HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

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

Tekijä/Författare – Author
Joonas Havukainen

Työn nimi / Arbetets titel – Title
Berry-Esseen-lauseen todistus Steinin menetelmällä

Oppiaine /Läroämne – Subject
Soveltava matematiikka

Työn laji/Arbetets art – Level	Aika/Datum – Month and year	Sivumäärä/ Sidoantal – Number of pages
Pro gradu-tutkielma	Joulukuu 2017	40 s.

Tiivistelmä/Referat – Abstract
Tutkielma perehdyttää lukijan Steinin menetelmään normaaliapproksimaatiolle sekä esittää tämän avulla todistuksen Berry-Esseen-lauseelle. Steinin menetelmä on todennäköisyysteorian piiriin kuuluva nykyaikainen ja tehokas tapa tuottaa ylärajoja kahden eri todennäköisyysjakauman väliselle etäisyydelle.

Tutkielmassa esitetään todennäköisyysjakaumien etäisyydelle kolme eniten käytettyä mittaa, jotka ovat Total variation, Kolmogorov sekä Wasserstein-mitat. Tämän jälkeen käydään läpi Steinin menetelmä aloittaen Steinin lemmasta, joka karakterisoi normaalijakauman Steinin operaattorin avulla siten, että operaattorin arvon ollessa nolla, on tarkasteltava jakauma normaali. Seuraavaksi esitetään Steinin yhtälöt, joiden ratkaisujen avulla saadaan Steinin rajoitukset jokaiselle käytetylle kolmelle mitalle. Näiden rajoitusten avulla voidaan päätellä asymptoottinen normaalijakautuneisuus myös silloin, kun Steinin operaattorin arvo on lähellä nollaa.

Berry-Esseen-lause on keskeinen raja-arvolause, johon on erityisesti lisätty suppenemisnopeus Kolmogorov-etäisyyden suhteen. Tämä suppenemisnopeus todistetaan tutkielmassa käyttäen hyväksi Steinin menetelmää.

Lopuksi käsitellään vielä ylimalkaisesti Steinin menetelmää moniulotteisen jakauman tapauksessa. Huomataan sen olevan hyvin paljon samankaltaista kuin yksiulotteisessa tapauksessa.

Avainsanat – Nyckelord – Keywords
Steinin menetelmä, Steinin lemma, Steinin yhtälö, jakaumien etäisyys, Total variation, Kolmogorov, Wasserstein, Berry-Esseen-lause

Säilytyspaikka – Förvaringställe – Where deposited
Kumpulan tiedekirjasto, Helda

Muita tietoja – Övriga uppgifter – Additional information

Berry-Esseen-lauseen todistus Steinin menetelmällä

Joonas Havukainen

9. joulukuuta 2017

Sisältö

1 Johdanto

Tämän tutkielman tarkoituksena on perehdyttää lukija Steinin menetelmään normaaliapproksimaatiolle, joka on todennäköisyysteorian piiriin kuuluva nykyaikainen ja tehokas tapa arvioida kahden eri todennäköisyysjakauman etäisyyttä. Alkutilanne voi siis olla esimerkiksi summa riippuvaisia satunnaismuuttujia, joiden jakaumaa ei tunneta tarkasti. Tämän summan jakaumaa voidaan Steinin menetelmän avulla verrata normaalijakaumaa vasten ja sen tiheyden supetessa kohti normaalijakauman tiheyttä, kun n kasvaa rajatta, voidaan sen olettaa olevan likimain normaalijakautunut. Näin ollen käytännön laskuissa voidaan käyttää normaalijakauman ominaisuuksia, mikä tekee laskuista mullistavan paljon helpompia.

Tutkielma aloitetaan esittämällä joitakin yleisiä todennäköisyysteoreettisia oletuksia, jotka ovat voimassa koko tutkielman läpi, ellei toisin mainita. Lisäksi esitetään, sekä tarpeen tullen todistetaan, muutamia aputuloksia, joiden avulla jatkossa todistetaan menetelmään kuuluvia tärkeitä lauseita.

Seuraavaksi käsitellään todennäköisyysjakaumien etäisyyden käsite. Tätä käsitettä laajennetaan koskemaan kolmea erikoistapausta. Ne ovat Total variation-etäisyys, Kolmogorov-etäisyys sekä Wasserstein-etäisyys. Erikoistapaukset tietyllä tapaa täsmentävät ja kohdistavat paremmin jakaumien etäisyyttä tarkasteltavaa tilannetta varten. Kolmogorov-etäisyyttä tullaan jatkossa hyödyntämään niin kutsutussa Berry-Esseen-epäyhtälössä.

Luvussa 4 aloitetaan itse asia, eli Steinin menetelmän käsitteleminen. Perusajatuksena on aluksi valita tarkoituksenmukainen etäisyysmitta, jonka ominaisuuksia sovelletaan etäisyyden arvioimiseen. Seuraavaksi käytetään hyväksi Steinin lemmaa, joka karakterisoi normaalijakauman Steinin operaattorin avulla siten, että operaattorin arvon ollessa nolla, on tarkasteltava jakauma normaali. Steinin lemmasta herää kysymys, voidaanko asymptoottinen normaalijakauma päätellä myös silloin, kun Steinin operaattorin arvo on lähellä nollaa. Steinin yhtälön avulla voidaan päätellä, että näin todella on. Yhtälön ratkaisun avulla saadaankin niin kutsutut Steinin rajoitukset, jotka ovat Steinin menetelmän keskeisimmässä osassa. Niiden avulla saadaan esitettyä jakaumien etäisyydelle ylärajat. Rajoitukset esitetään ja todistetaan kaikille kolmelle esitetylle etäisyysmitalle erikseen.

Luvussa 5 esitetään ja todistetaan Berry-Esseen-lause. Se on klassinen todennäköisyysteorian tulos, jonka ensimmäinen osa on keskeinen raja-arvolause. Kiinnostavampi osa on kuitenkin Berry-Esseen-raja, eli suppenemisnopeus keskeiselle raja-arvolauseelle. Tämän rajan todistukseen käytetään Steinin menetelmää, joka on siis myös tehokas todistusväline.

Lopuksi käsitellään vielä ylimalkaisesti sen enempää todistelematta Steinin menetelmää moniulotteisessa tapauksessa. Huomataan sen olevan hyvin paljon samankaltaista kuin yksiulotteisessa versiossa.

Päälähteenä tutkielmassa on käytetty Nourdinin ja Peccatin Normal Approximations With Malliavin Calculus [8].

2 Oletuksia ja aputuloksia

Tässä työssä oletetaan aina, ellei toisin mainita, että on olemassa todennäköisyysavaruus $(\Omega, \mathcal{F}, \mathbb{P})$, jossa pari $(\Omega, \mathcal{F})$ on mitallinen avaruus. Perusjoukkona on siis Ω, jonka eräs sigma-algebra on $\mathcal{F}$. Joukot $A_n \in \mathcal{F}$ ovat mitallisia joukkoja, eli tapahtumia. Lisäksi $\mathbb{P} : \mathcal{F} \to [0,1]$ on todennäköisyysmitta, jolle pätee $\mathbb{P}(\Omega) = 1$ sekä erillisille tapahtumille $A_n \in \mathcal{F}$ on voimassa $\mathbb{P}\left(\bigcup A_n\right) = \sum \mathbb{P}(A_n)$.

Lemma 2.1 (Fubinin lause)**.** *Olkoon $A = [a,b] \times [c,d] \in \mathbb{R}^2$ kaksiulotteisen avaruuden osajoukko ja lisäksi $f : A \to \mathbb{R}$ integroituva funktio. Tällöin pätee*

$$\int_A f(x,y)dxdy = \int_a^b \left(\int_c^d f(x,y)dy \right) dx = \int_c^d \left(\int_a^b f(x,y)dx \right) dy.$$

Lemman 2.1 mukaan siis kaksiulotteinen integraali on palautettavissa kahteen yksiulotteiseen integraaliin, mikä luonnollisesti on paljon helpompi laskea.

Fubinin lause on niin yleinen tulos todennäköisyysteoriassa, että sen todistus tässä sivuutetaan. Eräs todistus löytyy esimerkiksi Sottisen todennäköisyysteorian luentomonisteesta [13].

Seuraava lemma on hyvin yksinkertainen, mutta hyödyllinen, standardinormaalijakauman ominaisuuksiin liittyvä pieni aputulos, jota käytetään myöhemmin.

Lemma 2.2. *Olkoon $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ standardinormaalijakauman tiheysfunktio. Tällöin pätee*

$$\phi(x) = \int_x^\infty y\phi(y)dy.$$

Todistus.

$$\begin{aligned}
\int_x^\infty y\phi(y)dy &= \lim_{z\to\infty} \int_x^z \frac{1}{\sqrt{2\pi}} y e^{-y^2/2} dy \\
&= \frac{1}{\sqrt{2\pi}} \lim_{z\to\infty} \Big/_{x}^{z} -e^{-y^2/2} \\
&= \frac{1}{\sqrt{2\pi}} (0 + e^{-x^2/2}) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \phi(x).
\end{aligned}$$

□

Tutkielmassa käytetään indikaattorifunktiolle merkintää

$$1_A(w) = \begin{cases} 1 & \text{kun } w \in A, \\ 0 & \text{kun } w \notin A \end{cases}$$

Indikaattorifunktiolle tunnetusti pätee $\mathbb{E}(1_A(w)) = \mathbb{P}(w \in A)$.

Määritellään seuraavaksi niin kutsutut Borel-funktiot.

Määritelmä 2.1. Olkoon (X, τ) topologinen avaruus. Olkoon $\mathcal{F} \subset \mathcal{P}(X)$. Kokoelma

$$Bor(X) = \bigcap \mathcal{F},$$

jossa $\mathcal{F}$ on sigma-algebra, jolle pätee $\tau \subset \mathcal{F}$ (eli $\mathcal{F}$ sisältää joukon X avoimet joukot), on niin sanottu Borelin perhe joukolle X. Perheet alkiot ovat Borel-joukkoja. Nyt, jos otetaan joukko $A \subset X$, kutsutaan funktiota $f : A \to \mathbb{R}$ Borel-funktioksi jos, ja vain jos kaikkien avointen joukkojen alkukuvat funktiolle f ovat Borel-joukkoja.

Määritelmä 2.2. Satunnaismuuttuja X on Beta-jakautunut, jos sillä on tiheysfunktio

$$f_X(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

jossa B on beta-funktio, joka määritellään kaavalla

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Tässä $\Gamma(t)$ on tunnetusti gamma-funktio, joka määräytyy kaavalla

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx.$$

Gamma-funktiolle pätee

$$\Gamma(\beta+1) = \beta\Gamma(\beta) \qquad \textit{sekä} \qquad \Gamma(1) = 1,$$

jolloin erityisesti, kun $\alpha = 1$, pätee

$$\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \frac{\Gamma(\beta)}{\Gamma(\beta+1)} = \frac{1}{\beta}. \tag{2.1}$$

2.1 Aputulokset Wasserstein-rajoitukselle

Seuraavat tulokset ovat kirjasta [8]. Todistukset sekä yksityiskohtaisemmat esitykset jätetään väliin. Ne auttavat todistamaan Wasserstein-rajoitusta koskevan Lemman 4.2.

Lemma 2.3. *[8] Olkoon $f : \mathbb{R} \to \mathbb{R}$ kaikkialla jatkuva funktio, jolle pätee $\int_{\mathbb{R}} |f'| d\phi < \infty$. Tällöin pätee myös, että $\int_{\mathbb{R}} |xf| d\phi < \infty$ sekä lisäksi*

$$\int_{\mathbb{R}} xf(x)\phi(x)dx = \int_{\mathbb{R}} f'(x)\phi(x)dx, \tag{2.2}$$

jossa siis ϕ on standardinormaalijakauman tiheysfunktio.

Olkoon $\mathcal{S}$ joukko funktioita $f : \mathbb{R} \to \mathbb{R}$, jotka ovat äärettömän monta kertaa jatkuvasti derivoituvia. Lisäksi funktiot f sekä kaikki sen derivaatat ovat korkeintaan polynomisesti kasvavia.

Määritelmä 2.3. [8] Ornstein-Uhlenbeck puoliryhmä määritellään kaavalla

$$P_t f(x) = \int_{\mathbb{R}} f(e^{-t}x + \sqrt{1 - e^{-2t}}y)\phi(y)dy, \tag{2.3}$$

kun $f \in \mathcal{S}$ sekä $t \geq 0$.

Nähdään, että erityisesti

$$P_0 f(x) = \int_{\mathbb{R}} f(x)\phi(y)dy = f(x) \int_{\mathbb{R}} \phi(y)dy = f(x).$$

Lemma 2.4. *[8] Kaikille funktioille $f \in \mathcal{S}$ pätee*

$$Lf(x) = -xf'(x) + f''(x)$$

jossa $L = P_t'(0)$.

3 Etäisyysmittoja

Tässä luvussa esitellään kolme ominaisuuksiltaan erilaista mittaa jakaumien etäisyydelle. Käytön vaikeusasteessa on eroja, jolloin luonnollisesti saadut rajoitukset paranevat suhteessa vaikeuden kasvaessa. Steinin menetelmällä saadaan käyttökelpoisia rajoituksia kaikille näistä etäisyysmitoista.

