tabela 3.3.

Tabela 3.3: Estatísticas sumárias das variáveis do exemplo da massa gorda¹.

variável	definição	média	d.p.	mín.	máx.
y	massa gorda (percentagem)	18.89	7.72	0	45.10
x_1	idade (anos)	44.89	12.63	22	81
x_2	peso (libras ¹)	178.82	29.40	118.5	363.15
x_3	altura (polegadas ¹)	70.31	2.61	64.0	77.75
x_4	circunferência do pescoço (cm)	38.00	2.43	31.1	51.20
x_5	circunferência do peito (cm)	100.80	8.44	79.3	136.20
x_6	circunferência do abdómen (cm)	92.51	10.78	69.4	148.10
x_7	circunferência do quadril (cm)	99.84	7.11	85.0	147.70
x_8	circunferência da coxa (cm)	59.36	5.21	47.2	87.30
x_9	circunferência do joelho (cm)	38.58	2.40	33.0	49.10
x_{10}	circunferência do tornozelo (cm)	23.10	1.70	19.1	33.90
x_{11}	circunferência do bíceps em extensão (cm)	32.27	3.03	24.8	45.00
x_{12}	circunferência do antebraço (cm)	28.66	2.02	21.0	34.90
x_{13}	circunferência do pulso (cm)	18.23	0.93	15.8	21.40

¹1 libra \approx 454 gr e 1 polegada \approx 2.54 cm.

A abordagem bayesiana não dispensa o primeiro contacto com os dados através de uma análise exploratória. A eliciação da distribuição *a priori* do vector de parâmetros deve ser feita antes desta análise, de modo a não ser influenciada pela informação amostral. Na figura 3.2 apresenta-se, através de caixas de bigodes, outras medidas de estatística descritiva do conjunto de dados da massa gorda. Destaca-se a simetria da maior parte das variáveis e a existência de valores discordantes para algumas delas.

¹Na realidade a hidrodensitometria permite determinar a densidade corporal e, através de equações desenvolvidas para esse efeito (equação de Siri e de Brozek, por exemplo), estimar a percentagem de massa gorda (*e.g.*, Costa, 2001).

²Estimada através da equação de Brozek e com recurso à técnica de hidrodensitometria.

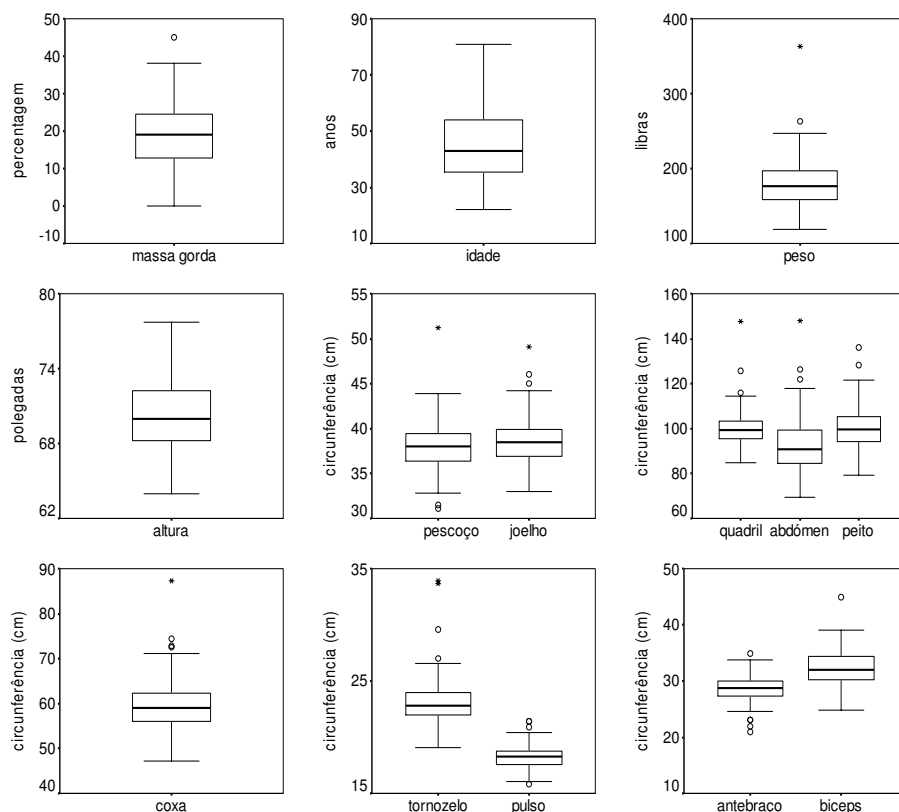


Figura 3.2: Caixas de bigodes das variáveis do exemplo da massa gorda.

Na figura 3.3 são apresentados os diagramas de dispersão da percentagem de massa gorda *versus* cada uma das variáveis explicativas e os respectivos valores do coeficiente de correlação amostral de Pearson. Refira-se que a correlação linear entre a percentagem de massa gorda e as variáveis idade, altura, circunferência do tornozelo, circunferência do antebraço e circunferência do pulso é fraca e, no caso da altura, é negativa. Relativamente à correlação linear da percentagem de massa gorda com as variáveis peso, circunferência do peito, circunferência do abdômen e circunferência do quadril, regista-se valores mais elevados. Como seria de esperar, devido à natureza das próprias covariáveis, o coeficiente de correlação amostral de Pearson entre algumas delas é também bastante elevado.

Na análise a este conjunto de dados, Hoeting *et al.* (1999) começam por efectuar a estimação, por mínimos quadrados, do modelo de regressão linear múltipla com todas as covariáveis. O *output* do SPSS com as estimativas dos coeficientes de regressão é apresentado na tabela 3.4. Para este modelo obtém-se $R^2 = 0.746$, $R^2_{aj} = 0.732$ e um erro standard da estimativa igual a 3.996. O valor observado da estatística F com 13 e 237 graus de liberdade é 53.620, ao qual corresponde um $p\text{-value} < 0.0001$.

Algumas das covariáveis incluídas no modelo de regressão linear normal podem ser supérfluas, isto é, a sua inclusão no modelo aumenta a complexidade do mesmo sem um acréscimo significativo na explicação da variabilidade da variável resposta. Para encontrar um modelo mais parcimonioso e que possa ainda ser considerado um bom modelo, pode usar-se qualquer um dos métodos de selecção de covariáveis utilizados na

abordagem clássica. Tendo em conta as opções existentes no SPSS, utiliza-se 3 métodos de selecção de covariáveis: o progressivo, o regressivo e o passo a passo. Para um F de entrada e de saída de 0.05, os resultados obtidos por estes três métodos são iguais, sendo o modelo seleccionado constituído pelas covariáveis circunferência do abdómen, peso, circunferência do pulso e circunferência do antebraço. O *output* do SPSS com as estimativas dos coeficientes de regressão, obtidas por mínimos quadrados, é apresentado na tabela 3.5. Para este modelo obtém-se $R^2 = 0.733$, $R_{aj}^2 = 0.729$ e um erro standard da estimativa igual a 4.021. O valor observado da estatística F com 4 e 246 graus de liberdade é 169.124, ao qual corresponde um $p\text{-value} < 0.0001$.

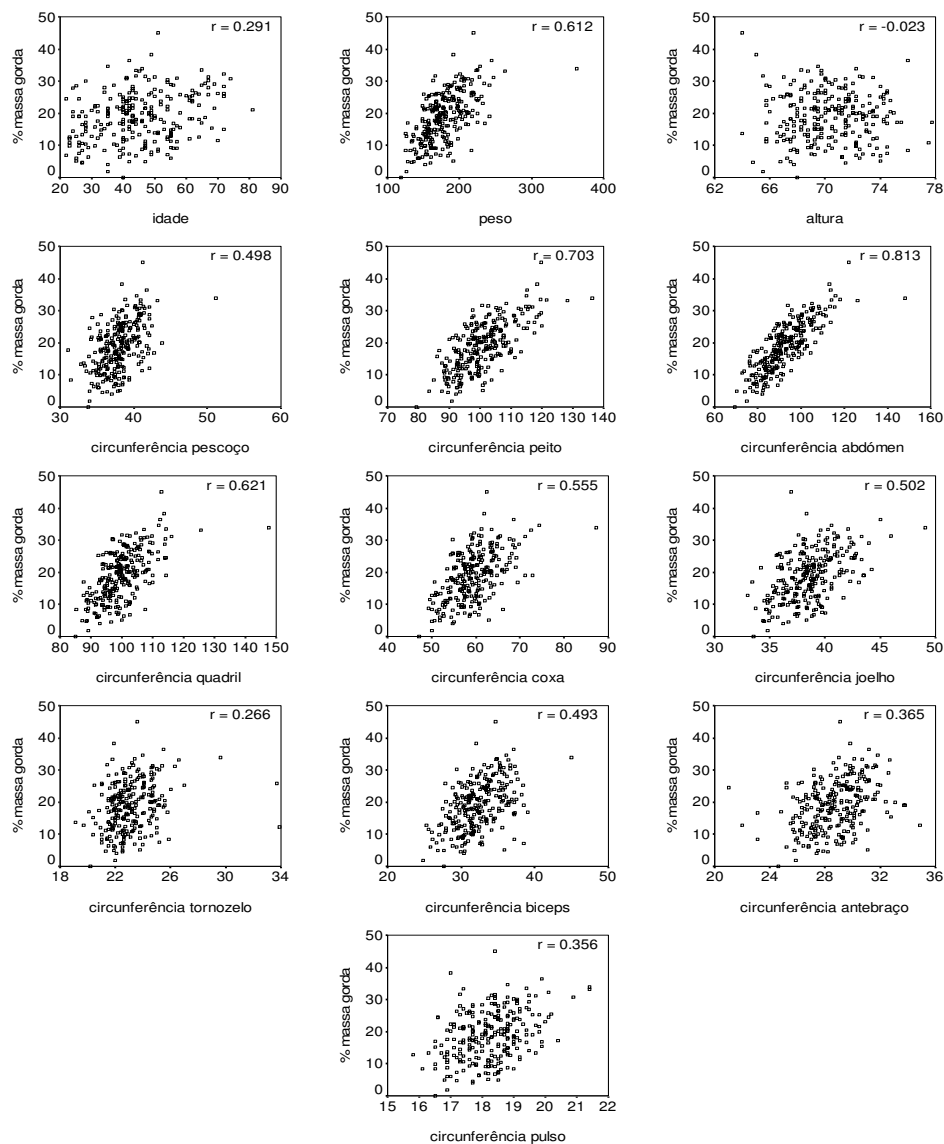


Figura 3.3: Diagramas de dispersão da percentagem de massa gorda *versus* cada uma das covariáveis.

Tabela 3.4: *Output* do SPSS com as estimativas dos coeficientes do modelo de regressão linear com 13 covariáveis, para o exemplo da massa gorda.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-17.801	20.604		-.864	.388
idade	.057	.030	.093	1.893	.060
peso	-.086	.057	-.327	-1.496	.136
altura	-.037	.166	-.013	-.225	.822
circunferência pescoço	-.430	.219	-.136	-1.965	.051
circunferência peito	-.018	.096	-.020	-.193	.847
circunferência abdómen	.890	.084	1.242	10.624	.000
circunferência quadril	-.196	.136	-.180	-1.440	.151
circunferência coxa	.236	.136	.160	1.739	.083
circunferência joelho	-.021	.230	-.007	-.093	.926
circunferência tornozelo	.167	.206	.037	.810	.419
circunferência biceps	.157	.160	.061	.978	.329
circunferência antebraço	.429	.185	.113	2.322	.021
circunferência pulso	-1.474	.497	-.178	-2.969	.003

Tabela 3.5: *Output* do SPSS com as estimativas dos coeficientes do modelo de regressão linear com 4 covariáveis, para o exemplo da massa gorda.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-31.995	6.743		-4.745	.000
circunferência abdómen	.920	.052	1.284	17.713	.000
peso	-.127	.023	-.483	-5.530	.000
circunferência pulso	-1.335	.414	-.161	-3.228	.001
circunferência antebraço	.448	.168	.117	2.661	.008

3.5.3 Métodos de selecção via medidas de discrepância

A aplicação dos três métodos de selecção de covariáveis via medidas de discrepância, ao exemplo da massa gorda, foi efectuada com quatro medidas de discrepância: DIC, \widehat{BIC} , LLP e SlnCPO. Para qualquer um dos métodos e qualquer uma das medidas de discrepância, o ajustamento de cada um dos modelos é efectuado: (i) considerando distribuições *a priori* difusas para todos os parâmetros do modelo, nomeadamente, b_0, b_1, \dots, b_p independentes com distribuição normal de valor médio nulo e variância 10000, e τ com distribuição gama com parâmetro de forma e inverso do parâmetro de escala iguais a 0.001; (ii) com base nos resultados obtidos em 10000 iterações do método de amostragem, após um período de aquecimento de 1000 iterações.

Devido ao elevado número de modelos ajustados, a monitorização da convergência é apenas efectuada por análise dos gráficos das séries temporais e das funções de autocorrelação empírica dos valores simulados. Para cada um dos modelos ajustados e para cada um dos parâmetros monitorizados, a série temporal das iterações não sugere a existência de qualquer anomalia no seu comportamento estocástico, e a função de autocorrelação empírica permite considerar válida a suposição de independência entre os valores simulados.

Método SIMD

Começa por ajustar-se um modelo de regressão linear normal com as 13 covariáveis e calcula-se os valores das medidas de discrepância. O código do WinBUGS, para a aplicação da 1ª iteração do método SIMD ao exemplo da massa gorda, é apresentado no anexo D.1.1. Na figura 3.4 são apresentados os gráficos das densidades marginais *a posteriori* dos coeficientes de regressão. O *output* com as estimativas dos parâmetros do modelo de regressão linear normal com as 13 covariáveis é apresentado na tabela 3.6. Os valores das medidas de discrepância são apresentados na tabela 3.7, na coluna referente à 1ª iteração. Da análise da figura 3.4 e da tabela 3.6 destaca-se que:

- os intervalos de credibilidade a 95% dos parâmetros de regressão b_6 , b_{12} e b_{13} são os únicos que não contêm o zero. As covariáveis x_6 , x_{12} e x_{13} não são, portanto, candidatas a sair do modelo;
- embora os intervalos de credibilidade a 95% dos parâmetros de regressão b_1 , b_2 , b_4 , b_7 e b_8 contenham o zero, os gráficos das densidades marginais *a posteriori* destes parâmetros não estão concentrados em torno de zero. As covariáveis x_1 , x_2 , x_4 , x_7 e x_8 não são também candidatas a sair do modelo;
- dos restantes parâmetros de regressão, apenas os intervalos de credibilidade a 95% de b_3 , b_5 e b_9 são aproximadamente centrados em zero. Entre as covariáveis x_3 , x_5 e x_9 opta-se por excluir do modelo a covariável x_5 , uma vez que a densidade marginal *a posteriori* do parâmetro de regressão b_5 é a que se encontra mais concentrada em torno de zero.

Tabela 3.6: *Output* do WinBUGS com as estimativas dos parâmetros do modelo de regressão linear com 13 covariáveis, para o exemplo da massa gorda.

node	mean	sd	MC error	2.50%	median	97.50%	start	sample
b0	-17.74	20.5	0.2171	-57.18	-17.85	23.02	1001	10000
b[1]	0.05642	0.03021	3.02E-04	-0.002479	0.05591	0.1164	1001	10000
b[2]	-0.08546	0.05741	5.83E-04	-0.1969	-0.08563	0.02877	1001	10000
b[3]	-0.03523	0.1654	0.001747	-0.3644	-0.03225	0.295	1001	10000
b[4]	-0.4297	0.2194	0.001956	-0.8613	-0.4293	2.43E-04	1001	10000
b[5]	-0.01832	0.09592	0.001019	-0.2042	-0.01822	0.1714	1001	10000
b[6]	0.8905	0.08493	8.63E-04	0.7246	0.8908	1.057	1001	10000
b[7]	-0.1969	0.1348	0.001605	-0.4615	-0.1979	0.06486	1001	10000
b[8]	0.2347	0.1365	0.001308	-0.03321	0.2344	0.5064	1001	10000
b[9]	-0.0232	0.231	0.002165	-0.4799	-0.02384	0.4331	1001	10000
b[10]	0.1647	0.2056	0.002127	-0.2433	0.1654	0.5674	1001	10000
b[11]	0.1575	0.1624	0.001519	-0.161	0.156	0.4817	1001	10000
b[12]	0.4265	0.1873	0.00191	0.06563	0.4265	0.7985	1001	10000
b[13]	-1.473	0.4997	0.005138	-2.451	-1.477	-0.48	1001	10000
tau	0.06248	0.005743	5.80E-05	0.05169	0.06233	0.0742	1001	10000

Na 2ª iteração ajusta-se um modelo de regressão linear normal com 12 covariáveis, uma vez que x_5 foi excluída do modelo. O código do WinBUGS para o ajustamento deste novo modelo de regressão linear normal é idêntico ao apresentado para a 1ª iteração, exceptuando o facto de $p = 12$ e ser necessário excluir do modelo o termo respeitante à covariável x_5 . Os valores obtidos para as medidas de discrepância são apresentados na tabela 3.7, na coluna referente à 2ª iteração. Comparando a 1ª com a 2ª iteração, verifica-se a existência de valores mais favoráveis para as medidas de discrepância na 2ª

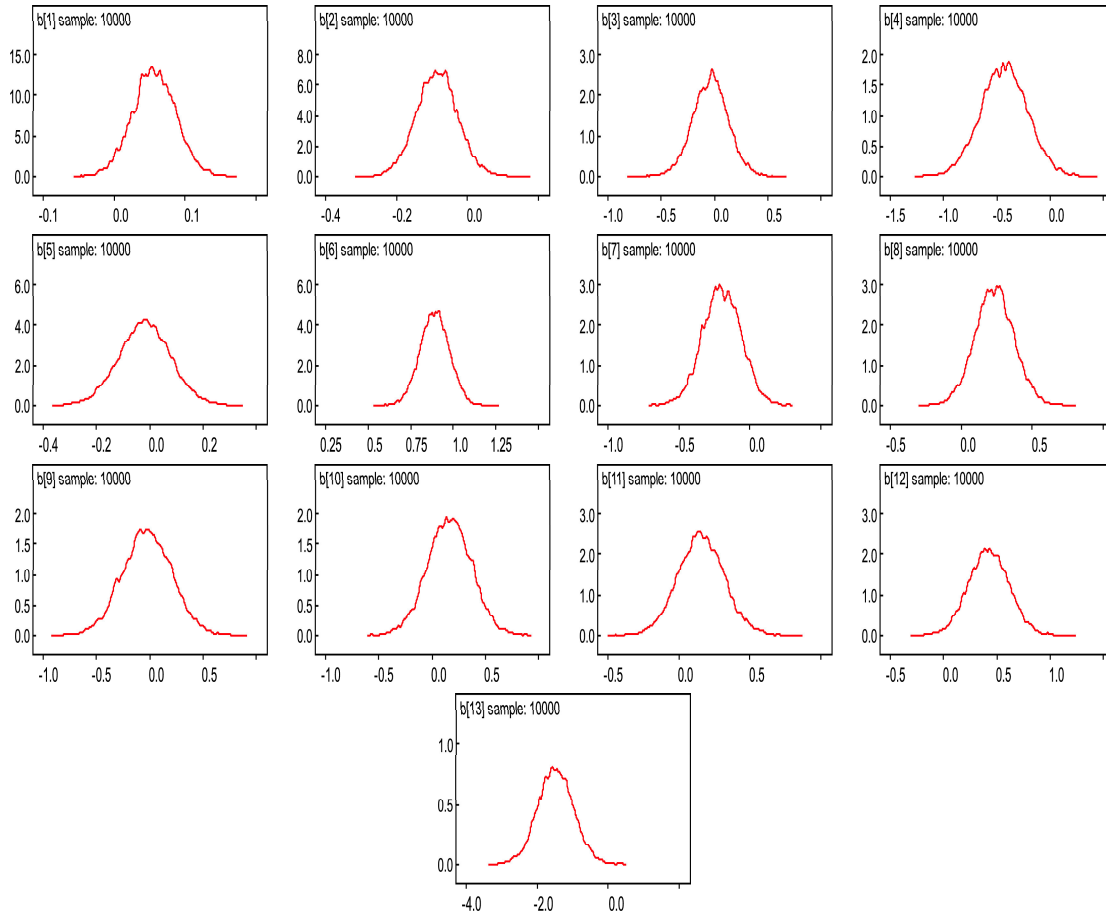


Figura 3.4: Gráficos das densidades marginais *a posteriori* dos coeficientes do modelo de regressão linear com 13 covariáveis, para o exemplo da massa gorda.

iteração, o que quer dizer que o método SIMD prossegue. Analisando as estimativas dos parâmetros da regressão e respectivos gráficos das densidades marginais *a posteriori* do modelo ajustado na 2ª iteração — que opta-se por não apresentar para não alongar demasiado a ilustração do método SIMD —, conclui-se que x_3 é a próxima covariável a ser excluída do modelo. Na 3ª iteração ajusta-se um modelo de regressão linear normal com 11 covariáveis, calcula-se os valores das medidas de discrepância e averigua-se qual a próxima covariável a sair do modelo.

Procede-se deste modo até já não existir uma evidência favorável à exclusão de mais alguma covariável ou, caso exista, se essa exclusão originar valores mais desfavoráveis para as medidas de discrepância consideradas.

Na figura 3.5 são apresentados os gráficos das densidades marginais *a posteriori* dos coeficientes de regressão e na tabela 3.8 os *outputs* do WinBUGS com as estimativas dos parâmetros, referentes à 6ª iteração, em que é ajustado um modelo de regressão linear múltipla com 8 covariáveis, mais precisamente, com $x_1, x_2, x_4, x_6, x_7, x_8, x_{12}$ e x_{13} . Da análise da figura 3.5 e da tabela 3.8, conclui-se que nenhuma das densidades marginais *a posteriori* dos parâmetros de regressão está concentrada em torno de zero. Assim, o

método SIMD pára na 6^a iteração, pois mais nenhuma covariável é candidata a sair do modelo. O resumo da aplicação do método SIMD é apresentado na tabela 3.7.

Tabela 3.7: Resultados do método SIMD, para o exemplo da massa gorda.

iteração	1 ^a it.	2 ^a it.	3 ^a it.	4 ^a it.	5 ^a it.	6 ^a it.
covariáveis	todas	sai x_5	sai x_3	sai x_9	sai x_{11}	sai x_{10}
DIC	1423.8	1421.7	1419.8	1417.8	1416.8	1415.4
$\widehat{\text{BIC}}$	1492	1485	1479	1472	1467	1461
LLP	1408.8	1407.7	1406.8	1405.7	1405.8	1405.3
SlnCPO	-713.1	-711.8	-710.6	-709.5	-709.1	-708.0

Tabela 3.8: *Output* do WinBUGS com as estimativas dos parâmetros do modelo de regressão linear com 8 covariáveis, para o exemplo da massa gorda.

node	mean	sd	MC error	2.50%	median	97.50%	start	sample
b0	-19.38	10.87	0.1152	-40.72	-19.25	1.832	1001	10000
b[1]	0.0568	0.0285	2.80E-04	0.001822	0.05659	0.1133	1001	10000
b[2]	-0.08182	0.03694	3.72E-04	-0.1546	-0.08174	-0.008863	1001	10000
b[4]	-0.424	0.2102	0.002129	-0.8358	-0.423	-0.01139	1001	10000
b[6]	0.8808	0.06649	7.16E-04	0.7492	0.8818	1.012	1001	10000
b[7]	-0.2032	0.1305	0.001159	-0.4516	-0.2037	0.05312	1001	10000
b[8]	0.2794	0.1217	0.001194	0.04066	0.2798	0.5159	1001	10000
b[12]	0.4743	0.173	0.001809	0.1344	0.4737	0.8131	1001	10000
b[13]	-1.367	0.4749	0.004294	-2.279	-1.37	-0.4377	1001	10000
tau	0.06352	0.005754	5.89E-05	0.05273	0.06333	0.07542	1001	10000

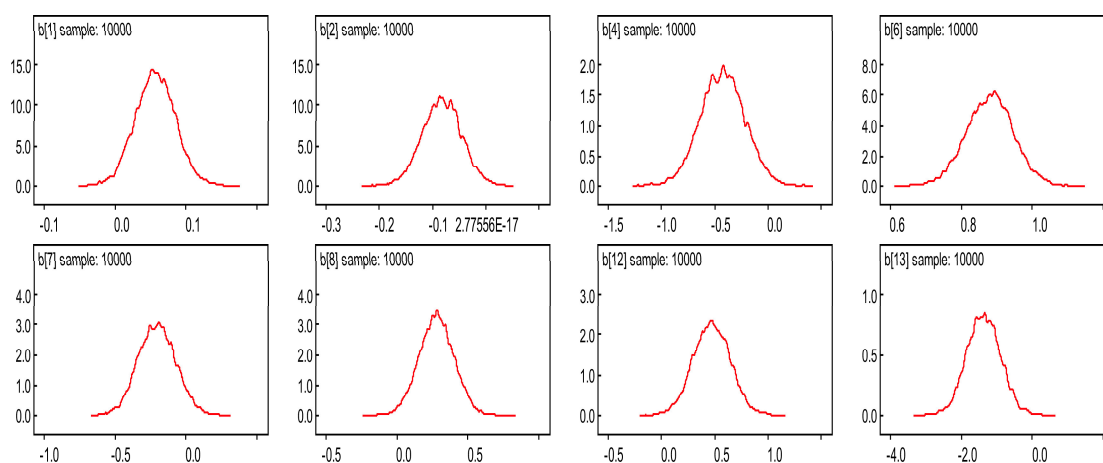


Figura 3.5: Gráficos das densidades marginais *a posteriori* dos coeficientes do modelo de regressão linear com 8 covariáveis, para o exemplo da massa gorda.

Método FVS

O modo de implementação, no WinBUGS, do método FVS para selecção do melhor subconjunto de covariáveis do modelo de regressão linear normal, para os dados da massa gorda, está descrito no anexo D.1.2. Os resultados da aplicação deste método ao referido conjunto de dados, para as medidas de discrepância DIC, \widehat{BIC} , LLP e SlnCPO, estão sumariados nas tabelas 3.9, 3.10, 3.11 e 3.12, respectivamente.

Procede-se, de seguida, à interpretação da tabela 3.9. O modelo só com a constante, ajustado na 1ª iteração, apresenta DIC = 1741.6. Na 2ª iteração são ajustados 13 modelos de regressão linear normal, cada um com uma única covariável, e calcula-se os valores do DIC de cada um deles. O valor do DIC do modelo de regressão com a covariável x_6 encontra-se sublinhado para evidenciar que é o menor dos valores do DIC obtidos nesta iteração e que, além disso, é menor que o valor do DIC da iteração anterior, o que origina a inclusão da covariável x_6 no modelo.

Tabela 3.9: Resultados do método FVS via DIC, para o exemplo da massa gorda¹.

cov.	2ª it.	3ª it.	4ª it.	5ª it.	6ª it.	7ª it.	8ª it.	9ª it.	10ª it.	11ª it.
x_1	1721.4	1465.8	1430.0	1424.1	1418.1	1417.2	<u>1416.8</u>	◇	◇	◇
x_2	1626.0	<u>1428.1</u>	◇	◇	◇	◇	◇	◇	◇	◇
x_3	1743.5	1448.6	1429.8	1424.7	1419.8	1419.2	1419.2	1418.8	1418.7	1418.4
x_4	1672.0	1448.5	1426.4	1423.9	<u>1417.4</u>	◇	◇	◇	◇	◇
x_5	1572.4	1466.3	1430.1	1424.9	1419.8	1419.5	1419.2	1418.8	1418.7	1418.5
x_6	<u>1472.3</u>	◇	◇	◇	◇	◇	◇	◇	◇	◇
x_7	1621.5	1446.9	1430.0	1424.0	1419.4	1418.5	1418.1	1418.4	<u>1416.5</u>	◇
x_8	1651.1	1466.0	1426.5	1423.0	1418.5	1418.2	1418.6	<u>1416.6</u>	◇	◇
x_9	1670.6	1459.1	1429.9	1424.3	1419.2	1419.1	1418.9	1418.7	1418.7	1418.5
x_{10}	1725.2	1464.4	1430.0	1424.2	1418.9	1418.9	1418.5	1418.0	1417.9	1417.8
x_{11}	1673.7	1468.5	1426.9	1420.4	1418.2	<u>1417.2</u>	◇	◇	◇	◇
x_{12}	1707.7	1472.3	1426.3	<u>1417.8</u>	◇	◇	◇	◇	◇	◇
x_{13}	1709.6	1444.0	<u>1423.0</u>	◇	◇	◇	◇	◇	◇	◇
DIC	1472.3	1428.1	1423.0	1417.8	1417.4	1417.2	1416.8	1416.6	1416.5	◇
decr.	269.3	44.2	5.1	5.2	0.4	0.2	0.4	0.2	0.1	◇

¹Na primeira iteração, em que é ajustado um modelo só com a constante, tem-se DIC = 1741.6.

Em cada iteração o valor sublinhado corresponde ao valor mais favorável para a medida de discrepância, indicando qual a covariável a ser incluída no modelo. O símbolo “◇”, associado a uma determinada covariável a partir da t -ésima iteração, indica que essa covariável foi incluída no modelo na iteração $t - 1$. Na penúltima e última linhas da tabela 3.9 são apresentados, respectivamente, o valor mínimo do DIC obtido na iteração correspondente e o decréscimo registado no valor do DIC relativamente à iteração anterior.

Na 3ª iteração, ajusta-se 12 modelos de regressão linear normal, cada um deles com as covariáveis x_6 e x_k , para $k = 1, \dots, 13$ e $k \neq 6$; como a maior redução no DIC é obtida com a inclusão de x_2 , esta é a próxima covariável a ser incluída no modelo. Procede-se deste modo até à 11ª iteração, em que já estão no modelo 9 covariáveis e em que a inclusão de qualquer uma das outras não produz um decréscimo no DIC.

O conjunto de covariáveis seleccionadas pelo método FVS, quando se utiliza como medida de discrepância o DIC, é constituído pelas covariáveis: $x_6, x_2, x_{13}, x_{12}, x_4, x_{11}, x_1, x_8$ e x_7 . Uma vez que os decréscimos no valor do DIC são, a partir da 5ª iteração, bastante pequenos, não seria despropositado escolher o modelo obtido na 5ª iteração. Neste caso, o conjunto de covariáveis seleccionadas seria constituído por: x_6, x_2, x_{13} e x_{12} . A escolha de qualquer subconjunto de covariáveis que contenha estas 4 covariáveis e que esteja contido no subconjunto das 9 covariáveis já referido, não deve ser totalmente descartada como escolha para melhor subconjunto de covariáveis.

A interpretação dos resultados do método FVS via $\widehat{\text{BIC}}$, apresentados na tabela 3.10, é idêntica à efectuada para a tabela 3.9. O modelo só com a constante, ajustado na 1ª iteração, apresenta $\widehat{\text{BIC}} = 1751$. Na 2ª iteração são ajustados 13 modelos de regressão linear normal, cada um deles com uma das covariáveis, e calcula-se os valores do $\widehat{\text{BIC}}$ destes modelos. Como o mínimo dos valores do $\widehat{\text{BIC}}$ é obtido com a inclusão de x_6 no modelo e este valor é inferior ao da iteração anterior, então a covariável x_6 é incluída no modelo. Na 3ª iteração entra para o modelo a covariável x_2 pois é a que origina o maior decréscimo no valor do $\widehat{\text{BIC}}$. O método FVS via $\widehat{\text{BIC}}$ pára na 4ª iteração pois a inclusão de qualquer uma das outras covariáveis no modelo não origina nenhum decréscimo no valor da medida de discrepância, sendo o modelo seleccionado constituído pelas covariáveis x_6 e x_2 .

Tabela 3.10: Resultados do método FVS via $\widehat{\text{BIC}}$, para o exemplo da massa gorda¹.

cov.	2ª it.	3ª it.	4ª it.
x_1	1735	1484	1453
x_2	1640	<u>1446</u>	◇
x_3	1757	1467	1452
x_4	1686	1467	1449
x_5	1586	1484	1453
x_6	<u>1486</u>	◇	◇
x_7	1635	1465	1453
x_8	1665	1484	1449
x_9	1684	1477	1452
x_{10}	1739	1482	1453
x_{11}	1687	1487	1449
x_{12}	1721	1490	1449
x_{13}	1723	1462	<u>1446</u>
$\widehat{\text{BIC}}$	1486	1446	◇
decr.	265	40	◇

¹Na primeira iteração, em que é ajustado um modelo só com a constante, tem-se $\widehat{\text{BIC}} = 1751$.

O conjunto de covariáveis seleccionadas pelo método FVS quando se utiliza o LLP ou o SlnCPO é o mesmo que o obtido quando se utiliza o DIC. De modo análogo ao efectuado quando se usa o DIC como medida de discrepância, e uma vez que a partir da 5ª iteração os decréscimos do LLP (última linha da tabela 3.11) e os acréscimos do SlnCPO (última linha da tabela 3.12) são muito pequenos, pode optar-se por parar nesta iteração e considerar como melhor modelo aquele que tem apenas as 4 primeiras covariáveis.

Tabela 3.11: Resultados do método FVS via LLP, para o exemplo da massa gorda¹.

cov.	2 ^a it.	3 ^a it.	4 ^a it.	5 ^a it.	6 ^a it.	7 ^a it.	8 ^a it.	9 ^a it.	10 ^a it.	11 ^a it.
x_1	1718.4	1461.8	1425.0	1418.1	1411.1	1409.2	<u>1407.8</u>	◇	◇	◇
x_2	1623.0	<u>1424.1</u>	◇	◇	◇	◇	◇	◇	◇	◇
x_3	1740.5	1444.6	1424.8	1418.7	1412.8	1411.2	1410.2	1408.7	1407.6	1406.4
x_4	1669.0	1444.5	1421.4	1417.9	<u>1410.4</u>	◇	◇	◇	◇	◇
x_5	1569.4	1462.2	1425.1	1418.9	1412.8	1411.5	1410.2	1408.8	1407.6	1406.4
x_6	<u>1469.3</u>	◇	◇	◇	◇	◇	◇	◇	◇	◇
x_7	1618.5	1442.9	1425.0	1418.1	1412.4	1410.5	1409.1	1408.4	<u>1405.4</u>	◇
x_8	1648.1	1462.0	1421.5	1417.0	1411.6	1410.2	1409.6	<u>1406.6</u>	◇	◇
x_9	1667.6	1455.1	1424.9	1418.3	1412.2	1411.1	1409.9	1408.7	1407.6	1406.4
x_{10}	1722.2	1460.4	1425.0	1418.2	1411.9	1410.9	1409.6	1407.9	1406.8	1405.7
x_{11}	1670.7	1464.5	1421.9	1414.4	1411.2	<u>1409.2</u>	◇	◇	◇	◇
x_{12}	1704.7	1468.3	1421.3	<u>1411.9</u>	◇	◇	◇	◇	◇	◇
x_{13}	1706.6	1440.0	<u>1418.0</u>	◇	◇	◇	◇	◇	◇	◇
LLP	1469.3	1424.1	1418.0	1411.9	1410.4	1409.2	1407.8	1406.6	1405.4	◇
decr.	270.3	45.2	6.1	6.1	1.5	1.2	1.4	1.2	1.2	◇

¹Na primeira iteração, em que é ajustado um modelo só com a constante, tem-se LLP = 1739.6.

Tabela 3.12: Resultados do método FVS via SlnCPO, para o exemplo da massa gorda¹.

cov.	2 ^a it.	3 ^a it.	4 ^a it.	5 ^a it.	6 ^a it.	7 ^a it.	8 ^a it.	9 ^a it.	10 ^a it.	11 ^a it.
x_1	-860.6	-733.7	-715.3	-712.8	-709.7	-709.1	<u>-708.8</u>	◇	◇	◇
x_2	-813.5	<u>-714.5</u>	◇	◇	◇	◇	◇	◇	◇	◇
x_3	-871.9	-725.5	-715.9	-713.6	-710.9	-710.6	-710.3	-710.3	-710.3	-709.8
x_4	-835.9	-724.9	-713.6	-712.6	<u>-709.2</u>	◇	◇	◇	◇	◇
x_5	-786.1	-734.3	-715.7	-713.3	-710.5	-710.3	-710.0	-710.0	-709.9	-709.6
x_6	<u>-737.1</u>	◇	◇	◇	◇	◇	◇	◇	◇	◇
x_7	-811.6	-724.0	-715.5	-712.7	-710.2	-709.6	-709.3	-709.6	<u>-708.5</u>	◇
x_8	-825.6	-733.8	-713.6	-712.1	-709.7	-709.4	-709.6	<u>-708.7</u>	◇	◇
x_9	-835.4	-730.4	-715.4	-712.8	-710.0	-709.9	-709.7	-709.8	-709.8	-709.5
x_{10}	-863.2	-734.5	-716.0	-713.3	-710.5	-710.4	-710.0	-709.7	-709.7	-709.5
x_{11}	-836.9	-735.1	-713.8	-710.8	-709.6	<u>-709.0</u>	◇	◇	◇	◇
x_{12}	-854.0	-737.7	-713.5	<u>-709.4</u>	◇	◇	◇	◇	◇	◇
x_{13}	-854.7	-723.0	<u>-712.2</u>	◇	◇	◇	◇	◇	◇	◇
SlnCPO	-737.1	-714.5	-712.2	-709.4	-709.2	-709.0	-708.8	-708.7	-708.5	◇
acr.	133.6	22.6	2.3	2.8	0.2	0.2	0.2	0.1	0.2	◇

¹Na primeira iteração, em que é ajustado um modelo só com a constante, tem-se SlnCPO = -870.7.

Método BVE

A implementação, no WinBUGS, do método BVE para selecção do melhor subconjunto de covariáveis do modelo de regressão linear normal, para os dados da massa gorda, está sumariada no anexo D.1.3. Os resultados da aplicação do método BVE a este conjunto de dados, com as medidas de discrepância DIC, $\widehat{\text{BIC}}$, LLP e SlnCPO, são apresentados nas tabelas 3.13, 3.14, 3.15 e 3.16, respectivamente.

Procede-se, de seguida, à interpretação da tabela 3.13. O modelo de regressão linear normal com todas as covariáveis, ajustado na 1ª iteração, apresenta $\text{DIC} = 1423.8$. Na 2ª iteração são ajustados 13 modelos de regressão linear normal, cada um deles resultante da exclusão de uma das covariáveis. O valor do DIC do modelo de regressão sem a covariável x_9 encontra-se sublinhado para evidenciar que é o menor dos valores do DIC obtidos nessa iteração e que, além disso, é menor que o valor da iteração anterior, o que origina a exclusão da covariável x_9 do modelo. Na 3ª iteração, ajusta-se 13 modelos de regressão linear normal, cada um deles resultante da exclusão das covariáveis x_9 e x_k , para $k = 1, \dots, 13$ e $k \neq 9$; o mínimo do DIC é obtido com a exclusão da covariável x_5 e é menor do que o valor do DIC da iteração anterior, logo esta covariável sai do modelo. Procede-se deste modo até à 7ª iteração, em que já foram excluídas do modelo 5 covariáveis e em que a exclusão de qualquer uma das outras covariáveis não produz um decréscimo no valor do DIC.

Tabela 3.13: Resultados do método BVE via DIC, para o exemplo da massa gorda¹.

cov.	2ª it.	3ª it.	4ª it.	5ª it.	6ª it.	7ª it.
x_1	1425.5	1423.7	1421.6	1419.7	1418.1	1417.3
x_2	1424.0	1422.2	1421.2	1422.2	1420.1	1418.4
x_3	1421.7	1419.8	<u>1417.8</u>	◇	◇	◇
x_4	1425.7	1423.8	1421.8	1419.8	1418.9	1417.4
x_5	1421.7	<u>1419.8</u>	◇	◇	◇	◇
x_6	1519.5	1517.5	1525.6	1552.6	1551.0	1549.0
x_7	1423.9	1422.0	1419.9	1417.9	1416.6	1415.8
x_8	1424.9	1423.1	1421.4	1419.6	1418.4	1418.8
x_9	<u>1421.7</u>	◇	◇	◇	◇	◇
x_{10}	1422.4	1420.4	1418.4	<u>1416.5</u>	◇	◇
x_{11}	1422.7	1420.7	1418.7	1416.8	<u>1415.4</u>	◇
x_{12}	1427.3	1425.4	1423.3	1421.4	1419.9	1421.1
x_{13}	1430.9	1429.0	1427.0	1425.1	1423.0	1421.9
DIC	1421.7	1419.8	1417.8	1416.5	1415.4	◇
decr.	2.1	1.9	2.0	1.3	1.1	◇

¹Na primeira iteração, em que é ajustado um modelo com as 13 covariáveis, tem-se $\text{DIC} = 1423.8$.

A interpretação das tabelas que sumarizam os resultados obtidos para o método BVE com as outras medidas de discrepância é idêntica.

As covariáveis excluídas do modelo pelo método de pesquisa BVE quando se utiliza o $\widehat{\text{DIC}}$, o LLP e o SlnCPO são: x_9 , x_5 , x_3 , x_{10} e x_{11} ; quando se utiliza o método BVE via $\widehat{\text{BIC}}$, além das covariáveis excluídas pelos outros métodos, são também excluídas x_7 , x_4 , x_8 e x_1 .

Tabela 3.14: Resultados do método BVE via $\widehat{\text{BIC}}$, para o exemplo da massa gorda¹.

cov.	2 ^a it.	3 ^a it.	4 ^a it.	5 ^a it.	6 ^a it.	7 ^a it.	8 ^a it.	9 ^a it.	10 ^a it.	11 ^a it.
x_1	1489	1482	1476	1469	1463	1458	1454	1450	<u>1445</u>	◇
x_2	1487	1481	1475	1472	1465	1459	1463	1464	1457	1468
x_3	1485	1479	<u>1472</u>	◇	◇	◇	◇	◇	◇	◇
x_4	1489	1483	1476	1470	1464	1458	<u>1453</u>	◇	◇	◇
x_5	1485	<u>1479</u>	◇	◇	◇	◇	◇	◇	◇	◇
x_6	1583	1576	1580	1602	1596	1590	1586	1580	1586	1645
x_7	1487	1481	1474	1468	1462	<u>1456</u>	◇	◇	◇	◇
x_8	1488	1482	1476	1469	1464	1460	1453	<u>1450</u>	◇	◇
x_9	<u>1485</u>	◇	◇	◇	◇	◇	◇	◇	◇	◇
x_{10}	1486	1479	1473	<u>1466</u>	◇	◇	◇	◇	◇	◇
x_{11}	1486	1480	1473	1467	<u>1461</u>	◇	◇	◇	◇	◇
x_{12}	1491	1484	1478	1471	1465	1462	1459	1454	1451	1446
x_{13}	1494	1488	1481	1475	1468	1463	1459	1458	1455	1449
$\widehat{\text{BIC}}$	1485	1479	1472	1466	1461	1456	1453	1450	1445	◇
decr.	7	6	7	6	5	5	3	3	5	◇

¹Na primeira iteração, em que é ajustado um modelo com as 13 covariáveis, tem-se $\widehat{\text{BIC}} = 1492$.

Tabela 3.15: Resultados do método BVE via LLP, para o exemplo da massa gorda¹.

cov.	2 ^a it.	3 ^a it.	4 ^a it.	5 ^a it.	6 ^a it.	7 ^a it.
x_1	1411.4	1410.6	1409.6	1408.6	1408.0	1408.4
x_2	1410.0	1409.2	1409.2	1411.2	1410.2	1409.4
x_3	1407.8	1406.8	<u>1405.8</u>	◇	◇	◇
x_4	1411.8	1410.8	1409.8	1408.8	1408.8	1408.4
x_5	1407.8	<u>1406.8</u>	◇	◇	◇	◇
x_6	1505.4	1504.4	1513.6	1541.6	1541.0	1540.0
x_7	1409.8	1409.0	1407.8	1406.8	1406.6	1406.8
x_8	1410.8	1410.0	1409.4	1408.6	1408.4	1409.8
x_9	<u>1407.8</u>	◇	◇	◇	◇	◇
x_{10}	1408.4	1407.4	1406.4	<u>1405.4</u>	◇	◇
x_{11}	1408.8	1407.8	1406.6	1405.8	<u>1405.4</u>	◇
x_{12}	1413.4	1412.4	1411.2	1410.4	1409.8	1412.2
x_{13}	1416.8	1416.0	1415.0	1414.0	1413.0	1412.9
LLP	1407.8	1406.8	1405.8	1405.4	1405.4	◇
decr.	1.0	1.0	1.0	0.4	0.0	◇

¹Na primeira iteração, em que é ajustado um modelo com as 13 covariáveis, tem-se $\text{LLP} = 1408.8$.

Tabela 3.16: Resultados do método BVE via SlnCPO, para o exemplo da massa gorda¹.

cov.	2 ^a it.	3 ^a it.	4 ^a it.	5 ^a it.	6 ^a it.	7 ^a it.
x_1	-713.9	-712.6	-711.5	-710.5	-709.4	-708.8
x_2	-713.3	-713.3	-712.8	-712.0	-711.5	-709.7
x_3	<u>-711.7</u>	◇	◇	◇	◇	◇
x_4	-714.3	-713.0	-711.8	-710.6	-710.1	-709.1
x_5	-711.8	<u>-710.6</u>	◇	◇	◇	◇
x_6	-762.7	-764.7	-778.6	-778.4	-778.7	-776.8
x_7	-713.4	-712.1	-710.8	-709.7	-708.7	-708.3
x_8	-714.0	-713.1	-711.5	-710.6	-709.6	-709.8
x_9	-712.2	-711.0	<u>-709.5</u>	◇	◇	◇
x_{10}	-712.2	-711.2	-709.6	<u>-708.5</u>	◇	◇
x_{11}	-713.0	-711.7	-710.1	-709.1	<u>-708.0</u>	◇
x_{12}	-715.8	-714.3	-712.7	-711.7	-710.6	-711.1
x_{13}	-716.7	-715.4	-714.1	-713.1	-711.5	-710.9
SlnCPO	-710.0	-710.6	-709.5	-708.5	-708.0	◇
acr.	1.4	0.6	1.1	1.0	0.5	◇

¹Na primeira iteração, em que é ajustado um modelo com as 13 covariáveis, tem-se SlnCPO = -713.1.

3.5.4 Método Gibbs de selecção de covariáveis

A aplicação do método GVS ao conjunto das 13 covariáveis impossibilita a monitorização das probabilidades *a posteriori* dos 8192 possíveis modelos. Assim, opta-se por monitorizar apenas as probabilidades *a posteriori* de inclusão das covariáveis no modelo e usar os resultados obtidos para: (i) determinar o modelo de probabilidade *a posteriori* mediana; (ii) seleccionar um subconjunto de 8 covariáveis ao qual se volta a aplicar o método GVS.

Considera-se que não existe informação *a priori* sobre a plausibilidade relativa dos diferentes modelos, o que pode ser expresso através de probabilidades *a priori* iguais a $1/8192$ para cada modelo identificado pelo vector $\gamma = (\gamma_1, \dots, \gamma_{13})$. Assim, os indicadores γ_j , para $j = 1, \dots, 13$, são considerados independentes com distribuição *a priori* Bernoulli com probabilidade de sucesso $1/2$. Considera-se que a distribuição *a priori* para b_0 é normal de valor médio nulo e variância 10000, que a distribuição *a priori* de b_j , condicional a γ_j , é normal com valor médio nulo e variância $0.01^{1-\gamma_j} \times 1000$, para $j = 1, \dots, 13$ (Ntzoufras, 2002) e que a distribuição *a priori* de τ é gama com parâmetro de forma e inverso do parâmetro de escala iguais a 0.001.

Nas duas aplicações do método GVS, as estimativas são obtidas em 10000 iterações do método de amostragem, após um período de aquecimento de 1000 iterações. Como a correlação entre as observações geradas consecutivamente é elevada, é necessário considerar um espaçamento de amostragem de 10 iterações. Por análise dos gráficos das funções de autocorrelação empírica dos valores simulados, o espaçamento de 10 iterações permite, sem aumentar demasiado o número total de iterações, considerar válida a suposição de independência entre os valores simulados.

O código do WinBUGS, referente à aplicação do método GVS ao conjunto de 13

covariáveis, é apresentado no anexo D.1.4 e os resultados obtidos são apresentados na tabela 3.17. O modelo com probabilidade *a posteriori* mediana é constituído por x_2 e x_6 , pois apenas estas covariáveis têm probabilidade *a posteriori* de inclusão da covariável no modelo superior a $1/2$. Selecciona-se, para uma nova aplicação do método GVS, as 8 covariáveis que apresentam estimativas da probabilidade *a posteriori* de inclusão da covariável no modelo mais elevadas, isto é, $x_6, x_2, x_{13}, x_{12}, x_4, x_{11}, x_8$ e x_7 .

Tabela 3.17: Estimativas das probabilidades *a posteriori* de inclusão das covariáveis no modelo, para o exemplo da massa gorda.

ind. covar.	média	d.p.	erro MC	2.5%	mediana	97.5%
$\gamma_1 = 1$	0.0150	0.1216	0.001275	0.0	0.0	0.0
$\gamma_2 = 1$	0.9757	0.1540	0.003465	1.0	1.0	1.0
$\gamma_3 = 1$	0.0153	0.1227	0.001314	0.0	0.0	0.0
$\gamma_4 = 1$	0.0812	0.2731	0.002801	0.0	0.0	1.0
$\gamma_5 = 1$	0.0238	0.1524	0.001619	0.0	0.0	0.0
$\gamma_6 = 1$	1.0000	0.0000	1.0E-12	1.0	1.0	1.0
$\gamma_7 = 1$	0.0548	0.2276	0.003716	0.0	0.0	1.0
$\gamma_8 = 1$	0.0713	0.2573	0.002766	0.0	0.0	1.0
$\gamma_9 = 1$	0.0172	0.1300	0.001296	0.0	0.0	0.0
$\gamma_{10} = 1$	0.0128	0.1124	0.001181	0.0	0.0	0.0
$\gamma_{11} = 1$	0.0749	0.2632	0.002603	0.0	0.0	1.0
$\gamma_{12} = 1$	0.1262	0.3321	0.003142	0.0	0.0	1.0
$\gamma_{13} = 1$	0.3441	0.4751	0.004847	0.0	0.0	1.0

Aplica-se o método GVS ao subconjunto de 8 covariáveis seleccionadas. Considera-se ainda que não existe informação *a priori* sobre a plausibilidade relativa dos diferentes modelos, atribuindo probabilidades *a priori* iguais a $1/256$ a cada modelo identificado por $\gamma = (\gamma_1, \dots, \gamma_8)$, o que equivale a considerar que os indicadores $\gamma_j, j = 1, \dots, 8$, são independentes com distribuição *a priori* Bernoulli com probabilidade de sucesso $1/2$. As distribuições *a priori* dos restantes parâmetros do modelo são iguais às consideradas na primeira aplicação do método GVS. Neste caso, além da monitorização das probabilidades *a posteriori* de inclusão das covariáveis no modelo, pode também monitorizar-se as probabilidades *a posteriori* dos modelos, pois o número total de modelos é agora 256, bastante inferior ao número anterior. O código do WinBUGS, referente à 2ª aplicação do método GVS, é apresentado no anexo D.1.4 e os resultados obtidos estão sumariados nas tabelas 3.18 e 3.19.

As estimativas das probabilidades *a posteriori* de inclusão das covariáveis no modelo, quando se considera o subconjunto de 8 covariáveis, são muito idênticas às obtidas quando se considera todas as covariáveis. As covariáveis x_6 e x_2 continuam a ter estimativas das probabilidades *a posteriori* de inclusão das covariáveis no modelo próximas de 1, logo seguidas por x_{13} e x_{12} com estimativas das probabilidades *a posteriori* de inclusão das covariáveis no modelo já mais pequenas. Estas covariáveis são as que têm uma presença mais constante nos modelos com probabilidade *a posteriori* mais elevada.

O modelo de regressão linear normal com maior probabilidade *a posteriori* é constituído pelas covariáveis x_2 e x_6 e tem probabilidade *a posteriori* aproximadamente igual a 0.46. O modelo que inclui as covariáveis x_2, x_6 e x_{13} tem a segunda maior probabilidade *a*

posteriori, com valor próximo de 0.19. O modelo constituído pelas covariáveis x_2 , x_6 , x_{12} e x_{13} corresponde ao terceiro modelo com maior probabilidade *a posteriori*, com um valor de 0.07. A soma das probabilidades *a posteriori* destes 3 modelos é, aproximadamente, 0.72. Os 7 modelos seguintes, que prefazem um total de 0.20 para a soma das probabilidades *a posteriori*, têm em comum o facto de todos eles incluírem as covariáveis x_2 e x_6 . Refira-se, ainda, que a soma das probabilidades dos 20 modelos com maior probabilidade *a posteriori* é aproximadamente igual a 0.98. A existência de muitos modelos com probabilidades *a posteriori* relevantes pode ser devida às correlações elevadas entre as covariáveis e ao facto da dimensão da amostra não ser suficientemente grande quando comparada com o número de covariáveis.

Tabela 3.18: Estimativas das probabilidades *a posteriori* de inclusão das covariáveis no modelo, para o exemplo da massa gorda, quando se considera um subconjunto de 8 covariáveis.

ind. covar.	média	d.p.	erro MC	2.5%	mediana	97.5%
$\gamma_2 = 1$	0.9854	0.1199	0.002359	1.0	1.0	1.0
$\gamma_4 = 1$	0.0828	0.2756	0.002582	0.0	0.0	1.0
$\gamma_6 = 1$	1.0000	0.0000	1.0E-12	1.0	1.0	1.0
$\gamma_7 = 1$	0.0464	0.2103	0.002855	0.0	0.0	1.0
$\gamma_8 = 1$	0.0744	0.2624	0.002630	0.0	0.0	1.0
$\gamma_{11} = 1$	0.0721	0.2587	0.002422	0.0	0.0	1.0
$\gamma_{12} = 1$	0.1359	0.3427	0.003514	0.0	0.0	1.0
$\gamma_{13} = 1$	0.3381	0.4731	0.005008	0.0	0.0	1.0

Tabela 3.19: Dez modelos com maior probabilidade *a posteriori*, para o exemplo da massa gorda, quando se considera um subconjunto de 8 covariáveis.

covariáveis								probabilidades <i>a posteriori</i>		
x_2	x_4	x_6	x_7	x_8	x_{11}	x_{12}	x_{13}	média	d.p.	erro MC
×		×						0.4556	0.4980	4.71E-03
×		×					×	0.1946	0.3959	3.85E-03
×		×				×	×	0.0700	0.2551	2.55E-03
×		×		×				0.0474	0.2125	2.17E-03
×	×	×						0.0421	0.2008	1.96E-03
×		×				×		0.0351	0.1840	1.69E-03
×		×			×			0.0290	0.1678	1.51E-03
×		×			×		×	0.0266	0.1609	1.61E-03
×	×	×				×		0.0128	0.1124	9.86E-04
×		×	×					0.0115	0.1066	1.02E-03

3.5.5 Comparação de resultados

Os modelos seleccionados por cada um dos métodos e para cada uma das medidas de discrepância são apresentados na tabela 3.20. Os valores das medidas de discrepância e as probabilidades *a posteriori* de alguns destes modelos estão sumariados na tabela 3.21. Como os modelos m_1 e m_2 incluem a covariável x_1 , que não pertence ao subconjunto de

covariáveis seleccionadas para a 2ª aplicação do método GVS, não foram estimadas as probabilidades *a posteriori* destes modelos.

O modelo m_1 parece ser o melhor modelo, uma vez que apresenta o menor valor do DIC e do LLP e o maior valor de SlnCPO. Este modelo foi seleccionado pelo método SIMD e pelo método BVE quando se considera as medidas de discrepância DIC, LLP e SlnCPO. O modelo m_2 , além de ser menos parcimonioso que m_1 , tem valores das medidas de discrepância mais desfavoráveis. Este modelo, seleccionado pelo método FVS quando se considera as medidas DIC, LLP e SlnCPO, não parece ser a melhor opção.

Os dois modelos com maior probabilidade *a posteriori* obtidos pelo método GVS, m_4 e m_5 , têm os valores das medidas de discrepância bastante mais desfavoráveis que o modelo m_1 . O modelo m_3 , com a terceira maior probabilidade *a posteriori*, apresenta já um valor da probabilidade muito pequeno. Este modelo, também seleccionado pela abordagem clássica já referida e pelo método BVE via $\widehat{\text{BIC}}$, apresenta valores das medidas de discrepância próximos dos valores obtidos para m_1 .

Tabela 3.20: Covariáveis incluídas nos modelos de regressão linear seleccionados por cada um dos métodos, no caso do exemplo da massa gorda.

método	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	modelo
SIMD	×	×		×		×	×	×				×	×	m_1
FVS - DIC, LLP, SlnCPO	×	×		×		×	×	×			×	×	×	m_2
FVS - $\widehat{\text{BIC}}$		×				×								m_4
BVE - DIC, LLP, SlnCPO	×	×		×		×	×	×				×	×	m_1
BVE - $\widehat{\text{BIC}}$		×				×						×	×	m_3
GVS		×				×								m_4
		×				×							×	m_5
		×				×						×	×	m_3
Análise clássica ¹		×			×						×	×	m_3	

¹O modelo final obtido pela selecção progressiva, pela eliminação regressiva e pelo método passo a passo, através da realização de testes F parciais, é o mesmo.

Tabela 3.21: Valores das medidas de discrepância dos modelos de regressão linear m_1 a m_5 , para o exemplo da massa gorda.

modelo	DIC	$\widehat{\text{BIC}}$	LLP	SlnCPO	$\text{Pr}\{m_i \mathcal{D}\}$
m_1	1415.4	1461	1405.4	-708.0	—
m_2	1416.5	1473	1405.4	-708.5	—
m_3	1417.8	1445	1411.9	-709.4	0.0700
m_4	1428.1	1446	1424.1	-714.5	0.4556
m_5	1423.0	1446	1418.0	-712.2	0.1946

Os cinco modelos seleccionados são comparados através do método de Carlin e Chib. Compara-se, por um lado, o modelo m_1 com o modelo m_2 e, por outro, os modelos m_3 , m_4 e m_5 . O modo de implementação deste método no WinBUGS está sumariado no anexo D.1.5. Todos os resultados do método de Carlin e Chib são obtidos em 10000 iterações do método de amostragem, com um espaçamento de 10 iterações, e após um período de aquecimento de 5000 iterações.

Começa por utilizar-se o método de Carlin e Chib para comparar os modelos m_1 e m_2 . Para garantir que ambos os modelos são frequentemente visitados, considera-se $\Pr\{m_1\} = 0.01$ e $\Pr\{m_2\} = 0.99$, obtendo-se $\Pr\{m_1|\mathcal{D}\} = 0.5664$ e $\Pr\{m_2|\mathcal{D}\} = 0.4336$; assim o factor Bayes a favor de m_1 e contra m_2 é $\text{BF}_{12} = 129.3$, o que se traduz numa evidência forte a favor de m_1 .

Aplica-se também o método de Carlin e Chib para comparar os modelos m_3 , m_4 e m_5 . Considerando $\Pr\{m_3\} = 0.90$, $\Pr\{m_4\} = 0.01$ e $\Pr\{m_5\} = 0.09$, obtém-se $\Pr\{m_3|\mathcal{D}\} = 0.3804$, $\Pr\{m_4|\mathcal{D}\} = 0.2826$ e $\Pr\{m_5|\mathcal{D}\} = 0.3370$; assim $\text{BF}_{45} = 7.5$, o que se traduz numa evidência a favor de m_4 quando comparado com m_5 , e $\text{BF}_{43} = 66.9$, o que se traduz numa evidência forte a favor de m_4 quando comparado com m_3 .

O método de Carlin e Chib pode ser novamente utilizado para comparar os modelos m_1 e m_4 . Para garantir que ambos os modelos são frequentemente visitados, considera-se $\Pr\{m_1\} = 0.9999999$ e $\Pr\{m_4\} = 0.0000001$, obtendo-se $\Pr\{m_1|\mathcal{D}\} = 0.0246$ e $\Pr\{m_4|\mathcal{D}\} = 0.9754$; assim, $\text{BF}_{41} = 3.97 \times 10^8$, o que se traduz numa evidência muito forte a favor do modelo m_4 .

3.6 Exemplo de regressão gama

3.6.1 Introdução

Considere-se o modelo de regressão gama

$$Y_i = \exp\{b_0 + b_1x_{i1} + \dots + b_px_{ip}\} \times \epsilon_i, \quad i = 1, \dots, n,$$

onde Y_i é uma variável resposta, b_0, b_1, \dots, b_p são parâmetros desconhecidos, x_{i1}, \dots, x_{ip} são observações de p covariáveis e ϵ_i é o erro aleatório, que tem distribuição gama com valor médio 1 e variância $1/\nu$, $\epsilon_i \sim \text{Ga}(\nu, \nu)$. Pondo $b = (b_0, b_1, \dots, b_p)$ e $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$, então

$$Y_i | b, \nu, x_i \sim \text{Ga}(\nu, \nu/\mu_i), \quad i = 1, \dots, n,$$

cuja função densidade de probabilidade é dada por

$$f(y_i | b, \nu, x_i) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu_i}\right)^\nu \exp\left\{-\frac{\nu}{\mu_i} y_i\right\} y_i^{\nu-1}, \quad y_i > 0 \text{ e } \nu > 0,$$

com valor médio $\mu_i = \exp\{b_0 + b_1x_{i1} + \dots + b_px_{ip}\}$ e variância μ_i^2/ν . Deste modo, $\log \mu_i = b_0 + b_1x_{i1} + \dots + b_px_{ip}$ e, na terminologia dos Modelos Lineares Generalizados, diz-se que a função de ligação do modelo de regressão gama é o logaritmo.

Devido à impossibilidade de usar distribuições *a priori* impróprias no WinBUGS, em vez da distribuição *a priori* não informativa para o vector de parâmetros (b, ν) , utiliza-se distribuições *a priori* difusas para os parâmetros do modelo. Nomeadamente, nas aplicações dos métodos de selecção de covariáveis apresentadas nas secções 3.6.3 e 3.6.4, considera-se que a distribuição *a priori* de cada um dos coeficientes de regressão é $N(0, 100^2)$ e que $\nu \sim \text{Ga}(0.001, 0.001)$.

3.6.2 Dados da homocisteinemia

É do conhecimento geral que a hipertensão arterial, a hipercolesterolemia e os hábitos tabágicos são factores de risco para a ocorrência de enfartes ou acidentes vasculares cerebrais. Palma dos Reis (1994) salienta que, além dos factores de risco clássicos, a ocorrência precoce de doenças vasculares pode ser altamente influenciada pela hiperhomocisteinemia, que se traduz na ocorrência de valores elevados da homocisteinemia basal (HBP) e da homocisteinemia após sobrecarga com metionina oral (HSP). De modo a determinar a relação destas variáveis com os outros factores de risco das doenças vasculares, Teles (1995) utiliza a metodologia dos Modelos Lineares Generalizados e ajusta modelos de regressão lognormal e gama à variável HBP e de regressão normal à variável HSP.

O conjunto de dados original, recolhido por Palma dos Reis (1994) e listado em Teles (1995, anexo A), refere-se a 145 indivíduos, 121 homens e 24 mulheres, com idades compreendidas entre os 26 e os 56 anos, que sofreram enfarte ou acidente vascular cerebral. Para cada indivíduo foram registados os valores de 30 variáveis clínico-laboratoriais, 8 qualitativas e 22 quantitativas, cuja descrição detalhada pode ser vista em Teles (1995).

Tabela 3.22: Estatísticas sumárias das variáveis do exemplo da homocisteinemia.

variáveis qualitativas		frequência absoluta			
x_1	sexo (masc.=1 e fem.=0)	$n_1 = 106$	$n_0 = 20$		
x_2	fumar1 (fumador=1 e c.c.=0)	$n_1 = 23$	$n_0 = 103$		
x_3	fumar2 (ex-fumador=1 e c.c.=0)	$n_1 = 58$	$n_0 = 68$		
x_4	história familiar (sim=1 e não=0)	$n_1 = 72$	$n_0 = 54$		
variáveis quantitativas		média	d.p.	mín.	máx.
x_5	glucose	105.79	45.18	63	397
x_6	ureia	36.60	14.54	17	167
x_7	creatinina	1.10	1.00	0.6	12.1
x_8	gama-glutamyl-transpeptidase (GGT)	46.06	38.26	10	193
x_9	fosfatase alcalina	81.71	28.26	29	249
x_{10}	colesterol das LDL	159.90	38.91	78.0	274.4
x_{11}	colesterol das HDL	46.93	13.42	19	104
x_{12}	pressão arterial sistólica	136.90	17.22	106	190
x_{13}	pressão arterial diastólica	87.22	10.07	60	110
y	homocisteinemia basal (HBP)	10.31	4.66	5.40	38.47

Exemplifica-se a selecção bayesiana de covariáveis em modelos de regressão não normal, através do ajustamento de modelos de regressão gama à variável HBP (variável resposta). Para facilitar a ilustração da aplicação dos métodos SIMD, FVE, BVE e GVS, são consideradas apenas 12 covariáveis para o conjunto das possíveis variáveis explicativas, tal como foi efectuado em Paulino *et al.* (2003). Das covariáveis escolhidas, correspondentes às que em estudos anteriores — Teles (1995) e Amaral Turkman e Silva (2000) — foram consideradas mais relevantes, 9 são quantitativas e 3 qualitativas. Uma das covariáveis qualitativas é o fumar que tem 3 categorias, nomeadamente, fumador, ex-fumador e não fumador; esta covariável deu origem a duas variáveis *dummy*, fumar1 e fumar2. Do conjunto inicial de 145 indivíduos, 19 foram excluídos pelo facto de terem valores omissos

de pelo menos uma das covariáveis. O conjunto de dados utilizado, referente às observações dos restantes 126 indivíduos, encontra-se listado na tabela C.2 do anexo C; a definição das variáveis e algumas medidas sumárias são apresentadas na tabela 3.22.

Da análise exploratória efectuada, destaca-se a existência de observações discordantes para algumas covariáveis e a existência de correlações lineares elevadas apenas entre dois pares de covariáveis, nomeadamente entre a creatinina e a ureia (0.827) e entre a pressão arterial sistólica e a diastólica (0.804).

3.6.3 Métodos de selecção via medidas de discrepância

A aplicação dos três métodos de selecção de covariáveis via medidas de discrepância ao exemplo da homocisteinemia foi efectuada com quatro medidas de discrepância: DIC, \widehat{BIC} , LLP e $SlnCPO$. Para qualquer um dos métodos e qualquer uma das medidas de discrepância, o ajustamento de cada um dos modelos é efectuada: (i) considerando distribuições *a priori* difusas para todos os parâmetros do modelo, nomeadamente, b_0, b_1, \dots, b_p independentes com distribuição normal de valor médio nulo e variância 10000, e ν com distribuição gama com parâmetro de forma e inverso do parâmetro de escala iguais a 0.001; (ii) com base em 5000 iterações do método de amostragem, com um espaçamento de 5 iterações e após um período de aquecimento de 4000 iterações.

Devido ao elevado número de modelos ajustados, a monitorização da convergência é apenas efectuada por análise dos gráficos das funções de autocorrelação empírica e das séries temporais dos valores simulados. Por análise dos gráficos das funções de autocorrelação empírica dos valores simulados, o espaçamento de 5 iterações permite, sem aumentar demasiado o número total de iterações, considerar válida a suposição de independência entre os valores simulados.

Método SIMD

Ajusta-se um modelo de regressão gama com 13 covariáveis e calcula-se os valores das medidas de discrepância. O código do WinBUGS, para a aplicação da 1ª iteração do método SIMD aos dados da homocisteinemia, é apresentado no anexo D.2.1. Na figura 3.6 são apresentados os gráficos das densidades marginais *a posteriori* dos coeficientes de regressão. O *output* com as estimativas dos parâmetros do modelo de regressão gama é apresentado na tabela 3.23. Os valores das medidas de discrepância constam da tabela 3.24. Da análise da figura 3.6 e da tabela 3.23, destaca-se que:

- os intervalos de credibilidade a 95% dos parâmetros de regressão $b_1, b_2, b_4, b_7, b_8, b_9, b_{10}$ e b_{13} são os únicos que não contêm o zero. As covariáveis $x_1, x_2, x_4, x_7, x_8, x_9, x_{10}$ e x_{13} não são, portanto, candidatas a sair do modelo;
- embora o intervalo de credibilidade a 95% do coeficiente de regressão b_3 contenha o zero, a covariável x_3 (fumar2) não é excluída do modelo, pois a outra variável *dummy* que define a covariável fumar não é candidata a sair do modelo;
- dos restantes parâmetros de regressão, apenas os intervalos de credibilidade a 95% de b_5 e b_6 são aproximadamente centrados em zero. Entre as covariáveis x_5 e x_6 ,

opta-se por excluir do modelo a covariável x_5 , uma vez que a densidade marginal *a posteriori* de b_5 é a que se encontra mais concentrada em torno de zero.

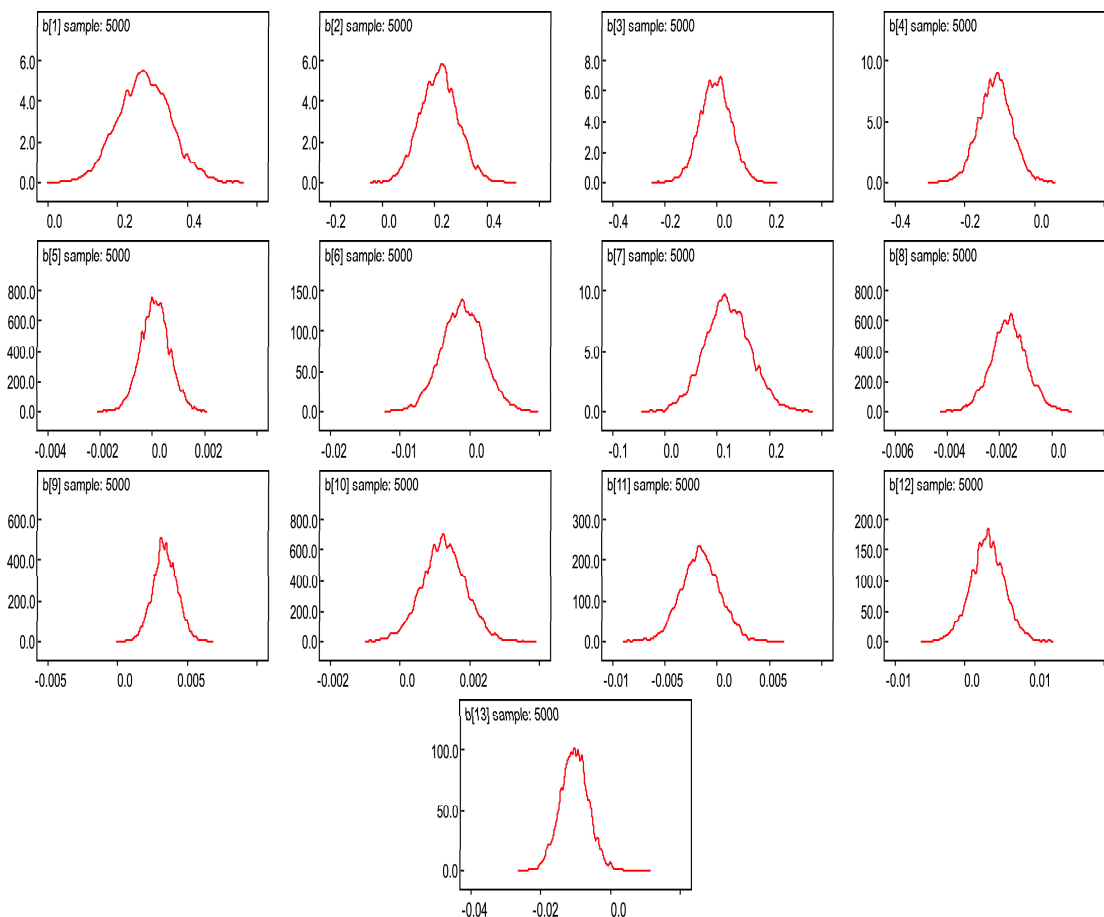


Figura 3.6: Gráficos das densidades marginais *a posteriori* dos coeficientes do modelo de regressão gama com 13 covariáveis, para o exemplo da homocisteinemia.

Tabela 3.23: *Output* do WinBUGS com as estimativas dos parâmetros do modelo de regressão gama com 13 covariáveis, para o exemplo da homocisteinemia.

node	mean	sd	MC error	2.50%	median	97.50%	start	sample
b0	2.116	0.3108	0.004113	1.518	2.117	2.732	4001	5000
b[1]	0.2769	0.07467	0.001101	0.1342	0.2763	0.4267	4001	5000
b[2]	0.2163	0.07158	0.001003	0.08042	0.2171	0.361	4001	5000
b[3]	-0.007219	0.05822	7.74E-04	-0.1217	-0.006765	0.1058	4001	5000
b[4]	-0.1122	0.04669	7.12E-04	-0.2021	-0.1114	-0.02038	4001	5000
b[5]	1.12E-04	5.49E-04	7.57E-06	-9.59E-04	1.08E-04	0.001206	4001	5000
b[6]	-0.001017	0.002956	5.08E-05	-0.006816	-0.001011	0.004773	4001	5000
b[7]	0.1202	0.04357	7.02E-04	0.03456	0.1189	0.2063	4001	5000
b[8]	-0.001639	6.73E-04	9.97E-06	-0.002951	-0.001637	-3.02E-04	4001	5000
b[9]	0.003411	8.68E-04	1.31E-05	0.001733	0.003388	0.005141	4001	5000
b[10]	0.001259	6.18E-04	9.12E-06	3.69E-05	0.001259	0.002451	4001	5000
b[11]	-0.001583	0.001843	2.67E-05	-0.005107	-0.001597	0.002051	4001	5000
b[12]	0.003287	0.002403	3.66E-05	-0.001487	0.003292	0.008086	4001	5000
b[13]	-0.01032	0.003992	6.62E-05	-0.01816	-0.01035	-0.002452	4001	5000
nu	15.63	2.079	0.03209	11.93	15.51	20.01	4001	5000

Tabela 3.24: Resultados do método SIMD, para o exemplo da homocisteinemia.

iteração	1 ^a it.	2 ^a it.	3 ^a it.	4 ^a it.	5 ^a it.
covariáveis	todas	sai x_5	sai x_6	sai x_{11}	sai x_{12}
DIC	603.2	601.3	599.1	597.9	597.9
$\widehat{\text{BIC}}$	660.6	654.8	648.8	643.9	640.0
LLP	-294.0	-293.6	-293.0	-293.0	-293.4
SlnCPO	-285.6	-286.1	-283.8	-284.5	-283.7

Na 2^a iteração ajusta-se um modelo de regressão gama com 12 covariáveis, uma vez que x_5 foi excluída do modelo, e calcula-se os valores das medidas de discrepância. Comparando a 1^a com a 2^a iteração, verifica-se a existência de valores mais favoráveis para as medidas de discrepância na 2^a iteração, o que quer dizer que o método SIMD prossegue. Analisando as estimativas dos coeficientes do modelo de regressão gama ajustado na 2^a iteração e os respectivos gráficos das densidades marginais *a posteriori*, conclui-se que x_6 é a próxima covariável a sair do modelo. Na 3^a iteração é ajustado um modelo de regressão gama com 11 covariáveis, uma vez que x_5 e x_6 foram excluídas do modelo, e calcula-se os valores das medidas de discrepância deste modelo. Proceda-se deste modo até que não exista mais nenhuma covariável candidata a sair do modelo ou até que a exclusão de uma covariável origine valores mais desfavoráveis para as medidas de discrepância. As variáveis excluídas em cada iteração, assim como os valores obtidos para as medidas de discrepância, são apresentados na tabela 3.24.

