Markovin ketju Monte Carlo -simulointi ja Peskunin järjestys

Santeri Parkkinen

Matematiikan pro gradu -tutkielma

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
Kevät 2019

Tiivistelmä

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
Matematiikan pro gradu -tutkielma
Markovin ketju Monte Carlo -simulointi ja Peskunin järjestys, 42 s., liitteitä 19 s.
Santeri Parkkinen
Toukokuu 2019

Tämän tutkielman tavoitteena on esitellä Markovin ketju Monte Carlo -simulointi äärellisessä tila-avaruudessa ja käsitellä simulointialgoritmien vertailuun liittyvää ongelmaa. Markovin ketju Monte Carlo -simuloinnissa funktion odotusarvolle lasketaan approksimaatioita simuloitua Markovin ketjua apuna käyttäen. Usein simulointi voidaan tehdä useammalla kuin yhdellä algoritmilla ja halutaan selvittää, mikä algoritmi soveltuu tehtävään parhaiten.

Kun simulaatioaskelten määrä lähestyy ääretöntä, Markovin ketju Monte Carlo -estimaatti suppenee melkein varmasti kohti odotusarvon oikeaa arvoa käytetystä simulointialgoritmista riippumatta. Käytännössä Markovin ketju Monte Carlo -simulointi tuottaa kuitenkin vain odotusarvon approksimaation, koska mikään todellinen simulointi ei voi jatkua äärettömän pitkään. Vain äärellisen monen simulaatioaskeleen käyttö aiheuttaa virheen, joka on luonteeltaan satunnainen: virhe kuuluu tietylle välille jollakin todennäköisyydellä. On luonnollista kysyä, miten approksimaation tarkkuus riippuu simulointialgoritmista. Kaikki algoritmit antavat odotusarvolle arvion halutulla tarkkuudella, jos simulointia jatketaan riittävän kauan. Jotkut algoritmit tuottavat kuitenkin tarkkoja approksimaatioita nopeammin kuin toiset, mikä on motivaatio algoritmien vertailulle. Ei kuitenkaan ole aivan selvää, miten vertailu pitäisi käytännössä tehdä. Halutun tarkkuuden saavuttamiseksi vaadittavaa aikaa on vaikea arvioida, eivätkä pelkät arviot edes ole riittävä perusta algoritmien vertailulle. Voi esimerkiksi käydä niin, että arvioitu alaraja algoritmin P vaatimalle ajalle on pienempi kuin alaraja algoritmin Q vaatimalle ajalle, mutta silti algoritmi Q pääsee haluttuun tarkkuuteen nopeammin kuin algoritmi P. Pelkät alarajat eivät siis kerro tarpeeksi vaadittavien aikojen todellisista arvoista. Arvioiden sijaan tarvitaan jokin eksakti riippuvuus simulointivirheen ja algoritmin välille.

Kriteeri, jonka suhteen algoritmien vertailu tehdään, johdetaan Markovin ketjujen keskeistä raja-arvolausetta käyttäen. Keskeinen raja-arvolause mahdollistaa simulointivirheiden todennäköisyyksien raja-arvon eksaktin laskemisen tietyssä erikoistapauksessa. Tarkastelu osoittaa, että algoritmien vertailemiseksi kannattaa tutkia niiden asymptoottisia variansseja: kahdesta algoritmista se, jonka asymptoottinen varianssi annetulle funktiolle on pienempi, soveltuu paremmin kyseisen funktion odotusarvon simulointiin. Kriteerin ongelmana on, että sen soveltaminen ei ole ihan suoraviivaista. Käytännössä algoritmeista tiedetään vain niiden siirtymätodennäköisyysmatriisit, joten algoritmien vertailemiseksi täytyy selvittää, miten asymptoottinen varianssi riippuu siirtymätodennäköisyysmatriisista. Tässä tutkielmassa tarkastelu rajataan kääntyviin Markovin ketjuihin, jolloin kyseinen riippuvuus voidaan selvittää funktionaalianalyysin tuloksia hyödyntäen. Tämän jälkeen vertailu onnistuu tietyssä erikoistapauksessa Peskunin järjestystä käyttämällä. Peskunin järjestyksen sovelluksena osoitetaan, että Metropolis–Hastings -algoritmi on parempi kuin Barkerin algoritmi.

Sisältö

Johdanto

Stokastiikassa ja sen sovelluksissa usein vastaan tuleva ongelma on satunnaismuuttujan odotusarvon laskeminen. Aina odotusarvoa ei käytännössä ole mahdollista laskea suoraan satunnaismuuttujan jakaumaa käyttäen edes tapauksessa, jossa satunnaismuuttujan tila-avaruus tiedetään äärelliseksi. Ongelmia ilmenee, jos tila-avaruus on hyvin suuri tai jos jakauman normitusvakiota ei tunneta. Tällaisissa tilanteissa on kuitenkin usein mahdollista approksimoida odotusarvoa Markovin ketju Monte Carlo -simuloinniksi kutsuttua menetelmää käyttäen vieläpä niin, että simulointialgoritmi voidaan valita useasta eri vaihtoehdosta. Tämän tutkielman tavoitteena on esitellä Markovin ketju Monte Carlo -simulointi äärellisessä tila-avaruudessa sekä käsitellä algoritmien vertailuun liittyvää ongelmaa.

Kun simulaatioaskelten määrä lähestyy ääretöntä, Markovin ketju Monte Carlo -estimaatti suppenee melkein varmasti kohti odotusarvon oikeaa arvoa käytetystä simulointialgoritmista riippumatta. Käytännössä Markovin ketju Monte Carlo -simulointi tuottaa kuitenkin vain odotusarvon approksimaation, koska mikään todellinen simulointi ei voi jatkua äärettömän pitkään. Vain äärellisen monen simulaatioaskeleen käyttö aiheuttaa virheen, joka on luonteeltaan satunnainen: virhe kuuluu tietylle välille jollakin todennäköisyydellä. On luonnollista kysyä, miten approksimaation tarkkuus riippuu simulointialgoritmista. Kaikki algoritmit antavat odotusarvolle arvion halutulla tarkkuudella, jos simulointia jatketaan riittävän kauan. Jotkut algoritmit tuottavat kuitenkin tarkkoja approksimaatioita nopeammin kuin toiset, mikä on motivaatio algoritmien vertailulle. Ei kuitenkaan ole aivan selvää, miten vertailu pitäisi käytännössä tehdä. Halutun tarkkuuden saavuttamiseksi vaadittavaa aikaa on vaikea arvioida, eivätkä pelkät arviot edes ole riittävä perusta algoritmien vertailulle. Voi esimerkiksi käydä niin, että arvioitu alaraja algoritmin P vaatimalle ajalle on pienempi kuin alaraja algoritmin Q vaatimalle ajalle, mutta silti algoritmi Q pääsee haluttuun tarkkuuteen nopeammin kuin algoritmi P. Pelkät alarajat eivät siis kerro tarpeeksi vaadittavien aikojen todellisista arvoista. Arvioiden sijaan tarvitaan jokin eksakti riippuvuus simulointivirheen ja algoritmin välille.

Osoittautuu, että tietyssä erikoistapauksessa Markovin ketjujen keskeinen raja-arvolause mahdollistaa simulointivirheiden todennäköisyyksien raja-arvon eksaktin määrittämisen. Tarkastelun seurauksena nähdään, että algoritmien vertailemiseksi kannattaa tutkia niiden asymptoottisia variansseja: jos algoritmin P asymptoottinen varianssi funktiolle f on pienempi tai yhtä suuri kuin algoritmin Q asymptoottinen varianssi funktiolle f, niin algoritmi P on parempi funktion f odotusarvon simulointiin kuin algoritmi Q. Tällaisen vertailun tekeminen ei kuitenkaan ole aivan ongelmatonta. Lähtökohtaisesti simulointialgoritmeista tiedetään nimittäin vain niiden siirtymätodennäköisyysmatriisit, joten täytyy selvittää, miten algoritmin asymptoottinen varianssi funktiolle f riippuu siirtymätodennäköisyysmatriisista. Tässä tutkielmassa tarkastelu rajataan kääntyviin Markovin ketjuihin, jolloin kyseinen riippuvuus voidaan selvittää funktionaalianalyysin työkaluja apuna käyttäen. Päädytään lauseeseen, joka on sikäli ongelmallinen, että tyypillisessä Markovin ketju Monte Carlo -simuloinnin sovelluksessa käytössä olevat tiedot eivät riitä algoritmien vertailuun. Ongelman ratkaisi Peter Peskun, joka keksi, miten kahden algoritmin asymptoottisten varianssien suuruusjärjestys voidaan erikoistapauksessa päätellä suoraan siirtymätodennäköisyysmatriisien alkioita tutkimalla. Oletetaan, että algoritmeja P ja Q vastaavat siirtymätodennäköisyysmatriisit ovat kääntyviä saman invariantin jakauman suhteen. Peskun osoitti, että jos algoritmia P vastaavan matriisin kaikki alkiot – diagonaalialkioita lukuun ottamatta – ovat suurempia tai yhtä suuria kuin algoritmia Q vastaavan matriisin alkiot, niin algoritmin P asymptoottinen varianssi funktiolle f on pienempi tai yhtä suuri kuin algoritmin Q asymptoottinen varianssi funktiolle f. Johtopäätös on sama kaikille funktioille, joten tästä seuraa, että algoritmi P on parempi kuin algoritmi Q minkä tahansa funktion odotusarvon approksimointiin. Nykyisin matriisin alkioita koskevaa ehtoa kutsutaan Peskunin järjestykseksi. Lauseensa sovelluksena Peskun osoitti, että Metropolis–Hastings -algoritmi on parempi kuin Barkerin algoritmi.

Tutkielman esitietoina lukijalta edellytetään todennäköisyysteorian, lineaarialgebran ja funktionaalianalyysin perusteiden tunteminen. Markovin ketjujen tunteminen sen sijaan ei ole aivan välttämätöntä, sillä tarvittava määrä teoriaa sisältyy tutkielmaan. Kannattaa kuitenkin huomata, että Markovin ketjujen perusominaisuuksia käsitellään vain liitteessä A, ja varsinainen teksti keskittyy niihin tuloksiin, jotka tarvitaan Markovin ketju Monte Carlo -simuloinnin esittämiseen. Myös funktionaalianalyysistä on melko kattava liite.

Rakenteeltaan tutkielma on seuraavanlainen: Luvussa 1 määritellään siirtymätodennäköisyysmatriisi ja Markovin ketju äärellisessä tila-avaruudessa. Luvussa 2 käydään läpi Markovin ketju Monte Carlo -simuloinnin esittämiseen tarvittavia pohjatietoja. Keskeisiä käsitteitä ovat invariantti jakauma ja tasainen ergodisuus. Luvussa 3 perustellaan, miksi Markovin ketju Monte Carlo -simulointi toimii. Lisäksi määritellään Metropolis–Hastings -algoritmi ja Barkerin algoritmi ja osoitetaan, että ne ovat kääntyviä invariantin jakaumansa suhteen. Lopuksi pohjustetaan algoritmien vertailuun liittyvää ongelmaa. Luku 4 käsittelee siirtymätodennäköisyysmatriisien ominaisvektoreita ja -arvoja. Tämä tarkastelu on tarpeen, jotta luvussa 5 voidaan hyödyntää funktionaalianalyysin tuloksia asymptoottisen varianssin tutkimiseen. Luvussa 5 tehtävien tarkastelujen tuloksena saadaan lause, joka kertoo, miten asymptoottinen varianssi riippuu simulointialgoritmista. Luvussa 5 esitellään myös Peskunin järjestys ja selvitetään, miten se liittyy asymptoottisten varianssien ja algoritmien vertailuun. Lopuksi Peskunin järjestyksen sovelluksena osoitetaan, että Metropolis–Hastings -algoritmi on parempi kuin Barkerin algoritmi.

Merkintöjä

- $\mathbb{N}_0 = \{0, 1, 2, \dots\}$
- $\mathbb{N} = \{1, 2, 3, \dots\}$
- $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$
- Olkoon $P = [p_{i,j}]_{i,j=1,2,\dots,n}$ neliömatriisi sekä $k \in \mathbb{N}$. Potenssimatriisin P^k alkio paikassa (i, j) on $p_{i,j}^{(k)}$.
- Olkoon $P = [p_{i,j}]$ matriisi.
  - Jos $p_{i,j} \geq 0$ kaikilla i ja j, merkitään $P \geq 0$.
  - Jos $p_{i,j} > 0$ kaikilla i ja j, merkitään $P > 0$.
- Olkoon $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ vektori.
  - Jos $x_i \geq 0$ kaikilla $i = 1, 2, \dots, n$, merkitään $x \geq 0$.
  - Jos $x_i > 0$ kaikilla $i = 1, 2, \dots, n$, merkitään $x > 0$.
- Avaruuden $\mathbb{R}^n$ luonnollisia kantavektoreita merkitään tunnuksilla e_i, $i = 1, 2, \dots, n$. Vektorin e_i j. komponentti on siis

$$e_{i,j} = \begin{cases} 1, \text{ jos } j = i \\ 0, \text{ jos } j \neq i \end{cases} .$$

- Avaruuden $\mathbb{R}^n$ nollavektorille käytetään merkintää $\bar{0}$.
- Olkoon S joukko. Joukon S osajoukkojen kokoelmaa merkitään

$$\mathcal{P}(S) = \{A | A \subset S\}.$$

- Jos $x \in \mathbb{R}$, niin

$$\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \leq x\}.$$

1 Markovin ketju äärellisessä tila-avaruudessa

Tässä luvussa määritellään siirtymätodennäköisyysmatriisi ja Markovin ketju äärellisessä tila-avaruudessa. Karkeasti ottaen Markovin ketju on jono satunnaismuuttujia $X_k : \Omega \to S$, jolla on se ominaisuus, että jokaisella ajanhetkellä $k \in \mathbb{N}_0$ todennäköisyys siirtyä tilasta $X_k = x$ tilaan $X_{k+1} = y$ riippuu vain ketjun nykytilasta $X_k = x$, mutta ei siitä, miten nykytilaan on päädytty. Satunnaismuuttujat $X_0, X_1, \ldots, X_{k-1}$ eivät siis vaikuta siirtymän

$$X_k = x \to X_{k+1} = y$$

todennäköisyyteen. Siirtymätodennäköisyysmatriisi puolestaan kirjaa kyseisten siirtymien todennäköisyydet matriisin muotoon. Tämä on mahdollista, kun ketjun tila-avaruus S on äärellinen, ja erittäin hyödyllistä, koska siirtymätodennäköisyysmatriisin avulla Markovin ketjujen tutkiminen palautuu pitkälti matriisien tutkimiseen. Seuraavat määritelmät ovat lähteestä [1].

Määritelmä 1.1. *Neliömatriisi $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ on siirtymätodennäköisyysmatriisi, jos*

(i) $P \geq 0$

(ii) $\sum_{j=1}^{n} p_{i,j} = 1$ *kaikilla* $i = 1, 2, \ldots, n$.

Jos $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ on siirtymätodennäköisyysmatriisi, niin induktiolla on helppo todeta, että kaikilla $k \in \mathbb{N}$ myös potenssimatriisi $P^k = \left[p_{i,j}^{(k)}\right]_{i,j=1,2,\ldots,n}$ on siirtymätodennäköisyysmatriisi.

Määritelmä 1.2. *Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ todennäköisyysavaruus, $S = \{s_1, s_2, \ldots, s_n\} \subset \mathbb{R}$ ja $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ siirtymätodennäköisyysmatriisi. Jono $(X_k)_{k=0}^{\infty}$ satunnaismuuttujia*

$$X_k : \Omega \to S$$

on Markovin ketju, jonka siirtymätodennäköisyysmatriisi on P, jos

(i) $\mathbb{P}(X_{k+1} = x_{k+1} | X_k = x_k, X_{k-1} = x_{k-1}, \ldots, X_0 = x_0) = \mathbb{P}(X_{k+1} = x_{k+1} | X_k = x_k)$ *kaikilla $k \in \mathbb{N}_0$ ja kaikilla $x_0, x_1, \ldots, x_{k+1} \in S$, joille*

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_k = x_k) > 0$$

(ii) $\mathbb{P}(X_{k+1} = s_j | X_k = s_i) = p_{i,j}$ *kaikilla $i, j = 1, 2, \ldots, n$ ja $k \in \mathbb{N}_0$, joille $\mathbb{P}(X_k = s_i) > 0$.*

Satunnaismuuttujan X_0 jakaumaa kutsutaan Markovin ketjun $(X_k)_{k=0}^{\infty}$ alkujakaumaksi ja joukkoa S kutsutaan ketjun tila-avaruudeksi.

Jatkossa oletetaan, että Markovin ketjun tila-avaruus on reaalilukujen osajoukko, jossa on aina vähintään kaksi alkiota. Erityisesti siirtymätodennäköisyysmatriisi on aina vähintään kokoa 2×2.

Olkoon $(X_k)_{k=0}^{\infty}$ Markovin ketju tila-avaruudella $S = \{s_1, s_2, \ldots, s_n\}$ ja olkoon ketjun siirtymätodennäköisyysmatriisi $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$. Samaistetaan satunnaismuuttujan X_k jakauma

$$\mu_k : S \to [0,1],\ \mu_k(x) = \mathbb{P}(X_k = x),$$

vektorin

$$\mu^{(k)} = \left(\mu_1^{(k)}, \mu_2^{(k)}, \ldots, \mu_n^{(k)}\right) \in \mathbb{R}^n$$

kanssa, missä

$$\mu_i^{(k)} = \mu_k(s_i) = \mathbb{P}\left(X_k = s_i\right) \text{ kaikilla } i = 1, 2, \ldots, n.$$

Olkoon J niiden indeksien $j = 1, 2, \ldots, n$ joukko, joille pätee, että

$$\mathbb{P}\left(X_k = s_j\right) > 0.$$

Tällöin kaikilla $l = 1, 2, \ldots, n$

$$\begin{aligned}
\mu_l^{(k+1)} &= \mathbb{P}\left(X_{k+1} = s_l\right) = \mathbb{P}\left(\{X_{k+1} = s_l\} \cap \left(\bigcup_{j=1}^{n} \{X_k = s_j\}\right)\right) \\
&= \mathbb{P}\left(\bigcup_{j=1}^{n} \left(\{X_{k+1} = s_l\} \cap \{X_k = s_j\}\right)\right) = \sum_{j=1}^{n} \mathbb{P}\left(X_{k+1} = s_l, X_k = s_j\right) \\
&= \sum_{j \in J} \mathbb{P}\left(X_{k+1} = s_l, X_k = s_j\right) = \sum_{j \in J} \mathbb{P}\left(X_k = s_j\right) \mathbb{P}\left(X_{k+1} = s_l | X_k = s_j\right) \\
&= \sum_{j \in J} \mathbb{P}\left(X_k = s_j\right) p_{j,l} = \sum_{j=1}^{n} \mathbb{P}\left(X_k = s_j\right) p_{j,l} = \sum_{j=1}^{n} \mu_j^{(k)} p_{j,l} = \left[\mu^{(k)} P\right]_l. \qquad (1)
\end{aligned}$$

Näin ollen vektorimuodossa pätee, että

$$\mu^{(k+1)} = \mu^k P. \qquad (2)$$

Yhtälöstä (2) seuraa helposti induktiolla, että

$$\mu^{(k)} = \mu^{(0)} P^k \qquad (3)$$

kaikilla $k \in \mathbb{N}$.

2 Invariantti jakauma ja tasainen ergodisuus

Tässä luvussa käydään läpi Markovin ketju Monte Carlo -simuloinnin esittämiseen tarvittavia lauseita. Luvun päätuloksia ovat Lause 2.1 ja Lause 2.2. Tulosten motivoimiseksi tarkastellaan lyhyesti seuraavaa ongelmaa lähteen [1] esitystä mukaillen: Olkoon $(X_k)_{k=0}^{\infty}$ Markovin ketju ja $\mu^{(k)}$ satunnaismuuttujan X_k jakauma. Halutaan selvittää, millä ehdoilla on olemassa jakauma π siten, että

$$\mu^{(k)} \to \pi, \text{ kun } k \to \infty.$$

Ongelman ratkaiseminen lähtee liikkeelle rajajakaumakandidaatin etsimisestä. Tämä onnistuu helposti: yhtälön (2) mukaan

$$\mu^{(k+1)} = \mu^{(k)} P$$

kaikilla $k \in \mathbb{N}_0$, missä P on ketjun siirtymätodennäköisyysmatriisi. Näin ollen nähdään, että jos rajajakauma $\pi = \lim_{k\to\infty} \mu^{(k)}$ on olemassa, se toteuttaa yhtälön

$$\pi = \lim_{k\to\infty} \mu^{(k+1)} = \lim_{k\to\infty} \left(\mu^{(k)} P\right) = \left(\lim_{k\to\infty} \mu^{(k)}\right) P = \pi P.$$

Tämä johtaa invariantin jakauman käsitteeseen:

Määritelmä 2.1. *Jakauma $\pi \in \mathbb{R}^n$ on siirtymätodennäköisyysmatriisin $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ invariantti jakauma, jos*

$$\pi = \pi P.$$

Seuraavaksi osoitetaan, että jos Markovin ketjun siirtymätodennäköisyysmatriisille P pätee, että $P^m > 0$ jollakin $m \in \mathbb{N}$, niin yksikäsitteinen invariantti jakauma π on olemassa ja sille pätee, että $\pi > 0$. Erotetaan osa todistuksesta lemmaksi myöhempää käyttöä varten. Lemman 2.1 tulos on lähteestä [2] ja todistus mukailee lähteen todistusta.

Lemma 2.1. *Olkoon $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ siirtymätodennäköisyysmatriisi siten, että $P > 0$. Jos $x \in \mathbb{R}^n$ siten, että $xP = x$, niin $x \geq 0$ tai $x \leq 0$.*

Todistus. Jos väite on väärä, on olemassa indeksit $i, j \in \{1, 2, \ldots, n\}$ siten, että $x_i < 0$ ja $x_j > 0$. Kaikilla $k = 1, 2 \ldots, n$ $x_k \leq |x_k|$ ja $x_i < 0 < |x_i|$, joten kaikilla $l = 1, 2, \ldots, n$ pätee, että

$$\sum_{k=1}^{n} x_k p_{k,l} = \sum_{k\neq i} x_k p_{k,l} + x_i p_{i,l} < \sum_{k\neq i} |x_k|\, p_{k,l} + |x_i|\, p_{i,l} = \sum_{k=1}^{n} |x_k|\, p_{k,l}.$$

Vastaavasti $x_k \geq -|x_k|$ kaikilla $k = 1, 2, \ldots, n$ ja $x_j > 0 > -|x_j|$, joten

$$\sum_{k=1}^{n} x_k p_{k,l} > -\sum_{k=1}^{n} |x_k|\, p_{k,l} \text{ kaikilla } l = 1, 2, \ldots, n.$$

Näin ollen saadaan, että

$$\left|\sum_{k=1}^{n} x_k p_{k,l}\right| < \sum_{k=1}^{n} |x_k|\, p_{k,l} \text{ kaikilla } l = 1, 2, \ldots, n.$$

Toisaalta oletuksen $xP = x$ nojalla $x_l = [xP]_l = \sum_{k=1}^{n} x_k p_{k,l}$ kaikilla $l = 1, 2, \ldots, n$, joten

$$\sum_{l=1}^{n} |x_l| = \sum_{l=1}^{n} \left|\sum_{k=1}^{n} x_k p_{k,l}\right| < \sum_{l=1}^{n} \sum_{k=1}^{n} |x_k|\, p_{k,l} = \sum_{k=1}^{n} \sum_{l=1}^{n} |x_k|\, p_{k,l} = \sum_{k=1}^{n} |x_k| \sum_{l=1}^{n} p_{k,l} = \sum_{k=1}^{n} |x_k|\,,$$

mikä on ristiriita. $\square$

Lause 2.1. *Olkoon $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ siirtymätodennäköisyysmatriisi siten, että $P^m > 0$ jollakin $m \in \mathbb{N}$. Tällöin matriisilla P on yksikäsitteinen invariantti jakauma $\pi \in \mathbb{R}^n$. Lisäksi pätee, että $\pi > 0$.*

Todistus. Olkoon $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ siten, että $y_k = 1$ kaikilla $k = 1, 2, \ldots, n$. Tällöin

$$[Py]_i = \sum_{k=1}^{n} p_{i,k} y_k = \sum_{k=1}^{n} p_{i,k} = 1 = y_i \text{ kaikilla } i = 1, 2, \ldots, n,$$

joten $Py = y$. Erityisesti luku 1 on matriisin P oikeanpuoleinen ominaisarvo. Koska matriisin vasemmanpuoleiset ja oikeanpuoleiset ominaisarvot ovat samat, on 1 myös matriisin P vasemmanpuoleinen ominaisarvo. On siis olemassa vektori $x \in \mathbb{R}^n \backslash \{\bar{0}\}$ siten, että $xP = x$. Tästä seuraa induktiolla, että $xP^k = x$ kaikilla $k \in \mathbb{N}$. Erityisesti $xP^m = x$, joten soveltamalla Lemmaa 2.1 siirtymätodennäköisyysmatriisiin $P^m > 0$ saadaan, että $x \leq 0$ tai $x \geq 0$. Korvaamalla vektori x tarvittaessa vektorilla $-x$ voidaan olettaa, että $x \geq 0$. Koska x ei ole nollavektori, on olemassa indeksi $i \in \{1, 2, \ldots, n\}$ siten, että $x_i > 0$. Tästä seuraa, että $x_j > 0$ kaikilla $j = 1, 2, \ldots, n$, sillä

$$x_j = [xP^m]_j = \sum_{k=1}^{n} x_k p_{k,j}^{(m)} \geq x_i p_{i,j}^{(m)} > 0.$$

Olkoon

$$\pi = \frac{x}{\sum_{j=1}^{n} x_j},$$

jolloin siis $\pi = (\pi_1, \pi_2, \ldots, \pi_n) \in \mathbb{R}^n$ siten, että

$$\pi_i = \frac{x_i}{\sum_{j=1}^{n} x_j} \text{ kaikilla } i = 1, 2, \ldots, n.$$

Tällöin

(i) $\pi P = \left(\frac{x}{\sum_{j=1}^{n} x_j} \right) P = \frac{1}{\sum_{j=1}^{n} x_j} (xP) = \frac{x}{\sum_{j=1}^{n} x_j} = \pi$,

(ii) $\sum_{i=1}^{n} \pi_i = \sum_{i=1}^{n} \frac{x_i}{\sum_{j=1}^{n} x_j} = \frac{1}{\sum_{j=1}^{n} x_j} \sum_{i=1}^{n} x_i = 1$ ja

(iii) $\pi_i = \frac{x_i}{\sum_{j=1}^{n} x_j} > 0$ kaikilla $i = 1, 2, \ldots, n$, koska $x_i > 0$ kaikilla $i = 1, 2, \ldots, n$,

joten $\pi \in \mathbb{R}^n$ on etsitty jakauma.