Tunnetusti mille tahansa joukolle $\mathcal{H}$ reaalilukuarvoisia funktioita, etäisyys $d_{\mathcal{H}}(F, G)$ kahden reaalilukuarvoisen satunnaismuuttujan F ja G välillä on

$$d_{\mathcal{H}}(F, G) = \sup_{h \in \mathcal{H}} \{ |\mathbb{E}[h(F)] - \mathbb{E}[h(G)]| \}.$$

Seuraavat etäisyysmitan erikoistapaukset koskevat tilanteita, joissa funktiojoukko $\mathcal{H}$ on tarkasteltavaan tilanteeseen erityisesti sopiva. Nämä kolme tapausta ovat yleisimmin käytettyjä mittoja todennäköisyysteoriassa.

3.1 Total variation-etäisyys

Aloitetaan Total variation-etäisyydellä. Se on suurin mahdollinen etäisyys kahden todennäköisyysjakauman välillä samalle tapahtumalle. Se määritellään kaavalla

$$\mathrm{d}_{TV}(F, G) = \sup_{h \in \mathcal{H}} \left| \int h dF - \int h dG \right| = \sup_{A \subseteq Z_+} |\big(F(A) - G(A)\big)|,$$

jossa $\mathcal{H} = \{1_A : A \ mitallinen\}$. Total variation-etäisyys on hyvin yleinen tilastollinen etäisyysmitta. Sen suhteellinen helppokäyttöisyys tekee siitä mukavan operoida. Sitä käytetään yleensä äärellisten jakaumien kanssa, sillä se voidaan määritellä myös kaavalla

$$\mathrm{d}_{TV}(F, G) = \frac{1}{2} \sum_{k=0}^{\infty} |F(k) - G(k)|.$$

Huomautus. Steinin menetelmästä on olemassa versioita muihinkin jakaumiin kuin vain normaalijakaumaan. Normaalijakauman ohella tunnetuin näistä on varmaankin Poisson-jakauman tapaus, jossa varsinkin käytetään Total variation-etäisyyttä Poisson-jakauman diskreettisyyden vuoksi. Ks. lähde [3].

3.2 Kolmogorov-etäisyys

Kolmogorov-etäisyys on Berry-Esseen-lauseen kannalta tärkein näistä mitoista, sillä Berry-Esseen-suppenemisnopeus keskeiselle raja-arvolauseelle on määritelty Kolmogorov-etäisyyden avulla. Kolmogorov-etäisyys määritellään kaavalla

$$\mathrm{d}_K(F,G) = \sup_{h\in\mathcal{H}} \left| \int hdF - \int hdG \right| = \sup_{z\in\mathbb{R}} |F(-\infty,z] - G(-\infty,z]|,$$

jossa $\mathcal{H} = \{1_{(-\infty,z]} : z \in \mathbb{R}\}$.

3.3 Wasserstein-etäisyys

Wasserstein-etäisyyden käsitteleminen lienee näistä erikoistapauksista hankalinta. Sitä varten on määriteltävä niin sanottu Lipschitz-jatkuva funktio. Lipschitz-jatkuvuus on tietyllä tapaa vahvempi ja rajoittavampi ehto kuin pelkkä jatkuvuus.

Määritelmä 3.1. Funktio $f : X \to Y$ on Lipschitz-funktio, jos on olemassa luku $L > 0$, jolle pätee

$$|f(y) - f(x)| \leq L|y - x|.$$

Pienintä epäyhtälön toteuttavaa lukua L kutsutaan funktion f Lipschitz-vakioksi.

Nyt, Wasserstein-etäisyys määritellään kaavalla

$$\mathrm{d}_W(F,G) = \sup_{h\in\mathcal{H}} \left| \int hdF - \int hdG \right|,$$

jossa $\mathcal{H} = \{h : \mathbb{R} \to \mathbb{R} : |h(y) - h(x)| \leq |y - x|\}$. Joukkona $\mathcal{H}$ on siis kaikki Lipschitz-funktiot h Lipschitz-vakiolla 1. Käytetään tällaiselle joukolle merkintää Lip(1). Wasserstein-etäisyydelle saadaan näin esitys [11]

$$\mathrm{d}_W(F,G) = \sup_{h\in Lip(1)} |\mathbb{E}h(Y) - \mathbb{E}h(X)| = \inf |Y - X|,$$

jossa satunnaismuuttujat X ja Y ovat sellaisia, että kaikilla $h \in Lip(1)$ on voimassa $h(X) = F$ ja $h(Y) = G$.

Näille kolmelle mitalle tullaan siis myöhemmin johtamaan käyttökelpoiset rajoitukset jakaumien etäisyydelle. Vaikka Kolmogorov-etäisyys onkin tätä tutkielmaa varten keskeisimmässä roolissa, käydään ne kuitenkin läpi kaikissa näissä tapauksissa, sillä ne ovat todennäköisyysteoriassa hyödyllisimmät ja eniten käytetyt.

4 Steinin menetelmä

Tässä luvussa esitetään Steinin menetelmä normaalijakaumalle. Se on tapa tuottaa täsmällisiä estimaatteja kahden todennäköisyysjakauman etäisyydelle. Menetelmä siis antaa tälle etäisyydelle täsmällisen ylärajan. Etäisyydelle käytettävä mitta on tässä työssä aina joku edellisessä luvussa mainituista mitoista.

Menetelmän rakenne on seuraavanlainen. Aluksi tarvitsemme niin kutsutun Steinin operaattorin $\mathcal{A}$, jonka avulla pystytään karakterisoimaan kohdejakaumamme eli normaalijakauma. Steinin lemma antaa meille tämän operaattorin. Osoittautuu, että operaattori on muotoa $\mathcal{A} = f'(x) - xf(x)$. Seuraavaksi esitellään Steinin yhtälö, jonka ratkaiseminen osoittautuu helpoksi, sillä se tiedetään jo valmiiksi. Tästä ratkaisusta saamme edellisen luvun mittojen avulla rajoituksen jakaumien etäisyydelle.

4.1 Steinin lemma

Esitetään aluksi normaalijakauman karakterisoiva Steinin operaattori $\mathcal{A}$. Tämä operaattori ei suinkaan ole ainoa normaalijakauman karakterisoiva operaattori. Niitä voi olla olemassa hyvinkin paljon, jolloin voi joskus olla vaikeaa tehdä päätös, mitä niistä käyttää. Kuitenkin monille jakaumille on olemassa "paras"tällainen operaattori, ja on osoittautunut, että nyt tarkasteltava $\mathcal{A}$ on normaalijakaumalle paras mahdollinen.

Perusajatus on, että reaalilukuarvoinen satunnaismuuttuja W on standardinormaalijakautunut jos, ja vain jos, operaattorille $\mathcal{A}$ pätee kaikilla f

$$\mathbb{E}(\mathcal{A}_f(W)) = 0.$$

Tämä ei vielä ole käyttökelpoinen kaava, sillä tarvitsemme operaattorille tarkan esityksen. Seuraava lemma osoittaa, että Steinin operaattori on muotoa $\mathcal{A}_f(x) = f'(x) - xf(x)$.

Lemma 4.1 (Steinin lemma)**.** *Olkoon W reaalilukuarvoinen satunnaismuuttuja. W on standardinormaalijakautunut jos, ja vain jos, kaikilla f pätee*

$$\mathbb{E}(f'(W)) = \mathbb{E}(Wf(W)).$$

Todistus. [1] Tunnetusti, jos satunnaismuuttuja W on standardinormaalijakautunut, pätee Lemman 2.2 nojalla kaikille f

$$\begin{aligned}
\mathbb{E}(f'(W)) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f'(w)e^{-w^2/2}dw \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} f'(w)\left(\int_{-\infty}^{w}(-x)e^{-x^2/2}dx\right)dw \\
&+ \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} f'(w)\left(\int_{w}^{\infty}xe^{-x^2/2}dx\right)dw.
\end{aligned}$$

Nyt, Lemman 2.1 (Fubinin lause) nojalla pätee

$$\begin{aligned}
\mathbb{E}(f'(W)) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}\left(\int_{x}^{0} f'(w)dw\right)(-x)e^{-x^2/2}dx \\
&+ \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty}\left(\int_{0}^{x} f'(w)dw\right)xe^{-x^2/2}dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}[f(x)-f(0)]xe^{-x^2/2}dx \\
&= \mathbb{E}(Wf(W)),
\end{aligned}$$

sillä $\{(x,w)\in\mathbb{R}^2 : x\geq 0, w\geq x\} = \{(x,w)\in\mathbb{R}^2 : w\geq 0, 0\leq x\leq w\}$.

Kiinnitetään nyt $z\in\mathbb{R}$ ja olkoon $f_z(w)$ ratkaisu differentiaaliyhtälölle

$$f'(w) - wf(w) = 1_{(-\infty,z]}(w) - \Phi(z), \tag{4.1}$$

jossa $\Phi(z)$ on siis standardinormaalijakauman kertymäfunktio. Huomataan, että (4.1) voidaan kirjoittaa muodossa

$$\frac{d}{dw}[e^{-w^2/2}f(w)] = e^{-w^2/2}(1_{(-\infty,z]}(w) - \Phi(z)), \tag{4.2}$$

sillä

$$\frac{d}{dw}[e^{-w^2/2}f(w)] = -we^{-w^2/2}f(w) + e^{-w^2/2}f'(w) = e^{-w^2/2}[f'(w) - wf(w)].$$

Siten $f_z(w)$ on muotoa

$$\begin{aligned}
f_z(w) &= e^{w^2/2}\int_{-\infty}^{w}[1_{(-\infty,z]}(x) - \Phi(z)]e^{-x^2/2}dx \\
&= -e^{w^2/2}\int_{w}^{\infty}[1_{(-\infty,z]}(x) - \Phi(z)]e^{-x^2/2}dx \\
&= \begin{cases} \sqrt{2\pi}e^{w^2/2}\Phi(w)[1-\Phi(z)] & \text{jos } w\leq z, \\ \sqrt{2\pi}e^{w^2/2}\Phi(z)[1-\Phi(w)] & \text{jos } w\geq z. \end{cases}
\end{aligned}$$

Huomataan, että $f_z(w)$ on rajoitettu, jatkuva sekä paloittain jatkuva, differentioituva funktio. Näin ollen pätee

$$0 = \mathbb{E}[f_z'(W) - Wf_z(W)] = \mathbb{E}[1_{(-\infty,z]}(w) - \Phi(z)] = \mathbb{P}(W \leq z) - \Phi(z),$$

joten W on standardinormaalijakautunut. $\square$

On osoitettu, että yhtälöstä $\mathbb{E}(f'(W)) = \mathbb{E}(Wf(W))$ seuraa satunnaismuuttujan W standardinormaalijakautuneisuus. Sen päteminen yhtälönä ei kuitenkaan ole riittävän mielenkiintoista, sillä monessa käytännön tapauksessa näin ei tule tapahtumaan. Kysymys kuuluukin, voidaanko päätellä satunnaismuuttujan W olevan likimain normaalijakautunut, jos $\mathbb{E}(f'(W)) - \mathbb{E}(Wf(W))$ on likimain 0. Suppeneminen kohti nollaa on kaikista paras vaihtoehto, sillä tällöin tarkasteltava jakauma myös suppenee täysin kohti normaalijakaumaa. Useimmissa tapauksissa näin todella voidaan päätellä, joka erityisesti motivoi käyttämään Steinin menetelmää. Osoitusta varten on esiteltävä Steinin yhtälö, jonka ratkaisusta saadaan Steinin rajoitukset.

4.2 Steinin yhtälö

Määritelmä 4.1. [8] Olkoon $W \sim N(0,1)$ standardinormaalijakautunut. Olkoon lisäksi $h : \mathbb{R} \to \mathbb{R}$ Borel-funktio siten, että $\mathbb{E}|h(W)| < \infty$. Steinin yhtälö on tällöin tavallinen ensimmäisen kertaluvun differentiaaliyhtälö

$$f'(x) - xf(x) = h(x) - \mathbb{E}[h(W)]. \tag{4.3}$$

Yhtälön ratkaisu saattaa vaikuttaa hankalalta, mutta sitä ei kannata säikähtää. Osoittautuu, että käytännön tilanteissa ratkaisua ei tarvitse tehdä, sillä seuraava lause ratkaisee sen meille valmiiksi.

Lause 4.1. *[8] Kaikki Steinin yhtälön (4.3) ratkaisut ovat muotoa*

$$f(x) = ce^{x^2/2} + e^{x^2/2}\int_{-\infty}^{x}\{h(y) - \mathbb{E}[h(W)]\}e^{-y^2/2}dy, \qquad x \in \mathbb{R}, \tag{4.4}$$

jossa $c \in \mathbb{R}$. Erityisesti,

$$f_h(x) := e^{x^2/2}\int_{-\infty}^{x}\{h(y) - \mathbb{E}[h(W)]\}e^{-y^2/2}dy \tag{4.5}$$

on yksikäsitteinen ratkaisu yhtälölle (4.3) toteuttaen ehdon

$$\lim_{x\to\pm\infty} e^{-x^2/2}f(x) = 0.$$

Todistus. [8] Kuten kohdassa (4.2), voidaan yhtälö (4.3) kirjoittaa muodossa

$$e^{x^2/2}\frac{d}{dx}[e^{-x^2/2}f(x)] = h(x) - \mathbb{E}[h(W)].$$

Tämä saadaan muotoon

$$e^{-x^2/2}f(x) = \int_{-\infty}^{x}\{h(y) - \mathbb{E}[h(W)]\}e^{-y^2/2}dy + c,$$

jossa $c \in \mathbb{R}$. On siis selvää, että kaikki ratkaisut yhtälölle (4.3) ovat muotoa (4.4).

