Painotusotanta vakioivilla muuttujilla ja sekoitejakaumilla

Elviira Pohjanheimo

Matematiikan pro gradu

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
Syksy 2023

Tiivistelmä: Elviira Pohjanheimo, *Painotusotanta vakioivilla muuttujilla ja sekoitejakaumilla* (engl. *Importance sampling with control variates and mixture distributions*), matematiikan pro gradu -tutkielma, 50 sivua, Jyväskylän yliopisto, Matematiikan ja tilastotieteen laitos, syksy 2023.

Tämän tutkielman tarkoituksena on perehtyä painotusotantaan ja sen varianssin pienentämiseen vakioivien muuttujien ja sekoitejakaumien avulla. Sekä painotusotanta että vakioivat muuttujat ovat Monte Carlo -menetelmiä, ja tässä tutkielmassa keskitytään niiden käyttämiseen odotusarvon integraalin stokastisessa simuloinnissa. Tutkielmassa tarkastellaan kirjallisuudessa esitettyä painotusotannan parannusehdotusta, joka mahdollisesti parantaa menetelmän tehokkuutta pienentämällä varianssia ja samalla suojaa menetelmää epäonnistumiselta.

Painotusotanta perustuu todennäköisyysmitan vaihtamiseen tehokkuuden parantamiseksi. Sen estimaattoriin voidaan lisätä vakioivat muuttujat, joilla varianssin pienennys perustuu korrelaatioon alkuperäisen muuttujan ja vakioivan muuttujan välillä. Vakioivien muuttujien kertoimet täytyy estimoida, sillä varianssin pienentämisen kannalta optimaalisia kertoimia ei yleensä voida laskea. Tämän tutkielman päätuloksena onkin lause, jonka mukaan muutaman ehdon pätiessä optimaalisten kertoimien korvaaminen pienimmän neliösumman estimaattorilla on vaikutukseltaan asymptoottisesti merkityksetöntä. Sen yksityiskohtaisessa todistuksessa käytetään todennäköisyyslaskennan perustuloksia, stokastista kertaluokkaa, matriisilaskentaa sekä analyysin differentiaali- ja integraalilaskentaa.

Vakioivien muuttujien lisäksi painotusotannan parannusehdotukseen liittyy sekoitejakauman käyttäminen estimaattorissa. Sekoitejakaumalla tarkoitetaan todennäköisyysjakaumaa, jonka tiheysfunktio on painotettu summa tiheysfunktioista. Yhdistämällä vakioivat muuttujat ja sekoitejakaumaa käyttävä painotusotanta saadaan estimaattori, jonka varianssi todistetaan ylhäältä rajoitetuksi. Lisäksi osoitetaan, että menetelmän varianssi ei ole suurempi kuin vastaavan painotusotannan varianssi parhaimmillaan. Täten, koska estimaattori välttää painotusotannan mahdollisen äärettömän varianssin eikä myöskään suurenna varianssia, voidaan todeta, että parannusehdotusta käyttävä menetelmä on turvallinen ja tehokas versio painotusotannasta.

Sisällys

Johdanto

Eri tieteenaloilla eteen voi tulla ongelma, jonka matemaattinen ratkaiseminen on hyvin työlästä tai mahdotonta. Eräs esimerkki tällaisesta ongelmasta on todennäköisyyksien ja odotusarvojen yhteydessä ilmenevien integraalien laskeminen. Teknologian ja etenkin tietokoneiden kehittyminen on mahdollistanut näiden ongelmien ratkaisemisen simuloinnilla, jossa ongelmaan saadaan likimääräinen ratkaisu. Erästä simulointimenetelmien joukkoa kutsutaan Monte Carlo -menetelmiksi. Monte Carlo -menetelmät ovat otantamenetelmiä, joista saatujen estimaattoreiden arvo perustuu useasti toistetun satunnaisen simulaation keskiarvon laskemiseen.

Yleisesti ottaen hyvä estimaattori lähestyy alkuperäistä arvoa ja on varianssiltaan pieni, mikä tarkoittaa sitä, että kohtuullisella otoskoolla saadaan luotettava tulos. Varianssin pienentäminen mielletäänkin menetelmän tehokkuuden parantamiseksi, ja se on ollut tärkeä motivaattori Monte Carlo -menetelmien tutkimiselle. Fishman [5] toteaa, että aikana, jolloin tietokoneet olivat laskentatehokkuudeltaan hitaita, oli tärkeää, että luotettava tulos saadaan pienimmällä mahdollisella otoskoolla. Nykypäivänä tietokoneet ovat kehittyneempiä, mutta kun simuloitavana on useampiulotteinen tapaus, on tärkeää, että otoskokoa voidaan pienentää ja saadaan edelleen halutulla luottamusvälillä oleva tulos [5].

Art Owen ja Yi Zhou esittelevät artikkelissaan *Safe and effective importance sampling* [9] vuodelta 2000 kaksi keinoa parantaa painotusotannaksi kutsuttua Monte Carlo -menetelmää. Tämä tutkielma seuraa vahvasti heidän artikkeliaan ja perehtyy syvällisesti toiseen näistä parannusehdotuksista. Painotusotannan yleinen versio esitellään tutkielman luvussa 2, mutta ennen sitä luvussa 1 annetaan tarvittavat taustatiedot. Lukijan esitietoina on, että hän tuntee stokastiikan ja lineaarialgebran perustiedot, mutta tarkemmat taustatiedot ovat tarpeen luvun 3 todistuksiin.

Tarkasteltavaan Owenin ja Zhoun parannusehdotukseen liittyy vakioivat muuttujat -menetelmä, ja luku 3 keskittyy siihen. Luvussa myös todistetaan yksityiskohtaisesti Owenin ja Zhoun painotusotantaan ja vakioiviin muuttujiin liittyvä lause [9, Theorem 1]. Heidän lauseensa muotoillaan uudelleen lauseissa 3.13 ja 3.14, ja ne ovat tämän tutkielman päätulokset. Parannusehdotuksen käytännön kannalta tärkeä tulos on lause 4.8 [9, Theorem 2]. Se on luettavissa luvussa 4, jossa esitellään myös muita hyödyllisiä ja varianssia pienentäviä versioita painotusotannasta.

Painotusotannasta löytyy vähän suomenkielistä aineistoa, eikä kaikille englanninkielisille termeille ole vakiintunutta suomenkielistä vastinetta. Tästä syystä käsitteen määrittelyn yhteydessä esitellään myös niiden englanninkieliset nimet. Tutkielman todistukset, joihin ei ole mainittu lähdettä, on tehty yhteistyössä tutkielman ohjaajan Matti Viholan kanssa.

LUKU 1

Taustatiedot ja merkinnät

Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ todennäköisyysavaruus, missä Ω on epätyhjä joukko, $\mathcal{F}$ σ-algebra joukossa Ω ja kuvaus $\mathbb{P} : \mathcal{F} \to [0,1]$ todennäköisyysmitta. Olkoon lisäksi satunnaismuuttujien avaruus $\mathbb{X}$ mitallinen osajoukko avaruudesta $\mathbb{R}^d$ ja satunnaismuuttuja $X = X(\omega)$, $X : \Omega \to \mathbb{X}$, jolla on jatkuva todennäköisyysjakauma tiheysfunktiolla p, sekä funktio $f : \mathbb{X} \to \mathbb{R}$ mitallinen. Mikäli muuttujan $f(X)$ odotusarvo on hyvin määritelty, voidaan se kirjoittaa muodossa

$$\mathbb{E}[f(X)] = \int_{\mathbb{X}} f(X(\omega))\, d\mathbb{P}(\omega) = \int_{\mathbb{X}} f(x)\, d\mathbb{P}_x(x) = \int_{\mathbb{X}} f(x)p(x)\, d\lambda(x),$$

missä λ on Lebesguen mitta. Täten nojaten Lebesgue-integroituvuuteen sanaparilla "hyvin määritelty" tarkoitetaan, että $\int_{\mathbb{X}} |f(x)|\, d\lambda(x) < \infty$. Selkeyden vuoksi merkinnän $d\lambda(x)$ sijasta käytetään Riemann-integraalin merkintää dx.

Jatkossa tiheysfunktiosta p puhutaan todennäköisyysjakaumana tai jakaumana, koska kontekstista on selvää, mitä tarkoitetaan. Lisäksi merkintää $\mathbb{E}_p[\cdot]$ käytetään tarkoittamaan sitä, että odotusarvo lasketaan todennäköisyysjakauman p suhteen, toisin sanoen odotusarvon sisällä oleva satunnaisluku noudattaa jakaumaa p. Tämän tutkielman kiinnostuksen kohteena on odotusarvon $\mathbb{E}_p[f(X)]$ ratkaiseminen Monte Carlo -menetelmillä. Oletetaan siis jatkossa reaaliarvoinen ja mitallinen funktio f ja tiheysfunktio p kiinteiksi ja käytetään odotusarvolle, jonka oletetaan jatkossa olevan hyvin määritelty, merkintää

$$I := \mathbb{E}_p[f(X)] = \int_{\mathbb{X}} f(x)p(x)\, dx. \tag{1.1}$$

Määritellään perinteisen Monte Carlo -menetelmän estimaattori. Jatkossa sanomalla $X_1, \ldots, X_n$ otos jakaumasta p tarkoitetaan riippumattomia ja samoin jakautuneita satunnaismuuttujia todennäköisyysjakaumasta p joukossa $\mathbb{X}$.

MÄÄRITELMÄ 1.1 (Monte Carlo). Olkoon $X_1, \ldots, X_n$ otos jakaumasta p ja funktio $f : \mathbb{X} \to \mathbb{R}$. Tällöin odotusarvon I Monte Carlo -estimaattori on

$$I_p^{(n)} := \frac{1}{n} \sum_{i=1}^{n} f(X_i). \tag{1.2}$$

Monte Carlo -menetelmä on siis otantamenetelmä, jolla saadaan likimääräinen ratkaisu ongelmaan. Menetelmä perustuu vahvaan suurten lukujen lakiin.

LAUSE 1.2 (Vahva suurten lukujen laki). *Olkoon $X_1, X_2, \dots$ riippumattomia ja samoin jakautuneita satunnaismuuttujia, joille $\mathbb{E}[|X_1|] < \infty$. Tällöin, kun $n \to \infty$,*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}[X_1] \quad \textit{melkein varmasti.}$$

TODISTUS. Katso [14, Theorem 1.13 (ii) s. 62]. □

Määritelmässä 1.1 satunnaismuuttujat X_i, $i = 1, \dots, n$ ovat riippumattomia ja samoin jakautuneita, joten sama pätee myös satunnaismuuttujiin $Y_i = f(X_i)$. Täten jos $\mathbb{E}_p[|f(X)|] < \infty$, niin vahvan suurten lukujen lain nojalla Monte Carlo -estimaattorille pätee $I_p^{(n)} \to I$ melkein varmasti, kun $n \to \infty$. Todistetaan estimaattori $I_p^{(n)}$ harhattomaksi ja todetaan sen varianssi.

LAUSE 1.3. *Olkoon satunnaismuuttujat $X_1, \dots, X_n$ ja estimaattori $I_p^{(n)}$ kuten määritelmässä 1.1. Tällöin estimaattori $I_p^{(n)}$ on harhaton eli*

$$\mathbb{E}\left[I_p^{(n)}\right] = I.$$

Lisäksi kun oletetaan, että $\sigma_p^2 = \mathrm{Var}[f(X_1)] < \infty$, niin

$$\mathrm{Var}\left[I_p^{(n)}\right] = \frac{\sigma_p^2}{n}.$$

TODISTUS. Koska satunnaismuuttujat $X_1, \dots, X_n$ ovat riippumattomia ja samoin jakautuneita satunnaismuuttujia, myös satunnaismuuttujat $f(X_1), \dots, f(X_n)$ ovat. Tällöin odotusarvolle $\mathbb{E}[I_p^{(n)}]$ pätee

$$\mathbb{E}\left[I_p^{(n)}\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_p\left[f\left(X\right)\right] = \frac{n}{n}\mathbb{E}_p\left[f\left(X\right)\right] = I.$$

Riippumattomuudesta seuraa myös, että $\mathrm{Var}\left[\sum_{i=1}^{n} f(X_i)\right] = \sum_{i=1}^{n}\mathrm{Var}[f(X_i)]$. Täten

$$\begin{aligned}\mathrm{Var}\left[I_p^{(n)}\right] &= \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left[f(X_i)\right] \\ &= \frac{n}{n^2}\mathrm{Var}\left[f(X_1)\right] = \frac{\sigma_p^2}{n}\end{aligned}$$ □

Varianssista nähdään, että Monte Carlo -estimaattorin varianssi $\mathrm{Var}[I_p^{(n)}]$ lähestyy nollaa, kun näytteiden lukumäärän n lähestyy ääretöntä ja $\sigma_p^2 < \infty$. Mitä pienempi satunnaismuuttujan $f(X)$ varianssi σ_p^2 on, sitä vähemmän näytteitä tarvitaan, jotta varianssi $\mathrm{Var}[I_p^{(n)}]$ saadaan haluttua pienemmäksi. Owen ja Zhou [9] käyttävät varianssista σ_p^2 nimitystä asymptoottinen varianssi. Tässä tutkielmassa asymptoottinen varianssi σ_*^2 mille tahansa estimaattorille $I_*^{(n)}$ määritellään seuraavasti:

MÄÄRITELMÄ 1.4 (Asymptoottinen varianssi). Jos jono $\sqrt{n}\left(I_*^{(n)} - I\right)$ suppenee jakaumaltaan normaalijakaumaan eli, kun $n \to \infty$,

$$\sqrt{n}\left(I_*^{(n)} - I\right) \to \mathcal{N}(0, \sigma_*^2) \quad \text{jakaumaltaan,}$$

on σ_*^2 estimaattorin $I_*^{(n)}$ asymptoottinen varianssi.

Yllä oleva määritelmä perustuu keskeiseen raja-arvolauseeseen.

LAUSE 1.5 (Keskeinen raja-arvolause). *Olkoon $X_1, X_2, \dots$ riippumattomia ja samoin jakautuneita satunnaismuuttujia siten, että $0 < \sigma^2 = \mathrm{Var}[X_1] < \infty$. Tällöin, kun $\mu = \mathbb{E}[X_1]$ ja $n \to \infty$,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \to N(0, \sigma^2) \quad \textit{jakaumaltaan.}$$

TODISTUS. Katso [14, Corollary 1.2 s. 69] □

Tämän tutkielman matemaattisten päätuloksien eli lauseiden 3.13 ja 3.14 yksityiskohtaiseen todistamiseen tarvitaan määritelmiä ja apulauseita liittyen matriiseihin ja stokastiseen kertaluokkaan. Konkretian vuoksi jatkossa tulevat todistukset tehdään ympäristössä, jossa satunnaismuuttujien avaruus $\mathbb{X} = \mathbb{R}$ ja p on tiheysfunktio. Tulokset yleistyvät myös tapauksiin, joissa mitallinen $\mathbb{X} \subset \mathbb{R}^d$ ja p on diskreetti jakauma. Matemaattisten symboleiden osalta pienillä aakkosilla ja kreikkalaisilla kirjaimilla merkitään yleensä joko vakioita, vektoreita tai kuvauksia, isoilla aakkosilla matriiseja, estimaattoreita tai satunnaismuuttujia ja lihavoiduilla isoilla aakkosilla satunnaisvektoreita tai -matriiseja.

1.1. Matriisien ominaisuuksia

Määritellään matriiseille Frobenius-normi $\|\cdot\|_F$, josta käytetään myös nimityksiä Euklidinen normi, l_2-normi, Schur-normi ja Hilbert-Schimidt -normi [8].

MÄÄRITELMÄ 1.6. Frobenius-normi $m \times n$ -matriisille A on

$$\|A\|_F := \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}$$

Kun Frobenius-normia käytetään vektorille $b \in \mathbb{R}^m$, se samaistetaan $m \times 1$ -matriisiksi ja saadaan

$$\|b\|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{1} |b_i|^2 = \|b\|_2^2.$$

Vektorien tapauksessa Frobenius-normi vastaa siis vektorien euklidista normia $\|\cdot\|_2$. Vastaava yhteys voidaan muodostaa myös $m \times n$ -matriisille A: Kootaan matriisin A alkiot a_{ij} sarakevektoreiksi $a_j = (a_{1j}, \dots, a_{mj})^\top \in \mathbb{R}^m$, jolloin $A = [a_1, \dots, a_n]$. Tällöin

$$\|A\|_F^2 = \sum_{j=1}^{n} \|a_j\|_2^2.$$

Frobenius-normi täyttää normin ehdot: $\|A\|_F \geq 0$ ja $\|A\|_F = 0$ jos ja vain jos $A = 0$, $\|cA\|_F = |c|\|A\|_F$ kaikille vakioille $c \in \mathbb{R}$ sekä kolmioepäyhtälö $\|A + B\|_F \leq \|A\|_F + \|B\|_F$ on voimassa. Todistetaan näistä jälkimmäisin, sillä muut ehdot ovat triviaaleja.

APULAUSE 1.7. *Frobenius-normi toteuttaa kolmioepäyhtälö eli matriiseille $A_{m\times n}$ ja $B_{m\times n}$ pätee*

$$\|A+B\|_F \leq \|A\|_F + \|B\|_F\,.$$

TODISTUS. Cauchy-Schwarzin epäyhtälön nojalla

$$\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}b_{ij}| \leq \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|b_{ij}|^2\right)^{1/2}.$$

Täten

$$\begin{aligned}
\|A+B\|_F^2 &= \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}+b_{ij}|^2 \\
&\leq \sum_{i=1}^{m}\sum_{j=1}^{n}\left(|a_{ij}|+|b_{ij}|\right)^2 \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2 + 2|a_{ij}||b_{ij}| + |b_{ij}|^2 \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2 + 2\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}b_{ij}| + \sum_{i=1}^{m}\sum_{j=1}^{n}|b_{ij}|^2 \\
&\leq \|A\|_F^2 + 2\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|b_{ij}|^2\right)^{1/2} + \|B\|_F^2 \\
&= \|A\|_F^2 + 2\,\|A\|_F\,\|B\|_F + \|B\|_F^2 \\
&= \left(\|A\|_F + \|B\|_F\right)^2.
\end{aligned}$$

Ottamalla neliöjuuri molemmilta puolilta on kolmioepäyhtälö todistettu. □

Frobenius-normilla on myös tärkeä ominaisuus: se on submultiplikatiivinen.

APULAUSE 1.8. *Frobenius-normi on submultiplikatiivinen eli matriiseille $A_{m\times p}$ ja $B_{p\times n}$ pätee*

$$\|AB\|_F \leq \|A\|_F\,\|B\|_F\,.$$

TODISTUS. Katso [15, s. 23]. □

Perehdytään seuraavaksi kääntyvän matriisin ominaisuuksiin. Tavoite on osoittaa, että kääntyvän matriisin läheisyydessä olevat matriisit ovat myös kääntyviä. Tämä väite on muotoiltu lauseeseen 1.10, mutta sen todistamista varten todistetaan ensin apulause 1.9. Sekä apulause 1.9 että lause 1.10 ovat peräisin Campbellilta ja Meyeriltä [4, Proposition 10.3.2 ja Theorem 10.3.1 s. 214–215], mutta eroavat normin suhteen. Campbell ja Meyer todistavat väitteet normille $\|\cdot\|$, joka on myös submultiplikatiivinen mutta jossa Frobenius-normista poiketen pätee $\|I_m\| = 1$, kun I_m on $m \times m$ -identiteettimatriisi. Apulauseen 1.9 ja lauseen 1.10 todistukset siis mukailevat heidän todistuksiaan, mutta tulos osoitetaan pätevän Frobenius-normille.

Apulause 1.9. *Olkoon I_m $m \times m$ -identiteettimatriisi ja A $m \times m$ -matriisi siten, että $\|A\|_F < 1$. Tällöin matriisi $(I_m - A)$ on kääntyvä ja*

$$\left\|(I_m - A)^{-1}\right\|_F \leq \sqrt{m} - 1 + \frac{1}{1 - \|A\|_F}.$$

Todistus. Oletuksesta $\|A\|_F < 1$ seuraa, että matriisit $A^n \to 0$ Frobenius-normin suhteen, kun $n \to \infty$, ja

$$\sum_{n=1}^{\infty} \|A^n\|_F \leq \sum_{n=1}^{\infty} \|A\|_F^n < \infty,$$

missä on hyödynnetty submultiplikatiivisuutta. Täten sarja $\sum_{n=0}^{\infty} A^n$ suppenee alkioittain itseisesti ja rajamatriisi $B = \sum_{n=0}^{\infty} A^n$ on olemassa sekä hyvin määritelty.

Olkoon $S_k = \sum_{n=0}^{k} A^n$. Tällöin $(I_m - A)S_k \to (I_m - A)B$, kun $k \to \infty$. Matriisille $(I_m - A)S_k$ pätee myös $(I_m - A)S_k = I_m - A^{k+1} \to I_m$, kun $k \to \infty$. Täten rajamatriisille $B = \lim_{k\to\infty} S_k$ täytyy päteä $B = (I_m - A)^{-1}$ eli matriisi $(I_m - A)$ on kääntyvä. Lisäksi, koska sarja on suppeneva, normin voi kolmioepäyhtälön nojalla viedä summan sisälle, ja hyödyntämällä submultiplikaattisuutta saadaan

$$\begin{aligned}
\|B\|_F &= \left\|\sum_{n=0}^{\infty} A^n\right\|_F \\
&\leq \sum_{n=0}^{\infty} \|A^n\|_F \\
&\leq \|I_m\|_F + \sum_{n=1}^{\infty} \|A\|_F^n \\
&= \sqrt{m} + \frac{1}{1 - \|A\|_F} - 1.
\end{aligned}$$

□

Lause 1.10. *Olkoon A ja B $m \times m$ -matriiseja ja matriisi A kääntyvä. Jos $||B||_F < 1/\left\|A^{-1}\right\|_F$ niin matriisi $(A + B)$ on kääntyvä. Lisäksi*

$$\left\|A^{-1} - (A + B)^{-1}\right\|_F \leq \frac{\sqrt{m}\,\|A^{-1}\|_F^2\,\|B\|_F}{1 - \|A^{-1}\|_F\,\|B\|_F}.$$

Todistus. Koska matriisi A on oletuksen nojalla kääntyvä, matriisi $(A + B) = A(I + A^{-1}B)$ on kääntyvä, jos matriisi $(I + A^{-1}B) = (I - (-A^{-1}B))$ on kääntyvä. Oletusta $||B||_F < 1/||A^{-1}||_F$ hyödyntäen saadaan

$$\| - A^{-1}B\|_F = \|A^{-1}B\|_F \leq \left\|A^{-1}\right\|_F \|B\|_F < \left\|A^{-1}\right\|_F \frac{1}{\|A^{-1}\|_F} = 1.$$

Apulauseen 1.9 nojalla matriisi $(I + A^{-1}B)$ on siis kääntyvä ja täten myös matriisi $(A + B)$ on kääntyvä.

Todistetaan seuraavaksi lauseen epäyhtälö. Havaitaan, että

$$\begin{aligned}
A^{-1} - (A+B)^{-1} &= \left[A^{-1}(A+B) - I\right](A+B)^{-1} \\
&= \left[I + A^{-1}B - I\right](A+B)^{-1} \\
&= A^{-1}B(A+B)^{-1} \\
&= A^{-1}B\left(A[I + A^{-1}B]\right)^{-1} \\
&= A^{-1}B\left(I + A^{-1}B\right)^{-1}A^{-1}.
\end{aligned}$$

Täten

$$\begin{aligned}
\left\|A^{-1} - (A+B)^{-1}\right\|_F &= \left\|A^{-1}B\left(I + A^{-1}B\right)^{-1}A^{-1}\right\|_F \\
&\le \left\|A^{-1}\right\|_F^2 \|B\|_F \left\|\left(I + A^{-1}B\right)^{-1}\right\|_F \\
&\le \left\|A^{-1}\right\|_F^2 \|B\|_F \left(\sqrt{m} - 1 + \frac{1}{1 - \|A^{-1}B\|_F}\right) \\
&\le \left\|A^{-1}\right\|_F^2 \|B\|_F \left(\sqrt{m} - 1 + \frac{1}{1 - \left\|A^{-1}\right\|_F \|B\|_F}\right) \\
&= \left\|A^{-1}\right\|_F^2 \|B\|_F \left(\frac{(\sqrt{m} - 1)\left(1 - \left\|A^{-1}\right\|_F \|B\|_F\right) + 1}{1 - \left\|A^{-1}\right\|_F \|B\|_F}\right) \\
&\le \left\|A^{-1}\right\|_F^2 \|B\|_F \left(\frac{\sqrt{m} - 1 + 1}{1 - \left\|A^{-1}\right\|_F \|B\|_F}\right) \\
&= \frac{\sqrt{m}\left\|A^{-1}\right\|_F^2 \|B\|_F}{1 - \left\|A^{-1}\right\|_F \|B\|_F},
\end{aligned}$$

missä on hyödynnetty apulauseen 1.9 epäyhtälöä, submultiplikaattisuutta $\|A^{-1}B\|_F \le \|A^{-1}\|_F \|B\|_F$ sekä epäyhtälöä $\left\|A^{-1}\right\|_F \|B\|_F < 1$. $\square$

Lause 1.10 osoittaa, että kun matriisi A on kääntyvä ja matriisi B on norminsa suhteen riittävän pieni, niin matriisi $C = A + B$ on kääntyvä. Toisin sanoen, jos matriisi A on kääntyvä ja matriisi C riittävän lähellä matriisia A eli matriisin $B = A - C$ normi $\|B\|_F$ on tarpeeksi pieni, niin matriisi C on kääntyvä. Kääntyvien matriisien joukko $\{A \in \mathbb{R}^{m\times m} : A \textit{ kääntyvä } \}$ on siis avoin. Lisäksi lauseesta seuraa käänteiskuvauksen jatkuvuus Frobenius-normin suhteen.

