Sisältö

Luku 7

Ehdollinen jakauma

7.1 Ehdolliset jakaumat

Jos sv:lla (X, Y) on diskreetti jakauma, niin sm:n X ehdollinen ptnf ehdolla $Y = y$ määritellään ehdollisen todennäköisyyden kaavan avulla,

$$f_{X|Y}(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \qquad (7.1)$$

kun y on sellainen piste, jossa $f_Y(y) > 0$. Funktio $f_{X|Y}(\cdot \mid y)$ on satunnaismuuttujan X ptnf, kun tiedetään, että $Y = y$. Ehdollinen ptnf $f_{Y|X}(y \mid x)$ määritellään analogisesti.

Kertolaskusääntö eli *ketjusääntö* kertoo, että

$$f_{X,Y}(x, y) = f_X(x)\, f_{Y|X}(y \mid x) = f_Y(y)\, f_{X|Y}(x \mid y),$$

ja *Bayesin kaava* kertoo, että

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_X(x)\, f_{Y|X}(y \mid x)}{f_Y(y)}.$$

Näissä kaavoissa oletetaan, että x ja y ovat sellaisia pisteitä, joissa ei jouduta jakamaan nollalla. Jos $X \perp\!\!\!\perp Y$, niin

$$f_{X|Y}(x \mid y) = f_X(x), \qquad f_{Y|X}(y \mid x) = f_Y(y)$$

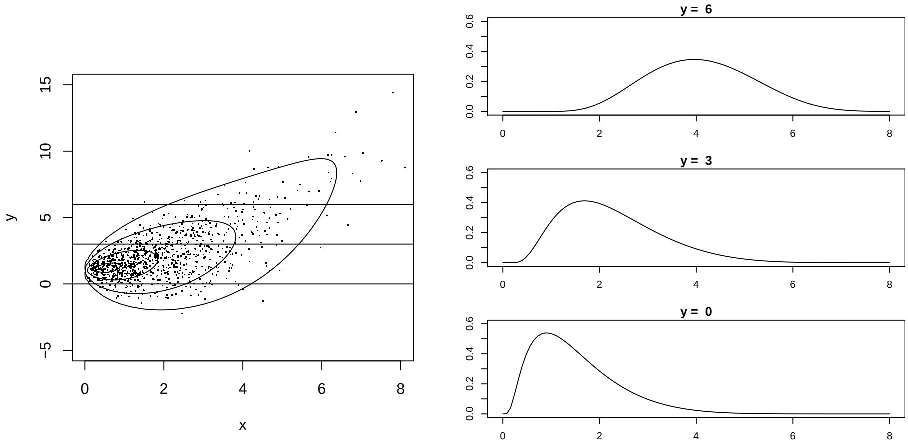
kaikilla kyseeseen tulevilla argumenteilla. Kun $X \perp\!\!\!\perp Y$, niin tieto sm:n Y arvosta ei anna tietoa X:n jakaumasta eikä tieto X:n arvosta anna tietoa Y:n jakaumasta.

Kaavaa (7.1) pidetään mallina, kun määritellään ehdollinen tiheysfunktio jatkuvan yhteisjakauman tapauksessa.

Määritelmä 7.1. Olkoon (X, Y):llä jatkuva jakauma tiheysfunktiolla $f_{X,Y}$. Tällöin sm:n X ehdollinen tiheysfunktio ehdolla $Y = y$ on

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \qquad x \in \mathbb{R},$$

Kuva 7.1 Yhteistiheysfunktio ja ehdollisia tiheysfunktioita $x \mapsto f_{X|Y}(x \mid y)$.



kun y on sellainen, että $f_Y(y) > 0$. Vastaavasti määritellään

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \qquad y \in \mathbb{R},$$

kun $f_X(x) > 0$.

Huomaa, että $x \mapsto f_{X|Y}(x \mid y)$ on tiheysfunktio, sillä se on ei-negatiivinen ja

$$\int_{-\infty}^{\infty} f_{X|Y}(x \mid y)\,\mathrm{d}x = \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x = 1.$$

Funktio $x \mapsto f_{X|Y}(x \mid y)$ saadaan yhteistiheysfunktiosta $f_{X,Y}$ normalisoimalla sen vaakasuora “leike” $x \mapsto f_{X,Y}(x,y)$ tiheysfunktioksi. Vastaavasti funktio $y \mapsto f_{Y|X}(y \mid x)$ saadaan normalisoimalla pystysuora leike $y \mapsto f_{X,Y}(x,y)$. Kuva 7.1 havainnollistaa sitä, miten ehdollinen tiheysfunktio saadaan yhteistiheysfunktiosta.

Jatkuvan yhteisjakauman voi esittää useamman kuin yhden tiheysfunktion avulla kunhan ne vain yhtyvät melkein kaikkialla. Ehdollisen tf:n määrittelyyn saadaan käyttää mitä tahansa ytf:n määritelmää, joten myöskään ehdollinen tf ei ole yksikäsitteinen. Toisaalta osoittautuu, ettei tästä monikäsitteisyydestä synny sen suurempia ongelmia, minkä takia sitä ei oteta huomioon puhetavassa: puhutaan ehdollisesta tiheysfunktiosta (eikä esim. ehdollisen tiheysfunktion tietystä versiosta).

Jatkuvan yhteisjakauman tapauksessa ehdon $\{Y = y\}$ todennäköisyys on aina nolla. Tämän takia ehdolliselle tiheysfunktiolle ei voida antaa suoraan tulkintaa ehdollisen todennäköisyyden avulla, vaan kyseessä on määritelmä, jota yritämme seuraavaksi motivoida.

Ehdollisen tiheysfunktion kaavalle voidaan antaa infinitesimaalinen tulkinta

$$f_{X|Y}(x \mid y)\,\mathrm{d}x = \frac{f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y}{f_Y(y)\,\mathrm{d}y} = \frac{P(X \in \mathrm{d}x, Y \in \mathrm{d}y)}{P(Y \in \mathrm{d}y)} = P(X \in \mathrm{d}x \mid Y \in \mathrm{d}y)$$

Tämä ei ole pelkästään formaalia leikkiä symboleilla, vaan sille voidaan antaa tulkinta raja-arvojen kautta. Jos ytf on riittävän sileä, niin voidaan tulkita, että

$$P(X \in A \mid Y \in \mathrm{d}y) = \lim_{h \to 0+} P(X \in A \mid y \le Y \le y + h).$$

Jos $f_{X,Y}$ ja f_Y ovat riittävän sileitä, ja $h > 0$ on pieni, niin suurpiirteisesti laskien nähdään, että

$$\begin{aligned} P(X \in A \mid y \le Y \le y + h) &= \frac{\frac{1}{h} P(X \in A, y \le Y \le y + h)}{\frac{1}{h} P(y \le Y \le y + h)} \\ &= \frac{\int_A \mathrm{d}x \, \frac{1}{h} \int_y^{y+h} f_{X,Y}(x, t) \, \mathrm{d}t}{\frac{1}{h} \int_y^{y+h} f_Y(u) \, \mathrm{d}u} \approx \frac{\int_A f_{X,Y}(x, y) \, \mathrm{d}x}{f_Y(y)} \end{aligned}$$

ja tässä tarkkuus kasvaa,kun $h \to 0+$. Näin saadaan ehdolliselle tf:lle tulkinta raja-arvona, ts.

$$P(X \in A \mid y \le Y \le y + h) \xrightarrow[h \to 0+]{} \int_A f_{X \mid Y}(x \mid y) \, \mathrm{d}x, \qquad \text{kaikilla } A \subset \mathbb{R}.$$

(Jätämme auki, mitä ehtoja ytf:lle ja reunatiheydelle f_Y pitäisi asettaa, jotta lasku saataisiin vietyä täsmällisesti läpi.) Jatkuvan sm:n arvo havaitaan vain tietyllä tarkkuudella, joten sovelluksissa tehdään tämän tapainen tulkinta, kun käsitellään ehdollista tf:a ehdolla $Y = y$, jossa y on havaittu arvo. Näiden tarkastelujen avulla myös ehdolliselle tiheysfunktiolle olisi mahdollista antaa frekvenssitulkinta.

7.2 Kertolaskusääntö eli ketjusääntö

Ehdollisen ptnf/tf:n määritelmästä saadaan *kertolaskusääntö* eli *ketjusääntö*,

$$f_{X,Y}(x, y) = f_X(x) \, f_{Y \mid X}(y \mid x) = f_Y(y) \, f_{X \mid Y}(x \mid y), \qquad \text{kaikilla } x, y. \tag{7.2}$$

Tämä kaava vaatii hieman tulkintoja jatkuvan jakauman tapauksessa, koska ytf ei ole yksikäsitteinen. Kaavan voidaan ajatella tarkoittavan sitä, että jos reunatiheysfunktio $f_X(x)$ ja ehdollinen tf $f_{Y \mid X}(y \mid x)$ johdetaan lähtemällä liikkeelle joistakin ytf:n (mahdollisesti toisistaan eriävistä) versioista, niin niiden tulo kelpaa yhteistiheysfunktioksi.

Lisäksi kaavassa on se ongelma, että esim. $f_{Y \mid X}(y \mid x)$ ei ole välttämättä määritelty kaikilla x. Sovitaan, että kertolaskusäännön yhteydessä ehdollisen tf:n tai ptnf:n määritelmää laajennetaan (tarvittaessa) jollakin tavalla niihin pisteisiin, joissa ehtomuuttujan tiheys on nolla, esim. sopimalla, että

$$f_{Y \mid X}(y \mid x) = \begin{cases} \dfrac{f_{X,Y}(x, y)}{f_X(x)}, & \text{kun } f_X(x) > 0, \\ 0, & \text{muuten.} \end{cases} \tag{7.3}$$

(Yhtä hyvin voitaisiin käyttää jotakin muuta johdonmukaista sopimusta.) Toiselle ehdolliselle tf:lle $f_{X \mid Y}(x \mid y)$ käytetään tietenkin samanlaista laajennusta. Tämän jälkeen kertolaskusääntö pitää ongelmattomasti paikkansa kaikilla $x, y \in \mathbb{R}$. Nämä komplikaatiot eivät aiheuta todellisia ongelmia käytännön laskuissa.

Kertolaskusäännöstä voidaan ratkaista toinen ehdollisista tiheysfunktioista, jos reunatiheydet sekä toinen ehdollinen tf tunnetaan. Esim., kun $f_X(x) > 0$ (ja muut x-arvot eivät ole relevantteja), on

$$f_{Y|X}(x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_Y(y)\, f_{X|Y}(x \mid y)}{f_X(x)}. \tag{7.4}$$

Tämä on *Bayesin kaava* tiheysfunktioille.

Huomaa, että Bayesin kaavasta seuraa, että kiinteällä x

$$f_{Y|X}(y \mid x) = \frac{1}{f_X(x)} f_Y(y)\, f_{X|Y}(x \mid y) \propto f_Y(y)\, f_{X|Y}(x \mid y) = f_{X,Y}(x,y).$$

Verrannollisuus $g(y) \propto h(y)$ tarkoittaa sitä, että

$$g(y) = c\, h(y)$$

jollakin verrannollisuusvakiolla c (joka saa riippua parametreista, ja muista muuttujista paitsi muuttujasta y). Tämän ansiosta ehdollinen jakauma on toisinaan helppo tunnistaa tarkastelemalla ytf:n lauseketta.

Esimerkki 7.1. Esimerkissä 6.4 ytf:llä on lauseke

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}}\, 1\{0 < y < \exp(-\frac{1}{2}x^2)\}.$$

Ehdolliset tf:t voidaan laskea joko jakamalla ytf esimerkissä johdetuilla reunatf:oilla, tai seuraavalla päättelyllä.

a) $f_{Y|X}$: Kiinteällä x on ytf:llä $f_{X,Y}(x,y)$ nollaa suurempi vakioarvo, kun $0 < y < \exp(-\frac{1}{2}x^2)$. Tämän takia ehdollinen jakauma on tasajakauma,

$$Y \mid (X = x) \sim \mathrm{U}(0, \exp(-\frac{1}{2}x^2)).$$

Kun muistetaan, että reunajakaumassaan $X \sim N(0,1)$, niin yhteisjakauma voidaan esittää muodossa

$$\begin{aligned} Y \mid X &\sim \mathrm{U}(0, \exp(-\frac{1}{2}X^2)), \\ X &\sim N(0,1). \end{aligned}$$

b) $f_{X|Y}$: Kiinteällä $0 < y < 1$ on ytf:llä nolla suurempi vakioarvo tietyllä x-akselin välillä, joka saadaan ratkaisemalla epäyhtälö

$$0 < y < \exp(-\frac{1}{2}x^2)$$

muuttujan x suhteen. Tämä lasku ratkaistiin esimerkissä 6.4. Tuloksesta tunnistetaan, että ehdollinen jakauma on tasajakauma

$$X \mid (Y = y) \sim \mathrm{U}(-\sqrt{-2\ln y}, \sqrt{-2\ln y}), \qquad 0 < y < 1.$$

△

7.3 Diskreetin ja jatkuvan muuttujan yhteisjakauma

Tarkastellaan nyt samalla perusjoukolla määriteltyjen sm:ien X ja Y yhteisjakaumaa siinä tapauksessa, kun X:n jakauma on diskreetti Y:n jatkuva. Olkoon X:n mahdolliset arvot $x_1, x_2, \ldots$ ja olkoon sen (reuna-)ptnf f_X. Voidaan osoittaa, että tällöin on olemassa ehdolliset tiheysfunktiot $y \mapsto f_{Y|X}(y \mid x_i), i \geq 1$ siten, että

$$P(X = x_i, Y \in B) = f_X(x_i) \int_B f_{Y|X}(y \mid x_i) \, \mathrm{d}y, \qquad \text{kaikilla } i \text{ ja kaikilla } B \subset \mathbb{R},$$

mutta todistuksessa tarvitaan mittateoriaa (tulos seuraa ns. Radonin-Nikodymin lauseesta). Tällöin yhteisjakauman esittää funktio

$$f_{X,Y}(x, y) = f_X(x) \, f_{Y|X}(y \mid x)$$

siinä mielessä, että seuraavantyyppiset todennäköisyydet saadaan laskettua summaamalla diskreetin muuttujan ja integroimalla jatkuvan muuttujan suhteen,

$$P(X \in A, Y \in B) = \sum_{x \in A} \int_B f_{X,Y}(x, y) \, \mathrm{d}y, \qquad A, B \subset \mathbb{R}\,.$$

Voimme myös tässä tapauksessa kutsua yhteisjakauman esitystä $f_{X,Y}$ tiheydeksi tai tiheysfunktioksi, mutta on tärkeää pitää mielessä, että yhden muuttujan suhteen summataan ja toisen suhteen integroidaan.

Diskreetin muuttujan X ptnf ehdolla $Y = y$ määritellään Bayesin kaavalla, siis

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)},$$

kun $f_Y(y) > 0$. Tämän kaavan voisi myös motivoida rajankäynnin kautta samaan tapaan kuin jatkuvan yhteisjakauman tapauksessa tehtiin.

Kertolaskusääntö ja Bayesin kaava ovat voimassa myös tässä tapauksessa, ja tarvittaessa ehdollisen tiheyden $f_{Y|X}(y \mid x)$ ja ehdollisen ptnf:n $f_{X|Y}(x \mid y)$ määritelmää voidaan laajentaa myös sellaisille argumenteille, joilla ne eivät vielä tulleet määriteltyä.

Siinä tapauksessa jossa toisen jakauma on diskreetti ja toisen jatkuva, odotusarvoja laskettaessa diskreetin muuttujan suhteen summataan ja jatkuvan suhteen integroidaan.

Lause 7.1. *Olkoon (X, Y) sv, jossa X:llä on diskreetti ja Y:llä jatkuva jakauma, ja olkoon $Z = g(X, Y)$ jokin sen reaaliarvoinen muunnos. Tällöin*

$$EZ = \sum_x \int g(x, y) \, f_{X,Y}(x, y) \, \mathrm{d}y = \int \sum_x g(x, y) \, f_{X,Y}(x, y) \, \mathrm{d}y$$

mikäli

$$\sum_x \int |g(x, y)| \, f_{X,Y}(x, y) \, \mathrm{d}y < \infty.$$

7.4 Ehdollinen odotusarvo

Määritelmä 7.2 (Ehdollinen odotusarvo ja ehdollinen varianssi ehdolla sm:n arvo)**.** Olkoon (X, Y) sv, $g(X, Y)$ jokin sen muunnos, ja oletetaan, että osaamme määritellä sm:n Y ehdollisen jakauman ehdolla $X = x$. Sm:n $g(X, Y)$ ehdollinen odotusarvo ehdolla $X = x$,

$$E(g(X, Y) \mid X = x),$$

on satunnaismuuttujan $g(x, Y)$ odotusarvo, kun Y:n jakaumana käytetään sen ehdollista jakaumaa ehdolla $X = x$. Sm:n $g(X, Y)$ ehdollinen varianssi ehdolla $X = x$,

$$\operatorname{var}(g(X, Y) \mid X = x),$$

on satunnaismuuttujan $g(x, Y)$ varianssi, kun Y:n jakaumana käytetään sen ehdollista jakaumaa ehdolla $X = x$.

Jos Y:n jakauma on jatkuva, on

$$E(g(X, Y) \mid X = x) = \int g(x, y)\, f_{Y|X}(y \mid x)\, \mathrm{d}y,$$

ja

$$\operatorname{var}(g(X, Y) \mid X = x) = \int [g(x, y) - E(g(X, Y) \mid X = x)]^2\, f_{Y|X}(y \mid x)\, \mathrm{d}y.$$

Jos jälkimmäisessä kaavassa binomin neliö kerrotaan auki ja järjestellään termejä, nähdään että

$$\operatorname{var}(g(X, Y) \mid X = x) = E[(g(X, Y))^2 \mid X = x] - \{E[g(X, Y) \mid X = x]\}^2 . \quad (7.5)$$

Tämä on tutun kaavan $\operatorname{var} Z = EZ^2 - (EZ)^2$ vastine ehdolliselle varianssille. Jos Y on diskreetti sm, ylläolevissa kaavoissa tarvitaan summia integraalien sijasta, ja identiteetti (7.5) pysyy voimassa.

Jos funktio g on muotoa

$$g(x, y) = g_1(x)\, g_2(x, y),$$

niin tietenkin

$$E(g_1(X)\, g_2(X, Y) \mid X = x) = g_1(x)\, E(g_2(X, Y) \mid X = x). \quad (7.6)$$

Tämä voidaan ilmaista sanomalla, että tunnetut tekijät saadaan vetää ulos ehdollisesta odotusarvosta.

Valintaa $g(x, y) = y$. vastaava ehdollinen odotusarvo on nimeltään Y:n ehdollinen odotusarvo ehdolla $X = x$.

Määritelmä 7.3 (Y:n ehdollinen odotusarvo ehdolla $X = x$; regressiofunktio)**.** Sm:n Y ehdollinen odotusarvo ehdolla $X = x$ on sen ehdollisen jakauman odotusarvo,

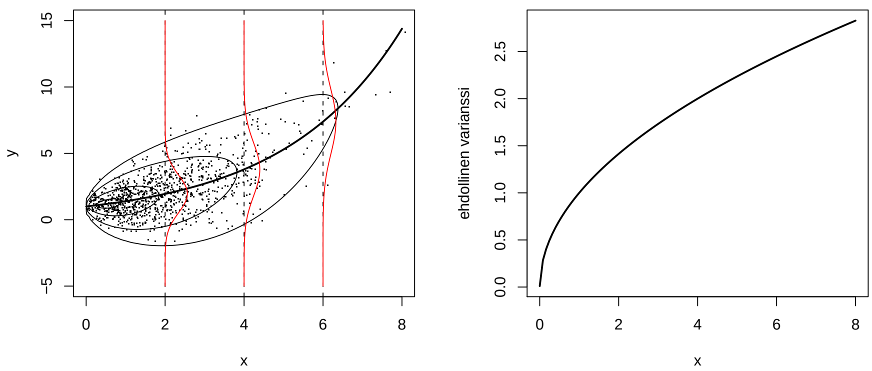
$$E(Y \mid X = x)$$

Funktiota

$$x \mapsto E(Y \mid X = x)$$

kutsutaan Y:n *regressiofunktioksi* X:n suhteen (engl. *regression function of* Y *on* X).

Kuva 7.2 (a) Yhteisjakauma, ehdollisia tiheysfunktioita $y \mapsto f_{Y|X}(y \mid x)$ ja ehdollinen odotusarvo $E(Y \mid X = x)$ (b) ehdollinen varianssi $\operatorname{var}(Y \mid X = x)$.



Jos Y:llä on jatkuva jakauma, on

$$E(Y \mid X = x) = \int y\, f_{Y|X}(y \mid x)\, \mathrm{d}y,$$

ja jos Y:llä on diskreetti jakauma, on

$$E(Y \mid X = x) = \sum_y y\, f_{Y|X}(y \mid x).$$

Kuvassa 7.2 on piirretty ehdollinen odotusarvo $E(Y \mid X = x)$ eli regressiofunktio sekä ehdollinen varianssi $\operatorname{var}(Y \mid X = x)$ kuvan 7.1 yhteisjakaumalle.

Määritelmä 7.4 (Ehdollinen odotusarvo ehdolla satunnaismuuttuja)**.** Merkitään väliaikaisesti

$$m(x) = E(g(X, Y) \mid X = x).$$

Sovitaan, että $m(x) = 0$ niillä x, joilla $m(x)$ se ei muuten tule määritellyksi, eli joilla ehdollinen jakauma $Y \mid (X = x)$ ei ole määritelty. Tämän jälkeen voidaan vapaasti puhua satunnaismuuttujasta $m(X)$. Sitä kutsutaan sm:n $g(X, Y)$:n ehdolliseksi odotusarvoksi ehdolla sm X. Sille käytetään merkintää

$$E(g(X, Y) \mid X) = m(X).$$

Huomautus. Mieleen saattaa juolahtaa merkintä $E(g(X, Y) \mid X = X)$, mutta se on järjetön. Ehdollinen odotusarvo $E(g(X, Y) \mid X)$ on sellainen sm, joka saa arvon $E(g(X, Y) \mid X = x)$ (todennäköisyydellä yksi) silloin, kun X saa arvon x.

Lause 7.2. *Jos $E[g(X, Y)^2] < \infty$, niin $E(g(X, Y) \mid X)$ on keskineliövirheen mielessä paras sm:n $g(X, Y)$ ennuste sm:n X funktion avulla, ts.*

$$E[(g(X, Y) - E(g(X, Y) \mid X))^2] \le E[(g(X, Y) - h(X))^2]$$

valitaan funktio $h : \mathbb{R} \to \mathbb{R}$ miten tahansa.