Osoitetaan vielä invariantin jakauman yksikäsitteisyys. Oletetaan vastoin väitettä, että on olemassa invariantti jakauma $\mu \in \mathbb{R}^n$ siten, että $\mu \neq \pi$. Tällöin

$$(\pi - \mu) P^m = \pi P^m - \mu P^m = \pi - \mu,$$

joten Lemman 2.1 nojalla $\pi - \mu \geq 0$ tai $\pi - \mu \leq 0$. Oletetaan, että $\pi - \mu \geq 0$. Koska $\mu \neq \pi$, on olemassa indeksi $j \in \{1, 2, \ldots, n\}$ siten, että $\mu_j \neq \pi_j$. Tällöin ehdon $\pi - \mu \geq 0$ nojalla on $\mu_j < \pi_j$, joten saadaan, että

$$1 = \sum_{i=1}^{n} \mu_i < \sum_{i=1}^{n} \pi_i = 1,$$

mikä on ristiriita. Tapaus $\pi - \mu \leq 0$ käsitellään vastaavasti. $\square$

Lause 2.1 ratkaisee invariantin jakauman olemassaoloa koskevan kysymyksen. Voidaan myös osoittaa, että

$$\lim_{k\to\infty} \mu^{(k)} = \pi,$$

jos $P^m > 0$ jollakin $m \in \mathbb{N}$. Tulos ei kuitenkaan ole tämän tutkielman kannalta kovin oleellinen; ongelmaa tarkasteltiin lähinnä invariantin jakauman käsitteen motivoimiseksi. Lisäksi nyt yhtälön (3) perusteella on luonnollista kysyä, mitä tapahtuu matriiseille P^k, missä $k \in \mathbb{N}$, kun $k \to \infty$. Osoittautuu, että kaikilla $i, j = 1, 2, \ldots, n$

$$p_{i,j}^{(k)} \to \pi_j, \text{ kun } k \to \infty,$$

vieläpä niin, että

$$\left| p_{i,j}^{(k)} - \pi_j \right| \leq C\beta^k$$

kaikilla $k \in \mathbb{N}$, missä $C \geq 0$ ja $0 \leq \beta < 1$. Huomaa, että yläraja ei riipu indekseistä i ja j, mikä on oleellista seuraavassa luvussa. Tulosta kutsutaan tasaiseksi ergodisuudeksi ja se todistetaan Lauseessa 2.2. Tarvitaan kuitenkin useita lemmoja, jotta luvun

$$\left| p_{i,j}^{(k)} - \pi_j \right|$$

yläraja voidaan kirjoittaa haluttuun muotoon. Lemma 2.2 todistuksineen on lähteestä [3]. Myös joissain luvun loppuosan tuloksissa on epäsuorasti hyödynnetty lähdettä [3]

Lemma 2.2. *Olkoon $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ siirtymätodennäköisyysmatriisi. Tällöin kaikilla $x \in \mathbb{R}^n$*

$$\max_i [Px]_i - \min_i [Px]_i \leq \left(1 - 2\min_{i,j} p_{i,j}\right)\left(\max_i x_i - \min_i x_i\right).$$

Todistus. Olkoon $x \in \mathbb{R}^n$ ja $j_0 \in \{1, 2, \ldots, n\}$ siten, että $x_{j_0} = \min_j x_j$. Tällöin kaikilla $i = 1, 2, \ldots, n$

$$\begin{aligned}
[Px]_i &= \sum_{j=1}^{n} p_{i,j}x_j = p_{i,j_0}x_{j_0} + \sum_{j\neq j_0} p_{i,j}x_j \leq p_{i,j_0}x_{j_0} + \max_j x_j \sum_{j\neq j_0} p_{i,j} \\
&= p_{i,j_0}\min_j x_j + \max_j x_j\,(1 - p_{i,j_0}) = \max_j x_j - p_{i,j_0}\left(\max_j x_j - \min_j x_j\right) \\
&\leq \max_j x_j - \min_{i,j} p_{i,j}\left(\max_j x_j - \min_j x_j\right).
\end{aligned}$$

Erityisesti

$$\max_i [Px]_i \leq \max_j x_j - \min_{i,j} p_{i,j}\left(\max_j x_j - \min_j x_j\right). \tag{4}$$

Vastaavasti arvioimalla nähdään, että kaikilla $i = 1, 2, \ldots, n$

$$\min_i [Px]_i \geq \min_j x_j + \min_{i,j} p_{i,j}\left(\max_j x_j - \min_j x_j\right). \tag{5}$$

Väite nähdään todeksi yhdistämällä arviot (4) ja (5):

$$\begin{aligned}\max_i [Px]_i - \min_i [Px]_i &\leq \left(\max_j x_j - \min_{i,j} p_{i,j}\left(\max_j x_j - \min_j x_j\right)\right)\\ &\quad - \left(\min_j x_j + \min_{i,j} p_{i,j}\left(\max_j x_j - \min_j x_j\right)\right)\\ &= \left(\max_j x_j - \min_j x_j\right) - 2\min_{i,j}\left(\max_j x_j - \min_j x_j\right)\\ &= \left(1 - 2\min_{i,j} p_{i,j}\right)\left(\max_j x_j - \min_j x_j\right). \end{aligned}$$

□

Seuraus 2.1. *Olkoon $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ siirtymätodennäköisyysmatriisi. Tällöin kaikilla $x \in \mathbb{R}^n$ ja $k \in \mathbb{N}$*

$$\max_i \left[P^k x\right]_i - \min_i \left[P^k x\right]_i \leq \left(1 - 2\min_{i,j} p_{i,j}\right)^k \left(\max_i x_i - \min_i x_i\right).$$

Todistus. Lemman 2.2 nojalla väite pätee, kun $k = 1$. Oletetaan induktio-oletuksena, että

$$\max_i \left[P^k x\right]_i - \min_i \left[P^k x\right]_i \leq \left(1 - 2\min_{i,j} p_{i,j}\right)^k \left(\max_i x_i - \min_i x_i\right)$$

jollakin $k \in \mathbb{N}$. Lemman 2.2 ja induktio-oletuksen nojalla saadaan, että

$$\begin{aligned}\max_i \left[P^{k+1} x\right]_i - \min_i \left[P^{k+1} x\right]_i &= \max_i \left[P(P^k x)\right]_i - \min_i \left[P(P^k x)\right]_i\\ &\leq \left(1 - 2\min_{i,j} p_{i,j}\right)\left(\max_i \left[P^k x\right]_i - \min_i \left[P^k x\right]_i\right)\\ &\leq \left(1 - 2\min_{i,j} p_{i,j}\right)\left(1 - 2\min_{i,j} p_{i,j}\right)^k \left(\max_i x_i - \min_i x_i\right)\\ &\leq \left(1 - 2\min_{i,j} p_{i,j}\right)^{k+1} \left(\max_i x_i - \min_i x_i\right). \end{aligned}$$

□

Lemma 2.3. *Olkoon $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ siirtymätodennäköisyysmatriisi siten, että $P^m > 0$ jollakin $m \in \mathbb{N}$. Merkitään*

$$\alpha = 1 - 2\min_{i,j} p_{i,j}^{(m)}.$$

Tällöin

(i) $\alpha \in [0, 1[$ *ja*

(ii) on olemassa luvut $C \geq 0$ ja $0 \leq \beta < 1$ siten, että

$$\alpha^{\left\lfloor \frac{k}{m} \right\rfloor} \leq C\beta^k$$

kaikilla $k \in \mathbb{N}$.

Todistus. Koska $P^m > 0$, niin selvästi $\alpha < 1$. Väite $\alpha \geq 0$ seuraa osoittamalla, että

$$\min_{i,j} p_{i,j}^{(m)} \leq \frac{1}{2}.$$

Tehdään antiteesi: $\min_{i,j} p_{i,j}^{(m)} > \frac{1}{2}$. Olkoot $i_0, j_0 \in \{1, 2, \ldots, n\}$ siten, että

$$p_{i_0,j_0}^{(m)} = \min_{i,j} p_{i,j}^{(m)}.$$

Tällöin

$$1 = \sum_{k=1}^{n} p_{i_0,k}^{(m)} \geq \min_k p_{i_0,k}^{(m)} + \max_k p_{i_0,k}^{(m)} \geq 2 \min_k p_{i_0,k}^{(m)} \geq 2 \min_{i,j} p_{i,j}^{(m)} > 1,$$

mikä on ristiriita. Siispä $\alpha \in [0, 1[$.

Kaikilla $k \in \mathbb{N}$

$$k = \left\lfloor \frac{k}{m} \right\rfloor m + \underbrace{\left(k - \left\lfloor \frac{k}{m} \right\rfloor m \right)}_{\leq m-1} \leq \left\lfloor \frac{k}{m} \right\rfloor m + (m-1),$$

joten

$$\frac{k}{m} \leq \left\lfloor \frac{k}{m} \right\rfloor + \frac{m-1}{m}$$

kaikilla $k \in \mathbb{N}$. Edelleen, koska $\alpha \in [0, 1[$, niin kaikilla $k \in \mathbb{N}$

$$\left(\alpha^{\frac{1}{m}} \right)^k = \alpha^{\frac{k}{m}} \geq \alpha^{\lfloor \frac{k}{m} \rfloor + \frac{m-1}{m}} = \alpha^{\lfloor \frac{k}{m} \rfloor} \alpha^{\frac{m-1}{m}},$$

mistä

$$\alpha^{\lfloor \frac{k}{m} \rfloor} \leq \frac{1}{\alpha^{\frac{m-1}{m}}} \left(\alpha^{\frac{1}{m}} \right)^k$$

kaikilla $k \in \mathbb{N}$. Voidaan siis valita $C = \frac{1}{\alpha^{\frac{m-1}{m}}}$ ja $\beta = \alpha^{\frac{1}{m}}$. □

Lemma 2.4. *Olkoon $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ siirtymätodennäköisyysmatriisi siten, että $P^m > 0$ jollakin $m \in \mathbb{N}$. Tällöin on olemassa luvut $C \geq 0$ ja $0 \leq \beta < 1$ siten, että*

$$\max_i p_{i,j}^{(k)} - \min_i p_{i,j}^{(k)} \leq C\beta^k$$

kaikilla $j = 1, 2, \ldots, n$ ja kaikilla $k \in \mathbb{N}$.

Todistus. Merkitään

$$\alpha = 1 - 2 \min_{i,j} p_{i,j}^{(m)}.$$

Lemman 2.3 nojalla on olemassa luvut $C \geq 0$ ja $0 \leq \beta < 1$ siten, että

$$\alpha^{\lfloor \frac{k}{m} \rfloor} \leq C\beta^k$$

kaikilla $k \in \mathbb{N}$. Kiinnitetään $j \in \{1, 2, \ldots, n\}$ ja $k \in \mathbb{N}$. Koska kaikilla $i = 1, 2, \ldots, n$

$$\left[P^k e_j \right]_i = \sum_{l=1}^{n} p_{i,l}^{(k)} e_{j,l} = p_{i,j}^{(k)},$$

niin

$$\min_i p_{i,j}^{(k)} = \min_i \left[P^k e_j \right]_i \text{ ja } \max_i p_{i,j}^{(k)} = \max_i \left[P^k e_j \right]_i.$$

Siispä Lausetta A.3 käyttämällä saadaan, että

$$
\begin{aligned}
\max_i p_{i,j}^{(k)} - \min_i p_{i,j}^{(k)} &= \max_i \left[P^k e_j\right]_i - \min_i \left[P^k e_j\right]_i \\
&= \max_i \left[P^{\lfloor \frac{k}{m} \rfloor m + \left(k - \lfloor \frac{k}{m} \rfloor m\right)} e_j\right]_i - \min_i \left[P^{\lfloor \frac{k}{m} \rfloor m + \left(k - \lfloor \frac{k}{m} \rfloor m\right)} e_j\right]_i \\
&= \max_i \left[P^{\lfloor \frac{k}{m} \rfloor m} P^{k - \lfloor \frac{k}{m} \rfloor m} e_j\right]_i - \min_i \left[P^{\lfloor \frac{k}{m} \rfloor m} P^{k - \lfloor \frac{k}{m} \rfloor m} e_j\right]_i \\
&= \max_i \left[(P^m)^{\lfloor \frac{k}{m} \rfloor} \left(P^{k - \lfloor \frac{k}{m} \rfloor m} e_j\right)\right]_i - \min_i \left[(P^m)^{\lfloor \frac{k}{m} \rfloor} \left(P^{k - \lfloor \frac{k}{m} \rfloor m} e_j\right)\right]_i \\
&\leq \left(1 - 2 \min_{i,j} p_{i,j}^{(m)}\right)^{\lfloor \frac{k}{m} \rfloor} \left(\max_i \left[P^{k - \lfloor \frac{k}{m} \rfloor m} e_j\right]_i - \min_i \left[P^{k - \lfloor \frac{k}{m} \rfloor m} e_j\right]_i\right) \\
&= \left(1 - 2 \min_{i,j} p_{i,j}^{(m)}\right)^{\lfloor \frac{k}{m} \rfloor} \underbrace{\left(\max_i p_{i,j}^{\left(k - \lfloor \frac{k}{m} \rfloor m\right)} - \min_i p_{i,j}^{\left(k - \lfloor \frac{k}{m} \rfloor m\right)}\right)}_{\leq 1} \\
&\leq \left(1 - 2 \min_{i,j} p_{i,j}^{(m)}\right)^{\lfloor \frac{k}{m} \rfloor} \\
&= \alpha^{\lfloor \frac{k}{m} \rfloor} \\
&\leq C\beta^k.
\end{aligned}
$$

□

Lause 2.2. *Olkoon $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ siirtymätodennäköisyysmatriisi siten, että $P^m > 0$ jollakin $m \in \mathbb{N}$, ja olkoon $\pi \in \mathbb{R}^n$ matriisin P invariantti jakauma. Tällöin on olemassa luvut $C \geq 0$ ja $0 \leq \beta < 1$ siten, että*

$$\left| p_{i,j}^{(k)} - \pi_j \right| \leq C\beta^k$$

kaikilla $i, j = 1, 2, \ldots, n$ ja $k \in \mathbb{N}$.

Todistus. Koska π on siirtymätodennäköisyysmatriisin P invariantti jakauma, niin

$$\pi = \pi P.$$

Tästä seuraa induktiolla, että

$$\pi = \pi P^k$$

kaikilla $k \in \mathbb{N}$. Näin ollen kaikilla $j = 1, 2, \ldots, n$ ja $k \in \mathbb{N}$ pätee, että

$$\pi_j = \left[\pi P^k\right]_j = \sum_{i=1}^n \pi_i p_{i,j}^{(k)} \leq \max_i p_{i,j}^{(k)} \sum_{i=1}^n \pi_i = \max_i p_{i,j}^{(k)}$$

ja

$$\pi_j = \left[\pi P^k\right]_j = \sum_{i=1}^n \pi_i p_{i,j}^{(k)} \geq \min_i p_{i,j}^{(k)} \sum_{i=1}^n \pi_i = \min_i p_{i,j}^{(k)}.$$

Lemman 2.4 nojalla on olemassa luvut $C \geq 0$ ja $0 \leq \beta < 1$ siten, että

$$\max_i p_{i,j}^{(k)} - \min_i p_{i,j}^{(k)} \leq C\beta^k$$

kaikilla $j = 1, 2, \ldots, n$ ja $k \in \mathbb{N}$. Olkoot $i, j = 1, 2, \ldots, n$ ja $k \in \mathbb{N}$ mitä tahansa. Koska

$$\min_i p_{i,j}^{(k)} \leq \pi_j \leq \max_i p_{i,j}^{(k)} \leq \min_i p_{i,j}^{(k)} + C\beta^k$$

ja

$$\min_i p_{i,j}^{(k)} \leq p_{i,j}^{(k)} \leq \max_i p_{i,j}^{(k)} \leq \min_i p_{i,j}^{(k)} + C\beta^k,$$

niin $\pi_j, p_{i,j}^{(k)} \in \left[\min_i p_{i,j}^{(k)}, \min_i p_{i,j}^{(k)} + C\beta^k\right]$. Siispä

$$\left|p_{i,j}^{(k)} - \pi_j\right| \leq C\beta^k. \qquad \square$$

3 Markovin ketju Monte Carlo -simulointi

Olkoon $\pi \in \mathbb{R}^n$ jakauma joukolla $S = \{s_1, s_2, \dots, s_n\}$, $f : S \to \mathbb{R}$ funktio ja $X : \Omega \to S$ satunnaismuuttuja, joka noudattaa jakaumaa π. Halutaan selvittää odotusarvo

$$\mathbb{E}_\pi f := \mathbb{E}_\pi f(X) = \sum_{i=1}^{n} f(s_i)\pi_i,$$

missä $f(X)$ on kuvausten $f : S \to \mathbb{R}$ ja $X : \Omega \to S$ yhdistetty kuvaus

$$f(X) : \Omega \to \mathbb{R},\ f(X)(\omega) = f(X(\omega)).$$

Jos tieto jakaumasta π on puutteellista tai n on hyvin suuri, niin odotusarvoa ei voi selvittää suoraan laskemalla. Tällöin on kuitenkin usein mahdollista konstruoida Markovin ketju $(X_k)_{k=0}^{\infty}$, jonka siirtymätodennäköisyysmatriisin P invariantti jakauma on π. Jos $P^m > 0$ jollakin $m \in \mathbb{N}$, voi odotusarvoa $\mathbb{E}_\pi f$ approksimoida satunnaisluvulla

$$\frac{1}{k}\sum_{i=1}^{k} f\left(X_i\right),$$

kun $k \in \mathbb{N}$ on riittävän suuri. Tätä kutsutaan Markovin ketju Monte Carlo -simuloinniksi ja sen käyttö perustellaan Lauseessa 3.1.

Otetaan käyttöön seuraavat merkinnät: Olkoon $(X_k)_{k=0}^{\infty}$ Markovin ketju äärellisellä tila-avaruudella $S = \{s_1, s_2, \dots, s_n\}$ ja olkoon ketjun taustalla oleva todennäköisyysavaruus $(\Omega, \mathcal{F}, \mathbb{P})$. Jos $i \in \{1, 2, \dots, n\}$ siten, että $\mathbb{P}\left(X_0 = s_i\right) > 0$ ja $A \in \mathcal{F}$, merkitään

$$\mathbb{P}_i\left(A\right) = \mathbb{P}\left(A | X_0 = s_i\right).$$

Lisäksi, jos $g : \Omega \to \mathbb{R}$ on funktio, merkitään

$$\mathbb{E}_i g = \mathbb{E}\left[g | X_0 = s_i\right].$$

Lemma 3.1. *Olkoon $(X_k)_{k=0}^{\infty}$ Markovin ketju tila-avaruudella $S = \{s_1, s_2, \dots, s_n\}$ ja $f : S \to \mathbb{R}$ funktio. Oletetaan, että ketjun siirtymätodennäköisyysmatriisille $P = [p_{i,j}]_{i,j=1,2,\dots,n}$ pätee $P^m > 0$ jollakin $m \in \mathbb{N}$ ja että $\pi \in \mathbb{R}^n$ on matriisin P invariantti jakauma. Tällöin, jos $\mathbb{P}\left(X_0 = s_i\right) > 0$, niin*

$$\lim_{k\to\infty} \frac{1}{k^2}\sum_{j=1}^{k} \mathbb{E}_i\left(f(X_j) - \mathbb{E}_\pi f\right)^2 = 0.$$

Todistus. Kaikilla $j \in \mathbb{N}$

$$\begin{aligned}\mathbb{E}_i\left(f(X_j) - \mathbb{E}_\pi f\right)^2 &= \sum_{l=1}^{n}\left(f(s_l) - \mathbb{E}_\pi f\right)^2 \mathbb{P}_i\left(X_j = s_l\right)\\ &\le \max_{m=1,2,\dots,n}\left(f(s_m) - \mathbb{E}_\pi f\right)^2 \sum_{l=1}^{n} P_i\left(X_j = s_l\right) = \max_{m=1,2,\dots,n}\left(f(s_m) - \mathbb{E}_\pi f\right)^2,\end{aligned}$$

joten

$$\begin{aligned}
0 &\leq \frac{1}{k^2}\sum_{j=1}^{k}\mathbb{E}_i\left(f(X_j)-\mathbb{E}_\pi f\right)^2 \leq \frac{1}{k^2}\sum_{j=1}^{k}\max_{m=1,2,\ldots,n}\left(f(s_m)-\mathbb{E}_\pi f\right)^2 \\
&= \frac{1}{k^2}k\cdot\max_{m=1,2,\ldots,n}\left(f(s_m)-\mathbb{E}_\pi f\right)^2 = \frac{1}{k}\cdot\max_{m=1,2,\ldots,n}\left(f(s_m)-\mathbb{E}_\pi f\right)^2 \to 0,
\end{aligned}$$

kun $k \to \infty$. □

Lemma 3.2. *Olkoon $(X_k)_{k=0}^{\infty}$ Markovin ketju tila-avaruudella $S = \{s_1, s_2, \ldots, s_n\}$ ja $f : S \to \mathbb{R}$ funktio. Oletetaan, että ketjun siirtymätodennäköisyysmatriisille $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ pätee $P^m > 0$ jollakin $m \in \mathbb{N}$ ja että $\pi \in \mathbb{R}^n$ on matriisin P invariantti jakauma. Tällöin, jos $\mathbb{P}(X_0 = s_i) > 0$, niin*

$$\lim_{k\to\infty}\frac{2}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{j-1}\mathbb{E}_i\left(f(X_j)-\mathbb{E}_\pi f\right)\left(f(X_l)-\mathbb{E}_\pi f\right) = 0.$$

Todistus. Merkitään

$$\bar{f} = f - \mathbb{E}_\pi f,$$

jolloin väitteenä on

$$\lim_{k\to\infty}\frac{2}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{j-1}\mathbb{E}_i\bar{f}(X_j)\bar{f}(X_l) = 0.$$

Huomataan, että kaikilla $l \in \mathbb{N}$ ja $i = 1, 2, \ldots, n$

$$\begin{aligned}
\sum_{v=1}^{n}\sum_{u=1}^{n}\bar{f}(s_v)\bar{f}(s_u)\pi_v p_{i,u}^{(l)} &= \sum_{v=1}^{n}\bar{f}(s_v)\pi_v\sum_{u=1}^{n}\bar{f}(s_u)p_{i,u}^{(l)} \\
&= \sum_{v=1}^{n}\left(f(s_v)-\mathbb{E}_\pi f\right)\pi_v\sum_{u=1}^{n}\bar{f}(s_u)p_{i,u}^{(l)} \\
&= \left(\sum_{v=1}^{n}f(s_v)\pi_v - \mathbb{E}_\pi f\sum_{v=1}^{n}\pi_v\right)\sum_{u=1}^{n}\bar{f}(s_u)p_{i,u}^{(l)} \\
&= \left(\mathbb{E}_\pi f - \mathbb{E}_\pi f\right)\sum_{u=1}^{n}\bar{f}(s_u)p_{i,u}^{(l)} \\
&= 0.
\end{aligned}$$

Olkoon $i \in \{1, 2, \ldots, n\}$ mikä tahansa siten, että $\mathbb{P}(X_0 = s_i) > 0$. Olkoot edelleen $j, l \in \mathbb{N}$ mitä tahansa siten, että $l < j$. Lemman A.1 nojalla

$$\mathbb{P}(X_j = s_v, X_l = s_u | X_0 = s_i) = \mathbb{P}(X_j = s_v | X_l = s_u, X_0 = s_i)\,\mathbb{P}(X_l = s_u | X_0 = s_i),$$

kun $\mathbb{P}(X_l = s_u, X_0 = s_i) > 0$. Lauseen A.2 nojalla puolestaan

$$\mathbb{P}(X_j = s_v | X_l = s_u, X_0 = s_i) = \mathbb{P}(X_j = s_v | X_l = s_u),$$

kun $\mathbb{P}(X_l = s_u, X_0 = s_i) > 0$. Kun vielä huomataan, että ehdosta $\mathbb{P}(X_l = s_u | X_0 = s_i) > 0$ seuraa, että $\mathbb{P}(X_l = s_u, X_0 = s_i) = \mathbb{P}(X_l = s_u | X_0 = s_i)\,\mathbb{P}(X_0 = s_i) > 0$, saadaan, että

$$\begin{aligned}
&\mathbb{E}_i\left(\bar{f}(X_j)\bar{f}(X_l)\right) = \sum_{v=1}^{n}\sum_{u=1}^{n} \bar{f}(s_v)\bar{f}(s_u)\mathbb{P}_i\left(X_j = s_v, X_l = s_u\right) \\
&= \sum_{v=1}^{n}\sum_{u=1}^{n} \bar{f}(s_v)\bar{f}(s_u)\mathbb{P}\left(X_j = s_v, X_l = s_u | X_0 = s_i\right) \\
&= \sum_{v=1}^{n}\sum_{u:\mathbb{P}(X_l=s_u|X_0=s_i)>0} \bar{f}(s_v)\bar{f}(s_u)\mathbb{P}\left(X_j = s_v, X_l = s_u | X_0 = s_i\right) \\
&= \sum_{v=1}^{n}\sum_{u:\mathbb{P}(X_l=s_u|X_0=s_i)>0} \bar{f}(s_v)\bar{f}(s_u)\mathbb{P}\left(X_j = s_v | X_l = s_u, X_0 = s_i\right)\mathbb{P}\left(X_l = s_u | X_0 = s_i\right) \\
&= \sum_{v=1}^{n}\sum_{u:\mathbb{P}(X_l=s_u|X_0=s_i)>0} \bar{f}(s_v)\bar{f}(s_u)\mathbb{P}\left(X_j = s_v | X_l = s_u\right)\mathbb{P}\left(X_l = s_u | X_0 = s_i\right) \\
&\overset{(*)}{=} \sum_{v=1}^{n}\sum_{u:\mathbb{P}(X_l=s_u|X_0=s_i)>0} \bar{f}(s_v)\bar{f}(s_u)p_{u,v}^{(j-l)}\mathbb{P}\left(X_l = s_u | X_0 = s_i\right) \\
&= \sum_{v=1}^{n}\sum_{u=1}^{n} \bar{f}(s_v)\bar{f}(s_u)p_{u,v}^{(j-l)}\mathbb{P}\left(X_l = s_u | X_0 = s_i\right) \\
&\overset{(**)}{=} \sum_{v=1}^{n}\sum_{u=1}^{n} \bar{f}(s_v)\bar{f}(s_u)p_{u,v}^{(j-l)}p_{i,u}^{(l)} \\
&= \sum_{v=1}^{n}\sum_{u=1}^{n} \bar{f}(s_v)\bar{f}(s_u)p_{u,v}^{(j-l)}p_{i,u}^{(l)} - \underbrace{\sum_{v=1}^{n}\sum_{u=1}^{n} \bar{f}(s_v)\bar{f}(s_u)\pi_v p_{i,u}^{(l)}}_{=0} \\
&= \sum_{v=1}^{n}\sum_{u=1}^{n} \bar{f}(s_v)\bar{f}(s_u)\left(p_{u,v}^{(j-l)} - \pi_v\right)p_{i,u}^{(l)}. \qquad (6)
\end{aligned}$$

