Helsingin yliopisto

Matematiikan ja tilastotieteen laitos

Pro gradu -tutkielma

Markovin ketjut jatkuvalla tila-avaruudella sekä Metropolisin ja Hastingsin algoritmi

Tekijä:
Petri Aaltonen

Ohjaaja:
Prof. Esa Nummelin

19. kesäkuuta 2016

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Matemaattis-luonnontieteellinen	Matematiikan ja tilastotieteen laitos

Tekijä — Författare — Author
Petri Aaltonen

Työn nimi — Arbetets titel — Title
Markovin ketjut jatkuvalla tila-avaruudella sekä Metropolisin ja Hastingsin algoritmi

Oppiaine — Läroämne — Subject
Soveltava matematiikka

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Pro gradu -tutkielma	Kesäkuu 2016	60 s.

Tiivistelmä — Referat — Abstract

Työssä rakennetaan yleisten Markovin ketjujen teoriaa tila-avaruudella, joka on euklidisen avaruuden $\mathbb{R}^d$ osajoukko. Määrittelemme uusiutumisprosessit ja rakennamme regeneroituvien Markovin ketjujen teorian. Näytämme, että ergodisuusoletuksen toteuttava Markovin ketju on Harris-palautuva, positiivisesti palautuva ja sen tasapainojakauma on yksikäsitteinen. Regeneroituvalla Markovin ketjulla on tila-avaruuden osajoukko, johon osuessaan sillä on mahdollisuus regeneroitua positiivisella todennäköisyydellä. Regeneraation tapahtuessa Markovin ketju unohtaa historiansa ja sitä voidaan tarkastella kuten se käynnistyisi uudestaan tietyllä regeneraatiokonstruktion määrittämällä alkujakaumalla. Harris-palautuvuus ja positiivinen palautuvuus ovat vahvoja regeneraatioajan äärellisyyttä koskevia tuloksia. Teoriaa hyväksikäyttämällä todistetaan kolme keskeistä konvergenssitulosta Markovin ketjuille: suurten lukujen laki, jakauman suppeneminen kokonaisvariaatioetäisyydessä sekä keskeinen raja-arvolause.

Markovin ketjujen teoria rakennetaan siinä laajuudessa, kuin sen avulla on mahdollista ymmärtää Metropolisin ja Hastingsin algoritmin toiminta. On annettu jonkin todennäköisyysjakauman mahdollisesti normalisoimaton tiheysfunktio π ja tehtävänä on muodostaa satunnaisotos kyseisestä jakaumasta. Metropolisin ja Hastingsin algoritmi konstruoi Markovin ketjun, jonka tasapainojakauma on π. Markovin ketjua simuloimalla saadaan siten haluttu otos. Mikäli Markovin ketju toteuttaa riittävät säännöllisyysominaisuudet, on muun muassa suurten lukujen laki ja keskeinen raja-arvolause voimassa, mikä merkitsee, että saatu otos on käytännössä hyödyllinen.

Metropolisin ja Hastingsin algoritmi on esimerkki Markovin ketju Monte Carlo eli MCMC-menetelmistä. Ne mahdollistavat simuloinnin hyvin monimutkaisista jakaumista, joiden hallinta muita menetelmiä käyttäen on vaikeaa tai mahdotonta. Bayesiläinen tilastotiede ja tilastollinen mekaniikka ovat esimerkkejä MCMC-menetelmien tärkeistä sovellusaloista. Esittelemme lyhyesti MCMC-menetelmien soveltamisen perusteet ja suoritamme lyhyen katsauksen menetelmien historiaan. Lopuksi esittelemme soveltavan esimerkin, jossa Metropolisin ja Hastingsin algoritmia käytetään salakirjoitetun tekstin selventämiseen.

Avainsanat — Nyckelord — Keywords
Yleinen Markovin ketju, Markovin ketju Monte Carlo, Metropolisin ja Hastingsin algoritmi

Säilytyspaikka — Förvaringsställe — Where deposited
Kumpulan tiedekirjasto

Muita tietoja — Övriga uppgifter — Additional information

Sisältö

LUKU 1

Johdanto

Diskreetillä eli joko äärellisellä tai numeroituvasti äärettömällä joukolla eli niin sanotulla *tila-avaruudella* määriteltyjen *Markovin ketjujen* teoria tunnetaan nykyään varsin perusteellisesti. Teoria hyödyntää hyvin elegantilla tavalla elementaarista todennäköisyysmatematiikkaa ja matriisilaskentaa. *Yleiset* Markovin ketjut on määritelty sitä vastoin millä tahansa mitallisella tila-avaruudella ja niiden teoria on kehitetty pitkälti viimeisen puolen vuosisadan aikana. Yleisen teorian ymmärtäminen vaatii mittateoreettisen todennäköisyyslaskennan hyvää hallintaa ja johtaa moniin teoreettisiin ja käsitteellisiin vaikeuksiin, jotka puuttuvat diskreetistä teoriasta. Tämän tutkielman lähtökohtana on rakentaa yleisten Markovin ketjujen teoriaa niiden soveltamiseksi MCMC-menetelmiä varten olettaen, että tila-avaruudella on tietty rakenne, mikä yksinkertaistaa esitystä täysin yleiseen tilanteeseen verrattuna. Tämän lähtökohdan innoittajana toimii artikkeli [**Num02**].

Markovin ketju Monte Carlo -menetelmät (lyhyemmin MCMC-menetelmät) ovat joukko satunnaisotantaan perustuvia simulointimenetelmiä, joiden toiminta nojaa sopivan Markovin ketjun simulointiin. On annettu todennäköisyysjakauma, josta haluamme suorittaa satunnaisotannan. Suora otanta jakaumasta saattaa olla vaikeaa tai tehotonta tai jakauma saattaa olla tunnettu normalisoivaa tekijää vailla. Yleensä on suhteellisen helppo muodostaa sellainen Markovin ketju, jolle kiinnostuksen kohteena oleva jakauma on niin sanottu tasapainojakauma. Mikäli Markovin ketju toteuttaa riittävät säännöllisyysominaisuudet, yksittäisten iteraattien jakaumat suppenevat tietyssä mielessä kohti tasapainojakaumaa, kun Markovin ketjua ajetaan riittävän monta iteraatiota. Markovin ketjun havaitut realisaatiot muodostavat siten likimääräisen otoksen tasapainojakaumasta. Otos ei ole riippumaton, mutta tulemme osoittamaan, että tiettyjen säännöllisyysominaisuuksien vallitessa otos tai jokin sen funktio toteuttaa suurten lukujen lain ja keskeisen raja-arvolauseen.

MCMC-menetelmillä on lukuisia sovelluskohteita, joista hyvä katsaus löytyy artikkelista [**Dia09**] sekä kirjoista [**GRS96**], [**AG07**] ja [**BGJM11**]. Fysiikka ja erityisesti tilastollinen mekaniikka on yksi keskeinen MCMC-menetelmien sovellusalue [**Lan05**]. Bayesiläinen tilastotiede on ollut hyvin merkittävä sovellusalue 80- ja 90-luvun taiteesta lähtien. Laajat Bayesiläiset mallit johtavat nimittäin integrointiin yli avaruuden, jonka dimensio voi olla kymmeniä, satoja tai jopa tuhansia. Matalissa dimensioissa integraaleja on mahdollista hallita käyttämällä numeerisia integroimismenetelmiä, asymptoottisia approksimaatioita tai riippumattomaan otokseen perustuvia *klassisia Monte Carlo -menetelmiä* [**ES95**]. Nämä johtavat kuitenkin ongelmiin korkeadimensioisissa malleissa, joiden tehokkaan käsittelyn MCMC-menetelmät mahdollistavat. MCMC-menetelmät ovatkin mullistaneet Bayesiläisen tilastotieteen tehden siitä käytännöllisen tavan tehdä tilastotiedettä. Kirjallisuudessa puhutaan MCMC-*vallankumouksesta*, kts. esim. [**CR00**] ja [**Dia09**].

Tutkielman lähestymistapa on teoreettinen. Rakennamme *yleisten* Markovin ketjujen teorian siinä laajuudessa kuin se riittää MCMC-menetelmien ymmärtämiseksi sovellettaessa sitä erityisesti parametrisiin tilastollisiin malleihin. Esittelemme Metropolisin ja Hastingsin algoritmin ja todistamme edellä rakennettua teoriaa käyttäen, että se toimii halutulla tavalla. Esimerkkien tarkoitus on valaista teoriaa ja siten ne ovat yksinkertaisia ja helposti analyyttisesti käsiteltäviä. Lopuksi tarkastelemme soveltavaa esimerkkiä, joka on hyvä osoitus MCMC-menetelmien monipuolisista käyttökohteista.

Oletamme, että lukija tuntee Markovin ketjujen teorian perusteet diskreetillä tila-avaruudella. Elementaarisen todennäköisyyslaskennan peruskäsitteet oletetaan luonnollisesti tunnetuiksi. Lisäksi käytämme jonkin verran mittateorian käsitteitä, mutta niiden puutteellinen tuntemus ei oleellisesti haittaa esityksen seuraamista. Liitteessä A esitetään joitakin hyödyllisiä tuloksia.

Tutkielma rakentuu seuraavasti. Luku 2 esittelee yleisten Markovin ketjujen teorian perusteet, uusiutumisprosessit ja Markovin ketjujen regeneraation. Luvussa 3 muotoillaan ja todistetaan kolme keskeistä Markovin ketjujen konvergenssitulosta: suurten lukujen laki, jakauman suppeneminen kokonaisvariaatioetäisyydessä sekä keskeinen raja-arvolause. Luku 4 soveltaa edellisessä kahdessa luvussa rakennettua teoriaa Metropolisin ja Hastingsin algoritmin ymmärtämiseksi. Luvussa 5 suoritetaan lyhyt yhteenveto ja tarkastellaan tutkielman ulkopuolelle jääneitä kysymyksiä. Tutkielman lopussa liitteessä A on esitetty joitakin matemaattisia tuloksia ja liitteessä B on esitetty R-kielinen ohjelmakoodi, jota on käytetty esimerkkien Markovin ketjujen simulointiin.

Haluan esittää kiitokseni työn ohjaajalle *Esa Nummelinille* johdatuksesta Markovin ketjujen kiehtovaan maailmaan sekä erinomaisesta ajatuksesta lähteä tutkimaan MCMC-menetelmiä ja yleisiä Markovin ketjuja artikkelin [**Num02**] viitoittamalla tiellä. Lisäksi haluan kiittää *Jukka Rantaa* ja *Jukka Coranderia* johdatuksesta Bayesiläiseen tilastotieteeseen ja MCMC-menetelmien käyttöön.

LUKU 2

Markovin ketjut

1. Markovin ketjut jatkuvalla tila-avaruudella

Markovin ketju on *stokastinen prosessi* eli ajassa etenevä satunnaisilmiö, jonka muisti ulottuu ainoastaan yhden aika-askeleen päähän menneisyyteen. Markovin ketjujen teoria jaetaan tyypillisesti *diskreettiin* ja *yleiseen* teoriaan. Diskreetissä teoriassa Markovin ketjun tila-avaruus on äärellinen tai numeroituvasti ääretön joukko. Yleisessä teoriassa tila-avaruus voi olla mielivaltainen mitta-avaruus eikä siltä tarvitse vaatia esimerkiksi mitään topologisia ominaisuuksia. Tässä tutkielmassa tarkastellaan yleistä teoriaa sillä rajoituksella, että tila-avaruus on jokin euklidisen avaruuden $\mathbb{R}^d$ osajoukko. Nimitämme tällaista tila-avaruutta *jatkuvaksi*. Tutkielmassa laadittava teoria käsittää erikoistapauksenaan myös diskreetit Markovin ketjut. Lisäksi vaadimme pian esiteltäviltä todennäköisyysytimiltä ja todennäköisyysjakaumilta tiettyjä säännöllisyysominaisuuksia. Viittaamme toisinaan diskreettiin tai yleiseen teoriaan, kun haluamme esittää tietyn jatkuvassa teoriassa esitellyn ominaisuuden rajoitetummassa tai yleisemmässä kontekstissa. Markovin ketjun aikamuuttuja on diskreetti. *Markovin prosessiksi* puolestaan tyypillisesti nimitetään sellaista Markovin ketjuille läheistä sukua olevaa stokastista prosessia, jonka aikamuuttuja on jatkuva [**MT93**] s. 21.

Luvun 1 tarkoituksena on esitellä Markovin ketjujen teorian perusteet jatkuvalla tila-avaruudella. Rajoitumme käsittelemään sellaisia ominaisuuksia, jotka riittävät Metropolisin ja Hastingsin algoritmin ymmärtämiseksi. Luvun materiaali perustuu ensisijaisesti artikkeliin [**Num02**] ja kirjaan [**Num84**].

Käydään ensin läpi joitakin tutkielmassa käytettäviä merkintöjä. Luonnollisten lukujen joukkoa merkitään $\mathbb{N} = \{0, 1, 2, \dots\}$. Positiivisten kokonaislukujen joukkoa merkitään puolestaan $\mathbb{N}^+$. Laajennettu luonnollisten lukujen joukko on $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$. Positiivisten reaalilukujen joukkoa merkitään $\mathbb{R}^+$ ja laajennettua positiivisten reaalilukujen joukkoa $\bar{\mathbb{R}}^+ = \mathbb{R}^+ \cup \{\infty\}$. Avointa x-keskistä ja r-säteistä kuulaa merkitään $B(x, r) = \{y : |y - x| < r\}$.

1.1. Todennäköisyysytimet.

Merkitään kirjaimella d euklidisen avaruuden $\mathbb{R}^d$ dimensiota. Olkoon joukko $E \subset \mathbb{R}^d$ sekä $\mathcal{E}$ eräs sen σ-algebra eli sellainen joukon E osajoukkojen kokoelma, joka sisältää joukon E ja on suljettu komplementin oton sekä numeroituvien yhdisteiden suhteen. Välttämätön tekninen vaatimus on, että $\mathcal{E}$ on numeroituvasti generoitu [**Num84**] s. 1. Tämä vaatimus ei tuota ongelmia, eikä siihen kiinnitetä jatkossa huomiota. Mitallista avaruutta $(E, \mathcal{E})$ nimitetään *tila-avaruudeksi*. Myös joukkoa E voidaan lyhyemmin nimittää tila-avaruudeksi, mutta pidämme mielessä, että sen rinnalla on aina määritelty jokin σ-algebra. Alkio $x \in E$ on *tila*. Tyypillisesti E on tuloavaruus, joka muodostuu jatkuvista väleistä ja diskreeteistä joukoista. $\mathcal{E}$ on tällöin tulo-σ-algebra. Mitta $\mathrm{vol}(\cdot) : \mathcal{E} \to \bar{\mathbb{R}}^+$ määritellään tulomittana. Jatkuvalla joukolla määritellään Lebesguen mitta ja diskreetillä joukolla lukumäärämitta.

Joukko E kirjoitetaan tarvittaessa jatkuvan ja diskreetin osan eli *komponentin* tulona $E = E_C \times E_D$, jossa $E_C \subset \mathbb{R}^{d_C}$ on numeroituvasti ääretön joukko ja $E_D \subset \mathbb{R}^{d_D}$ on diskreetti joukko sekä pätee $d_C + d_D = d$. Tarvittaessa tila $x \in E$ voidaan kirjoittaa muodossa $x = (x_C, x_D)$, jossa $x \in E_C$ ja $x_D \in E_D$. Näin menetellen pystymme helposti määrittelemään topologian tila-avaruudella E käyttäen tulotopologiaa. Jatkuvalla komponentilla E_C määritellään tavanomainen euklidisen metriikan indusoima topologia. Diskreetillä komponentilla E_D määritellään diskreetti topologia, jossa jokainen $A \subset E_D$ on avoin. Tarvitsemme myöhemmin jatkuvuuden ja kompaktisuuden käsitteitä, joten oletetaan, että tila-avaruudella E on aina määritelty topologia edellä kuvattua menettelyä käyttäen.

Esitetään ytimien perusominaisuudet perustuen kirjaan [**Num84**].

MÄÄRITELMÄ 2.1. Kuvaus $K : E \times \mathcal{E} \to \bar{\mathbb{R}}^+$ on *ydin*, jos kuvaus $x \mapsto K(x, A)$ on mitallinen jokaisella $A \in \mathcal{E}$ ja kuvaus $A \mapsto K(x, A)$ on mitta jokaisella $x \in E$. Sanomme, että K on *subtodennäköisyysydin* (tai *substokastinen* ydin), jos $K(x, E) \leq 1$ jokaisella $x \in E$. Sanomme, että K on *todennäköisyysydin* (tai *stokastinen* ydin), jos $K(x, E) = 1$ jokaisella $x \in E$.

Määritellään *identiteettiydin* id asettamalla $\mathrm{id}(x, A) = \mathbf{1}_A(x)$ jokaisella $x \in E$ ja jokaisella $A \in \mathcal{E}$.

Olkoon λ mitta. Määritellään mitta λK asettamalla

$$\lambda K(A) \triangleq \int \lambda(dx) K(x, A), \quad A \in \mathcal{E}. \tag{1}$$

Mikäli λ on todennäköisyysmitta ja ydin K on stokastinen, on λK todennäköisyysmitta. On ilmeistä, että $\lambda \mathrm{id} = \lambda$.

Olkoon f mikä tahansa epänegatiivinen, $\mathcal{E}$-mitallinen funktio. Epänegatiivinen ja $\mathcal{E}$-mitallinen funktio Kf määritellään asettamalla

$$Kf(x) \triangleq \int K(x, dy) f(y), \quad x \in E.$$

Jos f on mikä tahansa $\mathcal{E}$-mitallinen funktio, voidaan se kirjoittaa positiivisen ja negatiivisen osan erotuksena $f = f^+ - f^-$, kun f^+ ja f^- ovat molemmat epänegatiivisia ja $\mathcal{E}$-mitallisia, edellyttäen, että ei ole sellaista $x \in E$, jossa f^+ ja f^- ovat molemmat äärettömiä. $\mathcal{E}$-mitallinen funktio Kf voidaan määritellä asettamalla $Kf = Kf^+ - Kf^-$ edellyttäen, että Kf^+ ja Kf^- eivät ole äärettömiä samassa pisteessä.

Ytimien K_1 ja K_2 *tuloydin* $K_1 K_2$ määritellään asettamalla

$$K_1 K_2(x, A) \triangleq \int K_1(x, dy) K_2(y, A), \quad x \in E, A \in \mathcal{E}. \tag{2}$$

On ilmeistä, että jos K_1 ja K_2 ovat stokastisia, niin myös niiden tulo $K_1 K_2$ on stokastinen. Yhtä lailla substokastisien ytimien tulo on substokastinen.

LAUSE 2.1. Mitta λK, funktio Kf ja tuloydin $K_1 K_2$ ovat hyvin määriteltyjä aina, kun $\lambda : E \to \bar{\mathbb{R}}^+$ on mitta ja f on epänegatiivinen ja $\mathcal{E}$-mitallinen funktio.

TODISTUS. Olkoon λ mitta. Osoitetaan, että λK on σ-additiivinen. Mikäli $(A_i : i \geq 1)$ ovat erillisiä $\mathcal{E}$-joukkoja, käyttämällä mitan $A \mapsto K(x, A)$ σ-additiivisuutta sekä vaihtamalla integroinnin ja summan järjestys

$$\begin{aligned}\lambda K\left(\bigcup_{i=1}^{\infty} A_i\right) &= \int \lambda(dx) K\left(x, \bigcup_{i=1}^{\infty} A_i\right) \\ &= \sum_{i=1}^{\infty} \int \lambda(dx) K(x, A_i) = \sum_{i=1}^{\infty} \lambda K(A_i).\end{aligned}$$

Olkoon $f = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}$ yksinkertainen funktio, jossa a_i ovat positiivisia ja $A_i \in \mathcal{E}$ ovat erillisiä. Nyt

$$Kf(x) = \int K(x, dy) f(y) = \sum_{i=1}^{n} a_i \int K(x, dy) \mathbf{1}_{A_i}(y) = \sum_{i=1}^{n} a_i K(x, A_i),$$

joka on mitallinen, sillä kuvaus $x \mapsto K(x, A_i)$ on mitallinen kullakin i. Löytyy kasvava jono $(f_n : n \geq 1)$ yksinkertaisia funktioita siten, että $f_n \to f$ melkein varmasti. Monotonisen konvergenssin lauseen nojalla

$$\lim_{n\to\infty} Kf_n(x) = \lim_{n\to\infty} \int K(x, dy) f_n(y) = \int K(x, dy) f(y) \quad \text{m.v.}$$

Siten [**Cin11**] teoreeman 2.15 nojalla Kf on mitallisten funktioiden rajana mitallinen.

Olemme osoittaneet, että mitta λK ja epänegatiivinen funktio Kf ovat hyvin määriteltyjä. Tästä seuraa, että $x \mapsto K_1 K_2(x, A)$ on hyvin määritelty mitallinen funktio ja $A \mapsto K_1 K_2(x, A)$ on hyvin määritelty mitta. Siten myös tuloydin $K_1 K_2$ on hyvin määritelty. □

Lauseesta seuraa, että Kf on hyvin määritelty myös yleisellä $\mathcal{E}$-mitallisella f edellyttäen, että Kf^+ ja Kf^- eivät ole molemmat äärettömiä samassa pisteessä. Mikäli f on äärellinen ja integroituva eli $\int |f(x)| dx < \infty$, sekä lisäksi ydin K on stokastinen tai substokastinen, seuraa, että Kf on välttämättä hyvin määritelty.

Käyttämällä tuloytimen määritelmää voidaan määritellä ytimen K iteraatit K^t kullakin $t \geq 0$. Nimittäin asetetaan $K^0 \triangleq \mathrm{id}$ ja $K^t \triangleq KK^{t-1}$ jokaisella $t \geq 1$. Edellä todetun nojalla stokastisen ytimen iteraatit ovat stokastisia ja substokastisen ytimen iteraatit ovat substokastisia.

Kirjan [**Num84**] sivu 9 mukaisesti määrittelemme ytimen K *potentiaalin*

$$G \triangleq \sum_{t=0}^{\infty} K^t.$$

LAUSE 2.2. Ytimen K potentiaali G on ydin.

TODISTUS. Osasummat $S_t(x, A) = \sum_{s=0}^{t} K^s(x, A)$ muodostavat kullakin $A \in \mathcal{E}$ jonon $\mathcal{E}$-mitallisia funktioita, joille $\lim_{t\to\infty} S_t(x, A) = G(x, A)$. Siten [**Cin11**] teoreeman 2.15 nojalla kuvaus $x \mapsto G(x, A)$ on $\mathcal{E}$-mitallinen. On helppo näyttää, että kuvaus $A \mapsto G(x, A)$ on mitta jokaisella $x \in E$, sillä olennaisesti riittää todeta, että kuvaus on σ-additiivinen. Olkoon $(A_n : n \geq$

1) jono erillisiä $\mathcal{E}$-mitallisia joukkoja. Kiinnitetään mielivaltainen $x \in E$. Nyt

$$\begin{aligned} G\left(x, \bigcup_{n=1}^{\infty} A_n\right) &= \sum_{t=0}^{\infty} K^t \left(x, \bigcup_{n=1}^{\infty} A_n\right) \\ &= \sum_{t=0}^{\infty} \sum_{n=1}^{\infty} K^t(x, A_n) \\ &= \sum_{n=1}^{\infty} \sum_{t=0}^{\infty} K^t(x, A_n) = \sum_{n=1}^{\infty} G(x, A_n). \qquad \square \end{aligned}$$

Lause 2.3. Ytimen K potentiaali G toteuttaa $G = \mathrm{id} + GK = \mathrm{id} + KG$.

Todistus. Käyttämällä iteraattien määritelmää

$$\begin{aligned} G(x, A) &= \mathrm{id}(x, A) + \sum_{t=1}^{\infty} K^{t-1} K(x, A) \\ &= \mathrm{id}(x, A) + \sum_{t=0}^{\infty} \int K^t(x, dy) K(y, A). \end{aligned}$$

Vaihtamalla summan ja integraalin järjestys

$$\begin{aligned} G(x, A) &= \mathrm{id}(x, A) + \int \sum_{t=0}^{\infty} K^t(x, dy) K(y, A) \\ &= \mathrm{id}(x, A) + \int G(x, dy) K(y, A) = \mathrm{id}(x, A) + GK(x, A). \qquad \square \end{aligned}$$

On ilmeistä, että ytimen K substokastisuus tai stokastisuus ei siirry potentiaalille G.