Os gráficos das densidades marginais *a posteriori* dos coeficientes do modelo de regressão gama ajustado na 5^a iteração são apresentados na figura 3.7. O *output* do WinBUGS com as estimativas dos parâmetros do modelo de regressão gama ajustado na 5^a iteração é apresentado na tabela 3.25. Da análise da figura 3.7 e da tabela 3.25, conclui-se que apenas a covariável x_3 (fumar2) é candidata a sair do modelo; no entanto, esta covariável não é excluída, uma vez que a outra variável *dummy* que define a covariável fumar não é candidata a sair do modelo, e o método SIMD pára.

Tabela 3.25: *Output* do WinBUGS com as estimativas dos parâmetros do modelo de regressão gama com 9 covariáveis, para o exemplo da homocisteinemia.

node	mean	sd	MC error	2.50%	median	97.50%	start	sample
b0	2.077	0.255	0.003221	1.582	2.078	2.567	4001	5000
b[1]	0.2732	0.07065	9.97E-04	0.137	0.2741	0.4098	4001	5000
b[2]	0.2285	0.06837	1.17E-03	0.09351	0.2277	0.3633	4001	5000
b[3]	0.002827	0.0549	8.39E-04	-0.1051	0.002533	0.1094	4001	5000
b[4]	-0.1108	0.04701	7.38E-04	-0.2033	-0.1111	-0.01856	4001	5000
b[7]	0.1125	0.02484	3.87E-04	0.06603	0.1114	0.164	4001	5000
b[8]	-0.001762	6.44E-04	9.24E-06	-0.002969	-0.001768	-4.80E-04	4001	5000
b[9]	0.003448	8.65E-04	1.15E-05	0.001802	0.003443	0.005175	4001	5000
b[10]	0.001211	6.12E-04	9.26E-06	1.14E-05	0.00121	0.002404	4001	5000
b[13]	-0.005705	0.002351	2.97E-05	-0.0105	-0.005673	-0.001052	4001	5000
nu	15.81	2.084	0.02987	12.04	15.75	20.2	4001	5000

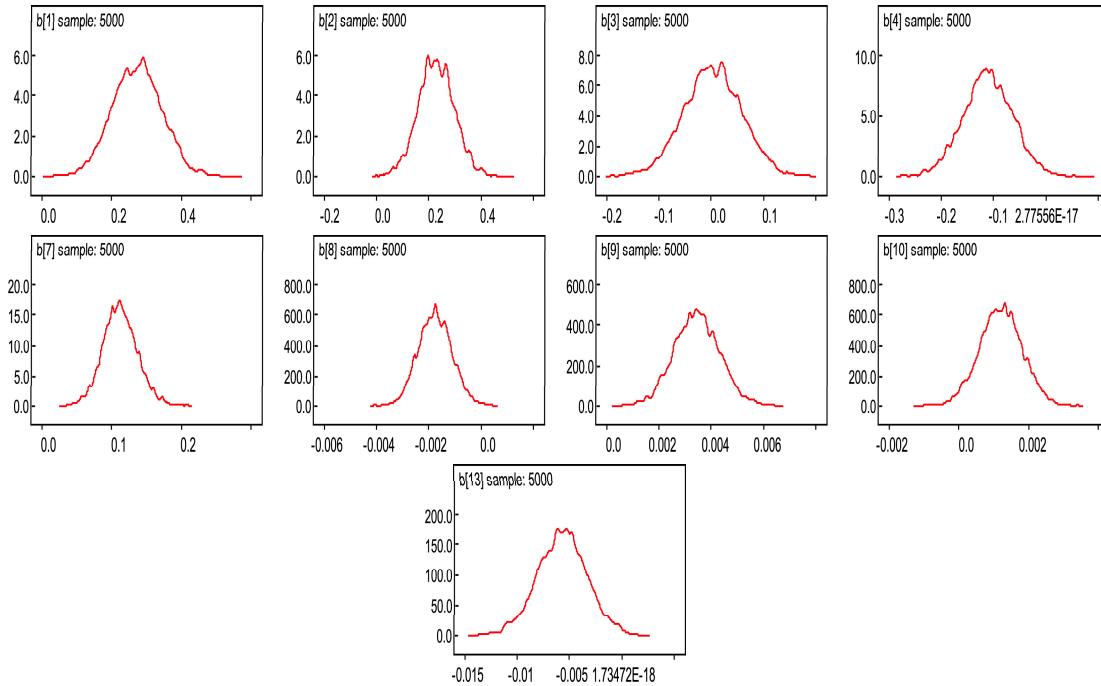


Figura 3.7: Gráficos das densidades marginais *a posteriori* dos coeficientes do modelo de regressão gama com 9 covariáveis, para o exemplo da homocisteinemia.

Método FVS

O modo de implementação do método FVS no WinBUGS está detalhada no anexo D.2.2. Os resultados obtidos com o método FVS via DIC estão sumariados na tabela 3.26. Na 1ª iteração é ajustado um modelo de regressão gama só com a constante, isto é, o modelo $HBP_i = \exp\{b_0\} \times \epsilon_i$, com $\epsilon_i \sim \text{Ga}(\nu, \nu)$, para o qual se tem $\text{DIC} = 667.2$.

Na 2ª iteração são ajustados 13 modelos de regressão gama, cada um deles com uma das covariáveis. Selecciona-se x_7 para entrar no modelo pois é a covariável que, ao ser incluída, origina o maior decréscimo no valor do DIC, mais concretamente, produz um decréscimo de 18.2.

Em cada iteração o valor sublinhado corresponde ao valor mais favorável para a medida de discrepância, indicando qual a covariável a ser incluída no modelo. O símbolo \diamond a partir da t -ésima iteração indica que a correspondente covariável foi incluída no modelo na iteração $t - 1$. Na penúltima e última linhas da tabela 3.26 são apresentados o mínimo do DIC obtido nessa iteração e o decréscimo entre iterações consecutivas, respectivamente.

Na 3ª iteração ajusta-se 12 modelos de regressão gama, cada um deles com as covariáveis x_7 e x_k , com $k = 1, \dots, 13$ e $k \neq 7$. A maior redução do DIC é obtida com a inclusão de x_2 (fumar1) e, como se trata de uma das variáveis *dummy* que permite definir a variável qualitativa fumar, na iteração seguinte entra obrigatoriamente x_3 (fumar2). Assim, na 4ª iteração só é calculado o valor do DIC relativo à inclusão de x_3 no modelo, sendo as células relativas às outras covariáveis que não estão no modelo indicadas com \star . O valor do DIC após a inclusão de x_2 e x_3 é 628.4 o que corresponde a uma redução de 20.6 no valor desta medida de discrepância.

Procede-se deste modo até à 10ª iteração em que já estão no modelo 9 covariáveis e em que a inclusão de qualquer uma das outras covariáveis já não produz um decréscimo no valor do DIC. As covariáveis seleccionadas por este método são as mesmas do método SIMD: $x_1, x_2, x_3, x_4, x_7, x_8, x_9, x_{10}$ e x_{13} .

A aplicação do método FVS com a medida de discrepância $\widehat{\text{BIC}}$ — resultados na tabela 3.27 — proporciona a entrada das covariáveis no modelo pela ordem obtida com o DIC. No entanto, o processo de selecção de covariáveis pára na 6ª iteração quando estão no modelo as covariáveis x_1, x_2, x_3, x_7 e x_9 , dando origem a um modelo mais parcimonioso.

Tabela 3.26: Resultados do método FVS via DIC, para o exemplo da homocisteinemia¹.

cov.	2ª it.	3ª it.	4ª it.	5ª it.	6ª it.	7ª it.	8ª it.	9ª it.	10ª it.	11ª it.
x_1	654.9	637.9	★	<u>619.5</u>	◇	◇	◇	◇	◇	◇
x_2	652.7	<u>627.4</u>	◇	◇	◇	◇	◇	◇	◇	◇
x_3	668.9	649.1	628.4	◇	◇	◇	◇	◇	◇	◇
x_4	661.9	645.9	★	625.7	615.7	607.1	603.4	<u>599.8</u>	◇	◇
x_5	668.3	650.8	★	630.5	621.5	612.2	608.7	605.2	601.8	599.9
x_6	660.7	646.5	★	629.2	621.0	612.4	608.5	604.6	601.7	599.8
x_7	<u>649.0</u>	◇	◇	◇	◇	◇	◇	◇	◇	◇
x_8	668.4	650.9	★	630.2	620.1	606.8	<u>602.9</u>	◇	◇	◇
x_9	661.1	640.9	★	621.3	<u>610.4</u>	◇	◇	◇	◇	◇
x_{10}	669.0	648.3	★	629.6	621.4	611.2	606.9	601.9	<u>597.6</u>	◇
x_{11}	660.5	647.1	★	628.6	621.1	611.9	607.9	604.4	601.1	599.1
x_{12}	664.4	644.9	★	627.0	620.3	611.3	606.3	603.7	600.4	597.9
x_{13}	660.4	640.0	★	623.3	615.6	<u>606.4</u>	◇	◇	◇	◇
DIC	649.0	627.4	628.4	619.5	610.4	606.4	602.9	599.8	597.6	◇
decr.	18.2	★	20.6	7.9	9.1	4.0	3.5	3.1	2.2	◇

¹Na 1ª iteração, em que é ajustado um modelo só com a constante, tem-se DIC = 667.2.

Tabela 3.27: Resultados do método FVS via $\widehat{\text{BIC}}$, para o exemplo da homocisteinemia¹.

cov.	2ª it.	3ª it.	4ª it.	5ª it.	6ª it.	7ª it.
x_1	666.4	653.2	★	<u>642.5</u>	◇	◇
x_2	664.2	<u>642.8</u>	◇	◇	◇	◇
x_3	680.3	664.3	647.5	◇	◇	◇
x_4	673.5	661.2	★	648.7	642.6	637.8
x_5	679.8	666.2	★	653.4	648.3	642.8
x_6	672.2	661.8	★	652.1	647.7	643.0
x_7	<u>660.5</u>	◇	◇	◇	◇	◇
x_8	679.9	666.3	★	653.2	646.8	637.5
x_9	672.7	656.3	★	644.3	<u>637.2</u>	◇
x_{10}	680.6	663.6	★	652.7	648.1	641.8
x_{11}	672.0	662.4	★	651.5	648.0	642.6
x_{12}	675.9	660.3	★	650.0	647.0	642.0
x_{13}	671.9	655.3	★	646.4	642.5	637.2
$\widehat{\text{BIC}}$	660.5	642.8	647.5	642.5	637.2	◇
decr.	14.3	★	13.0	5.0	5.3	◇

¹Na 1ª iteração, em que é ajustado um modelo só com a constante, tem-se $\widehat{\text{BIC}}$ = 674.8.

Relativamente à medida LLP as covariáveis também entram no modelo pela mesma ordem do DIC, mas é efectuada mais uma iteração, correspondente à entrada da covariável x_{12} no modelo, conforme se pode observar pela análise da tabela 3.28. Quando se utiliza a medida SlnCPO, a ordem de entrada das covariáveis no modelo é diferente da das medidas anteriores, mas as covariáveis escolhidas acabam por ser as mesmas do DIC, conforme se pode ver na tabela 3.29. No caso do SlnCPO a procura do melhor modelo corresponde à maximização da medida de discrepância e o valor sublinhado indica o máximo obtido em cada iteração; a comparação dos sucessivos modelos é efectuada à custa do acréscimo do valor da medida de discrepância.

Tabela 3.28: Resultados do método FVS via LLP, para o exemplo da homocisteinemia¹.

cov.	2 ^a it.	3 ^a it.	4 ^a it.	5 ^a it.	6 ^a it.	7 ^a it.	8 ^a it.	9 ^a it.	10 ^a it.	11 ^a it.	12 ^a it.
x_1	651.9	633.9	*	<u>613.5</u>	◇	◇	◇	◇	◇	◇	◇
x_2	649.6	<u>623.4</u>	◇	◇	◇	◇	◇	◇	◇	◇	◇
x_3	665.9	645.1	623.4	◇	◇	◇	◇	◇	◇	◇	◇
x_4	658.9	641.9	*	619.7	608.7	599.0	594.4	<u>589.8</u>	◇	◇	◇
x_5	665.2	646.8	*	624.4	614.4	604.2	599.6	595.1	590.8	587.8	586.9
x_6	657.7	642.5	*	623.2	613.9	604.3	599.4	594.6	590.6	587.7	586.8
x_7	<u>646.0</u>	◇	◇	◇	◇	◇	◇	◇	◇	◇	◇
x_8	665.4	646.9	*	624.1	613.0	598.8	<u>593.9</u>	◇	◇	◇	◇
x_9	658.1	636.9	*	615.3	<u>603.3</u>	◇	◇	◇	◇	◇	◇
x_{10}	666.0	644.3	*	623.6	614.3	603.1	597.9	591.8	<u>586.6</u>	◇	◇
x_{11}	657.4	643.0	*	622.5	614.1	603.9	598.9	594.3	590.1	587.0	586.1
x_{12}	661.4	640.9	*	621.0	613.2	603.3	597.2	593.7	589.4	<u>585.8</u>	◇
x_{13}	657.4	636.0	*	617.3	608.6	<u>598.5</u>	◇	◇	◇	◇	◇
LLP	646.0	623.4	*	613.5	603.3	598.5	593.9	589.8	586.6	585.8	◇
decr.	19.2	*	22.6	9.9	10.2	4.8	4.6	4.1	3.2	0.8	◇

¹Na 1^a iteração, em que é ajustado um modelo só com a constante, tem-se LLP = 665.2.

Tabela 3.29: Resultados do método FVS via SlnCPO, para o exemplo da homocisteinemia¹.

cov.	2 ^a it.	3 ^a it.	4 ^a it.	5 ^a it.	6 ^a it.	7 ^a it.	8 ^a it.	9 ^a it.	10 ^a it.	11 ^a it.
x_1	-311.2	-305.5	*	-296.0	<u>-289.9</u>	◇	◇	◇	◇	◇
x_2	-309.5	<u>-298.9</u>	◇	◇	◇	◇	◇	◇	◇	◇
x_3	-317.0	-309.0	-306.0	◇	◇	◇	◇	◇	◇	◇
x_4	-313.6	-309.6	*	-298.4	-299.0	-289.0	-286.3	<u>-285.2</u>	◇	◇
x_5	-316.3	-310.3	*	-302.0	-296.4	-290.9	-289.3	-287.4	-285.2	-285.2
x_6	-316.9	-309.1	*	-300.2	-297.2	-290.7	-288.5	-286.6	-285.5	-285.5
x_7	<u>-308.9</u>	◇	◇	◇	◇	◇	◇	◇	◇	◇
x_8	-316.1	-311.3	*	-300.8	-295.1	-288.0	<u>-286.2</u>	◇	◇	◇
x_9	-313.1	-305.7	*	<u>-295.6</u>	◇	◇	◇	◇	◇	◇
x_{10}	-317.6	-309.1	*	-300.2	-294.6	-290.2	-287.9	-285.7	<u>-283.8</u>	◇
x_{11}	-313.1	-308.8	*	-299.4	-295.6	-290.2	-287.7	-286.4	-286.5	-284.4
x_{12}	-314.6	-306.8	*	-298.5	-294.1	-290.2	-287.9	-286.8	-286.1	-284.1
x_{13}	-312.8	-304.6	*	-297.0	-292.2	<u>-287.7</u>	◇	◇	◇	◇
SlnCPO	-308.9	-298.9	-306.0	-295.6	-289.9	-287.7	-286.2	-285.2	-283.8	◇
acr.	7.2	*	2.9	10.4	5.7	2.2	1.5	1.0	1.4	◇

¹Na 1^a iteração, em que é ajustado um modelo só com a constante, tem-se SlnCPO = -316.1.

Método BVE

Os pormenores de implementação do método BVE no WinBUGS estão sumariados no anexo D.2.3. O método BVE via DIC, cujos resultados são apresentados na tabela 3.30, conduzem à selecção do mesmo conjunto de covariáveis do método FVS via DIC. Começa por ajustar-se o modelo completo,

$$\text{HBP}_i = \exp\{b_0 + b_1x_{1,i} + \dots + b_{13}x_{13,i}\} \times \epsilon_i, \quad \text{com } \epsilon_i \sim \text{Ga}(\nu, \nu),$$

para o qual se tem $\text{DIC} = 603.2$.

Na 2ª iteração são ajustados 13 modelos de regressão gama, com 12 covariáveis, resultantes da exclusão de cada uma das covariáveis. O valor mais favorável da medida de discrepância encontra-se sublinhado; neste caso, o desempate entre as covariáveis x_3 , x_5 e x_6 foi fácil de efectuar pois a covariável x_3 (fumar2) não pode sair do modelo uma vez que x_2 (fumar1) está no modelo, tendo-se optado, atendendo aos resultados obtidos com as outras medidas de discrepância, por tirar do modelo a covariável x_5 .

Nas três iterações seguintes o mínimo do DIC refere-se à exclusão da covariável fumar2; no entanto, pelos motivos já mencionados, esta covariável não é excluída do modelo. Assim, saem do modelo as covariáveis x_6 , x_{11} e x_{12} , na 3ª, 4ª e 5ª iterações, respectivamente.

Não obstante o decréscimo no valor do DIC resultante da exclusão da covariável x_{12} ser nulo, o método BVE não pára ainda nesta iteração, pois é sempre preferível o modelo mais parcimonioso. Na 6ª iteração, a exclusão de qualquer uma das covariáveis já não produz um decréscimo no valor do DIC e o método BVE pára.

Tabela 3.30: Resultados do método BVE via DIC, para o exemplo da homocisteinemia¹.

cov.	2ª it.	3ª it.	4ª it.	5ª it.	6ª it.
x_1	615.8	614.1	612.4	613.3	611.3
x_2	611.0	608.9	607.7	607.0	607.1
x_3	601.3	599.0	597.1	595.9	595.7
x_4	607.5	605.2	603.6	602.4	601.9
x_5	<u>601.3</u>	◇	◇	◇	◇
x_6	601.3	<u>599.1</u>	◇	◇	◇
x_7	609.3	607.3	622.5	624.9	624.1
x_8	607.9	605.6	603.5	602.3	603.5
x_9	618.1	616.1	615.1	613.9	613.3
x_{10}	605.6	603.6	601.6	600.4	599.9
x_{11}	602.0	599.8	<u>597.9</u>	◇	◇
x_{12}	603.3	601.1	599.1	<u>597.9</u>	◇
x_{13}	608.3	606.2	604.0	602.7	602.1
DIC	601.3	599.1	597.9	597.9	◇
decr.	1.9	2.2	1.2	0.0	◇

¹Na 1ª iteração, em que é ajustado um modelo com todas as covariáveis, tem-se $\text{DIC} = 603.2$.

Os resultados obtidos com $\widehat{\text{BIC}}$, apresentados na tabela 3.31, são idênticos aos do DIC até à 5ª iteração, conduzindo à exclusão das mesmas covariáveis pela mesma ordem. No entanto, o BVE via $\widehat{\text{BIC}}$ continua a excluir covariáveis até à 9ª iteração — x_{10} , x_4 , x_8 e x_{13}

também saem do modelo. Este método pára na 10ª iteração pois a exclusão de qualquer uma das covariáveis do modelo já não origina uma redução no valor do $\widehat{\text{BIC}}$.

Relativamente ao método BVE via LLP, cujos resultados constam da tabela 3.32, o processo de exclusão de covariáveis pára bastante mais cedo, mais precisamente na 4ª iteração — foram excluídas do modelo as covariáveis x_5 , x_6 e x_{11} . Da análise da tabela 3.33 verifica-se que, embora a ordem de exclusão das duas primeiras covariáveis não seja coincidente, as covariáveis excluídas quando se aplica o método BVE via SlnCPO são as mesmas do DIC.

Tabela 3.31: Resultados do método BVE via $\widehat{\text{BIC}}$, para o exemplo da homocisteinemia¹.

cov.	2ª it.	3ª it.	4ª it.	5ª it.	6ª it.	7ª it.	8ª it.	9ª it.	10ª it.
x_1	669.3	663.9	658.5	655.4	649.6	648.4	646.6	643.0	644.4
x_2	664.7	658.6	653.6	649.1	645.4	645.5	644.7	644.6	647.1
x_3	654.9	648.7	643.1	638.1	634.0	632.2	631.5	631.3	631.3
x_4	661.0	655.1	649.5	644.4	640.2	<u>637.5</u>	◇	◇	◇
x_5	<u>654.8</u>	◇	◇	◇	◇	◇	◇	◇	◇
x_6	654.9	<u>648.8</u>	◇	◇	◇	◇	◇	◇	◇
x_7	662.9	657.1	668.4	667.0	662.4	657.0	658.9	661.1	659.2
x_8	661.4	655.4	649.5	644.5	641.8	638.0	<u>637.1</u>	◇	◇
x_9	671.7	665.9	661.1	656.0	651.6	647.0	646.9	642.5	642.6
x_{10}	659.2	653.4	647.6	642.7	<u>638.1</u>	◇	◇	◇	◇
x_{11}	655.7	649.7	<u>643.9</u>	◇	◇	◇	◇	◇	◇
x_{12}	656.9	650.9	645.0	<u>640.0</u>	◇	◇	◇	◇	◇
x_{13}	661.9	655.9	650.0	644.8	640.3	638.0	637.6	<u>637.1</u>	◇
$\widehat{\text{BIC}}$	654.8	648.8	643.9	640.0	638.1	637.5	637.1	637.1	◇
decr.	5.8	6.0	4.9	3.9	1.9	0.6	0.4	0.0	◇

¹Na 1ª iteração, em que é ajustado um modelo com todas as covariáveis, tem-se $\widehat{\text{BIC}} = 660.6$.

Tabela 3.32: Resultados do método BVE via LLP, para o exemplo da homocisteinemia¹.

cov.	2ª it.	3ª it.	4ª it.	5ª it.
x_1	601.6	601.0	600.3	602.2
x_2	596.9	595.9	595.7	595.9
x_3	587.1	586.0	585.0	584.8
x_4	593.4	592.2	591.5	591.3
x_5	<u>587.1</u>	◇	◇	◇
x_6	587.2	<u>586.0</u>	◇	◇
x_7	595.2	594.2	610.4	613.8
x_8	593.7	592.5	591.4	591.2
x_9	604.0	603.0	603.0	602.8
x_{10}	591.6	590.5	589.6	589.4
x_{11}	587.9	586.8	<u>585.9</u>	◇
x_{12}	589.2	588.0	587.0	586.8
x_{13}	594.2	593.1	592.0	591.6
LLP	587.1	586.0	585.9	◇
decr.	1.0	1.1	0.1	◇

¹Na 1ª iteração, em que é ajustado um modelo com todas as covariáveis, tem-se LLP = 588.1.

Tabela 3.33: Resultados do método BVE via SlnCPO, para o exemplo da homocisteinemia¹.

cov.	2 ^a it.	3 ^a it.	4 ^a it.
x_1	-291.6	-292.0	-289.9
x_2	-289.7	-288.8	-288.4
x_3	-285.4	-284.1	-283.3
x_4	-288.0	-287.7	-286.8
x_5	-286.1	<u>-283.8</u>	◇
x_6	<u>-285.3</u>	◇	◇
x_7	-289.2	-296.5	-295.3
x_8	-288.1	-288.1	-286.1
x_9	-295.0	-293.2	-293.7
x_{10}	-286.9	-286.5	-285.8
x_{11}	-286.4	-285.1	-284.5
x_{12}	-286.5	-284.7	-284.9
x_{13}	-288.3	-287.1	-288.2
SlnCPO	-285.3	-283.8	◇
acr.	0.3	1.5	◇

¹Na 1^a iteração, em que é ajustado um modelo com todas as covariáveis, tem-se SlnCPO = -285.6.

3.6.4 Método Gibbs de selecção de covariáveis

Devido à impossibilidade de monitorizar as probabilidades *a posteriori* dos 8192 possíveis modelos, começa por aplicar-se o método GVS apenas com o intuito de monitorizar as probabilidades *a posteriori* de inclusão das covariáveis no modelo. Os resultados obtidos são usados para restringir o número inicial de covariáveis, escolhendo-se as que apresentam probabilidades *a posteriori* de inclusão no modelo mais elevadas.

Este método é aplicado em duas situações distintas: sem restrição no espaço dos modelos e com uma restrição nesse espaço. A restrição considerada obriga a que as duas variáveis *dummy* que definem a covariável fumar estejam, ou ambas presentes, ou ambas ausentes do modelo.

Como a estrutura de correlação da cadeia é elevada, foram efectuadas 5000 iterações, depois de um período de aquecimento de 4000 iterações e com um espaçamento de 10 iterações. Os código do WinBUGS para aplicação deste método ao conjunto das 13 covariáveis, sem restrição no espaço dos modelos e com restrição no espaço dos modelos, são apresentados no anexo D.2.4.

Os resultados obtidos para as estimativas das probabilidades *a posteriori* de inclusão das covariáveis no modelo nas duas situações já referidas são apresentados na tabela 3.34.

Analisa-se em primeiro lugar a situação “sem restrição no espaço dos modelos”. As estimativas das probabilidades *a posteriori* de inclusão das covariáveis no modelo mais elevadas são as referentes às covariáveis x_7 e x_2 , com valores próximos de 1. A estimativa da probabilidade *a posteriori* de inclusão da covariável x_1 no modelo é 0.23, seguindo-se x_9 e x_{13} , com estimativas da probabilidade *a posteriori* de inclusão da covariável no modelo de 0.0733 e 0.0272, respectivamente. Para as restantes covariáveis, as estimativas da probabilidade *a posteriori* de inclusão das covariáveis no modelo são muito pequenas

(inferiores a 0.01). Como a monitorização das probabilidades *a posteriori* de 256 modelos não exige um grande esforço computacional, além das covariáveis já referidas junta-se x_3 , x_4 e x_{12} ao conjunto de covariáveis ao qual vai ser novamente aplicado o método GVS. O código do WinBUGS para aplicação deste método ao subconjunto das 8 covariáveis é apresentado no anexo D.2.4. Os resultados obtidos, referentes às estimativas das probabilidades *a posteriori* de inclusão das covariáveis no modelo e às estimativas das probabilidades *a posteriori* dos próprios modelos, são apresentados nas tabelas 3.35 e 3.36, respectivamente.

Tabela 3.34: Estimativas das probabilidades *a posteriori* de inclusão das covariáveis no modelo, para o exemplo da homocisteinemia.

Sem restrição no espaço dos modelos				Com restrição no espaço dos modelos			
termos	média	d.p.	erro MC	termos	média	d.p.	erro MC
$\gamma_1 = 1$	0.2308	0.4213	0.006179	$\gamma_1 = 1$	0.3714	0.4832	0.009494
$\gamma_2 = 1$	0.9816	0.1343	0.007898	$\gamma_{2,3} = 1$	0.1182	0.3228	0.017610
$\gamma_3 = 1$	0.0016	0.0402	1.833E-4	$\gamma_4 = 1$	0.0158	0.1247	0.001888
$\gamma_4 = 1$	0.0078	0.0878	4.725E-4	$\gamma_5 = 1$	0.0004	0.0245	3.409E-4
$\gamma_5 = 1$	9.8E-4	0.0313	1.421E-4	$\gamma_6 = 1$	0.0140	0.1175	0.001975
$\gamma_6 = 1$	0.0028	0.0530	2.535E-4	$\gamma_7 = 1$	0.9376	0.2419	0.016090
$\gamma_7 = 1$	0.9986	0.0374	8.117E-4	$\gamma_8 = 1$	0.0042	0.0647	8.707E-4
$\gamma_8 = 1$	0.0017	0.0414	1.847E-4	$\gamma_9 = 1$	0.1672	0.3732	0.008180
$\gamma_9 = 1$	0.0733	0.2606	0.003336	$\gamma_{10} = 1$	0.0024	0.0489	6.970E-4
$\gamma_{10} = 1$	0.0013	0.0358	1.548E-4	$\gamma_{11} = 1$	0.0046	0.0677	0.001054
$\gamma_{11} = 1$	0.0025	0.0497	2.208E-4	$\gamma_{12} = 1$	0.0112	0.1052	0.001548
$\gamma_{12} = 1$	0.0040	0.0633	3.002E-4	$\gamma_{13} = 1$	0.1110	0.3141	0.005306
$\gamma_{13} = 1$	0.0272	0.1627	0.001249				

Tabela 3.35: Estimativas das probabilidades *a posteriori* de inclusão das covariáveis no modelo, para o exemplo da homocisteinemia, quando se excluem algumas covariáveis.

Sem restrição no espaço dos modelos				Com restrição no espaço dos modelos			
termos	média	d.p.	erro MC	termos	média	d.p.	erro MC
$\gamma_1 = 1$	0.2297	0.4206	0.006327	$\gamma_1 = 1$	0.3616	0.4805	0.010810
$\gamma_2 = 1$	0.9948	0.0721	0.001649	$\gamma_{2,3} = 1$	0.0646	0.2458	0.011490
$\gamma_3 = 1$	0.0014	0.0369	1.761E-4	$\gamma_4 = 1$	0.0142	0.1183	0.001685
$\gamma_4 = 1$	0.0090	0.0947	5.229E-4	$\gamma_6 = 1$	0.0154	0.1231	0.002084
$\gamma_7 = 1$	0.9999	0.0100	4.434E-5	$\gamma_7 = 1$	0.9516	0.2146	0.014430
$\gamma_9 = 1$	0.0736	0.2612	0.002852	$\gamma_9 = 1$	0.1766	0.3813	0.006979
$\gamma_{12} = 1$	0.0038	0.0617	3.014E-4	$\gamma_{12} = 1$	0.0114	0.1062	0.001715
$\gamma_{13} = 1$	0.0267	0.1613	0.001157	$\gamma_{13} = 1$	0.1228	0.3282	0.005350

Conforme já foi referido, aplica-se também o método GVS na situação “com restrição no espaço dos modelos”. Assim, começa por aplicar-se o método GVS ao conjunto de 13 covariáveis com a restrição de x_2 e x_3 serem ambas incluídas ou ambas excluídas do modelo. Com esta imposição, o número de modelos a monitorizar, no caso em que se considera 9 covariáveis, é também 256, uma vez que duas das covariáveis estão, simultaneamente, presentes ou ausentes do modelo. Neste caso, as estimativas mais elevadas das probabilidade *a posteriori* de inclusão das covariáveis no modelo são referentes

às covariáveis x_7 , x_1 , x_9 , x_2 e x_3 , x_{13} , x_4 , x_6 e x_{12} (tabela 3.34). A estimativa da probabilidade *a posteriori* de inclusão simultânea de x_2 e x_3 no modelo é 0.1182, valor bastante inferior ao obtido para x_2 quando não se efectua a restrição no espaço dos modelos. Refira-se, ainda, que as estimativas das probabilidades *a posteriori* de inclusão das covariáveis no modelo são menos díspares do que no caso “sem restrição no espaço dos modelos”. O código do WinBUGS, para aplicação do método GVS ao subconjunto de 9 covariáveis com a restrição no espaço dos modelos, é apresentado no anexo D.2.4; os resultados obtidos estão sumariados nas tabelas 3.35 e 3.37.

Tabela 3.36: Dez modelos com maior probabilidade *a posteriori* obtidos pelo método GVS sem restrição no espaço dos modelos, para o exemplo da homocisteinemia, quando se considera um subconjunto de 8 covariáveis.

covariáveis								probabilidades <i>a posteriori</i>		
x_1	x_2	x_3	x_4	x_7	x_9	x_{12}	x_{13}	média	d.p.	erro MC
	×			×				0.6940	0.4608	6.27E-03
×	×			×				0.1917	0.3936	5.21E-03
	×			×	×			0.0396	0.1951	1.45E-03
×	×			×	×			0.0298	0.1700	2.18E-03
	×			×			×	0.0214	0.1448	9.88E-03
	×		×	×				0.0056	0.0745	3.94E-03
	×			×		×		0.0031	0.0560	2.70E-03
×	×		×	×				0.0266	0.0519	3.15E-03
×	×			×			×	0.0128	0.0483	2.49E-04
				×				0.0115	0.0460	5.77E-03

Tabela 3.37: Dez modelos com maior probabilidade *a posteriori* obtidos pelo método GVS com restrição no espaço dos modelos, para o exemplo da homocisteinemia, quando se considera um subconjunto de 9 covariáveis.

covariáveis									probabilidades <i>a posteriori</i>		
x_1	x_2 e x_3	x_4	x_6	x_7	x_9	x_{12}	x_{13}		média	d.p.	erro MC
				×					0.3948	0.4888	0.011810
×				×					0.2012	0.4009	0.008285
×				×	×				0.0960	0.2946	0.005821
				×			×		0.0826	0.2753	0.004490
	×			×					0.0514	0.2208	0.009721
				×	×				0.0496	0.2171	0.003567
×									0.0182	0.1337	0.005401
									0.0160	0.1255	0.005668
×				×			×		0.0150	0.1216	0.001794
				×	×		×		0.0128	0.1124	0.001511

No caso em que não há restrição no espaço dos modelos, os dez modelos com maior probabilidade *a posteriori* incluem a covariável x_7 e os nove primeiros modelos também incluem x_2 . A soma das probabilidades *a posteriori* destes dez modelos é 0.9924. O modelo com maior probabilidade *a posteriori*, que inclui as covariáveis x_2 e x_7 , tem probabilidade *a posteriori* igual a 0.7. O segundo modelo com maior probabilidade *a posteriori*, que inclui as covariáveis x_1 , x_2 , x_7 , tem probabilidade bastante inferior.

Quando se considera a restrição no espaço dos modelos, a soma das probabilidades *a posteriori* dos dez primeiros modelos é, aproximadamente, 0.94, e a dos vinte primeiros é 0.98. Neste caso, existe uma maior incerteza na selecção do modelo pois os seis modelos mais “visitados”, constituídos pelos subconjuntos de covariáveis $\{x_7\}$, $\{x_1, x_7\}$, $\{x_1, x_7, x_9\}$, $\{x_7, x_{13}\}$, $\{x_2, x_3, x_7\}$ e $\{x_7, x_9\}$, têm probabilidades *a posteriori* de 0.4, 0.2, 0.1, 0.08, 0.05 e 0.05, respectivamente.

3.6.5 Comparação de resultados

Na tabela 3.38 são indicadas as covariáveis incluídas nos modelos seleccionados por cada um dos métodos e para cada uma das medidas de discrepância. Os modelos de regressão gama seleccionados, os correspondentes valores das medidas de discrepância e as estimativas das probabilidades *a posteriori* estão sumariados na tabela 3.39.

Tabela 3.38: Covariáveis incluídas nos modelos de regressão gama seleccionados por cada um dos métodos, no caso do exemplo da homocisteinemia.

método	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	modelo
SIMD - DIC, $\widehat{\text{BIC}}$, SlnCPO	×	×	×	×			×	×	×	×			×	m_1
SIMD - LLP	×	×	×	×			×	×	×	×		×	×	m_2
FVS - DIC, SlnCPO	×	×	×	×			×	×	×	×			×	m_1
FVS - $\widehat{\text{BIC}}$	×	×	×				×		×					m_3
FVS - LLP	×	×	×	×			×	×	×	×		×	×	m_2
BVE - DIC, SlnCPO	×	×	×	×			×	×	×	×			×	m_1
BVE - $\widehat{\text{BIC}}$	×	×	×				×		×					m_3
BVE - LLP	×	×	×	×			×	×	×	×		×	×	m_2
GVS sem restrição		×					×							m_4
	×	×					×							m_5
GVS com restrição							×							m_6
	×						×							m_7

Tabela 3.39: Valores das medidas de discrepância dos modelos de regressão gama m_1 a m_7 , para o exemplo da homocisteinemia.

modelo	DIC	$\widehat{\text{BIC}}$	LLP	SlnCPO	$\text{Pr}\{m_i \mathcal{D}\}$
m_1	597.9	640.0	586.6	-283.7	—
m_2	597.9	643.9	585.8	-284.5	—
m_3	610.4	637.2	603.3	-289.9	—
m_6	649.0	660.5	646.0	-308.9	0.3948
m_7	653.3	637.9	633.9	-304.3	0.2012
m_4	627.4	642.8	623.4	-298.9	0.6940
m_5	617.4	636.6	612.4	-296.7	0.1917

Os métodos SIMD, FVS e BVE com as medidas de discrepância DIC e SlnCPO levam à selecção do mesmo modelo, indicado por m_1 , que parece ser o melhor modelo. Embora o mínimo dos valores do DIC dos 7 modelos seleccionados seja atingido para os modelos m_1 e m_2 , o modelo m_1 deve ser preferido ao modelo m_2 pois tem menos uma covariável.

Os restantes modelos apresentam valores do DIC bastante mais elevados.

Os métodos FVS e BVE via $\widehat{\text{BIC}}$ dão origem ao modelo m_3 que tem apenas 5 covariáveis incluídas no preditor linear, confirmando-se a tendência desta medida para privilegiar os modelos com menos covariáveis. O modelo m_3 tem, no entanto, um desempenho bastante inferior para as outras medidas de discrepância.

Qualquer um dos três métodos de selecção de covariáveis via LLP dá origem ao modelo menos parcimonioso, que se indica por m_2 .

Os valores das medidas de discrepância referentes aos modelos m_1 e m_2 são bastante idênticos. Os restantes modelos têm valores mais desfavoráveis para todas as medidas de discrepância, com excepção do $\widehat{\text{BIC}}$.

O modelo com maior probabilidade *a posteriori* obtido pelo método GVS, quando não se considera qualquer restrição no espaço dos modelos, é m_4 , constituído pelas covariáveis x_2 e x_7 , com $\Pr\{m_4|\mathcal{D}\} \approx 0.7$; o segundo modelo com maior probabilidade *a posteriori*, m_5 , constituído pelas covariáveis x_1 , x_2 e x_7 , tem $\Pr\{m_5|\mathcal{D}\} \approx 0.2$. Quando se impõe a restrição “as covariáveis x_2 e x_3 estarem ambas presentes ou ambas ausentes do modelo”, os dois modelos com maior probabilidade *a posteriori*, obtidos pelo método GVS, são m_6 (que inclui a covariável x_7) e m_7 (que inclui as covariáveis x_1 e x_7), com $\Pr\{m_6|\mathcal{D}\} \approx 0.4$ e $\Pr\{m_7|\mathcal{D}\} \approx 0.2$. Estes modelos têm valores bastante desfavoráveis das medidas de discrepância DIC, LLP e SlnCPO; no entanto, o modelo m_5 é o que apresenta o valor mais favorável do $\widehat{\text{BIC}}$.

Paulino *et al.* (2003) utilizam os métodos de selecção de covariáveis SIMD e SSVS para ajustar modelos de regressão gama a este conjunto de dados da homocisteinemia, obtendo resultados idênticos aos apresentados neste trabalho. Com o método SIMD, o modelo escolhido é o que aqui se identifica por m_1 , logo seguido do modelo m_2 . Quando utilizam o método SSVS, as estimativas das probabilidades *a posteriori* de inclusão das covariáveis no modelo são praticamente nulas, com excepção das probabilidades *a posteriori* de inclusão das covariáveis x_1 , x_2 e x_7 , sendo o modelo seleccionado por este método constituído por estas covariáveis e por x_3 .

3.7 Comentários

O método GVS tem a vantagem de possibilitar a estimação das probabilidades *a posteriori* dos modelos resultantes de todas as combinações possíveis de covariáveis. Se existir um modelo com probabilidade *a posteriori* muito superior à dos outros, pode optar-se por seleccionar apenas esse modelo; neste caso, o modelo com probabilidade *a posteriori* máxima pode ser entendido como se tivesse sido o modelo seleccionado por pesquisa exhaustiva. Os resultados obtidos com os exemplos da massa gorda (regressão normal) e da homocisteinemia (regressão gama) caracterizam-se pelo facto dos modelos com maior probabilidade *a posteriori* terem poucas covariáveis, e apresentarem valores bastante desfavoráveis (quando comparados com outros) das medidas de discrepância, nomeadamente do DIC. Este facto sugere que, para estes dois exemplos, o método GVS pode, por algum motivo, não ser o mais adequado para escolher o melhor modelo. O motivo mais plausível parece estar relacionado com o facto do número de observações ser muito reduzido quando comparado com o número de modelos cuja probabilidade *a*

posteriori se pretende estimar.

A aplicação do método GVS fica comprometida, mesmo para um número moderado de covariáveis, devido ao elevado número de modelos a monitorizar. Uma alternativa, para os modelos de regressão linear normal, consiste em determinar o modelo de probabilidade *a posteriori* mediana que, de um modo geral, coincide com o modelo de probabilidade *a posteriori* máxima.

O método de Carlin e Chib não é especialmente indicado para a selecção de covariáveis em modelos de regressão pois, para assegurar uma convergência rápida, a especificação das distribuições *pseudoprior* tem que ser feita de modo adequado, o que dificulta a aplicação deste método, mesmo quando o número de modelos em comparação não é muito elevado.

Com os métodos de pesquisa FVS e BVE, o número máximo de modelos a ajustar é bastante inferior ao número de modelos ajustados quando se efectua uma pesquisa exaustiva, mesmo para um número moderado de covariáveis. Quando existe multicolinearidade entre as covariáveis ou quando existem covariáveis conjuntamente, mas não individualmente, significativas, a garantia de sucesso do método BVE, na pesquisa do melhor subconjunto de covariáveis, é superior à do método FVS.

Nos exemplos da massa gorda e da homocisteinemia já referidos, cada um dos quais com um conjunto de 13 possíveis covariáveis, o número de iterações do método de pesquisa BVE via DIC, LLP e SlnCPO é inferior ao do método de pesquisa FVS com a correspondente medida de discrepância. No caso da medida de discrepância considerada ser o \widehat{BIC} sucede precisamente o contrário: o número de iterações do método FVS é inferior ao do método BVE. A utilização da medida de discrepância \widehat{BIC} com os métodos de pesquisa FVS ou BVE proporciona a selecção de modelos com um número de covariáveis inferior ao dos modelos obtidos com as outras medidas de discrepância consideradas.

No exemplo da massa gorda, os métodos BVE e FVS via medidas de discrepância conduzem à selecção de modelos (subconjuntos de covariáveis) distintos. Este facto pode ser devido à existência de correlações elevadas entre a maioria das covariáveis. Esta situação não ocorre no exemplo da homocisteinemia, para o qual os referidos métodos seleccionam, para uma dada medida de discrepância, o mesmo modelo.

Uma das desvantagens do método SIMD é o facto de não poder ser computacionalmente implementado, pois necessita a intervenção do investigador ao longo de todo o processo, para averiguar se as distribuições *a posteriori* dos coeficientes de regressão estão ou não suficientemente concentradas em torno de zero. A outra desvantagem deve-se à subjectividade inerente ao processo de avaliar se a distribuição *a posteriori* de um determinado coeficiente de regressão está, ou não, suficientemente concentrada em torno de zero. As principais virtudes do método SIMD relacionam-se com o facto de ser um método de fácil e rápida aplicação e de possibilitar a obtenção de resultados bastante razoáveis.

Comparando os métodos SIMD e BVE com o método FVS, os dois primeiros têm a vantagem de permitir avaliar o efeito de inclusão de todas as covariáveis no modelo, o que pode ser bastante importante quando as covariáveis são muito correlacionadas ou quando existem covariáveis que, embora não sejam individualmente significativas, o sejam conjuntamente.

Capítulo 4

Estudos de Simulação para Comparação de Critérios

4.1 Introdução

Algumas perguntas foram deixadas sem resposta nos capítulos anteriores: (i) Existem diferenças significativas na performance dos vários critérios de selecção de modelos? Em caso de resposta afirmativa, qual ou quais os critérios de selecção de modelos com melhor performance? (ii) De que modo a dimensão da amostra influencia a medida de selecção de modelos e a performance do respectivo critério? (iii) Que alteração no valor da medida de selecção de modelos, em particular que alteração no valor do DIC, pode ser considerada relevante? (iv) Como estimar a variância do DIC? (v) De que modo a correlação entre covariáveis influencia a performance dos critérios de selecção de covariáveis em modelos de regressão? (vi) A razão n/p , onde n é a dimensão da amostra e p é o número de possíveis covariáveis, influencia a performance dos critérios de selecção de covariáveis utilizados? (vii) O método GVS deve ou não ser preferido aos outros métodos de selecção de covariáveis referidos? (viii) De que modo as medidas de selecção de modelos podem ser utilizadas na validação desses mesmos modelos?

Tendo em vista dar resposta a algumas destas questões, efectua-se dois estudos de simulação. O primeiro deles, com dados de contagem, tem como objectivo a comparação da performance de alguns dos critérios de selecção de modelos apresentados no capítulo 2. O segundo estudo, com dados de regressão linear normal, tem como objectivo a comparação da performance dos critérios de selecção de modelos no caso particular da selecção de covariáveis, assim como a avaliação da influência da correlação linear entre covariáveis, da dimensão da amostra n e da razão n/p nos resultados obtidos. Tanto no caso do primeiro como do segundo estudo de simulação, utiliza-se algumas representações gráficas e medidas de estatística descritiva para sumariar os valores das medidas de selecção de modelos.

No primeiro estudo simula-se dados de contagem, quer do modelo beta/binomial negativo, quer do modelo gama/Poisson, apresentados nas secções 2.5.1 e 2.5.2, respectivamente. A estes dados são ajustados tanto o modelo beta/binomial negativo como o modelo gama/Poisson, sendo calculados os valores da distribuição preditiva *a priori*, do BIC, do $\widehat{\text{BIC}}$ e do DIC, para cada um dos modelos. A performance dos critérios de selecção

de modelos BF, ΔBIC , $\Delta\widehat{\text{BIC}}$ e ΔDIC é avaliada pela percentagem de classificação correcta registada. As médias e variâncias dos valores de cada uma das medidas de selecção de modelos são também analisadas. Como se conhece as expressões analíticas dos critérios de selecção de modelos (secção 2.5.3), a programação em Fortran limita-se à geração das amostras de cada um dos modelos, ao cálculo das medidas de selecção de modelos e de algumas estatísticas sumárias, e à determinação da percentagem de classificação correcta das amostras geradas.

No segundo estudo simula-se dados de regressão linear normal de 16 modelos distintos, cada um deles resultante da inclusão no modelo de uma das combinações possíveis de 4 covariáveis. Para as amostras simuladas de cada um dos 16 modelos, são ajustados os 16 modelos de regressão possíveis, sendo calculados os valores do DIC, do $\widehat{\text{BIC}}$ e do LLP para cada um deles. A performance dos critérios de selecção de modelos ΔDIC , $\Delta\widehat{\text{BIC}}$ e ΔLLP é avaliada pela frequência de classificação correcta registada. As médias e variâncias dos valores de cada uma das medidas de selecção de modelos são também analisadas. Como é possível determinar analiticamente a distribuição *a posteriori* do vector de parâmetros do modelo de regressão linear normal, gera-se amostras desta distribuição e estima-se os valores das medidas de selecção de modelos, utilizando o método de Monte Carlo directo.

4.2 Dados de contagem

4.2.1 Planeamento da simulação

Considere-se os modelos beta/binomial negativo e gama/Poisson, definidos por

$$m_1 : X_i | \theta_1 \stackrel{\text{iid}}{\sim} \text{BiNeg}(r, \theta_1), \quad \theta_1 \sim \text{Be}(a, b), \quad i = 1, 2, \dots, n, \quad \text{e}$$

$$m_2 : X_i | \theta_2 \stackrel{\text{iid}}{\sim} \text{Poi}(\theta_2), \quad \theta_2 \sim \text{Ga}(c, d), \quad i = 1, 2, \dots, n,$$

respectivamente, para ajustar a uma amostra $x = (x_1, x_2, \dots, x_n)$, constituída por observações de contagem. Recorre-se à simulação para avaliar e comparar a performance de alguns dos critérios de selecção de modelos apresentados — factor Bayes, critério de informação *bayesiano*, critério de informação *bayesiano* estimado e critério de informação da *deviance* — na escolha entre os modelos m_1 e m_2 . Além disso, tentando dar resposta à questão “que alteração no valor da medida de selecção de modelos pode ser considerada relevante”, são também determinadas as médias e variâncias das medidas de selecção de modelos.

Este estudo de simulação tem duas etapas. Numa primeira etapa, gera-se observações independentes e identicamente distribuídas $\text{BiNeg}(r, \theta_1)$, com $\theta_1 \sim \text{Be}(a, b)$, e ajusta-se m_1 e m_2 à amostra assim obtida, averiguando se os diferentes critérios de selecção de modelos optam, de facto, pelo modelo m_1 em detrimento do modelo m_2 . Numa segunda etapa, gera-se observações independentes e identicamente distribuídas $\text{Poi}(\theta_2)$, com $\theta_2 \sim \text{Ga}(c, d)$, e ajusta-se m_1 e m_2 à amostra obtida, averiguando se os diferentes critérios de selecção de modelos seleccionam o modelo m_2 em vez de m_1 . De modo a averiguar qual a influência da dimensão da amostra na performance dos critérios de selecção de modelos, incrementa-se n de 10 em 10, desde 10 até 100. Estas duas etapas do estudo de simulação estão sumariadas no algoritmo 4.1.

Algoritmo 4.1

1. Gerar um valor para o parâmetro
 - θ_1 da distribuição $\text{Be}(a, b)$, na 1ª etapa do estudo de simulação,
 - θ_2 da distribuição $\text{Ga}(c, d)$, na 2ª etapa do estudo de simulação;
2. Para $k = 1, 2, \dots, 10$,
 - (a) para $t = 1, 2, \dots, T$,
 - i. gerar uma amostra de dimensão $n = 10k$, $x_k^{(t)} = (x_{1k}^{(t)}, x_{2k}^{(t)}, \dots, x_{nk}^{(t)})$, da distribuição
 - $\text{BiNeg}(r, \theta_1)$, na 1ª etapa do estudo de simulação,
 - $\text{Poi}(\theta_2)$, na 2ª etapa do estudo de simulação;
 - ii. determinar os valores da distribuição preditiva *a priori*, do BIC, do $\widehat{\text{BIC}}$ e do DIC, para os modelos m_1 e m_2 (as expressões analíticas estão deduzidas na secção 2.5.3);
 - iii. determinar os valores do factor Bayes, do ΔBIC , do $\Delta\widehat{\text{BIC}}$ e do ΔDIC relativos à comparação de m_1 e m_2 (as expressões analíticas são apresentadas na secção 2.5.3);
 - (b) determinar as médias e variâncias dos valores da distribuição preditiva *a priori*, do BIC, do $\widehat{\text{BIC}}$ e do DIC, obtidos com o ajustamento dos modelos m_1 e m_2 ;
 - (c) determinar as percentagens de classificação correcta — obtidas com o factor Bayes, o ΔBIC , o $\Delta\widehat{\text{BIC}}$ e o ΔDIC — das amostras geradas. ■

Os programa em Fortran referentes à execução do algoritmo 4.1 na 1ª e 2ª etapas do estudo de simulação, designam-se, respectivamente, `selmod_dg_bbn` e `selmod_dg_gp`, e são apresentados em ficheiros com os mesmos nomes, no CD anexo a este documento.

O primeiro parâmetro da distribuição binomial negativa fixa-se em $r = 1$, o que corresponde ao caso particular da distribuição geométrica. Os hiperparâmetros dos dois modelos são escolhidos de modo que as médias preditivas de X_i de cada um desses modelos sejam iguais; atendendo a (2.22) e (2.27) considera-se $bd = (a - 1)c$. Relativamente às variâncias preditivas de X_i para os modelos m_1 e m_2 , cujas expressões são dadas por (2.23) e (2.28), respectivamente, considera-se dois casos distintos: o caso 1, em que as variâncias preditivas de X_i para os dois modelos são idênticas, e o caso 2, em que elas são distintas. O estudo de simulação pode ser esquematizado do seguinte modo:

Caso 1. Compara-se o ajustamento dos modelos m_1 com $r = 1$, $a = 3$ e $b = 3$, e m_2 com $c = 0.225$ e $d = 0.150$, a amostras de dados de contagem. As médias preditivas de X_i são iguais a 1.5 para ambos os modelos, sendo a variância preditiva de X_i no caso de m_1 igual a 11.25 e a variância preditiva de X_i no caso de m_2 igual a 11.5.

Amostras geradas de m_1 . O programa `selmod_dg_bbn`, referente à implementação do algoritmo 4.1 na 1ª etapa do estudo de simulação, é executado 6 vezes (isto é, para 6 valores de θ_1) com $T = 500$. Alguns dos resultados obtidos são apresentados na tabela 4.1 e na tabela E.3 do anexo E.

Amostras geradas de m_2 . O programa `selmod_dg_gp`, referente à implementação do algoritmo 4.1 na 2ª etapa do estudo de simulação, é executado 6 vezes (isto é, para 6 valores de θ_2) com $T = 500$. Alguns dos resultados obtidos são apresentados na tabela 4.2 e na tabela E.4 do anexo E.

Caso 2. Compara-se o ajustamento dos modelos m_1 com $r = 1$, $a = 2$ e $b = 2$, e m_2 com $c = 2$ e $d = 1$ (Bernardo e Smith, 2000, exemplo 6.5), a amostras de dados de contagem. As médias preditivas de X_i são iguais a 2 para ambos os modelos, não estando definida a variância preditiva de X_i no caso de m_1 e sendo a variância preditiva de X_i no caso de m_2 igual a 4.

Amostras geradas de m_1 . O programa `selmod_dg_bbn`, referente à implementação do algoritmo 4.1 na 1ª etapa do estudo de simulação, é executado 6 vezes (isto é, para 6 valores de θ_1) com $T = 500$. Alguns dos resultados obtidos são apresentados nas tabelas E.1 e E.5 do anexo E.

Amostras geradas de m_2 . O programa `selmod_dg_gp`, referente à implementação do algoritmo 4.1 na 2ª etapa do estudo de simulação é executado 6 vezes (isto é, para 6 valores de θ_2) com $T = 500$. Alguns dos resultados obtidos são apresentados nas tabelas E.2 e E.6 do anexo E.

4.2.2 Performance dos critérios de selecção de modelos

A performance de cada um dos critérios de selecção de modelos é avaliada pelas percentagens de classificação correcta das amostras geradas dos modelos m_1 e m_2 .

Analisa-se, em primeiro lugar, os resultados obtidos no **caso 1**.

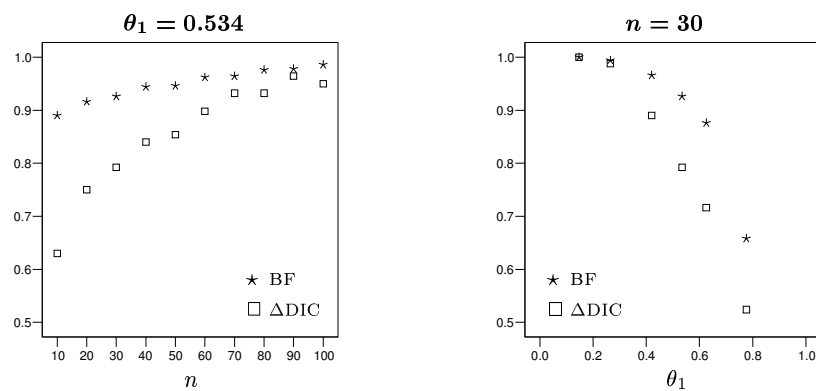
Na tabela 4.1 apresenta-se a percentagem de classificação correcta das **amostras geradas do modelo m_1** . Como seria de esperar, a percentagem de classificação correcta destas amostras, com qualquer um dos critérios de selecção de modelos considerados, aumenta à medida que a dimensão da amostra aumenta, e diminui à medida que θ_1 se aproxima de 1 pois, neste caso, a distribuição $\text{BiNeg}(1, \theta_1)$ pode confundir-se com a distribuição Poisson. Observa-se também que a performance do factor Bayes é superior à dos outros critérios e que, à medida que n aumenta ou θ_1 se aproxima de zero, as percentagens de classificação correcta dos restantes critérios aproximam-se entre si.

Alguns destes aspectos estão também patentes nos gráficos da figura 4.1, onde se representa a percentagem de classificação correcta das amostras geradas do modelo m_1 , obtida com o factor Bayes e com o ΔDIC , para $\theta_1 = 0.534$ e para $n = 30$. Os resultados obtidos com o ΔBIC e o $\Delta\widehat{\text{BIC}}$ não são representados, pois são muito idênticos aos do ΔDIC . Para $\theta_1 = 0.534$, regista-se que o factor Bayes tem um bom desempenho, enquanto o ΔDIC só têm uma boa performance para as amostras de dimensão mais elevada. No caso $n = 30$, destaca-se as frequências de classificação correcta mais reduzidas para valores de θ_1 próximos de 1.

Na tabela 4.2 apresenta-se a percentagem de classificação correcta das **amostras geradas do modelo m_2** . Como seria de esperar, a percentagem de classificação correcta destas amostras, com qualquer um dos critérios de selecção de modelos, aumenta à medida que a dimensão da amostra aumenta. Observa-se também que as performances do ΔBIC ,

Tabela 4.1: Percentagem de classificação correcta das amostras geradas de m_1 , para o caso 1.

θ_1	critério	dimensão da amostra									
		10	20	30	40	50	60	70	80	90	100
0.148	BF	95.8	99.6	100	100	100	100	100	100	100	100
	Δ BIC	93.8	99.4	100	100	100	100	100	100	100	100
	$\Delta\widehat{\text{BIC}}$	93.4	99.4	100	100	100	100	100	100	100	100
	Δ DIC	93.8	99.4	100	100	100	100	100	100	100	100
0.265	BF	94.6	96.4	99.4	99.4	100	99.8	100	100	100	100
	Δ BIC	81.4	92.4	98.8	99.4	99.6	99.8	100	100	100	100
	$\Delta\widehat{\text{BIC}}$	81.6	92.4	98.8	99.4	99.6	99.8	100	100	100	100
	Δ DIC	82.8	92.6	98.8	99.4	99.6	99.8	100	100	100	100
0.421	BF	92.4	93.8	96.6	98.0	98.6	99.2	99.0	99.4	99.6	99.8
	Δ BIC	63.8	82.2	88.4	94.0	95.8	98.2	98.8	98.0	99.2	99.8
	$\Delta\widehat{\text{BIC}}$	71.0	82.4	88.8	94.0	96.2	98.2	98.8	98.0	99.2	99.8
	Δ DIC	72.8	83.2	89.0	94.0	96.2	98.2	98.8	98.0	99.2	99.8
0.534	BF	89.0	91.6	92.6	94.4	94.6	96.2	96.4	97.6	97.8	98.6
	Δ BIC	52.8	71.8	78.2	83.8	85.0	89.8	93.0	92.8	96.4	95.0
	$\Delta\widehat{\text{BIC}}$	59.8	73.0	78.2	84.0	85.2	89.8	93.2	92.8	96.4	95.0
	Δ DIC	63.0	75.0	79.2	84.0	85.4	89.8	93.2	93.2	96.4	95.0
0.626	BF	84.4	87.8	87.6	88.0	90.0	91.6	94.6	92.6	95.2	95.6
	Δ BIC	42.2	55.4	68.4	71.6	75.6	80.4	84.8	84.2	88.6	88.8
	$\Delta\widehat{\text{BIC}}$	51.4	56.2	68.4	71.8	75.8	80.4	85.0	84.2	88.6	88.8
	Δ DIC	54.0	62.4	71.6	72.8	76.2	80.8	85.2	84.6	88.6	88.8
0.776	BF	58.4	67.0	65.8	60.8	68.4	71.8	71.0	71.4	73.6	78.0
	Δ BIC	26.0	50.6	50.8	53.6	57.6	64.4	64.4	67.8	67.6	72.8
	$\Delta\widehat{\text{BIC}}$	35.0	50.6	50.8	50.2	53.6	63.6	64.0	67.0	66.8	71.4
	Δ DIC	36.6	53.2	52.4	57.0	60.0	64.4	64.2	68.4	68.4	72.8

**Figura 4.1:** Percentagens de classificação correcta (no caso 1), das amostras geradas do modelo m_1 , para $\theta_1 = 0.534$ (esquerda) e para $n = 30$ (direita).

do $\Delta\widehat{\text{BIC}}$ e do ΔDIC são superiores à do factor Bayes, e que a performance do factor Bayes é mais influenciada pela dimensão da amostra do que a dos outros critérios. Além disso, à medida que n ou θ_2 aumentam, as performances dos critérios ΔBIC , $\Delta\widehat{\text{BIC}}$ e ΔDIC aproximam-se.

Estes aspectos são também evidenciados nos gráficos da figura 4.2, onde se representa a percentagem de classificação correcta das amostras geradas de m_2 , obtida com o factor Bayes e com o ΔDIC , para $\theta_2 = 1.128$ e para $n = 30$. No caso $\theta_2 = 1.128$, regista-se que a performance do factor Bayes é mais influenciada pela dimensão da amostra do que a do ΔDIC . Para $n = 30$, destaca-se a pior performance dos critérios de selecção de modelos para os valores de θ_2 próximos de zero.

Tabela 4.2: Percentagem de classificação correcta das amostras geradas de m_2 , para o caso 1.

θ_2	critério	dimensão da amostra									
		10	20	30	40	50	60	70	80	90	100
0.309	BF	40.4	55.4	59.2	62.6	62.8	65.4	69.0	70.6	73.8	74.2
	ΔBIC	86.8	79.4	78.8	80.0	77.2	78.2	80.8	82.0	82.2	82.4
	$\Delta\widehat{\text{BIC}}$	77.0	79.4	78.8	81.6	79.4	79.8	80.8	82.2	82.4	82.6
	ΔDIC	75.2	76.2	77.8	77.6	76.0	77.8	80.8	81.4	81.4	82.4
0.632	BF	26.8	49.2	57.2	64.6	73.0	75.8	78.8	81.6	84.4	88.6
	ΔBIC	86.6	84.6	84.8	87.0	90.8	91.0	90.8	93.8	94.4	94.4
	$\Delta\widehat{\text{BIC}}$	77.0	83.2	84.8	86.2	90.8	90.8	90.8	93.6	94.4	94.4
	ΔDIC	72.4	79.8	82.2	85.2	90.4	90.6	90.4	93.4	94.4	94.4
1.128	BF	45.0	67.6	79.0	84.6	91.8	91.4	94.2	96.4	98.0	98.0
	ΔBIC	87.4	90.8	91.6	95.2	97.0	96.4	98.4	98.6	98.6	99.4
	$\Delta\widehat{\text{BIC}}$	84.2	90.4	91.6	95.0	96.8	96.4	98.2	98.6	98.6	99.2
	ΔDIC	82.6	90.2	91.4	94.8	96.8	96.4	98.0	98.6	98.6	99.2
1.465	BF	58.2	75.0	87.6	90.2	94.8	95.4	97.8	98.6	99.4	99.8
	ΔBIC	89.6	93.2	95.4	97.0	97.2	98.0	99.2	99.0	99.8	100
	$\Delta\widehat{\text{BIC}}$	87.6	92.8	95.4	97.0	97.2	98.0	99.2	99.0	99.8	100
	ΔDIC	87.2	92.0	95.2	97.0	97.2	97.8	99.0	99.0	99.8	100
1.628	BF	64.2	81.2	92.2	96.6	97.8	98.6	98.8	99.4	99.8	99.4
	ΔBIC	90.2	94.0	97.6	98.2	99.4	99.6	99.2	99.8	99.8	99.8
	$\Delta\widehat{\text{BIC}}$	88.6	93.6	97.6	98.0	99.4	99.6	99.2	99.8	99.8	99.8
	ΔDIC	86.4	93.4	97.6	98.0	99.4	99.6	99.2	99.8	99.8	99.8
2.741	BF	91.4	96.2	99.0	99.8	99.8	100	100	100	100	100
	ΔBIC	96.0	98.2	99.8	100	100	100	100	100	100	100
	$\Delta\widehat{\text{BIC}}$	96.0	98.2	99.8	100	100	100	100	100	100	100
	ΔDIC	96.0	98.0	99.8	100	100	100	100	100	100	100

As percentagens de classificação correcta das amostras geradas de m_1 e m_2 , para o **caso 2**, são apresentadas, respectivamente, nas tabelas E.1 e E.2 do anexo E. Destaca-se que, no caso dos dados gerados de m_2 , e exceptuando os valores mais pequenos de n e θ_2 , os quatro critérios têm performances bastante idênticas. Não se regista outras diferenças relevantes entre os casos 1 e 2.

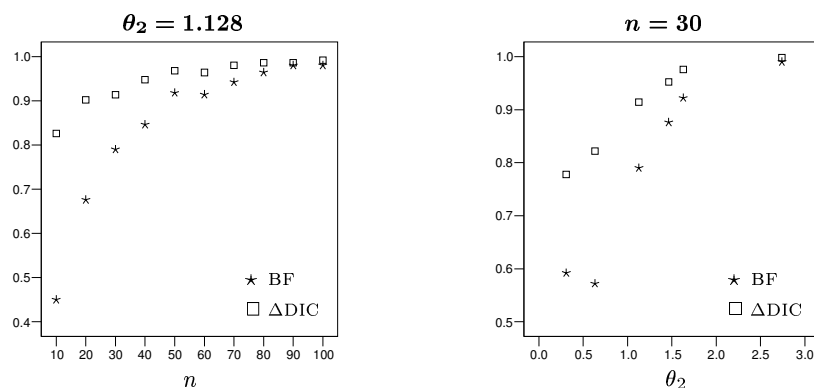


Figura 4.2: Percentagens de classificação correcta (no caso 1), das amostras geradas do modelo m_2 , para $\theta_2 = 1.128$ (esquerda) e para $n = 30$ (direita).

Os resultados obtidos sobre a performance dos critérios de selecção de modelos justificam que se faça alguns comentários.

A distribuição $\text{BiNeg}(1, \theta_1)$, com θ_1 perto de 1, confunde-se com a distribuição $\text{Poi}(\theta_2)$, com θ_2 próximo de 0. Por este motivo, no caso de dados gerados de m_1 com valores de θ_1 próximos de 1, e no caso de dados gerados de m_2 com valores de θ_2 próximos de zero, a percentagem de classificação correcta dos critérios de selecção de modelos é menor.

Neste estudo de simulação, compara-se a performance do factor Bayes com a do critério de informação da *deviance*. Conforme referem Berg *et al.* (2002), este facto pode ser criticável na medida em que estes dois critérios não estão a avaliar os mesmos aspectos do modelo. Do mesmo modo, Spiegelhalter *et al.* (2002, resposta aos avaliadores) alertam contra o uso indevido do factor Bayes como um modo de validar os resultados obtidos com o ΔDIC : “*We emphasize that these techniques are intended to answer different questions and cannot be expected to give the same conclusions*”.

Na realidade, o factor Bayes dá indicação de qual o modelo cuja distribuição preditiva *a priori* tem maior capacidade para prever os dados observados, enquanto o ΔDIC possibilita a identificação do modelo cuja distribuição preditiva *a posteriori* tem maior capacidade para prever observações futuras, geradas pelo mesmo “mecanismo” que deu origem aos dados observados. A perspectiva “preditiva *a posteriori*” do ΔDIC é, intuitivamente, mais apelativa, pois na maior parte das aplicações práticas o objectivo final do modelo seleccionado é a predição de observações futuras, obtidas nas mesmas condições dos dados observados. Refira-se que o $\widehat{\Delta\text{BIC}}$ partilha a mesma perspectiva “preditiva *a posteriori*” do ΔDIC .

4.2.3 Valores das medidas de selecção de modelos

Os valores das medidas de selecção de modelos, obtidas com o ajustamento de m_1 e m_2 aos dados gerados de cada um destes modelos, são também analisados. Para o caso 1, as médias e variâncias dos valores destas medidas, referentes às amostras geradas de m_1 e m_2 são apresentadas, respectivamente, nas tabelas E.3 e E.4 do anexo E; para o caso 2, os resultados obtidos são apresentados nas tabelas E.5 e E.6. A realização de alguns gráficos possibilita um melhor entendimento dos resultados obtidos.

Na interpretação dos resultados considera-se apenas o **caso 1**, uma vez que o caso 2 não apresenta diferenças relevantes. Começa por analisar-se os resultados obtidos com as **amostras geradas do modelo m_1** .

Com o objectivo de ilustrar o “comportamento” dos valores do DIC e do $\widehat{\text{BIC}}$ considera-se os dados gerados de m_1 com $\theta_1 = 0.534$. Na tabela 4.3 estão listados os valores das médias e desvios padrão destas medidas de selecção de modelos, obtidos com o ajustamento de m_1 e m_2 . Como seria de esperar, regista-se que as médias e os desvios padrão do DIC e do $\widehat{\text{BIC}}$ aumentam com a dimensão da amostra, e que os valores obtidos para as médias destas medidas de discrepância referentes ao ajustamento de m_1 são inferiores aos obtidos com o ajustamento de m_2 (devido ao facto dos dados serem gerados de m_1).

Tabela 4.3: Médias e desvios padrão do DIC e do $\widehat{\text{BIC}}$, obtidos com o ajustamento dos modelos m_1 e m_2 , a dados gerados de m_1 com $\theta_1 = 0.534$, para o caso 1.

n	DIC_{m_1}		DIC_{m_2}		$\widehat{\text{BIC}}_{m_1}$		$\widehat{\text{BIC}}_{m_2}$	
	méd.	d.p.	méd.	d.p.	méd.	d.p.	méd.	d.p.
10	26.9	6.1	28.4	8.0	28.5	6.0	29.7	8.0
20	52.4	9.1	56.1	12.1	54.6	9.1	58.1	12.1
30	78.5	10.7	83.2	14.0	81.0	10.7	85.6	14.0
40	104.7	12.3	111.9	16.9	107.4	12.3	114.6	16.9
50	129.7	13.8	138.3	18.8	132.7	13.8	141.2	18.8
60	156.9	15.4	167.4	20.9	160.0	15.4	170.5	20.9
70	180.7	17.2	193.6	23.1	184.0	17.2	196.8	23.1
80	207.2	16.3	221.9	22.2	210.6	16.3	225.3	22.2
90	234.3	18.7	251.2	25.5	237.8	18.7	254.7	25.5
100	258.4	19.2	276.3	26.6	262.1	19.2	280.0	26.6

As quantidades apresentadas na tabela 4.3 sugerem que as médias dos valores do DIC e do $\widehat{\text{BIC}}$ aumentam linearmente com a dimensão da amostra. Além disso, uma análise mais atenta, evidencia que os desvios padrão dos valores do DIC e do $\widehat{\text{BIC}}$ aumentam linearmente com \sqrt{n} . Deste modo, deve ter-se em conta a dimensão n quando se analisa estas quantidades.

Divide-se as médias dos valores do DIC e do $\widehat{\text{BIC}}$ por n e os desvios padrão destes valores por \sqrt{n} ; os resultados obtidos são apresentados na tabela 4.4. Os valores obtidos, tanto para o DIC como para o $\widehat{\text{BIC}}$, mantêm-se aproximadamente constantes, independentemente da dimensão da amostra. Para ilustrar esta situação representa-se, na figura 4.3, as médias/ $n \pm$ desvios padrão/ \sqrt{n} dos valores do DIC, referentes ao ajustamento de cada um dos modelos a dados gerados do modelo m_1 com $\theta_1 = 0.534$, *versus* a dimensão da amostra.

Na figura 4.4 representa-se graficamente as médias dos valores do DIC, obtidas com o ajustamento de m_1 e m_2 aos dados gerados do modelo m_1 , *versus* θ_1 , para diversos valores de n . Os gráficos para as medidas de selecção de modelos BIC e $\widehat{\text{BIC}}$ não são apresentados, uma vez que são idênticos aos obtidos para o DIC. Da análise desta figura e da tabela E.3 conclui-se que:

Tabela 4.4: Médias/ n e desvios padrão/ \sqrt{n} do DIC e do $\widehat{\text{BIC}}$, obtidos com o ajustamento dos modelos m_1 e m_2 , a dados gerados de m_1 com $\theta_1 = 0.534$, para o caso 1.

n	DIC $_{m_1}$		DIC $_{m_2}$		$\widehat{\text{BIC}}_{m_1}$		$\widehat{\text{BIC}}_{m_2}$	
	méd. n	d.p. \sqrt{n}	méd. n	d.p. \sqrt{n}	méd. n	d.p. \sqrt{n}	méd. n	d.p. \sqrt{n}
10	2.7	1.9	2.8	2.5	2.8	1.9	3.0	2.5
20	2.6	2.0	2.8	2.7	2.7	2.0	2.9	2.7
30	2.6	2.0	2.8	2.6	2.7	1.9	2.9	2.6
40	2.6	2.0	2.8	2.7	2.7	2.0	2.9	2.7
50	2.6	2.0	2.8	2.7	2.7	1.9	2.8	2.7
60	2.6	2.0	2.8	2.7	2.7	2.0	2.8	2.7
70	2.6	2.1	2.8	2.8	2.6	2.1	2.8	2.8
80	2.6	1.8	2.8	2.5	2.6	1.8	2.8	2.5
90	2.6	2.0	2.8	2.7	2.6	2.0	2.8	2.7
100	2.6	1.9	2.8	2.7	2.6	1.9	2.8	2.7

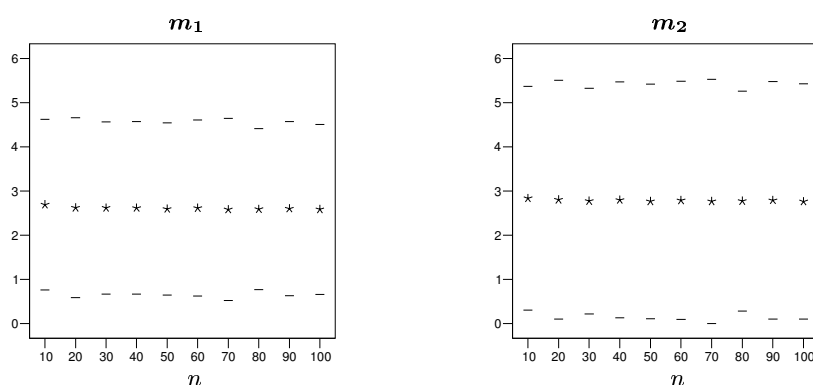


Figura 4.3: Médias/ $n \pm$ desvios padrão/ \sqrt{n} dos valores do DIC (no caso 1), referentes ao ajustamento dos modelos m_1 (esquerda) e m_2 (direita) a dados gerados do modelo m_1 com $\theta_1 = 0.534$, *versus* n .

- as médias dos valores do DIC, BIC e $\widehat{\text{BIC}}$ obtidas com o ajustamento de m_1 e m_2
 - aumentam linearmente com a dimensão da amostra, e
 - diminuem à medida que θ_1 aumenta. Não esquecer que X_i é a variável aleatória que representa o número de insucessos até à ocorrência de 1 sucesso, sendo a probabilidade de sucesso, θ_1 , constante. Assim, se θ_1 for elevado, o número de insucessos até à ocorrência de 1 sucesso é pequeno. Consequentemente, a amostra gerada de m_1 é essencialmente constituída pelas observações 0, 1, 2, o que se reflecte nos valores das medidas de selecção de modelos.
- as médias dos valores do DIC, BIC e $\widehat{\text{BIC}}$ referentes ao ajustamento de m_1 são inferiores às do modelo m_2 , devido ao facto dos dados serem gerados de m_1 .
- as variâncias dos valores do BIC, $\widehat{\text{BIC}}$ e DIC, obtidas com o ajustamento de cada um dos modelos, são muito idênticas.

