Todennäköisyyslaskenta, syksy 2012

Petri Koistinen
Matematiikan ja tilastotieteen laitos
Helsingin yliopisto

27. toukokuuta 2013

Sisältö

Luku 6

Epäyhtälöitä

Epäyhtälöt ovat yksi matemaatikon voimakkaimmista työvälineistä. Siinä missä yhtälö $a = b$ kertoo sen, että kaksi ehkä näennäisesti erilaista asiaa ovatkin samoja, epäyhtälö $a \leq b$ saattaa antaa keinon analysoida monimutkaista tilannetta paljon helpommin ymmärrettävän asian avulla.

Todennäköisyyslaskennan epäyhtälöitä käytetään tyypillisesti teorian apuna, kuten esim. raja-arvotulosten johtamiseen, mutta niille löytyy toisenlaisiakin sovelluksia. Esimerkiksi niitä voidaan käyttää työkaluna satunnaistettujen algoritmien vaativuuden analysoinnissa.

Kuten yhtälöiden kanssa, myös epäyhtälöiden kanssa työskennellessä täytyy kiinnittää huomiota paitsi epäyhtälön kummankin puolen muotoon myös siihen minkä ehtojen vallitessa se on voimassa. Näiden asioiden lisäksi epäyhtälöiden kohdalla pitää kiinnittää erityistä huomiota siihen, kumpi suuruusjärjestys puolien välillä vallitsee. Usein matemaattinen argumentointi perustuu siihen, että muotoa $a \leq b$ oleva tulos johdetaan pitkän epäyhtälöketjun

$$a \leq a_1 \leq \cdots \leq a_k \leq b$$

avulla. Jos tässä ketjussa yksikin $\leq$-merkki on tullut kirjoitettua vahingossa väärin päin, niin koko argumentilta putoaa pohja pois.

6.1 Markovin ja Tšebyševin epäyhtälöt

Lause 6.1 (Markovin epäyhtälö)**.** *Olkoon $X \geq 0$ sm, jonka odotusarvo on olemassa. Tällöin*

$$P(X \geq a) \leq \frac{EX}{a}, \qquad \forall a > 0$$

Todistus. Määritellään sm Y kaavalla

$$Y = a1_{[a,\infty)}(X) = \begin{cases} 0, & \text{kun } 0 \leq X < a, \\ a, & \text{kun } X \geq a. \end{cases}$$

Koska $Y \leq X$, on

$$EX \geq EY = 0 \cdot P(Y = 0) + a \cdot P(Y = a) = a\,P(X \geq a). \qquad \square$$

Lause 6.2 (Tšebyševin epäyhtälö)**.** *Olkoon X sm, jonka odotusarvo on μ ja varianssi σ^2 on olemassa Tällöin*

$$P[|X - \mu| \geq t] \leq \frac{\sigma^2}{t^2} \qquad \forall t > 0.$$

Erityisesti, jos $\sigma^2 > 0$, niin

$$P[|X - \mu| \geq k\,\sigma] \leq \frac{1}{k^2} \qquad \forall k > 0.$$

Todistus. Sovelletaan Markovin epäyhtälöä sm:aan $Y = (X - \mu)^2$. Jos $t > 0$, niin

$$P[|X - \mu| \geq t] = P[(X - \mu)^2 \geq t^2] \leq \frac{\sigma^2}{t^2}.$$

Toinen epäyhtälö seuraa valitsemalla $t = k\sigma$. □

Tšebyševin epäyhtälön sovelluksena seuraavaksi todistetaan heikko suurten lukujen laki (engl. *weak law of large numbers, WLLN*). Lauseessa esiintyvää satunnaismuuttujien suppenemisen lajia kutsutaan stokastiseksi suppenemiseksi (eli stokastiseksi konvergenssiksi).

Lause 6.3 (Heikko suurten lukujen laki)**.** *Olkoon $X_1, X_2, \ldots$ jono riippumattomia sm:ia, joilla on sama odotusarvo $\mu = EX_i$ ja varianssi $\sigma^2 = \operatorname{var} X_i < \infty$. Tällöin keskiarvojen*

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

muodostama jono $(\bar{X}_n)$ konvergoi stokastisesti kohti arvoa μ, eli

$$P(|\bar{X}_n - \mu| \geq \epsilon) \xrightarrow[n\to\infty]{} 0, \qquad \textit{kaikilla } \epsilon > 0.$$

Todistus. Nyt

$$E\bar{X}_n = \mu, \qquad \operatorname{var} \bar{X}_n = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Tšebyševin epäyhtälön nojalla

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$ □

6.2 Konveksit funktiot ja Jensenin epäyhtälö

Funktio on konveksi, jos sen kuvaaja jää jokaisen jänteensä alapuolelle. Funktio on konkaavi, jos sen kuvaaja on jokaisen jänteensä yläpuolella. Tämä asia muotoillaan täsmällisemmin alla olevassa määritelmässä. Yleensä funktiot eivät ole konvekseja eikä konkaaveja.

Seuraavassa määritelmässä sana väli tarkoittaa mitä tahansa suljettua, avointa tai puoliavointa äärellistä tai ääretöntä väliä, ts. väli on muotoa (a, b), $[a, b)$, $(a, b]$ tai $[a, b]$ oleva joukko, jossa $a \leq b$ ja tapaukset $a = -\infty$ tai $b = \infty$ sallitaan. Jos I on väli ja $x, y \in I$, niin myös pisteitä x ja y yhdistävän janan pisteet $\lambda x + (1 - \lambda)y$, $0 \leq \lambda \leq 1$ kuuluvat välille I. Konveksin joukon karakterisoi juuri tämä ominaisuus, ja reaaliakselin konveksit joukot ovat välejä.

Määritelmä 6.1 (Konveksi funktio, konkaavi funktio)**.** Olkoon $I \subset \mathbb{R}$ väli. Funktio $g : I \to \mathbb{R}$ on *konveksi*, jos

$$g(\lambda x + (1 - \lambda)y) \leq \lambda g(x) + (1 - \lambda)g(y), \qquad \text{kaikilla } x, y \in I \text{ ja kaikilla } 0 \leq \lambda \leq 1.$$

Funktio $g : I \to \mathbb{R}$ on *konkaavi*, jos $-g$ on konveksi.

Konveksisuuden määritelmässä voitaisiin yhtä hyvin rajoittua tarkastelemaan väliä $0 < \lambda < 1$ koska tapaukset $\lambda = 0$ ja $\lambda = 1$ ovat triviaaleja. Funktiot x^2, e^x ja $|x|$ ovat konvekseja koko reaaliakselilla; $1/x$ on konveksi, kun $x > 0$, ja $\log x$ on konkaavi funktio, kun $x > 0$. Affiini funktio $a + bx$ on sekä konveksi että konkaavi koko reaaliakselilla.

Seuraavassa lauseessa esiintyy erotusosamääriä, jotka kannattaa tulkita geometrisesti g:n jänteiden kulmakertoimien avulla. Piirrä itse kuva. Lauseen tarkistus jätetään lukijalle.

Lause 6.4. *Olkoon I avoin väli. Funktio $g : I \to \mathbb{R}$ on konveksi, jos ja vain jos kaikilla $a < b < c$, joille $a, b, c \in I$ pätee epäyhtälö*

$$\frac{g(b) - g(a)}{b - a} \le \frac{g(c) - g(b)}{c - b}. \tag{6.1}$$

Lause 6.5 (Konveksisuustarkistimia)**.** *Olkoon $I \subset \mathbb{R}$ avoin väli, ja $g : I \to \mathbb{R}$ funktio.*

(a) Jos g on jatkuvasti derivoituva ja g' on kasvava (ja molemmat ominaisuudet ovat voimassa koko I:llä), niin g on konveksi.

(b) Jos g on kahdesti derivoituva I:llä, ja $g''(x) \ge 0$ kaikilla $x \in I$, niin g on konveksi.

Todistus. **(a):** Jos $a, b, c \in I$ ja $a < b < c$, niin kaikilla $a \le x \le b$ ja kaikilla $b \le y \le c$ pätee

$$g'(x) \le g'(b) \le g'(y),$$

josta saadaan derivaatan integraaleille arviot

$$\int_a^b g'(x)\,\mathrm{d}x \le (b - a)\,g'(b), \qquad \int_b^c g'(y)\,\mathrm{d}y \ge (c - b)\,g'(b).$$

Tämän takia

$$\begin{aligned} \frac{g(b) - g(a)}{b - a} &= \frac{1}{b - a} \int_a^b g'(x)\,\mathrm{d}x \le g'(b) \\ &\le \frac{1}{c - b} \int_b^c g'(y)\,\mathrm{d}y = \frac{g(c) - g(b)}{c - b}. \end{aligned}$$

joten g on konveksi.

(b): Koska $g'' \ge 0$, niin g' on kasvava I:llä. □

Jos $g : I \to \mathbb{R}$ on konveksi, niin konveksin funktion määritelmästä saadaan helposti induktiolla johdettua epäyhtälö

$$g(\sum_{i=1}^n p_i\, x_i) \le \sum_{i=1}^n p_i\, g(x_i), \tag{6.2}$$

joka pitää paikkansa, kun kukin $x_i \in I$ ja kukin $p_i \ge 0$ ja $\sum_i p_i = 1$. Tälläistä lineaarikombinaatiota $x = \sum_i p_i x_i$, jossa painot p_i ovat ei-negatiivisia ja summautuvat ykköseksi, kutsutaan pisteiden $x_1, \ldots, x_n$ *konveksiksi kombinaatioksi.* Tietenkin myös konveksi kombinaatio $x \in I$.

Jensenin epäyhtälö on edellisen epäyhtälön yleistys, kun epäyhtälössä esiintyvä diskreetti jakauma ptnf:llä $f(x_i) = p_i$ korvataan mielivaltaisella jakaumalla. Jensenin epäyhtälön todistaminen on helppoa, kunhan ensin todistamme seuraavan apulauseen. Siinä todetaan, että konveksin funktion kuvaaja on jokaisen tangenttinsa yläpuolella.

Lause 6.6. *Jos $I \subset \mathbb{R}$ on avoin väli ja $g : I \to \mathbb{R}$ on konveksi funktio, ja $m \in I$, niin voidaan valita luku k siten, että*

$$g(x) \geq g(m) + k(x - m), \qquad \forall x \in I.$$

Todistus. Rajoitutaan todistamaan väite siinä tapauksessa, jossa g on jatkuvasti derivoituva I:llä. Tällöin g' on ei-vähenevä funktio. Jos $x \geq m$ ja $x \in I$, niin

$$g(x) - g(m) = \int_m^x g'(u)\,\mathrm{d}u \geq g'(m)(x - m).$$

Jos taas $x < m$ ja $x \in I$, niin

$$g(m) - g(x) = \int_x^m g'(u)\,\mathrm{d}u \leq g'(m)(m - x).$$

Väite pitää paikkansa, kun kulmakertoimeksi valitaan $k = g'(m)$.

Tämä lause pitää paikkansa myös ilman oletusta g:n derivoituvuudesta. Todistus löytyy useista analyysin oppikirjoista, kuten esim. Billingsleyn teoksesta *Probability and Measure.* □

Lause 6.7 (Jensenin epäyhtälö)**.** *Olkoon $I \subset \mathbb{R}$ avoin väli, ja $g : I \to \mathbb{R}$ konveksi funktio. Jos X on sm, jonka arvot ovat I:ssä (tn:llä yksi), ja EX sekä $Eg(X)$ ovat olemassa, niin*

$$g(EX) \leq Eg(X) \tag{6.3}$$

Todistus. Koska $\mu = EX \in I$, niin voidaan valita luku k siten, että

$$g(x) \geq g(\mu) + k(x - \mu), \qquad \forall x \in I.$$

Siis todennäköisyydellä yksi

$$g(X) \geq g(\mu) + k(X - \mu),$$

ja Jensenin epäyhtälö seuraa ottamalla odotusarvo puolittain. □

Esimerkki 6.1. Funktio x^2 on konveksi, joten Jensenin epäyhtälön mukaan

$$(EX)^2 \leq E(X^2),$$

mikäli kyseiset momentit ovat olemassa. Tästä seuraa arvio

$$\operatorname{var} X = E(X^2) - (EX)^2 \geq 0,$$

minkä tietenkin tiesimme ennestään. △

6.3 Hölderin epäyhtälö

Lause 6.8. *Olkoot $p, q > 1$ sellaisia lukuja, että*

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Tällöin

$$E|XY| \leq [E|X|^p]^{\frac{1}{p}}\,[E|Y|^q]^{\frac{1}{q}}$$

Todistus. Väite on triviaalisti tosi, jos $E|X|^p = \infty$ tai $E|Y|^q = \infty$ tai jos $E|X|^p = 0$ tai $E|Y|^q = 0$. Tämän takia oletamme, että nämä molemmat odotusarvot ovat äärellisiä ja nollaa suurempia.

Ensin todistetaan positiivisia reaalilukuja koskeva epäyhtälö. Olkoot $a, b > 0$ mielivaltaisia, ja määritellään s ja t siten, että

$$a = \exp(\frac{1}{p}s), \qquad b = \exp(\frac{1}{q}t).$$

Koska $\exp(x) = \mathrm{e}^x$ on konveksi, on

$$ab = \exp(\frac{1}{p}s + \frac{1}{q}t) \le \frac{1}{p}\mathrm{e}^s + \frac{1}{q}\mathrm{e}^t = \frac{a^p}{p} + \frac{b^q}{q}.$$

Sama epäyhtälö pätee myös, jos $a = 0$ tai $b = 0$, kun sovimme, että nolla korotettuna positiiviseen potenssiin on nolla.

Valitaan

$$u = [E|X|^p]^{\frac{1}{p}}, \qquad v = [E|Y|^q]^{\frac{1}{q}},$$

ja otetaan puolittain odotusarvo epäyhtälöstä

$$\left|\frac{XY}{uv}\right| \le \frac{1}{p}\left|\frac{X}{u}\right|^p + \frac{1}{q}\left|\frac{Y}{v}\right|^q,$$

jolloin saadaan

$$E\left|\frac{XY}{uv}\right| \le \frac{1}{p}\frac{E|X|^p}{E|X|^p} + \frac{1}{q}\frac{E|Y|^q}{E|Y|^q} = 1.$$

□

Hölderin epäyhtälön helppona sovelluksena (HT) voidaan todistaa, että

$$[E|X|^r]^{\frac{1}{r}} \le [E|X|^s]^{\frac{1}{s}}, \qquad \text{kaikilla } 0 < r \le s \tag{6.4}$$

Tätä tulosta kutsutaan *Ljapunovin epäyhtälöksi.*

Minkowskin epäyhtälö toteaa, että

$$[E|X+Y|^p]^{\frac{1}{p}} \le [E|X|^p]^{\frac{1}{p}} + [E|Y|^p]^{\frac{1}{p}}, \qquad \text{kaikilla } p \ge 1. \tag{6.5}$$

(Jos jokin odotusarvoista ei ole reaalilukuna olemassa, niin vastaavalle termille käytetään arvoa ∞). Minkowskin epäyhtälö seuraa kohtalaisen helposti kolmioepäyhtälöstä sekä Hölderin epäyhtälöstä (HT).