Kohdissa $(*)$ ja $(**)$ käytettiin Lausetta A.1. Huomaa kohdassa $(*)$, että

$$\mathbb{P}(X_l = s_u) \geq \mathbb{P}(X_l = s_u, X_0 = s_i) > 0,$$

joten Lauseen A.1 käyttö on mahdollista. Lauseen 2.2 nojalla on olemassa luvut $C \geq 0$ ja $0 \leq \beta < 1$ siten, että

$$\left|p_{i,j}^{(k)} - \pi_j\right| \leq C\beta^k$$

kaikilla $i, j = 1, 2, \ldots, n$ ja $k \in \mathbb{N}$, joten yhtälöstä (6) seuraa, että

$$\begin{aligned}
\left|\mathbb{E}_i\bar{f}(X_j)\bar{f}(X_l)\right| &\leq \sum_{v=1}^{n}\sum_{u=1}^{n} \left|\bar{f}(s_v)\right|\left|\bar{f}(s_u)\right|\left|p_{u,v}^{(j-l)} - \pi_v\right| p_{i,u}^{(l)} \\
&\leq \underbrace{\left(\sum_{v=1}^{n}\left|\bar{f}(s_v)\right|\right)\left(\sum_{u=1}^{n}\left|\bar{f}(s_u)\right|\right)C}_{=:M}\beta^{j-l} \\
&= M\beta^{j-l}.
\end{aligned}$$

Näin ollen

$$-\frac{2}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{j-1} M\beta^{j-l} \leq \frac{2}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{j-1} \mathbb{E}_i\left(\bar{f}(X_j)\bar{f}(X_l)\right) \leq \frac{2}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{j-1} M\beta^{j-l},$$

joten riittää osoittaa, että

$$\lim_{k\to\infty}\frac{2}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{j-1} M\beta^{j-l} = 0.$$

Koska $0 \leq \beta < 1$, niin

$$\sum_{l=1}^{\infty}\beta^l < \infty.$$

Kaikilla $k \in \mathbb{N}$

$$\begin{aligned}0 \leq \frac{2}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{j-1} M\beta^{j-l} &= \frac{2M}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{j-1}\beta^l \\ &= \frac{2M}{k^2}\sum_{l=1}^{k-1}(k-l)\beta^l \leq \frac{2M}{k^2}\sum_{l=1}^{k-1}k\beta^l \leq \frac{2M}{k}\sum_{l=1}^{\infty}\beta^l \to 0,\end{aligned}$$

kun $k \to \infty$. □

Seuraava lause kertoo, että Markovin ketju Monte Carlo -simulointi toimii. Lauseen muotoiluun on otettu mallia lähteestä [1], mutta todistus ei seuraa lainkaan lähteen todistusta. Itse asiassa lähteessä [1] todistetaan vahvempi tulos: Lauseen 3.1 tilanteessa

$$\mathbb{P}_i\left(\lim_{k\to\infty}\frac{1}{k}\sum_{j=1}^{k} f(X_j) = \mathbb{E}_\pi f\right) = 1.$$

Lause 3.1. *Olkoon* $(X_k)_{k=0}^{\infty}$ *Markovin ketju tila-avaruudella* $S = \{s_1, s_2, \ldots, s_n\}$ *ja olkoon* $f : S \to \mathbb{R}$ *funktio. Jos ketjun siirtymätodennäköisyysmatriisille* $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ *pätee, että* $P^m > 0$ *jollakin* $m \in \mathbb{N}$*, niin kaikilla* $\varepsilon > 0$ *ja kaikilla* $i = 1, 2, \ldots, n$*, joille* $\mathbb{P}(X_0 = s_i) > 0$*, pätee, että*

$$\lim_{k\to\infty}\mathbb{P}_i\left(\left|\frac{1}{k}\sum_{j=1}^{k} f(X_j) - \mathbb{E}_\pi f\right| > \varepsilon\right) = 0,$$

missä π *on matriisin* P *invariantti jakauma.*

Todistus. Olkoon $\varepsilon > 0$ mikä tahansa. Merkitään $\bar{f} = f - \mathbb{E}_\pi f$ ja huomataan, että kaikilla $k \in \mathbb{N}$

$$\begin{aligned}
\left|\frac{1}{k}\sum_{j=1}^{k} f(X_j) - \mathbb{E}_\pi f\right|^2 &= \left(\frac{1}{k}\sum_{j=1}^{k}(f(X_j) - \mathbb{E}_\pi f)\right)^2 = \frac{1}{k^2}\left(\sum_{j=1}^{k}\bar{f}(X_j)\right)^2 \\
&= \frac{1}{k^2}\sum_{j=1}^{k}\bar{f}(X_j)\sum_{l=1}^{k}\bar{f}(X_l) = \frac{1}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{k}\bar{f}(X_j)\bar{f}(X_l) \\
&= \frac{1}{k^2}\sum_{j=1}^{k}\left(\sum_{l=1}^{j-1}\bar{f}(X_j)\bar{f}(X_l) + \bar{f}(X_j)^2 + \sum_{l=j+1}^{k}\bar{f}(X_j)\bar{f}(X_l)\right) \\
&= \frac{1}{k^2}\sum_{j=1}^{k}\bar{f}(X_j)^2 + \frac{1}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{j-1}\bar{f}(X_j)\bar{f}(X_l) + \frac{1}{k^2}\sum_{j=1}^{k}\sum_{l=j+1}^{k}\bar{f}(X_j)\bar{f}(X_l) \\
&= \frac{1}{k^2}\sum_{j=1}^{k}\bar{f}(X_j)^2 + \frac{2}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{j-1}\bar{f}(X_j)\bar{f}(X_l).
\end{aligned}$$

Näin ollen lemmojen 3.1 ja 3.2 ja Markovin epäyhtälön[1] nojalla

$$\begin{aligned}
\mathbb{P}_i\left(\left|\frac{1}{k}\sum_{j=1}^{k} f(X_j) - \mathbb{E}_\pi f\right| > \varepsilon\right) &= \mathbb{P}_i\left(\left|\frac{1}{k}\sum_{j=1}^{k} f(X_j) - \mathbb{E}_\pi f\right|^2 > \varepsilon^2\right) \\
&\leq \frac{\mathbb{E}_i\left|\frac{1}{k}\sum_{j=1}^{k} f(X_j) - \mathbb{E}_\pi f\right|^2}{\varepsilon^2} \\
&= \frac{\mathbb{E}_i\left(\frac{1}{k^2}\sum_{j=1}^{k}\bar{f}(X_j)^2 + \frac{2}{k^2}\sum_{j=1}^{k}\sum_{l=1}^{j-1}\bar{f}(X_j)\bar{f}(X_l)\right)}{\varepsilon^2} \\
&= \frac{1}{k^2\varepsilon^2}\sum_{j=1}^{k}\mathbb{E}_i\bar{f}(X_j)^2 + \frac{2}{k^2\varepsilon^2}\sum_{j=1}^{k}\sum_{l=1}^{j-1}\mathbb{E}_i\left(\bar{f}(X_j)\bar{f}(X_l)\right) \\
&\to 0,
\end{aligned}$$

kun $k \to \infty$. □

Jotta Lause 3.1 olisi hyödyllinen, täytyy annetulle jakaumalle $\pi \in \mathbb{R}^n$ osata konstruoida siirtymätodennäköisyysmatriisi $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$, jolle pätee, että

(i) $\pi P = \pi$ ja

(ii) $P^m > 0$ jollakin $m \in \mathbb{N}$.

[1] Markovin epäyhtälö: Jos X on ei-negatiivinen satunnaismuuttuja ja $\lambda > 0$, niin $\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}$.

Ehto (i) saadaan voimaan, kun vaaditaan, että matriisi P on kääntyvä jakauman π suhteen.

Määritelmä 3.1. *Siirtymätodennäköisyysmatriisi* $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ *on kääntyvä jakauman* $\pi \in \mathbb{R}^n$ *suhteen, jos*

$$\pi_i p_{i,j} = \pi_j p_{j,i} \text{ kaikilla } i, j = 1, 2, \ldots, n. \tag{7}$$

Markovin ketju $(X_k)_{k=0}^{\infty}$ *on kääntyvä jakauman* π *suhteen, jos sen siirtymätodennäköisyysmatriisi on kääntyvä jakauman* π *suhteen.*

Lause 3.2. *Olkoon* $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ *siirtymätodennäköisyysmatriisi, joka on kääntyvä jakauman* $\pi \in \mathbb{R}^n$ *suhteen. Tällöin* π *on matriisin* P *invariantti jakauma.*

Todistus. Kaikilla $j = 1, 2, \ldots, n$

$$[\pi P]_j = \sum_{i=1}^{n} \pi_i p_{i,j} = \sum_{i=1}^{n} \pi_j p_{j,i} = \pi_j \sum_{i=1}^{n} p_{j,i} = \pi_j,$$

joten $\pi P = \pi$. □

Yhtälö (7) ei kiinnitä matriisin P alkioita, vaan ne täytyy määrätä muilla keinoilla. Ajatellaan Markovin ketjun siirtymä tilasta toiseen kaksivaiheisena tapahtumana: Jos ollaan tilassa s_i, ehdotetaan uutta tilaa $s_j \neq s_i$ todennäköisyydellä $g_{i,j}$. Ehdotettu tila s_j hyväksytään todennäköisyydellä $a_{i,j}$. Jos uuden tilan ehdottaminen ja hyväksyminen ovat toisistaan riippumattomia, niin todennäköisyys siirtyä tilasta s_i tilaan s_j on

$$p_{i,j} = a_{i,j} g_{i,j}.$$

Todennäköisyys pysyä tilassa s_i on

$$1 - \sum_{j \neq i} p_{i,j} = 1 - \sum_{j \neq i} a_{i,j} g_{i,j}.$$

Oletetaan, että ehdokastilojen jakaumat $g_i = (g_{i,1}, g_{i,2}, \ldots, g_{i,n}) \in \mathbb{R}^n$, $i = 1, 2, \ldots, n$, on jo valittu siten, että $g_i > 0$ kaikilla $i = 1, 2, \ldots, n$. Tehtäväksi jää valita tilojen hyväksymistodennäköisyydet $a_{i,j}$, $i, j = 1, 2, \ldots, n$, niin, että kääntyvyysehto (7) toteutuu. Jakaumasta π oletetaan, että $\pi > 0$.

Määritelmä 3.2. *Valitsemalla*

$$a_{i,j} = a_{i,j}^{MH} = \min \left\{ 1, \frac{\pi_j g_{j,i}}{\pi_i g_{i,j}} \right\}$$

kaikilla $i, j = 1, 2, \ldots, n$, *joille* $i \neq j$, *saadaan Metropolis–Hastings -algoritmi, jonka siirtymätodennäköisyysmatriisi on* $P_{MH} = \left[p_{i,j}^{MH}\right]_{i,j=1,2,\ldots,n}$, *missä*

$$p_{i,j}^{MH} = a_{i,j}^{MH} g_{i,j}, \text{ kun } i \neq j,$$

ja

$$p_{i,i}^{MH} = 1 - \sum_{j \neq i} p_{i,j}^{MH}.$$

Määritelmä 3.3. *Valitsemalla*

$$a_{i,j} = a^B_{i,j} = \frac{\pi_j g_{j,i}}{\pi_i g_{i,j} + \pi_j g_{j,i}}$$

kaikilla $i, j = 1, 2, \ldots, n$, *joille* $i \neq j$, *saadaan Barkerin algoritmi, jonka siirtymätodennäköisyysmatriisi on* $P_B = \left[p^B_{i,j}\right]_{i,j=1,2,\ldots,n}$, *missä*

$$p^B_{i,j} = a^B_{i,j} g_{i,j}, \text{ kun } i \neq j,$$

ja

$$p^B_{i,i} = 1 - \sum_{j \neq i} p^B_{i,j}.$$

Lause 3.3. *Metropolis–Hastings -algoritmin siirtymätodennäköisyysmatriisi* P_{MH} *on kääntyvä jakauman* $\pi \in \mathbb{R}^n$ *suhteen.*

Todistus. Olkoot $i, j \in \{1, 2, \ldots, n\}$ mitä tahansa siten, että $i \neq j$. Täytyy osoittaa, että

$$\pi_i p^{\mathrm{MH}}_{i,j} = \pi_j p^{\mathrm{MH}}_{j,i}.$$

Koska kaikilla $a, b, c \geq 0$ pätee, että

$$a \min\{b, c\} = \min\{ab, ac\},$$

niin

$$\pi_i p^{\mathrm{MH}}_{i,j} = \pi_i \min\left\{1, \frac{\pi_j g_{j,i}}{\pi_i g_{i,j}}\right\} g_{i,j} = \min\left\{\pi_i g_{i,j}, \pi_j g_{j,i}\right\} = \pi_j \min\left\{\frac{\pi_i g_{i,j}}{\pi_j g_{j,i}}, 1\right\} g_{j,i} = \pi_j p^{\mathrm{MH}}_{j,i}. \quad \square$$

Lause 3.4. *Barkerin algoritmin siirtymätodennäköisyysmatriisi* P_B *on kääntyvä jakauman* $\pi \in \mathbb{R}^n$ *suhteen.*

Todistus. Olkoot $i, j \in \{1, 2, \ldots, n\}$ siten, että $i \neq j$. Tällöin

$$\pi_i p^{\mathrm{B}}_{i,j} = \pi_i \frac{\pi_j g_{j,i}}{\pi_i g_{i,j} + \pi_j g_{j,i}} g_{i,j} = \pi_j \frac{\pi_i g_{i,j}}{\pi_j g_{j,i} + \pi_i g_{i,j}} g_{j,i} = \pi_j p^{\mathrm{B}}_{j,i}. \quad \square$$

Metropolis–Hastings algoritmia ja Barkerin algoritmia sovelletaan seuraavalla tavalla:

1. Kiinnitetään alkutila $s_{i_0} \in S$.
2. Jos ollaan tilassa s_{i_j}, arvotaan ehdokastila $\hat{s}_{i_{j+1}}$ jakaumasta $g_{i_j} \in \mathbb{R}^n$ ja riippumaton satunnaisluku α välin $[0, 1]$ tasajakaumasta. Jos luku α on pienempi tai yhtä suuri kuin hyväksymistodennäköisyys $a_{i_j, i_{j+1}}$, niin hyväksytään ehdotettu tila, eli asetetaan uudeksi tilaksi $s_{i_{j+1}} = \hat{s}_{i_{j+1}}$. Jos taas luku α on aidosti suurempi kuin hyväksymistodennäköisyys $a_{i_j, i_{j+1}}$, hylätään ehdotettu tila ja pysytään tilassa s_{i_j}, eli asetetaan $s_{i_{j+1}} = s_{i_j}$.
3. Toistetaan kohtaa 2.

Saadaan satunnaisluvut $s_{i_k} \sim e_{i_0} P^k$ - eli Markovin ketju $(s_{i_k})_{k=0}^{\infty}$ - ja Lauseen 3.1 nojalla satunnaisluku $\frac{1}{k} \sum_{j=1}^{k} f(s_{i_j})$ approksimoi lukua $\mathbb{E}_\pi f$ suurilla $k \in \mathbb{N}$.

Huomautus 3.1. *Kuten alussa todettiin, ei odotusarvon*

$$\mathbb{E}_\pi f = \sum_{i=1}^{n} f(s_i)\pi_i$$

laskeminen ole käytännössä mahdollista, jos n *on suuri tai tieto jakaumasta* $\pi \in \mathbb{R}^n$ *on puutteellista. Jos kuitenkin tunnetaan vektori* $\mu \in \mathbb{R}^n$*, joka on verrannollinen jakaumaan* π*, eli* $\mu = c\pi$ *jollakin vakiolla* $c > 0$*, niin Markovin ketju Monte Carlo -simulointi onnistuu sekä Metropolis–Hastings -algoritmilla että Barkerin algoritmilla. Toisin sanoen, vektori* μ *riittää algoritmien siirtymätodennäköisyysmatriisien määrittämiseen: kun* $i \neq j$*, niin*

$$\min\left\{1, \frac{\mu_j g_{j,i}}{\mu_i g_{i,j}}\right\} g_{i,j} = \min\left\{1, \frac{c\pi_j g_{j,i}}{c\pi_i g_{i,j}}\right\} g_{i,j} = \min\left\{1, \frac{\pi_j g_{j,i}}{\pi_i g_{i,j}}\right\} g_{i,j} = p_{i,j}^{MH}$$

ja vastaavasti Barkerin algoritmin tapauksessa

$$\frac{\mu_j g_{j,i}}{\mu_i g_{i,j} + \mu_j g_{j,i}} g_{i,j} = \frac{g_{j,i}}{\frac{\mu_i}{\mu_j} g_{i,j} + g_{j,i}} g_{i,j} = \frac{g_{j,i}}{\frac{\pi_i}{\pi_j} g_{i,j} + g_{j,i}} g_{i,j} = \frac{\pi_j g_{j,i}}{\pi_i g_{i,j} + \pi_j g_{j,i}} g_{i,j} = p_{i,j}^{B}.$$

Huomaa myös, ettei verrannollisuuskerrointa c *tarvitse tuntea: riittää tietää, että jokin verrannollisuuskerroin on olemassa.*

Metropolis–Hastings -algoritmi ja Barkerin algoritmi antavat siis käyttökelpoisen tavan soveltaa Lausetta 3.1 eli Markovin ketju Monte Carlo -simulointia. Myös muita algoritmeja on tietysti olemassa. Lause 3.1 ei kuitenkaan kerro, kuinka suuri luvun $k \in \mathbb{N}$ täytyy olla, jotta todennäköisyys

$$\mathbb{P}_i\left(\left|\frac{1}{k}\sum_{i=1}^{k} f(X_i) - \mathbb{E}_\pi f\right| > \varepsilon\right)$$

on pieni. Tämän arvioiminen on periaatteessa mahdollista, mutta käytännössä vaikeaa. Esimerkiksi lähteessä [1] on osoitettu, että kääntyvälle Markovin ketjulle

$$\mathbb{P}_i\left(\left|\frac{1}{t}\sum_{s=0}^{t-1} f(X_{r+s}) - \mathbb{E}_\pi f\right| \geq \eta\right) \leq \varepsilon,$$

kun

$$r \geq t_{\mathrm{mix}}\left(\frac{\varepsilon}{2}\right) \tag{8}$$

ja

$$t \geq \frac{4\mathrm{Var}_\pi f}{\eta^2 \varepsilon} t_{\mathrm{rel}}. \tag{9}$$

Englanninkielisessä kirjallisuudessa luvut t_{mix} ja t_{rel} ovat nimeltään mixing time ja relaxation time ja niiden määritelmät löytyvät lähteestä [1], mutta tässä tutkielmassa niitä ei tarvita. Epäyhtälöt (8) ja (9) vastaavat *periaatteessa* kysymykseen, miten annetun funktion odotusarvon simulointi toteutetaan halutulla tarkkuudella. *Käytännössä* epäyhtälöiden (8) ja (9) antamat alarajat eivät välttämättä ole laskettavissa tai luotettavasti arvioitavissa. Markovin ketju Monte Carlo -simuloinnin virheelle on johdettu myös eksplisiittiset virherajat; tutustu esimerkiksi lähteeseen [4]. Näiden tarkastelujen sijaan tässä tutkielmassa keskitytään jatkossa seuraavanlaiseen ongelmaan: kahdella siirtymätodennäköisyysmatriisilla P ja Q voi olla sama invariantti jakauma π (esimerkiksi P_{MH} ja P_{B}) ja halutaan selvittää, kumpi matriiseista on parempi annetun funktion odotusarvon Markovin ketju Monte Carlo -simulointiin. Lauseesta 3.1 seuraa, että kumpikin matriiseista P ja Q antaa odotusarvolle arvion halutulla tarkkuudella, jos simulointia jatketaan riittävän kauan. Oletettavasti halutun tarkkuuden saavuttamiseksi vaadittava aika kuitenkin riippuu simulointiin käytettävästä matriisista, joten voisi ajatella, että vertailun lähtökohtana on kyseisten aikojen arvioiminen. Käytännössä näin ei ole, koska simulointiin tarvittavan ajan arvioiminen on vaikeaa, eivätkä pelkät arviot edes mahdollista matriisien luotettavaa vertailua. Voi esimerkiksi käydä niin, että arvioitu simulointiajan alaraja matriisille P on pienempi kuin matriisille Q, mutta silti matriisia Q käyttäen haluttu tarkkuus saavutetaan nopeammin kuin matriisia P käyttäen. Pelkät alarajat eivät kerro tarpeeksi simulointiaikojen todellisista arvoista. Kriteeri, jonka mukaan algoritmien keskinäistä paremmuutta vertaillaan, johdetaan Markovin ketjujen keskeistä raja-arvolausetta apuna käyttäen. Aloitetaan siis keskeisestä raja-arvolauseesta:

Lause 3.5. *Olkoon $(X_k)_{k=0}^{\infty}$ Markovin ketju tila-avaruudella $S = \{s_1, s_2, \ldots, s_n\}$, jonka siirtymätodennäköisyysmatriisille $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ pätee, että $P^m > 0$ jollakin $m \in \mathbb{N}$. Olkoon $\pi \in \mathbb{R}^n$ matriisin P invariantti jakauma. Tällöin kaikille funktioille $f : S \to \mathbb{R}$ on olemassa matriisista P ja funktiosta f riippuva luku $\sigma(P, f)^2 \in [0, \infty[$ siten, että kaikilla ketjun $(X_k)_{k=0}^{\infty}$ alkujakaumilla satunnaismuuttujan*

$$\sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k} f(X_i) - \mathbb{E}_\pi f\right)$$

jakauma suppenee kohti normaalijakaumaa $N(0, \sigma(P, f)^2)$, jonka odotusarvo on 0 ja varianssi on $0 \le \sigma(P, f)^2 < \infty$.

Todistus. Väite seuraa lähteen [5] tuloksesta Corollary 4.2 (ii). □

Lauseen 3.5 tilanteessa varianssi $\sigma(P, f)^2$ riippuu siis vain siirtymätodennäköisyysmatriisista P ja funktiosta f, mutta ei Markovin ketjusta $(X_k)_{k=0}^{\infty}$. Luku $\sigma(P, f)^2$ on siis sama kaikille Markovin ketjuille, joilla on yhteinen siirtymätodennäköisyysmatriisi P. Näin ollen seuraava määritelmä on hyvin asetettu:

Määritelmä 3.4. *Lauseessa 3.5 varianssia $\sigma(P, f)^2$ kutsutaan siirtymätodennäköisyysmatriisin P asymptoottiseksi varianssiksi funktiolle f. Vastaavasti luku $\sigma(P, f) = \sqrt{\sigma(P, f)^2}$ on siirtymätodennäköisyysmatriisin P asymptoottinen keskihajonta funktiolle f.*

Lauseen 3.5 seurauksena voidaan nyt todistaa seuraava tulos, joka on perustana algoritmien vertailulle. Seurauksen 3.1 todistus mukailee lähdettä [6], mutta se käsitellään tässä yksityiskohtaisemmin.

Seuraus 3.1. *Olkoon $(X_k)_{k=0}^{\infty}$ Markovin ketju tila-avaruudella $S = \{s_1, s_2, \ldots, s_n\}$, jonka siirtymätodennäköisyysmatriisille $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ pätee, että $P^m > 0$ jollakin $m \in \mathbb{N}$. Olkoon $\pi \in \mathbb{R}^n$ matriisin P invariantti jakauma, $f : S \to \mathbb{R}$ funktio ja $\varepsilon > 0$. Jos matriisin P asymptoottinen varianssi funktiolle f on $\sigma(P, f)^2 > 0$, niin*

$$\lim_{k\to\infty} \mathbb{P}\left(\left|\frac{1}{k}\sum_{i=1}^{k} f(X_i) - \mathbb{E}_\pi f\right| \ge \frac{\varepsilon}{\sqrt{k}}\right) = 2\Phi\left(-\frac{\varepsilon}{\sigma(P, f)}\right), \tag{10}$$

missä Φ on standardinormaalijakauman $N(0, 1)$ kertymäfunktio.