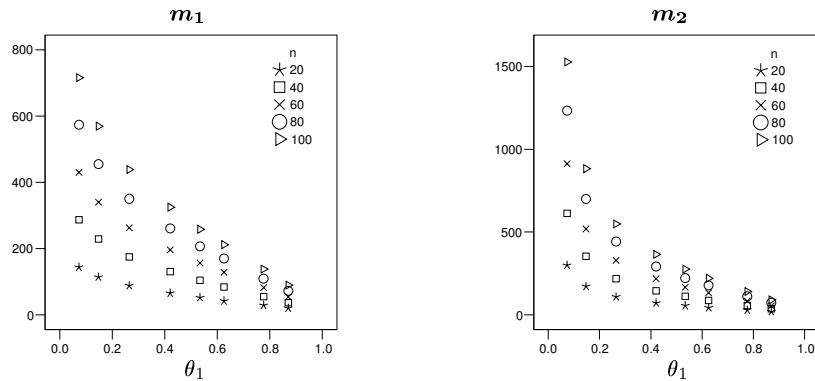


Figura 4.4: Médias dos valores do DIC (no caso 1), obtidas com o ajustamento de m_1 (esquerda) e m_2 (direita) a dados gerados do modelo m_1 , versus θ_1 , para diferentes valores de n .

Na figura 4.5 representa-se graficamente as médias dos valores do DIC, obtidas com o ajustamento de m_1 e m_2 aos dados gerados do modelo m_1 , versus a dimensão da amostra, para $\theta_1 = 0.073$ e para $\theta_1 = 0.421$. Destaca-se que à medida que θ_1 se aproxima de 1, as médias dos valores do DIC obtidas com o ajustamento de m_1 e m_2 aproximam-se, sendo mais difícil a identificação do modelo correcto. A selecção do modelo correcto é mais evidente para amostras de dimensão mais elevada.

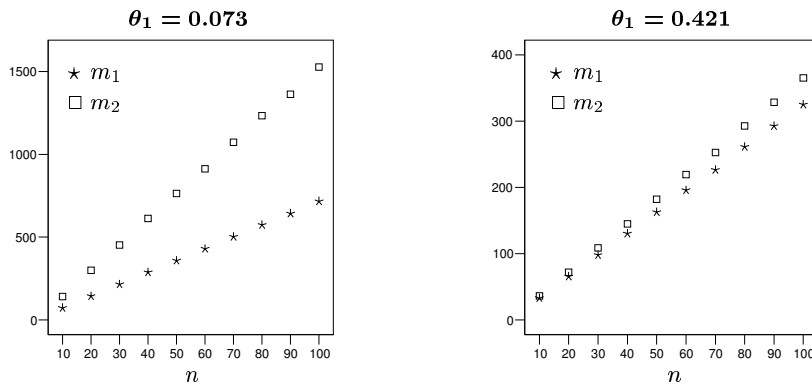


Figura 4.5: Médias dos valores do DIC (no caso 1), obtidas com o ajustamento de m_1 e m_2 a dados gerados do modelo m_1 , versus n , para $\theta_1 = 0.073$ (esquerda) e para $\theta_1 = 0.421$ (direita).

Analisa-se, agora, os resultados obtidos no **caso 1**, com as **amostras geradas do modelo m_2** .

Começa por considerar-se os dados gerados do modelo m_2 com $\theta_2 = 1.128$, com os quais se ilustra o “comportamento” do DIC e do \widehat{BIC} . Na tabela 4.5 estão listados os valores das médias e desvios padrão destas medidas de selecção de modelos, obtidos com o ajustamento de m_1 e m_2 . Na tabela 4.6 apresenta-se as médias divididas por n e os desvios padrão divididos por \sqrt{n} dos valores do DIC e do \widehat{BIC} ; tanto para o DIC como para o \widehat{BIC} , os valores obtidos mantêm-se aproximadamente constantes, independentemente da dimensão da amostra. Na figura 4.6 representa-se as médias/ $n \pm$ desvios padrão/ \sqrt{n} dos valores do DIC, referentes ao ajustamento de cada um dos modelos a dados gerados do modelo m_2 com $\theta_2 = 1.128$, versus n .

Tabela 4.5: Médias e desvios padrão do DIC e do $\widehat{\text{BIC}}$, obtidos com o ajustamento dos modelos m_1 e m_2 , a dados gerados de m_2 com $\theta_2 = 1.128$, para o caso 1.

n	DIC_{m_1}		DIC_{m_2}		$\widehat{\text{BIC}}_{m_1}$		$\widehat{\text{BIC}}_{m_2}$	
	méd.	d.p.	méd.	d.p.	méd.	d.p.	méd.	d.p.
10	30.7	4.2	28.6	4.0	32.2	4.2	29.9	4.0
20	60.1	6.3	56.0	5.9	62.2	6.3	58.0	5.9
30	89.8	7.4	83.6	7.4	92.3	7.4	86.0	7.4
40	119.3	8.7	111.5	8.4	122.1	8.7	114.1	8.4
50	149.2	9.4	138.8	9.3	152.1	9.4	141.7	9.3
60	178.5	10.4	166.5	10.2	181.6	10.4	169.6	10.2
70	207.5	10.8	193.5	10.9	210.8	10.8	196.8	10.9
80	236.2	12.3	220.1	11.5	239.6	12.3	223.5	11.5
90	266.1	12.4	248.1	11.7	269.6	12.4	251.6	11.7
100	295.6	13.8	275.8	12.7	299.2	13.8	279.4	12.7

Tabela 4.6: Médias/ n e desvios padrão/ \sqrt{n} do DIC e do $\widehat{\text{BIC}}$, obtidos com o ajustamento dos modelos m_1 e m_2 , a dados gerados de m_2 com $\theta_2 = 1.128$, para o caso 1.

n	DIC_{m_1}		DIC_{m_2}		$\widehat{\text{BIC}}_{m_1}$		$\widehat{\text{BIC}}_{m_2}$	
	$\frac{\text{méd.}}{n}$	$\frac{\text{d.p.}}{\sqrt{n}}$	$\frac{\text{méd.}}{n}$	$\frac{\text{d.p.}}{\sqrt{n}}$	$\frac{\text{méd.}}{n}$	$\frac{\text{d.p.}}{\sqrt{n}}$	$\frac{\text{méd.}}{n}$	$\frac{\text{d.p.}}{\sqrt{n}}$
10	3.1	1.3	2.9	1.3	3.2	1.3	3.0	1.3
20	3.0	1.4	2.8	1.3	3.1	1.4	2.9	1.3
30	3.0	1.3	2.8	1.3	3.1	1.3	2.9	1.3
40	3.0	1.4	2.8	1.3	3.1	1.4	2.9	1.3
50	3.0	1.3	2.8	1.3	3.0	1.3	2.8	1.3
60	3.0	1.3	2.8	1.3	3.0	1.3	2.8	1.3
70	3.0	1.3	2.8	1.3	3.0	1.3	2.8	1.3
80	3.0	1.4	2.8	1.3	3.0	1.4	2.8	1.3
90	3.0	1.3	2.8	1.2	3.0	1.3	2.8	1.2
100	3.0	1.4	2.8	1.3	3.0	1.4	2.8	1.3

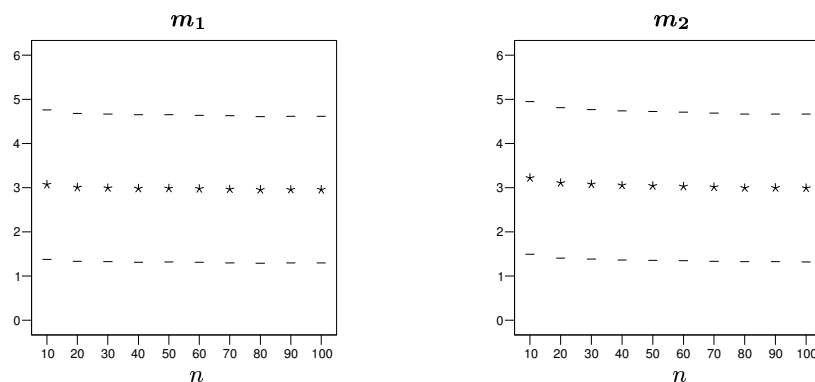


Figura 4.6: Médias/ $n \pm$ desvios padrão/ \sqrt{n} dos valores do DIC (no caso 1), referentes ao ajustamento dos modelos m_1 (esquerda) e m_2 (direita) a dados gerados do modelo m_2 com $\theta_2 = 1.128$, versus n .

Na figura 4.7 representa-se graficamente as médias dos valores do DIC, obtidas com o ajustamento de m_1 e m_2 aos dados gerados do modelo m_2 , para diversos valores de n e θ_2 ; os correspondentes gráficos obtidos com as medidas BIC e \widehat{BIC} são idênticos. Da análise desta figura e da tabela E.4 destaca-se que:

- as médias dos valores do DIC, BIC e \widehat{BIC} obtidas com o ajustamento dos modelos m_1 e m_2
 - aumentam linearmente com a dimensão da amostra, e
 - aumentam à medida que θ_2 aumenta. Neste caso, como as amostras são geradas de m_2 , as observações são provenientes de uma distribuição Poisson com valor médio θ_2 . Deste modo, se θ_2 for elevado, os valores das observações são também, de um modo geral, elevados, o que se reflecte nos valores das medidas de selecção de modelos.
- as médias dos valores do DIC, BIC e \widehat{BIC} referentes ao ajustamento do modelo m_2 são inferiores às do modelo m_1 , devido ao facto de os dados serem gerados de m_2 .
- as variâncias dos valores do BIC, \widehat{BIC} e DIC, obtidas com o ajustamento de cada um dos modelos, são muito idênticas.

Na figura 4.8 representa-se graficamente as médias dos valores do DIC obtidas com o ajustamento dos modelos m_1 e m_2 aos dados gerados do modelo m_2 , *versus* a dimensão da amostra, com $\theta_2 = 0.632$ e $\theta_2 = 2.741$. Destaca-se que, à medida que o valor de θ_2 aumenta, as médias dos valores do DIC referentes ao ajustamento dos modelos m_1 e m_2 distanciam-se, sendo mais fácil a selecção do modelo correcto. A identificação do modelo correcto é também mais evidente para as amostras de dimensão mais elevada.

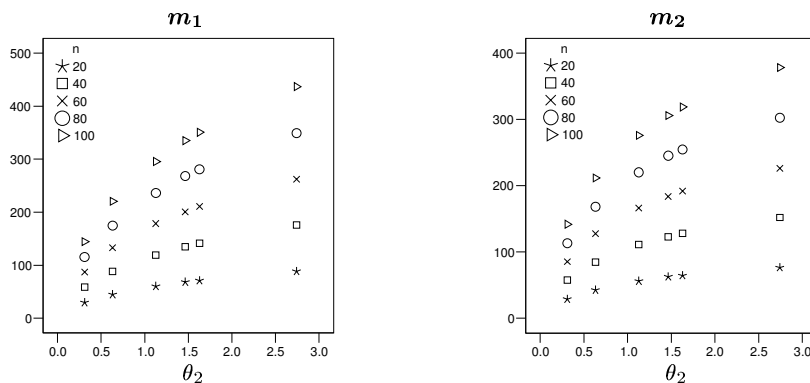


Figura 4.7: Médias dos valores do DIC (no caso 1), obtidas com o ajustamento de m_1 (esquerda) e m_2 (direita) a dados gerados do modelo m_2 , *versus* θ_2 , para diferentes valores de n .

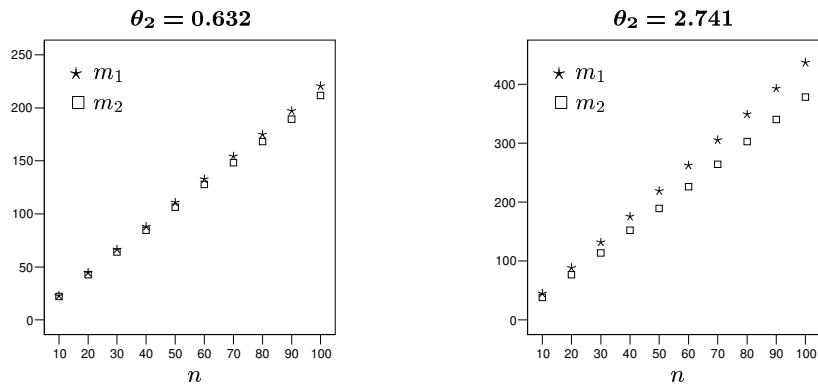


Figura 4.8: Médias dos valores do DIC (no caso 1), obtidas com o ajustamento de m_1 e m_2 a dados gerados do modelo m_2 , versus n , para $\theta_2 = 0.632$ (esquerda) e para $\theta_2 = 2.741$ (direita).

Após a análise dos valores das medidas de selecção de modelos é oportuno efectuar os comentários que se apresenta de seguida.

Conforme já foi referido, Zhu e Carlin (2000) sugerem que se efectue a replicação dos valores do DIC de modo a estimar $Var[\text{DIC}]$ através da sua variância amostral. Estes autores referem que, para já, este é o modo mais adequado para obter uma indicação da variância do DIC. No caso particular deste estudo de simulação, em que se compara os modelos beta/binomial negativo e gama/Poisson, regista-se que as variâncias amostrais obtidas por replicação dos valores do DIC e do $\widehat{\text{BIC}}$ são muito idênticas. Este facto sugere que, quando se utiliza os métodos MCMC na comparação de modelos, pode estimar-se a variância do DIC à custa da variância do $\widehat{\text{BIC}}$, a qual pode ser facilmente obtida quando se utiliza este método de simulação estocástica.

A sugestão de utilização da variância do $\widehat{\text{BIC}}$ para estimar a variância do DIC, quando se utiliza os métodos MCMC na comparação de modelos, carece de um estudo mais aprofundado: ver o que acontece quando se compara outro tipo de modelos — o caso da selecção de covariáveis em modelos de regressão vai ser visto na secção 4.3 — e, em particular, analisar o que se passa quando os modelos em comparação tiverem mais níveis de hierarquia.

Na selecção de modelos é necessário ter em conta qual o propósito final do modelo, de modo a utilizar o critério mais adequado. Se o objectivo do modelo for a predição de observações futuras, obtidas nas mesmas condições dos dados observados, então o DIC e o $\widehat{\text{BIC}}$ são as medidas de selecção de modelos mais adequadas. Caso não se pretenda efectuar a predição de observações futuras, mas apenas encontrar o modelo que melhor descreve um determinado conjunto de dados, então deve dar-se preferência ao factor Bayes.

Qualquer que seja o critério de selecção de modelos utilizado, a escolha do modelo pode ser mais, ou menos, óbvia, dependendo da dimensão da amostra.

Uma eventual medida de adequabilidade do modelo, construída à custa do DIC ou do $\widehat{\text{BIC}}$, deve ter em conta que o valor destas medidas deve ser dividido por n .

4.3 Dados de regressão linear normal

4.3.1 Planeamento da simulação

A avaliação e a comparação da performance dos critérios ΔDIC , $\Delta\widehat{\text{BIC}}$ e ΔLLP , em problemas de selecção de covariáveis, é efectuada recorrendo à simulação de dados de regressão linear normal. Alguns aspectos do planeamento da simulação são inspirados num estudo efectuado por Kuo e Mallick (1998, exemplo 4.1.1).

Considera-se amostras de 4 possíveis covariáveis, x_1, x_2, x_3 e x_4 , de dimensão n , cujas observações são obtidas por geração de normais padrão independentes. Com excepção da covariável x_3 nos casos 3 e 4 apresentados mais à frente, as observações das covariáveis são mantidas fixas em todo o processo de simulação.

Seja δ_j o indicador da presença da covariável x_j no modelo, isto é, $\delta_j = 1$ se a covariável x_j está incluída no modelo, e $\delta_j = 0$ caso contrário, para $j = 1, \dots, 4$. Os modelos possíveis são identificados por m_1, \dots, m_{16} , e as covariáveis incluídas em cada um deles estão listadas na tabela 4.7.

Simula-se uma amostra para o erro aleatório ϵ , de dimensão n , constituída por observações independentes e identicamente distribuídas $N(0, 2.5^2)$. Com esta amostra do erro aleatório, e para cada uma das possíveis combinações de covariáveis incluídas no modelo, obtém-se amostras da variável resposta Y de cada um dos modelos m_1, \dots, m_{16} . Ajusta-se todos os modelos possíveis, m_1, \dots, m_{16} , à amostra da variável resposta Y gerada do modelo m_k , com $k = 1, \dots, 16$. Este procedimento é repetido $T = 1000$ vezes, isto é, simula-se, ao todo, 1000 amostras do erro aleatório ϵ , e repete-se o esquema descrito. O procedimento utilizado na simulação está resumido no algoritmo 4.2.

Tabela 4.7: Modelos de regressão linear para um conjunto de 4 covariáveis.

modelo	k	δ_1	δ_2	δ_3	δ_4	covariáveis
m_1	1	1	1	1	1	x_1, x_2, x_3, x_4
m_2	2	1	1	1	0	x_1, x_2, x_3
m_3	3	1	1	0	1	x_1, x_2, x_4
m_4	4	1	0	1	1	x_1, x_3, x_4
m_5	5	0	1	1	1	x_2, x_3, x_4
m_6	6	1	1	0	0	x_1, x_2
m_7	7	1	0	1	0	x_1, x_3
m_8	8	1	0	0	1	x_1, x_4
m_9	9	0	1	1	0	x_2, x_3
m_{10}	10	0	1	0	1	x_2, x_4
m_{11}	11	0	0	1	1	x_3, x_4
m_{12}	12	1	0	0	0	x_1
m_{13}	13	0	1	0	0	x_2
m_{14}	14	0	0	1	0	x_3
m_{15}	15	0	0	0	1	x_4
m_{16}	16	0	0	0	0	nenhuma

Algoritmo 4.2

1. Para $t = 1, 2, \dots, 1000$,
 - (a) simular a t -ésima amostra do erro aleatório ϵ ;
 - (b) para $k = 1, 2, \dots, 16$,
 - i. obter a t -ésima amostra da variável resposta Y gerada do k -ésimo modelo de regressão linear normal;
 - ii. ajustar os 16 modelos de regressão linear normal possíveis à amostra gerada do k -ésimo modelo, e calcular o DIC, o $\widehat{\text{BIC}}$ e o LLP de cada um dos modelos ajustados;
 - iii. determinar qual o modelo ajustado que minimiza cada uma das medidas de discrepância, isto é, determinar qual o modelo seleccionado por cada uma delas. ■

O procedimento descrito no algoritmo 4.2 é repetido em 6 casos distintos:

Caso 1. As amostras das covariáveis x_1, x_2, x_3 e x_4 e do erro aleatório ϵ têm dimensão $n = 200$, e as amostras da variável resposta Y são obtidas de acordo com o modelo

$$Y = \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \delta_4 x_4 + \epsilon.$$

Caso 2. Difere do caso 1 apenas pelo facto das amostras da variável resposta Y serem obtidas de acordo com o modelo

$$Y = \delta_1 x_1 + 10\delta_2 x_2 + 2\delta_3 x_3 + 5\delta_4 x_4 + \epsilon.$$

Caso 3. Difere do caso 1 apenas porque a covariável x_3 é obtida à custa da igualdade $x_3 = x_1 + 1.7z$, sendo z uma amostra da $N(0, 1)$. Deste modo é induzida uma correlação linear moderada entre as covariáveis x_1 e x_3 ($r_{x_1, x_3} = 0.51$).

Caso 4. Difere do caso 1 apenas porque a covariável x_3 é obtida à custa da igualdade $x_3 = x_1 + 0.2z$, sendo z uma amostra da $N(0, 1)$. Deste modo é induzida uma correlação linear elevada entre as covariáveis x_1 e x_3 ($r_{x_1, x_3} = 0.98$).

Caso 5. Difere do caso 1 apenas pelo facto das amostras terem dimensão $n = 50$.

Caso 6. Difere do caso 1 apenas pelo facto das amostras terem dimensão $n = 100$.

Com o caso 2 pretende averiguar-se qual o impacto produzido no número de vezes que cada modelo é seleccionado, e nos valores do DIC, $\widehat{\text{BIC}}$ e LLP pelo facto dos coeficientes das covariáveis não serem todos iguais. A eventual ocorrência de alterações na presença de correlação linear moderada e extrema entre as covariáveis pode ser estudada, respectivamente, com os casos 3 e 4. A comparação dos casos 5 e 6 com o caso 1 permite tirar conclusões sobre a performance e as alterações produzidas nos valores das medidas de discrepância, com a variação da dimensão da amostra.

Resumindo, a execução do algoritmo 4.2 nos 6 casos enunciados permite:

- comparar e avaliar a performance dos critérios ΔDIC , $\Delta\widehat{\text{BIC}}$ e ΔLLP na selecção de covariáveis em modelos de regressão linear normal em diversas situações;
- identificar os factores que influenciam os valores do DIC, do $\widehat{\text{BIC}}$ e do LLP.

Avalia-se também a performance do método GVS na selecção de covariáveis em modelos de regressão, nos 6 casos apresentados. Como a replicação da experiência não é facilmente exequível, é apenas efectuada uma réplica.

4.3.2 Modelo de regressão linear normal

O modelo de regressão linear normal, apresentado na secção 3.5.1, pode ser representado em notação matricial por $Y = Xb + \epsilon$, onde $Y = (Y_1, Y_2, \dots, Y_n)'$, $b = (b_0, b_1, \dots, b_p)'$, X é a matriz $n \times (p + 1)$ de especificação do modelo associada ao vector b e $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n) \sim N_n(0, \sigma^2 I_n)$, sendo I_n a matriz identidade $n \times n$. Assim,

$$Y | b, \sigma^2, X \sim N_n(Xb, \sigma^2 I_n),$$

sendo a sua função densidade de probabilidade¹

$$f(y | b, \sigma^2) \propto (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (y - Xb)'(y - Xb) \right\},$$

para $y = (y_1, y_2, \dots, y_n)'$. Se a matriz X tiver característica $p + 1$ então

$$f(y | b, \sigma^2) \propto (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} [ks^2 + (b - \hat{b})'X'X(b - \hat{b})] \right\},$$

onde

$$s^2 = \frac{1}{k} (y - X\hat{b})'(y - X\hat{b}) \quad (4.1)$$

é a soma de quadrados residual, sendo $k = n - p - 1$, e

$$\hat{b} = (X'X)^{-1}X'y \quad (4.2)$$

é a estimativa de mínimos quadrados de b .

Considera-se a distribuição *a priori* não informativa para o vector de parâmetros do modelo, $h(b, \sigma^2) = (\sigma^2)^{-1}$, para $b \in \mathbb{R}^{p+1}$ e $\sigma^2 > 0$, obtida sobre o pressuposto de independência *a priori* entre b e σ^2 por aplicação da regra de Jeffreys para especificação das distribuições *a priori* marginais. Assim, o núcleo da densidade conjunta *a posteriori* de (b, σ^2) é dado por

$$\begin{aligned} h(b, \sigma^2 | y) &\propto (\sigma^2)^{-1} (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} [ks^2 + (b - \hat{b})'X'X(b - \hat{b})] \right\} \\ &\propto (\sigma^2)^{-(\frac{k}{2}+1)} \exp \left\{ -\frac{ks^2}{2\sigma^2} \right\} \times (\sigma^2)^{-\frac{p+1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} [(b - \hat{b})'X'X(b - \hat{b})] \right\}. \end{aligned}$$

¹Para simplificar a notação, suprime-se aqui e na notação subsequente a dependência em X .

Deste modo, a distribuição *a posteriori* do vector de parâmetros (b, σ^2) pode ser factorizada, $h(b, \sigma^2 | y) = h(\sigma^2 | y) h(b | \sigma^2, y)$, com

$$h(b | \sigma^2, y) \propto (\sigma^2)^{-\frac{p+1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[(b - \hat{b})' X' X (b - \hat{b}) \right] \right\},$$

e

$$h(\sigma^2 | y) \propto (\sigma^2)^{-\left(\frac{k}{2} + 1\right)} \exp \left\{ -\frac{ks^2}{2\sigma^2} \right\}.$$

Assim, a distribuição *a posteriori* condicional do vector de parâmetros b , dado σ^2 , é normal multivariada de dimensão $p + 1$,

$$b | \sigma^2, y \sim N_{p+1}(\hat{b}, \sigma^2 V_b),$$

com

$$V_b = (X' X)^{-1}, \quad (4.3)$$

e a distribuição *a posteriori* marginal de σ^2 é qui-quadrado invertida,

$$\sigma^2 | y \sim \chi^2\text{-Inv}(k, s^2).$$

Mais pormenores sobre a análise bayesiana do modelo de regressão linear normal podem ser vistos em Gelman *et al.* (1995, secção 8.3) e Paulino *et al.* (2003, secção 4.3).

Como é possível determinar analiticamente a distribuição *a posteriori* $h(b, \sigma^2 | y)$, pode facilmente obter-se uma amostra da distribuição *a posteriori* de (b, σ^2) . Assim, caso se pretenda fazer inferência por simulação utiliza-se o método de Monte Carlo directo; para caracterizar a distribuição *a posteriori* conjunta de (b, σ^2) , gera-se amostras de σ^2 e depois de $b | \sigma^2$. O procedimento de geração de amostras da distribuição $h(b, \sigma^2 | y)$ pode ser resumido do seguinte modo:

1. calcula-se \hat{b} através da expressão (4.2) e V_b através da expressão (4.3);
2. calcula-se s^2 , através da expressão (4.1);
3. gera-se σ^2 da distribuição $\chi^2\text{-Inv}(k, s^2)$;
4. condicional ao valor de σ^2 simulado, gera-se b da distribuição normal multivariada $N_{p+1}(\hat{b}, \sigma^2 V_b)$.

Para que a simulação de amostras da distribuição *a posteriori* de (b, σ^2) seja computacionalmente eficiente é necessário (*e.g.*, Gelman *et al.*, 1995, pág. 238): (i) Determinar a factorização QR , $X = QR$, onde Q é uma matriz ortogonal $n \times (p + 1)$, $Q'Q = I$, e R é uma matriz $(p + 1) \times (p + 1)$ triangular superior. (ii) Calcular R^{-1} , o que é bastante fácil uma vez que R é uma matriz triangular superior. R^{-1} é o factor de Cholesky da matriz de covariância V_b , pois $R^{-1}(R^{-1})' = (X' X)^{-1} = V_b$. (iii) Calcular \hat{b} , resolvendo o sistema linear $R\hat{b} = Q'y$ e usando o facto de R ser uma matriz triangular superior.

A simulação de um valor para σ^2 da distribuição $\chi^2\text{-Inv}(k, s^2)$ pode ser feita simulando um valor λ da distribuição χ_k^2 e fazendo $\sigma^2 = k s^2 / \lambda$ (*e.g.*, Gelman *et al.*, 1995, pág. 480).

Depois de simular σ^2 , b pode ser facilmente simulado da distribuição normal multivariada apropriada, usando o factor de Cholesky e a geração de amostras normais padrão independentes (*e.g.*, Gelman *et al.*, 1995, pág. 478).

Os programas em Fortran para execução do algoritmo 4.2 nos casos 1 a 6 designam-se por `selcov_reglin_caso1`, ..., `selcov_reglin_caso6`, respectivamente, e são apresentados em ficheiros com o mesmo nome, no CD anexo a este documento. São utilizadas algumas subrotinas do Fortran², nomeadamente R2SE que possibilita o ajustamento do modelo de regressão linear múltipla por mínimos quadrados, RCOVB que calcula a matriz de covariância dos coeficientes de regressão estimados, CHFAC que permite determinar uma factorização triangular superior de uma matriz simétrica definida positiva, e RNNOR e RNCHI que possibilitam a geração de números pseudo-aleatórios com distribuição, respectivamente, normal padrão e qui-quadrado.

4.3.3 Performance dos critérios de selecção de modelos

A performance de cada um dos critérios de selecção de modelos pode ser determinada pela capacidade que esse critério tem de seleccionar o modelo “correcto”, isto é, o modelo do qual os dados são gerados.

Os resultados referentes ao número de vezes que cada um dos modelos é seleccionado podem ser apresentados numa matriz 16×16 , em que o índice da linha indica o modelo do qual os dados são gerados e o índice da coluna o modelo escolhido pelo critério de selecção de modelos considerado. Assim, na diagonal principal da matriz de resultados estão as frequências de classificação correcta.

As matrizes de resultados obtidas, no **caso 1**, com as medidas DIC, \widehat{BIC} e LLP são apresentadas nas tabelas 4.8, 4.9 e 4.10, respectivamente. Para facilitar a interpretação destas matrizes de resultados, os modelos geradores e os modelos ajustados são separados de acordo com o número de covariáveis que os constituem.

Os resultados apresentados na tabela 4.8, obtidos com a medida de discrepância DIC, podem ser facilmente interpretados. Veja-se, por exemplo, como se interpreta a linha da tabela associada a m_8 . Das 1000 amostras de Y geradas do modelo m_8 (que inclui as covariáveis x_1 e x_4), o DIC proporciona a selecção do modelo “correcto” 700 vezes, sendo as 300 classificações incorrectas distribuídas do seguinte modo: m_1 é seleccionado 29 vezes, m_3 é escolhido 130 vezes, m_4 é seleccionado 140 vezes e, por fim, m_{12} é escolhido 1 vez. Relativamente ao critério de informação da *deviance*, salienta-se ainda que:

- a parte triangular superior da matriz de resultados é quase toda constituída por zeros, o que indica que o modelo seleccionado raramente tem um número de covariáveis inferior ao do modelo gerador;
- para os dados gerados do modelo com 4 covariáveis a frequência observada de classificação correcta é 1000, para os dados gerados dos modelos com 3 covariáveis a frequência observada de classificação correcta varia entre 831 e 857, e para os dados gerados dos modelos com 2 covariáveis a frequência observada de classificação

²Estas subrotinas estão descritas no *IMSL Fortran Numerical Libraries - Stat/Library*.

correcta varia entre 687 e 714. Para os dados gerados dos modelos com apenas 1 covariável ou nenhuma covariável as frequências observadas de classificação correcta são inferiores. Numa situação deste tipo, a diagonal principal da matriz de resultados pode ser classificada como “decrecente”.

- o modelo seleccionado é, na maior parte das vezes, o modelo gerador ou modelos com mais uma covariável que as incluídas no modelo gerador.

Tabela 4.8: Número de vezes que cada modelo é seleccionado pelo DIC, por modelo gerador dos dados e modelo ajustado, para o caso 1.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	1000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_2	169	831	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_3	166	0	834	0	0	0	0	0	0	0	0	0	0	0	0	0
m_4	169	0	0	831	0	0	0	0	0	0	0	0	0	0	0	0
m_5	143	0	0	0	857	0	0	0	0	0	0	0	0	0	0	0
m_6	25	133	128	0	0	714	0	0	0	0	0	0	0	0	0	0
m_7	27	142	0	124	0	0	707	0	0	0	0	0	0	0	0	0
m_8	29	0	130	140	0	0	0	700	0	0	0	1	0	0	0	0
m_9	28	132	0	0	136	0	0	0	704	0	0	0	0	0	0	0
m_{10}	32	0	144	0	137	0	0	0	0	687	0	0	0	0	0	0
m_{11}	29	0	0	138	142	0	0	0	0	0	690	0	0	1	0	0
m_{12}	7	23	20	22	0	118	116	108	0	0	0	586	0	0	0	0
m_{13}	3	24	20	0	16	122	0	0	108	116	0	0	591	0	0	0
m_{14}	6	19	0	18	26	0	118	0	101	0	133	0	0	579	0	0
m_{15}	6	0	20	18	18	0	0	119	0	110	118	0	0	0	591	0
m_{16}	1	3	3	5	1	19	17	12	20	21	18	102	90	93	113	482

A análise da tabela 4.9, relativa ao critério de informação bayesiano estimado ($\widehat{\text{BIC}}$), permite destacar que:

- a diagonal principal da matriz de resultados tem valores bastante elevados, variando entre 943 e 981, o que constitui uma frequência de classificação correcta muito boa para os dados gerados de qualquer um dos modelos;
- a frequência de classificação correcta mantém-se sensivelmente constante, independentemente do número de covariáveis do modelo gerador;
- quando o modelo escolhido não é o modelo gerador, há uma ligeira tendência para seleccionar modelos com mais uma covariável que as do modelo gerador (pois as classificações incorrectas mais elevadas estão abaixo da diagonal principal da matriz de resultados);
- o modelo escolhido é, de um modo geral, o modelo do qual os dados são gerados. As frequências de classificação incorrecta mais elevadas referem-se à selecção de modelos com mais uma ou menos uma covariável que o modelo gerador.

Tabela 4.9: Número de vezes que cada modelo é seleccionado pelo $\widehat{\text{BIC}}$, por modelo gerador dos dados e modelo ajustado, para o caso 1.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	980	5	3	10	0	0	2	0	0	0	0	0	0	0	0	0
m_2	17	974	0	0	0	3	6	0	0	0	0	0	0	0	0	0
m_3	12	0	969	0	0	7	0	11	0	0	0	1	0	0	0	0
m_4	14	1	0	978	0	0	5	2	0	0	0	0	0	0	0	0
m_5	9	0	1	1	981	0	0	0	5	1	2	0	0	0	0	0
m_6	0	16	6	0	0	972	0	1	0	0	0	5	0	0	0	0
m_7	0	11	0	15	0	0	971	0	0	0	0	2	0	1	0	0
m_8	1	0	11	11	0	0	0	968	0	0	0	8	0	0	1	0
m_9	1	9	0	0	12	0	0	0	972	0	0	0	3	3	0	0
m_{10}	0	0	14	0	20	1	0	0	0	943	1	0	11	0	10	0
m_{11}	0	0	0	8	9	0	0	0	0	0	977	0	0	2	4	0
m_{12}	0	0	1	0	0	12	11	12	0	0	0	964	0	0	0	0
m_{13}	0	0	0	0	0	17	0	0	14	11	0	0	951	0	0	7
m_{14}	0	1	0	1	0	0	14	0	13	0	9	0	0	960	0	2
m_{15}	0	0	1	0	2	0	0	16	0	16	13	0	0	0	945	7
m_{16}	0	0	0	0	0	0	1	0	0	0	0	17	9	12	11	950

Relativamente ao critério que utiliza como medida a log-verosimilhança *a posteriori* (LLP), pela análise da tabela 4.10, salienta-se que:

- a matriz de resultados evidencia uma performance bastante inferior à do DIC e do $\widehat{\text{BIC}}$, que se traduz por frequências de classificação correcta bastante pequenas (com excepção para os dados gerados de m_1);
- a diagonal principal da matriz de resultados é “decrecente”, isto é, as frequências de classificação correcta são mais elevadas para os dados gerados de modelos com um maior número de covariáveis;
- no caso do modelo escolhido não ser o modelo gerador, o modelo seleccionado tem mais covariáveis que o modelo do qual os dados são gerados (pois as frequências não nulas de classificação incorrecta estão localizadas abaixo da diagonal principal da matriz);
- a tendência em seleccionar modelos com mais covariáveis que as incluídas no modelo gerador é bastante mais evidente que no caso do DIC.

Os resultados obtidos para os casos 2 a 6, relativos ao número de vezes que cada um dos modelos é seleccionado pelas medidas DIC, $\widehat{\text{BIC}}$ e LLP estão sumariados nas tabelas E.7 a E.21 do anexo E.

Ao contrário do que acontece no caso 1, onde os dados são gerados de modelos em que os coeficientes das covariáveis são todos unitários, no **caso 2** os dados são gerados de modelos

Tabela 4.10: Número de vezes que cada modelo é seleccionado pelo LLP, por modelo gerador dos dados e modelo ajustado, para o caso 1.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	1000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_2	326	674	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_3	306	0	694	0	0	0	0	0	0	0	0	0	0	0	0	0
m_4	317	0	0	683	0	0	0	0	0	0	0	0	0	0	0	0
m_5	312	0	0	0	688	0	0	0	0	0	0	0	0	0	0	0
m_6	97	230	217	0	0	456	0	0	0	0	0	0	0	0	0	0
m_7	100	232	0	185	0	0	483	0	0	0	0	0	0	0	0	0
m_8	109	0	204	217	0	0	0	470	0	0	0	0	0	0	0	0
m_9	106	224	0	0	215	0	0	0	455	0	0	0	0	0	0	0
m_{10}	104	0	239	0	195	0	0	0	0	462	0	0	0	0	0	0
m_{11}	104	0	0	218	233	0	0	0	0	0	445	0	0	0	0	0
m_{12}	33	68	63	65	0	144	156	155	0	0	0	316	0	0	0	0
m_{13}	31	62	75	0	59	163	0	0	141	155	0	0	314	0	0	0
m_{14}	28	66	0	59	67	0	153	0	145	0	158	0	0	324	0	0
m_{15}	32	0	57	62	71	0	0	163	0	146	159	0	0	0	310	0
m_{16}	18	26	17	19	27	40	48	40	55	43	44	100	100	90	129	204

com coeficientes distintos. O coeficiente associado à covariável x_1 mantém-se unitário, mas os coeficientes associados às covariáveis x_2 , x_3 e x_4 são 10, 2 e 5, respectivamente. Esta modificação dos coeficientes das covariáveis do modelo gerador não origina alterações relevantes nas frequências de selecção dos modelos. A performance obtida com as medidas DIC e LLP é idêntica à do caso 1. Relativamente ao $\widehat{\text{BIC}}$ regista-se que a parte triangular superior da matriz de resultados é quase toda constituída por zeros; isto quer dizer que quando o modelo escolhido não é o gerador dos dados, há uma tendência mais acentuada, que no caso 1, para seleccionar modelos com mais uma covariável que as do modelo gerador (tabelas E.7, E.8 e E.9).

A introdução, no **caso 3**, de uma correlação linear moderada entre duas covariáveis não produz alterações relevantes no desempenho dos critérios de selecção de modelos (tabelas E.10, E.11 e E.12), quando comparado com o do caso 1.

No **caso 4**, a correlação linear elevada entre as covariáveis x_1 e x_3 produz alterações significativas na performance dos critérios de selecção de modelos, quando comparada com a dos casos 1 e 3. Da análise das tabelas E.13, E.14 e E.15, tanto para o DIC como para o $\widehat{\text{BIC}}$ e o LLP, regista-se que:

- no caso dos dados gerados de m_7 (que inclui as covariáveis x_1 e x_3), a frequência de classificação correcta é bastante inferior à dos casos 1 e 3, sendo m_{12} (com a covariável x_1) e m_{14} (com a covariável x_3) os modelos seleccionados mais frequentemente.
- as frequências de classificação correcta dos dados gerados dos modelos m_8 (que inclui x_1 e x_4), m_9 (que inclui x_2 e x_3) e m_{11} (que inclui x_3 e x_4) são inferiores às dos casos

1 e 3, pois os correspondentes modelos de duas covariáveis que resultam da troca de x_1 por x_3 , ou vice-versa, são também seleccionados com alguma frequência.

- as frequências de classificação correcta dos dados gerados dos modelos m_{12} (que inclui x_1) e m_{14} (que inclui x_3) são inferiores às dos casos 1 e 3, pois os modelos m_{14} e m_{12} , respectivamente, são também seleccionados com alguma frequência.

A introdução de uma correlação linear moderada entre covariáveis não tem grande impacto na performance destes critérios de selecção de modelos. O mesmo já não acontece no caso de uma correlação linear elevada. Mas, quando existe uma correlação linear elevada entre covariáveis será que é possível dizer que apenas o modelo do qual os dados são gerados é o modelo “correcto”? Considere-se, novamente, o caso 4 em que existe uma correlação linear de 0.98 entre as covariáveis x_1 e x_3 . Para os dados gerados do modelo m_7 (com as covariáveis x_1 e x_3), tanto o modelo m_7 como os modelos m_{12} (com x_1) e m_{14} (com x_3) podem ser considerados o modelo “correcto”. Assim, a frequência de classificação correcta dos dados gerados do modelo m_7 deve ser entendida como a soma das frequências de selecção dos modelos m_7 , m_{12} e m_{14} e não apenas o número de vezes que o modelo m_7 é escolhido. Pode mesmo considerar-se que os modelos m_{12} e m_{14} são melhores que o modelo m_7 pois são mais parcimoniosos, uma vez que a inclusão no modelo da covariável x_1 (respectivamente, x_3), torna a inclusão de x_3 (respectivamente, x_1) redundante.

No **caso 5**, em que as amostras têm dimensão $n = 50$, a performance dos métodos de selecção de modelos diminui substancialmente, quando comparada com o caso 1 ($n = 200$). Da análise das tabelas E.16, E.17 e E.18, salienta-se os seguintes aspectos:

- no caso do $\widehat{\text{BIC}}$, a diagonal principal da matriz de resultados é “crescente”, isto é, os dados gerados do modelo com 4 covariáveis têm uma frequência de classificação correcta inferior à dos dados gerados dos modelos de 3 covariáveis, e estes, por sua vez, têm uma frequência de classificação correcta inferior à dos dados gerados dos modelos com 2 covariáveis, e assim sucessivamente. No caso do LLP, a diagonal principal da matriz de resultados é “decrecente”. Por sua vez, no caso do DIC, as frequências de classificação correcta são muito idênticas, independentemente do número de covariáveis do modelo gerador;
- a matriz de resultados obtida com o DIC apresenta poucos zeros e tem um aspecto mais “simétrico” que no caso 1, na medida em que as frequências de classificação incorrecta acima e abaixo da diagonal principal são bastante idênticas. No caso de amostras de dimensão mais reduzida, quando o DIC selecciona um modelo incorrecto, tanto pode ser um modelo com mais ou menos covariáveis que as do modelo gerador;
- a matriz de resultados obtida com o $\widehat{\text{BIC}}$ tem um aspecto bastante assimétrico, pois as frequências de classificação incorrecta acima da diagonal principal são, de um modo geral, mais elevadas. No caso de amostras de dimensão mais reduzida, quando o $\widehat{\text{BIC}}$ selecciona um modelo incorrecto, tem tendência a seleccionar modelos com menos covariáveis que as incluídas no modelo gerador;
- a matriz de resultados obtida com o LLP tem um aspecto assimétrico, pois as frequências de classificação incorrecta abaixo da diagonal principal são, de um modo geral, mais elevadas. No caso de amostras de dimensão mais reduzida, quando o

LLP selecciona um modelo incorrecto, tem tendência a seleccionar modelos com mais covariáveis que as incluídas no modelo gerador;

- as frequências de classificação correcta obtidas com o DIC são mais homogêneas que as obtidas com as outras duas medidas de discrepância.

No **caso 6**, em que as amostras têm dimensão $n = 100$, da análise das tabelas E.19, E.20 e E.21, destaca-se que:

- a diagonal principal da matriz de resultados é “crescente” no caso do $\widehat{\text{BIC}}$ e “decrecente” no caso do DIC e LLP;
- a performance do DIC é inferior à do $\widehat{\text{BIC}}$ para os dados gerados dos modelos com 2, 1 ou nenhuma covariável.
- com excepção dos dados gerados de m_1 , o desempenho do LLP na selecção do modelo “correcto” é bastante inferior ao dos outros critérios.

Na tabela 4.11 estão sumariadas as frequências de classificação correcta obtidas com os critérios DIC, $\widehat{\text{BIC}}$ e LLP, para os casos 1, 5 e 6.

Tabela 4.11: Frequências de classificação correcta dos dados gerados dos modelos m_1, \dots, m_{16} , nos casos 1, 5 e 6, para cada uma das medidas de selecção de modelos.

modelo	DIC			$\widehat{\text{BIC}}$			LLP		
	$n = 200$	$n = 100$	$n = 50$	$n = 200$	$n = 100$	$n = 50$	$n = 200$	$n = 100$	$n = 50$
m_1	1000	964	507	980	722	136	1000	986	706
m_2	831	804	472	974	804	158	674	693	519
m_3	834	802	472	969	713	184	694	671	532
m_4	831	829	571	978	800	273	683	693	580
m_5	857	821	536	981	757	216	688	688	555
m_6	714	681	399	972	813	250	456	436	338
m_7	707	683	549	971	860	374	483	461	417
m_8	700	666	562	968	834	439	470	439	437
m_9	704	696	476	972	833	279	455	467	380
m_{10}	687	661	475	943	770	305	462	423	389
m_{11}	690	703	572	977	870	436	445	467	439
m_{12}	586	604	483	964	900	557	316	324	306
m_{13}	591	566	425	951	850	420	314	327	265
m_{14}	579	564	485	960	895	556	324	321	292
m_{15}	591	555	527	945	876	672	310	292	316
m_{16}	482	490	469	950	921	869	204	224	210

Para as amostras de dimensão $n = 200$, o $\widehat{\text{BIC}}$ é o que apresenta uma melhor performance, com frequências de classificação correcta bastante elevadas; neste caso, a performance do DIC é inferior à do $\widehat{\text{BIC}}$, especialmente para dados gerados de modelos com poucas covariáveis. Para amostras de dimensão $n = 100$, o desempenho do DIC

aproxima-se do desempenho do $\widehat{\text{BIC}}$, e para amostras de dimensão $n = 50$, a performance do DIC ultrapassa a do $\widehat{\text{BIC}}$. Com a excepção do caso $n = 50$, onde não se pode dizer que a performance do LLP seja muito inferior à performance do DIC e do $\widehat{\text{BIC}}$, o LLP é a medida de selecção de modelos com pior desempenho.

A penalização ao número de parâmetros do modelo imposta pelo $\widehat{\text{BIC}}$ parece ser demasiado pesada pois, no caso de selecção incorrecta, à medida que n diminui existe tendência para seleccionar modelos com menos covariáveis que as incluídas no modelo gerador. No caso do DIC a situação é diferente: no caso de selecção incorrecta, à medida que n aumenta existe tendência para seleccionar modelos com mais covariáveis que as incluídas no modelo gerador; esta situação pode dever-se ao facto da penalização ao número de parâmetros do modelo ser demasiado fraca. Esta constatação sugere que a performance dos critérios de selecção de modelos só é boa se for atingido um determinado “equilíbrio” entre a dimensão da amostra e a penalização em relação ao número de covariáveis do modelo.

4.3.4 Performance do método GVS

A avaliação do desempenho do método GVS, na identificação do modelo do qual os dados são gerados, é efectuada de um modo idêntico ao utilizado para estudar a performance das medidas de discrepância. Considera-se as covariáveis x_1, x_2, x_3, x_4 já referidas na secção 4.3.1 e simula-se uma amostra para o erro aleatório ϵ , constituída por observações independentes e identicamente distribuídas $N(0, 2.5^2)$. Com esta amostra do erro aleatório obtém-se 16 amostras da variável resposta Y , cada uma gerada de um dos modelos m_1, \dots, m_{16} . Aplica-se o método GVS a cada uma destas amostras, estimando as probabilidades *a posteriori* dos modelos m_1, \dots, m_{16} e as probabilidades *a posterior* de inclusão das covariáveis x_1, x_2, x_3 e x_4 no modelo. O procedimento descrito é repetido para os 6 casos apresentados na secção 4.3.1. O modo de implementação deste procedimento no WinBUGS é descrito na secção D.3 do anexo D.

Considera-se que não existe informação *a priori* sobre a plausibilidade relativa dos diferentes modelos, o que pode ser expresso através de probabilidades *a priori* iguais a $1/16$ para cada modelo identificado pelo vector³ $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)$. Deste modo, as variáveis γ_j , para $j = 1, \dots, 4$, são consideradas independentes com distribuição *a priori* Bernoulli com probabilidade de sucesso $1/2$. Considera-se que a distribuição *a priori* para b_0 é normal de valor médio nulo e variância 10000, que a distribuição *a priori* de b_j , condicional a γ_j , é normal com valor médio nulo e variância $0.01^{1-\gamma_j} \times 1000$, para $j = 1, \dots, 4$ (Ntzoufras, 2002), e que a distribuição *a priori* de τ é gama com parâmetro de forma e inverso do parâmetro de escala iguais a 0.001.

Nas 16 aplicações do método GVS, as estimativas são obtidas em 10000 iterações do método de amostragem, com um espaçamento de amostragem de 5 iterações, após um período de aquecimento de 5000 iterações. Os resultados obtidos para os casos 1 a 6 são apresentados nas tabelas 4.12 a 4.17, respectivamente. Ao contrário das tabelas com as matrizes de resultados 16×16 da secção 4.3.3, nas tabelas desta secção o índice de coluna

³O identificador do modelo, $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)$, sugerido por Ntzoufras (2002), e utilizado nas aplicações do método GVS para codificação dos possíveis modelos (ver secção 3.3.4), não coincide com o identificador $(\delta_1, \delta_2, \delta_3, \delta_4)$ utilizado ao longo da secção 4.3.

indica o modelo do qual os dados são gerados, e o índice de linha identifica o modelo ao qual se refere cada uma das probabilidades *a posteriori* dos modelos — na parte superior das tabelas —, ou a covariável à qual se refere cada uma das probabilidades *a posteriori* de inclusão das covariáveis no modelo — na parte inferior das tabelas.

No **caso 1**, em que se considera $n = 200$, o método GVS atribui probabilidades *a posteriori* bastante elevadas ao modelo do qual os dados são gerados. Por exemplo, para os dados gerados do modelo m_4 — modelo que inclui as covariáveis x_1, x_3 e x_4 — obtém-se $\Pr\{m_4|\mathcal{D}\} = 0.9271$, $\Pr\{m_{11}|\mathcal{D}\} = 0.0661$ (m_{11} inclui as covariáveis x_3 e x_4), $\Pr\{m_1|\mathcal{D}\} = 0.0062$ (m_1 inclui todas as covariáveis), $\Pr\{m_5|\mathcal{D}\} = 0.0006$ (m_5 inclui as covariáveis x_2, x_3 e x_4) e probabilidades *a posteriori* nulas para os restantes modelos; além disso, regista-se probabilidades *a posteriori* de inclusão das covariáveis no modelo elevadas para x_1, x_3 e x_4 , nomeadamente $\Pr\{\gamma_1|\mathcal{D}\} = 0.9333$, $\Pr\{\gamma_3|\mathcal{D}\} = 1$ e $\Pr\{\gamma_4|\mathcal{D}\} = 1$, e probabilidade *a posteriori* de inclusão da covariável x_2 no modelo pequena, $\Pr\{\gamma_2|\mathcal{D}\} = 0.0068$.

Neste primeiro caso, tanto o modelo com maior probabilidade *a posteriori* como o modelo de probabilidade *a posteriori* mediana coincidem com o modelo do qual os dados são gerados. O método GVS tem uma performance bastante boa, permitindo a identificação do modelo do qual os dados são gerados. Mas, o que acontece se a dimensão da amostra diminuir? Ou se as covariáveis forem correlacionadas? Ou, ainda, se os coeficientes das covariáveis forem distintos? Para tentar dar resposta a estas questões, analisa-se os resultados obtidos nos casos 2 a 6.

Tabela 4.12: Probabilidades *a posteriori* dos modelos e probabilidades *a posteriori* de inclusão das covariáveis no modelo, para o caso 1.

modelo gerador dos dados																
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
probabilidades <i>a posteriori</i> dos modelos																
m_1	0.8984	0.0073	0.0098	0.0062	0.0328	0.0001	0.0001	0	0.0003	0.0003	0	0	0	0	0	0
m_2	0	0.8729	0	0	0	0.0089	0.0067	0	0.0343	0	0	0	0.0003	0	0	0
m_3	0	0	0.865	0	0	0.0064	0	0.0064	0	0.0358	0	0.0001	0.0003	0	0	0
m_4	0.0197	0.0001	0.0001	0.9271	0.0005	0	0.0081	0.0106	0	0	0.032	0.0001	0	0.0003	0.0003	0
m_5	0.0751	0.0004	0.0007	0.0006	0.9526	0	0	0	0.0078	0.0119	0.006	0	0.0001	0.0001	0.0001	0
m_6	0	0	0	0	0	0.8415	0	0	0	0	0	0.0067	0.0378	0	0	0
m_7	0	0.0271	0	0	0	0.0002	0.9084	0	0.001	0	0	0.0103	0	0.0331	0	0.0003
m_8	0	0	0.0231	0	0	0.0002	0	0.9022	0	0.0009	0	0.0078	0	0	0.0348	0.0003
m_9	0	0.0835	0	0	0	0.0009	0.0007	0	0.9354	0	0	0	0.0113	0.007	0	0.0001
m_{10}	0	0	0.0934	0	0	0.0005	0	0.0007	0	0.9342	0	0	0.0077	0	0.0063	0.0001
m_{11}	0.0068	0.0003	0.0001	0.0661	0.0141	0	0.0004	0.0004	0.0001	0.0001	0.962	0	0	0.0087	0.0119	0.0001
m_{12}	0	0	0	0	0	0.0298	0	0	0	0	0	0.8838	0.0012	0	0	0.0362
m_{13}	0	0	0	0	0	0.1004	0	0	0	0	0	0.0009	0.9178	0	0	0.007
m_{14}	0	0.0084	0	0	0	0.0001	0.0756	0	0.0211	0	0	0.0004	0.0003	0.9508	0	0.0117
m_{15}	0	0	0.0078	0	0	0.0003	0	0.0797	0	0.0168	0	0.0006	0.0002	0	0.9466	0.0087
m_{16}	0	0	0	0	0	0.0107	0	0	0	0	0	0.0893	0.023	0	0	0.9355
probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo																
γ_1	0.9181	0.9074	0.898	0.9333	0.0333	0.8871	0.9233	0.9192	0.0356	0.037	0.032	0.9088	0.0396	0.0334	0.0351	0.0368
γ_2	0.9735	0.9641	0.9689	0.0068	0.9854	0.9587	0.0075	0.0071	0.9778	0.9822	0.006	0.0077	0.9753	0.0071	0.0064	0.0072
γ_3	1	1	0.0107	1	1	0.0102	1	0.011	1	0.0123	1	0.0108	0.012	1	0.0123	0.0122
γ_4	1	0.0081	1	1	1	0.0075	0.0086	1	0.0082	1	1	0.0086	0.0083	0.0091	1	0.0092

Se os dados são gerados de modelos em que as covariáveis têm coeficientes distintos, como acontece no **caso 2**, o modelo com maior probabilidade *a posteriori* é o modelo gerador dos dados ou o modelo que apenas inclui as covariáveis do modelo gerador com coeficientes mais elevados. Por exemplo, para os dados gerados de m_1 — modelo que inclui as covariáveis x_1, x_2, x_3 e x_4 com coeficientes 1, 10, 2 e 5, respectivamente —

verifica-se que o modelo m_5 (que inclui as covariáveis x_2, x_3 e x_4) é o que apresenta maior probabilidade *a posteriori*, $\Pr\{m_5|\mathcal{D}\} = 0.5912$. Veja-se ainda outro exemplo: para os dados gerados de m_4 — que inclui as covariáveis x_1, x_3 e x_4 com coeficientes 1, 2 e 5, respectivamente — verifica-se que o modelo m_{11} (que inclui as covariáveis x_3 e x_4) é o que apresenta maior probabilidade *a posteriori*, $\Pr\{m_{11}|\mathcal{D}\} = 0.5644$, logo seguido de m_4 com $\Pr\{m_4|\mathcal{D}\} = 0.4217$.

Tabela 4.13: Probabilidades *a posteriori* dos modelos e probabilidades *a posteriori* de inclusão das covariáveis no modelo, para o caso 2.

modelo gerador dos dados																
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
probabilidades <i>a posteriori</i> dos modelos																
m_1	0.4088	0.0048	0.0081	0.0061	0.013	0.0001	0	0	0	0.0001	0	0	0	0	0	0
m_2	0	0.3672	0	0	0	0.0071	0.0052	0	0.0124	0	0	0	0.0001	0.0001	0	0
m_3	0	0	0.4549	0	0	0.0055	0	0.0063	0	0.0133	0	0	0	0	0	0
m_4	0	0	0	0.4217	0	0	0.0041	0.0083	0	0	0.0133	0.0001	0	0	0.0001	0
m_5	0.5912	0.0086	0.0081	0.0078	0.987	0.0001	0.0001	0	0.0132	0.0138	0.0127	0	0.0001	0.0001	0	0
m_6	0	0	0	0	0	0.4134	0	0	0	0	0	0.0055	0.0125	0	0	0.0001
m_7	0	0	0	0	0	0	0.4106	0	0	0	0	0.008	0	0.0118	0	0.0001
m_8	0	0	0	0	0	0	0	0.4695	0	0	0	0.0046	0	0	0.0135	0
m_9	0	0.6194	0	0	0	0.0089	0.0079	0	0.9744	0	0	0.0001	0.0138	0.0116	0	0.0001
m_{10}	0	0	0.5289	0	0	0.0077	0	0.0074	0	0.9728	0	0.0001	0.013	0	0.013	0.0002
m_{11}	0	0	0	0.5644	0	0	0.0075	0.0078	0	0	0.974	0.0001	0	0.0113	0.0139	0.0001
m_{12}	0	0	0	0	0	0	0	0	0	0	0	0.4605	0	0	0	0.0123
m_{13}	0	0	0	0	0	0.5572	0	0	0	0	0	0.0072	0.9605	0	0	0.0114
m_{14}	0	0	0	0	0	0	0.5646	0	0	0	0	0.008	0	0.9651	0	0.0136
m_{15}	0	0	0	0	0	0	0	0.5007	0	0	0	0.0067	0	0	0.9595	0.0109
m_{16}	0	0	0	0	0	0	0	0	0	0	0	0.4991	0	0	0	0.9512
probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo																
γ_1	0.4088	0.372	0.463	0.4278	0.013	0.4261	0.4199	0.4841	0.0124	0.0134	0.0133	0.4787	0.0126	0.0119	0.0136	0.0125
γ_2	1	1	1	0.0139	1	1	0.0132	0.0137	1	1	0.0127	0.0129	1	0.0118	0.013	0.0118
γ_3	1	1	0.0162	1	1	0.0162	1	0.0161	1	0.0139	1	0.0163	0.014	1	0.014	0.0139
γ_4	1	0.0134	1	1	1	0.0134	0.0117	1	0.0132	1	1	0.0116	0.0131	0.0114	1	0.0112

Com uma correlação linear moderada entre duas covariáveis, como acontece no **caso 3**, o método GVS nem sempre permite identificar o modelo gerador. Por exemplo, para os dados gerados de m_4 — modelo que inclui as covariáveis x_1, x_3 e x_4 com uma correlação linear moderada entre x_1 e x_3 —, tem-se $\Pr\{m_4|\mathcal{D}\} = 0.2205$, sendo o modelo m_{11} (modelo que inclui as covariáveis x_3 e x_4) o que apresenta maior probabilidade *a posteriori*, $\Pr\{m_{11}|\mathcal{D}\} = 0.7724$. Ainda para os dados gerados do modelo m_4 , apenas as covariáveis x_3 e x_4 têm probabilidades *a posteriori* de inclusão no modelo elevadas, mais concretamente $\Pr\{\gamma_3|\mathcal{D}\} = 1$ e $\Pr\{\gamma_4|\mathcal{D}\} = 1$, sendo a probabilidade *a posteriori* de inclusão da covariável x_1 (que tem o menor dos coeficientes) no modelo bastante pequena, $\Pr\{\gamma_1|\mathcal{D}\} = 0.2219$.

No **caso 4**, em que existe uma correlação linear elevada entre duas covariáveis, os modelos dos quais os dados são gerados têm, por vezes, probabilidades *a posteriori* bastante pequenas. Por exemplo, para os dados gerados de m_4 — modelo que inclui as covariáveis x_1, x_3 e x_4 , com uma correlação linear elevada entre x_1 e x_3 —, tem-se $\Pr\{m_4|\mathcal{D}\} = 0.0306$, sendo o modelo m_{11} (que inclui as covariáveis x_3 e x_4) o que apresenta probabilidade *a posteriori* mais elevada, $\Pr\{m_{11}|\mathcal{D}\} = 0.8833$.

Conforme se ilustra nos casos 3 e 4, a existência de covariáveis linearmente correlacionadas altera os resultados do método GVS. Mas, será que esta alteração significa uma performance inferior deste método de selecção de covariáveis em situações de correlação? Não parece ser esse o caso! Se os dados são gerados de um modelo em

4.3. Dados de regressão linear normal

que duas das covariáveis são bastante correlacionadas, não é de estranhar que o método GVS atribua probabilidade *a posteriori* mais elevada ao modelo que inclui apenas uma dessas covariáveis, pois a presença de uma delas no modelo torna a inclusão da outra desnecessária.

Tabela 4.14: Probabilidades *a posteriori* dos modelos e probabilidades *a posteriori* de inclusão das covariáveis no modelo, para o caso 3.

modelo gerador dos dados																
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
probabilidades <i>a posteriori</i> dos modelos																
m_1	0.2022	0.0018	0.0119	0.0014	0.0848	0.0004	0	0	0.001	0.0005	0.0003	0	0	0	0	0
m_2	0	0.1899	0	0	0	0.011	0.0015	0	0.0886	0	0	0	0.0004	0.0004	0	0
m_3	0	0	0.8209	0	0	0.0055	0	0.0062	0	0.0357	0	0.0001	0.0003	0	0	0
m_4	0.0042	0	0	0.2205	0.0026	0	0.0017	0.0127	0	0	0.0844	0.0003	0	0.0012	0.0004	0
m_5	0.7632	0.0056	0.0479	0.0057	0.8962	0.0008	0.0001	0.0004	0.0068	0.0053	0.0057	0	0.0001	0.0001	0	0
m_6	0	0	0	0	0	0.7984	0	0	0	0	0	0.0065	0.0378	0	0	0
m_7	0	0.006	0	0	0	0.0001	0.2112	0	0.0028	0	0	0.0123	0	0.0857	0	0.0004
m_8	0	0	0.0219	0	0	0.0002	0	0.8571	0	0.001	0	0.0069	0	0	0.0347	0.0003
m_9	0	0.7579	0	0	0	0.0472	0.0062	0	0.8776	0	0	0.0004	0.0051	0.0066	0	0
m_{10}	0	0	0.0878	0	0	0.0002	0	0.0005	0	0.9407	0	0	0.0077	0	0.0064	0.0001
m_{11}	0.0304	0.0004	0.0015	0.7724	0.0164	0	0.0064	0.0468	0.0002	0.0002	0.9096	0.0008	0	0.0077	0.0054	0.0001
m_{12}	0	0	0	0	0	0.029	0	0	0	0	0	0.838	0.0012	0	0	0.0361
m_{13}	0	0	0	0	0	0.0941	0	0	0	0	0	0.0007	0.9239	0	0	0.0071
m_{14}	0	0.0384	0	0	0	0.0021	0.7729	0	0.023	0	0	0.0483	0.0003	0.8983	0	0.0056
m_{15}	0	0	0.0081	0	0	0.0003	0	0.0763	0	0.0166	0	0.0004	0.0002	0	0.9531	0.0087
m_{16}	0	0	0	0	0	0.0107	0	0	0	0	0	0.0853	0.023	0	0	0.9416
probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo																
γ_1	0.2064	0.1977	0.8547	0.2219	0.0874	0.8446	0.2144	0.876	0.0924	0.0372	0.0847	0.8641	0.0397	0.0873	0.0351	0.0368
γ_2	0.9654	0.9552	0.9685	0.0071	0.981	0.9576	0.0078	0.0071	0.974	0.9822	0.006	0.0077	0.9753	0.0071	0.0064	0.0072
γ_3	1	1	0.0613	1	1	0.0616	1	0.0599	1	0.006	1	0.0621	0.0059	1	0.0058	0.0061
γ_4	1	0.0078	1	1	1	0.0074	0.0082	1	0.008	1	1	0.0085	0.0083	0.009	1	0.0092

Tabela 4.15: Probabilidades *a posteriori* dos modelos e probabilidades *a posteriori* de inclusão das covariáveis no modelo, para o caso 4.

modelo gerador dos dados																
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
probabilidades <i>a posteriori</i> dos modelos																
m_1	0.0292	0.0002	0.0198	0.0001	0.0818	0.0001	0	0.0001	0.0007	0.0016	0.0008	0	0	0	0	0
m_2	0	0.0281	0	0	0	0.0187	0.0001	0	0.0802	0	0	0.0001	0.0006	0.0008	0	0
m_3	0.0778	0.0004	0.2946	0.001	0.1594	0.0025	0	0.0017	0.0013	0.0354	0.0011	0	0.0003	0	0	0
m_4	0.0009	0	0.0006	0.0306	0.002	0	0.0002	0.0226	0	0.0001	0.0936	0.0001	0	0.0006	0.0009	0
m_5	0.87	0.0072	0.6262	0.0058	0.7071	0.0044	0.0001	0.0049	0.0052	0.0227	0.0046	0.0001	0.0001	0.0001	0.0001	0
m_6	0	0.0674	0	0	0	0.3017	0.0011	0	0.1577	0	0	0.0019	0.0375	0.0011	0	0
m_7	0	0.001	0	0	0	0.0004	0.0295	0	0.0023	0	0	0.0211	0	0.0848	0	0.0018
m_8	0.0023	0	0.01	0.0792	0.0061	0.0001	0.0004	0.2956	0.0001	0.001	0.1595	0.003	0	0.0017	0.0344	0.0003
m_9	0	0.8673	0	0	0	0.6001	0.0063	0	0.6916	0	0	0.005	0.0245	0.0053	0	0.0001
m_{10}	0	0	0.0335	0	0.0241	0.0003	0	0.0005	0.0001	0.9225	0.0006	0	0.0077	0	0.0063	0.0001
m_{11}	0.0198	0.0003	0.0125	0.8833	0.0177	0	0.0078	0.6484	0	0.0008	0.7208	0.0051	0	0.0057	0.0222	0.0001
m_{12}	0	0.0025	0	0	0	0.0127	0.0783	0	0.0085	0	0	0.2907	0.0012	0.1617	0	0.0358
m_{13}	0	0	0	0	0	0.0376	0	0	0.0259	0	0	0.0006	0.9044	0.0005	0	0.007
m_{14}	0	0.0256	0	0	0	0.017	0.8762	0	0.0239	0	0	0.6417	0.0012	0.7163	0	0.024
m_{15}	0	0	0.0028	0	0.0018	0.0002	0	0.0262	0.0002	0.0159	0.019	0.0002	0.0002	0.0002	0.9361	0.0086
m_{16}	0	0	0	0	0	0.0042	0	0	0.0023	0	0	0.0304	0.0223	0.0212	0	0.9222
probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo																
γ_1	0.1102	0.0996	0.325	0.1109	0.2493	0.3362	0.1096	0.32	0.2508	0.0381	0.255	0.3169	0.0396	0.2507	0.0353	0.0379
γ_2	0.977	0.9706	0.9741	0.0069	0.9724	0.9654	0.0076	0.0072	0.9627	0.9822	0.0071	0.0077	0.9751	0.0078	0.0064	0.0072
γ_3	0.9199	0.9297	0.6591	0.9198	0.8086	0.6407	0.9202	0.676	0.8039	0.0252	0.8198	0.6732	0.0264	0.8136	0.0232	0.026
γ_4	1	0.0081	1	1	1	0.0076	0.0085	1	0.0076	1	1	0.0085	0.0083	0.0083	1	0.0091

No **caso 5**, em que se considera amostras de dimensão $n = 50$, o método GVS tem má performance, atribuindo probabilidades *a posteriori* bastante pequenas aos modelos dos quais os dados são gerados. Por exemplo, para os dados gerados de m_3 — que inclui x_1 , x_2 e x_4 — tem-se $\Pr\{m_3|\mathcal{D}\} = 0.0187$, sendo o modelo m_{16} (só com a constante) o que apresenta maior probabilidade *a posteriori*, $\Pr\{m_{16}|\mathcal{D}\} = 0.7017$.

Tabela 4.16: Probabilidades *a posteriori* dos modelos e probabilidades *a posteriori* de inclusão das covariáveis no modelo, para o caso 5.

modelo gerador dos dados																
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
probabilidades <i>a posteriori</i> dos modelos																
m_1	0.0088	0.0017	0	0.0019	0.0006	0	0	0	0	0	0	0	0	0	0	0
m_2	0.003	0.0944	0.0002	0.0004	0.0003	0.0026	0.004	0	0.0035	0	0	0	0.0001	0.0001	0	0
m_3	0.0032	0.0005	0.0187	0.0006	0.0002	0.0021	0	0.0031	0	0.0013	0.0001	0	0	0	0	0
m_4	0.0147	0.0017	0.0006	0.1105	0.0013	0.0001	0.0032	0.003	0	0	0.0064	0.0001	0	0	0	0
m_5	0.0114	0.0017	0.0003	0.0025	0.0638	0	0.0001	0	0.0033	0.0018	0.0073	0	0	0.0001	0	0
m_6	0.0024	0.0295	0.005	0.0003	0.0003	0.156	0.001	0.0006	0.0012	0.0007	0	0.0055	0.005	0	0	0.0001
m_7	0.0645	0.1795	0.0024	0.0401	0.0064	0.0052	0.3216	0.0014	0.0067	0	0.0032	0.008	0	0.0097	0	0.0001
m_8	0.0093	0.001	0.0268	0.0325	0.0009	0.0028	0.001	0.1878	0	0.0022	0.0027	0.0046	0	0	0.0095	0
m_9	0.0271	0.1559	0.001	0.0037	0.0202	0.003	0.0057	0.0002	0.2679	0.0007	0.0016	0.0001	0.0053	0.0091	0.0002	0.0001
m_{10}	0.0037	0.0005	0.0202	0.0005	0.017	0.0027	0	0.0032	0.0007	0.1087	0.0017	0.0001	0.0046	0	0.01	0.0002
m_{11}	0.0321	0.0029	0.0008	0.1497	0.1008	0.0001	0.0044	0.0041	0.0044	0.0027	0.5413	0.0001	0.0002	0.0082	0.0109	0.0001
m_{12}	0.0493	0.0972	0.1149	0.0283	0.0047	0.284	0.0901	0.0641	0.0037	0.0097	0.0017	0.4605	0.0095	0.0032	0.0037	0.0123
m_{13}	0.0153	0.0441	0.0433	0.0033	0.0136	0.2029	0.0024	0.0053	0.0645	0.0328	0.0025	0.0072	0.3691	0.0026	0.0026	0.0114
m_{14}	0.453	0.2594	0.0131	0.3847	0.4177	0.0052	0.4467	0.0085	0.448	0.0099	0.1809	0.008	0.0095	0.788	0.003	0.0136
m_{15}	0.0173	0.0039	0.051	0.045	0.046	0.0047	0.0031	0.2008	0.0045	0.1609	0.1304	0.0067	0.0076	0.0037	0.7209	0.0109
m_{16}	0.2849	0.1261	0.7017	0.196	0.3062	0.3286	0.1167	0.5179	0.1916	0.6686	0.1202	0.4991	0.5891	0.1753	0.2392	0.9512
probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo																
γ_1	0.1552	0.4055	0.1686	0.2146	0.0147	0.4528	0.4209	0.26	0.0151	0.0139	0.0141	0.4787	0.0146	0.013	0.0132	0.0125
γ_2	0.0749	0.3283	0.0887	0.0132	0.116	0.3693	0.0132	0.0124	0.3411	0.146	0.0132	0.0129	0.3841	0.0119	0.0128	0.0118
γ_3	0.6146	0.6972	0.0184	0.6935	0.6111	0.0162	0.7857	0.0172	0.7338	0.0151	0.7407	0.0163	0.0151	0.8152	0.0141	0.0139
γ_4	0.1005	0.0139	0.1184	0.3432	0.2306	0.0125	0.0118	0.402	0.0129	0.2776	0.6899	0.0116	0.0124	0.012	0.7513	0.0112

Relativamente ao **caso 6**, em que se considera amostras de dimensão $n = 100$, a performance do método GVS é superior à do caso 5 ($n = 50$), embora nem sempre o modelo do qual os dados são gerados seja o que apresenta probabilidade *a posteriori* máxima, como acontece no caso 1 ($n = 200$).

A comparação dos resultados obtidos nos casos 1, 5 e 6 evidencia que, tal como acontece com as medidas de discrepância, também a performance do método GVS depende da dimensão da amostra. Mas, será que não depende essencialmente da razão entre a dimensão da amostra e o número de covariáveis, ou da razão entre a dimensão da amostra e o número de modelos a monitorizar?

A repetição, num trabalho futuro, deste estudo de simulação com três possíveis covariáveis e amostras de dimensão 150 e 100, e com duas possíveis covariáveis e amostras de dimensão 100 e 50, pode acrescentar alguma informação adicional sobre os factores que influenciam a performance do método GVS. Se a performance do GVS com três covariáveis e amostras de dimensão 150, e com duas covariáveis e amostras de dimensão 100, for idêntica à do caso 1, então a performance do GVS será essencialmente influenciada pela razão entre a dimensão da amostra e o número de covariáveis. Mas, se o desempenho do GVS com três covariáveis e amostras de dimensão 100, e com duas covariáveis e amostras de dimensão 50, for idêntica à do caso 1, então a performance do GVS será essencialmente influenciada pela razão entre a dimensão da amostra e o número de modelos a monitorizar. Pode também acontecer que nenhuma destas duas situações ocorra!

Tabela 4.17: Probabilidades *a posteriori* dos modelos e probabilidades *a posteriori* de inclusão das covariáveis no modelo, para o caso 6.

modelo gerador dos dados																
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
probabilidades <i>a posteriori</i> dos modelos																
m_1	0.6412	0.0382	0.0097	0.0062	0.0114	0.0008	0.0003	0.0002	0.0015	0.0001	0	0	0	0	0	0
m_2	0	0.3525	0	0	0	0.0039	0.0021	0	0.0119	0	0	0.0001	0.0002	0	0	0
m_3	0	0	0.5491	0	0	0.0246	0	0.0045	0	0.0117	0	0	0.0014	0	0	0
m_4	0	0	0	0.6252	0	0	0.0365	0.0089	0	0	0.0116	0.0007	0	0.0016	0.0001	0
m_5	0.3588	0.0202	0.0061	0.0035	0.9886	0.0005	0.0001	0	0.1254	0.0185	0.0099	0	0.0025	0.0011	0.0002	0.0001
m_6	0	0	0	0	0	0.3133	0	0	0	0	0	0.0019	0.0131	0	0	0
m_7	0	0	0	0	0	0	0.3658	0	0	0	0	0.0038	0	0.0121	0	0.0002
m_8	0	0	0	0	0	0	0	0.5382	0	0	0	0.0247	0	0	0.0119	0.0013
m_9	0	0.5891	0	0	0	0.008	0.0054	0	0.8612	0	0	0.0001	0.0128	0.0068	0	0.0002
m_{10}	0	0	0.4351	0	0	0.019	0	0.0044	0	0.9697	0	0.0003	0.0938	0	0.0094	0.0009
m_{11}	0	0	0	0.3651	0	0	0.019	0.0063	0	0	0.9785	0.0004	0	0.1077	0.0183	0.0017
m_{12}	0	0	0	0	0	0	0	0	0	0	0	0.322	0	0	0	0.0132
m_{13}	0	0	0	0	0	0.6299	0	0	0	0	0	0.0055	0.8762	0	0	0.0067
m_{14}	0	0	0	0	0	0	0.5708	0	0	0	0	0.0078	0	0.8707	0	0.0133
m_{15}	0	0	0	0	0	0	0	0.4375	0	0	0	0.0182	0	0	0.9601	0.0825
m_{16}	0	0	0	0	0	0	0	0	0	0	0	0.6145	0	0	0	0.8799
probabilidades <i>a posteriori</i> de inclusão das covariáveis no modelo																
γ_1	0.6412	0.3907	0.5588	0.6314	0.0114	0.3426	0.4047	0.5518	0.0134	0.0118	0.0116	0.3532	0.0147	0.0137	0.012	0.0147
γ_2	1	1	1	0.0097	1	1	0.0079	0.0091	1	1	0.0099	0.0079	1	0.0079	0.0096	0.0079
γ_3	1	1	0.0158	1	1	0.0132	1	0.0154	1	0.0186	1	0.0129	0.0155	1	0.0186	0.0155
γ_4	1	0.0584	1	1	1	0.0449	0.0559	1	0.1269	1	1	0.0443	0.0977	0.1104	1	0.0865

4.3.5 Valores das medidas de selecção de modelos

A representação dos valores do DIC obtidos com o ajustamento dos modelos m_1, \dots, m_{16} , aos dados gerados de um dos modelos, através de caixas de bigodes em paralelo, possibilita uma melhor compreensão das variações dos valores desta medida de discrepância. Considera-se apenas as 1000 amostras geradas de m_8 (com as covariáveis x_1 e x_4) nos seis casos em análise; as caixas de bigodes em paralelo dos valores do DIC referentes ao ajustamento de m_1, \dots, m_{16} são apresentadas na figura 4.9. Da análise desta figura, destaca-se que:

Caso 1. Os valores do DIC referentes ao ajustamento de m_1 (com todas as covariáveis), m_3 (que inclui x_1 , x_2 e x_4) e m_4 (que inclui x_1 , x_3 e x_4) são idênticos aos do ajustamento do modelo “correcto”, o que se deve ao facto de todos eles incluírem as covariáveis x_1 e x_4 . O ajustamento de m_9 , m_{13} , m_{14} e m_{16} origina valores mais elevados do DIC; estes modelos são os que pior ajustam os dados, pois não incluem x_1 nem x_4 . Os restantes modelos, que incluem apenas uma das covariáveis x_1 ou x_4 , apresentam valores do DIC inferiores aos dos que não incluem qualquer delas, mas superiores aos dos modelos que incluem as duas covariáveis.