Tarkastellaan esitystä $e^{-x^2/2}f(x)$, jossa $f(x)$ on kuten kohdassa (4.4). Saadaan

$$\begin{aligned} e^{-x^2/2}f(x) &= c + \int_{-\infty}^{x}\{h(y) - \mathbb{E}[h(W)]\}e^{-y^2/2}dy \\ &= c + \int_{-\infty}^{x}\{h(y)e^{-y^2/2} - \mathbb{E}[h(W)]e^{-y^2/2}\}dy \\ &= c + \mathbb{E}[h(W)1_{(-\infty,x)}(W)] - \mathbb{E}[h(W)]\mathbb{P}(W < x). \end{aligned}$$

Nyt, kun $x \to \infty$, pätee $\mathbb{E}[h(W)1_{(-\infty,x)}(W)] \to \mathbb{E}[h(W)]$ sekä $\mathbb{E}[h(W)]\mathbb{P}(W < x) \to \mathbb{E}[h(W)]$. Lauseen ominaisuus $\lim_{x\to\pm\infty} e^{-x^2/2}f(x) = 0$ toteutuu nyt siis jos, ja vain jos, $c = 0$. $\square$

Tarkastellaan lisää Steinin yhtälöä (4.3). Oletetaan nyt, että kaikille funktioille $h : \mathbb{R} \to \mathbb{R}$ odotusarvot $\mathbb{E}|h(W)|$ ja $\mathbb{E}|h(X)|$ ovat äärellisiä. Nyt, funktion f_h ollessa yhtälön (4.3) ratkaisu sekä ottamalla odotusarvot puolittain, saadaan

$$\mathbb{E}[f_h'(X) - Xf_h(X)] = \mathbb{E}[h(X)] - \mathbb{E}[h(W)].$$

Tämä tarkoittaa sitä, että kun $\mathcal{H}$ on funktiojoukko, jolle odotusarvot $\mathbb{E}|h(W)|$ ja $\mathbb{E}|h(X)|$ ovat äärellisiä kaikilla $h \in \mathcal{H}$, on W:n ja X:n välinen etäisyys

$$\mathrm{d}_{\mathcal{H}}(W, X) = \sup_{h\in\mathcal{H}} |\mathbb{E}[f_h'(X) - Xf_h(X)]|.$$

Huomataan, että tämä etäisyys ei riipu standardinormaalijakaumasta, vaan ainoastaan jakaumasta, jota halutaan arvioida. Funktio f_h on kuitenkin vielä liian monimutkainen käytettäväksi suoraan, joten seuraavaksi johdetaan täsmällisiä rajoituksia funktioille f_h tietyillä funktiojoukoilla $\mathcal{H}$. Rajat johdetaan kolmelle etäisyysmitalle, jotka esiteltiin luvussa 3.

4.3 Steinin rajoitukset

4.3.1 Total variation

Merkinnällä $\|\|_\infty$ tarkoitetaan niin kutsuttua supremum normia eli esimerkiksi $\|f_h\|_\infty = \sup\{|f_h|\}$. Steinin rajoitus kaavan (4.5) funktiolle f_h Total-variation etäisyydellä saadaan seuraavasta lauseesta.

Lause 4.2. *[8] Olkoon $h : \mathbb{R} \to [0,1]$ Borel-funktio. Silloin kaavan (4.5) funktiolla f_h on ominaisuudet*

$$\|f_h\|_\infty \leq \sqrt{\frac{\pi}{2}} \qquad ja \qquad \|f_h'\|_\infty \leq 2. \tag{4.6}$$

Toisin sanoen, jos $W \sim N(0,1)$, pätee jokaiselle integroituvalle satunnaismuuttujalle X, että

$$d_{TV}(W,X) \leq \sup_{f_h \in \mathcal{F}_{TV}} |\mathbb{E}[f_h'(X)] - \mathbb{E}[Xf_h(X)]|,$$

jossa $\mathcal{F}_{TV}$ on niiden funktioiden f_h joukko, jotka toteuttavat ehdot (4.6).

Todistus. [8] Selvästi lauseen funktiolle h pätee $|h(x) - \mathbb{E}[h(W)]| \leq 1$ kaikilla $x \in \mathbb{R}$, sillä sekä $h(x)$ että $\mathbb{E}[h(W)]$ ovat pienempiä tai yhtä suuria kuin 1. Käyttämällä yhtälön (4.5) kaavaa funktiolle f_h todetaan, että

$$\begin{aligned} |f_h(x)| &\leq e^{x^2/2} \min\left(\int_{-\infty}^{x} e^{-y^2/2} dy, \int_{x}^{\infty} e^{-y^2/2} dy\right) \\ &= e^{x^2/2} \int_{|x|}^{\infty} e^{-y^2/2} dy \leq \sqrt{\frac{\pi}{2}}, \end{aligned}$$

jonka viimeinen kohta osoitetaan seuraavassa. Merkitään nyt funktiolla $s(x)$ edellisen epäyhtälön toiseksi viimeistä kohtaa, eli

$$s(x) = e^{x^2/2} \int_{|x|}^{\infty} e^{-y^2/2} dy.$$

Tutkitaan funktion $s(x)$ ääriarvoja. Nyt, kun $x > 0$, pätee Lemman 2.2 nojalla

$$\begin{aligned} s'(x) &= xe^{x^2/2} \int_{x}^{\infty} e^{-y^2/2} dy - 1 \\ &\leq e^{x^2/2} \int_{x}^{\infty} ye^{-y^2/2} dy - 1 = 0, \end{aligned}$$

jolloin siis $s(x)$ on välttämättä laskeva, kun $x > 0$. Vastaavasti, kun $x < 0$, pätee

$$s'(x) \geq 0,$$

jolloin $s(x)$ on kasvava, kun $x < 0$. Tästä seuraa väistämättä, että funktio $s(x)$ saavuttaa maksimiarvonsa, kun $x = 0$. Maksimiarvoksi määräytyy siis $s(0) = \sqrt{\frac{\pi}{2}}$, jonka nojalla lauseen ensimmäinen väite on siis todistettu.

Toista väitettä varten huomataan, että koska f_h on Steinin yhtälön (4.3) ratkaisu, pätee kaikilla $x \in \mathbb{R}$

$$\begin{aligned} f_h' &= [h(x) - \mathbb{E}[h(W)] + xe^{x^2/2} \int_{-\infty}^{x} \{h(y) - \mathbb{E}[h(W)]\} e^{-y^2/2} dy \\ &= [h(x) - \mathbb{E}[h(W)] - xe^{x^2/2} \int_{x}^{\infty} \{h(y) - \mathbb{E}[h(W)]\} e^{-y^2/2} dy. \end{aligned}$$

Ylläolevasta seuraa, että

$$|f_h'| \leq 1 + |x| e^{x^2/2} \int_{|x|}^{\infty} e^{-y^2/2} dy \leq 1 + e^{x^2/2} \int_{|x|}^{\infty} y e^{-y^2/2} dy = 2. \tag{4.7}$$

□

4.3.2 Kolmogorov

Seuraava lause on vastaava tulos Kolmogorov-etäisyydelle. Sitä varten muutetaan hieman merkintöjä, sillä funktio h on nyt muotoa $h = 1_{(-\infty,z]}$. Merkitäänkin siis $f_z = f_{1_{(-\infty,z]}}$ kaikilla $z \in \mathbb{R}$. Tällöin tunnetusti pätee $\mathbb{E}[h(W)] = \mathbb{E}[1_{(-\infty,z]}(W)] = \mathbb{P}(W \leq z) = \Phi(z)$. Sijoittamalla $h = 1_{(-\infty,z]}$ yhtälöön (4.5), saadaan

$$f_z(x) := \begin{cases} \sqrt{2\pi} e^{x^2/2} \Phi(x)[1 - \Phi(z)] & \text{jos } x \leq z, \\ \sqrt{2\pi} e^{x^2/2} \Phi(z)[1 - \Phi(x)] & \text{jos } x \geq z. \end{cases}$$

Lause 4.3. *[8] Olkoon $z \in \mathbb{R}$ reaaliluku. Funktio f_z toteuttaa ehdot*

$$\|f_z\|_\infty \leq \frac{\sqrt{2\pi}}{4} \qquad \text{ja} \qquad \|f_z'\|_\infty \leq 1. \tag{4.8}$$

Näin ollen, kun $W \sim N(0,1)$, pätee jokaiselle integroituvalle satunnaismuuttujalle X, että

$$d_{Kol}(W, X) \leq \sup_{f_z \in \mathcal{F}_{Kol}} |\mathbb{E}[f_z'(X)] - \mathbb{E}[X f_z(X)]|,$$

jossa $\mathcal{F}_{Kol}$ on niiden funktioiden f_z joukko, jotka toteuttavat ehdot (4.8).

Todistus. [1][15] Aluksi, koska

$$f_{-z}(-x) = \begin{cases} \sqrt{2\pi}e^{x^2/2}\Phi(-x)[1-\Phi(-z)] & \text{jos } -x \leq -z, \\ \sqrt{2\pi}e^{x^2/2}\Phi(-z)[1-\Phi(-x)] & \text{jos } -x \geq -z \end{cases}$$

on yhtäpitävää lausekkeen

$$f_{-z}(-x) = \begin{cases} \sqrt{2\pi}e^{x^2/2}[1-\Phi(x)][1-(1-\Phi(z))] & \text{jos } -x \leq -z, \\ \sqrt{2\pi}e^{x^2/2}[1-\Phi(z)][(1-1-\Phi(x))] & \text{jos } -x \geq -z \end{cases}$$

$$= \begin{cases} \sqrt{2\pi}e^{x^2/2}[1-\Phi(x)]\Phi(z) & \text{jos } x \geq z, \\ \sqrt{2\pi}e^{x^2/2}[1-\Phi(z)]\Phi(x) & \text{jos } x \leq z \end{cases} = f_z(x)$$

kanssa, voimme olettaa, että $z \geq 0$. Lisäksi käytämme tulosta, jonka mukaan kaikilla $x \geq 0$ pätee

$$\begin{aligned} 1-\Phi(x) &= \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-z^2/2}dz \\ &\leq \frac{1}{\sqrt{2\pi}}\int_x^\infty \frac{z}{x}e^{-z^2/2}dz \\ &= \frac{e^{-x^2/2}}{x\sqrt{2\pi}}. \end{aligned} \tag{4.9}$$

Vastaavasti nähdään, että kun $x \leq 0$, pätee tulos

$$\Phi(x) \leq \frac{e^{-x^2/2}}{|x|\sqrt{2\pi}}. \tag{4.10}$$

Nyt, väitteen $\|f_z'\|_\infty \leq 1$ todistus jaetaan kolmeen osaan. Ensin oletetaan, että $x \leq 0 \leq z$. Huomataan funktion $f_z(x)$ kaavasta, että $f_z(x)$ on kasvava, kun $x \leq z$, joten on oltava $f_z'(x) \geq 0$. Kun $x \leq 0$, pätee tuloksen (4.10) nojalla, että

$$-1 \leq \sqrt{2\pi}xe^{x^2/2}\Phi(x) \leq 0,$$

jolloin suoraan laskemalla saadaan

$$\begin{aligned} f_z'(x) &= [1-\Phi(z)](\sqrt{2\pi}xe^{x^2/2}\Phi(x) + \sqrt{2\pi}e^{x^2/2}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}) \\ &= [1-\Phi(z)](\sqrt{2\pi}xe^{x^2/2}\Phi(x)+1) \\ &\leq [1-\Phi(z)] \leq 1, \end{aligned}$$

sillä $0 \leq [1-\Phi(z)] \leq 1$. Nyt, koska $0 \leq f_z'(x) \leq 1$, on oltava $|f_z'(x)| \leq 1$.

Oletetaan nyt, että $0 \leq x \leq z$. Tällöin käyttämällä tulosta (4.9), saadaan

$$\begin{aligned} 0 \leq f_z'(x) &= [1-\Phi(z)](\sqrt{2\pi}xe^{x^2/2}\Phi(x)+1) \\ &= 1-\Phi(z)+[1-\Phi(z)]\sqrt{2\pi}xe^{x^2/2}\Phi(x) \\ &\leq 1-\Phi(z)+[1-\Phi(x)]\sqrt{2\pi}xe^{x^2/2}\Phi(z) \\ &\leq 1-\Phi(z)+\Phi(z)=1. \end{aligned}$$

Viimeisenä oletetaan, että $0 \leq z \leq x$. Tuloksen (4.9) avulla saadaan

$$\begin{aligned} f_z'(x) &= \Phi(z)[\sqrt{2\pi}xe^{x^2/2}-\sqrt{2\pi}xe^{x^2/2}\Phi(x)-1] \\ &= \Phi(z)[\sqrt{2\pi}xe^{x^2/2}(1-\Phi(x))-1] \leq 0, \end{aligned}$$

sekä lisäksi

$$0 \geq f_z'(x) \geq -\Phi(z) \geq -1.$$

Väite $\|f_z'\|_\infty \leq 1$ on nyt todistettu.