1.2. Markovin ketjut ja siirtymätodennäköisyydet.

Aikaindeksejä merkitään $s, t \in \mathbb{N}$. Kirjaimet S, T ja τ varataan myös merkitsemään aikaa. Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ todennäköisyysavaruus. Paria $(\Omega, \mathcal{F})$ nimitetään *polkuavaruudeksi*. Jono $(X_t : t \in \mathbb{N})$ $\mathcal{F}$-mitallisia satunnaismuuttujia $X_t : \Omega \to E$ on *Markovin ketju*, jos

$$\mathbb{P}(X_{t+1} \in A \mid X_t, X_{t-1}, \ldots, X_0) = \mathbb{P}(X_{t+1} \in A \mid X_t) \tag{3}$$

kullakin $A \in \mathcal{E}$. *Siirtymätodennäköisyydet* $\mathbb{P}(X_{t+1} \in A \mid X_t)$ määrittelevät *siirtymätodennäköisyysytimen* $P(X_t, A)$. Todennäköisyysjakaumaa $\mathbb{P}(X_0 \in A)$ nimitetään Markovin ketjun *alkujakaumaksi* ja sitä merkitään usein kirjaimella λ.

Kaikki alkujakaumat λ määrittelevät todennäköisyysmittojen perheen $\{\mathbb{P}_\lambda : \lambda \text{ on todennäköisyysjakauma}\}$. Jos alkujakauma on keskittynyt tilaan $x \in E$ eli $\mathbb{P}(X_0 = x) = 1$, merkitään todennäköisyysmittaa $\mathbb{P}_x$. Tilaa x nimitetään *alkutilaksi*. Todennäköisyysjakauman $\mathbb{P}_\lambda$ tai $\mathbb{P}_x$ suhteen määritettyä odotusarvoa merkitään vastaavasti $\mathbb{E}_\lambda$ tai $\mathbb{E}_x$.

Edellä oletettiin, että Markovin ketju eli stokastinen prosessi $(X_t : t \geq 0)$ ja todennäköisyysavaruus $(\Omega, \mathcal{F}, \mathbb{P})$ ovat ennalta annettuja. Toisinaan, kuten konstruoitaessa Metropolisin ja Hastingsin algoritmiin perustuva Markovin ketju, lähtökohtana on se, että siirtymätodennäköisyysydin P ja alkujakauma λ ovat annettuja. Siten kysymys kuuluukin, onko olemassa sellaista todennäköisyysavaruutta $(\Omega, \mathcal{F}, \mathbb{P}_\lambda)$ ja Markovin ketjua $(X_t : t \geq 0)$, että ydin

P määrittelee Markovin ketjun siirtymätodennäköisyydet ja λ sen alkujakauman. Vastaus kysymykseen on positiivinen. Lukija voi halutessaan lukea konstruktion idean kirjasta [**Num84**] sivu 6, [**MT93**] luku 3.4.1 tai [**Rev75**] sivut 16–18.

Epänegatiivista funktiota λ nimitetään *subtodennäköisyystiheysfunktioksi*, jos $\int \lambda(x)dx \leq 1$. Mikäli $\int \lambda(x)dx = 1$, on λ *todennäköisyystiheysfunktio*. Tulemme jatkossa olettamaan, että jakaumat ja ytimet ovat riittävän säännöllisiä, jotta ne voidaan lausua subtodennäköisyystiheysfunktion ja pistemassan avulla. Merkitään todennäköisyyttä pysyä tilassa x

$$r(x) \triangleq \mathbb{P}(X_{t+1} = x \mid X_t = x),$$

Markovin ketjun siirtymää tilasta $X_t = x$ tilaan $X_{t+1} = y \neq x$ nimitetään *hypyksi*. Todennäköisyys siirtyä tilajoukkoon $A \in \mathcal{E}$ lausutaan subtodennäköisyystiheysfunktion p avulla muodossa

$$\mathbb{P}(X_{t+1} \in A \mid X_t = x) = \int_A p(x,y)dy.$$

Hypyn kokonaistodennäköisyys eli todennäköisyys hypätä tilasta x mihin tahansa tilaan y on

$$\int p(x,y)dy = 1 - r(x).$$

Edellisten avulla ja merkitsemällä pistemassaa $\delta_x(A)$ Markovin ketjun siirtymätodennäköisyysydin voidaan esittää muodossa

$$P(x,A) = \int_A p(x,y)dy + r(x)\delta_x(A).$$

Pidetään mielessä, että subtodennäköisyystiheysfunktio p voi yleisesti koostua jatkuvasta ja diskreetistä osasta. Integrointi suoritetaan suhteessa tulomittaan.

Oletetaan, että satunnaismuuttujalla X_t on tiheysfunktio λ. Tällöin satunnaismuuttuja X_{t+1} jakautuu kaavan (1) nojalla kuten

$$\begin{aligned}
\lambda P(A) &= \int \lambda(dx)P(x,A) \\
&= \int \lambda(x)\left\{\int_A p(x,y)dy + r(x)\delta_x(A)\right\}dx \\
&= \int\int_A \lambda(x)p(x,y)dydx + \int \lambda(x)r(x)\delta_x(A)dx
\end{aligned}$$

Integrointijärjestystä vaihtamalla ja vaihtamalla jälkimmäisessä integraalissa muuttujan nimeä

$$\begin{aligned}
\lambda P(A) &= \int_A\int \lambda(x)p(x,y)dxdy + \int_A \lambda(y)r(y)dy \\
&= \int_A\left\{\int \lambda(x)p(x,y)dx + \lambda(y)r(y)\right\}dy. \qquad (4)
\end{aligned}$$

Voimme siten määritellä jakauman λP tiheysfunktion

$$\lambda P(y) \triangleq \int \lambda(x)p(x,y)dx + \lambda(y)r(y).$$

Lause 2.4. Mikäli λ on alkujakauma ja $t \geq 0$, siirtymätodennäköisyysytimen P iteraatit antavat *yleistetyt* siirtymätodennäköisyydet

$$\lambda P^t(A) = \mathbb{P}_\lambda(X_t \in A).$$

Todistus. Todistetaan väite induktiolla. Tapauksessa $t = 0$ iteraatti P^0 on identiteettiydin ja täten $\lambda P^0(A) = \lambda(A)$. Tämä on yhtä kuin $\mathbb{P}_\lambda(X_0 \in A)$. Oletetaan, että väite pätee yleisellä $t \geq 0$. Käyttämällä kokonaistodennäköisyyden kaavaa ja ehdollistamalla

$$\begin{aligned}
\mathbb{P}_\lambda(X_{t+1} \in A) &= \int \mathbb{P}_\lambda(X_{t+1} \in A, X_t \in dx) \\
&= \int \mathbb{P}(X_{t+1} \in A \mid X_t = x)\mathbb{P}_\lambda(X_t \in dx) \\
&= \int P(x, A)\lambda P^t(dx).
\end{aligned}$$

Käyttämällä kaavoja (1) ja (2) sekä vaihtamalla integrointijärjestys seuraa

$$\begin{aligned}
\int P(x, A)\lambda P^t(dx) &= \int P(x, A)\left\{\int \lambda(dy)P^t(y, dx)\right\} \\
&= \int \lambda(dy)\left\{\int P^t(y, dx)P(x, A)\right\} \\
&= \int \lambda(dy)P^{t+1}(y, A) \\
&= \lambda P^{t+1}(A).
\end{aligned}$$

□

Esimerkki 2.1. Oletetaan, että jono $(b_x : x \in \mathbb{N}^+)$ määrittelee todennäköisyysjakauman. Määritellään diskreetti Markovin ketju $(V_t : t \geq 0)$ tila-avaruudella $E = \mathbb{N}^+$ määrittelemällä diskreetistä teoriasta tuttu äärettömän kokoinen siirtymätodennäköisyysmatriisi

$$P = (p_{x,y}) = \begin{pmatrix} b_1 & b_2 & b_3 & \cdots \\ 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Oletetaan lisäksi, että jono $(a_x : i \geq 1)$ on todennäköisyysjakauma, joka määrittelee Markovin ketjun alkujakauman eli $\mathbb{P}(V_0 = x) = a_x$. Markovin ketjun dynamiikka on hyvin yksinkertainen. Mikäli Markovin ketju on ajan hetkellä t tilassa 1, hyppää Markovin ketju tilaan x todennäköisyydellä b_x. Mikäli Markovin ketju on tilassa $x > 1$, etenee ketju deterministisesti tilaan $x - 1$.

Esimerkki 2.2. Oletetaan, että $Z_0 = 0$ todennäköisyydellä yksi. Olkoon $(X_t : t \geq 1)$ jono riippumattomia ja samoin jakautuneita satunnaismuuttujia. Määritellään

$$Z_t = Z_0 + \sum_{s=1}^{t} X_s, \quad \text{kun } t \geq 1.$$

Stokastista prosessia $(Z_t : t \geq 0)$ nimitetään *satunnaiskuluksi* ja satunnaismuuttujia X_t *inkrementeiksi*. Konstruktion nojalla pätee $Z_t = Z_{t-1} + X_t$, mistä seuraa, että (Z_t) on Markovin ketju. Sen siirtymätodennäköisyysydin on

$$\begin{aligned} P(x, A) &= \mathbb{P}(Z_t \in A \mid Z_{t-1} = x) \\ &= \mathbb{P}(x + X_t \in A \mid Z_{t-1} = x) = \mathbb{P}(X_t \in A - x), \end{aligned}$$

jossa määritellään $A - x \triangleq \{y \in E : x + y \in A\}$.

1.3. Tasapainojakauma.

MÄÄRITELMÄ 2.2. Todennäköisyysjakauma π on Markovin ketjun (X_t) *tasapainojakauma*, jos on voimassa *tasapainoehto*

$$\pi P(A) = \pi(A), \quad \text{jokaisella } A \in \mathcal{E}.$$

Diskreetillä tila-avaruudella E tasapainojakauma π voidaan tulkita vaakavektoriksi $\pi = (\pi_x : x \in E)$ ja tasapainoehto matriisiyhtälöksi $\pi P = \pi$ tai komponenteittain

$$\pi_y = \sum_{x \in E} \pi_x p_{xy} \quad \text{kullakin } y \in E. \tag{5}$$

Kysymys tasapainojakauman olemassaolosta on itsessään mielenkiintoinen, mutta tämän tutkielman kannalta epäoleellinen, sillä sovellettaessa Markovin ketjujen teoriaa Metropolisin ja Hastingsin algoritmiin tasapainojakauma on jo lähtökohtaisesti olemassa. Luvussa 3.4 muotoillaan tasapainojakauman yksikäsitteisyyttä koskeva tulos. Tasapainojakaumia käsittelee laajemmin [**Num84**] luku 5.2 tai [**MT93**] luku 10.

Esittelemme seuraavaksi erään käytännössä hyödylliseksi osoittautuvan ehdon, jonka avulla annettu jakauma on tietyssä erikoistapauksessa helppo osoittaa tasapainojakaumaksi.

MÄÄRITELMÄ 2.3. Markovin ketju (X_t) on *kääntyvä*, jos löytyy sellainen tilajakauma, jolla on tiheysfunktio λ ja kaikilla tiloilla x ja y pätee

$$\lambda(x)p(x, y) = \lambda(y)p(y, x). \tag{6}$$

LAUSE 2.5. Kääntyvällä Markovin ketjulla on tasapainojakauma, nimittäin se jakauma, joka toteuttaa kääntyvyysehdon (6).

TODISTUS. Olkoon λ kääntyvyysehdon toteuttavan jakauman tiheysfunktio. Integroimalla kääntyvyysehto puolittain yli tila-avaruuden

$$\int \lambda(x)p(x, y)dx = \lambda(y) \int p(y, x)dx = \lambda(y)[1 - r(y)],$$

josta seuraa

$$\lambda(y) = \int \lambda(x)p(x, y)dx + \lambda(y)r(y) = \lambda P(y). \qquad \square$$

ESIMERKKI 2.3. Jatketaan esimerkin 2.1 Markovin ketjun (V_t) tarkastelua. Tasapainoehdosta (5) saadaan kullakin $x \geq 1$ yhtälö $\pi_x = b_x \pi_1 + \pi_{x+1}$. Näistä voidaan ratkaista kullakin $x \geq 2$

$$\pi_x = \pi_1(1 - b_1 - \cdots - b_{x-1}) = \pi_1 \sum_{y=x}^{\infty} b_y = \pi_1 B_{x-1},$$

jossa on määritelty jakauman b häntätodennäköisyydet $B_x = \sum_{y=x+1}^{\infty} b_y$. Sovitaan lisäksi $B_0 = 1$. Jotta π olisi todennäköisyysjakauma, tulee olla $(1 + B_1 + B_2 + \cdots)\pi_1 = 1$. Mutta $(B_x : x \geq 0)$ ovat todennäköisyysjakauman häntätodennäköisyydet, joten niiden summa on jakauman odotusarvo $M = \sum_{x=1}^{\infty} x b_x$. Tämä tarkoittaa, että Markovin ketjulla (V_t) on tasapainojakauma π, jolle $\pi_x = M^{-1} B_{x-1}$ jokaisella $x \geq 1$.

1.4. Ergodisuusoletus.

Tila-avaruuden osajoukko $A \in \mathcal{E}$ on *saavutettavissa* tilasta $x \in E$ käsin ja merkitsemme $x \to A$, mikäli $P^t(x, A) > 0$ jollakin $t \geq 1$.

Määritelmä 2.4. Markovin ketju (X_t) toteuttaa *pelkistymättömyys- ja minorisaatio-oletuksen*, jos löytyy positiivismittainen joukko $C \in \mathcal{E}$, positiivismittainen joukko $D \in \mathcal{E}$ ja luku $\beta > 0$, joille $p(x, y) \geq \beta$ kaikilla tiloilla $x \in C$ ja $y \in D$, sekä $x \to C$ jokaisella $x \in E$. Joukkoa C nimitetään *pieneksi*.

Yleisen teorian mukaan Markovin ketju on pelkistymätön, jos löytyy sellainen mitta φ, että $x \to A$ jokaisella $x \in E$ aina, kun $\varphi(A) > 0$. Pelkistymättömyys- ja minorisaatio-oletuksen toteuttava Markovin ketju on pelkistymätön yleisen teorian mielessä valitsemalla

$$\varphi(A) = \mathrm{vol}(A \cap C).$$

Yleisen teorian mielessä määritelty pelkistymättömyys johtaa tarkasteluihin, jotka eivät ole olennaisia tämän tutkielman kannalta. Lukija voi tutustua aiheeseen kirjasta [**Num84**] luku 2.2 tai [**MT93**] luku 4.

Määritelmä 2.5. Markovin ketju (X_t) toteuttaa *ergodisuusoletuksen*, mikäli se toteuttaa pelkistymättömyys- ja minorisaatio-oletuksen ja löytyy tasapainojakauma π. Ergodisuusoletuksen toteuttavaa Markovin ketjua nimitetään *ergodiseksi*.

Määritelmä on yleisen teorian määritelmää vahvempi, kts. [**Num84**] s. 114. Jokainen ergodisuusoletuksen toteuttava Markovin ketju on ergodinen yleisen teorian mielessä.

Olennainen osa ergodisuuden määritelmää yleisessä teoriassa on, että Markovin ketju on jaksoton. Markovin ketju (X_t) on jaksollinen, mikäli tila-avaruus voidaan osittaa m erilliseen joukkoon $E_1, \ldots, E_m$ siten, että Markovin ketju siirtyy todennäköisyydellä yksi joukosta E_i joukkoon E_{i+1}, kun $1 \leq i < m$ ja todennäköisyydellä yksi joukosta E_m joukkoon E_1. Pelkistymättömyys- ja minorisaatio-oletuksen muotoilu on riittävän vahva, jotta sen toteuttava Markovin ketju on *jaksoton* eli ei löydy edellä kuvattua tila-avaruuden hajotelmaa millään $m \geq 2$. Pelkistymättömyys- ja minorisaatio-oletuksen toteuttava Markovin ketju nimittäin toteuttaa [**Num84**] s. 15 minorisaatioehdon $P(x, A) \geq \beta \mathbf{1}_C(x)\varphi(A)$ kaikilla $x \in E$ ja $A \in \mathcal{E}$ valitsemalla $\varphi(A) = \mathrm{vol}(A \cap D)$, jolloin [**Num84**] s. 20–21 nojalla Markovin ketju on jaksoton.

Ergodisuusoletus on riittävän vahva, jotta siitä seuraa hyvin rikas jatkuvien Markovin ketjujen teoria. Siitä huolimatta hyvin laaja joukko Markovin ketjuja toteuttaa oletuksen ehdot.

Lause 2.6. Mikäli Markovin ketjulla (X_t) on siirtymätodennäköisyysydin P, jonka subtodennäköisyystiheysfunktio p on aidosti positiivinen ja jatkuva jossakin pisteessä $(x_0, y_0) \in E \times E$, ja lisäksi $x \to B(x_0, r)$ jokaisella $x \in E$ ja jokaisella $r > 0$, niin Markovin ketju toteuttaa pelkistymättömyys- ja minorisaatio-oletuksen.

Todistus. Merkitään $p(x_0, y_0) = \gamma > 0$ ja valitaan luku $\varepsilon > 0$ siten, että $\gamma - \varepsilon > 0$. Avoin väli $(\gamma - \varepsilon, \gamma + \varepsilon)$ on pisteen γ ympäristö, joten jatkuvuuden määritelmän nojalla alkukuva $N = p^{-1}(\gamma - \varepsilon, \gamma + \varepsilon)$ on pisteen (x_0, y_0) ympäristö tuloavaruudella $E \times E$. [**Men90**] s. 98 nojalla löytyy pisteiden x_0 ja y_0 ympäristöt N_1 ja N_2, joille $N_1 \times N_2 \subset N$. Edelleen löytyy avoimet joukot $C \subset N_1$ sekä $D \subset N_2$ siten että, $x_0 \in C$ ja $y_0 \in D$. Joukot C ja D ovat avoimina positiivismittaisia. Mikäli $x \in C$ ja $y \in D$, pätee $(x, y) \in N$ ja siten $p(x, y) > \gamma - \varepsilon$. Oletuksesta $x \to B(x_0, r)$ jokaisella $x \in E$ ja $r > 0$ seuraa, että $x \to C$. Pelkistymättömyys- ja minorisaatio-oletus on siten voimassa valitsemalla $\beta = \gamma - \varepsilon$. □

Esimerkki 2.4. Valitaan esimerkissä 2.3 Markovin ketjun (V_t) joukoksi C yksiö $\{1\}$. Todennäköisyysjakauman b tuki $\operatorname{supp} b$ muodostuu kaikista niistä $x \in E$, joille $b_x > 0$. Olkoon D jokin joukon $\operatorname{supp} b$ äärellinen osajoukko. Tämä tarkoittaa, että löytyy luku $\beta = \min_{x \in D} b_x > 0$. Nyt $p(x, y) \geq \beta$, kun $x \in C$ ja $y \in D$. Lisäksi Markovin ketjun konstruktion nojalla $x \to C$ millä hyvänsä $x \in E$. Edellisessä luvussa näytettiin, että Markovin ketjulla (V_t) on tasapainojakauma, joten (V_t) on ergodinen.

Esimerkki 2.5. Valitaan satunnaiskulun (Z_t) inkrementtien X_t jakaumaksi esimerkissä 2.2 standardinormaalijakauma. Tämä merkitsee, että Markovin ketjun (Z_t) siirtymätodennäköisyysydin on

$$P(x, A) = \mathbb{P}(X_t \in A - x) = \frac{1}{\sqrt{2\pi}} \int_A e^{-(y-x)^2/2} dy.$$

Siirtymätodennäköisyystiheys on siten

$$p(x, y) = \frac{1}{\sqrt{2\pi}} e^{-(y-x)^2/2}.$$

Olkoon $\varepsilon > 0$ mielivaltainen ja $C = [-\varepsilon, \varepsilon]$ sekä D mielivaltainen äärellinen suljettu väli. Joukko $C \times D$ on kompakti, joten tiheysfunktion p jatkuvuuden nojalla löytyy sellainen $\beta > 0$, että $p(x, y) \geq \beta$ kaikilla $x \in C$ ja $y \in D$. Koska p on aidosti positiivinen kaikilla $x, y \in \mathbb{R}$, seuraa $P(x, C) > 0$ jokaisella $x \in \mathbb{R}$. Siispä olemme näyttäneet suoraan määritelmästä, että Markovin ketju (Z_t) toteuttaa pelkistymättömyys- ja minorisaatio-oletuksen. Toisaalta väite seuraa lauseesta 2.6 valitsemalla esimerkiksi $x_0 = y_0 = 0$. Sen sijaan esimerkin Markovin ketjulla ei ole tasapainojakaumaa [**MT93**] s. 251–252. Välin $\mathbb{R}$ Lebesguen mitta toteuttaa kylläkin kääntyvyysehdon, mutta kyseessä ei ole todennäköisyysjakauma. Markovin ketju (Z_t) ei siis ole ergodinen.

2. Uusiutumisprosessit

Markovin ketjujen teoria tulee pitkälti nojaamaan diskreettiaikaisten uusiutumisprosessien teoriaan. Sekä diskreetti- että jatkuva-aikaiset uusiutumisprosessit ovat yksi keskeinen stokastisten prosessien luokka. Uusiutumisprosessi voidaan ymmärtää havainnollisesti tarkastelemalla kuvitteellisen komponentin elinikää. Oletetaan, että komponentti on ajan hetkellä $t = 0$ ehjä ja sen tila tarkastetaan kunakin diskreettinä ajan hetkenä $t \in \mathbb{N}$. Mikäli komponentti havaitaan vikaantuneeksi, se korvataan välittömästi uudella komponentilla ja uuden komponentin elinikä on riippumaton siitä kuinka pitkään edellinen komponentti kesti. Otamme huomioon sen, että ensimmäinen komponentti ei ole välttämättä uusi, joten sen elinikä voi noudattaa jakaumaa, joka on eri kuin uuden komponentin. Luvun materiaali nojaa pitkälti kirjaan [**Num84**] s. 47–51.

2.1. Uusiutumisprosessi.

Olkoon $\bar{\mathbb{N}}$-arvoinen satunnaismuuttuja T_0 ja oletetaan, että sen jakauma on $\mathbb{P}(T_0 = t) = a_t$, kun $a = (a_t : t \geq 0)$ on substokastinen jono eli jono, jonka jäsenet ovat epänegatiivisia ja joiden summa on korkeintaan yksi. Merkitään $a_\infty = 1 - \sum_{t=0}^{\infty} a_t$, jolloin $\mathbb{P}(T_0 = \infty) = a_\infty$. Satunnaismuuttuja T_0 tulkitaan ensimmäiseksi vikaantumishetkeksi tai *ensimmäiseksi uusiutumisajaksi*.

Määritellään jono $(\tau_i : i \geq 1)$ riippumattomia ja samoin jakautuneita $\bar{\mathbb{N}}$-arvoisia satunnaismuuttujia, joiden jakauma $\mathbb{P}(\tau = t) = b_t$, kun $b = (b_t : t \geq 0)$ on substokastinen jono, jolle $b_t = 0$ ja $b_\infty = 1 - \sum_{t=0}^{\infty} b_t$. Määritellään $\bar{\mathbb{N}}$-arvoiset satunnaismuuttujat $(T_i : i \geq 1)$ asettamalla

$$T_i = T_{i-1} + \tau_i \quad \text{jokaisella } i \geq 1.$$

Jonoa $(T_i : i \geq 0)$ nimitetään *uusiutumisprosessiksi*. Mikäli $T_0 = 0$ todennäköisyydellä yksi eli $a_0 = 1$, uusiutumisprosessia nimitetään *viiveettömäksi*. Muutoin uusiutumisprosessi on *viiveellinen*. Oletuksesta $b_0 = 0$ seuraa, että satunnaismuuttujat $(T_i : i \geq 0)$ muodostavat aidosti kasvavan jonon eli $T_i > T_{i-1}$ todennäköisyydellä yksi kaikilla $i \geq 1$.