Caso 2. As caixas de bigodes referentes ao ajustamento dos modelos que não incluem a covariável x_4 destacam-se claramente das restantes caixas de bigodes, pois os valores do DIC são muito superiores aos dos restantes modelos ajustados (recorde-se que os coeficientes das covariáveis x_1 e x_4 no modelo gerador são 1 e 5, respectivamente). Ao ajustamento de m_1 , m_3 e m_4 correspondem valores do DIC idênticos aos do ajustamento do modelo “correcto”. Os valores do DIC dos modelos ajustados que não incluem as covariáveis do modelo gerador com coeficientes mais expressivos são superiores aos dos outros modelos.

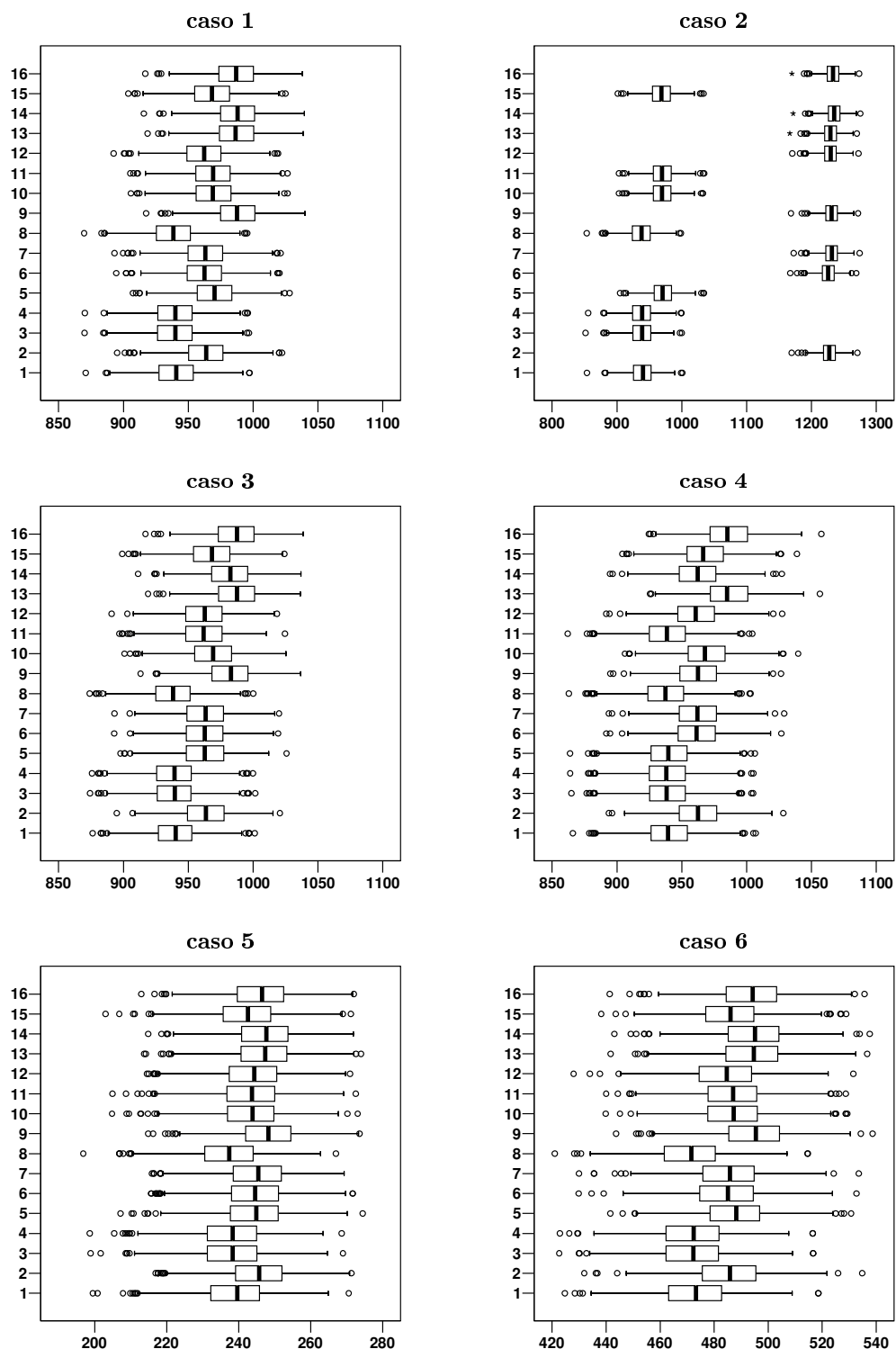


Figura 4.9: Caixas de bigodes dos valores do DIC, obtidos com os dados gerados do modelo m_8 , para os 16 modelos ajustados, nos casos 1 a 6.

Caso 3. As caixas de bigodes são muito idênticas às do caso 1, registando-se apenas uma ligeira diminuição nos valores do DIC dos modelos ajustados que não incluem x_1 mas que incluem x_3 . A existência de uma correlação linear moderada entre as covariáveis x_1 e x_3 não produz alterações significativas nos valores do DIC.

Caso 4. O ajustamento de m_1, m_3, m_4, m_5 e m_{11} , que incluem x_4 e pelo menos uma das covariáveis x_1 ou x_3 , origina valores do DIC idênticos aos do ajustamento do modelo “correcto”. Os valores do DIC correspondentes ao ajustamento de m_{13} e m_{16} , que não incluem qualquer das covariáveis x_1, x_3 e x_4 , são muito superiores aos do ajustamento do modelo “correcto”. O ajustamento dos restantes modelos dá origem a valores do DIC superiores aos do modelo “correcto”, mas inferiores aos de m_{13} e m_{16} .

Caso 5. Com $n = 50$ é mais difícil a identificação do modelo “correcto”, pois as diferenças entre as caixas de bigodes apresentam-se mais esbatidas.

Caso 6. São válidos os comentários efectuados para o caso 1, embora se deva referir que, com $n = 100$, a distinção entre os valores do DIC relativos ao ajustamento de m_1, \dots, m_{16} não é tão acentuada como no caso 1 ($n = 200$), mas é bastante mais perceptível que no caso 5 ($n = 50$).

As representações dos valores do $\widehat{\text{BIC}}$ e do LLP, através de caixas de bigodes em paralelo, para os dados gerados do modelo m_8 nos 6 casos considerados, são idênticas às obtidas para o DIC.

A replicação, neste estudo de simulação, dos valores do DIC, do $\widehat{\text{BIC}}$ e do LLP possibilita a determinação das médias e dos desvios padrão destas medidas de discrepância, para cada caso, por modelo gerador e por modelo ajustado. A análise destas estatísticas descritivas permite fazer algumas considerações sobre o “comportamento” destas medidas de discrepância.

Considera-se, em primeiro lugar, o **caso 1**. As médias e os desvios padrão do DIC, as médias do $\widehat{\text{BIC}}$ e as médias do LLP, referentes ao ajustamento dos modelos m_1, \dots, m_{16} às 1000 amostras de dados gerados de cada um dos possíveis modelos, são apresentados, respectivamente, nas tabelas E.22, E.23, E.24 e E.25, do anexo E. Destaca-se que:

- as médias do DIC e do LLP ao longo da diagonal principal (correspondentes ao ajustamento dos modelos “correctos”) são muito idênticas — variam entre 936 e 941, no caso do DIC, e entre 932 e 935, no caso do LLP; as médias do $\widehat{\text{BIC}}$ ao longo desta diagonal apresentam uma maior dispersão — variam entre 939 e 961;
- os desvios padrão do DIC são muito semelhantes em toda a tabela (variando entre 18.9 e 21.0); para o número de casas decimais consideradas, os desvios padrão do $\widehat{\text{BIC}}$ e do LLP são iguais aos registados para o DIC;
- os valores mais pequenos das médias destas medidas de discrepância localizam-se na parte triangular inferior da matriz de resultados.

Os valores do DIC não têm significado intrínseco; apenas as diferenças, entre modelos, dos valores destas medidas de discrepância têm significado. Assim, tem mais interesse os resultados da tabela 4.18, com as diferenças, por modelo gerador e modelo ajustado, entre as médias do DIC e a menor média do DIC do modelo gerador. Tal como acontece com o critério de informação da *deviance*, também os valores do $\widehat{\text{BIC}}$ e do LLP não têm significado intrínseco. As diferenças, por modelo gerador e modelo ajustado, entre as médias do $\widehat{\text{BIC}}$ e a menor média do $\widehat{\text{BIC}}$ do modelo gerador, é apresentada na tabela 4.19; para o LLP as correspondentes diferenças são apresentadas na tabela 4.20. Da análise destas tabelas salienta-se que:

- no caso do DIC apenas a diagonal principal é constituída por zeros, e os modelos ajustados que incluem outras covariáveis, além das incluídas no modelo gerador, têm valores do DIC muito idênticos aos dos modelos “correctos” — diferenças de 1 para os modelos com mais uma covariável, diferenças de 2 para os modelos com mais duas covariáveis, e diferenças de 3 para os modelos com mais três covariáveis;
- no caso do $\widehat{\text{BIC}}$ também apenas a diagonal principal é constituída por zeros, e os modelos ajustados que incluem outras covariáveis, além das incluídas no modelo gerador, apresentam valores do $\widehat{\text{BIC}}$ superiores aos dos modelos “correctos” — diferenças de 5 ou 6 para os modelos com mais uma covariável, diferenças de 10 ou 11 para os modelos com mais duas covariáveis, e diferenças de 16 para os modelos com mais três covariáveis;
- no caso do LLP, tanto a diagonal principal como as células referentes aos modelos ajustados que incluem outras covariáveis, para além das incluídas no modelo gerador, são constituídas por zeros.

Os resultados referentes aos casos 2 a 6 são apresentados nas tabelas E.26 a E.41 do anexo E. Da análise das tabelas E.26, E.30, E.33, E.36 e E.39, referentes às médias do DIC dos modelos ajustados nos casos 2 a 6, destaca-se que:

- em qualquer um dos casos, as médias do DIC mantêm-se muito semelhantes ao longo da diagonal principal da matriz de resultados;
- quando diminui a dimensão das amostras geradas, as médias do DIC referentes ao ajustamento dos modelos “correctos” também diminuem: nos casos 1 a 4 ($n = 200$) situam-se no intervalo $[935, 941]$, no caso 6 ($n = 100$) localizam-se no intervalo $[469, 473]$ e no caso 5 ($n = 50$) situam-se intervalo $[236, 239]$.

A constatação da relação existente entre as médias dos valores do DIC dos modelos “correctos” e a dimensão das amostras ajustadas sugere que uma eventual medida de adequabilidade do modelo, construída à custa do DIC, tenha em conta que este valor depende da dimensão da amostra — dividir o valor do DIC pela dimensão da amostra parece ser uma solução razoável. Analisando as tabelas do anexo E referentes ao $\widehat{\text{BIC}}$ e LLP, observa-se que uma eventual medida de adequabilidade do modelo construída à custa destas medidas de discrepância deve, também, ter em conta a sua dependência da dimensão da amostra.

Tabela 4.18: Diferenças, por modelo gerador e modelo ajustado para o caso 1, entre as médias do DIC e a menor média do modelo gerador.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	0	22	25	22	30	44	36	43	47	50	47	55	64	57	63	73
m_2	1	0	28	24	31	27	24	46	30	52	48	46	51	48	66	65
m_3	1	24	0	23	31	23	37	22	49	30	47	36	48	58	46	57
m_4	1	25	27	0	31	47	25	26	49	51	30	46	66	49	50	66
m_5	1	24	27	24	0	46	38	46	23	26	23	57	45	36	45	56
m_6	2	1	1	24	33	0	24	23	32	32	49	23	31	49	48	48
m_7	2	1	28	1	32	27	0	27	31	52	31	26	51	30	51	50
m_8	2	25	1	1	32	24	25	0	50	31	31	24	49	50	30	49
m_9	2	1	28	24	1	27	24	46	0	27	23	46	26	23	45	45
m_{10}	2	25	1	24	1	24	38	23	24	0	23	37	23	37	22	36
m_{11}	2	25	28	1	1	47	25	27	24	27	0	47	46	24	26	46
m_{12}	3	2	2	2	33	1	1	1	32	32	32	0	31	31	31	30
m_{13}	3	2	2	25	2	1	25	24	1	1	24	23	0	24	23	22
m_{14}	3	2	29	2	2	28	1	28	1	28	1	27	27	0	27	26
m_{15}	3	26	2	2	2	25	25	1	25	1	1	24	24	24	0	23
m_{16}	4	3	3	3	3	2	2	2	2	2	2	1	1	1	1	0

Tabela 4.19: Diferenças, por modelo gerador e modelo ajustado para o caso 1, entre as médias do \widehat{BIC} e a menor média do modelo gerador.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	0	18	21	18	26	36	28	35	39	42	39	43	52	45	51	56
m_2	6	0	28	24	31	23	19	42	26	48	44	37	42	39	57	53
m_3	6	24	0	23	32	19	33	18	45	26	43	28	39	50	38	45
m_4	5	24	26	0	30	42	20	21	44	46	25	37	57	40	40	53
m_5	6	24	28	24	0	42	33	41	19	22	19	49	37	28	36	43
m_6	11	5	5	29	37	0	24	23	32	32	49	19	27	45	44	39
m_7	11	5	32	5	37	27	0	27	31	52	31	22	47	26	47	41
m_8	11	30	5	5	36	24	25	0	50	31	31	20	44	45	26	40
m_9	10	5	32	28	5	27	24	46	0	27	23	41	21	18	41	36
m_{10}	10	29	5	28	5	24	38	23	24	0	23	33	18	33	18	27
m_{11}	10	29	32	5	5	47	25	27	24	27	0	42	42	19	22	37
m_{12}	16	11	11	11	42	5	5	5	36	36	36	0	31	31	31	26
m_{13}	16	10	10	33	10	5	28	28	5	5	28	23	0	23	22	18
m_{14}	16	11	38	11	11	33	6	33	6	33	6	27	28	0	28	22
m_{15}	16	35	11	11	11	30	30	6	30	6	6	25	24	25	0	19
m_{16}	22	16	16	16	16	11	11	11	11	11	11	6	6	6	6	0

Tabela 4.20: Diferenças, por modelo gerador e modelo ajustado para o caso 1, entre as médias do LLP e a menor média do modelo gerador.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	0	24	27	24	32	46	38	46	50	52	49	58	68	61	67	77
m_2	0	0	28	24	31	28	24	47	31	52	49	48	52	50	67	68
m_3	0	24	0	23	31	24	38	23	50	31	48	38	50	60	48	60
m_4	0	25	27	0	31	47	25	27	50	51	31	48	68	51	51	69
m_5	0	24	27	23	0	47	38	46	24	27	23	59	47	38	46	59
m_6	0	0	0	23	32	0	24	23	32	32	49	24	32	50	49	50
m_7	0	0	27	0	31	27	0	27	31	52	31	27	52	31	52	52
m_8	0	24	0	0	31	24	25	0	50	31	31	25	50	51	31	51
m_9	0	0	27	24	0	27	24	46	0	27	23	47	27	24	46	47
m_{10}	0	24	0	23	0	24	39	23	24	0	23	38	24	39	23	39
m_{11}	0	24	27	0	0	47	25	27	24	27	0	48	47	25	27	48
m_{12}	0	0	0	0	31	0	0	0	31	31	31	0	31	31	31	31
m_{13}	0	0	0	23	0	0	24	23	0	0	23	23	0	24	23	23
m_{14}	0	0	27	0	0	27	0	27	0	27	0	27	27	0	27	27
m_{15}	0	24	0	0	0	24	25	0	24	0	0	25	24	25	0	25
m_{16}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Embora não sejam apresentadas as tabelas, salienta-se que as variâncias do DIC são idênticas às do $\widehat{\text{BIC}}$ e LLP, não só no caso 1, mas também nos restantes casos estudados. Isto sugere que, quando se pretende efectuar comparações de modelos (e, em particular, seleccionar covariáveis) no WinBUGS, utilizando o DIC para esse efeito, pode usar-se as estimativas obtidas para a variância do $\widehat{\text{BIC}}$ ou do LLP como uma aproximação razoável da estimativa da variância do DIC.

Apresenta-se, também, na tabela 4.21, para o caso 5 ($n = 50$), as diferenças, por modelo gerador e modelo ajustado, entre as médias do DIC e a menor média do DIC do modelo gerador. Os quadros correspondentes para o $\widehat{\text{BIC}}$ e o LLP são apresentados nas tabelas 4.22 e 4.23, respectivamente.

Analisa-se, em primeiro lugar, a tabela referente ao DIC. Tal como acontece no caso 1 ($n = 200$), regista-se que a diagonal principal é constituída por zeros, e os modelos ajustados que incluem outras covariáveis, além das incluídas no modelo gerador, têm valores do DIC muito idênticos aos dos modelos “correctos” — diferenças de 0, 1 ou 2 para os modelos com mais uma covariável, diferenças de 2 ou 3 para os modelos com mais duas covariáveis, e diferenças de 3 para os modelos com mais três covariáveis. É curioso o facto das diferenças entre as médias do DIC e a menor média do DIC do modelo gerador não serem influenciadas pela dimensão da amostra, quando o modelo ajustado inclui as covariáveis contidas no modelo gerador; o mesmo acontece no caso do LLP.

No caso do critério de informação *bayesiano* estimado já não acontece o mesmo: quando o modelo ajustado inclui as covariáveis contidas no modelo gerador, as diferenças entre as médias do $\widehat{\text{BIC}}$ e a menor média do $\widehat{\text{BIC}}$ do modelo gerador são influenciadas pela dimensão da amostra.

Tabela 4.21: Diferenças, por modelo gerador e modelo ajustado para o caso 5, entre as médias do DIC e a menor média do modelo gerador.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	0	7	6	3	5	11	7	8	9	11	8	11	14	9	13	14
m_2	1	0	6	4	6	5	3	9	5	12	9	8	11	9	14	14
m_3	1	7	0	4	6	6	7	3	10	5	9	6	8	10	8	9
m_4	1	8	7	0	6	13	8	6	10	12	5	12	15	10	11	14
m_5	1	7	7	4	0	12	7	10	6	6	3	12	11	6	8	11
m_6	2	1	1	5	8	0	5	4	7	7	10	4	6	10	9	9
m_7	3	1	8	2	8	7	0	7	7	14	7	6	12	6	12	11
m_8	2	8	1	1	7	7	8	0	11	6	6	7	10	10	5	9
m_9	2	1	8	6	1	7	5	11	0	7	5	10	6	4	10	9
m_{10}	2	8	0	5	1	7	8	4	7	0	4	7	6	7	3	6
m_{11}	2	9	8	1	1	13	8	7	8	7	0	13	12	8	6	12
m_{12}	3	2	2	2	8	1	1	1	7	7	7	0	6	6	6	5
m_{13}	3	2	2	6	2	1	6	5	1	1	5	5	0	5	4	4
m_{14}	3	2	8	2	2	7	1	7	1	7	1	6	6	0	6	5
m_{15}	3	9	2	2	2	8	9	1	8	1	1	8	7	8	0	7
m_{16}	3	2	2	2	2	1	2	2	2	1	2	1	1	1	1	0

Tabela 4.22: Diferenças, por modelo gerador e modelo ajustado para o caso 5, entre as médias do \widehat{BIC} e a menor média do modelo gerador.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	0	3	2	0	2	5	1	2	3	5	2	2	5	0	4	2
m_2	3	0	6	4	6	2	0	6	2	8	6	2	4	3	8	5
m_3	4	7	0	4	6	3	4	0	6	2	6	0	2	4	2	0
m_4	4	8	7	0	6	10	4	3	7	9	2	6	9	3	5	5
m_5	4	7	7	4	0	9	4	6	3	3	0	6	5	0	2	2
m_6	8	4	4	8	10	0	5	4	6	6	10	1	3	7	6	3
m_7	8	4	11	4	10	7	0	7	7	13	6	3	9	2	9	5
m_8	8	11	4	4	10	7	8	0	11	6	6	4	7	7	2	3
m_9	8	4	11	8	4	7	5	11	0	7	4	7	3	1	7	3
m_{10}	8	12	4	9	14	8	9	5	8	0	5	5	4	5	1	1
m_{11}	8	11	11	4	4	13	8	7	8	7	0	10	9	4	3	6
m_{12}	11	8	8	7	14	4	4	4	10	10	10	0	6	6	6	2
m_{13}	12	8	8	12	8	4	9	8	4	4	8	5	0	5	4	1
m_{14}	11	8	14	8	8	10	4	10	4	10	4	6	6	0	6	2
m_{15}	12	16	8	8	8	12	13	4	12	4	4	9	8	9	0	5
m_{16}	15	11	11	11	11	7	7	7	7	7	7	3	3	4	3	0

Tabela 4.23: Diferenças, por modelo gerador e modelo ajustado para o caso 5, entre as médias do LLP e a menor média do modelo gerador.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	0	8	7	5	6	13	9	10	11	13	10	14	17	12	16	18
m_2	0	0	6	4	6	6	4	10	6	13	10	10	13	11	16	17
m_3	0	7	0	5	6	7	8	4	11	6	10	8	10	12	10	12
m_4	0	8	7	0	7	14	9	7	11	13	6	14	17	12	13	18
m_5	0	7	7	4	0	13	8	11	7	7	4	14	13	8	11	14
m_6	0	0	0	5	7	1	5	5	7	7	10	5	7	11	10	11
m_7	0	0	6	0	6	6	0	6	6	13	6	6	12	6	12	12
m_8	0	8	0	0	6	7	8	0	11	6	6	8	11	11	6	11
m_9	1	0	7	5	1	7	5	11	0	7	5	11	7	5	11	11
m_{10}	0	7	0	4	0	7	8	4	7	0	4	8	7	8	4	8
m_{11}	0	8	7	0	0	13	8	7	8	7	0	14	13	9	7	14
m_{12}	0	0	0	0	6	0	0	0	6	6	6	0	6	6	6	6
m_{13}	0	0	0	4	0	0	5	4	0	0	4	5	0	5	4	5
m_{14}	0	0	6	0	0	6	0	6	0	6	0	6	6	0	6	6
m_{15}	0	7	0	0	0	7	8	0	7	0	0	8	7	8	0	8
m_{16}	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1

Capítulo 5

Validação de Modelos

5.1 Introdução

Após a selecção de um modelo, que se pretende descreva o melhor possível um determinado fenómeno aleatório, é necessário efectuar a validação do mesmo, isto é, averiguar se esse modelo é adequado para realizar as inferências necessárias de modo a dar resposta às questões de interesse. A validação de modelos não tem como objectivo encontrar uma resposta para a pergunta “*Is our model true or false?*”. A questão relevante, para a qual interessa encontrar uma resposta, é “*Do the model’s deficiencies have a noticeable effect on the substantive inferences?*” (Gelman *et al.*, 1995, pág. 162). Esta ideia é também preconizada por Box (1976) ao referir que todos os modelos são errados mas alguns são úteis.

Convém relembrar que a referência a modelo inclui a distribuição amostral condicional ao vector de parâmetros, a distribuição *a priori* do vector de parâmetros, a estrutura hierárquica (caso exista) e, no caso de modelos de regressão, as variáveis explicativas. A validação do modelo engloba, portanto, a validação de todas estas componentes.

Gelman *et al.* (1996) referem que na Estatística Bayesiana a validação de modelos deve contemplar, pelo menos, três aspectos: (i) examinar a sensibilidade das inferências a alterações razoáveis da distribuição *a priori*, da verosimilhança ou de outros aspectos do modelo; (ii) averiguar se as inferências efectuadas são razoáveis dado o conhecimento substantivo do problema; (iii) verificar se o modelo ajusta bem os dados. O primeiro aspecto da validação de modelos, usualmente designado por análise de sensibilidade, não é contemplado neste trabalho. Alguns aspectos da análise de sensibilidade podem ser vistos em Carlin e Louis (2000, secção 6.1). O segundo aspecto da validação de modelos pode ser informalmente verificado através da análise cuidadosa dos resultados obtidos, verificando se são razoáveis, atendendo ao conhecimento substantivo do problema. Este capítulo debruça-se essencialmente sobre o terceiro aspecto da validação de modelos, que tem na literatura Estatística vários sinónimos, tais como estudo da adequabilidade do modelo e da bondade de ajustamento do modelo.

Embora sejam referidos diversos métodos de estudo da adequabilidade do modelo, é dado destaque ao método que se baseia na comparação, através do cálculo de um *p-value*, da amostra observada com amostras simuladas da distribuição preditiva *a posteriori*.

Alguns autores sugerem o uso da distribuição preditiva *a priori*, $f(x)$, na validação do modelo. Os valores observados x_k , do vector x , para os quais o valor da distribuição marginal $f_{X_k}(x_k)$ é demasiado pequeno, quando comparado com outros valores de $f_{X_i}(x_i)$, para $i \neq k$, são “improváveis”, devendo ser considerados *outliers* sob a validade do modelo proposto; se existirem muitas observações nestas condições, então o próprio modelo deve ser considerado inadequado e outros modelos devem ser propostos. O problema desta abordagem é a dificuldade em definir a partir de que valor $f_{X_i}(x_i)$ deve ser considerado pequeno, e que número de *outliers* deve ser considerado demasiado (Carlin e Louis, 2000, secção 2.4). Este procedimento de validação da adequabilidade do modelo tem o inconveniente de não poder ser usado quando a distribuição preditiva *a priori* $f(x)$ é imprópria, o que ocorre frequentemente quando se utilizam distribuições *a priori* não informativas.

Pelos motivos mencionados, a maior parte dos métodos de validação da qualidade do ajustamento de modelos usa a distribuição preditiva *a posteriori*. Destaca-se dois processos de validação que utilizam esta distribuição: uma validação informal, essencialmente gráfica, e uma validação formal através da determinação de medidas de surpresa, também designadas medidas de diagnóstico ou medidas de adequabilidade do modelo.

Tal como na abordagem clássica, a validação informal de modelos pode ser feita através da análise gráfica dos resíduos. Por exemplo, no caso de modelos de regressão linear normal, um diagrama de dispersão dos resíduos bayesianos (r_i ou d_i) ou dos resíduos bayesianos de eliminação (r'_i ou d'_i) *versus* os valores preditos das observações (\hat{y}_i) pode permitir, caso seja possível vislumbrar algum padrão no gráfico, a detecção de desvios à suposição de homogeneidade da variância dos resíduos. Estes gráficos de resíduos possibilitam também a identificação de eventuais *outliers*. A representação gráfica dos resíduos *versus* os números das observações é útil para detectar situações de falta de independência. Um diagrama de dispersão das ordenadas preditivas condicionais *versus* os números das observações permite a detecção de possíveis candidatos a *outliers*, através da identificação das observações às quais corresponde um pequeno valor da CPO.

Segundo Bayarri e Berger (1997), a questão essencial da validação da qualidade de ajustamento do modelo pode ser colocada nos seguintes termos: “formulado o modelo e observados os dados, serão esses mesmos dados surpreendentes?”. Muitos autores defendem que a quantificação da surpresa nos dados pode, e deve, ser feita à custa de medidas de surpresa — das quais se destaca os *p-values* — que quantificam o grau de incompatibilidade dos dados com o modelo proposto, sem recurso a modelos alternativos. Uma descrição pormenorizada de diversas medidas de surpresa pode ser encontrada em Bayarri e Berger (1997). Nesta trabalho é dado especial destaque ao *p-value* como medida de adequabilidade do modelo.

O maior inconveniente na utilização do *p-value* é a interpretação incorrecta do seu significado por parte de alguns utilizadores da Estatística. Interpretar o *p-value* como a probabilidade de rejeitar o modelo proposto, quando o modelo proposto é verdadeiro, ou mesmo como a probabilidade do modelo proposto ser verdadeiro, são dois erros muito frequentes; para obviar este problema, Bayarri e Berger (1997) e Sellke *et al.* (1999) propõem calibrações para o *p-value*.

5.2 *p-values*

Considere-se, para $X = (X_1, \dots, X_n)$, o modelo $m : X|\theta \sim f(x|\theta)$, com $\theta \in \Theta$. Suponha-se ainda que foi observada a amostra $x_{obs} = (x_{obs,1}, \dots, x_{obs,n})$. Pode investigar-se a compatibilidade do modelo com x_{obs} , recorrendo a uma quantidade de teste $T = t(X)$ e calculando a medida de surpresa ou adequabilidade do modelo

$$p(x_{obs}) = \Pr(t(X) \geq t(x_{obs})), \quad (5.1)$$

que se designa *p-value*. O *p-value* é a probabilidade de uma dada quantidade de teste T ser maior ou igual que o seu valor observado, sob a hipótese de validade do modelo m . A quantidade de teste T deve ser escolhida de modo que valores elevados de T indiquem uma menor compatibilidade dos dados com o modelo.

Se o modelo estiver completamente especificado, isto é, se θ for conhecido, o cálculo da probabilidade em (5.1) é feito com respeito à densidade de T , $f_T(t|\theta)$. No entanto, na maior parte dos casos, o modelo depende de um vector de parâmetros θ desconhecido. Assim, para determinar o *p-value*, é necessário “remover” θ ; várias sugestões têm sido apresentadas, dando origem a diferentes propostas de *p-values*.

Do lado frequentista, a solução mais usual consiste em substituir os parâmetros desconhecidos por estimativas. O *p-value* por estimativa directa e o *p-value* por estimativa directa condicional são obtidos deste modo (*e.g.*, Bayarri e Berger, 2000).

A abordagem bayesiana a este problema pressupõe que se considere um modelo hierárquico

$$m : X|\theta \sim f(x|\theta), \quad \theta \sim h(\cdot),$$

onde $h(\cdot)$ é a distribuição *a priori* de θ , que pode ser constituída por vários níveis de hierarquia. A solução tipicamente bayesiana para eliminação de parâmetros consiste em integrar no correspondente vector de parâmetros. Várias propostas têm sido apresentadas: o *p-value* preditivo *a priori* (Box, 1980), o *p-value* preditivo *a posteriori* (Guttman, 1967 e Rubin, 1984), o *p-value* de discrepância (Gelman *et al.*, 1995, Gelman *et al.*, 1996 e Meng, 1994).

Em Bayarri e Berger (1997, 1999, 2000) é dado destaque à utilização do *p-value* como medida de quantificação da surpresa nos dados, sendo descritos, comparados e criticados alguns dos *p-values* mencionados e propostos dois novos *p-values*: o *p-value* preditivo condicional e o *p-value* preditivo *a posteriori* parcial. Na maior parte das situações, estes *p-values* não são possíveis de determinar analiticamente e, além disso, os métodos de simulação requeridos para a sua determinação via métodos MCMC são bastante mais complicados que os necessários para determinar os *p-values* preditivos *a priori* e *a posteriori*. Por este motivo, esses *p-values* são pouco utilizados nas aplicações práticas e, conseqüentemente, não são objecto de estudo neste trabalho.

Uma questão essencial quando se utiliza os *p-values* na validação de modelos é a escolha adequada da quantidade de teste T . Gelman *et al.* (1995, pág. 172) defendem que a escolha de T deve ser feita de acordo com os aspectos do modelo que se pretende validar e, uma vez que o ajustamento do modelo pode falhar nos mais diversos aspectos, devem ser utilizadas diversas quantidades de teste de modo a avaliar várias possibilidades de falta de ajustamento. Além disso, Carlin e Louis (2000, pág. 47) referem que T não deve

focalizar-se em aspectos parametrizados pelo modelo, pois estes estão automaticamente ajustados na distribuição *a posteriori*.

p-value* preditivo *a priori

Conforme foi referido, a solução bayesiana para eliminação de θ consiste em integrar nesse vector de parâmetros. A distribuição preditiva *a priori* de T , $f(t) = \int f(t|\theta) h(\theta) d\theta$, não depende de θ e pode ser usada no cálculo do *p-value*, dando origem ao *p-value* preditivo *a priori*, que se representa por p_{prior} .

A crítica mais usual a este *p-value* é o facto de depender da distribuição *a priori* $h(\theta)$. No caso do modelo não ser validado, fica por esclarecer se esse facto é devido à inadequabilidade da distribuição de X , condicional ao vector de parâmetros θ , ou à inadequabilidade da distribuição *a priori* de θ . A utilização de distribuições *a priori* não informativas não é solução, pois estas são tipicamente impróprias, impossibilitando o cálculo do *p-value* preditivo *a priori*, tal como acontece no caso do exemplo 5.1. Outro inconveniente da utilização do *p-value* preditivo *a priori* é o facto da obtenção de estimativas via métodos MCMC não ser tão fácil como no caso do *p-value* preditivo *a posteriori* (secção B.1 do anexo B).

p-value* preditivo *a posteriori

Para eliminar o vector de parâmetros θ , Guttman (1967) e Rubin (1984) sugerem que o *p-value* seja calculado com recurso à distribuição preditiva *a posteriori* de T , $p(t|x_{obs}) = \int f(t|\theta) h(\theta|x_{obs}) d\theta$, dando origem ao *p-value* preditivo *a posteriori*, que se representa por p_{post} .

A utilização do *p-value* preditivo *a posteriori* permite que distribuições *a priori* não informativas (geralmente impróprias) possam ser usadas, uma vez que a distribuição *a posteriori* é, usualmente, própria. Como a distribuição preditiva *a posteriori* $p(t|x_{obs})$ é mais influenciada por $f(t|\theta)$ do que por $h(\theta)$, este *p-value* não é tão sensível a alterações da distribuição *a priori*. Outra vantagem deste *p-value* é a sua facilidade de cálculo através do *output* do método MCMC (Spiegelhalter *et al.*, secção 9.4, 1996a). No exemplo 5.1 ilustra-se, com o modelo normal, o modo de obter estimativas do *p-value* preditivo *a posteriori* via métodos MCMC.

Bayarri e Berger (2000) não recomendam a utilização do *p-value* preditivo *a posteriori* devido ao “duplo uso dos dados”. Conforme refere Stern (no comentário aos artigos de Bayarri e Berger, 2000 e Robins *et al.*, 2000) “*It seems as if they are suggesting an inappropriate use of the data, as occurs, for example, when the same data are used to identify a prior distribution and then again in the likelihood function*”, acrescentando que os referidos autores “*might prefer to use the posterior predictive distribution to calculate the probability of another event or might prefer to use another distribution in place of the posterior predictive distribution, but this does not make the posterior predictive approach invalid*”.

A crítica mais frequente a este *p-value* refere-se, na realidade, ao “duplo uso dos dados”. Os defensores desta ideia argumentam que não é apropriado utilizar os mesmos dados na

determinação da distribuição preditiva *a posteriori* $p(t|x_{obs})$ e no cálculo da área de cauda correspondente a $t(x_{obs})$. No caso da dimensão da amostra ser suficientemente elevada, Carlin e Louis (2000, pág. 48) defendem a adoção de um procedimento do tipo validação cruzada, de modo a evitar esta eventual dupla utilização dos dados: dividir a amostra x_{obs} em duas partes, $x_{obs} = (x_{obs}^{(1)}, x_{obs}^{(2)})$, sendo $x_{obs}^{(1)}$ a amostra de ajustamento e $x_{obs}^{(2)}$ a amostra de validação. A amostra de ajustamento é usada para determinar a distribuição preditiva *a posteriori*, $p(t|x_{obs}^{(1)})$ e a amostra de validação é usada para determinar a área de cauda correspondente a $t(x_{obs}^{(2)})$. No entanto, este procedimento nem sempre pode ser utilizado, como acontece no caso de amostras com pequena dimensão. Mas, mesmo que a dimensão da amostra seja elevada, a questão “como dividir a amostra?” pode inibir a utilização deste procedimento de validação cruzada.

Duas questões podem ser colocadas: (i) Existe, na realidade, uma dupla utilização dos dados? (ii) Caso exista, que tipo de impacto tem na estimativa do *p-value* preditivo *a posteriori*? Relativamente à questão (i), tal como refere Stern, a existir uma dupla utilização dos dados, essa situação não é comparável com o que acontece quando os mesmos dados são usados para identificar a distribuição *a priori* e na função de verosimilhança. Estudos de simulação podem ajudar a encontrar resposta para a questão (ii). Sugere-se, como trabalho futuro, a simulação de amostras (com dimensão apropriada) de um determinado modelo. Para cada uma das amostras propõe-se que se determine as estimativas do *p-value* preditivo *a posteriori* do modo usual, e usando o método de validação cruzada defendido por Carlin e Louis (2000, pág. 48). A análise dos resultados obtidos pode dar indicações importantes sobre a questão (ii).

Exemplo 5.1

Veja-se um exemplo utilizado por Bayarri e Berger (2000) para criticar o *p-value* preditivo *a posteriori*. Considere-se as variáveis aleatórias X_i , para $i = 1, \dots, n$, independentes e identicamente distribuídas $N(0, \sigma^2)$, com σ^2 desconhecido. Assim, a função densidade de probabilidade de $X = (X_1, \dots, X_n)$ é

$$f(x|\sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2\right\}, \quad x \in \mathbb{R}^n, \sigma > 0.$$

Como $\sum_{i=1}^n x_i^2/n = s^2 + \bar{x}^2$, então

$$f(x|\sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{n(s^2 + \bar{x}^2)}{2\sigma^2}\right\}, \quad x \in \mathbb{R}^n, \sigma > 0. \quad (5.2)$$

Considere-se ainda a distribuição *a priori* não informativa usual para σ^2 , isto é, $h(\sigma^2) = 1/\sigma^2$. O modelo bayesiano constituído pela distribuição de X , condicional a σ^2 , e pela distribuição *a priori* de σ^2 , representa-se por m_N .

A quantidade de teste $t(X) = |\bar{X}|$ é escolhida para investigar a compatibilidade de m_N com o vector de dados x_{obs} . Assim, conforme apresentado em (5.1), o *p-value* é dado por

$$p(x_{obs}) = \Pr(|\bar{X}| > |\bar{x}_{obs}|),$$

pois $t(x_{obs}) = |\bar{x}_{obs}|$.

Para determinar o p -value preditivo *a posteriori* é necessário calcular a distribuição *a posteriori* de σ^2 , $h(\sigma^2|x_{obs})$, e a distribuição preditiva *a posteriori* $p(\bar{x}|x_{obs})$. Determina-se, em primeiro lugar,

$$h(\sigma^2|x_{obs}) = \frac{f(x_{obs}|\sigma^2) h(\sigma^2)}{\int_0^{+\infty} f(x_{obs}|u^2) h(u^2) du^2}.$$

Atendendo a (5.2), a distribuição *a posteriori* de σ^2 é

$$h(\sigma^2|x_{obs}) = \frac{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{n(s_{obs}^2 + \bar{x}_{obs}^2)}{2\sigma^2}\right\} \frac{1}{\sigma^2}}{\int_0^{+\infty} \left(\frac{1}{u\sqrt{2\pi}}\right)^n \exp\left\{-\frac{n(s_{obs}^2 + \bar{x}_{obs}^2)}{2u^2}\right\} \frac{1}{u^2} du^2} \quad (5.3)$$

e pode multiplicar-se o numerador e o denominador de (5.3) por $\frac{1}{\Gamma(\frac{n}{2})} \left[\frac{n(s_{obs}^2 + \bar{x}_{obs}^2)}{2}\right]^{\frac{n}{2}}$, obtendo-se

$$h(\sigma^2|x_{obs}) = \frac{\frac{1}{\Gamma(\frac{n}{2})} \left[\frac{n(s_{obs}^2 + \bar{x}_{obs}^2)}{2}\right]^{\frac{n}{2}} (\sigma^2)^{-(\frac{n}{2}+1)} \exp\left\{-\frac{n(s_{obs}^2 + \bar{x}_{obs}^2)}{2\sigma^2}\right\}}{\int_0^{+\infty} \frac{1}{\Gamma(\frac{n}{2})} \left[\frac{n(s_{obs}^2 + \bar{x}_{obs}^2)}{2}\right]^{\frac{n}{2}} (u^2)^{-(\frac{n}{2}+1)} \exp\left\{-\frac{n(s_{obs}^2 + \bar{x}_{obs}^2)}{2u^2}\right\} du^2}.$$

Como a função integranda do integral do denominador é a densidade de probabilidade da gama invertida de parâmetro de forma $n/2$ e parâmetro de escala $n(s_{obs}^2 + \bar{x}_{obs}^2)/2$, este integral é unitário e

$$h(\sigma^2|x_{obs}) = \frac{\left[\frac{n(s_{obs}^2 + \bar{x}_{obs}^2)}{2}\right]^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} (\sigma^2)^{-(\frac{n}{2}+1)} \exp\left\{-\frac{n(s_{obs}^2 + \bar{x}_{obs}^2)}{2\sigma^2}\right\},$$

donde $\sigma^2|x_{obs} \sim \text{GI}(n/2, n(s_{obs}^2 + \bar{x}_{obs}^2)/2)$.

A distribuição preditiva *a posteriori* de \bar{X} é

$$p(\bar{x}|x_{obs}) = \int_0^{+\infty} f(\bar{x}|\sigma^2) h(\sigma^2|x_{obs}) d\sigma^2.$$

Como as variáveis aleatórias X_i são independentes e identicamente distribuídas $N(0, \sigma^2)$, para $i = 1, \dots, n$, logo $\bar{X} \sim N(0, \sigma^2/n)$, donde

$$f(\bar{x}|\sigma^2) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{n\bar{x}^2}{2\sigma^2}\right\}.$$

Deste modo, obtém-se

$$\begin{aligned} p(\bar{x}|x_{obs}) &= \\ &= \int_0^{+\infty} \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{n\bar{x}^2}{2\sigma^2}\right\} \frac{\left[\frac{n(s_{obs}^2 + \bar{x}_{obs}^2)}{2}\right]^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} (\sigma^2)^{-(\frac{n}{2}+1)} \exp\left\{-\frac{n(s_{obs}^2 + \bar{x}_{obs}^2)}{2\sigma^2}\right\} d\sigma^2 \\ &= \frac{\sqrt{n}}{\sqrt{2\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left[\frac{n(s_{obs}^2 + \bar{x}_{obs}^2 + \bar{x}^2)}{2}\right]^{-\frac{n+1}{2}} \left[\frac{n(s_{obs}^2 + \bar{x}_{obs}^2)}{2}\right]^{\frac{n}{2}} \times \\ &\quad \times \int_0^{+\infty} \frac{\left[\frac{n(s_{obs}^2 + \bar{x}_{obs}^2 + \bar{x}^2)}{2}\right]^{\frac{n+1}{2}}}{\Gamma\left(\frac{n+1}{2}\right)} (\sigma^2)^{-(\frac{n+1}{2}+1)} \exp\left\{-\frac{n(s_{obs}^2 + \bar{x}_{obs}^2 + \bar{x}^2)}{2\sigma^2}\right\} d\sigma^2. \end{aligned}$$

Como a função integranda do integral da expressão anterior é a densidade de probabilidade da gama invertida de parâmetros $(n+1)/2$ e $n(s_{obs}^2 + \bar{x}_{obs}^2 + \bar{x}^2)/2$, este integral é unitário e

$$\begin{aligned} p(\bar{x}|x_{obs}) &= \frac{\sqrt{n}}{\sqrt{2\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left[\frac{n(s_{obs}^2 + \bar{x}_{obs}^2 + \bar{x}^2)}{2} \right]^{-\frac{n+1}{2}} \left[\frac{n(s_{obs}^2 + \bar{x}_{obs}^2)}{2} \right]^{\frac{n}{2}} \\ &= \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2}) \sqrt{n\pi} \sqrt{\frac{s_{obs}^2 + \bar{x}_{obs}^2}{n}}} \left[\frac{s_{obs}^2 + \bar{x}_{obs}^2 + \bar{x}^2}{s_{obs}^2 + \bar{x}_{obs}^2} \right]^{-\frac{n+1}{2}} \\ &= \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2}) \sqrt{n\pi} \sqrt{\frac{s_{obs}^2 + \bar{x}_{obs}^2}{n}}} \left[1 + \frac{1}{n} \frac{\bar{x}^2}{\frac{s_{obs}^2 + \bar{x}_{obs}^2}{n}} \right]^{-\frac{n+1}{2}}. \end{aligned}$$

Pode assim concluir-se que a distribuição preditiva *a posteriori* de \bar{X} é a *t*-Student com n graus de liberdade, parâmetro de localização nulo e parâmetro de escala $(s_{obs}^2 + \bar{x}_{obs}^2)/n$. Deste modo, obtém-se

$$p_{post}(x_{obs}) = \Pr \left(\frac{\sqrt{n} |\bar{X}|}{\sqrt{s_{obs}^2 + \bar{x}_{obs}^2}} > \frac{\sqrt{n} |\bar{x}_{obs}|}{\sqrt{s_{obs}^2 + \bar{x}_{obs}^2}} \right) = 2 \left[1 - \Upsilon_n \left(\frac{\sqrt{n} |\bar{x}_{obs}|}{\sqrt{s_{obs}^2 + \bar{x}_{obs}^2}} \right) \right],$$

onde $\Upsilon_n(\cdot)$ representa a função distribuição da *t*-Student com n graus de liberdade, parâmetro de localização 0 e parâmetro de escala 1.

Um dos problemas deste *p-value* é o facto de $p_{post} \rightarrow 2[1 - \Upsilon_n(\sqrt{n})]$, uma constante positiva, à medida que $|\bar{x}_{obs}|/s_{obs} \rightarrow \infty$. Assim, para n fixo, mesmo que exista uma evidência muito forte contra o modelo m_N , o *p-value* preditivo *a posteriori* não tende para 0, conforme se pode constatar pela observação da tabela 5.1.

Tabela 5.1: Valores limite do p_{post} no modelo normal, à medida que $|\bar{x}_{obs}|/s_{obs} \rightarrow \infty$.

n	$2[1 - \Upsilon_n(\sqrt{n})]$
4	0.1161
5	0.0756
6	0.0498
7	0.0331
8	0.0222
9	0.0150
10	0.0101
15	0.0015
20	0.0002

Bayarri e Berger (2000) justificam que a inadequabilidade de p_{post} pode ser devida à dupla utilização dos dados, em particular pelo facto de \bar{x}_{obs} estar envolvido quer na determinação da distribuição *a posteriori* quer no cálculo da área de cauda.

Para concretizar a ilustração do cálculo do p -value preditivo *a posteriori* simula-se amostras x_1, \dots, x_7 , de dimensão $n = 20$, constituídas por observações independentes e identicamente distribuídas $N(0, 1)$, $N(0.3, 1)$, $N(0.5, 1)$, $N(1, 1)$, $N(2, 1)$, $N(0, 4)$ e $t_{10}(0, 1)$, respectivamente. Sob a validade do modelo m_N , e com base em cada uma destas amostras, calcula-se os p -values preditivos *a posteriori*. Os resultados obtidos são apresentados na tabela 5.2.

Tabela 5.2: Cálculo dos p -values preditivos *a posteriori* no modelo normal, para amostras geradas de várias distribuições.

Distribuição	N(0, 1)	N(0.3, 1)	N(0.5, 1)	N(1, 1)	N(2, 1)	N(0, 4)	$t_{10}(0, 1)$
Amostra	x_1	x_2	x_3	x_4	x_5	x_6	x_7
	0.890	0.476	0.675	1.671	2.370	-0.150	-4.651
	-0.509	0.535	2.085	2.595	3.494	1.356	-0.777
	-1.386	-0.516	0.376	1.659	2.106	-4.167	2.509
	1.128	0.740	0.800	-0.341	3.247	0.858	2.099
	0.760	-1.204	0.156	-0.069	2.492	2.160	-1.569
	-1.257	2.760	-0.042	0.991	1.201	-1.900	-0.355
	-0.103	-0.142	-0.838	0.934	2.244	5.503	-0.550
	1.682	2.206	-0.392	0.803	1.803	-2.048	-0.062
	0.229	0.267	-0.039	2.328	0.243	1.542	0.959
	-0.041	0.045	1.333	1.860	2.418	0.096	-0.493
Observações	0.689	-0.273	-0.046	0.744	2.278	-0.747	0.095
	1.326	0.256	0.801	0.599	2.158	-0.164	1.888
	0.114	-0.441	0.212	1.007	2.595	-1.483	0.286
	-0.791	-0.710	1.477	1.740	1.203	-2.564	0.185
	2.124	1.133	0.919	-1.123	1.573	1.660	-0.030
	0.776	1.755	0.098	1.602	0.541	-0.952	2.799
	-0.761	-0.281	2.669	1.721	1.174	2.722	0.508
	0.528	1.801	-0.387	0.052	1.992	-0.094	0.729
	0.153	0.802	0.696	1.080	0.989	1.397	-0.707
	0.278	1.228	0.656	1.085	1.590	-0.864	1.327
\bar{x}_{obs}	0.291	0.522	0.560	1.047	1.886	0.108	0.210
s_{obs}^2	0.827	1.029	0.694	0.797	0.659	4.400	2.535
p_{post}	0.188	0.054	0.021	0.003	5.47E-04	0.821	0.566

Como seria de esperar, os p -values obtidos para as amostras x_1 e x_6 são elevados, não havendo motivo para suspeitar da adequabilidade do modelo m_N . No caso da amostra x_2 o p -value está próximo de 0.05, pondo em dúvida a validade do modelo m_N . Por sua vez, os p -values obtidos para as amostras x_3 , x_4 e x_5 estão mais próximos de zero, sugerindo que o modelo proposto não é adequado. A amostra x_7 origina um p -value elevado, não sugerindo a rejeição do modelo proposto; este facto não é de estranhar uma vez que a distribuição $t_{10}(0, 1)$ é muito semelhante à distribuição $N(0, 1)$, com diferenças apenas nas caudas, que a quantidade de teste $t(X) = |\bar{X}|$ não permite detectar. ■

5.3 *p-value* preditivo *a posteriori*

Apenas em casos simples, como o do exemplo 5.1, é possível determinar analiticamente o *p-value* preditivo *a posteriori*. Na maior parte das situações a solução mais adequada, para validar o ajustamento do modelo aos dados, consiste em estimar p_{post} , simulando amostras da distribuição preditiva *a posteriori* e comparando, à custa de uma quantidade de teste $t(X)$, as amostras obtidas com a amostra observada.

Seja X^{rep} um vector independente e identicamente distribuído com X (isto é, uma réplica de X) e

$$p(x^{rep}|x_{obs}) = \int_{\Theta} f(x^{rep}|\theta) h(\theta|x_{obs}) d\theta, \quad (5.4)$$

a correspondente distribuição preditiva *a posteriori*. O vector $x^{rep} = (x_1^{rep}, x_2^{rep}, \dots, x_n^{rep})$, obtido precisamente nas mesmas condições de x_{obs} , constitui uma réplica de x_{obs} . O vector de dados replicados pode ser entendido como um vector de dados que poderia ter sido observado ou, numa perspectiva preditiva, que poderá ser futuramente observado se a experiência que originou a amostra x_{obs} for repetida no futuro em condições idênticas.

A falta de ajustamento do modelo aos dados pode ser avaliada comparando, à custa da quantidade de teste $T = t(X)$, a amostra observada com as amostras simuladas da distribuição preditiva *a posteriori*, através do cálculo do *p-value*

$$p_{post}(x_{obs}) = \Pr^{p(x^{rep}|x_{obs})}[t(X^{rep}) \geq t(x_{obs})],$$

onde o expoente $p(x^{rep}|x_{obs})$ na probabilidade indica que esta é calculada com respeito à distribuição preditiva *a posteriori* de X^{rep} dado x_{obs} . Deste modo,

$$p_{post}(x_{obs}) = \int I_A(x^{rep}) p(x^{rep}|x_{obs}) dx^{rep},$$

sendo $A = \{x^{rep} : t(x^{rep}) \geq t(x_{obs})\}$ e I_A a função indicatriz do conjunto A , isto é,

$$I_A(x^{rep}) = \begin{cases} 1 & \text{se } x^{rep} \in A \\ 0 & \text{se } x^{rep} \notin A \end{cases}.$$

Atendendo a (5.4) obtém-se

$$p_{post}(x_{obs}) = \iint I_A(x^{rep}) f(x^{rep}|\theta) h(\theta|x_{obs}) d\theta dx^{rep}.$$

Para estimar o *p-value* preditivo *a posteriori* procede-se do seguinte modo. Simula-se J vectores da distribuição *a posteriori* de θ e, para cada um dos $\theta^{rep,j}$ ($j = 1, \dots, J$) simula-se uma amostra $x^{rep,j}$ da distribuição $f(x|\theta^{rep,j})$. Obtém-se, deste modo, J amostras da distribuição preditiva *a posteriori* de X^{rep} e calcula-se $t(x^{rep,j})$. Os valores de $t(x^{rep,j})$ são comparados com o valor de $t(x_{obs})$, permitindo estimar o *p-value* preditivo *a posteriori* como sendo a proporção das J iterações para as quais $t(x^{rep,j})$ é maior ou igual a $t(x_{obs})$,

$$\hat{p}_{post}(x_{obs}) = \frac{1}{J} \sum_{j=1}^J I_A(x^{rep,j}).$$

Podem representar-se graficamente, através de um histograma, os valores simulados de $t(x^{rep,j})$ e localizar o valor de $t(x_{obs})$ nesse gráfico; deste modo tem-se uma perspectiva gráfica da estimativa do p -value preditivo *a posteriori*.

Exemplo 5.1 (continuação)

A obtenção, por simulação, de estimativas do p -value preditivo *a posteriori*, para cada uma das amostras x_1 a x_7 da tabela 5.2, é uma tarefa bastante simples. O código do WinBUGS para determinar \hat{p}_{post} é

```

model
{
  for (i in 1:20){
    X[i]~dnorm(0,tau)
  }
  tau~dgamma(0.0001,0.0001)
  sigma<-1/sqrt(tau)
  for (i in 1:20){
    X.rep[i]~dnorm(0,tau)
  }
  T<-abs(mean(X[]))
  T.rep<-abs(mean(X.rep[]))
  p<-step(T.rep-T)
}

```

Para estimar o valor do p -value preditivo *a posteriori* referente à amostra x_1 os valores de *input* são dados pela instrução `list(X=c(0.890,-0.509,...,0.278))`. A estimativa do p_{post} obtém-se por monitorização de `p`. Repete-se este procedimento para as outras amostras. Os resultados obtidos são apresentados na tabela 5.3. As estimativas obtidas por simulação são muito idênticas às obtidas analiticamente (ver tabela 5.2).

A distribuição preditiva *a posteriori* de $t(X^{rep})$ pode ser aproximada pelo histograma dos 10000 valores simulados de $t(X^{rep})$, permitindo a comparação do valor observado da quantidade de teste com a referida distribuição. Na figura 5.1 representa-se, para as amostras x_1 , x_5 e x_7 , o valor $t(x_{obs})$, que é comparado com a estimativa da distribuição preditiva *a posteriori* de $t(X^{rep})$.

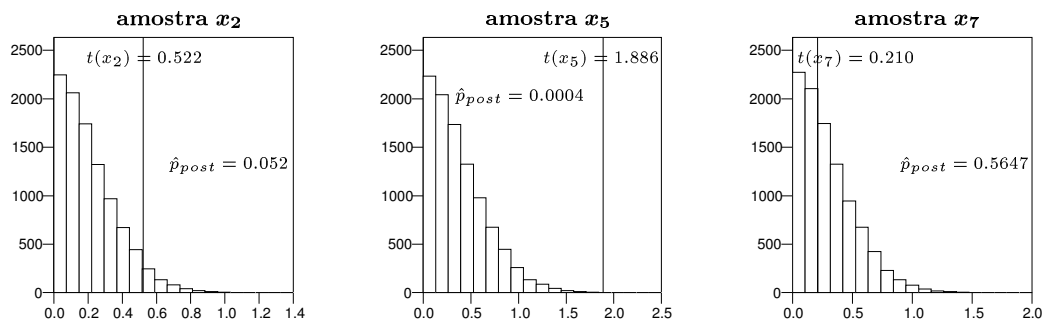


Figura 5.1: Valor observado de $t(X)$ (linha vertical), comparado com os 10000 valores simulados da distribuição preditiva *a posteriori* de $t(X^{rep})$, para as amostras x_2 , x_5 e x_7 .

Tabela 5.3: *Outputs* do WinBUGS com as estimativas dos *p-values* preditivos *a posteriori* para amostras geradas de várias distribuições.

amostra x1								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
T.rep	0.1768	0.1378	0.001456	0.006275	0.1468	0.5182	1001	10000
p	0.1885	0.3911	0.003614	0	0	1	1001	10000
amostra x2								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
T.rep	0.2112	0.1646	0.001739	0.007496	0.1753	0.619	1001	10000
p	0.052	0.222	0.002454	0	0	1	1001	10000
amostra x3								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
T.rep	0.1858	0.1449	0.001531	0.006596	0.1543	0.5447	1001	10000
p	0.0208	0.1427	0.001509	0	0	0	1001	10000
amostra x4								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
T.rep	0.2547	0.1986	0.002098	0.009041	0.2115	0.7466	1001	10000
p	0.0022	0.04685	4.40E-04	0	0	0	1001	10000
amostra x5								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
T.rep	0.38	0.2963	0.00313	0.01349	0.3155	1.114	1001	10000
p	4.00E-04	0.02	1.97E-04	0	0	0	1001	10000
amostra x6								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
T.rep	0.3888	0.3031	0.003203	0.0138	0.3228	1.14	1001	10000
p	0.82	0.3842	0.004008	0.00E+00	1	1	1001	10000
amostra x7								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
T.rep	0.2972	0.2317	0.002448	0.01055	0.2468	0.8713	1001	10000
p	0.5647	0.4958	0.005262	0	1	1	1001	10000

5.3.1 *p-value* de discrepância

Gelman (1995, 1996) e Meng (1994) defendem a generalização do *p-value* preditivo *a posteriori*, que consiste em considerar que a quantidade de teste T depende não só de X , mas também do vector de parâmetros do modelo, isto é, $T = t(X, \theta)$.

Esta generalização do *p-value* preditivo *a posteriori*, que Robins *et al.* (2000) designam *p-value* de discrepância e representam por p_{dis} , é definida como a probabilidade de $t(X^{rep}, \theta)$ ser tão ou mais extremo que $t(x_{obs}, \theta)$,

$$p_{dis}(x_{obs}) = \Pr^{p(x^{rep}, \theta | \cdot)}[t(X^{rep}, \theta) \geq t(x_{obs}, \theta)],$$

onde o expoente $p(x^{rep}, \theta | \cdot)$ indica que a probabilidade é obtida com respeito à distribuição conjunta *a posteriori* de X^{rep} e θ ,

$$p(x^{rep}, \theta | x_{obs}) = f(x^{rep} | x_{obs}, \theta) h(\theta | x_{obs}).$$

Como X^{rep} é independente de X , logo $f(x^{rep} | x_{obs}, \theta) = f(x^{rep} | \theta)$ e

$$p_{dis}(x_{obs}) = \iint I_B(x^{rep}, \theta) f(x^{rep} | \theta) h(\theta | x_{obs}) d\theta dx^{rep},$$

sendo $B = \{(x^{rep}, \theta) : t(x^{rep}, \theta) \geq t(x_{obs}, \theta)\}$ e I_B a função indicatriz do conjunto B .

A distribuição preditiva *a posteriori* apresentada em (5.4) não é mais do que a distribuição marginal de X_{rep} em relação à distribuição $p(x^{rep}, \theta | x_{obs})$.

Sob a validade do modelo proposto, o valor ideal para o p -value de discrepância é 0.5; se o valor observado do p -value de discrepância estiver próximo deste valor o modelo é validado. Um modelo é suspeito, não devendo portanto ser validado, se o valor observado deste p -value estiver próximo de 0 ou 1. Sobre este assunto, Gelman *et al.* (1995, pág. 173) salientam “Major failures of the model, typically corresponding to extreme tail-area probabilities (less than 0.01 or more than 0.99) ...” e, mais adiante, sugerem que o intervalo 0.05 e 0.95 pode ser considerado um intervalo razoável para o p -value de discrepância.

Na maior parte das aplicações não é possível obter analiticamente o p -value de discrepância. A solução mais simples consiste em estimá-lo, por simulação, através do procedimento que a seguir se apresenta. Simula-se J vectores da distribuição *a posteriori* de θ e, para cada um dos $\theta^{rep,j}$ ($j = 1, \dots, J$) simula-se uma amostra $x^{rep,j}$ da distribuição $f(x|\theta^{rep,j})$. Obtém-se, deste modo, J amostras da distribuição *a posteriori* conjunta $p(x^{rep}, \theta|x_{obs})$, calcula-se $t(x_{obs}, \theta^{rep,j})$ e $t(x^{rep,j}, \theta^{rep,j})$, e estima-se o p -value de discrepância como sendo a proporção dos J pares para os quais $t(x^{rep,j}, \theta^{rep,j})$ é maior ou igual a $t(x_{obs}, \theta^{rep,j})$. Este procedimento é esquematicamente representado na figura 5.2 e resumido através do algoritmo 5.1.

O diagrama de dispersão de $t(x^{rep,j}, \theta^{rep,j})$ versus $t(x_{obs}, \theta^{rep,j})$ permite ter uma perspectiva gráfica da adequabilidade do modelo; se o modelo for adequado, a distribuição dos pontos acima e abaixo da bissetriz do primeiro quadrante deve ser idêntica.

Algoritmo 5.1

1. Simula-se J valores, $\{\theta^{rep,1}, \theta^{rep,2}, \dots, \theta^{rep,J}\}$, da distribuição $h(\theta|x_{obs})$.
2. Para cada $\theta^{rep,j}$, com $j \in \{1, 2, \dots, J\}$,
 - (a) simula-se uma amostra $x^{rep,j}$ da distribuição $f(x|\theta^{rep,j})$;
 - (b) calcula-se $t(x_{obs}, \theta^{rep,j})$ e $t(x^{rep,j}, \theta^{rep,j})$.
3. Determina-se uma estimativa do p -value de discrepância:

$$\hat{p}_{dis}(x_{obs}) = \frac{1}{J} \sum_{j=1}^J I_B(x^{rep,j}, \theta^{rep,j}). \quad \blacksquare$$

A escolha da quantidade de teste deve ser feita de modo a quantificar os desvios dos dados em relação ao modelo nos aspectos mais relevantes, tendo em vista o uso final do modelo. Pode, no entanto, referir-se que a quantidade de teste

$$t(X, \theta) = \sum_{k=1}^n \frac{(X_k - E[X_k|\theta])^2}{Var[X_k|\theta]}$$

é muito utilizada para uma validação global do ajustamento. A solução mais correcta consiste em estimar p_{dis} com diferentes quantidades de teste, de modo a avaliar várias possibilidades de falta de ajustamento.

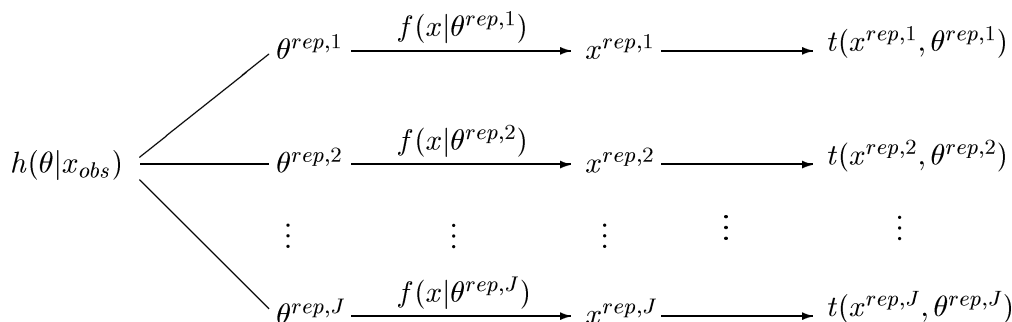


Figura 5.2: Esquema da simulação dos valores da distribuição preditiva *a posteriori* de $t(X^{rep}, \theta)$.

5.4 Aplicações

Exemplifica-se alguns dos métodos de validação informal e de validação formal de modelos, recorrendo ao exemplo 2.2 relativo ao ajustamento dos modelos beta/binomial negativo e gama/Poisson a dados de contagem e aos exemplos de regressão linear normal e de regressão gama, apresentados nas secções 3.5 e 3.6, respectivamente.

5.4.1 Modelos beta/binomial negativo e gama/Poisson — dados de contagem

Utiliza-se os modelos beta/binomial negativo (m_1) e gama/Poisson (m_2) ajustados às amostras x_u e x_v (exemplo 2.2 da secção 2.5.3) para exemplificar alguns aspectos da validação de modelos. Refira-se, novamente, que x_u e x_v são duas amostras de dimensão 20, constituídas por observações independentes e identicamente distribuídas BiNeg(2, 0.399) e BiNeg(2, 0.839), respectivamente.

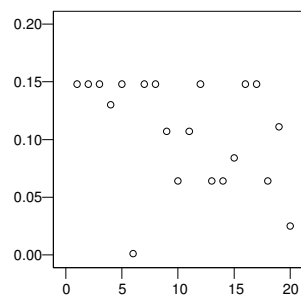
A representação gráfica das ordenadas preditivas condicionais possibilita a identificação de observações mal ajustadas pelo modelo. Não é necessário recorrer aos métodos de simulação de Monte Carlo para determinar as CPO, pois as distribuições marginais dos modelos beta/binomial negativo e gama/Poisson são fáceis de obter analiticamente (ver secções 2.5.1 e 2.5.2, respectivamente). Assim, para cada uma das amostras e para cada um dos modelos, determina-se as distribuições marginais $f(x)$ e $f(x_{(-i)})$, obtendo-se $CPO(x_i) = f(x)/f(x_{(-i)})$, com $i = 1, \dots, 20$. O programa em Fortran CPO_bbn_gp, apresentado no ficheiro com o mesmo nome no CD anexo a este documento, permite determinar as CPO dos modelos m_1 e m_2 , cujos valores são apresentados na tabela 5.4.

Na figura 5.3 representa-se graficamente, para cada uma das amostras, as ordenadas preditivas condicionais *versus* os números das observações, para os modelos m_1 e m_2 . Na amostra x_u parece existir uma observação ($x_{u,6} = 17$) mal ajustada quer pelo modelo beta/binomial negativo quer pelo modelo gama/Poisson. Na amostra x_v , e para qualquer um dos modelos, não parece existir qualquer observação que possa ser candidata a observação mal ajustada.

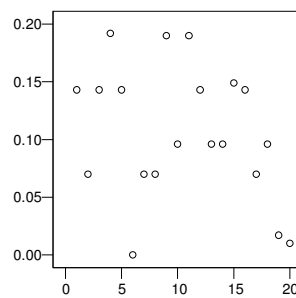
Tabela 5.4: Ordenadas preditivas condicionais das amostras x_u e x_v , para os modelos m_1 e m_2 .

número da observação	amostra x_u			amostra x_v		
	dados	CPO $_{m_1}$	CPO $_{m_2}$	dados	CPO $_{m_1}$	CPO $_{m_2}$
1	2	0.148	0.141	0	0.580	0.591
2	1	0.148	0.069	0	0.580	0.591
3	2	0.148	0.141	0	0.580	0.591
4	3	0.130	0.191	0	0.580	0.591
5	2	0.148	0.141	0	0.580	0.591
6	17	0.001	2.4E-07	1	0.263	0.288
7	1	0.148	0.069	1	0.263	0.288
8	1	0.148	0.069	1	0.263	0.288
9	4	0.107	0.190	0	0.580	0.591
10	6	0.064	0.097	0	0.580	0.591
11	4	0.107	0.190	1	0.263	0.288
12	2	0.148	0.141	1	0.263	0.288
13	6	0.064	0.097	0	0.580	0.591
14	6	0.064	0.097	0	0.580	0.591
15	5	0.084	0.150	2	0.083	0.063
16	2	0.148	0.141	1	0.263	0.288
17	1	0.148	0.069	0	0.580	0.591
18	6	0.064	0.097	1	0.263	0.288
19	0	0.111	0.017	0	0.580	0.591
20	9	0.025	0.010	1	0.263	0.288

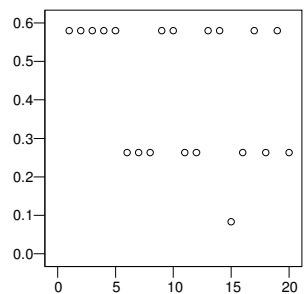
amostra x_u – modelo m_1



amostra x_u – modelo m_2



amostra x_v – modelo m_1



amostra x_v – modelo m_2

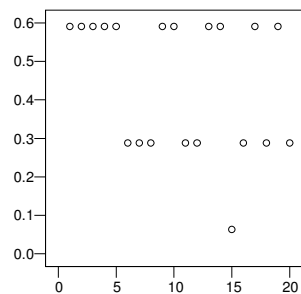


Figura 5.3: Ordenadas preditivas condicionais (ordenadas) *versus* números das observações (abcissas), obtidas com as amostras x_u e x_v , para os modelos m_1 e m_2 .

A validação formal dos modelos é efectuada através do cálculo de uma estimativa do p -value de discrepância, em que se utiliza a quantidade de teste

$$T = t(X, \theta) = \sum_{i=1}^{20} \frac{(X_i - E[X_i|\theta])^2}{Var[X_i|\theta]}.$$

No caso do modelo m_1 , tem-se $\theta = \theta_1$, e no caso de m_2 , tem-se $\theta = \theta_2$. Para o efeito, implementa-se em Fortran o algoritmo 5.1, cujo programa denominado `pvalue_dis_bbn_gp` é apresentado no ficheiro com o mesmo nome no CD anexo a este documento. Executa-se o referido programa considerando $J = 10000$ amostras replicadas. Na figura 5.4 e na tabela 5.5 apresenta-se um resumo dos resultados obtidos.

No caso da amostra x_u , apenas m_1 (modelo do qual os dados são gerados) é validado pois $\hat{p}_{dis}(x_u) = 0.2572$, enquanto para m_2 se obteve $\hat{p}_{dis}(x_u) = 0.0001$. No caso da amostra x_v , ambos os modelos são validados, pois obtém-se para m_1 e m_2 , respectivamente, $\hat{p}_{dis}(x_v) = 0.8306$ e $\hat{p}_{dis}(x_v) = 0.6509$. O facto da estimativa do p -value de discrepância sugerir que, no caso da amostra x_v , tanto m_1 como m_2 providenciam um ajustamento adequado dos dados, não é de estranhar uma vez que a distribuição $BiNeg(2, 0.839)$ é próxima da Poisson.

Na figura 5.4 ilustra-se, para as amostras x_u e x_v , o modo como se obtêm as estimativas do p -value de discrepância referentes ao ajustamento de m_1 e m_2 . Estas estimativas são dadas pela proporção de pontos que caem acima da bissectriz do primeiro quadrante.

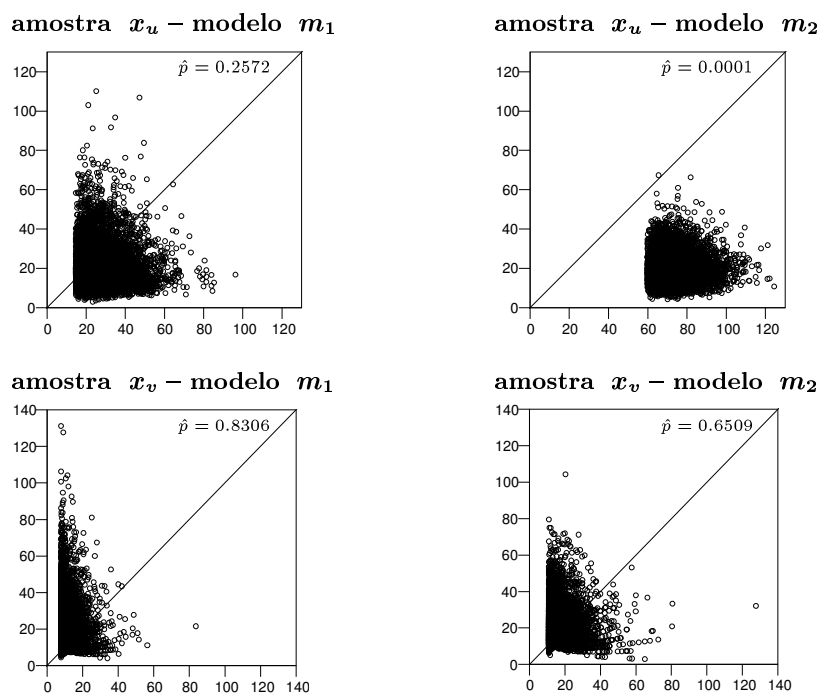


Figura 5.4: Diagramas de dispersão de $t(x^{rep,j}, \theta^{rep,j})$ (ordenadas) versus $t(x_{obs}, \theta^{rep,j})$ (abscissas), baseados em 10000 amostras simuladas da distribuição *a posteriori* conjunta de X^{rep} e θ , obtidas com as amostras x_u e x_v , para os modelos m_1 e m_2 .

Tabela 5.5: Estimativas dos p -values de discrepância, obtidas com as amostras x_u e x_v , para os modelos m_1 e m_2 .

amostra	x_u				x_v			
modelo	m_1		m_2		m_1		m_2	
	méd.	d.p.	méd.	d.p.	méd.	d.p.	méd.	d.p.
$t(x_{obs}, \theta^{rep,j})$	26.99	9.14	72.86	8.50	10.79	4.10	16.23	6.41
$t(x^{rep,j}, \theta^{rep,j})$	20.01	10.16	20.03	6.74	20.07	11.36	19.98	9.13
\hat{p}_{dis}	0.2572		0.0001		0.8306		0.6509	

5.4.2 Regressão linear normal — dados da massa gorda

Os modelos de regressão linear normal para os dados da massa gorda, m_1, \dots, m_5 , descritos na secção 3.5.5, são também objecto de validação. Para não alongar demasiado esta secção, apresenta-se apenas os gráficos referentes à validação dos modelos m_1 , m_3 e m_4 . Dos cinco modelos finais obtidos, m_1 é o que apresenta o menor valor do DIC, LLP e $-\text{SlnCPO}$, m_3 é o que tem o menor valor do BIC e m_4 é o modelo ao qual o método GVS atribui a maior probabilidade *a posteriori*. Os gráficos referentes à validação dos modelos m_2 e m_5 são semelhantes aos obtidos para os modelos m_1 e m_4 , respectivamente.

Na figura 5.5 representa-se graficamente os logaritmos das ordenadas preditivas condicionais *versus* os números das observações. O valor do logaritmo de $\text{CPO}_{m_4}(y_{39})$ é bastante pequeno, indicando que a observação y_{39} se encontra mal ajustada pelo modelo m_4 . Em relação aos modelos m_1 e m_3 parece não se destacar, de modo tão evidente, qualquer observação tendo um valor demasiado pequeno do logaritmo da ordenada preditiva condicional, e que possa ser considerada candidata a *outlier*.