Todistus. Merkintöjen yksinkertaistamiseksi olkoon

$$Y_k = \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k} f(X_i) - \mathbb{E}_\pi f\right) \text{ kaikilla } k \in \mathbb{N}.$$

Osoitetaan ensin aputuloksena, että kaikilla $\varepsilon > 0$

$$\lim_{k\to\infty} \mathbb{P}(Y_k < \varepsilon) = \Phi\left(\frac{\varepsilon}{\sigma(P,f)}\right).$$

Olkoon $\alpha > 0$ mikä tahansa. Koska normaalijakauman $N(0,1)$ kertymäfunktio Φ on jatkuva, niin on olemassa $\delta > 0$ siten, että

$$-\alpha < \Phi\left(\frac{\epsilon - \delta}{\sigma(P,f)}\right) - \Phi\left(\frac{\epsilon}{\sigma(P,f)}\right) < \alpha.$$

Olkoon Φ_σ normaalijakauman $N(0, \sigma(P,f)^2)$ kertymäfunktio. Lauseen 3.5 ja normaalijakauman ominaisuuksien nojalla

$$\lim_{k\to\infty} \mathbb{P}\left(Y_k \le \varepsilon - \delta\right) = \Phi_\sigma\left(\varepsilon - \delta\right) = \Phi\left(\frac{\varepsilon - \delta}{\sigma(P,f)}\right),$$

joten on olemassa $k_1 \in \mathbb{N}$ siten, että

$$-\alpha < \mathbb{P}\left(Y_k \le \varepsilon - \delta\right) - \Phi\left(\frac{\varepsilon - \delta}{\sigma(P,f)}\right) < \alpha$$

kaikilla $k \ge k_1$. Siispä kaikilla $k \ge k_1$ pätee, että

$$\begin{aligned}
\mathbb{P}\left(Y_k < \varepsilon\right) - \Phi\left(\frac{\varepsilon}{\sigma(P,f)}\right) &\ge \mathbb{P}\left(Y_k \le \varepsilon - \delta\right) - \Phi\left(\frac{\varepsilon}{\sigma(P,f)}\right) \\
&> \left(\Phi\left(\frac{\varepsilon - \delta}{\sigma(P,f)}\right) - \alpha\right) - \Phi\left(\frac{\varepsilon}{\sigma(P,f)}\right) \\
&= \left(\Phi\left(\frac{\varepsilon - \delta}{\sigma(P,f)}\right) - \Phi\left(\frac{\varepsilon}{\sigma(P,f)}\right)\right) - \alpha \\
&> -2\alpha. \qquad (11)
\end{aligned}$$

Toisaalta Lauseen 3.5 nojalla

$$\lim_{k\to\infty} \mathbb{P}\left(Y_k \le \varepsilon\right) = \Phi_\sigma\left(\varepsilon\right) = \Phi\left(\frac{\varepsilon}{\sigma(P,f)}\right),$$

joten on olemassa luku $k_2 \in \mathbb{N}$ siten, että

$$-\alpha < \mathbb{P}\left(Y_k \le \varepsilon\right) - \Phi\left(\frac{\varepsilon}{\sigma(P,f)}\right) < \alpha$$

kaikilla $k \ge k_2$. Erityisesti kaikilla $k \ge k_2$ pätee, että

$$\mathbb{P}\left(Y_k < \varepsilon\right) - \Phi\left(\frac{\varepsilon}{\sigma(P,f)}\right) \le \mathbb{P}\left(Y_k \le \varepsilon\right) - \Phi\left(\frac{\varepsilon}{\sigma(P,f)}\right) < \alpha < 2\alpha. \qquad (12)$$

Yhdistämällä arviot (11) ja (12) nähdään, että kaikilla $k \geq \max\{k_1, k_2\}$ pätee epäyhtälö

$$-2\alpha < \mathbb{P}\left(Y_k < \varepsilon\right) - \Phi\left(\frac{\varepsilon}{\sigma(P,f)}\right) < 2\alpha,$$

eli

$$\left|\mathbb{P}\left(Y_k < \varepsilon\right) - \Phi\left(\frac{\varepsilon}{\sigma(P,f)}\right)\right| < 2\alpha.$$

Näin ollen aputulos

$$\lim_{k\to\infty} \mathbb{P}(Y_k < \varepsilon) = \Phi\left(\frac{\varepsilon}{\sigma(P,f)}\right)$$

on todistettu ja voidaan siirtyä tarkastelemaan varsinaista väitettä, eli rajankäyntiä (10). Lauseesta 3.5 saadaan raja-arvo

$$\lim_{k\to\infty} \mathbb{P}\left(Y_k \leq -\varepsilon\right) = \Phi_\sigma\left(-\varepsilon\right) = \Phi\left(\frac{-\varepsilon}{\sigma(P,f)}\right),$$

mistä yhdessä aputuloksen kanssa seuraa, että

$$\begin{aligned}
\lim_{k\to\infty} \mathbb{P}\left(\left|\frac{1}{k}\sum_{i=1}^{k} f(X_i) - \mathbb{E}_\pi f\right| \geq \frac{\varepsilon}{\sqrt{k}}\right) &= \lim_{k\to\infty} \mathbb{P}\left(\sqrt{k}\left|\frac{1}{k}\sum_{i=1}^{k} f(X_i) - \mathbb{E}_\pi f\right| \geq \varepsilon\right) \\
&= \lim_{k\to\infty} \mathbb{P}\left(|Y_k| \geq \varepsilon\right) \\
&= \lim_{k\to\infty} \left(\mathbb{P}\left(Y_k \geq \varepsilon\right) + \mathbb{P}\left(Y_k \leq -\varepsilon\right)\right) \\
&= \lim_{k\to\infty} \left(1 - \mathbb{P}\left(Y_k < \varepsilon\right) + \mathbb{P}\left(Y_k \leq -\varepsilon\right)\right) \\
&= 1 - \Phi\left(\frac{\varepsilon}{\sigma(P,f)}\right) + \Phi\left(\frac{-\varepsilon}{\sigma(P,f)}\right) \\
&= \Phi\left(-\frac{\varepsilon}{\sigma(P,f)}\right) + \Phi\left(-\frac{\varepsilon}{\sigma(P,f)}\right) \\
&= 2\Phi\left(-\frac{\varepsilon}{\sigma(P,f)}\right). \qquad \square
\end{aligned}$$

Seurauksen 3.1 yhteys algoritmien vertailuun on seuraava: Olkoot $(X_k)_{k=0}^{\infty}$ ja $(Y_k)_{k=0}^{\infty}$ Markovin ketjuja tila-avaruudella $S = \{s_1, s_2, \ldots, s_n\}$ siten, että niiden siirtymätodennäköisyysmatriisit P ja Q toteuttavat Seurauksen 3.1 oletukset samalle invariantille jakaumalle $\pi \in \mathbb{R}^n$. Luku

$$\mathbb{P}\left(\left|\frac{1}{k}\sum_{i=1}^{k} f(X_i) - \mathbb{E}_\pi f\right| \geq \frac{\varepsilon}{\sqrt{k}}\right) \tag{13}$$

on todennäköisyys, että algoritmia P käytettäessä simulointivirheen itseisarvo hetkellä $k \in \mathbb{N}$ on suurempi tai yhtä suuri kuin $\frac{\varepsilon}{\sqrt{k}}$. Jos $\sigma(P,f)^2 \leq \sigma(Q,f)^2$, niin

$$\Phi\left(-\frac{\varepsilon}{\sigma(P,f)}\right) \leq \Phi\left(-\frac{\varepsilon}{\sigma(Q,f)}\right)$$

ja Seurauksen 3.1 nojalla todennäköisyys (13) on rajalla $k \to \infty$ korkeintaan yhtä suuri kuin vastaava todennäköisyys algoritmille Q. Toisin sanoen algoritmi P on asymptoottisesti parempi algoritmi Q.

Tehdään lopuksi yhteenveto tämän luvun keskeisistä asioista ja otetaan lyhyt katsaus tulevaan. Lauseessa 3.1 osoitettiin, miten odotusarvoa $\mathbb{E}_\pi f$ voi Markovin ketjun $(X_k)_{k=0}^\infty$ avulla approksimoida satunnaisluvulla

$$\frac{1}{k}\sum_{j=1}^{k} f(X_j).$$

Markovin ketjun $(X_k)_{k=0}^\infty$ simuloimiseksi on tarjolla useita simulointialgoritmeja, joista esimerkkeinä käsiteltiin Metropolis–Hastings -algoritmi ja Barkerin algoritmi. Algoritmista riippuu, kuinka hyvin $\frac{1}{k}\sum_{j=1}^{k} f(X_j)$ approksimoi lukua $\mathbb{E}_\pi f$. Approksimaation luotettavuutta hetkellä k voidaan kuvata todennäköisyydellä

$$\mathbb{P}\left(\left|\frac{1}{k}\sum_{j=1}^{k} f(X_j) - \mathbb{E}_\pi f\right| \geq \frac{\varepsilon}{\sqrt{k}}\right),$$

jolle johdettiin raja-arvo

$$2\Phi\left(-\frac{\varepsilon}{\sigma(P,f)}\right).$$

Tästä pääteltiin, että algoritmien asymptoottisen paremmuuden vertailemiseksi kannattaa tutkia asymptoottisia variansseja. Avoimeksi kysymykseksi kuitenkin jää, miten tämä vertailu käytännössä tehdään. Yleensä algoritmeista tiedetään vain niiden siirtymätodennäköisyysmatriisit, joten algoritmien vertailemiseksi täytyisi tietää, miten asymptoottinen varianssi riippuu siirtymätodennäköisyysmatriisista. Tähän ongelmaan palataan luvussa 5. Ratkaisun perustana on seuraava lause, joka antaa asymptoottiselle varianssille hieman konkreettisemman esityksen kuin Lause 3.5.

Lause 3.6. *Olkoon $(X_k)_{k=0}^\infty$ Markovin ketju tila-avaruudella $S = \{s_1, s_2, \ldots, s_n\}$, jonka siirtymätodennäköisyysmatriisille $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ pätee, että $P^m > 0$ jollakin $m \in \mathbb{N}$. Jos Markovin ketjun $(X_k)_{k=0}^\infty$ alkujakauma on matriisin P invariantti jakauma $\pi \in \mathbb{R}^n$, niin matriisin P asymptoottinen varianssi funktiolle $f : S \to \mathbb{R}$ on*

$$\sigma(P,f)^2 = \lim_{k\to\infty} \frac{1}{k}\mathbb{E}\left(\sum_{i=1}^{k}\left(f(X_i) - \mathbb{E}_\pi f\right)\right)^2. \tag{14}$$

Todistus. Väite seuraa lähteen [7] tuloksesta Theorem 17.0.1. □

Huomaa, että Lauseen 3.6 tulos ei sellaisenaan vielä riitä algoritmien vertailuun, koska siirtymätodennäköisyysmatriisin P ja asymptoottisen varianssin $\sigma(P,f)^2$ välinen eksplisiittinen riippuvuus ei ole selvillä. Vertailuun kyllä tarvitaan Lausetta 3.6, mutta erilaisessa muodossa. Oleellista on, että yhtälön (14) oikea puoli voidaan kirjoittaa funktionaalianalyysin tuloksia hyödyntäen sisätulon muotoon, jossa myös matriisia P vastaava lineaarikuvaus esiintyy. Tähän päästään kuitenkin vasta luvussa 5.

4 Siirtymätodennäköisyysmatriisin ominaisvektorit ja -arvot

Olkoon $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ siirtymätodennäköisyysmatriisi siten, että $P^m > 0$ jollakin $m \in \mathbb{N}$, ja olkoon $\pi \in \mathbb{R}^n$ matriisin P invariantti jakauma. Tässä luvussa osoitetaan, että jos siirtymätodennäköisyysmatriisi P on kääntyvä jakauman π suhteen, niin avaruudella $\mathbb{R}^n$ on ortonormaali kanta $u_1, u_2, \ldots, u_n \in \mathbb{R}^n$, jonka jokainen vektori on matriisin P ominaisvektori: on olemassa luvut $\lambda_i \in \mathbb{R}$ siten, että

$$Pu_i = \lambda_i u_i \text{ kaikilla } i = 1, 2, \ldots, n.$$

Lisäksi osoitetaan, että $\lambda_i = 1$ jollakin $i = 1, 2, \ldots, n$ ja että vastaavaksi ominaisvektoriksi voidaan valita

$$u_i = (1, 1, \ldots, 1) \in \mathbb{R}^n.$$

Muille ominaisarvoille pätee, että $|\lambda_j| < 1$ kaikilla $j \neq i$. Tulos on luonteeltaan sellainen, että sen yhteyttä asymptoottisten varianssien vertailuun on tässä vaiheessa valitettavan vaikea havainnollistaa muuten kuin toteamalla, että se on tärkeä tiettyjen "teknisten"tarkastelujen läpikäymiseksi. On myös syytä huomata, että vektoreiden $u_i, u_2, \ldots, u_n \in \mathbb{R}^n$ ortonormaalisuus riippuu käytetystä sisätulosta. Avaruuden $\mathbb{R}^n$ tavallisen sisätulon

$$\langle \cdot | \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \ \langle x|y \rangle = \sum_{i=1}^{n} x_i y_i$$

sijaan tarvitaan nyt painotettua sisätuloa

$$\langle \cdot | \cdot \rangle_\pi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \ \langle x|y \rangle_\pi = \sum_{i=1}^{n} x_i y_i \pi_i.$$

Sisätulot $\langle \cdot | \cdot \rangle$ ja $\langle \cdot | \cdot \rangle_\pi$ määräävät avaruuteen $\mathbb{R}^n$ normit

$$\|\cdot\| : \mathbb{R}^n \to \mathbb{R}, \|x\| = \sqrt{\langle x|x \rangle}$$

ja

$$\|\cdot\|_\pi : \mathbb{R}^n \to \mathbb{R}, \|x\| = \sqrt{\langle x|x \rangle_\pi}.$$

Lemmaa 4.1 lukuun ottamatta tämän luvun tulokset ovat lähteestä [1]. Lause 4.1 on esitetty lähteessä [1] ilman todistusta ja Lauseen 4.2 todistusta on tässä jonkin verran muutettu. Myös luvun rakenne seuraa tarkasti lähteen [1] esitystä.

Ennen päätuloksen 4.2 todistamista tarkastellaan siirtymätodennäköisyysmatriisin ominaisarvoja. Avuksi tarvitaan seuraava lemma.

Lemma 4.1. *Olkoon $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ siirtymätodennäköisyysmatriisi siten, että $P^m > 0$ jollakin $m \in \mathbb{N}$, ja olkoon $\pi \in \mathbb{R}^n$ matriisin P invariantti jakauma. Jos $x \in \mathbb{R}^n$ siten, että*

$$xP = x,$$

niin on olemassa $c \in \mathbb{R}$ siten, että

$$x = c\pi.$$

Todistus. Jos x on nollavektori, väite on tosi vakiolla $c = 0$. Oletetaan, että x ei ole nollavektori. Koska P^m on siirtymätodennäköisyysmatriisi siten, että $P^m > 0$ ja relaatiosta $xP = x$ seuraa, että $xP^m = x$, niin Lemman 2.1 nojalla $x \geq 0$ tai $x \leq 0$. Kummassakin tapauksessa vektorille

$$y = \frac{x}{\sum_{j=1}^{n} x_j} \in \mathbb{R}^n$$

pätee, että

$$y_i = \frac{x_i}{\sum_{j=1}^{n} x_j} \geq 0 \text{ kaikilla } i = 1, 2, \ldots, n$$

ja

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \frac{x_i}{\sum_{j=1}^{n} x_j} = \frac{1}{\sum_{j=1}^{n} x_j} \sum_{i=1}^{n} x_i = 1,$$

joten $y \in \mathbb{R}^n$ on jakauma. Lisäksi

$$yP = \left(\frac{x}{\sum_{j=1}^{n} x_j} \right) P = \frac{1}{\sum_{j=1}^{n} x_j}(xP) = \frac{1}{\sum_{j=1}^{n} x_j} x = y,$$

joten myös vektori y on matriisin P invariantti jakauma. Lauseen 2.1 yksikäsitteisyysosan nojalla on

$$\frac{x}{\sum_{j=1}^{n} x_j} = y = \pi,$$

mistä

$$x = \left(\sum_{j=1}^{n} x_j \right) \pi,$$

ja siten luvuksi $c \in \mathbb{R}$ voidaan ottaa

$$c = \sum_{j=1}^{n} x_j. \qquad \square$$

Lause 4.1. *Olkoon* $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ *siirtymätodennäköisyysmatriisi.*

(i) Jos $\lambda \in \mathbb{R}$ *on matriisin* P *ominaisarvo, niin* $|\lambda| \leq 1$.

(ii) Jos $P^m > 0$ *jollakin* $m \in \mathbb{N}$ *ja* $x \in \mathbb{R}$ *siten, että* $Px = x$, *niin* $x_i = x_j$ *kaikilla* $i, j = 1, 2, \ldots, n$. *Toisin sanoen ominaisarvoa* 1 *vastaavien ominaisvektoreiden avaruus*

$$\{x \in \mathbb{R}^n : Px = x\}$$

on vektorin $\mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$ *virittämä.*

(iii) Jos $P^m > 0$ *jollakin* $m \in \mathbb{N}$, *niin* -1 *ei ole matriisin* P *ominaisarvo.*

Todistus. (i) Olkoon $\lambda \in \mathbb{R}$ matriisin P ominaisarvo ja $x \in \mathbb{R}^n \setminus \{\bar{0}\}$ siten, että $Px = \lambda x$. Olkoon $i \in \{1, 2, \ldots, n\}$ se indeksi, jolle

$$|x_i| = \max\{|x_j| : j = 1, 2, \ldots, n\} > 0.$$

Tällöin

$$|\lambda|\,|x_i| = |\lambda x_i| = |[Px]_i| = \left|\sum_{j=1}^{n} p_{i,j}x_j\right| \le \sum_{j=1}^{n} p_{i,j}\,|x_j| \le |x_i| \sum_{j=1}^{n} p_{i,j} = |x_i|\,,$$

mistä väite seuraa jakamalla puolittain luvulla $|x_i| > 0$.

(ii) Jos väite on epätosi, on olemassa $x \in \mathbb{R}^n$ siten, että $Px = x$ ja

$$x_j = \min_{k=1,2,\ldots,n} x_k < \max_{k=1,2,\ldots,n} x_k = x_i.$$

Koska $P^m > 0$, niin

$$p_{i,j}^{(m)} x_j < p_{i,j}^{(m)} x_i$$

ja edelleen

$$\begin{aligned}[P^m x]_i &= \sum_{k=1}^{n} p_{i,k}^{(m)} x_k = p_{i,j}^{(m)} x_j + \sum_{k\neq j} p_{i,k}^{(m)} x_k < p_{i,j}^{(m)} x_i + x_i \sum_{k\neq j} p_{i,k}^{(m)} \\ &= x_i \left(p_{i,j}^{(m)} + \sum_{k\neq j} p_{i,k}^{(m)} \right) = x_i \sum_{k=1}^{n} p_{i,k}^{(m)} = x_i.\end{aligned}$$

Tämä on ristiriita, sillä oletuksesta $Px = x$ seuraa, että $P^m x = x$, ja siten $[P^m x]_i = x_i$.

(iii) Matriisin vasemman- ja oikeanpuoleiset ominaisarvot ovat samat, joten riittää osoittaa, että -1 ei ole matriisin P vasemmanpuoleinen ominaisarvo. Tehdään antiteesi: -1 on matriisin P vasemmanpuoleinen ominaisarvo. Tällöin on olemassa $x \in \mathbb{R}^n \setminus \{\bar{0}\}$ siten, että $xP = -x$. Koska $P^m > 0$, niin Lauseen 2.1 nojalla matriisilla P on invariantti jakauma $\pi \in \mathbb{R}^n$. Helposti nähdään, että jakauma π on myös siirtymätodennäköisyysmatriisin P^2 invariantti jakauma. Osoitetaan seuraavaksi, että $(P^2)^m > 0$. Lauseen A.3 nojalla $(P^2)^m = P^{2m}$, joten riittää osoittaa, että $P^{2m} > 0$. Tämä seuraa suoraan Lauseesta A.4: koska $P^m > 0$ ja $2m \ge m$, niin Lauseen A.4 nojalla myös $P^{2m} > 0$. Edelleen, koska

$$xP^2 = (xP)P = -xP = -(xP) = -(-x) = x,$$

niin soveltamalla Lemmaa 4.1 siirtymätodennäköisyysmatriisiin P^2 nähdään, että $x = c\pi$ jollakin $c \in \mathbb{R}$. Näin ollen

$$-x = xP = (c\pi)P = c(\pi P) = c\pi = x,$$

mistä seuraa, että $x = \bar{0}$. Tämä on ristiriidassa oletuksen $x \in \mathbb{R}^n \setminus \{\bar{0}\}$ kanssa, joten antiteesi on epätosi. □

Seuraavaa todistusta varten kannattaa kerrata itseadjungoidun kuvauksen määritelmä ja spektrilause eli Määritelmä B.7 ja Lause B.13.

Lause 4.2. *Olkoon $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ siirtymätodennäköisyysmatriisi siten, että $P^m > 0$ jollakin $m \in \mathbb{N}$. Jos matriisi P on kääntyvä invariantin jakaumansa $\pi \in \mathbb{R}^n$ suhteen, niin avaruudella $(\mathbb{R}^n, \langle\cdot|\cdot\rangle_\pi)$ on ortonormaali kanta $u_1, u_2, \ldots, u_n \in \mathbb{R}^n$, jonka jokainen vektori on matriisin P ominaisvektori: löytyy luvut $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$ siten, että*

$$Pu_i = \lambda_i u_i \text{ kaikilla } i = 1, 2, \ldots, n.$$

Lisäksi $\lambda_i = 1$ *jollakin* $i = 1, 2, \ldots, n$ *ja vastaavaksi ominaisvektoriksi voidaan valita*

$$u_i = (1, 1, \ldots, 1) \in \mathbb{R}^n.$$

Muille ominaisarvoille pätee, että $|\lambda_j| < 1$ *kaikilla* $j \neq i$.

Todistus. Osoitetaan, että siirtymätodennäköisyysmatriisia $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ vastaava lineaarikuvaus $P : \mathbb{R}^n \to \mathbb{R}^n$ on itseadjungoitu sisätulon $\langle \cdot | \cdot \rangle_\pi$ suhteen. Olkoot $x, y \in \mathbb{R}^n$ mitä tahansa. Siirtymätodennäköisyysmatriisin P kääntyvyyden nojalla

$$\pi_i p_{i,j} = \pi_j p_{j,i}$$

kaikilla $i, j = 1, 2, \ldots, n$, joten saadaan, että

$$\begin{aligned} \langle Px|y \rangle_\pi &= \sum_{i=1}^{n} [Px]_i \, y_i \pi_i = \sum_{i=1}^{n} \sum_{j=1}^{n} p_{i,j} x_j y_i \pi_i = \sum_{j=1}^{n} x_j \sum_{i=1}^{n} \pi_i p_{i,j} y_i = \sum_{j=1}^{n} x_j \sum_{i=1}^{n} \pi_j p_{j,i} y_i \\ &= \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{n} p_{j,i} y_i \right) \pi_j = \sum_{j=1}^{n} x_j \, [Py]_j \, \pi_j = \langle x|Py \rangle_\pi \, . \end{aligned}$$

Näin ollen lineaarikuvaus P on itseadjungoitu ja siten Lauseesta B.13 seuraa, että avaruudella $(\mathbb{R}^n, \langle \cdot | \cdot \rangle_\pi)$ on ortonormaali kanta $u_1, u_2, \ldots, u_n \in \mathbb{R}^n$, jonka jokainen vektori on kuvauksen P (ja siten myös matriisin P) ominaisvektori. Olkoot vektoreita $u_1, u_2, \ldots, u_n \in \mathbb{R}^n$ vastaavat ominaisarvot $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$. Koska vektorille $\mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$ pätee, että

$$[P\mathbf{1}]_i = \sum_{j=1}^{n} p_{i,j} \mathbf{1}_j = \sum_{j=1}^{n} p_{i,j} = 1 = \mathbf{1}_i \text{ kaikilla } i = 1, 2, \ldots, n,$$

on luku 1 matriisin P ominaisarvo. Tästä seuraa, että $\lambda_i = 1$ jollakin $i = 1, 2, \ldots, n$, mikä nähdään seuraavasti: Olkoon $x \in \mathbb{R}^n \backslash \{\bar{0}\}$ ominaisarvoa 1 vastaava ominaisvektori. Koska vektorit $u_1, u_2, \ldots, u_n$ ovat avaruuden $\mathbb{R}^n$ kantavektoreita, niin vektorilla x on esitys

$$x = \sum_{i=1}^{n} c_i u_i,$$

missä $c_i \in \mathbb{R}$ kaikilla $i = 1, 2, \ldots, n$. Näin ollen

$$Px = P \left(\sum_{i=1}^{n} c_i u_i \right) = \sum_{i=1}^{n} P (c_i u_i) = \sum_{i=1}^{n} c_i P u_i = \sum_{i=1}^{n} c_i \lambda_i u_i.$$

Nyt siis $Px = x$, eli

$$\sum_{i=1}^{n} c_i \lambda_i u_i = \sum_{i=1}^{n} c_i u_i.$$

Siirtämällä kaikki termit vasemmalle puolelle saadaan edelleen, että

$$\sum_{i=1}^{n} (\lambda_i - 1) c_i u_i = \bar{0},$$

mistä vektoreiden $u_1, u_2, \ldots, u_n$ lineaarisen riippumattomuuden nojalla seuraa, että $(\lambda_i - 1)c_i = 0$ kaikilla $i = 1, 2, \ldots, n$. Koska x ei ole nollavektori, on $c_i \neq 0$ jollakin $i = 1, 2, \ldots, n$, ja siten on oltava $\lambda_i = 1$. Osoitetaan seuraavaksi, että vektoriksi u_i voidaan valita

$$u_i = \mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n.$$

Koska

$$Pu_i = \lambda_i u_i = u_i,$$

niin Lauseen 4.1 kohdan (ii) nojalla vektori u_i on muotoa

$$u_i = c\mathbf{1} = (c, c, \ldots, c) \in \mathbb{R}^n$$

jollakin $c \in \mathbb{R}$. Vektoreiden $u_1, u_2, \ldots, u_n$ ortonormaalisuuden nojalla

$$1 = \|u_i\|_\pi^2 = \langle u_i | u_i \rangle_\pi = \sum_{j=1}^{n} c^2 \pi_j = c^2 \sum_{j=1}^{n} \pi_j = c^2,$$

mistä seuraa, että $c = 1$ tai $c = -1$. Edelleen kaikilla $j \neq i$

$$0 = \langle u_i | u_j \rangle_\pi = c \langle \mathbf{1} | u_j \rangle_\pi ,$$

joten on oltava $\langle \mathbf{1} | u_j \rangle_\pi = 0$ luvusta c riippumatta. Näin ollen voidaan valita $c = 1$. Erityisesti siis vektoriksi u_i voidaan valita $u_i = (1, 1, \ldots, 1) \in \mathbb{R}^n$. Osoitetaan vielä, että $|\lambda_j| < 1$ kaikilla $j \neq i$. Huomataan, että $\lambda_j \neq 1$ kaikilla $j \neq i$, sillä muutoin Lauseen 4.1 kohdasta (ii) seuraisi, että sekä u_i että u_j kuuluvat vektorin $\mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$ virittämään aliavaruuteen, mikä on vastoin vektoreiden $u_1, u_2, \ldots, u_n$ lineaarista riippumattomuutta. Lisäksi, koska $P^m > 0$, niin Lauseen 4.1 kohdan (iii) nojalla -1 ei ole matriisin P ominaisarvo. Näin ollen väite $|\lambda_j| < 1$ kaikilla $j \neq i$ seuraa Lauseen 4.1 kohdasta (i). □

5 Peskunin järjestys ja asymptoottisten varianssien vertailu

Tässä luvussa selvitetään, miten Määritelmän 3.4 asymptoottinen varianssi $\sigma(P,f)^2$ riippuu Markovin ketjun $(X_k)_{k=0}^{\infty}$ siirtymätodennäköisyysmatriisista $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$. Vastaus kysymykseen saadaan Lauseessa 5.1. Lisäksi määritellään Peskunin järjestys ja osoitetaan Lausetta 5.1 apuna käyttäen, että Peskunin järjestyksen avulla voidaan tietyin edellytyksin vertailla kahden siirtymätodennäköisyysmatriisin P ja Q asymptoottisia variansseja. Vertailu on hyödyllistä, sillä kahdesta algoritmista (matriisista) se, jonka asymptoottinen varianssi on pienempi, soveltuu paremmin Markovin ketju Monte Carlo -simulointiin. Peskunin järjestyksen voima puolestaan piilee siinä, että sen avulla asymptoottisten varianssien vertailu on helppo tehdä: riittää vertailla siirtymätodennäköisyysmatriisien alkioiden suuruuksia. Tätä havainnollistaa Lause 5.3, jossa osoitetaan, että Metropolis–Hastings -algoritmi on parempi kuin Barkerin algoritmi.

Tässä luvussa matriisista P oletetaan, että:

Oletus 5.1. *Matriisi $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ on siirtymätodennäköisyysmatriisi siten, että*

(i) $P^m > 0$ jollakin $m \in \mathbb{N}$

(ii) P on kääntyvä invariantin jakaumansa $\pi \in \mathbb{R}^n$ suhteen.