Todistus. Harjoitustehtävä. □

Erityisesti, jos tehtävänä on ennustaa sm:n Y arvo jollakin sm:n X arvon funktiolla, niin keskineliövirheen mielessä paras mahdollinen ennuste $m(x)$, jossa m on regressiofunktio $m(x) = E[Y \mid X = x]$.

Lause 7.3. *Odotusarvo voidaan laskea iteroituna odotusarvona, eli*

$$Eg(X,Y) = EE(g(X,Y) \mid X),$$

mikäli odotusarvo $Eg(X,Y)$ on olemassa laajennettuna reaalilukuna.

Todistus. Esitetään perustelu diskreetissä tapauksessa.

$$\begin{aligned}
Eg(X,Y) &= \sum_{x,y} g(x,y)\, f_{X,Y}(x,y) = \sum_{x,y} g(x,y)\, f_X(x)\, f_{Y|X}(y \mid x) \\
&= \sum_x f_X(x) \sum_y g(x,y)\, f_{Y|X}(y \mid x) \\
&= EE(g(X,Y) \mid X).
\end{aligned}$$

□

Esimerkki 7.2. Tarkastellaan yhteisjakaumaa

$$\begin{aligned}
X \mid Y &\sim \mathrm{Bin}(Y, \theta), \\
Y &\sim \mathrm{Poi}(\lambda)
\end{aligned}$$

jossa $\lambda > 0$ ja $0 < \theta < 1$. Nyt

$$E(X \mid Y) = Y\,\theta,$$

joten

$$EX = EE(X \mid Y) = E(Y\,\theta) = \theta EY = \theta\lambda.$$

△

Kuten ehdollinen odotusarvo, myös ehdollinen varianssi voidaan laskea ehtona satunnaismuuttuja X (eikä ehdolla sen arvo $X = x$). Ensin määritellään funktio

$$v(x) = \mathrm{var}(g(X,Y) \mid X = x),$$

ja määritelmää jatketaan koko reaaliakselille sopimalla, että $v(x) = 0$ niillä argumenteilla, joilla ehdollinen jakauma $Y \mid (X = x)$ ei ole luonnostaan määritelty. Tämän jälkeen määritellään, että

$$\mathrm{var}(g(X,Y) \mid X) = v(X).$$

Lause 7.4. *Sm:n varianssi on yhtä kuin sen ehdollisen varianssin odotusarvon sekä ehdollisen odotusarvon varianssin summa, eli*

$$\mathrm{var}(g(X,Y)) = E\,\mathrm{var}(g(X,Y) \mid X) + \mathrm{var}\,E(g(X,Y) \mid X). \tag{7.7}$$

Todistus. Harjoitustehtävä. □

Huomautus. Mittateoriaan pohjautuvassa todennäköisyysteoriassa ehdollinen odotusarvo $E(g(X,Y) \mid X)$ määritellään täysin erilaisella tekniikalla, kuin millä me sen teemme. Lisäksi oppikirjoissa se tavallisesti määritellään vain siinä tapauksessa, jossa $g(X,Y)$ on integroituva eli kun $Eg(X,Y) \in \mathbb{R}$, mikä on yhtäpitävää sen kanssa, että $E|g(X,Y)| < \infty$. Integroituvuus on tässä kuitenkin tarpeeton rajoitus; kvasi-integroituvuus, eli odotusarvon $Eg(X,Y)$ olemassaolo laajennettuna reaalilukuna on riittävää mielekkään ehdollisen odotusarvon käsitteen olemassaololle. Tätä laajennettua teoriaa on hieman hankala löytää oppikirjoista, mutta se esitellään esim. Ashin [1], tai Chown ja Teicherin [2] tai Jacodin ja Protterin teoksessa [3]. Tämän laajennetun teorian perusteella satunnaismuuttujan $g(X,Y)$ integroituvuuden saa tarkistaa laskemalla odotusarvon $E|g(X,Y)|$ kaavalla $EE(|g(X,Y)| \mid X)$. Mikäli tulos on äärellinen, niin sitten $g(X,Y)$ on integroituva, eli $Eg(X,Y)$ on reaaliluku, jonka puolestaan saa laskea kaavalla $EE(g(X,Y) \mid X)$. Integraalin äärellisyyden tarkistamisessa voidaan siis soveltaa samaa tekniikkaa kuin mitä käytetään Fubinin lauseen kohdalla.

Kirjallisuutta

[1] Robert B. Ash. *Probability and Measure Theory.* Academic Press, 2nd edition, 2000.

[2] Y. S. Chow and H. Teicher. *Probability Theory: Independence, Interchangeability, Martingales.* Springer-Verlag, 2nd edition, 1988.

[3] Jean Jacod and Philip Protter. *Probability Essentials.* Springer, 2nd edition, 2002.

7.5 Yhteisjakauman määrittely kertolaskukaavan avulla

Tilastollisia malleja spesifioidaan usein kertomalla, mikä on yhden sm:n reunajakauma ja toisen ehdollinen jakauma. Tällöin voidaan puhua *hierarkkisesta mallista.* Tarkastellaan tästä esimerkkejä.

Esimerkki 7.3. (Diskreetti ja diskreetti.) Hyönteisen munimien munien lkm $Y \sim \mathrm{Poi}(\lambda)$, jossa $\lambda > 0$ on jakauman odotusarvo. Kukin munista kehittyy toukaksi toisistaan riippumatta tn:llä $0 < \theta < 1$. Olkoon X toukaksi kehittyvien munien lkm. Tällöin

$$X \mid (Y = y) \sim \mathrm{Bin}(y, \theta).$$

Yhteisjakauma on diskreetti, ja sen yptnf on

$$f_{X,Y}(x,y) = f_Y(y)\, f_{X|Y}(x \mid y) = \mathrm{e}^{-\lambda} \frac{\lambda^y}{y!} \binom{y}{x} \theta^x (1-\theta)^{y-x}, \qquad 0 \le x \le y,$$

jossa x ja y ovat kokonaislukuja 0, 1, 2, ...

Tämä malli voitaisiin määritellä myös sanomalla, että

$$\begin{aligned} X \mid Y &\sim \mathrm{Bin}(Y, \theta), \\ Y &\sim \mathrm{Poi}(\lambda) \end{aligned}$$

jossa $\lambda > 0$ ja $0 < \theta < 1$ ovat vakioita.

Tässä mallissa voidaan esim. kysyä, mikä on X:n reunajakauma. Sen ptnf voidaan esittää summana

$$f_X(x) = \sum_y f_{X,Y}(x, y).$$

Joidenkin laskujen jälkeen havaitaan, että reunajakauma on $X \sim \text{Poi}(\lambda\theta)$. $\triangle$

Esimerkki 7.4. (Jatkuva ja diskreetti.) Olkoon sm:lla Θ betajakauma $\text{Be}(\alpha, \beta)$, jossa $\alpha, \beta > 0$ ovat vakioita. Kolikkoa heitetään n kertaa, ja lasketaan kuinka monta kertaa saadaan klaava, kun klaavan todennäköisyys on Θ. Ehdolla $\Theta = \theta$ klaavojen lukumäärällä X on jakauma $\text{Bin}(n, \theta)$. Tällöin yhteisjakauma voidaan esittää funktiolla

$$f_{\Theta,X}(\theta, x) = \frac{1}{B(\alpha, \beta)}\, \theta^{\alpha-1}\,(1-\theta)^{\beta-1} \binom{n}{x}\, \theta^x\,(1-\theta)^{n-x},$$

jossa $0 < \theta < 1$ ja $x = 0, 1, \ldots, n$.

Tämä malli voitaisiin spesifioida myös sanomalla, että

$$\begin{aligned} X \mid \Theta &\sim \text{Bin}(n, \Theta), \\ \Theta &\sim \text{Be}(\alpha, \beta) \end{aligned}$$

jossa $\alpha, \beta > 0$ ja $n \geq 0$ ovat vakioita.

Mikä on sm:n Θ jakauma, kun havaitaan, että $X = x$? Thomas Bayes esitti 1700-luvulla ratkaisun tähän kysymykseen. Me ratkaisemme kysymyksen tarkastelemalla hetken yhteisjakauman esitystä muuttujan θ funktiona, minkä jälkeen on selvää, että

$$\Theta \mid (X = x) \sim \text{Be}(\alpha + x, \beta + n - x).$$

Tämä lasku on esimerkki *Bayes-päättelystä* (eng. *Bayesian inference*). Tässä binomijakauman parametria pidettiin satunnaismuuttujana, jolla oli tietty priorijakauma (lat. *a priori*, ennen [havaintoa]). Havainnon jälkeen parametrilla on sen ehdollinen jakauma ehdolla havainto, ja tätä jakaumaa kutsutaan parametrin posteriorijakaumaksi (lat. *a posteriori*, [havainnon] jälkeen). Tässä esimerkissä kävi niin onnellisesti, että sekä priorijakauma että posteriorijakauma ovat samassa jakaumaperheessä: ne ovat molemmat beetajakaumia. Tällöin sanotaan, että sekä priori että posteriori ovat *konjugaatti-* eli *liittoperheessä* (engl. *conjugate family*) tarkasteltavan uskottavuusfunktion suhteen. Asia voidaan ilmaista myös sanomalla, että beetajakauma on binomiuskottavuuden *liitto-* eli *konjugaattipriori.* $\triangle$

Esimerkki 7.5. (Jatkuva ja jatkuva.) Useasta ohjelmistoista löytyy satunnaislukugeneraattori normaalijakaumalle. Tarkastellaan seuraavaa algoritmia, jossa $\sigma_X > 0$, μ_X ja $\sigma_Z > 0$ ovat annettuja lukuja ja m on jokin funktio, joka palauttaa reaaliluvun reaalilukuargumentilla.

1. Simuloi $X \sim N(\mu_X, \sigma_X^2)$.
2. Simuloi $Z \sim N(0, \sigma_Z^2)$.
3. Aseta $Y = m(X) + Z$.

Kuvaile sm:ien X ja Y yhteisjakauma.

Huomautus. Kun satunnaislukugeneraattoria kutsutaan monta kertaa (kuten edellä askelissa 1 ja 2), niin eri kerroilla palautettavia lukuja voidaan pitää keskenään riippumattomien satunnaismuuttujien arvoina, sillä todelliset satunnaislukugeneraattorit toimivat tällä tavoin. Tätä ominaisuutta pidetään kirjallisuudessa niin itsestään selvänä asiana, että sitä harvoin vaivaudutaan selittämään.

Ratkaisu: Yhteisjakauma voidaan esittää kaavoilla

$$\begin{aligned} Y \mid X &\sim N(m(X), \sigma_Z^2) \\ X &\sim N(\mu_X, \sigma_X^2). \end{aligned}$$

Ehdollinen jakauma $[Y \mid X = x]$ nähdään siitä, että simuloinnissa $X \perp\!\!\!\perp Z$, joten X:n arvolla x ehdollistaminen ei muuta Z:n jakaumaa. Kyseisessä ehdollisessa jakaumassa sm:n X arvo on vakio x, ja lopuksi vakion x:n tilalle kirjoitetaan satunnaismuuttuja X.

Koska X:n jakauma on jatkuva, ja Y:n ehdollinen jakauma ehdolla $X = x$ on jatkuva kaikilla x, on myös yhteisjakauma jatkuva, ja sen ytf on

$$\begin{aligned} f_{X,Y}(x, y) &= f_X(x)\, f_{Y|X}(y \mid x) \\ &= \frac{1}{\sigma_X\, \sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}\right)\, \frac{1}{\sigma_Z\, \sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(y - m(x))^2}{\sigma_Z^2}\right) \end{aligned}$$

Huomioita: X:n reunajakauma on $N(\mu_X, \sigma_X^2)$ ja regressiofunktio on

$$E[Y \mid X = x] = m(x), \qquad \operatorname{var}(Y \mid X = x) = \sigma_Z^2,$$

mutta esim. Y:n reunajakauma ei ole mikään tuttu jakauma, ellei regressiofunktion m muotoa rajoiteta. △

Esimerkki 7.6. (Jatkoa edelliselle esimerkille.) Valitaan regressiofunktiolle m lineaarinen muoto. Tällöin $m(X)$ antaa keskineliövirheen mielessä parhaan ennusteen satunnaismuuttujan Y arvolle. Koska $m(X)$ on lineaarinen, sen täytyy myös olla keskineliövirheen mielessä paras lineaarinen ennuste, joten jakson 6.9 kaavojen mukaan täytyy olla

$$m(x) = \mu_Y + \rho\, \frac{\sigma_Y}{\sigma_X}\, (x - \mu_X),$$

jossa μ_Y ja $\sigma_Y > 0$ (oletus) ovat sm:n Y odotusarvo ja keskihajonta ja $-1 < \rho < 1$ (oletus) on X:n ja Y:n korrelaatiokerroin. Lisäksi pätee

$$\sigma_Z^2 = \sigma_Y^2\, (1 - \rho^2).$$

Pitkähköjen mutta suoraviivaisten laskujen jälkeen ytf voidaan kirjoittaa muotoon

$$f_{X,Y}(x, y) = \frac{1}{2\pi}\, \frac{1}{\sqrt{\det \mathbf{C}}} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{z} - \boldsymbol{\mu})\right), \quad \mathbf{z} = \begin{bmatrix} x \\ y \end{bmatrix}$$

jossa

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \qquad \mathbf{C} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

Tässä $\boldsymbol{\mu}$ on satunnaisvektorin (X, Y) odotusarvovektori, ja $\mathbf{C}$ sen kovarianssimatriisi.

Johdettu jakauma on kaksiulotteinen normaalijakauma parametreilla $\boldsymbol{\mu}$ ja $\mathbf{C}$, mikä merkitään

$$(X, Y) \sim N(\boldsymbol{\mu}, \mathbf{C}).$$

Käsittelemme myöhemmin tarkemmin moniulotteista normaalijakaumaa eli multinormaalijakaumaa, joka on normaalijakauman yleistys mielivaltaiseen dimensioon. $\triangle$

Luku 8

Moniulotteiset jakaumat

Moniulotteinen satunnaisvektori on muuten samanlainen kuin kaksiulotteinen satunnaisvektori, paitsi että sillä voi olla useampi kuin kaksi komponenttia. Sen jakaumaa kuvaillaan samaan tapaan kuin kaksiulotteisen satunnaisvektorin jakaumaa. Tämän takia suurin osa tämän luvun ideoista on ennestään tuttuja, eikä tuttuja tuloksia enää perustella uudestaan. Tilastotieteilijä tarvitsee moniulotteisia jakaumia sen takia, että hän voisi määritellä ja analysoida tilastollisia malleja.

Tässä luvussa uutta on se, että reunajakauma voidaan määritellä mille tahansa komponenttien osajoukolle, ja ehdollisia jakaumia voidaan laskea ehdollistamalla millä tahansa komponenttien osajoukolla. Ehdollinen riippumattomuus on käsite, joka ei ole mielekäs kuin vasta dimensioissa kolmesta eteenpäin.

Moniulotteisten jakaumien havainnollistaminen on hankalaa. Kuvien piirtäminen ei onnistu korkeissa dimensioissa, vaan niiden sijasta pitää luottaa kaavojen manipulointiin.

8.1 Satunnaisvektori

Avaruuden $\mathbb{R}^n$ vektoria $(x_1, \ldots, x_n)$ merkitään $\mathbf{x} = (x_1, \ldots, x_n)$. Silloin, kun se esiintyy matriisien kanssa samassa lausekkeessa, se ymmäretään pystyvektoriksi eli $n \times 1$-matriisiksi,

$$\mathbf{x} = (x_1, \ldots, x_n) = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Määritelmä 8.1. Jos $X_1, \ldots, X_n$ ovat samalla perusjoukolla Ω määriteltyjä satunnaismuuttujia, niin $\mathbf{X} = (X_1, \ldots, X_n)$ on n-ulotteinen satunnaisvektori (lyh. sv). Se on kuvaus $\Omega \to \mathbb{R}^n$ siten, että

$$\mathbf{X}(\omega) = (X_1(\omega), \ldots, X_n(\omega)) = \begin{bmatrix} X_1(\omega) \\ \vdots \\ X_n(\omega) \end{bmatrix}$$

Määritelmä 8.2 (Yhteiskertymäfunktio)**.** Satunnaismuuttujien $X_1, \ldots, X_n$ yhteiskertymäfunktio (ykf) eli satunnaisvektorin $\mathbf{X} = (X_1, \ldots, X_n)$ kertymäfunktio

on
$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = F_{\mathbf{X}}(\mathbf{x}) = P(X_1 \le x_1,\ldots,X_n \le x_n),$$
jossa $\mathbf{x} = (x_1,\ldots,x_n) \in \mathbb{R}^n$.

Satunnaisvektorin jakauma määritellään samoin kuin aikaisemmin, eli se on funktio
$$B \mapsto P(\mathbf{X} \in B), \qquad B \subset \mathbb{R}^n .$$
Kuten kaksiulotteisessa tapauksessa, ykf määrää jakauman myös dimensiossa n. Korkeissa dimensioissa ykf:a käytetään harvoin konkreettisissa laskuissa, vaan se on pikemminkin teoreettinen apuväline.

Jos kaikki satunnaisvektorin $\mathbf{X}$ komponentit X_i ovat diskreettejä sm:ia, niin sen ptnf eli sm:ien X_i yhteispistetodennäköisyysfunktio (yptnf) on
$$f_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}) = P(X_1 = x_1,\ldots,X_n = x_n). \tag{8.1}$$
Jos $g(\mathbf{X})$ on sv:n $\mathbf{X}$ reaaliarvoinen muunnos, niin sen odotusarvo saadaan kaavalla
$$Eg(\mathbf{X}) = \sum_{\mathbf{x}} g(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x}), \tag{8.2}$$
mikäli kyseinen summa suppenee itseisesti.

Satunnaisvektorilla $\mathbf{X}$ on jatkuva jakauma, mikäli sillä on tiheysfunktio $f_{\mathbf{X}}$, mikä tarkoittaa sitä, että
$$P(\mathbf{X} \in B) = \int_B f_{\mathbf{X}} = \int_B f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x}, \qquad \text{kaikilla } B \subset \mathbb{R}^n. \tag{8.3}$$
Tällöin sen komponenteilla $X_1,\ldots,X_n$ on jatkuva yhteisjakauma, ja tiheysfunktiota
$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{\mathbf{X}}(\mathbf{x}), \qquad \mathbf{x} = (x_1,\ldots,x_n)$$
kutsutaan sm:ien $X_1,\ldots,X_n$ yhteistiheysfunktioksi. Edellä kyseessä on n-kertainen integraali,
$$\int_B f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int \cdots \int_B f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}x_1 \cdots \mathrm{d}x_n,$$
joka voidaan laskea iteroituna integraalina käyttämällä mielivaltaista integrointijärjestystä.

Jos $g(\mathbf{X})$ on sv sv:n $\mathbf{X}$ reaaliarvoinen muunnos, niin sen odotusarvo saadaan kaavalla
$$Eg(\mathbf{X}) = \int g(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{8.4}$$
mikäli kyseinen integraali suppenee itseisesti. Tällöin integraali voidaan laskea iteroituna integraalina käyttäen mielivaltaista integrointijärjestystä. Kun integrointijoukkoa ei merkitä näkyviin, tarkoitetaan koko avaruuden (tässä $\mathbb{R}^n$) yli laskettua integraalia. (Jos dimensio $n \ge 2$, niin tämä merkintä ei voi sekaantua määrämättömän integraalin (engl. *indefinite integral*) merkinnän kanssa, sillä määräämättömän integraalin käsitettä käytetään vain dimensiossa yksi.)

Jatkuvan (yhteis)jakauman tapauksessa ykf on
$$F_{\mathbf{X}}(\mathbf{x}) = \int_{s_1=-\infty}^{x_1} \cdots \int_{s_n=-\infty}^{x_n} f_{\mathbf{X}}(s_1,\ldots,s_n)\,\mathrm{d}s_1 \cdots \mathrm{d}s_n.$$

Kun tätä funktiota derivoidaan osittain jokaisen argumenttinsa suhteen, nähdään että

$$\frac{\partial^n F_{\mathbf{X}}(x_1, \dots, x_n)}{\partial x_1 \cdots \partial x_n} = f_{\mathbf{X}}(x_1, \dots, x_n),$$

ainakin niissä pisteissä $(x_1, \dots, x_n)$, joissa $f_{\mathbf{X}}$ on jatkuva. Voidaan osoittaa, että tämä funktio kelpaa jatkuvasti jakautuneen sv:n tiheysfunktioksi (kun se jatketaan mielivaltaisesti niissä pisteissä, joissa ko. derivaatta ei ole olemassa).

Jos $\mathbf{X}$ on diskreetti sv, ja $\mathbf{Y}$ on jatkuvasti jakautunut sv, ja ne on määritelty samalla perusjoukolla Ω, niin silloin niiden yhteisjakauma voidaan esittää funktiolla

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})\, f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}),$$

jossa $f_{\mathbf{X}}$ on sv:n $\mathbf{X}$ reunajakauman (yhteis-)ptnf, ja $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})$ on sv:n $\mathbf{Y}$ ehdollinen (yhteis-)tf ehdolla $\mathbf{X} = \mathbf{x}$. Jos $g(\mathbf{X}, \mathbf{Y})$ on reaaliarvoinen sm, niin sen odotusarvo saadaan kaavalla

$$Eg(\mathbf{X}, \mathbf{Y}) = \sum_{\mathbf{x}} \int g(\mathbf{x}, \mathbf{y})\, f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})\, \mathrm{d}\mathbf{y}, \tag{8.5}$$

mikäli kyseinen odotusarvo on olemassa.

Kahden satunnaisvektorin $\mathbf{X}$ ja $\mathbf{Y}$ riippumattomuus määritellään täysin vastaavasti kuin satunnaismuuttujien kohdalla, ks. jakso 2.9. Samalla päättelyllä kuin aikaisemmin,

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \quad \Rightarrow \quad g(\mathbf{X}) \perp\!\!\!\perp h(\mathbf{Y}),$$

kun g ja h ovat mielivaltaisia vektoriargumentin reaaliarvoisia funktioita.

8.2 Odotusarvovektori ja kovarianssimatriisi

Määritelmä 8.3. Asetamme seuraavat määritelmät, mikäli kyseessä olevat odotusarvot ovat olemassa.