Atendendo aos histogramas dos resíduos bayesianos, apresentados na figura 5.6, não existe motivo para suspeitar da suposição de normalidade dos resíduos dos modelos m_1 , m_3 e m_4 .

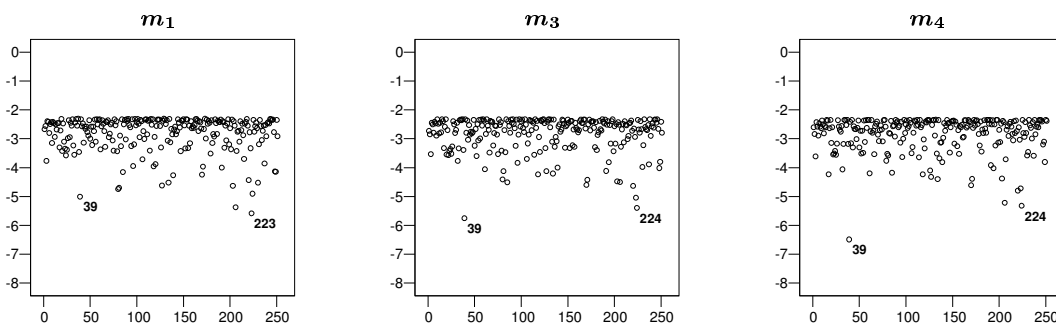


Figura 5.5: Logaritmos das ordenadas preditivas condicionais (ordenadas) *versus* números das observações (abcissas), para os modelos m_1 , m_3 e m_4 do exemplo da massa gorda.

Os gráficos dos resíduos bayesianos padronizados *versus* os valores preditos das observações são apresentados na figura 5.7. Os padrões de resíduos destes três gráficos são muito idênticos, não evidenciando qualquer desvio à suposição de homogeneidade de variâncias dos resíduos. Embora se opte por não os apresentar, foram também efectuados os gráficos dos resíduos bayesianos padronizados *versus* os números das observações, não havendo evidência para suspeitar da suposição de independência dos resíduos.

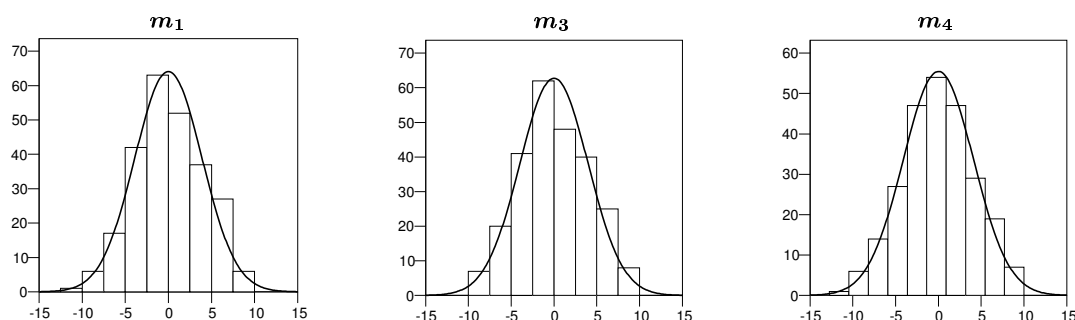


Figura 5.6: Histogramas dos resíduos bayesianos, para os modelos m_1 , m_3 e m_4 do exemplo da massa gorda, com sobreposição da densidade normal ajustada.

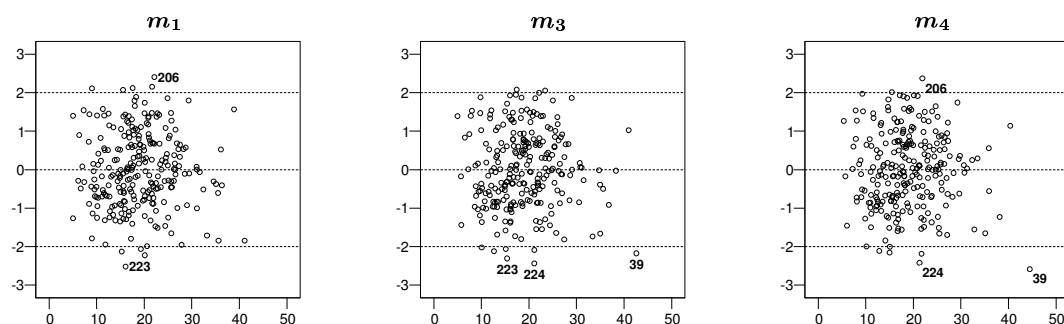


Figura 5.7: Resíduos bayesianos padronizados (ordenadas) *versus* valores preditos das observações (abscissas), para os modelos m_1 , m_3 e m_4 do exemplo da massa gorda.

A validação global do modelo pode ser feita à custa do p -value de discrepância. A discrepância entre a amostra observada e as amostras simuladas da distribuição preditiva *a posteriori* é quantificada à custa de três quantidades de teste. Spiegelhalter *et al.* (1996a, pág. 44) sugerem que se utilize uma medida de assimetria (T_3) e uma medida de curtose (T_4), definidas por

$$T_3 = t_3(Y, \theta) = \frac{1}{251} \sum_{i=1}^{251} \left(\frac{Y_i - E[Y_i|\theta]}{\sqrt{\text{Var}[Y_i|\theta]}} \right)^3 \quad \text{e} \quad T_4 = t_4(Y, \theta) = \frac{1}{251} \sum_{i=1}^{251} \left(\frac{Y_i - E[Y_i|\theta]}{\sqrt{\text{Var}[Y_i|\theta]}} \right)^4,$$

para testar a existência de afastamentos significativos dos valores de assimetria e curtose esperados sob a validade de normalidade dos resíduos. Utiliza-se também a quantidade de teste

$$T_6 = t_6(Y, \theta) = \sum_{i=1}^{251} |Y_i - E[Y_i|\theta]|$$

na validação da adequabilidade global destes modelos de regressão linear. As quantidades de teste $\sum_{i=1}^{251} (Y_i - E[Y_i|\theta])^2$ e $\sum_{i=1}^{251} (Y_i - E[Y_i|\theta])^2 / Var[Y_i|\theta]$ não devem ser usadas, pois reflectem aspectos parametrizados pelo modelo.

Veja-se como é estimado o *p-value* de discrepância obtido com T_3 . Para cada iteração, o valor da quantidade de teste T_3 calculado à custa da amostra replicada é comparado com o valor de T_3 calculado à custa da amostra observada. A proporção de iterações para as quais o valor de $t_3(x^{rep,j}, \theta^{rep,j})$ é maior ou igual ao valor $t_3(x_{obs}, \theta^{rep,j})$ constitui uma estimativa do *p-value* de discrepância para a quantidade de teste T_3 .

Quando se utiliza as quantidades de teste T_4 e T_6 para obter estimativas de *p-dis* procede-se de modo análogo.

Por exemplo, no caso do modelo m_1 , para determinar as estimativas do *p-value* de discrepância obtidas com as quantidades de teste T_3 , T_4 e T_6 é necessário acrescentar

```

for (i in 1:N){
  Y.rep[i]~dnorm(mu[i], tau)
  resid.rep[i]<-Y.rep[i]-mu[i]
  sresid.rep[i]<-resid.rep[i]*sqrt(tau)
  m3[i]<-pow(sresid[i], 3)
  m4[i]<-pow(sresid[i], 4)
  m3.rep[i]<-pow(sresid.rep[i], 3)
  m4.rep[i]<-pow(sresid.rep[i], 4)
  resid.abs[i]<-abs(resid[i])
  resid.abs.rep[i]<-abs(resid.rep[i])
}
T3.obs<-sum(m3[])/N
T3.rep<-sum(m3.rep[])/N
p.T3<-step(T3.rep-T3.obs)
T4.obs<-sum(m4[])/N
T4.rep<-sum(m4.rep[])/N
p.T4<-step(T4.rep-T4.obs)
T6.obs<-sum(resid.abs[])
T6.rep<-sum(resid.abs.rep[])
p.T6<-step(T6.rep-T6.obs)

```

ao código do WinBUGS apresentado na página 189, e monitorizar p.T3, p.T4 e p.T6. Na tabela 5.6 resume-se as estimativas dos *p-values*, para os modelos m_1 a m_5 , obtidas com as referidas quantidades de teste. Os *outputs* do WinBUGS com as estimativas de T_3 , T_4 e T_6 , obtidas com a amostra observada e as amostras replicadas, para os modelos m_1 , m_3 e m_4 , e as respectivas estimativas do *p-value* de discrepância, são apresentadas na tabela 5.7.

Tabela 5.6: Estimativas dos *p-values* de discrepância, obtidas com as quantidades de teste T_3 , T_4 e T_6 , para os modelos de regressão linear m_1 a m_5 , do exemplo da massa gorda.

modelos	T_3	T_4	T_6
m_1	0.4480	0.7095	0.3710
m_2	0.4739	0.7177	0.3696
m_3	0.5375	0.7535	0.3268
m_4	0.5730	0.7198	0.3761
m_5	0.5920	0.6952	0.3423

As estimativas dos p -values de discrepância, obtidas com qualquer uma das quantidades de teste, não permitem que se suspeite da validade de qualquer um dos modelos.

Ilustra-se, na figura 5.8, com o modelo m_1 , a determinação de estimativas do p -value de discrepância para as três quantidades de teste consideradas. Estas estimativas são dadas pela proporção de pontos que caem acima da bissetriz do primeiro quadrante.

Tabela 5.7: *Outputs* do WinBUGS com as estimativas dos p -values de discrepância obtidas com as quantidades de teste T_3 , T_4 e T_6 , para os modelos m_1 , m_3 e m_4 , do exemplo da massa gorda.

modelo m1								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
T3.obs	0.03907	0.1989	0.002106	-0.3484	0.03666	0.4281	1001	10000
T3.rep	-0.003292	0.2431	0.002241	-0.4818	-0.002831	0.4789	1001	10000
p.T3	0.448	0.4973	0.005678	0	0	1	1001	10000
T4.obs	2.576	0.468	0.004894	1.765	2.538	3.604	1001	10000
T4.rep	3.006	0.6216	0.006139	1.981	2.94	4.384	1001	10000
p.T4	0.7095	0.454	0.004617	0	1	1	1001	10000
T6.obs	812.4	8.355	0.087	799.7	811.2	831.7	1001	10000
T6.rep	796.9	52.29	0.5685	698.8	795.4	905	1001	10000
p.T6	0.371	0.4831	0.004825	0	0	1	1001	10000
modelo m3								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
T3.obs	-0.02863	0.1922	0.001777	-0.4028	-0.0306	0.352	1001	10000
T3.rep	-0.002876	0.2454	0.002462	-0.4869	-0.00341	0.476	1001	10000
p.T3	0.5375	0.4986	0.004496	0	1	1	1001	10000
T4.obs	2.482	0.444	0.005155	1.7	2.449	3.437	1001	10000
T4.rep	3.003	0.6214	0.006531	1.959	2.948	4.387	1001	10000
p.T4	0.7535	0.431	0.004324	0	1	1	1001	10000
T6.obs	830.1	7.288	0.0784	818.6	829.2	846.8	1001	10000
T6.rep	807.7	52.89	0.5981	709.8	805.7	916.8	1001	10000
p.T6	0.3268	0.469	0.004344	0	0	1	1001	10000
modelo m4								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
T3.obs	-0.0564	0.1919	0.00181	-0.4275	-0.05732	0.3264	1001	10000
T3.rep	-0.00347	0.2456	0.002106	-0.4911	-0.002264	0.4786	1001	10000
p.T3	0.573	0.4946	0.004653	0	1	1	1001	10000
T4.obs	2.539	0.467	0.004739	1.736	2.503	3.565	1001	10000
T4.rep	3.002	0.6258	0.006573	1.97	2.929	4.409	1001	10000
p.T4	0.7198	0.4491	0.004528	0	1	1	1001	10000
T6.obs	843.1	7.038	0.0641	834	841.6	860.3	1001	10000
T6.rep	827.1	54.4	0.6134	726.6	825.1	938.6	1001	10000
p.T6	0.3761	0.4844	0.004832	0	0	1	1001	10000

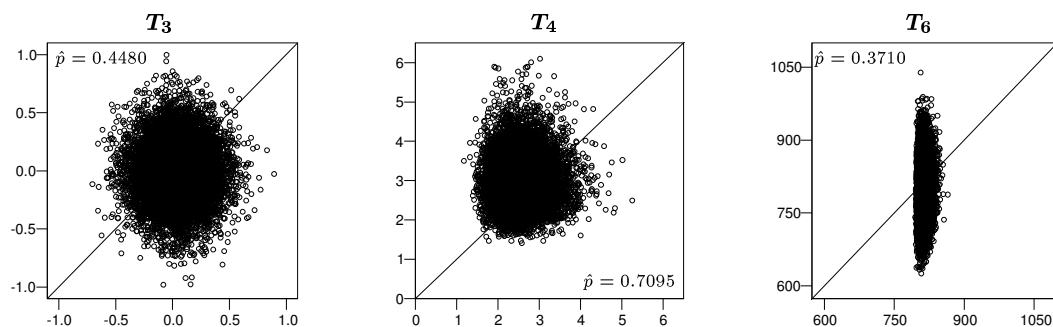


Figura 5.8: Valores simulados (ordenadas) das quantidades de teste *versus* valores observados (abscissas), para o modelo m_1 do exemplo da massa gorda.

5.4.3 Regressão gama — dados da homocisteinemia

Efectua-se representações gráficas para uma validação informal de alguns dos modelos — descritos na secção 3.6.5 — ajustados aos dados da homocisteinemia. Considerando a restrição que impõe que as duas variáveis *dummy* que definem a covariável fumar estão ambas presentes no modelo ou ambas ausentes do modelo, m_1 é o modelo que apresenta o menor valor do DIC e $-\text{SlnCPO}$, m_3 é o modelo com o menor valor do $\widehat{\text{BIC}}$ e m_6 é o modelo ao qual o método GVS atribui a maior probabilidade *a posteriori*. Para não alongar demasiado esta secção, os gráficos efectuados referem-se apenas a estes 3 modelos.

Na figura 5.9 representa-se os logaritmos das ordenadas preditivas condicionais *versus* os números das observações, para os modelos m_1 , m_3 e m_6 . Os valores dos logaritmos das ordenadas preditivas condicionais das observações 41 e 42 são demasiado pequenos, quando comparados com os logaritmos das CPO das restantes observações.

Na figura 5.10 representa-se os resíduos bayesianos padronizados *versus* os logaritmos dos valores preditos das observações, para os modelos m_1 , m_3 e m_6 . Destaca-se o facto das observações 41 e 42 serem mal ajustadas por qualquer um dos modelos. Estas observações são, muito possivelmente, *outliers*, devendo ser equacionada a sua exclusão do estudo.

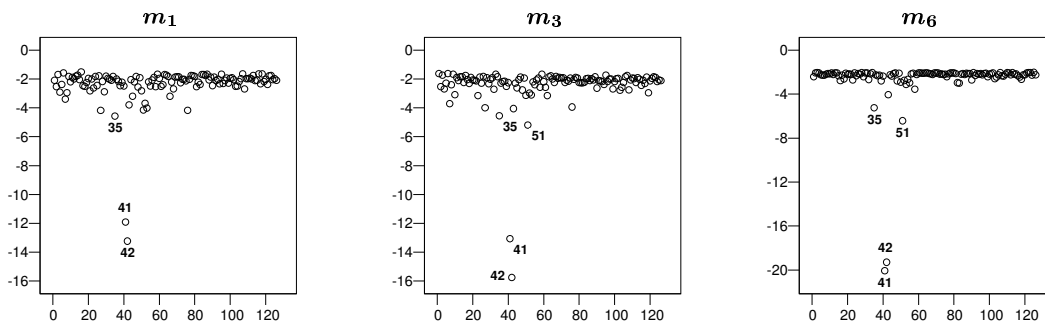


Figura 5.9: Logaritmos das ordenadas preditivas condicionais (ordenadas) *versus* números das observações (abscissas), para os modelos m_1 , m_3 e m_6 do exemplo da homocisteinemia.

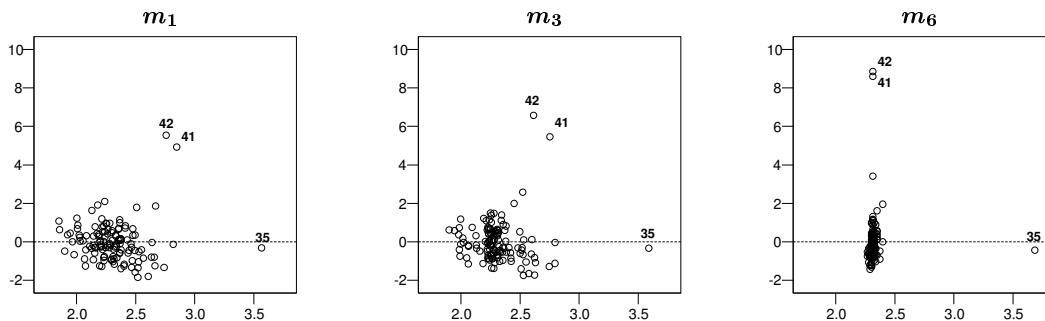


Figura 5.10: Resíduos bayesianos padronizados (ordenadas) *versus* logaritmos dos valores preditos das observações (abscissas), para os modelos m_1 , m_3 e m_6 do exemplo da homocisteinemia.

Além da análise gráfica, determina-se estimativas do p -value de discrepância, com vista ao estudo da adequabilidade dos modelos de regressão gama m_1 a m_7 . Utiliza-se as quantidades de teste

$$T_1 = t_1(Y, \theta) = \sum_{i=1}^{126} \left(\frac{Y_i - E[Y_i|\theta]}{\sqrt{\text{Var}[Y_i|\theta]}} \right)^2 \quad \text{e} \quad T_2 = t_2(Y, \theta) = \sum_{i=1}^{126} (Y_i - E[Y_i|\theta])^2,$$

também usadas em Paulino *et al.* (2003, secção 9.4.6) na validação de modelos de regressão gama. Além disso, considera-se também as quantidades de teste

$$T_5 = t_5(Y, \theta) = \sum_{i=1}^{126} \left| \frac{Y_i - E[Y_i|\theta]}{\sqrt{\text{Var}[Y_i|\theta]}} \right| \quad \text{e} \quad T_6 = t_6(Y, \theta) = \sum_{i=1}^{126} |Y_i - E[Y_i|\theta]|.$$

Para determinar as estimativas do p -value de discrepância é necessário seguir as indicações da secção D.2.1 e acrescentar

```
for (i in 1:N){
  resid[i]<-Y[i]-mu[i]
  sresid[i]<-resid[i]/(mu[i]/sqrt(nu))
  resid.quad[i]<-resid[i]*resid[i]
  sresid.quad[i]<-sresid[i]*sresid[i]
  resid.abs[i]<-abs(resid[i])
  sresid.abs[i]<-abs(sresid[i])
  Y.rep[i]~dgamma(nu,q[i])
  resid.rep[i]<-Y.rep[i]-mu[i]
  sresid.rep[i]<-resid.rep[i]/(mu[i]/sqrt(nu))
  resid.rep.quad[i]<-resid.rep[i]*resid.rep[i]
  sresid.rep.quad[i]<-sresid.rep[i]*sresid.rep[i]
  resid.rep.abs[i]<-abs(resid.rep[i])
  sresid.rep.abs[i]<-abs(sresid.rep[i])
}
T1.obs<-sum(sresid.quad)
T1.rep<-sum(sresid.rep.quad)
p.T1<-step(T1.rep-T1.obs)
T2.obs<-sum(resid.quad)
T2.rep<-sum(resid.rep.quad)
p.T2<-step(T2.rep-T2.obs)
T5.obs<-sum(sresid.abs[])
T5.rep<-sum(sresid.rep.abs[])
p.T5<-step(T5.rep-T5.obs)
T6.obs<-sum(resid.abs[])
T6.rep<-sum(resid.rep.abs[])
p.T6<-step(T6.rep-T6.obs)
```

ao código do WinBUGS apresentado nessa secção. Na tabela 5.8 resume-se as estimativas dos p -values, para os modelos m_1 a m_7 , obtidas com as quatro quantidades de teste. Os *outputs* do WinBUGS com as estimativas de T_1 , T_2 , T_5 e T_6 , obtidas com a amostra observada e as amostras replicadas, para os modelos m_1 , m_3 e m_6 , e as respectivas estimativas do p -value de discrepância, são apresentados na tabela 5.9.

Destaca-se o facto das diferentes quantidades de teste conduzirem a conclusões distintas. As estimativas dos p -values de discrepância obtidas com T_2 não nos permitem validar nenhum dos modelos, enquanto a utilização de T_1 sugere que apenas os modelos m_1 , m_2 e m_3 ajustam adequadamente os dados. Por sua vez, a utilização das quantidades de teste T_5 e T_6 dá uma indicação favorável à validação de qualquer um dos modelos. Esta diferença de resultados é devida à existência de observações discordantes, mal ajustadas

por qualquer um dos modelos. O impacto destas observações na estimativa do p -value é maior quando se utiliza as quantidades de teste T_1 e T_2 , pelo facto dos resíduos estarem ao “quadrado”. As quantidades de teste T_5 e T_6 são menos sensíveis à existência de *outliers*.

Tabela 5.8: Estimativas dos p -values de discrepância, obtidas com as quantidades de teste T_1 , T_2 , T_5 e T_6 para os modelos de regressão gama m_1 a m_7 , do exemplo da homocisteinemia.

modelos	T_1	T_2	T_5	T_6
m_1	0.2172	0.0110	0.6926	0.3680
m_2	0.2108	0.0082	0.6380	0.3322
m_3	0.1228	0.0078	0.6494	0.3446
m_4	0.0652	0.0128	0.6938	0.4370
m_5	0.0710	0.0088	0.7396	0.4532
m_6	0.0014	0.0182	0.8598	0.8220
m_7	0.0026	0.0138	0.9134	0.8704

Tabela 5.9: *Outputs* do WinBUGS com as estimativas dos p -values de discrepância obtidas com as quantidades de teste T_1 , T_2 , T_5 e T_6 , para os modelos m_1 , m_3 e m_6 , do exemplo da homocisteinemia.

modelo m1								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
T1.obs	147	20.96	0.3224	109.7	145.9	191	4001	5000
T1.rep	126	17.37	0.238	95.04	125.1	162.9	4001	5000
p.T1	0.2172	0.4123	0.005959	0	0	1	4001	5000
T2.obs	1725	311.9	4.053	1535	1649	2450	4001	5000
T2.rep	961.3	259	3.776	603.2	919.5	1607	4001	5000
p.T2	0.011	0.1043	0.001394	0	0	0	4001	5000
T5.obs	95.3	6.22	0.0865	83.36	95.21	108.1	4001	5000
T5.rep	99.99	6.834	0.09345	86.82	99.93	113.8	4001	5000
p.T5	0.6926	0.4614	0.00592	0	1	1	4001	5000
T6.obs	270.2	11.68	0.1719	252.2	268.5	297.9	4001	5000
T6.rep	262.4	26.52	0.3744	213.7	261.1	317.8	4001	5000
p.T6	0.368	0.4823	0.006617	0	0	1	4001	5000
modelo m3								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
T1.obs	159.6	23.57	0.3389	117.9	158.3	209.3	4001	5000
T1.rep	125.7	17.47	0.2342	94.39	124.4	163.4	4001	5000
p.T1	0.1228	0.3282	0.004893	0	0	1	4001	5000
T2.obs	2.00E+03	444.2	6.395	1803	1884	3011	4001	5000
T2.rep	1081	319.9	4.212	700.3	1033	1729	4001	5000
p.T2	0.0078	0.08797	0.001299	0	0	0	4001	5000
T5.obs	96.2	6.372	0.08171	83.86	95.98	109	4001	5000
T5.rep	99.8	6.908	0.08689	86.8	99.61	113.9	4001	5000
p.T5	0.6494	0.4772	0.006214	0	1	1	4001	5000
T6.obs	289.6	12.95	0.1696	269.3	287.9	319.8	4001	5000
T6.rep	279.4	27.8	0.3626	229.8	277.5	338.1	4001	5000
p.T6	0.3446	0.4752	0.007503	0	0	1	4001	5000
modelo m6								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
T1.obs	223.1	30.68	0.4103	166.8	222.3	287	4001	5000
T1.rep	125.8	17.73	0.2251	94.49	124.6	163.6	4001	5000
p.T1	0.0014	0.03739	5.04E-04	0	0	0	4001	5000
T2.obs	2611	974.6	12.86	2270	2347	4530	4001	5000
T2.rep	1491	555.5	8.155	921	1394	2589	4001	5000
p.T2	0.0182	0.1337	0.002113	0	0	0	4001	5000
T5.obs	90.28	5.678	0.07516	79.24	90.24	101.5	4001	5000
T5.rep	99.66	6.765	0.09306	86.85	99.41	113.4	4001	5000
p.T5	0.8598	0.3472	0.004815	0	1	1	4001	5000
T6.obs	299.4	14.41	0.1783	281.5	296.3	335.4	4001	5000
T6.rep	328.8	32.76	0.4593	269.5	326.8	396.8	4001	5000
p.T6	0.822	0.3825	0.005425	0	1	1	4001	5000

Para ilustrar a obtenção de estimativas do p -value de discrepância, representa-se, na figura 5.11, os valores simulados das quantidades de teste *versus* os valores observados, para o modelo m_1 . Conforme já foi referido, apenas o p -value obtido à custa da quantidade de teste T_2 sugere que o modelo m_1 não ajusta adequadamente os dados.

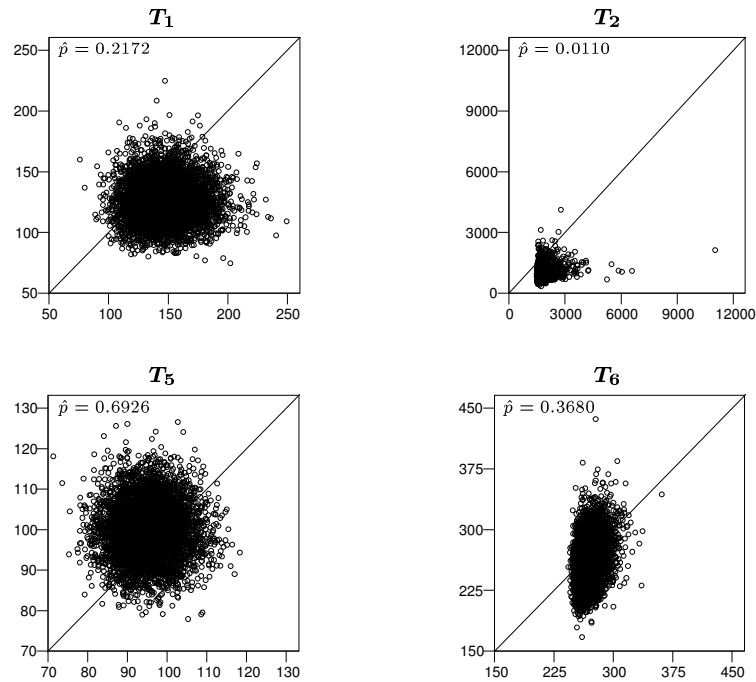


Figura 5.11: Valores simulados (ordenadas) das quantidades de teste *versus* valores observados (abscissas), para o modelo m_1 do exemplo da homocisteinemia.

Note-se o que acontece com o ajustamento de m_1 . Na figura 5.12 representa-se os desvios padrão de Y_i *versus* os resíduos não padronizados. Destaca-se a existência de duas observações (a 41 e a 42) mal ajustadas pelo modelo, e uma observação (a 35) cujo desvio padrão do resíduo é bastante elevado. Esta situação influencia substancialmente as estimativas dos parâmetros do modelo e dos p -values de discrepância.

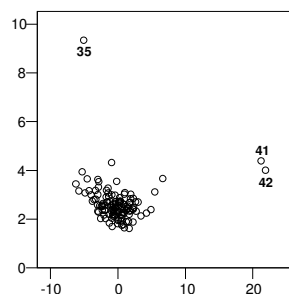


Figura 5.12: Desvios padrão de Y_i (ordenadas) *versus* resíduos bayesianos não padronizados (abscissas), para o modelo m_1 do exemplo da homocisteinemia.

A situação referida no parágrafo anterior também ocorre com o ajustamento dos outros modelos. Assim, retira-se as observações 35, 41 e 42 e ajusta-se, novamente, os modelos m_1 a m_7 . As estimativas dos p -values de discrepância destes modelos são apresentadas na tabela 5.10. Após retirar as três observações mencionadas, as estimativas obtidas para os p -values indicam que qualquer um dos modelos m_1 a m_7 ajusta adequadamente os dados.

Tabela 5.10: Estimativas dos p -values de discrepância, obtidas com as quantidades de teste T_1 , T_2 , T_5 e T_6 para os modelos de regressão gama m_1 a m_7 , do exemplo da homocisteinemia, quando se retiram as observações 35, 41 e 42.

modelos	T_1	T_2	T_5	T_6
m_1	0.4314	0.3288	0.4920	0.4564
m_2	0.4312	0.3432	0.4960	0.4608
m_3	0.3874	0.2916	0.4266	0.3828
m_4	0.3900	0.2860	0.4576	0.3994
m_5	0.3946	0.2840	0.4628	0.3988
m_6	0.3098	0.2858	0.5002	0.4726
m_7	0.3194	0.2886	0.5314	0.5030

Capítulo 6

Discussão e Trabalho Futuro

Neste trabalho estuda-se diversos aspectos da abordagem bayesiana à determinação de modelos, nomeadamente critérios de selecção de modelos, métodos de selecção de covariáveis em modelos de regressão e métodos de validação de modelos. Recorre-se a estudos de simulação para avaliar e comparar a performance de alguns destes métodos.

Os métodos utilizados na abordagem clássica deste problema estão bastante divulgados. Tal não acontece com a abordagem bayesiana: a maior parte dos utilizadores da estatística não está familiarizada com os métodos bayesianos de selecção e validação de modelos. Espera-se, com esta dissertação, ajudar a colmatar esta lacuna.

A selecção de covariáveis em modelos de regressão é um problema bastante usual nas aplicações estatísticas. Propõe-se dois métodos, inspirados nos seus homólogos clássicos: os métodos de eliminação regressiva e de selecção progressiva de covariáveis via medidas de discrepância. Para além dos métodos propostos, dá-se também destaque, neste trabalho, ao método Gibbs de selecção de covariáveis e ao método de selecção informal por pesquisa manual via medidas de discrepância. Utiliza-se diversas medidas de discrepância (também designadas medidas de selecção de modelos), nomeadamente o critério de informação da *deviance*, o critério de informação bayesiano estimado, a log-verosimilhança *a posteriori* e a soma dos logaritmos das ordenadas preditivas condicionais.

O método SIMD é bastante fácil de utilizar e permite avaliar o efeito de inclusão de todas as covariáveis no modelo. As críticas mais usuais são relativas à pequena percentagem de modelos examinados, à impossibilidade de implementação computacional do método, uma vez que é requerida a intervenção do investigador ao longo de todo o processo, e à subjectividade inerente ao facto do investigador ter que avaliar se as distribuições marginais *a posteriori* dos coeficientes de regressão estão ou não suficientemente concentradas em torno de zero.

O método GVS tem a vantagem de possibilitar a estimação das probabilidades *a posteriori* dos modelos resultantes de todas as combinações possíveis de covariáveis. A probabilidade *a posteriori* de um modelo tem uma interpretação bastante directa: é a probabilidade desse modelo ser a melhor aproximação, entre os modelos considerados, do fenómeno em estudo.

Nos dois exemplos de selecção de covariáveis apresentados — massa gorda e

homocisteinemia — os modelos com maior probabilidade *a posteriori* têm poucas covariáveis, e apresentam valores do DIC, LLP e SlnCPO bastante desfavoráveis, quando comparados com outros modelos. Em qualquer um dos exemplos, a dimensão da amostra é bastante pequena, quando comparada com o número de possíveis covariáveis ou modelos. Por este motivo, nestes dois exemplos, o método GVS não é, eventualmente, o mais adequado para identificação das covariáveis mais significativas.

Num estudo de simulação com dados de regressão linear, com quatro possíveis covariáveis, o método GVS apresenta uma performance bastante boa com amostras de dimensão 200: aos modelos “correctos” são atribuídas probabilidades *a posteriori* que se situam entre 0.84 e 0.96. No entanto, com amostras de dimensão 100 ou 50, este método nem sempre possibilita a identificação dos modelos “correctos”: os modelos com probabilidades *a posteriori* mais elevadas têm, na maior parte dos casos, menos covariáveis que as incluídas no modelo gerador dos dados.

Torna-se, portanto, evidente que a performance do método Gibbs de selecção de covariáveis é bastante influenciada pela dimensão da amostra. Mas, mais importante ainda, deve essencialmente depender da razão entre a dimensão da amostra e o número de covariáveis, ou da razão entre a dimensão da amostra e o número de possíveis modelos.

A repetição do estudo de simulação com, por exemplo, três possíveis covariáveis e amostras de dimensão 150 e 100, e com duas possíveis covariáveis e amostras de dimensão 100 e 50, pode acrescentar alguma informação adicional sobre o que realmente influencia a performance do método GVS. Se, com três possíveis covariáveis e amostras de dimensão 150, e com duas possíveis covariáveis e amostras de dimensão 100, se obtiver performances idênticas às obtidas com quatro possíveis covariáveis e amostras de dimensão 200, então a performance do método GVS deverá depender da razão entre a dimensão da amostra e o número de covariáveis. Se, com três possíveis covariáveis e amostras de dimensão 100, e com duas possíveis covariáveis e amostras de dimensão 50, se obtiver performances idênticas às obtidas com quatro possíveis covariáveis e amostras de dimensão 200, então a performance deste método deverá depender da razão entre a dimensão da amostra e o número de possíveis modelos. Este é um dos aspectos a investigar em trabalhos futuros.

A questão referida no parágrafo anterior é, na realidade, importante. Considere-se, por exemplo, um problema de selecção de covariáveis em modelos de regressão linear com oito possíveis covariáveis. Se a performance do método GVS depender da razão n/p então, eventualmente, com uma amostra de dimensão 400 já se pode utilizar os resultados obtidos por este método com alguma confiança. Caso o desempenho do método Gibbs de selecção de covariáveis dependa da razão $n/2^p$ então a amostra deve ter uma dimensão bastante superior, devendo rondar as 3200 observações, para garantir uma boa performance do método GVS.

Ainda em relação ao estudo de simulação com os dados de regressão linear, constata-se que a medida $\widehat{\text{BIC}}$ tem, para amostras de dimensão 200, uma performance bastante superior à do DIC e à do LLP. Mesmo para amostras de dimensão 100, regista-se frequências de classificação correcta bastante elevadas com o $\widehat{\text{BIC}}$. Para amostras de dimensão 100, a performance do DIC aproxima-se da performance do $\widehat{\text{BIC}}$, ultrapassando-a para amostras de dimensão 50. Com excepção do caso $n = 50$, em relação ao qual não se pode dizer que o desempenho do LLP seja muito inferior ao do DIC e do $\widehat{\text{BIC}}$, a log-verosimilhança *a posteriori* é a medida de selecção de modelos que tem o pior

desempenho.

A penalização imposta pelo $\widehat{\text{BIC}}$ ao número de parâmetros do modelo parece ser demasiado pesada pois, no caso de selecção incorrecta, à medida que n diminui existe tendência para seleccionar modelos com menos covariáveis que as incluídas no modelo gerador. No caso do DIC a situação é distinta: no caso de selecção incorrecta, à medida que n aumenta existe tendência para seleccionar modelos com mais covariáveis que as incluídas no modelo gerador; esta situação pode dever-se ao facto da penalização imposta pelo DIC ao número de parâmetros do modelo ser demasiado fraca. Esta constatação sugere que, para que um critério de selecção de modelos tenha uma boa performance, é necessário atingir um determinado “equilíbrio” entre a dimensão da amostra e a penalização em relação ao número de covariáveis do modelo.

O estudo de simulação efectuado aponta para a superioridade do critério de informação bayesiano estimado, relativamente ao DIC e LLP. No entanto, a estrutura tipicamente hierárquica da modelação bayesiana dificulta a determinação do número de parâmetros do modelo, podendo mesmo inviabilizar a utilização de algumas medidas de selecção de modelos, nomeadamente do $\widehat{\text{BIC}}$. O critério de informação da *deviance* surgiu como uma alternativa adequada para a comparação de modelos nesta situação. A superioridade do $\widehat{\text{BIC}}$, relativamente ao DIC, pode dever-se ao facto dos modelos considerados neste estudo terem uma pequena estrutura hierárquica. Num trabalho futuro pretende averiguar-se a influência do número de níveis da estrutura hierárquica na performance destes dois critérios de selecção de modelos.

Como já foi referido, a utilização do método Gibbs de selecção de covariáveis pode não ser a solução mais adequada no caso de amostras de dimensão reduzida. Além disso, podem surgir problemas de índole computacional: o elevado número de modelos a monitorizar pode inviabilizar a utilização do método GVS, mesmo para um número moderado de possíveis covariáveis. Os dois métodos de pesquisa do melhor subconjunto de covariáveis propostos neste trabalho — os métodos de pesquisa progressiva e de eliminação regressiva de covariáveis via medidas de discrepância — podem ser uma alternativa ao método GVS.

Quando existe multicolinearidade entre as covariáveis ou quando existem covariáveis conjuntamente, mas não individualmente, significativas, o método de eliminação regressiva via medidas de discrepância permite obter resultados bastante melhores que os da pesquisa progressiva e, além disso, possibilita a avaliação do efeito da inclusão de todas as covariáveis no modelo.

A utilização da medida de discrepância $\widehat{\text{BIC}}$ com os métodos de pesquisa FVS ou BVE proporciona a selecção de modelos com um número de covariáveis inferior ao dos modelos obtidos por estes métodos com as outras medidas de discrepância consideradas. A penalização imposta pelo $\widehat{\text{BIC}}$ ao número de parâmetros do modelo é mais “pesada” que a das medidas DIC ou LLP.

Uma constatação importante, resultante dos dois estudos de simulação, é o facto das variâncias amostrais obtidas por replicação dos valores do DIC e do $\widehat{\text{BIC}}$ serem muito idênticas. Após alguma investigação que conduziu a resultados infrutíferos, Zhu e Carlin (2000) sugerem que se efectue a replicação dos valores do DIC de modo a estimar $\text{Var}[\text{DIC}]$ através da sua variância amostral. Esta sugestão e a constatação referida sugerem que, quando se utiliza os métodos MCMC na selecção de modelos, pode estimar-se a variância

do DIC à custa da variância do $\widehat{\text{BIC}}$, a qual pode ser facilmente obtida quando se utiliza este método de simulação estocástica. Mas, que acontecerá quando se considera modelos com mais níveis de hierarquia? Serão as variâncias amostrais do $\widehat{\text{BIC}}$ e do DIC idênticas? Eis outro aspecto que necessita ser investigado.

Um aspecto também bastante importante na escolha da medida de selecção de modelos é o conhecimento do uso final do modelo. Caso se pretenda utilizar o modelo na predição de observações futuras, obtidas nas mesmas condições dos dados observados, então deve dar-se preferência ao DIC e ao $\widehat{\text{BIC}}$, relativamente ao factor Bayes. Caso não se pretenda efectuar a predição de observações futuras, mas apenas encontrar o modelo que melhor descreva um determinado conjunto de dados, então pode ser mais adequada, para a selecção de modelos, a utilização do factor Bayes.

Qualquer que seja o critério de selecção de modelos utilizado, a escolha do modelo pode ser mais, ou menos, óbvia, dependendo da dimensão da amostra. Com amostras de dimensão reduzida é difícil seleccionar um modelo pois, na maior parte das situações, existem diversos modelos com valores semelhantes da medida de discrepância. No caso particular da selecção de covariáveis em modelos de regressão, os resultados obtidos no estudo de simulação indicam que a identificação das covariáveis a incluir no modelo só pode ser efectuada com sucesso se a dimensão da amostra for razoável.

Se a selecção de modelos é um dos aspectos chave da modelação estatística, não menos importante é a validação de modelos. Existem diversos métodos, na sua maioria gráficos, para avaliar se o modelo ajusta adequadamente os dados.

Neste trabalho dá-se destaque ao *p-value* preditivo *a posteriori* como método de validação formal de modelos. Em particular, utiliza-se uma generalização deste *p-value* — o *p-value* de discrepância — que consiste em considerar que a quantidade de teste T depende não só dos dados, mas também do vector de parâmetros do modelo. Embora sejam objecto de algumas críticas, relativas a uma eventual “dupla utilização dos dados”, estes *p-values* são muito utilizados no estudo da adequabilidade dos modelos.

Através de estudos de simulação, pode avaliar-se se essa eventual “dupla utilização dos dados” tem impacto nas estimativas dos *p-values* preditivos *a posteriori* e de discrepância. Mais concretamente, a obtenção, por simulação, de amostras de um determinado modelo e a estimação destes *p-values* — para uma determinada quantidade de teste T — do modo usual e por validação cruzada (Carlin e Louis, 2000, pág. 48), permite avaliar se o facto dos dados serem usadas na determinação da distribuição preditiva *a posteriori* e no cálculo da área de cauda tem impacto nas estimativas destes *p-values*.

A constatação da relação existente entre as médias dos valores do DIC dos modelos “correctos” e a dimensão das amostras ajustadas sugere que uma eventual medida de adequabilidade absoluta do modelo, construída à custa do DIC, tenha em conta que os valores desta medida aumentam linearmente com a dimensão da amostra — dividir o valor do DIC pela dimensão da amostra parece ser uma solução a ter em conta. A sugestão efectuada também é válida para o BIC, $\widehat{\text{BIC}}$ e LLP.

Anexo A

Distribuições de Probabilidade

Neste anexo apresenta-se um breve sumário de algumas distribuições de probabilidade usadas no decurso da tese.

Binomial

A variável aleatória X tem distribuição binomial com parâmetros $n \in \mathbb{N}$ e $p \in]0, 1[$, simbolicamente $X \sim \text{Bi}(n, p)$, se a sua função de probabilidade for dada por

$$f(x|n, p) = \binom{n}{x} p^x (1-p)^{n-x},$$

para $x = 0, 1, 2, \dots, n$. O valor médio e a variância de X são dados, respectivamente, por: $E(X) = np$ e $Var(X) = np(1-p)$.

A distribuição Bernoulli, $X \sim \text{Bern}(p)$, é um caso particular da distribuição binomial, que resulta de considerar $n = 1$.

Binomial negativa

A variável aleatória X tem distribuição binomial negativa com parâmetros $r \in \mathbb{N}$ e $p \in]0, 1[$, simbolicamente $X \sim \text{BiNeg}(r, p)$, se a sua função de probabilidade for dada por

$$f(x|r, p) = \binom{x+r-1}{x} p^r (1-p)^x,$$

para $x = 0, 1, 2, \dots$. O valor médio e a variância de X são dados, respectivamente, por: $E(X) = r(1-p)/p$ e $Var(X) = r(1-p)/p^2$.

A distribuição geométrica, $X \sim \text{Geom}(p)$, é um caso particular da distribuição binomial negativa, que resulta de considerar $r = 1$.

Poisson

A variável aleatória X tem distribuição Poisson com parâmetro $\lambda > 0$, simbolicamente $X \sim \text{Poi}(\lambda)$, se a sua função de probabilidade for dada por

$$f(x|\lambda) = \frac{1}{x!} \lambda^x e^{-\lambda},$$

para $x = 0, 1, 2, \dots$. O valor médio e a variância de X são dados, respectivamente, por: $E(X) = \lambda$ e $Var(X) = \lambda$.

Uniforme

A variável aleatória X tem distribuição uniforme com parâmetros $a, b \in \mathbb{R}$, com $a < b$, simbolicamente $X \sim U(a, b)$, se a sua função densidade de probabilidade for dada por

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

para $x \in (a, b)$. O valor médio e a variância de X são dados, respectivamente, por: $E(X) = (a+b)/2$ e $Var(X) = (b-a)^2/12$.

Normal

A variável aleatória X tem distribuição normal com parâmetro de localização $\mu \in \mathbb{R}$ e parâmetro de escala $\sigma > 0$, simbolicamente $X \sim N(\mu, \sigma^2)$, se a sua função densidade de probabilidade for dada por

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\},$$

para $x \in \mathbb{R}$. Se $\mu = 0$ e $\sigma = 1$ diz-se que X tem distribuição normal padrão. O valor médio e a variância de X são dados, respectivamente, por: $E(X) = \mu$ e $Var(X) = \sigma^2$.

Gama

A variável aleatória X tem distribuição gama com parâmetro de forma $\alpha > 0$ e inverso do parâmetro de escala $\beta > 0$, simbolicamente $X \sim \text{Ga}(\alpha, \beta)$, se a sua função densidade de probabilidade for dada por

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},$$

para $x > 0$. A função Γ é definida por

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx = \int_0^\infty \beta^\alpha x^{\alpha-1} e^{-\beta x} dx,$$

para $\alpha > 0$, e designa-se por função gama. O valor médio e a variância de X são dados, respectivamente, por: $E(X) = \alpha/\beta$ e $Var(X) = \alpha/\beta^2$. O valor médio de $\ln X$ é $E(\ln X) = \psi(\alpha) - \ln \beta$, onde ψ designa a função digama, definida por $\psi(\alpha) = \frac{d \ln \Gamma}{d \alpha}(\alpha)$.

A distribuição gama é a distribuição *a priori* conjugada para o inverso da variância da normal e para o valor médio da Poisson. Uma distribuição não informativa é obtida no limite quando $\alpha, \beta \rightarrow 0$.

A distribuição qui-quadrado com parâmetro ν , simbolicamente $X \sim \chi^2(\nu)$, é um caso particular da distribuição gama, que resulta de considerar $\alpha = \frac{\nu}{2}$ e $\beta = \frac{1}{2}$. O parâmetro ν designa-se por graus de liberdade e é usual considerar que ν é um inteiro positivo (embora se obtenha uma distribuição própria para todo $\nu > 0$).

Gama invertida

A variável aleatória X tem distribuição gama invertida com parâmetro de forma $\alpha > 0$ e parâmetro de escala $\beta > 0$, simbolicamente $X \sim \text{GI}(\alpha, \beta)$, se a sua função densidade de probabilidade for dada por

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\beta/x},$$

para $x > 0$. O valor médio e a variância de X são dados, respectivamente, por:

$$E(X) = \frac{\beta}{\alpha - 1}, \quad \text{para } \alpha > 1,$$

$$\text{Var}(X) = \frac{\beta^2}{(\alpha - 1)^2 (\alpha - 2)}, \quad \text{para } \alpha > 2.$$

Se X^{-1} tem distribuição gama de parâmetros α, β , então X tem distribuição gama invertida com os mesmos parâmetros. Uma distribuição não informativa é obtida fazendo $\alpha, \beta \rightarrow 0$. A distribuição gama invertida é a distribuição *a priori* conjugada da variância da normal.

A distribuição qui-quadrado invertida com ν graus de liberdade, simbolicamente $X \sim \chi^2\text{-Inv}(\nu)$, é um caso particular da distribuição gama invertida, que resulta de considerar $\alpha = \frac{\nu}{2}$ e $\beta = \frac{1}{2}$.

A distribuição qui-quadrado invertida com ν graus de liberdade e parâmetro de escala $s > 0$, simbolicamente $X \sim \chi^2\text{-Inv}(\nu, s^2)$, é um caso particular da distribuição gama invertida, que resulta de considerar $\alpha = \frac{\nu}{2}$ e $\beta = \frac{\nu}{2}s^2$.

t-Student

A variável aleatória X tem distribuição t-Student com $n > 0$ graus de liberdade, parâmetro de localização $\mu \in \mathbb{R}$ e parâmetro de escala $\sigma > 0$, simbolicamente $X \sim t_n(\mu, \sigma^2)$, se a sua função densidade de probabilidade for dada por

$$f(x|\mu, \sigma^2, n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \sqrt{n\pi} \sigma} \left[1 + \frac{1}{n} \left(\frac{x - \mu}{\sigma}\right)^2\right]^{-\frac{n+1}{2}},$$

para $x \in \mathbb{R}$. O valor médio e a variância de X são dados, respectivamente, por:

$$E(X) = \mu, \quad \text{para } n > 1,$$

$$\text{Var}(X) = \frac{n}{n-2} \sigma^2, \quad \text{para } n > 2.$$

Beta

A variável aleatória X tem distribuição beta, com parâmetros $\alpha > 0$ e $\beta > 0$, simbolicamente $X \sim \text{Be}(\alpha, \beta)$, se a sua função densidade de probabilidade for dada por

$$f(x|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1},$$

para $x \in (0, 1)$. Como a função beta, definida por

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx,$$

para $\alpha > 0$ e $\beta > 0$, verifica a igualdade

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)},$$

então $f(x)$ pode ser representada por

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}.$$

O valor médio e a variância de X são dados, respectivamente, por: $E(X) = \alpha/(\alpha + \beta)$ e $Var(X) = \alpha\beta/[(\alpha + \beta)^2(\alpha + \beta + 1)]$.

Os valores médios de $\ln X$ e $\ln(1 - X)$ são dados por:

$$\begin{aligned} E(\ln X) &= \psi(\alpha) - \psi(\alpha + \beta), \\ E[\ln(1 - X)] &= \psi(\beta) - \psi(\alpha + \beta), \end{aligned}$$

sendo ψ a função digama.

A distribuição $U(0, 1)$ é um caso particular da beta, correspondente a considerar $\alpha = \beta = 1$. Os casos $\alpha = \beta = 0.5$ e $\alpha = \beta = 0$ são muitas vezes utilizados como distribuições *a priori* não informativas.

Para estimar os parâmetros da beta é útil usar as seguintes igualdades:

$$\begin{aligned} \alpha + \beta &= \frac{E(X)(1 - E(X))}{Var(X)} - 1, \\ \alpha &= (\alpha + \beta)E(X), \\ \beta &= (\alpha + \beta)(1 - E(X)). \end{aligned}$$

Normal multivariada

O vector aleatório X tem distribuição normal multivariada de dimensão k , com vector médio $\mu = (\mu_1, \dots, \mu_k)' \in \mathbb{R}^k$ e matriz de covariância Σ , sendo Σ uma matriz $k \times k$ simétrica e definida positiva, simbolicamente $X \sim N_k(\mu, \Sigma)$, se a sua função densidade de probabilidade for dada por

$$f(x | \mu, \Sigma) = \frac{|\Sigma^{-1}|^{1/2}}{(2\pi)^{k/2}} \exp \left[-\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right],$$

para $x \in \mathbb{R}^k$. $E(X) = \mu$, isto é, $E(X_i) = \mu_i$, e $Var(X) = \Sigma$, isto é, $Var(X_i) = \sigma_{ii}$ e $Cov(X_{ij}) = \sigma_{ij}$, onde $\Sigma = (\sigma_{ij})$.

Anexo B

Determinação de Modelos via Métodos MCMC

Na maior parte das situações, a abordagem bayesiana à selecção e validação de modelos só é possível de implementar recorrendo aos métodos MCMC. Considere-se que, por simulação estocástica via métodos MCMC foi ajustado um modelo, e que se obteve a amostra $\theta^* = (\theta_1^*, \dots, \theta_N^*)$ da distribuição *a posteriori* $h(\theta|x)$. Para calcular algumas das medidas de determinação de modelos apresentadas nesta dissertação é necessário saber como estimar uma distribuição preditiva de interesse, e calcular valores esperados preditivos. Outros aspectos referentes à determinação de modelos via métodos MCMC podem ser vistos em Gelfand (1996).

B.1 Estimação de distribuições preditivas

Estimação da distribuição preditiva *a posteriori*

A estimação da distribuição preditiva *a posteriori*, $p(y|x) = \int f(y|\theta) h(\theta|x) d\theta$, é o caso mais simples de estimação de distribuições preditivas. Se $\theta^* = (\theta_1^*, \dots, \theta_N^*)$ for uma amostra da distribuição *a posteriori* $h(\theta|x)$, então $p(y|x)$ pode ser estimada por

$$\hat{p}(y|x) = \frac{1}{N} \sum_{j=1}^N f(y|\theta_j^*).$$

Estimação da distribuição preditiva condicional

Pode estimar-se a distribuição preditiva condicional, $p(x_i|x_{(-i)}) = \int f(x_i|x_{(-i)}, \theta) h(\theta|x_{(-i)}) d\theta$, sem recorrer ao cálculo de $h(\theta|x_{(-i)})$. A distribuição $p(x_i|x_{(-i)}) = f(x)/f(x_{(-i)})$ pode ser escrita do seguinte modo:

$$p(x_i|x_{(-i)}) = \frac{1}{\frac{f(x_{(-i)})}{f(x)}} = \frac{1}{\int \frac{f(x_{(-i)}|\theta) h(\theta)}{f(x)} d\theta} = \frac{1}{\int \frac{f(x_{(-i)}|\theta) h(\theta) h(\theta|x)}{f(x) h(\theta|x)} d\theta}. \quad (\text{B.1})$$

Além disso

$$f(x) h(\theta|x) = h(\theta) f(x|\theta) = h(\theta) f(x_i|x_{(-i)}, \theta) f(x_{(-i)}|\theta),$$

donde se conclui que

$$\frac{1}{f(x_i|x_{(-i)}, \theta)} = \frac{f(x_{(-i)}|\theta) h(\theta)}{f(x) h(\theta|x)}. \quad (\text{B.2})$$

Deste modo, das expressões (B.1) e (B.2), obtém-se

$$p(x_i|x_{(-i)}) = \frac{1}{\int \frac{1}{f(x_i|x_{(-i)}, \theta)} h(\theta|x) d\theta}.$$

Se $\theta^* = (\theta_1^*, \dots, \theta_N^*)$ for uma amostra da distribuição *a posteriori* $h(\theta|x)$, então

$$\hat{p}(x_i|x_{(-i)}) = \left[\frac{1}{N} \sum_{j=1}^N \frac{1}{f(x_i|x_{(-i)}, \theta_j^*)} \right]^{-1};$$

se, além disso, (x_1, \dots, x_n) forem condicionalmente independentes dado θ , então $f(x_i|x_{(-i)}, \theta) = f(x_i|\theta)$ e

$$\hat{p}(x_i|x_{(-i)}) = \left[\frac{1}{N} \sum_{j=1}^N \frac{1}{f(x_i|\theta_j^*)} \right]^{-1}.$$

Estimação da distribuição preditiva *a priori*

Considere-se que a distribuição *a priori* $h(\theta)$ é própria. A distribuição preditiva *a priori* $f(x) = \int f(x|\theta) h(\theta) d\theta$ pode ser aproximada por

$$\hat{f}(x) = \frac{1}{N} \sum_{j=1}^N f(x|\theta_j^\dagger),$$

sendo $\theta^\dagger = (\theta_1^\dagger, \dots, \theta_N^\dagger)$ uma amostra da distribuição *a priori* $h(\theta)$. Esta estimativa é, no entanto, muito ineficiente (*e.g.*, Paulino *et al.*, 2003, secção 8.4.3). Além disso, tem o inconveniente de não ser possível obter através do *output* de uma simulação via métodos MCMC.

Como a distribuição *a priori* é própria, então $\int h(\theta) d\theta = 1$ e

$$\frac{1}{f(x)} = \int \frac{h(\theta)}{f(x) h(\theta|x)} h(\theta|x) d\theta.$$

Além disso, atendendo a que $f(x) h(\theta|x) = f(x|\theta) h(\theta)$, obtém-se

$$\frac{1}{f(x)} = \int \frac{1}{f(x|\theta)} h(\theta|x) d\theta.$$

Fica assim justificada a utilização da média harmónica de $\{f(x|\theta_j^*)\}_{j=1}^N$ para obter uma estimativa de $f(x)$,

$$\hat{f}(x) = \left[\frac{1}{N} \sum_{j=1}^N \frac{1}{f(x|\theta_j^*)} \right]^{-1}, \quad (\text{B.3})$$

sendo $\theta^* = (\theta_1^*, \dots, \theta_N^*)$ uma amostra da distribuico *a posteriori* $h(\theta|x)$. Se as distribuices $h(\theta)$ e $h(\theta|x)$ forem muito diferentes, o que acontece com alguma frequncia, a estimativa (B.3) tem tendncia a ser bastante instvel. Gelfand e Dey (1994) propem uma generalizao de (B.3) que d origem a estimativas mais estveis (*e.g.*, Paulino *et al.*, 2003, seco 8.4.3).

B.2 Estimaco do valor mdio preditivo e da varincia preditiva

O valor mdio preditivo de cada componente de Y , $E_{Y_i|x}[Y_i] = \int y_i p(y_i|x) dy_i$, pode ser reescrito

$$\begin{aligned} E_{Y_i|x}[Y_i] &= \int y_i \left(\int f(y_i|\theta) h(\theta|x) d\theta \right) dy_i \\ &= \int \left(\int y_i f(y_i|\theta) dy_i \right) h(\theta|x) d\theta \\ &= \int E_{Y_i|\theta}[Y_i] h(\theta|x) d\theta \end{aligned}$$

e o $E_{Y_i|x}[Y_i^2] = \int y_i^2 p(y_i|x) dy_i$  dado por

$$\begin{aligned} E_{Y_i|x}[Y_i^2] &= \int y_i^2 \left(\int f(y_i|\theta) h(\theta|x) d\theta \right) dy_i \\ &= \int \left(\int y_i^2 f(y_i|\theta) dy_i \right) h(\theta|x) d\theta \\ &= \int E_{Y_i|\theta}[Y_i^2] h(\theta|x) d\theta \\ &= \int \left[\text{Var}_{Y_i|\theta}[Y_i] + (E_{Y_i|\theta}[Y_i])^2 \right] h(\theta|x) d\theta. \end{aligned}$$

Assim, dada uma amostra $\theta^* = (\theta_1^*, \dots, \theta_N^*)$ da distribuico *a posteriori* $h(\theta|x)$, o $E_{Y_i|x}[Y_i]$ e o $E_{Y_i|x}[Y_i^2]$ podem ser estimados por:

$$\begin{aligned} \widehat{E}_{Y_i|x}[Y_i] &= \frac{1}{N} \sum_{j=1}^N E_{Y_i|\theta_j^*}[Y_i], \\ \widehat{E}_{Y_i|x}[Y_i^2] &= \frac{1}{N} \sum_{j=1}^N \left[\text{Var}_{Y_i|\theta_j^*}[Y_i] + (E_{Y_i|\theta_j^*}[Y_i])^2 \right], \end{aligned}$$

respectivamente. A obteno de uma estimativa de $\text{Var}_{Y_i|x}[Y_i]$  imediata:

$$\widehat{\text{Var}}_{Y_i|x}[Y_i] = \widehat{E}_{Y_i|x}[Y_i^2] - \left(\widehat{E}_{Y_i|x}[Y_i] \right)^2.$$

Anexo C

Conjuntos de Dados

Tabela C.1: Dados da massa gorda.

Y[]	x[,1]	x[,2]	x[,3]	x[,4]	x[,5]	x[,6]	x[,7]	x[,8]	x[,9]	x[,10]	x[,11]	x[,12]	x[,13]
12.6	23	154.25	67.75	36.2	93.1	85.2	94.5	59.0	37.3	21.9	32.0	27.4	17.1
6.9	22	173.25	72.25	38.5	93.6	83.0	98.7	58.7	37.3	23.4	30.5	28.9	18.2
24.6	22	154.00	66.25	34.0	95.8	87.9	99.2	59.6	38.9	24.0	28.8	25.2	16.6
10.9	26	184.75	72.25	37.4	101.8	86.4	101.2	60.1	37.3	22.8	32.4	29.4	18.2
27.8	24	184.25	71.25	34.4	97.3	100.0	101.9	63.2	42.2	24.0	32.2	27.7	17.7
20.6	24	210.25	74.75	39.0	104.5	94.4	107.8	66.0	42.0	25.6	35.7	30.6	18.8
19.0	26	181.00	69.75	36.4	105.1	90.7	100.3	58.4	38.3	22.9	31.9	27.8	17.7
12.8	25	176.00	72.50	37.8	99.6	88.5	97.1	60.0	39.4	23.2	30.5	29.0	18.8
5.1	25	191.00	74.00	38.1	100.9	82.5	99.9	62.9	38.3	23.8	35.9	31.1	18.2
12.0	23	198.25	73.50	42.1	99.6	88.6	104.1	63.1	41.7	25.0	35.6	30.0	19.2
7.5	26	186.25	74.50	38.5	101.5	83.6	98.2	59.7	39.7	25.2	32.8	29.4	18.5
8.5	27	216.00	76.00	39.4	103.6	90.9	107.7	66.2	39.2	25.9	37.2	30.2	19.0
20.5	32	180.50	69.50	38.4	102.0	91.6	103.9	63.4	38.3	21.5	32.5	28.6	17.7
20.8	30	205.25	71.25	39.4	104.1	101.8	108.6	66.0	41.5	23.7	36.9	31.6	18.8
21.7	35	187.75	69.50	40.5	101.3	96.4	100.1	69.0	39.0	23.1	36.1	30.5	18.2
20.5	35	162.75	66.00	36.4	99.1	92.8	99.2	63.1	38.7	21.7	31.1	26.4	16.9
28.1	34	195.75	71.00	38.9	101.9	96.4	105.2	64.8	40.8	23.1	36.2	30.8	17.3
22.4	32	209.25	71.00	42.1	107.6	97.5	107.0	66.9	40.0	24.4	38.2	31.6	19.3
16.1	28	183.75	67.75	38.0	106.8	89.6	102.4	64.2	38.7	22.9	37.2	30.5	18.5
16.5	33	211.75	73.50	40.0	106.2	100.5	109.0	65.8	40.6	24.0	37.1	30.1	18.2
19.0	28	179.00	68.00	39.1	103.3	95.9	104.9	63.5	38.0	22.1	32.5	30.3	18.4
15.3	28	200.50	69.75	41.3	111.4	98.8	104.8	63.4	40.6	24.6	33.0	32.8	19.9
15.7	31	140.25	68.25	33.9	86.0	76.4	94.6	57.4	35.3	22.2	27.9	25.9	16.7
17.6	32	148.75	70.00	35.5	86.7	80.0	93.4	54.9	36.2	22.1	29.8	26.7	17.1
14.2	28	151.25	67.75	34.5	90.2	76.3	95.8	58.4	35.5	22.9	31.1	28.0	17.6
4.6	27	159.25	71.50	35.7	89.6	79.7	96.5	55.0	36.7	22.5	29.9	28.2	17.7
8.5	34	131.50	67.50	36.2	88.6	74.6	85.3	51.7	34.7	21.4	28.7	27.0	16.5
22.4	31	148.00	67.50	38.8	97.4	88.7	94.7	57.5	36.0	21.0	29.2	26.6	17.0
4.7	27	133.25	64.75	36.4	93.5	73.9	88.5	50.1	34.5	21.3	30.5	27.9	17.2
9.4	29	160.75	69.00	36.7	97.4	83.5	98.7	58.9	35.3	22.6	30.1	26.7	17.6
12.3	32	182.00	73.75	38.7	100.5	88.7	99.8	57.5	38.7	33.9	32.5	27.7	18.4
6.5	29	160.25	71.25	37.3	93.5	84.5	100.6	58.5	38.8	21.5	30.1	26.4	17.9
13.4	27	168.00	71.25	38.1	93.0	79.1	94.5	57.3	36.2	24.5	29.0	30.0	18.8

Tabela C.1: Dados da massa gorda (continuação).

Y[]	x[,1]	x[,2]	x[,3]	x[,4]	x[,5]	x[,6]	x[,7]	x[,8]	x[,9]	x[,10]	x[,11]	x[,12]	x[,13]
20.9	41	218.50	71.00	39.8	111.7	100.5	108.3	67.1	44.2	25.2	37.5	31.5	18.7
31.1	41	247.25	73.50	42.1	117.0	115.6	116.1	71.2	43.3	26.3	37.3	31.7	19.7
38.2	49	191.75	65.00	38.4	118.5	113.1	113.8	61.9	38.3	21.9	32.0	29.8	17.0
23.6	40	202.25	70.00	38.5	106.5	100.9	106.2	63.5	39.9	22.6	35.1	30.6	19.0
27.5	50	196.75	68.25	42.1	105.6	98.8	104.8	66.0	41.5	24.7	33.2	30.5	19.4
33.8	46	363.15	72.25	51.2	136.2	148.1	147.7	87.3	49.1	29.6	45.0	29.0	21.4
31.3	50	203.00	67.00	40.2	114.8	108.1	102.5	61.3	41.1	24.7	34.1	31.0	18.3
33.1	45	262.75	68.75	43.2	128.3	126.2	125.6	72.5	39.6	26.6	36.4	32.7	21.4
30.4	48	217.00	70.00	37.3	113.3	111.2	114.1	67.7	40.9	25.0	36.7	29.8	18.4
30.8	41	212.00	71.50	41.5	106.6	104.3	106.0	65.0	40.2	23.0	35.8	31.5	18.8
8.4	39	125.25	68.00	31.5	85.1	76.0	88.2	50.0	34.7	21.0	26.1	23.1	16.1
14.1	43	164.25	73.25	35.7	96.6	81.5	97.2	58.4	38.2	23.4	29.7	27.4	18.3
11.2	40	133.50	67.50	33.6	88.2	73.7	88.5	53.3	34.5	22.5	27.9	26.2	17.3
6.4	39	148.50	71.25	34.6	89.8	79.5	92.7	52.7	37.5	21.9	28.8	26.8	17.9
13.4	45	135.75	68.50	32.8	92.3	83.4	90.4	52.0	35.8	20.6	28.8	25.5	16.3
5.0	47	127.50	66.75	34.0	83.4	70.4	87.2	50.6	34.4	21.9	26.8	25.8	16.8
10.7	47	158.25	72.25	34.9	90.2	86.7	98.3	52.6	37.2	22.4	26.0	25.8	17.3
7.4	40	139.25	69.00	34.3	89.2	77.9	91.0	51.4	34.9	21.0	26.7	26.1	17.2
8.7	51	137.25	67.75	36.5	89.7	82.0	89.1	49.3	33.7	21.4	29.6	26.0	16.9
7.1	49	152.75	73.50	35.1	93.3	79.6	91.6	52.6	37.6	22.6	38.5	27.4	18.5
4.9	42	136.25	67.50	37.8	87.6	77.6	88.6	51.9	34.9	22.5	27.7	27.5	18.5
22.2	54	198.00	72.00	39.9	107.6	100.0	99.6	57.2	38.0	22.0	35.9	30.2	18.9
20.1	58	181.50	68.00	39.1	100.0	99.8	102.5	62.1	39.6	22.5	33.1	28.3	18.5
27.1	62	201.25	69.50	40.5	111.5	104.2	105.8	61.8	39.8	22.7	37.7	30.9	19.2
30.4	54	202.50	70.75	40.5	115.4	105.3	97.0	59.1	38.0	22.5	31.6	28.8	18.2
24.0	61	179.75	65.75	38.4	104.8	98.3	99.6	60.6	37.7	22.9	34.5	29.6	18.5
25.4	62	216.00	73.25	41.4	112.3	104.8	103.1	61.6	40.9	23.1	36.2	31.8	20.2
28.8	56	178.75	68.50	35.6	102.9	94.7	100.8	60.9	38.0	22.1	32.5	29.8	18.3
29.6	54	193.25	70.25	38.0	107.6	102.4	99.4	61.0	39.4	23.6	32.7	29.9	19.1
25.1	61	178.00	67.00	37.4	105.3	99.7	99.7	60.8	40.1	22.7	33.6	29.0	18.8
31.0	57	205.50	70.00	40.1	105.3	105.5	108.3	65.0	41.2	24.7	35.3	31.1	18.4
28.9	55	183.50	67.50	40.9	103.0	100.3	104.2	64.8	40.2	22.7	34.8	30.1	18.7
21.1	54	151.50	70.75	35.6	90.0	83.9	93.9	55.0	36.1	21.7	29.6	27.4	17.4
14.0	55	154.75	71.50	36.9	95.4	86.6	91.8	54.3	35.4	21.5	32.8	27.4	18.7
7.1	54	155.25	69.25	37.5	89.3	78.4	96.1	56.0	37.4	22.4	32.6	28.1	18.1
13.2	55	156.75	71.50	36.3	94.4	84.6	94.3	51.2	37.4	21.6	27.3	27.1	17.3
23.7	62	167.50	71.50	35.5	97.6	91.5	98.5	56.6	38.6	22.4	31.5	27.3	18.6
9.4	55	146.75	68.75	38.7	88.5	82.8	95.5	58.9	37.6	21.6	30.3	27.3	18.3
9.1	56	160.75	73.75	36.4	93.6	82.9	96.3	52.9	37.5	23.1	29.7	27.3	18.2
13.7	55	125.00	64.00	33.2	87.7	76.0	88.6	50.9	35.4	19.1	29.3	25.7	16.9
12.0	61	143.00	65.75	36.5	93.4	83.3	93.0	55.5	35.2	20.9	29.4	27.0	16.8
18.3	61	148.25	67.50	36.0	91.6	81.8	94.8	54.5	37.0	21.4	29.3	27.0	18.3
9.2	57	162.50	69.50	38.7	91.6	78.8	94.3	56.7	39.7	24.2	30.2	29.2	18.1
21.7	69	177.75	68.50	38.7	102.0	95.0	98.3	55.0	38.3	21.8	30.8	25.7	18.8
21.1	81	161.25	70.25	37.8	96.4	95.4	99.3	53.5	37.5	21.5	31.4	26.8	18.3
18.6	66	171.25	69.25	37.4	102.7	98.6	100.2	56.5	39.3	22.7	30.3	28.7	19.0
30.2	67	163.75	67.75	38.4	97.7	95.8	97.1	54.8	38.2	23.7	29.4	27.2	19.0
26.0	64	150.25	67.25	38.1	97.1	89.0	96.9	54.8	38.0	22.0	29.9	25.2	17.7
18.2	64	190.25	72.75	39.3	103.1	97.8	99.6	58.9	39.0	23.0	34.3	29.6	19.0
26.2	70	170.75	70.00	38.7	101.8	94.9	95.0	56.0	36.5	24.1	31.2	27.3	19.2
26.1	72	168.00	69.25	38.5	101.4	99.8	96.2	56.3	36.6	22.0	29.7	26.3	18.0
25.8	67	167.00	67.50	36.5	98.9	89.7	96.2	54.7	37.8	33.7	32.4	27.7	18.2

Tabela C.1: Dados da massa gorda (continuação).

Y[]	x[,1]	x[,2]	x[,3]	x[,4]	x[,5]	x[,6]	x[,7]	x[,8]	x[,9]	x[,10]	x[,11]	x[,12]	x[,13]
15.0	72	157.75	67.25	37.7	97.5	88.1	96.9	57.2	37.7	21.8	32.6	28.0	18.8
22.6	64	160.00	65.75	36.5	104.3	90.9	93.8	57.8	39.5	23.3	29.2	28.4	18.1
8.8	46	176.75	72.50	38.0	97.3	86.0	99.3	61.0	38.4	23.8	30.2	29.3	18.8
14.3	48	176.00	73.00	36.7	96.7	86.5	98.3	60.4	39.9	24.4	28.8	29.6	18.7
20.2	46	177.00	70.00	37.2	99.7	95.6	102.2	58.3	38.2	22.5	29.1	27.7	17.7
18.1	44	179.75	69.50	39.2	101.9	93.2	100.6	58.9	39.7	23.1	31.4	28.4	18.8
9.2	47	165.25	70.50	37.5	97.2	83.1	95.4	56.9	38.3	22.1	30.1	28.2	18.4
24.2	46	192.50	71.75	38.0	106.6	97.5	100.6	58.9	40.5	24.5	33.3	29.6	19.1
9.6	47	184.25	74.50	37.3	99.6	88.8	101.4	57.4	39.6	24.6	30.3	27.9	17.8
17.3	53	224.50	77.75	41.1	113.2	99.2	107.5	61.7	42.3	23.2	32.9	30.8	20.4
10.1	38	188.75	73.25	37.5	99.1	91.6	102.4	60.6	39.4	22.9	31.6	30.1	18.5
11.1	50	162.50	66.50	38.7	99.4	86.7	96.2	62.1	39.3	23.3	30.6	27.8	18.2
17.7	46	156.50	68.25	35.9	95.1	88.2	92.8	54.7	37.3	21.9	31.6	27.5	18.2
21.7	47	197.00	72.00	40.0	107.5	94.0	103.7	62.7	39.0	22.3	35.3	30.9	18.3
20.8	49	198.50	73.50	40.1	106.5	95.0	101.7	59.0	39.4	22.3	32.2	31.0	18.6
20.1	48	173.75	72.00	37.0	99.1	92.0	98.3	59.3	38.4	22.4	27.9	26.2	17.0
19.8	41	172.75	71.25	36.3	96.7	89.2	98.3	60.0	38.4	23.2	31.0	29.2	18.4
21.9	49	196.75	73.75	40.7	103.5	95.5	101.6	59.1	39.8	25.4	31.0	30.3	19.7
24.7	43	177.00	69.25	39.6	104.0	98.6	99.5	59.5	36.1	22.0	30.1	27.2	17.7
17.8	43	165.50	68.50	31.1	93.1	87.3	96.6	54.7	39.0	24.8	31.0	29.4	18.8
19.1	43	200.25	73.50	38.6	105.2	102.8	103.6	61.2	39.3	23.5	30.5	28.5	18.1
18.2	52	203.25	74.25	42.0	110.0	101.6	100.7	55.8	38.7	23.4	35.1	29.6	19.1
17.2	43	194.00	75.50	38.5	110.1	88.7	102.1	57.5	40.0	24.8	35.1	30.7	19.2
21.0	40	168.50	69.25	34.2	97.8	92.3	100.6	57.5	36.8	22.8	32.1	26.0	17.3
19.5	43	170.75	68.50	37.2	96.3	90.6	99.3	61.9	38.0	22.3	33.3	28.2	18.1
27.1	43	183.25	70.00	37.1	108.0	105.0	103.0	63.7	40.0	23.6	33.5	27.8	17.4
21.6	47	178.25	70.00	40.2	99.7	95.0	98.6	62.3	38.1	23.9	35.3	31.1	19.8
20.9	42	163.00	70.25	35.3	93.5	89.6	99.8	61.5	37.8	21.9	30.7	27.6	17.4
25.9	48	175.25	71.75	38.0	100.7	92.4	97.5	59.3	38.1	21.8	31.8	27.3	17.5
16.7	40	158.00	69.25	36.3	97.0	86.6	92.6	55.9	36.3	22.1	29.8	26.3	17.3
19.8	48	177.25	72.75	36.8	96.0	90.0	99.7	58.8	38.4	22.8	29.9	28.0	18.1
14.1	51	179.00	72.00	41.0	99.2	90.0	96.4	56.8	38.8	23.3	33.4	29.8	19.5
25.1	40	191.00	74.00	38.3	95.4	92.4	104.3	64.6	41.1	24.8	33.6	29.5	18.5
17.9	44	187.50	72.25	38.0	101.8	87.5	101.0	58.5	39.2	24.5	32.1	28.6	18.0
27.0	52	206.50	74.50	40.8	104.3	99.2	104.1	58.5	39.3	24.6	33.9	31.2	19.5
24.6	44	185.25	71.50	39.5	99.2	98.1	101.4	57.1	40.5	23.2	33.0	29.6	18.4
14.8	40	160.25	68.75	36.9	99.3	83.3	97.5	60.5	38.7	22.6	34.4	28.0	17.6
16.0	47	151.50	66.75	36.9	94.0	86.1	95.2	58.1	36.5	22.1	30.6	27.5	17.6
14.0	50	161.00	66.50	37.7	98.9	84.1	94.0	58.5	36.6	23.5	34.4	29.2	18.0
17.4	46	167.00	67.00	36.6	101.0	89.9	100.0	60.7	36.0	21.9	35.6	30.2	17.6
26.4	42	177.50	68.75	38.9	98.7	92.1	98.5	60.7	36.8	22.2	33.8	30.3	17.2
17.4	43	152.25	67.75	37.5	95.9	78.0	93.2	53.5	35.8	20.8	33.9	28.2	17.4
20.4	40	192.25	73.25	39.8	103.9	93.5	99.5	61.7	39.0	21.8	33.3	29.6	18.1
15.0	42	165.25	69.75	38.3	96.2	87.0	97.8	57.4	36.9	22.2	31.6	27.8	17.7
18.0	49	171.75	71.50	35.5	97.8	90.1	95.8	57.0	38.7	23.2	27.5	26.5	17.6
22.2	40	171.25	70.50	36.3	94.6	90.3	99.1	60.3	38.5	23.0	31.2	28.4	17.1
23.1	47	197.00	73.25	37.8	103.6	99.8	103.2	61.2	38.1	22.6	33.5	28.6	17.9
25.3	50	157.00	66.75	37.8	100.4	89.4	92.3	56.1	35.6	20.5	33.6	29.3	17.3
23.8	41	168.25	69.50	36.5	98.4	87.2	98.4	56.0	36.9	23.0	34.0	29.8	18.1
26.3	44	186.00	69.75	37.8	104.6	101.1	102.1	58.9	37.9	22.7	30.9	28.8	17.6
21.4	39	166.75	70.75	37.0	92.9	86.1	95.6	58.8	36.1	22.4	32.7	28.3	17.1
28.4	43	187.75	74.00	37.7	97.8	98.6	100.6	63.6	39.2	23.8	34.3	28.4	17.7

Tabela C.1: Dados da massa gorda (continuação).