6.4 Cauchyn-Schwarzin epäyhtälö, kovarianssi ja korrelaatio

Cauchyn-Schwarzin epäyhtälöksi sanotaan Hölderin epäyhtälön sitä tapausta, jossa $p = q = 2$, jolloin saadaan

$$E|XY| \le \sqrt{EX^2}\,\sqrt{EY^2}.$$

Tämä epäyhtälö voi joissakin tapauksissa olla voimassa muodossa $E|XY| \le \infty$, mikä ei kerro mitään mielenkiintoista. Jos sekä $EX^2 < \infty$ että $EY^2 < \infty$, niin silloin yläraja on äärellinen, ja saadaan tulos

$$|E(XY)| \le E|XY| \le \sqrt{EX^2}\,\sqrt{EY^2}. \tag{6.6}$$

Tässä ensimmäinen epäyhtälö seurasi lauseesta 4.4.

Oletetaan nyt, että satunnaismuuttujilla X ja Y on äärelliset varianssit, $\sigma_X^2 = \operatorname{var} X$ ja $\sigma_Y^2 = \operatorname{var} Y$. Niiden kovarianssi on

$$\operatorname{cov}(X, Y) = E[(X - EX)(Y - EY)].$$

Tämä odotusarvo on äärellinen, sillä kun Cauchyn-Schwarzin epäyhtälöä sovelletaan satunnaismuuttujien $X - EX$ ja $Y - EY$ tuloon, niin saadaan epäyhtälö

$$|\operatorname{cov}(X, Y)| \leq \sigma_X \, \sigma_Y. \tag{6.7}$$

Tässä ylärajana on muuttujien keskihajontojen tulo.

Mikäli molemmat keskihajonnat σ_X ja σ_Y ovat aidosti positiivisia, niin X:n ja Y:n **korrelaatiokerroin** (jota merkitään usein $\operatorname{corr}(X, Y)$ tai ρ_{XY}) määritellään kaavalla

$$\operatorname{corr}(X, Y) = \rho_{XY} = \rho = \frac{\operatorname{cov}(X, Y)}{\sigma_X \, \sigma_Y}. \tag{6.8}$$

Cauchyn-Scwarzin epäyhtälön nojalla

$$|\operatorname{corr}(X, Y)| \leq 1. \tag{6.9}$$

Jos $\operatorname{corr}(X, Y) > 0$, jolloin myös $\operatorname{cov}(X, Y) > 0$, niin sanotaan että X ja Y ovat *positiivisesti korreloituneita*. Jos $\operatorname{corr}(X, Y) < 0$, jolloin myös $\operatorname{cov}(X, Y) < 0$, niin sanotaan että X ja Y ovat *negatiivisesti korreloituneita*. Jos $\operatorname{corr}(X, Y) = 0$, jolloin myös $\operatorname{cov}(X, Y) = 0$, niin sanotaan, että X ja Y eivät korreloi (lineaarisesti) tai että ne ovat (lineaarisesti) korreloimattomia.

6.5 Momenttiemäfunktiota koskevia epäyhtälöitä

Palautetaan mieleen määritelmät. Satunnaismuuttujan X momenttiemäfunktio M_X ja sen kumulanttiemäfunktio ovat

$$M_X(t) = E\mathrm{e}^{tX}, \quad K_X(t) = \ln M_X(t)$$

niissä pisteissä, joissa odotusarvo on äärellinen.

Lause 6.9 (Emäfunktioiden ominaisuuksia)**.** *a) Häntätodennäköisyyksille pätevät seuraavat ylärajat kaikilla a,*

$$P(X \geq a) \leq \inf_{t>0} \mathrm{e}^{-ta} M_X(t), \qquad \text{ja} \qquad P(X \leq a) \leq \inf_{t<0} \mathrm{e}^{-ta} M_X(t).$$

b) Kumulanttiemäfunktio ja momenttiemäfunktio ovat konvekseja funktioita.

Todistus. Kohta (a): kun $t > 0$ niin funktio $x \mapsto \mathrm{e}^{tx}$ on aidosti kasvava, joten Markovin epäyhtälön avulla

$$P(X \geq a) = P(\mathrm{e}^{tX} \geq \mathrm{e}^{ta}) \leq \frac{E\mathrm{e}^{tX}}{\mathrm{e}^{ta}}.$$

Koska tämä arvio pätee kaikilla $t > 0$, niin se pätee vielä silloinkin, kun yläraja korvataan sen minimillä alueessa $t > 0$. Toinen epäyhtälöistä todistetaan soveltamalla ensimmäistä sm:an $Y = -X$.

Kumulanttiemäfunktion konveksisuus seuraa Hölderin epäyhtälöstä (HT). Tämän jälkeen momenttiemäfunktion konveksisuus seuraa siitä, että e^x on kasvava ja konveksi funktio. □

Luku 7

Kaksiulotteinen jakauma

7.1 Jatkuva kaksiulotteinen jakauma

Määritelmä 7.1 (Jatkuva (yhteis)jakauma, yhteistiheysfunktio)**.** Satunnaisvektorilla (X, Y) on jatkuva jakauma, eli satunnaismuuttujilla X ja Y on jatkuva yhteisjakauma, mikäli on olemassa $\mathbb{R}^2$:lla määritelty funktio $f \geq 0$ siten, että

$$P((X, Y) \in B) = \iint_B f(x, y)\, \mathrm{d}x\, \mathrm{d}y, \qquad \text{kaikille } B \subset \mathbb{R}^2. \tag{7.1}$$

Tällöin $f = f_{X,Y}$ on sv:n (X, Y) tiheysfunktio, eli satunnaismuuttujien X ja Y *yhteistiheysfunktio* (lyh. ytf) (engl. *joint (probability) density function, joint pdf*).

Kaava (7.1) ilmaistaan toisinaan heuristisella merkinnällä

$$P(X \in \mathrm{d}x,\, Y \in \mathrm{d}y) = f(x, y)\, \mathrm{d}x\, \mathrm{d}y.$$

Tässä merkinnän $P(X \in \mathrm{d}x,\, Y \in \mathrm{d}y)$ voidaan ajatella tarkoittavan samaa kuin $P(x \leq X \leq x + \mathrm{d}x,\, y \leq Y \leq y + \mathrm{d}y)$, jossa $\mathrm{d}x$ ja $\mathrm{d}y$ ovat hyvin pieniä positiivisia lukuja. Kuten yksiulotteisessa myös kaksiulotteisessa tapauksessa tiheysfunktion arvolle pisteessä (x, y) voidaan antaa frekvenssitulkinta.

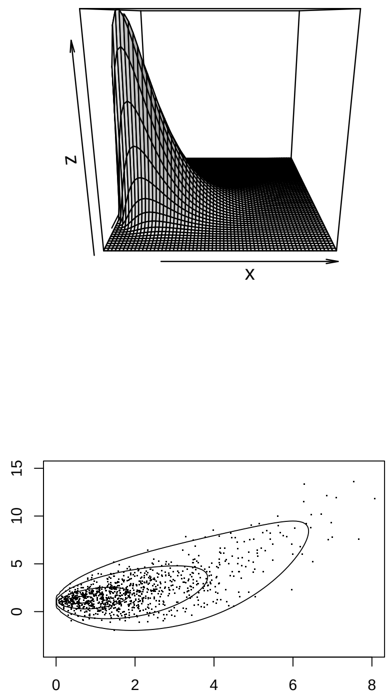
Kaksiulotteista jatkuvaa jakaumaa voidaan havainnollistaa erilaisilla tavoilla. Yksi mahdollisuus on piirtää perspektiivikuva funktiosta $z = f(x, y)$. Toinen mahdollisuus on piirtää xy-tasoon funktion *tasa-arvokäyriä* (engl. *level curve*, *contour line*). Tasoon c liittyvä funktion f tasa-arvokäyrä koostuu niistä pisteistä (x, y), joissa $f(x, y) = c$. Eri vakion c arvoilla saadaan erilaisia tasa-arvokäyriä. Kolmas mahdollisuus on simuloida pisteitä $(X_i, Y_i), i = 1, \ldots, N$ kyseisestä jakaumasta ja piirtää pisteet tasoon. Kuvassa 7.1 esitellään näitä tapoja.

Kaksiulotteisen jatkuvan jakauman tapauksessa joudutaan laskemaan tasointegraaleja annettujen integrointijoukkojen yli. Joukon B yli laskettu tasointegraali funktiosta g tarkoittaa funktion $1_B\, g$ integraalia koko tason $\mathbb{R}^2$ yli. Sille käytetään mm. seuraavia merkintöjä,

$$\int_B g = \int_B g(x, y)\, \mathrm{d}(x, y) = \iint_B g(x, y)\, \mathrm{d}x\, \mathrm{d}y = \iint_{R^2} 1_B(x, y)\, g(x, y)\, \mathrm{d}x\, \mathrm{d}y.$$

Tavallisesti tasointegraalit lasketaan iteroituina integraaleina. Tässä yhteydessä sovelletaan Fubinin lausetta (jota ei todisteta tällä kurssilla). Seuraavan lauseen muotoilu on pätevä Lebesguen integraalille.

Kuva 7.1 Erilaisia tapoja havainnollistaa jatkuvaa jakaumaa: (a) perspektiivikuva, (b) tiheysfunktion tasa-arvokäyriä sekä jakaumasta simuloitu pisteparvi.



Lause 7.1 (Fubini)**.** *Seuraavissa kahdessa tapauksissa tasointegraali voidaan laskea iteroituna integraalina, eli alla olevat yhtälöt ovat voimassa,*

$$\begin{aligned}\int_B g(x,y)\,\mathrm{d}(x,y) &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} 1_B(x,y)\,g(x,y)\,\mathrm{d}x\right)\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} 1_B(x,y)\,g(x,y)\,\mathrm{d}y\right)\mathrm{d}x\end{aligned} \tag{7.2}$$

(a) Jos $g \ge 0$, jolloin integraalien yhteinen arvo voi olla ∞.

(b) Jos $\int_B |g| < \infty$, jolloin integraalien yhteinen arvo on äärellinen. (Huomaa, että integraali $\int_B |g|$ voidaan aina laskea a-kohdan perusteella iteroituna integraalina.)

Esimerkiksi, jos B voidaan esittää muodossa

$$B = \{(x,y) : a < x < b,\ c(x) < y < d(x)\},$$

ja $\int_B |g| < \infty$ tai $g \ge 0$, niin

$$\int_B g = \iint_B g(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_a^b \left(\int_{c(x)}^{d(x)} g(x,y)\,\mathrm{d}y\right)\mathrm{d}x.$$

Sama tulos saadaan, vaikka joukon B määritelmässä yksi tai useampi relaatioista $<$ korvataan relaatiolla $\le$. Tälle iteroidulle integraalille voidaan käyttää myös muita merkintöjä, kuten esim.

$$\int_B g = \int_{x=a}^b \int_{y=c(x)}^{d(x)} g(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_a^b \mathrm{d}x \int_{c(x)}^{d(x)} g(x,y)\,\mathrm{d}y$$

Jos taas B voidaan esittää muodossa

$$B = \{(x,y) : c < y < d,\ a(y) < x < b(y)\},$$

ja $\int_B |g| < \infty$ tai $g \ge 0$, niin tasointegraali voidaan laskea iteroituna integraalina

$$\begin{aligned}\iint_B g(x,y)\,\mathrm{d}x\,\mathrm{d}y &= \int_c^d \left(\int_{a(y)}^{b(y)} g(x,y)\,\mathrm{d}x\right)\mathrm{d}y \\ &= \int_{y=c}^d \int_{x=a(y)}^{b(y)} g(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_c^d \mathrm{d}y \int_{a(y)}^{b(y)} g(x,y)\,\mathrm{d}x\end{aligned}$$

Useimmat sovelluksissa esiintyvät integrointijoukot voidaan osittaa erillisiksi paloiksi, joille voidaan käyttää jompaa kumpaa esitystä. Koko alkuperäisen joukon yli laskettu tasointegraali voidaan laskea summana näiden palojen yli lasketuista tasointegraaleista.

Huomautus. Myös merkintä

$$\int_a^b \int_c^d g(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

on yleinen. Sen yhteydessä ei ole itsestään selvää, kumpaan muuttujaan kumpikin integrointijoukko liitty: joissakin lähteissä merkinnällä tarkoitetaan joukon $(x,y) \in (a,b) \times (c,d)$ yli laskettua integraalia, ja toisissa taas joukon $(x,y) \in (c,d) \times (a,b)$ yli laskettua integraalia.

Kaksiulotteisen jakauman tiheysfunktio ei ole yksikäsitteinen, mutta melkein yksikäsitteinen. Jos f on tiheysfunktio, ja g on saman jakauman tiheysfunktio, niin $f = g$ melkein kaikkialla, eli $f(x,y) = g(x,y)$ kaikkialla muualla paitsi nollamittaisessa joukossa. Kaksiulotteisessa avaruudessa esim. kaikkien yksiulotteisten käyrien kuvaajat ovat nollamittaisia.

Lause 7.2. *Olkoon satunnaisvektorilla (X,Y) jatkuva jakauma. Tällöin reunajakaumat ovat myös jatkuvia, ja niiden tiheysfunktiot saadaan integroimalla toinen muuttuja pois yhteistiheysfunktiosta,*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y, \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x.$$

Todistus. Jos $B \subset \mathbb{R}$, niin

$$P(X \in B) = P((X,Y) \in B \times \mathbb{R}) = \int_B \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y \right) \mathrm{d}x.$$

Y:n tiheysfunktion lauseke johdetaan vastaavasti. □

Esimerkki 7.1. Siitä, että reunajakaumat ovat jatkuvia, ei seuraa, että yhteisjakauma olisi jatkuva. Otetaan sm X, jolla on jatkuva jakauma tf:lla f_X. Määritellään sm Y siten, että $Y = X$, ts.

$$Y(\omega) = X(\omega), \qquad \text{kaikilla } \omega \in \Omega.$$

Tällöin Y:n jakauma on sama kuin X:n jakauma, joten molemmat reunajakaumat ovat jatkuvia. Sv:n (X,Y) jakauma on keskittynyt lävistäjälle

$$B = \{(x,y) : x = y\},$$

mutta mille tahansa funktiolle $f : \mathbb{R}^2 \to \mathbb{R}$ on $\int_B f = 0$, sillä lävistäjä B on yksiulotteisena joukkona nollamittainen $\mathbb{R}^2$:ssa. Tämän takia sv:lla (X,Y) ei voi olla jatkuva jakauma, eli yhteisjakauma ei ole jatkuva. △

Jatkuvan yhteisjakauman tapauksessa ykf:llä on esitys,

$$F_{X,Y}(x,y) = P((X,Y) \in (-\infty,x] \times (-\infty,y]) = \int_{-\infty}^{x} \mathrm{d}s \int_{-\infty}^{y} f_{X,Y}(s,t)\,\mathrm{d}t.$$

Derivoidaan tämä yhtälö ensin muuttujan x suhteen,

$$\frac{\partial}{\partial x} F_{X,Y}(x,y) = \int_{-\infty}^{y} f_{X,Y}(x,t)\,\mathrm{d}t,$$

ja sitten vielä muuttujan y suhteen,

$$\frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x,y) = f_{X,Y}(x,y),$$

ja tämä tulos pätee ainakin niissä pisteissä (x,y), joissa $f_{X,Y}$ on jatkuva. Seuraavan lauseen todistus vaatisi työkaluja, joita meillä ei ole käytettävissä.