- Satunnaisvektorin $\mathbf{X} = (X_1, \dots, X_n)$ *odotusarvo(vektori)* on n-komponenttinen vakiovektori

$$E\mathbf{X} = E(\mathbf{X}) = (EX_1, \dots, EX_n) = \begin{bmatrix} EX_1 \\ \vdots \\ EX_n \end{bmatrix}.$$

- Dimensioltaan $m \times n$ olevan satunnaismatriisin $\mathbf{Z} = [Z_{ij}]$ odotusarvo(matriisi) on se $m \times n$-vakiomatriisi, jonka alkio (i, j) on alkion Z_{ij} odotusarvo, ts.

$$E\mathbf{Z} = E \begin{bmatrix} Z_{11} & Z_{12} & \dots & Z_{1n} \\ Z_{21} & Z_{22} & \dots & Z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{m1} & Z_{m2} & \dots & Z_{mn} \end{bmatrix} = \begin{bmatrix} EZ_{11} & EZ_{12} & \dots & EZ_{1n} \\ EZ_{21} & EZ_{22} & \dots & EZ_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ EZ_{m1} & EZ_{m2} & \dots & EZ_{mn} \end{bmatrix}.$$

- Satunnaisvektorien $\mathbf{X} = (X_1, \ldots, X_n)$ ja $\mathbf{Y} = (Y_1, \ldots, Y_k)$ (välinen) *kovarianssi* on $n \times k$-matriisi

$$\begin{aligned}\operatorname{cov}(\mathbf{X}, \mathbf{Y}) &= E[(\mathbf{X} - E\mathbf{X})(\mathbf{Y} - E\mathbf{Y})^T] \\ &= \begin{bmatrix} \operatorname{cov}(X_1, Y_1) & \operatorname{cov}(X_1, Y_2) & \ldots & \operatorname{cov}(X_1, Y_k) \\ \operatorname{cov}(X_2, Y_1) & \operatorname{cov}(X_2, Y_2) & \ldots & \operatorname{cov}(X_2, Y_k) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}(X_n, Y_1) & \operatorname{cov}(X_n, Y_2) & \ldots & \operatorname{cov}(X_n, Y_k) \end{bmatrix}.\end{aligned}$$

  Mikäli $\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}$, niin sanotaan, että satunnaisvektorit $\mathbf{X}$ ja $\mathbf{Y}$ *eivät korreloi* eli että ne ovat *korreloimattomat* (engl. *uncorrelated*).

- Satunnaisvektorin $\mathbf{X} = (X_1, \ldots, X_n)$ *kovarianssimatriisi* on $n \times n$-matriisi

$$\begin{aligned}\operatorname{Cov}(\mathbf{X}) = \operatorname{cov}(\mathbf{X}, \mathbf{X}) &= E[(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^T] \\ &= \begin{bmatrix} \operatorname{cov}(X_1, X_1) & \operatorname{cov}(X_1, X_2) & \ldots & \operatorname{cov}(X_1, X_n) \\ \operatorname{cov}(X_2, X_1) & \operatorname{cov}(X_2, X_2) & \ldots & \operatorname{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}(X_n, X_1) & \operatorname{cov}(X_n, X_2) & \ldots & \operatorname{cov}(X_n, X_n) \end{bmatrix}.\end{aligned}$$

Määritelmästä seuraa suoraan, että

$$E(\mathbf{Z}^T) = (E\mathbf{Z})^T. \tag{8.6}$$

Laskusääntö

$$E[\mathbf{AZB} + \mathbf{C}] = \mathbf{A}(E\mathbf{Z})\mathbf{B} + \mathbf{C}, \tag{8.7}$$

on voimassa kun $\mathbf{Z}$ on satunnaismatriisi ja $\mathbf{A}$, $\mathbf{B}$ ja $\mathbf{C}$ ovat vakiomatriiseja, joiden dimensiot ovat sellaiset, että lauseke on määritelty. Ts. vakiomatriisit saadaan vetää ulos odotusarvosta, jos ne sijaitsevat matriisitulossa äärimmäisenä vasemmalla tai äärimmäisenä oikealla.

Mikäli $\mathbf{X}$ koostuu osavektoreista $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$, ja $\mathbf{G}(\mathbf{Y})$ ja $\mathbf{H}(\mathbf{Z})$ ovat matriisiarvoisia lausekkeita, niin

$$\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \quad \Rightarrow \quad \mathbf{G}(\mathbf{Y}) \perp\!\!\!\perp \mathbf{H}(\mathbf{Z}). \tag{8.8}$$

(Koska tämä kaava pätee matrisiarvoisille funktioille, se pätee tietenkin myös vektori- tai skalaariarvoisille funktioille.) Toisin sanoen *riippumattomien satunnaisvektorien funktiot ovat keskenään riippumattomia.*

Tässä matriisilausekkeiden riippumattomuus tarkoittaa tietenkin sitä, että kaikki matriisin $\mathbf{G}(\mathbf{Y})$ komponentit ovat riippumattomia kaikista matriisin $\mathbf{H}(\mathbf{Z})$ komponenteista. Riippumattomuudesta ja matriisikertolaskun määritelmästä seuraa, että

$$\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \quad \Rightarrow \quad E[\mathbf{G}(\mathbf{Y})\,\mathbf{H}(\mathbf{Z})] = E[\mathbf{G}(\mathbf{Y})]\,E[\mathbf{H}(\mathbf{Z})] \tag{8.9}$$

mikäli dimensiot ovat matriisitulossa $\mathbf{G}(\mathbf{Y})\,\mathbf{H}(\mathbf{Z})$ yhteensopivat ja mikäli odotusarvot ovat olemassa.

Lause 8.1 (Kovarianssin ominaisuuksia)**.**

(a) Jos $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$*, ja* $\mathbf{X}$ *on* m*-komponenttinen ja* $\mathbf{Y}$ *on* n*-komponenttinen, niin*

$$\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}_{m\times n}.$$

(b) Sv:ien $\mathbf{X}$ *ja* $\mathbf{Y}$ *kovarianssi voidaan laskea myös kaavalla*

$$\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = E[\mathbf{X}\mathbf{Y}^T] - (E\mathbf{X})(E\mathbf{Y})^T.$$

(c) Jos $\mathbf{v} \in \mathbb{R}^m$ *on vakiovektori, ja* $\mathbf{X}$ *on* n*-komponenttinen sv, niin*

$$\operatorname{cov}(\mathbf{v}, \mathbf{X}) = \mathbf{0}_{m\times n}, \qquad \operatorname{cov}(\mathbf{X}, \mathbf{v}) = \mathbf{0}_{n\times m}.$$

(d) $\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = \operatorname{cov}(\mathbf{Y}, \mathbf{X})^T$.

(e) Jos $\mathbf{X}_1$*,* $\mathbf{X}_2$ *ja* $\mathbf{Y}$ *ovat satunnaisvektoreita, niin*

$$\begin{aligned}\operatorname{cov}(\mathbf{X}_1 + \mathbf{X}_2, \mathbf{Y}) &= \operatorname{cov}(\mathbf{X}_1, \mathbf{Y}) + \operatorname{cov}(\mathbf{X}_2, \mathbf{Y}),\\ \operatorname{cov}(\mathbf{Y}, \mathbf{X}_1 + \mathbf{X}_2) &= \operatorname{cov}(\mathbf{Y}, \mathbf{X}_1) + \operatorname{cov}(\mathbf{Y}, \mathbf{X}_2).\end{aligned}$$

(f) Jos $\mathbf{A}$ *ja* $\mathbf{B}$ *ovat vakiomatriiseja ja* $\mathbf{v}$ *ja* $\mathbf{w}$ *ovat vakiovektoreja, niin*

$$\operatorname{cov}(\mathbf{A}\mathbf{X} + \mathbf{v}, \mathbf{B}\mathbf{Y} + \mathbf{w}) = \mathbf{A}\,\operatorname{cov}(\mathbf{X}, \mathbf{Y})\,\mathbf{B}^T. \tag{8.10}$$

Erityisesti

$$\operatorname{Cov}(\mathbf{A}\mathbf{X} + \mathbf{v}) = \mathbf{A}\operatorname{Cov}(\mathbf{X})\mathbf{A}^T. \tag{8.11}$$

Todistus. Kaikki kohdat seuraavat helpoilla laskuilla kovarianssin määritelmästä sekä laskusäännoistä (8.7) ja (8.9). □

Huomaa, että satunnaisvektorin kovarianssimatriisi on symmetrinen ja että sen päälävistäjällä on komponenttisatunnaismuuttujien varianssit.

Palautetaan mieleen, että neliömatriisi $\mathbf{B}$ on

- *säännöllinen* eli *kääntyvä* eli *ei-singulaarinen*, jos sillä on olemassa käänteismatriisi $\mathbf{B}^{-1}$;
- *symmetrinen*, jos $\mathbf{B}^T = \mathbf{B}$;
- *positiivisesti semidefiniitti*, jos $\mathbf{v}^T\mathbf{B}\mathbf{v} \ge 0$ kaikilla $\mathbf{v}$;
- *positiivisesti definiitti*, jos $\mathbf{v}^T\mathbf{B}\mathbf{v} > 0$ kaikilla $\mathbf{v} \ne 0$.

Lause 8.2. *Sv:n* $\mathbf{X}$ *kovarianssimatriisi on symmetrinen ja positiivisesti semidefiniitti matriisi. Jos se ei ole positiivisesti definiitti, niin sv:n* $\mathbf{X}$ *komponenttien välillä on lineaarisia sidosehtoja, eli* $\mathbf{X}$ *saa todennäköisyydellä yksi arvoja vain tietyltä kiinteältä hypertasolta.*

Todistus. Symmetrisyys seuraa kovarianssioperaattorin symmetrisyydestä. Positiivinen semidefiniittisyys seuraa siitä, että

$$\begin{aligned}\mathbf{v}^T\operatorname{Cov}(\mathbf{X})\mathbf{v} &= \mathbf{v}^T E[(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^T]\,\mathbf{v}\\ &= E[\mathbf{v}^T(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^T\mathbf{v}] = E\left[\left(\mathbf{v}^T(\mathbf{X} - E\mathbf{X})\right)^2\right] \ge 0,\end{aligned}$$

sillä ei-negatiivisen sm:n odotusarvo on ei-negatiivinen. Jos kovarianssimatriisi ei ole positiivisesti definiitti, niin on olemassa $\mathbf{v} \neq \mathbf{0}$ siten, että

$$0 = \mathbf{v}^T \operatorname{Cov}(\mathbf{X})\, \mathbf{v} = E\left[\left(\mathbf{v}^T(\mathbf{X} - E\mathbf{X})\right)^2\right],$$

joten todennäköisyydellä yksi on

$$\mathbf{v}^T(\mathbf{X} - E\mathbf{X}) = 0. \qquad \square$$

Huomautus. Jos $\operatorname{Cov} \mathbf{X}$ ei ole positiivisesti definiitti, niin $\mathbf{X}$:llä ei voi olla jatkuva jakauma, sillä hypertaso on alempiulotteisena pintana nollamittainen.

8.3 Ehdolliset jakaumat, kertolaskusääntö ja ehdollinen odotusarvo

Tarkastellaan sv:ia $\mathbf{Z}$, jonka r ensimmäistä koordinaattia muodostavat sv:n $\mathbf{X}$ ja loput s koordinaattia sv:n $\mathbf{Y}$,

$$\mathbf{Z} = (\mathbf{X}, \mathbf{Y}) = (X_1, \ldots, X_r, Y_1, \ldots, Y_s).$$

Oletetaan, että sv:n $\mathbf{Z}$ yhteisjakauman esittää tiheys

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}), \qquad \mathbf{z} = (\mathbf{x}, \mathbf{y}).$$

Nyt sallimme sen tilanteen, että jotkut $\mathbf{Z}$:n koordinaatit ovat diskreettejä sm:ia ja lopuilla on jatkuva yhteisjakauma. Tällöin $\mathbf{X}$:n reunajakauman esittää tiheys $f_{\mathbf{X}}$, joka saadaan summaamalla pois tiheydestä $f_{\mathbf{X},\mathbf{Y}}$ sv:n $\mathbf{Y}$ diskreetit komponentit ja integroimalla pois sen jatkuvat komponentit. Merkintöjä väärinkäyttämällä tämän idean voi esittää kaavalla

$$f_{\mathbf{X}}(\mathbf{x}) = \int f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})\, \mathrm{d}\mathbf{y}, \tag{8.12}$$

missä integraalimerkki tarkoittaa summaamista diskreettien komponenttien kohdalla. Tässä yhteydessä voidaan sanoa, että $\mathbf{X}$:n jakauma marginalisoidaan yhteisjakaumasta. Vastaavasti $f_{\mathbf{Y}}$ saadaan kaavalla

$$f_{\mathbf{Y}}(\mathbf{y}) = \int f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})\, \mathrm{d}\mathbf{x}.$$

Se tilanne, missä pitää esittää mielivaltaiselle osalle komponenteista reunajakauma saadaan aina palautettua edelliseen tilanteeseen permutoimalla koordinaatteja.

Esimerkki 8.1. Olkoot U ja V diskreettejä sm:ia ja (X, Y) jatkuvasti jakautunut sv. Tällöin esim. sm:ien V ja Y yhteisjakauman esittää tiheys

$$f_{V,Y}(v, y) = \sum_u \int f_{U,V,X,Y}(u, v, x, y)\, \mathrm{d}x,$$

sillä (permutoidaan V ja Y ensimmäisiksi)

$$f_{V,Y,U,X}(v, y, u, x) = f_{U,V,X,Y}(u, v, x, y),$$

ja (summataan tai integroidaan muut pois)

$$f_{V,Y}(v,y) = \sum_u \int f_{V,Y,U,X}(v,y,u,x)\,\mathrm{d}x.$$

△

Sv:n $\mathbf{Y}$ ehdollisen jakauman ehdolla $\mathbf{X} = \mathbf{x}$ esittää tiheys

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}) = \frac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})},$$

kun $f_{\mathbf{X}}(\mathbf{x}) > 0$. Tarvittaessa tämä lauseke voidaan määritellä esim. nollaksi niillä $\mathbf{x}$, joilla $f_{\mathbf{X}}(\mathbf{x}) = 0$. Vastaavasti määritellään

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} \mid \mathbf{y}) = \frac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})}.$$

Kertolaskukaava (eli ketjusääntö) on voimassa, eli

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})\, f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}) = f_{\mathbf{Y}}(\mathbf{y})\, f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} \mid \mathbf{y}).$$

Kertolaskukaavaa voidaan iteroida. Oletetaan, että sv $\mathbf{X} = (\mathbf{U}, \mathbf{V})$. Tällöin

$$f_{\mathbf{U},\mathbf{V}}(\mathbf{u},\mathbf{v}) = f_{\mathbf{U}}(\mathbf{u})\, f_{\mathbf{V}|\mathbf{U}}(\mathbf{v} \mid \mathbf{u}),$$

minkä takia

$$f_{\mathbf{U},\mathbf{V},\mathbf{Y}}(\mathbf{u},\mathbf{v}) = f_{\mathbf{U}}(\mathbf{u})\, f_{\mathbf{V}|\mathbf{U}}(\mathbf{v} \mid \mathbf{u})\, f_{\mathbf{Y}|\mathbf{U},\mathbf{V}}(\mathbf{y} \mid \mathbf{u},\mathbf{v}).$$

Tätä prosessia voidaan jatkaa, kunnes ollaan päästy skalaarikomponentteihin asti. Esimerkiksi neljän sm:n U, V, X, Y tiheys voidaan esittää muodossa

$$f_{U,V,X,Y}(u,v,x,y) = f_U(u)\, f_{V|U}(v \mid u)\, f_{X|U,V}(x \mid u,v)\, f_{Y|X,U,V}(y \mid x,u,v).$$

Kertolaskusääntöä voidaan soveltaa yhtä hyvin käyttämällä jotakin muuta sm:ien permutaatiota.

Kertolaskusääntö pätee myös ehdollisille jakaumille. Esim.

$$\begin{aligned} f_{\mathbf{U},\mathbf{V}|\mathbf{Y}}(\mathbf{u},\mathbf{v} \mid \mathbf{y}) &= \frac{f_{\mathbf{U},\mathbf{V},\mathbf{Y}}(\mathbf{u},\mathbf{v},\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})} = \frac{f_{\mathbf{U},\mathbf{Y}}(\mathbf{u},\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})}\, \frac{f_{\mathbf{U},\mathbf{V},\mathbf{Y}}(\mathbf{u},\mathbf{v},\mathbf{y})}{f_{\mathbf{U},\mathbf{Y}}(\mathbf{u},\mathbf{y})} \\ &= f_{\mathbf{U}|\mathbf{Y}}(\mathbf{u} \mid \mathbf{y})\, f_{\mathbf{V}|\mathbf{U},\mathbf{Y}}(\mathbf{v} \mid \mathbf{u},\mathbf{y}). \end{aligned}$$

Ehdollisen jakauman reunajakauman voi laskea marginalisoimalla ehdollista yhteisjakaumaa. Esim.

$$\begin{aligned} f_{\mathbf{U}|\mathbf{Y}}(\mathbf{u} \mid \mathbf{y}) &= \frac{f_{\mathbf{U},\mathbf{Y}}(\mathbf{u},\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})} = \frac{1}{f_{\mathbf{Y}}(\mathbf{y})} \int f_{\mathbf{U},\mathbf{V},\mathbf{Y}}(\mathbf{u},\mathbf{v},\mathbf{y})\,\mathrm{d}\mathbf{v} \\ &= \int f_{\mathbf{U},\mathbf{V}|\mathbf{Y}}(\mathbf{u},\mathbf{v} \mid \mathbf{y})\,\mathrm{d}\mathbf{v}. \end{aligned}$$

Huomautus. Kun jakaumien välisiä yhteyksiä johdetaan tähän tapaan kertolaskukaavan ja marginalisoinnin kautta, niin tiheysfunktioiden alaindeksit jätetään usein kirjoittamatta, koska ne selviävät tiheysfunktion argumenteista. Tämä merkintöjen väärinkäyttö ei johda sekaannukseen, kunhan pidetään mielessä, että esim. $f(x)$, $f(y)$, $f(x \mid y)$ ja $f(y \mid x)$ ovat tyypillisesti kaikki eri funktioita.

Satunnaisvektorin $\mathbf{g}(\mathbf{X},\mathbf{Y})$ ehdollinen odotusarvo(vektori) ehdolla $\mathbf{X} = \mathbf{x}$ määritellään siten, että se on satunnaisvektorin $\mathbf{g}(\mathbf{x},\mathbf{Y})$ odotusarvovektori, kun $\mathbf{Y}$:n jakaumana käytetään ehdollista jakaumaa $f_{\mathbf{Y}|\mathbf{X}}(\cdot \mid \mathbf{x})$. Merkintöjä väärinkäyttämällä

$$E(\mathbf{g}(\mathbf{X},\mathbf{Y}) \mid \mathbf{X} = \mathbf{x}) = \int \mathbf{g}(\mathbf{x},\mathbf{y})\, f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x})\, \mathrm{d}\mathbf{y}.$$

Tässä integraalimerkki voi tarkoittaa summausta joidenkin vektorin $\mathbf{y}$ komponenttien suhteen. Jos ehdollistetaan satunnaisvektorilla $\mathbf{X}$, niin satunnaisvektori $E(\mathbf{g}(\mathbf{X},\mathbf{Y}) \mid \mathbf{X})$ määritellään niin, että se on $\mathbf{m}(\mathbf{X})$, kun

$$\mathbf{m}(\mathbf{x}) = E(\mathbf{g}(\mathbf{X},\mathbf{Y}) \mid \mathbf{X} = \mathbf{x}),$$

ja funktion $\mathbf{m}$ määritelmä jatketaan tarvittaessa nollavektoriksi niillä argumentin arvoilla, jolla se ei luonnostaan ole määritelty.

Satunnaisvektorin $\mathbf{g}(\mathbf{X},\mathbf{Y})$ ehdollinen kovarianssimatriisi ehdolla $\mathbf{X} = \mathbf{x}$ määritellään vastaavasti satunnaisvektorin $\mathbf{g}(\mathbf{x},\mathbf{Y})$ kovarianssimatriisina, kun $\mathbf{Y}$:n jakaumana käytetään ehdollista jakaumaa $f_{\mathbf{Y}|\mathbf{X}}(\cdot \mid \mathbf{x})$, ja sille voidaan käyttää merkintää

$$\operatorname{Cov}(\mathbf{g}(\mathbf{X},\mathbf{Y}) \mid \mathbf{X} = \mathbf{x}).$$

Jos ehdollistetaan satunnaisvektorilla $\mathbf{X}$, niin ehdollinen kovarianssimatriisi on satunnaismatriisi, ja sitä merkitään

$$\operatorname{Cov}(\mathbf{g}(\mathbf{X},\mathbf{Y}) \mid \mathbf{X}).$$

8.4 Ehdollinen riippumattomuus

Satunnaisvektorit $\mathbf{X}$ ja $\mathbf{Y}$ ovat riippumattomia, jos

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})\, f_{\mathbf{Y}}(\mathbf{y}).$$

kaikilla argumenteilla $\mathbf{x},\mathbf{y}$. Vastaavasti satunnaisvektorit $\mathbf{X}_1,\dots,\mathbf{X}_n$ ovat riippumattomia, jos

$$f_{\mathbf{X}_1,\dots,\mathbf{X}_n}(\mathbf{x}_1,\dots,\mathbf{x}_n) = f_{\mathbf{X}_1}(\mathbf{x}_1)\cdots f_{\mathbf{X}_n}(\mathbf{x}_n)$$

kaikilla argumenteilla.

Määritelmä 8.4. Satunnaisvektorit $\mathbf{X}$ ja $\mathbf{Y}$ ovat riippumattomia ehdolla $\mathbf{Z}$, jos ne ovat riippumattomia niiden ehdollisessa yhteisjakaumassa ehdolla $\mathbf{Z} = \mathbf{z}$ kaikilla $\mathbf{z}$, ts. mikäli

$$f_{\mathbf{X},\mathbf{Y}|\mathbf{Z}}(\mathbf{x},\mathbf{y} \mid \mathbf{z}) = f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x} \mid \mathbf{z})\, f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y} \mid \mathbf{z}), \qquad \text{kaikilla } \mathbf{x},\, \mathbf{y} \text{ ja } \mathbf{z} \tag{8.13}$$

Vastaavasti, satunnaisvektorit $\mathbf{X}_1,\dots,\mathbf{X}_n$ ovat riippumattomia ehdolla $\mathbf{Z}$, mikäli

$$f_{\mathbf{X}_1,\dots,\mathbf{X}_n|\mathbf{Z}}(\mathbf{x}_1,\dots,\mathbf{x}_n \mid \mathbf{z}) = \prod_{i=1}^{n} f_{\mathbf{X}_i|\mathbf{Z}}(\mathbf{x}_i \mid \mathbf{z})$$

kaikilla $\mathbf{x}_1,\dots,\mathbf{x}_n$ ja kaikilla $\mathbf{z}$.

Ehdollisesti riippumattomat satunnaisvektorit eivät tyypillisesti ole marginaalisesti riippumattomia, eli riippumattomia niiden (reuna-)yhteisreunajakaumassa. Esim. jos $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y}) \mid \mathbf{Z}$, niin

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = \int f_{\mathbf{X},\mathbf{Y}|\mathbf{Z}}(\mathbf{x},\mathbf{y} \mid \mathbf{z})\, f_{\mathbf{Z}}(\mathbf{z})\, \mathrm{d}\mathbf{z} = \int f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x} \mid \mathbf{z})\, f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y} \mid \mathbf{z})\, f_{\mathbf{Z}}(\mathbf{z})\, \mathrm{d}\mathbf{z},$$

eikä tämä tyypillisesti enää faktoroidu.