Y[]	x[,1]	x[,2]	x[,3]	x[,4]	x[,5]	x[,6]	x[,7]	x[,8]	x[,9]	x[,10]	x[,11]	x[,12]	x[,13]
21.8	40	168.25	71.25	34.3	98.3	88.5	98.3	58.1	38.4	22.5	31.7	27.4	17.6
20.1	49	212.75	75.00	40.8	104.7	106.6	107.7	66.5	42.5	24.5	35.5	29.8	18.7
24.3	40	176.75	71.00	37.4	98.6	93.1	101.6	59.1	39.6	21.6	30.8	27.9	16.6
18.1	40	173.25	69.50	36.5	99.5	93.0	99.3	60.4	38.2	22.0	32.0	28.5	17.8
22.7	52	167.00	67.75	37.5	102.7	91.0	98.9	57.1	36.7	22.3	31.6	27.5	17.9
9.9	23	159.75	72.25	35.5	92.1	77.1	93.9	56.1	36.1	22.7	30.5	27.2	18.2
10.8	23	188.15	77.50	38.0	96.6	85.3	102.5	59.1	37.6	23.2	31.8	29.7	18.3
14.4	24	156.00	70.75	35.7	92.7	81.9	95.3	56.4	36.5	22.0	33.5	28.3	17.3
19.0	24	208.50	72.75	39.2	102.0	99.1	110.1	71.2	43.5	25.2	36.1	30.3	18.7
28.6	25	206.50	69.75	40.9	110.9	100.5	106.2	68.4	40.8	24.6	33.3	29.7	18.4
6.1	25	143.75	72.50	35.2	92.3	76.5	92.1	51.9	35.7	22.0	25.8	25.2	16.9
24.5	26	223.00	70.25	40.6	114.1	106.8	113.9	67.6	42.7	24.7	36.0	30.4	18.4
9.9	26	152.25	69.00	35.4	92.9	77.6	93.5	56.9	35.9	20.4	31.6	29.0	17.8
19.1	26	241.75	74.50	41.8	108.3	102.9	114.4	72.9	43.5	25.1	38.5	33.8	19.6
10.6	27	146.00	72.25	34.1	88.5	72.8	91.1	53.6	36.8	23.8	27.8	26.3	17.4
16.5	27	156.75	67.25	37.9	94.0	88.2	95.2	56.8	37.4	22.8	30.6	28.3	17.9
20.5	27	200.25	73.50	38.2	101.1	100.1	105.0	62.1	40.0	24.9	33.7	29.2	19.4
17.2	28	171.50	75.25	35.6	92.1	83.5	98.3	57.3	37.8	21.7	32.2	27.7	17.7
30.1	28	205.75	69.00	38.5	105.6	105.0	106.4	68.6	40.0	25.2	35.2	30.7	19.1
10.5	28	182.50	72.25	37.0	98.5	90.8	102.5	60.8	38.5	25.0	31.6	28.0	18.6
12.8	30	136.50	68.75	35.9	88.7	76.6	89.8	50.1	34.8	21.8	27.0	34.9	16.9
22.0	31	177.25	71.50	36.2	101.1	92.4	99.3	59.4	39.0	24.6	30.1	28.2	18.2
9.9	31	151.25	72.25	35.0	94.0	81.2	91.5	52.5	36.6	21.0	27.0	26.3	16.5
14.8	33	196.00	73.00	38.5	103.8	95.6	105.1	61.4	40.6	25.0	31.3	29.2	19.1
13.3	33	184.25	68.75	40.7	98.9	92.1	103.5	64.0	37.3	23.5	33.5	30.6	19.7
15.2	34	140.00	70.50	36.0	89.2	83.4	89.6	52.4	35.6	20.4	28.3	26.2	16.5
26.5	34	218.75	72.00	39.5	111.4	106.0	108.8	63.8	42.0	23.4	34.0	31.2	18.5
19.0	35	217.00	73.75	40.5	107.5	95.1	104.5	64.8	41.3	25.6	36.4	33.7	19.4
21.4	35	166.25	68.00	38.5	99.1	90.4	95.6	55.5	34.2	21.9	30.2	28.7	17.7
20.0	35	224.75	72.25	43.9	108.2	100.4	106.8	63.3	41.7	24.6	37.2	33.1	19.8
34.7	35	228.25	69.50	40.4	114.9	115.9	111.9	74.4	40.6	24.0	36.1	31.8	18.8
16.5	35	172.75	69.50	37.6	99.1	90.8	98.1	60.1	39.1	23.4	32.5	29.8	17.4
4.1	35	152.25	67.75	37.0	92.2	81.9	92.8	54.7	36.2	22.1	30.4	27.4	17.7
1.9	35	125.75	65.50	34.0	90.8	75.0	89.2	50.0	34.8	22.0	24.8	25.9	16.9
20.2	35	177.25	71.00	38.4	100.5	90.3	98.7	57.8	37.3	22.4	31.0	28.7	17.7
16.8	36	176.25	71.50	38.7	98.2	90.3	99.9	59.2	37.7	21.5	32.4	28.4	17.8
24.6	36	226.75	71.75	41.5	115.3	108.8	114.4	69.2	42.4	24.0	35.4	21.0	20.1
10.4	37	145.25	69.25	36.0	96.8	79.4	89.2	50.3	34.8	22.2	31.0	26.9	16.9
13.4	37	151.00	67.00	35.3	92.6	83.2	96.4	60.0	38.1	22.0	31.5	26.6	16.7
28.8	37	241.25	71.50	42.1	119.2	110.3	113.9	69.8	42.6	24.8	34.4	29.5	18.4
22.0	38	187.25	69.25	38.0	102.7	92.7	101.9	64.7	39.5	24.7	34.8	30.3	18.1
16.8	39	234.75	74.50	42.8	109.5	104.5	109.9	69.5	43.1	25.8	39.1	32.5	19.9
25.8	39	219.25	74.25	40.0	108.5	104.6	109.8	68.1	42.8	24.1	35.6	29.0	19.0
0.0	40	118.50	68.00	33.8	79.3	69.4	85.0	47.2	33.5	20.2	27.7	24.6	16.5
11.9	40	145.75	67.25	35.5	95.5	83.6	91.6	54.1	36.2	21.8	31.4	28.3	17.2
12.4	40	159.25	69.75	35.3	92.3	86.8	96.1	58.0	39.4	22.7	30.0	26.4	17.4
17.4	40	170.50	74.25	37.7	98.9	90.4	95.5	55.4	38.9	22.4	30.5	28.9	17.7
9.2	40	167.50	71.50	39.4	89.5	83.7	98.1	57.3	39.7	22.6	32.9	29.3	18.2
23.0	41	232.75	74.25	41.9	117.5	109.3	108.8	67.7	41.3	24.7	37.2	31.8	20.0
20.1	41	210.50	72.00	38.5	107.4	98.9	104.1	63.5	39.8	23.5	36.4	30.4	19.1
20.2	41	202.25	72.50	40.8	109.2	98.0	101.8	62.8	41.3	24.8	36.6	32.4	18.8
23.8	41	185.00	68.25	38.0	103.4	101.2	103.1	61.5	40.4	22.9	33.4	29.2	18.5

Tabela C.1: Dados da massa gorda (continuação).

Y[]	x[,1]	x[,2]	x[,3]	x[,4]	x[,5]	x[,6]	x[,7]	x[,8]	x[,9]	x[,10]	x[,11]	x[,12]	x[,13]
11.8	41	153.00	69.25	36.4	91.4	80.6	92.3	54.3	36.3	21.8	29.6	27.3	17.9
36.5	42	244.25	76.00	41.8	115.2	113.7	112.4	68.5	45.0	25.5	37.1	31.2	19.9
16.0	42	193.50	70.50	40.7	104.9	94.1	102.7	60.6	38.6	24.7	34.0	30.1	18.7
24.0	42	224.75	74.75	38.5	106.7	105.7	111.8	65.3	43.3	26.0	33.7	29.9	18.5
22.3	42	162.75	72.75	35.4	92.2	85.6	96.5	60.2	38.9	22.4	31.7	27.1	17.1
24.8	42	180.00	68.25	38.5	101.6	96.6	100.6	61.1	38.4	24.1	32.9	29.8	18.8
21.5	42	156.25	69.00	35.5	97.8	86.0	96.2	57.7	38.6	24.0	31.2	27.3	17.4
17.6	42	168.00	71.50	36.5	92.0	89.7	101.0	62.3	38.0	22.3	30.8	27.8	16.9
7.3	42	167.25	72.75	37.6	94.0	78.0	99.0	57.5	40.0	22.5	30.6	30.0	18.5
22.6	43	170.75	67.50	37.4	103.7	89.7	94.2	58.5	39.0	24.1	33.8	28.8	18.8
12.5	43	178.25	70.25	37.8	102.7	89.2	99.2	60.2	39.2	23.8	31.7	28.4	18.6
21.7	43	150.00	69.25	35.2	91.1	85.7	96.9	55.5	35.7	22.0	29.4	26.6	17.4
27.7	43	200.50	71.50	37.9	107.2	103.1	105.5	68.8	38.3	23.7	32.1	28.9	18.7
6.8	44	184.00	74.00	37.9	100.8	89.1	102.6	60.6	39.0	24.0	32.9	29.2	18.4
33.4	44	223.00	69.75	40.9	121.6	113.9	107.1	63.5	40.3	21.8	34.8	30.7	17.4
16.6	44	208.75	73.00	41.9	105.6	96.3	102.0	63.3	39.8	24.1	37.3	23.1	19.4
31.7	44	166.00	65.50	39.1	100.6	93.9	100.1	58.9	37.6	21.4	33.1	29.5	17.3
31.5	47	195.00	72.50	40.2	102.7	101.3	101.7	60.7	39.4	23.3	36.7	31.6	18.4
10.1	47	160.50	70.25	36.0	99.8	83.9	91.8	53.0	36.2	22.5	31.4	27.5	17.7
11.3	47	159.75	70.75	34.5	92.9	84.4	94.0	56.0	38.2	22.6	29.0	26.2	17.6
7.8	49	140.50	68.00	35.8	91.2	79.4	89.0	51.1	35.0	21.7	30.9	28.8	17.4
26.4	49	216.25	74.50	40.2	115.6	104.0	109.0	63.7	40.3	23.2	36.8	31.0	18.9
19.3	49	168.25	71.75	38.3	98.3	89.7	99.1	56.3	38.8	23.0	29.5	27.9	18.6
18.5	50	194.75	70.75	39.0	103.7	97.6	104.2	60.0	40.9	25.5	32.7	30.0	19.0
19.3	50	172.75	73.00	37.4	98.7	87.6	96.1	57.1	38.1	21.8	28.6	26.7	18.0
45.1	51	219.00	64.00	41.2	119.8	122.1	112.8	62.5	36.9	23.6	34.7	29.1	18.4
13.8	51	149.25	69.75	34.8	92.8	81.1	96.3	53.8	36.5	21.5	31.3	26.3	17.8
8.2	51	154.50	70.00	36.9	93.3	81.5	94.4	54.7	39.0	22.6	27.5	25.9	18.6
23.9	52	199.25	71.75	39.4	106.8	100.0	105.0	63.9	39.2	22.9	35.7	30.4	19.2
15.1	53	154.50	69.25	37.6	93.9	88.7	94.5	53.7	36.2	22.0	28.5	25.7	17.1
12.7	54	153.25	70.50	38.5	99.0	91.8	96.2	57.7	38.1	23.9	31.4	29.9	18.9
25.3	54	230.00	72.25	42.5	119.9	110.4	105.5	64.2	42.7	27.0	38.4	32.0	19.6
11.9	54	161.75	67.50	37.4	94.2	87.6	95.6	59.7	40.2	23.4	27.9	27.0	17.8
6.1	55	142.25	67.25	35.2	92.7	82.8	91.9	54.4	35.2	22.5	29.4	26.8	17.0
11.3	55	179.75	68.75	41.1	106.9	95.3	98.2	57.4	37.1	21.8	34.1	31.1	19.2
12.8	55	126.50	66.75	33.4	88.8	78.2	87.5	50.8	33.0	19.7	25.3	22.0	15.8
14.9	55	169.50	68.25	37.2	101.7	91.1	97.1	56.6	38.5	22.6	33.4	29.3	18.8
24.5	55	198.50	74.25	38.3	105.3	96.7	106.6	64.0	42.6	23.4	33.2	30.0	18.4
15.0	56	174.50	69.50	38.1	104.0	89.4	98.4	58.4	37.4	22.5	34.6	30.1	18.8
16.9	56	167.75	68.50	37.4	98.6	93.0	97.0	55.4	38.8	23.2	32.4	29.7	19.0
11.1	57	147.75	65.75	35.2	99.6	86.4	90.1	53.0	35.0	21.3	31.7	27.3	16.9
16.1	57	182.25	71.75	39.4	103.4	96.7	100.7	59.3	38.6	22.8	31.8	29.1	19.0
15.5	58	175.50	71.50	38.0	100.2	88.1	97.8	57.1	38.9	23.6	30.9	29.6	18.0
25.9	58	161.75	67.25	35.1	94.9	94.9	100.2	56.8	35.9	21.0	27.8	26.1	17.6
25.5	60	157.75	67.50	40.4	97.2	93.3	94.0	54.3	35.7	21.0	31.3	28.7	18.3
18.4	62	168.75	67.50	38.3	104.7	95.6	93.7	54.4	37.1	22.7	30.3	26.3	18.3
24.0	62	191.50	72.25	40.6	104.0	98.2	101.1	59.3	40.3	23.0	32.6	28.5	19.0
26.4	63	219.15	69.50	40.2	117.6	113.8	111.8	63.4	41.1	22.3	35.1	29.6	18.5
12.7	64	155.25	69.50	37.9	95.8	82.8	94.5	61.2	39.1	22.3	29.8	28.9	18.3
28.8	65	189.75	65.75	40.8	106.4	100.5	100.5	59.2	38.1	24.0	35.9	30.5	19.1
17.0	65	127.50	65.75	34.7	93.0	79.7	87.6	50.7	33.4	20.1	28.5	24.8	16.5
33.6	65	224.50	68.25	38.8	119.6	118.0	114.3	61.3	42.1	23.4	34.9	30.1	19.4

Tabela C.1: Dados da massa gorda (continuação).

Y[]	x[,1]	x[,2]	x[,3]	x[,4]	x[,5]	x[,6]	x[,7]	x[,8]	x[,9]	x[,10]	x[,11]	x[,12]	x[,13]
29.3	66	234.25	72.00	41.4	119.7	109.0	109.1	63.7	42.4	24.6	35.6	30.7	19.5
31.4	67	227.75	72.75	41.3	115.8	113.4	109.8	65.6	46.0	25.4	35.3	29.8	19.5
28.1	67	199.50	68.50	40.7	118.3	106.1	101.6	58.2	38.8	24.1	32.1	29.3	18.5
15.3	68	155.50	69.25	36.3	97.4	84.3	94.4	54.3	37.5	22.6	29.2	27.3	18.5
29.1	69	215.50	70.50	40.8	113.7	107.6	110.0	63.3	44.0	22.6	37.5	32.6	18.8
11.5	70	134.25	67.00	34.9	89.2	83.6	88.8	49.6	34.8	21.5	25.6	25.7	18.5
32.3	72	201.00	69.75	40.9	108.5	105.0	104.5	59.6	40.8	23.2	35.2	28.6	20.1
28.3	72	186.75	66.00	38.9	111.1	111.5	101.7	60.3	37.3	21.5	31.3	27.2	18.0
25.3	72	190.75	70.50	38.9	108.3	101.3	97.8	56.0	41.6	22.7	30.5	29.4	19.8
30.7	74	207.50	70.00	40.8	112.4	108.5	107.1	59.3	42.2	24.6	33.7	30.0	20.9

Tabela C.2: Dados da homocisteinemia.

Y[]	x[,1]	x[,2]	x[,3]	x[,4]	x[,5]	x[,6]	x[,7]	x[,8]	x[,9]	x[,10]	x[,11]	x[,12]	x[,13]
6.44	0	0	1	0	74	23	0.7	11	78	106.2	53	110	70
9.65	0	0	0	1	101	38	0.9	15	73	164.0	69	136	86
8.00	1	0	1	1	163	28	0.6	61	40	186.4	36	140	95
9.81	1	1	0	0	98	25	1.1	64	98	158.2	96	120	80
11.07	1	1	0	1	97	39	0.9	51	94	215.8	46	116	76
7.08	0	0	1	1	90	40	0.9	21	74	200.0	63	166	100
7.27	1	1	0	1	90	41	1.0	33	82	167.4	29	126	90
8.16	1	0	1	0	86	32	1.3	16	118	107.6	46	136	86
7.34	0	0	0	0	94	22	0.6	129	110	104.6	58	150	100
7.60	1	1	0	1	93	41	0.9	20	62	95.8	68	140	100
9.67	1	0	1	1	78	36	1.0	52	75	176.4	36	116	70
10.88	1	1	0	1	131	27	1.0	22	48	128.2	44	136	86
9.14	1	0	0	0	90	36	1.1	36	54	144.6	46	150	90
8.39	1	0	1	1	92	33	0.9	29	69	197.2	56	166	100
11.60	1	0	1	0	146	38	1.0	21	88	171.0	38	140	90
5.83	0	0	0	1	101	30	0.9	34	71	205.8	36	180	106
7.13	1	0	0	0	95	42	0.9	69	62	189.4	58	126	80
6.16	1	0	0	1	209	48	0.9	19	47	191.8	50	136	86
8.04	0	0	0	0	91	39	0.9	31	54	118.6	46	166	106
7.36	1	0	1	1	82	31	0.9	29	76	135.6	34	130	86
7.81	1	0	1	0	84	42	1.0	21	89	152.8	42	126	80
10.91	1	0	0	0	285	36	1.0	17	74	204.8	46	160	100
13.26	1	0	0	0	76	30	1.2	24	58	193.6	47	130	90
9.40	1	0	1	0	117	43	1.1	41	59	156.4	29	150	90
11.74	1	0	0	1	115	35	1.3	31	74	144.4	51	150	90
8.24	1	0	0	0	97	17	1.1	38	52	167.0	33	160	100
7.33	1	1	0	1	111	20	0.8	67	108	232.8	28	106	86
9.05	0	0	0	1	273	37	0.9	27	120	119.0	36	180	100
6.95	1	0	0	0	97	36	1.2	54	71	143.0	30	120	76
9.26	1	0	1	1	83	38	1.2	88	108	102.0	53	140	86
9.71	1	0	0	1	397	50	1.1	16	65	104.6	56	140	90
6.05	1	0	0	1	130	27	0.9	189	74	212.4	41	120	80

Tabela C.2: Dados da homocisteinemia (continuação).

Y[.]	x[,1]	x[,2]	x[,3]	x[,4]	x[,5]	x[,6]	x[,7]	x[,8]	x[,9]	x[,10]	x[,11]	x[,12]	x[,13]
7.43	0	0	1	1	99	36	0.8	27	87	213.4	66	110	60
8.99	1	0	1	0	95	37	0.8	45	68	158.4	56	140	90
30.35	1	0	1	0	63	167	12.1	14	75	83.2	19	140	90
11.12	1	0	0	0	92	40	1.0	19	61	135.8	56	106	76
11.40	1	0	1	0	103	45	0.9	79	96	137.4	32	140	90
11.29	1	0	1	1	113	29	1.0	27	164	146.4	50	120	70
13.16	1	0	1	0	78	31	0.9	49	119	169.0	38	130	86
11.72	1	0	1	0	76	25	1.0	53	91	157.8	47	180	110
38.47	1	1	0	0	127	21	1.0	56	146	168.6	28	126	80
37.65	1	1	0	0	87	32	1.0	31	96	182.8	34	126	76
17.76	1	0	1	1	89	41	1.7	27	111	159.8	27	130	80
11.13	1	0	1	0	76	34	1.1	22	55	131.0	35	120	80
11.82	1	0	1	1	95	34	1.1	20	50	95.2	53	136	86
10.32	1	0	1	1	77	21	0.9	108	66	191.2	48	110	80
9.13	1	0	1	1	88	45	1.2	39	64	105.2	44	150	90
13.43	1	0	0	0	98	46	1.1	126	95	190.0	54	110	70
9.98	1	0	0	1	100	37	0.9	37	81	144.8	38	130	86
13.63	1	0	1	1	82	25	1.0	30	78	250.4	44	150	90
21.04	1	1	0	0	91	22	1.0	39	63	232.6	57	130	80
13.05	1	0	1	1	104	34	0.9	20	74	86.2	36	130	86
14.29	1	0	0	1	109	42	1.0	84	101	145.2	44	110	86
12.86	1	1	0	1	89	32	1.0	25	66	167.8	45	140	86
13.81	1	1	0	1	89	32	1.0	23	54	260.4	53	130	100
10.59	1	0	1	1	91	34	1.0	74	72	227.4	47	130	90
10.02	1	0	1	0	97	28	1.1	33	75	93.8	38	130	80
15.83	1	0	0	0	84	34	1.3	87	249	78.0	64	140	86
7.43	0	0	0	1	84	42	1.0	13	92	145.0	60	170	110
9.72	0	0	0	1	85	59	1.1	22	100	203.4	58	120	80
9.16	1	1	0	0	80	32	1.1	63	69	153.0	55	120	86
10.12	1	1	0	1	100	25	1.1	115	140	168.0	46	130	90
7.97	1	0	1	1	98	25	0.9	55	61	178.8	45	130	90
7.92	0	0	0	0	80	35	0.7	12	75	128.6	68	120	70
8.95	1	0	1	0	92	34	1.1	193	100	141.6	41	120	80
10.20	1	1	0	0	85	30	1.0	28	64	226.6	35	110	70
7.48	0	0	0	0	79	26	0.8	21	136	121.6	54	126	80
7.50	1	0	1	0	91	46	1.0	51	60	178.4	36	110	70
9.54	1	0	1	1	89	35	1.1	26	60	125.4	49	136	86
8.58	1	0	1	1	97	44	1.1	192	84	180.8	47	140	80
9.47	1	0	1	1	93	39	1.0	61	94	177.8	68	130	82
10.94	1	0	1	1	96	42	1.7	25	31	145.8	44	130	80
10.59	1	0	1	1	110	49	1.1	20	63	140.8	29	126	80
10.85	1	1	0	0	96	36	0.9	59	64	149.8	46	140	90
7.83	1	0	1	0	87	32	1.1	26	64	127.4	46	136	86
6.61	1	1	0	0	115	29	0.9	43	69	146.6	31	136	88
8.09	1	0	1	1	95	35	1.2	93	81	204.6	51	120	76
7.75	0	0	0	0	93	26	0.7	10	46	132.8	41	140	90
8.16	1	0	1	1	94	30	0.8	160	84	206.4	47	140	100

Tabela C.2: Dados da homocisteinemia (continuação).

Y[]	x[,1]	x[,2]	x[,3]	x[,4]	x[,5]	x[,6]	x[,7]	x[,8]	x[,9]	x[,10]	x[,11]	x[,12]	x[,13]
8.66	1	0	1	1	118	38	1.1	139	126	118.6	39	120	80
7.76	1	0	0	0	87	32	0.9	34	108	151.4	40	150	96
13.80	1	1	0	1	83	40	1.0	26	91	154.6	31	130	70
5.40	0	0	0	0	84	34	0.8	10	101	147.0	64	156	106
7.64	1	0	1	1	100	32	1.0	17	69	110.6	29	170	100
8.07	1	0	1	1	115	32	1.0	26	77	138.8	47	170	106
7.33	1	0	1	1	85	37	0.9	84	78	167.6	42	140	90
7.72	1	0	0	1	263	29	0.8	108	78	222.8	48	130	100
10.40	1	0	0	1	148	30	1.0	64	60	219.0	56	140	86
9.07	1	0	1	0	88	55	1.5	49	46	141.6	58	146	80
13.03	0	1	0	0	110	46	1.0	17	97	98.6	65	130	76
9.45	1	1	0	1	91	36	0.9	76	29	145.0	42	140	76
7.00	1	0	0	1	91	48	0.9	25	68	156.8	38	130	86
10.18	1	1	0	1	98	47	0.9	22	59	165.4	30	136	86
7.38	0	1	0	1	92	26	0.8	13	94	191.6	41	140	90
7.86	0	0	0	1	89	28	1.0	23	76	171.2	60	140	80
11.15	1	0	1	0	94	28	1.0	114	78	200.2	45	140	96
8.20	1	0	0	1	92	46	1.1	41	83	181.4	38	140	80
10.14	1	0	1	1	94	37	1.0	24	124	153.8	30	160	110
12.24	1	0	1	0	85	53	1.3	40	51	197.6	30	130	86
10.70	1	0	0	0	88	36	1.0	24	56	159.2	48	120	80
10.36	1	0	0	1	76	37	1.1	17	86	158.6	30	120	80
9.69	1	0	0	0	271	43	0.9	33	109	146.0	39	130	90
6.10	1	0	0	0	112	29	0.8	27	85	109.8	51	146	110
12.11	1	0	0	1	81	32	1.0	34	69	169.2	55	130	80
11.15	1	0	1	0	86	24	1.0	41	67	153.2	56	150	96
9.08	1	0	1	0	83	19	0.8	25	60	128.8	45	116	70
7.63	0	0	0	0	140	48	1.5	91	97	146.2	39	156	100
11.84	1	0	0	1	134	59	1.2	17	37	162.6	38	130	80
10.89	1	0	1	0	80	56	1.3	20	76	120.4	46	140	90
9.00	1	0	1	1	96	20	0.9	39	100	164.8	32	120	80
9.60	1	0	0	1	91	31	1.1	23	103	173.2	45	130	90
11.10	1	0	1	0	96	40	1.0	25	66	177.4	54	136	86
8.55	1	0	1	1	78	25	0.9	47	85	188.8	33	150	100
10.85	1	0	1	0	94	38	1.3	40	102	193.8	80	146	86
10.66	1	1	0	1	121	46	1.0	44	99	160.4	43	176	96
7.10	1	0	0	1	138	37	1.1	41	57	131.4	51	170	100
12.38	1	0	0	1	90	35	1.3	29	98	115.6	53	130	80
5.98	0	0	0	1	108	44	0.8	12	96	130.6	58	146	96
11.33	1	1	0	1	144	30	1.3	150	149	184.2	36	145	95
11.05	1	0	0	1	93	40	0.9	45	87	186.0	42	130	80
8.87	1	0	1	1	130	22	1.1	25	94	229.4	56	130	80
9.05	1	0	1	1	87	37	1.0	26	70	168.2	80	170	98
8.68	1	0	1	1	79	35	0.8	34	86	96.2	104	126	84
11.30	1	0	1	0	85	43	1.0	43	84	274.4	60	150	98
8.67	0	0	0	0	97	50	0.7	21	77	157.2	67	190	100
11.11	1	0	1	1	148	46	1.0	21	88	135.2	49	136	86

Anexo D

Códigos do WinBUGS

D.1 Dados da massa gorda

D.1.1 Método SIMD

A 1ª iteração do método SIMD, para os dados da massa gorda, começa com o ajustamento do modelo de regressão linear normal com as 13 covariáveis e o cálculo do valor das medidas de discrepância deste modelo. O código do WinBUGS para o ajustamento do modelo referente a esta iteração do método SIMD é

```
model massagorda_SIMD_13covar;
const
  pt=13,          # número total de covariáveis
  p=13,          # número de covariáveis no modelo
  N=251,         # número de observações
  PI=3.141593;
var
  d[N,pt]       # todas as covariáveis
  x[N,p],       # covariáveis consideradas
  z[N,p],       # covariáveis standardizadas
  Y[N],         # variável resposta
  mu[N],        # valor médio
  beta0, beta[p], # constante e coeficientes standardizados
  b0, b[p],     # constante e coeficientes não standardizados
  resid[N], sresid[N], # resíduos e resíduos standardizados
  ver[N], log.ver[N], # verosimilhança e log-verosimilhança
  inv.ver[N],   # ordenada preditiva condicional
  num.par,     # número de parâmetros do modelo
  BIC.est,     # BIC estimado
  soma.log.ver; # LLP
{
  # leitura dos dados
  for (i in 1:N){
    x[i,1]<-d[i,in.mod[1]]
    x[i,2]<-d[i,in.mod[2]]
    x[i,3]<-d[i,in.mod[3]]
    x[i,4]<-d[i,in.mod[4]]
    x[i,5]<-d[i,in.mod[5]]
    x[i,6]<-d[i,in.mod[6]]
    x[i,7]<-d[i,in.mod[7]]
    x[i,8]<-d[i,in.mod[8]]
    x[i,9]<-d[i,in.mod[9]]
    x[i,10]<-d[i,in.mod[10]]
    x[i,11]<-d[i,in.mod[11]]
  }
}
```

```

    x[i,12]<-d[i,in.mod[12]]
    x[i,13]<-d[i,in.mod[13]]
  }
  # standardização dos x's e dos coeficientes
  for (j in 1:p){
    b[j]<-beta[j]/sd(x[,j])
    for (i in 1:N){
      z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
    }
  }
  b0<-beta0-b[1]*mean(x[,1])-b[2]*mean(x[,2])-b[3]*mean(x[,3])-b[4]*mean(x[,4])
    -b[5]*mean(x[,5])-b[6]*mean(x[,6])-b[7]*mean(x[,7])-b[8]*mean(x[,8])
    -b[9]*mean(x[,9])-b[10]*mean(x[,10])-b[11]*mean(x[,11])
    -b[12]*mean(x[,12])-b[13]*mean(x[,13])
  # modelo
  for (i in 1:N){
    mu[i]<-beta0+beta[1]*z[i,1]+beta[2]*z[i,2]+beta[3]*z[i,3]+beta[4]*z[i,4]
      +beta[5]*z[i,5]+beta[6]*z[i,6]+beta[7]*z[i,7]+beta[8]*z[i,8]
      +beta[9]*z[i,9]+beta[10]*z[i,10]+beta[11]*z[i,11]
      +beta[12]*z[i,12]+beta[13]*z[i,13]
    Y[i]~dnorm(mu[i],tau)
  }
  # priors
  beta0~dnorm(0.0,1.0E-4)
  for(j in 1:p){
    beta[j]~dnorm(0.0,1.0E-4)
  }
  tau~dgamma(0.001,0.001)
  sigma<-1/sqrt(tau)
  # selecção de covariáveis
  for (i in 1:N){
    resid[i]<-Y[i]-mu[i]
    sresid[i]<-resid[i]*sqrt(tau)
    ver[i]<-sqrt(tau/(2*PI))*exp(-0.5*pow(sresid[i],2))
    log.ver[i]<-log(ver[i])
    inv.ver[i]<-1/ver[i]
  }
  num.par<-p+2
  BIC.est<--2*sum(log.ver[])+num.par*log(N)
  soma.log.ver<--2*sum(log.ver[])
}

```

Os dados do exemplo da massa gorda estão no ficheiro `bodyfat.dat.odc`, onde Y representa a variável resposta e $d[j]$ as diferentes covariáveis. Os restantes valores de *input* são dados pela instrução `list(N=251, p=13, PI=3.141593, in.mod=c(1,2,3,4,5,6,7,8,9,10,11,12,13))`. A inicialização dos parâmetros do modelo é dada pela instrução `list(beta0=10, beta=c(0,0,0,0,0,0,0,0,0,0,0,0), tau=0.1)`.

As estimativas dos parâmetros b_0, b_1, \dots, b_{13} e $\tau = 1/\sigma^2$, do modelo de regressão linear normal, são obtidas através da monitorização de b_0, b e τ . Para obter o valor do DIC é necessário seleccionar a opção DIC do menu *Inference* e seleccionar a opção *set* da caixa de diálogo; depois de efectuadas as iterações consideradas necessárias, a selecção da opção *stats* da caixa de diálogo do DIC possibilita a obtenção de estimativas desta medida de discrepância. A monitorização de $\widehat{BIC.est}$ e $soma.log.ver$ permite a obtenção de estimativas das medidas de discrepância \widehat{BIC} e LLP, respectivamente. As ordenadas preditivas condicionais, $p(x_i|x_{(-i)})$, são obtidas através da monitorização de $inv.ver$ e $SlnCPO = \sum_{i=1}^n \ln p(x_i|x_{(-i)})$ obtém-se de modo imediato.

Os códigos do WinBUGS para efectuar as restantes iterações do método SIMD são idênticos ao anterior, sendo apenas necessário alterar o valor de p e adaptar o modelo, uma vez que em cada iteração é excluída uma covariável. Por exemplo, o código do

WinBUGS para o ajustamento do modelo de regressão linear normal da 6^a iteração é

```

model massagorda_SIMD_8covar;
const
  pt=13,           # número total de covariáveis
  p=8,            # número de covariáveis no modelo
  N=251,          # número de observações
  PI=3.141593;
var
  d[N,pt]         # todas as covariáveis
  x[N,p],         # covariáveis consideradas
  z[N,p],         # covariáveis standardizadas
  Y[N],           # variável resposta
  mu[N],          # valor médio
  beta0, beta[p], # constante e coeficientes standardizados
  b0, b[p],       # constante e coeficientes não standardizados
  resid[N], sresid[N], # resíduos e resíduos standardizados
  ver[N], log.ver[N], # verosimilhança e log-verosimilhança
  inv.ver[N],     # ordenada preditiva condicional
  num.par,        # número de parâmetros do modelo
  BIC.est,        # BIC estimado
  soma.log.ver;   # LLP
{
  # leitura dos dados
  for (i in 1:N){
    x[i,1]<-d[i,in.mod[1]]
    x[i,2]<-d[i,in.mod[2]]
    x[i,3]<-d[i,in.mod[3]]
    x[i,4]<-d[i,in.mod[4]]
    x[i,5]<-d[i,in.mod[5]]
    x[i,6]<-d[i,in.mod[6]]
    x[i,7]<-d[i,in.mod[7]]
    x[i,8]<-d[i,in.mod[8]]
  }
  # standardização dos x's e dos coeficientes
  for (j in 1:p){
    b[j]<-beta[j]/sd(x[,j])
    for (i in 1:N){
      z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
    }
  }
  b0<-beta0-b[1]*mean(x[,1])-b[2]*mean(x[,2])-b[3]*mean(x[,3])-b[4]*mean(x[,4])
    -b[5]*mean(x[,5])-b[6]*mean(x[,6])-b[7]*mean(x[,7])-b[8]*mean(x[,8])
  # modelo
  for (i in 1:N){
    mu[i]<-beta0+beta[1]*z[i,1]+beta[2]*z[i,2]+beta[3]*z[i,3]+beta[4]*z[i,4]
      +beta[5]*z[i,5]+beta[6]*z[i,6]+beta[7]*z[i,7]+beta[8]*z[i,8]
    Y[i]~dnorm(mu[i],tau)
  }
  # priors
  beta0~dnorm(0.0,1.0E-4)
  for(j in 1:p){
    beta[j]~dnorm(0.0,1.0E-4)
  }
  tau~dgamma(0.001,0.001)
  sigma<-1/sqrt(tau)
  # selecção de covariáveis
  for (i in 1:N){
    resid[i]<-Y[i]-mu[i]
    sresid[i]<-resid[i]*sqrt(tau)
    ver[i]<-sqrt(tau/(2*PI))*exp(-0.5*pow(sresid[i],2))
    log.ver[i]<-log(ver[i])
    inv.ver[i]<-1/ver[i]
  }
  num.par<-p+2
  BIC.est<--2*sum(log.ver[])+num.par*log(N)
  soma.log.ver<--2*sum(log.ver[])
}

```

As observações da variável resposta e das covariáveis estão no ficheiro `bodyfat.dat.odc` e os restantes valores de *input* são dados pela instrução `list(N=251, p=8, PI=3.141593, in.mod=c(1,2,4,6,7,8,12,13))`. O ajustamento de modelos de regressão linear normal com 8 covariáveis, quaisquer que elas sejam, pode ser efectuado com este mesmo código, bastando para isso alterar o vector `in.mod` que indica quais as covariáveis incluídas no modelo. A inicialização dos parâmetros é dada pela instrução `list(beta0=10, beta=c(0,0,0,0,0,0,0), tau=0.1)`.

D.1.2 Método FVS

A execução do método FVS é facilitada pela possibilidade de efectuar uma série de comandos através da linguagem de scripts do WinBUGS (versão 1.4), em alternativa aos menus e às caixas de diálogo.

A 1ª iteração do método FVS é de fácil execução pois consiste em ajustar o modelo só com a constante.

Na 2ª iteração são ajustados 13 modelos de regressão linear, cada um deles com uma das covariáveis, o que constitui uma tarefa árdua, caso se utilize o sistema de menus e as caixas de diálogo. Para contornar este problema, utiliza-se um ficheiro de script em que são repetidos 13 vezes os comandos

```
display('log')
check('forw_lvar_model.odc')
data('forw_lvar_data.odc')
data('in_mod1.odc')
data('bodyfat_dat.odc')
compile(1)
inits(1, 'forw_lvar_inits.odc')
beg(1001)
set(BIC.est)
set(deviance)
set(soma.log.ver)
set(inv.ver)
dic.set()
update(11000)
stats(BIC.est)
stats(deviance)
stats(soma.log.ver)
stats(inv.ver)
dic.stats()
```

sendo, apenas, substituído o ficheiro `in_mod1.odc`, que tem a instrução `list(in.mod=c(1))`, pelos ficheiros `in_mod2.odc`, ..., `in_mod13.odc`, que têm as instruções `list(in.mod=c(2))`, ..., `list(in.mod=c(13))`, respectivamente.

O ficheiro `forw_lvar_model.odc` contém o código do WinBUGS

```
model
{
  # leitura dos dados
  for (i in 1:N){
    x[i,1]<-d[i,in.mod[1]]
  }
}
```

```

# standardização dos x's e dos coeficientes
for (j in 1:p){
  b[j]<-beta[j]/sd(x[,j])
  for (i in 1:N){
    z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
  }
}
b0<-beta0-b[1]*mean(x[,1])
# modelo
for (i in 1:N ){
  mu[i]<-beta0+beta[1]*z[i,1]
  Y[i]~dnorm(mu[i],tau)
}
# priors
beta0~dnorm(0.0,1.0E-4)
for (j in 1:p){
  beta[j]~dnorm(0.0,1.0E-4)
}
tau~dgamma(0.001,0.001)
sigma<-1/sqrt(tau)
# selecção de covariáveis
for (i in 1:N){
  resid[i]<-Y[i]-mu[i]
  sresid[i]<-resid[i]*sqrt(tau)
  ver[i]<-sqrt(tau/(2*PI))*exp(-0.5*pow(sresid[i],2))
  log.ver[i]<-log(ver[i])
  inv.ver[i]<-1/ver[i]
}
num.par<-p+2
BIC.est<--2*sum(log.ver[]) + num.par*log(N)
soma.log.ver<--2*sum(log.ver[])
}

```

que permite ajustar modelos de regressão linear normal com uma única covariável. No ficheiro `forw_1var_data.odc` está a instrução `list(N=251, p=1, PI=3.141593)`. As observações das variáveis estão no ficheiro `bodyfat_dat.odc`. No ficheiro `forw_1var_inits.odc` está a instrução de inicialização dos parâmetros, `list(beta0=10, beta=c(0), tau=0.1)`. No ficheiro `in_mod1.odc` encontra-se a instrução que indica que x_1 é a covariável que está ser incluída no modelo. Os comandos de script apresentados possibilitam o ajustamento do modelo de regressão linear normal com a covariável x_1 . De modo a contemplar todos os modelos de regressão linear normal com uma covariável, é necessário repetir os comandos de script apresentados mais 12 vezes, considerando os ficheiros `in_mod2.odc`, ..., `in_mod13.odc`.

Na 3ª iteração são ajustados 12 modelos de regressão linear normal cada um deles com duas covariáveis, x_6 e x_k , para $k = 1, \dots, 13$ e $k \neq 6$. De modo a facilitar a tarefa de ajustar todos estes modelos, é efectuado um ficheiro de script em que são repetidos 12 vezes os comandos

```

display('log')
check('forw_2var_model.odc')
data('forw_2var_data.odc')
data('in_mod6_1.odc')
data('bodyfat_dat.odc')
compile(1)
inits(1,'forw_2var_inits.odc')
beg(1001)
set(BIC.est)
set(deviance)
set(soma.log.ver)
set(inv.ver)
dic.set()

```

```

update(11000)
stats(BIC.est)
stats(deviance)
stats(soma.log.ver)
stats(inv.ver)
dic.stats()

```

sendo, apenas, substituído o ficheiro `in_mod6_1.odc`, que tem a instrução `list(in.mod=c(6,1))`, pelos ficheiros `in_mod6_2.odc`, ..., `in_mod6_13.odc`, que têm as instruções `list(in.mod=c(6,2))`, ..., `list(in.mod=c(6,13))`, respectivamente.

O ficheiro `forw_2var_model.odc` contém o código do WinBUGS

```

model
{
  # leitura dos dados
  for (i in 1:N){
    x[i,1]<-d[i,in.mod[1]]
    x[i,2]<-d[i,in.mod[2]]
  }
  # standardização dos x's e dos coeficientes
  for (j in 1:p){
    b[j]<-beta[j]/sd(x[,j])
    for (i in 1:N){
      z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
    }
  }
  b0<-beta0-b[1]*mean(x[,1])-b[2]*mean(x[,2])
  # modelo
  for (i in 1:N){
    mu[i]<-beta0+beta[1]*z[i,1]+beta[2]*z[i,2]
    Y[i]~dnorm(mu[i],tau)
  }
  # priors
  beta0~dnorm(0.0,1.0E-4)
  for (j in 1:p){
    beta[j]~dnorm(0.0,1.0E-4)
  }
  tau~dgamma(0.001,0.001)
  sigma<-1/sqrt(tau)
  # selecção de covariáveis
  for (i in 1:N){
    resid[i]<-Y[i]-mu[i]
    sresid[i]<-resid[i]*sqrt(tau)
    ver[i]<-sqrt(tau/(2*PI))*exp(-0.5*pow(sresid[i],2))
    log.ver[i]<-log(ver[i])
    inv.ver[i]<-1/ver[i]
  }
  num.par<-p+2
  BIC.est<--2*sum(log.ver[])+num.par*log(N)
  soma.log.ver<--2*sum(log.ver[])
}

```

que permite ajustar modelos de regressão linear normal com duas covariáveis. No ficheiro `forw_2var_data.odc` está a instrução `list(N=251, p=2, PI=3.141593)`. A matriz de dados está no ficheiro `bodyfat.dat.odc`. No ficheiro `forw_2var_inits.odc` está a instrução de inicialização dos parâmetros do modelo, `list(beta0=10, beta=c(0,0), tau=0.1)`. O ficheiro `in_mod6_1.odc` tem a instrução `list(in.mod=c(6,1))`, que indica que as covariáveis incluídas no modelo são x_6 e x_1 . Para averiguar qual a covariável que, conjuntamente com x_6 , deve ser incluída no modelo, é necessário repetir os comandos script apresentados mais 11 vezes, com os ficheiros `in_mod6_2.odc`, ..., `in_mod6_13.odc`, respectivamente.

As restantes iterações do método FVS são efectuadas de modo análogo à 2ª e 3ª iterações.

D.1.3 Método BVE

A implementação do método BVE é também efectuada com recurso aos ficheiros de script do WinBUGS.

A 1ª iteração do método BVE consiste em ajustar um modelo de regressão linear normal com 13 covariáveis, tal como foi efectuado para a 1ª iteração do método SIMD, cujo código do WinBUGS é apresentado na secção D.1.1.

Na 2ª iteração são ajustados 13 modelos de regressão linear, cada um deles resultante da exclusão de uma das covariáveis. Recorre-se a um ficheiro de script, no qual são repetidos 13 vezes os comandos

```
display('log')
check('back_12var_model.odc')
data('back_12var_data.odc')
data('out_mod1.odc')
data('bodyfat_dat.odc')
compile(1)
inits(1,'back_12var_inits.odc')
beg(1001)
set(BIC.est)
set(deviance)
set(soma.log.ver)
set(inv.ver)
dic.set()
update(11000)
stats(BIC.est)
stats(deviance)
stats(soma.log.ver)
stats(inv.ver)
dic.stats()
```

sendo, apenas, substituído o ficheiro out_mod1.odc, que tem a instrução list(in.mod=c(2,3,4,5,6,7,8,9,10,11,12,13)), pelos ficheiros out_mod2.odc, ..., out_mod13.odc que têm, respectivamente, as instruções list(in.mod=c(1,3,4,5,6,7,8,9,10,11,12,13)), ..., list(in.mod=c(1,2,3,4,5,6,7,8,9,10,11,12)).

No ficheiro back_12var_model.odc está o código do WinBugs

```
model
{
  # leitura dos dados
  for (i in 1:N){
    x[i,1]<-d[i,in.mod[1]]
    x[i,2]<-d[i,in.mod[2]]
    x[i,3]<-d[i,in.mod[3]]
    x[i,4]<-d[i,in.mod[4]]
    x[i,5]<-d[i,in.mod[5]]
    x[i,6]<-d[i,in.mod[6]]
    x[i,7]<-d[i,in.mod[7]]
    x[i,8]<-d[i,in.mod[8]]
    x[i,9]<-d[i,in.mod[9]]
    x[i,10]<-d[i,in.mod[10]]
    x[i,11]<-d[i,in.mod[11]]
    x[i,12]<-d[i,in.mod[12]]
  }
}
```

```

# standardização dos x's e dos coeficientes
for (j in 1:p){
  b[j]<-beta[j]/sd(x[,j])
  for (i in 1:N){
    z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
  }
}
b0<-beta0-b[1]*mean(x[,1])-b[2]*mean(x[,2])-b[3]*mean(x[,3])-b[4]*mean(x[,4])
-b[5]*mean(x[,5])-b[6]*mean(x[,6])-b[7]*mean(x[,7])-b[8]*mean(x[,8])
-b[9]*mean(x[,9])-b[10]*mean(x[,10])-b[11]*mean(x[,11])-b[12]*mean(x[,12])
# modelo
for (i in 1:N){
  mu[i]<-beta0+beta[1]*z[i,1]+beta[2]*z[i,2]+beta[3]*z[i,3]+beta[4]*z[i,4]
+beta[5]*z[i,5]+beta[6]*z[i,6]+beta[7]*z[i,7]+beta[8]*z[i,8]
+beta[9]*z[i,9]+beta[10]*z[i,10]+beta[11]*z[i,11]+beta[12]*z[i,12]
  Y[i]~dnorm(mu[i],tau)
}
# priors
beta0~dnorm(0.0,1.0E-4)
for (j in 1:p){
  beta[j]~dnorm(0.0,1.0E-4)
}
tau~dgamma(0.001,0.001)
sigma<-1/sqrt(tau)
# seleção de covariáveis
for (i in 1:N){
  resid[i]<-Y[i]-mu[i]
  sresid[i]<-resid[i]*sqrt(tau)
  ver[i]<-sqrt(tau/(2*PI))*exp(-0.5*pow(sresid[i],2))
  log.ver[i]<-log(ver[i])
  inv.ver[i]<-1/ver[i]
}
num.par<-p+2
BIC.est<--2*sum(log.ver[])+num.par*log(N)
soma.log.ver<--2*sum(log.ver[])
}

```

que permite ajustar modelos de regressão linear normal com 12 covariáveis. No ficheiro `back_12var_data.odc` está a instrução `list(N=251, p=12, PI=3.141593)` e no ficheiro `bodyfat_dat.odc` está a matriz de dados. A instrução de inicialização dos parâmetros do modelo, `list(beta0=10, beta=c(0,0,0,0,0,0,0,0,0,0,0,0), tau=0.1)`, encontra-se no ficheiro `back_12var_inits.odc`. No ficheiro `out_mod1.odc` está a instrução que indica que apenas a covariável x_1 é excluída do modelo. Para ajustar todos os possíveis modelos de regressão linear normal com 12 covariáveis, correspondentes à exclusão do modelo de cada uma das covariáveis x_2, \dots, x_{13} , é necessário repetir os comandos de script apresentados, com os ficheiros `out_mod2.odc, \dots, out_mod13.odc`, respectivamente.

Na 3ª iteração são ajustados 12 modelos de regressão linear normal, cada um deles resultante da exclusão de duas covariáveis, x_9 e x_k , para $k = 1, \dots, 13$ e $k \neq 9$. Como já foi referido, para facilitar a tarefa de ajustar todos estes modelos é efectuado um ficheiro de script em que são repetidos 12 vezes os comandos

```

display('log')
check('back_11var_model.odc')
data('back_11var_data.odc')
data('out_mod9_1.odc')
data('bodyfat_dat.odc')
compile(1)
inits(1,'back_11var_inits.odc')
beg(1001)

```



```

set(BIC.est)
set(deviance)
set(soma.log.ver)
set(inv.ver)
dic.set()
update(11000)
stats(BIC.est)
stats(deviance)
stats(soma.log.ver)
stats(inv.ver)
dic.stats()

```

sendo, apenas, substituído o ficheiro `out_mod9_1.odc`, que tem a instrução `list(in.mod=c(2,3,4,5,6,7,8,10,11,12,13))`, pelos ficheiros `out_mod9_2.odc`, ..., `out_mod9_13.odc`, que têm, respectivamente, as instruções `list(in.mod=c(1,3,4,5,6,7,8,10,11,12,13))`, ..., `list(in.mod=c(1,2,3,4,5,6,7,8,10,11,12))`.

No ficheiro `back_11var_model.odc` está o código do WinBUGS

```

model
{
  # leitura dos dados
  for (i in 1:N){
    x[i,1]<-d[i,in.mod[1]]
    x[i,2]<-d[i,in.mod[2]]
    x[i,3]<-d[i,in.mod[3]]
    x[i,4]<-d[i,in.mod[4]]
    x[i,5]<-d[i,in.mod[5]]
    x[i,6]<-d[i,in.mod[6]]
    x[i,7]<-d[i,in.mod[7]]
    x[i,8]<-d[i,in.mod[8]]
    x[i,9]<-d[i,in.mod[9]]
    x[i,10]<-d[i,in.mod[10]]
    x[i,11]<-d[i,in.mod[11]]
  }
  # standardização dos x's e dos coeficientes
  for (j in 1:p){
    b[j]<-beta[j]/sd(x[,j])
    for (i in 1:N){
      z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
    }
  }
  b0<-beta0-b[1]*mean(x[,1])-b[2]*mean(x[,2])-b[3]*mean(x[,3])-b[4]*mean(x[,4])
    -b[5]*mean(x[,5])-b[6]*mean(x[,6])-b[7]*mean(x[,7])-b[8]*mean(x[,8])
    -b[9]*mean(x[,9])-b[10]*mean(x[,10])-b[11]*mean(x[,11])
  # modelo
  for (i in 1:N){
    mu[i]<-beta0+beta[1]*z[i,1]+beta[2]*z[i,2]+beta[3]*z[i,3]+beta[4]*z[i,4]
      +beta[5]*z[i,5]+beta[6]*z[i,6]+beta[7]*z[i,7]+beta[8]*z[i,8]
      +beta[9]*z[i,9]+beta[10]*z[i,10]+beta[11]*z[i,11]
    Y[i]~dnorm(mu[i],tau)
  }
  # priors
  beta0~dnorm(0.0,1.0E-4)
  for (j in 1:p){
    beta[j]~dnorm(0.0,1.0E-4)
  }
  tau~dgamma(0.001,0.001)
  sigma<-1/sqrt(tau)
  # selecção de covariáveis
  for (i in 1:N){
    resid[i]<-Y[i]-mu[i]
    sresid[i]<-resid[i]*sqrt(tau)
  }
}

```

```

    ver[i]<-sqrt(tau/(2*PI))*exp(-0.5*pow(sresid[i],2))
    log.ver[i]<-log(ver[i])
    inv.ver[i]<-1/ver[i]
  }
  num.par<-p+2
  BIC.est<--2*sum(log.ver[]) + num.par*log(N)
  soma.log.ver<--2*sum(log.ver[])
}

```

que permite ajustar modelos de regressão linear normal com 11 covariáveis. No ficheiro `back_11var.data.odc` está a instrução `list(N=251, p=11, PI=3.141593)`. A matriz de dados está no ficheiro `bodyfat.dat.odc`. No ficheiro `back_11var.inits.odc` está a instrução de inicialização dos parâmetros do modelo, `list(beta0=10, beta=c(0,0,0,0,0,0,0,0,0,0), tau=0.1)`. No ficheiro `out_mod9.1.odc` indica-se que o modelo de 11 covariáveis ajustado é o que resulta da exclusão de x_9 e x_1 . Para averiguar qual a covariável que, conjuntamente com x_9 , deve ser excluída do modelo, é necessário repetir esta série de comandos de script com os ficheiros `out_mod9.2.odc`, ..., `out_mod9.13.odc`.

As restantes iterações do método BVE são efectuadas de modo análogo à 2ª e 3ª iterações.

D.1.4 Método GVS

O código do WinBUGS para a aplicação do método GVS ao conjunto das 13 covariáveis é

```

model massagorda_GVS_13var;
const
  pt=13,           # número total de covariáveis
  p=13,           # número de covariáveis consideradas
  N=251,          # número de observações
  models=8192;    # número de modelos
var
  d[N,pt],        # todas as covariáveis
  x[N,p],         # covariáveis consideradas
  z[N,p],         # covariáveis standardizadas
  Y[N],           # variável resposta
  mu[N],          # valor médio
  beta0, beta[p], # constante e coeficientes standardizados
  b0, b[p],       # constante e coeficientes não standardizados
  tau, sigma,     # precisão e desvio padrão
  g[p],           # vector dos indicadores das covariáveis
  mdl,           # identificador do modelo
  pmdl[models],  # vector dos indicadores dos modelos
  bprior[p],     # média a priori dos coeficientes
  tprior[p];     # precisão a priori dos coeficientes
{
  # leitura dos dados
  for (i in 1:N){
    for (j in 1:p){
      x[i,j]<-d[i,j]
    }
  }
  # standardização dos x's e dos coeficientes
  for (j in 1:p){
    b[j]<-beta[j]/sd(x[,j])
    for (i in 1:N){
      z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
    }
  }
}

```

```

b0<-beta0-b[1]*mean(x[,1])-b[2]*mean(x[,2])-b[3]*mean(x[,3])-b[4]*mean(x[,4])
-b[5]*mean(x[,5])-b[6]*mean(x[,6])-b[7]*mean(x[,7])-b[8]*mean(x[,8])
-b[9]*mean(x[,9])-b[10]*mean(x[,10])-b[11]*mean(x[,11])
-b[12]*mean(x[,12])-b[13]*mean(x[,13])

# modelo
for (i in 1:N){
  Y[i]~dnorm(mu[i],tau)
  mu[i]<-beta0+g[1]*beta[1]*z[i,1]+g[2]*beta[2]*z[i,2]+g[3]*beta[3]*z[i,3]
  +g[4]*beta[4]*z[i,4]+g[5]*beta[5]*z[i,5]+g[6]*beta[6]*z[i,6]
  +g[7]*beta[7]*z[i,7]+g[8]*beta[8]*z[i,8]+g[9]*beta[9]*z[i,9]
  +g[10]*beta[10]*z[i,10]+g[11]*beta[11]*z[i,11]
  +g[12]*beta[12]*z[i,12]+g[13]*beta[13]*z[i,13]
}

# identificador do modelo
mdl<-1+g[1]*1+g[2]*2+g[3]*4+g[4]*8+g[5]*16+g[6]*32+g[7]*64+g[8]*128+g[9]*256
+g[10]*512+g[11]*1024+g[12]*2048+g[13]*4096

# vector dos indicadores dos modelos
for (j in 1:models){
  pmdl[j]<-equals(mdl,j)
}

# priors
beta0~dnorm(0, 0.00001)
for (j in 1:p){
  bprior[j]<-0.0
  tprior[j]<-pow(100,1-g[j])*0.001
  beta[j]~dnorm(bprior[j],tprior[j])
}

tau~dgamma(0.001,0.001)
sigma<-sqrt(1/tau)
# priors para os indicadores das covariáveis
for (j in 1:p){
  g[j]~dbern(0.5)
}
}

```

O código do WinBUGS para a aplicação do método GVS ao subconjunto de 8 covariáveis é

```

model massagorda_GVS_8var;
const
  pt=13,          # número total de covariáveis
  p=8,           # número de covariáveis consideradas
  N=251,         # número de observações
  models=256;    # número de modelos
var
  d[N,pt],       # todas as covariáveis
  x[N,p],        # covariáveis consideradas
  z[N,p],        # covariáveis standardizadas
  Y[N],          # variável resposta
  mu[N],         # valor médio
  beta0, beta[p], # constante e coeficientes standardizados
  b0, b[p],      # constante e coeficientes não standardizados
  tau, sigma,    # precisão e desvio padrão
  g[p],          # vector dos indicadores das covariáveis
  mdl,           # identificador do modelo
  pmdl[models], # vector dos indicadores dos modelos
  bprior[p],     # média a priori dos coeficientes
  tprior[p];     # precisão a priori dos coeficientes
{
  # leitura dos dados
  for (i in 1:N){
    x[i,1]<-d[i,2]
    x[i,2]<-d[i,4]
    x[i,3]<-d[i,6]
    x[i,4]<-d[i,7]
    x[i,5]<-d[i,8]
    x[i,6]<-d[i,11]

```

```

        x[i,7]<-d[i,12]
        x[i,8]<-d[i,13]
    }
    # standardização dos x's e dos coeficientes
    for (j in 1:p){
        b[j]<-beta[j]/sd(x[,j])
        for (i in 1:N){
            z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
        }
    }
    b0<-beta0-b[1]*mean(x[,1])-b[2]*mean(x[,2])-b[3]*mean(x[,3])-b[4]*mean(x[,4])
        -b[5]*mean(x[,5])-b[6]*mean(x[,6])-b[7]*mean(x[,7])-b[8]*mean(x[,8])
    # modelo
    for (i in 1:N){
        Y[i]~dnorm(mu[i],tau)
        mu[i]<-beta0+g[1]*beta[1]*z[i,1]+g[2]*beta[2]*z[i,2]+g[3]*beta[3]*z[i,3]
            +g[4]*beta[4]*z[i,4]+g[5]*beta[5]*z[i,5]+g[6]*beta[6]*z[i,6]
            +g[7]*beta[7]*z[i,7]+g[8]*beta[8]*z[i,8]
    }
    # identificador do modelo
    mdl<-1+g[1]*1+g[2]*2+g[3]*4+g[4]*8+g[5]*16+g[6]*32+g[7]*64+g[8]*128
    # vector dos indicadores dos modelos
    for (j in 1:models){
        pmdl[j]<-equals(mdl,j)
    }
    # priors
    beta0~dnorm(0,0.00001)
    for (j in 1:p){
        bprior[j]<-0.0
        tprior[j]<-pow(100,1-g[j])*0.001
        beta[j]~dnorm(bprior[j], tprior[j])
    }
    tau~dgamma(0.001,0.001)
    sigma<-sqrt(1/tau)
    # priors dos indicadores das covariáveis
    for (j in 1:p){
        g[j]~dbern(0.5)
    }
}

```

As probabilidades *a posteriori* de inclusão das covariáveis no modelo podem ser obtidas através da monitorização de g . Se o número de possíveis modelos não for demasiado elevado, as probabilidades *a posteriori* desses modelos podem ser obtidas através da monitorização de $pmdl$.

D.1.5 Método de Carlin e Chib

Todos os pormenores sobre o modo de implementação, no WinBUGS, do método de Carlin e Chib podem ser encontrados em Spiegelhalter *et al.* (1996c). Carlin e Chib (1995) sugerem que se comece por ajustar cada um dos modelos separadamente, de modo a utilizar as distribuições *a posteriori* obtidas como distribuições *pseudoprior* dos outros modelos. Depois, ajusta-se simultaneamente os modelos, monitorizando as respectivas probabilidades *a posteriori* e, se for necessário, ajustando as probabilidades *a priori* dos modelos, de modo a garantir que todos os modelos são frequentemente visitados.

Comparação de m_1 e m_2

Começa por ajustar-se cada um dos modelos separadamente, sendo os resultados apresentados na tabela D.1. Para ajustar m_1 , basta substituir no model CarlinChib_m1m2

apresentado nesta secção, as instruções $pp[1] \leftarrow 0.01$, $pp[2] \leftarrow 0.99$ por $pp[1] \leftarrow 1$, $pp[2] \leftarrow 0$; de modo análogo, para ajustar m_2 , deve considerar-se $pp[1] \leftarrow 0$, $pp[2] \leftarrow 1$.

Tabela D.1: *Outputs* do WinBUGS com as estimativas dos parâmetros dos modelos de regressão linear m_1 (em cima) e m_2 (em baixo), para o exemplo da massa gorda.

node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha0	18.88	0.2511	0.002533	18.4	18.88	19.38	1001	10000
alpha[1]	0.7172	0.3599	0.003541	0.023	0.7145	1.431	1001	10000
alpha[2]	-2.406	1.086	0.01093	-4.545	-2.403	-0.2606	1001	10000
alpha[3]	-1.032	0.5116	0.005182	-2.035	-1.03	-0.02773	1001	10000
alpha[4]	9.494	0.7167	0.007713	8.075	9.505	10.91	1001	10000
alpha[5]	-1.445	0.928	0.008239	-3.211	-1.448	0.3777	1001	10000
alpha[6]	1.456	0.6342	0.006224	0.212	1.458	2.689	1001	10000
alpha[7]	0.9603	0.3504	0.003662	0.2722	0.9591	1.646	1001	10000
alpha[8]	-1.276	0.4436	0.00401	-2.128	-1.28	-0.4088	1001	10000
tau	0.06352	0.005754	5.89E-05	0.05273	0.06333	0.07542	1001	10000

node	mean	sd	MC error	2.50%	median	97.50%	start	sample
beta0	18.89	0.252	0.002712	18.39	18.89	19.39	1001	10000
beta[1]	0.6911	0.3636	0.003415	-0.01413	0.6902	1.395	1001	10000
beta[2]	-2.616	1.123	0.01098	-4.825	-2.623	-0.4229	1001	10000
beta[3]	-1.076	0.5092	0.005595	-2.079	-1.076	-0.06673	1001	10000
beta[4]	9.525	0.7244	0.006389	8.108	9.517	10.94	1001	10000
beta[5]	-1.357	0.938	0.009755	-3.218	-1.357	0.4808	1001	10000
beta[6]	1.274	0.6529	0.006218	-0.01335	1.267	2.53	1001	10000
beta[7]	0.4671	0.4825	0.004832	-0.4803	0.4608	1.422	1001	10000
beta[8]	0.8527	0.368	0.003405	0.1419	0.8522	1.57	1001	10000
beta[9]	-1.287	0.4444	0.004512	-2.156	-1.29	-0.4198	1001	10000
tau	0.06341	0.00575	6.12E-05	0.05265	0.06326	0.07524	1001	10000

As médias e os desvios padrão *a posteriori* dos parâmetros de cada um dos modelos são usados para determinar as distribuições *pseudoprior*. Por exemplo, como a média e o desvio padrão de:

- alpha[1] são dados por 0.7172 e 0.3599 (isto é, a precisão é $1/0.3599^2 \cong 7.7$), respectivamente, considera-se $\mu.alpha1[2] \leftarrow 0.7$ e $\tau.alpha1[2] \leftarrow 7.7$;
- tau são dados por 0.06352 e 0.005754, respectivamente, considera-se $r1[2] \leftarrow 122$ e $l1[2] \leftarrow 1919$, pois $122/1919 \cong 0.06352$ e $122/1919^2 \cong 0.005754^2$.

Utiliza-se o método de Carlin e Chib para estimar as probabilidades *a posteriori* dos modelos m_1 e m_2 , não esquecendo que as probabilidades *a priori* destes modelos devem ser convenientemente ajustadas (por isso considera-se $pp[1] \leftarrow 0.01$, $pp[2] \leftarrow 0.99$) de modo a garantir que ambos os modelos são frequentemente visitados.

O código do WinBUGS para comparação de m_1 e m_2 pelo método de Carlin e Chib é

```

model CarlinChib_m1m2
const
  N=251, # número de observações
  p=13, # número total de covariáveis
  M=2, # número de modelos
  p1=8,p2=9; # número de covariáveis em cada modelo
var
  d[N,p] # todas as covariáveis
  xM1[N,p1],xM2[N,p2], # covariáveis consideradas em cada modelo
  zM1[N,p1],zM2[N,p2], # covariáveis standardizadas
  Y[N], # variável resposta
  mu[N,M], # valor médio para cada modelo
  tau[M], # precisão para cada modelo

```

Anexo D. Códigos do WinBUGS

```
alpha0,alpha[p1],beta0,beta[p2], # constantes e coeficientes standardizados
a0,a[p1],b0,b[p2], # constantes e coeficientes não standardizados
mu.alpha0[M],...,mu.alpha8[M], # priors e pseudopriors dos parâmetros
tau.alpha0[M],...,tau.alpha8[M],
r1[M],l1[M]
mu.beta0[M],...,mu.beta9[M],
tau.beta0[M],...,tau.beta9[M],
r2[M],l2[M]
pp[M], # probabilidades a priori dos modelos
pM1,pM2, # probabilidades a posteriori dos modelos
k; # modelo "verdadeiro"
{
# leitura dos dados
for (i in 1:N){
# modelo 1
xM1[i,1]<-d[i,1]
xM1[i,2]<-d[i,2]
xM1[i,3]<-d[i,4]
xM1[i,4]<-d[i,6]
xM1[i,5]<-d[i,7]
xM1[i,6]<-d[i,8]
xM1[i,7]<-d[i,12]
xM1[i,8]<-d[i,13]
# modelo 2
xM2[i,1]<-d[i,1]
xM2[i,2]<-d[i,2]
xM2[i,3]<-d[i,4]
xM2[i,4]<-d[i,6]
xM2[i,5]<-d[i,7]
xM2[i,6]<-d[i,8]
xM2[i,7]<-d[i,11]
xM2[i,8]<-d[i,12]
xM2[i,9]<-d[i,13]
}
# standardização dos x's e dos coeficientes
for (j in 1:p1){
a[j]<-alpha[j]/sd(xM1[,j])
for (i in 1:N){
zM1[i,j]<-(xM1[i,j]-mean(xM1[,j]))/sd(xM1[,j])
}
}
a0<-alpha0-a[1]*mean(xM1[,1])-a[2]*mean(xM1[,2])-a[3]*mean(xM1[,3])-a[4]*mean(xM1[,4])
-a[5]*mean(xM1[,5])-a[6]*mean(xM1[,6])-a[7]*mean(xM1[,7])-a[8]*mean(xM1[,8])
for (j in 1:p2){
b[j]<-beta[j]/sd(xM2[,j])
for (i in 1:N){
zM2[i,j]<-(xM2[i,j]-mean(xM2[,j]))/sd(xM2[,j])
}
}
}
b0<-beta0-b[1]*mean(xM2[,1])-b[2]*mean(xM2[,2])-b[3]*mean(xM2[,3])-b[4]*mean(xM2[,4])
-b[5]*mean(xM2[,5])-b[6]*mean(xM2[,6])-b[7]*mean(xM2[,7])-b[8]*mean(xM2[,8])
-b[9]*mean(xM2[,9])
# nodos do modelo
k~dcat(pp[])
pp[1]<-0.01
pp[2]<-0.99
pM1<-equals(k,1)
pM2<-equals(k,2)
# estrutura do modelo
for (i in 1:N){
Y[i]~dnorm(mu[i,k],tau[k])
mu[i,1]<-alpha0+alpha[1]*zM1[i,1]+alpha[2]*zM1[i,2]+alpha[3]*zM1[i,3]
+alpha[4]*zM1[i,4]+alpha[5]*zM1[i,5]+alpha[6]*zM1[i,6]
+alpha[7]*zM1[i,7]+alpha[8]*zM1[i,8]
mu[i,2]<-beta0+beta[1]*zM2[i,1]+beta[2]*zM2[i,2]+beta[3]*zM2[i,3]+beta[4]*zM2[i,4]
+beta[5]*zM2[i,5]+beta[6]*zM2[i,6]+beta[7]*zM2[i,7]+beta[8]*zM2[i,8]
+beta[9]*zM2[i,9]
}
}
```

```

# modelo 1
alpha0~dnorm(mu.alpha0[k],tau.alpha0[k])
alpha[1]~dnorm(mu.alpha1[k],tau.alpha1[k])
alpha[2]~dnorm(mu.alpha2[k],tau.alpha2[k])
alpha[3]~dnorm(mu.alpha3[k],tau.alpha3[k])
alpha[4]~dnorm(mu.alpha4[k],tau.alpha4[k])
alpha[5]~dnorm(mu.alpha5[k],tau.alpha5[k])
alpha[6]~dnorm(mu.alpha6[k],tau.alpha6[k])
alpha[7]~dnorm(mu.alpha7[k],tau.alpha7[k])
alpha[8]~dnorm(mu.alpha8[k],tau.alpha8[k])
tau[1]~dgamma(r1[k],l1[k])
sigma[1]<-1/sqrt(tau[1])
# priors do modelo 1
mu.alpha0[1]<-0
tau.alpha0[1]<-1.0E-6
mu.alpha1[1]<-0
tau.alpha1[1]<-1.0E-4
mu.alpha2[1]<-0
tau.alpha2[1]<-1.0E-4
mu.alpha3[1]<-0
tau.alpha3[1]<-1.0E-4
mu.alpha4[1]<-0
tau.alpha4[1]<-1.0E-4
mu.alpha5[1]<-0
tau.alpha5[1]<-1.0E-4
mu.alpha6[1]<-0
tau.alpha6[1]<-1.0E-4
mu.alpha7[1]<-0
tau.alpha7[1]<-1.0E-4
mu.alpha8[1]<-0
tau.alpha8[1]<-1.0E-4
r1[1]<-0.0001
l1[1]<-0.0001
# pseudopriors do modelo 1
mu.beta0[1]<-18.9
tau.beta0[1]<-15.7
mu.beta1[1]<-0.7
tau.beta1[1]<-7.6
mu.beta2[1]<--2.6
tau.beta2[1]<-0.8
mu.beta3[1]<--1.1
tau.beta3[1]<-3.9
mu.beta4[1]<-9.5
tau.beta4[1]<-1.9
mu.beta5[1]<--1.4
tau.beta5[1]<-1.1
mu.beta6[1]<-1.3
tau.beta6[1]<-2.3
mu.beta7[1]<-0.5
tau.beta7[1]<-4.3
mu.beta8[1]<-0.9
tau.beta8[1]<-7.4
mu.beta9[1]<--1.3
tau.beta9[1]<-5.1
r2[1]<-122
l2[1]<-1931
# modelo 2
beta0~dnorm(mu.beta0[k],tau.beta0[k])
beta[1]~dnorm(mu.beta1[k],tau.beta1[k])
beta[2]~dnorm(mu.beta2[k],tau.beta2[k])
beta[3]~dnorm(mu.beta3[k],tau.beta3[k])
beta[4]~dnorm(mu.beta4[k],tau.beta4[k])
beta[5]~dnorm(mu.beta5[k],tau.beta5[k])
beta[6]~dnorm(mu.beta6[k],tau.beta6[k])
beta[7]~dnorm(mu.beta7[k],tau.beta7[k])
beta[8]~dnorm(mu.beta8[k],tau.beta8[k])
beta[9]~dnorm(mu.beta9[k],tau.beta9[k])
tau[2]~dgamma(r2[k],l2[k])
sigma[2]<-1/sqrt(tau[2])

```

```

# priors do modelo 2
mu.beta0[2]<-0
tau.beta0[2]<-1.0E-6
mu.beta1[2]<-0
tau.beta1[2]<-1.0E-4
mu.beta2[2]<-0
tau.beta2[2]<-1.0E-4
mu.beta3[2]<-0
tau.beta3[2]<-1.0E-4
mu.beta4[2]<-0
tau.beta4[2]<-1.0E-4
mu.beta5[2]<-0
tau.beta5[2]<-1.0E-4
mu.beta6[2]<-0
tau.beta6[2]<-1.0E-4
mu.beta7[2]<-0
tau.beta7[2]<-1.0E-4
mu.beta8[2]<-0
tau.beta8[2]<-1.0E-4
mu.beta9[2]<-0
tau.beta9[2]<-1.0E-4
r2[2]<-0.0001
l2[2]<-0.0001
# pseudopriors do modelo 2
mu.alpha0[2]<-18.9
tau.alpha0[2]<-15.9
mu.alpha1[2]<-0.7
tau.alpha1[2]<-7.7
mu.alpha2[2]<--2.4
tau.alpha2[2]<-0.8
mu.alpha3[2]<--1.0
tau.alpha3[2]<-3.8
mu.alpha4[2]<-9.5
tau.alpha4[2]<-1.9
mu.alpha5[2]<--1.4
tau.alpha5[2]<-1.2
mu.alpha6[2]<-1.5
tau.alpha6[2]<-2.5
mu.alpha7[2]<-1.0
tau.alpha7[2]<-8.1
mu.alpha8[2]<--1.3
tau.alpha8[2]<-5.1
r1[2]<-122
l1[2]<-1919
}

```

As observações das variáveis estão no ficheiro `bodyfat.dat.odc`, onde Y representa a variável resposta e $d[i,j]$ as diferentes covariáveis. Os restantes valores de *input* são dados pela instrução `list(N=251, p1=8, p2=9)` e a inicialização dos parâmetros do modelo é dada pela instrução `list(k=1, tau=c(1,1), alpha0=0, alpha=c(0,0,0,0,0,0,0), beta0=0, beta=c(0,0,0,0,0,0,0,0))`.

Comparação de m_3 , m_4 e m_5

Começa por ajustar-se cada um dos modelos separadamente; os resultados são apresentados na tabela D.2. Para ajustar m_3 , m_4 e m_5 basta substituir no model `CarlinChib_m3m4m5`, apresentado nesta secção, as instruções `pp[1] <- 0.90`, `pp[2] <- 0.01`, `pp[3] <- 0.09` por (i) `pp[1] <- 1`, `pp[2] <- 0`, `pp[3] <- 0`, (ii) `pp[1] <- 0`, `pp[2] <- 1`, `pp[3] <- 0` e (iii) `pp[1] <- 0`, `pp[2] <- 0`, `pp[3] <- 1`, respectivamente.