Nyt, koska $f_z(x)$ on kasvava, kun $x < z$ ja laskeva, kun $x > z$, saavuttaa se maksiminsa kohdassa $x = z$, jolloin saadaan

$$f_z(x) \leq f_z(z) = \sqrt{2\pi}e^{z^2/2}\Phi(z)[1-\Phi(z)] \leq \frac{\sqrt{2\pi}}{4}.$$

Viimeisen epäyhtälön yksityiskohtainen todistus löytyy esimerkiksi lähteestä [1] sivuilta 54-55. □

4.3.3 Wasserstein

Käsitellään vielä Wasserstein-etäisyys. Näytetään aluksi, että kohdan (4.5) funktio f_h on mahdollista esittää vaihtoehtoisella (hieman monimutkaisemmalla) tavalla, jonka avulla johdetaan tulokset $f_h \in \mathcal{C}^1$ sekä $\|f_h'\|_\infty \leq \sqrt{\frac{2}{\pi}}K$.

Lemma 4.2. *[8] Olkoon $h : \mathbb{R} \to \mathbb{R}$ Lipschitz-funktio vakiolla $K > 0$. Silloin kohdassa (4.5) esitetty funktio f_h voidaan esittää myös muodossa*

$$f_h(x) = -\int_0^\infty \frac{e^{-t}}{\sqrt{1-e^{-2t}}}\mathbb{E}[h(e^{-t}x+\sqrt{1-e^{-2t}}W)W]dt. \tag{4.11}$$

Lisäksi pätee tulokset $f_h \in \mathcal{C}^1$ sekä $\|f_h'\|_\infty \leq \sqrt{\frac{2}{\pi}}K$.

Todistus. Todistus esitetään lähteessä [8].

Väite $f_h \in \mathcal{C}^1$ seuraa suoraan funktion h Lipschitz-ominaisuuksista sekä esityksestä (4.5). Merkitään nyt

$$\widetilde{f}_h(x) = -\int_0^\infty \frac{e^{-t}}{\sqrt{1-e^{-2t}}}\mathbb{E}[h(e^{-t}x + \sqrt{1-e^{-2t}}W)W]dt,$$

eli sama kuin f_h esityksessä (4.11). Tästä saadaan suoraan laskemalla

$$\widetilde{f}_h'(x) = -\int_0^\infty \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}\mathbb{E}[h'(e^{-t}x + \sqrt{1-e^{-2t}}W)W]dt,$$

kun $x \in \mathbb{R}$. Näin ollen, dominoidun konvergenssin nojalla, saadaan

$$|\widetilde{f}_h'(x)| \leq K\mathbb{E}(|W|)\int_0^\infty \frac{e^{-2t}dt}{\sqrt{1-e^{-2t}}} = K\sqrt{\frac{2}{\pi}}\int_0^1 \frac{dv}{2\sqrt{1-v}} = \sqrt{\frac{2}{\pi}}K,$$

jossa integraalit $\int_0^\infty \frac{e^{-2t}}{\sqrt{1-e^{-2t}}}$ ja $\int_0^1 \frac{dv}{2\sqrt{1-v}}$ vastaavat muuttujanvaihdolla $v = e^{-2t}$. Tunnetusti pätee myös $\mathbb{E}(|W|) = \sqrt{\frac{2}{\pi}}$. Lisäksi $\int_0^1 \frac{dv}{\sqrt{1-v}}$ on niin kutsuttu beta-funktio parametreilla $\alpha = 1$ ja $\beta = 1/2$, jolloin se saa arvon $\int_0^1 \frac{dv}{\sqrt{1-v}} = 2$. (Ks. kaava (2.1).) Näin ollen siis $\int_0^1 \frac{dv}{2\sqrt{1-v}} = 1$.

Lopuksi on enää osoitettava, että $\widetilde{f}_h(x)$ on sama funktio kuin kohdan (4.5) funktio $f_h(x)$. Sitä varten määritellään funktio $\widetilde{F}_h(x) : \mathbb{R} \to \mathbb{R}$ esityksellä

$$\widetilde{F}_h(x) = \int_0^\infty \Big(\mathbb{E}[h(W) - h(e^{-t}x + \sqrt{1-e^{-2t}}W)]\Big)\, dt,$$

kun $x \in \mathbb{R}$. Koska h on Lipschitz-funktio, pätee sille

$$\begin{aligned} |h(W) - h(e^{-t}x + \sqrt{1-e^{-2t}}W)| &\leq K|W - e^{-t}x - \sqrt{1-e^{-2t}}W| \\ &= K|W|(1 - \sqrt{1-e^{-2t}} + Ke^{-t}|x| \\ &\leq K|W|e^{-2t} + Ke^{-t}|x|, \end{aligned}$$

jolloin $\widetilde{F}_h(x)$ on tosiaan integroituva. Lisäksi huomataan, että

$$\widetilde{F}_h'(x) = -\int_0^\infty e^{-t}\mathbb{E}[h'(e^{-t}x + \sqrt{1-e^{-2t}}W)]dt.$$

Nyt, Lemman 2.3 nojalla on selvää, että $\widetilde{F}_h'(x) = \widetilde{f}_h(x)$. Siten Lemman

2.4 nojalla voidaan kirjoittaa

$$\begin{aligned}
\widetilde{f}_h'(x) - x\widetilde{f}_h(x) &= L\widetilde{F}_h(x) \\
&= -\int_0^\infty LP_t h(x)dt \\
&= -\int_0^\infty \frac{d}{dt}P_t h(x)dt \\
&= P_0 h(x) - P_\infty h(x) = h(x) - \mathbb{E}[h(W)],
\end{aligned}$$

joten $\widetilde{f}_h$ toteuttaa ehdon (4.3). Vielä olisi todistettava, että $\widetilde{f}_h = f_h$, jossa f_h on siis kuten kohdassa (4.5). Huomataan, että

$$\begin{aligned}
|\widetilde{f}_h(x)| &\leq \int_0^\infty \frac{e^{-t}}{\sqrt{1-e^{-2t}}}|\mathbb{E}[h(e^{-t}x + \sqrt{1-e^{-2t}}W)W]|dt \\
&\leq |h(0)|\int_0^\infty \frac{e^{-t}}{\sqrt{1-e^{-2t}}}dt \\
&+ K\int_0^\infty \frac{e^{-t}}{\sqrt{1-e^{-2t}}}(e^{-t}|x|\mathbb{E}[|W|] + \sqrt{1-e^{-2t}}\mathbb{E}[W^2])dt \\
&= \alpha|x| + \beta,
\end{aligned}$$

joillekin äärellisille vakioille $\alpha, \beta \geq 0$, jotka riippuvat ainoastaan parametreista $h(0)$ sekä K. Näin ollen päätellään, että

$$\lim_{x\to\pm\infty} e^{-x^2/2}\widetilde{f}_h(x) = 0,$$

josta kohdan (4.5) nojalla päätellään, että $\widetilde{f}_h = f_h$. $\square$

Steinin rajoitukset Wasserstein-etäisyyden suhteen saadaan seuraavasti.

Lause 4.4. *[8] Satunnaismuuttujalle $W \sim N(0,1)$ sekä mille tahansa neliöintegroituvalle satunnaismuuttujalle X pätee*

$$d_W(X,W) \leq \sup_{f\in\mathcal{F}_W} |\mathbb{E}[f'(X)] - \mathbb{E}[Xf(X)]|, \tag{4.12}$$

jossa $\mathcal{F}_W := \{f : \mathbb{R} \to \mathbb{R} \in \mathcal{C}^1; \|f'\|_\infty \leq \sqrt{2/\pi}\}$.

Huomautus. Tässä ei pidä sekoittaa satunnaismuuttujaa W sekä Wasserstein-etäisyydelle käytettyä merkintää W.

Huomautus. Merkinnällä $f \in \mathcal{C}^1$ tarkoitetaan, että funktio f on ensinnäkin derivoituva sekä toiseksi tämä derivaatta on kaikkialla jatkuva.

Todistus. [8] Jo todistetusta tuloksesta $\|f'\|_\infty \leq \sqrt{2/\pi}$ seuraa suoraan Lipschitz-funktioiden määritelmästä, että $\mathcal{F}_W \subset Lip(\sqrt{2/\pi})$. Nyt, etäisyyksien määritelmästä seuraa

$$\begin{aligned} d_W(X,W) &= \sup_{h\in Lip(1)} |\mathbb{E}[h(W)] - \mathbb{E}[h(X)]| \\ &= \sup_{h\in Lip(1)} |\mathbb{E}[f_h'(X)] - \mathbb{E}[Xf_h(X)]| \\ &\leq \sup_{f\in\mathcal{F}_W} |\mathbb{E}[f'(X)] - \mathbb{E}[Xf(X)]|. \end{aligned}$$

□

Olemme siis johtaneet jakaumien etäisyydessä

$$\mathrm{d}_{\mathcal{H}}(W,X) = \sup_{h\in\mathcal{H}} |\mathbb{E}[f_h'(X) - Xf_h(X)]|$$

käytettäville funktioille f_h käyttökelpoisia ominaisuuksia kolmessa käsitellyssä erikoistapauksessa. Steinin menetelmässä on lopulta kyse vain juuri näiden ominaisuuksien hyödyntämisestä funktioille f_h. Myöhemmin luvussa 6 näytetään konkreettisesti, miten helposti Steinin menetelmällä muodostetaan rajoitukset etäisyydelle, ja samalla siis tutkitaan, onko haluttu jakauma likimain normaali vai ei. Steinin menetelmästä on kuitenkin myös muuta hyötyä. Sitä voidaan käyttää myös tehokkaana todistusvälineenä, kuten seuraavassa kappaleessa tullaan osoittamaan. Itse asiassa Stein kehitti menetelmänsä juuri keskeisen raja-arvolauseen todistamista varten.

5 Berry-Esseen-lause

Palautetaan mieleen yksi matematiikan klassisimmista tuloksista, keskeinen raja-arvolause. Sen mukaan siis keskiarvo riittävän suuresta määrästä toisistaan riippumattomia satunnaismuuttujia, joilla kaikilla on hyvin määritellyt odotusarvo ja varianssi, on likimain normaalisti jakautunut. Tulokseen ei vaikuta satunnaismuuttujien omat eksaktit jakaumat.

Lause 5.1 (Keskeinen raja-arvolause)**.** *Oletetaan, että $n \to \infty$. Silloin pätee*

$$\sqrt{n} \cdot \left(\frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right) \longrightarrow^{d} N(0, \sigma^2).$$

Lauseelle on olemassa useita erilaisia todistuksia. Eräs karakteristisiin funktioihin nojaava todistus löytyy esimerkiksi Sottisen todennäköisyysteorian luentomonisteesta [13].

Steinin menetelmä antaa kuitenkin vaihtoehtoisen tavan todistaa tämä klassinen tulos Kolmogorov-etäisyysmitan avulla. Erityisesti on löydettävissä rajoitus suppenemisnopeudelle. Tämä rajoitus on keskeisessä osassa Berry-Esseen-lausetta.

Lause 5.2 (Berry-Esseen-lause)**.** *Olkoon $\{Y_k : k \geq 1\}$ jono riippumattomia, samoin jakautuneita satunnaismuuttujia siten, että $\mathbb{E}(Y_1) = 0$ ja $\mathbb{E}(Y_1^2) = 1$, jolloin siis $Var(Y_1) = 1$. Olkoon lisäksi*

$$V_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Y_k, \qquad n \geq 1. \tag{5.1}$$

Tällöin pätee Berry-Esseen-epäyhtälö

$$d_{Kol}(V_n, W) \leq \frac{C\mathbb{E}[|Y_1|^3]}{\sqrt{n}}, \qquad n \geq 1, \tag{5.2}$$

jossa $W \sim N(0,1)$ ja vakio $C > 0$ on riippumaton sekä n:stä että Y:stä.

Erityisesti, kun $n \to \infty$, seuraa Berry-Esseen-epäyhtälöstä suoraan keskeinen raja-arvolause

$$V_n \longrightarrow^{d} W \sim N(0,1), \qquad n \to \infty. \tag{5.3}$$

Berry-Esseen-lause on tämän tutkielman päätulos, jonka todistus käydään läpi yksityiskohtaisesti. Todistuksessa keskitytään Berry-Esseen-epäyhtälöön, sillä keskeisen raja-arvolauseen seuraaminen siitä on ilmeistä. Todistus löytyy myös lähteestä [8].

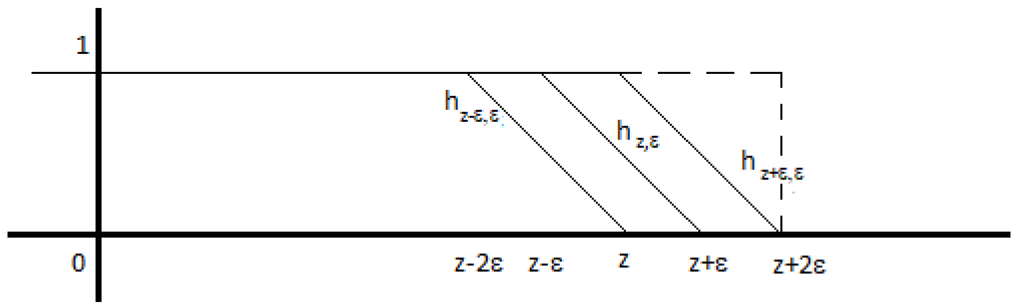
Todistus. Olkoon $n \geq 2$ ja $C_n > 0$ paras mahdollinen vakio, joka toteuttaa ehdon

$$d_{Kol}(V_n, W) \leq \frac{C_n \mathbb{E}[|Y_1|^3]}{\sqrt{n}}, \tag{5.4}$$

jossa $Y_1, ..., Y_n$ ovat riippumattomia, samoin jakautuneita satunnaismuuttujia, joille pätee $\mathbb{E}[|Y_1|^3] < \infty$ sekä $Var(Y_1) = 1$ ja $\mathbb{E}[Y_1] = 0$. Muodostetaan aluksi karkea estimaatti vakiolle C_n. Koska $\mathbb{E}[Y_1^2] = Var(Y_1) - \mathbb{E}[Y_1]^2 = 1 - 0 = 1$, nähdään, että $\mathbb{E}[|Y_1|^3] \geq \mathbb{E}[Y_1^2]^{\frac{3}{2}} = 1$. Tällöin siis välttämättä pätee $C_n \leq \sqrt{n}$. Tämä arvio ei tietenkään ole missään tapauksessa tarpeeksi laadukas, sillä on osoitettava, että $C_n < \infty$, mutta se auttaa meitä jatkossa. Myöhemmin tullaankin osoittamaan, että $C_n \leq 33$.