Satunnaismuuttuja T_0 voidaan tulkita ensimmäisen komponentin eliniäksi ja kukin τ_i edellisen rikkoontuneen komponentin tilalle asennetun uuden komponentin eliniäksi. Mikäli uusiutumisprosessi on viiveetön, voidaan konstruktion valossa ajatella, että ensimmäinen komponentti rikkoontuu saman tien ja korvataan uudella, jolloin ajan hetkellä $t = 0$ prosessi käynnistyy tilasta, jossa komponentti on uusi.

Määritellään jono $(Y_t : t \geq 0)$ $\{0, 1\}$-arvoisia satunnaismuuttujia määrittelemällä

$$Y_t \triangleq \begin{cases} 1, & \text{jos } T_i = t \text{ jollakin } i \geq 0 \\ 0, & \text{muuten} \end{cases}.$$

Stokastista prosessia (Y_t) nimitetään uusiutumisprosessin $(T_i : i \geq 0)$ *insidenssiprosessiksi*. Satunnaismuuttuja Y_t saa arvon yksi, jos ajan hetkellä t tapahtuu uusiutuminen. Mikäli uusiutumisprosessi on viiveetön, merkitään $u_t \triangleq \mathbb{P}(Y_t = 1)$, ja mikäli prosessi on viiveellinen, merkitään $v_t \triangleq \mathbb{P}(Y_t = 1)$. Sovitaan $u_0 = 1$. Määritetään u_t, kun $t \geq 1$. Pätee

$$\begin{aligned} u_t &= \mathbb{P}(\tau = t) + \sum_{s=1}^{t-1} \mathbb{P}(Y_{t-s} = 1, \tau = s) \\ &= b_t + \sum_{s=1}^{t-1} \mathbb{P}(Y_{t-s} = 1)\mathbb{P}(\tau = s) = b_t + \sum_{s=1}^{t-1} u_{t-s} b_s. \end{aligned}$$

Käyttämällä hyväksi tietoa $u_0 = 1$ ja $b_0 = 0$ voidaan edellinen kirjoittaa konvoluution (kts. liite A) avulla muodossa $u_t = b \star u_t$. Käyttämällä yksikköjonoa $\delta = (1, 0, 0, \dots)$ voidaan ottaa huomioon myös tapaus $t = 0$. Seuraa niin sanottu *viiveetön uusiutumisyhtälö*

$$u = \delta + b \star u.$$

Edellistä periaatetta noudattaen on helppo näyttää, että v toteuttaa *viiveellisen uusiutumisyhtälön*

$$v = a + b \star v.$$

Jonoa u nimitetään *viiveettömäksi uusiutumisjonoksi* ja jonoa v nimitetään *viiveelliseksi uusiutumisjonoksi*.

Uusiutumisprosessi $(T_i : i \geq 0)$ on *jaksoton*, jos siihen liittyvälle inkrementtijakaumalle b pätee $\gcd\{t \geq 1 : b_t > 0\} = 1$. Tässä tutkielmassa tarkastellaan ainoastaan jaksottomia uusiutumisprosesseja.

MÄÄRITELMÄ 2.6. Viiveetön uusiutumisprosessi (T_i) on *palautuva*, jos jokaisella $i \geq 0$ pätee $T_i < \infty$ todennäköisyydellä yksi. Uusiutumisprosessi on positiivisesti palautuva, jos $\mathbb{E}(\tau) = \sum_{t=1}^{\infty} t b_t < \infty$.

LAUSE 2.7. Viiveetön uusiutumisprosessi (T_i) on palautuva, jos ja vain jos $\sum_{t=0}^{\infty} b_t = 1$ tai yhtäpitävästi $\sum_{t=0}^{\infty} u_t = \infty$.

TODISTUS. Palautuvuus on yhtäpitävää sen kanssa, että $\mathbb{P}(\tau < \infty) = 1$ tai yhtäpitävästi $\sum_{t=0}^{\infty} b_t = 1$. Riittää siis osoittaa, että $\sum_{i=1}^{\infty} b_t = 1$ tasan siinä tapauksessa, että $\sum_{i=1}^{\infty} u_t = \infty$. Oletetaan nyt, että $\sum_{t=0}^{\infty} b_t = 1$. Määritellään jonon b generoiva funktio $B(x) = \sum_{t=0}^{\infty} b_t x^t$ sekä jonon u generoiva funktio $U(x) = \sum_{t=0}^{\infty} u_t x^t$. Käyttämällä hyväksi uusiutumisyhtälöä ja liitteen A tulosta konvoluution generoivalle funktiolle seuraa

$$U(x) = 1 + B(x)U(x).$$

Mutta oletuksen nojalla $B(1) = \sum_{t=0}^{\infty} b_t = 1$, joten yhtälö $U(1) = 1 + U(1)$ voi pitää paikkansa vain siinä tapauksessa, että $U(1) = \infty$. Toisaalta pätee $B(1) = 1 - U(1)^{-1}$, joten mikäli $U(1) = \infty$, seuraa $B(1) = 1$. □

Määritellään ajan hetkeen $t \geq 0$ mennessä sattuneiden insidenssien lukumäärä $N(t)$ asettamalla

$$N(t) \triangleq \sum_{s=0}^{t} Y_s.$$

LAUSE 2.8. Olkoon $(T_i : i \geq 0)$ positiivisesti palautuva viiveetön uusiutumisprosessi. Sille on voimassa

$$\lim_{t \to \infty} \frac{N(t)}{t} = \mathbb{E}(\tau)^{-1} \quad \text{m.v.}$$

Lauseen todistus nojaa kirjan [**MO14**] teoreeman 1.1 todistukseen.

TODISTUS. Viiveettömän uusiutumisprosessin määritelmän ja suurten lukujen lain nojalla

$$\frac{T_i}{i} = \frac{1}{i} \sum_{j=1}^{i} \tau_j \overset{\text{m.v.}}{\to} \mathbb{E}(\tau). \tag{7}$$

Pätee $T_{N(t)} \leq t < T_{N(t)+1}$. Jakamalla puolittain satunnaismuuttujalla $N(t)$ seuraa

$$\frac{T_{N(t)}}{N(t)} \leq \frac{t}{N(t)} < \frac{T_{N(t)+1}}{N(t)+1} \frac{N(t)+1}{N(t)} \tag{8}$$

Palautuvuuden nojalla $N(t) \to \infty$ melkein varmasti, kun $t \to \infty$. Siten $\frac{N(t)+1}{N(t)} \to 1$ melkein varmasti ja yhtälön (7) nojalla epäyhtälön (8) vasen ja oikea puoli suppenevat kohti odotusarvoa $\mathbb{E}(\tau)$ melkein varmasti, kun $t \to \infty$. □

2.2. Keskeinen konvergenssilause.

Tarkastellaan *kytkentäargumenttia* nojaten kirjan [**MT93**] sivuihin 320–322. Oletetaan, että $(T_i : i \geq 0)$ on viiveetön uusiutumisprosessi inkrementtijakaumanaan b sekä $(\tilde{T}_i : i \geq 0)$ viiveellinen uusiutumisprosessi viivejakaumanaan a ja inkrementtijakaumanaan b. Viiveettömän ja viiveellisen uusiutumisprosessin insidenssiprosessit ovat vastaavasti $(Y_t : t \geq 0)$ ja $(\tilde{Y}_t : t \geq 0)$. Uusiutumisprosessit *kytkeytyvät*, jos löytyy sellainen $t \geq 0$, jolle $Y_t = \tilde{Y}_t$. Määritellään *kytkentäaika*

$$T_{ab} = \min\{t \geq 0 : Y_t = \tilde{Y}_t\}.$$

Lemma 2.9. Mikäli uusiutumisprosessi $(T_i : i \geq 0)$ ja jaksoton ja positiivisesti palautuva ja a on inkrementtijakauma, niin $\mathbb{P}(T_{ab} < \infty) = 1$.

Lemman todistus on jossakin määrin tekninen eikä ole erityisen kiinnostava esityksen kannalta. Todistuksen eräs versio on esitetty teoksessa [**Num84**] sivut 99–100 ja toinen diskreettien Markovin ketjujen teoriaan perustuva versio kirjassa [**MT93**] sivu 321.

Lemma 2.10. Mikäli uusiutumisprosessi $(T_i : i \geq 0)$ on jaksoton ja positiivisesti palautuva ja a on inkrementtijakauma, pätee

$$\lim_{t \to \infty} |a \star u_t - u_t| = 0.$$

Todistus. Määritellään stokastinen prosessi $(Z_t : t \geq 0)$ asettamalla

$$Z_t = \begin{cases} \tilde{Y}_t, & t < T_{ab} \\ Y_t, & t \geq T_{ab} \end{cases} .$$

Koska kytkennän jälkeen Y_t ja $\tilde{Y}_t$ ovat samoin jakautuneet, seuraa, että Z_t on jakautunut kuten $\tilde{Y}_t$ kullakin $t \geq 0$. Nyt

$$\begin{aligned} |a \star u_t - u_t| &= |\mathbb{P}(\tilde{Y}_t = 1) - \mathbb{P}(Y_t = 1)| \\ &= |\mathbb{P}(Z_t = 1) - \mathbb{P}(Y_t = 1)| \\ &= |\mathbb{P}(Z_t = 1, T_{ab} > t) + \mathbb{P}(Z_t = 1, T_{ab} \leq t) \\ &\qquad - \mathbb{P}(Y_t = 1, T_{ab} > t) - \mathbb{P}(Y_t = 1, T_{ab} \leq t)| \end{aligned}$$

Mutta nyt $\mathbb{P}(Z_t = 1, T_{ab} \leq t) = \mathbb{P}(Y_t = 1, T_{ab} \leq t)$, joten seuraa

$$\begin{aligned} |a \star u_t - u_t| &= |\mathbb{P}(\tilde{Y}_t = 1, T_{ab} > t) - \mathbb{P}(Y_t = 1, T_{ab} > t)| \\ &= |\mathbb{P}(\tilde{Y}_t = 1 \mid T_{ab} > t) - \mathbb{P}(Y_t = 1 \mid T_{ab} > t)|\mathbb{P}(T_{ab} > t) \\ &\leq \mathbb{P}(T_{ab} > t) \end{aligned}$$

Raja-arvoa koskeva väite seuraa lemmasta 2.9, sillä $\mathbb{P}(T_{ab} > t) \to 0$ rajalla $t \to \infty$. $\square$

Seuraava tulos on erikoistapaus [**MT93**] lemmasta D.7.1.

LEMMA 2.11. Oletetaan, että $(x_i : i \geq 0)$ on epänegatiivinen jono, jolle $x_i \to 0$ sekä $(y_i : i \geq 0)$ on epänegatiivinen jono, jolle $\sum_{i=0}^{\infty} y_i < \infty$. Tällöin $x \star y_i \to 0$, kun $i \to \infty$.

TODISTUS. Konvoluution raja-arvo voidaan esittää muodossa

$$\lim_{i\to\infty} x \star y_i = \lim_{i\to\infty} \sum_{j=0}^{\infty} x_{i-j} \mathbf{1}(i \geq j) y_j.$$

Mutta oletuksen nojalla $x_{i-j}\mathbf{1}(i \geq j)$ on välttämättä rajoitettu, joten dominoidun konvergenssin nojalla

$$\lim_{i\to\infty} x \star y_i = \sum_{j=0}^{\infty} y_j \lim_{i\to\infty} x_{i-j} \mathbf{1}(i \geq j) = 0. \qquad \square$$

Määritellään nyt satunnaismuuttujan τ häntätodennäköisyydet määrittelevä jono $B = (B_t : t \geq 0)$ asettamalla $B_t \triangleq \mathbb{P}(\tau > t)$.

Seuraava lause on sama kuin teoksen [**Num84**] teoreema 6.1 sillä lisäoletuksella, että inkrementtijakaumalta vaaditaan positiivinen palautuvuus pelkän palautuvuuden sijaan. Lauseen todistus on helppo edellä käsitellyn kytkentäargumentin nojalla.

LAUSE 2.12. Oletetaan, että inkrementtijakauma b on jaksoton ja positiivisesti palautuva sekä a on viivejakauma. Tällöin

$$\lim_{t\to\infty} |a \star u - u| \star B_t = 0.$$

TODISTUS. Väite seuraa lemmoista 2.10 ja 2.11 havaitsemalla, että oletuksen nojalla

$$\infty > \mathbb{E}(\tau) = \sum_{t=1}^{\infty} \mathbb{P}(\tau > t) = \sum_{t=1}^{\infty} B_t. \qquad \square$$

3. Regeneraatio

Tässä luvussa tarkastellaan teoreettisia tuloksia, jotka yhdessä edellisen luvun uusiutumisprosessien kanssa mahdollistavat kolmen keskeisen konvergenssituloksen todistamisen luvussa 3. Esitetään seuraavaksi välttämättömät Markovin ketjua koskevat oletukset, jotka ovat voimassa sekä tässä että tulevassa luvussa. Luvun materiaali perustuu artikkeliin [**Num02**].

Oletetaan, että on annettu ergodinen Markovin ketju (X_t), jolla on siirtymätodennäköisyysydin P. Erityisesti oletamme, että positiivismittaiset joukot $C \in \mathcal{E}$ ja $D \in \mathcal{E}$ sekä luku $\beta > 0$ on kiinnitetty. Lukua β tai joukkoa D voi tarvittaessa pienentää, joten oletamme lisäksi $0 < \beta \mathrm{vol}(D) \leq 1$.

Oletetaan lisäksi, että tasapainojakauman tuki $\operatorname{supp} \pi = E$. Tämä merkitsee, että jokaisen positiivismittaisen joukon $A \in \mathcal{E}$ tasapainotodennäköisyys $\pi(A) > 0$. Tämä ei ole millään muotoa rajoittava vaatimus, sillä mikäli Markovin ketjulla on tasapainojakauma, jonka tuki on tila-avaruuden E aito osajoukko ja lisäksi millään $x \in \operatorname{supp} \pi$ ei ole mahdollista päästä joukkoon $E \setminus \operatorname{supp} \pi$, voi tarkastella Markovin ketjun rajoittumaa joukkoon $\operatorname{supp} \pi$.

3.1. Bivariaatti Markovin ketju ja regeneraatio.

Määritellään funktio $s : E \to \mathbb{R}^+$ asettamalla $s(x) \triangleq \beta \mathrm{vol}(D)\mathbf{1}_C(x)$. Joukkoon D tasaisesti jakautuneen satunnaismuuttujan jakaumaa merkitään ν. Sen tiheysfunktio on $\nu(y) = \mathrm{vol}(D)^{-1}\mathbf{1}_D(y)$ ja tapahtuman $A \in \mathcal{E}$ todennäköisyys on $\nu(A) = \mathrm{vol}(D)^{-1}\mathrm{vol}(A \cap D)$. Ergodisuusoletuksen nojalla kaikilla $x, y \in E$ on voimassa

$$p(x,y) \geq s(x)\nu(y) = \beta \mathbf{1}_C(x)\mathbf{1}_D(y). \tag{9}$$

Määritellään substokastinen ydin

$$Q(x,A) \triangleq P(x,A) - s(x)\nu(A).$$

Määritelmä on hyvin asetettu, sillä ytimen Q mitallisuusominaisuudet seuraavat suoraan siitä, että s on mitallinen ja ν on mitta, ja lisäksi kaavan (9) nojalla

$$\begin{aligned} Q(x,A) &= \int_A p(x,y)dy + r(x)\delta_x(A) - s(x)\int_A \nu(y)dy \\ &\geq \int_A \{p(x,y) - s(x)\nu(y)\}\, dy \geq 0. \end{aligned}$$

Olkoon $\{0,1\}$-arvoinen stokastinen prosessi $(Y_t : t \geq 0)$ ja määritellään stokastinen prosessi $(X_t, Y_t : t \geq 0)$ antamalla sen siirtymätodennäköisyydet

$$\mathbb{P}(X_t \in A, Y_t = 1 \mid X_{t-1}, Y_{t-1}, \dots, X_0, Y_0) = s(X_{t-1})\nu(A), \tag{10}$$

$$\mathbb{P}(X_t \in A, Y_t = 0 \mid X_{t-1}, Y_{t-1}, \dots, X_0, Y_0) = Q(X_{t-1}, A). \tag{11}$$

Siirtymätodennäköisyyksien määritelmästä nähdään välittömästi, että stokastinen prosessi (X_t, Y_t) on Markovin ketju. Siitä käytetään nimitystä *bivariaatti* Markovin ketju. Suora laskutoimitus antaa bivariaatin Markovin ketjun ehdolliset marginaalijakaumat

$$\begin{aligned} &\mathbb{P}(X_t \in A \mid X_{t-1} = x, Y_{t-1} = y) = P(x,A), \\ &\mathbb{P}(Y_t = 1 \mid X_{t-1} = x, Y_{t-1} = y) = s(x), \\ &\mathbb{P}(Y_t = 0 \mid X_{t-1} = x, Y_{t-1} = y) = 1 - s(x). \end{aligned} \tag{12}$$

Tarkastellaan seuraavaksi tilannetta, jossa bivariaatti Markovin ketju (X_t, Y_t) saa ajan hetkellä t arvon $Y_t = 1$. Tarkastellaan bivariaatin Markovin ketjun kehittymistä tästä hetkestä lähtien.

Lemma 2.13. Pätee

$$\mathbb{P}(X_t \in A \mid X_{t-1} = x, Y_{t-1} = y, Y_t = 1) = \nu(A),$$

sekä

$$\begin{aligned} &\mathbb{P}(X_t \in A_0, X_{t+1} \in A_1, Y_{t+1} = y_1, \cdots \mid X_{t-1} = x, Y_{t-1} = y, Y_t = 1) \\ &\quad = \mathbb{P}_\nu(X_0 \in A_0, X_1 \in A_1, Y_1 = y_1, \dots). \end{aligned}$$

Todistus. Käyttämällä ehdollisen todennäköisyyden kaavaa ja yhtälöitä (10) sekä (12)

$$\begin{aligned} &\mathbb{P}(X_t \in A \mid X_{t-1} = x, Y_{t-1} = y, Y_t = 1) \\ &\quad = \frac{\mathbb{P}(X_t \in A, Y_t = 1 \mid X_{t-1} = x, Y_{t-1} = y)}{\mathbb{P}(Y_t = 1 \mid X_{t-1} = x, Y_{t-1} = y)} = \nu(A). \end{aligned}$$

Merkitään $B \triangleq \{X_{t+1} \in A_1, Y_{t+1} = y_1, \dots\}$. Hyödyntämällä ehdollisen todennäköisyyden kaavaa, bivariaatin Markovin ketjun siirtymätodennäköisyyksien määritelmää sekä ehdollista marginaalijakaumaa

$$
\begin{aligned}
&\mathbb{P}(X_t \in A_0, X_{t+1} \in A_1, Y_{t+1} = y_1, \cdots \mid X_{t-1} = x, Y_{t-1} = y, Y_t = 1) \\
&\quad = \frac{\mathbb{P}(X_t \in A_0, Y_t = 1, B \mid X_{t-1} = x, Y_{t-1} = y)}{\mathbb{P}(Y_t = 1 \mid X_{t-1} = x, Y_{t-1} = y)} \\
&\quad = \mathbb{P}(B \mid X_{t-1} = x, Y_{t-1} = y, X_t \in A_0, Y_t = 1) \\
&\qquad \times \frac{\mathbb{P}(X_t \in A_0, Y_t = 1 \mid X_{t-1} = x, Y_{t-1} = y)}{s(x)} \\
&\quad = \mathbb{P}(B \mid X_t \in A_0)\nu(A_0).
\end{aligned}
$$

Toisaalta

$$\mathbb{P}(B \mid X_t \in A_0) = \mathbb{P}(X_1 \in A_1, Y_1 = y_1, \cdots \mid X_0 \in A_0)$$

ja $\nu(A_0) = \mathbb{P}_\nu(X_0 \in A_0)$, joten väite seuraa ehdollisen todennäköisyyden kaavasta. $\square$

Lemma merkitsee, että bivariaatti Markovin ketju unohtaa historiansa aina niinä hetkinä, kun $Y_t = 1$. Tästä hetkestä lähtien bivariaatti Markovin ketju käyttäytyy kuten se käynnistyisi uudelleen alkujakaumanaan ν. Sanomme, että Markovin ketju *regeneroituu*. Satunnaismuuttujan Y_t ehdollisesta marginaalijakaumasta nähdään, että regeneraatio voi tapahtua ainoastaan sellaisena ajan hetkenä t, jolle $X_{t-1} \in C$. Regeneraatiotodennäköisyys $s(x) = \beta \mathrm{vol}(D)$ on riippumaton tilasta $x \in C$.

Määritellään bivariaatin Markovin ketjun (X_t, Y_t) perättäiset *regeneraatiohetket* asettamalla

$$
\begin{aligned}
T_0 &\triangleq \min\{t \geq 1 : Y_t = 1\}, \\
T_i &\triangleq \min\{t > T_{i-1} : Y_t = 1\}, \quad \text{kun } i \geq 1.
\end{aligned}
$$

Peräkkäisten regeneraatiohetkien väliset ajat $T_1 - T_0, T_2 - T_1, \dots$ ovat konstruktion nojalla riippumattomia ja samoin jakautuneita. Otetaan käyttöön merkintä $T = T_1 - T_0$.

Regeneraatiohetkille saadaan luonnollinen tulkinta uusiutumisprosessien avulla. Mikäli Markovin ketjun alkutila on x, merkitään ensimmäisen regeneraatiohetken T_0 jakaumaa $\mathbb{P}_x(T_0 = t) = a_t(x)$. Kahden peräkkäisen regeneraation välisen ajan T jakaumaa merkitään $\mathbb{P}_\nu(T = t) = b_t$. Satunnaismuuttujat Y_t määrittelevät uusiutumisprosessin insidenssijakauman.

3.2. Päättyvä Markovin ketju.

Muodostamme seuraavaksi apuneuvon, joka osoittautuu hyödylliseksi tulevissa teoreettisissa tarkasteluissa. Olkoon Δ sellainen *aputila*, joka ei kuulu tila-avaruuteen E. Määritellään *päättyvä* Markovin ketju $(\tilde{X}_t)$ asettamalla $\tilde{X}_t = X_t$, kun $t < T$ ja $\tilde{X}_t = \Delta$, kun $t \geq T$. Markovin ketjun annetaan siis kehittyä tavalliseen tapaan kunnes tapahtuu regeneraatio, jolloin Markovin ketju jää tilaan Δ. Päättyvän Markovin ketjun siirtymätodennäköisyydet ovat yhtälön (11) nojalla muotoa

$$\begin{aligned}
&\mathbb{P}(\tilde{X}_t \in A \mid \tilde{X}_{t-1} = x) \\
&\qquad = \mathbb{P}(X_t \in A, Y_t = 0 \mid X_{t-1} = x, Y_{t-1} = 0) = Q(x, A). \qquad (13)
\end{aligned}$$

Yllä määritellyt siirtymätodennäköisyydet määrittelevät substokastisen ytimen. Stokastinen ydin voidaan muodostaa seuraavasti. Muodostetaan laajennettu tila-avaruus $(E_\Delta, \mathcal{E}_\Delta)$ asettamalla $E_\Delta = E \cup \{\Delta\}$ ja $\mathcal{E}_\Delta = \sigma(\mathcal{E}, \{\Delta\})$. Asettamalla $\mathbb{P}(\tilde{X}_t = \Delta \mid \tilde{X}_{t-1} = \Delta) = 1$ sekä

$$\mathbb{P}(\tilde{X}_t = \Delta \mid \tilde{X}_{t-1} = x) = \mathbb{P}(Y_t = 1 \mid X_{t-1} = x, Y_{t-1} = 0) = s(x) \qquad (14)$$

nähdään, että $\mathbb{P}(\tilde{X}_t \in E_\Delta \mid \tilde{X}_{t-1} = x) = 1$ jokaisella $x \in E_\Delta$, sillä erityisesti, jos $x \in E$,

$$\begin{aligned}
\mathbb{P}(\tilde{X}_t \in E_\Delta \mid \tilde{X}_{t-1} = x) &= \mathbb{P}(\tilde{X}_t \in E \mid \tilde{X}_{t-1} = x) + \mathbb{P}(\tilde{X}_t = \Delta \mid \tilde{X}_{t-1} = x) \\
&= \mathbb{P}(X_t \in E, \Psi_t = 0 \mid X_{t-1} = x, \Psi_{t-1} = 0) + s(x) \\
&= Q(x, E) + s(x) \\
&= 1.
\end{aligned}$$

Stokastinen prosessi $(\tilde{X})$ on hyvin määritelty Markovin ketju, joten siirtymätodennäköisyyksien iteraatteja koskevan tuloksen nojalla saadaan välittömästi seuraava lemma.