8.5 Tilastollisia malleja

Tilastolliseen päättelyyn on kaksi pääasiallista lähestymistapaa: ns. frekventistinen päättely ja Bayes-päättely. Ne molemmat perustuvat uskottavuusfunktion käyttöön. Suurpiirteiseti selittäen kysymyksenasettelu on seuraava.

Meillä on käsillä numeerinen aineisto vektorin $\mathbf{y} = (y_1, \ldots, y_n)$ muodossa. Ennen havaintojen tekoa aineiston arvot ovat epävarmoja (mittausvirheiden, populaation luonnollisen vaihtelun tms. syyn takia). Tämän takia mallinnamme tilanteen niin, että $\mathbf{y}$ on jollakin perusjoukolla määritellyn satunnaisvektorin $\mathbf{Y}$ havaittu arvo, ts.

$$\mathbf{y} = \mathbf{Y}(\omega^{\mathrm{act}}),$$

jossa ω^{act} on todennäköisyysmallissa aktualisoitunut alkeistapaus.

Tyypillisesti vektorin $\mathbf{Y}$ jakauma mallinnetaan parametrisella mallilla, jossa on yksi tai useampia parametreja $\theta_1, \ldots, \theta_p$, jotka kootaan parametrivektoriksi $\boldsymbol{\theta}$. Kun parametrivektorin arvo on kiinnitetty, niin sv:n $\mathbf{Y}$ jakauman esittää tiheys

$$\mathbf{y} \mapsto f(\mathbf{y} \mid \boldsymbol{\theta}).$$

Kiinnostuksen kohteena on sv:n $\mathbf{Y}$ jakauma. Sitä voidaan arvoida, jos ensin arvioidaan eli estimoidaan tuntematonta parametrivektoria $\boldsymbol{\theta}$.

Kun aineisto $\mathbf{y}$ on havaittu, ja havaittua arvoa käytetään funktion $f(\mathbf{y} \mid \boldsymbol{\theta})$ ensimmäisenä argumenttina, niin funktiota

$$\boldsymbol{\theta} \mapsto f(\mathbf{y} \mid \boldsymbol{\theta})$$

kutsutaan uskottavuusfunktioksi (engl. *likelihood function*). Tässä yhteydessä tiheysfunktiosta saatetaan jättää pois kertoimia, jotka eivät riipu parametrivektorista $\boldsymbol{\theta}$, ja silti kyseistä funktiota edelleen kutsutaan uskottavuusfunktioksi.

Ns. klassisessa eli frekventistisessä tilastotieteessä parametrivektoria $\boldsymbol{\theta}$ pidetään tuntemattomana vakiona, josta tiedetään vain, missä joukossa (eli parametriavaruudessa) sen arvot voivat olla. Tällöin merkintään $f(\mathbf{y} \mid \boldsymbol{\theta})$ ei liitetä tulkintaa ehdollisena jakaumana, koska parametrivektorille ei ole olemassa mitään todennäköisyysjakaumaa. Tyypillisempi merkintä tässä tilanteessa olisikin $f(\mathbf{y}; \boldsymbol{\theta})$. Tunnetuin estimointiperiaate on ns. *suurimman uskottavuuden*, eli SU-periaate (engl. *maximum likelihood, ML*), jonka mukaan parametrivektorin parhaana estimaattina pidetään sitä arvoa $\hat{\boldsymbol{\theta}}^{\mathrm{ML}}$ parametriavaruudessa, joka maksimoi uskottavuusfunktion. Sitä kutsutaan suurimman uskottavuuden estimaatiksi (eli SU-estimaatiksi) (engl. *maximum likelihood estimate, MLE*).

Bayes-päättelyssä parametrivektoria pidetään satunnaisvektorin $\boldsymbol{\Theta}$ arvona. Funktio $f(\mathbf{y} \mid \boldsymbol{\theta})$ ja ehdollinen jakauma $f_{\mathbf{Y}|\boldsymbol{\Theta}}(\mathbf{y} \mid \boldsymbol{\theta})$ samastetaan. Bayes-päättelyssä tilastollisessa mallissa tarvitaan uskottavuusfunktion lisäksi reunajakauma vektorille $\boldsymbol{\Theta}$. Tätä kutsutaan priorijakaumaksi. Parametrivektorin ja havaintovektorin yhteisjakauman esittää tiheys

$$(\mathbf{y}, \boldsymbol{\theta}) \mapsto f_{\boldsymbol{\Theta},\mathbf{Y}}(\boldsymbol{\theta}, \mathbf{y}) = f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\, f_{\mathbf{Y}|\boldsymbol{\Theta}}(\mathbf{y} \mid \boldsymbol{\theta}).$$

Kun on saatu havainto $\mathbf{Y} = \mathbf{y}$, niin tilastolliset päätelmät tehdään karakterisoimalla parametrivektorin posteriorijakaumaa, joka on ehdollinen jakauma

$$f_{\boldsymbol{\Theta}|\mathbf{Y}}(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{f_{\boldsymbol{\Theta},\mathbf{Y}}(\boldsymbol{\theta}, \mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})} = \frac{f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\, f_{\mathbf{Y}|\boldsymbol{\Theta}}(\mathbf{y} \mid \boldsymbol{\theta})}{\int f_{\boldsymbol{\Theta}}(\mathbf{t})\, f_{\mathbf{Y}|\boldsymbol{\Theta}}(\mathbf{y} \mid \mathbf{t})\, \mathrm{d}\mathbf{t}}$$

Kummassakin tilastollisen päättelyn lähestymistavassa tarvitaan konkreettinen kaava uskottavuusfunktiolle. Tämän takia tilastotieteilijän pitää osata johtaa moniulotteisten satunnaisvektoreiden tiheyksiä.

Esimerkki 8.2. Tehdas valmistaa komponentteja, jotka ovat joko toimivia tai viallisia. Määritellään

$$Y_i = \begin{cases} 1, & \text{jos } i\text{:s komponentti on viallinen,} \\ 0, & \text{jos } i\text{:s komponentti toimii.} \end{cases}$$

Tässä tilanteessa tilastollinen malli voisi yksinkertaisimmillaan olla seuraava. Kun parametrin arvo on $0 \le \theta \le 1$, niin satunnaismuuttujat Y_i ovat riippumattomia ja samoin jakautuneita, ja niillä on jakauma Bernoulli(θ). ("Onnistuminen" tarkoittaa nyt sitä, että komponentti ei toimi.) Tällöin uskottavuusfunktio on

$$f(y_1, \ldots, y_n \mid \theta) = \prod_{i=1}^{n} \theta^{y_i} (1 - \theta)^{1 - y_i}, \qquad \mathbf{y} = (y_1, \ldots, y_n) \in \{0, 1\}^n.$$

Tässä tilanteessa SU-estimaatti löytyy helposti: se on k/n, jossa $k = \sum y_i$ on viallisten komponenttien lukumäärä.

Bayes-päättelyä harrastava tilastotieteilijä antaisi parametrille Θ priorijakauman, esim. välin $(0, 1)$ tasajakauman Be$(1, 1)$, ja johtaisi sitten posteriorijakauman käyttämällä samaa uskottavuusfunktiota. Tällöin siis sm:t Y_i ovat riippumattomia ehdolla Θ, ja niillä on jakauma Bernoulli(Θ). Yhteisjakauma on

$$f_{\Theta,\mathbf{Y}}(\theta, \mathbf{y}) = f(y_1, \ldots, y_n \mid \theta), \qquad 0 \le \theta \le 1, \quad \mathbf{y} \in \{0, 1\}^n$$

ja posteriorijakauma on Be$(1+k, 1+n-k)$. Tämän laskun olemme jo ratkaisseet esimerkissä 7.4.

Ajatellaan seuraavaksi, että komponentin i valmistuksen yhteydessä mitataan jokin tieto x_i, joka on jossakin yhteydessä sen seikan kanssa, tuleeko komponentista i toimiva vai viallinen. Tällöin x_i:tä kutsutaan selittäväksi muuttujaksi (engl. *explanatory variable*) tai kovariaatiksi (engl. *covariate*). Eräs tapa ottaa kovariaatit mukaan malliin on seuraava.

Otetaan parametreiksi α ja β, ja yritetään selittää onnistumistodennäköisyyttä i:nnessä Bernoullin kokeessa lineaarisen lausekkeen $\alpha + \beta x_i$ kautta. Tämä lauseke voi saada mielivaltaisia arvoja, joten se ei suoraan sovi Bernoullin jakauman todennäköisyysparametriksi. Sen sijaan kaavalla

$$p(\alpha, \beta, x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)},$$

saadaan luku joka on nollan ja yhden välillä oli arvot α, β ja x mitä tahansa. Kun parametrit ovat α, β, niin mallin mukaan

$$Y_i \sim \text{Bernoulli}(p(\alpha, \beta, x_i)), \qquad i = 1, \ldots, n$$

riippumattomasti. Tällöin uskottavuusfunktio on

$$f(y_1, \ldots, y_n \mid \alpha, \beta) = \prod_{i=1}^{n} \left(\frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} \right)^{y_i} \left(\frac{1}{1 + \exp(\alpha + \beta x_i)} \right)^{1-y_i}.$$

Tämä on esimerkki logistisesta regressiomallista, joka puolestaan on erikoistapaus yleistetystä lineaarisesta mallista (engl. *generalized linear model, GLM*). Monesta tilastollisesta ohjelmistosta löytyy rutiini, joka ratkaisee SU-estimaatit iteratiivisesti logistisessa regressiossa.

Bayes-päättelyssä parametreille tarvitaan priorijakauma $f_{A,B}(\alpha, \beta)$, joka voisi olla esim. jokin kaksiulotteinen normaalijakauma. Mallin mukaan sm:t Y_i ovat riippumattomia ehdolla $(A, B) = (\alpha, \beta)$, ja sm:n Y_i ehdollinen jakauma on Bernoulli($p(\alpha, \beta, x_i)$). Parametrien ja aineiston yhteisjakauma on

$$f_{A,B}(\alpha, \beta)\, f(y_1, \ldots, y_n \mid \alpha, \beta),$$

josta posteriorijakauman ominaisuudet joudutaan käytännössä selvittämään numeerisesti. $\triangle$

Esimerkki 8.3. Jossain tapauksissa havaittu aikasarja $y_1, \ldots, y_n$ voidaan mallintaa satunnaismuuttujilla $Y_0, Y_1, \ldots, Y_n$, jossa $Y_0 = y_0$ on jokin tunnettu vakio, ja muuten pätee autoregressiivinen malli

$$Y_t = h(Y_{t-1}, \beta) + \epsilon_t, \qquad t \geq 1.$$

Tässä $h(y, \beta)$ on jokin tunnettu funktio, ja sm:t $\epsilon_t \sim N(0, \sigma^2)$ riippumattomasti. Parametreja ovat β sekä σ.

Tässä mallissa on voimassa Markov-ominaisuus

$$f(y_t \mid y_1, \ldots, y_{t-1}) = f(y_t \mid y_{t-1}), \qquad \forall t \geq 1.$$

Kertolaskusääntö antaa yhteisjakaumalle esityksen

$$\begin{aligned} f(y_1, \ldots, y_n) &= f(y_1)\, f(y_2 \mid y_1)\, f(y_3 \mid y_1, y_2) \cdots f(y_n \mid y_1, \ldots, y_{n-1}) \\ &= f(y_1)\, f(y_2 \mid y_1)\, f(y_3 \mid y_2) \cdots f(y_n \mid y_{n-1}). \end{aligned}$$

Tästä saadaan uskottavuusfunktioksi

$$f(y_1, \ldots, y_n \mid \beta, \sigma) = \prod_{t=1}^{n} \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(y_t - h(y_{t-1}, \beta))^2}{\sigma^2} \right)$$

Tilanteessa voidaan soveltaa joko frekventistä tai Bayes-päättelyä. Jos erityisesti $h(y, \beta) = \beta y$, niin tilanteeseen löytyy paljon teoriaa ja valmisohjelmia, mutta yleisessä tapauksessa joudutaan itse kirjoittamaan ohjelmat mallin estimointia varten. $\triangle$

Esimerkki 8.4. (Puuttuva aineisto) Käytännön tilastotieteilijä joutuu usein tekemisiin sellaisten aineistojen kanssa, joista puuttuu jokin alunperin suunniteltu havainto. Toisinaan taas todennäköisyysmalli on kätevämpi muotoilla siten, että se sisältää satunnaismuuttujia, joiden arvoja ei voida havaita. Kummassakin tapauksessa ei-havaittuja satunnaismuuttujia voidaan kutsua puuttuvaksi aineistoksi (engl. *missing data*) tai niitä voidaan kutsua latenteiksi muuttujiksi (engl. *latent variables*) tai voidaan puhua apumuuttujista (engl *auxiliary variables*). Tällaisessa mallissa havaitun aineiston uskottavuusfunktio saadaan marginalisoimalla havaitun aineiston $\mathbf{Y}$ ja puuttuvan aineiston $\mathbf{U}$ yhteistiheyttä

$$f_{\mathbf{Y}|\mathbf{\Theta}}(\mathbf{y} \mid \boldsymbol{\theta}) = \int f_{\mathbf{U},\mathbf{Y}|\Theta}(\mathbf{u}, \mathbf{y} \mid \boldsymbol{\theta})\, \mathrm{d}\mathbf{u}.$$

Usein tätä integraalia ei osata käsitellä analyyttisesti, minkä takia tähän puuttuvan aineiston tilanteen analysointiin on kehitetty erityismenetelmiä (esim. ns. EM-algoritmi). $\triangle$

8.6 Multinomijakauma

Olkoon $(p_1, \dots, p_n)$ todennäköisyysvektori, eli kukin $p_i \ge 0$, ja

$$p_1 + \cdots + p_n = 1.$$

Multinomijakauman $\mathrm{Mult}(k, (p_1, \dots, p_n))$ ptnf on

$$f(x_1, \dots, x_n) = \binom{k}{x_1, \dots, x_n} p_1^{x_1} \dots p_n^{x_n}, \tag{8.14}$$

kun $x_1 + \cdots + x_n = k$ ja nolla muuten.

Multinomijakauma syntyy, kun tarkastellaan k-kertaista toistokoetta, jossa kussakin toistossa toteutuu yksi n:sta vaihtoehdosta, jossa toistot ovat toisistaan riippumattomia. Muuttujat $x_1, \dots, x_n$ ovat eri lopputulosten frekvenssit k:ssa toistossa. Kaava (8.14) perustellaan samalla tavalla kuin trinomijakauman tapauksessa (vrt. jakso 6.3); viimekädessä vedotaan jaksossa 1.5 esitettyyn multinomikertoimien kombinatoriseen luonnehdintaan.

Esimerkki 8.5. (Ryhmitellyt normaaliset havainnot.) Olkoot $Z_1, \dots Z_k \sim N(\mu, \sigma^2)$ riippumattomasti. Tarkastellaan katkaisupisteitä $t_1 < t_2 < \cdots < t_{n-1}$ sekä välejä

$$B_1 = (-\infty, t_1], \quad B_n = (t_{n-1}, \infty), \quad B_j = (t_{j-1}, t_j], \qquad j = 2, \dots, n-1.$$

Määritellään sm X_j siten, että se kertoo, kuinka moni sm:sta Z_i saa arvon joukosta B_j. Sv:n $(X_1, \dots, X_n)$ jakauma on multinomijakauma otoskoolla k ja todennäköisyysvektorilla $(p_1, \dots, p_n)$, jossa

$$p_1 = \Phi(t_1 \mid \mu, \sigma), \quad p_n = 1 - \Phi(t_{n-1} \mid \mu, \sigma),$$
$$p_j = \Phi(t_j \mid \mu, \sigma) - \Phi(t_{j-1} \mid \mu, \sigma), \quad j = 2, \dots, n-1.$$

Tässä $\Phi(x \mid \mu, \sigma)$ on jakauman $N(\mu, \sigma^2)$ kf.

Satunnaismuuttuja Z_i voisi esim. edustaa miespuolisen yliopisto-opiskelijan pituutta. Frekvenssit X_j saattaisivat olla peräisin kyselylomaketutkimuksesta, jossa pyydetään laittamaan rasti ruutuun sen mukaan, onko oma pituus pienempi kuin 160, välillä 160–170, välillä 170–180, välillä 180–190 vai suurempi kuin 190 senttimetriä. $\triangle$

Useimmiten multinomijakaumaa pidetään n-ulotteisena jakauma, mutta joissakin yhteyksissä on mielekkäämpää pitää sitä $(n-1)$-ulotteisena jakauma (jolloin $X_n = k - X_1 - \cdots - X_{n-1}$). Tällä konventiolla binomijakauma ja trinomijakauma ovat multinomijakauman erikoistapauksia.

Se, että (8.14) on ptnf seuraa suoraan *multinomikaavasta*, jonka mukaan kaikilla reaaliluvuilla $a_1, \ldots, a_n$ pätee

$$(a_1 + \cdots + a_n)^k = \sum \binom{k}{x_1, \ldots, x_n} a_1^{x_1} \cdots a_n^{x_n}, \tag{8.15}$$

jossa summataan yli kaikkien $(x_1, \ldots, x_n)$, joille $x_i \geq 0$ ovat kokonaislukuja, joiden summa on k. Kaikki multinomijakauman reunajakaumat ovat multinomijakaumia: esimerkiksi yksiulotteiset reunajakaumat ovat binomijakaumia. Jos joidenkin komponenttien arvoilla ehdollistetaan, niin multinomijakauman ehdolliset jakaumat ovat multinomijakaumia.

Multinomijakauman odotusarvovektori on

$$k(p_1, \ldots, p_n),$$

ja sen kovarianssimatriisin alkiot ovat

$$\operatorname{var} X_i = kp_i(1 - p_i), \qquad \operatorname{cov}(X_i, X_j) = -k\, p_i\, p_j, \quad i \neq j.$$

8.7 Tiheysfunktion muuntokaava

Tarkastellaan ensin diffeomorfisimia $\mathbf{g} : A \to B$, jossa $A, B \subset \mathbb{R}^n$ ovat avoimia joukkoja. Tällöin $\mathbf{g}$:llä on käänteisfunktio $\mathbf{h} : B \to A$, ja sekä $\mathbf{g}$ että $\mathbf{h}$ ovat jatkuvasti derivoituvia. Näiden kuvausten komponenttifunktiot ovat

$$\mathbf{g} = (g_1, \ldots, g_n), \qquad \mathbf{h} = (h_1, \ldots, h_n).$$

Otetaan käyttöön merkintä $D_i\, h_j(\mathbf{y})$ tarkoittamaan reaaliarvoisen funktion h_j osittaisderivaattaa sen i:nnen argumentin suhteen pisteessä $\mathbf{y}$, ts.

$$D_i\, h_j(\mathbf{y}) = \frac{\partial}{\partial y_i} h_j(\mathbf{y}).$$

Tarkastelemme bijektiivistä vastaavuutta

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) \quad \Longleftrightarrow \quad \mathbf{x} = \mathbf{h}(\mathbf{y}),$$

eli

$$\mathbf{y} = \mathbf{y}(\mathbf{x}) \quad \Longleftrightarrow \quad \mathbf{x} = \mathbf{x}(\mathbf{y}).$$

Kuvauksen $\mathbf{h} = (h_1, \ldots, h_n)$ derivaatta(matriisi) (eli Jacobin matriisi) pisteessä $\mathbf{y} \in B$ on

$$\begin{bmatrix} D_1 h_1(\mathbf{y}) & D_2 h_1(\mathbf{y}) & \ldots & D_n h_1(\mathbf{y}) \\ D_1 h_2(\mathbf{y}) & D_2 h_2(\mathbf{y}) & \ldots & D_n h_2(\mathbf{y}) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 h_n(\mathbf{y}) & D_2 h_n(\mathbf{y}) & \ldots & D_n h_n(\mathbf{y}) \end{bmatrix}$$

Kuvauksen $\mathbf{h}$ jacobiaani, eli Jacobin determinantti eli funktionaalideterminantti

$$J_{\mathbf{h}}(\mathbf{y}) = \frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)}$$

on Jacobin matriisin determinantti.

Tiheysfunktion muuntokaava n-ulotteisessa tilanteessa todistetaan täsmälleen samalla tekniikalla kuin kaksiulotteisessa tilanteessa (vrt. lause 6.13).

Lause 8.3. *Jos* $\mathbf{g} : A \to B$ *on diffeomorfismi,* $\mathbf{h} : B \to A$ *on sen käänteiskuvaus, eli* $\mathbf{h} = \mathbf{g}^{-1}$*, ja sv:lla* $\mathbf{X}$ *on jatkuva jakauma, ja* $P(\mathbf{X} \in A) = 1$*, niin sv:lla* $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ *on jatkuva jakauma tf:lla*

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{h}(\mathbf{y}))\, |J_{\mathbf{h}}(\mathbf{y})|, \qquad \textit{kun } \mathbf{y} \in B,$$

ja nolla muualla.

Tämä tulos kannattaa pitää mielessään muodossa

$$f_{\mathbf{X}}(\mathbf{x})\, |\partial \mathbf{x}| = f_{\mathbf{Y}}(\mathbf{y})\, |\partial \mathbf{y}|, \qquad \text{kun} \tag{8.16}$$

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) \quad \iff \quad \mathbf{x} = \mathbf{h}(\mathbf{y}), \qquad \mathbf{x} \in A, \mathbf{y} \in B \tag{8.17}$$

Esimerkki 8.6. (Affiini muunnos.) Jos $\mathbf{A}$ on vakiomatriisi ja $\mathbf{b}$ on vakiovektori siten, että seuraavassa kaavassa dimensiot ovat yhteensopivia, niin funktiota

$$\mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}, \qquad \mathbf{x} \in \mathbb{R}^n .$$

kutsutaan affiiniksi muunnokseksi. Jos edellä $\mathbf{A}$ on säännöllinen matriisi (eli kääntyvä matriisi), jolloin sen täytyy olla neliömatriisi, niin tämän kuvauksen käänteiskuvaus on myös affiini muunnos, sillä

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b} \quad \iff \quad \mathbf{x} = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) = \mathbf{h}(\mathbf{y})$$

Kuvauksen $\mathbf{h}$ jacobiaaniksi saadaan helpolla laskulla

$$J_{\mathbf{h}}(\mathbf{y}) = \frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})}.$$

Jos $\mathbf{X}$ on n-ulotteinen sv, jolla on jatkuva jakauma, ja sv $\mathbf{Y}$ määritellään kaavalla $\mathbf{Y} = \mathbf{g}(\mathbf{X})$, niin muistisäännöstä

$$f_{\mathbf{X}}(\mathbf{x})\, |\partial \mathbf{x}| = f_{\mathbf{Y}}(\mathbf{y})\, |\partial \mathbf{y}|$$

saadaan tulos

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| = f_{\mathbf{X}}(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}))\, \frac{1}{|\det(\mathbf{A})|}.$$

△

Muistathan, että diffeomorfismi on aina samandimensioisten avaruuksien välinen kuvaus. Jos tavoitteena on johtaa alempidimensioisen sv:n tf, niin sitten kuvaus ensin täydennetään bijektioksi (mikäli tämä on mahdollista), johdetaan muunnoksen tf, ja lopuksi kiinnostuksen kohteena olevan osavektorin tf johdetaan integroimalla.