Määritellään seuraavaksi funktio $h_{z,\epsilon}(x)$ kaikille $z \in \mathbb{R}$ ja $\epsilon > 0$ kaavalla

$$h_{z,\epsilon}(x) = \begin{cases} 1 & \text{jos } x \leq z - \epsilon, \\ \textit{lineaarinen} & \text{jos } z - \epsilon < x < z + \epsilon, \\ 0 & \text{jos } x \geq z + \epsilon. \end{cases}$$



Kuvassa siis funktioiden $h_{z,\epsilon}(x)$, $h_{z-\epsilon,\epsilon}(x)$ sekä $h_{z+\epsilon,\epsilon}(x)$ kuvaajat.

Kohtuullisin ponnistuksin päätellään kuvasta katsomalla, että

$$h_{z-\epsilon,\epsilon}(x) \leq 1_{\{x \leq z\}} \leq h_{z+\epsilon,\epsilon}(x) \leq 1_{\{x \leq z+2\epsilon\}} \leq h_{z-\epsilon,\epsilon}(x) + 1_{\{z-2\epsilon<x<z+2\epsilon\}}.$$

Sijoittamalla $x = W$ sekä ottamalla odotusarvot, saadaan

$$\mathbb{E}[h_{z+\epsilon,\epsilon}(W)] \leq \mathbb{E}[h_{z-\epsilon,\epsilon}(W)] + \mathbb{P}(z - 2\epsilon < W < z + 2\epsilon). \tag{5.5}$$

Nyt, koska standardinormaalijakaumalle pätee

$$\mathbb{P}(z-2\epsilon < W < z+2\epsilon) = \frac{1}{\sqrt{2\pi}} \int_{z-2\epsilon}^{z+2\epsilon} e^{-x^2/2} dx \leq \frac{1}{\sqrt{2\pi}} [z+2\epsilon-(z-2\epsilon)] = \frac{4\epsilon}{\sqrt{2\pi}},$$

muokkautuu kaava (5.5) muotoon

$$\mathbb{E}[h_{z+\epsilon,\epsilon}(W)] \leq \mathbb{E}[h_{z-\epsilon,\epsilon}(W)] + \frac{4\epsilon}{\sqrt{2\pi}}.$$

Näin ollen, kun $n \geq 2$ sekä $\epsilon > 0$, saadaan Kolmogorov-etäisyydelle esitys

$$d_{Kol}(V_n, W) \leq \sup_{z\in\mathbb{R}} |\mathbb{E}[h_{z,\epsilon}(V_n) - \mathbb{E}[h_{z,\epsilon}(W)]| + \frac{4\epsilon}{\sqrt{2\pi}}. \tag{5.6}$$

Seuraavaksi esitetään eräs rajoitus, joka osoitetaan hieman myöhemmin todeksi käyttäen Steinin menetelmää. Tällä hetkellä se vain oletetaan todeksi, kun $\epsilon > 0$.

$$\sup_{z\in\mathbb{R}} |\mathbb{E}[h_{z,\epsilon}(V_n) - \mathbb{E}[h_{z,\epsilon}(W)]| \leq \frac{6\mathbb{E}[|Y_1|^3]}{\sqrt{n}} + \frac{3C_{n-1}\mathbb{E}[|Y_1|^3]^2}{\epsilon n}. \tag{5.7}$$

Kohtien (5.6) ja (5.7) nojalla on siis selvää, että

$$d_{Kol}(V_n, W) \leq \frac{6\mathbb{E}[|Y_1|^3]}{\sqrt{n}} + \frac{3C_{n-1}\mathbb{E}[|Y_1|^3]^2}{\epsilon n} + \frac{4\epsilon}{\sqrt{2\pi}}.$$

Valitsemalla $\epsilon = \sqrt{\frac{C_{n-1}}{n}}\mathbb{E}[|Y_1|^3]$, saadaan

$$d_{Kol}(V_n, W) \leq \frac{\mathbb{E}[|Y_1|^3]}{\sqrt{n}} \left[6 + \left(3 + \frac{4}{\sqrt{2\pi}}\right)\sqrt{C_{n-1}}\right],$$

sillä tällöin

$$\begin{aligned}
\frac{6\mathbb{E}[|Y_1|^3]}{\sqrt{n}} + \frac{3C_{n-1}\mathbb{E}[|Y_1|^3]^2}{\epsilon n} + \frac{4\epsilon}{\sqrt{2\pi}} &= \frac{6\mathbb{E}|Y_1|^3}{\sqrt{n}} + \frac{3C_{n-1}\mathbb{E}(|Y_1|^3)^2}{\sqrt{\frac{C_{n-1}}{n}}\mathbb{E}|Y_1|^3 n} + \frac{4\sqrt{\frac{C_{n-1}}{n}}\mathbb{E}|Y_1|^3}{\sqrt{2\pi}} \\
&= \frac{6\mathbb{E}|Y_1|^3}{\sqrt{n}} + 3\frac{C_{n-1}}{\sqrt{C_{n-1}}}\mathbb{E}|Y_1|^3\frac{\sqrt{n}}{n} + \frac{4}{\sqrt{2\pi}}\sqrt{C_{n-1}}\frac{\mathbb{E}|Y_1|^3}{\sqrt{n}} \\
&= \frac{\mathbb{E}|Y_1|^3}{\sqrt{n}}\left[6 + 3\sqrt{C_{n-1}} + \frac{4\sqrt{C_{n-1}}}{\sqrt{2\pi}}\right].
\end{aligned}$$

Näin ollen pätee $C_n \leq 6 + \left(3 + \frac{4}{\sqrt{2\pi}}\right)\sqrt{C_{n-1}}$. Todistetaan seuraavaksi induktion avulla, että $C_n \leq 33$, josta Berry-Esseen-epäyhtälö seuraa. Muistetaan alusta, että $C_n \leq \sqrt{n}$. Osoitetaan ensin perusaskel. Koska $n \geq 2$, riittää näyttää, että $C_2 \leq \sqrt{2} \approx 1.4 \leq 33$. Seuraavaksi otetaan induktioaskel. Oletetaan, että $C_n \leq 33$. Silloin

$$C_{n+1} \leq 6 + \left(3 + \frac{4}{\sqrt{2\pi}}\right)\sqrt{C_n} \leq 6 + \left(3 + \frac{4}{\sqrt{2\pi}}\right)\sqrt{33} \approx 32.4 \leq 33.$$

Johtopäätös on siis, että $C_n \leq 33$. Tämä ei luonnollisestikaan ole optimaalisin arvo lauseen vakiolle C. Se on kuitenkin riittävä Berry-Esseen-lauseen todistamiseksi, sillä nyt on osoitettu, että vakio $C < \infty$ riippumatta parametreista n ja Y.

Tämä johtopäätös saatiin siis olettamalla, että (5.7) pätee. Seuraavaksi osoitetaankin se todeksi. Sitä varten kiinnitetään $z \in \mathbb{R}$ sekä $\epsilon > 0$. Olkoon $f = f_{z,\epsilon}$ Steinin yhtälön (4.3) ratkaisu, kun $h = h_{z,\epsilon}$. Muistetaan Lauseesta 4.2, että $\|f\|_\infty \leq \sqrt{\pi/2}$ ja $\|f'\|_\infty \leq 2$. Nyt pätee

$$|xf(x)-yf(y)| = |f(x)(x-y)+(f(x)-f(y))y| \leq \left(\sqrt{\frac{\pi}{2}} + 2|y|\right)|x-y|. \quad (5.8)$$

Määritellään seuraavaksi satunnaismuuttuja $V_n^i \perp Y_i$ kaavalla

$$V_n^i = V_n - \frac{Y_i}{\sqrt{n}}, \qquad i = 1, ..., n.$$

Näin ollen, saadaan

$$\begin{aligned}
\mathbb{E}[h(V_n)] - \mathbb{E}[h(W)] &= \mathbb{E}[f'(V_n) - V_n f(V_n)] \\
&= \sum_{i=1}^n \mathbb{E}\left[f'(V_n)\frac{1}{n} - f(V_n)\frac{Y_i}{\sqrt{n}}\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[f'(V_n)\frac{1}{n} - (f(V_n) - f(V_n^i))\frac{Y_i}{\sqrt{n}}\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[f'(V_n)\frac{1}{n} - f'\left(V_n^i + \theta\frac{Y_i}{\sqrt{n}}\right)\frac{Y_i^2}{n}\right],
\end{aligned}$$

jossa satunnaismuuttuja θ on jakautunut tasaisesti välillä $[0, 1]$ sekä riippumaton satunnaismuuttujista Y_i. Viimeinen yhtälö seuraa soveltamalla tietoa, että

$$f(b)-f(a) = \int_a^b f'(u)du = (b-a)\int_0^1 f'(a+(b-a)v)dv = (b-a)\mathbb{E}(f'(a+\theta(b-a))).$$

Muistetaan Steinin yhtälöstä (4.3), että $f'(x) = xf(x) + h(x) - \mathbb{E}[h(W)]$, jolloin

$$\mathbb{E}[h(V_n)] - \mathbb{E}[h(W)] = \sum_{i=1}^n \left(a_i(xf) - b_i(xf) + a_i(h) - b_i(h)\right), \qquad (5.9)$$

jossa

$$\begin{aligned}
a_i(g) &= \mathbb{E}[g(V_n) - g(V_n^i)]\frac{1}{n} \\
b_i(g) &= \mathbb{E}\left[\left(g\left(V_n^i + \theta\frac{Y_i}{\sqrt{n}}\right) - g(V_n^i)\right) Y_i^2\right]\frac{1}{n}.
\end{aligned}$$

Olemme siis tilanteessa, jossa meidän on rajoitettava nämä kaavan (5.9) neljä termiä. Ensimmäistä termiä $a_i(xf)$ varten muistetaan, että $Y_i \sim Y_1$. Lisäksi pätee $\mathbb{E}[|Y_1|] \leq \mathbb{E}[Y_1^2]^{\frac{1}{2}} = 1$ sekä $\mathbb{E}[|V_n^i|] \leq \mathbb{E}[(V_n^i)^2]^{\frac{1}{2}} \leq 1$. Käyttämällä tulosta (5.8), saadaan

$$\begin{aligned}
|a_i(xf)| = |\mathbb{E}[xf(V_n) - xf(V_n^i)]\frac{1}{n}| &\leq \frac{1}{n}\left(\sqrt{\frac{\pi}{2}} + 2\mathbb{E}|V_n^i|\right)\mathbb{E}|V_n - V_n^i| \\
&\leq \frac{1}{n}\left(\sqrt{\frac{\pi}{2}} + 2\right)\mathbb{E}|Y_1|\frac{1}{\sqrt{n}} \\
&\leq \frac{1}{n\sqrt{n}}\left(\sqrt{\frac{\pi}{2}} + 2\right).
\end{aligned}$$

Toista termiä $b_i(xf)$ varten muistetaan, että $\mathbb{E}[\theta] = 1/2$. Nyt vastaavasti pätee

$$\begin{aligned}
|b_i(xf)| &\leq \frac{1}{n\sqrt{n}}\left(\mathbb{E}[\theta]\mathbb{E}[|Y_1|^3]\sqrt{\frac{\pi}{2}} + 2\mathbb{E}[\theta]\mathbb{E}[|Y_1|^3]\mathbb{E}[|V_n^i|]\right) \\
&\leq \left(\frac{1}{2}\sqrt{\frac{\pi}{2}} + 1\right)\frac{\mathbb{E}[|Y_1|^3]}{n\sqrt{n}}.
\end{aligned}$$