Lause 7.3. *Jos satunnaisvektorilla (X,Y) on jatkuva jakauma, niin jakauman tiheysfunktioksi voidaan valita*

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x,y).$$

Tämä toisen kertaluvun sekaderivaatta on olemassa melkein kaikilla (x,y), ja niissä pisteissä, joissa sekaderivaatta ei ole määritelty, voidaan käyttää mielivaltaista määrittelyä $f_{X,Y}$:lle.

Huomautus. Mielivaltaisesta annetusta kaksiulotteisen jakauman kertymäfunktiosta voi olla hankala tunnistaa, onko se jatkuvan jakauman kertymäfunktio. Periaatteessa pitäisi tutkia, kelpaako ko. toisen kertauluvun sekaderivaattaa jakauman tiheysfunktioksi.

Lause 7.4. *Jos $f : \mathbb{R}^2 \to \mathbb{R}$ on ei-negatiivinen, ja sen integraali koko tason yli on yksi, niin se on jonkin kaksiulotteisen satunnaisvektorin tiheysfunktio.*

7.2 Tasajakauma alueessa

Määritelmä 7.2 (Alueen A tasajakauma)**.** Olkoon $A \subset \mathbb{R}^2$ joukko, jonka pinta-ala

$$m(A) = \iint_{R^2} 1_A(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

toteuttaa ehdot $0 < m(A) < \infty$. Sv (X,Y) noudattaa tasajakaumaa joukossa A, mikäli sillä on ytf

$$f(x,y) = 1_A(x,y)/m(A).$$

Tasajakauman tiheysfunktio häviää kyseisen alueen A ulkopuolella, ja sen pisteissä $(x,y) \in A$ sillä on nollasta poikkeava vakioarvo. Olkoon vektorilla (X,Y) tasajakauma alueessa A. Jos $B \subset \mathbb{R}^2$, niin

$$P((X,Y) \in B) = \iint_B f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \frac{m(A \cap B)}{m(A)}.$$

Erityisesti $P((X,Y) \in \mathbb{R}^2) = 1$, kuten pitää ollakin.

Huomaa, että esimerkiksi suljetun neliön $[0,1] \times [0,1]$ ja avoimen neliön $(0,1) \times (0,1)$ tasajakaumat ovat samat, sillä neliön reunan pinta-ala on nolla eli se on nollamittainen.

Seuraavassa esimerkissä tarkastellaan tasajakaumaa ei-negatiivisen ja integroituvan funktion h kuvaajan ja x-akselin väliin jäävässä alueessa. Huomaa, että x-koordinaatin reunajakaumalla on tällöin tiheysfunktio, joka on verrannollinen funktioon h. Toisin sanoen, sm:lla X on jatkuva jakauma, jonka määrää normalisoimaton tiheysfunktio h.

Esimerkki 7.2. Olkoon $h(x) = \exp(-\frac{1}{2}x^2)$, ja olkoon sv:lla (X,Y) tasajakauma h:n kuvaajan alla, eli joukossa

$$A = \{(x,y) : 0 < y < h(x)\}.$$

Tämän joukon pinta-ala on

$$m(A) = \iint 1_A(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{\infty} \mathrm{d}x \int_0^{h(x)} 1\ \mathrm{d}y = \int_{-\infty}^{\infty} h(x)\,\mathrm{d}x = \sqrt{2\pi},$$

joten ytf on

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

jossa käytetään merkintää

$$1\{0 < y < h(x)\} = 1_{\{(u,v):0<v<h(u)\}}(x,y) = \begin{cases} 1, & \text{kun } 0 < y < h(x) \\ 0, & \text{muuten.} \end{cases}$$

X:n reunatiheysfunktio on

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ \mathrm{d}y = \int_0^{h(x)} \frac{1}{\sqrt{2\pi}}\ \mathrm{d}y = \frac{1}{\sqrt{2\pi}}\ \exp(-\frac{1}{2}x^2)$$

joten X:n reunajakauma on $N(0,1)$.

Mikä on Y:n reunatiheysfunktio? Kaikilla x on $0 < h(x) \le 1$, joten epäyhtälö

$$0 < y < h(x)$$

voi toteutua vain, kun $0 < y < 1$. Jos $0 < y < 1$, niin epäyhtälö

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

saadaan helposti ratkaistua x:n suhteen, ja tulos on

$$-\sqrt{-2\ln y} < x < \sqrt{-2\ln y}.$$

(Huomaa, että $-\ln y > 0$, kun $0 < y < 1$.) Siis ytf voidaan esittää myös kaavalla

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

josta on helppo laskea Y:n reunatiheysfunktio,

$$f_Y(y) = \frac{2}{\sqrt{\pi}}\sqrt{-\ln y}, \qquad 0 < y < 1.$$

$\triangle$

7.3 Riippumattomuus

Palautetaan mieleen määritelmä. Sm:t X ja Y ovat riippumattomia (merkintä $X \perp\!\!\!\perp Y$), jos

$$P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B), \qquad \text{kaikilla } A, B \subset \mathbb{R}.$$

Jaksossa 3.5 todettiin, että

$$X \perp\!\!\!\perp Y \quad \Longleftrightarrow \quad F_{X,Y}(x,y) = F_X(x)\, F_Y(y) \text{ kaikilla } (x,y).$$

Seuraavaksi näytetään, että riippumattomien satunnaismuuttujien yptnf tai ytf voidaan esittää reunajakaumien vastaavien funktioiden tulona, jos yhteisjakauma on joko diskreetti tai jatkuva.

Huomautus. Huomaa, että riippumattomien satunnaismuuttujien yhteiskertymäfunktio ei ole funktioiden F_X ja F_Y tavanomainen tulo

$$u \mapsto F_X(u)\, F_Y(u),$$

vaan niiden ns. *tensoritulo*

$$(u,v) \mapsto F_X(u)\, F_Y(v).$$

Samaan tapaan, riippumattomien satunnaismuuttujien yptnf/ytf voidaan esittää reunajakaumien ptnf/tf:ien tulona, mikäli ymmärretään, että kyseessä on tällainen funktioiden tensoritulo.

Lause 7.5. *Olkoon satunnaisvektorilla (X,Y) joko diskreetti tai jatkuva yhteisjakauma, jonka yptnf tai ytf on $f_{X,Y}$.*

a) Jos $X \perp\!\!\!\perp Y$, niin ytf:ksi kelpaa $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$.

b) Jos $f_{X,Y}$ voidaan esittää tulomuodossa

$$f_{X,Y}(x,y) = g(x)\,h(y), \qquad \forall x, y,$$

jossa $g \geq 0$ ja $h \geq 0$ ovat yhden muuttujan funktioita, niin $X \perp\!\!\!\perp Y$.

Todistus. **(a):** Diskreetti tapaus todistettiin jaksossa 3.5, ja nyt todistetaan jatkuva tapaus. Oletetaan, että yhteisjakauma on jatkuva ja että $X \perp\!\!\!\perp Y$. Tarkistamme, että ykf:n arvo mielivaltaisessa pisteessä (x, y) saadaan integroimalla väitteessä ilmoitettua ytf:a. Väite seuraa tästä sillä perusteella, että ykf määrää jakauman. Riippumattomuuden nojalla kaikilla $x, y \in \mathbb{R}$ on

$$\begin{aligned} F_{X,Y}(x,y) = F_X(x)\,F_Y(y) &= \int_{-\infty}^{x} f_X(s)\,\mathrm{d}s \int_{-\infty}^{y} f_Y(t)\,\mathrm{d}t \\ &= \int_{-\infty}^{x} f_X(s) \left(\int_{-\infty}^{y} f_Y(t)\,\mathrm{d}t \right) \mathrm{d}s \\ &= \int_{-\infty}^{x} \mathrm{d}s \int_{-\infty}^{y} f_X(s)\,f_Y(t)\,\mathrm{d}t. \end{aligned}$$

Siis $f_X(x)\,f_Y(y)$ on (X, Y):n ytf.

(b): Esitetään todistus jatkuvassa tapauksessa; todistuksen idea pysyy samana diskreetissä tapauksessa. Nyt oletetaan, että ytf voidaan esittää tulomuodossa $f_{X,Y}(x, y) = g(x)\,h(y)$. Olkoon

$$c = \int_{-\infty}^{\infty} g(x)\,\mathrm{d}x, \qquad d = \int_{-\infty}^{\infty} h(y)\,\mathrm{d}y.$$

Tällöin

$$1 = \iint_{R^2} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{\infty} g(x)\,\mathrm{d}x \int_{-\infty}^{\infty} h(y)\,\mathrm{d}y = cd.$$

X:n reunatiheysfunktio on

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y = g(x)d = g(x)/c,$$

ja Y:n reunatiheysfunktio on

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x = ch(y) = h(y)/d.$$

Ts. reunatiheysfunktiot f_X ja f_Y ovat tekijät g ja h normalisoituna tiheysfunktioiksi. Kaikilla x, y ykf on reunajakaumien kertymäfunktioiden tulo, sillä

$$\begin{aligned} F_{X,Y}(x,y) &= \int_{-\infty}^{x} \mathrm{d}s \int_{-\infty}^{y} f_{X,Y}(s,t)\,\mathrm{d}t = \int_{-\infty}^{x} \mathrm{d}s \int_{-\infty}^{y} \frac{g(s)}{c}\,\frac{h(t)}{d}\,\mathrm{d}t \\ &= \int_{-\infty}^{x} f_X(s)\,\mathrm{d}s \int_{-\infty}^{y} f_Y(t)\,\mathrm{d}t = F_X(x)\,F_Y(y), \end{aligned}$$

joten $X \perp\!\!\!\perp Y$. □

Toisinaan tahdotaan todistaa, että sm:t X ja Y eivät ole riippumattomia. Mikäli X ja Y ovat diskreettejä, niin niiden riippuvuuden voi todistaa esittämällä yksi piste (x, y), jossa kertolaskuominaisuus ei pidä paikkaansa, eli

$$f_{X,Y}(x,y) \neq f_X(x)\,f_Y(y).$$

Tällöin ollaan näytetty, että $P(X \in A, Y \in B)$ on erisuuri kuin $P(X \in A)\,P(Y \in B)$, kun valitaan $A = \{x\}$ ja $B = \{y\}$.

Jatkuvan yhteisjakauman tilanteessa tarkistus on mutkikkaampaa, sillä moni hieman erilainen funktio voi olla saman jakauman tiheysfunktio.

Esimerkki 7.3. Olkoon vektorilla (X, Y) tasajakauma kolmiossa

$$C = \{(x, y) : x \geq 0, y \geq 0, x + y \leq 1\}.$$

Ovatko X ja Y riippumattomia?

Tämän kolmion pinta-ala on $\frac{1}{2}$, joten ytf on

$$f_{X,Y}(x, y) = 2, \quad \text{kun } (x, y) \in C.$$

Äkkiseltään voisi näyttää, että ytf on tulomuotoa $g(x)\,h(y)$, mutta totuus paljastuu, kun kirjoitamme ehdon $(x, y) \in C$ indikaattorifunktioiden avulla:

$$f_{X,Y}(x, y) = 2\,1_{\{x\geq 0\}}\,1_{\{y\geq 0\}}\,1_{\{x+y\leq 1\}}$$

Tässä viimeinen termi ei todellakaan ole tulomuotoa.

Jos valitaan $A = B = (\frac{1}{2}, 1)$, niin $(A \times B) \cap C = \emptyset$, joten

$$P((X, Y) \in A \times B) = 0,$$

mutta $P(X \in A) > 0$ ja $P(Y \in B) > 0$, joten

$$0 = P(X \in A, Y \in B) \neq P(X \in A)\,P(Y \in B) > 0.$$

Olemme näyttäneet suoraan riipumattomuuden määritelmän perusteella, että X ja Y eivät ole riippumattomia. △

7.4 Satunnaisvektorin muunnoksen odotusarvo

Lause 7.6. *Olkoon (X, Y) sv, jolla on diskreetti jakauma, ja olkoon $Z = g(X, Y)$ jokin sen reaaliarvoinen muunnos. Tällöin*

$$EZ = \sum_{x,y} g(x, y)\,f_{X,Y}(x, y),$$

mikäli summa suppenee itseisesti.

Todistus. Sovellamme apulausetta 4.1 seuraavaan perusjoukon ositukseen. Olkoon $A = \{(x_i, y_i) : i \geq 1\}$ sv:n (X, Y) korkeintaan numeroituva arvojoukko. Tällöin joukot

$$\{(X, Y) \notin A\}, \quad (\{X = x_i,\, Y = y_i\})_{i\geq 1}$$

osittavat perusjoukon. Nyt

$$g(X, Y)(\omega) = g(X(\omega), Y(\omega)) = g(x_i, y_i), \qquad \text{kun } \omega \in \{X = x_i,\, Y = y_i\},$$

joten

$$EZ = \sum_{i\geq 1} g(x_i, y_i)\,P(X = x_i,\, Y = y_i) = \sum_{i\geq 1} g(x_i, y_i)\,f_{X,Y}(x_i, y_i). \qquad \square$$

Jatkuvan yhteisjakauman tapauksessa odotusarvo $Eg(X,Y)$ saadaan laskettua integraalina. Sivuutamme tämän tapauksen todistuksen.

Lause 7.7. *Olkoon (X,Y) sv, jolla on jatkuva yhteisjakauma, ja olkoon $Z = g(X,Y)$ jokin sen reaaliarvoinen muunnos. Tällöin*

$$EZ = \iint_{R^2} g(x,y)\, f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

mikäli integraali suppenee itseisesti.

Yhteenvetona kahdesta edellisestä lauseesta saadaan kaksiulotteinen versio *tiedostamattoman tilastotieteilijän laista* (vrt. jakso 4.4), nimittäin

$$Eg(X,Y) = \begin{cases} \sum_{x,y} g(x,y)\, f_{X,Y}(x,y), & \text{jos diskreetti yhteisjakauma} \\ \iint_{R^2} g(x,y)\, f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y, & \text{jos jatkuva yhteisjakauma.} \end{cases}$$

Tämä kaava pitää paikkansa sillä oletuksella, että kyseinen odotusarvo on äärellisenä olemassa, eli että kyseinen summa tai integraali suppenee itseisesti.

Tämä kaksiulotteinen versio on yhteensopiva aikaisemman yksiulotteisen version kanssa. Esim. jos käsitellään diskreettiä yhteisjakaumaa sekä yhden muuttujan funktiota g, niin $Eg(X)$ voidaan laskea tiedostamattoman tilastotieteilijän lain kaksiulotteisen version avulla seuraavasti

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

jossa viimeinen kaava on jaksossa 4.4 kerrottu tiedostamattoman tilastotieteilijän lain yksiulotteinen versio.

Esimerkki 7.4. Jos satunnaismuuttujilla X ja Y on jatkuva yhteisjakauma, ja ne ovat integroituvia, niin kaikilla vakioilla $a, b \in \mathbb{R}$

$$\begin{aligned} E(aX + bY) &= \iint (ax + by)\, f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= a \int x \left(\int f_{X,Y}(x,y)\,\mathrm{d}y \right) \mathrm{d}x + b \int y \left(\int f_{X,Y}(x,y)\,\mathrm{d}x \right) \mathrm{d}y \\ &= a \int x\, f_X(x)\,\mathrm{d}x + b \int y\, f_Y(y)\,\mathrm{d}y = a\,EX + b\,EY. \end{aligned}$$

Tässä integraali tarkoittaa koko reaaliakselin yli laskettua integraalia. Olemme johtaneet tässä tapauksessa sen ennestään tutun tosiseikan, että odotusarvo on lineaarinen operaattori. $\triangle$

Huomautus. Funktiolle $g(x,y) = ax + by$ pätee $Eg(X,Y) = g(EX, EY)$, mutta muille funktioille tällainen identiteetti ei yleensä päde.