Kun matriisi P on kuten Oletuksessa 5.1, niin Lauseen 4.2 nojalla avaruudella $(\mathbb{R}^n, \langle\cdot|\cdot\rangle_\pi)$ on ortonormaali kanta $u_1, u_2, \ldots, u_n \in \mathbb{R}^n$, jonka jokainen vektori on matriisin P ominaisvektori:

$$Pu_i = \lambda_i u_i \text{ kaikilla } i = 1,2,\ldots,n,$$

missä $\lambda_i \in \mathbb{R}$ kaikilla $i = 1,2,\ldots,n$. Lisäksi $\lambda_i = 1$ jollakin $i = 1,2,\ldots,n$ ja sitä vastaavaksi ominaisvektoriksi voidaan valita

$$u_i = (1,1,\ldots,1) \in \mathbb{R}^n.$$

Tarvittaessa indeksointia muuttamalla voidaan olettaa, että $\lambda_1 = 1$ ja

$$u_1 = (1,1,\ldots,1) \in \mathbb{R}^n.$$

Tällöin Lauseen 4.2 nojalla pätee, että

$$\lambda_* = \max\{|\lambda_i| : i = 2,3,\ldots,n\} < 1.$$

Merkitään vielä

$$u_1^{\perp} = \{x \in \mathbb{R}^n : \langle x|u_1\rangle_\pi = 0\}$$

ja ryhdytään todistamaan asymptoottisen varianssin määritykseen tarvittavia aputuloksia. Todistuksissa hyödynnetään funktionaalianalyysin tuloksia, joita on koottu Liitteeseen B. Erityisesti tarvitaan käsitteitä Banachin avaruus (Määritelmä B.3), lineaarikuvauksen operaattorinormi (Määritelmä B.1) ja sarjan itseinen suppeneminen (Määritelmä B.2).

Oletetaan tunnetuksi, että $(\mathbb{R}^n, \langle\cdot|\cdot\rangle)$ on Banachin avaruus ja osoitetaan, että myös $(\mathbb{R}^n, \langle\cdot|\cdot\rangle_\pi)$ on Banachin avaruus[2].

[2] Väite seuraa triviaalisti myös siitä, että äärellisulotteisessa avaruudessa kaikki normit ovat keskenään ekvivalentteja, mutta normien ekvivalenssia ei tässä tutkielmassa todisteta.

Lemma 5.1. *Avaruus* $(\mathbb{R}^n, \langle\cdot|\cdot\rangle_\pi)$ *on Banachin avaruus.*

Todistus. Olkoon $(x_k)_{k=1}^\infty$ mikä tahansa Cauchyn jono siten, että $x_k \in \mathbb{R}^n$ kaikilla $k \in \mathbb{N}$. Täytyy osoittaa, että on olemassa $x \in \mathbb{R}^n$ siten, että $\|x_k - x\|_\pi \to 0$, kun $k \to \infty$. Koska $(\mathbb{R}^n, \langle\cdot|\cdot\rangle)$ on Banachin avaruus, on olemassa $x \in \mathbb{R}^n$ siten, että $\|x_k - x\| \to 0$, kun $k \to \infty$. Koska funktio $[0, \infty[\to \mathbb{R}$: $y \mapsto \sqrt{y}$ on kasvava ja $0 \leq \pi_i \leq 1$ kaikilla $i = 1, 2, \dots, n$, niin

$$
\begin{aligned}
0 \leq \|x_k - x\|_\pi &= \sqrt{\langle x_k - x|x_k - x\rangle_\pi} = \sqrt{\sum_{i=1}^n (x_{k,i} - x_i)^2 \pi_i} \\
&\leq \sqrt{\sum_{i=1}^n (x_{k,i} - x_i)^2} = \sqrt{\langle x_k - x|x_k - x\rangle} = \|x_k - x\| \to 0,
\end{aligned}
$$

mistä väite seuraa. □

Huomautus 5.1. *Soveltamalla Lausetta B.11 joukkoon* $\{u_1\}$ *nähdään, että* $u_1^\perp = \{u_1\}^\perp$ *on avaruuden* $(\mathbb{R}^n, \langle\cdot|\cdot\rangle_\pi)$ *suljettu aliavaruus. Koska* $(\mathbb{R}^n, \langle\cdot|\cdot\rangle_\pi)$ *on Lemman 5.1 nojalla Banachin avaruus, seuraa Lauseesta B.5, että myös* $u_1^\perp$ *on Banachin avaruus. Lauseen B.4 nojalla jokainen avaruuden* $(u_1^\perp, \langle\cdot|\cdot\rangle_\pi)$ *itseisesti suppeneva sarja suppenee.*

Usein tarkastellaan Oletuksen 5.1 matriisia P vastaavaa lineaarikuvausta niin, että kuvauksen määrittelyjoukoksi rajataan avaruuden $(\mathbb{R}^n, \langle\cdot|\cdot\rangle_\pi)$ aliavaruus $(u_1^\perp, \langle\cdot|\cdot\rangle_\pi)$. Seuraavaksi osoitetaan, että tällöin kyseessä on lineaarikuvaus $P : u_1^\perp \to u_1^\perp$, eli että $Px \in u_1^\perp$ aina, kun $x \in u_1^\perp$. Matriisille ja sitä vastaavalle lineaarikuvaukselle käytetään samaa merkintää.

Lemma 5.2. *Olkoon matriisi* $P = [p_{i,j}]_{i,j=1,2,\dots,n}$ *kuten Oletuksessa 5.1. Kaikilla* $x \in u_1^\perp$ *myös* $Px \in u_1^\perp$.

Todistus. Olkoon $x \in u_1^\perp$ mikä tahansa. Koska $u_1^\perp \subset \mathbb{R}^n$ ja vektorit $u_1, u_2, \dots, u_n$ ovat sisätuloavaruuden $(\mathbb{R}^n, \langle\cdot|\cdot\rangle_\pi)$ ortogonaalisia kantavektoreita, niin Lauseen B.9 nojalla

$$x = \sum_{i=1}^n \langle x|u_i\rangle_\pi u_i.$$

Koska $\langle x|u_1\rangle_\pi = 0$, niin

$$x = \sum_{i=1}^n \langle x|u_i\rangle_\pi u_i = \sum_{i=2}^n \langle x|u_i\rangle_\pi u_i. \tag{15}$$

Erityisesti

$$
\begin{aligned}
Px &= P\left(\sum_{i=2}^n \langle x|u_i\rangle_\pi u_i\right) = \sum_{i=2}^n P\left(\langle x|u_i\rangle_\pi u_i\right) \\
&= \sum_{i=2}^n \langle x|u_i\rangle_\pi Pu_i = \sum_{i=2}^n \langle x|u_i\rangle_\pi \lambda_i u_i, \qquad (16)
\end{aligned}
$$

ja siten

$$\langle Px|u_1\rangle_\pi = \left\langle \sum_{i=2}^{n} \langle x|u_i\rangle_\pi \lambda_i u_i \middle| u_1 \right\rangle_\pi = \sum_{i=2}^{n} \langle x|u_i\rangle_\pi \lambda_i \underbrace{\langle u_i|u_1\rangle_\pi}_{=0\;\forall\; i=2,3,\ldots,n} = 0.$$

Näin ollen $Px \in u_1^\perp$. □

Seuraavaksi tarkastellaan lineaarikuvauksen $P : u_1^\perp \to u_1^\perp$ operaattorinormia.

Lemma 5.3. *Olkoon matriisi $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ kuten Oletuksessa 5.1. Lineaarikuvauksen $P : u_1^\perp \to u_1^\perp$ operaattorinormi on $\|P\| = \lambda_*$.*

Todistus. Osoitetaan ensin, että $\|P\| \leq \lambda_*$. Olkoon $x \in u_1^\perp$ mikä tahansa siten, että $\|x\|_\pi \leq 1$. Koska $\langle x|u_1\rangle_\pi = 0$, niin samoin kuin Lemman 5.2 todistuksessa saadaan, että

$$x = \sum_{i=1}^{n} \langle x|u_i\rangle_\pi u_i = \sum_{i=2}^{n} \langle x|u_i\rangle_\pi u_i \tag{17}$$

ja

$$\begin{aligned} Px &= P\left(\sum_{i=2}^{n} \langle x|u_i\rangle_\pi u_i\right) = \sum_{i=2}^{n} P\left(\langle x|u_i\rangle_\pi u_i\right) \\ &= \sum_{i=2}^{n} \langle x|u_i\rangle_\pi Pu_i = \sum_{i=2}^{n} \langle x|u_i\rangle_\pi \lambda_i u_i. \end{aligned} \tag{18}$$

Käyttämällä yhtälöitä (17) ja (18) sekä soveltamalla Pythagoraan lausetta B.10 ortogonaalisiin vektoreihin $u_1, u_2, \ldots, u_n \in \mathbb{R}^n$ saadaan, että

$$\begin{aligned} \|Px\|_\pi^2 &= \left\| \sum_{i=2}^{n} \langle x|u_i\rangle_\pi \lambda_i u_i \right\|_\pi^2 = \sum_{i=2}^{n} \|\langle x|u_i\rangle_\pi \lambda_i u_i\|_\pi^2 = \sum_{i=2}^{n} |\lambda_i|^2 \, \|\langle x|u_i\rangle_\pi u_i\|_\pi^2 \\ &\leq \lambda_*^2 \sum_{i=2}^{n} \|\langle x|u_i\rangle_\pi u_i\|_\pi^2 = \lambda_*^2 \left\| \sum_{i=2}^{n} \langle x|u_i\rangle_\pi u_i \right\|_\pi^2 = \lambda_*^2 \|x\|_\pi^2 \leq \lambda_*^2. \end{aligned}$$

Tästä seuraa, että $\|Px\|_\pi \leq \lambda_*$, ja siten $\|P\| = \sup_{\|x\|_\pi \leq 1} \|Px\|_\pi \leq \lambda_*$. Osoitetaan, että myös $\|P\| \geq \lambda_*$. Olkoon $i \in \{2, 3, \ldots, n\}$ se indeksi, jolle $|\lambda_i| = \lambda_*$. Koska $u_i \in u_1^\perp$ siten, että $\|u_i\|_\pi = 1$ ja

$$\|Pu_i\|_\pi = \|\lambda_i u_i\|_\pi = |\lambda_i| \|u_i\|_\pi = \lambda_*,$$

niin myös $\|P\| = \sup_{\|x\|_\pi \leq 1} \|Px\|_\pi \geq \lambda_*$. □

Jos matriisi $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ on kuten Oletuksessa 5.1, niin Lemman 5.2 nojalla voidaan tarkastella lineaarikuvausta $P : u_1^\perp \to u_1^\perp$. Erityisesti kaikille $k \in \mathbb{N}_0$ voidaan määritellä lineaarikuvaukset $P^k : u_1^\perp \to u_1^\perp$ rekursiivisesti asettamalla

$$\begin{aligned} &P^0 x = x, \\ &P^1 = P \text{ ja} \\ &P^{k+1} x = P\left(P^k x\right). \end{aligned}$$

Seuraavassa lemmassa tarkastellaan sarjan

$$\sum_{k=0}^{\infty} P^k x$$

suppenemista, kun $x \in u_1^{\perp}$.

Lemma 5.4. *Olkoon matriisi $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ kuten Oletuksessa 5.1. Lineaarikuvaukselle $P : u_1^{\perp} \to u_1^{\perp}$ pätee, että*

(i) Avaruuden $\left(u_1^{\perp}, \langle\cdot|\cdot\rangle_{\pi}\right)$ sarja $\sum_{k=0}^{\infty} P^k x$ suppenee itseisesti kaikilla $x \in u_1^{\perp}$.

(ii) Sarja $\sum_{k=0}^{\infty} \left\langle x|P^k x\right\rangle_{\pi}$ suppenee itseisesti kaikilla $x \in u_1^{\perp}$. Lisäksi

$$\sum_{k=0}^{\infty} \left\langle x|P^k x\right\rangle_{\pi} = \left\langle x \middle| \sum_{k=0}^{\infty} P^k x \right\rangle_{\pi}$$

kaikilla $x \in u_1^{\perp}$.

Todistus. Todistetaan ensin sarjojen itseistä suppenemista koskevat väitteet. Olkoon $x \in u_1^{\perp}$ mikä tahansa. Koska osasummien jonot

$$\left(\sum_{k=0}^{m} \left|\left\langle x|P^k x\right\rangle_{\pi}\right|\right)_{m=1}^{\infty} \quad \text{ja} \quad \left(\sum_{k=0}^{m} \left\|P^k x\right\|_{\pi}\right)_{m=1}^{\infty}$$

ovat selvästi kasvavia, riittää osoittaa, että ne ovat myös ylhäältä rajoitettuja. Huomataan, että geometrinen sarja $\sum_{k=0}^{\infty} \lambda_*^k$ suppenee, koska $0 \leq \lambda_* < 1$. Näin ollen saadaan, että kaikilla $m \in \mathbb{N}$ on

$$\begin{aligned}
\sum_{k=0}^{m} \left|\left\langle x|P^k x\right\rangle_{\pi}\right| &\leq \sum_{k=0}^{m} \|x\|_{\pi} \left\|P^k x\right\|_{\pi} \leq \sum_{k=0}^{m} \|x\|_{\pi} \left\|P^k\right\| \|x\|_{\pi} \\
&\leq \sum_{k=0}^{m} \|x\|_{\pi}^2 \|P\|^k \leq \|x\|_{\pi}^2 \sum_{k=0}^{\infty} \lambda_*^k < \infty
\end{aligned}$$

ja

$$\sum_{k=0}^{m} \left\|P^k x\right\|_{\pi} \leq \sum_{k=0}^{m} \left\|P^k\right\| \|x\|_{\pi} \leq \|x\|_{\pi} \sum_{k=0}^{m} \|P\|^k \leq \|x\|_{\pi} \sum_{k=0}^{\infty} \lambda_*^k < \infty,$$

mistä sarjojen itseistä suppenemista koskevat väitteet seuraavat. Arvioissa hyödynnettiin Lausetta B.7, Lausetta B.1 ja Lausetta B.3.

Osoitetaan seuraavaksi, että

$$\sum_{k=0}^{\infty} \left\langle x|P^k x\right\rangle_{\pi} = \left\langle x \middle| \sum_{k=0}^{\infty} P^k x \right\rangle_{\pi}.$$

Tiedetään jo, että sarja

$$\sum_{k=0}^{\infty} P^k x$$

suppenee itseisesti avaruudessa $\left(u_1^\perp, \langle\cdot|\cdot\rangle_\pi\right)$. Huomautuksen 5.1 nojalla tästä seuraa, että sarja

$$\sum_{k=0}^{\infty} P^k x$$

suppenee avaruudessa $\left(u_1^\perp, \langle\cdot|\cdot\rangle_\pi\right)$. Näin ollen $\sum_{k=0}^{\infty} P^k x \in u_1^\perp$, joten sisätulon jatkuvuuden nojalla (Lause B.8) saadaan

$$\begin{aligned}\sum_{k=0}^{\infty} \left\langle x|P^k x\right\rangle_\pi &= \lim_{m\to\infty} \sum_{k=0}^{m} \left\langle x|P^k x\right\rangle_\pi = \lim_{m\to\infty} \left\langle x \middle| \sum_{k=0}^{m} P^k x \right\rangle_\pi \\ &= \left\langle x \middle| \lim_{m\to\infty} \sum_{k=0}^{m} P^k x \right\rangle_\pi = \left\langle x \middle| \sum_{k=0}^{\infty} P^k x \right\rangle_\pi . \qquad \square\end{aligned}$$

Seuraavan lemman todistusta varten kannattaa kerrata Kroneckerin lemma Lauseesta C.1.

Lemma 5.5. *Olkoon matriisi $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ kuten Oletuksessa 5.1. Kaikilla $x \in u_1^\perp$*

$$\lim_{k\to\infty} \sum_{m=0}^{k-1} \frac{k-m}{k} \langle x|P^m x\rangle_\pi = \sum_{m=0}^{\infty} \langle x|P^m x\rangle_\pi .$$

Todistus. Olkoon $x \in u_1^\perp$ mikä tahansa. Lemman 5.4 nojalla sarja $\sum_{k=0}^{\infty} \left\langle x|P^k x\right\rangle_\pi$ suppenee, joten

$$\lim_{k\to\infty} \left\langle x|P^k x\right\rangle_\pi = 0.$$

Lisäksi Lauseen C.1 nojalla

$$\lim_{k\to\infty} \frac{1}{k} \sum_{m=1}^{k} m \langle x|P^m x\rangle_\pi = 0,$$

joten saadaan, että

$$\begin{aligned}\lim_{k\to\infty} \frac{1}{k} \sum_{m=0}^{k-1} m \langle x|P^m x\rangle_\pi &= \lim_{k\to\infty} \left(\frac{1}{k} \sum_{m=0}^{k} m \langle x|P^m x\rangle_\pi - \frac{1}{k} k \left\langle x|P^k x\right\rangle_\pi \right) \\ &= \lim_{k\to\infty} \frac{1}{k} \sum_{m=1}^{k} m \langle x|P^m x\rangle_\pi - \lim_{k\to\infty} \left\langle x|P^k x\right\rangle_\pi \\ &= 0.\end{aligned}$$

Edelleen

$$\begin{aligned}\lim_{k\to\infty} \sum_{m=0}^{k-1} \frac{k-m}{k} \langle x|P^m x\rangle_\pi &= \lim_{k\to\infty} \left(\sum_{m=0}^{k-1} \langle x|P^m x\rangle_\pi - \frac{1}{k} \sum_{m=0}^{k-1} m \langle x|P^m x\rangle_\pi \right) \\ &= \lim_{k\to\infty} \sum_{m=0}^{k-1} \langle x|P^m x\rangle_\pi - \lim_{k\to\infty} \frac{1}{k} \sum_{m=0}^{k-1} m \langle x|P^m x\rangle_\pi \\ &= \sum_{m=0}^{\infty} \langle x|P^m x\rangle_\pi . \qquad \square\end{aligned}$$

Jatkossa funktio $f : \{s_1, s_2, \ldots, s_n\} \to \mathbb{R}$ samaistetaan vektorin

$$f = (f(s_1), f(s_2), \ldots, f(s_n)) \in \mathbb{R}^n$$

kanssa, jolloin voidaan tarkastella sisätuloa $\langle f|g\rangle_\pi$, missä myös $g \in \mathbb{R}^n$.

Lemma 5.6. *Kaikilla funktioilla $f : S = \{s_1, s_2, \ldots, s_n\} \to \mathbb{R}$ pätee, että $\bar{f} = f - \mathbb{E}_\pi f \in u_1^\perp$, missä*

$$\mathbb{E}_\pi f = \sum_{i=1}^{n} f(s_i)\pi_i.$$

Todistus. Todistus on suora lasku:

$$\begin{aligned}\left\langle \bar{f}|u_1\right\rangle_\pi &= \langle f - \mathbb{E}_\pi f|u_1\rangle_\pi = \sum_{i=1}^{n} \left(f(s_i) - \mathbb{E}_\pi f\right) \underbrace{u_{1,i}}_{=1} \pi_i \\ &= \sum_{i=1}^{n} f(s_i)\pi_i - \sum_{i=1}^{n} \left(\mathbb{E}_\pi f\right)\pi_i = \mathbb{E}_\pi f - \mathbb{E}_\pi f = 0. \qquad \square\end{aligned}$$

Lauseen 5.1 todistuksessa hyödynnetään lähteestä [8] lainattuja ideoita.

Lause 5.1. *Olkoon $S = \{s_1, s_2, \ldots, s_n\} \subset \mathbb{R}$ joukko, $f : S \to \mathbb{R}$ funktio ja $(X_k)_{k=0}^\infty$ Markovin ketju tila-avaruudella S. Jos Markovin ketjun $(X_k)_{k=0}^\infty$ siirtymätodennäköisyysmatriisi $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ on kuten Oletuksessa 5.1, niin kaikilla alkujakaumilla $X_0 \sim \mu^{(0)} \in \mathbb{R}^n$ matriisin P asymptoottinen varianssi funktiolle f (Määritelmä 3.4) on*

$$\sigma(P, f)^2 = 2\left\langle \bar{f}|(I - P)^{-1}\bar{f}\right\rangle_\pi - \left\langle \bar{f}|\bar{f}\right\rangle_\pi, \tag{19}$$

missä

(i) $\mathbb{E}_\pi f = \sum_{i=1}^{n} f(s_i)\pi_i$,

(ii) $\bar{f} = f - \mathbb{E}_\pi f$,

(iii) *I on identtinen kuvaus $I : u_1^\perp \to u_1^\perp$, $Ix = x$ ja*

(iv) *$(I - P)^{-1} : u_1^\perp \to u_1^\perp$ on funktion $(I - P) : u_1^\perp \to u_1^\perp$ käänteisfunktio.*

Todistus. Lauseen 3.5 nojalla asymptoottinen varianssi on sama kaikilla alkujakaumilla, joten voidaan olettaa, että satunnaismuuttujan X_0 jakauma on $\mu^{(0)} = \pi$. Tällöin Lauseen 3.5 nojalla

$$\sigma(P, f)^2 = \lim_{k\to\infty} \frac{1}{k}\mathbb{E}\left(\sum_{i=1}^{k} \left(f(X_i) - \mathbb{E}_\pi f\right)\right)^2.$$

Kaikilla $k \in \mathbb{N}$

$$
\begin{aligned}
\frac{1}{k}\mathbb{E}\left(\sum_{i=1}^{k}(f(X_i)-\mathbb{E}_\pi f)\right)^2 &= \frac{1}{k}\mathbb{E}\left[\left(\sum_{i=1}^{k}\bar{f}(X_i)\right)\left(\sum_{j=1}^{k}\bar{f}(X_j)\right)\right] \\
&= \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{k}\mathbb{E}\left(\bar{f}(X_i)\bar{f}(X_j)\right) \\
&= \frac{1}{k}\sum_{i=1}^{k}\left(\sum_{j=1}^{i-1}\mathbb{E}\left(\bar{f}(X_i)\bar{f}(X_j)\right)+\mathbb{E}\bar{f}(X_i)^2+\sum_{j=i+1}^{k}\mathbb{E}\left(\bar{f}(X_i)\bar{f}(X_j)\right)\right) \\
&= \frac{1}{k}\sum_{i=1}^{k}\mathbb{E}\bar{f}(X_i)^2+\frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{i-1}\mathbb{E}\left(\bar{f}(X_i)\bar{f}(X_j)\right) \\
&\quad+\frac{1}{k}\sum_{i=1}^{k}\sum_{j=i+1}^{k}\mathbb{E}\left(\bar{f}(X_i)\bar{f}(X_j)\right) \\
&= \frac{1}{k}\sum_{i=1}^{k}\mathbb{E}\bar{f}(X_i)^2+\frac{2}{k}\sum_{i=1}^{k}\sum_{j=1}^{i-1}\mathbb{E}\left(\bar{f}(X_i)\bar{f}(X_j)\right). \qquad (20)
\end{aligned}
$$

Koska $\mu^{(0)} = \pi$, niin kaikilla $k \in \mathbb{N}$ satunnaismuuttujan X_k jakauma on

$$
\mu^{(k)} = \mu^{(0)}P^k = \pi P^k = \pi.
$$

Näin ollen käyttämällä Lausetta A.1 saadaan, että kaikilla $i \geq j$

$$
\begin{aligned}
\mathbb{E}\left(\bar{f}(X_i)\bar{f}(X_j)\right) &= \sum_{v=1}^{n}\sum_{u=1}^{n}\bar{f}(s_v)\bar{f}(s_u)\mathbb{P}\left(X_i=s_v,X_j=s_u\right) \\
&= \sum_{v=1}^{n}\sum_{u:\mathbb{P}(X_j=s_u)>0}\bar{f}(s_v)\bar{f}(s_u)\mathbb{P}\left(X_i=s_v,X_j=s_u\right) \\
&= \sum_{v=1}^{n}\sum_{u:\mathbb{P}(X_j=s_u)>0}\bar{f}(s_v)\bar{f}(s_u)\underbrace{\mathbb{P}\left(X_i=s_v|X_j=s_u\right)}_{=p_{u,v}^{(i-j)}}\mathbb{P}\left(X_j=s_u\right) \\
&= \sum_{v=1}^{n}\sum_{u:\mathbb{P}(X_j=s_u)>0}\bar{f}(s_v)\bar{f}(s_u)p_{u,v}^{(i-j)}\mathbb{P}\left(X_j=s_u\right) \\
&= \sum_{v=1}^{n}\sum_{u=1}^{n}\bar{f}(s_v)\bar{f}(s_u)p_{u,v}^{(i-j)}\underbrace{\mathbb{P}\left(X_j=s_u\right)}_{=\pi_u} \\
&= \sum_{v=1}^{n}\sum_{u=1}^{n}\bar{f}(s_v)\bar{f}(s_u)p_{u,v}^{(i-j)}\pi_u \\
&= \sum_{u=1}^{n}\bar{f}(s_u)\pi_u\underbrace{\sum_{v=1}^{n}p_{u,v}^{(i-j)}\bar{f}(s_v)}_{=\left[P^{i-j}\bar{f}\right]_u} \\
&= \left\langle\bar{f}|P^{i-j}\bar{f}\right\rangle_\pi. \qquad (21)
\end{aligned}
$$

Lisäksi kaikilla $i \in \mathbb{N}$

$$\mathbb{E}\bar{f}(X_i)^2 = \sum_{j=1}^{n} \bar{f}(s_j)^2 \mathbb{P}(X_i = s_j) = \sum_{j=1}^{n} \bar{f}(s_j)^2 \pi_j = \langle \bar{f} | \bar{f} \rangle_\pi ,$$

joten yhtälöistä (20) ja (21) nähdään, että

$$\begin{aligned}
\frac{1}{k}\mathbb{E}\left(\sum_{i=1}^{k}(f(X_i) - \mathbb{E}_\pi f)\right)^2 &= \frac{1}{k}\sum_{i=1}^{k}\langle \bar{f}|\bar{f}\rangle_\pi + \frac{2}{k}\sum_{i=1}^{k}\sum_{j=1}^{i-1}\langle \bar{f}|P^{i-j}\bar{f}\rangle_\pi \\
&= \langle \bar{f}|\bar{f}\rangle_\pi + \frac{2}{k}\sum_{i=1}^{k}\sum_{j=1}^{i-1}\langle \bar{f}|P^{i-j}\bar{f}\rangle_\pi \\
&= \langle \bar{f}|\bar{f}\rangle_\pi + \frac{2}{k}\sum_{i=1}^{k}\sum_{j=1}^{i-1}\langle \bar{f}|P^{j}\bar{f}\rangle_\pi \\
&= \langle \bar{f}|\bar{f}\rangle_\pi + \frac{2}{k}\sum_{j=1}^{k-1}(k-j)\langle \bar{f}|P^{j}\bar{f}\rangle_\pi \\
&= \langle \bar{f}|\bar{f}\rangle_\pi + 2\sum_{j=1}^{k-1}\frac{k-j}{k}\langle \bar{f}|P^{j}\bar{f}\rangle_\pi . \qquad (22)
\end{aligned}$$

Lemman 5.6 nojalla $\bar{f} \in u_1^\perp$, joten soveltamalla Lemmaa 5.5 saadaan yhtälöstä (22), että

$$\begin{aligned}
\sigma(P,f)^2 &= \lim_{k\to\infty}\frac{1}{k}\mathbb{E}\left(\sum_{i=1}^{k}(f(X_i) - \mathbb{E}_\pi f)\right)^2 = \langle \bar{f}|\bar{f}\rangle_\pi + 2\lim_{k\to\infty}\sum_{j=1}^{k-1}\frac{k-j}{k}\langle \bar{f}|P^{j}\bar{f}\rangle_\pi \\
&= \langle \bar{f}|\bar{f}\rangle_\pi + 2\sum_{j=1}^{\infty}\langle \bar{f}|P^{j}\bar{f}\rangle_\pi \\
&= \langle \bar{f}|\bar{f}\rangle_\pi + 2\left(\sum_{j=0}^{\infty}\langle \bar{f}|P^{j}\bar{f}\rangle_\pi - \underbrace{\langle \bar{f}|\bar{f}\rangle_\pi}_{=\langle \bar{f}|P^0\bar{f}\rangle_\pi}\right) \\
&= 2\sum_{j=0}^{\infty}\langle \bar{f}|P^{j}\bar{f}\rangle_\pi - \langle \bar{f}|\bar{f}\rangle_\pi ,
\end{aligned}$$

ja Lemman 5.4 nojalla edelleen, että

$$\sigma(P,f)^2 = 2\left\langle \bar{f}\,\middle|\,\sum_{j=0}^{\infty}P^j\bar{f}\right\rangle_\pi - \langle \bar{f}|\bar{f}\rangle_\pi . \qquad (23)$$

Lemman 5.3 nojalla kuvauksen $P : u_1^\perp \to u_1^\perp$ operaattorinormi on $\|P\| = \lambda_* < 1$, joten Lauseen B.6 nojalla yhtälö (23) voidaan kirjoittaa muotoon

$$\sigma(P,f)^2 = 2\langle \bar{f}|(I-P)^{-1}\bar{f}\rangle_\pi - \langle \bar{f}|\bar{f}\rangle_\pi . \qquad \square$$

Asymptoottisen varianssin $\sigma(P,f)^2$ laskeminen yhtälöä (19) käyttäen on työlästä yhtälössä esiintyvän termin $(I-P)^{-1}\bar{f}$ vuoksi. Tietyin edellytyksin yhtälö (19) mahdollistaa kuitenkin asymptoottisten varianssien $\sigma(P,f)^2$ ja $\sigma(Q,f)^2$ suuruuksien *vertailun*. Tähän tarvitaan Lausetta B.12, jonka avulla yhtälö (19) voidaan kirjoittaa muotoon

$$\sigma(P,f)^2 = 2\sup_{g\in u_1^\perp}\left(2\left\langle \bar{f}|g\right\rangle_\pi - \langle g|(I-P)g\rangle_\pi\right) - \left\langle \bar{f}|\bar{f}\right\rangle_\pi.$$

Näin päästään eroon ongelmallisesta termistä $(I-P)^{-1}\bar{f}$ ja asymptoottisten varianssien vertailuun riittää termien

$$\langle g|(I-P)g\rangle_\pi \text{ ja } \langle g|(I-Q)g\rangle_\pi$$

tutkiminen. Tämä puolestaan on helpompaa kuin äkkiseltään katsottuna voisi kuvitella: Lemmassa 5.8 osoitetaan, että riittä tutkia matriisien P ja Q alkioita. Aloitetaan kuitenkin tarkistamalla Lauseen B.12 oletusten voimassaolo. Asymptoottisten varianssien vertailu tehdään Lauseessa 5.2.