Toisinaan tiheysfunktion muuntokaavaa tarvitaan tilanteessa, jossa $\mathbf{g} : A \to B$ ei ole diffeomorfismi, mutta A voidaan osittaa paloihin A_0 ja $A_i, i \geq 1$ siten, että seuraavat ehdot ovat voimassa

- $P(\mathbf{X} \in A_0) = 0$.
- Kun $i \geq 1$, niin kuvaus $\mathbf{g}$ rajoitettuna joukolle A_i, eli $\mathbf{g}_{|A_i}$, on diffeomorfismi $A_i \to B_i$, jossa B_i on kyseisen funktio kuvajoukko. Olkoon $\mathbf{h}_i : B_i \to A_i$ tämän funktion käänteisfunktio.

Tällöin sv:lla $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ on jatkuva jakauma tf:lla

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{i \geq 1} 1_{B_i}(\mathbf{y})\, f_{\mathbf{X}}(\mathbf{h}_i(\mathbf{y}))\, |J_{\mathbf{h}_i}(\mathbf{y})|. \tag{8.18}$$

Tämä tulos on yksiulotteisen tuloksen (lause 2.15) moniulotteinen yleistys.

8.8 Satunnaisvektorin momenttiemäfunktio

Määritelmä 8.5. Satunnaisvektorin $\mathbf{X} = (X_1, \ldots, X_n)$ momenttiemäfunktio, eli satunnaismuuttujien $X_1, \ldots, X_n$ yhteismomenttiemäfunktio on

$$M(\mathbf{t}) = M_{\mathbf{X}}(\mathbf{t}) = E \exp(\mathbf{t}^T \mathbf{X}) = E \exp\left(\sum_{j=1}^{n} t_j X_j\right),$$

niillä $\mathbf{t} = (t_1, \ldots, t_n) \in \mathbb{R}^n$, joilla odotusarvo on määritelty. Sv:n $\mathbf{X}$ kumulanttiemäfunktio (eli sm:ien $X_1, \ldots X_n$ yhteiskumulanttiemäfunktio) on

$$K(\mathbf{t}) = K_{\mathbf{X}}(\mathbf{t}) = \ln M_{\mathbf{X}}(\mathbf{t}),$$

niillä $\mathbf{t} \in \mathbb{R}^n$, joilla $M(\mathbf{t})$ on määritelty.

Määritelmä 8.6. Satunnaisvektorin $\mathbf{X} = (X_1, \ldots, X_n)$ karaketeristinen funktio, eli satunnaismuuttujien $X_1, \ldots, X_n$ yhteisjakauman karakteristinen funktio on

$$\phi(\mathbf{t}) = \phi_{\mathbf{X}}(\mathbf{t}) = E \exp(i\mathbf{t}^T \mathbf{X}) = E \exp\left(i \sum_{j=1}^{n} t_j X_j\right),$$

jossa $i = \sqrt{-1}$.

Merkintä $D_i g$ tarkoittaa funktion $g : \mathbb{R}^n \to \mathbb{R}$ osittaisderivaattaa sen i:nnen argumentin suhteen,

$$D_i\, g(\mathbf{x}) = \frac{\partial}{\partial x_i}\, g(x_1, \ldots, x_n).$$

Merkintä D_i^k tarkoittaa k:nnetta osittaisderivaattaa i:nnen argumentin suhteen, ja jos $k_1, \ldots, k_n$ ovat ei-negatiivisia kokonaislukuja, niin

$$D_1^{k_1} D_2^{k_2} \ldots D_n^{k_n} g(\mathbf{x}) = \frac{\partial^{k_1 + k_2 + \cdots + k_n}}{\partial x_1^{k_1}\, \partial x_2^{k_2} \cdots \partial x_n^{k_n}}\, g(\mathbf{x}).$$

Jos g on äärettömän monta kertaa jatkuvasti derivoituva, niin edellä osittaisderivaattojen laskujärjestyksellä ei ole väliä.

Momenttiemäfunktiolla ja kumulanttiemäfunktiolla on samanlaiset ominaisuudet kuin skalaaritapauksessa.

Lause 8.4. *Jos $M_{\mathbf{X}}(\mathbf{t})$ on olemassa jossakin epätyhjässä origon ympäristössä, niin $M_{\mathbf{X}}$ on origon ympäristössä äärettömän monta kertaa jatkuvasti derivoituva ja $M_{\mathbf{X}}$ voidaan esittää origon ympäristössä suppenevana potenssisarjana. Lisäksi*

a) Kertaluvun $(k_1, \ldots, k_n)$ momentti saadaan derivaattana

$$E\left(X_1^{k_1} X_2^{k_2} \cdots X_n^{k_n}\right) = D_1^{k_1} D_2^{k_2} \cdots D_n^{k_n} M_{\mathbf{X}}(\mathbf{0}),$$

jossa $(k_1, \ldots, k_n)$ on vektori, jonka komponentit ovat ei-negatiivisia kokonaislukuja.

b) $M_{\mathbf{X}}$ määrää sv:n $\mathbf{X}$ jakauman.

c) Karakteristinen funktio saadaan sijoituksella

$$\phi_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{X}}(i\mathbf{t}).$$

Määritelmä 8.7 (Gradientti, Hessen matriisi)**.** (Pysty)vektori

$$\nabla g(\mathbf{x}) = (D_1\, g(\mathbf{x}), D_2\, g(\mathbf{x}), \ldots, D_n\, g(\mathbf{x}))$$

on funktion $g : \mathbb{R}^n \to \mathbb{R}$ pisteessä $\mathbf{x}$ laskettu *gradientti* (engl. *gradient*). Funktion g pisteessä $\mathbf{x}$ laskettu Hessen matriisi $\mathbf{H}_g(\mathbf{x})$ (engl. *Hessian (matrix)*) on sen toisen kertaluvun osittaisderivaatoista koottu $n \times n$-matriisi

$$\mathbf{H}_g(\mathbf{x}) = \begin{bmatrix} D_1 D_1\, g(\mathbf{x}) & D_1 D_2\, g(\mathbf{x}) & \ldots & D_1 D_n\, g(\mathbf{x}) \\ D_2 D_1\, g(\mathbf{x}) & D_2 D_2\, g(\mathbf{x}) & \ldots & D_2 D_n\, g(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ D_n D_1\, g(\mathbf{x}) & D_n D_2\, g(\mathbf{x}) & \ldots & D_n D_n\, g(\mathbf{x}) \end{bmatrix}$$

Huomautuksia. Hessen matriisin käsitteen esitti 1800-luvulla preussilainen matemaatikon Ludwig Otto Hesse. Sitä varten on virheellistä kutsua sitä Hessin matriisiksi (jota nimitystä myös käytetään). Monissa lähteissä gradientti ymmärretään vaakavektoriksi, mutta tässä monisteessa se on pystyvektori. Hessen matrisille käytetään usein myös merkintöjä $\nabla^2 g(\mathbf{x})$ tai $g''(\mathbf{x})$.

Lause 8.5. *Olkoot sv:n $\mathbf{X}$ momenttiemäfunktio M ja kumulanttiemäfunktio K määriteltyjä jossakin origon ympäristössä. Tällöin*

$$E\mathbf{X} = \nabla M(\mathbf{0}) = \nabla K(\mathbf{0}), \qquad \mathrm{Cov}(\mathbf{X}) = \mathbf{H}_K(\mathbf{0}).$$

Todistus. Tulokset johdetaan yhdistetyn funktion derivointikaavalla, aivan kuten yksiulotteisessa tapauksessa. □

Esimerkki 8.7. Multinomijakaumalle $\mathbf{X} \sim \mathrm{Mult}(k, \mathbf{p})$ momenttiemäfunktion saa laskettua multinomikaavan avulla,

$$\begin{aligned} M(\mathbf{t}) = E \exp(\mathbf{t}^T \mathbf{X}) &= \sum \binom{k}{x_1, \ldots, x_n} (p_1 \mathrm{e}^{t_1})^{x_1} \cdots (p_n \mathrm{e}^{t_n})^{x_n} \\ &= (p_1 \mathrm{e}^{t_1} + \cdots + p_n \mathrm{e}^{t_n})^k \end{aligned}$$

Tätä tai sen logaritmia derivoimalla nähdään helposti, että

$$E\mathbf{X} = k\,\mathbf{p}, \qquad \mathrm{Cov}(\mathbf{X}) = k(\mathrm{diag}(\mathbf{p}) - \mathbf{p}\,\mathbf{p}^T).$$

Tässä $\mathrm{diag}(\mathbf{p})$ on lävistäjämatriisi, jonka lävistäjällä on luvut $(p_1, \ldots, p_n)$. △

Lause 8.6. *Olkoon* $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$ *sellainen sv, että sen momenttiemäfunktio* $M_{\mathbf{X}}(\mathbf{t})$ *on olemassa jossakin origon ympäristössä. Olkoon* $\mathbf{t} = (\mathbf{u}, \mathbf{v})$ *ositettu samandimensioisiin osiin kuin* $\mathbf{Y}$ *and* $\mathbf{Z}$*. Tällöin*

(a) Sv:ien $\mathbf{Y}$ *ja* $\mathbf{Z}$ *momenttiemäfunktiot ovat*

$$M_{\mathbf{Y}}(\mathbf{u}) = M_{\mathbf{X}}(\mathbf{u}, \mathbf{0}), \qquad M_{\mathbf{Z}}(\mathbf{v}) = M_{\mathbf{X}}(\mathbf{0}, \mathbf{v}).$$

(b) $\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}$ *jos ja vain jos*

$$M_{\mathbf{X}}(\mathbf{u}, \mathbf{v}) = M_{\mathbf{Y}}(\mathbf{u})\, M_{\mathbf{Z}}(\mathbf{v}).$$

kaikilla riittävän pienillä $\mathbf{u}$ *ja* $\mathbf{v}$*.*

Todistus. **(a):**

$$M_{\mathbf{Y}}(\mathbf{u}) = E \exp(\mathbf{u}^T \mathbf{Y}) = E \exp(\mathbf{u}^T \mathbf{Y} + \mathbf{0}^T \mathbf{Z}) = M_{\mathbf{X}}(\mathbf{u}, \mathbf{0}),$$

ja $M_{\mathbf{Z}}(\mathbf{v})$:n kaava nähdään oikeaksi samalla tavalla.

(b): Olkoon ensin $\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}$. Nyt

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{u}, \mathbf{v}) &= E \exp(\mathbf{u}^T \mathbf{Y} + \mathbf{v}^T \mathbf{Z}) = E \left(\exp(\mathbf{u}^T \mathbf{Y})\ \exp(\mathbf{v}^T \mathbf{Z}) \right) \\ &= M_{\mathbf{Y}}(\mathbf{u})\, M_{\mathbf{Z}}(\mathbf{v}), \end{aligned}$$

missä käytettiin tietoa $\exp(\mathbf{u}^T \mathbf{Y}) \perp\!\!\!\perp \exp(\mathbf{v}^T \mathbf{Z})$. Käänteistä implikaatiota varten oletetaan, että $M_{\mathbf{X}}(\mathbf{u}, \mathbf{v}) = M_{\mathbf{Y}}(\mathbf{u})\, M_{\mathbf{Z}}(\mathbf{v})$ kaikilla riittävän pienillä argumenteilla. Olkoot $\mathbf{Y}'$ ja $\mathbf{Z}'$ sellaisia sv:eita, että

$$\mathbf{Y}' \stackrel{\mathrm{d}}{=} \mathbf{Y}, \quad \mathbf{Z}' \stackrel{\mathrm{d}}{=} \mathbf{Z}, \quad \mathbf{Y}' \perp\!\!\!\perp \mathbf{Z}'.$$

Tällöin

$$M_{\mathbf{Y}', \mathbf{Z}'}(\mathbf{u}, \mathbf{v}) = M_{\mathbf{Y}'}(\mathbf{u})\, M_{\mathbf{Z}'}(\mathbf{v}) = M_{\mathbf{Y}}(\mathbf{u})\, M_{\mathbf{Z}}(\mathbf{v}) = M_{\mathbf{Y}, \mathbf{Z}}(\mathbf{u}, \mathbf{v}),$$

joten sv:n $(\mathbf{Y}', \mathbf{Z}')$ ja sv:n $(\mathbf{Y}, \mathbf{Z})$ momenttiemäfunktiot yhtyvät jossakin origon ympäristössä, joten niillä on sama jakauma. □

Luku 9

Moniulotteinen normaalijakauma

Tässä luvussa tarkastellaan normaalijakauman moniulotteista yleistystä eli moniulotteista (eli monimuuttujaista) normaalijakaumaa (engl. *multivariate normal distribution*). Sitä kutsutaan myös multinormaalijakaumaksi. Paitsi että tässä luvussa tutustutaan tähän sovelluksissa usein esiintyvään moniulotteiseen jakaumaan, tarkoituksena on lisäksi demonstroida, miten moniulotteisia jakaumia käsitellään vektori- ja matriisimerkinnöillä.

9.1 Standardinormaalijakauma $N_n(\mathbf{0}, \mathbf{I})$

Määritelmä 9.1. Sv:lla $\mathbf{U} = (U_1, \ldots, U_n)$ on n-ulotteinen standardinormaalijakauma eli normaalijakauma $N_n(\mathbf{0}, \mathbf{I})$ täsmälleen silloin, kun sen komponentit ovat riippumattomia $N(0, 1)$-jakaumaa noudattavia satunnaismuuttujia.

Satunnaisvektorin $\mathbf{U} \sim N_n(\mathbf{0}, \mathbf{I})$ tf on

$$\begin{aligned} f_{\mathbf{U}}(\mathbf{u}) &= \prod_{i=1}^{n} f_{U_i}(u_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-\frac{1}{2}u_i^2} \\ &= (2\pi)^{-n/2} \exp(-\frac{1}{2}(u_1^2 + \cdots + u_n^2)) = (2\pi)^{-n/2} \, \exp(-\frac{1}{2}\mathbf{u}^T\mathbf{u}). \end{aligned} \tag{9.1}$$

Sv:n $\mathbf{U}$ odotusarvovektori on n-dimensionen nollavektori, ja sen kovarianssimatriisi on dimensiota $n \times n$ oleva yksikkömatriisi

$$E\mathbf{U} = \mathbf{0}_n, \qquad \operatorname{Cov}\mathbf{U} = \mathbf{I}_n. \tag{9.2}$$

Sv:n $\mathbf{U}$ momenttiemäfunktio on

$$M_{\mathbf{U}}(\mathbf{t}) = E\exp(t^T\mathbf{U}) = E\prod_{i=1}^{n}\exp(t_iU_i) = \prod_{i=1}^{n}\mathrm{e}^{\frac{1}{2}t_i^2} = \exp(\frac{1}{2}\mathbf{t}^T\mathbf{t}). \tag{9.3}$$

Jos sv $\mathbf{X}$ määritellään kaavalla

$$\mathbf{X} = \mathbf{A}\mathbf{U} + \boldsymbol{\mu}, \tag{9.4}$$

jossa $\mathbf{A}$ on $m \times n$-vakiomatriisi, $\boldsymbol{\mu} \in \mathbb{R}^m$ on vakiovektori, ja $\mathbf{U} \sim N_n(\mathbf{0}, \mathbf{I}_n)$, niin

$$E\mathbf{X} = \boldsymbol{\mu}, \qquad \operatorname{Cov} \mathbf{X} = \mathbf{A}\,\mathbf{I}_n\mathbf{A}^T = \mathbf{A}\mathbf{A}^T. \tag{9.5}$$

Merkitään $\mathbf{X}$:n kovarianssimatriisia $\boldsymbol{\Sigma} = \operatorname{Cov} \mathbf{X} = \mathbf{A}\mathbf{A}^T$. Sv:n $\mathbf{X}$ momenttiemäfunktio saadaan helpolla laskulla, sillä

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{t}) &= E\exp(\mathbf{t}^T\mathbf{X}) = E\exp(\mathbf{t}^T(\mathbf{A}\mathbf{U} + \boldsymbol{\mu})) = \exp(\mathbf{t}^T\boldsymbol{\mu})\, M_{\mathbf{U}}(\mathbf{A}^T\mathbf{t}) \\ &= \exp(\mathbf{t}^T\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^T\mathbf{A}\mathbf{A}^T\mathbf{t}) = \exp(\mathbf{t}^T\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}) \end{aligned} \tag{9.6}$$

Määrittelemme kohta multinormaalijakauman esityksen (9.4) avulla. Sitä ennen tarkastelmme kysymystä, kuinka mielivaltainen kovarianssimatriisi $\boldsymbol{\Sigma}$ voidaan esittää tulona $\mathbf{A}\mathbf{A}^T$.

Eräs (mutta ei suinkaan ainoa) mahdollisuus hajotelman $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$ löytämiseksi on käyttää *Choleskyn hajotelmaa* (engl. *Cholesky decomposition*). Choleskyn hajotelmassa symmetrinen ja positiivisesti semidefiniitti matriisi $\boldsymbol{\Sigma}$ esitetään tulona

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T, \tag{9.7}$$

jossa $\mathbf{L}$ on alakolmiomatriisi. (Alakolmiomatriisi on neliömatriisi, jonka yläkolmion alkiot $(i, j), j > i$ ovat kaikki nollia.) Choleskyn hajotelma on saatavilla matriisilaskennan ohjelmakirjastoissa (mutta tyypillisesti ohjelmat palauttavat jälkimmäisen tekijän eli yläkolmiomatriisin $\mathbf{L}^T$).

Huomautus. Jos $\boldsymbol{\Sigma}$ on positiivisesti definiitti matriisi, niin se on kääntyvä matriisi. Jos se esitetään muodossa

$$\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T,$$

jossa $\mathbf{A}$ on neliömatriisi, niin matriisi $\mathbf{A}$ on välttämättä kääntyvä matriisi.

9.2 Yleinen multinormaalijakauma

Määritelmä 9.2. Sv:lla $\mathbf{X}$ on multinormaalijakauma, jos se voidaan esittää muodossa

$$\mathbf{X} = \mathbf{A}\mathbf{U} + \boldsymbol{\mu} \tag{9.8}$$

jossa $\mathbf{A}$ on $m \times n$-vakiomatriisi, $\boldsymbol{\mu} \in \mathbb{R}^m$ on vakiovektori, ja $\mathbf{U} \sim N_n(\mathbf{0}, \mathbf{I}_n)$ jollakin n.

Yksiulotteinen normaalijakauma $N(\mu, \sigma^2)$ on multinormaalijakauman erikoistapaus, sillä

$$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad (X = \mu + \sigma U, \quad U \sim N(0, 1)).$$

Määritelmästä sekä kaavoista (9.5) ja (9.6) seuraa, että

$$\begin{aligned} E\mathbf{X} = \boldsymbol{\mu}, \qquad \boldsymbol{\Sigma} = \operatorname{Cov} \mathbf{X} = \mathbf{A}\mathbf{A}^T, \\ M_{\mathbf{X}}(\mathbf{t}) = \exp(\mathbf{t}^T\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}) \end{aligned}$$

Lause 9.1. *Määritelmästä 9.2 seuraa, että* $E\mathbf{X} = \boldsymbol{\mu}$ *ja* $\operatorname{Cov} \mathbf{X} = \mathbf{A}\mathbf{A}^T$*. Sv:n* $\mathbf{X}$ *jakauma riippuu vain sen odotusarvovektorista ja kovarianssimatriisista, ei siis esitysdimensiosta* n *eikä esitysmatriisin* $\mathbf{A}$ *muista ominaisuuksista.*

Kääntäen, jos $\boldsymbol{\mu} \in \mathbb{R}^m$ *on vakiovektori ja* $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}$ *on positiivisesti semidefiniitti matriisi, niin on olemassa sv* $\mathbf{X}$*, jolla on multinormaalijakauma odotusarvovektorilla* $\boldsymbol{\mu}$ *ja kovarianssimatriisilla* $\boldsymbol{\Sigma}$*.*

Todistus. Odotusarvovektori ja kovarianssimatriisi on jo johdettu aikaisemmin. Koska $\mathbf{X}$:n momenttiemäfunktio riippuu vain sen odotusarvovektorista ja kovarianssimatriisista, niin myös sen jakauma riippuu vain näistä parametreista.

Käänteisen tuloksen todistamiseksi positiivisesti semidefiniitti matriisi $\boldsymbol{\Sigma}$ jaetaan tekijöihin $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$, ja sen jälkeen jakaumaa noudattava sv konstruoidaan kaavalla (9.8). □

Tästä lähtien multinormaalijakaumaa merkitään tunnuksella $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, jossa $\boldsymbol{\mu}$ on jakauman odotusarvo ja $\boldsymbol{\Sigma}$ sen kovarianssimatriisi. Alaindeksi m voidaan jättää pois, jos satunnaisvektorin $\mathbf{X}$ dimensiosta ei tarvitse pitää kirjaa.