Lemma 2.14. Millä tahansa alkujakaumalla λ ja jokaisella $t \geq 1$ sekä $A \in \mathcal{E}$ on voimassa

$$\mathbb{P}_\lambda(\tilde{X}_t \in A) = \lambda Q^t(A).$$

Päättyvää Markovin ketjua hyödyntämällä saadaan tulos

$$\mathbb{P}_\lambda(T > t) = \mathbb{P}_\lambda(\tilde{X}_t \in E) = \lambda Q^t(E). \qquad (15)$$

3.3. Potentiaali.

Substokastisen ytimen Q potentiaali G on ydin (kts. lause 2.2)

$$G = \sum_{t=0}^{\infty} Q^t.$$

Lemma 2.15. Alkujakaumalla λ varustetun Markovin ketjun odotettu vierailujen lukumäärä tilajoukkoon $A \in \mathcal{E}$ ensimmäiseen regeneraatiohetkeen T mennessä on

$$\mathbb{E}_\lambda \left[\sum_{t=0}^{T-1} \mathbf{1}_A(X_t) \right] = \lambda G(A).$$

Yleisemmin millä tahansa mitallisella funktiolla f on voimassa

$$\mathbb{E}_\lambda \left[\sum_{t=0}^{T-1} f(X_t) \right] = \int f(x) \lambda G(dx).$$

Todistus. Pätee

$$\mathbb{E}_\lambda \left[\sum_{t=0}^{T-1} \mathbf{1}_A(X_t) \right] = \sum_{t=0}^{T-1} \mathbb{P}_\lambda(X_t \in A) = \sum_{t=0}^{\infty} \mathbb{P}_\lambda(X_t \in A, T > t).$$

Hyödyntämällä päättyvää Markovin ketjua ja lemmaa 2.14

$$\begin{aligned}\mathbb{E}_\lambda\left[\sum_{t=0}^{T-1}\mathbf{1}_A(X_t)\right] &= \sum_{t=0}^{\infty}\mathbb{P}_\lambda(\tilde{X}_t\in A)\\ &= \sum_{t=0}^{\infty}\int_A \lambda Q^t(x)dx = \int_A \lambda G(dx).\end{aligned}$$

Todistetaan seuraavaksi yleistys. Olkoon funktio g pieni. Toisin sanoen joillakin positiivisilla $a_0,\dots,a_n$ ja mitallisilla $A_0,\dots,A_n$ g on määritelty summana $g(x)=\sum_{i=0}^n a_i\mathbf{1}_{A_i}(x)$. Edellisen nojalla

$$\begin{aligned}\mathbb{E}_\lambda\left[\sum_{t=0}^{T-1}g(X_t)\right] &= \mathbb{E}_\lambda\left[\sum_{t=0}^{T-1}\sum_{i=0}^{n}a_i\mathbf{1}_{A_i}(X_t)\right]\\ &= \sum_{i=0}^{n}a_i\mathbb{E}_\lambda\left[\sum_{t=0}^{T-1}\mathbf{1}_{A_i}(X_t)\right]\\ &= \sum_{i=0}^{n}a_i\int_{A_i}\lambda G(dx)\\ &= \int\sum_{i=0}^{n}a_i\mathbf{1}_{A_i}(x)\lambda G(dx) = \int g(x)\lambda G(dx).\end{aligned}$$

Jos f on mielivaltainen epänegatiivinen ja mitallinen funktio, löytyy sellainen kasvava jono $g_0, g_1, \dots$ yksinkertaisia funktioita, että $g_i \to f$. Väite seuraa monotonisen konvergenssin lauseesta. Yleisellä f väite seuraa edellisestä käyttämällä esitystä $f = f^+ - f^-$. □

Markovin ketjun alkujakauma ν on erityisasemassa regeneroituvien ketjujen teoriaa kehiteltäessä, joten otamme käyttöön seuraavan merkinnän.

Määritelmä 2.7. Mikäli G on ytimen Q potentiaali, määritellään mitta $\mu = \nu G$.

Lemman 2.15 väittämät voidaan kirjoittaa uuttaa merkintää käyttäen muodossa

$$\mathbb{E}_\nu\left[\sum_{t=0}^{T-1}\mathbf{1}_A(X_t)\right] = \mu(A) \quad \text{ja} \quad \mathbb{E}_\nu\left[\sum_{t=0}^{T-1}f(X_t)\right] = \int f(x)\mu(dx).$$

Valitsemalla ensimmäisessä yhtälössä joukoksi A koko tila-avaruus E saadaan seuraava tulos.

Seurauslause 2.16. Pätee $\mathbb{E}_\nu(T) = \int \mu(dx)$.

Lemma 2.17. Pätee

$$\mathbb{P}_\nu(T<\infty) = \int s(x)\mu(dx).$$

Todistus. Tapahtumat $\{T=t\}$, $t\geq 1$, muodostavat tapahtuman $\{T<\infty\}$ erillisen osituksen. Tämän avulla

$$\mathbb{P}_\nu(T<\infty) = \sum_{t=1}^{\infty}\mathbb{P}_\nu(T=t) = \sum_{t=1}^{\infty}\mathbb{P}_\nu(\tilde{X}_{t-1}\in E, \tilde{X}_t = \Delta)$$

Osittamalla tapahtuma $\{\tilde{X}_{t-1} \in E\}$ saadaan

$$\mathbb{P}_\nu(\tilde{X}_{t-1} \in E, \tilde{X}_t = \Delta) = \int_E \mathbb{P}_\nu(\tilde{X}_t = \Delta, \tilde{X}_{t-1} \in dx).$$

Ehdollistamalla ja käyttämällä kaavaa (14) sekä lemmaa 2.14

$$\begin{aligned}\mathbb{P}_\nu(T < \infty) &= \sum_{t=1}^{\infty} \int_E \mathbb{P}_\nu(\tilde{X}_t = \Delta \mid \tilde{X}_{t-1} = x)\mathbb{P}_\nu(\tilde{X}_{t-1} \in dx) \\ &= \sum_{t=1}^{\infty} \int_E s(x)\nu Q^{t-1}(dx).\end{aligned}$$

Väite seuraa vaihtamalla summan ja integraalin järjestys. □

3.4. Palautuvuus ja tasapainojakauma.

Bivariaatti Markovin ketju (X_t, Y_t) on *Harris-palautuva*, jos $\mathbb{P}_x(T < \infty) = 1$ jokaisella $x \in E$. Bivariaatti Markovin ketju on *positiivisesti palautuva*, jos siihen liittyvä uusiutumisprosessi on positiivisesti palautuva eli $\mathbb{E}_\nu(T) < \infty$. Harris-palautuvuuden määritelmässä on olennaista, että $\mathbb{P}_x(T < \infty) = 1$ jokaisella alkutilalla $x \in E$. Mikäli väite pätee pelkästään melkein jokaisella x, sanotaan, että Markovin ketju on palautuva.

Lemma 2.18. Bivariaatti Markovin ketju on positiivisesti palautuva eli on voimassa $\mathbb{E}_\nu(T) < \infty$.

Todistus. Määritellään luvut $L_t \triangleq \min\{0 \leq s < t : Y_{t-s} = 1\}$. Jos bivariaattia Markovin ketjua tarkastellaan ajan hetkellä t, antaa luku L_t ajan, joka on kulunut edellisestä regeneraatiosta. Tapahtuma $\{T \leq t\}$ voidaan esittää nyt erillisenä yhdisteenä

$$\{T \leq t\} = \bigcup_{s=0}^{t-1} \{L_t = s\}.$$

Olkoon $A \in \mathcal{E}$. Nyt

$$\begin{aligned}\mathbb{P}_\pi(X_t \in A) &= \mathbb{P}_\pi(X_t \in A, T > t) + \mathbb{P}_\pi(X_t \in A, T \leq t) \\ &= \mathbb{P}_\pi(X_t \in A, T > t) + \sum_{s=0}^{t-1} \mathbb{P}_\pi(X_t \in A, L_t = s).\end{aligned}$$

Toisaalta käyttämällä lemmaa 2.13

$$\begin{aligned}&\mathbb{P}_\pi(X_t \in A, L_t = s) \\ &\quad = \mathbb{P}_\pi(Y_{t-s} = 1, Y_{t-s+1} = 0, \ldots, Y_t = 0, X_t \in A) \\ &\quad = \mathbb{P}(Y_{t-s+1} = 0, \ldots, Y_t = 0, X_t \in A \mid Y_{t-s} = 1)\mathbb{P}_\pi(Y_{t-s} = 1) \\ &\quad = \mathbb{P}_\nu(Y_1 = 0, \ldots, Y_{s-1} = 0, Y_s = 0, X_s \in A)\mathbb{P}_\pi(Y_{t-s} = 1),\end{aligned}$$

jossa kaavan (12) ja tasapainojakauman määritelmän nojalla

$$\begin{aligned}\mathbb{P}_\pi(Y_{t-s} = 1) &= \int \mathbb{P}(Y_{t-s} = 1 \mid X_{t-s-1} = x)\mathbb{P}_\pi(X_{t-s-1} \in dx) \\ &= \int s(x)\pi(dx).\end{aligned}$$

Käyttämällä kaavaa (13) ja lemmaa 2.14

$$\begin{aligned}
&\mathbb{P}(Y_1 = 0, \ldots, Y_{s-1} = 0, X_s \in A, Y_s = 0) \\
&\quad = \int \mathbb{P}_\nu(\tilde{X}_{s-1} \in dx)\mathbb{P}(\tilde{X}_s \in A \mid \tilde{X}_{s-1} = x) \\
&\quad = \int \nu Q^{s-1}(dx) Q(x, A) \\
&\quad = \nu Q^s(A).
\end{aligned}$$

Yhdistämällä tulokset

$$\mathbb{P}_\pi(X_t \in A) = \mathbb{P}_\pi(X_t \in A, T > t) + \int s(x)\pi(dx) \sum_{s=0}^{t-1} \nu Q^s(A). \tag{16}$$

Valinta $A = E$ tuottaa rajalla $t \to \infty$

$$1 = \mathbb{P}_\pi(T = \infty) + \int s(x)\pi(dx) \int \mu(dx).$$

Käyttämällä hyväksi funktion s määritelmää ja tasapainojakauman π tukea koskevaa oletusta

$$\int s(x)\pi(dx) = \int \beta \mathrm{vol}(D) \mathbf{1}_C(x)\pi(dx) = \beta \mathrm{vol}(D) \int_C \pi(dx) > 0, \tag{17}$$

Siten seurauslauseen 2.16 nojalla

$$\int \mu(dx) = \mathbb{E}_\nu(T) < \infty. \qquad \square$$

Lause 2.19. Tasapainojakauma π on yksikäsitteinen ja sillä on esitys

$$\pi(A) = \frac{\mu(A)}{\mathbb{E}_\nu(T)} \quad \text{jokaisella } A \in \mathcal{E}. \tag{18}$$

Todistus. Käyttämällä hyväksi lauseen 2.3 tulosta $G = \mathrm{id} + GQ$ kullakin $A \in \mathcal{E}$ on voimassa

$$\mu(A) = \nu G(A) = \nu(A) + \nu GQ(A) = \nu(A) + \mu Q(A).$$

Käyttämällä ytimen Q määritelmää ja lemmaa 2.17 sekä sitä, että positiivisen palautuvuuden nojalla $\mathbb{P}_\nu(T < \infty) = 1$,

$$\mu(A) = \nu(A) + \mu P(A) - \nu(A) \int s(x)\mu(dx) = \mu P(A).$$

Siispä μ toteuttaa tasapainoehdon, joten normittamalla siitä saadaan tasapainojakauma.

Osoitetaan seuraavaksi, että tasapainojakauma on yksikäsitteinen. Oletetaan, että π on mikä tahansa todennäköisyysjakauma, joka toteuttaa tasapainoehdon. Yhtälössä (16) rajalla $t \to \infty$

$$\pi(A) = \mathbb{P}_\pi(T = \infty) + \mu(A) \int s(x)\pi(dx) \geq \mu(A) \int s(x)\pi(dx).$$

Toisin sanoen

$$\pi(A) \geq \mu(A) \int s(x)\pi(dx).$$

Tämän nojalla voidaan määritellä mitta $\lambda(A) \triangleq \pi(A) - \mu(A) \int s(x)\pi(dx)$. Nyt joko $\lambda(E) = 0$ tai $\lambda(E) > 0$. Ensimmäisessä tapauksessa tulee olla

$$\pi(A) = \mu(A) \int s(x)\pi(dx) \tag{19}$$

kaikilla $A \in \mathcal{E}$. Mutta $\int s(x)\pi(dx) > 0$, joten π on välttämättä muotoa (18). Tarkastellaan seuraavaksi tapausta $\lambda(E) > 0$. Ensinnäkin tämä tarkoittaa, että λ voidaan normittaa siten, että se määrittelee todennäköisyysjakauman tiheysfunktion. Toisaalta

$$\begin{aligned}\lambda P(A) &= \pi P(A) - \mu P(A) \int s(x)\pi(dx) \\ &= \pi(A) - \mu(A) \int s(x)\pi(dx) = \lambda(A),\end{aligned}$$

sillä sekä π, että μ toteuttavat tasapainoehdon. Siten $\lambda(A)/\lambda(E)$ määrittelee tasapainojakauman. Minkä tahansa tasapainojakauman tulee toteuttaa ehto (17). Kuitenkin

$$\int s(y)\lambda(dy) = \int s(y)\pi(dy) - \int s(y)\mu(dy) \int s(x)\pi(dx) = 0,$$

mikä on ristiriita. Siispä on välttämättä oltava $\lambda(E) = 0$ ja edellä todetun nojalla tämä tarkoittaa, että tasapainojakauma π on yksikäsitteistä yhtälön (18) määräämää muotoa. □

Muistetaan, että edellisen lauseen oletuksena on, että tasapainojakauma on olemassa. Kyseessä ei ole siis tasapainojakauman olemassaolotodistus.

Lause 2.20. Markovin ketju (X_t) on Harris-palautuva eli jokaisella $x \in E$ pätee $\mathbb{P}_x(T < \infty) = 1$.

Todistus. Lemman 2.18 todistuksessa osoitettiin, että

$$1 = \mathbb{P}_\pi(T = \infty) + \mu(E) \int s(x)\pi(dx).$$

Toisaalta yhtälöstä (18), lemmasta 2.17 ja positiivisesta palautuvuudesta seuraa

$$\int s(x)\pi(dx) = \mu(E)^{-1},$$

joten tulee olla $\mathbb{P}_\pi(T = \infty) = 0$. Integroimalla yli kaikkien lähtötilojen

$$\mathbb{P}_\pi(T = \infty) = \int \mathbb{P}_x(T < \infty)\pi(dx) = 0,$$

josta seuraa, että melkein jokaisella $x \in E$ on voimassa $\mathbb{P}_x(T = \infty) = 0$ tai yhtäpitävästi $\mathbb{P}_x(T < \infty) = 1$. Osoitamme seuraavaksi, että ei löydy sellaista nollamittaista joukkoa, jossa väite ei olisi voimassa. Määritellään $h(x) \triangleq \mathbb{P}_x(T = \infty)$. Tapahtumat $\{T > 1\}, \{T > 2\}, \ldots$ muodostavat laskevan jonon ja pätee

$$h(x) = \lim_{t\to\infty} \mathbb{P}_x(T > t) = \lim_{t\to\infty} \mathbb{P}_x(\tilde{X}_t \in E) = \lim_{t\to\infty} Q^t(x, E).$$

Käyttämällä monotonisen konvergenssin lausetta

$$\begin{aligned} h(x) &= \lim_{t\to\infty} \int Q(x,dy)Q^{t-1}(y,E) \\ &= \int Q(x,dy)h(y) \\ &= \int P(x,dy)h(y) - s(x)\int h(y)\nu(dy) \\ &= \int P(x,dy)h(y), \end{aligned}$$

sillä

$$\int h(y)\nu(dy) = \frac{1}{\mathrm{vol}(D)}\int_D \mathbb{P}_y(T=\infty)dy = 0,$$

koska muutoin tämä olisi ristiriidassa yhtälön $\mathbb{P}_\pi(T=\infty)=0$ kanssa. Käyttämällä ytimen P esitystä tiheysfunktion ja pistemassan summana

$$\begin{aligned} h(x) &= \int p(x,y)h(y)dy + \int r(x)\delta_x(dy)h(y) \\ &= \int p(x,y)h(y)dy + r(x)h(x), \end{aligned}$$

toisin sanoen

$$\int p(x,y)h(y)dy = [1-r(x)]h(x).$$

Oletetaan, että löytyy sellainen x, jolla $h(x)>0$. Koska pätee $1-r(x) = \int p(x,y)dy > 0$, on oltava

$$\int p(x,y)h(y)dy = \int p(x,y)\mathbb{P}_x(T=\infty)dy > 0.$$

Mutta tästä seuraa, että on oltava $\mathbb{P}_x(T=\infty)>0$ positiivismittaisessa joukossa, mikä on ristiriita. $\square$

Mikäli λ on mielivaltainen alkujakauma, lemmasta seuraa, että ensimmäinen regeneraatioaika on äärellinen todennäköisyydellä yksi, sillä

$$\mathbb{P}_\lambda(T<\infty) = \int \mathbb{P}_x(T<\infty)\lambda(dx) = 1.$$

LUKU 3

Kolme Markovin ketjujen konvergenssitulosta

1. Suurten lukujen laki

Tämän luvun tavoitteena on todistaa suurten lukujen laki ergodiselle Markovin ketjulle, jonka tasapainojakauma on π. Todistus perustuu artikkeliin [**Num02**]. Käytämme lisäksi summan satunnaisen ylärajan arviointiin samankaltaista menettelyä kuin artikkelin [**Num02**] todistuksessa keskeiselle raja-arvolauseelle. Otetaan käyttöön merkintä

$$\hat{\pi}_t(f) \triangleq \frac{1}{t}\sum_{s=0}^{t-1} f(X_s).$$

Sanomme, että funktio $f : E \to \mathbb{R}$ on π-*integroituva*, jos

$$\pi(|f|) = \int |f(x)|\pi(dx) < \infty.$$

Lause 3.1 (Suurten lukujen laki). Oletetaan, että Markovin ketju (X_t) on ergodinen ja sillä on tasapainojakauma π, jonka tuki $\operatorname{supp}\pi = E$. Tällöin jokaisella alkutilalla $x \in E$ ja jokaisella π-integroituvalla funktiolla $f : E \to \mathbb{R}$ on voimassa

$$\lim_{t\to\infty} \hat{\pi}_t(f) = \pi(f) \quad \text{m.v.}$$

Bivariaatti Markovin ketju (X_t, Y_t) voidaan jakaa *satunnaislohkoihin*. Ensimmäinen lohko alkaa Markovin ketjun alkutilasta ja päättyy ensimmäistä regeneraatiohetkeä T_0 edeltävään hetkeen. Toinen lohko alkaa ensimmäisestä regeneraatiosta ja päättyy toista regeneraatiota edeltävään hetkeen ja niin edelleen. Olkoon $f : E \to \mathbb{R}$ jokin π-integroituva funktio. Määritellään *satunnaissummat* yli satunnaislohkojen asettamalla

$$\zeta_0 \triangleq \sum_{t=0}^{T_0-1} f(X_t)$$
$$\zeta_i \triangleq \sum_{t=T_i}^{T_{i+1}-1} f(X_t), \quad \text{kun } i \geq 1.$$

Konstruktion nojalla on ilmeistä, että satunnaissummat $(\zeta_i : i \geq 0)$ ovat riippumattomia. Satunnaissumman ζ_0 jakauma riippuu Markovin ketjun alkujakaumasta λ. Sen sijaan satunnaissummat $(\zeta_i : i \geq 1)$ ovat samoin jakautuneita yhteisenä jakaumanaan $\mathbb{P}_\nu(\zeta \in A)$.

Suurten lukujen lain todistus. Olkoon $N(t) = \sum_{s=0}^{t} Y_s$ ajan hetkeen t mennessä tapahtuneiden regeneraatioiden lukumäärä. Summalauseke

$\sum_{s=0}^{t-1} f(X_s)$ voidaan esittää satunnaissummien avulla muodossa

$$\sum_{s=0}^{t-1} f(X_s) = \zeta_0 + \sum_{i=1}^{N(t)-1} \zeta_i + \zeta^*_{N(t)}, \tag{20}$$

jossa

$$\zeta^*_{N(t)} \triangleq \begin{cases} \sum_{s=T_{N(t)}}^{t-1} f(X_s), & T_{N(t)} < t, \\ 0, & T_{N(t)} = t \end{cases}.$$

Lauseessa 2.20 on osoitettu, että ensimmäinen regeneraatioaika on äärellinen todennäköisyydellä yksi riippumatta Markovin ketjun alkutilasta. Lisäksi oletuksen nojalla f on äärellinen todennäköisyydellä yksi, mistä seuraa, että ensimmäinen satunnaissumma ζ_0 on äärellinen todennäköisyydellä yksi. Siten

$$\lim_{t\to\infty} \frac{\zeta_0}{t} = 0 \quad \text{m.v.}$$

Arvioidaan

$$\zeta^*_{N(t)} \leq \sum_{s=T_{N(t)}}^{T_{N(t)+1}-1} |f(X_s)|.$$

Mutta epäyhtälön oikean puolen summa on äärellinen todennäköisyydellä yksi, joten

$$\lim_{t\to\infty} \frac{\zeta^*_{N(t)}}{t} = 0 \quad \text{m.v.}$$

Yhtälön (20) oikean puolen toinen termi sisältää satunnaismuuttujan summan ylärajassa. Pyritään muodostamaan arvio, jolla satunnainen yläraja pystytään korvaamaan deterministisellä luvulla. Lauseen 2.8 nojalla

$$\lim_{t\to\infty} \frac{N(t)}{t} = \mathbb{E}_\nu(T)^{-1} \quad \text{m.v.}$$

Merkitään $\alpha = \mathbb{E}_\nu(T)^{-1}$. Mikäli $\varepsilon > 0$ on mielivaltainen, löytyy sellainen $t_\varepsilon \geq 1$, että jokaisella $t \geq t_\varepsilon$

$$\left| \frac{N(t)}{t} - \alpha \right| < \varepsilon \quad \text{m.v.},$$

josta seuraa arvio

$$t(\alpha - \varepsilon) - 1 < N(t) - 1 < t(\alpha + \varepsilon) - 1 \quad \text{m.v.}$$

Olkoon nyt t_* suurin sellainen kokonaisluku, joka on pienempi tai yhtäsuuri kuin $t(\alpha - \varepsilon) - 1$ ja t^* pienin sellainen kokonaisluku, jonka on suurempi tai yhtäsuuri kuin $t(\alpha + \varepsilon) - 1$. Saamme muodostettua arvion

$$\frac{t_*}{t} \left(\frac{1}{t_*} \sum_{i=1}^{t_*} \zeta_i \right) \leq \frac{1}{t} \sum_{i=1}^{N(t)-1} \zeta_i \leq \frac{t^*}{t} \left(\frac{1}{t^*} \sum_{i=1}^{t^*} \zeta_i \right) \quad \text{m.v.}$$

Epäyhtälön vasemman ja oikean puolen summiin voi soveltaa tavanomaista suurten lukujen lakia, sillä t_* ja t^* ovat deterministisiä lukuja, jotka kasvavat rajatta, kun $t \to \infty$. Valitsemalla luku $\varepsilon > 0$ mielivaltaisen pieneksi luvut t_*/t ja t^*/t saadaan mielivaltaisen lähelle lukua α. Siten

$$\lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{N(t)-1} \zeta_i = \alpha \mathbb{E}_\nu(\zeta_0) \quad \text{m.v.}$$

Lemman 2.15 nojalla

$$\begin{aligned}\mathbb{E}_\nu(\zeta_0) &= \int f(x)\mu(dx) \\ &= \mathbb{E}_\nu(T) \int f(x)\pi(dx) = \mathbb{E}_\nu(T)\pi(f).\end{aligned}$$

Yhdistämällä saadut tulokset

$$\lim_{t\to\infty} \hat{\pi}_t(f) = \lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{N(t)-1} \zeta_i = \frac{\mathbb{E}_\nu(\zeta_0)}{\mathbb{E}_\nu(T)} = \pi(f) \quad \text{m.v.}$$ □

2. Kokonaisvariaatioetäisyys

2.1. Suppeneminen kokonaisvariaatioetäisyydessä.

Asetetaan seuraava määritelmä [**Tie96**] s. 61 mukaisesti.