Seuraus 1.11. *Käänteiskuvaus $A \mapsto A^{-1}$ on jatkuva kääntyvien $m\times m$ -matriisien joukossa.*

Lauseella 1.10 on myös merkittävä osa seuraavan apulauseen todistuksessa. Sen väittämät on muotoiltu niin, että sitä voidaan käyttää tulevan apulauseen 1.23 todistuksessa.

Apulause 1.12. *Olkoon matriisi $A_* \in \mathbb{R}^{m\times m}$ kääntyvä. Tällöin on olemassa vakiot $s = s(A_*) > 0$ ja $c = c(A_*) < \infty$ siten, että matriisit $A \in \mathbb{R}^{m\times m}$, joille pätee $||A - A_*||_F \le s$, ovat kääntyviä ja kaikille vektoreille b, $b_* \in \mathbb{R}^m$ pätee*

$$\left\|A^{-1}b - A_*^{-1}b_*\right\|_F \le c\left\|A - A_*\right\|_F \|b\|_F + \left\|A_*^{-1}\right\|_F \|b - b_*\|_F .$$

TODISTUS. Todistetaan ensin vakion s olemassaolo ja matriisin A kääntyvyys. Koska matriisi A_* on kääntyvä, voidaan valita $s = \frac{1}{2||A_*^{-1}||_F} < \frac{1}{||A_*^{-1}||_F}$. Olkoon matriisi A siten, että $||A - A_*||_F \leq s$, ja merkitään $B = A - A_*$. Tällöin lauseen 1.10 nojalla matriisi $A = A_* + (A - A_*) = A_* + B$ on kääntyvä ja, koska $||B||_F = ||A - A_*||_F < 1/||A_*^{-1}||_F$, niin

$$\left\|A_*^{-1} - A^{-1}\right\|_F \leq \frac{\sqrt{m}\,\|A_*^{-1}\|_F^2\,\|A - A_*\|_F}{1 - \left\|A_*^{-1}\right\|_F\|A - A_*\|_F}.$$

Koska valittiin $s = \frac{1}{2||A_\infty^{-1}||_F}$, pätee tällöin

$$\frac{\sqrt{m}\,\|A_*^{-1}\|_F^2\,\|A - A_*\|_F}{1 - \left\|A_*^{-1}\right\|_F\|A - A_*\|_F} \leq \frac{\sqrt{m}\,\|A_*^{-1}\|_F^2\,\|A - A_*\|_F}{1 - \left\|A_*^{-1}\right\|_F \frac{1}{2||A_*^{-1}||_F}} = 2\sqrt{m}\left\|A_*^{-1}\right\|_F^2\|A - A_*\|_F.$$

Valitsemalla täten $c = 2\sqrt{m}\,\|A_*^{-1}\|_F^2 < \infty$ saadaan

$$\left\|A_*^{-1} - A^{-1}\right\|_F \leq c\|A - A_*\|_F.$$

Arvioidaan seuraavaksi väitteessä olevaa erotusta käyttämällä Frobenius-normin kolmioepäyhtälöä ja submultiplikatiivisuutta sekä yllä olevaa tulosta.

$$\begin{aligned}
\left\|A^{-1}b - A_*^{-1}b_*\right\|_F &= \left\|(A^{-1} - A_*^{-1})b - A_*^{-1}(b - b_*)\right\|_F \\
&\leq \left\|(A^{-1} - A_*^{-1})b\right\|_F + \left\|A_*^{-1}(b - b_*)\right\|_F \\
&\leq \left\|A^{-1} - A_*^{-1}\right\|_F\|b\|_F + \left\|A_*^{-1}\right\|_F\|b - b_*\|_F \\
&\leq c\,\|A - A_*\|_F\,\|b\|_F + \left\|A_*^{-1}\right\|_F\|b - b_*\|_F\,.
\end{aligned}$$ □

Käänteismatriisi on määritelty kaikille epäsingulaarisille neliömatriiseille. Kuitenkin tutkielman päätuloksessa lauseessa 3.13 ja sen todistamiseen tarvittavassa apulauseessa 1.23 ei haluta rajoittua pelkästään epäsingulaarisiin neliömatriiseihin. Määritellään tätä varten matriisille A yleistetty käänteismatriisi A^+.

MÄÄRITELMÄ 1.13 (Moore-Penrose käänteismatriisi). Olkoon A $m \times n$-matriisi. Sen Moore-Penrose käänteismatriisi on $n \times m$-matriisi A^+, jolle pätee ehdot

$$AA^+A = A \tag{1.3}$$

$$A^+AA^+ = A^+ \tag{1.4}$$

$$(AA^+)^\top = AA^+ \tag{1.5}$$

$$(A^+A)^\top = A^+A. \tag{1.6}$$

Yleistettyjä käänteismatriiseja on erilaisia, mutta tämän tutkielman kannalta Mooren vuonna 1935 ja Penrosen vuonna 1955 [4] kehittämä versio on riittävä. Moore-Penrose käänteismatriisin tärkeimpiä ominaisuuksia on sen yksikäsitteisyys ja yhtenevyys käänteismatriisiin silloin, kun matriisi A on kääntyvä.

LAUSE 1.14. *Jokaiselle $m \times n$-matriisille A on olemassa yksi ja vain yksi $n \times m$-matriisi A^+, joka täyttää ehdot* (1.3)-(1.6).

TODISTUS. Katso [13, Theorem 5.1 s. 202]. □

LAUSE 1.15. *Olkoon A $m \times m$-matriisi. Jos A on kääntyvä, niin $A^+ = A^{-1}$.*

TODISTUS. Käänteismatriisi A^{-1} toteuttaa ehdot (1.3)-(1.6) ja Moore-Penrose käänteismatriisi on yksikäsitteinen matriisille A. $\square$

1.2. Stokastinen kertaluokka ja rajoittuneisuus

Owenin ja Zhoun lauseessa [9, Theorem 1], joka vastaa tämän tutkielman lauseita 3.13 ja 3.14, käytetään stokastista kertaluokkaa (engl. *stochastic order*) ilmaisemaan satunnaismuuttujista muodostetun estimaattorin tarkkuutta. Määritellään stokastinen kertaluokka ja rajoittuneisuus satunnaismuuttujalle X sekä satunnaisvektoreille ja -matriiseille $\boldsymbol{X}$. Käytetään jatkossa merkintää $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ ei-negatiivisten reaalilukujen joukolle.

MÄÄRITELMÄ 1.16 (Stokastinen kertaluokka). Olkoon jono $(a_n)_{n\geq 1} \subset \mathbb{R}_+$. Satunnaismuuttujien jono $(X_n)_{n\geq 1}$ on stokastista kertaluokkaa $O_p(a_n)$, mitä merkitään $X_n = O_p(a_n)$, jos kaikille $\epsilon > 0$ on olemassa $s_\epsilon \in [0, \infty)$ ja $n_\epsilon \in \mathbb{N}$ siten, että kaikille $n > n_\epsilon$ pätee

$$\mathbb{P}(|X_n| > a_n s_\epsilon) < \epsilon.$$

Satunnaismuuttujien joukko $\{X_n\}$ on stokastisesti rajoitettu, jos $X_n = O_p(1)$.

MÄÄRITELMÄ 1.17 (Stokastinen kertaluokka vektoreille ja matriiseille). Olkoon $(\boldsymbol{X}_n)_{n\geq 1}$ satunnaisten vektorien tai matriisien jono ja vakioiden jono $(a_n)_{n\geq 1} \subset \mathbb{R}_+$. Jonolle $(\boldsymbol{X}_n)_{n\geq 1}$ merkitään $\boldsymbol{X}_n = O_p(a_n)$, jos $\|\boldsymbol{X}_n\|_F = O_p(a_n)$.

Stokastisen kertaluokan avulla stokastisista muuttujista muodostetun jonon asymptotiikkaa voidaan luokitella ja vertailla muihin jonoihin. Jono $(a_n)_{n\geq 1}$ viittaa satunnaismuuttujien suppenemis- tai kasvunopeuden kertaluokkaan, ja sen alkiot a_n noudattavat yleensä jotain näytteiden lukumäärän n funktiota $g(n)$. Esimerkiksi $a_n = n^{-1}$, $a_n = n^{-1/2}$, $a_n = n$ ja $a_n = n\log(n)$ ovat tärkeitä vertailujonoja [3, s. 459].

Todistetaan stokastinen kertaluokka satunnaismuuttujien keskiarvolle.

APULAUSE 1.18. *Olkoon $X_1, \ldots, X_n$ riippumattomia ja samoin jakautuneita satunnaismuuttujia, joille $E[X_1] = 0$ ja $0 < Var[X_1] < \infty$. Tällöin*

$$\frac{1}{n}\sum_{i=1}^{n} X_i = O_p\left(\frac{1}{\sqrt{n}}\right).$$

TODISTUS. Koska satunnaismuuttujat $X_1, \ldots, X_n$ ovat riippumattomia ja samoin jakautuneita, saadaan $\sigma_n^2 = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\mathrm{Var}\left[X_1\right]$. Oletuksen nojalla $0 < \sigma_n^2 < \infty$ kaikilla $n \in \mathbb{N}$. Olkoon $\epsilon > 0$. Tällöin löytyy reaaliluku $s > 0$, jolle $\frac{1}{s^2} < \epsilon$. Käyttämällä Chebyshevin epäyhtälöä [5, Theorem 2.1 s. 21], saadaan

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| > \frac{s\sqrt{\mathrm{Var}\left[X_1\right]}}{\sqrt{n}}\right) \leq \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq s\sigma_n\right) \leq \frac{1}{s^2} < \epsilon,$$

kun $n > 0$. Koska luku $s\sqrt{\mathrm{Var}\left[X_1\right]} > 0$ vastaa määritelmän 1.16 vakiota s_ϵ, väite pätee. $\square$

Lauseiden 3.13 ja 3.14 todistuksia varten perehdytään tarkemmin stokastisen kertaluokan hyödyllisiin ominaisuuksiin ja todistetaan ne.

APULAUSE 1.19. *Olkoon* a_n, $b_n > 0$ *kaikilla* n *ja reaaliarvoinen vakio* $c > 0$. *Jos satunnaismuuttujille* $X_n = O_p(a_n)$ *ja* $Y_n = O_p(b_n)$, *niin*

(1) $cX_n = O_p(ca_n) = O_p(a_n)$,
(2) $X_n + Y_n = O_p(a_n + b_n)$ *ja*
(3) $X_nY_n = O_p(a_nb_n)$.

TODISTUS. Todistetaan väitteen (1) ensimmäinen yhtäsuuruus $cX_n = O_p(ca_n)$. Olkoon $\epsilon > 0$. Koska $X_n = O_p(a_n)$, on olemassa $s_\epsilon^X > 0$ ja $n_\epsilon^X \in \mathbb{N}$ siten, että

$$\mathbb{P}\left(|X_n| > a_n s_\epsilon^X\right) < \epsilon$$

kaikille $n \geq n_\epsilon^X$. Valitaan $s_1 = s_\epsilon^X$. Tällöin

$$\mathbb{P}\left(|cX_n| > ca_n s_1\right) = \mathbb{P}\left(c|X_n| > ca_n s_1\right) = \mathbb{P}\left(|X_n| > a_n s_\epsilon^X\right) < \epsilon$$

kaikille $n > n_\epsilon^X$, joten $cX_n = O_p(ca_n)$. Jos taas valitaan $s_2 = cs_\epsilon^X$, havaitaan, että

$$\mathbb{P}\Big(|cX_n| > a_n s_2\Big) = \mathbb{P}\Big(c|X_n| > a_n c s_\epsilon^X\Big) < \epsilon$$

kaikille $n > n_\epsilon^X$. Osoitettiin siis, että $cX_n = O_p(a_n)$. Koska muuttuja cX on stokastiselta kertaluokaltaan sekä $O_p(ca_n)$ että $O_p(a_n)$, todetaan, ettei vakiolla c kertominen vaikuta kertaluokkaan. Täten myös väitteen toinen yhtäsuuruus $O_p(ca_n) = O_p(a_n)$ pätee.

Todistetaan seuraavaksi väite (2). Olkoon $\epsilon > 0$. Tällöin on olemassa $s_{\epsilon/2}^X > 0$ ja $n_{\epsilon/2}^X$ sekä $s_{\epsilon/2}^Y > 0$ ja $n_{\epsilon/2}^Y$ siten, että

$$\begin{aligned} \mathbb{P}\Big(|X_n| > a_n s_{\epsilon/2}^X\Big) &< \frac{\epsilon}{2} \quad \text{kaikille } n > n_{\epsilon/2}^X \quad \text{ja} \\ \mathbb{P}\Big(|Y_n| > b_n s_{\epsilon/2}^Y\Big) &< \frac{\epsilon}{2} \quad \text{kaikille } n > n_{\epsilon/2}^Y. \end{aligned} \tag{1.7}$$

Merkitään $s_3 = 2\max\{s_{\epsilon/2}^X, s_{\epsilon/2}^Y\}$ ja $n_\epsilon = \max\{n_{\epsilon/2}^X, n_{\epsilon/2}^Y\}$. Koska a_n ja b_n ovat positiivisia, pätee $\{|X_n| > (a_n + b_n)\frac{s_3}{2}\} \subset \{|X_n| > a_n s_{\epsilon/2}^X\}$. Täten

$$\begin{aligned} \mathbb{P}\Big(|X_n + Y_n| > (a_n + b_n)s_3\Big) &\leq \mathbb{P}\left(|X_n| > (a_n + b_n)\frac{s_3}{2}\right) + \mathbb{P}\left(|Y_n| > (a_n + b_n)\frac{s_3}{2}\right) \\ &\leq \mathbb{P}\Big(|X_n| > a_n s_{\epsilon/2}^X\Big) + \mathbb{P}\Big(|Y_n| > b_n s_{\epsilon/2}^Y\Big) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \end{aligned}$$

kun $n > n_\epsilon$, ja väite (2) on todistettu.

Todistetaan vielä viimeinen väite (3). Olkoon edelleen $\epsilon > 0$. Tällöin on olemassa $s_{\epsilon/2}^X > 0$ ja $n_{\epsilon/2}^X$ sekä $s_{\epsilon/2}^Y > 0$ ja $n_{\epsilon/2}^Y$ kuten kohdassa (1.7). Merkitään $s_4 = s_{\epsilon/2}^X s_{\epsilon/2}^Y$ ja $n_\epsilon = \max\{n_{\epsilon/2}^X, n_{\epsilon/2}^Y\}$. Täten

$$\begin{aligned} \mathbb{P}\Big(|X_nY_n| > a_nb_ns_4\Big) &= \mathbb{P}\left(|X_n||Y_n| > a_nb_ns_4, |X_n| \leq a_n s_{\epsilon/2}^X\right) \\ &\quad + \mathbb{P}\left(|X_nY_n| > a_nb_ns_4, |X_n| > a_n s_{\epsilon/2}^X\right) \\ &\leq \mathbb{P}\left(a_n s_{\epsilon/2}^X |Y_n| > a_nb_ns_4\right) + \mathbb{P}\left(|X_n| > a_n s_{\epsilon/2}^X\right) \\ &\leq \mathbb{P}\left(|Y_n| > b_n s_{\epsilon/2}^Y\right) + \mathbb{P}\left(|X_n| > a_n s_{\epsilon/2}^X\right) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \end{aligned}$$

kun $n > n_\epsilon$, sillä

$$\{|X_n||Y_n| > a_n b_n s_4\} \cap \{|X_n| \le a_n s^X_{\epsilon/2}\} \subset \{a_n s^X_{\epsilon/2}|Y_n| > a_n b_n s_4\}. \qquad \square$$

APULAUSE 1.20. *Olkoon $a_n > 0$ kaikilla n ja $(X_n)_{n\ge1}$ ja $(Y_n)_{n\ge1}$ satunnaismuuttujien jonoja. Tällöin*

(1) Jos $Y_n = O_p(a_n)$ ja $|X_n| \le |Y_n|$ kaikilla n, niin $X_n = O_p(a_n)$.
(2) Jos $X_n = O_p(a_n)$ ja $Y_n = O_p(a_n)$, niin $\max\{|X_n|, |Y_n|\} = O_p(a_n)$.

TODISTUS. Todistetaan ensin väite (1). Olkoon $\epsilon > 0$. Koska $Y_n = O_p(a_n)$, on olemassa $s^Y_\epsilon > 0$ ja $n^Y_\epsilon \in \mathbb{N}$ siten, että

$$\mathbb{P}\big(|Y_n| > a_n s^Y_\epsilon\big) < \epsilon$$

kaikille $n > n^Y_\epsilon$. Lisäksi, koska $|X_n| \le |Y_n|$ kaikilla n, niin kaikille vakiolle $c \in \mathbb{R}_+$ pätee

$$\{|X_n| > c\} \subseteq \{|Y_n| > c\}.$$

Valitaan $s^X_\epsilon = s^Y_\epsilon$ ja $n^X_\epsilon = n^Y_\epsilon$. Tällöin kaikille $n > n^X_\epsilon$ pätee

$$\mathbb{P}\big(|X_n| > a_n s^X_\epsilon\big) = \mathbb{P}\big(|X_n| > a_n s^Y_\epsilon\big) \le \mathbb{P}\left(|Y_n| > a_n s^Y_\epsilon\right) < \epsilon,$$

ja väite on todistettu.

Todistetaan seuraavaksi väite (2). Havaitaan, että kaikille n pätee

$$\max\{|X_n|, |Y_n|\} = \left|\max\left\{|X_n|, |Y_n|\right\}\right| \le \big||X_n| + |Y_n|\big| = |X_n| + |Y_n|. \tag{1.8}$$

Apulauseen 1.19 kohtien (2) ja (1) nojalla

$$|X_n| + |Y_n| = O_p(2a_n) = O_p(a_n).$$

Täten epäyhtälön (1.8) ja ensimmäisenä todistetun väitteen (1) perusteella väite $\max\{|X_n|, |Y_n|\} = O_p(a_n)$ on todistettu. $\square$

Yllä olevaa apulausetta 1.20 tarvitaan etenkin seuraavan apulauseen todistuksessa, joka liittyy stokastisen kertaluokan määritelmiin 1.16 ja 1.17. Satunnaisvektoreille ja -matriiseille stokastinen kertaluokka määritellään Frobenius-normin kautta, mutta koska niiden alkiot ovat satunnaismuuttujia, voidaan näiden kahden määritelmän välille osoittaa yhteys.

APULAUSE 1.21. *Satunnaismatriisin stokastinen kertaluokka on sama kuin sen alkioiden. Toisin sanoen $m \times l$ -satunnaismatriisien jonolle $(\boldsymbol{A})_{n\ge1}$ pätee*

$$\boldsymbol{A}_n = O_p(a_n) \Leftrightarrow [\boldsymbol{A}_n]_{ij} = O_p(a_n) \text{ kaikille } i \in \{1,\dots,m\} \text{ ja } j \in \{1,\dots,l\},$$

missä $[\boldsymbol{A}_n]_{ij}$ tarkoittaa matriisin $\boldsymbol{A}_n$ alkiota kohdassa i, j ja $(a_n)_{n\ge1} \subset \mathbb{R}_+$ on vakioiden jono.

TODISTUS. Todistetaan ensin vasemmalta oikealle ($\Rightarrow$). Oletetaan, että $\boldsymbol{A}_n = O_p(a_n)$. Toisin sanoen määritelmän 1.17 mukaan $\|\boldsymbol{A}_n\|_F = O_p(a_n)$. Koska Frobeniusnormille pätee

$$\max_{i,j}\left|[\boldsymbol{A}_n]_{ij}\right| \le \|\boldsymbol{A}_n\|_F$$

kaikilla n, niin apulauseen 1.20 kohdan (1) nojalla

$$\max_{i,j} \left| [\boldsymbol{A}_n]_{ij} \right| = O_p(a_n).$$

Vastaavasti koska $|\,[\boldsymbol{A}_n]_{ij}\,| \leq \max_{i,j} |\,[\boldsymbol{A}_n]_{ij}\,|$ kaikilla n sekä $i \in \{1, \dots, m\}$ ja $j \in \{1, \dots, l\}$, niin edelleen apulauseen 1.20 kohdan (1) nojalla kaikille i ja j pätee

$$[\boldsymbol{A}_n]_{ij} = O_p(a_n).$$

Todistetaan seuraavaksi väite oikealta vasemmalle ($\Leftarrow$). Oletetaan täten, että kaikille i ja j pätee $[\boldsymbol{A}_n]_{ij} = O_p(a_n)$. Käyttämällä apulauseen 1.20 kohtaa (2) rekursiivisesti saadaan

$$\max_{i,j} \left| [\boldsymbol{A}_n]_{ij} \right| = O_p(a_n).$$

Koska Frobenius-normille pätee

$$\|\boldsymbol{A}_n\|_F \leq \sqrt{ml} \max_{i,j} \left| [\boldsymbol{A}_n]_{ij} \right|,$$

saadaan käyttämällä apulauseen 1.19 kohtaa (1) ja apulauseen 1.20 kohtaa (1)

$$\|\boldsymbol{A}_n\|_F = O_p(a_n),$$

mikä määritelmän 1.17 nojalla tarkoittaa $\boldsymbol{A}_n = O_p(a_n)$. □

Apulauseen 1.21 mukaan satunnaismatriisin jokaisen alkion tai osamatriisin on siis supettava vähintään samaa kertaluokkaa kuin matriisin. Lisäksi lauseesta saadaan, että stokastisen kertaluokan ominaisuudet, jotka pätevät satunnaismuuttujille, pätevät myös satunnaismatriiseille. Koska vektori on $m \times 1$ -matriisi, yllä oleva tulos pätee myös satunnaisvektoreille.

Tarvittavat apulauseet lauseiden 3.13 ja 3.14 todistukseen on lähes kaikki muotoiltu ja todistettu. Lauseen 3.13 todistus tarvitsee vielä tärkeän apulauseen 1.23, jota varten todistetaan seuraava tulos.

Apulause 1.22. *Olkoon $(\boldsymbol{B}_n)_{n\geq 1}$ avaruuden $\mathbb{R}^m$ satunnaisvektorien jono ja $b \in \mathbb{R}^m$ vektori siten, että $\boldsymbol{B}_n - b = O_p(n^{-1/2})$. Tällöin $(\boldsymbol{B}_n)_{n\geq 1}$ on stokastisesti rajoitettu eli*

$$\boldsymbol{B}_n = O_p(1).$$

Todistus. Määritelmän 1.17 nojalla $\boldsymbol{B}_n - b = O_p(n^{-1/2})$ tarkoittaa samaa kuin $\|\boldsymbol{B}_n - b\|_F = O_p(n^{-1/2})$. Frobenius-normille pätee

$$\|\boldsymbol{B}_n\|_F \leq \|\boldsymbol{B}_n - b\|_F + \|b\|_F, \tag{1.9}$$

joten kiinnostuksen kohteena on selvittää vakion $\|b\|_F$ stokastinen rajoittuneisuus. Koska vektorijono $(b)_{n\geq 1}$ on triviaali satunnaisvektorien jono, valitsemalla $s = \|b\|_F$ saadaan kaikille $\epsilon > 0$ pätemään

$$\mathbb{P}\left(\|b\|_F > s\right) = \mathbb{P}\left(\|b\|_F > \|b\|_F\right) = 0 < \epsilon,$$

kun $n > 0$. Täten $\|b\|_F = O_p(1)$ ja apulauseen 1.19 kohdan (2) mukaan $\|\boldsymbol{B}_n - b\|_F + \|b\|_F = O_p\left(n^{-1/2} + 1\right)$. Näin epäyhtälön (1.9) ja apulauseen 1.20 kohdan (1) nojalla $\|\boldsymbol{B}_n\|_F = O_p\left(n^{-1/2} + 1\right)$.