Lemma 5.7. *Olkoon matriisi $P=[p_{i,j}]_{i,j=1,2,\dots,n}$ kuten Oletuksessa 5.1. Lineaarikuvaus $(I-P)$: $u_1^\perp \to u_1^\perp$ on bijektio siten, että kaikilla $x,y\in u_1^\perp$*

(i) $\langle x|(I-P)y\rangle_\pi = \langle (I-P)x|y\rangle_\pi$ *ja*

(ii) $\langle x|(I-P)x\rangle_\pi \geq 0.$

Todistus. Tarkistetaan ensin, että kuvaus $(I-P):u_1^\perp \to u_1^\perp$ on hyvin määritelty, eli että $(I-P)x\in u_1^\perp$ kaikilla $x\in u_1^\perp$. Olkoon $x\in u_1^\perp$ mikä tahansa. Lemman 5.2 nojalla $Px\in u_1^\perp$ ja Huomautuksen 5.1 nojalla $u_1^\perp$ on avaruuden $(\mathbb{R}^n, \langle\cdot|\cdot\rangle_\pi)$ suljettu aliavaruus, joten saadaan, että

$$(I-P)x = Ix - Px = x - Px \in u_1^\perp.$$

Huomautuksen 5.1 nojalla $u_1^\perp$ on Banachin avaruus ja Lemman 5.3 nojalla kuvauksen $P:u_1^\perp \to u_1^\perp$ operaattorinormi on $\|P\| = \lambda_* < 1$, joten kuvauksen $I-P$ bijektiivisyys seuraa Lauseesta B.6.

Olkoot $x,y\in u_1^\perp$ mitä tahansa. Samoin kuin Lauseen 4.2 todistuksessa nähdään, että

$$\langle x|Py\rangle_\pi = \langle Px|y\rangle_\pi.$$

Näin ollen saadaan, että

$$\langle x|(I-P)y\rangle_\pi = \langle x|y\rangle_\pi - \langle x|Py\rangle_\pi = \langle x|y\rangle_\pi - \langle Px|y\rangle_\pi = \langle x-Px|y\rangle_\pi = \langle (I-P)x|y\rangle_\pi.$$

Osoitetaan lopuksi, että $\langle x|(I-P)x\rangle_\pi \geq 0$ kaikilla $x\in u_1^\perp$. Lauseen B.1 ja Lemman 5.3 nojalla

$$\langle x|Px\rangle_\pi \leq |\langle x|Px\rangle_\pi| \leq \|x\|_\pi \|Px\|_\pi \leq \|x\|_\pi\|P\|\|x\|_\pi = \|x\|_\pi^2\lambda_* \leq \|x\|_\pi^2,$$

joten saadaan, että

$$\langle x|(I-P)x\rangle_\pi = \langle x|x\rangle_\pi - \langle x|Px\rangle_\pi \geq \langle x|x\rangle_\pi - \|x\|_\pi^2 = \|x\|_\pi^2 - \|x\|_\pi^2 = 0. \qquad \square$$

Kirjoitetaan seuraavaksi sisätulon $\langle x|(I-P)x\rangle_\pi$ lauseke vertailun kannalta käyttökelpoisempaan muotoon.

Lemma 5.8. *Olkoon matriisi* $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ *kuten Oletuksessa 5.1. Tällöin kaikilla* $x \in \mathbb{R}^n$

$$\langle x|(I-P)x\rangle_\pi = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2 \pi_i p_{i,j}.$$

Todistus. Kaikilla $x \in \mathbb{R}^n$ pätee, että

$$\begin{aligned}
\langle x|(I-P)x\rangle_\pi &= \langle x|x\rangle_\pi - \langle x|Px\rangle_\pi \\
&= \sum_{i=1}^{n} x_i^2 \pi_i - \sum_{i=1}^{n} x_i \underbrace{\left(\sum_{j=1}^{n} p_{i,j}x_j\right)}_{=[Px]_i} \pi_i \\
&= \sum_{i=1}^{n}\left(x_i^2\pi_i - x_i \sum_{j=1}^{n} p_{i,j}x_j\pi_i\right) \\
&= \sum_{i=1}^{n}\left(\sum_{j=1}^{n} p_{i,j}x_i^2\pi_i - \sum_{j=1}^{n} x_i x_j p_{i,j}\pi_i\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i^2 - x_i x_j)p_{i,j}\pi_i.
\end{aligned}$$

Toisaalta

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2 p_{i,j}\pi_i &= \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i^2 - 2x_i x_j + x_j^2)p_{i,j}\pi_i \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i^2 - x_i x_j)\pi_i p_{i,j} + \sum_{i=1}^{n}\sum_{j=1}^{n}(x_j^2 - x_i x_j)\underbrace{\pi_i p_{i,j}}_{=\pi_j p_{j,i}} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i^2 - x_i x_j)\pi_i p_{i,j} + \sum_{j=1}^{n}\sum_{i=1}^{n}(x_j^2 - x_j x_i)\pi_j p_{j,i} \\
&= 2\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i^2 - x_i x_j)\pi_i p_{i,j},
\end{aligned}$$

joten

$$\begin{aligned}
\langle x|(I-P)x\rangle_\pi &= \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i^2 - x_i x_j)p_{i,j}\pi_i \\
&= \frac{1}{2}\cdot 2\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i^2 - x_i x_j)\pi_i p_{i,j} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2 p_{i,j}\pi_i. \qquad \square
\end{aligned}$$

Asymptoottisten varianssien vertailuun tarvittavat aputulokset on nyt todistettu. Määritellään vielä Peskunin järjestys ennen vertailutuloksen esittämistä.

Määritelmä 5.1. *Olkoot* $P = [p_{i,j}]_{i,j=1,2,\dots,n}$ *ja* $Q = [q_{i,j}]_{i,j=1,2,\dots,n}$ *siirtymätodennäköisyysmatriiseja. Jos* $p_{i,j} \geq q_{i,j}$ *kaikilla* $i \neq j$*, niin merkitään*

$$P \succeq Q.$$

Relaatiota $\succeq$ *kutsutaan Peskunin järjestykseksi.*

Huomautus 5.2. *Peskunin järjestys* $\succeq$ *on osittainen järjestys siirtymätodennäköisyysmatriiseille:*

(i) $P \succeq P$

(ii) jos $P \succeq Q$ *ja* $Q \succeq P$*, niin* $P = Q$

(iii) jos $P \succeq Q$ *ja* $R \succeq P$*, niin* $R \succeq Q$*.*

Seuraava lause kertoo, miten Peskunin järjestys liittyy asymptoottisten varianssien vertailuun. Tuloksen todisti ensimmäisenä Peter H. Peskun ja se julkaistiin ensimmäisen kerran vuonna 1973 lähteessä [9]. Seuraavaksi esitettävä todistus poikkeaa Peskunin alkuperäisestä todistuksesta.

Lause 5.2. *Olkoon* $S = \{s_1, s_2, \dots, s_n\} \subset \mathbb{R}$*,* $f : S \to \mathbb{R}$ *funktio sekä* $P = [p_{i,j}]_{i,j=1,2,\dots,n}$ *ja* $Q = [q_{i,j}]_{i,j=1,2,\dots,n}$ *siirtymätodennäköisyysmatriiseja. Oletetaan, että*

(i) $P^l > 0$ *ja* $Q^m > 0$ *joillakin* $l, m \in \mathbb{N}$

(ii) matriiseilla P *ja* Q *on sama invariantti jakauma* $\pi \in \mathbb{R}^n$

(iii) kumpikin matriiseista P *ja* Q *on kääntyvä invariantin jakauman* π *suhteen.*

Tällöin, jos

$$P \succeq Q,$$

niin

$$\sigma(P, f)^2 \leq \sigma(Q, f)^2.$$

Todistus. Lemman 5.8 nojalla oletuksesta

$$P \succeq Q$$

seuraa, että

$$\begin{aligned}
\langle f|(I-Q)f\rangle_\pi &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(f(s_i)-f(s_j))^2\pi_i q_{i,j} \\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}(f(s_i)-f(s_j))^2\pi_i q_{i,j} \\
&\leq \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}(f(s_i)-f(s_j))^2\pi_i p_{i,j} \\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(f(s_i)-f(s_j))^2\pi_i p_{i,j} \\
&= \langle f|(I-P)f\rangle_\pi\,,
\end{aligned}$$

joten käyttämällä lauseita 5.1 ja B.12 saadaan, että

$$\begin{aligned}\sigma(P,f)^2 &= 2\left\langle \bar{f}|(I-P)^{-1}\bar{f}\right\rangle_\pi - \left\langle \bar{f}|\bar{f}\right\rangle_\pi \\ &= 2\sup_{g\in u_1^\perp}\left(2\left\langle \bar{f}|g\right\rangle_\pi - \langle g|(I-P)g\rangle_\pi\right) - \left\langle \bar{f}|\bar{f}\right\rangle_\pi \\ &\le 2\sup_{g\in u_1^\perp}\left(2\left\langle \bar{f}|g\right\rangle_\pi - \langle g|(I-Q)g\rangle_\pi\right) - \left\langle \bar{f}|\bar{f}\right\rangle_\pi \\ &= 2\left\langle \bar{f}|(I-Q)^{-1}\bar{f}\right\rangle_\pi - \left\langle \bar{f}|\bar{f}\right\rangle_\pi \\ &= \sigma(Q,f)^2.\end{aligned}$$

Huomaa, että Lemman 5.7 nojalla kuvaukset $I-P : u_1^\perp \to u_1^\perp$ ja $I-Q : u_1^\perp \to u_1^\perp$ toteuttavat Lauseen B.12 oletukset ja lisäksi Lemman 5.6 nojalla $\bar{f} \in u_1^\perp$. Näin ollen Lauseen B.12 käyttö edellä on sallittua. □

Lauseen 5.2 sovelluksena osoitetaan, että Metropolis–Hastings -algoritmi on parempi Markovin ketju Monte Carlo -simulointiin kuin Barkerin algoritmi. Tulos on sama, jonka Peskun esitti lähteessä [9], vaikka käsittely poikkeaa hieman Peskunin esityksestä.

Lause 5.3. *Olkoon $S = \{s_1, s_2, \dots, s_n\} \subset \mathbb{R}$ joukko, $f : S \to \mathbb{R}$ funktio ja $\pi \in \mathbb{R}^n$ jakauma siten, että $\pi > 0$. Kaikilla ehdokastilojen jakaumilla $g_i = (g_{i,1}, g_{i,2}, \dots, g_{i,n}) \in \mathbb{R}^n$, joille $g_i > 0$ kaikilla $i = 1, 2, \dots, n$ pätee, että*

$$\sigma(P_{MH}, f)^2 \le \sigma(P_B, f)^2,$$

missä $P_{MH} = \left[p_{i,j}^{MH}\right]_{i,j=1,2,\dots,n}$ on kuten määritelmässä 3.2 ja $P_B = \left[p_{i,j}^{B}\right]_{i,j=1,2,\dots,n}$ on kuten määritelmässä 3.3.

Todistus. Kiinnitetään ehdokastilojen jakaumat $g_i = (g_{i,1}, g_{i,2}, \dots, g_{i,n}) \in \mathbb{R}^n$ siten, että $g_i > 0$ kaikilla $i = 1, 2, \dots, n$. Lauseen 5.2 nojalla riittää osoittaa, että $P_{\mathrm{MH}} \succeq P_{\mathrm{B}}$. Olkoot $i, j \in \{1, 2, \dots, n\}$ mitä tahansa siten, että $i \ne j$. Määritelmän 3.2 mukaan

$$p_{i,j}^{\mathrm{MH}} = \min\left\{1, \frac{\pi_j g_{j,i}}{\pi_i g_{i,j}}\right\} g_{i,j}$$

ja määritelmän 3.3 mukaan

$$p_{i,j}^{\mathrm{B}} = \frac{\pi_j g_{j,i}}{\pi_i g_{i,j} + \pi_j g_{j,i}} g_{i,j},$$

joten riittää osoittaa, että

$$\min\left\{1, \frac{\pi_j g_{j,i}}{\pi_i g_{i,j}}\right\} \ge \frac{\pi_j g_{j,i}}{\pi_i g_{i,j} + \pi_j g_{j,i}}.$$

Jos

$$\min\left\{1, \frac{\pi_j g_{j,i}}{\pi_i g_{i,j}}\right\} = 1,$$

niin

$$\frac{\pi_j g_{j,i}}{\pi_i g_{i,j} + \pi_j g_{j,i}} \le \frac{\pi_j g_{j,i}}{\pi_j g_{j,i}} = 1 = \min\left\{1, \frac{\pi_j g_{j,i}}{\pi_i g_{i,j}}\right\}.$$

Jos taas

$$\min\left\{1, \frac{\pi_j g_{j,i}}{\pi_i g_{i,j}}\right\} = \frac{\pi_j g_{j,i}}{\pi_i g_{i,j}},$$

niin yhtä triviaalisti

$$\frac{\pi_j g_{j,i}}{\pi_i g_{i,j} + \pi_j g_{j,i}} \le \frac{\pi_j g_{j,i}}{\pi_i g_{i,j}} = \min\left\{1, \frac{\pi_j g_{j,i}}{\pi_i g_{i,j}}\right\}.$$

□

Liitteet

A Markovin ketjujen perustuloksia

Lemma A.1. *Olkoon* $(\Omega, \mathcal{F}, \mathbb{P})$ *todennäköisyysavaruus,* $n \in \mathbb{N}$ *ja*

$$A_0, A_1, \dots, A_n \in \mathcal{F}$$

siten, että

$$\mathbb{P}\left(A_{n-1} \cap A_{n-2} \cap ... \cap A_1 \cap A_0\right) > 0.$$

Tällöin

$$\mathbb{P}\left(\cap_{i=1}^n A_i | A_0\right) = \prod_{i=1}^n \mathbb{P}\left(A_i | A_{i-1}, A_{i-2}, \dots, A_0\right). \tag{24}$$

Todistus. Kaikilla $i = 1, 2, \dots, n$

$$\mathbb{P}\left(A_{i-1}, A_{i-2}, \dots, A_1, A_0\right) \geq \mathbb{P}\left(A_{n-1}, A_{n-2}, \dots, A_1, A_0\right) > 0,$$

joten kaikki väitteessä esiintyvät ehdolliset todennäköisyydet on määritelty. Jos $n = 1$, niin väite pätee triviaalisti.

Oletetaan induktio-oletuksena, että

$$\mathbb{P}\left(\cap_{i=1}^n A_i | A_0\right) = \prod_{i=1}^n \mathbb{P}\left(A_i | A_{i-1}, A_{i-2}, \dots, A_1, A_0\right)$$

jollakin $n \in \mathbb{N}$ aina, kun $A_0, A_1, A_2, \dots, A_n \in \mathcal{F}$ siten, että

$$\mathbb{P}\left(A_{n-1} \cap A_{n-2} \cap ... \cap A_1 \cap A_0\right) > 0.$$

Olkoot

$$A_0, A_1, A_2, \dots, A_{n+1} \in \mathcal{F}$$

siten, että

$$\mathbb{P}\left(A_n \cap A_{n-1} \cap ... \cap A_1 \cap A_0\right) > 0.$$

Saadaan, että

$$\begin{aligned}
\mathbb{P}\left(\cap_{i=1}^{n+1} A_i | A_0\right) &= \frac{\mathbb{P}\left(A_{n+1} \cap \left(\cap_{i=0}^n A_i\right)\right)}{\mathbb{P}\left(A_0\right)} \\
&= \frac{\mathbb{P}\left(A_{n+1} | A_n, A_{n-1}, \dots, A_1, A_0\right) \mathbb{P}\left(\cap_{i=0}^n A_i\right)}{\mathbb{P}\left(A_0\right)} \\
&= \mathbb{P}\left(A_{n+1} | A_n, A_{n-1}, \dots, A_1, A_0\right) \frac{\mathbb{P}\left(\cap_{i=0}^n A_i\right)}{\mathbb{P}\left(A_0\right)} \\
&= \mathbb{P}\left(A_{n+1} | A_n, A_{n-1}, \dots, A_1, A_0\right) \mathbb{P}\left(\cap_{i=1}^n A_i | A_0\right) \\
&= \mathbb{P}\left(A_{n+1} | A_n, A_{n-1}, \dots, A_1, A_0\right) \prod_{i=1}^n \mathbb{P}\left(A_i | A_{i-1}, A_{i-2}, \dots, A_1, A_0\right) \\
&= \prod_{i=1}^{n+1} \mathbb{P}\left(A_i | A_{i-1}, A_{i-2}, \dots, A_1, A_0\right),
\end{aligned}$$

joten väite seuraa induktioperiaatteesta. □

Lemma A.2. *Olkoon* $(X_k)_{k=0}^{\infty}$ *Markovin ketju tila-avaruudella* $S = \{s_1, s_2, \dots, s_n\}$. *Tällöin myös* $(X_{m+k})_{k=0}^{\infty}$ *on Markovin ketju kaikilla* $m \in \mathbb{N}_0$.

Todistus. Tapauksessa $m = 0$ väite on selvä oletuksen nojalla. Oletetaan induktio-oletuksena, että $(X_{m+k})_{k=0}^{\infty}$ on Markovin ketju jollakin $m \in \mathbb{N}_0$. Induktioväitteenä on, että $\left(X_{(m+1)+k}\right)_{k=0}^{\infty}$ on Markovin ketju. Olkoon $k \in \mathbb{N}$ ja $x_{m+1}, x_{m+2}, \dots, x_{m+k+1} \in S$ siten, että

$$\mathbb{P}\left(X_{m+k+1} = x_{m+k+1}, X_{m+k} = x_{m+k}, \dots, X_{m+1} = x_{m+1}\right) > 0.$$

Merkitään

$$T = \{s \in S : \mathbb{P}\left(X_{m+k+1} = x_{m+k+1}, X_{m+k} = x_{m+k}, \dots, X_m = s\right) > 0\}$$

ja

$$B = \bigcap_{l=1}^{k+1} \{X_{m+l} = x_{m+l}\}.$$

Koska

$$\begin{aligned} 0 < \mathbb{P}(B) &= \mathbb{P}\left(B \cap \left(\bigcup_{s \in S} \{X_m = s\}\right)\right) = \mathbb{P}\left(\bigcup_{s \in S} (B \cap \{X_m = s\})\right) \\ &= \sum_{s \in S} \mathbb{P}\left(B \cap \{X_m = s\}\right) = \sum_{s \in T} \mathbb{P}\left(B \cap \{X_m = s\}\right), \end{aligned}$$

niin $T \neq \varnothing$ ja $\sum_{s \in T} \mathbb{P}\left(B \cap \{X_m = s\}\right) = \mathbb{P}(B)$.

Kiinnitetään $y \in S$ ja merkitään $A = \{X_{m+k+2} = y\}$. Koska $(X_{m+k})_{k=0}^{\infty}$ on induktio-oletuksen nojalla Markovin ketju, niin kaikilla $s \in T$

$$\begin{aligned} \mathbb{P}\left(A | B \cap \{X_m = s\}\right) &= \mathbb{P}\left(X_{m+k+2} = y | X_{m+k+1} = x_{m+k+1}, \dots, X_m = s\right) \\ &= \mathbb{P}\left(X_{m+k+2} = y | X_{m+k+1} = x_{m+k+1}\right), \end{aligned}$$

sillä $\mathbb{P}\left(X_{m+k+1} = x_{m+k+1}, \dots, X_m = s\right) > 0$.

Näin ollen saadaan, että

$$
\begin{aligned}
&\mathbb{P}\left(X_{m+k+2}=y|X_{m+k+1}=x_{m+k+1},X_{m+k}=x_{m+k},\ldots,X_{m+1}=x_{m+1}\right)\\
=&\frac{\mathbb{P}\left(X_{m+k+2}=y,X_{m+k+1}=x_{m+k+1},X_{m+k}=x_{m+k},\ldots,X_{m+1}=x_{m+1}\right)}{\mathbb{P}\left(X_{m+k+1}=x_{m+k+1},X_{m+k}=x_{m+k},\ldots,X_{m+1}=x_{m+1}\right)}\\
=&\frac{\mathbb{P}\left(A\cap B\right)}{\mathbb{P}\left(B\right)}\\
=&\frac{\mathbb{P}\left(\bigcup_{s\in S}\left(A\cap B\cap\{X_m=s\}\right)\right)}{\mathbb{P}\left(B\right)}\\
=&\frac{\sum_{s\in S}\mathbb{P}\left(A\cap B\cap\{X_m=s\}\right)}{\mathbb{P}\left(B\right)}\\
=&\frac{\sum_{s\in T}\mathbb{P}\left(A\cap B\cap\{X_m=s\}\right)}{\mathbb{P}\left(B\right)}\\
=&\frac{\sum_{s\in T}\mathbb{P}\left(A|B\cap\{X_m=s\}\right)\mathbb{P}\left(B\cap\{X_m=s\}\right)}{\mathbb{P}\left(B\right)}\\
=&\frac{\sum_{s\in T}\mathbb{P}\left(X_{m+k+2}=y|X_{m+k+1}=x_{m+k+1}\right)\mathbb{P}\left(B\cap\{X_m=s\}\right)}{\mathbb{P}\left(B\right)}\\
=&\mathbb{P}\left(X_{m+k+2}=y|X_{m+k+1}=x_{m+k+1}\right)\frac{\sum_{s\in T}\mathbb{P}\left(B\cap\{X_m=s\}\right)}{\mathbb{P}\left(B\right)}\\
=&\mathbb{P}\left(X_{m+k+2}=y|X_{m+k+1}=x_{m+k+1}\right)\frac{\mathbb{P}\left(B\right)}{\mathbb{P}\left(B\right)}\\
=&\mathbb{P}\left(X_{m+k+2}=y|X_{m+k+1}=x_{m+k+1}\right).
\end{aligned}
$$

□

Lemma A.3. *Olkoon* $(X_k)_{k=0}^{\infty}$ *Markovin ketju tila-avaruudella* $S=\{s_1,s_2,\ldots,s_n\}$*. Jos*

$$x_m,x_{m+1},\ldots,x_{m+k}\in S$$

siten, että

$$\mathbb{P}\left(X_m=x_m\right)>0$$

ja

$$\mathbb{P}\left(X_{m+k}=x_{m+k},X_{m+(k-1)}=x_{m+(k-1)},\ldots,X_m=x_m\right)=0,$$

niin on olemassa $i\in\{1,2,\ldots,k\}$ *siten, että*

$$\mathbb{P}\left(X_{m+i}=x_{m+i}|X_{m+(i-1)}=x_{m+(i-1)}\right)=0.$$

Todistus. Jos $k=1$, niin

$$0=\mathbb{P}\left(X_{m+1}=x_{m+1},X_m=x_m\right)=\mathbb{P}\left(X_{m+1}=x_{m+1}|X_m=x_m\right)\mathbb{P}\left(X_m=x_m\right),$$

mistä

$$\mathbb{P}\left(X_{m+1}=x_{m+1}|X_m=x_m\right)=0.$$

Oletetaan induktio-oletuksena, että väite pätee jollakin $k\in\mathbb{N}$. Olkoot $x_m,x_{m+1},\ldots,x_{m+(k+1)}\in S$ siten, että

$$\mathbb{P}\left(X_m=x_m\right)>0$$

ja

$$\mathbb{P}\left(X_{m+(k+1)} = x_{m+(k+1)}, X_{m+k} = x_{m+k}, \ldots, X_m = x_m\right) = 0.$$

Jos

$$\mathbb{P}\left(X_{m+k} = x_{m+k}, X_{m+(k-1)}, \ldots, X_m = x_m\right) = 0,$$

niin väite seuraa suoraan induktio-oletuksesta. Oletetaan, että

$$\mathbb{P}\left(X_{m+k} = x_{m+k}, X_{m+(k-1)}, \ldots, X_m = x_m\right) > 0.$$

Käyttämällä Lemmaa A.2 saadaan

$$\begin{aligned}
0 &= \mathbb{P}\left(X_{m+(k+1)} = x_{m+(k+1)}, X_{m+k} = x_{m+k}, \ldots, X_m = x_m\right) \\
&\mathbb{P}\left(X_{m+(k+1)} = x_{m+(k+1)} | X_{m+k} = x_{m+k}, \ldots, X_m = x_m\right) \mathbb{P}\left(X_{m+k} = x_{m+k}, \ldots, X_m = x_m\right) \\
&\mathbb{P}\left(X_{m+(k+1)} = x_{m+(k+1)} | X_{m+k} = x_{m+k}\right) \mathbb{P}\left(X_{m+k} = x_{m+k}, \ldots, X_m = x_m\right),
\end{aligned}$$

mistä

$$\mathbb{P}\left(X_{m+(k+1)} = x_{m+(k+1)} | X_{m+k} = x_{m+k}\right) = 0$$

ja väite seuraa. □

Lause A.1. *Olkoon* $(X_k)_{k=0}^{\infty}$ *Markovin ketju tila-avaruudella* $S = \{s_1, s_2, \ldots, s_n\}$ *ja olkoon* $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ *ketjun siirtymätodennäköisyysmatriisi. Jos*

$$\mathbb{P}\left(X_m = s_i\right) > 0,$$

niin

$$\mathbb{P}\left(X_{m+k} = s_j | X_m = s_i\right) = p_{i,j}^{(k)} \text{ kaikilla } k \in \mathbb{N} \text{ ja kaikilla } j = 1, 2, \ldots, n.$$

Todistus. Voidaan olettaa, että $k > 1$. Olkoot $s_{i_m}, s_{i_{m+1}}, \ldots, s_{i_{m+k}} \in S$ mitä tahansa siten, että

$$\mathbb{P}\left(X_m = s_{i_m}\right) > 0.$$

Oletetaan ensin, että

$$\mathbb{P}\left(X_{m+k} = s_{i_{m+k}}, X_{m+(k-1)} = s_{i_{m+(k-1)}}, \ldots, X_m = s_{i_m}\right) > 0.$$

Tällöin Lemmojen A.1 ja A.2 nojalla

$$\begin{aligned}
&\mathbb{P}\left(X_{m+k} = s_{i_{m+k}}, X_{m+(k-1)} = s_{i_{m+(k-1)}}, \ldots, X_{m+1} = s_{i_{m+1}} | X_m = s_{i_m}\right) \\
&= \prod_{j=1}^{k} \mathbb{P}\left(X_{m+j} = s_{i_{m+j}} | X_{m+(j-1)} = s_{i_{m+(j-1)}}, \ldots, X_m = s_{i_m}\right) \\
&= \prod_{j=1}^{k} \mathbb{P}\left(X_{m+j} = s_{i_{m+j}} | X_{m+(j-1)} = s_{i_{m+(j-1)}}\right) \\
&= \prod_{j=1}^{k} p_{i_{m+(j-1)}, i_{m+j}}.
\end{aligned} \tag{25}$$

Huomataan, että yhtälö (25) pätee kaikilla $s_{i_m}, s_{i_{m+1}}, \ldots, s_{i_{m+k}} \in S$, joille

$$\mathbb{P}\left(X_m = s_{i_m}\right) > 0:$$

Oletetaan, että onkin

$$\mathbb{P}\left(X_{m+k} = s_{i_{m+k}}, X_{m+(k-1)} = s_{i_{m+(k-1)}}, \ldots, X_m = s_{i_m}\right) = 0.$$

Tällöin yhtälön (25) vasen puoli on selvästi yhtä suuri kuin nolla. Toisaalta Lemman A.3 nojalla on olemassa $j = 1, 2, \ldots, k$ siten, että

$$p_{i_{m+(j-1)}, i_{m+j}} = \mathbb{P}\left(X_{m+j} = s_{i_{m+j}} | X_{m+(j-1)} = s_{i_{m+(j-1)}}\right) = 0,$$

joten myös yhtälön (25) oikea puoli on nolla.