Kolmannen termin $a_i(h)$ kohdalla saadaan funktion h määritelmän nojalla

$$\begin{aligned}
h(y) - h(x) &= (y - x)\int_0^1 h'(x + s(y - x))ds \\
&= -\frac{y - x}{2\epsilon}\mathbb{E}[1_{[z-\epsilon,z+\epsilon]}(x + \widehat{\theta}(y - x))],
\end{aligned}$$

jossa $\widehat{\theta}$ on tasaisesti jakautunut välillä $[0,1]$ ja riippumaton sekä satunnaismuuttujista θ että Y_i. Näin ollen päätellään, että

$$\begin{aligned}
|a_i(h)| &\leq \frac{1}{2\epsilon n\sqrt{n}}\mathbb{E}\left[|Y_i|1_{z-\epsilon,z+\epsilon]}\left(V_n^i + \widehat{\theta}\frac{Y_i}{\sqrt{n}}\right)\right] \\
&= \frac{1}{2\epsilon n\sqrt{n}}\mathbb{E}\left[|Y_i|\mathbb{P}\left(z - t\frac{y}{\sqrt{n}} - \epsilon \leq V_n^i \leq z - t\frac{y}{\sqrt{n}} + \epsilon\right)|_{t=\widehat{\theta},y=Y_i}\right] \\
&\leq \frac{1}{2\epsilon n\sqrt{n}}\sup_{t\in[0,1]}\sup_{y\in\mathbb{R}}\mathbb{P}\left(z - t\frac{y}{\sqrt{n}} - \epsilon \leq V_n^i \leq z - t\frac{y}{\sqrt{n}} + \epsilon\right).
\end{aligned}$$

On siis rajoitettava $\mathbb{P}(a \leq V_n^i \leq b)$ kaikille $a, b \in \mathbb{R}$, joille pätee $a \leq b$. Sitä varten asetetaan $\widetilde{V}_n^i = \frac{1}{\sqrt{n-1}} \sum_{j \neq i} Y_j$, jolloin siis $V_n^i = \sqrt{1 - \frac{1}{n}} \widetilde{V}_n^i$. Käyttämällä epäyhtälöä (5.4) sekä tietoa, että standardinormaalijakauman tiheysfunktiolla on yläraja $1/\sqrt{2\pi}$, saadaan

$$\begin{aligned}
\mathbb{P}(a \leq V_n^i \leq b) &= \mathbb{P}\left(\frac{a}{\sqrt{1-\frac{1}{n}}} \leq \widetilde{V}_n^i \leq \frac{b}{\sqrt{1-\frac{1}{n}}}\right) \\
&= \mathbb{P}\left(\frac{a}{\sqrt{1-\frac{1}{n}}} \leq W \leq \frac{b}{\sqrt{1-\frac{1}{n}}}\right) + \mathbb{P}\left(\frac{a}{\sqrt{1-\frac{1}{n}}} \leq \widetilde{V}_n^i \leq \frac{b}{\sqrt{1-\frac{1}{n}}}\right) \\
&- \mathbb{P}\left(\frac{a}{\sqrt{1-\frac{1}{n}}} \leq W \leq \frac{b}{\sqrt{1-\frac{1}{n}}}\right) \leq \frac{b-a}{\sqrt{2\pi}\sqrt{1-\frac{1}{n}}} + \frac{2C_{n-1}\mathbb{E}[|Y_1|^3]}{\sqrt{n-1}}.
\end{aligned}$$

Tässä siis $a = z - t\frac{y}{\sqrt{n}} - \epsilon$ ja $b = z - t\frac{y}{\sqrt{n}} + \epsilon$, jolloin $b - a = 2\epsilon$. Näin ollen saadaan kolmannen termin rajoitukseksi

$$\begin{aligned}
|a_i(h)| &\leq \frac{1}{2\epsilon n\sqrt{n}}\left(\frac{2\epsilon}{\sqrt{2\pi}\sqrt{1-\frac{1}{n}}} + \frac{2C_{n-1}\mathbb{E}[|Y_1|^3]}{\sqrt{n-1}}\right) \\
&= \frac{1}{\sqrt{2\pi}n\sqrt{n-1}} + \frac{C_{n-1}\mathbb{E}[|Y_1|^3]}{n\sqrt{n}\sqrt{n-1}\epsilon}.
\end{aligned}$$

Neljännen termin $b_i(h)$ rajoitus saadaan vastaavasti kaavalla

$$\begin{aligned}
|b_i(h)| &= \frac{1}{2n\sqrt{n}\epsilon}\left|\mathbb{E}\left[Y_i^3\theta 1_{[z-\epsilon, z+\epsilon]}\left(V_n^i + \widehat{\theta}\theta\frac{Y_i}{\sqrt{n}}\right)\right]\right| \\
&\leq \frac{\mathbb{E}[|Y_1|^3]}{4n\sqrt{n}\epsilon}\sup_{t\in[0,1]}\sup_{y\in\mathbb{R}}\mathbb{P}\left(z - t\frac{y}{\sqrt{n}} - \epsilon \leq V_n^i \leq z - t\frac{y}{\sqrt{n}} + \epsilon\right) \\
&\leq \frac{\mathbb{E}[|Y_1|^3]}{2\sqrt{2\pi}n\sqrt{n-1}} + \frac{C_{n-1}(\mathbb{E}[|Y_1|^3])^2}{2n\sqrt{n}\sqrt{n-1}\epsilon}.
\end{aligned}$$

Ei ole täysin ilmiselvää, että väite seuraisi suoraan edellisten laskujen perusteella kaavojen ollessa hyvin sotkuisen näköisiä. Seuraava askel onkin sijoittaa saadut rajoitukset kaavaan (5.9). Suoraan sijoittamalla nähdään, että

$$
\begin{aligned}
|\mathbb{E}[h(V_n)] - \mathbb{E}[h(W)]| &= \sum_{i=1}^{n} \left(|a_i(xf)| + |b_i(xf)| + |a_i(h)| + |b_i(h)|\right) \\
&= n\left(|a_i(xf)| + |b_i(xf)| + |a_i(h)| + |b_i(h)|\right) \\
&= n \times [\left(\sqrt{\frac{\pi}{2}} + 2\right)\frac{1}{n\sqrt{n}} + \left(\frac{1}{2}\sqrt{\frac{\pi}{2}} + 1\right)\frac{\mathbb{E}|Y_1|^3}{n\sqrt{n}} + \frac{1}{\sqrt{2\pi}n\sqrt{n-1}} \\
&+ \frac{C_{n-1}\mathbb{E}|Y_1|^3}{n\sqrt{n}\sqrt{n-1}\epsilon} + \frac{\mathbb{E}|Y_1|^3}{2\sqrt{2\pi}n\sqrt{n-1}} + \frac{C_{n-1}\left(\mathbb{E}|Y_1|^3\right)^2}{2n\sqrt{n}\sqrt{n-1}\epsilon}] \\
&= \left(\sqrt{\frac{\pi}{2}} + 2\right)\frac{1}{\sqrt{n}} + \left(\frac{1}{2}\sqrt{\frac{\pi}{2}} + 1\right)\frac{\mathbb{E}|Y_1|^3}{\sqrt{n}} + \frac{1}{\sqrt{2\pi}\sqrt{n-1}} \\
&+ \frac{\mathbb{E}|Y_1|^3}{2\sqrt{2\pi}\sqrt{n-1}} + \frac{C_{n-1}\mathbb{E}|Y_1|^3}{\sqrt{n}\sqrt{n-1}\epsilon} + \frac{C_{n-1}\left(\mathbb{E}|Y_1|^3\right)^2}{2\sqrt{n}\sqrt{n-1}\epsilon}.
\end{aligned}
$$

Jaetaan tämä osiin niin, että ensimmäisen osan kattaa neljä ensimmäistä termiä ja toisen osan kattaa kaksi viimeistä termiä. Käytetään lisäksi hyödyksi tietoa, että $\mathbb{E}|Y_1|^3 \geq 1$ sekä $n-1 \geq n/2$, kun $n \geq 2$. Näin ollen voidaan kirjoittaa

$$
\begin{aligned}
\left(\sqrt{\frac{\pi}{2}} + 2\right)\frac{1}{\sqrt{n}} &+ \left(\frac{1}{2}\sqrt{\frac{\pi}{2}} + 1\right)\frac{\mathbb{E}|Y_1|^3}{\sqrt{n}} + \frac{1}{\sqrt{2\pi}\sqrt{n-1}} + \frac{\mathbb{E}|Y_1|^3}{2\sqrt{2\pi}\sqrt{n-1}} \\
&= \mathbb{E}|Y_1|^3[\frac{\sqrt{\frac{\pi}{2}} + 2}{\sqrt{n}\mathbb{E}|Y_1|^3} + \frac{\frac{1}{2}\sqrt{\frac{\pi}{2}} + 1}{\sqrt{n}} + \frac{1}{\sqrt{2\pi}\sqrt{n-1}\mathbb{E}|Y_1|^3} \\
&+ \frac{1}{2\sqrt{2\pi}\sqrt{n-1}}] \\
&\leq \mathbb{E}|Y_1|^3[\frac{\sqrt{\frac{\pi}{2}} + 2}{\sqrt{n}} + \frac{\frac{1}{2}\sqrt{\frac{\pi}{2}} + 1}{\sqrt{n}} + \frac{1}{\sqrt{2\pi}\sqrt{\frac{n}{2}}} \\
&+ \frac{1}{2\sqrt{2\pi}\sqrt{\frac{n}{2}}}] \\
&= \mathbb{E}|Y_1|^3[\frac{\sqrt{\frac{\pi}{2}} + 2}{\sqrt{n}} + \frac{\frac{1}{2}\sqrt{\frac{\pi}{2}} + 1}{\sqrt{n}} + \frac{3}{2\sqrt{2\pi}\sqrt{\frac{n}{2}}}] \\
&\leq \mathbb{E}|Y_1|^3[\frac{\sqrt{\frac{\pi}{2}} + 2}{\sqrt{n}} + \frac{\frac{1}{2}\sqrt{\frac{\pi}{2}} + 1}{\sqrt{n}} + \frac{1}{\sqrt{n}}],
\end{aligned}
$$

sillä $3\sqrt{2}/2\sqrt{2\pi} \approx 0.85 < 1$. Lisäksi, koska

$$
\sqrt{\frac{\pi}{2}} + 2 + \frac{1}{2}\sqrt{\frac{\pi}{2}} + 1 + 1 \approx 5.88 < 6,
$$

pätee

$$\begin{aligned}\left(\sqrt{\frac{\pi}{2}}+2\right)\frac{1}{\sqrt{n}} \quad &+ \quad \left(\frac{1}{2}\sqrt{\frac{\pi}{2}}+1\right)\frac{\mathbb{E}|Y_1|^3}{\sqrt{n}}+\frac{1}{\sqrt{2\pi}\sqrt{n-1}}+\frac{\mathbb{E}|Y_1|^3}{2\sqrt{2\pi}\sqrt{n-1}} \\ &\leq \quad \frac{6\mathbb{E}|Y_1|^3}{\sqrt{n}},\end{aligned}$$

jolloin ensimmäinen osa on valmis. Käyttämällä taas hyväksi tietoa $n-1 \geq \frac{n}{2}$, kun $n \geq 2$ saadaan toiselle osalle arvio

$$\begin{aligned}\frac{C_{n-1}\mathbb{E}|Y_1|^3}{\sqrt{n}\sqrt{n-1}\epsilon}+\frac{C_{n-1}\left(\mathbb{E}|Y_1|^3\right)^2}{2\sqrt{n}\sqrt{n-1}\epsilon} \quad &= \quad \frac{C_{n-1}\left(\mathbb{E}|Y_1|^3\right)^2}{\sqrt{n}\sqrt{n-1}\epsilon\mathbb{E}|Y_1|^3}+\frac{C_{n-1}\left(\mathbb{E}|Y_1|^3\right)^2}{2\sqrt{n}\sqrt{n-1}\epsilon} \\ &\leq \quad \frac{C_{n-1}\left(\mathbb{E}|Y_1|^3\right)^2}{\sqrt{n}\sqrt{n-1}\epsilon}+\frac{C_{n-1}\left(\mathbb{E}|Y_1|^3\right)^2}{2\sqrt{n}\sqrt{n-1}\epsilon} \\ &= \quad \frac{3C_{n-1}\left(\mathbb{E}|Y_1|^3\right)^2}{2\sqrt{n}\sqrt{n-1}\epsilon} \\ &\leq \quad \frac{3C_{n-1}\left(\mathbb{E}|Y_1|^3\right)^2}{2\sqrt{n}\sqrt{\frac{n}{2}}\epsilon} \\ &= \quad \frac{3C_{n-1}\left(\mathbb{E}|Y_1|^3\right)^2}{\sqrt{2}n\epsilon} \leq \frac{3C_{n-1}\left(\mathbb{E}|Y_1|^3\right)^2}{n\epsilon},\end{aligned}$$

mikä todistaa toisen osan. Yhdistämällä tulokset voidaan todeta suoraan, että (5.7) pitää paikkansa, josta lause seuraa. □

Huomautus. Vakiolle C on vuosien saatossa löydetty aina vaan pienempiä, eli tarkempia, arvoja. Esseen [4] itse käytti alunperin vuonna 1942 arvoa $C < 7.59$, joka on 70 vuoden kuluessa, vuoteen 2012 mennessä pienentynyt aina jopa arvoon $C < 0.4748$ Irina Shevtsovan [12] toimesta. Esseen [5] todisti vuonna 1956 vakiolle C myös alarajan $C \geq \frac{\sqrt{10}+3}{6\sqrt{2\pi}} \approx 0.40973$.

6 Esimerkkejä

6.1 Esimerkki 1

[2] Ensimmäisenä esimerkkinä näytetään kuinka melko yksinkertaisen satunnaismuuttujan jakauma voidaan osoittaa olevan likimain normaalijakautunut. Olkoon $X_1, X_2, ..., X_n$ riippumattomia, samoin jakautuneita satunnaismuuttujia, joille pätee $\mathbb{E}(X_i) = 0$ sekä $\mathbb{E}(X_i^2) = 1$, jolloin siis myös $Var(X_i) = 1$. Oletetaan, että $n \to \infty$.