Vakiolle $\epsilon > 0$ on siis olemassa $s_\epsilon^B > 0$ ja n_ϵ^B siten, että

$$\mathbb{P}\left(\|\boldsymbol{B}_n\|_F > (n^{-1/2}+1)s_\epsilon^B\right) < \epsilon$$

kaikille $n > n_\epsilon^B$. Koska $2 \geq n^{-1/2}+1$ kaikilla n, pätee

$$\mathbb{P}\left(\|\boldsymbol{B}_n\|_F > 2s_\epsilon^B\right) \leq \mathbb{P}\left(\|\boldsymbol{B}_n\|_F > (n^{-1/2}+1)s_\epsilon^B\right) < \epsilon,$$

kaikille $n > n_\epsilon^B$. Täten $\|\boldsymbol{B}_n\|_F = O_p(2) = O_p(1)$, missä viimeinen yhtäsuuruus saadaan apulauseen 1.19 kohdasta (1). □

APULAUSE 1.23. *Olkoon $(\boldsymbol{A}_n)_{n\geq 1}$ $m \times m$ -satunnaismatriisien ja $(\boldsymbol{B}_n)_{n\geq 1}$ m-ulotteisten satunnaisvektoreiden jonoja, jotka suppenevat kohti kääntyvää matriisia $A_* \in \mathbb{R}^{m\times m}$ ja rajavektoria $b_* \in \mathbb{R}^m$ siten, että $\boldsymbol{A}_n - A_* = O_p(n^{-1/2})$ ja $\boldsymbol{B}_n - b_* = O_p(n^{-1/2})$. Tällöin*

$$\boldsymbol{A}_n^+\boldsymbol{B}_n - A_*^{-1}b_* = O_p(n^{-1/2})$$

TODISTUS. Lauseen 1.14 mukaan on olemassa yksikäsitteinen Moore-Penrose käänteismatriisijono $(\boldsymbol{A}_n^+)_{n\geq 1}$ ja matriisi A_*^+. Matriisin A_* kääntyvyydestä seuraa lauseen 1.15 nojalla, että $A_*^+ = A_*^{-1}$. Lisäksi apulauseen 1.12 nojalla on olemassa $s^* = s^*(A_*)$ siten, että matriisit A, joille pätee $\|A - A_*\|_F \leq s^*$, ovat kääntyviä. Täten satunnaismatriiseille $\boldsymbol{A}_n$, joille pätee $\|\boldsymbol{A}_n - A_*\|_F \leq s^*$, pätee myös $\boldsymbol{A}_n^+ = \boldsymbol{A}_n^{-1}$ eli ne ovat kääntyviä. Kaikille vakiosta n riippuvaisille $t_n \in (0,\infty)$ voidaan siis suorittaa jako

$$\begin{aligned}
&\mathbb{P}(\|\boldsymbol{A}_n^+\boldsymbol{B}_n - A_*^{-1}b_*\|_F > t_n) \\
&\quad = \mathbb{P}(\|\boldsymbol{A}_n^{-1}\boldsymbol{B}_n - A_*^{-1}b_*\|_F > t_n, \|\boldsymbol{A}_n - A_*\|_F \leq s^*) \\
&\quad\quad + \mathbb{P}(\|\boldsymbol{A}_n^+\boldsymbol{B}_n - A_*^{-1}b_*\|_F > t_n, \|\boldsymbol{A}_n - A_*\|_F > s^*)
\end{aligned} \tag{1.10}$$

Tutkitaan ensin jälkimmäistä termiä, joka kattaa mahdollisesti singulaariset matriisit $\boldsymbol{A}_n$. Olkoon $\epsilon > 0$. Koska jono $(\boldsymbol{A}_n - A_*)_{n\geq 1}$ on stokastisesti rajoitettu jonolla $(n^{-1/2})_{n\geq 1}$, niin on olemassa $n_{\epsilon/2}^A$, $s_{\epsilon/2}^A$ siten, että

$$\mathbb{P}\left(\|\boldsymbol{A}_n - A_*\|_F > s_{\epsilon/2}^A n^{-1/2}\right) < \frac{\epsilon}{2}$$

kaikille $n > n_{\epsilon/2}^A$. Täten valitsemalla $\tilde{n}_{\epsilon/2}^A \geq n_{\epsilon/2}^A$ siten, että kaikilla $n > \tilde{n}_{\epsilon/2}^A$ pätee $s^* > s_{\epsilon/2}^A n^{-1/2}$, niin kaikille t_n

$$\begin{aligned}
&\mathbb{P}\left(\|\boldsymbol{A}_n^+\boldsymbol{B}_n - A_*^{-1}b_*\|_F > t_n, \|\boldsymbol{A}_n - A_*\|_F > s^*\right) \\
&\quad \leq \mathbb{P}\left(\|\boldsymbol{A}_n - A_*\|_F > s^*\right) \leq \mathbb{P}\left(\|\boldsymbol{A}_n - A_*\|_F > s_{\epsilon/2}^A n^{-1/2}\right) < \frac{\epsilon}{2},
\end{aligned} \tag{1.11}$$

kun $n > \tilde{n}_{\epsilon/2}^A$.

Yhtälön (1.10) ensimmäisessä termissä kaikki satunnaismatriisit $\boldsymbol{A}_n$ ovat kääntyviä. Matriiseille $\boldsymbol{A}_n$, joille $\|\boldsymbol{A}_n - A_*\|_F \leq s^*$, voidaan käyttää apulausetta 1.12: on olemassa $c = c(A_*) < \infty$ siten, että kaikille n pätee

$$\|\boldsymbol{A}_n^{-1}\boldsymbol{B}_n - A_*^{-1}b_*\|_F \leq c\,\|\boldsymbol{A}_n - A_*\|_F\,\|\boldsymbol{B}_n\|_F + \|A_*^{-1}\|_F\,\|\boldsymbol{B}_n - b_*\|_F\,. \tag{1.12}$$

Koska oletetaan, että $\boldsymbol{B}_n - b_* = O_p(n^{-1/2})$, niin apulauseesta 1.22 saadaan, että $\boldsymbol{B}_n = O_p(1)$. Vastaavasti vakiojonolle $(A_*^{-1})_{n\geq 1}$ pätee $A_*^{-1} = O_p(1)$, kuten osoitettiin apulauseen 1.22 todistuksessa. Lisäksi oletuksen mukaan $\boldsymbol{A}_n - A_* = O_p(n^{-1/2})$. Täten apulauseen 1.19 kohtien (1)-(3) perusteella $c\,\|\boldsymbol{A}_n - A_*\|_F\,\|\boldsymbol{B}_n\|_F + \|A_*^{-1}\|_F\,\|\boldsymbol{B}_n - b_*\|_F = O_p(n^{-1/2})$.

Kun $\epsilon > 0$, on siis olemassa $n^\alpha_{\epsilon/2}$ ja $s^\alpha_{\epsilon/2}$ siten, että

$$\mathbb{P}\left(c\,\|\boldsymbol{A}_n - A_*\|_F\,\|\boldsymbol{B}_n\|_F + \|A_*^{-1}\|_F\,\|\boldsymbol{B}_n - b_*\|_F > s^\alpha_{\epsilon/2} n^{-1/2}\right) < \frac{\epsilon}{2}$$

kaikille $n > n^\alpha_{\epsilon/2}$. Täten hyödyntäen epäyhtälöä (1.12) kaikille $t_n \geq s^\alpha_{\epsilon/2} n^{-1/2}$ saadaan, että yhtälön (1.10) ensimmäiselle termille pätee

$$\begin{aligned}
&\mathbb{P}(\|\boldsymbol{A}_n^{-1}\boldsymbol{B}_n - A_*^{-1}b_*\|_F > t_n, \|\boldsymbol{A}_n - A_*\|_F \leq s^*)\\
&\quad\leq \mathbb{P}(c\,\|\boldsymbol{A}_n - A_*\|_F\,\|\boldsymbol{B}_n\|_F + \|A_*^{-1}\|_F\,\|\boldsymbol{B}_n - b_*\|_F > t_n, \|\boldsymbol{A}_n - A_*\|_F \leq s^*)\\
&\quad\leq \mathbb{P}(c\,\|\boldsymbol{A}_n - A_*\|_F\,\|\boldsymbol{B}_n\|_F + \|A_*^{-1}\|_F\,\|\boldsymbol{B}_n - b_*\|_F > t_n) \qquad (1.13)\\
&\quad\leq \mathbb{P}(c\,\|\boldsymbol{A}_n - A_*\|_F\,\|\boldsymbol{B}_n\|_F + \|A_*^{-1}\|_F\,\|\boldsymbol{B}_n - b_*\|_F > s^\alpha_{\epsilon/2} n^{-1/2}) < \frac{\epsilon}{2}
\end{aligned}$$

kaikille $n > n^\alpha_{\epsilon/2}$.

Valitaan $n_\epsilon = \max\{\tilde{n}^A_{\epsilon/2}, n^\alpha_{\epsilon/2}\}$ ja $s_\epsilon = \max\{s^A_{\epsilon/2}, s^\alpha_{\epsilon/2}\}$. Koska yhtälö (1.10) pätee kaikille vakiosta n riippuvaisille t_n, niin se pätee myös, kun valitaan $t_n = s_\epsilon n^{-1/2}$. Tällöin koska $t_n \geq s^\alpha_{\epsilon/2} n^{-1/2}$ ja n_ϵ valittu siten, että $s^* > s^A_{\epsilon/2} n^{-1/2}$ kaikille $n > n_\epsilon$, voidaan käyttää epäyhtälöitä (1.11) ja (1.13) yhtälöön (1.10). Täten on osoitettu, että kaikille $\epsilon > 0$ on olemassa n_ϵ ja $0 < s_\epsilon < \infty$ siten, että pätee

$$\mathbb{P}(\|\boldsymbol{A}_n^+\boldsymbol{B}_n - A_*^{-1}b_*\|_F > s_\epsilon n^{-1/2}) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

kun $n > n_\epsilon$. Stokastisen rajoittuneisuuden määritelmän 1.17 perusteella on siis todistettu väite

$$\boldsymbol{A}_n^+\boldsymbol{B}_n - A_*^{-1}b_* = O_p(n^{-1/2}). \qquad \square$$

LUKU 2

Painotusotanta ja sen varianssi

> A general Monte Carlo tenet is: *never sample from a distribution merely because it arises in the physical context of a problem, for we may be able to use a better distribution in the computations and still get the right answer.* [6, s. 42]

Yllä oleva Hammersleyn ja Handscombin toteamus Monte Carlo -menetelmien toimintaperiaatteesta sopii hyvin painotusotantaan (engl. *importance sampling*). Muiden Monte Carlo -menetelmien tavoin, painotusotanta on suunniteltu estimoimaan integraalia tai tarkemmin todettuna satunnaismuuttujaan liittyvää odotusarvoa. Kun X on jatkuva satunnaismuuttuja tunnetulla todennäköisyysjakaumalla p joukossa $\mathbb{X}$, kiinnostuksen kohteena on odotusarvon integraali

$$I = \mathbb{E}_p[f(X)] = \int_{\mathbb{X}} f(x)p(x)\,dx,$$

missä $f : \mathbb{X} \to \mathbb{R}$ on mitallinen funktio, jolle yllä oleva odotusarvo on äärellinen.

Perinteisessä Monte Carlo -menetelmässä käytetään estimaattoria $I_p^{(n)}$ (1.2), mutta se ei ole aina sovellettavissa tai se ei ole paras mahdollinen. Ongelmana voi olla, että jakaumasta p on käytännössä mahdotonta tai hankala poimia otosta, jolloin Monte Carlo -estimaattia ei voida laskea. Toinen mahdollinen ongelma juontuu funktion f käyttäytymisestä siten, että Monte Carlo -estimaattorin varianssi on hyvin suuri. Menetelmä on tällöin tehoton, sillä luotettavan estimaatin saamiseksi tarvittaisiin hyvin suuri määrä näytteitä. Painotusotanta tarjoaa ratkaisun näihin molempiin tapauksiin.

2.1. Painotusotanta

Painotusotannan idea pohjautuu havaintoon

$$I = \int_{\mathbb{X}} f(x)p(x)dx = \int_{\mathbb{X}} \frac{f(x)p(x)}{q(x)} q(x)dx = \mathbb{E}_q\left[\frac{f(X)p(X)}{q(X)}\right], \tag{2.1}$$

missä joukon $\mathbb{X}$ todennäköisyysjakaumalle q oletetaan pätevän $q(x) > 0$, kun funktio $f(x)p(x) \neq 0$. Yllä ja jatkossa määritellään $\frac{a}{b} = 0$, kun $a = 0 = b$. Odotusarvoa laskettaessa voidaan siis vaihtaa todennäköisyysmittaa, kunhan tehdään tarvittavat korjaukset integroitavaan. Simuloinnissa tämä tarkoittaa sitä, että todennäköisyysjakauman p sijasta otos voidaan poimia todennäköisyysjakaumasta q. Se, että otos poimitaan eri jakaumasta, korjataan painottamalla integroitavaa funktiota $f(x)$ painolla $p(x)/q(x)$. Näin saadaan painotusotannan estimaattori.

MÄÄRITELMÄ 2.1 (Painotusotanta). Oletetaan joukon $\mathbb{X}$ todennäköisyysjakaumalle q, että kaikille $x \in \mathbb{X}$ pätee $q(x) > 0$, kun $f(x)p(x) \neq 0$. Olkoon $X_1, \ldots, X_n$ otos jakaumasta q. Tällöin odotusarvon I painotusotannan estimaattori on

$$I_{p,q}^{(n)} := \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)p(X_i)}{q(X_i)}. \tag{2.2}$$

Todennäköisyysjakaumaa q kutsutaan tässä tutkielmassa ehdotusjakaumaksi (engl. *proposal distibution* [17]). Muita nimityksiä on painotusjakauma (engl. *importance distribution* [10]) tai otantajakauma (engl. *sampling distribution* [9]). Jakaumasta p puhutaan puolestaan kohdejakaumana (engl. *nominal distribution* [9, 10]) ja suhdetta $\frac{p(x)}{q(x)}$ kutsutaan otantapainoksi tai painoksi (engl. *importance weight* [17], *likehood ratio* [10]).

Painokertoimen takia tämä tutkielma käyttää menetelmälle nimitystä *painotusotanta*. Toinen suomenkielen käännös *tärkeysotanta* [1] tuo esille englanninkielisen nimen *importance sampling* alkuperän. Bartonin, Nakayaman ja Schrubenin [2] mukaan englanninkielinen nimitys on peräisin menetelmän varhaisesta käyttötarkoituksesta: menetelmässä yritetään poimia otos satunnaismuuttujien avaruudesta $\mathbb{X}$ siten, että näytteiden lukumäärä kustakin avaruuden alueesta on suhteessa alueen "tärkeyden" (engl. *importance*) ja sieltä poimimisen todennäköisyyden tuloon.

Barton ym. [2] ajoittavat painotusotantamenetelmän syntymisen haluun parantaa simuloinnin tilastollista tehokkuutta 1940- ja 1950-lukujen vaihteessa. Tällöin tehokkuuden parantamista motivoi kolmen tahon yhdistyminen: (1) halu ymmärtää atomin fuusio- ja fissioreaktioita Monte Carlo -menetelmän avulla, (2) numeeristen tietokoneiden kehittyminen ja (3) silloiset edistysaskeleet soveltavassa todennäköisyyslaskennassa ja tilastotieteessä kuten harvinaisen tapahtuman todennäköisyyden estimointi (engl. *rare-event-probability estimation*), jonomallit (engl. *queueing models*) ja Monte Carlo -menetelmät. Myös yhdysvaltalaisen RAND-organisaation tutkimussopimukset Yhdysvaltojen ilmavoimien kanssa tuki menetelmien tilastollisen tehokkuuden varhaista kehittämistä. Esimerkiksi juuri painotusotannan konsepti esitellään vuoden 1945 RAND raportissaan. [2]

Painotusotannan alkuperää ei pystytä tarkasti määrittelemään. Barton ym. [2] lainaa G. Goertzelin vuoden 1949 julkaisua, jossa kerrotaan painotusotannan tutkimustyön olevan loogista jatkumoa H. Kahn ja T. E. Harrisin RAND raporteista sekä keskusteluista G. Goertzelin ja H. Kahn välillä. Alussa painotusotannan tutkiminen keskittyi etupäässä todennäköisyysmittojen vaihtamisen empiiriseen arviointiin, mutta myöhemmät tutkimukset perehtyivät myös teoreettisiin ominaisuuksiin [2].

Painotusotannan estimaattorilla $I_{p,q}^{(n)}$ (2.2) on hyviä ominaisuuksia.

APULAUSE 2.2. *Estimaattori $I_{p,q}^{(n)}$ määritelmässä 2.1 on harhaton eli*

$$\mathbb{E}_q \left[I_{p,q}^{(n)} \right] = I,$$

ja vahvasti tarkentuva eli, kun $n \to \infty$,

$$I_{p,q}^{(n)} \to I \quad \text{melkein varmasti.}$$

TODISTUS. Olkoon $X_1, \dots, X_n$ otos todennäköisyysjakaumasta q. Todistetaan ensin harhattomuus. Hyödyntämällä yhtälöä (2.1) sekä satunnaismuuttujien X_i samoin jakautuneisuutta kaikilla $i = 1, \dots, n$, saadaan

$$\begin{aligned}\mathbb{E}\left[I_{p,q}^{(n)}\right] &= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\frac{f(X_i)p(X_i)}{q(X_i)}\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_q\left[\frac{f(X)p(X)}{q(X)}\right] \\ &= \frac{n}{n}\mathbb{E}_p\left[f(X)\right] = I.\end{aligned}$$

Todistetaan seuraavaksi vahva tarkentuvuus. Koska satunnaismuuttujien X_i riippumattomuudesta ja samoin jakautuneisuudesta seuraa, että myös satunnaismuuttujat $Y_i = \varphi(X_i) = \frac{f(X_i)p(X_i)}{q(X_i)}$ ovat muuttujan X_i deterministisinä muunnoksina riippumattomia ja samoin jakautuneita. Koska I on hyvin määritelty, niin $\mathbb{E}_q[|Y_1|] = \mathbb{E}_p[|f(X)|] < \infty$ ja suurten lukujen lain 1.2 sekä yhtälön (2.1) mukaan

$$I_{p,q}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} Y_i \to \mathbb{E}_q\left[Y_i\right] = \mathbb{E}_q\left[\frac{f(X)p(X)}{q(X)}\right] = I$$

melkein varmasti, kun $n \to \infty$. □

Harhattoman estimaattorin $I_{p,q}^{(n)}$ harha eli odotusarvon poikkeama estimoitavan parametrin I arvosta on siis nolla. Se ei siis tuota keskimäärin liian pieniä tai liian suuria estimaatteja. Estimaattorin $I_{p,q}^{(n)}$ tarkentuvuus puolestaan tarkoittaan sitä, että kasvattamalla otoksen kokoa estimaatti suppenee kohti tuntemattoman parametrin oikeaa arvoa. Toisin sanoen otosmäärää kasvattamalla estimaattori antaa luotettavan tuloksen. Harhattomuus ja tarkentuvuus ei kuitenkaan takaa hyvää estimaattoria. Esimerkiksi Monte Carlo -estimaattori (1.2) on myös harhaton ja tarkentuva, mutta ongelmallinen tietyissä tilanteissa.

Monte Carlo -menetelmän tavoin painotusotannalla on kaksi käytännön rajoitetta: todennäköisyysjakaumasta q täytyy pystyä poimimaan satunnaismuuttuja X ja poimituille muuttujille laskemaan $f(X)p(X)/q(X)$. Jos $f(X)$ voidaan laskea, on jälkimmäisen rajoituksen kannalta riittävää pystyä laskemaan suhde $p(X)/q(X)$, mikä voi olla joskus yksinkertaisempaa kuin arvojen $p(X)$ ja $q(X)$ laskeminen erikseen. Painotusotannalla on kuitenkin ominaisuuksia, joiden takia se toimii estimaattorina paremmin kuin perinteinen Monte Carlo. Esimerkiksi, kun jakaumasta p ei voida tai on hankala poimia muuttujia, niin painotusotannan todennäköisyysmitan vaihtaminen selvästi ratkaisee ongelman. Painotusotanta voi myös olla parempi estimaattori tilanteissa, joissa ongelmana on funktion f käyttäytymisestä johtuva korkea varianssi.

Owen ja Zhou [9] nostavat esimerkiksi tilanteen, jossa funktio f on piikikäs siinä mielessä, että sen arvo on lähellä nollaa (tai vakiota) suurimman osan alueesta $\mathbb{X}$, mutta pienellä alueella $A \subset \mathbb{X}$ funktion f arvon vaihtelu on suurta. Lisäksi jos osajoukolle A pätee, että sillä on pieni todennäköisyys todennäköisyysjakauman p suhteen, kyse on harvinaisesta tapahtumasta. Yleisesti ajateltuna, kun harvinaisen tapahtuman $X \in A$ todennäköisyys on pieni, jakaumasta p poimittu otos ei välttämättä sisällä yhtään muuttujaa X alueesta A, jolloin Monte Carlo -estimaatti on nolla. Otoskoon n olisi siis oltava hyvin suuri luotettavan estimaatin saamiseksi.

Jos yllä kuvaillussa tilanteessa käytetään kohdejakauman p sijasta ehdotusjakaumaa q, jolle tapahtuman $X \in A$ todennäköisyys on suurempi, voidaan välttää suuren otoskoon ongelma. Koska jakauma q antaa enemmän informaatiota tärkeästä alueesta A, missä funktion f arvojen vaihtelu sijaitsee, on tällöin todennäköisempää otoksen sisältävän muuttujia alueesta A ja pienempi otoskoko riittää takaamaan luotettavan estimaatin. Owenin [10] mukaan tällaisia tapauksia nousee esimerkiksi hiukkasfysiikassa, bayesiläisessä tilastotieteessä, finanssi- ja vakuutusalan harvinaisen tapahtuman simuloinnissa sekä tietokonegrafiikan renderöinnissa.

2.2. Varianssi

Menetelmän tehokkuuden tarkastelussa otetaan huomioon estimaattorin muodostamiseen käytetty työmäärä ja estimaattorin varianssi. Etenkin jos estimaattorin varianssi on korkea, menetelmä on ongelmallinen. Hammersley ja Handscomb [6] toteavatkin, että monimutkaisempaan estimaattoriin käytetty työmäärä maksaa itsensä takaisin, jos sen varianssi on huomattavasti pienempi kuin yksinkertaisemman estimaattorin.

Koska Monte Carlo -menetelmät pohjautuvat suurten lukujen lakiin ja keskeiseen raja-arvolauseeseen, varianssi määrää menetelmän estimaattorin asymptoottiseen tehokkuuteen. Toisin sanoen keskeisen raja-arvolauseen 1.5 mukaan, kun $n \to \infty$, niin

$$\frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \to N(0,1) \quad \text{jakaumaltaan.}$$

Tämän tiedon nojalla voidaan määrittää haluttu tarkkuus, joka riippuu satunnaismuuttujan X_1 asymptoottisen varianssin σ^2 neliöjuuresta σ sekä luvusta $1/\sqrt{n}$. Tästä puolestaan nähdään, miksi varianssin σ^2 pienentäminen parantaa menetelmän tehokuutta: mitä pienempi σ on, sitä pienempi $\sqrt{n}$ eli otoskoko n voi olla.