Esimerkki 7.5. Jos $X \perp\!\!\!\perp Y$, niin tällöin $g(X) \perp\!\!\!\perp h(Y)$ kaikilla funktioilla $g, h : \mathbb{R} \to \mathbb{R}$, joten

$$E(g(X)\,h(Y)) = E(g(X))\,E(h(Y)),$$

mikäli kyseiset odotusarvot ovat olemassa. Jos sv:lla (X,Y) on jatkuva jakauma, niin $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$, joten asian näkee myös laskulla

$$\begin{aligned} E(g(X)\,h(Y)) &= \iint g(x)\,h(y) f_X(x)\,f_Y(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \left(\int g(x)\,f_X(x)\,\mathrm{d}x\right)\left(\int h(y)\,f_Y(y)\,\mathrm{d}y\right) = E(g(X))\,E(h(Y)). \end{aligned}$$

△

7.5 Kovarianssi ja muita yhteismomentteja

Määritelmä 7.3. Olkoot $m,n \geq 0$ kokonaislukuja. Mikäli odotusarvo

$$E[X^m Y^n]$$

on olemassa, sitä kutsutaan kertalukujen (m,n) yhteismomentiksi (tai tulomomentiksi). Mikäli odotusarvo

$$E[(X-EX)^m (Y-EY)^n]$$

on olemassa, sitä sanotaan kertalukujen (m,n) keskusmomentiksi.

Näistä momenteista ehdottomasti tärkeimmät ovat reunajakaumien momentit EX^m sekä EY^n (eli kertalukujen $(m,0)$ ja $(0,n)$ yhteismomentit) sekä kertaluvun $(1,1)$ keskusmomentti, joka tunnetaan paremmin nimellä X:n ja Y:n kovarianssi,

$$\operatorname{cov}(X,Y) = E[(X-EX)(Y-EY)].$$

Huomautus. Kovarianssia käsiteltiin jo jaksossa 4.6. Kovarianssi on symmetrinen, $\operatorname{cov}(X,Y) = \operatorname{cov}(Y,X)$, ja varianssi saadaan lausuttua kovarianssin avulla, sillä $\operatorname{var} X = \operatorname{cov}(X,X)$. Odotusarvon lineaarisuuden avulla nähtiin, että

$$\operatorname{cov}(X,Y) = E(XY) - (EX)\,(EY).$$

Lause 7.8. *Kovarianssi on bilineaarinen, eli molempien argumenttiensa suhteen lineaarinen operaattori. Ts. jos $a_1, a_2 \in \mathbb{R}$ ovat vakioita ja X_1, X_2 ja Z ovat satunnaismuuttujia, niin*

$$\begin{aligned} \operatorname{cov}(a_1X_1 + a_2X_2, Z) &= a_1 \operatorname{cov}(X_1, Z) + a_2 \operatorname{cov}(X_2, Z) \\ \operatorname{cov}(Z, a_1X_1 + a_2X_2) &= a_1 \operatorname{cov}(Z, X_1) + a_2 \operatorname{cov}(Z, X_2) \end{aligned}$$

Jos lisäksi $b_1, b_2 \in \mathbb{R}$ ja Y_1, Y_2 ovat satunnaismuuttujia, niin

$$\operatorname{cov}(a_1X_1 + a_2X_2, b_1Y_1 + b_2Y_2) = \sum_{i=1}^{2}\sum_{j=1}^{2} a_i\,b_j\,\operatorname{cov}(X_i, Y_j).$$

Nämä kaavat pitävät paikkansa, mikäli kyseiset odotusarvot ovat olemassa.

Todistus. Odotusarvon lineaarisuuden nojalla

$$E(a_1X_1 + a_2X_2) = a_1EX_1 + a_2EX_2,$$

joten
$$a_1X_1 + a_2X_2 - E(a_1X_1 + a_2X_2) = a_1(X_1 - EX_1) + a_2(X_2 - EX_2).$$
Siis odotusarvon lineaarisuuden nojalla,

$$\begin{aligned}\operatorname{cov}(a_1X_1 + a_2X_2, Z) &= E\left[(a_1(X_1 - EX_1) + a_2(X_2 - EX_2))\,(Z - EZ)\right] \\ &= a_1E[(X_1 - EX_1)(Z - EZ)] + a_2E[(X_2 - EX_2)(Z - EZ)] \\ &= a_1 \operatorname{cov}(X_1, Z) + a_2 \operatorname{cov}(X_2, Z).\end{aligned}$$

Kovarianssin symmetrisyyden nojalla

$$\operatorname{cov}(Z, \sum_i a_iX_i) = \operatorname{cov}(\sum_i a_iX_i, Z) = \sum_i a_i \operatorname{cov}(X_i, Z) = \sum_i a_i \operatorname{cov}(Z, X_i).$$

Kovarianssi on nyt todistettu molempien argumenttiensa suhteen lineaariseksi. Näin ollen

$$\begin{aligned}\operatorname{cov}(\sum_i a_iX_i, \sum_j b_jY_j) &= \sum_i a_i \operatorname{cov}(X_i, \sum_j b_jY_j) = \sum_i a_i \sum_j b_j \operatorname{cov}(X_i, Y_j) \\ &= \sum_i \sum_j a_ib_j \operatorname{cov}(X_i, Y_j).\end{aligned}$$ □

Esimerkki 7.6. Lineaarikombinaation varianssi.

$$\begin{aligned}\operatorname{var}(aX + bY) &= \operatorname{cov}(aX + bY, aX + bY) \\ &= a \operatorname{cov}(X, aX + bY) + b \operatorname{cov}(Y, aX + bY) \\ &= a^2 \operatorname{cov}(X, X) + ab \operatorname{cov}(X, Y) + ab \operatorname{cov}(Y, X) + b^2 \operatorname{cov}(Y, Y) \\ &= a^2 \operatorname{var}(X) + 2ab \operatorname{cov}(X, Y) + b^2 \operatorname{var}(Y).\end{aligned}$$

△

7.6 Paras lineaarinen ennuste

Joissakin tilanteissa sm:n Y arvoa ei havaita suoraan, mutta jollakin tavalla sm:an Y liittyvän sm:n X arvo on mahdollista havaita. Tällöin saatetaan haluta ennustaa Y:n arvo X:n arvon avulla. Ennusteelle voidaan esimerkiksi valita lineaarinen muoto $a + bX$, tai yhtäpitävästi muoto $\alpha + \beta(X - EX)$. Tällöin jää vielä ratkaistavaksi kysymys, miten kertoimet pitäisi valita jotta ennuste olisi mahdollisimman hyvä. Mikäli ennusteen hyvyyden mittana pidetään mahdollisimman pientä keskineliövirhettä
$$E[(Y - (\alpha + \beta(X - EX)))^2] = \min!$$
niin osoittautuu, että paras lineaarinen ennuste löytyy yksinkertaisella laskulla, jota ryhdymme seuraavaksi laskemaan. Oletamme, että

$$EX^2 < \infty, \quad EY^2 < \infty, \quad \operatorname{var} X > 0, \quad \operatorname{var} Y > 0.$$

Merkitään $\mu_X = EX$ ja $\mu_Y = EY$. Jos α ja β ovat vakioita, niin ennuste $\alpha + \beta(X - \mu_X)$ poikkeaa totuudesta Y määrän

$$Z = Y - \alpha - \beta(X - \mu_X).$$

Ennustevirheen odotusarvo ja varianssi ovat

$$\begin{aligned} EZ &= \mu_Y - \alpha - \beta\,(\mu_X - \mu_X) = \mu_Y - \alpha \\ \operatorname{var} Z &= \operatorname{var} Y + \beta^2 \operatorname{var} X - 2\beta \operatorname{cov}(X, Y) \end{aligned}$$

joten keskineliövirhe on

$$\begin{aligned} EZ^2 &= (EZ)^2 + \operatorname{var} Z \\ &= (\mu_Y - \alpha)^2 + \operatorname{var} Y + \beta^2 \operatorname{var} X - 2\beta \operatorname{cov}(X, Y). \end{aligned}$$

Seuraavaksi täydennämme kerrointa β sisältävät termit neliöksi,

$$\begin{aligned} EZ^2 &= (\mu_Y - \alpha)^2 + \operatorname{var} Y + \operatorname{var} X \left(\beta^2 - 2\beta \frac{\operatorname{cov}(X, Y)}{\operatorname{var} X} + \frac{(\operatorname{cov}(X, Y))^2}{(\operatorname{var} X)^2}\right) - \frac{(\operatorname{cov}(X, Y))^2}{\operatorname{var} X} \\ &= \operatorname{var} Y + (\mu_Y - \alpha)^2 + \operatorname{var} X \left(\beta - \frac{\operatorname{cov}(X, Y)}{\operatorname{var} X}\right)^2 - \frac{(\operatorname{cov}(X, Y))^2}{\operatorname{var} X} \end{aligned}$$

Viimeisessä muodossa vain toinen ja kolmas termi riippuvat muuttujien α tai β arvoista, ja nämä termit ovat molemmat ei-negatiivisia. Valitaan α ja β miten tahansa, niin keskineliövirhe on aina suurempi tai yhtä kuin

$$\operatorname{var} Y - \frac{(\operatorname{cov}(X, Y))^2}{\operatorname{var} X} = (1 - \rho^2) \operatorname{var} Y, \tag{7.3}$$

ja keskineliövirheen minimi saavutetaan valitsemalla

$$\alpha = \mu_Y, \qquad \beta = \frac{\operatorname{cov}(X, Y)}{\operatorname{var} X}.$$

Ts. keskineliövirheen mielessä paras Y:n lineaarinen ennuste on

$$EY + \frac{\operatorname{cov}(X, Y)}{\operatorname{var}(X)}\,(X - EX) \tag{7.4}$$

Edellä ρ on korrelaatiokerroin

$$\rho = \operatorname{corr}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{var} X}\,\sqrt{\operatorname{var} Y}},$$

jonka arvot ovat (Cauchyn–Schwarzin epäyhtälön mukaan) välillä $-1 \le \rho \le 1$. Koska ennusteen odotusarvo on EY, on keskineliövirhe sama kuin ennustusvirheen

$$Y - EY - \frac{\operatorname{cov}(X, Y)}{\operatorname{var}(X)}\,(X - EX)$$

varianssi.

Ratkaisemme myöhemmin vielä kunnianhimoisemman tehtävän, jossa kysytään, mikä on keskineliövirheen mielessä paras Y:n ennuste lausekkeella $m(X)$, jossa funktio m saadaan valita vapaasti.

Parhaan lineaarisen ennusteen keskineliövirheen kaavasta (7.3) nähdään, kuinka *korrelaatiokerroin mittaa muuttujien välisen lineaarisen riippuvuuden voimakkuutta.* Jos korrelaatiokerroin saavuttaa jommankumman ääriarvonsa $\rho = \pm 1$, niin tällöin keskineliövirhe on nolla, joten tällöin ennustusvirhe on nolla melkein varmasti (ks. lause 4.3), eli

$$Y = EY + \frac{\operatorname{cov}(X, Y)}{\operatorname{var} X}\,(X - EX) \quad (\text{m.v.}), \tag{7.5}$$

Edellä lyhenne m.v. on lyhenne sanoille melkein varmasti, joka taas tarkoittaa samaa kuin todennäköisyydellä yksi. Siis ominaisuudesta $|\rho| = 1$ seuraa, että

$$Y(\omega) = EY + \frac{\operatorname{cov}(X,Y)}{\operatorname{var} X}(X(\omega) - EX)$$

kaikilla alkeistapauksilla ω, paitsi sellaisilla, jotka kuuluvat poikkeusjoukkoon, jonka tn on nolla.

Huomaa myös, että korrelaatikertoimen etumerkki on sama kuin kovarianssin $\operatorname{cov}(X,Y)$ etumerkki ja se on edelleen sama kuin parhaan lineaarisen ennusteen kulmakertoimen etumerkki. Jos $\rho > 0$ eli $\operatorname{cov}(X,Y) > 0$, niin Y:n arvolla on taipumus kasvaa X:n arvon kasvaessa, koska parhaan lineaarisen ennusteen kulmakerroin on tällöin positiivinen. Jos taas $\rho < 0$ eli $\operatorname{cov}(X,Y) < 0$, niin Y:n arvoilla on taipumus vähetä X:n arvon kasvaessa.

7.7 Odotusarvovektori ja kovarianssimatriisi

Silloin kuin satunnaisvektori $\mathbf{V} = (X,Y)$ esiintyy vektori- ja matriisilausekkeissa, sitä pidetään pystyvektorina. Varoitus: joissakin muissa lähteissä paria (X,Y) pidetään vaakavektorina. Näissä luentomuistiinpanoissa vektorit ja matriisit ovat paksunnettuja ja skalaarit paksuntamattomia. Edistyneemmissä esityksissä tällaista konventioita ei välttämättä käytetä, vaan lukijan pitää päätellä asiayhteydestä, minkätyyppistä oliota kukin symboli tarkoittaa.

Määritelmä 7.4 (Satunnaisvektorin odotusarvo)**.** Satunnaisvektorin (X,Y) odotusarvo(vektori) määritellään ottamalla satunnaisvektorista odotusarvot komponentti komponentilta, eli

$$E\mathbf{V} = E(X,Y) = E\begin{bmatrix} X \\ Y \end{bmatrix} = (EX, EY) = \begin{bmatrix} EX \\ EY \end{bmatrix}$$

mikäli kaikkien komponenttien odotusarvot ovat olemassa.

Koska kaksikomponenttista vektoria (X,Y) voidaan pitää myös pitää 2×1-matriisina, niin satunnaisvektorin odotusarvon määritelmä on erikoistapaus seuraavaksi esitettävästä satunnaismatriisin odotusarvon määritelmästä.

Määritelmä 7.5 (Satunnaismatriisi ja sen odotusarvo)**.** Satunnaismatriisi on matriisi, jonka alkiot ovat satunnaismuuttujia. Sen odotusarvo on se vakiomatriisi, jonka alkio kohdassa (i,j) on ko. satunnaismatriisin alkion (i,j) odotusarvo (mikäli kaikkien alkoiden odotusarvot ovat olemassa).

Määritelmästä seuraa välittömästi, että

$$E(\mathbf{Z}^T) = (E\mathbf{Z})^T \tag{7.6}$$

mikäli $\mathbf{Z}$ on satunnaismatriisi.

Odotusarvon lineaarisuuden avulla nähdään helposti, että

$$E(\mathbf{AZB} + \mathbf{C}) = \mathbf{A}(E\mathbf{Z})\mathbf{B} + \mathbf{C}, \tag{7.7}$$

kun $\mathbf{Z}$ on satunnaismatriisi ja $\mathbf{A}$, $\mathbf{B}$ ja $\mathbf{C}$ ovat vakiomatriiseja, joiden dimensiot ovat sellaiset, että lauseke on määritelty. Vakiomatriisit saadaan vetää ulos odotusarvon alta, jos ne sijaitsevat matriisitulossa äärimmäisenä vasemmalla tai äärimmäisenä oikealla. Sen sijaan lausekkeen keskellä olevia vakiomatriiseja ei pystytä vetämään ulos tällä kaavalla. Kaava (7.7) pätee myös satunnaisvektorille $\mathbf{Z}$, sillä d-komponenttinen pystyvektori voidaan tulkita $d \times 1$-matriisiksi.