Lopuksi huomataan, että

$$\begin{aligned}
&\mathbb{P}\left(X_{m+k} = s_{i_{m+k}} | X_m = s_{i_m}\right) \\
&= \sum_{i_{m+(k-1)}=1}^{n} \cdots \sum_{i_{m+1}=1}^{n} \mathbb{P}\left(X_{m+k} = s_{i_{m+k}}, X_{m+(k-1)} = s_{i_{m+(k-1)}}, \ldots, X_{m+1} = s_{i_{m+1}} | X_m = s_{i_m}\right) \\
&= \sum_{i_{m+(k-1)}=1}^{n} \cdots \sum_{i_{m+1}=1}^{n} \prod_{j=1}^{k} p_{i_{m+(j-1)}, i_{m+j}} \\
&= p^{(k)}_{i_m, i_{m+k}},
\end{aligned}$$

missä viimeinen yhtäsuuruus todistetaan induktiolla luvun k suhteen. □

Lemma A.4. *Olkoon $(X_k)_{k=0}^{\infty}$ Markovin ketju tila-avaruudella $S = \{s_1, s_2, \ldots, s_n\}$. Tällöin*

$$\begin{aligned}
&\mathbb{P}\left(X_j = x_j, X_{j-1} = x_{j-1}, \ldots, X_{i+1} = x_{i+1} | X_i = x_i, X_{i-1} = x_{i-1}, \ldots, X_0 = x_0\right) \\
=&\mathbb{P}\left(X_j = x_j, X_{j-1} = x_{j-1}, \ldots, X_{i+1} = x_{i+1} | X_i = x_i\right)
\end{aligned}$$

kaikilla $x_0, x_1, \ldots, x_j \in S$, joille

$$\mathbb{P}\left(X_{j-1} = x_{j-1}, X_{j-2} = x_{j-2}, \ldots, X_0 = x_0\right) > 0.$$

Todistus. Lemman A.1 nojalla

$$\begin{aligned}
&\mathbb{P}\left(X_j = x_j, X_{j-1} = x_{j-1}, \ldots, X_{i+1} = x_{i+1} | X_i = x_i, X_{i-1} = x_{i-1}, \ldots, X_0 = x_0\right) \\
&= \prod_{k=1}^{j-i} \mathbb{P}\left(X_{i+k} = x_{i+k} | X_{i+(k-1)} = x_{i+(k-1)}, \ldots, X_i = x_i, \ldots, X_0 = x_0\right) \\
&= \prod_{k=1}^{j-i} \mathbb{P}\left(X_{i+k} = x_{i+k} | X_{i+(k-1)} = x_{i+(k-1)}\right).
\end{aligned}$$

Toisaalta Lemman A.2 nojalla myös $(X_{i+k})_{k=0}^{\infty}$ on Markovin ketju, joten Lemman A.1 avulla nähdään, että myös

$$\begin{aligned}
&\mathbb{P}\left(X_j = x_j, X_{j-1} = x_{j-1}, \ldots, X_{i+1} = x_{i+1} | X_i = x_i\right) \\
&= \prod_{k=1}^{j-i} \mathbb{P}\left(X_{i+k} = x_{i+k} | X_{i+(k-1)} = x_{i+(k-1)}, X_{i+(k-2)} = x_{i+(k-2)}, \ldots, X_i = x_i\right) \\
&= \prod_{k=1}^{j-i} \mathbb{P}\left(X_{i+k} = x_{i+k} | X_{i+(k-1)} = x_{i+(k-1)}\right).
\end{aligned}$$

□

Lause A.2. *Olkoon* $(X_k)_{k=0}^{\infty}$ *Markovin ketju tila-avaruudella* $S = \{s_1, s_2, \dots, s_n\}$. *Tällöin*

$$\mathbb{P}(X_j = x_j | X_i = x_i, X_0 = x_0) = \mathbb{P}(X_j = x_j | X_i = x_i)$$

kaikilla $1 \leq i < j$ *ja* $x_0, x_i, x_j \in S$, *joille*

$$\mathbb{P}(X_i = x_i, X_0 = x_0) > 0.$$

Todistus. Olkoon

$$\bar{\mathbb{P}} : \mathcal{P}\left(S^{i-1}\right) \to [0,1]$$

kuvaus siten, että

$$\bar{\mathbb{P}}(A_1 \times A_2 \times \dots \times A_{i-1}) = \mathbb{P}(X_1 \in A_1, X_2 \in A_2, \dots, X_{i-1} \in A_{i-1} | X_i = x_i, X_0 = x_0).$$

Olkoon edelleen

$$\begin{aligned} B &= B_1 \times B_2 \times \dots \times B_{i-1} \\ &= \{(z_1, z_2, \dots, z_{i-1}) \in S^{i-1} : \bar{P}(\{z_1\} \times \{z_2\} \times \dots \times \{z_{i-1}\}) > 0\}. \end{aligned}$$

Koska

$$\begin{aligned} 0 < \mathbb{P}(X_i = x_i, X_0 = x_0) &= \sum_{x_1 \in S} \sum_{x_2 \in S} \cdots \sum_{x_{i-1} \in S} \mathbb{P}(X_i = x_i, X_{i-1} = x_{i-1}, \dots, X_0 = x_0) \\ &= \sum_{x_1 \in S} \sum_{x_2 \in S} \cdots \sum_{x_{i-1} \in S} \mathbb{P}(X_{i-1} = x_{i-1}, \dots, X_1 = x_1 | X_i = x_i, X_0 = x_0) \mathbb{P}(X_i = x_i, X_0 = x_0) \\ &= \sum_{x_1 \in S} \sum_{x_2 \in S} \cdots \sum_{x_{i-1} \in S} \mathbb{P}(X_1 = x_1, \dots, X_{i-1} = x_{i-1} | X_i = x_i, X_0 = x_0) \mathbb{P}(X_i = x_i, X_0 = x_0) \\ &= \sum_{x_1 \in S} \sum_{x_2 \in S} \cdots \sum_{x_{i-1} \in S} \bar{\mathbb{P}}(\{x_1\} \times \{x_2\} \times \dots \times \{x_{i-1}\}) \mathbb{P}(X_i = x_i, X_0 = x_0), \end{aligned}$$

niin $B \neq \varnothing$. Merkitään $A = \{X_i = x_i, X_0 = x_0\}$. Joukon B määritelmästä seuraa, että

$$\begin{aligned} &\mathbb{P}(X_j = x_j | X_i = x_i, X_0 = x_0) \\ &= \sum_{z_1 \in S} \cdots \sum_{z_{i-1} \in S} \sum_{z_{i+1} \in S} \cdots \sum_{z_{j-1} \in S} \mathbb{P}(X_j = x_j, \dots, X_{i+1} = z_{i+1}, X_{i-1} = z_{i-1}, \dots, X_1 = z_1 | A) \\ &= \sum_{z_1 \in B_1} \cdots \sum_{z_{i-1} \in B_{i-1}} \sum_{z_{i+1} \in S} \cdots \sum_{z_{j-1} \in S} \mathbb{P}(X_j = x_j, \dots, X_{i+1} = z_{i+1}, X_{i-1} = z_{i-1}, \dots, X_1 = z_1 | A). \end{aligned}$$

Lemmaa A.4 käyttämällä nähdään, että kun $z_k \in B_k$ kaikilla $k = 1, 2, \dots, i-1$, niin

$$\begin{aligned} &\mathbb{P}(X_j = x_j, \dots, X_{i+1} = z_{i+1}, X_{i-1} = z_{i-1}, \dots, X_1 = z_1 | A) \\ =&\mathbb{P}(X_j = x_j, \dots, X_{i+1} = z_{i+1}, X_{i-1} = z_{i-1}, \dots, X_1 = z_1 | X_i = x_i, X_0 = x_0) \\ =&\frac{\mathbb{P}(X_j = x_j, \dots, X_{i+1} = z_{i+1}, X_{i-1} = z_{i-1}, \dots, X_1 = z_1, X_i = x_i, X_0 = x_0)}{\mathbb{P}(X_i = x_i, X_0 = x_0)} \\ =&\frac{\mathbb{P}(X_j = x_j, \dots, X_{i+1} = z_{i+1} | X_i = x_i, \dots, X_0 = x_0) \mathbb{P}(X_i = x_i, \dots, X_0 = x_0)}{\mathbb{P}(X_i = x_i, X_0 = x_0)} \\ =&\frac{\mathbb{P}(X_j = x_j, \dots, X_{i+1} = z_{i+1} | X_i = x_i) \mathbb{P}(X_i = x_i, \dots, X_0 = x_0)}{\mathbb{P}(X_i = x_i, X_0 = x_0)}. \end{aligned}$$

Näin ollen saadaan, että

$$\begin{aligned}
&\mathbb{P}\left(X_j = x_j | X_i = x_i, X_0 = x_0\right) \\
=& \sum_{z_1 \in B_1} \cdots \sum_{z_{i-1} \in B_{i-1}} \sum_{z_{i+1} \in S} \cdots \sum_{z_{j-1} \in S} \mathbb{P}\left(X_j = x_j, \ldots, X_{i+1} = z_{i+1} | X_i = x_i\right) \\
&\times \frac{\mathbb{P}\left(X_i = x_i, \ldots, X_0 = x_0\right)}{\mathbb{P}\left(X_i = x_i, X_0 = x_0\right)} \\
=& \sum_{z_1 \in B_1} \cdots \sum_{z_{i-1} \in B_{i-1}} \frac{\mathbb{P}\left(X_i = x_i, \ldots, X_0 = x_0\right)}{\mathbb{P}\left(X_i = x_i, X_0 = x_0\right)} \\
&\underbrace{\sum_{z_{i+1} \in S} \cdots \sum_{z_{j-1} \in S} \mathbb{P}\left(X_j = x_j, \ldots, X_{i+1} = z_{i+1} | X_i = x_i\right)}_{=\mathbb{P}(X_j = x_j | X_i = x_i)} \\
=& \sum_{z_1 \in B_1} \cdots \sum_{z_{i-1} \in B_{i-1}} \mathbb{P}\left(X_{i-1} = z_{i-1}, \ldots, X_1 = z_1 | X_i = x_i, X_0 = x_0\right) \mathbb{P}\left(X_j = x_j | X_i = x_i\right) \\
=& \underbrace{\sum_{z_1 \in S} \cdots \sum_{z_{i-1} \in S} \mathbb{P}\left(X_{i-1} = z_{i-1}, \ldots, X_1 = z_1 | X_i = x_i, X_0 = x_0\right)}_{=1} \mathbb{P}\left(X_j = x_j | X_i = x_i\right) \\
=& \mathbb{P}\left(X_j = x_j | X_i = x_i\right). \qquad \square
\end{aligned}$$

Seuraava lause pätee kaikille neliömatriiseille. Muista, että matriisitulo on assosiatiivinen:

$$(AB)C = A(BC)$$

aina, kun A on $m \times n$ matriisi, B on $n \times p$ matriisi ja C on $p \times q$ matriisi.

Lause A.3. *Olkoon $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ matriisi. Tällöin kaikilla $k, m \in \mathbb{N}$*

(i) $P^k P^m = P^{k+m}$ *ja*

(ii) $(P^m)^k = P^{mk}$.

Todistus. (i) Olkoon $m \in \mathbb{N}$ mikä tahansa. Jos $k = 1$, niin potenssimatriisin määritelmän nojalla

$$P^k P^m = P P^m = P^{m+1} = P^{k+m}.$$

Oletetaan induktio-oletuksena, että $P^k P^m = P^{k+m}$ jollakin $k \in \mathbb{N}$. Tällöin

$$P^{k+1} P^m = (P P^k) P^m = P(P^k P^m) = P P^{k+m} = P^{(k+1)+m},$$

joten väite (i) on todistettu.

Todistetaan väite (ii). Olkoon $m \in \mathbb{N}$ mikä tahansa. Jos $k = 1$, niin väite pätee triviaalisti. Oletetaan induktio-oletuksena, että $(P^m)^k = P^{mk}$ jollakin $k \in \mathbb{N}$. Tällöin induktio-oletuksen ja kohdan (i) nojalla

$$(P^m)^{k+1} = P^m (P^m)^k = P^m P^{mk} = P^{m+mk} = P^{m(k+1)}. \qquad \square$$

Lause A.4. *Olkoon* $P = [p_{i,j}]_{i,j=1,2,\ldots,n}$ *siirtymätodennäköisyysmatriisi ja* $m, k \in \mathbb{N}$ *siten, että* $k \geq m$. *Jos* $P^m > 0$*, niin myös* $P^k > 0$.

Todistus. Jos $k = m$, väite pätee oletuksen nojalla. Oletetaan induktio-oletuksena, että $P^k > 0$ jollakin $k \in \mathbb{N}$. Olkoot $i, j \in \{1, 2, \ldots, n\}$ mitä tahansa. Täytyy osoittaa, että $p_{i,j}^{(k+1)} > 0$. Koska P on siirtymätodennäköisyysmatriisi, niin

$$\sum_{l=1}^{n} p_{i,l} = 1.$$

Näin ollen $p_{i,l_0} > 0$ jollakin $l_0 \in \{1, 2, \ldots, n\}$. Koska $P \geq 0$ ja $P^k > 0$, niin

$$p_{i,j}^{(k+1)} = \left[PP^k\right]_{i,j} = \sum_{l=1}^{n} p_{i,l} p_{l,j}^{(k)} \geq p_{i,l_0} p_{l_0,j}^{(k)} > 0. \qquad \square$$

B Funktionaalianalyysin tuloksia

Seuraava esitys perustuu lähteeseen [10]. Vektoriavaruuden E nollavektoria merkitään $\bar{0}_E$. Lisäksi oletetaan, että vektoriavaruuden kerroinkunta on aina reaalilukujen joukko $\mathbb{R}$.

Operaattorinormi

Määritelmä B.1. *Olkoot $(E, \|\cdot\|_E)$ ja $(F, \|\cdot\|_F)$ normiavaruuksia ja $L : E \to F$ lineaarikuvaus. Kuvauksen L operaattorinormi on*

$$\|L\| = \sup_{\|x\|_E \leq 1} \|Lx\|_F \,.$$

Lause B.1. *Olkoot $(E, \|\cdot\|_E)$ ja $(F, \|\cdot\|_F)$ normiavaruuksia ja $L : E \to F$ lineaarikuvaus. Tällöin kaikilla $x \in E$*

$$\|Lx\|_F \leq \|L\| \, \|x\|_E \,.$$

Todistus. $\|L(\bar{0}_E)\|_F = \|\bar{0}_F\|_F = 0 = \|L\| \, \|\bar{0}_E\|_E$, joten väite pätee tapauksessa $x = \bar{0}_E$. Olkoon siis $x \in E \setminus \{\bar{0}_E\}$. Tällöin $\frac{x}{\|x\|_E} \in E$ siten, että

$$\left\| \frac{x}{\|x\|_E} \right\|_E = \frac{1}{\|x\|_E} \|x\|_E = 1,$$

joten

$$\frac{1}{\|x\|_E} \|Lx\|_F = \left\| \frac{1}{\|x\|_E} Lx \right\|_F = \left\| L\left(\frac{x}{\|x\|_E} \right) \right\|_F \leq \|L\| \,,$$

mistä väite seuraa kertomalla saatu epäyhtälö puolittain luvulla $\|x\|_E > 0$. □

Lause B.2. *Olkoot $(E, \|\cdot\|_E)$ ja $(F, \|\cdot\|_F)$ normiavaruuksia ja $L : E \to F$ lineaarikuvaus. Tällöin seuraavat ehdot ovat yhtäpitäviä:*

(i) kuvaus L on jatkuva,

(ii) kuvaus L on jatkuva pisteessä $x_0 \in E$ ja

(iii) $\|L\| < \infty$.

Todistus. Implikaatio (i) $\Rightarrow$ (ii) on selvä määritelmän nojalla.

Todistetaan implikaatio (ii) $\Rightarrow$ (iii). Koska kuvaus L on jatkuva pisteessä $x_0 \in E$, on olemassa $\delta > 0$ siten, että

$$\|Lx - Lx_0\|_F \leq 1$$

kaikilla $x \in E$, joille $\|x - x_0\|_E \leq \delta$. Edelleen, jos $x \in E$ siten, että $\|x\|_E \leq \delta$, niin

$$\|Lx\|_F = \|L(x_0 - (x_0 - x))\|_F = \|Lx_0 - L(x_0 - x)\|_F \leq 1,$$

koska

$$\|x_0 - (x_0 - x)\|_E = \|x\|_E \leq \delta.$$

Jos sitten $x \in E \setminus \{\bar{0}_E\}$ on mikä tahansa, niin $\frac{\delta x}{\|x\|_E} \in E$ siten, että $\left\| \frac{\delta x}{\|x\|_E} \right\|_E = \delta$. Todistuksen alkuosan ja lineaarisuuden perusteella saadaan

$$\frac{\delta}{\|x\|_E}\|Lx\|_F = \left\| L\left(\frac{\delta x}{\|x\|_E}\right)\right\|_F \leq 1,$$

mistä edelleen

$$\|Lx\|_F \leq \frac{1}{\delta}\|x\|_E. \tag{26}$$

Lisäksi

$$\|L\bar{0}_E\|_F = \|\bar{0}_F\|_F = 0 = \frac{1}{\delta}\|\bar{0}_E\|_E,$$

joten yhtälö (26) pätee kaikilla $x \in E$. Operaattorinormin määritelmästä ja yhtälöstä (26) seuraa nyt, että

$$\|L\| \leq \frac{1}{\delta} < \infty.$$

Osoitetaan lopuksi implikaatio (iii) $\Rightarrow$ (i). Olkoot $\varepsilon > 0$ ja $x, x_0 \in E$ mitä tahansa siten, että

$$\|x - x_0\|_E < \frac{\varepsilon}{\|L\| + 1}.$$

Tällöin Lauseen B.1 nojalla

$$\|Lx - Lx_0\|_F = \|L(x - x_0)\|_F \leq \|L\|\|x - x_0\|_E < \varepsilon.$$ □

Olkoot E, F ja G vektoriavaruuksia ja $L : E \to F$ ja $M : F \to G$ lineaarikuvauksia. Otetaan yhdistetylle kuvaukselle $M \circ L : E \to G$ käyttöön merkintä

$$M \circ L = ML.$$

Edelleen, jos $P : E \to E$ on lineaarikuvaus, merkitään

$$P^k = \underbrace{P \circ P \circ P \circ \cdots \circ P}_{k \text{ kappaletta}}$$

kaikilla $k \in \mathbb{N}$. Lisäksi määritellään, että $P^0 = I$, missä $I : E \to E$ on identtinen kuvaus $Ix = x$.

Lause B.3. *Olkoon $(E, \|\cdot\|_E)$ normiavaruus ja $L : E \to E$ lineaarikuvaus. Kaikilla $k \in \mathbb{N}$ $L^k : E \to E$ on lineaarikuvaus ja $\left\|L^k\right\| \leq \|L\|^k$.*

Todistus. Väite on selvä, jos $k = 1$. Oletetaan induktio-oletuksena, että väite pätee jollakin $k \in \mathbb{N}$. Kaikilla $x, y \in E$ ja $\lambda \in \mathbb{R}$ pätee induktio-oletuksen ja kuvauksen L lineaarisuuden nojalla, että

$$L^{k+1}(x + y) = L\left(L^k(x + y)\right) = L\left(L^k x + L^k y\right) = L\left(L^k x\right) + L\left(L^k y\right) = L^{k+1}x + L^{k+1}y$$

ja

$$L^{k+1}(\lambda x) = L\left(L^k(\lambda x)\right) = L\left(\lambda L^k x\right) = \lambda L\left(L^k x\right) = \lambda L^{k+1}x,$$

joten kuvaus L^{k+1} on lineaarikuvaus. Todistetaan operaattorinormia koskeva väite kuvaukselle L^{k+1}. Olkoon $x \in E$ mikä tahansa siten, että $\|x\|_E \leq 1$. Käyttämällä Lausetta B.1 ja induktio-oletuksen antamaa epäyhtälöä $\left\|L^k\right\| \leq \|L\|^k$ saadaan, että

$$\left\|L^{k+1}x\right\|_E = \left\|L\left(L^k x\right)\right\|_E \leq \|L\|\left\|L^k x\right\|_E \leq \|L\|\left\|L^k\right\|\|x\|_E \leq \|L\|^{k+1},$$

mistä väite seuraa. □

Normiavaruuden sarjat

Määritelmä B.2. *Olkoon* $(E, \|\cdot\|_E)$ *normiavaruus ja* $x_k \in E$ *kaikilla* $k \in \mathbb{N}$*. Sarja*

$$\sum_{k=1}^{\infty} x_k$$

suppenee, jos jono

$$\left(\sum_{k=1}^{n} x_k\right)_{n=1}^{\infty}$$

suppenee. Sarja

$$\sum_{k=1}^{\infty} x_k$$

suppenee itseisesti, jos sarja

$$\sum_{k=1}^{\infty} \|x_k\|_E$$

suppenee.

Määritelmä B.3. *Normiavaruus* $(E, \|\cdot\|_E)$ *on Banachin avaruus, jos sen jokainen Cauchyn jono suppenee.*

Esimerkki B.1. *Joukko* $\mathbb{R}^n$ *varustettuna tavallisella Euklidisella normilla on Banachin avaruus.*

Lause B.4. *Normiavaruus* $(E, \|\cdot\|_E)$ *on Banachin avaruus, jos ja vain jos sen jokainen itseisesti suppeneva sarja suppenee.*

Todistus. Oletetaan ensin, että $(E, \|\cdot\|_E)$ on Banachin avaruus. Olkoon $\sum_{k=1}^{\infty} x_k$ itseisesti suppeneva sarja. Täytyy osoittaa, että sarja $\sum_{k=1}^{\infty} x_k$ suppenee. Koska $(E, \|\cdot\|_E)$ on Banachin avaruus, riittää osoittaa, että jono

$$\left(\sum_{k=1}^{n} x_k\right)_{n=1}^{\infty}$$

on Cauchyn jono. Olkoon $\varepsilon > 0$ mikä tahansa. Koska sarja $\sum_{k=1}^{\infty} x_k$ suppenee itseisesti, on olemassa $n_\varepsilon \in \mathbb{N}$ siten, että

$$\sum_{k=n_\varepsilon+1}^{\infty} \|x_k\|_E < \varepsilon.$$

Olkoot n ja m mitä tahansa luonnollisia lukuja siten, että $n_\varepsilon \leq n \leq m$. Tällöin

$$\left\|\sum_{k=1}^{n} x_k - \sum_{k=1}^{m} x_k\right\|_E = \left\|\sum_{k=n+1}^{m} x_k\right\|_E \leq \sum_{k=n+1}^{m} \|x_k\|_E \leq \sum_{k=n_\varepsilon+1}^{\infty} \|x_k\|_E < \varepsilon,$$

mistä väite seuraa.