LAUSE 2.3. *Painotusotannan estimaattorin* $I_{p,q}^{(n)}$ (2.2) *varianssi on*

$$\mathrm{Var}\left[I_{p,q}^{(n)}\right] = \frac{\sigma_{p,q}^2}{n},$$

missä

$$\sigma_{p,q}^2 = \mathbb{E}_p\left[\frac{f(X)^2 p(X)}{q(X)}\right] - I^2. \tag{2.3}$$

TODISTUS. Olkoon $X_1, \dots, X_n$ otos todennäköisyysjakaumasta q. Hyödyntämällä varianssin kaavaa $\mathrm{Var}\left[a(X+Y)\right] = a^2\left(\mathrm{Var}[X] + \mathrm{Var}[Y]\right)$ vakiolle $a \in \mathbb{R}$ ja riippumattomille satunnaismuuttujille X ja Y, saadaan

$$\mathrm{Var}\left[I_{p,q}^{(n)}\right] = \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n}\frac{f(X_i)p(X_i)}{q(X_i)}\right] = \frac{1}{n}\mathrm{Var}\left[\frac{f(X_1)p(X_1)}{q(X_1)}\right].$$

Vastaavasti hyödyntämällä yhtälöä (2.1) saadaan

$$\begin{aligned}
\sigma_{p,q}^2 &= \operatorname{Var}\left[\frac{f(X_1)p(X_1)}{q(X_1)}\right] \\
&= \mathbb{E}\left[\left(\frac{f(X_1)p(X_1)}{q(X_1)} - \mathbb{E}\left[\frac{f(X_1)p(X_1)}{q(X_1)}\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{f(X_1)p(X_1)}{q(X_1)}\right)^2 - 2\left(\frac{f(X_1)p(X_1)}{q(X_1)}\right)\mathbb{E}_q\left[\frac{f(X)p(X)}{q(X)}\right] + \mathbb{E}_q\left[\frac{f(X)p(X)}{q(X)}\right]^2\right] \\
&= \mathbb{E}\left[\left(\frac{f(X_1)p(X_1)}{q(X_1)}\right)^2\right] - 2\mathbb{E}_q\left[\frac{f(X)p(X)}{q(X)}\right] I + I^2 \\
&= \mathbb{E}_q\left[\left(\frac{f(X)^2p(X)}{q(X)}\right)\frac{p(X)}{q(X)}\right] - 2I^2 + I^2 \\
&= \mathbb{E}_p\left[\frac{f(X)^2p(X)}{q(X)}\right] - I^2. \qquad \square
\end{aligned}$$

Painotusotannan estimaattorin varianssi on siis riippuvainen otoskoosta n ja varianssista $\sigma_{p,q}^2$. Kuten yhtälöstä (2.3) havaitaan, varianssin $\sigma_{p,q}^2$ pienentämiseksi sopiva ehdotusjakauma q tulee valita funktiolle fp siten, että odotusarvo $\mathbb{E}_p\left[\frac{f(X)^2p(X)}{q(X)}\right]$ on mahdollisimman pieni.

Lause 2.4. *Olkoon funktio $f : \mathbb{X} \to \mathbb{R}$, jolle pätee $\mathbb{E}_p[|f(X)|] > 0$. Tällöin painotusotannan estimaattorin $I_{p,q}^{(n)}$ (2.2) ehdot täyttävien ehdotusjakaumien q joukosta ehdotusjakauma*

$$q^*(x) := \frac{|f(x)|p(x)}{\mathbb{E}_p[|f(X)|]} \propto |f(x)|p(x)$$

minimoi varianssin $\sigma_{p,q}^2$ (2.3).

Todistus. Katso [11, Theorem 3.3.4 s. 84]. □

Kuten yllä oleva lause osoittaa, varianssin $\sigma_{p,q}^2$ pienentämisen kannalta mikään muu jakauma kuin q^* ei ole paras mahdollinen valinta ehdotusjakaumaksi. Kun pohditaan yleistä erikoistapausta, jossa $f(x) \geq 0$ ja $\mathbb{E}_p[f(X)] > 0$, voidaan kirjoittaa ehdotusjakaumaksi $q^*(x) = f(x)p(x)/I$. Tällä ehdotusjakaumalla varianssi $\sigma_{p,q^*}^2 = 0$, mutta kyseinen jakauma ei ole hyödyllinen käytännössä, sillä I on tuntematon suure. Kuitenkin tieto siitä, että paras ehdotusjakauma q^* on verrannollinen funktioon $|f|p$, viittaa siihen, että hyvä ehdotusjakauma q on suurpiirteisesti verrannollinen funktioon $|f|p$. Esimerkiksi, jos joillakin avaruuden $\mathbb{X}$ alueilla D funktion f arvo on nolla tai sen itseisarvo on hyvin pieni, on hyvä valita sellainen q, joka ei lisää yhtään tai lisää hyvin vähän painoa kyseisellä alueella D.

Parhaimpaan tulokseen päästään siis, jos jakauma q on lähes verrannollinen funktioon $|f|p$. Tämä verrannollisuus ei kuitenkaan takaa hyvää estimaattoria. Owen ja Zhou osoittavan esimerkin [9, Example 1] avulla, miten painotusotanta epäonnistuu silmiinpistävästi, vaikka q on lähes verrannollinen funktioon fp.

Esimerkki 2.5. Olkoon $\mathbb{X} = (0,1)^5$ kohdejakaumalla $p = U(0,1)^5$ ja funktio

$$f_1(x) = 0,9 \times \prod_{j=1}^{5} B(x^j, 20, 20) + 0,1 \times \prod_{j=1}^{5} B(x^j, 2, 2).$$

Yllä funktio B on betajakauman tiheysfunktio parametreillä $a > 0$ ja $b > 0$

$$B(x, a, b) = \frac{x^{a-1}(1-x)^{b-1}}{\Gamma(a)\Gamma(b)/\Gamma(a+b)}, \qquad 0 < x < 1, \tag{2.4}$$

missä $\Gamma(x)$ on gammafunktio. Koska betafunktiolle $\mathrm{B}(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, voidaan suoralla laskulla osoittaa, että $\mathbb{E}_p[f_1(X)] = \int_{\mathbb{X}} f_1(x)p(x)\,dx = 1$. Valitaan ehdotusjakaumaksi $q(x) = \prod_{j=1}^{5} B(x^j, 20, 20)$. Ehdotusjakauma q on tällöin todella läheisesti verrannollinen jakaumaa $f_1 p$, mutta satunnaismuuttujan $f_1(X)p(X)/q(X)$ varianssi $\sigma^2_{p,q} = \infty$.

Syy yllä olevan esimerkin äärettömään varianssiin ja painotusotannan epäonnistumiseen on ehdotusjakauman q sijainti varianssin $\sigma^2_{p,q}$ lausekkeen (2.3) nimittäjässä. Jos nimittäjän jakauma $q(x)$ suppenee nopeammin nollaan kuin osoittajan funktio $f^2(x)p^2(x)$ jollain muuttujan x integroimisalueella, niin varianssi $\mathrm{Var}[f(X)p(X)/q(X)]$ voi olla ääretön. Esimerkissä 2.5 näin tapahtuu hännissä. Ääretön varianssi voi siis syntyä alueen $\mathbb{X}$ vähemmän tärkeässä osassa, jossa funktion $|f|p$ arvo on pieni, mutta jakauman q arvo on vielä pienimpi kuin funktion $|f|p$.

Hyvän ehdotusjakauman valitseminen vaatii valistuneita arvauksia ja mahdollisesti numeerisen haun (engl. *numerical search*). Tietämys funktion f käyttäytymisestä alueella $\mathbb{X}$ tai joku muu mallin oleellinen yksityiskohta voi auttaa ehdotusjakauman valinnassa. Esimerkiksi kun kohdejakauma p on likipitäen normaalijakauma, yleinen taktiikka on valita ehdotusjakaumaksi q Studentin t-jakauma, sillä tällöin jakauman q hännät ovat paksummat kuin jakauman p [10].

Luvussa 4 käsitellään parannusehdotuksia etenkin ehdotusjakauman valintaan ja esitetään menetelmiä, jotka toimivat huomattavasti paremmin esimerkin 2.5 tapauksessa. Ennen sitä perehdytään kuitenkin toisenlaiseen menetelmään - vakioiviin muuttujiin, jotka ovat tämän tutkielman kannalta keskeinen varianssin pienennystekniikka. Seuraava luku esittelee tämän menetelmän ja tuo esille sen, miten niistä voi olla hyötyä painotusotannassa.

LUKU 3

Painotusotanta ja vakioivat muuttujat

Samoihin aikoihin painotusotannan kanssa kehiteltiin myös muita menetelmiä Monte Carlon tehokkuuden parantamiseksi [2]. Fishman [5] toteaa, että varianssin pienentämiseen kehitettyjen menetelmien toimintaperiaatteet perustuvat yleensä toiseen kahdesta eri metodista.

Ensimmäisessä metodissa otoksen poimintaa ja estimaattoria muutetaan siten, että saadaan pienempi varianssi näytettä kohden. Painotusotanta ehdotusjakaumalla q ja estimaattorilla $I_{p,q}^{(n)}$ on esimerkki tästä. Toisessa metodissa ei puututa otoksen poimintaan vaan hyödynnetään apumuuttujaa. Lisäämällä sopivan apumuuttujan avulla saatu lisäinformaatio tuntemattoman parametrin estimaattoriin, voidaan pienentää varianssia. Vakioivat muuttujat (engl. *control variates*) on esimerkki tällaisesta menetelmästä. [5, s. 5–6]

Yhdistämällä samaan estimaattoriin kaksi menetelmää, jotka ovat erilaisia metodiltaan, voidaan hyötyä molemmista varianssin pienennysmetodeista. Painon $p(x)/q(x)$ käyttäminen vakioivana muuttujana painotusotannassa suojaa menetelmän epäonnistumiselta etenkin, kun otos poimitaan sekoitejakaumasta [9]. Tätä tarkastellaan seuraavassa luvussa 4, mutta ennen sitä perehdytään vakioiviin muuttujiin ja siihen, miten yhdistää ne painotusotantaan. Tutkielman päätulokset eli lauseet 3.13 ja 3.14 osoittavat, että estimoiduilla vakioivien muuttujien kertoimilla varustettu estimaattori on asymptoottisesti sama kuin estimaattori optimaalisilla kertoimilla.

3.1. Vakioivat muuttujat

Vakioivat muuttujat -menetelmä hyödyntää funktiosta $g : \mathbb{X} \to \mathbb{R}$ ja satunnaismuuttujasta $g(X)$ saatavaa lisäinformaatiota varianssin pienentämiseksi. Käyttämällä funktiota g voidaan kiinnostuksen kohteena oleva integraali I (1.1) esittää muodossa

$$I = \int_{\mathbb{X}} f(x)p(x)\,dx = \int_{\mathbb{X}} \big(f(x) - g(x)\big)p(x)\,dx + \int_{\mathbb{X}} g(x)p(x)\,dx.$$

Menetelmän ideana on estimoida ensimmäinen integraali Monte Carlo -menetelmällä ja laskea jälkimmäinen integraali matemaattisesti. Tätä varten funktion g on oltava tarpeeksi yksinkertainen, jotta funktion gp integraali yli alueen $\mathbb{X}$ on laskettavissa. Toisaalta, jotta saadaan aikaan haluttu varianssin pienentyminen, funktio g on matkittava funktiota f ja "selitettävä" suurin osa sen vaihtelusta. Satunnaismuuttujaa $g(X)$ kutsutaankin vakioivaksi muuttujaksi, jos sen odotusarvo $\mathbb{E}_p[g(X)] = \theta$ on tunnettu ja funktio g käyttäytyy likipitäen samoin kuin funktio f.

Vaatimus funktioiden g ja f samankaltaisesta käyttäytymisestä voidaan perustella, kun tutkitaan muuttujan $f(X) - g(X) + \theta$ Monte Carlo -estimaattorin (1.2) varianssia. Olkoon $X_1, \dots, X_n \sim p$. Tällöin

$$\begin{aligned}\operatorname{Var}\left[\frac{1}{n}\sum_{i=1}^{n} f(X_i) - g(X_i) + \theta\right] &= \frac{1}{n}\operatorname{Var}\big[f(X_1) - g(X_1)\big] \qquad (3.1)\\ &= \frac{1}{n}\Big(\operatorname{Var}\big[f(X_1)\big] + \operatorname{Var}\big[g(X_1)\big] - 2\operatorname{Cov}\big[f(X_1), g(X_1)\big]\Big).\end{aligned}$$

Havaitaan siis, että jos muuttujien väliselle kovarianssille ja muuttujan $g(X)$ varianssille pätee

$$\operatorname{Var}\big[g(X)\big] < 2\operatorname{Cov}\big[f(X), g(X)\big], \qquad \text{missä } X \sim p,$$

niin muuttujan $f(X)-g(X)+\theta$ Monte Carlo -estimaattorin varianssi on pienempi kuin yksittäisen muuttujan $f(X)$ Monte Carlo -estimaattorin varianssi. Toisin sanoen, mitä suurempi positiivinen lineaarinen riippuvuussuhde muuttujien $f(X)$ ja $g(X)$ välillä on, sitä pienempi varianssi saadaan vakioivat muuttujat -menetelmällä.

Hammersley ja Handscomb [6, s. 52, 60] antavat seuraavan teoreettisen esimerkin menetelmästä.

Esimerkki 3.1. Olkoon tutkittava funktio $f : \mathbb{X} \to \mathbb{R}$, $f(x) = \frac{e^x - 1}{e-1}$ ja satunnaismuuttuja X tasajakaumasta $U(0,1)$. Teoreettisesti laskemalla saadaan, että

$$\mathbb{E}[f(X)] = \int_0^1 f(x)\,dx = \Big|_0^1 \frac{e^x - x}{e-1} = \frac{e-2}{e-1} \approx 0{,}41802$$

ja

$$\operatorname{Var}[f(X)] = \int_0^1 \left(f(x) - \mathbb{E}\left[f(X)\right]\right)^2 dx = \frac{3-e}{2e-2} \approx 0{,}08198.$$

Valitsemalla vakioivaksi muuttujaksi $g(X) = X$, jolle $\mathbb{E}[g(X)] = 0{,}5$, saadaan

$$\begin{aligned}\mathbb{E}\left[f(X) - g(X) + 0{,}5\right] &= \int_0^1 f(x) - g(x)\,dx + 0{,}5\\ &= \frac{e-3}{2e-2} + 0{,}5 \approx 0{,}41802\end{aligned}$$

ja koska vakiolle $a \in \mathbb{R}$ pätee $\operatorname{Var}[X + a] = \operatorname{Var}[X]$, niin varianssi $\operatorname{Var}[f(X) - g(X) + 0{,}5]$ on sama kuin

$$\begin{aligned}\operatorname{Var}\left[f(X) - g(X)\right] &= \int_0^1 \left(f(x) - g(x) - \mathbb{E}\left[f(X) - g(X)\right]\right)^2 dx\\ &= \frac{7e-19}{12e-12} \approx 0{,}0013566.\end{aligned}$$

Havaitaan siis, että vakioivan muuttujan lisääminen ei vaikuta odotusarvoon, mutta se pienentää varianssia huomattavasti.

Tässä tutkielmassa käsitellään useaa vakioivaa muuttujaa samanaikaisesti. Määritellään tätä varten vakioiva muuttuja vektoriarvoiseksi.

Määritelmä 3.2 (Vakioiva muuttuja). Vakioiva muuttuja h on vektoriarvoinen funktio $h : \mathbb{X} \to \mathbb{R}^m$, $h(x) = (h_1(x), \dots, h_m(x))^T$, jolle $\int_{\mathbb{X}} h(x)\,dx = \theta$, missä $\theta \in \mathbb{R}^m$ on tunnettu arvo.

Yllä olevan määritelmän muotoilu johtuu siitä, että seuraavassa alaluvussa 3.2 vakioivilla muuttujilla varustettua estimaattoria on helpompi käsitellä, kun vakioivalle muuttujalle pätee $\int h(x)\,dx = \theta$. Merkitsemällä $h(x) = g(x)p(x)$ saadaan tässä alaluvussa käytetty vakioiva muuttuja g, jonka odotusarvolle pätee halutusti $\mathbb{E}_p[g(X)] = \int g(x)p(x)\,dx = \theta$. Määritellään täten odotusarvon I (1.1) vakioivien muuttujien estimaattori funktion g avulla.

Määritelmä 3.3 (Vakioivien muuttujien estimaattori). Olkoon $X_1, \dots, X_n$ otos jakaumasta p, vakioiva muuttuja $g(x) = (g_1(x), \dots, g_m(x))^\top$ ja sen odotusarvo $\theta = (\theta_1, \dots, \theta_m)^\top$ sekä kerroin $\beta \in \mathbb{R}^m$. Tällöin odotusarvon I vakioivien muuttujien estimaattori on

$$\tilde{\mu}_\beta := \frac{1}{n} \sum_{i=1}^{n} \left(f(X_i) - \beta^\top g(X_i) \right) + \beta^\top \theta. \tag{3.2}$$

Estimaattorin kerroin β on vektori, jonka optimoinnilla pyritään takaamaan menetelmälle pienin mahdollinen varianssi [10]. Yksiulotteisessa tapauksessa optimointiongelma on varsin yksinkertainen. Kun vakioiva muuttuja g ja vektori β ovat yksiulotteisia eli $m = 1$, suluissa oleva varianssi (3.1) voidaan esittää muodossa

$$\sigma^2_{\tilde{\mu},\beta} = \mathrm{Var}\big[f(X)\big] + \beta^2 \mathrm{Var}\big[g(X)\big] - 2\beta \mathrm{Cov}\big[f(X), g(X)\big].$$

Derivoimalla tämä kertoimen β suhteen ja ratkaisemalla nollakohdan saadaan varianssin minimoijaksi kerroin

$$\beta_{min} = \frac{\mathrm{Cov}\big[f(X), g(X)\big]}{\mathrm{Var}\big[g(X)\big]} = \frac{\mathbb{E}\left[\left(f(X) - \mathbb{E}\left[f(X)\right]\right)\left(g(X) - \theta\right)\right]}{\mathbb{E}\left[\left(g(X) - \theta\right)^2\right]}.$$

Käytännössä vakion β_{min} arvoa ei useinkaan voida laskea. Tällöin käytetään yleensä pienimmän neliösumman estimaattoria, jossa odotusarvot korvataan otoskeskiarvoilla:

$$\hat{\beta}_n = \left(\frac{1}{n} \sum_{i=1}^{n} \left(f(X_i) - \bar{f} \right) \left(g(X_i) - \bar{g} \right) \right) \left(\frac{1}{n} \sum_{i=1}^{n} \left(g(X_i) - \bar{g} \right)^2 \right)^{-1}, \tag{3.3}$$

missä $\bar{f} = \frac{1}{n} \sum_{i=1}^{n} f(X_i)$ ja $\bar{g} = \frac{1}{n} \sum_{i=1}^{n} g(X_i)$.

Kun yksiulotteisen tapauksen estimaattoriin $\tilde{\mu}_\beta$ lisätään samanmuotoisia vakioivia muuttujia $\sum_{i=1}^{n} \beta_j (g_j(X_i) - \theta_j)$, missä $j = 2, \dots, m$, päästään moniulotteiseen tapaukseen. Tällöinkään varianssin minimoivan vektorin $\beta_{min} \in \mathbb{R}^m$ tarkkaa arvoa ei yleensä voida laskea, joten se estimoidaan otoksen avulla. Tarkastelemalla moniulotteisen estimaattorin $\tilde{\mu}_\beta$ varianssia

$$\sigma^2_{\tilde{\mu},\beta} = \int \left(f(x) - \beta^\top g(x) + \beta^\top \theta - I \right)^2 p(x)\, dx$$

Owen ja Zhou [9] havaitsevat, että minimoijan β_{min} estimaattori $\hat{\beta}_n$ voidaan löytää regressioanalyysin avulla. Owen [10] nimittääkin estimaattoria $\tilde{\mu}_\beta$ (3.2) regressioestimaattoriksi (engl. *regression estimator*). Nimitys on osuva, sillä mallin voi mieltää muuttujan $f(X)$ regressioksi, missä $g(X)$ on selittävä muuttuja ja vektori β mallin regressiokerroin.

Kuten esimerkistä 3.1 havaitaan ja seuraavan alaluvun 3.2 apulauseessa 3.5 myös todistetaan, vakioivien muuttujien estimaattori $\tilde{\mu}_\beta$ on harhaton. On kuitenkin huomioitava, että estimoiduilla kertoimilla kuten estimaatilla $\hat{\beta}_n$ (3.3) varustettu estimaattori ei yleensä ole harhaton eli $\mathbb{E}[\tilde{\mu}_{\hat{\beta}_n}] \neq I$. Owenin [10] mukaan estimaatin $\hat{\beta}_n$ käyttämisestä johtuva harha on kuitenkin yleensä pieni ja merkityksetön. Jossain malleissa harhan voi jättää huomiotta, kun n on riittävän suuri, mutta jos estimaattorin on oltava harhaton, niin harha voidaan poistaa eri menetelmin (katso [10, luku 8 s. 32, 35]).

3.2. Painotusotanta vakioivilla muuttujilla ja varianssin minimointi

Sopivasti valitut vakioivat muuttujat voivat tehokkaasti pienentää tarvittavaa otoskokoa, mutta menetelmä ei välttämättä toimi riittävän hyvin esimerkiksi harvinaisten tapahtumien analysoimisessa. Yhdistämällä vakioivat muuttujat ja painotusotanta voidaan estimaattoriin saada molempien menetelmien hyödyt. Tarkastellaankin seuraavaksi vakioivilla muuttujilla varustetun painotusotannan estimaattorin ominaisuuksia sekä sen varianssin minimointia.

Määritelmä 3.4 (Painotusotanta vakioivilla muuttujilla). Olkoon $h : \mathbb{X} \to \mathbb{R}^m$ määritelmän 3.2 vakioiva muuttuja ja $\beta \in \mathbb{R}^m$. Oletetaan todennäköisyysjakaumalle $q(x)$, että kaikille $x \in \mathbb{X}$ pätee $q(x) > 0$ aina, kun $f(x)p(x) \neq 0$ tai $h(x) \neq 0$. Kun $X_1, \ldots, X_n$ on otos ehdotusjakaumasta q, niin odotusarvon I painotusotannan estimaattori $I_{p,q}^{(n)}$ (2.2) vakioivin muuttujin on

$$I_{q,\beta}^{(n)} := \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)p(X_i) - \sum_{j=1}^{m} \beta_j h_j(X_i)}{q(X_i)} + \sum_{j=1}^{m} \beta_j \theta_j. \tag{3.4}$$

Painotusotannan estimaattoriin $I_{p,q}^{(n)}$ (2.2) lisätään siis vakioiva muuttuja h, jonka integraali $\int_{\mathbb{X}} h(x)\,dx = \theta \in \mathbb{R}^m$ on tunnettu. Kuten painotusotannan estimaattori $I_{p,q}^{(n)}$, estimaattori $I_{q,\beta}^{(n)}$ on hyvin määritelty, kun ehdotusjakauma q toteuttaa sen ehdot. Myös harhattomuus pätee estimaattorille.

Apulause 3.5. *Estimaattori $I_{q,\beta}^{(n)}$ (3.4) on harhaton.*

Todistus. Tarkastellaan estimaattorin $I_{q,\beta}^{(n)}$ odotusarvoa.

$$\begin{aligned}
\mathbb{E}\left[I_{q,\beta}^{(n)}\right] &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\frac{f(X_i)p(X_i) - \sum_{j=1}^{m} \beta_j h_j(X_i)}{q(X_i)} + \sum_{j=1}^{m} \beta_j \theta_j\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(\mathbb{E}_q\left[\frac{f(X)p(X)}{q(X)}\right] - \sum_{j=1}^{m} \mathbb{E}_q\left[\frac{\beta_j h_j(X)}{q(X)}\right] + \sum_{j=1}^{m} \mathbb{E}\left[\beta_j \theta_j\right]\right) \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(\mathbb{E}_p\left[f(X)\right] - \sum_{j=1}^{m} \beta_j \int_{\mathbb{X}} h_j(x)\,dx + \sum_{j=1}^{m} \beta_j \theta_j\right) \\
&= \frac{n}{n} I + \frac{1}{n} \sum_{i=1}^{n} \left[-\sum_{j=1}^{m} \beta_j \theta_j + \sum_{j=1}^{m} \beta_j \theta_j\right] = I,
\end{aligned}$$

missä hyödynnetään todennäköisyysjakauman vaihtamisen yhtälöä (2.1) ja vakioivan muuttujan määritelmää $\int_{\mathbb{X}} h_j(x) = \theta_j$ sekä satunnaismuuttujan X_i samoin jakautuneisuutta kaikilla $i = 1, \ldots, n$. □

Estimaattorin $I_{q,\beta}^{(n)}$ määritelmästä 3.4 havaitaan, että jos ehdotusjakaumaksi asetetaan $q(x) = p(x)$ ja vakioivaksi muuttujaksi valitaan $h(x) = g(x)p(x)$, jossa $g : \mathbb{X} \to \mathbb{R}^m$ on vakioiva muuttuja, niin estimaattori $I_{q,\beta}^{(n)}$ (3.4) on sama kuin vakioivien muuttujien estimaattori $\tilde{\mu}_\beta$ (3.2). Täten apulauseessa 3.5 todistettiin myös estimaattorin $\tilde{\mu}_\beta$ harhattomuus.

Estimaattorin soveltaminen käytäntöön ei välttämättä ole helppoa. Kuten vakioivien muuttujien yhteydessä tarkastellusta varianssista (3.1) havaitaan, varianssin pienentyminen perustuu muuttujien $f(X)$ ja $h(X)$ väliseen korrelaation. Ongelmana on siis löytää sopiva vakioiva muuttuja h, joka korreloi halutulla tavalla funktion f kanssa. Lisäksi ongelmana on varianssin minimointi kertoimen β suhteen.