Määritelmä 9.2 voidaan tulkita simulointireseptiksi. Jos tahdotaan simuloida jakaumaa $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, jossa $\boldsymbol{\Sigma}$ on annettu symmetrinen ja positiivisesti semidefiniitti matriisi, niin ensin etsitään kovarianssimatriisille $\boldsymbol{\Sigma}$ hajotelma muodossa

$$\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T.$$

Tekijöihinjako voidaan tehdä esim. Choleskyn hajotelmalla. Tämän jälkeen simuloidaan (riippumattomasti) m arvoa $(u_1, \ldots, u_m)$ standardinormaalijakaumasta $N(0, 1)$, ja lopuksi lasketaan

$$\mathbf{x} = \mathbf{A}\mathbf{u} + \boldsymbol{\mu}, \qquad \text{jossa } \mathbf{u} = (u_1, \ldots, u_n).$$

9.3 Affiinin muunnoksen jakauma

Lause 9.2 (Multinormaalisuus säilyy affiinissa muunnoksessa)**.** *Olkoon* $\mathbf{X} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, ja olkoon* $\mathbf{B} \in \mathbb{R}^{p \times m}$ *vakiomatriisi ja* $\mathbf{b} \in \mathbb{R}^p$ *vakiovektori. Tällöin*

$$\mathbf{B}\mathbf{X} + \mathbf{b} \sim N_p(\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T).$$

Todistus. Käytetään esitystä (9.8):

$$\mathbf{B}\mathbf{X} + \mathbf{b} = \mathbf{B}(\mathbf{A}\mathbf{U} + \boldsymbol{\mu}) + \mathbf{b} = (\mathbf{B}\mathbf{A})\mathbf{U} + (\mathbf{B}\boldsymbol{\mu} + \mathbf{b}).$$

Määritelmän 9.2 mukaan sv:lla $\mathbf{B}\mathbf{X} + \mathbf{b}$ on multinormaalijakauma. Lisäksi

$$\begin{aligned} E(\mathbf{B}\mathbf{X} + \mathbf{b}) &= \mathbf{B}E(\mathbf{X}) + \mathbf{b} = \mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \\ \operatorname{Cov}(\mathbf{B}\mathbf{X} + \mathbf{b}) &= \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T. \quad \square \end{aligned}$$

Tämä lause selvittää myös kaikki reunajakaumat. Jos määritellään i:s yksikkövektori $\mathbf{e}_i$ siten, että sen i:s koordinaatti on yksi ja muut koordinaatit ovat nollia, niin

$$X_i = \mathbf{e}_i^T \mathbf{X}.$$

Jos sv $\mathbf{Y}$ koostuu sv:n $\mathbf{X}$ komponenteista $(X_{i_1}, \ldots, X_{i_k})$, niin $\mathbf{Y}$ saadaan kertomalla sv:a $\mathbf{X}$ vasemmalta vakiomatriisilla, jonka j:s vaakarivi on $\mathbf{e}_{i_j}$,

$$\mathbf{Y} = \begin{bmatrix} X_{i_1} \\ \vdots \\ X_{i_k} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{i_1}^T \\ \vdots \\ \mathbf{e}_{i_k}^T \end{bmatrix} \mathbf{X}.$$

Tästä seuraa, että $\mathbf{Y}$:llä on multinormaalijakauma, joten multinormaalijakauman kaikki reunajakaumat ovat multinormaalisia.

Sv:n $\mathbf{Y}$ jakauman parametrit saadaan joko lauseen 9.2 avulla, tai vaihtoehtoisesti poimimalla asiaankuuluva osa $\mathbf{X}$:n jakauman parametreista, sillä

$$E\mathbf{Y} = \begin{bmatrix} EX_{i_1} \\ \vdots \\ EX_{i_k} \end{bmatrix} = \begin{bmatrix} \mu_{i_1} \\ \vdots \\ \mu_{i_k} \end{bmatrix}, \quad (\operatorname{Cov} \mathbf{Y})(p, q) = \operatorname{cov}(X_{i_p}, X_{i_q}) = \mathbf{\Sigma}(i_p, i_q).$$

9.4 Tiheysfunktio

Lause 9.3 (Multinormaalijakauman tiheysfunktio)**.** *Jos* $\mathbf{X} \sim N_m(\boldsymbol{\mu}, \mathbf{\Sigma})$*, jossa* $\mathbf{\Sigma}$ *on positiivisesti definiitti matriisi, niin*

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-m/2} \left(\det(\mathbf{\Sigma})\right)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right). \tag{9.9}$$

Todistus. Olkoon $\mathbf{A}$ säännöllinen $m \times m$-matriisi siten, että $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T$. Tällainen $\mathbf{A}$ on mahdollista löytää, koska $\mathbf{\Sigma}$ on positiivisesti definiitti. Käytetään esitystä

$$\mathbf{X} = \mathbf{A}\mathbf{U} + \boldsymbol{\mu}, \quad \text{jossa} \quad \mathbf{U} \sim N(\mathbf{0}, \mathbf{I}_m),$$

sekä muuttujanvaihtoa

$$\mathbf{x} = \mathbf{A}\mathbf{u} + \boldsymbol{\mu} \quad \iff \quad \mathbf{u} = \mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu}).$$

(Jotta voitaisiin soveltaa tiheysfunktion muuntokaavaa, tarvitaan säännöllinen kerroinmatriisi $\mathbf{A}$, ja multinormaalijakauman esityksessä vektoreilla $\mathbf{X}$ ja $\mathbf{U}$ täytyy olla sama dimensio.) Tiheysfunktion muuntokaavalla saadaan

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= f_{\mathbf{U}}(\mathbf{u}) \left| \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \right| = f_{\mathbf{U}}(\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})) \left| \det(\mathbf{A}^{-1}) \right| \\ &= (2\pi)^{-m/2} \left| \det(\mathbf{A}^{-1}) \right| \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T (\mathbf{A}^{-1})^T \mathbf{A}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) \end{aligned}$$

Väite seuraa tästä kaavasta sekä seuraavista laskuista,

$$\begin{gathered} \mathbf{\Sigma}^{-1} = (\mathbf{A}\mathbf{A}^T)^{-1} = (\mathbf{A}^T)^{-1}\mathbf{A}^{-1} = (\mathbf{A}^{-1})^T \mathbf{A}^{-1}, \\ \det(\mathbf{\Sigma}) = \det(\mathbf{A}\mathbf{A}^T) = \det(\mathbf{A}) \det(\mathbf{A}^T) = (\det(\mathbf{A}))^2 > 0, \\ \left| \det(\mathbf{A}^{-1}) \right| = \left| \frac{1}{\det(\mathbf{A})} \right| = \frac{1}{\sqrt{\det(\mathbf{\Sigma})}}. \end{gathered}$$

□

Huomautus. Multinormaalijakaumalla $N_m(\boldsymbol{\mu}, \mathbf{\Sigma})$ on tiheysfunktio täsmälleen silloin, kun $\mathbf{\Sigma}$ on positiivisesti definiitti. Jos $\mathbf{\Sigma}$ on pelkästään positiivisesti semidefiniitti, mutta ei positiivisesti definiitti, niin lauseen 8.2 mukaan $\mathbf{X}$ saa arvoja tietyltä hypertasolta todennäköisyydellä yksi, joten tällöin $\mathbf{X}$:llä ei voi olla tiheysfunktiota. Myös tällainen *singulaarinen multinormaalijakauma* on tärkeä jakauma sovelluksissa: esim. lineaaristen mallien yhteydessä toisaalta sovitevektorin jakauma ja toisaalta residuaalivektorin eli jäännösvektorin jakauma ovat kumpikin singulaarisia multinormaalijakaumia.

9.5 Tiheysfunktion tasa-arvopinnat

Moniulotteisen normaalijakauman tiheysfunktion (9.9) tasa-arvopinnat ovat m-ulotteisia ellipsoideja. Tämä nähdään soveltamalla kovarianssimatriisiin ominaisarvohajotelmaa, jonka ensin kertaamme.

Olkoon $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ symmetrinen ja positiivisesti semidefiniitti matriisi. Symmetrisenä matriisina sillä on reaaliset ominaisarvot λ_i ja ominaisvektorit $\mathbf{v}_i$. Ominaisvektorit voidaan valita ortonormaaleiksi, jolloin

$$\mathbf{\Sigma} \mathbf{v}_i = \lambda_i \mathbf{v}_i, \qquad i = 1, \ldots, n, \tag{9.10}$$

$$\mathbf{v}_i^T \mathbf{v}_j = \begin{cases} 1, & \text{kun } i = j, \\ 0, & \text{kun } i \neq j. \end{cases} \tag{9.11}$$

Ominaisarvojen ja ominaisvektoreiden haku sisältyy kaikkiin matriisilaskennan ohjelmakirjastoihin.

Koska $\mathbf{\Sigma}$ on positiivisesti semidefiniitti, on

$$0 \le \mathbf{v}_i^T \mathbf{\Sigma} \mathbf{v}_i = \mathbf{v}_i^T (\lambda_i \mathbf{v}_i) = \lambda_i.$$

Siis kaikki ominaisarvot ovat ei-negatiivisia. Jos $\mathbf{\Sigma}$ on peräti positiivisesti definiitti, niin kaikki ominaisarvot ovat aidosti positiivisia.

Kootaan nyt ominaisvektoreista matriisi $\mathbf{V}$ laittamalla ne matriisin pystyriveksi, sekä muodostetaan ominaisarvoista lävistäjämatriisi $\mathbf{\Lambda}$,

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_n \end{bmatrix}, \qquad \mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

Tällöin ominaisarvon ja ominaisvektorin määritelmästä seuraa kaava

$$\mathbf{\Sigma V} = \mathbf{V \Lambda}$$

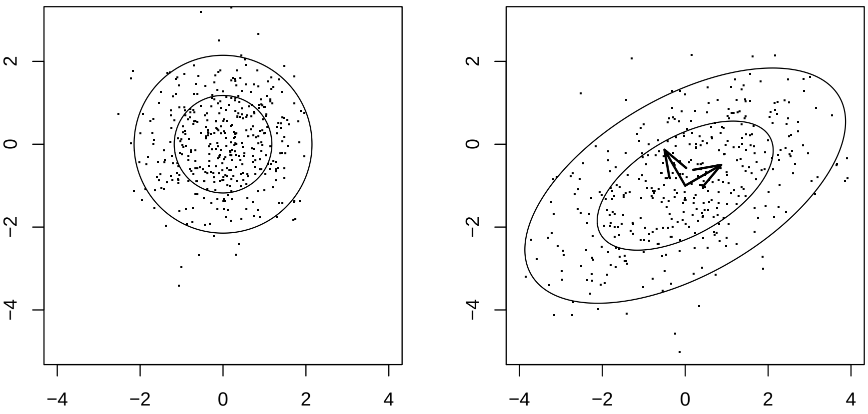
Siitä, että ominaisvektorit ovat ortonormaaleja seuraa, että $\mathbf{V}^T \mathbf{V} = \mathbf{I}_n$. Koska $\mathbf{V}$ on lisäksi neliömatriisi, niin $\mathbf{V}^{-1} = \mathbf{V}^T$. Tällaista matriisia, jonka käänteismatriisi on sen transpoosi, kutsutaan *ortogonaaliseksi matriisiksi.* Nyt ollaan saatu johdettua matriisin $\mathbf{\Sigma}$ *ominaisarvohajotelma*,

$$\mathbf{\Sigma} = \mathbf{V \Lambda V}^T, \quad \text{jossa} \quad \mathbf{V}^{-1} = \mathbf{V}. \tag{9.12}$$

Olkoon nyt $\mathbf{\Sigma}$ multinormaalijakauman kovarianssimatriisi, ja oletetaan että se on positiivisesti definiitti. Tällöin sen ominaisarvohajotelmasta saadaan hajotelma sen käänteismatriisille, nimittäin

$$\mathbf{\Sigma} = \mathbf{V \Lambda V}^T \quad \Rightarrow \quad \mathbf{\Sigma}^{-1} = \mathbf{V \Lambda}^{-1} \mathbf{V}^T.$$

Kuva 9.1 Kaksiulotteinen normaalijakauma: (a) $N(\mathbf{0}, \mathbf{I})$, (b) $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.



(a) (b)

Normaalijakauman $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ tiheysfunktio $f_{\mathbf{X}}(\mathbf{x})$ (9.9) saa vakioarvon niillä argumenteilla $\mathbf{x}$, joilla eksponenttifunktion argumenttissa oleva lauseke saa vakioarvon, mutta

$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{V} \boldsymbol{\Lambda}^{-1} \mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu}) = \frac{y_1^2}{\lambda_i} + \cdots + \frac{y_m^2}{\lambda_m},$$

jossa vektori $\mathbf{y}$ määritellään kaavalla

$$\mathbf{y} = \mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu}) \qquad \Longleftrightarrow \qquad \mathbf{x} = \boldsymbol{\mu} + \mathbf{V}\mathbf{y}.$$

Vektori $\mathbf{y}$ saadaan vektorista $\mathbf{x}$ koordinaatistonmuutoksella, jossa uudeksi origoksi valitaan $\boldsymbol{\mu}$ ja uusiksi koordinaattiakseleiksi vektorit $\mathbf{V}\mathbf{e}_i$ (jossa $\mathbf{e}_i$ on avaruuden $\mathbb{R}^n$ standardikannan i:s yksikkövektori). Nämä vektorit ovat ortonormaaleja, sillä

$$(\mathbf{V}\mathbf{e}_i)^T (\mathbf{V}\mathbf{e}_j) = \mathbf{e}_i^T \mathbf{V}^T \mathbf{V} \mathbf{e}_j = \mathbf{e}_i^T \mathbf{e}_j.$$

Koska kaikki $\lambda_i > 0$, tasa-arvopinnat toteuttavat yhtälön

$$\frac{y_1^2}{(\sqrt{\lambda_i})^2} + \cdots + \frac{y_m^2}{(\sqrt{\lambda_m})^2} = c, \qquad c > 0.$$

Tämä on m-ulotteisen ellipsoidin yhtälö uusille koordinaateille $\mathbf{y}$. Ellipsoidin puoliakselien pituudet ovat $\sqrt{c\lambda_1}, \ldots, \sqrt{c\lambda_m}$.

Esimerkki 9.1. (Kaksiulotteisen normaalijakauman havainnollistus.) Olkoon

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Kuvassa 9.1 a on piirretty otosvektoreita $\mathbf{u}_i$ kaksiulotteisesta normaalijakaumasta $N(\mathbf{0}, \mathbf{I})$ sekä piirretty sen tiheysfunktion tasa-arvokäyriä. Kuvassa 9.1 b tämä otos on muunnettu kaavalla

$$\mathbf{x}_i = \mathbf{A}\mathbf{u}_i + \boldsymbol{\mu}, \qquad \mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}^{1/2}$$

niin, että on saatu otos kaksiulotteisesta normaalijakaumasta $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, jossa

$$\boldsymbol{\Sigma} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T,$$

sekä piirretty jakauman tiheysfunktion tasa-arvokäyriä, kun $\theta = 30°$. Kuvaan on piirretty pisteeseen $\boldsymbol{\mu}$ myös vektorit $\mathbf{V}\mathbf{e}_1$ ja $\mathbf{V}\mathbf{e}_2$. $\triangle$

9.6 Korreloimattomuus ja riippumattomuus

Riippumattomat satunnaisvektorit $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$ eivät korreloi, sillä

$$\begin{aligned} \operatorname{cov}(\mathbf{X}, \mathbf{Y}) &= E[(\mathbf{X} - E\mathbf{X})(\mathbf{Y} - E\mathbf{Y})^T] \\ &= E[\mathbf{X} - E\mathbf{X}]\, E[(\mathbf{Y} - E\mathbf{Y})^T] = \mathbf{0}. \end{aligned}$$

Yleisesti ottaen korreloimattomuudesta ei seuraa riippumattomuus. Jos yhdistetyn satunnaisvektorin $(\mathbf{X}, \mathbf{Y})$ yhteisjakauma on multinormaalinen, niin tässä tapauksessa korreloimattomuudesta seuraa riippumattomuus, mikä asia todistetaan seuraavassa lauseessa.

Kirjataan myöhempää käyttöä varten, minkälaiset ovat osavektorien $\mathbf{X} = (X_1, \ldots, X_k)$ ja $\mathbf{Y} = (Y_1, \ldots, Y_m)$ multinormaalijakaumien parametrit, jos yhdistetty vektori $(\mathbf{X}, \mathbf{Y})$ noudattaa multinormaalijakaumaa $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Odotusarvovektori $\boldsymbol{\mu}$ koostuu osista

$$\boldsymbol{\mu} = E\mathbf{Z} = \begin{bmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{bmatrix} = \begin{bmatrix} E\mathbf{X} \\ E\mathbf{Y} \end{bmatrix}, \tag{9.13}$$

ja kovarianssimatriisi $\boldsymbol{\Sigma} = \operatorname{Cov} \mathbf{Z}$ voidaan myös osittaa, sillä

$$\begin{aligned} \boldsymbol{\Sigma} &= \begin{bmatrix} \operatorname{cov}(X_1, X_1) & \ldots & \operatorname{cov}(X_1, X_k) & \operatorname{cov}(X_1, Y_1) & \ldots & \operatorname{cov}(X_1, Y_m) \\ \vdots & & \vdots & \vdots & & \vdots \\ \operatorname{cov}(X_k, X_1) & \ldots & \operatorname{cov}(X_k, X_k) & \operatorname{cov}(X_k, Y_1) & \ldots & \operatorname{cov}(X_k, Y_m) \\ \operatorname{cov}(Y_1, X_1) & \ldots & \operatorname{cov}(Y_1, X_k) & \operatorname{cov}(Y_1, Y_1) & \ldots & \operatorname{cov}(Y_1, Y_m) \\ \vdots & & \vdots & \vdots & & \vdots \\ \operatorname{cov}(Y_m, X_1) & \ldots & \operatorname{cov}(Y_m, X_k) & \operatorname{cov}(Y_m, Y_1) & \ldots & \operatorname{cov}(Y_m, Y_m) \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \boldsymbol{\Sigma}_{YY} \end{bmatrix} = \begin{bmatrix} \operatorname{cov}(\mathbf{X}, \mathbf{X}) & \operatorname{cov}(\mathbf{X}, \mathbf{Y}) \\ \operatorname{cov}(\mathbf{Y}, \mathbf{X}) & \operatorname{cov}(\mathbf{Y}, \mathbf{Y}) \end{bmatrix} \end{aligned} \tag{9.14}$$

Tällöin reunajakaumat ovat jakson 9.3 kaavojen mukaan

$$\mathbf{X} \sim N(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_{XX}), \qquad \mathbf{Y} \sim N(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_{YY}). \tag{9.15}$$

Lause 9.4. *Jos vektorilla $(\mathbf{X}, \mathbf{Y})$ on multinormaalijakauma, niin*

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \quad \iff \quad \operatorname{cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}.$$

Todistus. Koska multinormaalisen yhteisjakauman momenttiemäfunktio on määritelty kaikilla argumenteilla, niin satunnaisvektorit $\mathbf{X}$ ja $\mathbf{Y}$ ovat riippumattomat silloin ja vain silloin, kun niiden yhteismomenttiemäfunktio faktoroituu komponenttien $\mathbf{X}$ ja $\mathbf{Y}$ momenttiemäfunktioiden tuloksi, eli silloin kun

$$M_{\mathbf{X},\mathbf{Y}}(\mathbf{u}, \mathbf{v}) = M_{\mathbf{X}}(\mathbf{u})\, M_{\mathbf{Y}}(\mathbf{v}) \quad \text{kaikilla } \mathbf{u}, \mathbf{v}.$$

Kun käytetään osituksia (9.13) ja (9.14), niin

$$
\begin{aligned}
&M_{\mathbf{X},\mathbf{Y}}(\mathbf{u},\mathbf{v}) \\
&= \exp\left(\mathbf{u}^T\boldsymbol{\mu}_X + \mathbf{v}^T\boldsymbol{\mu}_Y + \frac{1}{2}(\mathbf{u}^T\boldsymbol{\Sigma}_{XX}\mathbf{u} + \mathbf{u}^T\boldsymbol{\Sigma}_{XY}\mathbf{v} + \mathbf{v}^T\boldsymbol{\Sigma}_{YX}\mathbf{u} + \mathbf{v}^T\boldsymbol{\Sigma}_{YY}\mathbf{v})\right).
\end{aligned}
$$

Toisaalta reunajakaumien momenttiemäfunktioiden tulo on

$$M_{\mathbf{X}}(\mathbf{u})\; M_{\mathbf{Y}}(\mathbf{v}) = \exp\left(\mathbf{u}^T\boldsymbol{\mu}_X + \frac{1}{2}\mathbf{u}^T\boldsymbol{\Sigma}_{XX}\mathbf{u} + \mathbf{v}^T\boldsymbol{\mu}_Y + \frac{1}{2}\mathbf{v}^T\boldsymbol{\Sigma}_{YY}\mathbf{v}\right).$$

Edellä

$$\mathbf{v}^T\boldsymbol{\Sigma}_{YX}\mathbf{u} = (\mathbf{v}^T\boldsymbol{\Sigma}_{YX}\mathbf{u})^T = \mathbf{u}^T\boldsymbol{\Sigma}_{XY}\mathbf{v},$$

joten funktiot $M_{\mathbf{X},\mathbf{Y}}$ ja $M_{\mathbf{X}}\; M_{\mathbf{Y}}$ ovat samat silloin ja vain silloin, kun

$$\mathbf{u}^T\boldsymbol{\Sigma}_{XY}\mathbf{v} = 0 \quad \text{kaikilla } \mathbf{u},\mathbf{v},$$

eli silloin, kun

$$\boldsymbol{\Sigma}_{XY} = \operatorname{cov}(\mathbf{X},\mathbf{Y}) = \mathbf{0}.$$

□

Huomautus. Jos oletetaan multinormaaliset reunajakaumat

$$\mathbf{X} \sim N(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_{XX}), \qquad \mathbf{Y} \sim N(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_{YY}),$$

niin yhdistetyn vektorin $(\mathbf{X},\mathbf{Y})$ jakauma ei välttämättä ole multinormaalinen. Ei edes vaikka oletettaisiin esim. lisäehto $\operatorname{cov}(\mathbf{X},\mathbf{Y}) = \mathbf{0}$. Jos sen sijaan oletetaan, että $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$, niin sen jälkeen momenttiemäfunktiosta

$$M_{\mathbf{X},\mathbf{Y}}(\mathbf{u},\mathbf{v}) = M_{\mathbf{X}}(\mathbf{u})M_{\mathbf{Y}}(\mathbf{v})$$

nähdään (ks. todistuksen kaavoja), että $(\mathbf{X},\mathbf{Y}) \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, jossa

$$\boldsymbol{\mu} = \begin{bmatrix}\boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y\end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_{XX} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{YY}\end{bmatrix}.$$

9.7 Ehdolliset jakaumat

Lause 9.5. *Olkoon* $(\mathbf{X},\mathbf{Y}) \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, *ja käytetään odotusarvovektorille* $\boldsymbol{\mu}$ *ositusta* (9.13), *ja kovarianssimatriisille* $\boldsymbol{\Sigma}$ *ositusta* (9.14). *Jos osamatriisi* $\boldsymbol{\Sigma}_{XX}$ *on säännollinen, niin sv:n* $\mathbf{Y}$ *jakauma ehdolla* $\mathbf{X} = \mathbf{x}$ *on multinormaalinen, ja ehdollisen jakauman parametrit ovat*

$$\mathbf{Y} \mid (\mathbf{X} = \mathbf{x}) \sim N(\boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}(\mathbf{x} - \boldsymbol{\mu}_X),\; \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}).$$

Todistus. Vaikka kaavat ovat monimutkaisia, todistuksen idea on yksinkertainen. Ensin muodostetaan sv $\mathbf{V}$ kaavalla

$$\mathbf{V} = \mathbf{Y} - \mathbf{B}\mathbf{X}, \tag{9.16}$$

jossa $\mathbf{B}$ on vakiomatriisi, joka valitaan kohta. Yhdistetty vektori $(\mathbf{V},\mathbf{X})$ saadaan multinormaalisesta vektorista $(\mathbf{X},\mathbf{Y})$ lineaarisella muunnoksella, joten sillä on multinormaalinen yhteisjakauma. Kun valitaan

$$\mathbf{B} = \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1},$$

niin $\mathbf{V}$ ja $\mathbf{X}$ eivät korreloi, sillä

$$\operatorname{cov}(\mathbf{V}, \mathbf{X}) = \operatorname{cov}(\mathbf{Y} - \mathbf{B}\mathbf{X}, \mathbf{X}) = \mathbf{\Sigma}_{YX} - \mathbf{B}\mathbf{\Sigma}_{XX} = \mathbf{0}.$$

Koska vektorilla $(\mathbf{V}, \mathbf{X})$ on multinormaalijakauma, niin korreloimattomuudesta seuraa riippumattomuus, joten $\mathbf{V} \perp\!\!\!\perp \mathbf{X}$. Esityksen (9.16) mukaan

$$\mathbf{Y} = \mathbf{V} + \mathbf{B}\mathbf{X},$$

jossa $\mathbf{V}$ ja $\mathbf{X}$ ovat riippumattomia. Tällöin $\mathbf{Y}$:n jakauma ehdolla $\mathbf{X} = \mathbf{x}$ sama kuin sv:n

$$\mathbf{V} + \mathbf{B}\mathbf{x}$$

jakauma, sillä ehto $\mathbf{X} = \mathbf{x}$ ei muuta sv:n $\mathbf{V}$ jakaumaa, koska $\mathbf{V} \perp\!\!\!\perp \mathbf{X}$. Tästä nähdään, että ehdollinen jakauma on multinormaalinen. Tämän jälkeen ehdollisen jakauman parametrit saadaan laskemalla sv:n $\mathbf{V}+\mathbf{B}\mathbf{x}$ jakauman parametrit. Ehdollisen jakauman odotusarvovektori on

$$E(\mathbf{V} + \mathbf{B}\mathbf{x}) = \boldsymbol{\mu}_Y + \mathbf{B}(\mathbf{x} - \boldsymbol{\mu}_X) = \boldsymbol{\mu}_Y + \mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}(\mathbf{x} - \boldsymbol{\mu}_X),$$

ja kovarianssimatriisi on

$$\begin{aligned} \operatorname{Cov}(\mathbf{V} + \mathbf{B}\mathbf{x}) &= \operatorname{Cov}(\mathbf{V}) = \operatorname{cov}(\mathbf{Y} - \mathbf{B}\mathbf{X}, \mathbf{Y} - \mathbf{B}\mathbf{X}) \\ &= \mathbf{\Sigma}_{YY} - \mathbf{\Sigma}_{YX}\mathbf{B}^T - \mathbf{B}\mathbf{\Sigma}_{XY} + \mathbf{B}\mathbf{\Sigma}_{XX}\mathbf{B}^T \\ &= \mathbf{\Sigma}_{YY} - \mathbf{B}\mathbf{\Sigma}_{XX}^{-1}\mathbf{B}^T. \end{aligned}$$

□

9.8 Kaksiulotteinen normaalijakauma

Tarkastellaan kaksiulotteisen normaalijakauman $(X, Y) \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$ ominaisuuksia. Johdettavat kaavat ovat pitkiä, eikä niitä kannata yrittää painaa muistiinsa. Ne voi tarvittaessa johtaa itse tai tarkistaa kirjallisuudesta.