Määritelmä 7.6 (Kovarianssimatriisi)**.** Satunnaisvektorin $\mathbf{V} = (X, Y)$ kovarianssimatriisi määritellään kaavalla

$$\operatorname{Cov}(\mathbf{V}) = E[(\mathbf{V} - E\mathbf{V})(\mathbf{V} - E\mathbf{V})^T]. \tag{7.8}$$

Huomautuksia: Käytän satunnaisvektorin kovarianssimatriisille merkintää $\operatorname{Cov}(\mathbf{V})$ ja satunnaismuuttujien X ja Y kovarianssille merkintää $\operatorname{cov}(X, Y)$. Näemme pian, että nämä käsitteet ovat läheistä sukua toisilleen. Sv:n $\mathbf{V}$ kovarianssimatriisille käytetään kirjallisuudessa monia muitakin merkintöjä, kuten esim. $\operatorname{var}(\mathbf{V})$. Englannin kielellä kovarianssimatriisista käytetään ainakin nimityksiä *covariance matrix*, *variance–covariance matrix*, *variance matrix*, *dispersion matrix*.

Määritelmän perusteella

$$\begin{aligned}
\operatorname{Cov}(\mathbf{V}) &= E\left\{\begin{bmatrix} X - EX \\ Y - EY \end{bmatrix} \begin{bmatrix} X - EX & Y - EY \end{bmatrix}\right\} \\
&= E\begin{bmatrix} (X - EX)(X - EX) & (X - EX)(Y - EY) \\ (Y - EY)(X - EX) & (Y - EY)(Y - EY) \end{bmatrix} \\
&= \begin{bmatrix} \operatorname{cov}(X, X) & \operatorname{cov}(X, Y) \\ \operatorname{cov}(Y, X) & \operatorname{cov}(Y, Y) \end{bmatrix} = \begin{bmatrix} \operatorname{var}(X) & \operatorname{cov}(X, Y) \\ \operatorname{cov}(X, Y) & \operatorname{var}(Y) \end{bmatrix}
\end{aligned}$$

Kovarianssimatriisin päälävistäjällä on muuttujien varianssit ja muualla muuttujien väliset kovarianssit. Tästä syystä kovarianssimatriisia kutsutaan myös kömpelöllä nimellä varianssi-kovarianssimatriisi.

Jos merkitään $\sigma_X^2 = \operatorname{var}(X)$ ja $\sigma_Y^2 = \operatorname{var}(Y)$ ja $\sigma_X, \sigma_Y > 0$, niin satunnaismuuttujien X ja Y välinen korrelaatio(kerroin) $\rho = \operatorname{corr}(X, Y)$ määriteltiin kaavalla

$$\rho = \operatorname{corr}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{var} X}\sqrt{\operatorname{var} Y}} = \frac{\operatorname{cov}(X, Y)}{\sigma_X\, \sigma_Y}$$

Tämän ansiosta sv:n $\mathbf{V} = (X, Y)$ kovarianssimatriisi voidaan kirjoittaa myös muodossa

$$\operatorname{Cov}(\mathbf{V}) = \begin{bmatrix} \sigma_X^2 & \rho\, \sigma_X \sigma_Y \\ \rho\, \sigma_X \sigma_Y & \sigma_Y^2 \end{bmatrix}$$

Jos satunnaisvektori $\mathbf{W}$ saadaan satunnaisvektorista $\mathbf{V}$ affiinilla muunnoksella,

$$\mathbf{W} = \mathbf{A}\mathbf{V} + \mathbf{b},$$

jossa $\mathbf{A}$ on vakiomatriisi ja $\mathbf{b}$ on vakiovektori, niin

$$E\mathbf{W} = \mathbf{A}(E\mathbf{V}) + \mathbf{b} \quad \Rightarrow \quad (\mathbf{W} - E\mathbf{W})(\mathbf{W} - E\mathbf{W})^T = \mathbf{A}(\mathbf{V} - E\mathbf{V})(\mathbf{V} - E\mathbf{V})^T \mathbf{A}^T,$$

joten kaavan (7.7) nojalla

$$\operatorname{Cov}(\mathbf{A}\mathbf{V} + \mathbf{b}) = \mathbf{A} \operatorname{Cov}(\mathbf{V}) \mathbf{A}^T. \tag{7.9}$$

Sovelletaan lopuksi kaavaa (7.9) tapaukseen $\mathbf{V} = (X, Y)$ ja $\mathbf{A} = \mathbf{u}^T$, jossa $\mathbf{u} = (a, b)$. Tällöin saadaan johdettua tuttu kaava, nimittäin

$$\begin{aligned}
\operatorname{var}(aX + bY) = \operatorname{var}(\mathbf{u}^T \mathbf{V}) = \mathbf{u}^T \operatorname{Cov}(\mathbf{V})\, \mathbf{u} &= \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \operatorname{var}(X) & \operatorname{cov}(X, Y) \\ \operatorname{cov}(X, Y) & \operatorname{var}(Y) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \\
&= a^2 \operatorname{var}(X) + 2ab \operatorname{cov}(X, Y) + b^2 \operatorname{var}(Y).
\end{aligned}$$

7.8 Tiheysfunktion muuntokaava

Jos sv:illa (X, Y) on diskreetti jakauma, ja satunnaismuuttuja tai -vektori Z määritellään sen muunnoksena, $Z = g(X, Y)$, niin tällöin Z:lla on diskreetti jakauma, jonka ptnf voidaan laskea suoraan määritelmän perusteella, sillä

$$f_Z(z) = P(Z = z) = \sum_{(x,y):g(x,y)=z} f_{X,Y}(x, y).$$

Jatkuvan jakauman tapauksessa tarvitaan monimutkaisempia laskuja.

Tarkastelemme jatkuvasti jakautunutta sv:a (X, Y), jolla on tf $f_{X,Y}$, sekä funktiota $\mathbf{g} : A \to B$, jossa $A, B \subset \mathbb{R}^2$, ja A on sellainen avoin joukko, että

$$P((X, Y) \in A) = 1. \tag{7.10}$$

Määrittelemme kaksiulotteisen satunnaisvektorin (U, V) kaavalla

$$\begin{bmatrix} U \\ V \end{bmatrix} = \mathbf{g}(X, Y) = \begin{bmatrix} g_1(X, Y) \\ g_2(X, Y) \end{bmatrix}$$

Oletamme, että

$$\mathbf{g} : A \to B \quad \text{on diffeomorfismi} \tag{7.11}$$

eli että

- joukot A ja B ovat avoimia,
- funktio $\mathbf{g} : A \to B$ on jatkuvasti derivoituva bijektio,
- sen käänteisfunktio $\mathbf{h} = \mathbf{g}^{-1}$ on jatkuvasti derivoituva bijektio $B \to A$.

Käytännössä diffeomorfisuuden tarkistaminen sujuu tavallisesti seuraavasti. Ensin käänteisfunktiolle $\mathbf{h} = (h_1, h_2)$ johdetaan kaava ratkaisemalla yhtälöistä

$$g_1(x, y) = u, \quad g_2(x, y) = v, \qquad (u, v) \in B$$

muuttujat x ja y muuttujien u ja v funktiona. Mikäli ratkaisu

$$(x, y) = (h_1(u, v), h_2(u, v))$$

on yksikäsitteinen kaikilla $(u, v) \in B$ ja ratkaisu $(x, y) \in A$, niin tällöin ollaan tarkistettu, että kyseessä on bijektiivinen vastaavuus. Tämän jälkeen on tavallisesti suoraviivaista tarkistaa, ovatko lausekkeet $g_1(x, y)$ ja $g_2(x, y)$ jatkuvasti derivoituvia A:lla ja lausekkeet $h_1(u, v)$ ja $h_2(u, v)$ jatkuvasti derivoituvia B:llä. Esimerkiksi lausekkeen $g_1(x, y)$ kohdalla pitää miettiä, ovatko molemmat ensimmäisen kertaluvun osittaisderivaatat $\partial g_1(x, y)/\partial x$ ja $\partial g_2(x, y)/\partial y$ olemassa ja jatkuvia funktioita joukossa A.

Nyt ajattelemme bijektiivistä vastaavuutta

$$(u, v) = (g_1(x, y), g_2(x, y)) \quad \iff \quad (x, y) = (h_1(u, v), h_2(u, v)), \tag{7.12}$$

jota on mukavaa merkitä myös

$$(u, v) = (u(x, y), v(x, y)) \quad \iff \quad (x, y) = (x(u, v), y(u, v)). \tag{7.13}$$

Kuvauksen $\mathbf{h}$ derivaatta(matriisi) eli Jacobin matriisi pisteessä (u, v) on

$$\begin{bmatrix} \dfrac{\partial h_1(u,v)}{\partial u} & \dfrac{\partial h_1(u,v)}{\partial v} \\ \dfrac{\partial h_2(u,v)}{\partial u} & \dfrac{\partial h_2(u,v)}{\partial v} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{bmatrix} \tag{7.14}$$

Jacobin matriisin eli derivaattamatriisin determinanttia kutsutaan kuvauksen $\mathbf{h}$ *funktionaalideterminantiksi* eli *Jacobin determinantiksi* eli *jacobiaaniksi* (engl. *Jacobian determinant; Jacobian*). (Myös nimiä Jakobin determinantti tai jakobiaani käytetään; käsitteen esitti 1800-luvulla saksalainen matemaatikko Carl Gustav Jacobi.) Varoitus: joissakin muissa lähteissä termi jacobiaani saattaa tarkoittaa derivaattamatriisia eikä sen determinanttia.

Jacobiaanille eli Jacobin matriisin determinantille käytetään kirjallisuudessa useita erilaisia merkintöjä. Me käytämme sekä modernia eksplisiittistä merkintää

$$J_{\mathbf{h}}(u, v),$$

että 1800-luvun puolivälistä lähtien käytössä ollutta klassista merkintää

$$\frac{\partial(x, y)}{\partial(u, v)}.$$

Klassinen merkintä on modernia merkintää paljon kätevämpi silloin, kun funktiot on määritelty konkreettisten lausekkeiden avulla. Tässä

$$J_{\mathbf{h}}(u,v) = \frac{\partial(x,y)}{\partial(u,v)} = \det \begin{bmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{bmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}.$$

Funktion $\mathbf{g}$ ja sen käänteisfunktion $\mathbf{h}$ derivaattamatriisit ovat tunnetusti toistensa käänteismatriiseja, josta seuraa jacobiaaneille yhteys

$$1 = J_{\mathbf{h}}(u,v)\, J_{\mathbf{g}}(x,y) = \frac{\partial(x,y)}{\partial(u,v)} \frac{\partial(u,v)}{\partial(x,y)}. \tag{7.15}$$

Tässä kaavassa (x, y) ja (u, v) vastaavat toisiaan bijektiivisesti yhtälön (7.13) mukaan. Kaava (7.15) seuraa siitä, että toisaalta identiteettimatriisin determinantti on yksi ja toisaalta kaavasta $\det(\mathbf{A}\,\mathbf{B}) = \det(\mathbf{A})\,\det(\mathbf{B})$, joka pätee, kun $\mathbf{A}$ ja $\mathbf{B}$ ovat samankokoisia neliömatriiseja.

Lause 7.9. *Oletuksilla* (7.10), (7.11) *ja tämän jakson merkinnöillä satunnaisvektorilla* (U, V) *on jatkuva jakauma tf:llä*

$$f_{U,V}(u,v) = \begin{cases} f_{X,Y}(h_1(u,v), h_2(u,v))\, |J_{\mathbf{h}}(u,v)|, & \text{kun } (u,v) \in B \\ 0, & \text{muuten.} \end{cases} \tag{7.16}$$

Todistus. Oletusten nojalla $(U, V) \in B$ tn:llä yksi, sillä

$$P((U,V) \in B) = P(\mathbf{g}(X,Y) \in B) = P((X,Y) \in \mathbf{g}^{-1}(B)) = P((X,Y) \in A) = 1.$$

Olkoon $C \subset B$ mielivaltainen joukko. Nyt

$$\begin{aligned} P((U,V) \in C) &= P(\mathbf{g}(X,Y) \in C) = P((X,Y) \in \mathbf{h}(C)) \\ &= \iint_{\mathbf{h}(C)} f_{X,Y}(x,y)\, \mathrm{d}x\, \mathrm{d}y. \end{aligned}$$

Tehdään nyt muuttujanvaihto $(u,v) = \mathbf{g}(x,y)$, eli $(x,y) = \mathbf{h}(u,v)$, jolloin tasointegraali saadaan muotoon (tähän tarkoituksen riittävän voimakas Lebesguen integraalia koskeva muuttujanvaihtokaava löytyy esim. Billingsleyn teoksen [1] lauseesta 17.2)

$$\iint_{\mathbf{h}(C)} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \iint_C f_{X,Y}(h_1(u,v), h_2(u,v))\,|J_{\mathbf{h}}(u,v)|\,\mathrm{d}u\,\mathrm{d}v$$

Nämä tarkastelut todistavat väitteen. □

Diffeomorfismin tapauksessa muuttujanvaihtokaava (7.16) kannattaa pitää mielessään muodossa

$$f_{X,Y}(x,y)\,|\partial(x,y)| = f_{U,V}(u,v)\,|\partial(u,v)|, \quad \text{kun} \tag{7.17}$$

$$(u,v) = \mathbf{g}(x,y) \quad \Leftrightarrow \quad (x,y) = \mathbf{h}(u,v), \qquad (x,y) \in A, (u,v) \in B. \tag{7.18}$$

Tästä sitten tuntemattomat suureet ratkaistaan tunnettujen suureiden avulla. Jos $f_{X,Y}$ on annettu, niin

$$f_{U,V}(u,v) = f_{X,Y}(x,y)\,\left|\frac{\partial(x,y)}{\partial(u,v)}\right| = f_{X,Y}(h_1(u,v), h_2(u,v))\,|J_{\mathbf{h}}(u,v)|,$$

kun $(u,v) \in B$. Saman ytf:n voi ilmaista myös muodossa

$$f_{U,V}(u,v) = \frac{f_{X,Y}(x,y)}{\left|\dfrac{\partial(u,v)}{\partial(x,y)}\right|} = \frac{f_{X,Y}(h_1(u,v), h_2(u,v))}{|J_{\mathbf{g}}(h_1(u,v), h_2(u,v))|},$$

kun $(u,v) \in B$. Tämän tuloksen oikeellisuus perustuu kaavaan (7.15).

Tiheysfunktion muuntokaavaa (7.16) sovellettaessa on tärkeää pitää kirjaa siitä, missä joukossa B johdettu kaava on pätevä. Tämän kirjanpidon saa usein hoidettua automaattisesti joukkojen indikaattorien avulla, kuten seuraavassa esimerkissä tehdään. Valitettavasti jatkotarkasteluja varten (esim. reunajakauman johtamista varten) on usein tarpeen esittää alue B esim. muodossa

$$B = \{(u,v) : a < u < b, c(u) < v < d(u)\}$$

tai muodossa

$$B = \{(u,v) : c < v < d, a(v) < u < b(v)\}.$$

Tästä voi aiheutua paljon lisätyötä.