Apulause 3.6. *Estimaattorin* $I_{q,\beta}^{(n)}$ (3.4) *varianssi on* $\operatorname{Var}\left[I_{q,\beta}^{(n)}\right] = \frac{\sigma_{q,\beta}^2}{n}$, *missä*

$$\sigma_{q,\beta}^2 = \int_{\mathbb{X}} \left(\frac{f(x)p(x) - \sum_{j=1}^m \beta_j h_j(x)}{q(x)} + \sum_{j=1}^m \beta_j \theta_j - I \right)^2 q(x)\, dx. \tag{3.5}$$

Todistus. Merkitään $\varphi(x) = \frac{f(x)p(x) - \sum_{j=1}^m \beta_j h_j(x)}{q(x)} + \sum_{j=1}^m \beta_j \theta_j$. Tällöin voidaan kirjoittaa $I_{q,\beta}^{(n)} = \frac{1}{n} \sum_{i=1}^n \varphi(X_i)$ ja $\operatorname{Var}\left[I_{q,\beta}^{(n)}\right] = \frac{\operatorname{Var}[\varphi(X_1)]}{n}$. Täten

$$\begin{aligned}
\sigma_{q,\beta}^2 &= \operatorname{Var}\left[\varphi\left(X_1\right)\right] \\
&= \mathbb{E}\left[\left(\varphi\left(X_1\right) - \mathbb{E}\left[\varphi\left(X_1\right)\right]\right)^2\right] \\
&= \int_{\mathbb{X}} \left(\frac{f(x)p(x) - \sum_{j=1}^m \beta_j h_j(x)}{q(x)} + \sum_{j=1}^m \beta_j \theta_j - I \right)^2 q(x)\, dx,
\end{aligned}$$

missä on hyödynnetty estimaattorin $I_{q,\beta}^{(n)}$ harhattomuutta eli apulausetta 3.5. □

Toinen ongelmakohta on siis varianssin $\sigma_{q,\beta}^2$ (3.5) minimoivan kertoimen β_{min} löytäminen. Samoin kuin vakioivat muuttujat -menetelmässä, kerrointa β_{min} ei käytännössä voida aina laskea, joten se estimoidaan. Estimoinnilla saadun vektorin $\hat{\beta}_n$ käyttäminen puolestaan johtaa siihen, ettei estimaattori $I_{q,\hat{\beta}_n}^{(n)}$ ole välttämättä enää harhaton.

Owenin ja Zhoun tulos [9, Theorem 1] osoittaa, että tietyillä ehdoilla pienimmän neliösumman avulla määritelty esimaattori $\hat{\beta}_n$ suppenee varianssin minimoivaan tarkkaan arvoon β_{min} stokastisella kertaluokalla $O_p(n^{-1/2})$. He myös osoittavat, että estimoidulla arvolla varustettu estimaattori $I_{q,\hat{\beta}_n}^{(n)}$ suppenee estimaattoriin $I_{q,\beta_{min}}^{(n)}$ kertaluokalla $O_p(n^{-1})$. Toisin sanoen estimaattori $I_{q,\hat{\beta}_n}^{(n)}$ on likipitäen sama kuin estimaattori $I_{q,\beta_{min}}^{(n)}$. Tämän tarkentuvuuden ansiosta estimoidun kertoimen $\hat{\beta}_n$ käyttö ei aiheuta ongelmaa menetelmän kannalta, vaikka sen estimaattorilla olisikin tällöin harha. Owenin ja Zhoun osoittamat väitteet ehtoineen on muotoiltu tarkasti seuraavan alaluvun 3.3 oletuksessa 3.12 sekä lauseissa 3.13 ja 3.14. Näitä ennen perehdytään kuitenkin tarkemmin estimaattorin $I_{q,\beta}^{(n)}$ (3.4) varianssin $\sigma_{q,\beta}^2$ (3.5) minimoivan kertoimen β_{min} yksikäsitteisyyteen.

Kuten vakioivien muuttujien estimaattorin tapauksessa, varianssin $\sigma_{q,\beta}^2$ minimointi kertoimen β suhteen on optimointiongelma. Jotta optimointiongelma on matemaattisesti hyvin määritelty, on tutkittava, millä ehdoilla minimi on olemassa ja minimoija on yksikäsitteinen. Tästä syystä kertoimeen β_{min} perehdytään ensin teoreettiselta näkökulmalta ja vasta sitten alaluvussa 3.3 syvennytään minimoijan etsimisen empiiriseen puoleen. Varsinainen todistus varianssin globaalin minimoijan β_{min} olemassaolosta on lauseessa 3.10, mutta ennen sitä osoitetaan kaksi apulausetta. Määritellään näitä varten riittävät ehdot yksikäsitteisyyden takaamiseen.

Oletus 3.7. *Oletetaan seuraavat väittämät:*

(i) Olkoon mitallinen funktio $f : \mathbb{X} \to \mathbb{R}$, p tiheysfunktio joukossa $\mathbb{X}$, h määritelmän 3.4 vakioiva muuttuja, jonka integraali $\theta \in \mathbb{R}^m$, ja q ehdotusjakauma, jolle pätee $q(x) > 0$, kun $f(x)p(x) \neq 0$ tai $h(x) \neq 0$ kaikilla $x \in \mathbb{X}$.

(ii) Oletetaan, että kaikille $i, j = 1, \dots, m$ integraalit

$$\kappa_i = \int_{\mathbb{X}} \frac{f(x)p(x)h_i(x)}{q(x)} dx \quad ja \quad c_{ij} = \int_{\mathbb{X}} \frac{h_i(x)h_j(x)}{q(x)} dx \tag{3.6}$$

ovat hyvin määriteltyjä ja äärellisiä. Määritellään näiden integraalien avulla vektori κ, jossa $[\kappa]_i = \kappa_i$, ja $m \times m$-matriisi C, jossa $[C]_{ij} = c_{ij}$.

Oletetaan myös, että satunnaisvektorin $\boldsymbol{Z} = (Z^{(1)}, \dots, Z^{(m)})^\top$, missä $Z^{(i)} = \frac{h_i(X)}{q(X)}$, $i = 1, \dots, m$ ja $X \sim q$, kovarianssimatriisille $\mathrm{Cov}[\boldsymbol{Z}, \boldsymbol{Z}] = \mathrm{Cov}[\boldsymbol{Z}]$ pätee $\mathrm{Cov}[\boldsymbol{Z}] > 0$, eli $\mathrm{Cov}[\boldsymbol{Z}]$ on positiividefiniitti.

Oletuksen avulla halutaan siis varmistaa, että jatkossa todistuksissa käytettävät vektorin κ ja matriisin C alkiot (3.6) ovat hyvin määriteltyjä sekä satunnaismuuttujista $h_i(X)/q(X)$, $i = 1, \dots, m$, muodostuvan satunnaisvektorin $\boldsymbol{Z}$ kovarianssimatriisi $\mathrm{Cov}[\boldsymbol{Z}]$ on positiivisesti definiitti. Kovarianssimatriisi on tunnetusti aina positiivisesti semidefiniitti [14, s. 28–29], mutta todistukseemme tarvitsee tiukemman positiividefiniittisyyden. Tämän oletuksen voimassaoloa tutkitaan tarkemmin alaluvussa 3.4.

Aloitetaan minimin olemassaolon ja yksikäsitteisyyden osoittaminen selvittämällä varianssin $\sigma_{q,\beta}^2$ gradientti vektorin β suhteen ja osoittamalla sen Hessen matriisin positiivisesti definiittiseksi.

Apulause 3.8. *Mikäli oletus 3.7 on voimassa, varianssille $\sigma_{q,\beta}^2$ (3.5) pätee*

(1) $\nabla_\beta \sigma_{q,\beta}^2 = -2(\kappa - I\theta) - 2(\beta^\top\theta)\theta + 2\beta^\top C$,

(2) sen Hessen matriisi muuttujan β suhteen on $H = 2(C - \theta\theta^\top)$,

(3) ja matriisi H on symmetrinen ja positiivisesti definiitti.

TODISTUS. Osoitetaan ensin väite (1) eli selvitetään minimoitavan varianssin gradientti $\nabla_\beta \sigma^2_{q,\beta}$. Kiinnitetään vektorin β alkiot β_j, kun $j \neq i$, ja tutkitaan, miten varianssi $\sigma^2_{q,\beta}$ käyttäytyy, kun alkio β_i muuttuu. Kirjoitetaan

$$\sigma^2_{q,\beta} = \int_{\mathbb{X}} \big(\varphi(\beta_i, x)\big)^2 q(x)\, dx,$$

missä

$$\varphi(\beta_i, x) = \frac{f(x)p(x) - \sum_{j=1}^m \beta_j h_j(x)}{q(x)} + \sum_{j=1}^m \beta_j \theta_j - I.$$

Koska $\frac{d}{d\beta_i}\varphi(\beta_i, x) = \theta_i - \frac{h_i(x)}{q(x)}$, saadaan jatkuva osittaisderivaatta

$$\frac{\partial}{\partial \beta_i}\varphi^2(\beta_i, x) = 2\varphi(\beta_i, x)\left(\theta_i - \frac{h_i(x)}{q(x)}\right) =: \xi_i(\beta_i, x).$$

Mikäli integraalin $\int_{\mathbb{X}} \varphi^2(\beta_i, x)q(x)$ osittaisderivaatta on olemassa, niin

$$\begin{aligned}\frac{\partial}{\partial \beta_i}\int_{\mathbb{X}} \varphi^2(\beta_i, x)q(x)\, dx &= \lim_{\epsilon\to 0}\int_{\mathbb{X}} \frac{\varphi^2(\beta_i+\epsilon, x) - \varphi^2(\beta_i, x)}{\epsilon} q(x)\, dx \\ &= \lim_{\epsilon\to 0}\int_{\mathbb{X}} \xi_i(\beta_{i_\epsilon}, x)q(x)\, dx,\end{aligned}$$

missä $\beta_{i_\epsilon} \in (\beta_i, \beta_i + \epsilon)$ on olemassa väliarvolauseen nojalla kaikilla $\epsilon \in \mathbb{R}\backslash\{0\}$.

Kun $\beta_{i_\epsilon} \in (a, b) \subset \mathbb{R}$, niin kaikilla ϵ pätee

$$|\xi_i(\beta_{i_\epsilon}, x)| \leq \psi_i(\beta_{i_\epsilon}, x),$$

missä

$$\begin{aligned}\psi_i(\beta_{i_\epsilon}, x) = 2&\left(\frac{|f(x)p(x)|}{q(x)} + \sum_{j=1, j\neq i}^m \frac{|\beta_j h_j(x)|}{q(x)} + \sum_{j=1, j\neq i}^m |\beta_j\theta_j| + |I|\right)\left(|\theta_i| + \frac{|h_i(x)|}{q(x)}\right) \\ &+ 2\max\{|a|, |b|\}\left(\frac{|h_i(x)|}{q(x)} + |\theta_i| + |I|\right)^2.\end{aligned}$$

Oletuksen 3.7 (ii) nojalla $\int |f(x)p(x)h_i(x)|/q(x) < \infty$ ja $\int |h_j(x)h_i(x)|/q(x) < \infty$ kaikilla $i, j = 1, \ldots, m$, joten pätee $\int |\psi_i(\beta_{i_\epsilon}, x)q(x)|\, dx < \infty$ eli funktio $\psi_i(\beta_{i_\epsilon}, x)q(x)$ on integroituva. Lisäksi osittaisderivaatan jatkuvuudesta seuraa, että

$$\lim_{\epsilon\to 0}\xi_i(\beta_{i_\epsilon}, x)q(x) = \xi_i(\beta_i, x)q(x).$$

Täten dominoivan konvergenssin lauseen nojalla

$$\begin{aligned}\frac{d}{d\beta_i}\sigma^2_{q,\beta} &= \lim_{\epsilon\to 0}\int_{\mathbb{X}} \xi_i(\beta_{i_\epsilon}, x)q(x)\, dx = \int_{\mathbb{X}} \xi_i(\beta_i, x)q(x)\, dx \\ &= \int_{\mathbb{X}} 2\varphi(\beta_i, x)\left(\theta_i - \frac{h_i(x)}{q(x)}\right) q(x)\, dx.\end{aligned}$$

Estimaattorin harhattomuudesta seuraa, että $\int_{\mathbb{X}} \varphi(\beta_i, x)\theta_i q(x)\, dx = 0$, joten yhtälö sievenee muotoon

$$\begin{aligned}
\frac{d}{d\beta_i}\sigma^2_{q,\beta} &= -2\int_{\mathbb{X}} \varphi(\beta_i, x) h_i(x)\, dx \\
&= -2\int_{\mathbb{X}} \left(\frac{f(x)p(x)}{q(x)} - I\right) h_i(x)\, dx - 2\int_{\mathbb{X}} \sum_{j=1}^{m} \beta_j \left(\theta_j - \frac{h_j(x)}{q(x)}\right) h_i(x)\, dx. \\
&= -2\int_{\mathbb{X}} \frac{f(x)p(x)h_i(x)}{q(x)}\, dx + 2I\theta_i - 2\sum_{j=1}^{m} \beta_j \left(\theta_j\theta_i - \int_{\mathbb{X}} \frac{h_j(x)h_i(x)}{q(x)}\, dx\right) \\
&= -2\left(\kappa_i - I\theta_i\right) - 2\sum_{j=1}^{m} \beta_j \left(\theta_j\theta_i - c_{ij}\right) \\
&= -2\left(\kappa_i - I\theta_i\right) - 2\beta^\top \left(\theta\theta_i - c_i\right),
\end{aligned}$$

missä $c_i = (c_{i1}, \ldots, c_{im})^\top$ on matriisin C i. rivivektori. Huomioiden, että $\sum_{j=1}^{m} \beta_j c_{ij} = \beta^\top c_i = c_i^\top \beta$, saadaan

$$\nabla_\beta \sigma^2_{q,\beta} = -2(\kappa - I\theta) - 2(\beta^\top\theta)\theta + 2C\beta,$$

mikä vastaa väitteen (1) tulosta.

Osoitetaan seuraavaksi väite (2). Gradientin $\nabla_\beta \sigma^2_{q,\beta}$ derivaatta muuttujan β_j suhteen on

$$\frac{d}{d\beta_j}\nabla_\beta \sigma^2_{q,\beta} = -2(\theta_j)\theta + 2c_j,$$

missä on hyödynnetty määritelmästä (3.6) saatua tietoa $c_{ij} = c_{ji}$ kaikilla $i, j = 1, \ldots, m$. Täten Hessen matriisi on haluttua muotoa

$$H = -2\theta\theta^\top + 2C = 2\left(C - \theta\theta^\top\right).$$

Todistetaan lopuksi väite (3). Koska matriisi C on symmetrinen, niin matriisi H on myös selvästi symmetrinen. Matriisin H positiivisen definiittisyyden todistamista varten merkitään $\boldsymbol{Z} = (Z^{(1)}, \ldots, Z^{(m)})^\top$, missä $Z^{(i)} = \frac{h_i(X)}{q(X)}$. Tällöin $c_{ij} = \mathbb{E}[Z^{(i)}Z^{(j)}]$ ja $C = \mathbb{E}\left[\boldsymbol{Z}\boldsymbol{Z}^\top\right]$. Lisäksi $\theta_i = \mathbb{E}\left[Z^{(i)}\right]$ ja $\mathbb{E}[\boldsymbol{Z}] = \theta$. Täten

$$H = 2\left(\mathbb{E}\left[\boldsymbol{Z}\boldsymbol{Z}^\top\right] - \mathbb{E}[\boldsymbol{Z}]\mathbb{E}[\boldsymbol{Z}]^\top\right) = 2\mathrm{Cov}[\boldsymbol{Z}].$$

Koska oletuksen 3.7 (ii) nojalla $\mathrm{Cov}[\boldsymbol{Z}] > 0$, niin pätee $H > 0$ eli matriisi H on positiivisesti definiitti. □

Todistetun apulauseen 3.8 kohdasta (2) havaitaan, että varianssin $\sigma^2_{q,\beta}$ Hessen matriisi H on vakio kaikilla $\beta \in \mathbb{R}^m$. Matriisi H on siis positiivisesti definiitti kaikilla β. Koska varianssi $\sigma^2_{q,\beta}$ on toisen asteen lauseke kertoimen β suhteen, niin matriisin H positiivinen definiittisyys takaa kriittisen pisteen olevan globaali aito minimi. Toisin sanoen kerroin β_{min}, jolle gradientti $\nabla_\beta \sigma^2_{q,\beta_{min}} = 0$, on varianssin $\sigma^2_{q,\beta}$ yksikäsitteinen minimoija.

Määritellään vektorin β_{min} löytämistä varten ositetut matriisi A^* ja vektori γ^*.

$$A^* = \begin{bmatrix} 1 & \theta^\top \\ \theta & C \end{bmatrix} \quad \text{ja} \quad \gamma^* = \begin{bmatrix} I \\ \kappa \end{bmatrix}, \tag{3.7}$$

missä vektori $\theta = \mathbb{E}[\boldsymbol{Z}]$ ja matriisin C sekä vektorin κ alkiot ovat kuten oletuksen 3.7 yhtälöissä (3.6). Koska oletuksen 3.7 nojalla matriisi C ja vektori κ ovat hyvin määriteltyjä, niin matriisi A^* ja vektori γ^* ovat myös hyvin määriteltyjä oletuksen pätiessä. Osoitetaan, että matriisi A^* on tällöin myös kääntyvä.

Apulause 3.9. *Oletuksen 3.7 voimassaollessa matriisi A^* yhtälössä* (3.7) *on kääntyvä.*

Todistus. Osoitetaan symmetrinen matriisin A^* positiivisesti definiitiksi, mikä takaa sen kääntyvyyden. Merkitään $\widetilde{\boldsymbol{Z}} = \begin{bmatrix} 1 \\ \boldsymbol{Z} \end{bmatrix}$, jolloin $\mathbb{E}\left[\widetilde{\boldsymbol{Z}}\right] = \begin{bmatrix} 1 \\ \theta \end{bmatrix}$ ja $A^* = \mathbb{E}\left[\widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{Z}}^\top\right]$. Olkoon $a = \begin{bmatrix} c \\ b \end{bmatrix} \in \mathbb{R}^{1+m}$ siten, että $c \in \mathbb{R}$, $b \in \mathbb{R}^m$ ja $a \neq 0$. Tällöin

$$\begin{aligned} a^\top A^* a &= a^\top \mathbb{E}\left[\widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{Z}}^\top\right] a = \mathbb{E}\left[\left(a^\top \widetilde{\boldsymbol{Z}}\right)\left(\widetilde{\boldsymbol{Z}}^\top a\right)\right] = \mathbb{E}\left[\left(a^\top \widetilde{\boldsymbol{Z}}\right)^2\right] \\ &= \mathbb{E}\left[\left(c + b^\top \boldsymbol{Z}\right)^2\right] = \operatorname{Var}\left[b^\top \boldsymbol{Z}\right] + \left(c + b^\top \theta\right)^2, \end{aligned}$$

missä hyödynnetään satunnaisluvulle X ja vakiolle c pätevää kaava $\mathbb{E}[(X+c)^2] = \operatorname{Var}[X] + (c + \mathbb{E}[X])^2$. Kun $b \neq 0$,

$$\operatorname{Var}\left[b^\top \boldsymbol{Z}\right] = \operatorname{Cov}\left[b^\top \boldsymbol{Z}, b^\top \boldsymbol{Z}\right] = b^\top \operatorname{Cov}\left[\boldsymbol{Z}, \boldsymbol{Z}\right] b > 0,$$

sillä oletuksen 3.7 (ii) nojalla $\operatorname{Cov}[\boldsymbol{Z}] > 0$. Tällöin $a^\top A^* a \geq \operatorname{Var}\left[b^\top \boldsymbol{Z}\right] > 0$. Vastaavasti, kun $b = 0$, niin, koska $a \neq 0$, on oltava $c \neq 0$, jolloin

$$a^\top A^* a = c^2 > 0.$$

Täten $a^\top A^* a > 0$ kaikilla $a \in \mathbb{R}^{1+m}$, $a \neq 0$, joten matriisi A^* on positiivisesti definiitti ja näin myös kääntyvä. □

Oletus 3.7 takaa siis sekä varianssin $\sigma^2_{q,\beta}$ Hessen matriisin H positiivisen definiittisyyden että matriisin A^* kääntyvyyden. Todistetaan seuraavaksi, että oletuksen 3.7 pätiessä matriisin A^* ja vektorin γ^* avulla muodostetun vektorin $\tilde{\beta}_{min} = (A^*)^{-1}\gamma^*$ osavektori β_{min} on gradientin $\nabla_\beta \sigma^2_{q,\beta}$ nollakohta ja täten varianssin $\sigma^2_{q,\beta}$ yksikäsitteinen minimoija.

Lause 3.10. *Päteköön oletus 3.7. Olkoon A^* ja γ^* kuten yhtälössä* (3.7) *sekä $\tilde{\beta}_{min} = (A^*)^{-1}\gamma^*$. Tällöin merkitsemällä $\tilde{\beta}_{min} = \begin{bmatrix} \beta_0 \\ \beta_{min} \end{bmatrix}$, missä $\beta_0 \in \mathbb{R}$ ja $\beta_{min} \in \mathbb{R}^m$, vektori β_{min} on varianssin $\sigma^2_{q,\beta}$* (3.5) *yksikäsitteinen minimoija muuttujan β suhteen.*

Todistus. Apulauseen 3.9 nojalla A^* on kääntyvä, joten $\tilde{\beta}_{min} = (A^*)^{-1}\gamma^*$ on hyvin määritelty. Koska $A^*\tilde{\beta}_{min} = A^*(A^*)^{-1}\gamma^* = \gamma^*$, voidaan kirjoittaa

$$A^*\tilde{\beta}_{min} = \begin{bmatrix} 1 & \theta^\top \\ \theta & C \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_{min} \end{bmatrix} = \begin{bmatrix} \beta_0 + \theta^\top \beta_{min} \\ \beta_0 \theta + C\beta_{min} \end{bmatrix} = \begin{bmatrix} I \\ \kappa \end{bmatrix} = \gamma^*.$$

Saadaan täten yhtälöt $\beta_0 + \theta^\top \beta_{min} = I$ ja $\beta_0\theta + C\beta_{min} = \kappa$. Ratkaistaan ensimmäinen yhtälö $\beta_0 = I - \theta^\top \beta_{min}$ ja sijoitetaan se jälkimmäiseen yhtälöön.

$$\begin{aligned} &(I - \theta^\top \beta_{min})\theta + C\beta_{min} = \kappa \\ \Leftrightarrow \quad &-(\theta^\top \beta_{min})\theta + C\beta_{min} = \kappa - I\theta. \end{aligned}$$

Sijoitetaan $\kappa - I\theta = (-\theta^\top \beta_{min})\theta + C\beta_{min}$ puolestaan apulauseen 3.8 kohdassa (1) laskettuun varianssin yhtälöön.

$$\begin{aligned} \nabla_\beta \sigma^2_{q,\beta_{min}} &= -2(\kappa - I\theta) - 2(\beta_{min}^\top \theta)\theta + 2C\beta_{min} \\ &= -2\left(-(\beta_{min}^\top \theta)\theta + C\beta_{min}\right) - 2(\beta_{min}^\top \theta)\theta + 2C\beta_{min} \\ &= 2(\beta_{min}^\top \theta)\theta - 2C\beta_{min} - 2(\beta_{min}^\top \theta)\theta + 2C\beta_{min} \\ &= 0. \end{aligned}$$

Koska oletus 3.7 on voimassa, apulauseen 3.8 kohdan (3) mukaan varianssin $\sigma^2_{q,\beta}$ Hessen matriisi H on positiivisesti definiitti. Täten, koska pätee $\nabla_\beta \sigma^2_{q,\beta_{min}} = 0$, on vektori β_{min} varianssin yksikäsitteinen minimoija. □

Huomautus 3.11. Lauseessa 3.10 esiintyvän vektorin $\tilde{\beta}_{min} = \begin{bmatrix} \beta_0 \\ \beta_{min} \end{bmatrix}$ alkio $\beta_0 = I - \theta^\top \beta_{min}$ on vakiotermi, joka sisältää alkuperäisen ongelman ratkaisun eli odotusarvon $I = \mathbb{E}_p[f(X)]$.