Olkoon $-1 < \rho = \operatorname{corr}(X, Y) < 1$, ja merkitään jakauman parametrien komponentteja seuraavasti,

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \qquad \mathbf{\Sigma} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

Kovarianssimatriisin käänteismatriisi on

$$\mathbf{\Sigma}^{-1} = \frac{1}{\det(\mathbf{\Sigma})} \begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}, \quad \text{jossa} \quad \det(\mathbf{\Sigma}) = (1 - \rho^2)\sigma_X^2\sigma_Y^2.$$

Jakauman ytf on

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{1}{2\pi\sigma_X\,\sigma_Y\sqrt{1 - \rho^2}} \\ &\exp\left(-\frac{1}{2(1 - \rho^2)} \left[\left(\frac{x - \mu_X}{\sigma_X} \right)^2 - 2\rho\,\frac{x - \mu_X}{\sigma_X}\,\frac{y - \mu_Y}{\sigma_Y} + \left(\frac{y - \mu_Y}{\sigma_Y} \right)^2 \right] \right). \end{aligned}$$

Tässä eksponenttifunktion argumentin kaava saadaan kertomalla auki tulo

$$-\frac{1}{2}\begin{bmatrix} (x - \mu_X) & (y - \mu_Y) \end{bmatrix} \mathbf{\Sigma}^{-1} \begin{bmatrix} (x - \mu_X) \\ (y - \mu_Y) \end{bmatrix}.$$

Reunajakaumat ovat

$$X \sim N(\mu_X, \sigma_X^2), \qquad Y \sim N(\mu_Y, \sigma_Y^2).$$

Ehdolliset jakaumat saadaan lauseen 9.5 perusteella:

$$\begin{aligned}
Y \mid (X = x) &\sim N(\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X),\ (1-\rho^2)\sigma_Y^2), \\
X \mid (Y = y) &\sim N(\mu_X + \rho \frac{\sigma_X}{\sigma_Y}(y - \mu_Y),\ (1-\rho^2)\sigma_X^2).
\end{aligned}$$

Tässä esim. jakauman $Y \mid (X = x)$ parametrit saadaan laskuilla

$$\begin{aligned}
\boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}(\mathbf{x} - \boldsymbol{\mu}_X) &= \mu_Y + \frac{\rho\,\sigma_X\sigma_Y}{\sigma_X^2}(x - \mu_X) \\
\boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY} &= \sigma_Y^2 - \frac{(\rho\sigma_X\sigma_Y)^2}{\sigma_X^2} = (1-\rho^2)\sigma_Y^2.
\end{aligned}$$

On suoraviivaista mutta työlästä tarkistaa, että nämä ehdolliset jakaumat saadaan johdettua myös kaavoilla

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \qquad f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

9.9 Normaalijakauman otoskeskiarvon ja otosvarianssin yhteisjakauma

Tarkastellaan sm:ia $X_1, \ldots, X_n$. Niiden otoskeskiarvo $\bar{X}$ ja otosvarianssi S^2 määritellään kaavoilla

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2. \tag{9.17}$$

Jos sm:t X_i ovat riippumattomia, ja niillä on yhteinen odotusarvo μ ja yhteinen varianssi σ^2, niin helpohkoilla laskuilla voidaan näyttää, että

$$E\bar{X} = \mu, \qquad ES^2 = \sigma^2.$$

Ts. otoskeskiarvo ja otosvarianssi ovat vastaavien populaatioparametrien μ ja σ^2 harhattomia estimaattoreita.

Oletetaan tästä lähtien, että sm:t $X_1, \ldots, X_n$ ovat riippumattomia, ja niillä kaikilla on normaalijakauma $N(\mu, \sigma^2)$. Tällöin vektorilla $(X_1, \ldots, X_n)$ on multinormaalijakauma $N_n(\mathbf{m}, \boldsymbol{\Sigma})$ parametreilla

$$\mathbf{m} = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_n.$$

Seuraavaksi johdetaan vektorin $(\bar{X}, S^2)$ yhteisjakauma.

Otoskeskiarvon reunajakauma on helppo johtaa, nimittäin $\bar{X} \sim N(\mu, \frac{1}{n}\sigma^2)$, sillä se on multinormaalijakautuneen vektorin lineaarimuunnos, ja osaamme laskea sm:n $\bar{X}$ odotusarvon ja varianssin. Seuraavan lauseen todistuksessa osoitetaan, että otosvarianssilla S^2 on sopivan skaalauksen jälkeen khiin neliön jakauma vapausasteluvulla $n - 1$.

Tehdään ennen lauseen muotoilua se huomio, että mikäli satunnaisvektorilla $\mathbf{Z}$ on k-ulotteinen standardinormaalijakauma, ts. $\mathbf{Z} \sim N_k(\mathbf{0}, \mathbf{I})$, niin sen pituuden neliöllä $\|\mathbf{Z}\|^2 = \mathbf{Z}^T\mathbf{Z}$ on khiin neliön jakauma vapausasteluvulla k, sillä

$$\|\mathbf{Z}\|^2 = \mathbf{Z}^T\mathbf{Z} = \sum_{i=1}^{k} Z_i^2,$$

jossa $Z_i \sim N(0,1)$ riippumattomasti. Tällöin $E\|\mathbf{Z}\|^2 = k$.

Lause 9.6. *Olkoot $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$, ja määritellään $\bar{X}$ ja S^2 kaavoilla* (9.17)*. Tällöin*

a) $\bar{X} \sim N(\mu, \sigma^2/n)$,

b) $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, *joten* $ES^2 = \sigma^2$,

c) $\bar{X} \perp\!\!\!\perp S^2$.

Todistus. Kohta a todistettiin jo edellä, joten todistamme vain kohdat b ja c.

Jakaumaoletus sv:lle $\mathbf{X} = (X_1, \ldots, X_n)$ voidaan lausua muodossa

$$\mathbf{X} \sim N_n(\mu\,\mathbf{1}, \sigma^2\mathbf{I}_n),$$

jossa $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^n$ on n-komponenttinen vakiovektori, jonka kaikki komponentit ovat ykkösiä.

Määritellään n-komponenttinen vektori $\mathbf{u}$ kaavalla

$$\mathbf{u} = \frac{1}{\sqrt{n}}\,\mathbf{1},$$

jolloin $\mathbf{u}$ on ykkösvektorin $\mathbf{1}$ suuntainen yksikkövektori, ts. $\mathbf{u}^T\mathbf{u} = 1$. Huomaa, että otoskeskiarvo voidaan esittää kaavalla

$$\bar{X} = \frac{1}{n}\,\mathbf{1}^T\mathbf{X} = \frac{1}{\sqrt{n}}\,\mathbf{u}^T\mathbf{X}.$$

Määrittelemme residuaalivektorin (eli jäännösvektorin) $\mathbf{R}$ kaavalla

$$\mathbf{R} = \begin{bmatrix} X_1 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{bmatrix} = \mathbf{X} - \bar{X}\mathbf{1} = \mathbf{X} - \mathbf{u}\mathbf{u}^T\mathbf{X}.$$

(Tässä vektori $\bar{X}\mathbf{1}$ on nimeltään sovitevektori.) Otosvarianssi S^2 voidaan esittää kaavalla

$$(n-1)S^2 = \mathbf{R}^T\mathbf{R} = \|\mathbf{R}\|^2, \tag{9.18}$$

jossa residuaalivektori voidaan esittää kaavalla

$$\mathbf{R} = \mathbf{X} - \mathbf{u}\mathbf{u}^T\mathbf{X} = (\mathbf{I}_n - \mathbf{u}\mathbf{u}^T)\mathbf{X}, \tag{9.19}$$

Yhdistetty vektori $(\bar{X}, \mathbf{R})$ saadaan lineaarisella muunnoksella vektorista $\mathbf{X}$, sillä

$$\begin{bmatrix} \bar{X} \\ \mathbf{R} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{n}\,\mathbf{u}^T \\ \mathbf{I} - \mathbf{u}\mathbf{u}^T \end{bmatrix} \mathbf{X},$$

minkä takia sillä on multinormaalinen yhteisjakauma. Lisäksi komponentit $\bar{X}$ ja $\mathbf{R}$ ovat korreloimattomia, sillä

$$\operatorname{cov}(\bar{X}, \mathbf{R}) = \frac{1}{\sqrt{n}} \operatorname{cov}(\mathbf{u}^T \mathbf{X}, (\mathbf{I} - \mathbf{u}\mathbf{u}^T)\mathbf{X}) = \frac{1}{\sqrt{n}} \mathbf{u}^T \sigma^2 \mathbf{I}(\mathbf{I} - \mathbf{u}\mathbf{u}^T) = \mathbf{0}.$$

Lauseen 9.4 perusteella $\bar{X}$ ja $\mathbf{R}$ ovat riippumattomia. Koska S^2 saadaan laskettua kaavan (9.18) mukaan residuaalivektorin $\mathbf{R}$ funktiona, ovat myös $\bar{X}$ ja S^2 riippumattomia, mikä todistaa kohdan c.

Otosvarianssin jakauman selvittämiseksi määritellään $n \times n$ neliömatriisi $\mathbf{Q}$ siten, että sen ensimmäinen pystyrivi on $\mathbf{u}$. Muut matriisin $\mathbf{Q}$ pystyrivit $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ valitaan siten, että matriisin $\mathbf{Q}$ pystyrivit muodostavat yhdessä $\mathbb{R}^n$:n ortonormeeratun kannan. Tällöin $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_n$, ja koska $\mathbf{Q}$ on neliömatriisi, niin se on ortogonaalinen matriisi, eli $\mathbf{Q}^{-1} = \mathbf{Q}^T$. Tästä seuraa, että $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}_n$. Kun käytetään $\mathbf{Q}$:n ositusta $\mathbf{Q} = \begin{bmatrix}\mathbf{u} & \mathbf{V}\end{bmatrix}$ jossa $n \times (n-1)$-matriisin $\mathbf{V}$ pystyrivit ovat $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$, niin saadaan kaava

$$\mathbf{I}_n = \mathbf{Q}\mathbf{Q}^T = \mathbf{u}\mathbf{u}^T + \mathbf{V}\mathbf{V}^T.$$

Tästä nähdään kaavan (9.19) avulla, että

$$\mathbf{R} = (\mathbf{I}_n - \mathbf{u}\mathbf{u}^T)\mathbf{X} = \mathbf{V}\mathbf{V}^T\mathbf{X}.$$

Koska matriisin $\mathbf{V}$ pystyrivit ovat ortonormaalisia, on $\mathbf{V}^T\mathbf{V} = \mathbf{I}_{n-1}$, joten

$$\mathbf{R}^T\mathbf{R} = \left(\mathbf{V}\mathbf{V}^T\mathbf{X}\right)^T (\mathbf{V}\mathbf{V}^T\mathbf{X}) = \mathbf{X}^T\mathbf{V}\mathbf{V}^T\mathbf{X} = \|\mathbf{V}^T\mathbf{X}\|^2. \tag{9.20}$$

Vektorilla $\mathbf{V}^T\mathbf{X}$ on $(n-1)$-ulotteinen normaalijakauma, jonka odotusarvo on

$$E(\mathbf{V}^T\mathbf{X}) = \mathbf{V}^T(\mu\mathbf{1}) = \mathbf{0}$$

sillä perusteella, että $\mathbf{u}$ on kohtisuorassa jokaista matriisin $\mathbf{V}$ pystyriviä vastaan, ja $\mathbf{1} = \sqrt{n}\mathbf{u}$. Kovarianssimatriisi on

$$\operatorname{Cov}(\mathbf{V}^T\mathbf{X}) = \mathbf{V}^T(\sigma^2\mathbf{I}_n)\mathbf{V} = \sigma^2\mathbf{I}_{n-1}.$$

Tällöin

$$\frac{1}{\sigma}\mathbf{V}^T\mathbf{X} \sim N_{n-1}(\mathbf{0}, \mathbf{I}_{n-1}),$$

joten sen pituuden neliöllä on khiin neliön jakauma vapausasteluvulla $n-1$. Kaavojen (9.18) ja (9.20) avulla nähdään lopulta, että

$$\frac{n-1}{\sigma^2}S^2 = \frac{1}{\sigma^2}\mathbf{R}^T\mathbf{R} = (\frac{1}{\sigma}\mathbf{V}^T\mathbf{X})^T(\frac{1}{\sigma}\mathbf{V}^T\mathbf{X}) = \|\frac{1}{\sigma}\mathbf{V}^T\mathbf{X}\|^2 \sim \chi^2_{n-1}.$$

Tästä seuraa, että $E[(n-1)S^2/\sigma^2] = n-1$, josta $ES^2 = \sigma^2$. $\square$

Palautetaan mieleen, että t-jakauma vapausasteluvulla ν määritellään siten, että se on sm:n

$$\frac{Z}{\sqrt{Y/\nu}}$$

jakauma, kun $Z \sim N(0,1)$, $Y \sim \chi^2_\nu$ ja $Z \perp\!\!\!\perp Y$. Sen neliöllä

$$\frac{Z^2/1}{Y/\nu}$$

on F-jakauma $F_{1,\nu}$ osoittajan vapausasteilla 1 ja nimittäjän vapausasteilla ν.

Satunnaismuuttujalla $\bar{X} - \mu$ on normaalijakauma odotusarvolla nolla ja varianssilla σ^2/n, joten satunnaismuuttujalla

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

on standardinormaalijakauma. Jos tässä tuntemattoman keskihajontaparametrin σ tilalle sijoitetaan sen otosestimaatti, saadaan ns. t-testisuure

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

Tämä t-testisuure voidaan esittää yhtäpitävästi kaavalla

$$T = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{S^2/\sigma^2}}.$$

Edellisen lauseen mukaan tässä osoittaja ja nimittäjä ovat riipumattomia, osoittajan jakauma on $N(0,1)$ ja nimittäjässä on neliöjuuri satunnaismuuttujasta, jossa χ^2-jakaumaa noudattava satunnaismuuttuja jaetaan vapausasteluvullaan $n-1$. Tästä seuraa, että T:llä on t-jakauma $n-1$ vapausasteella.

Edeltävän lauseen tarkastelu on vähällä vaivalla laajennettavissa myös tilanteeseen, jossa tarkastellaan lineaarista mallia

$$\mathbf{X} \sim N_n(\mathbf{m}, \sigma^2\mathbf{I}),$$

jossa odotusarvovektorin $\mathbf{m}$ tiedetään kuuluvan tunnettuun $\mathbb{R}^n$:n lineaariseen aliavaruuteen L, jonka dimensio $p \le n$. Muodostetaan tälle aliavaruudelle ortonormaalit kantavektorit $\mathbf{u}_1, \ldots, \mathbf{u}_p$, ja asetetaan nämä vektorit matriisin $\mathbf{U}$ pystyriveiksi, $\mathbf{U} = \begin{bmatrix}\mathbf{u}_1 & \ldots & \mathbf{u}_p\end{bmatrix}$. Tällöin matriisi

$$\mathbf{H} = \mathbf{U}\mathbf{U}^T$$

on projektiomatriisi aliavaruudelle L, eli $\mathbf{H}$ on symmetrinen ja idempotentti matriisi (eli $\mathbf{HH} = \mathbf{H}$) ja $\mathbf{Hv} = \mathbf{v}$ aina kun $\mathbf{v} \in L$. Tämän jakson lauseen todistusta matkimalla on helppo tarkistaa, että

$$\mathbf{HX} \quad \perp\!\!\!\perp \quad (\mathbf{I} - \mathbf{H})\mathbf{X},$$

ja että

$$\frac{1}{\sigma^2} \|\mathbf{X} - \mathbf{HX}\|^2 \sim \chi^2_{n-p}.$$

Tämän takia satunnaismuuttuja

$$\frac{1}{n-p} \|\mathbf{X} - \mathbf{HX}\|^2$$

on varianssin σ^2 harhaton estimattori, ja sen jakauma on skaalattu khiin neliö. Tässä tilanteessa sovitevektori $\mathbf{HX}$ ja residuaalivektori $\mathbf{X} - \mathbf{HX} = (\mathbf{I} - \mathbf{H})\mathbf{X}$ ovat riippumattomia satunnaisvektoreita.

Luku 10

Raja-arvolauseita ja approksimaatioita

Tässä luvussa esitellään sellaisia kuuluisia todennäköisyysteorian raja-arvolauseita, joita sovelletaan usein tilastollisessa päättelyssä. Näiden raja-arvolauseiden tunteminen kuuluu jokaisen tilastotieteilijän yleissivistykseen. Valitettavasti päätulosten todistuksia ei voida käydä läpi tämän kurssin puitteissa.

10.1 Suurten lukujen laki

Palautetaan mieleen stokastisen suppenemisen ja melkein varman suppenemisen määritelmät.

Määritelmä 10.1. Jono satunnaismuuttujia $X_1, X_2, \ldots$ *suppenee stokastisesti* eli *konvergoi stokastisesti* (engl. *converges in probability*) kohti satunnaismuuttujaa Y, jos

$$P(|X_n - Y| \geq \epsilon) \xrightarrow[n\to\infty]{} 0,$$

kaikilla $\epsilon > 0$.

Stokastista suppenemista merkitään usein seuraavaan tapaan,

$$X_n \xrightarrow{\text{P}} Y, \quad \text{tai} \quad \operatorname{plim}_{n\to\infty} X_n = Y.$$

(Määritelmässä voitaisiin yhtäpitävästi vaatia, että $P(|X_n - Y| > \epsilon) \to 0$ kaikilla $\epsilon > 0$.)

Määritelmä 10.2. Jono satunnaismuuttujia $X_1, X_2, \ldots$ suppenee (eli konvergoi) melkein varmasti (engl. *converges almost surely*) kohti satunnaismuuttujaa Y, jos $X_n(\omega) \to Y(\omega)$ perusjoukon osajoukossa, jonka tn on yksi, eli jos

$$P(\lim_{n\to\infty} X_n = Y) = 1.$$

Melkein varmaa suppenemista merkitään esim. seuraavilla tavoilla,

$$X_n \xrightarrow{\text{m.v.}} Y, \quad \text{tai} \quad X_n \xrightarrow{\text{a.s.}} Y.$$

On mahdollista osoittaa, että melkein varmasta suppenemisesta seuraa stokastinen suppeneminen, eli että

$$X_n \xrightarrow{\text{a.s.}} Y \qquad \Rightarrow \qquad X_n \xrightarrow{\text{p}} Y \tag{10.1}$$

Todistimme luvussa 3.7 *i.i.d.*-jonoja (eli jonoa riippumattomia ja samoin jakautuneita satunnaismuuttujia) koskevan ns. heikon suurten lukujen lain (engl. *weak law of large numbers, WLLN*). Sen mukaan, tiettyjen momenttiehtojen ollessa voimassa, n ensimmäisen satunnaismuuttujan aritmeettinen keskiarvo suppenee stokastisesti kohti yhden jonon jäsenen keskiarvoa. Heikon suurten lukujen lain todistus oli suoraviivainen Tšebyševin epäyhtälön sovellus.

Kuuluisan vahvan suurten lukujen lain mukaan keskiarvon suppeneminen on peräti melkein varmaa.

Lause 10.1 (Vahva suurten lukujen laki)**.** *(Engl.* strong law of large numbers, SLLN.*) Olkoon $X_1, X_2, \ldots$ jono riippumattomia ja samoin jakautuneita satunnaismuuttujia, joiden odotusarvo $\mu = EX_1$ on olemassa. Tällöin keskiarvo*

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

suppenee melkein varmasti kohti arvoa μ, eli $\bar{X}_n \xrightarrow{\text{a.s.}} \mu$.