Esimerkki 7.7. Olkoon sv:lla (X,Y) jatkuva yhteisjakauma tf:lla

$$f_{X,Y}(x,y) = \begin{cases} k(x,y) & \text{kun } x > 0 \text{ ja } y > 0 \\ 0 & \text{muuten.} \end{cases} \tag{7.19}$$

Esitetään tämä ytf tason ensimmäisen neljänneksen indikaattorin $1\{x > 0, y > 0\}$ ja lausekkeen $k(x,y)$ tulona,

$$f_{X,Y}(x,y) = 1\{x > 0, y > 0\}\,k(x,y).$$

Tässä lauseke $k(x,y)$ ei ole välttämättä hyvin määritelty, jos $x \le 0$ tai $y \le 0$. Ylläoleva esitys pitää ymmärtää lyhennemerkintänä kaavalle (7.19). Käytämme tällaista konventiota, jossa joukon indikaattori tarvittaessa nollaa mahdollisesti huonosti määritellyn lausekkeen.

Tarkastellaan sv:a (U, V), jossa $U = X + Y$ ja $V = X - Y$. Tätä muunnosta vastaava diffeomorfismi on

$$\begin{cases} u = x + y \\ v = x - y \end{cases} \iff \begin{cases} x = \frac{1}{2}(u + v) \\ y = \frac{1}{2}(u - v) \end{cases}$$

joka on diffeomorfismi koko tason ja koko tason välillä. Tämän takia satunnaismuuttujien U ja V ytf on

$$\begin{aligned} f_{U,V}(u, v) &= f_{X,Y}(x, y) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| \\ &= 1\{u + v > 0, u - v > 0\}\, k(\frac{1}{2}(u + v), \frac{1}{2}(u - v))\, \frac{1}{2}. \end{aligned}$$

Lasketaan seuraavaksi sm:n U reunatiheys. Oikeiden integrointirajojen selvittämiseksi edellä esiintyvä epäyhtälöpari $u + v > 0$ ja $u - v > 0$ pitää onnistua ratkaisemaan muodossa

$$a < u < b, \quad c(u) < v < d(u).$$

Hetken pohtimisen jälkeen (piirrä kuva) nähdään, että ratkaisu on

$$u > 0, \quad -u < v < u.$$

Tämän takia U:n reunatiheys on

$$f_U(u) = \int_{-u}^{u} \frac{1}{2}\, k(\frac{1}{2}(u + v), \frac{1}{2}(u - v))\, \mathrm{d}v, \qquad u > 0.$$

△

Huomaa, että diffeomorfismi on aina kuvaus kahden samandimensioisen euklidisen avaruuden välillä. Toisinaan tavoitteena on johtaa jatkuvasti jakautuneen sv:n (X, Y) jonkin skalaariarvoisen muunnoksen $U = g_1(X, Y)$ tiheysfunktio. Tämä voidaan tehdä kahdella erilaisella tekniikalla.

1. Lasketaan ensin U:n kertymäfunktio

$$F_U(u) = P(g_1(X, Y) \le u) = \int_{\{(x,y): g_1(x,y) \le u\}} f_{X,Y}(x, y)\, \mathrm{d}x\, \mathrm{d}y$$

   integroimalla kyseisen joukon yli, ja sitten U:n tiheysfunktio lasketaan derivoimalla (mikäli U:n jakauma osoittautuu jatkuvaksi).

2. Voidaan yrittää täydentää muunnos (keinotekoisesti) bijektioksi valitsemalla $V = g_2(X, Y)$, sitten johdetaan sv:n (U, V) tiheysfunktio, ja lopuksi kiinnostuksen kohteena olevan muuttujan U tiheysfunktio lasketaan integroimalla

Näistä tavoista jälkimmäinen on usein suoraviivaisempi.

Esimerkki 7.8. Summan tiheysfunktio. Olkoon sv:lla (X, Y) tf $f_{X,Y}$. Olkoon

$$U = X + Y.$$

Johda sm:n U tf. Minkälaisen kaavan saat, jos $X \perp\!\!\!\perp Y$?

Ratkaisu bijektioksi täydentämällä: Tässä tehtävässä voisimme täydentää kuvauksen bijektioksi monella tavalla, mutta nyt teemme valinnan $V = X$. Tällöin

$$\begin{cases} u = x + y \\ v = x \end{cases} \iff \begin{cases} x = v \\ y = u - v \end{cases}$$

Siis

$$f_{U,V}(u,v) = f_{X,Y}(x,y) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| = f_{X,Y}(v, u-v).$$

Sm:n U reunatiheysfunktio saadaan tästä integroimalla v pois,

$$f_U(u) = \int_{-\infty}^{\infty} f_{X,Y}(v, u-v)\,\mathrm{d}v$$

Ratkaisu kertymäfunktiotekniikalla: Kun lasketaan U:n kertymäfunktiota pisteessä u, niin integrointijoukko on

$$\{(x,y) : x + y \le u\} = \{(x,y) : x \in \mathbb{R}, y \le u - x\}.$$

Tämän takia

$$F_U(u) = \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{u-x} f_{X,Y}(x,y)\,\mathrm{d}y.$$

Mikäli derivaatan vienti integraalin alle pystytään perustelemaan jotenkin, niin

$$\begin{aligned} f_U(u) &= F_U'(u) \\ &= \int_{-\infty}^{\infty} \mathrm{d}x \frac{\partial}{\partial u} \int_{-\infty}^{u-x} f_{X,Y}(x,y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x, u-x)\,\mathrm{d}x. \end{aligned}$$

Jälkimmäisessä laskussa jää arveluttamaan se, mitä ehtoja ytf:lle pitäisi asettaa, jotta tulos voitaisiin perustella täsmällisesti. Ensimmäisessä tavassa nähtiin, että mitään ehtoja ei tarvita.

Jos $X \perp\!\!\!\perp Y$, niin $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$ identtisesti, ja

$$f_U(u) = \int_{-\infty}^{\infty} f_X(v)\, f_Y(u-v)\,\mathrm{d}v,$$

eli f_U on tiheysfunktioiden f_X ja f_Y konvoluutio,

$$f_U = f_X * f_Y.$$

△

Kirjallisuutta

[1] P. Billingsley. *Probability and Measure.* John Wiley & Sons, Inc., 2nd edition, 1986.

7.9 t-jakauman ominaisuuksia

Studentin t-jakauma vapausasteluvulla $\nu > 0$ eli t_ν-jakauma määriteltiin niin, että se on sm:n

$$Y = \frac{Z}{\sqrt{X/\nu}} \tag{7.20}$$

jakauma, kun $X \sim \chi^2_\nu = \mathrm{Gam}(\frac{\nu}{2}, \frac{1}{2})$, $Z \sim N(0,1)$, ja $X \perp\!\!\!\perp Z$. Johdamme seuraavaksi t-jakauman tiheysfunktion esimerkkinä muuttujanvaihtotekniikan käytöstä. Lisäksi päättelemme jakauman odotusarvon ja varianssin käyttämällä tätä jakauman stokastista esitystä.

Riippumattomuuden nojalla

$$f_{X,Z}(x,z) = f_X(x)\, f_Z(z) = \frac{(\frac{1}{2})^{\nu/2}}{\Gamma(\frac{\nu}{2})}\, x^{\frac{\nu}{2}-1}\, \mathrm{e}^{-\frac{1}{2}x}\, \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-\frac{1}{2}z^2}, \qquad x > 0.$$

Täydennetään kuvaus bijektioksi valitsemalla $U = X$. Tällöin tarkasteltavana on diffeomorfismi

$$\begin{cases} u = x \\ y = \dfrac{z}{\sqrt{x/\nu}} \end{cases} \qquad \Longleftrightarrow \qquad \begin{cases} x = u \\ z = y\sqrt{u/\nu} \end{cases}$$

jossa $x > 0$ ja $u > 0$. Siis

$$f_{U,Y}(u,y) = f_{X,Z}(x,z) \left| \frac{\partial(x,z)}{\partial(u,y)} \right| = f_{X,Z}(u, y\sqrt{\frac{u}{\nu}})\, \sqrt{\frac{u}{\nu}}, \qquad u > 0.$$

Tästä saadaan Y:n reunatiheys integroimalla u pois,

$$f_Y(y) = \int_0^\infty f_{X,Z}(u, y\sqrt{u/\nu})\, \sqrt{\frac{u}{\nu}}\, \mathrm{d}u = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\, \sqrt{\nu\pi}}\, \frac{1}{(1 + y^2/\nu)^{(\nu+1)/2}}.$$

Edellä oleva määrätty integraali pystytään laskemaan integroimalla kuten tilastotieteilijä: koska tiedämme, mikä on gammajakauman tiheysfunktion lauseke, osaamme suoraan kirjoittaa tämänmuotoisen integraalin arvon.

Arvolla $\nu = 1$ tf saa muodon

$$f_Y(y) = \frac{1}{\pi} \frac{1}{1 + y^2},$$

joten t_1 on Cauchyn jakauma.

Studentin t-jakaumalla ei ole olemassa kaikkia absoluuttisia momentteja. Jos $a > 0$, niin

$$E|Y|^a < \infty \quad \Rightarrow \quad a < \nu.$$

Tämän seikan näkee esim. jakauman stokastiseta esityksestä (7.20), josta seuraa että

$$E|Y|^a = \nu^{a/2}\, E|Z|^a\, EX^{-a/2},$$

joka on äärellinen mikäli molemmat kaavan oikealla puolella olevat odotusarvot ovat äärellisiä. Normaalijakauman absoluuttiset momentit $E|Z|^a$ ovat äärellisiä, mutta tarkastelemalla jälkimmäistä odotusarvoa päädytään helposti ehtoon $a < \nu$. Esim. Cauchyn jakaumalla ei ole odotusarvoa; t-jakaumalla on olemassa varianssi vain silloin, kun $\nu > 2$. Siitä, että jakauman kaikki momentit eivät ole äärellisiä seuraa, että t-jakauman momenttiemäfunktio ei ole olemassa missään origon ympäristössä, joten momenttiemäfunktion on hyödytön työkalu tälle jakaumalle.

Studentin t-jakauman odotusarvo ja varianssi selviävät helposti jakauman stokastisesta esityksestä (7.20). Jos $\nu > 1$, niin t_ν-jakaumalla on odotusarvo, ja se on nolla. Tämä nähdään siitä, että koska $Z \perp\!\!\!\perp X$, niin

$$EY = EZ\, E\sqrt{\frac{\nu}{X}} = 0 \cdot E\sqrt{\frac{\nu}{X}} = 0.$$

Jos $\nu > 2$, niin t_ν-jakauman varianssi on

$$\operatorname{var} Y = EY^2 = EZ^2\, E\frac{\nu}{X} = 1 \cdot E\frac{\nu}{X}.$$

Tästä nähdään muutaman välivaiheen jälkeen (integroi kuten tilastotieteilijä ja käytä gammafunktion funktionaaliyhtälöä), että

$$\operatorname{var} Y = \frac{\nu}{\nu - 2}, \qquad \text{kun } \nu > 2.$$

Luku 8

Ehdollinen jakauma

8.1 Ehdolliset jakaumat

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

$$f_{X|Y}(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \tag{8.1}$$

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.

Kaavaa (8.1) pidetään mallina, kun määritellään ehdollinen tiheysfunktio jatkuvan yhteisjakauman tapauksessa.

Määritelmä 8.1. Olkoon (X, Y):llä jatkuva jakauma tiheysfunktiolla $f_{X,Y}$. Tällöin sm:n X ehdollinen tiheysfunktio ehdolla $Y = y$ on

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

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 $f_X(x) > 0$.

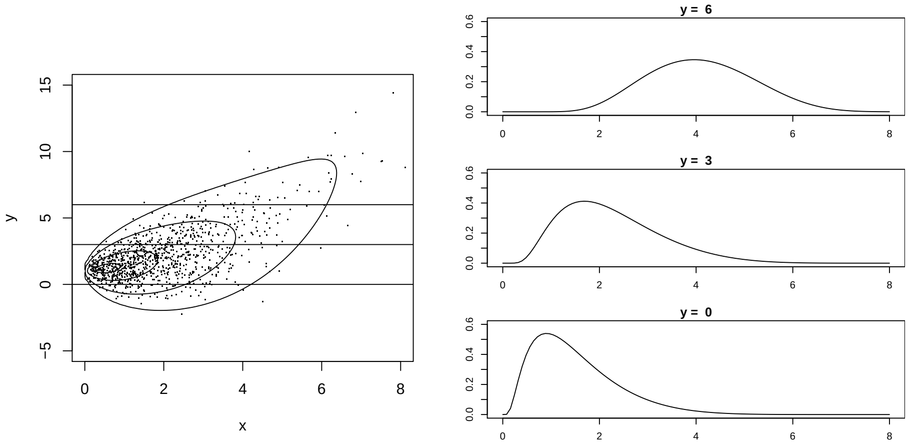
Huomaa, että $x \mapsto f_{X|Y}(x \mid y)$ on tiheysfunktio, sillä se on ei-negatiivinen ja

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

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 8.1 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

Kuva 8.1 Yhteistiheysfunktio ja ehdollisia tiheysfunktioita $x \mapsto f_{X|Y}(x \mid y)$.



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.

Teemme uskottavaksi, että jos ytf $f_{X,Y}$ on riittävän sileä (jätämme auki, mitä ehtoja täsmällisesti ottaen vaadittaisiin) ja $f_Y(y) > 0$, niin tällöin

$$P(X \in A \mid y \le Y \le y+h) \xrightarrow[h \to 0+]{} \int_A f_{X|Y}(x \mid y)\,\mathrm{d}x, \qquad \text{kaikilla } A \subset \mathbb{R}. \tag{8.2}$$

Toisin sanoen pienillä $h > 0$ ehdollinen todennäköisyys $P(X \in A \mid y \le Y \le y+h)$ saadaan osapuilleen selville integroimalla ehdollista tiheysfunktiota $f_{X|Y}(x \mid y)$ muuttujan x suhteen joukon A yli mielivaltaiselle $A \subset \mathbb{R}$. 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. Lisäksi tämän tuloksen perusteella olisi mahdollista antaa frekvenssitulkinta myös ehdolliselle tiheysfunktiolle.

Tulos (8.2) perustellaan lähtien liikkeelle esityksestä

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

Sekä osoittajassa että nimittäjässä esiintyy muotoa

$$\frac{1}{h} \int_y^{y+h} g(u)\,\mathrm{d}u$$

olevia integraalikeskiarvoja, jotka lähestyvät arvoa $g(y)$ aika yleisillä oletuksilla, kun $h \to 0+$. Esimerkiksi, jos funktio g on jatkuva pisteessä y ja integroituva jossakin sen ympäristössä, niin mielivaltaiselle $\epsilon > 0$ voidaan jatkuvuuden nojalla valita luku $\delta > 0$ siten, että $|g(u) - g(y)| \le \epsilon$ kaikilla u, joille $|u - y| < \delta$. Jos h on mikä tahansa luku siten, että $0 < h < \delta$, niin on voimassa

$$g(y) - \epsilon = \frac{1}{h}\int_y^{y+h} (g(y) - \epsilon)\, du \le \frac{1}{h}\int_y^{y+h} g(u)\, du \le \frac{1}{h}\int_y^{y+h} (g(y) + \epsilon)\, du = g(y) + \epsilon.$$

Tämä todistaa, että integraalikeskiarvo lähestyy raja-arvoa $g(y)$, kun $h \to 0+$. Kun integraalikeskiarvon raja-arvoa sovelletaan tarkasteltavaan ehdolliseen todennäköisyyteen, ja osoittajassa vaihdetaan huolettomasti rajankäynnin ja integroinnin järjestys, saadaan

$$P(X \in A \mid y \le Y \le y + h) \xrightarrow[h \to 0+]{} \frac{\int_A f_{X,Y}(x, y)\, dx}{f_y(y)} = \int_A \frac{f_{X,Y}(x, y)}{f_Y(y)}\, dx.$$

8.2 Kertolaskusääntö eli ketjusääntö

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

$$f_{X,Y}(x, y) = f_X(x)\, f_{Y|X}(y \mid x) = f_Y(y)\, f_{X|Y}(x \mid y), \qquad \text{kaikilla } x, y. \tag{8.3}$$

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 $f_{Y|X}(y \mid x)$ johdetaan lähtemällä liikkeelle joistakin ytf:n (mahdollisesti toisistaan eriävistä) versioista, niin niiden tulo kelpaa yhteistiheysfunktioksi.