Asetetaan satunnaismuuttuja S kaavalla $S = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$. Tutkitaan siis Steinin menetelmän avulla, onko satunnaismuuttuja S likimain normaalijakautunut vai ei.

Olkoon $S_i = S - \frac{1}{\sqrt{n}} X_i$ kaikilla $1 \leq i \leq n$. Tällöin satunnaismuuttujat S_i sekä X_i ovat riippumattomia.

Näin ollen kaikilla funktioilla f saadaan riippumattomuuden nojalla odotusarvoksi

$$\mathbb{E}(X_i f(S_i)) = \mathbb{E}(X_i)\mathbb{E}(f(S_i)) = 0.$$

Tästä seuraa Taylorin sarjan avulla

$$\mathbb{E}(X_i f(S)) = \mathbb{E}(X_i(f(S) - f(S_i))) \approx \mathbb{E}(X_i(S - S_i)f'(S)) = \frac{1}{\sqrt{n}}\mathbb{E}(X_i^2 f'(S)).$$

Nyt, satunnaismuuttujan S määritelmän mukaan

$$\mathbb{E}(Sf(S)) = \frac{1}{\sqrt{n}} \sum_{i}^{n} \mathbb{E}(X_i f(S)) \approx \frac{1}{n} \sum_{i}^{n} \mathbb{E}(X_i^2 f'(S)) = \mathbb{E}((\frac{1}{n} \sum_{i}^{n} X_i^2) f'(S)).$$

Koska $\mathbb{E}(X_i^2) = 1$ sekä $n \to \infty$, pätee suurten lukujen lain perusteella, että $\frac{1}{n} \sum_{i}^{n} X_i^2 \approx 1$. Tällöin saadaan $\mathbb{E}(Sf(S)) \approx \mathbb{E}(f'(S))$, jolloin Steinin lemman mukaan satunnaismuuttuja S on siis likimain normaalijakautunut.

6.2 Esimerkki 2

[8] Esimerkkinä Lauseiden 4.2, 4.3 ja 4.4 käytöstä verrataan kahta eri varianssin omaavaa normaalijakautunutta satunnaismuuttujaa toisiinsa. Olkoon N_1 ja N_2 kaksi normaalijakautunutta satunnaismuuttujaa, joille pätee $\mathbb{E}(N_1) = \mathbb{E}(N_2) = 0$ sekä $Var(N_1) = \sigma_1^2 > 0$ ja $Var(N_2) = \sigma_2^2 > 0$. Lasketaan näiden satunnaismuuttujien väliselle etäisyydelle rajoitukset käyttäen Total variation-, Kolmogorov- sekä Wasserstein-mittoja. Oletetaan, että $\sigma_1 \leq \sigma_2$ sekä merkitään standardinormaalijakautunutta satunnaismuuttujaa merkinnällä N_0.

Total variation-etäisyydelle huomataan, että

$$d_{TV}(N_1,N_2)=d_{TV}(N_1/\sigma_2,N_2/\sigma_2)=d_{TV}(N_1/\sigma_2,N_0).$$

Käyttämällä Lausetta 4.2 sekä tietoa, että reaaliarvoisella satunnaismuuttujalla X on jakauma $X\sim N(\mu,\sigma^2)$ jos, ja vain jos, $\mathbb{E}((X-\mu)f(X))=\sigma^2\mathbb{E}(f'(X))$ kaikille differentioituville funktioille f, saadaan

$$\begin{aligned}
d_{TV}(N_1/\sigma_2,N_0) &\leq \sup_{f\in\mathcal{F}_{TV}}|\mathbb{E}(f'(N_1/\sigma_2))-\mathbb{E}(N_1f(N_1/\sigma_2)/\sigma_2)| \\
&= \sup_{f\in\mathcal{F}_{TV}}|\mathbb{E}(f'(N_1/\sigma_2))-(\sigma_1^2/\sigma_2^2)\mathbb{E}(f'(N_1/\sigma_2))| \\
&= \sup_{f\in\mathcal{F}_{TV}}|\mathbb{E}(f'(N_1/\sigma_2)(1-\sigma_1^2/\sigma_2^2))| \\
&\leq 2|1-\sigma_1^2/\sigma_2^2|,
\end{aligned}$$

sillä $Var(N_1/\sigma_2)=\sigma_1^2/\sigma_2^2$.

Kolmogorov-etäisyydelle käsittely on lähes identtinen. Muistetaan vain Lauseesta 4.3, että $\|f_z'\|_\infty\leq 1$ (vertaus Lauseesta 4.2 $\|f_h'\|_\infty\leq 2$). Näin ollen

$$\begin{aligned}
d_{Kol}(N_1/\sigma_2,N_0) &\leq \sup_{f\in\mathcal{F}_{Kol}}|\mathbb{E}(f'(N_1/\sigma_2))-\mathbb{E}(N_1f(N_1/\sigma_2)/\sigma_2)| \\
&= \sup_{f\in\mathcal{F}_{Kol}}|\mathbb{E}(f'(N_1/\sigma_2)(1-\sigma_1^2/\sigma_2^2))| \\
&\leq |1-\sigma_1^2/\sigma_2^2|.
\end{aligned}$$

Wasserstein-etäisyydelle tilanne hankaloituu luonnollisesti hieman. Aluksi on huomattava, että

$$d_W(N_1,N_2)=\sup_{h\in Lip(\sigma_2)}|\mathbb{E}(h(N_1/\sigma_2))-\mathbb{E}(h(N_0))|.$$

Soveltamalla Lausetta 4.4 saadaan

$$\begin{aligned}
\sup_{h\in Lip(\sigma_2)}|\mathbb{E}(h(N_1/\sigma_2))-\mathbb{E}(h(N_0))| &= \sup_{h\in Lip(\sigma_2)}|\mathbb{E}(f_h'(N_1/\sigma_2)-N_1f_h(N_1/\sigma_2)/\sigma_2)| \\
&\leq \sup_{f\in\mathcal{C}^1\cap Lip(\sigma_2\sqrt{\frac{2}{\pi}})}|\mathbb{E}(f'(N_1/\sigma_2))-\mathbb{E}(N_1f(N_1/\sigma_2)/\sigma_2)| \\
&\leq \frac{\sigma_2\sqrt{2}}{\sqrt{\pi}}|1-\sigma_1^2/\sigma_2^2|.
\end{aligned}$$

7 Matriisilaskennasta

Tässä luvussa määritellään lyhyesti erilaisia lineaari- ja matriisilaskennasta tuttuja, tätä tutkielmaa varten hyödyllisiä käsitteitä. Niitä ei ole tarkoitus avata enempää, vain muistuttaa mieleen, jotta niiden käyttäminen seuraavassa luvussa olisi mielekkäämpää.

Määritelmä 7.1. Neliömatriisin $A \in d \times d$ jälki on sen päädiagonaalin alkioiden summa. Lyhyesti

$$Tr(A) = \sum_{i=1}^{d} a_{ii}.$$

Määritelmä 7.2. Matriisin A transpoosi määritellään kääntämällä matriisi ikään kuin ympäri, jolloin sen riveistä tulee sarakkeita, ja sarakkeista tulee rivejä. Erityisesti neliömatriisin transpoosi voidaan muodostaa peilaamalla alkiot päädiagonaalin suhteen. Lyhyesti, jos A on $m \times n$ matriisi, pätee

$$A^T_{ij} = A_{ji},$$

kaikilla $1 \leq i \leq m$ sekä $1 \leq j \leq n$.

Seuraavaksi määritellään matriisien tavallinen sisätulo sekä normi, joiden erikoistapauksena seuraavassa luvussa tarvittavat niin kutsutut Hilbert-Schmidt versiot niistä.

Määritelmä 7.3. Kahden vektorin $x \in \mathbb{R}^d$ sekä $y \in \mathbb{R}^d$ välinen sisätulo määritellään yksinkertaisesti kaavalla

$$\langle x, y \rangle = \sum_{i=1}^{d} x_i y_i.$$

Määritelmä 7.4. Sisätulon määritelmästä saamme kätevästi määriteltyä suoraan vektorin $x \in \mathbb{R}^d$ normin, joka määritellään neliöjuurena sen sisätulosta itsensä kanssa kaavalla

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Määritelmä 7.5. Matriisien $A \in m \times s$ ja $B \in s \times n$ tavallinen kertolasku määritellään ensimmäisen matriisin vaakarivivektorien ja toisen matriisin pystyrivivektorien sisätulona. Tulomatriisin AB dimensio on $m \times n$ ja se muodostuu siis kaavalla

$$AB = \begin{bmatrix} a_1^T b_1 & \dots & a_1^T b_n \\ \vdots & \ddots & \vdots \\ a_m^T b_1 & \dots & a_m^T b_n \end{bmatrix}.$$

Määritelmä 7.6. Hilbert-Schmidt sisätulo määritellään kaikille matriiseille A ja B kaavalla

$$\langle A, B\rangle_{HS} = Tr(AB^T).$$

Määritelmä 7.7. Vastaavasti siis Hilbert-Schmidt normi kaikille matriiseille A saadaan kaavasta

$$\|A\|_{HS} = \sqrt{\langle A, A\rangle_{HS}}.$$

Määritelmä 7.8. Neliömatriisin $A \in d \times d$ niin kutsuttu operaattorinormi määritellään kaavalla

$$\|A\|_{op} = \sup\left\{\|Ax\|_{\mathbb{R}^d} : x \in \mathbb{R}^d, \textit{jolle pätee } \|x\|_{\mathbb{R}^d} = 1\right\},$$

jossa $\|\|_{\mathbb{R}^d}$ on tavallinen d-ulotteinen normi.

Määritelmä 7.9 (Hessen matriisi)**.** Hessen matriisi on neliömatriisi $d \times d$, jonka alkiot koostuvat funktion $f : \mathbb{R}^d \to \mathbb{R}$ toisen kertaluvun osittaisderivaatoista. Toisin sanoen funktion f Hessen matriisi on

$$Hessf = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1} & \frac{\partial^2 f}{\partial x_d \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_d^2} \end{bmatrix}.$$

Esimerkki 7.1. *Yksinkertaisena esimerkkinä lasketaan funktion $f(x, y) = e^x y + y^3$ Hessen matriisi. Nyt pätee*

$$\frac{\partial f(x,y)}{\partial x^2} = e^x y, \qquad \frac{\partial f(x,y)}{\partial x \partial y} = \frac{\partial f(x,y)}{\partial y \partial x} = e^x, \qquad \frac{\partial f(x,y)}{\partial y^2} = 6y.$$

Hessen matriisiksi saadaan

$$Hessf = \begin{bmatrix} e^x y & e^x \\ e^x & 6y \end{bmatrix}.$$

Määritellään vielä lopuksi funktion gradientti, jota tullaan käyttämään seuraavassa kappaleessa.

Määritelmä 7.10. Olkoon $x = [x_1, x_2, ..., x_n]^T$ n muuttujan vektori. Nyt funktion f gradientti määrätään osittaisderivaattojen vektorina kaavalla

$$\nabla f(x) = [\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), ..., \frac{\partial f}{\partial x_n}(x)]^T.$$

8 Moniulotteisuudesta

Tässä luvussa käydään läpi (hieman ylimalkaisemmin kuin yksiulotteinen tapaus) Steinin menetelmää moniulotteisessa tapauksessa. Sitä varten käytetään luonnollisesti matriiseja, jolloin lausekkeet näyttävät sotkuisemmilta ja hankalammilta. Tulokset ovat kuitenkin hyvin paljon samankaltaisia kuin aiemmin esitetyt vastaavat tulokset yksiulotteisessa tapauksessa. Kuten aiemmin, menetelmän läpikäyminen aloitetaan Steinin lemmasta. Tarkemmin ja yksityiskohtaisemmin moniulotteisuudesta voi lukea lähteestä [8].

Lemma 8.1. *[8] Olkoon $C_{ij}, i, j = 1, ..., d$ ei-negatiivisesti definiitti $d \times d$ matriisi. Lisäksi olkoon $W \in \mathbb{R}^d = (W_1, ..., W_d)$ satunnaisvektori. Silloin satunnaisvektorilla W on normaalijakauma $W \sim N(0, C)$ jos, ja vain jos*

$$\mathbb{E}[\langle W, \nabla f(W)\rangle_{\mathbb{R}^d}] = \mathbb{E}[\langle C, Hessf(W)\rangle_{HS}], \tag{8.1}$$

kaikille funktioille $f : \mathbb{R}^d \to \mathbb{R} \in \mathcal{C}^2$.

Vertaamalla Lemmaa 8.1 yksiulotteisen version tulokseen

$$\mathbb{E}f'(W) = \mathbb{E}(Wf(W)),$$

joka funktion f derivaattaan soveltamalla voidaan kirjoittaa muodossa

$$\mathbb{E}(f''(W)) = \mathbb{E}(Wf'(W)),$$

huomataan niiden vastaavan toisiaan melko selvästi.