Määritelmä 3.1. Kahden tila-avaruudella $(E, \mathcal{E})$ määritellyn todennäköisyysjakauman λ ja μ välinen *kokonaisvariaatioetäisyys* on

$$||\lambda - \mu|| \triangleq 2 \sup_{A \in \mathcal{E}} |\lambda(A) - \mu(A)|.$$

Lause 3.2. Olkoon (X_t) ergodinen Markovin ketju. Mikäli λ ja μ ovat kaksi alkujakaumaa, pätee

$$\lim_{t\to\infty} ||\lambda P^t - \mu P^t|| = 0.$$

Valitsemalla lauseessa Markovin ketjun tasapainojakauma jakaumaksi μ saadaan seuraava tulos.

Seurauslause 3.3. Olkoon (X_t) ergodinen Markovin ketju, jonka tasapainojakauma on π. Mikäli λ on mielivaltainen alkujakauma, pätee

$$\lim_{t\to\infty} ||\lambda P^t - \pi|| = 0.$$

Valitsemalla alkujakaumaksi λ tilaan x keskittynyt pistemassafunktio saadaan seuraava tulos.

Seurauslause 3.4. Olkoon (X_t) ergodinen Markovin ketju, jonka tasapainojakauma on π. Mikäli x on mielivaltainen alkutila, pätee

$$\lim_{t\to\infty} ||P^t(x,\cdot) - \pi(\cdot)|| = 0.$$

Lauseen 3.2 todistamista varten muodostetaan siirtymätodennäköisyysytimen $P^t(x, A)$ hajotelma [**Num84**] s. 64 mukaisesti. Määritellään viimeisin aikaan t mennessä tapahtunut regeneraatiohetki satunnaisaikana

$$L_t \triangleq \max\{s : Y_s = 1, 0 \le s \le t\}.$$

Tarkastellaan Markovin ketjua ajan hetkellä t. Ensimmäinen vaihtoehto on, että ajan hetkeen $t-1$ mennessä ei tapahdu ensimmäistäkään regeneraatiota eli $T_0 \ge t$. Toisena vaihtoehtona ensimmäinen regeneraatio tapahtuu ajan hetkellä $0 \leq s < t$ eli $T_0 = s$ ja viimeinen aikaa t edeltävä regeneraatio tapahtuu ajan hetkellä $s \leq r < t$ eli $L_{t-1} = r$. Siten

$$P^t(x, A) = \mathbb{P}_x(X_t \in A, T_0 \geq t) + \sum_{s=0}^{t-1} \sum_{r=s}^{t-1} \mathbb{P}_x(X_t \in A, L_{t-1} = r, T_0 = s)$$

Ehdollistamalla

$$\mathbb{P}_x(X_t \in A, L_{t-1} = r, T_0 = s) = \mathbb{P}(X_t \in A, L_{t-1} = r \mid T_0 = s)\mathbb{P}_x(T_0 = s)$$

Toisaalta

$$\begin{aligned}
&\mathbb{P}(X_t \in A, L_{t-1} = r \mid T_0 = s) \\
&\quad = \mathbb{P}_\nu(X_{t-s} \in A, L_{t-s-1} = r - s) \\
&\quad = \mathbb{P}_\nu(X_{t-s} \in A, L_{t-s-1} = r - s, Y_{r-s} = 1) \\
&\quad = \mathbb{P}_\nu(X_{t-s} \in A, L_{t-s-1} = r - s \mid Y_{r-s} = 1)\mathbb{P}_\nu(Y_{r-s} = 1) \\
&\quad = \mathbb{P}_\nu(X_{t-r} \in A, T \geq t - r)\mathbb{P}_\nu(Y_{r-s} = 1).
\end{aligned}$$

Olemme saaneet esityksen

$$\begin{aligned}
P^t(x, A) &= \mathbb{P}_x(X_t \in A, T_0 \geq t) \\
&\quad + \sum_{s=0}^{t-1} \sum_{r=s}^{t-1} \mathbb{P}_x(T_0 = s)\mathbb{P}_\nu(Y_{r-s} = 1)\mathbb{P}_\nu(X_{t-r} \in A, T \geq t - r).
\end{aligned}$$

Uusiutumisprosessiksi tulkittuna $\mathbb{P}_x(T_0 = s) = a_s(x)$ ja $\mathbb{P}_\nu(Y_{r-s} = 1) = u_{r-s}$. Määritellään lisäksi

$$\sigma_t(A) \triangleq \mathbb{P}_\nu(X_{t+1} \in A, T \geq t + 1),$$

joten hajotelma voidaan kirjoittaa muodossa

$$P^t(x, A) = \mathbb{P}_x(X_t \in A, T_0 \geq t) + \sum_{s=0}^{t-1} \sum_{r=s}^{t-1} a_s(x) u_{r-s} \sigma_{t-1-r}(A).$$

Vaihtamalla summausjärjestystä kaksinkertainen summa voidaan tunnistaa kaksinkertaiseksi konvoluutioksi

$$\begin{aligned}
\sum_{s=0}^{t-1} \sum_{r=s}^{t-1} a_s(x) u_{r-s} \sigma_{t-1-r}(A) &= \sum_{r=0}^{t-1} \sigma_{t-1-r}(A) \sum_{s=0}^{r} a_s(x) u_{r-s} \\
&= \sum_{s=0}^{t-1} \sigma_{t-1-r}(A) a(x) \star u_r \\
&= a(x) \star u \star \sigma(A)_{t-1},
\end{aligned}$$

joten siirtymätodennäköisyysytimelle $P^t(x, A)$ on saatu hajotelma

$$P^t(x, A) = \mathbb{P}_x(X_t \in A, T_0 \geq t) + a(x) \star u \star \sigma(A)_{t-1}. \tag{21}$$

LAUSEEN 3.2 TODISTUS. Olkoon $A \in \mathcal{E}$ mielivaltainen joukko. Kolmioepäyhtälön nojalla

$$\begin{aligned} &|\lambda P^t(A) - \mu P^t(A)| \\ &\leq \int \lambda(dx)\mathbb{P}_x(X_t \in A, T_0 \geq t) + \int \mu(dx)\mathbb{P}_x(X_t \in A, T_0 \geq t) \\ &+ \left| \int \lambda(dx)a(x) \star u \star \sigma(A)_{t-1} - \int \mu(dx)a(x) \star u \star \sigma(A)_{t-1} \right| \end{aligned} \tag{22}$$

On ilmeistä, että $\mathbb{P}_x(X_t \in A, T_0 \geq t) \leq \mathbb{P}_x(T_0 \geq t)$ jokaisella $x \in E$. Toisaalta ensimmäinen regeneraatioaika on äärellinen todennäköisyydellä yksi riippumatta alkutilasta x. Siten

$$\sup_{A \in \mathcal{E}} \int \lambda(dx)\mathbb{P}_x(X_t \in A, T_0 \geq t) \leq \int \lambda(dx)\mathbb{P}_x(T_0 \geq t) \to 0,$$

kun $t \to \infty$. Vastaavasti summan (22) toinen termi häviää rajalla $t \to \infty$.

Merkitään $\lambda_t(a) = \int \lambda(dx)a_t(x)$ ja vastaavasti $\mu_t(a) = \int \mu(dx)a_t(x)$. $\lambda(a) = (\lambda_t(a) : t \geq 0)$ ja $\mu(a) = (\mu_t(a) : t \geq 0)$ määrittelevät todennäköisyysjakauman diskreetillä avaruudella $\mathbb{N}$, sillä

$$\sum_{t=0}^{\infty} \lambda_t(a) = \int \lambda(dx) \sum_{t=0}^{\infty} a_t(x) = \int \lambda(dx) = 1.$$

Vaihtamalla summan ja integraalin järjestystä konvoluutiossa yhtälön (22) kolmas termi voidaan esittää muodossa

$$|\lambda(a) \star u \star \sigma(A)_{t-1} - \mu(a) \star u \star \sigma(A)_{t-1}|.$$

Ilmeisesti pätee

$$\sigma_{t-1}(A) = \mathbb{P}_\nu(X_t \in A, T \geq t) \leq \mathbb{P}_\nu(T \geq t) = B_{t-1}.$$

Käyttämällä konvoluutiossa kolmioepäyhtälöä sekä edellistä arviota saadaan

$$\begin{aligned} &\sup_{A \in \mathcal{E}} |\lambda(a) \star u \star \sigma(A)_{t-1} - \mu(a) \star u \star \sigma(A)_{t-1}| \\ &\qquad \leq |\lambda(a) \star u - \mu(a) \star u| \star B_{t-1}. \end{aligned}$$

Edelleen kolmioepäyhtälön nojalla

$$\begin{aligned} &\sup_{A \in \mathcal{E}} |\lambda(a) \star u \star \sigma(A)_{t-1} - \mu(a) \star u \star \sigma(A)_{t-1}| \\ &\qquad \leq |\lambda(a) \star u - u| \star B_{t-1} + |\mu(a) \star u - u| \star B_{t-1}. \end{aligned}$$

Mutta nyt $\lambda(a)$ ja $\mu(a)$ ovat hyvin määriteltyjä viivejakaumia, joten lauseen 2.12 nojalla epäyhtälön oikea puoli häviää, kun $t \to \infty$. Olemme täten näyttäneet, että $||\lambda P^t - \mu P^t|| \to 0$. □

2.2. Geometrinen ergodisuus.

On mahdollista osoittaa, että edellisen pykälän ehto kokonaisvariaatiosuppenemiselle on välttämätön, kts. esimerkiksi [**Num84**] propositio 6.3. Tämän nojalla Markovin ketjun (X_t) ergodisuus voidaan määritellä yhtäpitävästi siten, että löytyy todennäköisyysjakauma π, jolle

$$\lim_{t \to \infty} ||P^t(x, \cdot) - \pi(\cdot)|| = 0 \quad \text{jokaisella } x \in E.$$

Suppeneminen kokonaisvariaatioetäisyydessä ei anna mitään tietoa suppenemisvauhdista. Sen sijaan ergodisuutta vahvemmat ehdot antavat keinon rajata suppenemisvauhtia, kts. esim. [**Num84**] luvut 6.3–6.6. Suppenemisvauhdin tarkastelu ei kuitenkaan kuulu tämän tutkielman aihepiiriin, joten esittelemme yhden vahvemman ergodisuusluokan siksi, että sen avulla saadaan käyttökelpoinen keskeinen raja-arvolause Markovin ketjuille.

MÄÄRITELMÄ 3.2. Ergodinen Markovin ketju (X_t) on *geometrisesti ergodinen*, jos löytyy luku $\gamma > 1$, jolle

$$\mathbb{E}_\nu(\gamma^T) < \infty.$$

Geometrinen ergodisuus rajoittaa kokonaisvariaatioetäisyyden suppenemisvauhtia. [**Num84**] teoreeman 6.14 mukaan geometrinen ergodisuus on yhtäpitävää sen kanssa, että on olemassa epänegatiivinen ja π-integroituva funktio M sekä vakio $0 < r < 1$, joille

$$||P^t(x, \cdot) - \pi|| \leq M(x)r^t \quad \text{kaikilla } x \in E \text{ ja } t \geq 0.$$

Regeneraatioaikaa koskevan ehdon toteen näyttämiseen ei ole olemassa mitään ilmeistä keinoa. Muotoilemme seuraavaksi [**Rob04**] teoreeman 9 tai [**MT96**] teoreeman 1.4 mukaisesti konkreettisen ehdon, jonka avulla on mahdollista näyttää, että annettu Markovin ketju on geometrisesti ergodinen. Kyseinen ehto tunnetaan englanninkielisessä kirjallisuudessa nimellä *drift condition*. Tarvitsemme ensin pienen joukon määritelmän.

MÄÄRITELMÄ 3.3. Joukko $C^* \in \mathcal{E}$ on *pieni*, jos löytyy $t_0 \geq 1$, luku $\beta > 0$ sekä mitta $\varphi : \mathcal{E} \to [0, +\infty]$, joille

$$P^{t_0}(x, A) \geq \beta\varphi(A) \quad \text{kaikilla } x \in C^* \text{ ja } A \in \mathcal{E}.$$

On helppo näyttää, että pelkistymättömyys- ja minorisaatio-oletuksen joukko C on pieni edellisen määritelmän mielessä.

LAUSE 3.5. Ergodinen Markovin ketju (X_t) on geometrisesti ergodinen, mikäli löytyy pieni joukko $C^* \in \mathcal{E}$, luvut $0 < \lambda < 1$ ja $b < \infty$ sekä mitallinen funktio $V \geq 1$, jolle $V(x) < \infty$ ainakin jollain $x \in E$, jotka toteuttavat ehdon

$$\int P(x, dy)V(y) \leq \lambda V(x) + b\mathbf{1}_{C^*}(x) \quad \text{jokaisella } x \in E.$$

ESIMERKKI 3.1. Jatketaan esimerkkien 2.3 ja 2.4 geometrisellä jakaumalla b varustetun ergodisen Markovin ketjun (V_t) tarkastelua ja osoitetaan, että se on geometrisesti ergodinen. Kiinnitetään pieni joukko $C^* = \{1\}$. Mikäli $V(y)$ on joukolla $\mathbb{N}^+$ määritelty funktio, pätee

$$\int P(x, dy)V(y) = \sum_{y=1}^{\infty} p_{xy}V(y) = \begin{cases} \sum_{y=1}^{\infty} b_y V(y), & x = 1 \\ V(x-1), & x > 1 \end{cases}.$$

Olkoon $\gamma > 1$ sellainen, että $(1 - r)\gamma < 1$. Määritellään $V(x) = \gamma^x$. Nyt

$$\sum_{y=1}^{\infty} b_y V(y) = \frac{r}{1-r} \sum_{y=1}^{\infty} [(1-r)\gamma]^y = \frac{r}{1-r} \left[\frac{1}{1-(1-r)\gamma} - 1 \right] < \infty.$$

Olkoon $\lambda = \gamma^{-1} < 1$. Nyt jokaisella $x \geq 2$ on voimassa

$$\int P(x, dy)V(y) = V(x-1) = \gamma^{x-1} \leq \lambda V(x) = \gamma^{x-1}.$$

Lisäksi valitsemalla $b < \infty$ riittävän suureksi on voimassa

$$\int P(1, dy)V(y) = \sum_{y=1}^{\infty} b_y V(y) \leq \lambda V(1) + b\mathbf{1}_{C^*}(1) = 1 + b.$$

Lauseen 3.5 nojalla Markovin ketju (V_t) on geometrisesti ergodinen.

3. Keskeinen raja-arvolause

Markovin ketjujen keskeinen raja-arvolause voidaan muotoilla eri tavoin riippuen siitä, mitkä oletukset Markovin ketju ja funktio f toteuttavat. Seuraava keskeisen raja-arvolauseen muotoilu perustuu ensimmäisen satunnaissumman ζ_0 toisen momentin äärellisyyteen. Kyseessä on tekninen oletus, joten seuraava tulos on muotoiltu lemmana. Muotoilemme myöhemmin keskeisen raja-arvolauseen käyttäen Markovin ketjua ja funktiota f koskevia ehtoja, jotka takaavat vaaditun toisen momentin äärellisyysehdon. Lemman todistus nojaa artikkeliin [**Num02**].

LEMMA 3.6. Olkoon (X_t) ergodinen Markovin ketju, jolla on tasapainojakauma π, sekä f π-integroituva funktio. Oletetaan, että pätee $\mathbb{E}_\nu(\zeta_0^2) < \infty$. Tällöin keskeinen raja-arvolause on voimassa eli

$$\sqrt{t}\,[\hat{\pi}_t(f) - \pi(f)] \xrightarrow{d} \mathrm{N}(0, \sigma_f^2) \quad \text{jollakin } \sigma_f^2 < \infty.$$

TODISTUS. Oletetaan, että $\pi(f) = 0$. Muutoin määritellään $\tilde{f} = f - \pi(f)$, jolle $\pi(\tilde{f}) = 0$ ja

$$\sqrt{t}\,[\hat{\pi}_t(f) - \pi(f)] = \sqrt{t}\left[\hat{\pi}_t(\tilde{f}) - \pi(\tilde{f})\right] \xrightarrow{d} \mathrm{N}(0, \sigma_{\tilde{f}}^2).$$

Satunnaissummat oletetaan määritellyiksi kuten suurten lukujen lain todistuksessa. Pätee

$$\begin{aligned}\sqrt{t}\,[\hat{\pi}_t(f) - \pi(f)] &= \sqrt{t}\left[\frac{1}{t}\sum_{i=0}^{N(t)-1}\zeta_i + \frac{1}{t}\zeta_{N(t)}^*\right] \\ &= \sqrt{t}\left[\frac{1}{t}\zeta_0 + \frac{1}{t}\sum_{i=1}^{N(t)}\zeta_i - \frac{1}{t}\zeta_{N(t)} + \frac{1}{t}\zeta_{N(t)}^*\right]\end{aligned}$$

Rajalla $t \to \infty$

$$\sqrt{t}\,[\hat{\pi}_t(f) - \pi(f)] \to \lim_{t\to\infty} \frac{1}{\sqrt{t}}\sum_{i=1}^{N(t)}\zeta_i \quad \text{m.v.}$$

Summan yläraja on satunnainen. Pyrimme korvaamaan sen deterministisellä ylärajalla käyttäen pitkälti samaa ideaa kuin suurten lukujen lain todistuksessa. Tarvitsemme nyt kuitenkin edellistä vahvemman keinon kontrolloida suppenemista.

Merkitään $\alpha \triangleq \mathbb{E}_\nu(T)^{-1}$. Tiedämme, että $N(t)/t \to \alpha$ melkein varmasti, kun $t \to \infty$, joten voimme soveltaa Egorovin lausetta (kts. liite A). Olkoon $\delta > 0$ mielivaltainen. Tällöin löytyy sellainen joukko $\Lambda_\delta \in \mathcal{E}$, että $\mathbb{P}(\Lambda_\delta^C) < \delta$

ja $N(t)/t \to \alpha$ tasaisesti joukossa Λ_δ. Toisin sanoen löytyy sellainen $t_\delta \geq 1$, että jokaisella $t \geq t_\delta$ joukossa Λ_δ on voimassa

$$\left| \frac{N(t)}{t} - \alpha \right| < \delta \quad \text{m.v.}$$

Yhtäpitävästi

$$t(\alpha - \delta) < N(t) < t(\alpha + \delta) \quad \text{m.v.}$$

Merkitään t_* suurinta kokonaislukua, joka on pienempi tai yhtäsuuri kuin luku $t(\alpha - \delta)$ ja t^* pienintä kokonaislukua, joka on suurempi tai yhtäsuuri kuin luku $t(\alpha + \delta)$. Voimme arvioida

$$\left| \sum_{i=1}^{N(t)} \zeta_i - \sum_{i=1}^{t_*} \zeta_i \right| \leq \max_{t_* < k < t^*} \left| \sum_{i=t_*+1}^{k} \zeta_i \right|.$$

Epäyhtälön oikeaan puoleen voi soveltaa Kolmogorovin epäyhtälöä (kts. liite A), sillä oletuksen nojalla $\mathbb{E}(\zeta_i) = \mathbb{E}_\nu(\zeta_0) = \mathbb{E}_\nu(T)\pi(f) = 0$ ja $\mathrm{Var}(\zeta_i) = \mathbb{E}_\nu(T^2) < \infty$ kullakin i. Olkoon $\varepsilon > 0$ kiinnitetty. Jokaisella $t \geq t_\delta$

$$\begin{aligned}
\mathbb{P}\left(\max_{t_* < k < t^*} \left| \sum_{i=t_*+1}^{k} \zeta_i \right| \geq \varepsilon\sqrt{t} \mid \Lambda_\delta \right) &\leq \frac{1}{\varepsilon^2 t} \sum_{i=t_*+1}^{t^*-1} \mathrm{Var}(\zeta_i) \\
&\leq \frac{\mathbb{E}_\nu(\zeta_0^2)}{\varepsilon^2 t} \sum_{i=t_*+1}^{t^*-1} 1 \leq \frac{2\delta}{\varepsilon^2} \mathbb{E}_\nu(\zeta_0^2),
\end{aligned}$$

sillä pätee

$$\sum_{i=t_*+1}^{t^*-1} 1 = (t^* - 1) - t_* \leq t(\alpha + \delta) - t(\alpha - \delta) = 2t\delta.$$

Merkitään

$$A \triangleq \left\{ \frac{1}{\sqrt{t}} \left| \sum_{i=1}^{N(t)} \zeta_i - \sum_{i=1}^{t_*} \zeta_i \right| > \varepsilon \right\}.$$

Käyttämällä hyväksi kokonaistodennäköisyyden kaavaa ja ehdollistamalla

$$\mathbb{P}(A) = \mathbb{P}(A \cap \Lambda_\delta^C) + \mathbb{P}(A \cap \Lambda_\delta) \leq \mathbb{P}(\Lambda_\delta^C) + \mathbb{P}(A \mid \Lambda_\delta)\mathbb{P}(\Lambda_\delta).$$

Käyttämällä hyväksi tapahtuman Λ_δ^C todennäköisyyttä ja edellä johdettua arviota sekä arvioimalla tapahtuman Λ_δ todennäköisyyttä ylhäältä luvulla yksi saamme

$$\mathbb{P}(A) \leq \delta + \mathbb{P}\left(\max_{t_* < k < t^*} \left| \sum_{i=t_*+1}^{k} \zeta_i \right| > \varepsilon\sqrt{t} \right) \leq \delta + \frac{2\delta}{\varepsilon^2} \mathbb{E}_\nu(\zeta_0^2).$$

Mutta δ voidaan valita mielivaltaisen pieniksi, joten on näytetty

$$\frac{1}{\sqrt{t}} \left| \sum_{i=1}^{N(t)} \zeta_i - \sum_{i=1}^{t_*} \zeta_i \right| \xrightarrow{\mathbb{P}} 0. \tag{23}$$

Sovelletaan riippumattomien ja samoin jakautuneiden satunnaismuuttujien keskeistä raja-arvolausetta summaan

$$\frac{1}{\sqrt{t}}\sum_{i=1}^{t_*}\zeta_i,$$

jonka yläraja on deterministinen. Satunnaissummat $(\zeta_i : i \geq 1)$ ovat riippumattomia ja samoin jakautuneita. Siten

$$\frac{1}{\sqrt{t}}\sum_{s=1}^{t_*}\zeta_s = \sqrt{\frac{t_*}{t}}\left(\frac{\sqrt{t_*}}{t_*}\sum_{i=1}^{t_*}\zeta_i\right) \xrightarrow{d} \mathrm{N}(0, \alpha\mathbb{E}_\nu(\zeta_0^2)),$$

sillä

$$\alpha - \delta - \frac{1}{t} < \frac{t_*}{t} \leq \alpha - \delta$$

ja tekemällä $\delta > 0$ mielivaltaisen pieneksi seuraa rajalla $t \to \infty$ tulos $t_*/t \to \alpha$. Tuloksen (23) nojalla pätee myös

$$\frac{1}{\sqrt{t}}\sum_{i=1}^{N(t)}\zeta_i \xrightarrow{d} \mathrm{N}(0, \alpha\mathbb{E}_\nu(\zeta_0^2)).$$

Kokoamalla saadut tulokset yhteen

$$\sqrt{t}\,[\hat{\pi}_t(f) - \pi(f)] \xrightarrow{d} \mathrm{N}(0, \alpha\mathbb{E}_\nu(\zeta_0^2)). \qquad \square$$

Voi osoittaa, kts. [**Num02**], että keskeisen raja-arvolauseen varianssilla $\sigma_f^2 = \mathbb{E}_\nu(T)^{-1}\mathbb{E}_\nu(\zeta_0^2)$ on esitys

$$\sigma_f^2 = \int [f(x) - \pi(f)]^2\pi(x)dx$$
$$+ 2\sum_{t=1}^{\infty}\int\int \pi(x)[f(x) - \pi(f)]Q^t(x, dy)[f(y) - \pi(f)]dx.$$

Varianssin laskeminen kaavasta on yleensä vaikeaa tai mahdotonta. MCMC-sovelluksissa varianssi estimoidaan simuloidusta Markovin ketjun realisaatiosta. Estimointimenetelmiä käsittelee esimerkiksi artikkeli [**Gey92**].