Lisäksi kaavassa on se ongelma, että esim. $f_{Y|X}(y \mid x)$ 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

$$f_{Y|X}(y \mid x) = \begin{cases} \dfrac{f_{X,Y}(x, y)}{f_X(x)}, & \text{kun } f_X(x) > 0, \\ 0, & \text{muuten.} \end{cases} \tag{8.4}$$

(Yhtä hyvin voitaisiin käyttää jotakin muuta johdonmukaista sopimusta.) Toiselle ehdolliselle tf:lle $f_{X|Y}(x \mid y)$ 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

$$f_{Y|X}(x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{f_Y(y)\, f_{X|Y}(x \mid y)}{f_X(x)}. \tag{8.5}$$

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 8.1. Esimerkissä 7.2 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 7.2. Tuloksesta tunnistetaan, että ehdollinen jakauma on tasajakauma

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

△

8.3 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).

Tämän seurauksena 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 8.1. *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

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

8.4 Ehdollinen odotusarvo

Määritelmä 8.2 (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 $X = x$,

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

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

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

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

$$\mathrm{var}(g(X,Y) \mid X = x) = E[(g(X,Y))^2 \mid X = x] - \{E[g(X,Y) \mid X = x]\}^2. \tag{8.6}$$

Tämä on tutun kaavan var $Z = EZ^2 - (EZ)^2$ vastine ehdolliselle varianssille. Jos Y on diskreetti sm, ylläolevissa kaavoissa tarvitaan summia integraalien sijasta, ja identiteetti (8.6) pysyy voimassa.

Jos funktio g on muotoa
$$g(x, y) = g_1(x)\, g_2(x, y),$$
niin tietenkin
$$E(g_1(X)\, g_2(X, Y) \mid X = x) = g_1(x)\, E(g_2(X, Y) \mid X = x). \tag{8.7}$$
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 8.3 (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).

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

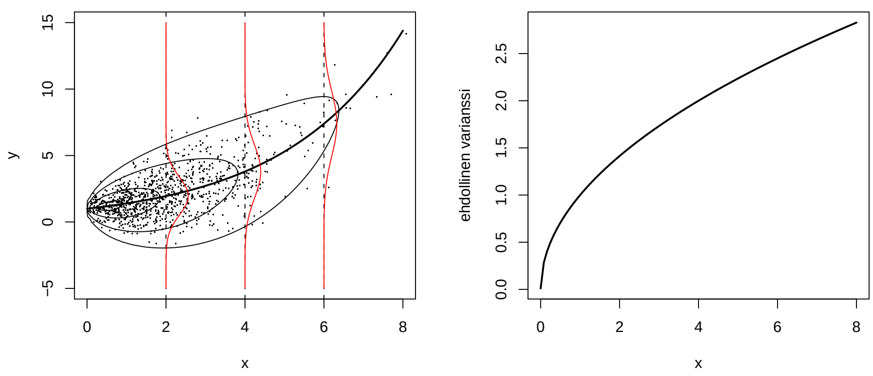
Kuvassa 8.2 on piirretty ehdollinen odotusarvo $E(Y \mid X = x)$ eli regressiofunktio sekä ehdollinen varianssi $\mathrm{var}(Y \mid X = x)$ kuvan 8.1 yhteisjakaumalle.

Määritelmä 8.4 (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 8.2. *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.

Kuva 8.2 (a) Yhteisjakauma, ehdollisia tiheysfunktioita $y \mapsto f_{Y|X}(y \mid x)$ ja ehdollinen odotusarvo $E(Y \mid X = x)$ (b) ehdollinen varianssi $\mathrm{var}(Y \mid X = x)$.



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 8.3. *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 8.2. 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

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

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

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

Lause 8.4. *Sm:n varianssi on yhtä kuin sen ehdollisen varianssin odotusarvon sekä ehdollisen odotusarvon varianssin summa, eli*

$$\operatorname{var}(g(X,Y)) = E \operatorname{var}(g(X,Y) \mid X) + \operatorname{var} E(g(X,Y) \mid X). \tag{8.8}$$

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.

8.5 Yhteisjakauman määrittely hierarkkisesti

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

Esimerkki 8.3. (Diskreetti ja diskreetti.) Hyönteisen munimien munien lkm $Y \sim \operatorname{Poi}(\lambda)$, jossa $\lambda > 0$ on jakauman odotusarvo. Kukin munista kehittyy toukaksi toisistaan riippumatta tn:llä $0 < \theta < 1$. Olkoon X toukaksi kehittyvien munien lkm. Tällöin

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

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 voidaan määritellä myös sanomalla, että

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

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 $X \sim \mathrm{Poi}(\lambda\theta)$. $\triangle$

Esimerkki 8.4. (Jatkuva ja diskreetti.) Olkoon sm:lla Θ betajakauma $\mathrm{Be}(\alpha, \beta)$, 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 $\mathrm{Bin}(n, \theta)$. 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

$$X \mid \Theta \sim \mathrm{Bin}(n, \Theta),$$
$$\Theta \sim \mathrm{Be}(\alpha, \beta)$$

jossa $\alpha, \beta > 0$ ja $n \ge 0$ 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

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

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 jakaumaperheessa: 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 8.5. (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 $N(\mu_X, \sigma_X^2)$. Y:n regressiofunktio sekä ehdollinen varianssi ovat

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

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

Esimerkki 8.6. (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 7.6 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. △

Luku 9

Moniulotteinen jakauma

Moniulotteisella satunnaisvektorilla 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.

9.1 Satunnaisvektori

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

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

Jos kaikki satunnaisvektorin $\mathbf{X}$ komponentit X_i ovat diskreettejä sm:ia, niin sen ptnf eli sm:ien X_i yhteispistetodennäköisyysfunktio (yptnf) on

$$f_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}) = P(X_1 = x_1, \ldots, X_n = x_n). \tag{9.1}$$

Jos $g(\mathbf{X})$ on sv:n $\mathbf{X}$ reaaliarvoinen muunnos, niin sen odotusarvo saadaan kaavalla

$$Eg(\mathbf{X}) = \sum_{\mathbf{x}} g(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x}), \tag{9.2}$$

mikäli kyseinen summa suppenee itseisesti.

Satunnaisvektorilla $\mathbf{X}$ on jatkuva jakauma, mikäli sillä on tiheysfunktio $f_{\mathbf{X}}$, mikä tarkoittaa sitä, että

$$P(\mathbf{X} \in B) = \int_B f_{\mathbf{X}} = \int_B f_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x}, \qquad \text{kaikilla } B \subset \mathbb{R}^n. \tag{9.3}$$

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

$$Eg(\mathbf{X}) = \int g(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x} \tag{9.4}$$

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 $n \geq 2$, 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

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

ainakin niissä pisteissä $(x_1, \ldots, x_n)$, 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

$$Eg(\mathbf{X}, \mathbf{Y}) = \sum_{\mathbf{x}} \int g(\mathbf{x}, \mathbf{y})\, f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})\, \mathrm{d}\mathbf{y}, \tag{9.5}$$

mikäli kyseinen odotusarvo on olemassa.

Kahden satunnaisvektorin $\mathbf{X}$ ja $\mathbf{Y}$ riippumattomuus määritellään täysin vastaavasti kuin satunnaismuuttujien kohdalla, ks. jakso 3.5. 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.

9.2 Odotusarvovektori ja kovarianssimatriisi

Määritelmä 9.1. Asetamme seuraavat määritelmät, mikäli kyseessä olevat odotusarvot ovat olemassa.

- Satunnaisvektorin $\mathbf{X} = (X_1, \ldots, X_n)$ *odotusarvo(vektori)* on n-komponenttinen vakiovektori
$$E\mathbf{X} = E(\mathbf{X}) = (EX_1, \ldots, EX_n) = \begin{bmatrix} EX_1 \\ \vdots \\ EX_n \end{bmatrix}.$$

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

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

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

Kuten aikaisemminkin, satunnaisvektorin kovarianssi $\operatorname{Cov}()$ on yhden argumentin funktio, ja kahden satunnaisvektorin kovarianssi $\operatorname{cov}(,)$ on kahden argumentin funktio. Huomaa, että kahden satunnaisvektorin kovarianssi $\operatorname{cov}(\mathbf{X}, \mathbf{Y})$ on matriisi.

Määritelmästä seuraa suoraan, että
$$E(\mathbf{Z}^T) = (E\mathbf{Z})^T. \tag{9.6}$$

Laskusääntö
$$E[\mathbf{A}\mathbf{Z}\mathbf{B} + \mathbf{C}] = \mathbf{A}(E\mathbf{Z})\mathbf{B} + \mathbf{C}, \tag{9.7}$$

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

$$\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \quad \Rightarrow \quad \mathbf{G}(\mathbf{Y}) \perp\!\!\!\perp \mathbf{H}(\mathbf{Z}). \tag{9.8}$$

(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

$$\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \quad \Rightarrow \quad E[\mathbf{G}(\mathbf{Y})\,\mathbf{H}(\mathbf{Z})] = E[\mathbf{G}(\mathbf{Y})]\,E[\mathbf{H}(\mathbf{Z})] \tag{9.9}$$

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

Lause 9.1 (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*

$$\operatorname{cov}(\mathbf{A}\mathbf{X} + \mathbf{v}, \mathbf{B}\mathbf{Y} + \mathbf{w}) = \mathbf{A}\,\operatorname{cov}(\mathbf{X}, \mathbf{Y})\,\mathbf{B}^T. \tag{9.10}$$

Erityisesti

$$\operatorname{Cov}(\mathbf{A}\mathbf{X} + \mathbf{v}) = \mathbf{A}\,\operatorname{Cov}(\mathbf{X})\mathbf{A}^T. \tag{9.11}$$

Todistus. Kaikki kohdat seuraavat helpoilla laskuilla kovarianssin määritelmästä sekä laskusäännoistä (9.7) ja (9.9). □

Esimerkki 9.1. (Summan kovarianssimatriisi) Jos satunnaisvektorit $\mathbf{X}$ ja $\mathbf{Y}$ ovat samanpituisia, niin

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

Jos $\mathbf{X}$ ja $\mathbf{Y}$ eivät korreloi, niin $\operatorname{cov}(\mathbf{X},\mathbf{Y}) = \mathbf{0}$ ja $\operatorname{cov}(\mathbf{Y},\mathbf{X}) = \mathbf{0}$, joten korreloimattomille satunnaisvektoreille summan kovarianssimatriisi on summattavien satunnaisvektoreiden kovarianssimatriisien summa. Tämä pitää erityisesti paikkansa riippumattomille satunnaisvektoreille. △

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 $\mathbf{v} \ne 0$.

Huomautus: Positiivinen (semi)definiittisyys ei seuraa siitä, että kaikki matriisin alkiot ovat positiivisia. Esim. matriisi

$$\mathbf{B} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

on symmetrinen, mutta $\mathbf{B}$ ei ole positiivisesti semidefiniitti, sillä esim. valitsemalla $\mathbf{v} = (-1, 1)$ saadaan $\mathbf{v}^T B\mathbf{v} = -2$. Definittisyys on helpointa ymmärtää ominaisarvojen etumerkkien avulla seuraavasti.

- Symmetrinen matriisi $\mathbf{B}$ on positiivisesti semidefiniitti jos ja vain jos kaikki sen ominaisarvot ovat suurempia tai yhtä kuin nolla.
- Symmetrinen matriisi $\mathbf{B}$ on positiivisesti definiitti jos ja vain jos kaikki sen ominaisarvot ovat aidosti suurempia kuin nolla.

Lause 9.2. *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 $\mathbf{v} \ne \mathbf{0}$ 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

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

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.

9.3 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

$$f_{\mathbf{X}}(\mathbf{x}) = \int f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) \,\mathrm{d}\mathbf{y}, \tag{9.12}$$

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 9.2. 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

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

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

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

9.4 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 $\mathbf{X}_1, \ldots, \mathbf{X}_n$ ovat riippumattomia, jos

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

kaikilla argumenteilla.

Määritelmä 9.2. 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

$$f_{\mathbf{X},\mathbf{Y}|\mathbf{Z}}(\mathbf{x}, \mathbf{y} \mid \mathbf{z}) = f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x} \mid \mathbf{z})\, f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y} \mid \mathbf{z}), \qquad \text{kaikilla } \mathbf{x},\ \mathbf{y} \text{ ja } \mathbf{z} \tag{9.13}$$

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

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

kaikilla $\mathbf{x}_1, \ldots, \mathbf{x}_n$ 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.

9.5 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 9.3. 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 $\mathrm{Be}(1,1)$, 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 $\mathrm{Be}(1+k, 1+n-k)$. Tämän laskun olemme jo ratkaisseet esimerkissä 8.4.

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

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

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 9.4. 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

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

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

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

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

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

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. △

Esimerkki 9.5. (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

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

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

9.6 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

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

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

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

eli

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

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

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

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 7.9).

Lause 9.3. *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

$$f_{\mathbf{X}}(\mathbf{x})\, |\partial \mathbf{x}| = f_{\mathbf{Y}}(\mathbf{y})\, |\partial \mathbf{y}|, \qquad \text{kun} \tag{9.14}$$

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) \iff \mathbf{x} = \mathbf{h}(\mathbf{y}), \qquad \mathbf{x} \in A, \mathbf{y} \in B \tag{9.15}$$

Esimerkki 9.6. (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

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

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

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

$\triangle$

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

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{i \geq 1} 1_{B_i}(\mathbf{y})\, f_{\mathbf{X}}(\mathbf{h}_i(\mathbf{y}))\, |J_{\mathbf{h}_i}(\mathbf{y})|. \tag{9.16}$$

Tämä tulos on yksiulotteisen tuloksen (lause 2.13) moniulotteinen yleistys.

9.7 Satunnaisvektorin momenttiemäfunktio

Määritelmä 9.3. 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.

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 9.4. *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.

Määritelmä 9.4 (Gradientti, Hessen matriisi)**.** (Pysty)vektori

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

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

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

Huomautuksia. Hessen matriisin käsitteen esitti 1800-luvulla preussilainen matemaatikon Ludwig Otto Hesse. Jotkut kutsuvat sitä nimellä Hessin matriisi, mutta tämä on virheellinen käännös englannikieliselle termille *Hessian matrix*. 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 9.5. *Olkoot sv:n* $\mathbf{X}$ *momenttiemäfunktio* M *ja kumulanttiemäfunktio* K *määriteltyjä jossakin origon ympäristössä. Tällöin*

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

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

Esimerkki 9.7. Multinomijakaumalle $\mathbf{X} \sim \operatorname{Mult}(k, \mathbf{p})$ momenttiemäfunktion saa laskettua multinomikaavan avulla,

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

Tätä tai sen logaritmia derivoimalla nähdään helposti, että

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

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

Lause 9.6. *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

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

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. $\square$

Luku 10

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.