Como já foi referido, as médias e os desvios padrão *a posteriori* dos parâmetros de cada um dos modelos são usados para determinar as distribuições *pseudoprior*.

Tabela D.2: *Outputs* do WinBUGS com as estimativas dos parâmetros dos modelos de regressão linear m_3 (em cima), m_4 (ao meio) e m_5 (em baixo), para o exemplo da massa gorda.

node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha0	18.89	0.2543	0.00243	18.39	18.89	19.4	1001	10000
alpha[1]	-3.723	0.6796	0.00562	-5.024	-3.721	-2.388	1001	10000
alpha[2]	9.906	0.5601	0.004625	8.799	9.904	10.99	1001	10000
alpha[3]	0.8982	0.3428	0.0032	0.2191	0.8961	1.583	1001	10000
alpha[4]	-1.242	0.3869	0.00435	-2.022	-1.237	-0.4826	1001	10000
tau	0.06168	0.00549	6.38E-05	0.05127	0.06152	0.07279	1001	10000

node	mean	sd	MC error	2.50%	median	97.50%	start	sample
beta0	18.88	0.2609	0.002762	18.37	18.88	19.4	1001	10000
beta[1]	-4.002	0.5628	0.006253	-5.112	-3.994	-2.903	1001	10000
beta[2]	9.829	0.5674	0.00591	8.735	9.825	10.95	1001	10000
tau	0.05885	0.005328	4.80E-05	0.04891	0.05865	0.06974	1001	10000

node	mean	sd	MC error	2.50%	median	97.50%	start	sample
gamma0	18.89	0.2575	0.002525	18.38	18.88	19.39	1001	10000
gamma[1]	-3.154	0.6414	0.005954	-4.42	-3.153	-1.92	1001	10000
gamma[2]	9.708	0.5571	0.005049	8.599	9.702	10.8	1001	10000
gamma[3]	-1.016	0.3785	0.003621	-1.763	-1.017	-0.2602	1001	10000
tau	0.06047	0.005447	5.61E-05	0.05017	0.06026	0.07167	1001	10000

O código do WinBUGS para comparação, pelo método de Carlin e Chib, dos modelos m_3 , m_4 e m_5 é

```

model CarlinChib_m3m4m5
const
  N=251,                # número de observações
  p=13,                # número total de covariáveis
  M=3,                 # número de modelos
  p1=4,p2=2,p3=3;    # número de covariáveis em cada modelo
var
  d[N,p],              # todas as covariáveis
  xM1[N,p1],xM2[N,p2],xM3[N,p3], # covariáveis consideradas em cada modelo
  zM1[N,p1],zM2[N,p2],zM3[N,p3], # covariáveis standardizadas
  Y[N],                # variável resposta
  mu[N,M],             # valor médio de cada modelo
  tau[M],              # precisão de cada modelo

  alpha0,alpha[p1],beta0, # constantes e coeficientes standardizados
  beta[p2],gamma0,gamma[p3],
  a0,a[p1],b0,b[p2],c0,c[p3], # constantes e coeficientes não standardizados
  mu.alpha0[M],...,mu.alpha4[M], # priors e pseudopriors dos parâmetros
  tau.alpha0[M],...,tau.alpha4[M],
  r1[M],l1[M]
  mu.beta0[M],...,mu.beta2[M],
  tau.beta0[M],...,tau.beta2[M],
  r2[M],l2[M]
  mu.gamma0[M],...,mu.gamma3[M],
  tau.gamma0[M],...,tau.gamma3[M],
  r3[M],l3[M]
  pp[M],               # probabilidades a priori dos modelos
  pM1,pM2,pM3,        # probabilidades a posteriori dos modelos
  k;                   # modelo "verdadeiro"
{
  # leitura dos dados
  for (i in 1:N){
    # modelo 3
    xM1[i,1]<-d[i,2]
    xM1[i,2]<-d[i,6]
    xM1[i,3]<-d[i,12]
    xM1[i,4]<-d[i,13]
  }
}

```

Anexo D. Códigos do WinBUGS

```
# modelo 4
xM2[i,1]<-d[i,2]
xM2[i,2]<-d[i,6]
# modelo 5
xM3[i,1]<-d[i,2]
xM3[i,2]<-d[i,6]
xM3[i,3]<-d[i,13]
}
# standardização dos x's e dos coeficientes
for (j in 1:p1){
  a[j]<-alpha[j]/sd(xM1[,j])
  for (i in 1:N){
    zM1[i,j]<-(xM1[i,j]-mean(xM1[,j]))/sd(xM1[,j])
  }
}
a0<-alpha0-a[1]*mean(xM1[,1])-a[2]*mean(xM1[,2])-a[3]*mean(xM1[,3])-a[4]*mean(xM1[,4])
for (j in 1:p2){
  b[j]<-beta[j]/sd(xM2[,j])
  for (i in 1:N){
    zM2[i,j]<-(xM2[i,j]-mean(xM2[,j]))/sd(xM2[,j])
  }
}
b0<-beta0-b[1]*mean(xM2[,1])-b[2]*mean(xM2[,2])
for (j in 1:p3){
  c[j]<-gamma[j]/sd(xM3[,j])
  for (i in 1:N){
    zM3[i,j]<-(xM3[i,j]-mean(xM3[,j]))/sd(xM3[,j])
  }
}
c0<-gamma0-c[1]*mean(xM3[,1])-c[2]*mean(xM3[,2])-c[3]*mean(xM3[,3])
# nodos do modelo
k~dcat(pp[])
pp[1]<-0.90
pp[2]<-0.01
pp[3]<-0.09
pM1<-equals(k,1)
pM2<-equals(k,2)
pM3<-equals(k,3)
# estrutura do modelo
for (i in 1:N){
  Y[i]~dnorm(mu[i,k],tau[k])
  mu[i,1]<-alpha0+alpha[1]*zM1[i,1]+alpha[2]*zM1[i,2]+alpha[3]*zM1[i,3]
  +alpha[4]*zM1[i,4]
  mu[i,2]<-beta0+beta[1]*zM2[i,1]+beta[2]*zM2[i,2]
  mu[i,3]<-gamma0+gamma[1]*zM3[i,1]+gamma[2]*zM3[i,2]+gamma[3]*zM3[i,3]
}
# modelo 3
alpha0~dnorm(mu.alpha0[k],tau.alpha0[k])
alpha[1]~dnorm(mu.alpha1[k],tau.alpha1[k])
alpha[2]~dnorm(mu.alpha2[k],tau.alpha2[k])
alpha[3]~dnorm(mu.alpha3[k],tau.alpha3[k])
alpha[4]~dnorm(mu.alpha4[k],tau.alpha4[k])
tau[1]~dgamma(r1[k],l1[k])
sigma[1]<-1/sqrt(tau[1])
# priors do modelo 3
mu.alpha0[1]<-0
tau.alpha0[1]<-1.0E-6
mu.alpha1[1]<-0
tau.alpha1[1]<-1.0E-4
mu.alpha2[1]<-0
tau.alpha2[1]<-1.0E-4
mu.alpha3[1]<-0
tau.alpha3[1]<-1.0E-4
mu.alpha4[1]<-0
tau.alpha4[1]<-1.0E-4
r1[1]<-0.0001
l1[1]<-0.0001
# pseudopriors do modelo 3
mu.beta0[1]<-18.9
```

```

tau.beta0[1]<-14.7
mu.beta1[1]<--4.0
tau.beta1[1]<-3.2
mu.beta2[1]<-9.8
tau.beta2[1]<-3.1
r2[1]<-122
l2[1]<-2073
mu.gamma0[1]<-18.9
tau.gamma0[1]<-15.1
mu.gamma1[1]<--3.2
tau.gamma1[1]<-2.4
mu.gamma2[1]<-9.7
tau.gamma2[1]<-3.2
mu.gamma3[1]<--1.0
tau.gamma3[1]<-7.0
r3[1]<-123
l3[1]<-2038
# modelo 4
beta0~dnorm(mu.beta0[k],tau.beta0[k])
beta[1]~dnorm(mu.beta1[k],tau.beta1[k])
beta[2]~dnorm(mu.beta2[k],tau.beta2[k])
tau[2]~dgamma(r2[k],l2[k])
sigma[2]<-1/sqrt(tau[2])
# priors do modelo 4
mu.beta0[2]<-0
tau.beta0[2]<-1.0E-6
mu.beta1[2]<-0
tau.beta1[2]<-1.0E-4
mu.beta2[2]<-0
tau.beta2[2]<-1.0E-4
r2[2]<-0.0001
l2[2]<-0.0001
# pseudopriors do modelo 4
mu.alpha0[2]<-18.9
tau.alpha0[2]<-15.5
mu.alpha1[2]<--3.7
tau.alpha1[2]<-2.2
mu.alpha2[2]<-9.9
tau.alpha2[2]<-3.2
mu.alpha3[2]<-0.9
tau.alpha3[2]<-8.5
mu.alpha4[2]<--1.2
tau.alpha4[2]<-6.7
r1[2]<-126
l1[2]<-2046
mu.gamma0[2]<-18.9
tau.gamma0[2]<-15.1
mu.gamma1[2]<--3.2
tau.gamma1[2]<-2.4
mu.gamma2[2]<-9.7
tau.gamma2[2]<-3.2
mu.gamma3[2]<--1.0
tau.gamma3[2]<-7.0
r3[2]<-123
l3[2]<-2038
# modelo 5
gamma0~dnorm(mu.gamma0[k],tau.gamma0[k])
gamma[1]~dnorm(mu.gamma1[k],tau.gamma1[k])
gamma[2]~dnorm(mu.gamma2[k],tau.gamma2[k])
gamma[3]~dnorm(mu.gamma3[k],tau.gamma3[k])
tau[3]~dgamma(r3[k],l3[k])
sigma[3]<-1/sqrt(tau[3])
# priors do modelo 5
mu.gamma0[3]<-0
tau.gamma0[3]<-1.0E-6
mu.gamma1[3]<-0
tau.gamma1[3]<-1.0E-4
mu.gamma2[3]<-0
tau.gamma2[3]<-1.0E-4

```

```
mu.gamma3[3]<-0
tau.gamma3[3]<-1.0E-4
r3[3]<-0.0001
l3[3]<-0.0001
# pseudopriors do modelo 5
mu.alpha0[3]<-18.9
tau.alpha0[3]<-15.5
mu.alpha1[3]<--3.7
tau.alpha1[3]<-2.2
mu.alpha2[3]<-9.9
tau.alpha2[3]<-3.2
mu.alpha3[3]<-0.9
tau.alpha3[3]<-8.5
mu.alpha4[3]<--1.2
tau.alpha4[3]<-6.7
r1[3]<-126
l1[3]<-2046
mu.beta0[3]<-18.9
tau.beta0[3]<-14.7
mu.beta1[3]<--4.0
tau.beta1[3]<-3.2
mu.beta2[3]<-9.8
tau.beta2[3]<-3.1
r2[3]<-122
l2[3]<-2073
}
```

As observações das variáveis estão no ficheiro `bodyfat_dat.odc`, onde Y representa a variável resposta e $d[j]$ as diferentes covariáveis. Os restantes valores de *input* são dados pela instrução `list(N=251, p1=4, p2=2, p3=3)` e a inicialização dos parâmetros é dada pela instrução `list(k=1, tau=c(1,1,1), alpha0=0, alpha=c(0,0,0,0), beta0=0, beta=c(0,0), gamma0=0, gamma=c(0,0,0))`.

Os dois modelos seleccionados em cada uma das aplicações do método de Carlin e Chib são, também, comparados por este mesmo método; a elaboração do respectivo código do WinBUGS não apresenta grande dificuldade, tendo em conta os códigos do WinBUGS das outras comparações.

D.2 Dados da homocisteinemia

D.2.1 Método SIMD

Na 1ª iteração do método SIMD, para os dados da homocisteinemia, ajusta-se um modelo de regressão gama com as 13 covariáveis e calcula-se as medidas de discrepância deste modelo. O código do WinBUGS para efectuar esta iteração do método SIMD é

```
model homocisteinemia_SIMD_13covar;
const
  pt=13,          # número total de covariáveis
  p=13,          # número de covariáveis no modelo
  N=126,         # número de observações
var
  d[N,pt],      # todas as covariáveis
  x[N,p],      # covariáveis consideradas
  z[N,p],      # covariáveis standardizadas
  Y[N],        # variável resposta
  mu[N],       # valor médio
  beta0, beta[p], # constante e coeficientes standardizados
```

```

b0, b[p],          # constante e coeficientes não standardizados
ver[N], log.ver[N], # verosimilhança e log-verosimilhança
inv.ver[N],        # ordenada preditiva condicional
num.par,          # número de parâmetros do modelo
BIC.est,          # BIC estimado
soma.log.ver;     # LLP
{
# leitura dos dados
for (i in 1:N){
  x[i,1]<-d[i,in.mod[1]]
  x[i,2]<-d[i,in.mod[2]]
  x[i,3]<-d[i,in.mod[3]]
  x[i,4]<-d[i,in.mod[4]]
  x[i,5]<-d[i,in.mod[5]]
  x[i,6]<-d[i,in.mod[6]]
  x[i,7]<-d[i,in.mod[7]]
  x[i,8]<-d[i,in.mod[8]]
  x[i,9]<-d[i,in.mod[9]]
  x[i,10]<-d[i,in.mod[10]]
  x[i,11]<-d[i,in.mod[11]]
  x[i,12]<-d[i,in.mod[12]]
  x[i,13]<-d[i,in.mod[13]]
}
# standardização dos x's e dos coeficientes
for (j in 1:p){
  b[j]<-beta[j]/sd(x[,j])
  for (i in 1:N){
    z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
  }
}
b0<-beta0-b[1]*mean(x[,1])-b[2]*mean(x[,2])-b[3]*mean(x[,3])-b[4]*mean(x[,4])
-b[5]*mean(x[,5])-b[6]*mean(x[,6])-b[7]*mean(x[,7])-b[8]*mean(x[,8])
-b[9]*mean(x[,9])-b[10]*mean(x[,10])-b[11]*mean(x[,11])
-b[12]*mean(x[,12])-b[13]*mean(x[,13])
# modelo
for(i in 1:N){
  log(mu[i])<-beta0+beta[1]*z[i,1]+beta[2]*z[i,2]+beta[3]*z[i,3]+beta[4]*z[i,4]
+beta[5]*z[i,5]+beta[6]*z[i,6]+beta[7]*z[i,7]+beta[8]*z[i,8]
+beta[9]*z[i,9]+beta[10]*z[i,10]+beta[11]*z[i,11]
+beta[12]*z[i,12]+beta[13]*z[i,13]
  Y[i]~dgamma(nu,q[i])
  q[i]<-nu/mu[i]
}
# priors
beta0~dnorm(0.0,1.0E-4)
for (j in 1:p){
  beta[j]~dnorm(0.0,1.0E-4)
}
nu~dgamma(0.001,0.001)
# selecção de covariáveis
for(i in 1:N){
  ver[i]<-pow((nu/mu[i]),nu)*(1/exp(loggam(nu)))*pow(Y[i],nu-1)*exp(-nu*Y[i]/mu[i])
  log.ver[i]<-log(ver[i])
  inv.ver[i]<-1/ver[i]
}
num.par<-p+2
BIC.est<--2*sum(log.ver[])+num.par*log(N)
soma.log.ver<--2*sum(log.ver[])
}

```

Os dados do exemplo da homocisteinemia estão no ficheiro HBP_dat.odc, onde Y representa a variável resposta e $d[j]$ as diferentes covariáveis. Os restantes valores de *input* são dados por `list(N=126, p=13, in.mod=c(1,2,3,4,5,6,7,8,9,10,11,12,13))`. A inicialização dos parâmetros do modelo é dada por `list(beta0=10, beta=c(0,0,0,0,0,0,0,0,0,0,0), nu=0.1)`.

As estimativas dos parâmetros b_0, b_1, \dots, b_{13} e ν são obtidas por monitorização de b_0, b e ν . Os valores das medidas de discrepância obtêm-se de modo idêntico ao efectuado para

os dados da massa gorda. Os códigos do WinBUGS para efectuar as restantes iterações do método SIMD são idênticos ao anterior, sendo apenas necessário alterar o valor de p e o vector `in.mod` nos dados de `input`, alterar a instrução de inicialização dos parâmetros e adaptar o modelo, uma vez que em cada iteração é excluída uma covariável.

D.2.2 Método FVS

A 1ª iteração do método FVS é de fácil execução pois consiste em ajustar o modelo só com a constante. As restantes iterações deste método foram efectuadas com recurso aos ficheiros de script. Por exemplo, na 2ª iteração, em que são ajustados 13 modelos de regressão gama, cada um deles com uma das covariáveis, é efectuado um ficheiro de script no qual se repete 13 vezes os comandos

```
display('log')
check('forw_lvar_model.odc')
data('forw_lvar_data.odc')
data('in_mod1.odc')
data('HBP_dat.odc')
compile(1)
inits(1,'forw_lvar_inits.odc')
set(BIC.est)
set(deviance)
set(soma.log.ver)
set(inv.ver)
beg(4001)
update(4001)
dic.set()
thin.samples(5)
thin.updater(5)
update(25000)
stats(BIC.est)
stats(deviance)
stats(soma.log.ver)
stats(inv.ver)
dic.stats()
```

substituindo-se o ficheiro `in_mod1.odc`, que tem a instrução `list(in.mod=c(1))`, pelos ficheiros `in_mod2.odc`, ..., `in_mod13.odc`, que têm as instruções `list(in.mod=c(2))`, ..., `list(in.mod=c(13))`, respectivamente. O ficheiro `forw_lvar_model.odc` contém o código do WinBUGS

```
model
{
  # leitura dos dados
  for (i in 1:N){
    x[i,1]<-d[i,in.mod[1]]
  }
  # standardização dos x's e dos coeficientes
  for (j in 1:p){
    b[j]<-beta[j]/sd(x[,j])
    for (i in 1:N){
      z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
    }
  }
  b0<-beta0-b[1]*mean(x[,1])
  # modelo
  for (i in 1:N){
```

```

    q[i]<-nu/mu[i]
  }
  # priors
  beta0~dnorm(0.0,1.0E-4)
  for (j in 1:p){
    beta[j]~dnorm(0.0,1.0E-4)
  }
  nu~dgamma(0.001,0.001)
  # selecção de covariáveis
  for (i in 1:N){
    ver[i]<-pow((nu/mu[i]),nu)*(1/exp(loggam(nu)))*pow(Y[i],nu-1)*exp(-nu*Y[i]/mu[i])
    log.ver[i]<-log(ver[i])
    inv.ver[i]<-1/ver[i]
  }
  num.par<-p+2
  BIC.est<--2*sum(log.ver[])+num.par*log(N)
  soma.log.ver<--2*sum(log.ver[])
}

```

que permite ajustar modelos de regressão gama com uma covariável. O ficheiro HBP_dat.odc contém a matriz de dados e no ficheiro forw_1var_data.odc está a instrução `list(N=126, p=1)`. No ficheiro forw_1var_inits.odc está a instrução de inicialização dos parâmetros do modelo, `list(beta0=10, beta=c(0), nu=0.1)`. No ficheiro in_mod1.odc está a instrução que indica que x_1 é a covariável que está a ser incluída no modelo. As restantes iterações do método FVS são efectuadas de modo análogo à 2ª iteração.

D.2.3 Método BVE

A 1ª iteração do método BVE é de fácil execução pois consiste em ajustar um modelo de regressão gama com todas as covariáveis, de modo análogo ao efectuado para a 1ª iteração do método SIMD, cujo código do WinBUGS é apresentado na secção D.2.1.

Na 2ª iteração são ajustados 13 modelos de regressão gama cada um deles resultante da exclusão de uma das covariáveis. Recorre-se, novamente, a um ficheiro de script, no qual são repetidos 13 vezes os comandos

```

display('log')
check('back_12var_model.odc')
data('back_12var_data.odc')
data('out_mod1.odc')
data('HBP_dat.odc')
compile(1)
inits(1,'back_12var_inits.odc')
set(BIC.est)
set(deviance)
set(soma.log.ver)
set(inv.ver)
beg(4001)
update(4001)
dic.set()
thin.samples(5)
thin.updater(5)
update(25000)
stats(BIC.est)
stats(deviance)
stats(soma.log.ver)
stats(inv.ver)
dic.stats()

```

sendo, apenas, substituído o ficheiro out_mod1.odc, que tem a instrução `list(in.mod=c(2,3,4,5,6,7,8,9,10,11,12,13))`, pelos ficheiros out_mod2.odc, ..., out_mod13.odc, que têm, respectivamente, as instruções `list(in.mod=c(1,3,4,5,6,7,8,9,10,11,12,13))`, ..., `list(in.mod=c(1,2,3,4,5,6,7,8,9,10,11,12))`. No ficheiro back_12var_model.odc está o código do WinBUGS

```

model
{
  # leitura dos dados
  for (i in 1:N){
    x[i,1]<-d[i,in.mod[1]]
    x[i,2]<-d[i,in.mod[2]]
    x[i,3]<-d[i,in.mod[3]]
    x[i,4]<-d[i,in.mod[4]]
    x[i,5]<-d[i,in.mod[5]]
    x[i,6]<-d[i,in.mod[6]]
    x[i,7]<-d[i,in.mod[7]]
    x[i,8]<-d[i,in.mod[8]]
    x[i,9]<-d[i,in.mod[9]]
    x[i,10]<-d[i,in.mod[10]]
    x[i,11]<-d[i,in.mod[11]]
    x[i,12]<-d[i,in.mod[12]]
  }
  # standardização dos x's e dos coeficientes
  for (j in 1:p){
    b[j]<-beta[j]/sd(x[,j])
    for (i in 1:N){
      z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
    }
  }
  b0<-beta0-b[1]*mean(x[,1])-b[2]*mean(x[,2])-b[3]*mean(x[,3])-b[4]*mean(x[,4])
  -b[5]*mean(x[,5])-b[6]*mean(x[,6])-b[7]*mean(x[,7])-b[8]*mean(x[,8])
  -b[9]*mean(x[,9])-b[10]*mean(x[,10])-b[11]*mean(x[,11])-b[12]*mean(x[,12])
  # modelo
  for (i in 1:N){
    log(mu[i])<-beta0+beta[1]*z[i,1]+beta[2]*z[i,2]+beta[3]*z[i,3]+beta[4]*z[i,4]
    +beta[5]*z[i,5]+beta[6]*z[i,6]+beta[7]*z[i,7]+beta[8]*z[i,8]
    +beta[9]*z[i,9]+beta[10]*z[i,10]+beta[11]*z[i,11]+beta[12]*z[i,12]
    Y[i]~dgamma(nu,q[i])
    q[i]<-nu/mu[i]
  }
  # priors
  beta0~dnorm(0.0,1.0E-4)
  for (j in 1:p){
    beta[j]~dnorm(0.0,1.0E-4)
  }
  nu~dgamma(0.001,0.001)
  # selecção de covariáveis
  for (i in 1:N){
    ver[i]<-pow((nu/mu[i]),nu)*(1/exp(loggam(nu)))*pow(Y[i],nu-1)*exp(-nu*Y[i]/mu[i])
    log.ver[i]<-log(ver[i])
    inv.ver[i]<-1/ver[i]
  }
  num.par<-p+2
  BIC.est<--2*sum(log.ver[])+num.par*log(N)
  soma.log.ver<--2*sum(log.ver[])
}

```

que permite ajustar modelos de regressão gama com 12 covariáveis. Em HBP_dat.odc está a matriz de dados e no ficheiro back_12var_data.odc está a instrução `list(N=126, p=12)`. No ficheiro back_12var_inits.odc está a instrução de inicialização dos parâmetros do modelo, `list(beta0=10, beta=c(0,0,0,0,0,0,0,0,0,0,0,0), nu=0.1)`. O ficheiro out_mod1.odc tem a instrução

`list(in.mod=c(2,3,4,5,6,7,8,9,10,11,12,13))`, que indica que a covariável x_1 não está no modelo. As restantes iterações do método BVE são efectuadas de modo análogo à 2ª iteração.

D.2.4 Método GVS

O código do WinBUGS para a aplicação do método GVS ao conjunto das 13 covariáveis, sem restrição no espaço dos modelos, é

```

model homocisteinemia_GVS_13var_sr;
const
  pt=13,          # número total de covariáveis
  p=13,           # número de covariáveis consideradas
  N=126,          # número de observações
  models=8192;    # número de modelos
var
  d[N,pt],        # todas as covariáveis
  x[N,p],         # covariáveis consideradas
  z[N,p],         # covariáveis standardizadas
  Y[N],           # variável resposta
  mu[N],          # valor médio
  beta0, beta[p], # constante e coeficientes standardizados
  b0, b[p],       # constante e coeficientes não standardizados
  g[p],           # vector dos indicadores das covariáveis
  mdl,            # identificador do modelo
  pmdl[models],  # vector dos indicadores dos modelos
  bprior[p],     # média a priori dos coeficientes
  tprior[p];     # precisão a priori dos coeficientes
{
  # leitura dos dados
  for (i in 1:N){
    for (j in 1:p){
      x[i,j]<-d[i,j]
    }
  }
  # standardização dos x's e dos coeficientes
  for (j in 1:p){
    b[j]<-beta[j]/sd(x[,j])
    for (i in 1:N){
      z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
    }
  }
  b0<-beta0-b[1]*mean(x[,1])-b[2]*mean(x[,2])-b[3]*mean(x[,3])-b[4]*mean(x[,4])
    -b[5]*mean(x[,5])-b[6]*mean(x[,6])-b[7]*mean(x[,7])-b[8]*mean(x[,8])
    -b[9]*mean(x[,9])-b[10]*mean(x[,10])-b[11]*mean(x[,11])
    -b[12]*mean(x[,12])-b[13]*mean(x[,13])
  # modelo
  for (i in 1:N){
    Y[i]~dgamma(nu,q[i])
    q[i]<-nu/mu[i]
    log(mu[i])<-beta0+g[1]*beta[1]*z[i,1]+g[2]*beta[2]*z[i,2]+g[3]*beta[3]*z[i,3]
      +g[4]*beta[4]*z[i,4]+g[5]*beta[5]*z[i,5]+g[6]*beta[6]*z[i,6]
      +g[7]*beta[7]*z[i,7]+g[8]*beta[8]*z[i,8]+g[9]*beta[9]*z[i,9]
      +g[10]*beta[10]*z[i,10]+g[11]*beta[11]*z[i,11]
      +g[12]*beta[12]*z[i,12]+g[13]*beta[13]*z[i,13]
  }
  # identificador do modelo
  mdl<-1+g[1]*1+g[2]*2+g[3]*4+g[4]*8+g[5]*16+g[6]*32+g[7]*64+g[8]*128+g[9]*256
    +g[10]*512+g[11]*1024+g[12]*2048+g[13]*4096
  # vector dos indicadores dos modelos
  for (j in 1:models){
    pmdl[j]<-equals(mdl,j)
  }
}

```

Anexo D. Códigos do WinBUGS

```
# priors
beta0~dnorm(0,0.00001)
for (j in 1:p){
  bprior[j]<-0.0
  tprior[j]<-pow(100,1-g[j])*0.001
  beta[j]~dnorm(bprior[j],tprior[j])
}
nu~dgamma(0.001,0.001)
# priors dos indicadores das covariáveis
for (j in 1:p){
  g[j]~dbern(0.5)
}
}
```

e, com restrição no espaço dos modelos, é

```
model homocisteinemia_GVS_13var_cr;
const
  pt=13,          # número total de covariáveis
  p=13,          # número de covariáveis consideradas
  N=126,         # número de observações
  models=4096;   # número de modelos
var
  d[N,pt],       # todas as covariáveis
  x[N,p],        # covariáveis consideradas
  z[N,p],        # covariáveis standardizadas
  Y[N],          # variável resposta
  mu[N],         # valor médio
  beta0, beta[p], # constante e coeficientes standardizados
  b0, b[p],      # constante e coeficientes não standardizados
  g[p-1],        # vector dos indicadores das covariáveis
  mdl,           # identificador do modelo
  pmdl[models], # vector dos indicadores dos modelos
  bprior[p],     # média a priori dos coeficientes
  tprior[p];     # precisão a priori dos coeficientes
{
  # leitura dos dados
  for (i in 1:N){
    for (j in 1:p){
      x[i,j]<-d[i,j]
    }
  }
  # standardização dos x's e dos coeficientes
  for (j in 1:p){
    b[j]<-beta[j]/sd(x[,j])
    for (i in 1:N){
      z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
    }
  }
  b0<-beta0-b[1]*mean(x[,1])-b[2]*mean(x[,2])-b[3]*mean(x[,3])-b[4]*mean(x[,4])
  -b[5]*mean(x[,5])-b[6]*mean(x[,6])-b[7]*mean(x[,7])-b[8]*mean(x[,8])
  -b[9]*mean(x[,9])-b[10]*mean(x[,10])-b[11]*mean(x[,11])
  -b[12]*mean(x[,12])-b[13]*mean(x[,13])
  # modelo
  for (i in 1 : N) {
    Y[i]~dgamma(nu,q[i])
    q[i]<-nu/mu[i]
    log(mu[i])<-beta0+g[1]*beta[1]*z[i,1]+g[2]*(beta[2]*z[i,2]+beta[3]*z[i,3])
    +g[3]*beta[4]*z[i,4]+g[4]*beta[5]*z[i,5]+g[5]*beta[6]*z[i,6]
    +g[6]*beta[7]*z[i,7]+g[7]*beta[8]*z[i,8]+g[8]*beta[9]*z[i,9]
    +g[9]*beta[10]*z[i,10]+g[10]*beta[11]*z[i,11]
    +g[11]*beta[12]*z[i,12]+g[12]*beta[13]*z[i,13]
  }
  # identificador do modelo
  mdl<-1+g[1]*1+g[2]*2+g[3]*4+g[4]*8+g[5]*16+g[6]*32+g[7]*64+g[8]*128+g[9]*256
  +g[10]*512+g[11]*1024+g[12]*2048
}
```

```

# vector dos indicadores dos modelos
for (j in 1:models){
  pmdl[j]<-equals(mdl,j)
}
# priors
beta0~dnorm(0, 0.00001)
for (j in 1:2){
  bprior[j]<-0.0
  tprior[j]<-pow(100,1-g[j])*0.001
  beta[j]~dnorm(bprior[j],tprior[j])
}
bprior[3]<-0.0
tprior[3]<-pow(100,1-g[2])*0.001
beta[3]~dnorm(bprior[3],tprior[3])
for (j in 4:p){
  bprior[j]<-0.0
  tprior[j]<-pow(100,1-g[j-1])*0.001
  beta[j]~dnorm(bprior[j],tprior[j])
}
nu~dgamma(0.001,0.001)
# priors dos indicadores das covariáveis
for (j in 1:(p-1)){
  g[j]~dbern(0.5)
}
}

```

O código do WinBUGS para a aplicação do método GVS ao subconjunto de 8 covariáveis sem restrição no espaço dos modelos é

```

model homocisteinemia_GVS_8var_sr;
const
  pt=13,           # número total de covariáveis
  p=8,            # número de covariáveis
  N=126,          # número de observações
  models=256;     # número de modelos
var
  d[N,pt],        # todas as covariáveis
  x[N,p],         # covariáveis consideradas
  z[N,p],         # covariáveis standardizadas
  Y[N],           # variável resposta
  mu[N],          # valor médio
  beta0, beta[p], # constante e coeficientes standardizados
  b0, b[p],       # constante e coeficientes não standardizados
  g[p],           # vector dos indicadores das covariáveis
  mdl,            # identificador do modelo
  pmdl[models],  # vector dos indicadores dos modelos
  bprior[p],     # média a priori dos coeficientes
  tprior[p];     # precisão a priori dos coeficientes
{
  # leitura dos dados
  for (i in 1:N){
    x[i,1]<-d[i,1]
    x[i,2]<-d[i,2]
    x[i,3]<-d[i,3]
    x[i,4]<-d[i,4]
    x[i,5]<-d[i,7]
    x[i,6]<-d[i,9]
    x[i,7]<-d[i,12]
    x[i,8]<-d[i,13]
  }
  # standardização dos x's e dos coeficientes
  for (j in 1:p){
    b[j]<-beta[j]/sd(x[,j])
    for (i in 1:N){
      z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
    }
  }
}

```

Anexo D. Códigos do WinBUGS

```
b0<-beta0-b[1]*mean(x[,1])-b[2]*mean(x[,2])-b[3]*mean(x[,3])-b[4]*mean(x[,4])
-b[5]*mean(x[,5])-b[6]*mean(x[,6])-b[7]*mean(x[,7])-b[8]*mean(x[,8])
# modelo
for (i in 1:N){
  Y[i]~dgamma(nu,q[i])
  q[i]<-nu/mu[i]
  log(mu[i])<-beta0+g[1]*beta[1]*z[i,1]+g[2]*beta[2]*z[i,2]+g[3]*beta[3]*z[i,3]
  +g[4]*beta[4]*z[i,4]+g[5]*beta[5]*z[i,5]+g[6]*beta[6]*z[i,6]
  +g[7]*beta[7]*z[i,7]+g[8]*beta[8]*z[i,8]
}
# identificador do modelo
mdl<-1+g[1]*1+g[2]*2+g[3]*4+g[4]*8+g[5]*16+g[6]*32+g[7]*64+g[8]*128
# vector dos indicadores dos modelos
for (j in 1:models){
  pmdl[j]<-equals(mdl,j)
}
# priors
beta0~dnorm(0,0.00001)
for (j in 1:p){
  bprior[j]<-0.0
  tprior[j]<-pow(100,1-g[j])*0.001
  beta[j]~dnorm(bprior[j],tprior[j])
}
nu~dgamma(0.001,0.001)
# priors dos indicadores das covariáveis
for (j in 1:p){
  g[j]~dbern(0.5)
}
}
```

O código do WinBUGS para a aplicação do método GVS a um subconjunto de 9 covariáveis com restrição no espaço dos modelos é

```
model homocisteinemia_GVS_9var_cr;
const
  pt=13,          # número total de covariáveis
  p=9,           # número de covariáveis consideradas
  N=126,         # número de observações
  models=256;    # número de modelos em consideração
var
  d[N,pt],      # todas as covariáveis
  x[N,p],       # covariáveis consideradas
  z[N,p],       # covariáveis standardizadas
  Y[N],         # variável resposta
  mu[N],        # valor médio
  beta0, beta[p], # constante e coeficientes standardizados
  b0, b[p],     # constante e coeficientes não standardizados
  g[p-1],       # vector dos indicadores das covariáveis
  mdl,          # identificador do modelo
  pmdl[models], # vector dos indicadores dos modelos
  bprior[p],    # média a priori dos coeficientes
  tprior[p];    # precisão a priori dos coeficientes
{
  # leitura dos dados
  for (i in 1:N){
    x[i,1]<-d[i,1]
    x[i,2]<-d[i,2]
    x[i,3]<-d[i,3]
    x[i,4]<-d[i,4]
    x[i,5]<-d[i,6]
    x[i,6]<-d[i,7]
    x[i,7]<-d[i,9]
    x[i,8]<-d[i,12]
    x[i,9]<-d[i,13]
  }
  # standardização dos x's e dos coeficientes
  for (j in 1:p){
```

```

    b[j]<-beta[j]/sd(x[,j])
    for (i in 1:N){
      z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
    }
  }
  b0<-beta0-b[1]*mean(x[,1])-b[2]*mean(x[,2])-b[3]*mean(x[,3])-b[4]*mean(x[,4])
    -b[5]*mean(x[,5])-b[6]*mean(x[,6])-b[7]*mean(x[,7])-b[8]*mean(x[,8])
  # modelo
  for (i in 1:N){
    Y[i]~dgamma(nu,q[i])
    q[i]<-nu/mu[i]
    log(mu[i])<-beta0+g[1]*beta[1]*z[i,1]+g[2]*(beta[2]*z[i,2]+beta[3]*z[i,3])
      +g[3]*beta[4]*z[i,4]+g[4]*beta[5]*z[i,5]+g[5]*beta[6]*z[i,6]
      +g[6]*beta[7]*z[i,7]+g[7]*beta[8]*z[i,8]+g[8]*beta[9]*z[i,9]
  }
  # identificador do modelo
  mdl<-1+g[1]*1+g[2]*2+g[3]*4+g[4]*8+g[5]*16+g[6]*32+g[7]*64+g[8]*128
  # vector dos indicadores dos modelos
  for (j in 1:models){
    pmdl[j]<-equals(mdl,j)
  }
  # priors
  beta0~dnorm(0,0.00001)
  for (j in 1:2){
    bprior[j]<-0.0
    tprior[j]<-pow(100,1-g[j])*0.001
    beta[j]~dnorm(bprior[j],tprior[j])
  }
  bprior[3]<-0.0
  tprior[3]<-pow(100,1-g[2])*0.001
  beta[3]~dnorm(bprior[3],tprior[3])
  for (j in 4:p){
    bprior[j]<-0.0
    tprior[j]<-pow(100,1-g[j-1])*0.001
    beta[j]~dnorm(bprior[j],tprior[j])
  }
  nu~dgamma(0.001,0.001)
  # priors dos indicadores das covariáveis
  for (j in 1:(p-1)){
    g[j]~dbern(0.5)
  }
}

```

As probabilidades *a posteriori* de inclusão das covariáveis no modelo podem ser obtidas através da monitorização de `g`. Se o número de possíveis modelos não for demasiado elevado, as suas probabilidades *a posteriori* podem ser obtidas por monitorização de `pmdl`.

D.3 Dados simulados de regressão linear normal — Método GVS

A estimação das probabilidades *a posteriori* dos modelos e das probabilidades *a posteriori* de inclusão das covariáveis no modelo pelo método GVS, no caso 1, pode ser feita com recurso a um ficheiro de script em que são repetidos 16 vezes os comandos

```

display('log')
check('modelo_reg_dgm1.odc')
data('dados_reg_casol.odc')
data('data_reg.odc')
compile(1)
inits(1,'inits_reg.odc')
beg(5001)
set(b0)

```

```
set (b)
set (beta0)
set (beta)
set (tau)
set (sigma)
set (g)
set (pmdl)
thin.samples (5)
thin.updater (5)
update (55000)
stats (b0)
stats (b)
stats (beta0)
stats (beta)
stats (tau)
stats (sigma)
stats (g)
stats (pmdl)
```

sendo, apenas, substituído o ficheiro modelo_reg_dgm1.odc pelos ficheiros modelo_reg_dgm2.odc, modelo_reg_dgm3.odc, . . . , modelo_reg_dgm16.odc. O ficheiro modelo_reg_dgm1.odc tem o código do WinBUGS

```
model
{
  # leitura dos dados
  for (i in 1:N){
    x[i,1]<-d[i,1]
    x[i,2]<-d[i,2]
    x[i,3]<-d[i,3]
    x[i,4]<-d[i,4]
  }
  # obtenção da amostra da variável resposta (d[i,5] é o erro aleatório)
  for (i in 1:N){
    Y[i]<-d[i,1]+d[i,2]+d[i,3]+d[i,4]+d[i,5]
  }
  # standardização dos x's e dos coeficientes
  for (j in 1:p){
    b[j]<-beta[j]/sd(x[,j])
    for (i in 1:N){
      z[i,j]<-(x[i,j]-mean(x[,j]))/sd(x[,j])
    }
  }
  b0<-beta0-b[1]*mean(x[,1])-b[2]*mean(x[,2])-b[3]*mean(x[,3])-b[4]*mean(x[,4])
  # modelo
  for (i in 1:N){
    Y[i]~dnorm(mu[i],tau)
    mu[i]<-beta0+g[1]*beta[1]*z[i,1]+g[2]*beta[2]*z[i,2]+g[3]*beta[3]*z[i,3]
      +g[4]*beta[4]*z[i,4]
  }
  # identificador do modelo
  mdl<-1+g[1]*1+g[2]*2+g[3]*4+g[4]*8
  # vector dos indicadores dos modelos
  for (j in 1:models){
    pmdl[j]<-equals(mdl,j)
  }
  # priors
  beta0~dnorm(0,0.00001)
  for (j in 1:p){
    bprior[j]<-0.0
    tprior[j]<-pow(100,1-g[j])*0.001
    beta[j]~dnorm(bprior[j],tprior[j])
  }
  tau~dgamma(0.001,0.001)
  sigma<-sqrt(1/tau)
}
```

```
# priors para os indicadores das covariáveis
for (j in 1:p){
  g[j]~dbern(0.5)
}
}
```

que permite estimar, através do método GVS, as probabilidades *a posteriori* dos modelos e as probabilidades *a posteriori* de inclusão das covariáveis no modelo, para os dados gerados do modelo 1. Os códigos do WinBUGS dos ficheiros `modelo_reg_dgm2.odc`, `modelo_reg_dgm3.odc`, ..., `modelo_reg_dgm16.odc` diferem do anterior no modo de obtenção da variável resposta que passa a ser $Y[i] \leftarrow d[i, 1] + d[i, 2] + d[i, 3] + d[i, 5]$, $Y[i] \leftarrow d[i, 1] + d[i, 2] + d[i, 4] + d[i, 5]$, ..., $Y[i] \leftarrow d[i, 5]$, respectivamente. No ficheiro `data_reg.odc` está a instrução `list(p=4, N=200, models=16)`. Os valores gerados para as covariáveis e para o erro estão no ficheiro `dados_reg_caso1.odc`, onde $d[1]$, $d[2]$, $d[3]$, $d[4]$ são as diferentes covariáveis e $d[5]$ são as observações do erro aleatório. No ficheiro `inits_reg.odc` está a instrução de inicialização dos parâmetros, `list(beta0=0, beta=c(0,0,0,0), tau=0.1, g=c(1,1,1,1))`.

No caso 2, para implementar o método GVS, basta alterar o modo de obtenção da variável resposta, substituindo a instrução $Y[i] \leftarrow d[i, 1] + d[i, 2] + d[i, 3] + d[i, 4] + d[i, 5]$ do ficheiro `modelo_reg_dgm1.odc` por $Y[i] \leftarrow d[i, 1] + 10 * d[i, 2] + 2 * d[i, 3] + 5 * d[i, 4] + d[i, 5]$. Obviamente, nos ficheiros `modelo_reg_dgm2.odc`, `modelo_reg_dgm3.odc`, ..., `modelo_reg_dgm16.odc`, as instruções para obtenção de observações da variável resposta são dadas, respectivamente, por $Y[i] \leftarrow d[i, 1] + 10 * d[i, 2] + 2 * d[i, 3] + d[i, 5]$, $Y[i] \leftarrow d[i, 1] + 10 * d[i, 2] + 5 * d[i, 4] + d[i, 5]$, ..., $Y[i] \leftarrow d[i, 5]$.

Para implementação do método GVS, nos casos 3 e 4, basta substituir o ficheiro onde estão os valores das covariáveis e do erro pelos ficheiros `dados_reg_caso3.odc` e `dados_reg_caso4.odc`, respectivamente.

Nos casos 5 e 6 é necessário substituir o ficheiro onde estão os valores das covariáveis e do erro pelos ficheiros `dados_reg_caso5.odc` e `dados_reg_caso6.odc`, respectivamente. Além disso, é necessário alterar a instrução do ficheiro `data_reg.odc`; no caso 5 essa instrução é dada por `list(p=4, N=50, models=16)`, e no caso 6 é dada por `list(p=4, N=100, models=16)`.

Anexo E

Resultados dos Estudos de Simulação

E.1 Dados de contagem

Tabela E.1: Percentagem de classificação correcta das amostras geradas de m_1 , para o caso 2.

θ_1	critério	dimensão da amostra										
		10	20	30	40	50	60	70	80	90	100	
0.171	BF	94.2	99.2	99.8	100	100	100	100	100	100	100	100
	Δ BIC	94.0	99.2	99.8	100	100	100	100	100	100	100	100
	$\Delta\widehat{\text{BIC}}$	94.0	99.2	99.8	100	100	100	100	100	100	100	100
	Δ DIC	94.2	99.2	99.8	100	100	100	100	100	100	100	100
0.265	BF	80.8	94.6	98.6	99.8	99.8	100	99.8	99.8	100	100	100
	Δ BIC	80.2	96.4	98.6	99.8	99.8	100	99.8	99.8	100	100	100
	$\Delta\widehat{\text{BIC}}$	80.4	96.4	98.6	99.8	99.8	100	99.8	99.8	100	100	100
	Δ DIC	80.4	96.4	98.6	99.8	99.8	100	99.8	99.8	100	100	100
0.404	BF	74.0	84.6	89.6	95.0	96.8	98.4	98.8	99.0	99.6	99.8	99.8
	Δ BIC	66.8	83.6	89.0	94.2	96.8	98.2	98.2	99.0	99.6	99.8	99.8
	$\Delta\widehat{\text{BIC}}$	70.0	83.6	89.2	94.4	96.8	98.2	98.2	99.0	99.6	99.8	99.8
	Δ DIC	70.0	83.6	89.2	94.4	96.8	98.2	98.2	99.0	99.6	99.8	99.8
0.525	BF	72.0	79.6	88.6	89.6	91.6	91.8	95.8	96.6	96.2	97.2	97.2
	Δ BIC	53.2	67.2	80.6	85.8	88.0	90.0	93.6	95.8	95.4	96.6	96.6
	$\Delta\widehat{\text{BIC}}$	57.2	68.8	81.0	86.0	88.2	90.0	93.6	95.8	95.4	96.6	96.6
	Δ DIC	57.2	68.8	81.0	86.0	88.0	90.0	93.6	95.8	95.4	96.6	96.6
0.651	BF	82.6	84.8	85.0	85.4	89.6	88.2	90.8	91.0	91.0	92.0	92.0
	Δ BIC	39.0	58.2	64.8	71.4	75.6	78.4	83.6	85.6	85.0	87.0	87.0
	$\Delta\widehat{\text{BIC}}$	50.4	61.8	65.2	72.2	76.2	78.6	84.0	85.6	85.0	87.0	87.0
	Δ DIC	50.4	61.2	65.2	72.2	75.8	78.4	84.0	85.6	85.0	87.0	87.0
0.743	BF	94.0	96.4	92.6	89.4	89.2	93.2	92.4	90.0	89.0	90.4	90.4
	Δ BIC	28.0	52.0	55.0	56.2	65.2	68.4	77.4	70.2	74.4	74.6	74.6
	$\Delta\widehat{\text{BIC}}$	49.2	59.2	55.8	59.4	65.8	68.4	77.8	71.4	74.8	74.6	74.6
	Δ DIC	49.2	56.6	55.4	58.4	65.6	68.4	77.8	70.6	74.8	74.6	74.6

Tabela E.2: Percentagem de classificação correcta das amostras geradas de m_2 , para o caso 2.

θ_2	critério	dimensão da amostra									
		10	20	30	40	50	60	70	80	90	100
0.657	BF	50.0	61.6	70.4	75.2	80.6	83.4	87.0	90.0	90.6	91.6
	Δ BIC	83.2	84.6	85.2	86.8	89.4	91.6	92.0	95.2	94.8	94.4
	$\widehat{\Delta}$ BIC	77.0	83.0	84.6	86.6	89.4	91.0	91.8	95.0	94.8	94.4
	Δ DIC	77.0	83.0	84.6	86.6	89.4	91.6	91.8	95.0	94.8	94.4
1.160	BF	78.8	89.8	93.2	95.2	95.6	97.0	98.0	98.4	98.6	100
	Δ BIC	86.0	91.8	95.6	96.0	96.0	97.8	98.4	98.8	98.8	100
	$\widehat{\Delta}$ BIC	83.4	91.8	95.6	96.0	96.0	97.8	98.4	98.8	98.8	100
	Δ DIC	83.4	91.8	95.6	96.0	96.0	97.8	98.4	98.8	98.8	100
1.409	BF	86.2	92.2	96.2	97.4	98.2	99.2	99.6	99.6	100	99.8
	Δ BIC	89.4	93.8	96.6	98.0	98.2	99.2	99.6	99.6	100	99.8
	$\widehat{\Delta}$ BIC	88.2	93.8	96.6	98.0	98.2	99.2	99.6	99.6	100	99.8
	Δ DIC	88.2	93.6	96.6	98.0	98.2	99.2	99.6	99.6	100	99.8
1.679	BF	89.6	94.4	96.4	99.0	99.6	99.4	99.4	100	99.8	99.8
	Δ BIC	91.4	94.6	96.6	99.0	99.6	99.4	99.4	100	99.8	99.8
	$\widehat{\Delta}$ BIC	90.6	94.6	96.6	99.0	99.6	99.4	99.4	100	99.8	99.8
	Δ DIC	90.6	94.6	96.6	99.0	99.6	99.4	99.4	100	99.8	99.8
2.688	BF	97.0	98.4	99.6	100	100	100	100	100	100	100
	Δ BIC	97.0	98.4	99.6	99.8	100	100	100	100	100	100
	$\widehat{\Delta}$ BIC	97.0	98.4	99.6	99.8	100	100	100	100	100	100
	Δ DIC	96.8	98.4	99.6	99.8	100	100	100	100	100	100
4.020	BF	97.6	99.4	100	100	100	100	100	100	100	100
	Δ BIC	97.8	99.6	100	100	100	100	100	100	100	100
	$\widehat{\Delta}$ BIC	97.8	99.6	100	100	100	100	100	100	100	100
	Δ DIC	97.4	99.6	100	100	100	100	100	100	100	100

Tabela E.3: Médias e variâncias das medidas de selecção de modelos, no caso 1, referentes às amostras geradas de m_1 .

θ_1	n	BIC_{m_1}		\widehat{BIC}_{m_1}		DIC_{m_1}		BIC_{m_2}		\widehat{BIC}_{m_2}		DIC_{m_2}	
		méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.
0.148	10	58.6	47.9	59.8	49.2	58.3	49.1	86.8	738.6	87.8	738.8	86.5	738.8
	20	115.2	90.1	116.3	90.7	114.2	90.7	173.1	1403.8	174.1	1403.9	172.1	1403.9
	30	173.0	122.1	174.0	122.5	171.6	122.5	261.3	2172.8	262.3	2172.8	259.9	2172.8
	40	230.6	146.3	231.6	146.6	228.9	146.6	356.4	3513.5	357.4	3513.6	354.7	3513.6
	50	286.2	211.1	287.3	211.3	284.3	211.5	440.0	4349.3	441.0	4349.4	438.1	4349.3
	60	342.1	240.3	343.1	240.7	340.0	240.5	521.2	4566.1	522.2	4566.2	519.1	4566.4
	70	400.1	279.3	401.1	279.8	397.8	280.0	613.8	5346.8	614.8	5347.0	611.6	5346.9
	80	457.1	302.2	458.1	302.6	454.7	302.6	702.6	7131.2	703.6	7131.2	700.2	7131.2
	90	514.1	353.6	515.1	353.1	511.6	353.7	796.9	6614.7	797.9	6614.4	794.4	6614.5
	100	572.1	393.0	573.1	392.8	569.5	392.3	884.9	7914.0	885.9	7914.2	882.3	7913.1
0.265	10	44.7	44.8	45.7	46.2	44.2	46.3	53.4	225.0	54.4	225.1	53.1	225.1
	20	89.5	78.3	90.5	79.0	88.4	79.0	110.1	501.0	111.1	501.0	109.1	501.0
	30	133.6	123.8	134.6	124.4	132.2	124.4	167.2	788.0	168.2	788.0	165.7	788.0
	40	176.6	162.2	177.6	162.6	174.8	162.6	219.5	1031.9	220.5	1031.9	217.8	1031.9
	50	221.1	208.1	222.1	208.4	219.1	208.5	277.3	1357.7	278.3	1357.6	275.4	1357.7
	60	264.6	244.0	265.6	244.3	262.5	244.2	331.2	1469.4	332.2	1469.3	329.1	1469.6
	70	309.6	269.4	310.6	269.4	307.3	269.3	389.4	1835.4	390.4	1835.5	387.2	1835.3
	80	352.8	346.7	353.8	347.1	350.4	346.9	444.9	2204.0	445.9	2203.9	442.6	2203.8
	90	396.6	356.6	397.7	356.8	394.1	356.8	498.0	2028.6	499.0	2028.4	495.5	2028.3
	100	441.1	384.3	442.1	384.0	438.5	383.7	552.1	2540.5	553.1	2540.3	549.5	2540.0
0.421	10	33.7	41.3	34.5	41.6	33.0	42.1	36.5	103.6	37.5	103.7	36.2	103.7
	20	66.7	76.8	67.7	77.3	65.5	77.5	73.2	182.8	74.2	182.8	72.2	182.8
	30	99.6	113.4	100.5	113.7	98.0	113.9	110.2	294.6	111.2	294.6	108.8	294.6
	40	132.1	142.3	133.0	142.5	130.3	142.7	146.5	375.7	147.5	375.7	144.9	375.7
	50	164.6	182.7	165.6	182.9	162.6	183.0	184.2	531.5	185.2	531.5	182.2	531.5
	60	198.0	234.6	199.0	234.7	195.8	234.7	221.6	617.5	222.6	617.5	219.6	617.5
	70	228.8	252.9	229.8	253.0	226.5	253.1	255.1	666.7	256.1	666.7	252.8	666.6
	80	263.8	311.6	264.7	311.9	261.3	312.1	294.9	852.9	295.9	852.8	292.5	853.0
	90	295.0	373.8	296.0	373.9	292.5	374.0	330.9	1003.4	331.9	1003.4	328.4	1003.5
	100	327.7	401.8	328.7	402.0	325.0	402.2	367.9	1118.3	368.9	1118.4	365.3	1118.4
0.534	10	27.6	37.3	28.5	36.5	26.9	37.3	28.7	63.9	29.7	63.9	28.4	64.0
	20	53.7	82.7	54.6	82.3	52.4	82.9	57.1	145.9	58.1	145.9	56.1	146.0
	30	80.1	113.9	81.0	113.8	78.5	114.2	84.6	195.6	85.6	195.6	83.2	195.6
	40	106.5	152.2	107.4	152.2	104.7	152.4	113.6	284.9	114.6	284.9	111.9	284.9
	50	131.7	190.2	132.7	190.1	129.7	190.3	140.2	353.1	141.2	353.2	138.3	353.1
	60	159.1	237.9	160.0	237.9	156.9	238.0	169.5	435.5	170.5	435.5	167.4	435.5
	70	183.0	296.7	184.0	296.7	180.7	296.9	195.8	535.6	196.8	535.6	193.6	535.6
	80	209.6	265.9	210.6	265.7	207.2	265.9	224.3	494.9	225.3	494.9	221.9	495.0
	90	236.8	348.3	237.8	348.3	234.3	348.3	253.7	648.9	254.7	648.9	251.2	648.9
	100	261.1	369.4	262.1	369.3	258.4	369.4	279.0	708.9	280.0	709.0	276.3	708.9
0.626	10	22.9	39.9	23.9	37.3	22.3	38.4	23.4	53.2	24.4	53.2	23.1	53.4
	20	43.3	73.6	44.3	72.0	42.1	72.8	44.4	98.6	45.4	98.6	43.4	98.7
	30	65.5	98.0	66.4	97.3	63.9	97.9	68.2	151.7	69.2	151.7	66.8	151.7
	40	86.4	144.0	87.3	143.4	84.6	143.9	89.7	196.5	90.7	196.5	88.0	196.5
	50	107.5	173.5	108.4	173.0	105.4	173.4	111.6	250.5	112.6	250.5	109.7	250.5
	60	130.7	213.9	131.7	213.5	128.5	213.8	136.2	315.4	137.2	315.4	134.1	315.4
	70	150.4	262.1	151.4	261.8	148.1	262.1	157.1	399.1	158.1	399.1	154.8	399.1
	80	172.9	287.1	173.9	286.8	170.4	287.1	180.9	451.4	181.9	451.4	178.5	451.4
	90	193.6	327.2	194.5	327.0	191.0	327.2	202.2	481.5	203.2	481.5	199.7	481.6
	100	214.9	425.0	215.9	424.9	212.2	425.1	224.0	626.0	225.0	626.0	221.4	626.0

Tabela E.3: Médias e variâncias das medidas de selecção de modelos, no caso 1, referentes às amostras geradas de m_1 (continuação).

θ_1	n	BIC_{m_1}		\widehat{BIC}_{m_1}		DIC_{m_1}		BIC_{m_2}		\widehat{BIC}_{m_2}		DIC_{m_2}	
		méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.
0.776	10	15.9	25.8	17.1	22.1	15.4	23.2	15.8	28.3	16.8	28.4	15.5	28.6
	20	29.9	60.6	31.1	56.1	28.8	57.6	30.3	69.4	31.3	69.4	29.3	69.6
	30	42.9	103.8	44.1	99.1	41.4	100.8	43.4	117.5	44.4	117.6	42.0	117.7
	40	57.3	136.4	58.4	133.2	55.5	134.5	58.0	155.7	59.0	155.7	56.3	155.7
	50	71.9	178.0	73.0	175.5	69.9	176.6	73.0	206.0	74.0	206.1	71.1	206.1
	60	84.7	185.6	85.8	183.7	82.6	184.6	86.3	215.6	87.3	215.6	84.2	215.6
	70	99.6	262.7	100.6	260.8	97.3	261.7	101.0	301.6	102.0	301.6	98.8	301.6
	80	111.9	285.4	113.0	283.8	109.5	284.6	113.7	326.0	114.7	326.0	111.3	326.0
	90	127.6	298.4	128.6	297.1	125.0	297.8	129.7	356.7	130.7	356.7	127.2	356.7
	100	141.2	312.4	142.2	311.2	138.6	311.8	143.7	360.3	144.7	360.3	141.1	360.3

Tabela E.4: Médias e variâncias das medidas de selecção de modelos, no caso 1, referentes às amostras geradas de m_2 .

θ_1	n	BIC_{m_1}		\widehat{BIC}_{m_1}		DIC_{m_1}		BIC_{m_2}		\widehat{BIC}_{m_2}		DIC_{m_2}	
		méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.
0.309	10	16.5	20.5	17.7	17.7	16.0	18.5	16.0	18.1	17.0	18.1	15.7	18.3
	20	30.6	55.6	31.8	52.0	29.5	53.2	29.8	50.4	30.8	50.4	28.8	50.5
	30	45.9	75.2	47.0	72.9	44.4	73.9	44.9	69.9	45.9	69.9	43.5	69.9
	40	60.7	97.5	61.8	95.7	58.9	96.5	59.5	90.4	60.5	90.4	57.8	90.4
	50	76.0	131.7	77.1	130.2	74.0	130.9	74.5	123.4	75.5	123.4	72.6	123.4
	60	89.0	168.0	90.1	166.6	86.9	167.3	87.5	157.8	88.5	157.8	85.5	157.8
	70	103.9	172.2	105.0	171.2	101.6	171.7	102.1	162.7	103.1	162.7	99.8	162.8
	80	117.8	191.8	118.9	190.9	115.4	191.3	115.8	178.3	116.8	178.3	113.4	178.3
	90	131.9	240.5	132.9	239.5	129.4	240.0	129.4	221.6	130.4	221.6	126.9	221.6
	100	147.1	257.8	148.1	256.9	144.4	257.3	144.4	235.5	145.4	235.5	141.8	235.5
0.632	10	23.6	21.0	24.4	20.0	22.9	20.6	22.4	17.8	23.3	17.9	22.0	17.9
	20	45.8	49.8	46.7	49.2	44.5	49.6	43.7	42.6	44.7	42.6	42.7	42.6
	30	68.1	70.7	69.0	70.3	66.5	70.6	65.3	61.1	66.3	61.1	63.9	61.1
	40	89.8	86.7	90.8	86.4	88.0	86.7	86.4	79.8	87.4	79.8	84.7	79.8
	50	112.8	110.6	113.8	110.4	110.8	110.6	108.1	101.1	109.1	101.1	106.2	101.1
	60	134.9	133.6	135.9	133.4	132.7	133.6	129.7	124.4	130.7	124.4	127.6	124.4
	70	156.5	165.5	157.5	165.3	154.2	165.5	150.4	137.1	151.4	137.1	148.2	137.1
	80	177.2	176.1	178.1	175.9	174.7	176.1	170.7	162.3	171.7	162.2	168.3	162.2
	90	199.6	202.6	200.6	202.4	197.0	202.7	191.9	182.5	192.9	182.5	189.4	182.4
	100	223.2	214.1	224.1	214.0	220.5	214.0	214.2	196.4	215.2	196.5	211.6	196.5
1.128	10	31.4	17.2	32.2	17.4	30.7	17.6	28.9	16.1	29.9	16.1	28.6	16.1
	20	61.3	39.7	62.2	39.8	60.1	40.0	57.0	35.0	58.0	35.1	56.0	35.1
	30	91.4	54.0	92.3	54.1	89.8	54.2	85.0	54.5	86.0	54.5	83.6	54.5
	40	121.1	75.7	122.1	75.8	119.3	75.9	113.1	70.9	114.1	70.9	111.4	70.9
	50	151.2	88.9	152.1	89.0	149.2	89.0	140.7	86.1	141.7	86.1	138.8	86.1
	60	180.7	108.2	181.6	108.2	178.5	108.2	168.6	104.4	169.6	104.5	166.5	104.5
	70	209.8	115.7	210.7	115.6	207.5	115.8	195.8	118.8	196.8	118.8	193.5	118.8
	80	238.7	152.2	239.6	152.1	236.2	152.2	222.5	132.1	223.5	132.1	220.1	132.1
	90	268.6	153.9	269.6	153.9	266.1	153.7	250.6	137.5	251.6	137.6	248.1	137.6
	100	298.2	191.0	299.2	190.8	295.6	191.1	278.4	160.6	279.4	160.7	275.8	160.5

Tabela E.4: Médias e variâncias das medidas de selecção de modelos, no caso 1, referentes às amostras geradas de m_2 (continuação).

θ_1	n	BIC_{m_1}		\widehat{BIC}_{m_1}		DIC_{m_1}		BIC_{m_2}		\widehat{BIC}_{m_2}		DIC_{m_2}	
		méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.
1.465	10	35.0	16.9	35.8	17.3	34.3	17.4	31.7	17.1	32.7	17.1	31.4	17.1
	20	69.5	31.7	70.5	32.0	68.4	32.0	63.6	32.9	64.6	32.9	62.6	32.9
	30	103.0	52.2	103.9	52.4	101.4	52.5	94.3	50.3	95.3	50.3	92.9	50.3
	40	136.3	59.0	137.3	59.1	134.6	59.2	124.7	65.3	125.7	65.3	123.0	65.3
	50	169.9	79.3	170.8	79.4	167.9	79.5	155.7	86.4	156.7	86.5	153.8	86.5
	60	203.0	93.2	203.9	93.2	200.8	93.2	185.9	98.8	186.9	98.8	183.8	98.8
	70	236.5	113.7	237.5	113.9	234.2	113.9	216.5	106.2	217.5	106.1	214.3	106.1
	80	270.7	126.7	271.7	126.8	268.3	127.0	247.5	121.3	248.5	121.2	245.1	121.3
	90	303.3	141.3	304.3	141.3	300.8	141.2	277.9	154.0	278.9	154.0	275.4	153.9
	100	337.6	157.0	338.6	157.2	335.0	157.1	308.5	168.8	309.5	168.8	305.9	168.9
1.628	10	37.0	16.6	37.9	17.0	36.4	17.1	33.2	16.5	34.2	16.5	32.9	16.5
	20	72.1	29.8	73.1	30.0	71.0	30.1	65.4	33.1	66.4	33.1	64.4	33.1
	30	108.1	43.0	109.1	43.1	106.6	43.2	98.0	49.8	99.0	49.8	96.6	49.8
	40	143.1	57.2	144.0	57.2	141.3	57.3	129.7	67.5	130.7	67.5	128.1	67.5
	50	177.7	71.2	178.6	71.4	175.7	71.5	161.4	82.3	162.4	82.3	159.5	82.3
	60	213.2	80.8	214.2	80.8	211.0	80.8	193.9	83.2	194.9	83.2	191.8	83.2
	70	248.3	95.0	249.3	95.3	246.0	95.1	225.4	124.5	226.4	124.5	223.2	124.5
	80	283.4	118.0	284.3	118.0	280.9	118.2	256.8	118.5	257.8	118.5	254.4	118.5
	90	318.5	132.9	319.5	132.8	316.0	132.7	289.6	143.3	290.6	143.4	287.1	143.3
	100	353.3	158.3	354.3	158.4	350.6	158.1	321.2	176.0	322.2	176.0	318.6	175.7
2.741	10	45.4	10.3	46.4	10.6	45.0	10.7	38.7	14.9	39.7	14.9	38.4	14.9
	20	89.5	20.2	90.5	20.5	88.4	20.4	77.6	34.6	78.6	34.5	76.6	34.6
	30	133.1	31.7	134.1	31.8	131.7	31.9	115.3	49.2	116.3	49.2	113.9	49.2
	40	177.3	38.7	178.3	38.8	175.5	38.8	153.7	73.5	154.7	73.5	152.0	73.5
	50	220.9	50.8	221.9	50.8	218.9	50.9	191.1	85.0	192.1	85.0	189.2	85.0
	60	264.4	65.8	265.4	65.9	262.3	65.8	228.3	105.4	229.3	105.3	226.2	105.3
	70	307.5	76.9	308.5	77.0	305.2	76.9	266.4	113.6	267.4	113.7	264.1	113.6
	80	351.4	77.2	352.4	77.5	349.0	77.4	304.8	129.9	305.8	129.9	302.5	130.0
	90	395.4	92.0	396.4	92.1	392.9	92.7	342.7	165.9	343.7	165.8	340.2	165.9
	100	439.4	96.3	440.4	96.4	436.7	96.0	380.8	166.3	381.8	166.2	378.2	166.3

Tabela E.5: Médias e variâncias das medidas de selecção de modelos, no caso 2, referentes às amostras geradas de m_1 .

θ_1	n	BIC_{m_1}		\widehat{BIC}_{m_1}		DIC_{m_1}		BIC_{m_2}		\widehat{BIC}_{m_2}		DIC_{m_2}	
		méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.
0.171	10	54.6	40.0	55.7	40.5	54.2	40.5	74.9	448.7	76.0	452.6	74.7	453.0
	20	108.6	83.8	109.6	84.0	107.5	84.0	154.1	1165.5	155.1	1167.7	153.1	1168.0
	30	163.2	132.1	164.2	132.2	161.7	132.2	234.2	2007.7	235.3	2009.7	232.9	2010.0
	40	216.4	156.9	217.4	157.1	214.7	157.1	312.7	2222.2	313.7	2223.3	311.0	2223.4
	50	270.6	205.8	271.6	205.9	268.7	205.8	393.9	2942.5	394.9	2943.5	392.0	2943.5
	60	323.7	224.4	324.7	224.4	321.6	224.4	469.1	3420.7	470.1	3421.4	467.0	3421.5
	70	377.8	262.3	378.8	262.6	375.5	262.2	551.4	4035.0	552.5	4036.4	549.2	4036.2
	80	430.9	330.6	431.9	330.9	428.5	331.3	629.5	5024.2	630.5	5025.1	627.1	5025.7
	90	484.7	351.7	485.7	351.9	482.1	351.8	702.9	5262.6	703.9	5262.6	700.4	5262.6
	100	538.2	386.5	539.2	386.9	535.6	386.3	785.5	5925.1	786.5	5926.6	782.9	5926.3

Tabela E.5: Médias e variâncias das medidas de selecção de modelos, no caso 2, referentes às amostras geradas de m_1 (continuação).

θ_1	n	BIC_{m_1}		\widehat{BIC}_{m_1}		DIC_{m_1}		BIC_{m_2}		\widehat{BIC}_{m_2}		DIC_{m_2}	
		méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.
0.265	10	44.9	43.8	45.9	44.5	44.5	44.5	53.8	223.4	54.7	224.7	53.4	225.3
	20	89.7	70.8	90.7	71.1	88.7	71.1	109.8	382.8	110.7	383.4	108.7	383.6
	30	133.4	123.3	134.4	123.5	132.0	123.6	165.5	688.7	166.5	689.2	164.1	689.4
	40	176.4	165.6	177.4	165.7	174.7	165.8	219.7	1001.2	220.7	1001.6	218.0	1001.8
	50	220.4	190.8	221.4	190.9	218.5	191.0	275.0	1215.3	276.0	1215.6	273.1	1215.8
	60	265.4	235.7	266.4	235.8	263.3	236.0	332.0	1522.3	333.0	1522.7	329.9	1522.7
	70	310.3	281.0	311.3	281.0	308.0	281.0	390.7	1999.1	391.7	1999.4	388.5	1999.3
	80	353.4	349.8	354.4	350.1	351.0	350.0	445.8	2261.5	446.8	2261.6	443.4	2261.9
	90	397.3	360.3	398.3	360.7	394.8	360.7	500.2	2532.5	501.2	2532.9	497.7	2532.9
	100	441.0	361.9	442.0	362.0	438.4	361.4	556.0	2617.1	557.0	2617.6	553.4	2617.2
0.404	10	34.8	42.9	35.7	43.3	34.2	43.7	37.9	109.3	38.9	109.0	37.4	109.9
	20	68.8	76.0	69.7	76.3	67.6	76.4	76.2	198.3	77.2	198.2	75.1	198.6
	30	102.6	125.1	103.5	125.3	101.1	125.4	114.5	353.9	115.4	353.9	113.0	354.2
	40	136.6	167.3	137.6	167.4	134.8	167.5	153.7	449.5	154.7	449.4	152.0	449.7
	50	169.2	195.0	170.2	195.3	167.2	195.3	189.6	512.0	190.6	512.0	187.6	512.1
	60	203.5	247.2	204.5	247.4	201.3	247.4	229.9	700.6	230.9	700.7	227.8	700.8
	70	236.0	284.8	237.0	285.0	233.8	285.1	264.6	810.3	265.6	810.3	262.3	810.5
	80	271.5	294.9	272.5	295.0	269.1	295.1	307.7	938.1	308.6	937.9	305.2	938.0
	90	303.8	355.1	304.8	354.9	301.3	355.0	342.1	1074.9	343.0	1074.8	339.5	1074.8
	100	337.1	359.2	338.1	358.9	334.5	359.2	381.7	1090.9	382.7	1091.1	379.1	1091.4
0.525	10	28.1	39.8	29.0	39.6	27.5	40.3	29.5	78.0	30.5	76.8	29.0	78.0
	20	55.1	80.8	56.0	80.8	53.9	81.2	57.9	134.0	58.9	133.3	56.8	134.1
	30	81.7	120.4	82.6	120.4	80.2	120.7	87.1	211.4	88.1	211.0	85.6	211.4
	40	108.1	163.7	109.0	163.7	106.3	164.0	115.7	325.9	116.7	325.6	113.9	326.0
	50	135.2	208.0	136.2	208.0	133.2	208.2	144.4	370.4	145.4	370.1	142.4	370.5
	60	161.2	208.1	162.2	208.2	159.1	208.2	172.4	387.8	173.4	387.7	170.3	387.9
	70	186.8	279.1	187.8	279.1	184.5	279.1	200.0	502.8	201.0	502.6	197.7	502.9
	80	214.4	284.7	215.4	284.7	212.0	284.7	229.6	506.6	230.6	506.6	227.2	506.6
	90	240.3	354.3	241.2	354.4	237.7	354.5	257.3	681.0	258.3	680.8	254.8	681.1
	100	268.7	358.9	269.7	358.9	266.1	359.2	289.1	682.4	290.1	682.3	286.5	682.5
0.651	10	21.5	35.1	22.4	33.9	20.8	35.0	21.7	43.7	22.9	41.4	21.3	42.9
	20	41.6	80.1	42.5	79.3	40.4	80.2	42.7	104.3	43.7	102.8	41.6	103.9
	30	62.5	111.8	63.5	111.4	61.0	111.9	64.6	157.6	65.6	156.6	63.1	157.4
	40	82.6	145.5	83.5	145.1	80.8	145.6	85.2	193.8	86.2	193.1	83.4	193.7
	50	102.2	178.6	103.2	178.3	100.2	178.6	106.0	255.3	107.0	254.8	104.0	255.3
	60	123.4	193.3	124.4	193.1	121.2	193.3	128.0	269.0	129.1	268.6	125.9	268.9
	70	141.5	273.6	142.4	273.4	139.2	273.7	147.2	394.5	148.2	394.0	145.0	394.4
	80	163.1	298.8	164.1	298.7	160.7	298.8	169.5	428.6	170.5	428.3	167.0	428.5
	90	182.6	355.3	183.6	355.3	180.0	355.5	189.4	478.8	190.4	478.5	186.9	478.7
	100	203.4	381.0	204.3	381.0	200.7	381.1	211.3	553.2	212.3	552.9	208.7	553.2
0.743	10	17.1	27.4	18.1	25.6	16.4	26.8	17.1	31.4	18.3	28.8	16.6	30.2
	20	32.7	67.7	33.7	66.0	31.5	67.2	33.2	81.3	34.4	78.8	32.2	80.4
	30	47.9	102.9	48.9	101.7	46.3	102.7	48.7	120.8	49.8	119.1	47.2	120.3
	40	64.4	142.7	65.4	141.8	62.6	142.6	65.4	165.8	66.5	164.5	63.6	165.5
	50	79.8	180.9	80.8	180.2	77.8	180.8	81.3	217.1	82.3	216.1	79.3	216.9
	60	94.5	200.2	95.5	199.7	92.4	200.2	96.7	243.0	97.7	242.2	94.5	242.9
	70	109.8	256.8	110.8	256.3	107.5	256.8	112.4	298.5	113.5	297.8	110.1	298.4
	80	126.1	280.6	127.1	280.3	123.7	280.7	128.7	326.3	129.7	325.7	126.2	326.2
	90	141.2	323.7	142.2	323.4	138.6	323.8	144.0	377.5	145.0	376.9	141.5	377.4
	100	155.6	351.9	156.6	351.5	152.9	351.8	158.7	417.3	159.8	416.8	156.1	417.2

Tabela E.6: Médias e variâncias das medidas de selecção de modelos, no caso 2, referentes às amostras geradas de m_2 .