Seuraava keskeisen raja-arvolauseen muotoilu on sama kuin artikkelin [**Num02**] teoreema 2.

LAUSE 3.7. Keskeinen raja-arvolause on voimassa ergodiselle Markovin ketjulle (X_t) ja rajoitetulle funktiolle f, jos $\mathbb{E}_\nu(T^2) < \infty$.

TODISTUS. Väite seuraa lemmasta 3.6, sillä

$$\mathbb{E}_\nu(\zeta_0^2) = \mathbb{E}_\nu\left[\left(\sum_{s=1}^{T} f(X_s)\right)^2\right] \leq \mathbb{E}_\nu\left[\left(\sup_{x\in E}|f(x)|\right)^2 T^2\right] < \infty. \qquad \square$$

Lauseen ehtoa $\mathbb{E}_\nu(T^2) < \infty$ nimitetään *toisen asteen ergodisuusehdoksi.* [**Tie94**] huomauttaa, että toisen asteen ergodisuusehdon toteen näyttäminen on usein vaikeaa ja useimmiten MCMC-algoritmien siirtymätodennäköisyysytimet toteuttavat jonkin vahvemman ergodisuusehdon. Sen vuoksi MCMC-menetelmien kirjallisuudessa keskeinen raja-arvolause muotoillaan yleensä geometrisesti ergodisille Markovin ketjuille. [**Tie96**] sivun 68 mukaisesti muotoillaan seuraava tulos.

LAUSE 3.8. Keskeinen raja-arvolause on voimassa geometrisesti ergodiselle Markovin ketjulle (X_t) ja funktiolle f, jos löytyy sellainen luku $\varepsilon > 0$, että seuraava integroituvuusehto on voimassa

$$\int |f(x)|^{2+\varepsilon} \pi(dx) < \infty.$$

Geometrisesti ergodisen Markovin ketjun keskeisessä raja-arvolauseessa funktion f integroituvuusehtoa voidaan lieventää Markovin ketjua koskevalla lisäoletuksella. Nimittäin, mikäli Markovin ketju (X_t) on geometrisesti ergodinen ja kääntyvä, on keskeinen raja-arvolause voimassa [**CG94**] teoreema 6 nojalla edellyttäen, että

$$\int |f(x)|^2 \pi(dx) < \infty.$$

Todistamme seuraavaksi lauseen 3.8 noudattaen artikkelissa [**Rob04**] sivu 59 esitettyä ideaa. Muotoilemme ja todistamme ensin tarpeellisen lemman.

LEMMA 3.9. Oletetaan, että Markovin ketju (X_t) on geometrisesti ergodinen eli löytyy $\gamma > 1$, jolle $\mathbb{E}_\nu(\gamma^T) < \infty$. Tällöin

$$\mathbb{E}_\pi\left(\gamma^T\right) < \infty.$$

TODISTUS. Odotusarvo $\mathbb{E}_\pi(\gamma^T)$ voidaan lausua muodossa

$$\mathbb{E}_\pi(\gamma^T) = \sum_{t=0}^{\infty} \gamma^t \mathbb{P}_\pi(T = t).$$

Toisaalta γ^t voidaan lausua geometrisen jonon summan avulla muodossa

$$\gamma^t = 1 + (\gamma - 1) \sum_{s=0}^{t-1} \gamma^s.$$

Sijoittamalla edelliseen seuraa

$$\mathbb{E}_\pi(\gamma^T) = 1 + (\gamma - 1) \sum_{t=0}^{\infty} \sum_{s=0}^{t-1} \gamma^s \mathbb{P}_\pi(T = t).$$

Siirrytään tarkastelemaan kaksinkertaista summalauseketta. Korvaamalla sisäsumman yläraja indikaattorifunktiolla ja käyttämällä regeneraatioaikaa koskevaa tulosta (15)

$$\begin{aligned}
\sum_{t=0}^{\infty} \sum_{s=0}^{t-1} \gamma^s \mathbb{P}_\pi(T = t) &= \sum_{t=0}^{\infty} \sum_{s=0}^{\infty} \gamma^t \mathbf{1}(t > s) \mathbb{P}_\pi(T = t) \\
&= \sum_{s=0}^{\infty} \gamma^s \mathbb{P}_\pi(T > s) \\
&= \sum_{s=0}^{\infty} \gamma^s \int \pi Q^s(dx).
\end{aligned}$$

Käyttämällä lausetta 2.19 ja mitan μ määritelmää seuraa

$$\begin{aligned}\sum_{s=0}^{\infty}\gamma^s\int \pi Q^s(dx) &= \mathbb{E}_\nu(T)^{-1}\sum_{s=0}^{\infty}\gamma^s\int \mu Q^s(dx)\\ &= \mathbb{E}_\nu(T)^{-1}\sum_{s=0}^{\infty}\gamma^s\int \nu\left(\sum_{t=0}^{\infty}Q^t\right)Q^s(dx)\\ &= \mathbb{E}_\nu(T)^{-1}\int \nu\sum_{s=0}^{\infty}\sum_{t=0}^{\infty}\gamma^s Q^{s+t}(dx).\end{aligned}$$

Pyritään kirjoittamaan kaksinkertainen summa toisessa muodossa. On ilmeistä, että summa sisältää tasan $\tau+1$ sellaista termiä, jolle $Q^{s+t}=Q^\tau$. Näiden termien kertoimet ovat γ^s, jossa $s=0,1,\dots,\tau$. Hyödyntämällä jälleen geometrisen jonon summan kaavaa ja suorittamalla integrointi

$$\begin{aligned}\sum_{s=0}^{\infty}\gamma^s\int \pi Q^s(dx) &= \mathbb{E}_\nu(T)^{-1}\int \nu\sum_{\tau=0}^{\infty}\sum_{s=0}^{\tau}\gamma^s Q^\tau(dx)\\ &= \mathbb{E}_\nu(T)^{-1}\int \nu\sum_{\tau=0}^{\infty}\frac{\gamma^{\tau+1}-1}{\gamma-1}Q^\tau(dx)\\ &= \frac{1}{(\gamma-1)\mathbb{E}_\nu(T)}\sum_{\tau=0}^{\infty}(\gamma^{\tau+1}-1)\mathbb{P}_\nu(T>\tau)\\ &= \frac{1}{(\gamma-1)\mathbb{E}_\nu(T)}\left[\sum_{\tau=0}^{\infty}\gamma^{\tau+1}\mathbb{P}_\nu(T>\tau)-\mathbb{E}_\nu(T)\right].\end{aligned}$$

Summa $\sum_{\tau=0}^{\infty}\gamma^{\tau+1}\mathbb{P}_\nu(T>\tau)$ on mahdollista määrittää käyttämällä samaa periaatetta kuin todistuksen alussa. Saamme

$$\sum_{\tau=0}^{\infty}\gamma^{\tau+1}\mathbb{P}_\nu(T>\tau)=\frac{\gamma}{\gamma-1}\left[\mathbb{E}_\nu(\gamma^T)-1\right]$$

Yhdistämällä saadut tulokset seuraa

$$\mathbb{E}_\pi(\gamma^T)=\frac{\gamma}{(\gamma-1)\mathbb{E}_\nu(T)}\left[\mathbb{E}_\nu(\gamma^T)-1\right]<\infty. \qquad \square$$

Lauseen 3.8 todistus. Lemman 3.6 nojalla riittää osoittaa $\mathbb{E}_\nu(\zeta_0^2)<\infty$. Arvioidaan

$$\begin{aligned}\pi(A)=\int \pi(dx)P(x,A) &\geq \int_C \pi(dx)\int_A p(x,y)dy\\ &\geq \beta\mathrm{vol}(A\cap D)\pi(C)=\beta\mathrm{vol}(D)\nu(A)\pi(C),\end{aligned}$$

josta seuraa

$$\nu(A)\leq\frac{\pi(A)}{\beta\mathrm{vol}(D)\pi(C)}\quad \text{millä tahansa } A\in\mathcal{E}.$$

Edellisen nojalla saadaan arvio

$$\begin{aligned}\mathbb{E}_\nu(\zeta_0^2) &= \int \mathbb{E}_x(\zeta_0^2)\nu(dx)\\ &\leq \frac{1}{\beta \mathrm{vol}(D)\pi(C)} \int \mathbb{E}_x(\zeta_0^2)\pi(dx) = \frac{\mathbb{E}_\pi(\zeta_0^2)}{\beta \mathrm{vol}(D)\pi(C)}.\end{aligned}$$

Tämä merkitsee, että riittää osoittaa $\mathbb{E}_\pi(\zeta_0^2) < \infty$. Käyttämällä hyväksi satunnaissumman määritelmää voidaan odotusarvo lausua muodossa

$$\begin{aligned}\mathbb{E}_\pi(\zeta_0^2) &\leq \mathbb{E}_\pi\left[\left(\sum_{t=0}^{T-1} |f(X_t)|\right)^2\right]\\ &= \mathbb{E}_\pi\left[\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} |f(X_s)|\mathbf{1}(T>s)|f(X_t)|\mathbf{1}(T>t)\right]\\ &= \sum_{s=0}^{\infty}\sum_{t=0}^{\infty} \mathbb{E}_\pi\left[|f(X_s)|\mathbf{1}(T>s)|f(X_t)|\mathbf{1}(T>t)\right].\end{aligned}$$

Odotusarvo voidaan tulkita funktioiden $|f(X_s)|\mathbf{1}(T > s)$ ja $|f(X_t)|\mathbf{1}(T > t)$ sisätuloksi. Mikäli kyseiset funktiot ovat neliöintegroituvia, Cauchyn epäyhtälöstä seuraa

$$\begin{aligned}\mathbb{E}_\pi(\zeta_0^2) &\leq \sum_{s=0}^{\infty}\sum_{t=0}^{\infty} \sqrt{\mathbb{E}_\pi[|f(X_s)|^2\mathbf{1}(T>s)]\mathbb{E}_\pi[|f(X_t)|^2\mathbf{1}(T>t)]}\\ &= \left\{\sum_{t=0}^{\infty} \sqrt{\mathbb{E}_\pi[|f(X_t)|^2\mathbf{1}(T>t)]}\right\}^2\end{aligned}$$

Pyrimme seuraavaksi arvioimaan odotusarvoa käyttämällä Hölderin epäyhtälöä. Merkitään $p = 1 + 2/\varepsilon$ ja $q = 1 + \varepsilon/2$. Tällä valinnalla p ja q ovat toistensa Hölder-konjugaatit eli $1/p + 1/q = 1$. Nyt

$$\mathbb{E}_\pi[|f(X_t)|^2\mathbf{1}(T>t)] \leq \mathbb{E}_\pi[\mathbf{1}(T>t)^p]^{1/p}\mathbb{E}_\pi[|f(X_t)|^{2q}]^{1/q}$$

Odotusarvo lasketaan alkujakaumana tasapainojakauma π, mikä merkitsee, että kukin X_t on jakautunut kuten π. Funktiota f koskevan integroituvuusoletuksen nojalla

$$\begin{aligned}\mathbb{E}_\pi[|f(X_t)|^{2q}]^{1/q} &= \mathbb{E}_\pi[|f(X_t)|^{2+\varepsilon}]^{1/q}\\ &= \left\{\int |f(x)|^{2+\varepsilon}\pi(dx)\right\}^{1/q} \triangleq K < \infty.\end{aligned}$$

Geometrisen ergodisuuden ja lemman 3.9 nojalla löytyy sellainen $\gamma > 1$, jolle $\mathbb{E}_\pi(\gamma^T) < \infty$. Markovin epäyhtälön (kts. liite A) nojalla

$$\mathbb{E}_\pi[\mathbf{1}(T>t)] = \mathbb{P}_\pi(T>t) = \mathbb{P}_\pi(\gamma^T > \gamma^t) \leq \frac{\mathbb{E}_\pi(\gamma^T)}{\gamma^t} < \infty.$$

Edellinen merkitsee sitä, että $|f(X_t)|\mathbf{1}(T > t)$ on neliöintegroituva funktio ja siten Cauchyn epäyhtälön käyttö edellä on perusteltua. Yhdistämällä saadut tulokset seuraa

$$\mathbb{E}_\pi(\zeta_0^2) \leq K\mathbb{E}_\pi(\gamma^T)^{1/p}\left\{\sum_{t=0}^{\infty}\gamma^{-t/2p}\right\}^2.$$

Mutta nyt $\gamma^{-1/2p} < 1$, joten sarja suppenee geometrisen sarjan summana. Olemme näyttäneet, että $\mathbb{E}_\pi(\zeta_0^2) < \infty$. $\square$

LUKU 4

Metropolisin ja Hastingsin algoritmi

1. Metropolisin ja Hastingsin algoritmin perusteet

Tässä luvussa sovelletaan edellä rakennettua yleisten Markovin ketjujen teoriaa jatkuvalla tila-avaruudella Metropolisin ja Hastingsin algoritmin ymmärtämiseksi. Metropolisin ja Hastingsin algoritmi on eräs varsin yleinen MCMC-algoritmien luokka ja sen ymmärtäminen luo hyvän perustan muiden MCMC-algoritmien ymmärtämiseksi.

1.1. MCMC-menetelmien peruskäsitteitä.

Oletetaan, että π on jokin epänegatiivinen funktio, jonka integraali yli sen määrittelyjoukon on aidosti positiivinen mutta kuitenkin äärellinen. Yleensä π on jonkin todennäköisyysjakauman mahdollisesti normalisoimaton tiheysfunktio ja tämän vuoksi funktiota π nimitetään jatkossa yleisesti jakaumaksi. Valitaan tila-avaruudeksi $E \subset \mathbb{R}^d$ jakauman π tuki eli se avaruuden $\mathbb{R}^d$ osajoukko, jossa π on aidosti positiivinen.

MCMC-menetelmien lähtökohtana on konstruoida sellainen siirtymätodennäköisyysydin P, että π on sen tasapainojakauma. Lisäksi tarvitaan alkutila x_0, joka voi olla mielivaltainen tila-avaruuden piste. Siirtymätodennäköisyysydin P ja alkutila x_0 määrittelevät Markovin ketjun $(X_t : t \geq 0)$, jolle $X_0 = x_0$ todennäköisyydellä yksi.

Kun Markovin ketjua (X_t) ajetaan t askelta, saadaan Markovin ketjun realisaatio eli jono $(x_0, \ldots, x_t)$, jonka jäsenten välillä on stokastista riippuvuutta. Mikäli Markovin ketju on ergodinen, satunnaisvektoreiden X_s jakaumat suppenevat kokonaisvariaatioetäisyyden mielessä kohti tasapainojakaumaa π. Aikaa, joka tarvitaan riittävän suppenemisen saavuttamiseksi, nimitetään *alustusajaksi* (*burn-in time*). Alustusajalle ei ole tarkkaa määritelmää vaan kyseessä on empiirinen käsite. Mikäli simulaation tuloksena saadaan suhteellisen lyhyt Markovin ketjun realisaatio, voi ketjun ensimmäisten iteraatioiden hylkääminen olla tarpeen, jotta ketjun alussa tasapainotilasta kaukana olevat tilat eivät vaikuta johtopäätöksiin. Tarvittavan alustusajan määrittäminen voi joissain tapauksissa olla mahdollista analyyttisiä keinoja käyttäen [**JH01**].

Ergodisuusoletuksen voimassa ollessa suurten lukujen laki takaa, että, jos Markovin ketjun realisaatiota $(x_0, x_1, \ldots, x_t)$ tarkastellaan riittävän pitkällä aikavälillä, niin $\hat{\pi}_t(f) \to \pi(f)$ todennäköisyydellä yksi riippumatta Markovin ketjun alkutilasta x_0. Käytännössä Markovin ketjua ajetaan aina jokin äärellinen aika, jolloin voi olla tarpeen hylätä alustusaikaa edeltävä osa realisaatiosta, sillä ennen tasapainotilan saavuttamista Markovin ketju saattaa viipyä suhteettoman pitkän ajan tiloissa, joiden tasapainotodennäköisyys on pieni. Ilman ketjun alkupään hylkäämistä luku $\hat{\pi}_t(f)$ saattaa olla epätarkka approksimaatio luvulle $\pi(f)$, mikäli tasapainojakauman saavuttamiseen kuluva aika on merkittävä osuus koko realisaation pituudesta.

Jotta MCMC-simulaatio tuottaisi luotettavia tuloksia on syytä kiinnittää huomiota Markovin ketjun *sekoittumiseen.* Tämä tarkoittaa, että Markovin ketju vierailee riittävän usein kaikissa tila-avaruuden osajoukoissa, jotta Markovin ketjun realisaatio muodostaisi kattavan otoksen tasapainojakauman tuesta. Käytännön simulaatioissa on usein kyse tasapainottelusta Markovin ketjun konvergenssivauhdin ja sekoittumisen välillä. Nopea konvergenssi saattaa johtaa hitaaseen sekoittumiseen ja kääntäen nopea sekoittuminen hitaaseen konvergenssiin. Moni MCMC-algoritmi sisältää parametreja, joilla voi vaikuttaa konvergenssivauhdin ja sekoittumisen väliseen suhteeseen, kts. esim. [**Ros11**]. Konvergenssivauhtiin ja sekoittumiseen voi vaikuttaa myös muuttujanvaihdolla tai ryhmittelemällä parametreja sopivasti, kts. esim [**GR96**].

Ei ole olemassa mitään yleistä sääntöä siihen, onko parempi muodostaa yksi pitkä realisaation vai useampi lyhyempää realisaatiota – [**Gey11**] luku 1.11 tosin argumentoi varsin jyrkin sanankääntein yhden pitkän otoksen puolesta myös alustusajan hyödyllisyyttä kritisoiden. Pitkä otos voi johtaa parempaan sekoittumiseen. Toisaalta useamman otoksen vertaaminen toisiinsa voi paljastaa puutteellisen suppenemisen kohti tasapainojakaumaa. Kumpikin strategia voi olla hyödyllinen, kun MCMC-algoritmi on muodostettu ja sen ominaisuuksia tutkitaan. Lopullisia johtopäätöksiä ei koskaan tule perustaa puutteellisesti testattuun algoritmiin.

1.2. Metropolisin ja Hastingsin algoritmi.

Olkoon q jokin ehdollinen muotoa $q(y \mid x)$ oleva tiheysfunktio, jossa $x, y \in E$. Tiheysfunktion q määrittelemää todennäköisyysjakaumaa nimitetään *ehdokasjakaumaksi* (*proposal distribution*). Määritellään kaikilla $x, y \in E$ *Metropolisin ja Hastingsin suhde* asettamalla

$$r(x,y) = \frac{\pi(y)q(x \mid y)}{\pi(x)q(y \mid x)}.$$

Edelleen määritellään *hyväksymistodennäköisyys*

$$\alpha(x,y) = \min\{1, r(x,y)\}.$$

Mikäli $\pi(x)q(y \mid x) = 0$, on osamäärä muotoa vakio jaettuna nollalla. Sovitaan, että osamäärä saa tällöin arvon $+\infty$, jolloin hyväksymistodennäköisyydeksi tulee yksi. Käytännössä tilanteeseen, jossa Metropolisin ja Hastingsin suhteen nimittäjä häviäisi, ei koskaan päädytä, sillä oletuksen $E = \operatorname{supp} \pi$ nojalla pätee aina $\pi(x) > 0$, ja seuraavaksi esitettävän Metropolisin ja Hastingsin algoritmin konstruktion nojalla koskaan ei päädytä tilanteeseen, jossa x ja y olisivat sellaiset, että $q(y \mid x) = 0$.

Oletetaan, että Markovin ketju on ajan hetkellä t tilassa $X_t = x$. Metropolisin ja Hastingsin algoritmin iteraatio etenee seuraavasti. Generoidaan ehdollisesta jakaumasta $q(\cdot \mid x)$ *kandidaatti* y. Kandidaatti hyväksytään todennäköisyydellä $\alpha(x,y)$. Jos kandidaatti hyväksytään, ketju siirtyy tilaan y eli $X_{t+1} = y$. Mikäli kandidaatti tulee hylätyksi, pysyy Markovin ketju tilassa x eli $X_{t+1} = x$.

Metropolisin ja Hastingsin algoritmin siirtymätodennäköisyystiheys on $p(x, y) = \alpha(x, y)q(y \mid x)$. Todennäköisyys sille, että ketju pysyy paikoillaan, on

$$r(x) = 1 - \int p(x, y)dy = 1 - \int \alpha(x, y)q(y \mid x)dy.$$

Metropolisin ja Hastingsin algoritmin määrittelemän Markovin ketjun siirtymätodennäköisyysydin on siten muotoa

$$P(x, A) = \int_A \alpha(x, y)q(y \mid x)dy + r(x)\mathbf{1}_A(x).$$

Siirtymätodennäköisyysydintä P nimitetään Metropolisin ja Hastingsin ytimeksi ja Markovin ketjua (X_t) Metropolisin ja Hastingsin ketjuksi.

Näemme, että Metropolisin ja Hastingsin algoritmi riippuu jakaumasta π ainoastaan osamäärän $\pi(y)/\pi(x)$ kautta. Täten on perusteltua käyttää normittamatonta jakaumaa π, sillä tuntematon normitustekijä supistuu pois.

LAUSE 4.1. Metropolisin ja Hastingsin ketju on kääntyvä. Jakauma π toteuttaa kääntyvyysehdon ja on siten Markovin ketjun tasapainojakauma.

TODISTUS. Väite seuraa lauseesta 2.5, sillä pätee

$$\begin{aligned}\pi(x)p(x, y) &= \pi(x)q(y \mid x) \min\left\{1, \frac{\pi(y)q(x \mid y)}{\pi(x)q(y \mid x)}\right\}\\ &= \min\left\{\pi(x)q(y \mid x), \pi(y)q(x \mid y)\right\}\\ &= \pi(y)q(x \mid y) \min\left\{\frac{\pi(x)q(y \mid x)}{\pi(y)q(x \mid y)}, 1\right\} = \pi(y)p(y, x).\end{aligned}$$

□

Seuraava lause antaa riittävän ehdon sille, että Metropolisin ja Hastingsin ketju on ergodinen.

LAUSE 4.2. Metropolisin ja Hastingsin ketju (X_t) on ergodinen, mikäli $q(y \mid x) > 0$ kaikilla $x, y \in E$ ja lisäksi π ja q ovat jatkuvia.

TODISTUS. Jatkuvuusoletuksesta seuraa, että Metropolisin ja Hastingsin suhde r on jatkuva ja edelleen hyväksymistodennäköisyys α on jatkuva kahden jatkuvan funktion miniminä. Siten Markovin ketjun (X_t) subtodennäköisyystiheysfunktio $p(x, y) = \alpha(x, y)q(y \mid x)$ on jatkuva kahden jatkuvan funktion tulona. Mikäli $x_0 \in E$ on mielivaltainen ja $r > 0$, löytyy positiivismittainen kompakti joukko $K \subset B(x_0, r)$. Valitaan mielivaltainen $x \in E$. Ehdokasjakaumaa koskevan positiivisuusoletuksen nojalla $p(x, y) > 0$ jokaisella $y \in E$ ja jatkuvuuden nojalla $p(x, y) \geq C$ jollakin $C > 0$ kullakin $y \in K$. Siten voimme arvioida

$$P(x, B(x_0, r)) \geq \int_K p(x, y)dy \geq C\text{vol}(K) > 0$$

eli $x \to B(x_0, r)$. Koska $r > 0$ ja $x \in E$ voitiin valita mielivaltaisiksi, seuraa lauseesta 2.6, että Markovin ketju toteuttaa pelkistymättömyys- ja minorisaatio-oletuksen ja edelleen ergodisuusoletuksen. □

Perusmuotoisesta Metropolisin ja Hastingsin algoritmista on olemassa lukuisia variaatioita. Erityisen yksinkertaiseen muotoon päästään valitsemalla ehdokasjakauma symmetrisesti siten, että $q(y \mid x) = q(x \mid y)$. Tällöin hyväksymistodennäköisyys saa muodon

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \right\}.$$

Symmetristä ehdokasjakaumaa käyttävää algoritmia nimitetään usein Metropolisin algoritmiksi.