3.3. Varianssin empiirinen minimointi

Estimaattorin $I^{(n)}_{q,\beta}$ (3.4) varianssin $\sigma^2_{q,\beta}$ (3.5) optimointiongelma on lauseen 3.10 nojalla hyvin asetettu: yksikäsitteinen minimoija β_{min} on olemassa, kun oletus 3.7 on voimassa. Minimoijaa β_{min} ei kuitenkaan yleensä voida laskea teoreettisesti, vaan se estimoidaan. Millainen on toimiva estimaattori $\hat{\beta}_n$ kertoimelle β_{min}?

Owen ja Zhou [9] toteavat, että varianssi $\sigma^2_{q,\beta}$ (3.5) viittaa muuttujan $\frac{f(X)p(X)}{q(X)}$ regressioon, jossa $\frac{h_j(X)}{q(X)}$ on selittävä muuttuja. He ehdottavatkin käyttämään kertoimen β estimaattorina pienimmän neliösumman ratkaisua. Tälle estimaattorille he todistavat lauseen [9, Theorem 1], joka on tässä tutkielmassa jaettu kahteen osaan ja muotoiltu tarkasti lauseisiin 3.13 ja 3.14.

Määritellään lauseiden muotoilua ja todistusta varten tarvittavat apumuuttujat. Olkoon funktio f, vakioiva muuttuja h, todennäköisyysfunktiot p ja q kuten oletuksessa 3.7 (i) sekä $X_i \sim q$ riippumattomia, kun $i = 1, \ldots, n$. Merkitään

$$Y_i = \frac{f(X_i)p(X_i)}{q(X_i)}, \qquad Z_i^{(j)} = \frac{h_j(X_i)}{q(X_i)} \qquad ja \qquad \boldsymbol{Z_i} = \begin{bmatrix} Z_i^{(1)} \\ \vdots \\ Z_i^{(m)} \end{bmatrix}, \tag{3.8}$$

missä $j = 1, \ldots, m$. Näiden avulla estimaattori $I^{(n)}_{q,\beta}$ (3.4) voidaan esittää muodossa

$$I^{(n)}_{q,\beta} = \frac{1}{n}\sum_{i=1}^{n} Y_i - \beta^\top \boldsymbol{Z_i} + \beta^\top \theta. \tag{3.9}$$

Apumuuttujien avulla voidaan myös muotoilla Owenin ja Zhoun väitteen [9, Theorem 1] todistamiseen tarvittava oletus täsmällisesti.

OLETUS 3.12. *Olkoon satunnaismuuttujat* $X_1, \dots X_n \sim q$ *riippumattomia sekä samoin jakautuneita ja olkoon* Y_i, $Z_i^{(j)}$ *ja* $\boldsymbol{Z_i}$ *määritelty kuten* (3.8). *Oletetaan, että* $\mathrm{Cov}[\boldsymbol{Z_1}] > 0$, *sekä*

$$\mathbb{E}\left[(Y_1)^2\right] < \infty, \qquad \mathbb{E}\left[\left(Z_1^{(j)}\right)^4\right] < \infty \qquad ja \qquad \mathbb{E}\left[\left(Y_1 Z_1^{(j)}\right)^2\right] < \infty,$$

kun $j = 1, \dots, m$.

Estimaattorin $\hat{\beta}_n$ täsmällistä määrittelyä varten määritellään lisäksi apumuuttujat $\boldsymbol{Y}_n$ ja $\boldsymbol{M}_n$:

$$\boldsymbol{Y_n} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \qquad ja \qquad \boldsymbol{M}_n = \begin{bmatrix} 1 & \boldsymbol{Z}_1^\top \\ \vdots & \vdots \\ 1 & \boldsymbol{Z}_n^\top \end{bmatrix} = \begin{bmatrix} 1 & Z_1^{(1)} & \cdots & Z_1^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & Z_n^{(1)} & \cdots & Z_n^{(m)} \end{bmatrix}.$$

Owenin ja Zhoun [9] ehdottama estimaattori $\hat{\beta}_n \in \mathbb{R}^m$ ja vakio $\hat{\beta}_0 \in \mathbb{R}$ voidaan näin esittää muodossa

$$\tilde{\beta}_n = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_n \end{bmatrix} = \left(\frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{M}_n\right)^+ \frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{Y_n}. \tag{3.10}$$

Koska matriisin $\boldsymbol{M}_n^\top \boldsymbol{M}_n$ kääntyvyydestä ei voida olla varmoja pienillä otoskoilla n, käytetään estimaattorissa määritelmän 1.13 Moore-Penrose käänteismatriisia. Kuten lauseessa 1.14 todetaan, käänteismatriisi $\left(\boldsymbol{M}_n^\top \boldsymbol{M}_n\right)^+$ on yksikäsitteinen matriisille $\boldsymbol{M}_n^\top \boldsymbol{M}_n$, joten estimaattori $\tilde{\beta}_n$ on hyvin määritelty. Tutkimalla tarkemmin matriisituloja

$$\frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{Y_n} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n Y_i \\ \frac{1}{n}\sum_{i=1}^n Z_i^{(1)} Y_i \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n Z_i^{(m)} Y_i \end{bmatrix} \tag{3.11}$$

ja

$$\frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{M}_n = \begin{bmatrix} 1 & \frac{1}{n}\sum_{i=1}^n Z_i^{(1)} & \cdots & \frac{1}{n}\sum_{i=1}^n Z_i^{(m)} \\ \frac{1}{n}\sum_{i=1}^n Z_i^{(1)} & \frac{1}{n}\sum_{i=1}^n \left(Z_i^{(1)}\right)^2 & \cdots & \frac{1}{n}\sum_{i=1}^n Z_i^{(1)} Z_i^{(m)} \\ \frac{1}{n}\sum_{i=1}^n Z_i^{(2)} & \frac{1}{n}\sum_{i=1}^n Z_i^{(2)} Z_i^{(1)} & \cdots & \frac{1}{n}\sum_{i=1}^n Z_i^{(2)} Z_i^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n}\sum_{i=1}^n Z_i^{(m)} & \frac{1}{n}\sum_{i=1}^n Z_i^{(m)} Z_i^{(1)} & \cdots & \frac{1}{n}\sum_{i=1}^n \left(Z_i^{(m)}\right)^2 \end{bmatrix} \tag{3.12}$$

havaitaan, että niiden alkiot ovat riippumattomien satunnaismuuttujien otoskeskiarvoja.

Oletuksen 3.12 ja estimaattorin $\tilde{\beta}_n$ (3.10) avulla Owenin ja Zhoun väitteen [9, Theorem 1] ensimmäinen osa voidaan muotoilla ja todistaa tarkasti.

LAUSE 3.13. *Olkoon vektorit* $\tilde{\beta}_n \in \mathbb{R}^{1+m}$ *ja* $\hat{\beta}_n \in \mathbb{R}^m$ *kuten yhtälössä* (3.10) *ja pätekööt oletukset 3.7 (i) ja 3.12. Tällöin on olemassa yksikäsitteinen vektori* $\beta_{min} \in \mathbb{R}^m$, *joka minimoi varianssin* $\sigma^2_{q,\beta}$ *yhtälössä* (3.5). *Lisäksi, kun* $n \to \infty$, *niin*

$$\hat{\beta}_n = \beta_{min} + O_q\left(n^{-1/2}\right).$$

TODISTUS. Oletuksen 3.7 (i) nojalla muuttujat $Y_i = \frac{f(X_i)p(X_i)}{q(X_i)}$ ja $Z_i^{(j)} = \frac{h_j(X_i)}{q(X_i)}$ on hyvin määritellyt. Koska satunnaismuuttujat X_i ovat oletuksen mukaan riippumattomia ja samoin jakautuneita kaikilla $i = 1, \ldots, n$, muuttujat Y_i ja $Z_i^{(j)}$ ovat myös riippumattomia ja samoin jakautuneita. Olettaen, että $\mathbb{E}_p[|f(X)|] < \infty$, saadaan

$$\begin{aligned}\mathbb{E}\left[|Y_1|\right] &= \int \left|\frac{f(x)p(x)}{q(x)}\right| q(x)\,dx = \int \frac{|f(x)|p(x)}{q(x)} q(x)\,dx \\ &= \int |f(x)|p(x)\,dx = \mathbb{E}_p\left[|f(X)|\right] < \infty.\end{aligned}$$

Käyttämällä Cauchy-Schwarzin epäyhtälöä ja oletusta $\mathbb{E}\big[\big(Y_1 Z_1^{(j)}\big)^2\big] < \infty$ saadaan, että

$$\mathbb{E}\left[\left|Y_1 Z_1^{(j)}\right|\right] \le \left(\mathbb{E}\left[\left|Y_1 Z_1^{(j)}\right|^2\right]\right)^{\frac{1}{2}} = \left(\mathbb{E}\left[\left(Y_1 Z_1^{(j)}\right)^2\right]\right)^{\frac{1}{2}} < \infty.$$

Vastaavasti hyödyntäen Cauchy-Schwarzin epäyhtälöä ja oletusta $\mathbb{E}\big[\big(Z_1^{(j)}\big)^4\big] < \infty$ saadaan

$$\mathbb{E}\left[\left|\left(Z_1^{(j)}\right)^2\right|\right] \le \left(\mathbb{E}\left[\left(Z_1^{(j)}\right)^4\right]\right)^{\frac{1}{2}} < \infty.$$

Tämän nojalla voidaan puolestaan todeta, että

$$\mathbb{E}\left[\left|Z_i^{(j)}\right|\right] \le \left(\mathbb{E}\left[\left(Z_i^{(j)}\right)^2\right]\right)^{\frac{1}{2}} < \infty$$

ja

$$\mathbb{E}\left[\left|Z_1^{(j)} Z_1^{(k)}\right|\right] \le \left(\mathbb{E}\left[\left(Z_1^{(j)}\right)^2\right]\right)^{\frac{1}{2}} \left(\mathbb{E}\left[\left(Z_1^{(k)}\right)^2\right]\right)^{\frac{1}{2}} < \infty.$$

Koska on osoitettu, että odotusarvot $\mathbb{E}\big[\big|Y_1 Z_1^{(j)}\big|\big]$ ja $\mathbb{E}\big[\big|Z_1^{(j)} Z_1^{(k)}\big|\big]$ ovat äärelliset kaikilla $j, k = 1, \ldots, m$, ja $\mathrm{Cov}[\boldsymbol{Z}_1] > 0$, on oletuksen 3.7 kohta (ii) voimassa. Täten vektori γ^* ja matriisi A^* määritelmässä (3.7) ovat hyvin määriteltyjä. Koko oletuksen 3.7 voimassaolosta seuraa myös, että lause 3.10 pätee eli varianssilla $\sigma^2_{q,\beta}$ (3.5) on olemassa yksikäsitteinen minimoija β_{min}, joka on muotoa

$$\tilde{\beta}_{min} = \begin{bmatrix} \beta_0 \\ \beta_{min} \end{bmatrix} = (A^*)^{-1}\gamma^*.$$

Todistetaan seuraavaksi, että $\tilde{\beta}_n = \tilde{\beta}_{min} + O_q\left(n^{-1/2}\right)$. Koska vektorin $\frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{Y_n}$ (3.11) ja matriisin $\frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{M}_n$ (3.12) alkiot ovat keskiarvoja riippumattomista ja samoin jakautuneista satunnaismuuttujista, joiden itseisarvon odotusarvot ovat äärellisiä, niin vahvan suurten lukujen lain 1.2 nojalla pätee, että

$$\frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{Y_n} \to \gamma^* \quad \text{ja} \quad \frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{M}_n \to A^* \quad \text{melkein varmasti,}$$

kun $n \to \infty$. Lisäksi, koska muuttujien varianssit ovat oletuksen 3.12 nojalla äärelliset, vektorin $\frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{Y_n} - \gamma^*$ ja matriisin $\frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{M}_n - A^*$ alkioille pätee apulauseen 1.18 nojalla, että

$$\left[\frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{Y_n} - \gamma^*\right]_j = O_p\left(n^{-1/2}\right) \quad \text{ja} \quad \left[\frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{M}_n - A^*\right]_{kj} = O_p\left(n^{-1/2}\right)$$

kaikilla $k, j = 1, \ldots, m+1$. Niiden alkioittainen suppeneminen tapahtuu siis stokastisella kertaluokalla $O_p(n^{-1/2})$.

Apulauseen 1.21 nojalla, koska sama kertaluokka pätee kaikille vektorin ja matriisin alkioille, niin se pätee myös niistä muodostetulle vektorille ja matriisille. Saadaan siis, että $\frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{Y_n} - \gamma^* = O_p\left(n^{-1/2}\right)$ ja $\frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{M}_n - A^* = O_p\left(n^{-1/2}\right)$. Näin voidaan käyttää apulausetta 1.23 ja saadaan

$$\tilde{\beta}_n - \tilde{\beta}_{min} = \left(\frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{M}_n\right)^+ \frac{1}{n}\boldsymbol{M}_n^\top \boldsymbol{Y_n} - (A^*)^{-1}\gamma^* = O_p\left(n^{-1/2}\right).$$

Vastaavasti, koska vektori $\tilde{\beta}_n - \tilde{\beta}_{min} = \begin{bmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_n - \beta_{min} \end{bmatrix}$ suppenee kertaluokalla $O_p(n^{-1/2})$, niin apulauseen 1.21 nojalla sama stokastinen kertaluokka pätee myös sen osavektorille $\hat{\beta}_n - \beta_{min}$. On siis osoitettu, että

$$\begin{aligned} & \hat{\beta}_n - \beta_{min} = O_p\left(n^{-1/2}\right) \\ \Leftrightarrow \quad & \hat{\beta}_n = \beta_{min} + O_p\left(n^{-1/2}\right). \end{aligned}$$

□

Todistuksessa käy ilmi, että oletus 3.12 implikoi oletuksen 3.7 kohdan (ii). Näin saadaan käyttöön alaluvussa 3.2 todistetut lauseet 3.8 - 3.10. Todistus osoittaa myös, että alkio $\hat{\beta}_0$ suppenee alkioon β_0, joten huomautuksen 3.11 nojalla alkio $\hat{\beta}_0$ on estimoitu vakiotermi, joka sisältää odotusarvon I ratkaisun. Kuten Owen ja Zhou [9] toteavat, estimoidun vakiotermin ja kertoimen avulla estimaattori $I_{q,\beta}^{(n)}$ (3.4) voidaan yksinkertaistaa muotoon $I_{q,\hat{\beta}}^{(n)} = \hat{\beta}_0 + \theta^\top \hat{\beta}_{min}$.

Lauseen 3.13 avulla voidaan todistaa Owenin ja Zhoun toinen väite, jonka mukaan estimoidulla arvolla $\hat{\beta}_n$ varustettu estimaattori $I_{q,\hat{\beta}_n}^{(n)}$ on suurilla otoskoilla n hyvin lähellä ideaalista estimaattoria $I_{q,\beta_{min}}^{(n)}$.

LAUSE 3.14. *Olkoon $\hat{\beta}_n$ kuten se esiintyy yhtälössä* (3.10) *ja β_{min} varianssin $\sigma_{q,\beta}^2$* (3.5) *yksikäsitteinen minimoija. Päteköön myös oletukset 3.7 (i) ja 3.12. Tällöin, kun $n \to \infty$, estimaattorille $I_{q,\beta_n}^{(n)}$* (3.4) *pätee*

$$I_{q,\hat{\beta}_n}^{(n)} = I_{q,\beta_{min}}^{(n)} + O_p(n^{-1}).$$

TODISTUS. Oletusten 3.7 (i) ja 3.12 voimassaollessa voidaan todeta lauseen 3.13 perusteella, että yksikäsitteinen minimoija β_{min} on olemassa ja hyvin määritelty. Tarkastellaan estimaattorien $I^{(n)}_{q,\hat{\beta}_n}$ ja $I^{(n)}_{q,\beta_{min}}$ erotusta esittämällä ne yhtälön (3.9) tavoin ja merkitään tässä vektorin a alkiota $a^{(i)}$.

$$
\begin{aligned}
I^{(n)}_{q,\hat{\beta}_n} - I^{(n)}_{q,\beta_{min}} &= \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \hat{\beta}_n^{\top}\boldsymbol{Z_i} + \hat{\beta}_n^{\top}\theta\right] - \left(\frac{1}{n}\sum_{i=1}^{n} Y_i - \beta_{min}^{\top}\boldsymbol{Z_i} + \beta_{min}^{\top}\theta\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[-\hat{\beta}_n^{\top}\boldsymbol{Z_i} + \hat{\beta}_n^{\top}\theta + \beta_{min}^{\top}\boldsymbol{Z_i} - \beta_{min}^{\top}\theta\right]\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[-\sum_{j=1}^{m}\hat{\beta}_n^{(j)}Z_i^{(j)} + \sum_{j=1}^{m}\beta_{min}^{(j)}Z_i^{(j)} + \sum_{j=1}^{m}\hat{\beta}_n^{(j)}\theta^{(j)} - \sum_{j=1}^{m}\beta_{min}^{(j)}\theta^{(j)}\right]\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[-\sum_{j=1}^{m}\left(\hat{\beta}_n^{(j)} - \beta_{min}^{(j)}\right)Z_i^{(j)} + \sum_{j=1}^{m}\left(\hat{\beta}_n^{(j)} - \beta_{min}^{(j)}\right)\theta^{(j)}\right]\\
&= -\sum_{j=1}^{m}\left(\hat{\beta}_n^{(j)} - \beta_{min}^{(j)}\right)\left(\frac{1}{n}\sum_{i=1}^{n}Z_i^{(j)}\right) + \sum_{j=1}^{m}\left(\hat{\beta}_n^{(j)} - \beta_{min}^{(j)}\right)\theta^{(j)}\\
&= \sum_{j=1}^{m}\left(\hat{\beta}_n^{(j)} - \beta_{min}^{(j)}\right)\left(\theta^{(j)} - \frac{1}{n}\sum_{i=1}^{n}Z_i^{(j)}\right),
\end{aligned}
$$

missä summat vaihdetaan toisin päin niiden äärellisyyden perusteella.

Lauseesta 3.13 saadaan, että vektorille $\hat{\beta}_n - \beta_{min} = O_p(n^{-1/2})$, joten apulauseen 1.21 nojalla sama stokastinen kertaluokka pätee myös sen alkioille eli $\hat{\beta}_n^{(j)} - \beta_{min}^{(j)} = O_p(n^{-1/2})$ kaikilla $j = 1, \ldots, m$. Vastaavasti, koska satunnaismuuttujalle $Z_i^{(j)}$ pätee $\mathbb{E}\left[\theta^{(j)} - Z_i^{(j)}\right] = 0$ ja oletuksesta 3.12 seuraa Cauchy-Schwarin epäyhtälön nojalla, että $\mathbb{E}\left[\left(Z_i^{(j)}\right)^2\right] < \infty$, niin apulauseen 1.18 mukaan kaikille $j = 1, \ldots, m$ pätee

$$\frac{1}{n}\sum_{i=1}^{n}\left(\theta^{(j)} - Z_i^{(j)}\right) = O_p\left(n^{-1/2}\right).$$

Apulauseen 1.19 kohdan (3) nojalla $\left(\hat{\beta}_n^{(j)} - \beta_{min}^{(j)}\right)\left(\theta^{(j)} - \frac{1}{n}\sum_{i=1}^{n}Z_i^{(j)}\right) = O_p\left(n^{-1/2}n^{-1/2}\right) = O_p\left(n^{-1}\right)$. Vastaavasti apulauseen kohtien (2) ja (1) perusteella saadaan, että koska m on vakioivien muuttujien lukumääränä kiinteä luku, niin

$$\sum_{j=1}^{m}\left(\hat{\beta}_n^{(j)} - \beta_{min}^{(j)}\right)\left(\theta^{(j)} - \frac{1}{n}\sum_{i=1}^{n}Z_i^{(j)}\right) = O_p(mn^{-1}) = O_p(n^{-1}),$$

kun $n \to \infty$. Täten on osoitettu, että

$$I^{(n)}_{q,\hat{\beta}_n} = I^{(n)}_{q,\beta_{min}} + O_p(n^{-1}). \qquad \square$$

Owen ja Zhou toteavat artikkelissaan [9] yllä olevien lauseiden osoittavan, että vaikka estimoitu kerroin $\hat{\beta}_n^{(j)}$ lähestyy optimaalista kerrointa $\beta_{min}^{(j)}$ tavanomaisella

Monte Carlo -nopeudella, tuntemattoman kertoimen $\beta_{min}^{(j)}$ korvaaminen estimaatilla $\hat{\beta}_n^{(j)}$ on vaikutukseltaan asymptoottisesti merkityksetön. Tästä syystä estimoituja kertoimia käyttävän estimaattorin $I_{q,\hat{\beta}_n}^{(n)}$ voidaan ajatella käyttäytyvän samalla tavalla kuin tuntemattomilla optimaalisilla kertoimilla olevan estimaattorin $I_{q,\beta_{min}}^{(n)}$, eikä estimaattorin $I_{q,\hat{\beta}_n}^{(n)}$ harhasta tarvitse välittää. Tämä käytäntö on yleensä järkevää Monte Carlo otannassa, jossa otoskoko n on huomattavan suuri verrattuna vakioivien muuttujien lukumäärään m [9].

3.4. Varianssin minimointiehtojen toteutuminen

Lauseiden 3.13 ja 3.14 tarvitsemat ehdot on muotoiltu oletuksiin 3.7 (i) ja 3.12. Kuten lauseen 3.13 todistuksessa havaitaan, oletuksen 3.12 ehdot $\mathbb{E}_q\left[\left(Z^{(j)}\right)^4\right] < \infty$ ja $\mathbb{E}_q\left[\left(YZ^{(j)}\right)^2\right] < \infty$ tarvitaan takamaan oleellisten odotusarvojen äärellisyys. Ehto $\mathrm{Cov}[\boldsymbol{Z}] > 0$ on kuitenkin abstraktimpi.

Oletuksen $\mathrm{Cov}[\boldsymbol{Z}] > 0$ tarkoitus on varmistaa varianssin $\sigma_{q,\beta}^2$ (3.5) minimoijan β_{min} olemassaolo ja yksikäsitteisyys. Apulauseessa 3.8 oletuksen avulla todistetaan mahdollinen minimoija yksikäsitteiseksi, ja vastaavasti apulauseessa 3.9 oletuksen avulla osoitetaan minimoijan olemassaolon todistamiseen tarvittavan matriisin A^* kääntyvyys. Matriisin kääntyvyys on yhtäpitävä sen kanssa, että matriisin sarakevektorit ovat lineaarisesti riippumattomat. Koska matriisi $A^* = \mathbb{E}\left[\widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{Z}}^\top\right]$, jossa $\widetilde{\boldsymbol{Z}} = [1, \boldsymbol{Z}]^\top$, sen sarakevektoreiden lineaarinen riippumattomuus antaa vihjeen siitä, mitä ehto $\mathrm{Cov}[\boldsymbol{Z}] > 0$ käytännössä tarkoittaa.

Määritellään asian havainnollistamista varten multikollineaarisuuden käsite.

Määritelmä 3.15 (Multikollineaarisuus). Satunnaismuuttujat $X_1, \ldots, X_n$ ovat multikollineaarisia, jos on olemassa $b \in \mathbb{R}^n \setminus \{0\}$ siten, että $\sum_{i=1}^n b_i X_i = c \in \mathbb{R}$ melkein varmasti.

Multikollineaarisuus tarkoittaa siis sitä, että yksi satunnaismuuttujista X_i voidaan esittää toisten muuttujien ja vakion avulla. Se eroaakin lineaarisesta riippuvuudesta siten, että siinä muuttuja voi olla riippuvainen myös vakiosta $c \neq 0$. Koska matriisin A^* esittämisessä käytetty $\widetilde{\boldsymbol{Z}}$ sisältää vakion 1, pelkästään satunnaisvektorin $\boldsymbol{Z}$ alkioiden lineaarinen riippumattomuus ei riitä takaamaan matriisi A^* on kääntyvyyttä. Tarvitaan siis oletus, että vektorin $\boldsymbol{Z}$ alkiot eivät ole multikollineaarisia. Osoitetaan, että ehto $\mathrm{Cov}[\boldsymbol{Z}] > 0$ kattaa kyseisen oletuksen.

Lause 3.16. *Olkoon* $\boldsymbol{Z} = \left(Z^{(1)}, \ldots, Z^{(m)}\right)^\top$ *äärellisvarianssinen satunnaisvektori. Tällöin kovarianssimatriisi* $\mathrm{Cov}[\boldsymbol{Z}] > 0$*, jos ja vain jos satunnaismuuttujat* $Z^{(1)}, \ldots, Z^{(m)}$ *eivät ole multikollineaarisia.*

Todistus. Koska kovarianssimatriisi on määritelmänsä nojalla aina positiivisesti semidefiniitti [14], apulauseen väite on yhtäpitävä negaationsa kanssa, eli $\mathrm{Cov}[\boldsymbol{Z}]$ ei ole positiivisesti definiitti, jos ja vain jos $Z^{(1)}, \ldots, Z^{(m)}$ ovat multikollineaarisia. Todistetaan tämä.