θ_1	n	BIC_{m_1}		\widehat{BIC}_{m_1}		DIC_{m_1}		BIC_{m_2}		\widehat{BIC}_{m_2}		DIC_{m_2}	
		méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.
0.657	10	23.8	24.8	24.7	24.3	23.2	24.9	22.6	21.1	23.6	20.1	22.1	20.8
	20	47.2	46.4	48.1	46.2	46.0	46.5	45.0	40.3	46.0	39.8	43.9	40.3
	30	69.7	69.6	70.7	69.5	68.2	69.8	66.7	59.3	67.7	59.0	65.2	59.3
	40	93.0	93.4	93.9	93.3	91.2	93.5	89.3	87.4	90.3	87.2	87.6	87.4
	50	115.4	122.0	116.3	121.9	113.4	122.1	110.6	109.8	111.6	109.6	108.7	109.8
	60	137.2	142.3	138.2	142.2	135.1	142.3	131.8	125.1	132.8	125.0	129.7	125.1
	70	159.5	171.8	160.5	171.7	157.2	171.8	153.0	149.2	154.0	149.1	150.7	149.3
	80	181.1	166.4	182.1	166.3	178.7	166.4	173.6	145.5	174.6	145.3	171.2	145.5
	90	204.1	191.1	205.1	191.1	201.6	191.2	196.2	169.6	197.2	169.5	193.6	169.6
	100	226.5	216.3	227.5	216.4	223.9	216.4	217.5	186.5	218.5	186.4	214.9	186.5
1.160	10	31.3	22.0	32.2	22.2	30.8	22.4	28.7	18.8	29.7	18.5	28.2	18.9
	20	62.1	34.6	63.1	34.7	61.0	34.8	57.3	30.8	58.2	30.7	56.1	30.9
	30	91.6	54.5	92.5	54.6	90.1	54.7	85.0	49.6	85.9	49.5	83.5	49.6
	40	122.7	62.8	123.6	62.8	120.9	63.0	113.9	59.0	114.9	59.0	112.2	59.0
	50	152.5	86.9	153.5	86.9	150.5	86.9	142.0	89.2	143.0	89.2	140.1	89.2
	60	182.9	116.9	183.9	117.0	180.7	117.0	170.5	109.0	171.4	109.0	168.3	109.1
	70	213.2	136.1	214.2	135.9	210.9	136.0	198.2	126.8	199.2	126.8	195.9	126.8
	80	243.0	149.6	243.9	149.5	240.5	149.6	225.9	145.7	226.9	145.6	223.5	145.7
	90	272.6	163.0	273.6	162.7	270.1	163.1	254.1	157.8	255.1	157.7	251.6	157.8
	100	302.1	171.8	303.1	172.0	299.5	172.0	281.5	166.5	282.5	166.6	278.9	166.7
1.409	10	34.5	17.9	35.3	18.1	33.9	18.2	31.2	16.7	32.1	16.6	30.7	16.8
	20	68.1	32.4	69.1	32.5	67.0	32.6	62.1	34.5	63.0	34.5	61.0	34.6
	30	100.9	55.0	101.9	55.1	99.5	55.1	92.5	54.2	93.5	54.1	91.0	54.2
	40	134.3	57.9	135.2	57.9	132.5	58.0	123.0	59.7	124.0	59.7	121.3	59.8
	50	167.0	82.8	168.0	83.0	165.0	82.9	153.4	85.2	154.3	85.2	151.4	85.2
	60	199.4	101.5	200.4	101.7	197.3	101.6	183.3	99.3	184.3	99.3	181.2	99.4
	70	232.8	106.0	233.8	105.9	230.5	106.0	214.5	111.4	215.5	111.4	212.2	111.4
	80	265.6	138.0	266.6	138.1	263.2	138.1	243.8	133.3	244.8	133.4	241.4	133.3
	90	297.6	156.7	298.6	156.6	295.1	156.7	273.2	161.9	274.2	161.9	270.7	161.7
	100	331.1	139.1	332.1	139.1	328.5	139.1	304.3	150.3	305.2	150.3	301.6	150.2
1.679	10	37.0	14.4	37.9	14.6	36.4	14.6	33.1	15.2	34.0	15.2	32.6	15.4
	20	73.9	29.0	74.9	29.2	72.8	29.2	66.5	33.3	67.4	33.3	65.4	33.3
	30	109.2	45.1	110.1	45.2	107.7	45.2	98.9	51.5	99.9	51.5	97.5	51.6
	40	144.7	54.1	145.7	54.1	142.9	54.1	131.2	55.9	132.1	55.9	129.4	56.0
	50	180.5	71.8	181.5	71.9	178.5	72.1	163.4	79.8	164.3	79.8	161.4	79.9
	60	216.1	82.9	217.0	83.0	213.9	83.0	195.8	97.1	196.8	97.0	193.7	97.1
	70	251.5	102.7	252.5	102.8	249.2	102.8	227.8	104.9	228.7	104.9	225.5	104.9
	80	287.2	117.1	288.2	117.4	284.8	117.1	260.7	136.2	261.7	136.2	258.3	136.2
	90	322.1	138.0	323.1	137.6	319.5	137.6	292.3	150.8	293.3	150.8	289.8	150.9
	100	358.5	150.1	359.5	149.7	355.8	150.0	325.0	167.0	326.0	167.0	322.4	166.9
2.688	10	45.2	10.2	46.2	10.3	44.7	10.3	38.8	15.6	39.8	15.7	38.4	15.8
	20	89.2	21.1	90.2	21.2	88.1	21.2	77.2	31.2	78.2	31.3	76.2	31.3
	30	132.8	32.3	133.8	32.3	131.4	32.4	115.2	50.8	116.2	50.8	113.8	50.8
	40	175.8	42.2	176.8	42.2	174.1	42.4	151.8	67.2	152.8	67.2	150.1	67.3
	50	219.6	58.7	220.6	58.8	217.7	58.7	190.4	85.9	191.4	85.9	188.4	85.9
	60	262.8	62.5	263.8	62.5	260.7	62.7	228.2	110.9	229.2	111.1	226.1	111.0
	70	305.8	75.1	306.8	75.1	303.6	75.2	265.7	130.4	266.7	130.4	263.4	130.4
	80	348.8	77.4	349.8	77.6	346.4	77.4	301.9	132.6	302.9	132.6	299.5	132.6
	90	392.5	89.2	393.5	89.7	390.0	89.7	341.0	152.7	342.0	152.6	338.5	152.7
	100	435.5	98.1	436.5	98.4	432.9	98.3	378.8	188.2	379.7	188.2	376.1	188.2

Tabela E.6: Médias e variâncias das medidas de selecção de modelos, no caso 2, referentes às amostras geradas de m_2 (continuação).

θ_1	n	BIC_{m_1}		\widehat{BIC}_{m_1}		DIC_{m_1}		BIC_{m_2}		\widehat{BIC}_{m_2}		DIC_{m_2}	
		méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.	méd.	var.
4.020	10	52.4	7.6	53.4	7.7	52.0	7.7	43.0	15.5	44.0	15.6	42.7	15.6
	20	103.3	14.8	104.3	14.9	102.2	14.8	86.0	37.4	87.1	37.5	85.0	37.5
	30	153.4	22.7	154.4	22.6	152.0	22.7	128.0	52.3	129.0	52.4	126.6	52.4
	40	203.9	32.5	204.9	32.4	202.1	32.4	168.9	71.4	169.9	71.4	167.2	71.4
	50	254.3	39.7	255.3	39.5	252.4	39.6	212.1	92.5	213.1	92.6	210.2	92.5
	60	305.7	48.5	306.7	48.8	303.5	48.7	254.8	109.3	255.8	109.4	252.7	109.4
	70	355.3	51.1	356.4	51.1	353.1	51.1	296.2	130.3	297.2	130.6	294.0	130.5
	80	405.1	62.6	406.1	62.4	402.7	62.7	337.2	153.9	338.2	153.8	334.9	153.8
	90	455.1	72.2	456.1	71.8	452.6	72.2	378.6	173.8	379.6	173.8	376.1	173.8
	100	506.0	72.3	507.0	72.6	503.3	73.2	421.8	190.0	422.8	189.9	419.2	189.8

E.2 Dados de regressão linear normal

Tabela E.7: Número de vezes que cada modelo é seleccionado pelo DIC, por modelo gerador dos dados e modelo ajustado, para o caso 2.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	1000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_2	156	844	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_3	163	0	837	0	0	0	0	0	0	0	0	0	0	0	0	0
m_4	167	0	0	833	0	0	0	0	0	0	0	0	0	0	0	0
m_5	153	0	0	0	847	0	0	0	0	0	0	0	0	0	0	0
m_6	36	116	130	0	0	718	0	0	0	0	0	0	0	0	0	0
m_7	28	152	0	121	0	0	699	0	0	0	0	0	0	0	0	0
m_8	26	0	141	133	0	0	0	700	0	0	0	0	0	0	0	0
m_9	24	157	0	0	130	0	0	0	689	0	0	0	0	0	0	0
m_{10}	31	0	139	0	144	0	0	0	0	686	0	0	0	0	0	0
m_{11}	23	0	0	139	136	0	0	0	0	0	702	0	0	0	0	0
m_{12}	4	27	24	13	0	102	118	100	0	0	0	612	0	0	0	0
m_{13}	3	24	22	0	30	110	0	0	108	116	0	0	587	0	0	0
m_{14}	0	21	0	19	27	0	110	0	115	0	114	0	0	594	0	0
m_{15}	6	0	30	20	22	0	0	93	0	116	118	0	0	0	595	0
m_{16}	2	3	6	4	8	24	28	18	17	19	21	85	83	100	82	500

Tabela E.8: Número de vezes que cada modelo é seleccionado pelo $\widehat{\text{BIC}}$, por modelo gerador dos dados e modelo ajustado, para o caso 2.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	998	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
m_2	19	981	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_3	9	0	990	0	0	0	0	0	0	1	0	0	0	0	0	0
m_4	15	0	0	985	0	0	0	0	0	0	0	0	0	0	0	0
m_5	14	0	0	0	986	0	0	0	0	0	0	0	0	0	0	0
m_6	0	15	10	0	0	975	0	0	0	0	0	0	0	0	0	0
m_7	0	21	0	12	0	0	967	0	0	0	0	0	0	0	0	0
m_8	0	0	14	15	0	0	0	971	0	0	0	0	0	0	0	0
m_9	0	20	0	0	18	0	0	0	962	0	0	0	0	0	0	0
m_{10}	0	0	18	0	13	0	0	0	0	969	0	0	0	0	0	0
m_{11}	1	0	0	16	9	0	0	0	0	0	974	0	0	0	0	0
m_{12}	0	0	0	0	0	10	12	9	0	0	0	968	0	0	0	1
m_{13}	0	0	0	0	0	7	0	0	14	13	0	0	966	0	0	0
m_{14}	0	1	0	0	0	0	15	0	10	0	14	0	0	960	0	0
m_{15}	0	0	0	0	0	0	0	11	0	9	11	0	0	0	969	0
m_{16}	0	0	0	0	0	0	0	0	0	0	1	8	16	16	12	947

Tabela E.9: Número de vezes que cada modelo é seleccionado pelo LLP, por modelo gerador dos dados e modelo ajustado, para o caso 2.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	1000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_2	349	651	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_3	323	0	677	0	0	0	0	0	0	0	0	0	0	0	0	0
m_4	322	0	0	678	0	0	0	0	0	0	0	0	0	0	0	0
m_5	293	0	0	0	707	0	0	0	0	0	0	0	0	0	0	0
m_6	115	200	208	0	0	477	0	0	0	0	0	0	0	0	0	0
m_7	116	207	0	205	0	0	472	0	0	0	0	0	0	0	0	0
m_8	105	0	222	220	0	0	0	453	0	0	0	0	0	0	0	0
m_9	96	230	0	0	203	0	0	0	471	0	0	0	0	0	0	0
m_{10}	100	0	220	0	239	0	0	0	0	441	0	0	0	0	0	0
m_{11}	101	0	0	223	213	0	0	0	0	0	463	0	0	0	0	0
m_{12}	33	76	70	64	0	149	164	128	0	0	0	316	0	0	0	0
m_{13}	28	62	78	0	58	164	0	0	149	146	0	0	315	0	0	0
m_{14}	27	68	0	55	76	0	160	0	154	0	158	0	0	302	0	0
m_{15}	39	0	64	67	58	0	0	116	0	156	149	0	0	0	351	0
m_{16}	11	24	21	21	27	55	63	37	41	42	32	87	110	94	116	219

Tabela E.10: Número de vezes que cada modelo é seleccionado pelo DIC, por modelo gerador dos dados e modelo ajustado, para o caso 3.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	1000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_2	147	853	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_3	163	0	837	0	0	0	0	0	0	0	0	0	0	0	0	0
m_4	149	0	0	851	0	0	0	0	0	0	0	0	0	0	0	0
m_5	157	0	0	0	843	0	0	0	0	0	0	0	0	0	0	0
m_6	24	133	152	0	0	690	0	0	1	0	0	0	0	0	0	0
m_7	26	140	0	138	0	0	696	0	0	0	0	0	0	0	0	0
m_8	25	0	105	149	0	0	0	720	0	0	1	0	0	0	0	0
m_9	25	139	0	0	133	0	0	0	702	0	0	0	0	1	0	0
m_{10}	34	0	119	0	128	0	0	0	0	718	1	0	0	0	0	0
m_{11}	33	0	0	121	133	0	0	0	0	0	713	0	0	0	0	0
m_{12}	4	24	26	19	0	100	116	106	0	0	0	604	0	1	0	0
m_{13}	9	33	25	0	18	100	0	0	97	115	0	0	603	0	0	0
m_{14}	5	22	0	17	29	0	112	0	116	0	115	0	0	584	0	0
m_{15}	4	0	17	39	14	0	0	107	0	117	96	0	0	0	606	0
m_{16}	0	4	3	8	3	15	26	15	11	21	13	107	102	75	92	505

Tabela E.11: Número de vezes que cada modelo é seleccionado pelo \widehat{BIC} , por modelo gerador dos dados e modelo ajustado, para o caso 3.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	970	8	0	11	10	0	1	0	0	0	0	0	0	0	0	0
m_2	14	972	0	0	0	0	7	0	7	0	0	0	0	0	0	0
m_3	18	0	965	0	1	3	0	11	0	0	0	2	0	0	0	0
m_4	7	0	0	981	0	0	3	0	0	0	9	0	0	0	0	0
m_5	9	0	0	1	978	0	0	0	5	0	6	0	0	1	0	0
m_6	1	10	14	0	0	967	0	0	2	0	0	5	1	0	0	0
m_7	0	13	0	12	0	0	971	0	0	0	0	0	0	4	0	0
m_8	1	0	12	16	0	0	0	964	0	0	3	4	0	0	0	0
m_9	1	11	0	0	13	0	0	0	971	0	1	0	0	3	0	0
m_{10}	1	0	8	0	6	0	0	0	1	968	0	0	8	0	7	1
m_{11}	1	0	0	11	18	0	0	0	0	0	964	0	0	6	0	0
m_{12}	0	0	0	0	0	11	20	12	0	0	0	956	0	1	0	0
m_{13}	0	1	0	0	1	16	0	0	7	12	0	0	953	0	0	10
m_{14}	0	0	0	0	0	0	9	0	10	0	10	0	0	971	0	0
m_{15}	0	0	0	0	0	0	0	9	0	10	10	0	0	0	967	4
m_{16}	0	0	0	1	0	0	0	0	0	0	0	13	11	12	16	947

Tabela E.12: Número de vezes que cada modelo é seleccionado pelo LLP, por modelo gerador dos dados e modelo ajustado, para o caso 3.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	1000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_2	308	692	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_3	311	0	689	0	0	0	0	0	0	0	0	0	0	0	0	0
m_4	302	0	0	698	0	0	0	0	0	0	0	0	0	0	0	0
m_5	323	0	0	0	677	0	0	0	0	0	0	0	0	0	0	0
m_6	100	208	227	0	0	465	0	0	0	0	0	0	0	0	0	0
m_7	103	234	0	225	0	0	438	0	0	0	0	0	0	0	0	0
m_8	93	0	200	219	0	0	0	488	0	0	0	0	0	0	0	0
m_9	110	213	0	0	217	0	0	0	460	0	0	0	0	0	0	0
m_{10}	123	0	196	0	210	0	0	0	0	471	0	0	0	0	0	0
m_{11}	119	0	0	203	198	0	0	0	0	0	480	0	0	0	0	0
m_{12}	31	77	64	66	0	130	162	149	0	0	0	321	0	0	0	0
m_{13}	34	84	64	0	59	137	0	0	133	148	0	0	341	0	0	0
m_{14}	32	64	0	72	80	0	152	0	147	0	133	0	0	320	0	0
m_{15}	38	0	65	87	55	0	0	142	0	161	120	0	0	0	332	0
m_{16}	12	20	30	20	16	48	46	52	43	49	36	114	99	84	103	228

Tabela E.13: Número de vezes que cada modelo é seleccionado pelo DIC, por modelo gerador dos dados e modelo ajustado, para o caso 4.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	999	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
m_2	161	839	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_3	179	0	821	0	0	0	0	0	0	0	0	0	0	0	0	0
m_4	155	0	0	845	0	0	0	0	0	0	0	0	0	0	0	0
m_5	154	0	0	0	846	0	0	0	0	0	0	0	0	0	0	0
m_6	31	129	141	0	0	699	0	0	0	0	0	0	0	0	0	0
m_7	0	3	13	1	20	64	6	81	70	0	63	329	0	350	0	0
m_8	18	0	106	69	54	0	0	518	0	0	235	0	0	0	0	0
m_9	13	64	55	0	93	252	0	0	523	0	0	0	0	0	0	0
m_{10}	66	0	79	0	83	0	0	0	0	772	0	0	0	0	0	0
m_{11}	23	0	41	66	101	0	0	235	0	0	534	0	0	0	0	0
m_{12}	4	15	14	9	6	94	58	76	38	0	40	442	0	204	0	0
m_{13}	16	60	12	0	13	72	0	0	64	118	0	0	645	0	0	0
m_{14}	3	11	3	12	14	35	65	42	100	0	76	178	0	461	0	0
m_{15}	12	0	11	68	14	0	0	63	0	116	62	0	0	0	654	0
m_{16}	2	13	3	10	4	13	42	12	6	21	9	53	102	61	104	545

Tabela E.14: Número de vezes que cada modelo é seleccionado pelo $\widehat{\text{BIC}}$, por modelo gerador dos dados e modelo ajustado, para o caso 4.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	982	3	0	9	5	0	1	0	0	0	0	0	0	0	0	0
m_2	15	964	0	1	1	0	8	0	11	0	0	0	0	0	0	0
m_3	10	0	973	0	1	6	0	10	0	0	0	0	0	0	0	0
m_4	14	0	0	976	0	0	2	0	0	0	8	0	0	0	0	0
m_5	15	0	0	0	972	0	0	0	6	0	5	0	0	2	0	0
m_6	1	12	12	0	0	964	0	1	1	0	0	9	0	0	0	0
m_7	0	0	0	0	0	7	0	7	8	0	7	478	0	493	0	0
m_8	0	0	6	5	5	0	0	694	0	0	285	2	0	2	1	0
m_9	0	5	2	0	10	314	0	0	663	0	0	0	0	6	0	0
m_{10}	0	0	11	0	12	0	0	0	0	968	0	0	3	0	6	0
m_{11}	1	0	6	10	13	0	1	276	0	0	690	1	0	2	0	0
m_{12}	0	0	0	0	0	11	5	7	4	0	3	686	0	284	0	0
m_{13}	0	0	0	0	0	9	0	0	9	16	0	0	958	0	1	7
m_{14}	0	0	0	0	0	4	6	6	10	0	7	257	0	710	0	0
m_{15}	0	0	0	1	1	0	0	7	0	14	3	0	0	0	971	3
m_{16}	0	0	0	0	0	0	1	0	0	0	0	13	10	6	23	947

Tabela E.15: Número de vezes que cada modelo é seleccionado pelo LLP, por modelo gerador dos dados e modelo ajustado, para o caso 4.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	1000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_2	299	701	0	0	0	0	0	0	0	0	0	0	0	0	0	0
m_3	344	0	656	0	0	0	0	0	0	0	0	0	0	0	0	0
m_4	315	0	0	685	0	0	0	0	0	0	0	0	0	0	0	0
m_5	311	0	0	0	689	0	0	0	0	0	0	0	0	0	0	0
m_6	100	202	218	0	0	480	0	0	0	0	0	0	0	0	0	0
m_7	14	14	52	21	55	97	44	108	96	0	95	185	0	219	0	0
m_8	59	0	174	125	82	0	0	367	0	0	193	0	0	0	0	0
m_9	57	113	91	0	170	197	0	0	372	0	0	0	0	0	0	0
m_{10}	219	0	116	0	124	0	0	0	0	541	0	0	0	0	0	0
m_{11}	68	0	86	123	173	0	0	182	0	0	368	0	0	0	0	0
m_{12}	15	37	66	35	27	116	74	107	61	0	69	272	0	121	0	0
m_{13}	61	137	44	0	41	79	0	0	91	172	0	0	375	0	0	0
m_{14}	15	42	24	49	54	49	81	48	128	0	128	117	0	265	0	0
m_{15}	52	0	32	150	42	0	0	94	0	163	81	0	0	0	386	0
m_{16}	25	43	12	42	13	21	91	34	31	58	26	60	119	59	112	254

Tabela E.16: Número de vezes que cada modelo é seleccionado pelo DIC, por modelo gerador dos dados e modelo ajustado, para o caso 5.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	507	28	71	172	63	5	45	26	15	10	23	8	3	20	2	2
m_2	105	472	9	48	17	71	135	7	82	2	10	17	8	14	1	2
m_3	97	9	472	21	19	26	7	141	1	89	0	61	13	5	15	24
m_4	121	3	15	571	14	1	33	94	8	2	102	2	0	26	5	3
m_5	98	12	9	29	536	0	12	8	20	50	146	1	1	46	25	7
m_6	25	112	93	8	2	399	31	46	13	19	0	145	76	3	12	16
m_7	26	117	2	92	4	13	549	8	20	0	16	67	1	75	3	7
m_8	24	4	94	116	3	18	6	562	3	13	19	33	8	5	73	19
m_9	16	101	1	6	94	15	22	1	476	9	38	2	64	128	6	21
m_{10}	22	1	121	2	82	9	1	22	3	475	18	12	21	10	141	60
m_{11}	28	4	1	106	97	0	11	8	18	16	572	1	2	34	96	6
m_{12}	9	22	23	20	0	102	101	111	2	3	6	483	8	20	14	76
m_{13}	6	22	16	2	13	94	7	8	90	81	5	28	425	21	40	142
m_{14}	5	24	0	23	35	3	102	2	102	2	102	13	16	485	14	72
m_{15}	5	0	20	20	23	3	2	115	1	94	115	9	19	5	527	42
m_{16}	4	8	8	7	8	23	17	17	17	36	15	96	94	88	93	469

Tabela E.17: Número de vezes que cada modelo é seleccionado pelo $\widehat{\text{BIC}}$, por modelo gerador dos dados e modelo ajustado, para o caso 5.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	136	8	75	108	62	4	90	61	34	22	37	73	7	155	22	106
m_2	9	158	4	19	2	93	187	3	123	1	17	175	64	91	2	52
m_3	6	0	184	3	2	28	4	139	1	79	1	154	37	6	51	305
m_4	14	4	1	273	5	2	52	166	11	1	169	33	3	143	47	76
m_5	5	5	5	4	216	0	12	4	21	106	165	2	11	227	77	140
m_6	0	12	10	2	1	250	15	14	6	5	2	320	177	3	19	164
m_7	1	9	1	8	0	8	374	6	10	0	8	221	4	251	5	94
m_8	1	0	12	11	0	12	3	439	0	7	11	87	8	3	171	235
m_9	0	11	0	1	10	5	9	0	279	1	19	7	166	304	6	182
m_{10}	0	0	12	0	11	5	0	10	2	305	5	13	25	4	242	366
m_{11}	1	1	0	14	11	0	11	7	15	6	436	4	7	128	253	106
m_{12}	0	1	1	0	1	20	22	22	0	0	1	557	2	13	17	343
m_{13}	0	0	0	0	0	13	0	1	21	14	0	8	420	15	31	477
m_{14}	0	1	0	2	2	1	26	0	21	2	20	9	11	556	11	338
m_{15}	0	0	1	2	1	1	0	23	1	22	14	6	21	4	672	232
m_{16}	0	1	0	0	0	2	1	3	2	2	1	31	30	29	29	869

Tabela E.18: Número de vezes que cada modelo é seleccionado pelo LLP, por modelo gerador dos dados e modelo ajustado, para o caso 5.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	706	18	48	132	38	1	23	9	7	2	11	2	0	3	0	0
m_2	232	519	11	53	18	38	67	3	42	1	5	7	2	2	0	0
m_3	231	7	532	26	17	15	8	92	0	41	0	20	1	0	7	3
m_4	259	1	20	580	12	0	20	39	8	0	51	1	0	5	4	0
m_5	238	10	15	34	555	0	9	4	9	28	75	1	0	12	10	0
m_6	85	195	173	11	9	338	25	45	13	17	1	54	25	2	4	3
m_7	88	207	2	177	10	8	417	9	10	0	18	23	1	30	0	0
m_8	84	8	173	205	7	14	5	437	2	13	16	4	1	2	26	3
m_9	62	195	4	15	185	11	20	1	380	12	40	1	23	48	2	1
m_{10}	86	2	214	13	152	6	3	18	4	389	24	7	11	5	55	11
m_{11}	96	6	2	176	183	0	7	15	14	15	439	0	1	13	32	1
m_{12}	38	66	70	63	2	110	140	134	1	7	7	306	8	15	6	27
m_{13}	22	62	56	8	58	141	11	10	122	118	12	21	265	18	32	44
m_{14}	31	64	2	71	74	2	134	7	144	5	129	9	8	292	9	19
m_{15}	22	1	65	73	72	6	2	130	1	146	138	5	10	4	318	7
m_{16}	13	25	37	22	23	50	43	41	48	55	42	92	107	99	93	210

Tabela E.19: Número de vezes que cada modelo é seleccionado pelo DIC, por modelo gerador dos dados e modelo ajustado, para o caso 6.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	964	14	2	13	4	0	2	0	1	0	0	0	0	0	0	0
m_2	174	804	3	3	1	2	8	0	4	0	1	0	0	0	0	0
m_3	163	1	802	0	2	9	0	16	0	4	0	3	0	0	0	0
m_4	152	5	0	829	1	0	3	6	0	0	3	0	0	0	1	0
m_5	149	2	0	3	821	0	0	0	12	2	10	0	0	0	1	0
m_6	29	134	139	0	0	681	0	6	0	1	0	7	3	0	0	0
m_7	35	121	0	147	0	1	683	0	1	0	2	4	0	6	0	0
m_8	31	0	145	140	1	4	1	666	0	1	1	7	0	0	3	0
m_9	36	113	1	2	139	2	1	0	696	0	5	1	0	3	1	0
m_{10}	31	2	138	0	137	3	0	6	1	661	3	0	5	0	12	1
m_{11}	29	0	0	126	127	0	3	0	4	1	703	0	0	5	2	0
m_{12}	6	18	27	25	0	94	107	113	0	0	1	604	0	1	1	3
m_{13}	4	22	24	0	24	120	0	0	113	116	0	1	566	3	6	1
m_{14}	7	21	0	25	27	0	109	0	129	0	115	0	0	564	1	2
m_{15}	3	0	28	32	18	0	1	137	0	103	119	1	2	1	555	0
m_{16}	2	5	7	6	8	23	16	18	18	21	16	99	95	82	94	490

Tabela E.20: Número de vezes que cada modelo é seleccionado pelo $\widehat{\text{BIC}}$, por modelo gerador dos dados e modelo ajustado, para o caso 6.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	722	48	22	54	31	12	48	3	11	3	8	11	2	15	3	7
m_2	15	804	2	5	1	35	83	1	28	0	1	8	4	10	0	3
m_3	24	1	713	1	1	54	2	75	2	31	1	56	12	1	5	21
m_4	15	5	0	800	0	0	65	38	1	0	43	6	1	8	7	11
m_5	12	2	0	2	757	2	2	0	29	31	71	1	8	60	6	17
m_6	0	13	18	0	0	813	2	2	1	0	0	99	44	0	2	6
m_7	1	15	0	20	0	1	860	0	1	0	1	38	0	47	1	15
m_8	0	0	14	17	0	4	2	834	0	2	1	68	1	1	42	14
m_9	0	14	0	2	16	0	2	0	833	3	7	1	31	86	0	5
m_{10}	0	1	12	0	12	2	0	2	2	770	1	4	42	0	85	67
m_{11}	0	0	0	17	12	0	3	0	6	0	870	0	0	49	33	10
m_{12}	0	0	0	0	0	13	20	18	0	0	0	900	0	2	3	44
m_{13}	0	0	0	0	1	17	0	0	13	13	0	2	850	4	7	93
m_{14}	0	0	0	0	0	0	17	0	15	0	18	0	1	895	0	54
m_{15}	0	0	0	0	0	0	0	20	0	23	18	3	2	2	876	56
m_{16}	0	0	0	0	0	0	0	1	1	2	0	21	20	16	18	921

Tabela E.21: Número de vezes que cada modelo é seleccionado pelo LLP, por modelo gerador dos dados e modelo ajustado, para o caso 6.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	986	4	1	7	2	0	0	0	0	0	0	0	0	0	0	0
m_2	301	693	1	3	0	0	0	0	2	0	0	0	0	0	0	0
m_3	315	2	671	2	2	2	0	4	0	1	0	1	0	0	0	0
m_4	302	1	0	693	0	0	2	0	0	0	2	0	0	0	0	0
m_5	300	1	0	4	688	0	0	0	3	2	2	0	0	0	0	0
m_6	115	211	233	1	0	436	0	2	0	0	0	1	1	0	0	0
m_7	113	197	0	223	1	1	461	0	1	0	2	1	0	0	0	0
m_8	92	3	232	228	0	2	0	439	0	0	2	1	0	0	1	0
m_9	118	183	0	1	225	1	1	0	467	0	3	0	0	1	0	0
m_{10}	116	2	225	1	226	2	0	2	0	423	1	0	0	0	1	1
m_{11}	106	0	0	221	203	0	0	0	2	1	467	0	0	0	0	0
m_{12}	35	62	81	66	0	132	153	147	0	0	0	324	0	0	0	0
m_{13}	30	73	85	0	67	142	2	3	127	143	0	0	327	1	0	0
m_{14}	42	60	0	73	73	0	123	0	154	0	153	0	0	321	0	1
m_{15}	40	0	82	91	62	0	0	146	0	151	134	0	1	1	292	0
m_{16}	12	26	24	27	27	46	45	40	45	48	43	94	100	98	101	224

Tabela E.22: Médias do DIC por modelo gerador dos dados e modelo ajustado, para o caso 1.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	941	963	966	963	971	985	977	984	988	991	988	996	1005	998	1004	1014
m_2	938	937	965	961	968	964	961	983	967	989	985	983	988	985	1003	1002
m_3	941	964	940	963	971	963	977	962	989	970	987	976	988	998	986	997
m_4	941	965	967	940	971	987	965	966	989	991	970	986	1006	989	990	1006
m_5	940	963	966	963	939	985	977	985	962	965	962	996	984	975	984	995
m_6	939	938	938	961	970	937	961	960	969	969	986	960	968	986	985	985
m_7	940	939	966	939	970	965	938	965	969	990	969	964	989	968	989	988
m_8	940	963	939	939	970	962	963	938	988	969	969	962	987	988	968	987
m_9	941	940	967	963	940	966	963	985	939	966	962	985	965	962	984	984
m_{10}	941	964	940	963	940	963	977	962	963	939	962	976	962	976	961	975
m_{11}	940	963	966	939	939	985	963	965	962	965	938	985	984	962	964	984
m_{12}	940	939	939	939	970	938	938	938	969	969	969	937	968	968	968	967
m_{13}	940	939	939	962	939	938	962	961	938	938	961	960	937	961	960	959
m_{14}	940	939	966	939	939	965	938	965	938	965	938	964	964	937	964	963
m_{15}	940	963	939	939	939	962	962	938	962	938	938	961	961	961	937	960
m_{16}	940	939	939	939	939	938	938	938	938	938	938	937	937	937	937	936

Tabela E.23: Desvios padrão do DIC por modelo gerador dos dados e modelo ajustado, para o caso 1.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	20.0	19.9	20.0	20.1	19.7	19.6	19.9	20.0	19.4	19.7	19.7	19.6	19.3	19.4	19.7	19.3
m_2	20.9	20.8	20.6	20.6	20.5	20.5	20.5	20.2	20.4	20.0	20.0	20.1	20.0	19.9	19.6	19.5
m_3	20.7	20.7	20.6	20.3	20.1	20.7	20.3	20.2	19.9	20.0	19.8	20.3	19.9	19.7	19.7	19.6
m_4	20.0	20.0	19.3	19.9	20.2	19.1	19.8	19.2	19.9	19.3	20.0	18.9	19.0	19.7	19.2	18.8
m_5	20.2	19.6	19.9	20.2	20.1	19.4	19.6	19.8	19.5	19.9	20.1	19.4	19.4	19.6	19.8	19.3
m_6	21.0	20.9	20.9	20.8	20.5	20.8	20.7	20.7	20.4	20.4	20.2	20.6	20.3	20.2	20.2	20.1
m_7	19.9	19.9	20.1	19.8	19.9	20.1	19.8	20.0	19.9	19.7	19.8	20.0	19.7	19.8	19.6	19.6
m_8	19.7	19.9	19.7	19.7	19.7	19.8	19.9	19.7	19.6	19.7	19.7	19.8	19.6	19.6	19.7	19.5
m_9	19.7	19.6	19.5	19.4	19.7	19.4	19.4	19.2	19.6	19.5	19.4	19.1	19.4	19.3	19.2	19.1
m_{10}	20.4	20.2	20.4	20.3	20.3	20.2	19.9	20.3	20.2	20.3	20.2	19.8	20.1	19.8	20.2	19.8
m_{11}	20.3	20.1	20.0	20.3	20.3	19.8	20.0	20.0	20.1	20.0	20.3	19.7	19.7	20.0	20.0	19.6
m_{12}	19.8	19.8	19.9	19.8	20.1	19.9	19.8	19.8	20.0	20.1	20.1	19.8	20.1	20.1	20.1	20.1
m_{13}	20.0	19.9	19.9	19.4	19.8	19.9	19.3	19.4	19.8	19.8	19.3	19.3	19.8	19.3	19.3	19.2
m_{14}	20.8	20.7	20.9	20.7	20.7	20.8	20.6	20.8	20.6	20.8	20.6	20.7	20.7	20.5	20.7	20.6
m_{15}	20.3	19.9	20.3	20.3	20.3	19.9	19.9	20.3	19.9	20.2	20.3	19.8	19.8	19.9	20.2	19.8
m_{16}	20.8	20.8	20.7	20.8	20.8	20.7	20.8	20.7	20.8	20.7	20.8	20.7	20.7	20.7	20.7	20.6

Tabela E.24: Médias do $\widehat{\text{BIC}}$ por modelo gerador dos dados e modelo ajustado, para o caso 1.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	961	979	982	979	987	997	989	996	1000	1003	1000	1004	1013	1006	1012	1017
m_2	959	953	981	977	984	976	972	995	979	1001	997	990	995	992	1010	1006
m_3	962	980	956	979	988	975	989	974	1001	982	999	984	995	1006	994	1001
m_4	962	981	983	957	987	999	977	978	1001	1003	982	994	1014	997	997	1010
m_5	961	979	983	979	955	997	988	996	974	977	974	1004	992	983	991	998
m_6	960	954	954	978	986	949	973	972	981	981	998	968	976	994	993	988
m_7	961	955	982	955	987	977	950	977	981	1002	981	972	997	976	997	991
m_8	961	980	955	955	986	974	975	950	1000	981	981	970	994	995	976	990
m_9	961	956	983	979	956	978	975	997	951	978	974	992	972	969	992	987
m_{10}	961	980	956	979	956	975	989	974	975	951	974	984	969	984	969	978
m_{11}	960	979	982	955	955	997	975	977	974	977	950	992	992	969	972	987
m_{12}	961	956	956	956	987	950	950	950	981	981	981	945	976	976	976	971
m_{13}	961	955	955	978	955	950	973	973	950	950	973	968	945	968	967	963
m_{14}	960	955	982	955	955	977	950	977	950	977	950	971	972	944	972	966
m_{15}	960	979	955	955	955	974	974	950	974	950	950	969	968	969	944	963
m_{16}	961	955	955	955	955	950	950	950	950	950	950	945	945	945	945	939

Tabela E.25: Médias do LLP por modelo gerador dos dados e modelo ajustado, para o caso 1.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	935	958	961	958	966	981	973	980	984	987	984	993	1002	995	1001	1012
m_2	932	932	960	956	963	960	957	979	963	985	981	980	985	982	1000	1000
m_3	935	959	935	958	966	959	973	958	985	966	983	973	985	995	983	995
m_4	935	960	962	935	966	983	961	962	985	987	966	983	1003	986	987	1004
m_5	934	958	961	958	934	981	973	981	958	961	958	993	981	973	981	993
m_6	933	933	933	956	965	933	957	956	965	965	982	957	965	983	982	983
m_7	934	934	961	934	965	961	934	961	965	986	965	961	986	965	986	986
m_8	934	958	934	934	965	958	959	934	984	965	965	959	984	985	965	985
m_9	935	935	962	958	935	962	959	981	935	962	958	982	962	959	981	982
m_{10}	935	959	935	958	935	959	973	958	959	935	958	973	959	973	958	973
m_{11}	934	958	961	934	934	981	959	961	958	961	934	982	981	959	961	982
m_{12}	934	934	934	934	965	934	934	934	965	965	965	934	965	965	965	965
m_{13}	934	934	934	957	934	934	958	957	934	934	957	957	934	958	957	957
m_{14}	934	934	961	934	934	961	934	961	934	961	934	961	961	934	961	961
m_{15}	934	958	934	934	934	958	958	934	958	934	934	958	958	958	934	958
m_{16}	934	934	934	934	934	934	934	934	934	934	934	934	934	934	934	934

Tabela E.26: Médias do DIC por modelo gerador dos dados e modelo ajustado, para o caso 2.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	939	1225	1030	1458	969	1247	1475	1462	1228	1046	1456	1479	1249	1474	1461	1477
m_2	940	939	1032	1458	971	1030	1462	1463	970	1047	1457	1467	1046	1461	1461	1465
m_3	941	1228	940	1458	971	1226	1475	1456	1231	970	1456	1474	1229	1474	1455	1472
m_4	941	1227	1031	940	971	1249	1231	1030	1231	1047	970	1252	1251	1234	1046	1254
m_5	941	1227	1031	1458	940	1248	1475	1462	1225	1030	1457	1479	1247	1474	1461	1478
m_6	941	940	940	1458	971	939	1462	1456	970	970	1456	1460	969	1461	1455	1459
m_7	940	939	1031	939	971	1030	938	1030	970	1047	970	1029	1046	969	1046	1045
m_8	939	1227	938	938	971	1226	1231	937	1231	970	970	1229	1229	1235	969	1233
m_9	941	940	1031	1458	940	1030	1462	1462	939	1030	1457	1467	1028	1461	1461	1465
m_{10}	940	1227	939	1458	939	1225	1476	1456	1225	938	1457	1474	1224	1475	1455	1473
m_{11}	941	1226	1032	940	940	1248	1230	1030	1225	1030	939	1251	1247	1228	1029	1250
m_{12}	940	939	939	939	970	938	938	938	969	969	969	937	968	968	968	967
m_{13}	940	939	939	1458	939	938	1462	1456	938	938	1457	1461	937	1461	1455	1459
m_{14}	939	938	1030	938	938	1028	937	1028	937	1029	937	1027	1027	936	1027	1026
m_{15}	939	1227	938	938	938	1225	1230	937	1225	937	937	1228	1224	1229	936	1227
m_{16}	940	939	939	939	939	938	938	938	938	938	938	937	937	937	937	936

Tabela E.27: Desvios padrão do DIC por modelo gerador dos dados e modelo ajustado, para o caso 2.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	20.4	13.3	19.1	7.6	20.3	12.8	7.3	7.4	13.1	18.6	7.6	7.2	12.6	7.3	7.4	7.2
m_2	20.4	20.4	19.3	7.8	20.0	19.3	7.6	7.6	20.0	18.9	7.8	7.4	18.9	7.6	7.6	7.4
m_3	20.6	13.0	20.5	8.0	20.4	13.0	7.6	8.0	12.9	20.3	8.0	7.6	12.8	7.6	8.0	7.6
m_4	20.6	12.9	19.3	20.5	20.1	12.5	12.8	19.3	12.8	18.7	20.1	12.5	12.5	12.8	18.6	12.4
m_5	19.6	13.6	18.2	7.7	19.5	12.8	7.4	7.5	13.7	18.2	7.6	7.3	12.8	7.4	7.5	7.3
m_6	19.9	19.8	19.7	8.2	20.1	19.6	8.0	8.2	20.0	20.0	8.2	8.0	19.9	8.0	8.2	8.0
m_7	20.1	20.1	18.8	20.2	20.2	18.7	20.2	18.8	20.2	18.4	20.2	18.8	18.3	20.2	18.3	18.3
m_8	20.2	13.6	20.2	20.2	20.1	13.6	13.4	20.2	13.3	20.1	20.1	13.4	13.3	13.1	20.1	13.1
m_9	20.3	20.3	18.9	7.5	20.2	18.9	7.4	7.4	20.1	18.8	7.5	7.3	18.8	7.4	7.4	7.3
m_{10}	20.0	13.8	20.0	7.4	20.0	13.8	7.2	7.4	13.8	20.0	7.4	7.2	13.7	7.2	7.4	7.2
m_{11}	20.3	13.2	19.4	20.3	20.2	12.9	13.1	19.3	13.2	19.3	20.2	12.7	12.8	13.1	19.2	12.7
m_{12}	21.2	21.1	21.1	21.2	21.0	21.1	21.1	21.1	20.9	20.9	21.0	21.0	20.8	20.9	20.9	20.8
m_{13}	20.0	19.9	20.0	7.5	20.0	19.9	7.4	7.5	19.9	20.0	7.5	7.4	19.9	7.4	7.5	7.4
m_{14}	19.7	19.7	18.1	19.7	19.7	18.1	19.7	18.1	19.7	18.0	19.6	18.1	18.0	19.6	18.0	18.0
m_{15}	20.4	13.0	20.3	20.4	20.3	13.0	12.8	20.2	13.0	20.2	20.3	12.8	12.9	12.8	20.2	12.8
m_{16}	20.2	20.2	20.2	20.2	20.2	20.1	20.1	20.2	20.1	20.1	20.1	20.1	20.1	20.0	20.1	20.0

Tabela E.28: Médias do $\widehat{\text{BIC}}$ por modelo gerador dos dados e modelo ajustado, para o caso 2.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	959	1241	1047	1474	985	1259	1487	1474	1240	1058	1468	1486	1257	1481	1468	1480
m_2	961	955	1048	1474	987	1042	1474	1474	982	1059	1469	1474	1054	1469	1469	1469
m_3	962	1244	957	1474	988	1238	1487	1468	1243	982	1468	1481	1237	1481	1462	1475
m_4	961	1244	1048	956	987	1261	1243	1042	1243	1059	982	1260	1258	1242	1053	1258
m_5	962	1243	1047	1474	957	1260	1487	1474	1237	1041	1468	1486	1254	1482	1469	1481
m_6	961	956	956	1474	987	951	1474	1468	982	982	1468	1468	976	1469	1462	1462
m_7	961	955	1048	955	987	1042	950	1042	982	1059	982	1036	1054	976	1054	1048
m_8	960	1244	955	955	987	1238	1243	949	1243	981	982	1237	1237	1242	976	1236
m_9	961	956	1047	1474	956	1042	1474	1474	951	1042	1468	1474	1036	1468	1469	1469
m_{10}	961	1243	955	1474	955	1237	1488	1468	1237	950	1469	1482	1231	1482	1463	1476
m_{11}	961	1243	1048	956	956	1260	1242	1042	1237	1042	951	1259	1255	1236	1037	1253
m_{12}	960	955	955	955	986	950	950	950	981	981	981	944	976	976	976	970
m_{13}	961	955	955	1474	955	950	1474	1468	950	950	1469	1468	945	1469	1463	1462
m_{14}	960	955	1046	955	955	1040	949	1040	949	1041	949	1035	1035	944	1035	1029
m_{15}	960	1243	954	954	954	1237	1242	949	1237	949	949	1236	1231	1236	944	1230
m_{16}	961	956	956	956	956	950	950	950	950	950	950	945	945	945	945	940

Tabela E.29: Médias do LLP por modelo gerador dos dados e modelo ajustado, para o caso 2.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	933	1220	1026	1453	964	1243	1471	1458	1224	1042	1452	1476	1246	1471	1458	1475
m_2	934	934	1027	1453	966	1026	1458	1459	966	1043	1453	1464	1043	1458	1458	1463
m_3	935	1223	935	1453	966	1222	1471	1452	1227	966	1452	1471	1226	1471	1452	1470
m_4	935	1222	1026	935	966	1245	1227	1026	1227	1043	966	1249	1248	1231	1043	1252
m_5	935	1222	1026	1453	935	1244	1471	1458	1221	1026	1453	1476	1244	1471	1458	1476
m_6	935	935	935	1453	966	935	1458	1452	966	966	1452	1457	966	1458	1452	1457
m_7	934	934	1026	934	966	1026	934	1026	966	1043	966	1026	1043	966	1043	1043
m_8	933	1222	933	933	966	1222	1227	934	1227	966	966	1226	1226	1232	966	1231
m_9	935	935	1026	1453	935	1026	1458	1458	935	1026	1453	1464	1025	1458	1458	1463
m_{10}	934	1222	934	1453	934	1221	1472	1452	1221	934	1453	1471	1221	1472	1452	1471
m_{11}	935	1221	1027	935	935	1244	1226	1026	1221	1026	935	1248	1244	1225	1026	1248
m_{12}	934	934	934	934	965	934	934	934	965	965	965	934	965	965	965	965
m_{13}	934	934	934	1453	934	934	1458	1452	934	934	1453	1458	934	1458	1452	1457
m_{14}	933	933	1025	933	933	1024	933	1024	933	1025	933	1024	1024	933	1024	1024
m_{15}	933	1222	933	933	933	1221	1226	933	1221	933	933	1225	1221	1226	933	1225
m_{16}	934	934	934	934	934	934	934	934	934	934	934	934	934	934	934	934

Tabela E.30: Médias do DIC por modelo gerador dos dados e modelo ajustado, para o caso 3.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	940	963	1010	963	963	1024	977	1024	982	1082	981	1033	1090	993	1089	1093
m_2	940	939	1009	962	963	1007	961	1022	962	1081	981	1021	1080	981	1087	1087
m_3	940	963	939	963	963	962	977	961	982	969	982	976	987	993	986	997
m_4	940	963	1009	939	962	1023	962	1007	981	1080	961	1023	1088	981	1079	1088
m_5	940	963	1010	963	939	1024	977	1024	962	1028	962	1033	1040	976	1039	1046
m_6	940	939	939	962	963	938	962	961	962	970	981	961	969	981	986	986
m_7	940	939	1009	939	963	1007	938	1007	962	1081	962	1006	1080	961	1080	1079
m_8	940	963	939	939	962	962	963	938	982	969	961	962	987	982	968	986
m_9	939	938	1008	962	938	1007	961	1022	938	1027	960	1021	1025	960	1037	1037
m_{10}	939	962	938	961	938	961	975	960	961	937	960	974	960	974	959	973
m_{11}	940	963	1010	939	939	1024	963	1009	962	1028	938	1024	1040	962	1027	1039
m_{12}	940	939	939	939	962	938	938	938	961	969	961	937	968	960	968	967
m_{13}	940	939	939	962	939	938	962	961	938	938	961	961	937	961	960	960
m_{14}	940	939	1009	939	939	1008	938	1008	938	1028	938	1007	1026	937	1026	1025
m_{15}	941	965	940	940	940	964	964	939	964	939	939	963	962	963	938	962
m_{16}	940	939	939	939	939	938	938	938	938	938	938	937	937	937	937	936

Tabela E.31: Médias do $\widehat{\text{BIC}}$ por modelo gerador dos dados e modelo ajustado, para o caso 3.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	961	979	1026	979	979	1036	989	1035	994	1094	993	1040	1097	1001	1096	1096
m_2	960	955	1025	978	979	1019	973	1034	974	1093	993	1029	1088	988	1095	1090
m_3	960	979	955	979	979	974	989	973	994	981	993	983	994	1001	993	1000
m_4	960	979	1025	955	978	1035	974	1019	993	1092	973	1031	1096	989	1087	1091
m_5	961	979	1026	979	955	1036	989	1035	974	1040	973	1040	1047	983	1046	1049
m_6	960	955	955	978	979	950	974	973	974	981	993	968	976	988	993	989
m_7	960	955	1025	955	979	1019	950	1019	974	1093	974	1014	1088	969	1088	1082
m_8	960	979	955	955	979	974	975	950	994	981	973	969	994	989	975	990
m_9	960	955	1025	978	955	1019	973	1034	949	1038	972	1029	1033	968	1045	1040
m_{10}	959	978	954	977	954	973	987	972	973	949	972	982	968	982	966	976
m_{11}	961	980	1026	955	955	1036	975	1020	974	1040	950	1032	1047	970	1034	1043
m_{12}	960	955	955	955	979	950	950	950	973	981	973	945	976	968	976	970
m_{13}	961	955	955	979	955	950	974	973	950	950	973	969	945	969	968	963
m_{14}	961	955	1026	955	955	1020	950	1020	950	1039	950	1015	1034	945	1034	1028
m_{15}	962	981	957	957	957	975	976	951	975	951	951	971	970	971	946	965
m_{16}	961	955	955	955	955	950	950	950	950	950	950	945	945	945	945	939

Tabela E.32: Médias do LLP por modelo gerador dos dados e modelo ajustado, para o caso 3.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	934	958	1005	958	958	1020	973	1020	978	1078	978	1030	1087	990	1086	1091
m_2	934	934	1004	957	958	1003	957	1018	958	1077	977	1018	1077	978	1085	1085
m_3	934	958	934	958	958	958	973	957	978	965	978	973	984	991	983	995
m_4	934	958	1004	934	957	1019	959	1003	978	1076	957	1020	1085	978	1076	1086
m_5	934	958	1005	958	934	1020	973	1020	958	1024	958	1030	1037	973	1036	1044
m_6	934	934	934	957	958	934	958	957	958	966	977	958	966	978	983	984
m_7	934	934	1004	934	958	1003	934	1003	958	1077	958	1003	1077	958	1077	1077
m_8	934	958	934	934	957	958	959	934	978	965	957	959	984	979	965	984
m_9	933	933	1003	957	934	1003	957	1018	934	1023	956	1018	1022	957	1034	1035
m_{10}	933	957	933	956	933	957	971	956	957	933	956	971	957	971	956	971
m_{11}	934	958	1005	934	934	1020	959	1005	958	1024	934	1021	1037	959	1024	1037
m_{12}	934	934	934	934	957	934	934	934	957	965	957	934	965	957	965	965
m_{13}	934	934	934	957	934	934	958	957	934	934	957	958	934	958	957	958
m_{14}	934	934	1004	934	934	1004	934	1004	934	1024	934	1004	1023	934	1023	1023
m_{15}	935	960	935	935	935	960	960	935	960	935	935	960	959	960	935	960
m_{16}	934	934	934	934	934	934	934	934	934	934	934	934	934	934	934	934

Tabela E.33: Médias do DIC por modelo gerador dos dados e modelo ajustado, para o caso 4.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	941	965	1010	963	964	1025	978	1023	983	1082	982	1033	1090	994	1088	1093
m_2	940	939	1009	962	963	1008	962	1023	961	1081	981	1022	1080	981	1088	1087
m_3	941	964	940	963	964	963	977	962	983	971	982	976	988	994	987	998
m_4	939	962	1008	938	962	1023	962	1007	981	1081	961	1023	1088	981	1080	1088
m_5	940	963	1009	962	939	1023	976	1023	962	1027	961	1032	1039	975	1038	1045
m_6	941	940	940	963	963	939	963	962	962	970	982	961	969	981	986	986
m_7	940	939	940	939	940	939	938	939	939	1042	939	938	1041	938	1041	1040
m_8	940	963	938	939	940	962	962	937	963	969	939	961	986	962	968	986
m_9	941	940	941	963	940	940	963	963	939	971	962	963	969	962	987	987
m_{10}	940	964	939	962	939	962	977	961	962	938	961	976	961	975	960	974
m_{11}	941	964	941	940	940	964	963	940	963	970	939	963	988	962	969	988
m_{12}	940	939	939	939	940	938	938	938	939	970	939	937	969	938	969	968
m_{13}	938	937	937	961	937	936	960	959	936	936	959	959	935	959	958	958
m_{14}	941	940	941	940	940	940	939	940	939	971	939	939	970	938	970	969
m_{15}	940	963	939	939	939	962	963	938	962	938	938	962	961	962	937	961
m_{16}	941	940	940	940	940	939	939	939	939	939	939	938	938	938	938	937

Tabela E.34: Médias do $\widehat{\text{BIC}}$ por modelo gerador dos dados e modelo ajustado, para o caso 4.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	961	981	1026	979	980	1037	990	1035	995	1094	994	1041	1097	1002	1096	1097
m_2	960	955	1025	978	979	1020	974	1035	973	1093	993	1030	1087	988	1095	1090
m_3	962	980	956	979	981	975	989	974	995	983	994	983	996	1002	995	1001
m_4	960	978	1025	954	978	1035	974	1019	993	1093	973	1030	1096	988	1087	1091
m_5	960	979	1025	979	955	1035	988	1035	974	1039	973	1039	1046	983	1045	1048
m_6	961	956	956	979	979	951	974	974	974	982	993	969	976	989	994	989
m_7	960	955	956	955	956	951	950	951	951	1054	951	945	1049	945	1049	1043
m_8	960	979	955	955	956	974	974	949	975	981	951	969	994	970	975	990
m_9	961	956	957	979	956	952	975	975	951	982	974	970	977	969	994	990
m_{10}	960	980	955	978	955	974	989	973	974	950	973	983	969	983	967	978
m_{11}	961	980	957	956	956	975	975	952	975	982	951	971	995	970	977	991
m_{12}	961	955	955	955	956	950	950	950	951	982	951	945	976	946	976	971
m_{13}	959	954	954	977	954	948	972	971	948	948	971	967	943	967	966	961
m_{14}	962	956	958	956	956	952	951	952	951	983	951	947	978	946	978	972
m_{15}	961	980	955	955	955	974	975	950	974	950	950	969	969	969	945	964
m_{16}	962	956	956	956	956	951	951	951	951	951	951	946	946	946	946	940

Tabela E.35: Médias do LLP por modelo gerador dos dados e modelo ajustado, para o caso 4.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	935	960	1005	958	959	1021	974	1019	979	1078	978	1030	1087	991	1085	1091
m_2	934	934	1004	957	958	1004	958	1019	957	1077	977	1019	1077	978	1085	1085
m_3	935	959	935	958	959	959	973	958	979	967	978	973	985	991	984	996
m_4	933	957	1003	933	957	1019	958	1003	977	1077	957	1020	1085	978	1077	1086
m_5	934	958	1004	957	934	1020	972	1019	958	1023	957	1029	1036	972	1035	1043
m_6	935	935	935	958	958	935	959	958	958	966	978	958	966	978	983	984
m_7	934	934	935	934	935	935	934	935	935	1038	935	935	1038	935	1038	1038
m_8	934	958	934	934	935	958	959	934	959	965	935	958	983	959	965	984
m_9	935	935	936	958	935	936	959	959	935	967	958	960	966	959	984	985
m_{10}	934	959	934	957	934	959	973	957	958	934	957	973	958	972	957	972
m_{11}	935	959	936	935	935	960	959	936	959	966	935	960	985	959	966	986
m_{12}	934	934	934	934	935	934	934	934	935	966	935	934	966	935	966	966
m_{13}	932	932	932	956	932	932	956	956	932	932	956	956	932	956	955	956
m_{14}	935	935	936	935	935	936	935	936	935	967	935	936	967	935	967	967
m_{15}	934	958	934	934	934	958	959	934	958	934	934	959	958	959	934	959
m_{16}	935	935	935	935	935	935	935	935	935	935	935	935	935	935	935	935

Tabela E.36: Médias do DIC por modelo gerador dos dados e modelo ajustado, para o caso 5.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	239	246	245	242	244	250	246	247	248	250	247	250	253	248	252	253
m_2	240	239	245	243	245	244	242	248	244	251	248	247	250	248	253	253
m_3	240	246	239	243	245	245	246	242	249	244	248	245	247	249	247	248
m_4	239	246	245	238	244	251	246	244	248	250	243	250	253	248	249	252
m_5	240	246	246	243	239	251	246	249	245	245	242	251	250	245	247	250
m_6	239	238	238	242	245	237	242	241	244	244	247	241	243	247	246	246
m_7	240	238	245	239	245	244	237	244	244	251	244	243	249	243	249	248
m_8	239	245	238	238	244	244	245	237	248	243	243	244	247	247	242	246
m_9	239	238	245	243	238	244	242	248	237	244	242	247	243	241	247	246
m_{10}	240	246	238	243	239	245	246	242	245	238	242	245	244	245	241	244
m_{11}	239	246	245	238	238	250	245	244	245	244	237	250	249	245	243	249
m_{12}	239	238	238	238	244	237	237	237	243	243	243	236	242	242	242	241
m_{13}	239	238	238	242	238	237	242	241	237	237	241	241	236	241	240	240
m_{14}	240	239	245	239	239	244	238	244	238	244	238	243	243	237	243	242
m_{15}	240	246	239	239	239	245	246	238	245	238	238	245	244	245	237	244
m_{16}	239	238	238	238	238	237	238	238	238	237	238	237	237	237	237	236

Tabela E.37: Médias do $\widehat{\text{BIC}}$ por modelo gerador dos dados e modelo ajustado, para o caso 5.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	253	256	255	253	255	258	254	255	256	258	255	255	258	253	257	255
m_2	253	250	256	254	256	252	250	256	252	258	256	252	254	253	258	255
m_3	254	257	250	254	256	253	254	250	256	252	256	250	252	254	252	250
m_4	253	257	256	249	255	259	253	252	256	258	251	255	258	252	254	254
m_5	254	257	257	254	250	259	254	256	253	253	250	256	255	250	252	252
m_6	253	249	249	253	255	245	250	249	251	251	255	246	248	252	251	248
m_7	253	249	256	249	255	252	245	252	252	258	251	248	254	247	254	250
m_8	253	256	249	249	255	252	253	245	256	251	251	249	252	252	247	248
m_9	253	249	256	253	249	252	250	256	245	252	249	252	248	246	252	248
m_{10}	253	257	249	254	259	253	254	250	253	245	250	250	249	250	246	246
m_{11}	253	256	256	249	249	258	253	252	253	252	245	255	254	249	248	251
m_{12}	252	249	249	248	255	245	245	245	251	251	251	241	247	247	247	243
m_{13}	253	249	249	253	249	245	250	249	245	245	249	246	241	246	245	242
m_{14}	253	250	256	250	250	252	246	252	246	252	246	248	248	242	248	244
m_{15}	253	257	249	249	249	253	254	245	253	245	245	250	249	250	241	246
m_{16}	253	249	249	249	249	245	245	245	245	245	245	241	241	242	241	238

Tabela E.38: Médias do LLP por modelo gerador dos dados e modelo ajustado, para o caso 5.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	233	241	240	238	239	246	242	243	244	246	243	247	250	245	249	251
m_2	234	234	240	238	240	240	238	244	240	247	244	244	247	245	250	251
m_3	234	241	234	239	240	241	242	238	245	240	244	242	244	246	244	246
m_4	233	241	240	233	240	247	242	240	244	246	239	247	250	245	246	251
m_5	234	241	241	238	234	247	242	245	241	241	238	248	247	242	245	248
m_6	233	233	233	238	240	234	238	238	240	240	243	238	240	244	243	244
m_7	234	234	240	234	240	240	234	240	240	247	240	240	246	240	246	246
m_8	233	241	233	233	239	240	241	233	244	239	239	241	244	244	239	244
m_9	234	233	240	238	234	240	238	244	233	240	238	244	240	238	244	244
m_{10}	234	241	234	238	234	241	242	238	241	234	238	242	241	242	238	242
m_{11}	233	241	240	233	233	246	241	240	241	240	233	247	246	242	240	247
m_{12}	233	233	233	233	239	233	233	233	239	239	239	233	239	239	239	239
m_{13}	233	233	233	237	233	233	238	237	233	233	237	238	233	238	237	238
m_{14}	234	234	240	234	234	240	234	240	234	240	234	240	240	234	240	240
m_{15}	234	241	234	234	234	241	242	234	241	234	234	242	241	242	234	242
m_{16}	233	234	233	234	234	234	234	234	234	234	234	234	234	234	234	234

Tabela E.39: Médias do DIC por modelo gerador dos dados e modelo ajustado, para o caso 6.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	473	485	487	484	487	495	490	496	494	497	495	499	501	498	503	505
m_2	473	473	488	485	488	487	484	496	487	497	496	496	497	496	504	505
m_3	474	485	473	484	488	484	490	483	494	487	495	489	494	498	494	497
m_4	473	485	488	472	487	495	485	486	494	497	486	495	502	494	496	501
m_5	473	485	487	484	472	495	490	496	485	486	483	499	495	489	495	499
m_6	473	472	472	484	488	471	484	483	487	487	495	483	486	496	494	495
m_7	473	472	487	472	487	486	471	486	487	497	486	485	497	485	496	495
m_8	473	485	472	472	488	484	485	471	495	487	487	484	494	495	486	494
m_9	472	471	487	484	471	486	484	495	470	486	483	496	485	482	494	495
m_{10}	473	485	472	484	472	484	490	483	484	471	483	489	483	489	482	488
m_{11}	474	485	488	473	473	495	485	487	485	487	472	495	495	484	486	495
m_{12}	473	472	472	472	487	471	471	471	486	486	486	470	485	485	485	484
m_{13}	472	471	471	483	471	470	483	482	470	470	482	482	469	482	481	481
m_{14}	474	473	488	473	473	487	472	486	472	487	472	486	486	471	485	485
m_{15}	473	485	472	472	472	484	485	471	484	471	471	484	483	484	470	483
m_{16}	473	472	472	472	472	471	471	471	471	471	471	470	470	470	470	469

Tabela E.40: Médias do $\widehat{\text{BIC}}$ por modelo gerador dos dados e modelo ajustado, para o caso 6.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	490	498	501	498	501	505	500	506	504	507	505	505	508	504	509	507
m_2	491	486	501	498	501	496	494	506	497	507	505	503	503	502	510	507
m_3	491	499	486	498	501	494	500	493	504	497	505	495	500	504	500	500
m_4	490	499	501	485	500	505	495	496	504	507	596	501	508	500	502	504
m_5	490	499	501	498	485	505	500	505	494	496	493	505	501	496	501	501
m_6	490	486	485	498	501	481	494	493	497	496	505	489	492	502	500	497
m_7	490	485	501	485	501	496	481	496	496	507	496	492	503	492	502	498
m_8	490	499	486	486	501	494	495	481	505	497	497	490	500	501	492	496
m_9	489	485	500	497	485	496	493	505	480	496	492	502	491	489	501	497
m_{10}	490	498	485	497	485	494	499	492	494	481	493	495	489	495	488	490
m_{11}	491	499	501	486	486	505	495	497	494	497	481	501	501	491	492	497
m_{12}	490	486	486	486	501	481	481	481	496	496	496	476	492	492	491	487
m_{13}	489	485	485	497	485	480	493	492	480	480	492	488	476	488	487	483
m_{14}	491	486	501	486	486	497	482	496	482	496	482	492	492	477	492	487
m_{15}	490	499	485	485	485	494	495	481	494	481	481	490	490	490	476	486
m_{16}	490	486	486	486	486	481	481	481	481	481	481	476	476	476	476	472

Tabela E.41: Médias do LLP por modelo gerador dos dados e modelo ajustado, para o caso 6.

modelo gerador	modelo ajustado															
	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}
m_1	467	480	482	479	482	491	486	492	490	493	491	496	498	495	500	503
m_2	468	468	483	480	483	483	480	492	483	493	492	493	494	493	501	503
m_3	468	480	468	479	483	480	486	479	490	483	491	486	491	495	491	495
m_4	467	480	483	467	482	491	481	482	490	493	482	492	499	491	493	499
m_5	467	480	482	479	467	491	486	492	481	482	479	496	492	486	492	497
m_6	467	467	467	479	483	467	480	479	483	483	491	480	483	493	491	493
m_7	467	467	482	467	482	482	467	482	483	493	482	482	494	482	493	493
m_8	467	480	467	467	483	480	481	467	491	483	483	481	491	492	483	492
m_9	466	466	482	479	466	482	480	491	466	482	479	493	482	479	492	493
m_{10}	467	480	467	479	467	480	486	479	480	467	479	486	480	486	479	486
m_{11}	468	480	483	468	468	491	481	483	481	483	468	492	492	481	483	493
m_{12}	467	467	467	467	482	467	467	467	483	482	482	467	482	482	482	482
m_{13}	466	466	466	478	466	466	479	478	466	466	478	479	466	479	478	479
m_{14}	468	468	483	468	468	483	468	483	468	483	468	483	483	468	483	483
m_{15}	467	480	467	467	467	480	481	467	480	467	467	481	480	481	467	481
m_{16}	467	467	467	467	467	467	467	467	467	467	467	467	467	467	467	467

Referências Bibliográficas

- Aitkin, M. (1991). Posterior Bayes factors (with discussion). *Journal of the Royal Statistical Society B* 53(1), 111–142.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. Em B. N. Petrov e F. Czaki (Eds.), *Proceedings of 2nd International Symposium on Information Theory*, pp. 267–281. Academia Kiado, Budapest.
- Amaral Turkman, M. A. e Silva, G. L. (2000). *Modelos Lineares Generalizados: da Teoria à Prática*. Edições SPE, Lisboa.
- Barbieri, M. M. e Berger, J. O. (2004). Optimal predictive model selection. *Annals of Statistics* 32(3), 870–897.
- Barnett, V. (1999). *Comparative Statistical Inference* (Third ed.). John Wiley and Sons, Chichester.
- Bayarri, M. J. e Berger, J. O. (1997). Measures of surprise in Bayesian analysis. ISDS Discussion Paper 97–46, Duke University.
- Bayarri, M. J. e Berger, J. O. (1999). Quantifying surprise in the data and model verification (with discussion). Em J. M. Bernardo, J. O. Berger, A. P. Dawid e A. F. M. Smith (Eds.), *Bayesian Statistics 6*, pp. 53–82. Clarendon Press, Oxford.
- Bayarri, M. J. e Berger, J. O. (2000). P-values for composite null models. *Journal of the American Statistical Association* 95, 1127–1142.
- Berg, A., Meyer, R. e Yu, J. (2002). Deviance information criterion for comparing stochastic volatility models. Economics Working Paper 228, University of Auckland.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (Second ed.). Springer-Verlag, New York.
- Berger, J. O. e Pericchi, L. R. (1996). The intrinsic Bayes factor for linear models. Em J. M. Bernardo, J. O. Berger, A. P. Dawid e A. F. M. Smith (Eds.), *Bayesian Statistics 5*, pp. 25–44. Oxford University Press, Oxford.
- Bernardo, J. M. e Smith, A. F. M. (2000). *Bayesian Theory*. Wiley, Chichester.
- Besag, J. E. (1975). Statistical analysis of non-lattice data. *The Statistician* 24, 179–195.
- Best, N., Cowles, M. K. e Vines, K. (1996). *CODA: Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output, Version 0.30*. MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK.
- Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association* 71, 791–799.

- Box, G. E. P. (1980). Sampling and Bayes inference in scientific modelling and robustness. *Journal of the Royal Statistical Society A* 143, 383–430.
- Burnham, K. P. e Anderson, D. R. (1998). *Model Selection and Inference: a Practical Information-Theoretic Approach*. Springer-Verlag, New York.
- Carlin, B. P. e Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B* 57, 473–484.
- Carlin, B. P. e Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis* (Second ed.). Chapman and Hall, New York.
- Cavalcante, M. I. A. (1998). *Comparação de Modelos Separados — uma Abordagem Bayesiana Usando Factores de Bayes com Distribuição a Priori Imprópria*. Tese de doutoramento, Universidade Federal do Rio de Janeiro.
- Congdon, P. (2001). *Bayesian Statistical Modelling*. Wiley, London.
- Costa, R. F. (2001). *Composição Corporal: Teoria e Prática da Avaliação*. Editora Manole Ltda, São Paulo.
- Cuzick, J. (1991). (In discussion following a paper by Aitkin) Posterior Bayes factors. *Journal of the Royal Statistical Society B* 53(1), 135–136.
- Dellaportas, P. e Forster, J. J. (1999). Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models. *Biometrika* 86, 615–633.
- Dellaportas, P., Forster, J. J. e Ntzoufras, I. (1997). On Bayesian model variable selection using MCMC. Technical report, Department of Statistics, Athens University of Economics and Business, Greece.
- Dellaportas, P., Forster, J. J. e Ntzoufras, I. (2000). Bayesian variable selection using the Gibbs sampler. Em D. K. Dey, S. K. Ghosh e B. K. Mallick (Eds.), *Generalized Linear Models: a Bayesian Perspective*, pp. 271–286. Marcel Dekker, New York.
- Dempster, A. P. (1974). The direct use of likelihood for significance testing. Em O. Barndorff-Nielsen, P. Blaesild e G. Schou (Eds.), *Proceedings of Conference on Foundational Questions in Statistical Inference*, pp. 335–352. University of Aarhus, Aarhus.
- Fragoso, I. e Vieira, F. (2000). *Morfologia e Crescimento. Curso Prático*. Faculdade de Motricidade Humana, Serviço de Edições, Cruz Quebrada.
- Gamerman, D. (1997). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall, New York.
- Geisser, S. e Eddy, W. (1979). A predictive approach to model selection. *Journal of the American Statistical Association* 74, 153–160.
- Gelfand, A. E. (1996). Model determination using sampling-based methods. Em W. R. Gilks, S. Richardson e D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 145–161. Chapman and Hall, New York.
- Gelfand, A. E. e Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society B* 56, 501–514.
- Gelfand, A. E., Dey, D. K. e Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with discussion).

-
- Em J. M. Bernardo, J. O. Berger, A. P. Dawid e A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 147–167. Oxford University Press, Oxford.
- Gelman, A., Carlin, J. B., Stern, H. S. e Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall, New York.
- Gelman, A., Meng, X.-L. e Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica* 6, 733–807.
- Gelman, A. e Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* 7, 457–511.
- George, E. I. e McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 88, 881–889.
- George, E. I. e McCulloch, R. E. (1996). Stochastic search variable selection. Em W. R. Gilks, S. Richardson e D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 203–214. Chapman and Hall, New York.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). Em J. M. Bernardo, J. O. Berger, A. P. Dawid e A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 169–193. Oxford University Press, Oxford.
- Gilks, W. R., Richardson, S. e Spiegelhalter, D. J. (Eds.) (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, New York.
- Gill, J. (2002). *Bayesian Methods: a Social and Behavioral Sciences Approach*. Chapman and Hall, New York.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82(4), 711–732.
- Guttman, I. (1967). The use of the concept of a future observation in goodness-of-fit problems. *Journal of the Royal Statistical Society B* 29, 83–100.
- Han, C. e Carlin, B. P. (2001). MCMC methods for computing Bayes factor: a comparative review. *Journal of the American Statistical Association* 96, 1122–1132.
- Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics* 32, 1–49.
- Hoeting, J. A., Madigan, D., Raftery, A. E. e Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical Science* 14(4), 382–417.
- Jeffreys, H. (1961). *Theory of Probability* (Third ed.). Oxford University Press, Oxford.
- Jobson, J. D. (1991). *Applied Multivariate Data Analysis, Volume 1: Regression and Experimental Design*. Springer-Verlag, New York.
- Johnson, N. L. e Kotz, S. (1992). *Distributions in Statistics — Discrete Distributions* (Second ed.). Wiley, New York.
- Johnson, R. W. (1996). Fitting percentage of body fat to simple body measurements. *Journal of Statistics Education* 4(1).
- Kass, R. E. e Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* 90(430), 773–795.
-

- Kuo, L. e Mallick, B. (1998). Variable selection for regression models. *Sankhya B* 60, 65–81.
- Lindley, D. V. (1957). A statistical paradox. *Biometrika* 44, 187–192.
- Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 2*. Cambridge University Press, Cambridge.
- Madigan, D. e Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam’s window. *Journal of the American Statistical Association* 89, 1535–1546.
- McCullagh, P. e Nelder, J. A. (1989). *Generalized Linear Models* (Second ed.). Chapman and Hall, London.
- Meng, X.-L. (1994). Posterior predictive p-values. *The Annals of Statistics* 22, 1142–1160.
- Miller, A. J. (1984). Selection of subsets of regression variables. *Journal of the Royal Statistical Society A* 147, 389–410.
- Miller, A. J. (1990). *Subset Selection in Regression*. Chapman and Hall, London.
- Montgomery, D. C. e Peck, E. A. (1992). *Introduction to Linear Regression Analysis* (Second ed.). Wiley, New York.
- Nelder, J. A. e Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A* 135, 370–384.
- Ntzoufras, I. (1999). *Aspects of Bayesian Model and Variable Selection Using MCMC*. Ph. D. thesis, Department of Statistics, Athens University of Economics and Business.
- Ntzoufras, I. (2002). Gibbs variable selection using BUGS. *Journal of Statistical Software* 7(7).
- O’Hagan, A. (1991). (In discussion following a paper by Aitkin) Posterior Bayes factors. *Journal of the Royal Statistical Society B* 53(1), 136.
- O’Hagan, A. (1995). Fractional Bayes factors for model comparison (with discussion). *Journal of the Royal Statistical Society B* 57, 99–138.
- O’Hagan, A. (1997). Properties of intrinsic and fractional Bayes factors. *Test* 6, 101–118.
- Palma dos Reis, R. (1994). *Homocisteinemia e Doença Vascular — Estudo Epidemiológico e Clínico*. Tese de doutoramento, Faculdade de Ciências Médicas da Universidade Nova de Lisboa.
- Paulino, C. D., Amaral Turkman, M. A. e Murteira, B. (2003). *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa.
- Penrose, K., Nelson, A. e Fisher, A. (1985). Generalized body composition prediction equation for men using simple measurement techniques (abstract). *Medicine and Science in Sports and Exercise* 17(2), 189.
- Raftery, A. E. (1996). Hypothesis testing and model selection. Em W. R. Gilks, S. Richardson e D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 163–187. Chapman and Hall, New York.

-
- Raftery, A. E. e Lewis, S. M. (1992). How many iterations in the Gibbs sampler? Em J. M. Bernardo, J. O. Berger, A. P. Dawid e A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 763–773. Oxford University Press, Oxford.
- Raftery, A. E., Madigan, D. e Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92, 179–191.
- Roberts, G. O. (1996). Markov chain concepts related to sampling algorithms. Em W. R. Gilks, S. Richardson e D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 45–57. Chapman and Hall, New York.
- Robins, J. M., van der Vaart, A. e Ventura, V. (2000). Asymptotic distribution of p values in composite null models (with discussion). *Journal of the American Statistical Association* 95(452), 1143–1172.
- Rowe, D. B. (2003). *Multivariate Bayesian Statistics — Models for Source Separation and Signal Unmixing*. Chapman and Hall/CRC, Boca Raton.
- Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics* 12, 1151–1172.
- Ryan, T. P. (1997). *Modern Regression Methods*. Wiley, New York.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* 6(2), 461–464.
- Sellke, T., Bayarri, M. J. e Berger, J. O. (1999). Calibration of p-values for testing precise null hypotheses. ISDS Discussion Paper 99–13, Duke University.
- Smith, A. F. M. e Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society B* 55, 3–23.
- Smith, B. J. (2005). *Bayesian Output Analysis Program (BOA), User’s Manual, Version 1.1*. Department of Biostatistics, College of Public Health, University of Iowa.
- Sorensen, D. e Gianola, D. (2002). *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer-Verlag, New York.
- Spiegelhalter, D., Thomas, A., Best, N. e Gilks, W. (1996a). *BUGS 0.5: Bayesian Inference Using Gibbs Sampling Manual (version ii)*. MRC Biostatistics Unit, Institute of Public Health, Cambridge.
- Spiegelhalter, D., Thomas, A., Best, N. e Gilks, W. (1996b). *BUGS Examples Volume 1, Version 0.5 (version ii)*. MRC Biostatistics Unit, Institute of Public Health, Cambridge.
- Spiegelhalter, D., Thomas, A., Best, N. e Gilks, W. (1996c). *BUGS Examples Volume 2, Version 0.5 (version ii)*. MRC Biostatistics Unit, Institute of Public Health, Cambridge.
- Spiegelhalter, D., Thomas, A., Best, N. e Lunn, D. (2001). *WinBUGS User Manual — version 1.4*. MRC Biostatistics Unit, Institute of Public Health, Cambridge.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. e van der Linden, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society B Part 4*(64), 583–639.
-

- Teles, J. (1995). *Modelos Lineares Generalizados: uma Aplicação à Medicina*. Tese de mestrado, Faculdade de Ciências da Universidade de Lisboa.
- Weisberg, S. (1985). *Applied Linear Regression* (Second ed.). Wiley, New York.
- Zhu, L. e Carlin, B. P. (2000). Comparing hierarchical models for spacio-temporally misaligned data using the deviance information criterion. *Statistics in Medicine* 19, 2265–2278.