Edelleen eräs erikoistapaus symmetrisestä ehdokasjakaumasta on muotoa $q(y \mid x) = g(x - y)$ oleva jakauma, kun g on jonkin origosymmetrisen jakauman tiheysfunktio. Näin muodostettu algoritmi on *satunnaiskulku-*Metropolisin algoritmi. Osoitetaan seuraavaksi satunnaiskulku-Metropolisin algoritmin ergodisuuden takaava ehto, joka luopuu lauseen 4.2 oletuksesta, jonka mukaan ehdokasjakauma on kaikkialla aidosti positiivinen. Lause on variaatio [**RC04**] lemmasta 7.6.

Lause 4.3. Satunnaiskulku-Metropolisin ketju (X_t) on ergodinen, jos tila-avaruus E on yhtenäinen, tasapainojakauma π on jatkuva ja g on jatkuva ja aidosti positiivinen joukossa $B(0, \varepsilon)$ jollakin $\varepsilon > 0$.

Todistus. Kiinnitetään jokin $x_0 \in E$. Argumentoimalla samoin kuin lauseen 4.2 todistuksessa voidaan näyttää, että siirtymätodennäköisyystiheysfunktio $p(x, y)$ on aidosti positiivinen ja jatkuva pisteessä (x_0, x_0). Kiinnitetään mielivaltainen $z_0 \in E$ ja $r > 0$. Haluamme osoittaa, että $z_0 \to B(x_0, r)$. Merkitään $\delta = \varepsilon/4$. Tila-avaruuden yhtenäisyyden nojalla löytyy jono $(z_i : 1 \leq i \leq t)$ jollakin $t \geq 1$ siten, että $B(z_i, \delta) \cap B(z_{i+1}, \delta) \neq \emptyset$ jokaisella $0 \leq i < t$ sekä $B(z_t, \delta) \cap B(x_0, r) \neq \emptyset$. Mikäli $x \in B(z_i, \delta)$, seuraa oletuksesta, että $p(x, y) > 0$ kaikilla $y \in B(z_{i+1}, \delta)$. Tämä merkitsee, että Markovin ketjulla on mahdollisuus päästä pisteestä x joukkoon $B(z_{i+1}, \delta)$ positiivisella todennäköisyydellä yhdellä aika-askeleella. Konstruktion nojalla pisteestä z_0 on mahdollista päästä kunkin kuulan $B(z_0, \delta), \ldots, B(z_t, \delta)$ läpi kulkemalla kuulaan $B(x_0, r)$ positiivisella todennäköisyydellä äärellisessä ajassa. Siispä mielivaltaisilla $x \in E$ ja $r > 0$ pätee $x \to B(x_0, r)$. Lauseen 2.6 nojalla Markovin ketju toteuttaa pelkistymättömyys- ja minorisaatio-oletuksen ja edelleen ergodisuusoletuksen. $\square$

Eräs yksinkertainen erikoistapaus Metropolisin ja Hastingsin algoritmista saadaan valitsemalla ehdokasjakauma q siten, että kandidaatin y jakauma on riippumaton Markovin ketjun edellisestä tilasta. Toisin sanoen q on muotoa $q(y \mid x) = q(y)$. Hyväksymistodennäköisyys saa muodon

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y) q(x)}{\pi(x) q(y)} \right\}.$$

Tällaisesta algoritmista käytetään kirjallisuudessa yleensä nimitystä *independence sampler*. Algoritmin yksinkertaisuuden vuoksi sen teoreettisia ominaisuuksia koskeva kirjallisuus on kattava, mutta käytännössä algoritmin sovellukset ovat varsin rajallisia.

Esimerkki 4.1. Oletetaan, että tila-avaruudella $E = \mathbb{R}$ on määritelty standardinormaalijakauman tiheysfunktio π. Simuloidaan jakaumasta käyttäen satunnaiskulku-Metropolisin algoritmia, jonka ehdokasjakauma on tasajakauma eli $q(y \mid x) = g(x - y) = \mathbf{1}_{(-c,c)}(x - y)$ jollakin $c > 0$. Mikäli

$\varepsilon < c$ seuraa, että g on jatkuva ja aidosti positiivinen joukossa $B(0, \varepsilon)$, mikä merkitsee lauseen 4.3 nojalla, että satunnaiskulku-Metropolisin ketju on ergodinen. Metropolisin ja Hastingsin suhde on muotoa

$$r(x, y) = \frac{\pi(y)}{\pi(x)} = \exp\left\{\frac{x^2 - y^2}{2}\right\}.$$

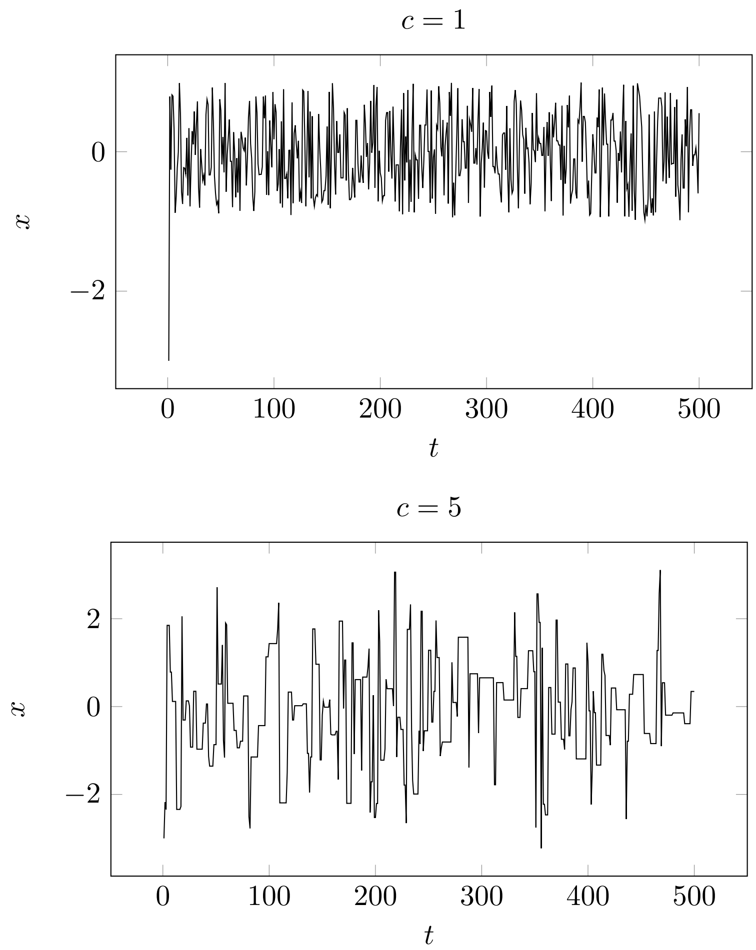
Esimerkki 4.2. Jatketaan esimerkin 4.1 käsittelyä suorittamalla numeerinen simulaatio. Valitaan alkutila $x_0 = -3$ ja simuloidaan kaksi 500 iteraation pituista Markovin ketjua parametrin c arvoilla 1 ja 5. Keskimääräinen hyväksymistodennäköisyys α voidaan estimoida kullakin algoritmin iteraatiolla laskettujen hyväksymistodennäköisyyksien keskiarvona. Simulaation tulokset on esitetty kuvassa 1. Parametrin arvolla $c = 1$ keskimääräiseksi hyväksymistodennäköisyydeksi saadaan 0,92 ja parametrin arvolla $c = 5$ havaitaan keskimääräinen hyväksymistodennäköisyys 0,33. Kuvasta nähdään välittömästi, että ensimmäisessä tapauksessa Markovin ketjun sekoittuminen on erittäin huonoa. Markovin ketju ei näytä pääsevän 500 iteraation aikana ensimmäistä aloitustilaa lukuun ottamatta kertaakaan lähellekään tiloja -2 tai 2. Kuitenkin lähes 5 % standardinormaalijakauman tuesta sisältyy alueeseen $|x| > 2$. Jälkimmäisessä tapauksessa nähdään, että sekoittuminen on huomattavasti edellistä parempaa. Havainnot ovat sopusoinnussa teoreettisten ja simulaatiotulosten kanssa, joiden mukaan satunnaiskulku-Metropolisin algoritmin hyväksymistodennäköisyyden optimaalinen arvo on alle luvun 0,5. Aihetta käsittelee tarkemmin [**Ros11**] luku 4.2.

1.3. Lyhyesti MCMC-menetelmien historiasta.

Monte Carlo -menetelmät saivat alkuna 1940-luvulla ensimmäisen tietokoneen, ENIACin keksimisen myötä. Ajatus Monte Carlo -menetelmien toiminnasta lienee käynyt monen ihmisen mielessä tätä aiemminkin, mutta laskentakapasiteetin puute esti niiden soveltamisen käytännön työhön – tosin tiettävästi teoreettinen fyysikko Enrico Fermi teki Monte Carlo -simulaatioita jo 1930-luvulla apunaan pelkkä mekaaninen laskukone. 1940-luvulla Monte Carlo -menetelmiä sovellettiin ydinfysiikassa esiintyneisiin ongelmiin, joita ilmeni mm. ensimmäisen fuusiopommin suunnittelun yhteydessä. [**Met87**]

Ensimmäinen MCMC-algoritmi julkaistiin artikkelissa [**MRR$^+$53**]. Kyseessä on satunnaiskulku-Metropolisin algoritmi, jota sovellettiin molekyylien välisten vuorovaikutusten mallintamiseen kahdessa dimensiossa. Molekyylejä simulaatiossa oli 224 kappaletta ja algoritmia ajettiin vajaa sata iteraatiota. Silloisella MANIAC-tietokoneella aikaa kului viitisen tuntia. MCMC-menetelmät saavuttivat pian vakiintuneen aseman osana fyysikoiden ja kemistien työkaluarsenaalia.

[**Has70**] ja [**Pes73**] tarkastelivat Metropolisin algoritmia tilastotieteellisessä kontekstissa varsin yleisellä tasolla ja esittivät siirtymätodennäköisyydet $p(x, y)$ tässä tutkielmassa määritellyssä muodossa. Artikkeli [**Has70**] ei aikanaan johtanut laajaan innostukseen, mikä johtunee siitä, että tietokone ei vielä tuolloin ollut jokaisen tilastotieteilijän ulottuvilla.



KUVA 1. Esimerkin 4.2 numeeristen simulaatioiden tulokset

Gibbsin algoritmia, joka voidaan nähdä erikoistapauksena Metropolisin ja Hastingsin algoritmista, käytettiin 1980-luvulla tietyillä erikoisaloilla kuten kuvankäsittelyssä – historiallisesti merkittävä rooli on erityisesti artikkelilla [**GG84**]. Edelleen Gibbsin algoritmiin perustuvat artikkelit [**GS90**] ja [**GHRPS90**] toivat MCMC-menetelmät tilastotieteilijöiden laajaan tietoisuuteen. Historia osoittaa, että aika oli kypsä, mikä johti laajaan innostukseen MCMC-menetelmiä kohtaan. Syvällisemmän katsauksen Metropolisin ja Hastingsin algoritmin sekä laajemmin MCMC-menetelmien historiaan tilastotieteen näkökulmasta muodostavat artikkelit [**Hit03**] ja [**RC11**].

2. Satunnaiskulku-Metropolisin algoritmin geometrinen ergodisuus

Tässä luvussa tarkastellaan esimerkinomaisesti satunnaiskulku-Metropolisin algoritmin geometristä ergodisuutta. Yksiulotteisella tila-avaruudella muotoillaan artikkelin [**MT96**] teoreema 3.2, joka antaa riittävän ehdon Markovin ketjun geometriselle ergodisuudelle. [**RT96**] sekä [**JH00**] yleistävät tuloksen koskemaan satunnaiskulku-Metropolisin algoritmia useammassa ulottuvuudessa, mutta tämä vaatii tasapainojakaumaa π koskevia lisäsäännöllisyysoletuksia.

MÄÄRITELMÄ 4.1. Jatkuvan tiheysfunktion π hännät ovat *log-konkaavit*, jos löytyy luvut $\alpha > 0$ ja $x_0 > 0$, joille

$$\log \frac{\pi(x)}{\pi(y)} \geq \alpha(y - x) \quad \text{aina, kun } y \geq x \geq x_0$$
$$\log \frac{\pi(x)}{\pi(y)} \geq \alpha(x - y) \quad \text{aina, kun } y \leq x \leq -x_0.$$

LAUSE 4.4. Olkoon π jatkuva tiheysfunktio, jonka hännät ovat log-konkaavit. Oletetaan, että satunnaiskulku-Metropolisin algoritmissa ehdokasjakauma $q(y \mid x) = g(x - y)$ on aidosti positiivinen ja jatkuva sekä pätee $g(x) \leq be^{-\alpha|x|}$ kaikilla x jollakin $0 < b < \infty$. Tällöin algoritmin määrittelemä Markovin ketju on geometrisesti ergodinen.

ESIMERKKI 4.3. Jatketaan esimerkin 4.1 tarkastelua ja osoitetaan Markovin ketju geometrisesti ergodiseksi. Kiinnitetään jokin $x_0 > 0$. Nyt

$$\log \frac{\pi(x)}{\pi(y)} = \frac{1}{2}y^2 - \frac{1}{2}x^2 = \frac{1}{2}(x + y)(y - x) \geq x_0(y - x),$$

kun $y \geq x \geq x_0$. Vastaavasti, jos $y \leq x \leq -x_0$ saadaan $x + y \leq -2x_0$. Toisaalta $y - x < 0$, joten seuraa

$$\log \frac{\pi(x)}{\pi(y)} \geq x_0(x - y).$$

Valitsemalla $b < \infty$ riittävän suureksi pätee $g(x) \leq be^{-x_0|x|}$ kaikilla x, joten satunnaiskulku-Metropolisin algoritmin määrittelemä Markovin ketju on geometrisesti ergodinen.

Mikäli kiinnostuksen kohteena on simuloitavan jakauman odotusarvo, on sen estimaatti $\hat{\mu}_t = \frac{1}{t}\sum_{s=0}^{t-1} x_s$, kun $x_0, x_1, \ldots, x_{n-1}$ on simulaation tuloksena saatu Markovin ketjun realisaatio. Suurten lukujen laki takaa, että $\hat{\mu}_t \to \mu$ todennäköisyydellä yksi, kun iteraatioiden lukumäärä t kasvaa rajatta. Suurten lukujen laissa käytettävä funktio f on identiteettifunktio, joka toteuttaa integroituvuusehdon

$$\int |f(x)|^{2+\varepsilon}\pi(dx) < \infty.$$

Odotusarvon estimaattori $\hat{\mu}_t$ toteuttaa siten keskeisen raja-arvolauseen, jonka mukaan

$$\sqrt{t}[\hat{\mu}_t - \mu] \xrightarrow{d} \mathrm{N}(0, \sigma^2)$$

jollakin $\sigma^2 > 0$. Saadusta tuloksesta voi päätellä, että odotusarvon estimaattorin tarkkuus esimerkiksi keskihajonnan mielessä on kääntäen verrannollinen otoskoon t neliöjuureen. Mikäli $t = 1000$ iteraatiolla pystytään saavuttamaan tarkkuus ensimmäisessä desimaalissa, tarvitaan tarkkuuden saavuttamiseen toisessa desimaalissa jo 100 000 iteraatiota.

Artikkelissa [**MT96**] s. 112 osoitetaan seuraava tulos. Oletetaan, että tiheysfunktio π on jatkuva ja symmetrinen. Mikäli $q(y \mid x) = g(x - y)$ on jatkuva satunnaiskulku-Metropolisin algoritmin ehdokasjakauma, jolle $\int |x|g(x)dx < \infty$, niin syntyvä Markovin ketju on geometrisesti ergodinen tasan siinä tapauksessa, että $\lim_{x\to\infty} \frac{d}{dx} \log \pi(x) < 0$ edellyttäen, että derivaatta on määritelty riittävän suurilla x. Tuloksen nojalla on helppo antaa esimerkki satunnaiskulku-Metropolisin algoritmista, joka ei ole geometrisesti ergodinen.

Esimerkki 4.4. Oletetaan, että tila-avaruudella $E = \mathbb{R}$ on määritelty standardi-Cauchy-jakauman tiheysfunktio, joka on muotoa

$$\pi(x) \propto \frac{1}{1+x^2}.$$

Suoralla laskutoimituksella nähdään, että

$$\lim_{x\to\infty} \frac{d}{dx} \log \pi(x) = -\lim_{x\to\infty} \frac{2x}{1+x^2} = 0,$$

Mikäli ehdokasjakauma q valitaan lisäksi siten, että se toteuttaa [**MT96**] tuloksen oletukset, seuraa, että satunnaiskulku-Metropolisin algoritmin tuottama Markovin ketju ei ole geometrisesti ergodinen.

3. Soveltava esimerkki

Tarkastellaan soveltavaa esimerkkiä, jonka idea on peräisin artikkelista [**Dia09**]. Eräs yksinkertaisimmista tavoista salakirjoittaa annettu teksti on muodostaa bijektio tekstissä käytettyjen merkkien joukolta samalle joukolle. Tällöin kutakin selkokielistä merkkiä vastaa tasan yksi salakirjoitusmerkki. Mikäli teksti sisältää n eri kirjoitusmerkkiä on mahdollisten salakirjoitusavainten määrä n-pituisen jonojen permutaatioiden lukumäärä $n!$. Kaikkien salakirjoitusavainten läpikäyminen on siten käytännössä mahdoton tapa purkaa salakirjoitus millään realistisen kokoisella n.

Tarkastellaan yksinkertaistettua tilannetta, jossa teksti koostuu 29 kirjoitusmerkistä, joihin sisältyy väli ja suomenkielen aakkoset kirjainta å lukuun ottamatta. Salakirjoitettava teksti puhdistetaan siten välimerkeistä ja isot kirjaimet muutetaan pieniksi. Säilytetään välit ennallaan ja salakirjoitetaan aakkoset käyttäen satunnaisesti muodostettua jonon $(1, 2, \ldots, 28)$ permutaatiota. Tekstiksi valittiin ensimmäinen kappale *Aleksis Kiven* romaanista *Seitsemän veljestä*. Seuraavassa on annettu selkokielinen teksti:

> jukolan talo eteläisessä hämeessä seisoo erään mäen pohjaisella rinteellä liki toukolan kylää sen läheisin ympäristö on kivinen tanner mutta alempana alkaa pellot joissa ennenkuin talo oli häviöön mennyt aaltoili teräinen vilja peltojen alla on niittu apilaäyräinen halkileikkaama monipolvisen ojan ja runsaasti antoi se heiniä ennenkuin joutui laitumeksi kylän karjalle muutoin on talolla avaria metsiä soita ja erämaita jotka tämän tilustan ensimmäisen perustajan oivallisen toiminnan kautta olivat langenneet sille

> osaksi jo isonjaon käydessä entisinä aikoina silloinpa jukolan isäntä pitäen enemmän huolta jälkeentulevainsa edusta kuin omasta parhaastansa otti vastaan osaksensa kulon polttaman metsän ja sai sillä keinolla seitsemän vertaa enemmän kuin toiset naapurinsa mutta kaikki kulovalkean jäljet olivat jo kadonneet hänen piiristänsä ja tuuhea metsä kasvanut sijaan ja tämä on niiden seitsemän veljen koto joiden elämänvaiheita tässä nyt käyn kertoilemaan

Salakirjoitetun tekstin alku on

> daewäkt bkäw öböäuosössu julöössu söosww öpuut luöt rwjdkosöääk potbööääu äoeo bwaewäkt emäuu söt äujöosot mlruposbq

Tavoitteena on käyttää Metropolisin ja Hastingsin algoritmia salakirjoituksen purkamiseen, kun salausavain on tuntematon. Tämä perustuu siihen, että tarkastellaan siirtymiä peräkkäisten kirjaimien välillä. Myös siirtymät kirjaimesta väliin ja välistä kirjaimeen huomioidaan. Edellytyksenä on, että meillä on käytössä mahdollisimman kattava aineisto suomenkielistä tekstiä. Tätä varten valittiin *Leo Tolstoin* romaanin *Anna Karenina* suomennos. Teksti käsiteltiin samoin kuin salakirjoitettavaksi tarkoitettu teksti. Tämä sisältää liki 2 miljoonaa merkkiä ja kahden merkin välistä siirtymää. Muodostetaan matriisi $A = (a_{ij} : 1 \leq i, j \leq 29)$, jonka alkio a_{ij} antaa todennäköisyyden siirtyä merkistä i merkkiin j. Oletetaan, että merkkien ja lukujen $1, \ldots, 29$ välillä on määritelty bijektio siten, että luku yksi vastaa merkkiä a, luku 2 merkkiä b ja niin edelleen kunnes luku 28 vastaa merkkiä ö ja luku 29 väliä. Muodostetaan A siten, että sen kunkin rivin summa on yksi. Tehdään lisäksi yksinkertaistava tekninen oletus, jonka mukaan $a_{ij} > 0$ kaikilla $i, j \in E$. Tämä saavutetaan esimerkiksi lisäämällä kaikkiin siirtymäfrekvensseihin luku yksi.

Tila-avaruutena E on nyt jonon $(1, 2, \ldots, 28)$ kaikkien permutaatioiden joukko. Tila-avaruuden dimensio $d = 28$ ja tila-avaruus on äärellinen muodostuen 28! tilasta. Sigma-algebra $\mathcal{E}$ on tila-avaruuden potenssijoukko eli kaikkien osajoukkojen kokoelma. Mittana $\mathrm{vol}(\cdot)$ on lukumäärämitta.

Merkitään σ jonon $(1, \ldots, 28)$ permutaatiota ja käytetään funktiomerkintää $\sigma(i)$ merkitsemään luvun i permutoitua arvoa. Sovitaan lisäksi $\sigma(29) = 29$, mikä vastaa oletusta, että välit pysyvät muuttumattomina.

Merkitään salakirjoitetun tekstin pituutta n ja itse salakirjoitettua tekstiä $w = (w_1, \ldots, w_n)$, jossa $1 \leq w_i \leq 29$ kullakin i. Määritellään permutaation σ uskottavuus

$$\pi(\sigma) = \prod_{i=1}^{n-1} a_{\sigma(w_i), \sigma(w_{i+1})}.$$

Tulo on ilmeisesti aidosti positiivinen ja lisäksi luvun n ja matriisin A alkioiden sekä tila-avaruuden äärellisyyden nojalla

$$\int \pi(d\sigma) = \sum_{\sigma \in E} \prod_{i=1}^{n-1} a_{\sigma(w_i), \sigma(w_{i+1})} < \infty$$

vaikkakin normitustekijä on käytännössä mahdoton laskea. Tämä merkitsee, että π voidaan tulkita todennäköisyysjakauman normittamattomaksi tiheysfunktioksi tai diskreettiä terminologiaa käyttäen pistemassafunktioksi. On ilmeistä, että π saa suhteessa suuria arvoja silloin, kun permutoidun salakirjoituksen siirtymätodennäköisyydet vastaavat hyvin referenssiaineiston siirtymätodennäköisyyksiä. Toivottavaa on, että jakauma π olisi yksimoodinen ja moodi olisi salakirjoitusavaimen käänteiskuvaus. Mikäli salakirjoitettu aineisto on riittävän pitkä, tämä lienee melko todennäköistä.