Aloitetaan osoittamalla, että multikollineaarisuudesta seuraa, ettei kovarianssimatriisi voi olla positiivisesti definiitti. Olkoon $b \in \mathbb{R}^m \backslash \{0\}$ siten, että $\sum_{i=1}^m b_i Z^{(i)} = c \in \mathbb{R}$ melkein varmasti. Tällöin

$$b^\top \mathrm{Cov}\left[\boldsymbol{Z}\right] b = \mathrm{Var}\left[b^\top \boldsymbol{Z}\right] = \mathrm{Var}\left[\sum_{i=1}^m b_i Z^{(i)}\right] = \mathrm{Var}[c] = 0,$$

sillä vakion varianssi on nolla. Koska on olemassa $b \neq 0$ siten, että $\mathrm{Cov}[\boldsymbol{Z}] = 0$, kovarianssimatriisi ei voi olla positiivisesti definiitti.

Todistetaan seuraavaksi, että mikäli $\mathrm{Cov}[\boldsymbol{Z}]$ ei ole positiivisesti definiitti tällöin $Z^{(1)}, \ldots, Z^{(m)}$ ovat multikollineaarisia. Koska $\mathrm{Cov}[\boldsymbol{Z}]$ ei ole positiivisesti definiitti on tällöin olemassa $b \in \mathbb{R}^m \backslash \{0\}$ siten, että $b^\top \mathrm{Cov}[\boldsymbol{Z}] b = 0$. Täten $\mathrm{Var}[b^\top \boldsymbol{Z}] = 0$, mistä puolestaan seuraa, että satunnaismuuttujan $X = b^\top \boldsymbol{Z}$ on oltava vakio $c \in \mathbb{R}$. Täten $b^\top \boldsymbol{Z} = \sum_{i=1}^m b_i Z^{(i)} = c$ eli $Z^{(1)}, \ldots, Z^{(m)}$ ovat multikollineaarisia. $\square$

Kovarianssimatriisin $\mathrm{Cov}[\boldsymbol{Z}]$ positiividefiniittisyys riittää siis takaamaan minimoijan olemassaolon ja yksikäsitteisyyden. Owenin ja Zhoun [9] eivät käytä lauseessaan kyseistä oletusta, vaan he olettavat suoraan minimoijan olevan yksikäsitteinen. Kuitenkin kirjassaan [10, luku 9 s. 26] Owen toteaa, ettei lause 3.13 päde, jos kaksi muuttujista Z_j, $j = 1, \ldots, m$, ovat kollineaariset. Tällaisessa tapauksessa ongelma voidaan ratkaista pudottamalla yksi tai useampi tarpeeton muuttuja pois siten, että taataan ei-multikollineaarisuus ja täten minimoijan yksikäsitteisyys.

LUKU 4

Painotusotanta ja sekoitejakauma

Monte Carlo -menetelmän tavoin myös painotusotantaan on historian saatossa kehitetty eri tekniikoita tehokkuuden parantamiseksi. Esimerkiksi Owen ja Zhou [9] pyrkivät parantamaan menetelmää valitsemalla ehdotusjakauma q ja vakioiva muuttuja h siten, että varianssin pienentyy ja estimaattoria on turvallista käyttää. Heidän parannusehdotus liittyy sekoitepainotusotantaan (engl. *mixture importance sampling*), josta löytyy varhaisin julkaisu 1970-luvulta [9]. Owen ja Zhou mainitsevat Hesterbergin artikkelin [7] vuodelta 1995, jossa osoitetaan ensimmäisten joukossa, että sekoitepainotusotannan erikoistapauksen, niin sanotun defensiivisen painotusotannan, varianssi on rajoitettu. Owen ja Zhou laajentavat artikkelissaan tämän teoreettisen tuloksen vakioivilla muuttujilla varustettuun sekoitepainotusotantaan [9, Theorem 2]. Tässä tutkielmassa kyseinen tulos on muotoiltu lauseeseen 4.8.

Sekoitepainotusotantaan voidaan lisätä myös ositettu otanta, joka painotusotannan tavoin muuttaa estimaattoria ja otoksen poimimistapaa menetelmän tehokkuuden parantamiseksi [5]. Owen ja Zhou [9] kutsuvat deterministiseksi sekoitepainotusotannaksi menetelmää, jossa sekoitepainotusotantaan yhdistetään sekä ositettu otanta että vakioivat muuttujat. Tätä menetelmää on käsitelty alaluvussa 4.2. Viimeisessä alaluvussa 4.3 perehdytään lyhyesti monipainotusotantaan, joka myös yhdistyy ositettu otanta ja sekoitepainotusotanta.

4.1. Sekoitepainotusotanta

Simuloinnissa voi olla tilanne, jossa todennäköisyysjakaumien q_j, $j = 1, \dots, m$, joukosta toivotaan löytyvän yksi tai useampi jakauma, joka on edes karkeasti verrannollinen funktioon $|f|p$. Tilanne voi olla myös sellainen, että kiinnostuksen kohteena on yhtä aikaa useampi integroitava funktio f_j, joille on hankala löytää yhtä sopivaa ehdotusjakaumaa q [9]. Tällaisissa tilanteissa sekoitepainotusotanta voi olla hyödyllinen. Kuten menetelmän nimi vihjaa, siinä painotusotannan ehdotusjakaumana q käytetään sekoitejakaumaa q_α (engl. *mixture distribution*).

MÄÄRITELMÄ 4.1 (Sekoitejakauma). Olkoon q_j, $j = 1, \dots, m$, todennäköisyysjakaumia joukossa $\mathbb{X}$ ja $\alpha_j \geq 0$ siten, että $\sum_{j=1}^m \alpha_j = 1$. Sekoitejakauma on

$$q_\alpha := \sum_{j=1}^{m} \alpha_j q_j. \tag{4.1}$$

Määritellään tämän avulla sekoitepainotusotanta.

MÄÄRITELMÄ 4.2 (Sekoitepainotusotanta). Olkoon $X_1, \ldots, X_n$ otos sekoitejakaumasta q_α (4.1), jolle pätee $q_\alpha(x) > 0$, kun $f(x)p(x) \neq 0$. Tällöin odotusarvon I (1.1) sekoitepainotusotannan estimaattori on

$$I_\alpha^{(n)} := \frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)p(X_i)}{q_\alpha(X_i)}.$$

Samoin kuin painotusotannassa menetelmään sopivat ehdotusjakaumat q_j voidaan löytää numeerisen haun tai aihepiiriin liittyvän tiedon avulla [9]. Yksinkertainen tapa muodostaa sopiva sekoitejakauma on käyttää kohdejakaumaa p sekä todennäköisyysjakaumaa q, joka on lähes verrannollinen funktioon $|f|p$ etenkin simuloinnin kannalta tärkeällä alueella. Valitsemalla α siten, että $0 < \alpha < 1$, saadaan sekoitejakauma

$$q_\alpha(x) = \alpha p(x) + (1-\alpha)q(x). \tag{4.2}$$

Menetelmä, joka käyttää tällaista ehdotusjakaumaa, kutsutaan defensiiviseksi painotusotannaksi (engl. *defensive importance sampling*) [7, 9].

MÄÄRITELMÄ 4.3 (Defensiivinen painotusotanta). Olkoon $X_1, \ldots, X_n$ otos sekoitejakaumasta q_α (4.2). Tällöin odotusarvon I defensiivisen painotusotannan estimaattori on

$$I_\alpha^{(n)} = \frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)p(X_i)}{\alpha p(X_i) + (1-\alpha)q(X_i)}. \tag{4.3}$$

Kun $f(x)p(x) \neq 0$, on $p(x) > 0$ ja täten myös $q_\alpha > 0$. Sekoitepainotusotannan ehtoa $q_\alpha(x) > 0$, kun $f(x)p(x) \neq 0$, ei siis tarvita. Defensiivinen painotusotanta onkin tärkeä erikoistapaus sekoitepainotusotannasta. Hesterberg [7] toteaa sen vakauttavan painotusotantaa kolmella tavalla. Ensinnäkin otantapaino on ylhäältä rajoitettu:

$$\frac{p(x)}{q_\alpha(x)} \leq \frac{p(x)}{\alpha p(x)} = \frac{1}{\alpha}. \tag{4.4}$$

Menetelmän nimi ”defensiivinen” (engl. *defensive*) viittaakin siihen, että kohdejakaumaa p käytetään sekoitejakaumassa rajoittamaan painoa ylhäältä ja näin suojaudutaan painotusotannan epäonnistumiselta.

Toinen menetelmää vakauttava tapa on se, että painon yläraja (4.4) takaa ylärajan myös menetelmän asymptoottiselle varianssille. Osoitetaan tämä käyttämällä painotusotannan asymptoottista varianssia (2.3) ehdotusjakaumalla q_α (4.2).

$$\begin{aligned}
\sigma^2_{p,q_\alpha} &= \mathbb{E}_p\left[\frac{f(X)^2 p(X)}{q_\alpha(X)}\right] - I^2 \\
&\leq \mathbb{E}_p\left[\frac{f(X)^2}{\alpha}\right] - I^2 \\
&= \frac{1}{\alpha}\left(\mathbb{E}_p\left[f\left(X\right)^2\right] - I^2 + \left(1-\alpha\right) I^2\right) \\
&= \frac{1}{\alpha}\left(\sigma_p^2 + \left(1-\alpha\right) I^2\right).
\end{aligned}$$

Jos siis kohdejakaumaa p käyttävän Monte Carlo -estimaattorin $I_p^{(n)}$ (1.2) asymptoottinen varianssi $\sigma_p^2 = \mathrm{Var}[f(X)]$ on äärellinen, näin on myös defensiivisen painotusotannan asymptoottinen varianssi. Kolmas vakauttava tapa on, että painotusotantaan verrattuna defensiivisen painotusotannan lopputulos on vähemmän riippuvainen todennäköisyysjakauman q muodosta. Kohdejakauman p käyttäminen ehdotusjakaumassa q_α tuo suojaa alueen $\mathbb{X}$ kaikilla osilla, joten vaikka jakauma q on jollain alueella huonosti verrannollinen funktioon $|f|p$, se ei aiheuta pahaa vahinkoa menetelmän tehokkuudelle. [7]

Luvun 2 esimerkki 2.5 kuvastaa hyvin sitä, miten defensiivinen painotusotanta toimii paremmin tilanteessa, jossa painotusotanta epäonnistuu. Esimerkissä painotusotannan estimaattorin varianssi on ääretön, koska käytetyn ehdotusjakauman q hännät ovat ohuemmat kuin funktion $f_1 p$ hännät. Tämä ongelma voidaan välttää käyttämällä ehdotusjakaumana sekoitejakaumaa $q_\alpha = \alpha p + (1-\alpha)q$, jonka hännät eivät ole liian ohuet. Defensiivisen painotusotannan varianssin yläraja ei kuitenkaan takaa tehokkuutta kaikissa tilanteissa. Menetelmän voi nimittäin antaa suuren varianssin tilanteessa, jossa painotusotanta toimii hyvin. Owen ja Zhou owen2000safe toteavat ongelman olevan siinä, että jos jakauma q on likipitäen verrannollinen funktioon $|f|p$, mutta kohdejakauma p ei ole, niin sekoitejakauma $q_\alpha = \alpha p + (1-\alpha)q$ ei myöskään ole verrannollinen.

Havainnollistetaan defensiivisen painotusotannan epäonnistumista Owenin ja Zhoun esimerkillä [9, Example 2]. Siinä funktion $f_2 p$ ja todennäköisyysjakauman q verrannollisuus menetetään, jos käytetään sekoitejakaumaa q_α, minkä seurauksena defensiivinen painotusotanta osoittautuu epätarkemmaksi kuin painotusotanta ehdotusjakaumalla q.

Esimerkki 4.4. Olkoon $\mathbb{X} = (0,1)^5$, kohdejakauma $p = U(0,1)^5$ ja funktio

$$f_2(x) = \prod_{j=1}^{5} B(x^j, 20, 20) + 0{,}1 \times \prod_{j=1}^{5} \sin(60\pi(x^j - 1/3))\mathbb{1}_{1/3 \le x^j \le 2/3},$$

missä B on betajakauman tiheysfunktio (2.4). Valitaan todennäköisyysjakaumaksi $q(x) = \prod_{j=1}^{5} B(x^j, 20, 20)$.

Koska kullekin $j = 1, \dots, 5$ pätee $\int_0^1 \sin(60\pi(x^j - 1/3))\mathbb{1}_{1/3 \le x^j \le 2/3}\, dx^j = 0$, saadaan $\mathbb{E}_p[f_2(x)] = 1$. Sinifunktion aikaansaama värähtely funktiossa f_2 on niin pientä, että funktiota $f_2(x)p(x)$ ja jakaumaa $q(x)$ on lähes mahdoton erottaa toisistaan etenkin, kun $x^j \notin [1/3, 2/3]$, jolloin ne yhtenevät.

Owen ja Zhou tutkivat simuloinnin avulla eri menetelmien toimivuutta esimerkkien 2.5 ja 4.4 tapauksissa [9, Section 6]. Defensiivisen painotusotannan kohdalla he käyttävät arvoja $\alpha = 0{,}1$ ja $\alpha = 0{,}5$. Vertailemalla menetelmistä saatuja normalisoituja keskineliövirheitä, he toteavat painotusotannan toimivan huonosti esimerkissä 2.5, mutta esimerkin 4.4 kohdalla se toimii erinomaisesti. Defensiivinen painotusotanta puolestaan toimii molemmilla vakion α arvoilla todella hyvin esimerkissä 2.5, mutta painotusotantaan verrattuna paljon huonommin esimerkin 4.4 tapauksessa. Vastaavasti Monte Carlo -menetelmä 1.1 toimii huonosti molemmissa esimerkeissä.

Owenin ja Zhoun [9] mukaan samaan estimaattoriin on mahdollista saada sekä painotusotannan tehokkuus että defensiivisen painotusotannan suoja epäonnistumiselta.

He toteavat, että käyttämällä jakaumia p ja q vakioivina muuttujina ja ehdotusjakaumana niiden sekoitejakaumaa q_α (4.2), saadaan turvallinen ja tehokas painotusotanta. Todetaan tämä yleiselle tapaukselle eli sekoitepainotusotannalle ja sekoitejakaumalle q_α (4.1).

Todennäköisyysjakauman määritelmän nojalla sekoitejakaman q_α (4.1) komponenteille pätee $\int_{\mathbb{X}} q_j(x)\,dx = 1$. Jakaumat q_j ovat siis määritelmän 3.2 mukaisia vakioivia muuttujia. Lisäksi, koska sekoitejakaumalle $q_\alpha(x) > 0$ aina, kun $q_j(x) > 0$, komponentteja q_j voidaan käyttää vakioivina muuttujina samalla tavalla kuin vakioivia muuttujia h_j painotusotannan määritelmässä 3.4.

MÄÄRITELMÄ 4.5 (Sekoitepainotusotanta vakioivilla muuttujilla). Olkoon $X_1, \ldots, X_n$ otos sekoitejakaumasta q_α (4.1), jolle $q_\alpha(x) > 0$, kun $f(x)p(x) \neq 0$. Olkoon lisäksi $\beta \in \mathbb{R}^m$. Tällöin odotusarvon I sekoitepainotusotannan estimaattori vakioivin muuttujin on

$$I_{\alpha,\beta}^{(n)} := \frac{1}{n}\sum_{i=1}^{n} \frac{f(X_i)p(X_i) - \sum_{j=1}^{m}\beta_j q_j(X_i)}{q_\alpha(X_i)} + \sum_{j=1}^{m}\beta_j. \tag{4.5}$$

LAUSE 4.6. *Olkoon estimaattori $I_{\alpha,\beta}^{(n)}$ kuten määritelmässä 4.5. Tällöin kaikilla $\beta \in \mathbb{R}^m$ pätee, että estimaattori $I_{\alpha,\beta}^{(n)}$ on harhaton.*

TODISTUS. Seuraa apulauseesta 3.5. □

Menetelmä vastaa siis vakioivilla muuttujilla varustettua painotusotantaa sijoituksilla $q = q_\alpha$, $h_j = q_j$ ja $\theta_j = 1$. Täten sen varianssi on sama kuin vakioivat muuttujat sisältävän painotusotannan estimaattorin $I_{q,\beta}^{(n)}$ (3.4) varianssi $\sigma_{q,\beta}^2$ (3.5).

APULAUSE 4.7. *Estimaattorin $I_{\alpha,\beta}^{(n)}$ (4.5) varianssi on*

$$\operatorname{Var}\left[I_{\alpha,\beta}^{(n)}\right] = \frac{\sigma_{\alpha,\beta}^2}{n},$$

missä

$$\sigma_{\alpha,\beta}^2 = \mathbb{E}_{q_\alpha}\left[\left(\frac{f(X)p(X) - \sum_{j=1}^{m}\beta_j q_j(X)}{q_\alpha(X)} + \sum_{j=1}^{m}\beta_j - I\right)^2\right] \tag{4.6}$$

Kuten painotusotannassakin vakioivien muuttujien kerroin $\beta = (\beta_1, \ldots, \beta_m)^\top$ halutaan valita siten, että varianssi $\sigma_{\alpha,\beta}^2$ minimoituu. Owen ja Zhou [9] nostavat esille mahdollisen tapauksen, jossa yksi tai useampi epäsopiva jakauma q_j häiritsee otoksella estimoitua kerrointa $\hat{\beta}_n$ tarpeeksi tuhotakseen estimaattoreiden $I_{\alpha,\hat{\beta}_n}^{(n)}$ ja $I_{\alpha,\beta_{min}}^{(n)}$ välisen asymptoottisen yhtenevyyden. Lauseen 3.14 mukaan näin ei kuitenkaan käy, kunhan tarvittavat oletukset 3.7 (i) ja 3.12 toteutuvat. Oletus 3.7 (i) toteutuu vakioivalla muuttujalla varustetun sekoitepainotusotannan määritelmän 4.5 nojalla. Oletus 3.12 sisältää puolestaan ehdot: kovarianssimatriisi $\operatorname{Cov}[\boldsymbol{Z}]$ on positiivisesti definiittinen sekä muuttujien $(Y)^2$, $(Z^{(j)})^4$ ja $(YZ^{(j)})^2$ odotusarvot ovat äärellisiä.

Sekoitepainotusotannan tapauksessa, kun $X \sim q_\alpha$, niin $Y = f(X)p(X)/q_\alpha(X)$ ja $Z^{(j)} = q_j(X)/q_\alpha(X)$. Koska $q_j^4(x)/q_\alpha^4(x) \leq \alpha_j^{-4}$ kaikille x ja $\alpha_j > 0$, niin ehto

$\mathbb{E}[(Z^{(j)})^4] < \infty$ seuraa helposti. Oletetaan, että vähintään yhdelle jakaumalle q_k varianssi σ^2_{p,q_k} (2.3) on äärellinen. Tällöin, kuten Owen ja Zhou [9] osoittavat,

$$\begin{aligned}
\mathbb{E}\left[\left(YZ^{(j)}\right)^2\right] &= \int_{\mathbb{X}} \frac{f^2(x)p^2(x)q_j^2(x)}{q_\alpha^3(x)}\,dx \\
&\leq \alpha_j^{-2} \int_{\mathbb{X}} \frac{f^2(x)p^2(x)}{q_\alpha(x)}\,dx \\
&\leq \alpha_j^{-2}\alpha_k^{-1} \int_{\mathbb{X}} \frac{f^2(x)p^2(x)}{q_k(x)}\,dx \\
&= \alpha_j^{-2}\alpha_k^{-1}\left(\sigma^2_{p,q_k} + I^2\right) < \infty.
\end{aligned}$$

Vastaavalla tavalla osoitetaan, että $\mathbb{E}\left[(Y)^2\right] < \infty$.

Tarkastellaan seuraavaksi, että toteutuuko ehto kovarianssimatriisi $\mathrm{Cov}[\boldsymbol{Z}]$ positiividefiniittisyydestä. Sekoitejakaumassa q_α (4.1) esiintyville kertoimille $\alpha_j \geq 0$ pätee $\sum_{j=1}^m \alpha_j Z^{(j)} = 1$. Jos $a_j \neq 0$ kaikilla $j = 1, \ldots, m$, niin satunnaismuuttujat $Z^1, \ldots, Z^m$ ovat määritelmän 3.15 mukaan multikollineaarisia. Täten lauseen 3.16 nojalla kovarianssimatriisi $\mathrm{Cov}[\boldsymbol{Z}]$ ei ole positiivisesti definiitti, mistä puolestaan seuraa se, ettei varianssin minimoija ole yksikäsitteinen. Owen ja Zhou [9] havainnollistavat asiaa seuraavasti: mille tahansa vakiolle $c \in \mathbb{R}$ pätee $I^{(n)}_{\alpha,\beta+c\alpha} = I^{(n)}_{\alpha,\beta}$ eli jos β_{min} minimoi varianssin $\sigma^2_{\alpha,\beta}$, niin samoin minimoi $\beta_{min} - c\alpha$. Jos valitaan $c = [\beta_{min}]_{j'}/\alpha_{j'}$, niin saadaan minimoija $\beta'_{min} = \beta_{min} - c\alpha$, jonka j'. alkio $[\beta_{min} - c\alpha]_{j'} = 0$. Joku minimoijan β_{min} alkioista voidaan siis valita nollaksi ja kyseessä on edelleen minimoija. Täten asettamalla kertoimen β j'. alkio $\beta_{j'}$ nollaksi, mikä vastaa jakauman $q_{j'}$ pois jättämistä vakioivana muuttujana, saadaan multikollineaarisuus poistettua ja varmistettua varianssin minimoijan yksikäsitteisyys.

Jos siis sekoitejakaumassa q_α vähintään yhdellä jakaumalla q_k on äärellinen painotusotannan varianssi ja varmistetaan, että vakioivilla muuttujilla muodostetut satunnaismuuttujat $Z^{(1)}, \ldots, Z^{(m)}$ eivät ole multikollineaarisia, niin vakioivilla muuttujilla varustettu sekoitejakauma toteuttaa oletukset 3.7 (i) ja 3.12. Täten lauseen 3.14 mukaan estimoidun kertoimen estimaattori $I^{(n)}_{\alpha,\hat{\beta}_n}$ käyttäytyy kuten estimaattori $I^{(n)}_{\alpha,\beta_{min}}$ optimaalisella kertoimella. Tällöin ei tarvitse huolehtia siitä, että huono valinta sekoitejakauman q_α komponenteissa q_j häiritsisi menetelmää liikaa.

Käyttämällä todennäköisyysjakaumia vakioivina muuttujina pystytään todistamaan, että sekoitepainotusotannan varianssi ei ole kovin paljon huonompi kuin painotusotannan varianssi parhaimmillaan. Optimaalisen kertoimen estimaattorin $I^{(n)}_{\alpha,\beta_{min}}$ varianssille saadaan nimittäin yläraja, kuten alla oleva lause osoittaa. Lauseen todistus on Owenin ja Zhoun artikkelista [9, Theorem 2].

Lause 4.8. *Olkoon estimaattori* $I^{(n)}_{\alpha,\beta}$ *kuten määritelmässä 4.5 ja* $\sigma^2_{\alpha,\beta}$ (4.6) *sen varianssi. Oletetaan, että on olemassa varianssin* $\sigma^2_{\alpha,\beta}$ *minimoija* β_{min}. *Tällöin*

$$\sigma^2_{q_\alpha,\beta_{min}} \leq \min_{1\leq j\leq m} \frac{\sigma^2_{p,q_j}}{\alpha_j}, \tag{4.7}$$

missä varianssi σ^2_{p,q_j}, *on annettu yhtälössä* (2.3).