Todistus. [8] Oletetaan ensin, että $W = (W_1, ..., W_d) \sim N_d(0, C)$ on d-ulotteinen normaalijakauma odotusarvolla 0 sekä kovarianssimatriisilla C. Nyt saadaan suoraan laskemalla, että

$$\begin{aligned}
\mathbb{E}[\langle W, \nabla f(W)\rangle_{\mathbb{R}^d}] &= \sum_{i=1}^{d} \mathbb{E}\left[W_i \frac{\partial f}{\partial x_i}(W_1, ..., W_d)\right] \\
&= \sum_{i,j=1}^{d} C_{ij}\mathbb{E}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}(W_1, ..., W_d)\right] \\
&= \mathbb{E}[\langle C, Hessf(W)\rangle_{HS}].
\end{aligned}$$

Seuraavaksi oletetaan, että satunnaisvektorille W pätee (8.1). Otetaan avuksi normaalijakautunut satunnaisvektori $G \sim N_d(0, C)$ ja osoitetaan, että

$\mathbb{E}[f(W)] = \mathbb{E}[f(G)]$ kaikille lauseen oletuksen mukaisille funktioille f. Nyt pätee

$$\begin{aligned}
\mathbb{E}[f(W)] &- \mathbb{E}[f(G)] = \int_0^1 \mathbb{E}[\langle \nabla f(\sqrt{t}W + \sqrt{1-t}G), W\rangle_{\mathbb{R}^d}] \frac{dt}{2\sqrt{t}} \\
&- \int_0^1 \mathbb{E}[\langle \nabla f(\sqrt{t}W + \sqrt{1-t}G), G\rangle_{\mathbb{R}^d}] \frac{dt}{2\sqrt{1-t}} \\
&= \int_0^1 \mathbb{E}\left[\mathbb{E}[\langle \nabla f(\sqrt{t}W + \sqrt{1-t}x), W\rangle_{\mathbb{R}^d}]|x = G\right] \frac{dt}{2\sqrt{t}} \\
&- \int_0^1 \mathbb{E}\left[\mathbb{E}[\langle \nabla f(\sqrt{t}x + \sqrt{1-t}G), G\rangle_{\mathbb{R}^d}]|x = W\right] \frac{dt}{2\sqrt{1-t}},
\end{aligned}$$

sillä kun

$$\gamma(t) = \mathbb{E}(f(\sqrt{t}W + \sqrt{1-t}G))$$

pätee tulos

$$\gamma(1) - \gamma(0) = \int_0^1 \gamma'(t)dt,$$

jolloin

$$\mathbb{E}(f(W)) = \gamma(1) \qquad \textit{sekä} \qquad \mathbb{E}(f(G)) = \gamma(0).$$

Nyt käyttämällä kaavaa (8.1) saadaan

$$\begin{aligned}
\mathbb{E}[\langle \nabla f(\sqrt{t}W + \sqrt{1-t}x), W\rangle_{\mathbb{R}^d}] &= \sqrt{t}\mathbb{E}[\langle C, Hessf(\sqrt{t}W + \sqrt{1-t}x)\rangle_{HS}] \\
&=: h_1(x).
\end{aligned}$$

Lisäksi vastaavasti saadaan

$$\begin{aligned}
\mathbb{E}[\langle \nabla f(\sqrt{t}x + \sqrt{1-t}G), G\rangle_{\mathbb{R}^d}] &= \sqrt{1-t}\mathbb{E}[\langle C, Hessf(\sqrt{t}x + \sqrt{1-t}G)\rangle_{HS}] \\
&=: h_2(x).
\end{aligned}$$

Lopuksi sijoitetaan $h_1(x)$ ja $h_2(x)$ yhtälöön

$$\begin{aligned}
\mathbb{E}[f(W)] - \mathbb{E}[f(G)] &= \int_0^1 \mathbb{E}\left[\mathbb{E}[\langle \nabla f(\sqrt{t}W + \sqrt{1-t}x), W\rangle_{\mathbb{R}^d}]|x = G\right] \frac{dt}{2\sqrt{t}} \\
&- \int_0^1 \mathbb{E}\left[\mathbb{E}[\langle \nabla f(\sqrt{t}x + \sqrt{1-t}G), G\rangle_{\mathbb{R}^d}]|x = W\right] \frac{dt}{2\sqrt{1-t}}.
\end{aligned}$$

Saadaan

$$\begin{aligned}
\mathbb{E}[f(W)] - \mathbb{E}[f(G)] &= \int_0^1 \mathbb{E}\left[\sqrt{t}\mathbb{E}[\langle C, Hessf(\sqrt{t}W + \sqrt{1-t}x)\rangle_{HS}]|x = G\right] \frac{dt}{2\sqrt{t}} \\
&- \int_0^1 \mathbb{E}\left[\sqrt{1-t}\mathbb{E}[\langle C, Hessf(\sqrt{t}x + \sqrt{1-t}G)\rangle_{HS}]|x = W\right] \frac{dt}{2\sqrt{1-t}} \\
&= \int_0^1 \mathbb{E}[\langle C, Hessf(\sqrt{t}W + \sqrt{1-t}G)\rangle_{HS}]dt \\
&- \int_0^1 \mathbb{E}[\langle C, Hessf(\sqrt{t}W + \sqrt{1-t}G)\rangle_{HS}]dt \\
&= 0.
\end{aligned}$$

□

Seuraavaksi esitellään Steinin yhtälö standardinormaalijakautuneelle satunnaisvektorille.

Määritelmä 8.1. Olkoon $W \sim N_d(0, I_d)$ standardinormaalijakautunut satunnaisvektori. Olkoon lisäksi funktio $h : \mathbb{R}^d \to \mathbb{R}$ sellainen, että odotusarvo $\mathbb{E}|h(W)|$ on äärellinen. Steinin yhtälö on tällöin muotoa

$$\Delta f(x) - \langle x, \nabla f(x)\rangle_{\mathbb{R}^d} = h(x) - \mathbb{E}(h(W)), \tag{8.2}$$

jonka ratkaisu $f : \mathbb{R}^d \to \mathbb{R}$ on kahdesti jatkuvasti derivoituva funktio.

Huomautus. Merkinnällä Δ tarkoitetaan niin kutsuttua Laplacen operaattoria, joka määritellään kaavalla

$$\Delta f(x) = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} f = \langle I_d, Hessf\rangle_{HS}.$$

Huomautus. Yksiulotteisessa tapauksessa Steinin yhtälö oli siis muotoa

$$f'(x) - xf(x) = h(x) - \mathbb{E}[h(W)],$$

joka jälleen voidaan kirjoittaa muodossa

$$f''(x) - xf'(x) = h(x) - \mathbb{E}[h(W)].$$

Kuten yksiulotteisessa tapauksessa, myös nyt Steinin yhtälön ratkaisu tiedetään valmiiksi. Seuraava lause kertoo erään tällaisen funktion f. Lauseen todistus löytyy esimerkiksi lähteestä [8].

Lause 8.1. *[8] Olkoon $W \sim N_d(0, I_d)$ standardinormaalijakautunut satunnaisvektori. Olkoon lisäksi $h : \mathbb{R}^d \to \mathbb{R}$ Lipschitz-funktio vakiolla $K = \sup_{x \neq y} \frac{|h(x)-h(y)|}{\|x-y\|_{\mathbb{R}^d}} > 0$. Määritellään funktio $f_h : \mathbb{R}^d \to \mathbb{R}$ kaavalla*

$$f_h(x) = \int_0^\infty \mathbb{E}(h(W) - h(e^{-t}x + \sqrt{1 - e^{-2t}}W))dt.$$

Nyt, funktio f_h on hyvin määritelty, kahdesti jatkuvasti derivoituva ratkaisu Steinin yhtälölle (8.2) kaikilla $x \in \mathbb{R}^d$. Lisäksi pätee

$$\sup_{x \in \mathbb{R}^d} \|Hessf_h(x)\|_{HS} \leq \sqrt{d}K.$$

Seuraavaksi esitellään versio Steinin yhtälöstä tapauksessa, jossa kovarianssimatriisi on identiteettimatriisin sijaan positiivisesti definiitti $d \times d$ matriisi C.

Määritelmä 8.2. Olkoon $W \sim N_d(0, C)$ normaalijakautunut satunnaisvektori. Olkoon lisäksi funktio $h : \mathbb{R}^d \to \mathbb{R}$ sellainen, että odotusarvo $\mathbb{E}|h(W)|$ on äärellinen. Steinin yhtälö on tällöin muotoa

$$\langle C, Hessf(x) \rangle_{HS} - \langle x, \nabla f(x) \rangle_{\mathbb{R}^d} = h(x) - \mathbb{E}(h(W)), \tag{8.3}$$

jonka ratkaisu $f : \mathbb{R}^d \to \mathbb{R}$ on kahdesti jatkuvasti derivoituva funktio.

Lauseen 8.1 tavoin, seuraava lause kertoo erään tällaisen ratkaisun f.

Lause 8.2. *[8] Olkoon $W \sim N_d(0, C)$ normaalijakautunut satunnaisvektori. Olkoon $h : \mathbb{R}^d \to \mathbb{R}$ Lipschitz-funktio vakiolla $K > 0$. Funktio*

$$f_h(x) = \int_0^\infty \mathbb{E}(h(W) - h(e^{-t}x + \sqrt{1 - e^{-2t}}W))dt$$

on hyvin määritelty, kahdesti jatkuvasti derivoituva ratkaisu Steinin yhtälölle (8.3) kaikilla $x \in \mathbb{R}^d$. Lisäksi on voimassa

$$\sup_{x \in \mathbb{R}^d} \|Hessf_h(x)\|_{HS} \leq \sqrt{d}K \times \|C^{-1}\|_{op} \times \|C\|_{op}^{1/2}.$$

Lause todistetaan esimerkiksi lähteessä [8].

Viimeinen askel Steinin menetelmässä on luonnollisesti johtaa Steinin yhtälön avulla käyttökelpoisia rajoituksia. Koska Steinin yhtälöissä määritellyt funktiot h ovat Lipschitz-funktioita, voidaan kolmesta esitellystä etäisyysmitasta käyttää ainoastaan Wasserstein-etäisyyttä.

Lause 8.3. *[8] Olkoon C positiivisesti definiitti $d \times d$ matriisi. Olkoon lisäksi $W \sim N_d(0, C)$ normaalijakautunut satunnaisvektori. Mille tahansa neliöintegroituvalle satunnaisvektorille $X \in \mathbb{R}^d$ pätee*

$$d_W(X, W) \leq \sup_{f \in \mathcal{F}_W^d(C)} |\mathbb{E}(\langle C, Hessf(X) \rangle_{HS} - \mathbb{E}(\langle X, \nabla f(X) \rangle_{\mathbb{R}^d})|,$$

jossa

$$\mathcal{F}_W^d(C) = \{f : \mathbb{R}^d \to \mathbb{R} \in \mathcal{C}^2 : \sup_{x \in \mathbb{R}^d} \|Hessf(x)\|_{HS} \leq \sqrt{d} \times \|C^{-1}\|_{op} \times \|C\|_{op}^{1/2}\}.$$

Todistus. Lauseen 8.2 nojalla päätellään, että

$$\begin{aligned} d_W(X, W) &= \sup_{h \in Lip(1)} |\mathbb{E}(h(W)) - \mathbb{E}(h(X))| \\ &= \sup_{h \in Lip(1)} |\mathbb{E}(\langle C, Hessf_h(X) \rangle_{HS}) - \mathbb{E}(\langle X, \nabla f_h(X) \rangle_{\mathbb{R}^d})| \\ &\leq \sup_{f \in \mathcal{F}_W^d(C)} |\mathbb{E}(\langle C, Hessf_h(X) \rangle_{HS}) - \mathbb{E}(\langle X, \nabla f_h(X) \rangle_{\mathbb{R}^d})|. \end{aligned}$$

□

Viitteet

[1] Barbour A.D., Chen Louis H. Y.: *An Introduction to Stein's Method, Singapore University Press ja World Scientific Publishing, 2005*

[2] Chatterjee Sourav: *A short survey of Stein's method, Stanford University*

[3] Dey Partha: *Stein-Chen Method For Poisson Approximation, University of Warwick, 2013-2014*

[4] Esseen Carl-Gustav: *On the Liapunoff limit of error in the theory of probability, Arkiv för matematik, astronomi och fysik, 1942*

[5] Esseen Carl-Gustav: *A moment inequality with an application to the central limit theorem, Skand. Aktuarietidskr, 1956*

[6] Honkasalo Hannu: *Lineaarialgebra I, Helsingin yliopisto, 2003*

[7] Nourdin Ivan: *Selected Aspects of Fractional Brownian Motion, Bocconi University Press, 2012*

[8] Nourdin Ivan, Peccati Giovanni: *Normal Approximations With Malliavin Calculus, Cambridge University Press, 2012*

[9] Peccati Giovanni: *Stein's method, Malliavin calculus and infinite-dimensional Gaussian analysis, University of Singapore Lecture notes, 2009*

[10] Peccati Giovanni, Reitzner Matthias: *Stochastic Analysis for Poisson Point Processes, Bocconi University Press, 2016*

[11] Reinert Gesine: *A Short Introduction to Stein's Method, Department of Statistics University of Oxford, 2011*

[12] Shevtsova Irina: *On the absolute constants in the Berry Esseen type inequalities for identically distributed summands, Lomonosov Moscow State University, 2011*

[13] Sottinen, Tommi: *Todennäköisyysteoria, Helsingin yliopisto, 2006*

[14] Stein Charles: *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, University of Stanford, 1972*

[15] Stein Charles: *Approximate Computation of Expectations, Stanford University, 1986*

[16] Stroock, Daniel W.: *Probability Theory, Cambridge University Press, 1994*