Oletetaan sitten, että jokainen avaruuden $(E, \|\cdot\|_E)$ itseisesti suppeneva sarja suppenee, ja osoitetaan, että avaruus $(E, \|\cdot\|_E)$ on Banachin avaruus. Olkoon $(x_k)_{k=1}^{\infty}$ mikä tahansa Cauchyn jono. Riittää osoittaa, että jonolla $(x_k)_{k=1}^{\infty}$ on suppeneva osajono $\left(x_{k_j}\right)_{j=1}^{\infty}$; on helppo nähdä, että

Cauchyn jono suppenee, jos sillä on suppeneva osajono. Olkoon $\varepsilon > 0$ mikä tahansa. Koska jono $(x_k)_{k=1}^{\infty}$ on Cauchyn jono, on olemassa luonnolliset luvut $k_1 \leq k_2 \leq k_3 \leq \ldots$ siten, että

$$\|x_k - x_l\|_E < \frac{\varepsilon}{2^{j+1}}$$

kaikilla $k, l \geq k_j$. Huomataan, että kaikilla $j \geq 2$

$$x_{k_j} = \sum_{i=2}^{j}(x_{k_i} - x_{k_{i-1}}) + x_{k_1},$$

missä

$$\sum_{i=2}^{j} \left\|x_{k_i} - x_{k_{i-1}}\right\|_E \leq \sum_{i=2}^{j} \frac{\varepsilon}{2^i} \leq \sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon.$$

Näin ollen sarja

$$\sum_{i=2}^{\infty}(x_{k_i} - x_{k_{i-1}})$$

suppenee itseisesti. Oletuksen nojalla jokainen itseisesti suppeneva sarja suppenee, joten saadaan, että

$$\lim_{j\to\infty} x_{k_j} = \lim_{j\to\infty} \left(\sum_{i=2}^{j}(x_{k_i} - x_{k_{i-1}}) + x_{k_1}\right) = \sum_{i=2}^{\infty}(x_{k_i} - x_{k_{i-1}}) + x_{k_1}. \qquad \square$$

Lause B.5. *Olkoon $(E, \|\cdot\|_E)$ Banachin avaruus. Aliavaruus $F \subset E$ on Banachin avaruus, jos ja vain jos joukko F on suljettu avaruudessa $(E, \|\cdot\|_E)$.*

Todistus. Oletetaan ensin, että aliavaruus $F \subset E$ on Banachin avaruus ja osoitetaan, että joukko F on suljettu avaruudessa $(E, \|\cdot\|_E)$. Tehdään antiteesi: on olemassa joukon F kasautumispiste $x \in E \setminus F$. Tällöin jokaiselle $k \in \mathbb{N}$ on olemassa $x_k \in F$ siten, että

$$\|x_k - x\|_E < \frac{1}{k},$$

joten $x = \lim_{k\to\infty} x_k$. Suppeneva jono on Cauchyn jono, ja koska aliavaruus F on Banachin avaruus, on olemassa $y \in F$ siten, että $\lim_{k\to\infty} x_k = y$. Raja-arvon yksikäsitteisyydestä seuraa, että $x = y \in F$, mikä on ristiriita.

Oletetaan sitten, että $F \subset E$ on suljettu avaruudessa $(E, \|\cdot\|_E)$ ja osoitetaan, että F on Banachin avaruus. Olkoon $(x_k)_{k=1}^{\infty}$ mikä tahansa Cauchyn jono siten, että $x_k \in F$ kaikilla $k \in \mathbb{N}$. Avaruus $(E, \|\cdot\|_E)$ on Banachin avaruus, joten jono $(x_k)_{k=1}^{\infty}$ suppenee avaruudessa $(E, \|\cdot\|_E)$. Olkoon $x \in E$ siten, että $\lim_{k\to\infty} x_k = x$. Koska $x_k \in F$ kaikilla $k \in \mathbb{N}$, niin x on joukon F kasautumispiste. Koska $F \subset E$ on suljettu, niin $x \in F$, ja siten jono $(x_k)_{k=0}^{\infty}$ suppenee avaruudessa F. $\square$

Lause B.6. *Olkoon $(E, \|\cdot\|_E)$ Banachin avaruus ja $L : E \to E$ lineaarikuvaus siten, että $\|L\| < 1$. Tällöin kuvauksella $I - L$ on käänteiskuvaus*

$$(I - L)^{-1} = \sum_{k=0}^{\infty} L^k.$$

Todistus. Olkoon $x \in E$ mikä tahansa. Käyttämällä Lausetta B.1 ja Lausetta B.3 saadaan, että

$$\sum_{k=0}^{\infty} \left\| L^k x \right\|_E \leq \sum_{k=0}^{\infty} \left\| L^k \right\| \|x\|_E \leq \sum_{k=0}^{\infty} \|L\|^k \|x\|_E < \infty,$$

missä geometrinen sarja $\sum_{k=0}^{\infty} \|L\|^k \|x\|_E$ suppenee, koska $0 \leq \|L\| < 1$. Lauseen B.4 nojalla tästä seuraa, että sarja

$$\sum_{k=0}^{\infty} L^k x$$

suppenee kaikilla $x \in E$. Helposti nähdään, että kuvaus $\sum_{k=0}^{\infty} L^k$ on lineaarinen. Käyttämällä Lausetta B.2 nähdään, että

$$\begin{aligned} L\left(\sum_{k=0}^{\infty} L^k x\right) &= L\left(\lim_{n\to\infty} \sum_{k=0}^{n} L^k x\right) = \lim_{n\to\infty} L\left(\sum_{k=0}^{n} L^k x\right) = \lim_{n\to\infty} \sum_{k=0}^{n} L^{k+1} x \\ &= \lim_{n\to\infty} \sum_{k=1}^{n+1} L^k x = \lim_{n\to\infty} \left(\sum_{k=0}^{n+1} L^k x - \underbrace{Ix}_{L^0 x}\right) = \sum_{k=0}^{\infty} L^k x - x, \end{aligned}$$

mistä

$$\underbrace{\sum_{k=0}^{\infty} L^k x}_{=I\left(\sum_{k=0}^{\infty} L^k x\right)} - L\left(\sum_{k=0}^{\infty} L^k x\right) = x,$$

eli

$$(I - L)\left(\sum_{k=0}^{\infty} L^k x\right) = x.$$

Näin ollen

$$(I - L)\left(\sum_{k=0}^{\infty} L^k\right) x = (I - L)\left(\sum_{k=0}^{\infty} L^k x\right) = x.$$

Toisaalta

$$\begin{aligned} \left(\sum_{k=0}^{\infty} L^k\right)(I - L)\, x &= \left(\sum_{k=0}^{\infty} L^k\right)(x - Lx) = \sum_{k=0}^{\infty} L^k x - \sum_{k=0}^{\infty} L^k (Lx) \\ &= \sum_{k=0}^{\infty} L^k x - \sum_{k=0}^{\infty} L^{k+1} x = \sum_{k=0}^{\infty} L^k x - \sum_{k=1}^{\infty} L^k x \\ &= \sum_{k=0}^{\infty} L^k x - \left(\sum_{k=0}^{\infty} L^k x - Ix\right) = Ix = x \end{aligned}$$

kaikilla $x \in E$, joten väite seuraa. $\square$

Sisätuloavaruudet

Määritelmä B.4. *Olkoon E vektoriavaruus. Kuvaus $\langle\cdot|\cdot\rangle : E \times E \to \mathbb{R}$ on sisätulo, jos*

(i) $\langle x|x\rangle \geq 0$ *kaikilla* $x \in E$ *ja* $\langle x|x\rangle = 0$*, jos ja vain jos* $x = \bar{0}_E$

(ii) kaikilla $x, y, z \in E$ *ja* $\lambda \in \mathbb{R}$ $\langle x + y|z\rangle = \langle x|z\rangle + \langle y|z\rangle$ *ja* $\langle \lambda x|y\rangle = \lambda \langle x|y\rangle$

(iii) $\langle x|y\rangle = \langle y|x\rangle$ *kaikilla* $x, y \in E$.

Paria $(E, \langle\cdot|\cdot\rangle)$ *kutsutaan sisätuloavaruudeksi.*

Vektoriavaruuden E sisätulo $\langle\cdot|\cdot\rangle : E \times E \to \mathbb{R}$ määrää normin

$$\|\cdot\| : E \to [0, \infty[\,,\ \|x\| = \sqrt{\langle x|x\rangle},$$

jolle pätee Cauchyn–Schwarzin epäyhtälö:

Lause B.7. *Olkoon* $(E, \langle\cdot|\cdot\rangle)$ *sisätuloavaruus ja* $\|\cdot\| : E \to [0, \infty[$ *sisätulon määräämä normi. Tällöin*

$$|\langle x|y\rangle| \leq \|x\|\|y\|$$

kaikilla $x, y \in E$.

Lause B.8. *Olkoon* $(E, \langle\cdot|\cdot\rangle)$ *sisätuloavaruus ja* $y \in E$ *mikä tahansa. Tällöin kuvaus*

$$E \to \mathbb{R} : x \mapsto \langle x|y\rangle$$

on jatkuva.

Todistus. Olkoot $\varepsilon > 0$ ja $x_0 \in E$ mitä tahansa. Tällöin, jos $x \in E$ on mikä tahansa siten, että

$$\|x - x_0\| < \frac{\varepsilon}{\|y\| + 1},$$

niin Lauseen B.7 nojalla

$$|\langle x|y\rangle - \langle x_0|y\rangle| = |\langle x - x_0|y\rangle| \leq \|x - x_0\|\,\|y\| \leq \frac{\varepsilon}{\|y\| + 1}\,\|y\| \leq \varepsilon. \qquad \square$$

Määritelmä B.5. *Olkoon* $(E, \langle\cdot|\cdot\rangle)$ *sisätuloavaruus. Vektorit* $x_1, x_2, \ldots, x_n \in E$ *ovat ortogonaalisia (sisätulon* $\langle\cdot|\cdot\rangle$ *suhteen), jos*

$$\langle x_i|x_j\rangle = 0$$

kaikilla $i \neq j$*. Jos lisäksi pätee, että*

$$\|x_i\| = 1$$

kaikilla $i = 1, 2, \ldots, n$*, niin sanotaan, että vektorit* x_i *ovat ortonormaaleja.*

Lause B.9. *Olkoon* $(E, \langle\cdot|\cdot\rangle)$ *sisätuloavaruus, jonka ortogonaaliset kantavektorit ovat* $u_1, u_2, \ldots, u_n \in E$*. Tällöin jokaisella vektorilla* $x \in E$ *on esitys*

$$x = \sum_{i=1}^{n} \langle x|u_i\rangle\, u_i.$$

Todistus. Koska vektorit $u_1, u_2, \ldots, u_n$ muodostavat avaruuden E kannan, on jokaiselle $x \in E$ olemassa luvut $c_1, c_2, \ldots, c_n \in \mathbb{R}$ siten, että

$$x = \sum_{i=1}^{n} c_i u_i.$$

Riittää siis osoittaa, että $c_i = \langle x|u_i \rangle$ kaikilla $i = 1, 2, \ldots, n$. Vektoreiden u_i ortogonaalisuuden nojalla saadaan kaikille $i = 1, 2, \ldots, n$

$$\langle x|u_i \rangle = \left\langle \sum_{j=1}^{n} c_j u_j \middle| u_i \right\rangle = \sum_{j=1}^{n} c_j \langle u_j|u_i \rangle = c_i,$$

mistä väite seuraa. □

Lause B.10. *[Pythagoraan lause] Olkoon $(E, \langle\cdot|\cdot\rangle)$ sisätuloavaruus, $n \in \mathbb{N}$ ja vektorit $x_1, x_2, \ldots, x_n \in E$ ortogonaalisia. Tällöin*

$$\left\| \sum_{i=1}^{n} x_i \right\|^2 = \sum_{i=1}^{n} \|x_i\|^2 .$$

Todistus. Jos $n = 1$, niin väite pätee triviaalisti. Oletetaan induktio-oletuksena, että väite pätee kaikilla $n \leq k \in \mathbb{N}$. Olkoot vektorit $x_1, x_2, \ldots, x_{k+1} \in E$ ortogonaalisia. Induktio-oletusta käyttäen saadaan, että

$$\begin{aligned}
\left\| \sum_{i=1}^{k+1} x_i \right\|^2 &= \left\| \sum_{i=1}^{k} x_i + x_{k+1} \right\|^2 = \left\langle \sum_{i=1}^{k} x_i + x_{k+1} \middle| \sum_{i=1}^{k} x_i + x_{k+1} \right\rangle \\
&= \left\langle \sum_{i=1}^{k} x_i \middle| \sum_{i=1}^{k} x_i \right\rangle + \left\langle \sum_{i=1}^{k} x_i \middle| x_{k+1} \right\rangle + \left\langle x_{k+1} \middle| \sum_{i=1}^{k} x_i \right\rangle + \langle x_{k+1}|x_{k+1} \rangle \\
&= \left\| \sum_{i=1}^{k} x_i \right\|^2 + 2 \sum_{i=1}^{k} \underbrace{\langle x_i|x_{k+1} \rangle}_{=0} + \|x_{k+1}\|^2 \\
&= \sum_{i=1}^{k} \|x_i\|^2 + \|x_{k+1}\|^2 \\
&= \sum_{i=1}^{k+1} \|x_i\|^2 ,
\end{aligned}$$

joten väite seuraa induktioperiaatteesta. □

Määritelmä B.6. *Olkoon $(E, \langle\cdot|\cdot\rangle)$ sisätuloavaruus ja $A \subset E$. Joukon A ortokomplementti on joukko*

$$A^{\perp} = \{x \in E : \langle x|a \rangle = 0 \text{ kaikilla } a \in A\} .$$

Lause B.11. *Olkoon $(E, \langle\cdot|\cdot\rangle)$ sisätuloavaruus ja $A \subset E$. Joukon A ortokomplementti on avaruuden $(E, \langle\cdot|\cdot\rangle)$ suljettu aliavaruus.*

Todistus. $\bar{0}_E \in A^{\perp}$, koska kaikilla $a \in A$

$$\langle \bar{0}_E|a \rangle = \langle 0 \cdot \bar{0}_E|a \rangle = 0 \langle \bar{0}_E|a \rangle = 0.$$

Olkoot $x, y \in A^{\perp}$ ja $\alpha \in \mathbb{R}$ mitä tahansa. Tällöin kaikilla $a \in A$

$$\langle x + y|a\rangle = \langle x|a\rangle + \langle y|a\rangle = 0 + 0 = 0$$

ja

$$\langle \alpha x|a\rangle = \alpha \langle x|a\rangle = \alpha \cdot 0 = 0,$$

joten $x + y \in A^{\perp}$ ja $\alpha x \in A^{\perp}$. Siispä $A^{\perp}$ on aliavaruus.

Olkoon $a \in A$ mikä tahansa ja $f : E \to \mathbb{R}$ siten, että $f(x) = \langle x|a\rangle$. Lauseen B.8 nojalla kuvaus f on jatkuva, joten joukko

$$a^{\perp} = \{x \in E : \langle x|a\rangle = 0\} = f^{\{-1\}}(\{0\})$$

on suljetun joukon $\{0\} \subset \mathbb{R}$ alkukuva jatkuvassa kuvauksessa f ja sellaisena itsekin suljettu. Koska

$$A^{\perp} = \bigcap_{a \in A} a^{\perp},$$

myös $A^{\perp}$ on suljettujen joukkojen leikkauksena suljettu. □

Määritelmä B.7. *Olkoon $(E, \langle\cdot|\cdot\rangle)$ sisätuloavaruus ja $L : E \to E$ lineaarikuvaus. Sanotaan, että kuvaus L on itseadjungoitu, jos*

$$\langle x|Ly\rangle = \langle Lx|y\rangle$$

kaikilla $x, y \in E$.

Määritelmä B.8. *Olkoon $(E, \langle\cdot|\cdot\rangle)$ sisätuloavaruus ja $L : E \to E$ lineaarikuvaus. Sanotaan, että kuvaus L on positiivinen, jos*

$$\langle x|Lx\rangle \geq 0$$

kaikilla $x \in E$.

Seuraavan lauseen yhtälö (27) on esitetty lähteessä [11, s. 44] ilman todistusta ja tietoa siitä, millä ehdoilla yhtälö tarkalleen ottaen pätee.

Lause B.12. *Olkoon $(E, \langle\cdot|\cdot\rangle)$ sisätuloavaruus ja $L : E \to E$ lineaarinen bijektio, joka on sekä positiivinen että itseadjungoitu. Tällöin*

$$\left\langle x|L^{-1}x\right\rangle = \sup_{y \in E} \left(2\langle x|y\rangle - \langle y|Ly\rangle\right) \tag{27}$$

kaikilla $x \in E$.

Todistus. Osoitetaan ensin, että $\left\langle x|L^{-1}x\right\rangle \geq 0$ kaikilla $x \in E$. Tehdään antiteesi: on olemassa $x \in E$ siten, että $\left\langle x|L^{-1}x\right\rangle < 0$. Koska kuvaus L on bijektio, on olemassa $y \in E$ siten, että $x = Ly$. Näin ollen

$$0 > \left\langle x|L^{-1}x\right\rangle = \left\langle Ly|L^{-1}(Ly)\right\rangle = \langle Ly|y\rangle = \langle y|Ly\rangle,$$

mikä on ristiriita.

Olkoon nyt $x \in E$ mikä tahansa. Tällöin kaikilla $y \in E$ pätee, että

$$\begin{aligned}
0 &\leq \left\langle x - Ly | L^{-1}(x - Ly) \right\rangle \\
&= \left\langle x - Ly | L^{-1}x - L^{-1}(Ly) \right\rangle \\
&= \left\langle x - Ly | L^{-1}x - y \right\rangle \\
&= \left\langle x | L^{-1}x \right\rangle - \langle x|y \rangle - \left\langle Ly | L^{-1}x \right\rangle + \langle Ly|y \rangle \\
&= \left\langle x | L^{-1}x \right\rangle - \langle x|y \rangle - \left\langle y | L\left(L^{-1}x\right) \right\rangle + \langle y|Ly \rangle \\
&= \left\langle x | L^{-1}x \right\rangle - \langle x|y \rangle - \langle y|x \rangle + \langle y|Ly \rangle \\
&= \left\langle x | L^{-1}x \right\rangle - 2\langle x|y \rangle + \langle y|Ly \rangle ,
\end{aligned}$$

ja siten

$$\left\langle x | L^{-1}x \right\rangle \geq 2\langle x|y \rangle - \langle y|Ly \rangle$$

kaikilla $y \in E$. Erityisesti siis

$$\left\langle x | L^{-1}x \right\rangle \geq \sup_{y \in E} \left(2\langle x|y \rangle - \langle y|Ly \rangle \right).$$

Toisaalta, valitsemalla $y = L^{-1}x$ saadaan, että $Ly = x$ ja

$$\begin{aligned}
\left\langle x | L^{-1}x \right\rangle &= 2\left\langle x | L^{-1}x \right\rangle - \left\langle x | L^{-1}x \right\rangle \\
&= 2\langle x|y \rangle - \langle Ly|y \rangle \\
&= 2\langle x|y \rangle - \langle y|Ly \rangle \\
&\leq \sup_{y \in E} \left(2\langle x|y \rangle - \langle y|Ly \rangle \right).
\end{aligned}$$

□

Lause B.13. *Olkoon $(E, \langle \cdot | \cdot \rangle)$, missä $E \neq \{\bar{0}_E\}$, sisätuloavaruus siten, että $n = \dim E \in \mathbb{N}$ ja olkoon kuvaus $L : E \to E$ lineaarinen ja itseadjungoitu. Tällöin avaruudella E on ortonormaali kanta $v_1, v_2, \ldots, v_n \in E$, jonka jokainen vektori v_i on kuvauksen L ominaisvektori, eli on olemassa luvut $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$ siten, että*

$$Lv_i = \lambda_i v_i$$

kaikilla $i = 1, 2, \ldots, n$.

Todistus. Katso esimerkiksi [12, Theorem 2.1]. □

C Kroneckerin lemma

Lemma C.1 (Abelin osittaissummauskaava)**.** *Olkoot* $(x_n)_{n=0}^{\infty}$ *ja* $(y_n)_{n=0}^{\infty}$ *lukujonoja ja*

$$Y_n = \sum_{k=0}^{n} y_k$$

kaikilla $n \in \mathbb{N}_0$*. Tällöin*

$$\sum_{k=0}^{n} x_k y_k = x_n Y_n - \sum_{k=0}^{n-1} (x_{k+1} - x_k) Y_k$$

kaikilla $n \in \mathbb{N}_0$*.*

Todistus. Kaikilla $k = 1, 2, 3, \ldots, n$ pätee, että $y_k = Y_k - Y_{k-1}$, joten

$$\sum_{k=0}^{n} x_k y_k = x_0 y_0 + x_n y_n + \sum_{k=1}^{n-1} x_k y_k = x_0 y_0 + x_n y_n + \sum_{k=1}^{n-1} x_k (Y_k - Y_{k-1}),$$

ja lisäämällä yhtälöön puolittain termi $\sum_{k=1}^{n-1} (x_{k+1} - x_k) Y_k$ saadaan, että

$$\sum_{k=0}^{n} x_k y_k + \sum_{k=1}^{n-1} (x_{k+1} - x_k) Y_k = x_0 y_0 + x_n y_n + \sum_{k=1}^{n-1} (x_{k+1} Y_k - x_k Y_{k-1}).$$

Yhtälön oikean puolen summatermi on teleskooppisumma

$$\sum_{k=1}^{n-1} (x_{k+1} Y_k - x_k Y_{k-1}) = \sum_{k=1}^{n-1} (z_k - z_{k-1}) = z_{n-1} - z_0,$$

missä $z_k = x_{k+1} Y_k$. Näin ollen saadaan, että

$$\begin{aligned} \sum_{k=0}^{n} x_k y_k + \sum_{k=1}^{n-1} (x_{k+1} - x_k) Y_k &= x_0 y_0 + x_n y_n + x_n Y_{n-1} - x_1 Y_0 \\ &= x_0 y_0 + x_n (y_n + Y_{n-1}) - x_1 y_0 \\ &= x_n Y_n - (x_1 - x_0) y_0 \\ &= x_n Y_n - (x_1 - x_0) Y_0, \end{aligned}$$

mistä edelleen

$$\begin{aligned} \sum_{k=0}^{n} x_k y_k &= x_n Y_n - (x_1 - x_0) Y_0 - \sum_{k=1}^{n-1} (x_{k+1} - x_k) Y_k \\ &= x_n Y_n - \sum_{k=0}^{n-1} (x_{k+1} - x_k) Y_k. \end{aligned}$$

□

Lause C.1 (Kroneckerin lemma). *Olkoon* $(y_n)_{n=1}^{\infty}$ *reaalilukujono siten, että sarja* $\sum_{n=1}^{\infty} y_n$ *suppenee. Olkoon lisäksi* $(x_n)_{n=1}^{\infty}$ *kasvava reaalilukujono siten, että*

(i) $x_1 > 0$

(ii) $\lim_{n\to\infty} x_n = \infty$.

Tällöin

$$\lim_{n\to\infty} \frac{1}{x_n} \sum_{k=1}^{n} x_k y_k = 0.$$

Todistus. Merkitään $x_0 = 0 = y_0$, $Y_n = \sum_{k=0}^{n} y_k$ ja $y = \sum_{k=0}^{\infty} y_n$. Olkoon $\varepsilon > 0$ mikä tahansa ja $N \in \mathbb{N}$ siten, että

$$|Y_n - y| \leq \varepsilon$$

kaikilla $n \geq N$. Kaikille $n \geq N+2$ saadaan, että

$$\begin{aligned}
\left| \frac{1}{x_n} \sum_{k=N+1}^{n-1} (x_{k+1} - x_k)(Y_k - y) \right| &\leq \frac{1}{|x_n|} \sum_{k=N+1}^{n-1} |x_{k+1} - x_k| \, |Y_k - y| \\
&\leq \frac{\varepsilon}{x_n} \sum_{k=N+1}^{n-1} (x_{k+1} - x_k) \\
&= \frac{\varepsilon}{x_n} (x_n - x_{N+1}) \\
&\leq \varepsilon.
\end{aligned}$$

Lauseen varsinainen väite voidaan nyt todistaa Lemman C.1 avulla: kun $n \geq N+2$, niin

$$\begin{aligned}
&\left| \frac{1}{x_n} \sum_{k=1}^{n} x_k y_k \right| = \left| \frac{1}{x_n} \sum_{k=0}^{n} x_k y_k \right| = \left| \frac{1}{x_n} \left(x_n Y_n - \sum_{k=0}^{n-1} (x_{k+1} - x_k) Y_k \right) \right| \\
&= \left| Y_n - \frac{1}{x_n} \sum_{k=0}^{n-1} (x_{k+1} - x_k) Y_k \right| \\
&= \left| Y_n - \frac{1}{x_n} \sum_{k=0}^{N} (x_{k+1} - x_k) Y_k - \frac{1}{x_n} \sum_{k=N+1}^{n-1} (x_{k+1} - x_k) Y_k \right| \\
&= \left| Y_n - \frac{1}{x_n} \sum_{k=0}^{N} (x_{k+1} - x_k) Y_k - \frac{1}{x_n} \sum_{k=N+1}^{n-1} (x_{k+1} - x_k) y - \frac{1}{x_n} \sum_{k=N+1}^{n-1} (x_{k+1} - x_k)(Y_k - y) \right| \\
&= \left| Y_n - \frac{1}{x_n} \sum_{k=0}^{N} (x_{k+1} - x_k) Y_k - \frac{y}{x_n} (x_n - x_{N+1}) - \frac{1}{x_n} \sum_{k=N+1}^{n-1} (x_{k+1} - x_k)(Y_k - y) \right| \\
&\leq \left| Y_n - \frac{1}{x_n} \sum_{k=0}^{N} (x_{k+1} - x_k) Y_k - \frac{y}{x_n} (x_n - x_{N+1}) \right| + \left| \frac{1}{x_n} \sum_{k=N+1}^{n-1} (x_{k+1} - x_k)(Y_k - y) \right| \\
&\leq \left| Y_n - \frac{1}{x_n} \sum_{k=0}^{N} (x_{k+1} - x_k) Y_k - \frac{y}{x_n} (x_n - x_{N+1}) \right| + \varepsilon \\
&\to |y - 0 - y| + \varepsilon = \varepsilon, \text{ kun } n \to \infty.
\end{aligned}$$

□

Viitteet

[1] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times.* American Mathematical Society, 2009.

[2] Richard Durrett. *Essentials of Stochastic Processes.* Springer Texts in Statistics. Springer, 2nd edition, 2012.

[3] John G. Kemeny and J. Laurie Snell. *Finite Markov Chains.* Princeton, NJ: Van Nostrand, 1965.

[4] Daniel Rudolf. Explicit error bounds for markov chain monte carlo, 2011. University Jena.

[5] Robert Cogburn. The central limit theorem for markov processes. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, pages 485–512. University of California Press, Berkeley, 1972.

[6] Matti Vihola. Lectures on stochastic simulation, March 2018.

[7] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability.* Springer Verlag, 1993.

[8] Antonietta Mira and Charles J. Geyer. Ordering monte carlo markov chains. Technical Report 632, School of Statistics, University of Minnesota, 1999.

[9] Peter H. Peskun. Optimum monte-carlo sampling using markov chains. *Biometrika*, 60:607–612, 1973.

[10] Lauri Kahanpää. Funktionaalianalyysi. Suoraviivaista ajattelua - osa II, 2004. Jyväskylän yliopisto. Matematiikan ja tilastotieteen laitos. Luentomoniste 51.

[11] Sergio Caracciolo, Andrea Pelissetto, and Alan D. Sokal. Nonlocal monte carlo algorithm for self-avoiding walks with fixed endpoints. *Journal of Statistical Physics*, 60, November 1990.

[12] George Bachman and Lawrence Narici. *Functional Analysis.* Dover Books on Mathematics. Dover Publications, 2012.