Todistus. Oletetaan kaikille $k \in \{1, \ldots, m\}$ pätevän, että jakauma $q_k(x) > 0$, kun $f(x)p(x) > 0$. Valitaan $k = 1$. Tarkastellaan varianssia $\sigma^2_{q_\alpha, \tilde{\beta}}$, jossa kertoimelle $\tilde{\beta} \in \mathbb{R}^m$ pätee $\tilde{\beta}_1 = 0$ ja $\tilde{\beta}_j = -I\alpha_j/\alpha_1$, kun $j = 2, \ldots, m$.

$$
\begin{aligned}
\sigma^2_{q_\alpha,\tilde{\beta}} &= \mathbb{E}_{q_\alpha}\left[\left(\frac{f(X)p(X) - \sum_{j=1}^m \tilde{\beta}_j q_j(X)}{q_\alpha(X)} + \sum_{j=1}^m \tilde{\beta}_j - I\right)^2\right] \\
&= \int_{\mathbb{X}} \left(\frac{f(x)p(x) + I\alpha_1^{-1}\sum_{j=2}^m \alpha_j q_j(x)}{q_\alpha(x)} - I\alpha_1^{-1}\sum_{j=2}^m \alpha_j - I\right)^2 q_\alpha(x)\,dx \\
&= \int_{\mathbb{X}} \left(\frac{f(x)p(x) + I\alpha_1^{-1}\left(q_\alpha(x) - \alpha_1 q_1(x)\right)}{q_\alpha(x)} - I\alpha_1^{-1}\left(\sum_{j=1}^m \alpha_j - \alpha_1\right) - I\right)^2 q_\alpha(x)\,dx \\
&= \int_{\mathbb{X}} \left(\frac{f(x)p(x) - Iq_1(x)}{q_\alpha(x)} + I\alpha_1^{-1} - I\alpha_1^{-1}(1-\alpha_1) - I\right)^2 q_\alpha(x)\,dx \\
&= \int_{\mathbb{X}} \left(\frac{f(x)p(x) - Iq_1(x)}{q_\alpha(x)} + I\alpha_1^{-1} - I\alpha_1^{-1} + I - I\right)^2 q_\alpha(x)\,dx \\
&= \int_{\mathbb{X}} \left(\frac{f(x)p(x) - Iq_1(x)}{q_\alpha(x)}\right)^2 q_\alpha(x)\,dx \\
&= \int_{\mathbb{X}} \frac{\left(f(x)p(x) - Iq_1(x)\right)^2}{q_\alpha(x)}\,dx \\
&\leq \int_{\mathbb{X}} \frac{\left(f(x)p(x) - Iq_1(x)\right)^2}{\alpha_1 q_1(x)}\,dx \\
&= \frac{1}{\alpha_1}\int_{\mathbb{X}} \left(\frac{f(x)p(x)}{q_1(x)} - I\right)^2 q_1(x)\,dx \\
&= \frac{\sigma^2_{p,q_1}}{\alpha_1},
\end{aligned}
$$

sillä $q_\alpha(x) = \sum_{j=1}^m \alpha_j q_j(x) \geq \alpha_1 q_1(x)$. Vastaava tulos saadaan kaikille $k \in \{2, \ldots, m\}$ valitsemalla vektorin $\tilde{\beta}$ alkioiksi $\tilde{\beta}_k = 0$ ja $\tilde{\beta}_j = -I\alpha_j/\alpha_k$, kun $j \neq k$. Koska vektori β_{min} oletettiin varianssin $\sigma^2_{\alpha,\beta}$ minimoijaksi, saadaan

$$\sigma^2_{q_\alpha,\beta_{min}} \leq \sigma^2_{q_\alpha,\tilde{\beta}} \leq \sigma^2_{p,q_j}/\alpha_j$$

kaikille $j = 1, \ldots, m$. Täten epäyhtälö (4.7) pätee. □

Lauseen 4.8 mukaan estimaattorin $I^{(n)}_{\alpha,\beta_{min}}$ varianssille pätee $\sigma^2_{\alpha,\beta_{min}}/n \leq \sigma^2_{p,q_j}/n\alpha_j$ kaikilla $j = 1, \ldots, m$. Varianssi $\sigma^2_{\alpha,\beta_{min}}$ on siis vähintään yhtä hyvä kuin estimaattorin $I^{(n)}_{q_j,\beta}$ (3.4) varianssi σ^2_{p,q_j}, joka saadaan $n\alpha_j$ näytteellä jakaumasta q_j. Owenin ja Zhoun [9] mukaan on vaikea kuvitella, että yleisesti ottaen voitaisiin muodostaa tehokkaampi mutta yhtä robusti estimaattori, sillä sekoitejakauman komponentista q_j saadaan keskimäärin $n\alpha_j$ näytettä ja jakauma q_j saattaa olla sekoitejakauman komponenteista ainoa, joka antaa äärellisen varianssin painotusotannassa. Toisin sanoen sekoitejakauman huonot komponentit q_k eivät tee estimaatista huonompaa kuin, mitä olisi saatu vain yhden hyvän komponentin q_j ”näytteiden” käytöstä.

Sekoitepainotusotanta vakioivilla muuttujilla on siis menetelmä, jossa yhdistyy painotusotannan tehokkuus ja defensiivisen painotusotannan suoja epäonnistumiselta. Owen ja Zhou esittelevätkin artikkelissaan [9] tämän menetelmän painotusotannan parannusehdotuksena ja kuvailevat sitä sanoilla turvallinen ja tehokas painotusotanta (engl. *safe and effective importance sampling*). Käytännön kannalta menetelmää voidaan vielä parantaa, ja perehdytään siihen seuraavaksi.

4.2. Deterministinen sekoitepainotusotanta

Sekoitejakaumasta $q_\alpha(x) = \sum_{j=1}^m \alpha_j q_j(x)$ poimitaan yleensä otos siten, että todennäköisyysjakaumasta q_j otetaan riippumattomia näytteitä X_i todennäköisyydellä α_j. Esimerkiksi defensiivisen painotusotannan (4.3) tapauksessa tämä käytännössä tapahtuu siten, että poimitaan muuttuja U_k tasajakaumasta $U(0,1)$. Jos $U_k \leq \alpha$, niin $X_k \sim p$, ja jos $U_k > \alpha$, niin $X_k \sim q$. Kustakin jakaumasta q_j poimittujen näytteiden lukumäärät ovat siis satunnaisia - tarkemmin sanottuna binomijakauman $Binom(\alpha_j, n)$ mukaisia. Tämä satunnaisuus voidaan kuitenkin poistaa.

Hesterberg [7] toteaa, että paras tapa poimia otos sekoitejakaumasta on osittaa näytteiden lukumäärä suhteessa jakaumien todennäköisyyteen. Tällainen otoskokojen deterministinen päättäminen kullekin sekoitejakauman komponentille q_j on varianssin pienennysmenetelmä, jota kutsutaan ositetuksi otannaksi (engl. *stratified sampling*). Yleisesti sen idea on jakaa integroimisalue pistevieraisiin osa-alueisiin, toteuttaa estimointi kullekin osa-alueelle erikseen ja summata estimaatit yhteen [6]. Kuitenkin kun ositettua otantaa sovelletaan sekoitejakaumaan q_α, kustakin jakaumasta q_j otetaan $n_j > 0$ riippumatonta näytettä X_{ij}. Lukumäärä n_j voidaan valita eri tavoin, mutta varianssin pienentämisen kannalta hyödytään suhteellisesta kiintiöinnistä (engl. *proportional allocation*). Tällöin näytteiden määrä n_j kiintiöidään suhteessa todennäköisyyteen α_j eli valitaan $n_j = n\alpha_j$ tai sen lähin kokonaisluku. Esimerkiksi defensiivisessä painotusotannassa otettaisiin $n\alpha$ näytettä jakaumasta p ja $n(1-\alpha)$ näytettä jakaumasta q.

Yhdistämällä ositettu otanta vakioivilla muuttujilla varustettuun sekoitepainotusotantaan saadaan menetelmä, jota Owen ja Zhou kutsuvat deterministiseksi sekoitepainotusotannaksi (engl. *deterministic mixture sampling*) [9].

MÄÄRITELMÄ 4.9 (Deterministinen sekoitepainotusotanta). Olkoon sekoitejakauma q_α (4.1) siten, että $q_\alpha(x) > 0$, kun $f(x)p(x) \neq 0$, ja $\beta \in \mathbb{R}^m$. Olkoon lisäksi $n_j = n\alpha_j \in \mathbb{N}$, jossa $j = 1, \ldots, m$, ja $(X_{ij})_{i=1,\ldots,n_j}$ otos todennäköisyysjakaumasta q_j. Odotusarvon I deterministisen sekoitepainotusotannan estimaattori on

$$\tilde{I}_{\alpha,\beta}^{(n)} := \frac{1}{n}\left(\sum_{j=1}^{m}\sum_{i=1}^{n_j} \frac{f(X_{ij})p(X_{ij}) - \sum_{k=1}^m \beta_k q_k(X_{ij})}{q_\alpha(X_{ij})}\right) + \sum_{j=1}^{m} \beta_j. \tag{4.8}$$

APULAUSE 4.10. *Estimaattori* $\tilde{I}_{\alpha,\beta}^{(n)}$ (4.8) *on harhaton.*

TODISTUS. Koska sekoitepainotusotannan estimaattori vakioivilla muuttujilla $I_{\alpha,\beta}^{(n)}$ (4.5) on lauseen 4.6 nojalla harhaton, riittää osoittaa, että $\mathbb{E}\left[\tilde{I}_{\alpha,\beta}^{(n)}\right] = \mathbb{E}\left[I_{\alpha,\beta}^{(n)}\right]$.

Merkitään

$$\varphi(X_{ij}) = \frac{f(X_{ij})p(X_{ij}) - \sum_{k=1}^{m} \beta_k q_k(X_{ij})}{q_\alpha(X_{ij})} + \sum_{j=1}^{m} \beta_j. \tag{4.9}$$

Tällöin estimaattori

$$\tilde{I}_{\alpha,\beta}^{(n)} = \frac{1}{n} \left(\sum_{j=1}^{m} \sum_{i=1}^{n_j} \varphi(X_{ij}) \right) \tag{4.10}$$

ja odotusarvolle saadaan

$$\begin{aligned} \mathbb{E}\left[\tilde{I}_{\alpha,\beta}^{(n)}\right] &= \frac{1}{n} \sum_{j=1}^{m} n_j \mathbb{E}\left[\varphi\left(X_{1j}\right)\right] = \frac{1}{n} \sum_{j=1}^{m} n\alpha_j \mathbb{E}_{q_j}\left[\varphi(X)\right] \\ &= \mathbb{E}_{q_\alpha}\left[\varphi(X)\right] = \mathbb{E}\left[I_{\alpha,\beta}^{(n)}\right]. \end{aligned}$$

□

Kuten todistuksessa osoitetaan, estimaattorin osittaminen ei muuta sen odotusarvoa. Owen ja Zhou [9] toteavat, että deterministisen estimaatin harhattomuus kattaa myös satunnaismuuttujien $(Z^{(j)})^4$ ja $(YZ^{(j)})^2$, missä $Z^{(j)} = q_k(X)/q_\alpha(X)$ ja $Y = f(X)p(X)/q_\alpha(X)$, odotusarvot, mistä seuraa, että vakioivin muuttujin varustetun sekoitepainotusotannan tavoin lauseet 3.13 ja 3.14 pätevät myös deterministiselle sekoitepainotusotannalle sopivin lisäoletuksin. Myös lause 4.8 pätee menetelmälle ja deterministisen sekoitepainotusotannan hyvin toteutettu ositus voi parantaa tehokkuutta lisää. Kun osituksessa käytetään määritelmän 4.9 kaltaista suhteellista kiintiöintiä, niin pätee, että ositetun estimaattorin varianssi ei ole suurempi kuin ilman ositusta olevan estimaattorin varianssi.

LAUSE 4.11. *Olkoon deterministisen sekoitepainotusotannan estimaattori $\tilde{I}_{\alpha,\beta}^{(n)}$ kuten määritelmässä 4.9 ja sekoitepainotusotannan estimaattori vakioivin muuttujin $I_{\alpha,\beta}^{(n)}$ kuten määritelmissä 4.5. Tällöin*

$$\operatorname{Var}\left[\tilde{I}_{\alpha,\beta}^{(n)}\right] \leq \operatorname{Var}\left[I_{\alpha,\beta}^{(n)}\right]$$

TODISTUS. Sovelletaan todistusta [12, Proposition 4.3.1 s. 133]. Olkoon $\varphi(X_{ij})$ kuten (4.9). Tällöin estimaattori $\tilde{I}_{\alpha,\beta}^{(n)}$ on muotoa (4.10), ja koska $X_{1j}, \ldots, X_{n_j j}$ ovat riippumattomia satunnaismuuttujia jakaumasta q_j, estimaattorin $\tilde{I}_{\alpha,\beta}^{(n)}$ varianssi voidaan esittää muodossa

$$\operatorname{Var}\left[\tilde{I}_{\alpha,\beta}^{(n)}\right] = \frac{\sum_{j=1}^{m} n_j \operatorname{Var}[\varphi(X_{1j})]}{n^2} = \frac{\sum_{j=1}^{m} \alpha_j \operatorname{Var}[\varphi(X_{1j})]}{n}. \tag{4.11}$$

Merkitään myös $\tilde{I}_j = \int_{\mathbb{X}} \varphi(x)\alpha_j q_j(x)\, dx$ eli

$$\tilde{I}_j = \alpha_j \mathbb{E}\left[\varphi\left(X_{1j}\right)\right].$$

Hyödyntämällä Cauchy-Schwarzin epäyhtälöä integraalille $\tilde{I} = \int_{\mathbb{X}} \varphi(x) q_\alpha(x)\, dx$ voidaan täten osoittaa, että

$$\tilde{I}^2 = \left(\sum_{j=1}^{m} \tilde{I}_j \right)^2 = \left(\sum_{j=1}^{m} \frac{\tilde{I}_j}{\sqrt{\alpha_j}} \sqrt{\alpha_j} \right)^2 \leq \sum_{j=1}^{m} \frac{\tilde{I}_j^2}{\alpha_j} \sum_{j=1}^{m} \alpha_j = \sum_{j=1}^{m} \frac{\tilde{I}_j^2}{\alpha_j}. \tag{4.12}$$

Koska

$$\begin{aligned}\mathrm{Var}\left[\varphi(X_{1j})\right] &= \mathbb{E}\left[\varphi(X_{1j})^2\right] - \mathbb{E}[\varphi(X_{1j})]^2 = \int_{\mathbb{X}} \varphi(x)^2 q_j(x)\,dx - \frac{\tilde{I}_j^2}{\alpha_j^2} \\ &= \frac{1}{\alpha_j}\int_{\mathbb{X}} \varphi^2(x)\alpha_j q_j(x)\,dx - \frac{\tilde{I}_j^2}{\alpha_j^2},\end{aligned}$$

niin

$$\sum_{j=1}^{m} \alpha_j \mathrm{Var}[\varphi(X_{1j})] = \int_{\mathbb{X}} \varphi(x)^2 q_\alpha(x)\,dx - \sum_{j=1}^{m} \frac{\tilde{I}_j^2}{\alpha_j}.$$

Täten hyödyntämällä epäyhtälöä (4.12) saadaan

$$\sum_{j=1}^{m} \alpha_j \mathrm{Var}[\varphi(X_{1j})] \le \int_{\mathbb{X}} \varphi^2(x) q_\alpha(x)\,dx - \tilde{I}^2 = n\mathrm{Var}[I_{\alpha,\beta}^{(n)}]$$

eli yhtälön (4.11) perusteella $\mathrm{Var}[\tilde{I}_{\alpha,\beta}^{(n)}] \le \mathrm{Var}[I_{\alpha,\beta}^{(n)}]$. □

Yllä oleva tulos voidaan yleistää myös Monte Carlo -estimaattorille sekä tapaukselle, jossa n_j ei ole kokonaisluku (katso [17]). Kiintiöimällä näytteiden lukumäärä voidaan siis poistaa otoskokojen satunnaisuutta, mikä näkyy varianssin mahdollisena pienentymisenä. Tästä nähdäänkin esimerkki Owenin ja Zhoun [9] suorittamassa simuloinnissa. Kuten alaluvussa 4.1 todettiin, painotusotanta ja defensiivinen painotusotanta toimi hyvin vain toisessa esimerkkien 2.5 ja 4.4 tapauksessa. Testatessaan determinististä sekoitepainotusotantaa, he käyttivät estimaattoria $\tilde{I}_{\alpha,\beta}^{(n)}$ (4.8) siten, että $m = 2$, $q_1 = p$, $q_2 = q$ ja $n_j = n\alpha_j$. He kokeilivat menetelmää kahdella kertoimen α arvolla, $\alpha_1 = 0,1$ ja $\alpha_2 = 0,9$ sekä $\alpha_1 = 0,5$ ja $\alpha_2 = 0,5$.

Owen ja Zhou [9] toteavat, että lauseen 4.8 tuloksen mukaisesti, deterministinen sekoitepainotusotanta toimii molemmissa esimerkeissä lähes yhtä hyvin kuin parempi painotusotannasta ja defensiivisestä painotusotannasta. Esimerkkien 2.5 ja 4.4 tapauksessa se siis osoittautuu tehokkaaksi ja turvalliseksi menetelmäksi. Lisäksi Owen ja Zhou saavat tulokseksi, että kertoimen α_1 valinnalla on vähemmän merkitystä kuin itse menetelmän valinnalla. He uskovat, että niin kauan, kun yksikään α_j ei ole liian pieni, hyöty, joka saataisiin menetelmän optimoinnista kertoimen α suhteen, on vähäinen verrattuna kertoimen β estimoinnista saatuun hyötyyn varianssin pienentämisessä.

Sekoitejakaumaa ja ositettua otantaa voidaan hyödyntää painotusotannassa myös toisella tavalla. Esitellään viimeisenä versio painotusotannan estimaattorista, jonka avulla on mahdollista estimoida vielä paremmat kertoimet etenkin ositetun otannan suhteen [9].

4.3. Monipainotusotanta

Sekoitejakauman q_α (4.1) käyttäminen painotusotannan ehdotusjakaumana mahdollistaa monipuolisen estimaattorivalikoiman. Samoin kuin deterministisessä sekoitepainotusotannassa, monipainotusotannassa (engl. *multiple importance sampling*) hyödynnetään ositettua otantaa poimimalla $n_j > 0$ näytettä jakaumasta q_j. Lisäksi menetelmässä käytetään ykkösen ositusta (engl. *partition of unity*).

MÄÄRITELMÄ 4.12 (Ykkösen ositus). Funktioperhe $\omega_j : \mathbb{X} \to [0,1]$, missä $j = 1,\dots,m$, on ykkösen ositus, jos kaikille $x \in \mathbb{X}$ pätee

$$\sum_{j=1}^{m} \omega_j(x) = 1.$$

MÄÄRITELMÄ 4.13 (Monipainotusotanta). Olkoon ω_j ykkösen ositus ja todennäköisyysjakaumat q_j, $j = 1\dots,m$, siten, että $q_j(x) > 0$, kun $\omega_j(x)f(x)p(x) \neq 0$. Olkoon myös $(X_{ij})_{i=1,\dots,n_j}$ otos jakaumasta q_j. Odotusarvon I monipainotusotannan estimaattori on

$$\tilde{I}_{\omega}^{(n)} := \sum_{j=1}^{m} \frac{1}{n_j} \sum_{i=1}^{n_j} \omega_j(X_{ij}) \frac{f(X_{ij})p(X_{ij})}{q_j(X_{ij})}. \tag{4.13}$$

Estimaattori $\tilde{I}_{\omega}^{(n)}$ on muiden estimaattorien tavoin harhaton [16] ja sen varianssin pienentämiseksi on kehitetty eri tapoja valita funktiot ω_j. Kun Veach ja Guibas esittelevät artikkelissaan [16] monipainotusotannan, he toteavat funktioiden ω_j valinnan olevan heuristinen eli ratkaisun löydetään yrityksen ja erheen kautta tai löyhästi määriteltyjen sääntöjen avulla. Artikkelissaan he esittelevät yhteensä neljä erilaista tapaa valita funktiot ω_j. Eräs niistä on tasapainoinen heuristiikka (engl. *balance heuristic*)

$$\omega_j(x) = \frac{n_j q_j(x)}{\sum_{k=1}^{m} n_k q_k(x)}.$$

Tällöin, kun valitaan $n_j = n\alpha_j$, estimaattori $\tilde{I}_{\omega}^{(n)}$ (4.13) yhtenee luvuilla n_j ositetun sekoitepainotusotannan kanssa, mikä puolestaan vastaa determinististä sekoitepainotusotantaa, kun $\beta_j = 0$ kaikilla $j = 1,\dots,m$.

Owen ja Zhou käyttävät myös monipainotusotantaa simuloidessaan eri menetelmien toimivuutta esimerkeissä 2.5 ja 4.4. He kokeilevat menetelmää neljällä Veachin ja Guibasin artikkelin [16] heuristiikalla. Eri heuristiikat antavat normalisoiduilta keskineliövirheiltään lähes samat tulokset molemmissa esimerkeissä. Esimerkin 2.5 tapauksessa heuristiikat toimivat huomattavasti paremmin kuin painotusotanta ja tilastollisen merkitsevyyden rajoissa yhtä hyvin kuin defensiivinen painotusotanta ja deterministinen sekoitepainotusotanta. Vastaavasti esimerkissä 4.4 menetelmät eivät onnistuneet yhtä hyvin kuin painotusotanta ja deterministinen sekoitepainotusotanta, mutta toimivat hieman paremmin kuin defensiivinen painotusotanta.

Veach ja Guibas [16] kehittävät monipainotusotantaa kuvien renderöintiin tietokonegrafiikassa. Heidän ideansa on yhdistää useampi painotusotannan strategia saadakseen menetelmä, joka toimii lähes yhtä hyvin kuin paras mahdollinen strategia [9]. Owen ja Zhou [9] kehittävät monipainotusotantaa vielä pidemmälle yhdistämällä siihen vakioivat muuttujat ja niin sanotun positivisaatiomenetelmän, jossa erotellaan funktion f positiiviset ja negatiiviset alueet. Heidän artikkelinsa toinen painotusotannan parannusehdotus liittyy tähän (katso [9]).

Varianssin pienentäminen voidaan siis nähdä tapana käyttää tunnettua tietoa ongelmasta [12]. Kuten luvun 2 alussa lainattu Hammersley ja Handscomb toteamus Monte Carlo -menetelmän periaatteesta tuo ilmi, otosta ei kannata simuloida yksistään siitä jakaumasta, joka nousee ongelmasta. Mitä enemmän tiedetään ongelmasta, sitä tehokkaampia menetelmiä voidaan käyttää.

Kirjallisuus

[1] Juha Alho et al. *Tilastotieteen sanasto*. Suomen tilastoseuran julkaisuja. Helsinki: Suomen Tilastoseura ry, 2021.

[2] Russell Barton, Marvin K. Nakayama ja Lee Schruben. "History of improving statistical efficiency". Teoksessa: *2017 Winter Simulation Conference (WSC)*. 2017, s. 158–180.

[3] Yvonne M. M. Bishop, Stephen E. Fienberg ja Paul W. Holland. *Discrete multivariate analysis : theory and practice*. 11. painos. Cambridge: MIT Press, 1991.

[4] Stephen L. Campbell ja Jr Meyer Carl D. *Generalized inverses of linear transformations*. Surveys and reference works in mathematics. Lontoo: Pitman, 1979.

[5] George S. Fishman. *Monte Carlo : concepts, algorithms, and applications*. 3. korjattu painos. Springer series in operations research. New York: Springer, 1999.

[6] J. M. Hammersley ja D. C. Handscomb. *Monte Carlo methods*. uudistettu painos. Lontoo: Chapman ja Hall, 1979.

[7] Tim Hesterberg. "Weighted Average Importance Sampling and Defensive Mixture Distributions". *Technometrics* 37.2 (1995), s. 185–194.

[8] Roger A. Horn ja Charles R. Johnson. *Matrix analysis*. 19. painos. Cambridge: Cambridge University Press, 2005.

[9] Art Owen ja Yi Zhou. "Safe and effective importance sampling". *Journal of the American Statistical Association* 95.449 (2000), s. 135–143.

[10] Art B. Owen. *Monte Carlo theory, methods and examples*. Viitattu 8.12.2023. `https://artowen.su.domains/mc/`, 2013.

[11] Christian P. Robert ja George Casella. *Monte Carlo statistical methods*. Springer texts in statistics. New York: Springer, 1999.

[12] Reuven Y. Rubinstein. *Simulation and the Monte Carlo method*. Wiley series in probability and mathematical statistics. New York: John Wiley & Sons, 1981.

[13] James R Schott. *Matrix analysis for statistics*. 3. painos. Hoboken, New Jersey: John Wiley & Sons, 2016.

[14] Jun Shao. *Mathematical statistics*. 2. painos. Springer texts in statistics. New York: Springer, 2003.

[15] Lloyd N. Trefethen ja David Bau. *Numerical linear algebra*. Philadelphia: SIAM, 1997.

[16] Eric Veach ja Leonidas Guibas. "Optimally Combining Sampling Techniques for Monte Carlo Rendering". Teoksessa: *SIGGRAPH '95: Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*. New York: Association for Computing Machinery, 1995, s. 419–428.

[17] Matti Vihola. *Lectures on stochastic simulation*. Luentomoniste, Jyväskylän yliopisto. 2020.