Konstruoidaan seuraavaksi Metropolisin algoritmi, jonka tasapainojakauma on π. Valitaan ehdokasjakauma q siten, että se arpoo kaksi erillistä indeksiä $1 \leq i, j \leq 28$ ja vaihtaa permutaation arvot $\sigma(i)$ ja $\sigma(j)$ päittäin. Näin muodostettua permutaatiota σ^* nimitetään permutaation σ *transpositioksi*. Ehdokasjakauma on symmetrinen eli pätee $q(\sigma^* \mid \sigma) = q(\sigma \mid \sigma^*)$, minkä nojalla Metropolisin suhde on

$$r(\sigma, \sigma^*) = \frac{\pi(\sigma^*)}{\pi(\sigma)}.$$

Metropolisin ytimen subtodennäköisyystiheysfunktio on

$$p(\sigma, \sigma^*) = \min\left\{1, \frac{\pi(\sigma^*)}{\pi(\sigma)}\right\} q(\sigma^* \mid \sigma).$$

Valitaan mielivaltainen permutaatio σ_C ja muodostetaan yksiö $C = \{\sigma_C\}$. Olkoon σ_D jokin permutaation σ_C transpositio ja muodostetaan yksiö $D = \{\sigma_D\}$. Tällöin $p(\sigma_C, \sigma_D) = \beta > 0$. Lisäksi on ilmeistä, että mikäli σ on mielivaltainen permutaatio, niin äärellisellä määrällä transpositioita on mahdollista päästä permutaatioon σ_C eli $\sigma \to C$. Metropolisin ketju toteuttaa siis pelkistymättömyys- ja minorisaatio-oletuksen ja koska sillä on lisäksi tasapainojakauma, toteuttaa se myös ergodisuusoletuksen. Markovin ketju toteuttaa siis suurten lukujen lain ja sen tilajakauma suppenee kokonaisvariaatioetäisyyden mielessä kohti tasapainojakaumaa π.

Toteutetaan Metropolisin algoritmi R-kielellä. Markovin ketjun alkutilaksi valitaan $\sigma_0 = (1, 2, \ldots, 28)$. Ajetaan Markovin ketjua 2000 iteraatiota ja tallennetaan Markovin ketjun tila iteraatioilla 100, 500, 1000 ja 2000. Seuraavassa on esitetty osa kutakin permutaatiota vastaavasta puretusta viestistä:

> 100: äkoumat ramu yrymeisysse jelyysse syisuu ypeet leyt zujäaisymma pitryymme mioi rukoumat odmee syt mejyisit dlzepisrf
>
> 500: jopulat kalu ekelyisessy dymeessy seisuu ehyyt myet budjaisella hitkeelly lipi kuopulat pälyy set lydeisit ämbyhiskx
>
> 1000: jukolan talo eteläisessä dämeessä seisoo ehään mäen podjaisella hinteellä liki toukolan kylää sen lädeisin ympähistö
>
> 2000: jukolan talo eteläisessä hämeessä seisoo erään mäen

pohjaisella rinteellä liki toukolan kylää sen läheisin ympäristö

Osoittautuu siis, että algoritmi kykenee palauttamaan alkuperäisen salaamattoman tekstin. Tämä ei kuitenkaan tarkoita, että algoritmi kykenisi löytämään täsmälleen alkuperäisen salausavaimen. Salausavaimena käytetty permutaatio nimittäin sisältää sellaisia merkkejä – kuten b ja c –, jotka eivät esiinny salakirjoitetussa tekstissä. Tämä havaitaan myös tarkastelemalla permutaation komponenttien otoskeskiarvoja

$$\hat{\sigma}(i) = \frac{1}{1000} \sum_{t=1001}^{2000} \sigma_t(i),$$

jotka suurten lukujen lain nojalla suppenevat kullakin $1 \leq i \leq 28$ kohti lukua

$$\int \sigma(i)\pi(d\sigma) = \sum_{\sigma \in E} \sigma(i)\pi(\sigma).$$

Nimittäin esimerkiksi kirjainta a vastaten $\hat{\sigma}(1) \approx 20{,}00$ ja kirjainta c vastaten $\hat{\sigma}(3) \approx 11{,}54$. Algoritmin onnistunut toiminta perustuu siihen, että kun Markovin ketju on löytänyt oikean permutaation tekstissä esiintymättömiä merkkejä lukuun ottamatta, transpositiot, jotka muuttaisivat tätä permutaatiota tekstissä esiintyvien merkkien kohdalla ovat hyvin epätodennäköisiä.

Esimerkki on hyvä osoitus Metropolisin ja Hastingsin algoritmin monipuolisuudesta. Algoritmin toteutukseen käytetty R-koodi on esitetty liitteessä B. Artikkelissa [**Dia09**] vastaavaa algoritmia sovelletaan vankien välisen salakirjoitetun viestin purkamiseen.

LUKU 5

Loppupäätelmät

Tässä tutkielmassa on rakennettu välttämätön teoreettinen pohja, joka mahdollistaa Metropolisin ja Hastingsin algoritmin ymmärtämisen parametriavaruudella, joka on euklidisen avaruuden $\mathbb{R}^d$ osajoukko. Samaa teoriaa soveltamalla voidaan yhtä hyvin todistaa Gibbsin algoritmia koskevia tuloksia. Gibbsin algoritmi on itse asiassa Metropolisin ja Hastingsin algoritmin erikoistapaus, mutta edellisen erityispiirteiden vuoksi osa kirjallisuudesta käsittelee Gibbsin algoritmia täysin itsenäisenä algoritmina. Gibbsin algoritmissa jokainen ehdotettu siirto hyväksytään ja siirrot generoidaan muotoa $\pi(\cdot|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$ olevista ehdollisista jakaumista tila-avaruuden E dimension ollessa d, ja kun merkitään $x = (x_1, \ldots, x_d)$. Hyvä johdatus Gibbsin algoritmiin löytyy esimerkiksi artikkelista [**CG92**] ja kirjasta [**GRS96**].

Tässä tutkielmassa on tarkasteltu yhden Metropolisin ja Hastingsin ytimen muodostamaa Markovin ketjua. Vähänkin monimutkaisemmissa sovelluksissa on yleensä tarpeen muodostaa useita Metropolisin ja Hastingsin tai Gibbsin ytimiä $P_1, \ldots, P_n$. Mikäli nämä ytimet ovat riittävän säännöllisiä, voi osoittaa, että yhdistetty ydin $P_1 \cdots P_n$ tai $\frac{1}{n}\sum_{i=1}^{n} P_i$ toteuttaa myös halutut säännöllisyysominaisuudet kuten ergodisuuden tai geometrisen ergodisuuden. MCMC-menetelmiä koskeva kirjallisuus on nykyään hyvin laaja, joten tutkielman ulkopuolelle jää useita erikoisempia algoritmeja kuten [**Gre95**]. Edistyneemmistä tekniikoista löytyy hyvin tietoa kirjoista [**BGJM11**] ja [**GCS$^+$14**].

MCMC-menetelmien säännöllisyyteen liittyvät kysymykset kuten kääntyvyys, ergodisuus ja geometrinen ergodisuus tunnetaan nykyään hyvin kvalitatiivisessa mielessä. Geometrinen ergodisuus on kvalitatiivinen ominaisuus sikäli, että se ei itsessään vielä takaa, että suppeneminen olisi jossakin mielessä nopeaa. Suppenemisvauhtia koskevien kvantitatiivisten tulosten etsiminen on tällä hetkellä yksi merkittävistä MCMC-menetelmiä koskevista tutkimusalueista. Artikkelit [**JH01**] ja [**Dia09**] tarjoavat hyvän johdatuksen aiheeseen.

Bayesiläinen tilastotiede on ollut merkittävä MCMC-menetelmien tutkimusta eteenpäin ajava sovellusalue. Tilastollisten mallien monimutkaisuus on vaatinut kehittyneitä algoritmeja ja toisaalta olemassa olevien algoritmien syvällistä ymmärtämistä. Voidaan uskoa, että mallien koko tulee vain kasvamaan tulevaisuudessa, mikä luo uusia haasteita MCMC-algoritmeille [**RS15**].

LIITE A

Todennäköisyysteorian tuloksia

Mikäli $a = (a_i : i \geq 0)$ ja $b = (b_i : i \geq 0)$ ovat kaksi jonoa, niin *konvoluutio* $a \star b$ on jono, jolle

$$a \star b_i = \sum_{j=0}^{i} a_j b_{i-j}.$$

Konvoluutio toteuttaa $a \star b = b \star a$. Jonon $a = (a_i : i \geq 0)$ *generoiva funktio* on muodollinen potenssisarja

$$A(z) = \sum_{i=0}^{\infty} a_i z^i, \quad z \in \mathbb{C}.$$

Olkoot a ja b kaksi jonoa sekä A ja B niiden generoivat funktiot. Konvoluution $c = a \star b$ generoiva funktio on [**Fel50**] s. 215 nojalla

$$C(z) = A(z)B(z).$$

[**KS07**] lemma 1.27 mukaan muotoillaan seuraava tulos, joka tunnetaan myös Chebyshevin epäyhtälönä.

Markovin epäyhtälö. Mikäli X on epänegatiivinen satunnaismuuttuja, jolle $\mathbb{E}(X) < \infty$, niin jokaisella $t > 0$ pätee

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

[**KS07**] teoreema 3.26 mukaan muotoillaan seuraava tulos.

Egorovin lause. Olkoon (g_t) jono mitallisia funktioita ja oletetaan, että on olemassa sellainen mitallinen funktio g, että $g_t \to g$ melkein varmasti. Tällöin kullakin $\delta > 0$ löytyy sellainen mitallinen joukko Λ_δ, jolle $\mu(\Lambda_\delta^C) < \delta$ sekä $g_t \to g$ tasaisesti joukossa Λ_δ eli

$$\lim_{t \to \infty} \sup_{\omega \in \Lambda_\delta} |g_t(\omega) - g(\omega)| = 0.$$

[**KS07**] teoreema 7.5 mukaan muotoillaan seuraava tulos.

Kolmogorovin epäyhtälö. Olkoon (X_t) jono riippumattomia satunnaismuuttujia, joille $\mu_t = \mathbb{E}(X_t) < \infty$ sekä $\mathrm{Var}(X_t) < \infty$ kullakin t. Tällöin

$$\mathbb{P}\left(\max_{1 \leq s \leq t} \left| \sum_{k=1}^{s} (X_k - \mu_k) \right| \geq \varepsilon \right) \leq \frac{1}{\varepsilon^2} \sum_{s=1}^{t} \mathrm{Var}(X_s).$$

LIITE B

Ohjelmakoodia

1. Esimerkki 4.2

Seuraavassa on annettu R-kielinen ohjelma, jota on käytetty esimerkin 4.2 Markovin ketjun simuloimiseen.

```
1  log.pi <- function(x) return(-0.5 * x^2)
2
3  simulate <- function(x.0, N, c) {
4
5    x <- rep(0, N)
6    a.vec <- rep(0, N-1)
7    x[1] <- x.0
8
9    for (i in 2:N) {
10     x.cand <- runif(1, -c, c)
11     a <- min(1, exp(log.pi(x.cand) - log.pi(x[i-1])))
12     a.vec[i-1] <- a
13     u <- runif(1)
14     if (u < a) {
15       x[i] <- x.cand
16     }
17     else {
18       x[i] <- x[i-1]
19     }
20   }
21
22   return (list(x=x, a.hat=mean(a.vec)))
23
24 }
25
26 set.seed(0)
27
28 run.1 <- simulate(-3, 500, 1)
29 run.2 <- simulate(-3, 500, 5)
```

Rivillä 1 määritellään tasapainojakauman logaritmi. Todennäköisyystiheyksiä koskevat numeeriset arvot on aina syytä laskea logaritmeina, sillä tämä parantaa algoritmin numeerista stabiiliutta. Funktio saa parametrinaan luvun x ja palauttaa standardinormaalijakauman tiheysfunktion eksponentissa olevan luvun $-x^2/2$.

Riveillä 3–22 määritellään funktio, joka saa parametreinaan Markovin ketjun alkuarvon `x.0`, iteraatioiden lukumäärän `N` sekä parametrin `c`, joka määrää ehdokasjakauman leveyden. Riveillä 5 ja 6 muodostetaan `N`-pituiset vektorit `x` ja `a.vec` Markovin ketjun iteraattien ja hyväksymistodennäköisyyksien tallentamista varten. Rivillä 7 vektorin `x` ensimmäiseksi alkioksi asetetaan Markovin ketjun alkuarvo.

Riveillä 9–20 Markovin ketjua simuloidaan $N-1$ iteraatiota. Rivillä 10 tuotetaan kandidaatti välin $(-c, c)$ tasajakaumasta, ja rivillä 11 lasketaan hyväksymistodennäköisyys. Rivillä 13 tuotetaan otos välin $(0, 1)$ tasajakaumasta, ja sitä käytetään rivillä 14 päättämään hyväksytäänkö kandidaatti `x.cand`. Rivillä 22 palautetaan Markovin ketjun iteraatit ja keskimääräinen hyväksymistodennäköisyys listana.

Rivillä 26 kiinnitetään satunnaislukugeneraattorin alkuarvo, jotta tulokset ovat toistettavissa. Riveillä 28 ja 29 on simuloitu kaksi Markovin ketjua, jotka on esitetty kuvassa 1. Markovin ketjun iteraatiot tallennetaan tiedostoon.

2. Soveltava esimerkki

Tekstin käsittely R-kielellä on jossakin määrin vaivalloista, joten kaikki tekstinkäsittely on tehty tätä varten kirjoitetuilla C-kielisillä ohjelmilla. Seuraavassa listattava R-ohjelma lukee tiedoston, joka sisältää referenssitekstin siirtymäfrekvenssit sekä salakirjoitetun tekstin siirtymäfrekvenssit. Oletetaan, että edelliset on tallennettu 29×29 -matriisiin `R` ja jälkimmäiset 29×29 -matriisiin C. Lisäksi matriisin `R` kaikkiin alkioihin lisätään luku 1 ja kukin rivi jaetaan rivisummallaan. Esitetään seuraavaksi log-tasapainojakauman laskemiseen käytettävä R-koodi.

```
1 log.pi <- function(sigma) {
2   S <- 0
3   for (i in 1:29) {
4     for (j in 1:29) {
5       S <- S + C[i,j] * log(R[sigma[i],sigma[j]])
6     }
7   }
8   return (S)
9 }
```

Funktio laskee riveillä 3 – 7 summan

$$S = \log \pi(\sigma) = \sum_{i=1}^{n-1} \log a_{\sigma(w_i),\sigma(w_{i+1})}.$$

Sen sijaan, että funktio lukisi merkit w_i ja w_{i+1} salakirjoitetusta tekstistä, on kaikki siirtymäfrekvenssit tallennettu valmiiksi matriisiin `C`, jolloin riittää kertoa kunkin siirtymän log-referenssitodennäköisyys siirtymän frekvenssillä salakirjoitetussa aineistossa.

Seuraavassa esitettään Markovin ketjun simulointiin tarkoitettu R-kielinen koodi.

```
set.seed(0)

```

```
3  N <- 2000
4  sigma <- matrix(0, nrow=N, ncol=29)
5  sigma[1,] <- 1:29
6
7  for (i in 2:N) {
8
9    ind <- sample(1:28, 2, replace=FALSE)
10   sigma.cand <- sigma[i-1,]
11   tmp <- sigma.cand[ind[1]]
12   sigma.cand[ind[1]] <- sigma.cand[ind[2]]
13   sigma.cand[ind[2]] <- tmp
14
15   alpha <- min(1, exp(log.pi(sigma.cand) -
16          log.pi(sigma[i-1,])))
17   u <- runif(1)
18
19   if (u < alpha) {
20     sigma[i,] <- sigma.cand
21   }
22   else {
23     sigma[i,] <- sigma[i-1,]
24   }
25
26 }
```

Riveillä 4 ja 5 muodostetaan $N \times 29$ -matriisi Markovin ketjun iteraatioiden tallentamista varten ja matriisin ensimmäiseen riviin tallennetaan Markovin ketjun alkuarvo $\sigma_0 = (1, 2, \dots, 29)$.

Rivillä 9 tuotetaan kaksi satunnaista ja erillistä lukua väliltä $1, \dots, 28$. Riveillä 10–13 muodostetaan kandidaattipermutaatio vaihtamalla edellisen iteraation permutaatiossa arvot päittäin edellä arvottujen kahden indeksin kohdalla.

Riveillä 15 ja 16 lasketaan hyväksymistodennäköisyys ja rivillä 17 tuotetaan otos välin $(0, 1)$ tasajakaumasta. Riveillä 19–24 kandidaatti joko hyväksytään tai hylätään.

Markovin ketjun valitut iteraatit eli matriisin `sigma` rivit tallennetaan tiedostoon ja C-kielinen ohjelma käyttää permutaatiota salakirjoitetun tekstin purkamiseen.

Kirjallisuutta

[AG07] Asmussen, Søren ja Glynn, Peter W.: *Stochastic Simulation: Algorithms and Analysis*, nide 57 sarjassa *Stochastic Modelling and Applied Probability.* Springer, 2007.

[BGJM11] Brooks, Steve, Gelman, Andrew, Jones, Galin L. ja Meng, Xiao Li: *Handbook of Markov Chain Monte Carlo.* Chapman & Hall/CRC, 2011.

[CG92] Casella, George ja George, Edward I.: *Explaining the Gibbs Sampler.* The American Statistician, 46(3):167–174, 1992.

[CG94] Chan, Kung Sik ja Geyer, Charler J.: *Discussion: Markov Chains for Exploring Posterior Distributions.* The Annals of Statistics, 22(4):1747–1758, 1994.

[Cin11] Cinlar, Erhan: *Probability and Stochastics.* Springer, 1. painos, 2011.

[CR00] Cappe, Olivier ja Robert, Christian P.: *Markov Chain Monte Carlo: 10 Years and Still Running!* Journal of the American Statistical Association, 95(452):1282–1286, 2000.

[Dia09] Diaconis, Persi: *The Markov Chain Monte Carlo Revolution.* Bulletin of the American Mathematical Society, 46(2):179–205, 2009.

[ES95] Evans, Michael ja Swartz, Tim: *Methods for Approximating Integrals in Statistics with Special Emphasis on Bayesian Integration Problems.* Statistical Science, 10(3):254–272, 1995.

[Fel50] Feller, William: *An Introduction to Probability Theory and its Applications.* Wiley Mathematical Statistics Series. John Wiley & Sons, Inc., 1950.

[GCS+14] Gelman, Andrew, Carlin, John B., Stern, Hal S., Dunson, David B., Vehtari, Aki ja Rubin, Donald B.: *Bayesian Data Analysis.* Chapman & Hall/CRC, 3. painos, 2014.

[Gey92] Geyer, Charles J.: *Practical Markov Chain Monte Carlo.* Statistical Science, 7(4):473–483, 1992.

[Gey11] Geyer, Charles J.: *Introduction to Markov Chain Monte Carlo.* Teoksessa *Handbook of Markov Chain Monte Carlo.* Chapman & Hall/CRC, 2011.

[GG84] Geman, Stuart ja Geman, Donald: *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 6(6):721–741, 1984.

[GHRPS90] Gelfand, Alan E., Hills, Susan E., Racine-Poon, Amy ja Smith, Adrian F. M.: *Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling.* Journal of the American Statistical Association, 85(412):972–985, 1990.

[GR96] Gilks, W. R. ja Roberts, G. O.: *Strategies for improving MCMC.* Teoksessa Gilks, W. R., Richardson, S. ja Spiegelhalter, D. J. (toimittajat): *Markov Chain Monte Carlo in Practice.* Chapman & Hall, 1996.

[Gre95] Green, Peter J.: *Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination.* Biometrika, 82(4):711–732, 1995.

[GRS96] Gilks, W. R., Richardson, S. ja Spiegelhalter, D. J.: *Markov Chain Monte Carlo in Practice.* Chapman & Hall, 1996.

[GS90] Gelfand, Alan E. ja Smith, Adrian F. M.: *Sampling-Based Approaches to Calculating Marginal Densities.* Journal of the American Statistical Association, 85(410):398–409, 1990.

[Has70] Hastings, W. K.: *Monte Carlo Sampling Methods Using Markov Chains and Their Applications.* Biometrika, 57(1):97–109, 1970.

[Hit03] Hitchcock, David B.: *A History of the Metropolis–Hastings Algorithm.* The American Statistician, 57(4):254–257, 2003.

[JH00] Jarner, Søren Fiig ja Hansen, Ernst: *Geometric ergodicity of Metropolis algorithms.* Stochastic Processes and their Applications, 85:341–361, 2000.

[JH01] Jones, Galin L. ja Hobert, James P.: *Honest Exploration of Intractable Probability Distributions via Markov Chain Monte Carlo.* Statistical Science, 16(4):312–334, 2001.

[KS07] Koralov, Leonid B. ja Sinai, Yakov G.: *Theory of Probability and Random Processes.* Springer, 2. painos, 2007.

[Lan05] Landau, D. P.: *An Introduction to Monte Carlo Methods in Statistical Physics.* Teoksessa Kendall, W. S., Liang, F. ja Wang, J. S (toimittajat): *Markov Chain Monte Carlo: Innovations and Applications.* World Scientific, 2005.

[Men90] Mendelson, Bert: *Introduction to Topology.* Dover, 3. painos, 1990.

[Met87] Metropolis, N.: *The beginning of the Monte Carlo method.* Los Alamos Science, (15):125–130, 1987.

[MO14] Mitov, Kosto V. ja Omey, Edward: *Renewal Processes.* Springer, 2014.

[MRR+53] Metropolis, Nicholas, Rosenbluth, Arianna W., Rosenbluth, Marshall N., Teller, Augusta H. ja Teller, Edward: *Equation of State Calculations by Fast Computing Machines.* The Journal of Chemical Physics, 21(6):1087–1092, 1953.

[MT93] Meyn, S. P. ja Tweedie, R. L.: *Markov Chains and Stochastic Stability.* Springer-Verlag, 1993.

[MT96] Mengersen, K. L. ja Tweedie, R. L.: *Rates of Convergence of the Hastings and Metropolis Algorithms.* The Annals of Statistics, 24(1):101–121, 1996.

[Num84] Nummelin, Esa: *General irreducible Markov chains and non-negative operators*, nide 83 sarjassa *Cambridge tracts in mathematics.* Cambridge University Press, 1984.

[Num02] Nummelin, Esa: *MC's for MCMC'ists.* International Statistical Review, 70(2):215–240, 2002.

[Pes73] Peskun, P. H.: *Optimum Monte-Carlo sampling using Markov chains.* Biometrika, 60(3):607–612, 1973.

[RC04] Robert, Christian P. ja Casella, George: *Monte Carlo Statistical Methods.* Springer Texts in Statistics. Springer, 2. painos, 2004.

[RC11] Robert, Christian ja Casella, George: *A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data.* Statistical Science, 26(1):102–115, 2011.

[Rev75] Revuz, D.: *Markov Chains*, nide 11 sarjassa *North-Holland Mathematical Library.* North-Holland Publishing Company, 1975.

[Rob04] Roberts, Gareth O.: *General state space Markov chains and MCMC algorithms.* Probability Surveys, 1:20–71, 2004.

[Ros11] Rosenthal, Jeffrey S.: *Optimal Proposal Distributions and Adaptive MCMC.* Teoksessa Brooks, Steve, Gelman, Andrew, Jones, Galin L. ja Meng, Xiao Li (toimittajat): *Handbook of Markov Chain Monte Carlo.* Chapman & Hall/CRC, 2011.

[RS15] Rajaratnam, Bala ja Sparks, Doug: *MCMC-Based Inference in the Era of Big Data: A Fundamental Analysis of the Convergence Complexity of High-Dimensional Chains.* `http://arxiv.org/pdf/1508.00947v2.pdf`, 2015.

[RT96] Roberts, G. O. ja Tweedie, R. L.: *Geometric Convergence and Central Limit Theorems for Multidimensional Hastings and Metropolis Algorithms.* Biometrika, 83(1):95–110, 1996.

[Tie94] Tierney, Luke: *Markov Chains for Exploring Posterior Distributions.* The Annals of Statistics, 22(4):1701–1728, 1994.

[Tie96] Tierney, Luke: *Introduction to general state-space Markov chain theory.* Teoksessa Gilks, W. R., Richardson, S. ja Spiegelhalter, D. J. (toimittajat): *Markov Chain Monte Carlo in Practice.* Chapman & Hall, 1996.