Vahvan suurten lukujen lain seurauksena jono $(\bar{X}_n)$ toteuttaa tietenkin myös heikon suurten lukujen lain.

Tällaisia (vahvoja tai heikkoja) suurten lukujen lakeja löytyy kirjallisuudesta myös muunkin tyyppisille (ei välttämättä *i.i.d*-oletukset täyttäville) satunnaismuuttujien, satunnaisvektoreitten ja muitten satunnaisalkioitten muodostamille jonoille.

10.2 Jakaumasuppeneminen

Määritelmä 10.3. Olkoon $X_1, X_2, \ldots$ jono satunnaismuuttujia, joiden kertymäfunktiot ovat $F_1, F_2, \ldots$. Olkoon Y satunnaismuuttuja, jonka kertymäfunktio on G, ts. kaikilla x

$$F_n(x) = P(X_n \le x), \qquad G(x) = P(Y \le x).$$

Jono (X_n) suppenee jakaumaltaan (engl. *converges in distribution* tai *converges in law*) kohti Y:tä, mikäli

$$\lim_{n \to \infty} F_n(x) = G(x)$$

kaikissa rajajakauman kertymäfunktion G jatkuvuuspisteissä x.

Usein jakaumasuppenemisessa rajajakauma on normaalijakauma $N(0,1)$ tai jokin muu jatkuva jakauma. Jatkuvan jakauman kertymäfunktio on jatkuva koko reaaliakselilla, joten tälläisen rajajakauman kohdalla kertymäfunktiot suppenevat jokaisessa reaaliakselin pisteessä.

Jakaumasuppenemista merkitään usein esim. seuraavasti:

$$X_n \xrightarrow{\text{d}} Y, \quad \text{tai} \quad X_n \xrightarrow{\mathcal{L}} Y, \quad \text{tai} \quad X_n \xrightarrow{\text{w}} Y.$$

Edellä yläindeksi d tulee sanasta *distribution*, $\mathcal{L}$ sanasta *law* ja yläindeksi w sanasta *weak*. Viimeinen termi johtuu siitä, että jakaumasuppenemista kutsutaan toisinaan jakaumien (tai niiden kertymäfunktioiden) heikoksi suppenemiseksi. Jos rajajakauma on jokin tuttu jakauma, kuten $N(0,1)$, niin jakaumasuppenemista voidaan merkitä siten, että satunnaismuuttujan sijasta käytetään rajajakauman tunnusta, seuraavaan tapaan,

$$X_n \xrightarrow{\mathrm{d}} N(0,1).$$

Voidaan osoittaa, että seuraavat implikaatiot pitävät paikkansa:

$$X_n \xrightarrow{\mathrm{a.s.}} Y \qquad \Rightarrow \qquad X_n \xrightarrow{\mathrm{p}} Y \qquad \Rightarrow \qquad X_n \xrightarrow{\mathrm{d}} Y. \tag{10.2}$$

Yksikään implikaatioista ei päde yleisesti käänteiseen suuntaan. Jos rajamuuttuja Y on deterministinen eli vakio c, jolloin Y:n jakauman sanotaan olevan degeneroitunut, niin tällöin osoittautuu, että jakaumasuppenemisesta seuraa stokastinen suppeneminen, eli

$$X_n \xrightarrow{\mathrm{d}} c \quad (c \text{ vakio}) \quad \Rightarrow \quad X_n \xrightarrow{\mathrm{p}} c. \tag{10.3}$$

Tilastotieteessä jakaumasuppenemista käytetään usein jakaumien approksimointiin. Jos tiedetään, että

$$X_n \xrightarrow{\mathrm{d}} Y, \qquad \text{kun } n \to \infty$$

niin jollakin äärellisellä indeksin n arvolla voidaan satunnaismuuttujan X_n jakaumaa approksimoida rajamuuttujan Y jakaumalla, mitä voidaan merkitä symbolisesti

$$X_n \overset{\mathrm{d}}{\approx} Y.$$

Edellisessä merkinnässä Y:n tilalla voidaan käyttää sen jakauman tunnusta.

Jos rajamuuttujalla Y on jatkuva jakauma, niin suppenemisesta $X_n \xrightarrow{\mathrm{d}} Y$ seuraa esimerkiksi, että

$$P(X_n \in I) \xrightarrow[n\to\infty]{} P(Y \in I),$$

kun $I \subset \mathbb{R}$ on mikä tahansa väli. Tämän ansiosta suurilla n

$$P(X_n \in I) \approx P(Y \in I).$$

Usein tilastollisessa päättelyssä joudutaan tarkkojen luottamusvälien sijasta soveltamaan tähän ajatukseen perustuvia asymptoottisia luottamusvälejä. Tämän takia jakaumasuppeneminen on tärkeä käsite tilastotieteessä.

10.3 Suppenemistuloksia

Kukin edellä määritellystä suppenemislajista voidaan helposti yleistää myös koskemaan jonoa satunnaisvektoreita $(\mathbf{X}_n)$, joiden raja on satunnaisvektori $\mathbf{Y}$. Stokastinen suppeneminen $\mathbf{X}_n \xrightarrow{\mathrm{p}} \mathbf{Y}$ määritellään vaatimalla, että

$$P(\|\mathbf{X}_n - \mathbf{Y}\| \geq \epsilon) \to 0, \quad \text{kaikilla } \epsilon > 0,$$

eli korvaamalla skalaaritapauksen määritelmässä erotuksen itseisarvo erotusvektorin normilla. Melkein varma suppeneminen määritellään satunnaisvektorien muodostamalle jonolle täysin samoin kuin skalaaritapauksessa. Jakaumasuppenemisen kohdalla vaaditaan moniulotteisten kertymäfunktioiden pisteittäistä suppenemista kaikissa rajamuuttujan kertymäfunktion jatkuvuuspisteissä.

Jatkuvan kuvauksen soveltaminen suppenevaan jonoon säilyttää suppenemisen kaikille kolmelle suppenemisen lajille.

Lause 10.2 (Jatkuvan kuvauksen lause). *Jos k-ulotteisten satunnaisvektorien muodostama jono $\mathbf{X}_n \to \mathbf{Y}$ joko melkein varmasti, stokastisesti tai jakaumaltaan, niin $\mathbf{g}(\mathbf{X}_n)$ suppenee samalla tavalla kohti satunnaisvektoria $\mathbf{g}(\mathbf{Y})$, mikäli funktio $\mathbf{g} : \mathbb{R}^k \to \mathbb{R}^d$ on jatkuva funktio.*

Tämä lause pitää paikkansa myös silloin, jos $\mathbf{g}$ on jatkuva pelkästään sellaisessa joukossa $A \subset \mathbb{R}^k$, jossa rajasatunnaisvektorin $\mathbf{Y}$ arvot ovat melkein varmasti, eli jos $P(\mathbf{Y} \in A) = 1$.

Otetaan käyttöön merkintä satunnaisvektoreiden komponenteille,

$$\mathbf{X}_n = (X_{n,1}, \ldots, X_{n,k}) \quad \text{ja} \quad \mathbf{Y} = (Y_1, \ldots, Y_k).$$

Osoittautuu, että sekä melkein varmalle että stokastiselle suppenemiselle satunnaisvektoreille komponenttien suppeneminen (eli marginaalinen suppeneminen) ja koko satunnaisvektorin suppeneminen (eli yhteissuppeneminen) ovat yhtäpitäviä, ts.

$$\begin{aligned} \mathbf{X}_n \xrightarrow{\text{a.s.}} \mathbf{Y} \quad &\Leftrightarrow \quad (X_{n,j} \xrightarrow{\text{a.s.}} Y_j, \quad \forall j = 1, \ldots, k), \\ \mathbf{X}_n \xrightarrow{\text{p}} \mathbf{Y} \quad &\Leftrightarrow \quad (X_{n,j} \xrightarrow{\text{p}} Y_j, \quad \forall j = 1, \ldots, k). \end{aligned}$$

Jakaumasuppenemisessa yhteissuppenemisen ja marginaalisen suppenemisen välinen yhteys on mutkikkaampaa, sillä marginaalisesta jakaumasuppenemisesta ei seuraa yhteissuppeneminen. Esimerkiksi, jos $X_n \xrightarrow{\text{d}} X$ ja $Y_n \xrightarrow{\text{d}} Y$, niin tavallisesti vektori (X_n, Y_n) ei suppene jakaumaltaan eikä täten missään muussakaan mielessä. Tämä ongelma liittyy siihen seikkaan, että reunajakaumat eivät määrää yhteisjakaumaa. Toisaalta, jos $(X_n, Y_n) \xrightarrow{\text{d}} (X, Y)$, niin tällöin jatkuvan kuvauksen lauseen perusteella kyllä $X_n \xrightarrow{\text{d}} X$ ja $Y_n \xrightarrow{\text{d}} Y$.

Seuraavassa tärkeässä erikoistapauksessa marginaalisesta jakaumasuppenemisesta seuraa yhteissuppeneminen. Tätä tulosta (tai sen sovellusta peruslaskutoimituksiin) kutsutaan Slutskyn lauseeksi tai Slutskyn lemmaksi.

Lause 10.3 (Slutsky). *Jos $X_n \xrightarrow{\text{d}} X$ ja $Y_n \xrightarrow{\text{d}} c$, jossa c on vakio, niin $(X_n, Y_n) \xrightarrow{\text{d}} (X, c)$. Tällöin myös*

a) $X_n + Y_n \xrightarrow{\text{d}} X + c$,

b) $X_n Y_n \xrightarrow{\text{d}} cX$,

c) $X_n / Y_n \xrightarrow{\text{d}} X/c$, *mikäli* $c \neq 0$.

Todistus. Suppenemisen $(X_n, Y_n) \xrightarrow{\text{d}} (X, c)$ todistus sivuutetaan. Loppuosa väitteistä seuraa jatkuvan kuvauksen lauseesta. □

10.4 Keskeinen raja-arvolause

Jos $X_1, X_2, \ldots$ on *i.i.d.*-jono, ja $\bar{X}_n$ on n ensimmäisen jonon muuttujan keskiarvo, niin tiedämme suurten lukujen lain nojalla, että $\bar{X}_n$ suppenee kohti muuttujien yhteistä odotusarvoa $\mu = EX_1$. Lisäksi tiedämme, että

$$E\bar{X}_n = \mu,$$

ja

$$\operatorname{var} \bar{X}_n = \frac{\sigma^2}{n}, \qquad \text{jossa } \sigma^2 = \operatorname{var} X_1.$$

Keskiarvon $\bar{X}_n$ jakauma keskittyy yhä tiiviimmin arvon μ ympärille, kun n kasvaa. Keskittymisvauhtia kuvaa tietyllä tavalla keskiarvon $\bar{X}_n$ keskihajonta $\sigma/\sqrt{n}$, joka suppenee nollaa kohti vauhdilla $1/\sqrt{n}$.

Jos keskiarvo $\bar{X}_n$ standardoidaan vähentämällä siitä keskiarvon odotusarvo ja jakamalla keskiarvon keskihajonnalla, saadaan lauseke

$$\frac{\bar{X}_n - E\bar{X}_n}{\sqrt{\operatorname{var} \bar{X}_n}} = \sqrt{n}\, \frac{\bar{X}_n - \mu}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}.$$

Tämän standardoidun keskiarvon odotusarvo on tietenkin nolla, ja sen varianssi on yksi kaikilla n. Keskeisen raja-arvolauseen mukaan standardoitu keskiarvo suppenee jakaumaltaan kohti standardinormaalijakaumaa.

Lause 10.4 (Keskeinen raja-arvolause)**.** *(Engl.* central limit theorem, CLT.*) Olkoon $X_1, X_2, \ldots$ jono riippumattomia ja samoin jakautuneita satunnaismuuttujia siten, että $0 < \sigma^2 < \infty$, jossa $\sigma^2 = \operatorname{var} X_1$. Merkitään*

$$\mu = EX_1, \qquad \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Tällöin

$$\sqrt{n}\, \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{\mathrm{d}} N(0, 1).$$

Jatkuvan kuvauksen lauseen mukaan edellisen lauseen oletuksilla pätee myös

$$\sqrt{n}\, (\bar{X}_n - \mu) \xrightarrow{\mathrm{d}} N(0, \sigma^2),$$

sillä kuvaus $x \mapsto \sigma x$ on jatkuva.

Tämä tulos antaa karakterisoinnin stokastisen suppenemisen $\bar{X}_n \xrightarrow{\mathrm{p}} \mu$ vauhdille. Jos kohti nollaa suppeneva erotus $\bar{X}_n - \mu$ kerrotaan kohti ääretöntä hajaantuvalla jonolla $\sqrt{n}$, niin yhä saadaan jotakin stabiilia, eli satunnaismuuttuja, jolla on (degeneroitumaton) rajajakauma. Jos kertoimena käytetään mitä tahansa hitaammin kasvavaa lukujonoa $a_n \sqrt{n}$, jossa $a_n \to 0$, niin Slutskyn lauseen perusteella

$$a_n \sqrt{n}\, (\bar{X} - \mu) \xrightarrow{\mathrm{p}} 0.$$

Keskeiselle raja-arvolauseelle löytyy kirjallisuudesta lukuisia yleistyksiä. Tilastotieteessä tarvitaan usein keskeisen raja-arvolauseen muotoilua satunnaisvektorien muodostamalle *i.i.d.*-jonolle. Jos $\mathbf{X}_1, \mathbf{X}_2, \ldots$ on *i.i.d.*-jono satunnaisvektoreita, joiden yhteinen odotusarvovektori on $\boldsymbol{\mu}$ ja kovarianssimatriisi on $\boldsymbol{\Sigma}$, niin keskeinen raja-arvolause on voimassa muodossa

$$\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{\mathrm{d}} N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{10.4}$$

jossa $\bar{\mathbf{X}}_n$ on n ensimmäisen satunnaisvektorin keskiarvo $\frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i$.

10.5 Normaaliapproksimaatio

Kun (X_i) on *i.i.d.*-jono, niin keskeiseen raja-arvolauseeseen tukeutuen toisinaan approksimoidaan äärellisellä otoskoolla n

$$\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \overset{\mathrm{d}}{\approx} N(0,1),$$

eli standardoidun otoskeskiarvon jakaumaa approksimoidaan sen rajajakaumalla. Tämä on sama asia kuin se, että käytetään approksimaatiota

$$\bar{X}_n \overset{\mathrm{d}}{\approx} N(\mu, \frac{\sigma^2}{n}),$$

jossa keskiarvon $\bar{X}_n$ jakaumaa approksimoidaan normaalijakaumalla, jonka odotusarvo ja varianssi ovat samat kuin keskiarvon $\bar{X}_n$ odotusarvo ja varianssi. Edelleen, tämä on sama asia kuin se, että käytetään approksimaatiota

$$\sum_{i=1}^{n} X_i \overset{\mathrm{d}}{\approx} N(n\mu, n\sigma^2),$$

jossa summan jakaumaa approksimoidaan normaalijakaumalla, jonka odotusarvo ja varianssi ovat samat kuin ko. summan odotusarvo ja varianssi.

Keskeinen raja-arvolause takaa, että nämä normaaliapproksimaatiot (eli normaaliset approksimaatiot tai normaalijakauma-approksimaatiot) saadaan mielivaltaisen tarkoiksi, kun otoskoko n valitaan riittävän suureksi. Milloin otoskoko n sitten on riittävän suuri? Tämä asia riippuu toisaalta halutusta tarkkuudesta ja approksimaation käyttötarkoituksesta ja toisaalta satunnaismuuttujien X_i yhteisen jakauman luonteesta. Symmetrisille ja yksihuippuisille jakaumille saavutetaan pienehköllä (muutaman kymmenen) otoskoolla useisiin tarpeisiin riittävä tarkkuus, mutta vinojen jakaumien kohdalla vastaavaan tarkkuuteen tarvittava otoskoko voi olla monta kertaluokkaa suurempi.

Esimerkki 10.1. (Binomijakauman normaaliapproksimaatio.) Olkoot $X_1, X_2, \ldots, X_n$ riippumattomia Bernoulli(p)-jakaumaa noudattavia sm:ia, jossa $0 < p < 1$. Tällöin

$$E\bar{X}_n = p, \qquad \operatorname{var} \bar{X}_n = \frac{1}{n}\,p(1-p).$$

Normaaliapproksimaatiolla saadaan

$$\bar{X}_n \overset{\mathrm{d}}{\approx} N(p, \frac{1}{n}\,p(1-p)), \quad \text{ja} \quad \sum_{i=1}^{n} X_i \overset{\mathrm{d}}{\approx} N(np, np(1-p)).$$

Tässä tapauksessa pätee tietenkin tarkka tulos

$$\sum_{i=1}^{n} X_i \sim \mathrm{Bin}(n,p).$$

Esimerkiksi, kun $n = 25$ ja $p = 0.6$, niin $np = 15$ ja $np(1-p) = 6$. Tällöin saadaan approksimaatio

$$P(\sum_{1}^{n} X_i \le 13) \approx P(15 + \sqrt{6}\,Z \le 13) = P(Z \le \frac{13-15}{\sqrt{6}}) \approx 0.207,$$

jossa $Z \sim N(0,1)$. Kun käytetään binomijakaumaa, eikä sen approksimaatiota, saadaan tulos

$$P(\sum_1^n X_i \le 13) \approx 0.268.$$

Näin pienellä otoskoolla normaalinen approksimaatio ei ole kovin tarkka, mutta sitä voidaan parantaa huomattavasti käyttämällä ns. *jatkuvuuskorjausta.* Koska summa on kokonaislukuarvoinen, niin seuraavat tapahtumat ovat samoja,

$$\{\sum_1^n X_i \le 13\} = \{\sum_1^n X_i \le 13 + 0.5\}.$$

Kun normaaliapproksimaatiossa yläraja 13 korvataan luvulla 13.5, niin saadaan jo alkuperäistä approksimaatiota paljon tarkempi approksimaatio

$$P(\sum_1^n X_i \le 13) \approx P(Z \le \frac{13.5 - 15}{\sqrt{6}}) \approx 0.270.$$

Nykyaikana tällaiset tunnettujen jakaumien approksimaatiot eivät ole niin tärkeitä kuin ennen. Nykyään standardijakaumien kertymäfunktiot pystytään helposti laskemaan tietokoneella. $\triangle$

10.6 Deltamenetelmä

Seuraava lause on yksinkertainen esimerkki ns. deltamenetelmästä (engl. *delta method*). Sen todistus käydään läpi lähinnä siksi, että lukija saisi kuvan siitä, miten suppenemiseen liittyvää koneistoa käytetään hyväksi.

Lause 10.5. *Jos*

$$\sqrt{n}(X_n - a) \xrightarrow{\mathrm{d}} N(0, \sigma^2),$$

jossa a on vakio, ja funktio g on derivoituva pisteessä a, niin

$$\sqrt{n}(g(X_n) - g(a)) \xrightarrow{\mathrm{d}} N(0, \sigma^2 (g'(a))^2). \tag{10.5}$$

Todistus. Sovelletaan ensimmäisen kertaluvun Taylorin kehitelmää

$$g(X_n) - g(a) = g'(a)(X_n - a) + r(X_n)(X_n - a),$$

jossa jäännöstermin funktio r on

$$r(x) = \begin{cases} \dfrac{g(x) - g(a)}{x - a} - g'(a), & x \ne a \\ 0 & x = a \end{cases}$$

Huomaa, että funktio r on jatkuva pisteessä $x = a$ (derivoituvuuden nojalla).

Nyt $X_n \xrightarrow{\mathrm{p}} a$, sillä

$$X_n - a = \frac{1}{\sqrt{n}} \sqrt{n}(X_n - a),$$

jossa termi $1/\sqrt{n} \to 0$ ja toinen termi suppenee jakaumaltaan, joten (Slutskyn lause) $X_n - a \xrightarrow{\text{d}} 0$, ja koska rajamuuttuja on vakio, niin suppeneminen on peräti stokastista (ks. kaava (10.3)). Siis $X_n - a \xrightarrow{\text{P}} 0$, eli $X_n \xrightarrow{\text{P}} a$. Lisäksi

$$\sqrt{n}(g(X_n) - g(a)) = g'(a)\,\sqrt{n}(X_n - a) + r(X_n)\,\sqrt{n}(X_n - a)$$

Tässä $r(X_n) \xrightarrow{\text{P}} 0$ (jatkuvan kuvauksen lause), ja koska jono $\sqrt{n}(X_n - a)$ suppenee jakaumaltaan, niin $r(X_n)\,\sqrt{n}(X_n - a) \xrightarrow{\text{d}} 0$ (Slutskyn lause). Tämä jäännöstermin suppeneminen on peräti stokastista, koska rajamuuttuja on vakio. Kun sovelletaan toistamiseen Slutskyn lausetta, niin nähdään, että

$$g'(a)\,\sqrt{n}(X_n - a) + r(X_n)\,\sqrt{n}(X_n - a) \xrightarrow{\text{d}} g'(a)Y,$$

jossa $Y \sim N(0, \sigma^2)$. Toisin sanoen jakaumasuppenemisen kannalta jäännöstermin saa unohtaa, ja rajajakauma määräytyy ensimmäisen kertaluvun Taylorin approksimaatiosta. Väite seuraa lopulta siitä, että $g'(a)Y \sim N(0, (g'(a))^2\sigma^2)$. □

Tätä raja-arvotulosta sovelletaan usein siten, että äärellisellä otoskoolla n approksimoidaan

$$\sqrt{n}(g(X_n) - g(a)) \overset{\text{d}}{\approx} N(0, \sigma^2(g'(a))^2).$$

Esimerkki 10.2. Tarkastellaan riippumattomia muuttujia $X_i \sim \text{Bernoulli}(p)$, $i = 1, \ldots, n$, jossa $0 < p < 1$. Tällöin parametrin p SU-estimaatti on $\bar{X}_n = s/n$, jossa s on onnistumisten lukumäärä. Oletetaan, että tahdotaan estimoida ns. *log-odds* -suuretta (vedonlyöntisuhteen logaritmia),

$$\theta = \ln \frac{p}{1-p}.$$

Sen luonteva estimaatti on

$$\hat{\theta}_n = \ln \frac{\bar{X}_n}{1 - \bar{X}_n} = \ln \frac{s}{n-s},$$

jonka jakauma ei ole mikään tunnettu jakauma.

Kun delta-menetelmässä valitaan

$$g(u) = \ln \frac{u}{1-u}, \qquad 0 < u < 1,$$

saadaan helposti approksimaatio

$$\sqrt{n}(\hat{\theta}_n - \theta) \overset{\text{d}}{\approx} N(0, \frac{1}{p(1-p)}).$$

Tällä perusteella on esim. mahdollista johtaa asymptoottinen luottamusväli parametrille θ. △