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

Määritelmä 10.1. 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

$$\begin{aligned} f_{\mathbf{U}}(\mathbf{u}) &= \prod_{i=1}^{n} f_{U_i}(u_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \,\mathrm{e}^{-\frac{1}{2}u_i^2} \\ &= (2\pi)^{-n/2} \exp(-\frac{1}{2}(u_1^2 + \cdots + u_n^2)) = (2\pi)^{-n/2} \;\exp(-\frac{1}{2}\mathbf{u}^T\mathbf{u}). \end{aligned} \tag{10.1}$$

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

$$E\mathbf{U} = \mathbf{0}_n, \qquad \operatorname{Cov}\mathbf{U} = \mathbf{I}_n. \tag{10.2}$$

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

$$M_{\mathbf{U}}(\mathbf{t}) = E\exp(t^T\mathbf{U}) = E\prod_{i=1}^{n}\exp(t_iU_i) = \prod_{i=1}^{n}\mathrm{e}^{\frac{1}{2}t_i^2} = \exp(\frac{1}{2}\mathbf{t}^T\mathbf{t}). \tag{10.3}$$

Jos sv $\mathbf{X}$ määritellään kaavalla

$$\mathbf{X} = \mathbf{A}\mathbf{U} + \boldsymbol{\mu}, \tag{10.4}$$

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

$$E\mathbf{X} = \boldsymbol{\mu}, \qquad \operatorname{Cov}\mathbf{X} = \mathbf{A}\,\mathbf{I}_n\mathbf{A}^T = \mathbf{A}\mathbf{A}^T. \tag{10.5}$$

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

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{t}) &= E \exp(\mathbf{t}^T \mathbf{X}) = E \exp(\mathbf{t}^T (\mathbf{A}\mathbf{U} + \boldsymbol{\mu})) = \exp(\mathbf{t}^T \boldsymbol{\mu}) \, M_{\mathbf{U}}(\mathbf{A}^T \mathbf{t}) \\ &= \exp(\mathbf{t}^T \boldsymbol{\mu} + \frac{1}{2} \mathbf{t}^T \mathbf{A}\mathbf{A}^T \mathbf{t}) = \exp(\mathbf{t}^T \boldsymbol{\mu} + \frac{1}{2} \mathbf{t}^T \mathbf{\Sigma} \mathbf{t}) \end{aligned} \tag{10.6}$$

Määrittelemme kohta multinormaalijakauman esityksen (10.4) avulla. Sitä ennen tarkastelmme kysymystä, kuinka mielivaltainen kovarianssimatriisi $\mathbf{\Sigma}$ voidaan esittää tulona $\mathbf{A}\mathbf{A}^T$.

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

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^T, \tag{10.7}$$

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

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

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

10.2 Yleinen multinormaalijakauma

Määritelmä 10.2. Sv:lla $\mathbf{X}$ on multinormaalijakauma, jos sillä on sama jakauma kuin vektorilla

$$\mathbf{A}\mathbf{U} + \boldsymbol{\mu} \tag{10.8}$$

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 (10.5) ja (10.6) seuraa, että

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

Lause 10.1. *Määritelmästä 10.2 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 $\mathbf{\Sigma} \in \mathbb{R}^{m \times m}$ on positiivisesti semidefiniitti matriisi, niin on olemassa sv $\mathbf{X}$, jolla on multinormaalijakauma odotusarvovektorilla $\boldsymbol{\mu}$ ja kovarianssimatriisilla $\mathbf{\Sigma}$.

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 $\mathbf{\Sigma}$ jaetaan tekijöihin $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^T$, ja sen jälkeen jakaumaa noudattava sv konstruoidaan kaavalla (10.8). □

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 10.2 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

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

10.3 Affiinin muunnoksen jakauma

Lause 10.2 (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 jakaumalle esitystä (10.8):

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

Määritelmän 10.2 mukaan sv:lla $\mathbf{B}\mathbf{X} + \mathbf{b}$ on multinormaalijakauma. Lisäksi

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

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 10.2 avulla, tai vaihtoehtoisesti poimimalla asiaankuuluva osa $\mathbf{X}$:n jakauman parametreista, sillä

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

10.4 Tiheysfunktio

Lause 10.3 (Multinormaalijakauman tiheysfunktio)**.** *Jos $\mathbf{\Sigma}$ on symmetrinen ja positiivisesti definiitti matriisi, niin jakauman $N_m(\boldsymbol{\mu}, \mathbf{\Sigma})$ tiheysfunktio on*

$$f(\mathbf{x}) = (2\pi)^{-m/2} \left(\det(\mathbf{\Sigma})\right)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right). \tag{10.9}$$

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. Tämän jälkeen $\mathbf{X} \sim N_m(\boldsymbol{\mu}, \mathbf{\Sigma})$, kun

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

Sitten sovelletaan 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,

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

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 9.2 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.

10.5 Tiheysfunktion tasa-arvopinnat

Moniulotteisen normaalijakauman tiheysfunktion (10.9) 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

$$\mathbf{\Sigma}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \qquad i = 1, \dots, n, \tag{10.10}$$

$$\mathbf{v}_i^T \mathbf{v}_j = \begin{cases} 1, & \text{kun } i = j, \\ 0, & \text{kun } i \neq j. \end{cases} \tag{10.11}$$

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 $\mathbf{\Lambda}$,

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

Tällöin ominaisarvon ja ominaisvektorin määritelmästä seuraa kaava

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

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*,

$$\mathbf{\Sigma} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T, \quad \text{jossa} \quad \mathbf{V}^{-1} = \mathbf{V}. \tag{10.12}$$

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

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

Normaalijakauman $N_m(\boldsymbol{\mu}, \mathbf{\Sigma})$ tiheysfunktio $f(\mathbf{x})$ (10.9) saa vakioarvon niillä argumenteilla $\mathbf{x}$, joilla eksponenttifunktion argumenttissa oleva lauseke saa vakioarvon, mutta

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

jossa vektori $\mathbf{y}$ määritellään kaavalla

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

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

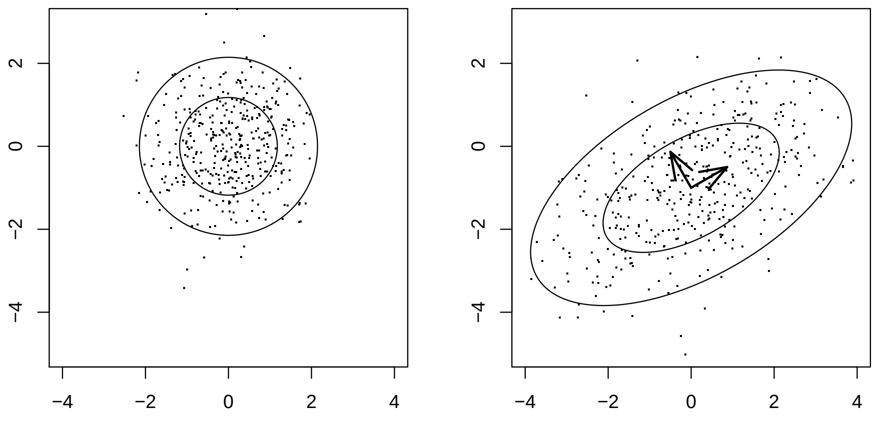
$$(\mathbf{V}\mathbf{e}_i)^T (\mathbf{V}\mathbf{e}_j) = \mathbf{e}_i^T \mathbf{V}^T \mathbf{V} \mathbf{e}_j = \mathbf{e}_i^T \mathbf{e}_j.$$

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

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

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

Kuva 10.1 Kaksiulotteinen normaalijakauma: (a) $N(\mathbf{0}, \mathbf{I})$, (b) $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.



(a) (b)

Esimerkki 10.1. (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 10.1 a on piirretty otosvektoreita $\mathbf{u}_i$ kaksiulotteisesta normaalijakaumasta $N(\mathbf{0}, \mathbf{I})$ sekä piirretty sen tiheysfunktion tasa-arvokäyriä. Kuvassa 10.1 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 $\mathbf{V}\mathbf{e}_2$. △

10.6 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 komponenteista $\mathbf{X}$ ja $\mathbf{Y}$ 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

$$\boldsymbol{\mu} = E\mathbf{Z} = \begin{bmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{bmatrix} = \begin{bmatrix} E\mathbf{X} \\ E\mathbf{Y} \end{bmatrix}, \tag{10.13}$$

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

$$\begin{aligned}
\mathbf{\Sigma} &= \begin{bmatrix}
\operatorname{cov}(X_1, X_1) & \dots & \operatorname{cov}(X_1, X_k) & \operatorname{cov}(X_1, Y_1) & \dots & \operatorname{cov}(X_1, Y_m) \\
\vdots & & \vdots & \vdots & & \vdots \\
\operatorname{cov}(X_k, X_1) & \dots & \operatorname{cov}(X_k, X_k) & \operatorname{cov}(X_k, Y_1) & \dots & \operatorname{cov}(X_k, Y_m) \\
\operatorname{cov}(Y_1, X_1) & \dots & \operatorname{cov}(Y_1, X_k) & \operatorname{cov}(Y_1, Y_1) & \dots & \operatorname{cov}(Y_1, Y_m) \\
\vdots & & \vdots & \vdots & & \vdots \\
\operatorname{cov}(Y_m, X_1) & \dots & \operatorname{cov}(Y_m, X_k) & \operatorname{cov}(Y_m, Y_1) & \dots & \operatorname{cov}(Y_m, Y_m)
\end{bmatrix} \\
&= \begin{bmatrix} \mathbf{\Sigma}_{XX} & \mathbf{\Sigma}_{XY} \\ \mathbf{\Sigma}_{YX} & \mathbf{\Sigma}_{YY} \end{bmatrix} = \begin{bmatrix} \operatorname{cov}(\mathbf{X}, \mathbf{X}) & \operatorname{cov}(\mathbf{X}, \mathbf{Y}) \\ \operatorname{cov}(\mathbf{Y}, \mathbf{X}) & \operatorname{cov}(\mathbf{Y}, \mathbf{Y}) \end{bmatrix}
\end{aligned} \tag{10.14}$$

Tällöin reunajakaumat ovat jakson 10.3 kaavojen mukaan

$$\mathbf{X} \sim N(\boldsymbol{\mu}_X, \mathbf{\Sigma}_{XX}), \qquad \mathbf{Y} \sim N(\boldsymbol{\mu}_Y, \mathbf{\Sigma}_{YY}). \tag{10.15}$$

Lause 10.4. *Jos vektorilla* $(\mathbf{X}, \mathbf{Y})$ *on multinormaalijakauma, niin*

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

Todistus. Koska multinormaalisen yhteisjakauman momenttiemäfunktio on määritelty kaikilla argumenteilla, niin lauseen 9.6 mukaan 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 (10.13) ja (10.14), niin

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

Toisaalta reunajakaumien momenttiemäfunktioiden tulo on

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

Edellä

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

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

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

eli silloin, kun

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

Jos oletetaan multinormaaliset reunajakaumat

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

niin yhdistetyn vektorin $(\mathbf{X}, \mathbf{Y})$ jakauma ei välttämättä ole multinormaalinen. Ei edes vaikka oletettaisiin lisäehto $\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}$. Edellisen lauseen kaavoja tarkastelemalla saadaan kuitenkin helposti todistettua seuraava tulos.

Lause 10.5. *Jos*

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

ja $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$*, niin vektori* $(\mathbf{X}, \mathbf{Y})$ *noudattaa multinormaalijakaumaa* $N(\mu, \boldsymbol{\Sigma})$*, jossa jakauman parametrit ovat*

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

Todistus. Riippumattomuuden nojalla yhteisjakauman momenttiemäfunktio on reunajakaumien momenttiemäfunktioiden tulo, joten kaikilla $\mathbf{u}, \mathbf{v}$.

$$\begin{aligned} M_{\mathbf{X},\mathbf{Y}}(\mathbf{u}, \mathbf{v}) &= M_{\mathbf{X}}(\mathbf{u})\, M_{\mathbf{Y}}(\mathbf{v}) \\ &= \exp\left(\mathbf{u}^T \boldsymbol{\mu}_X + \frac{1}{2} \mathbf{u}^T \boldsymbol{\Sigma}_{XX} \mathbf{u} + \mathbf{v}^T \boldsymbol{\mu}_Y + \frac{1}{2} \mathbf{v}^T \boldsymbol{\Sigma}_{YY} \mathbf{v} \right) \\ &= \exp\left(\begin{bmatrix} \mathbf{u}^T & \mathbf{v}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \mathbf{u}^T & \mathbf{v}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{XX} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{YY} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \right) \end{aligned}$$

ja tämä on multinormaalijakauman $N(\mu, \boldsymbol{\Sigma})$ momenttiemäfunktio. □

Esimerkki 10.2. Jos $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$ ja

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

ja $\mathbf{A}$ sekä $\mathbf{B}$ ovat vakiomatriiseja siten, että vektorit $\mathbf{A}\,\mathbf{X}$ ja $\mathbf{B}\,\mathbf{Y}$ ovat samanpituisia, niin satunnaisvektorilla

$$\mathbf{Z} = \mathbf{A}\,\mathbf{X} + \mathbf{B}\,\mathbf{Y}$$

on multinormaalijakauma.

Tämä nähdään siitä, että yhdistetyllä vektorilla $(\mathbf{X}, \mathbf{Y})$ on multinormaalijakauma, ja

$$\mathbf{A}\,\mathbf{X} + \mathbf{B}\,\mathbf{Y} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$$

Jakauman parametrit saadaan selville laskemalla vektorin odotusarvo ja kovarianssimatriisi,

$$\begin{aligned} E\mathbf{Z} &= E(\mathbf{A}\,\mathbf{X} + \mathbf{B}\,\mathbf{Y}) = \mathbf{A}\,\boldsymbol{\mu}_X + \mathbf{B}\,\boldsymbol{\mu}_Y \\ \mathrm{Cov}\,\mathbf{Z} &= \mathrm{Cov}(\mathbf{A}\,\mathbf{X}) + \mathrm{Cov}(\mathbf{B}\,\mathbf{Y}) = \mathbf{A}\,\boldsymbol{\Sigma}_{XX}\,\mathbf{A}^T + \mathbf{B}\,\boldsymbol{\Sigma}_{YY}\,\mathbf{B}^T \end{aligned}$$

△

10.7 Ehdolliset jakaumat

Lause 10.6. *Olkoon* $(\mathbf{X}, \mathbf{Y}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, ja käytetään odotusarvovektorille* $\boldsymbol{\mu}$ *ositusta* (10.13)*, ja kovarianssimatriisille* $\boldsymbol{\Sigma}$ *ositusta* (10.14)*. 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

$$\mathbf{V} = \mathbf{Y} - \mathbf{B}\mathbf{X}, \tag{10.16}$$

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

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

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 (10.16) 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

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

□

10.8 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

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

Tässä eksponenttifunktion argumentin kaava saadaan kertomalla auki tulo

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

Reunajakaumat ovat

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

Ehdolliset jakaumat saadaan lauseen 10.6 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

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

10.9 Normaalijakauman otoskeskiarvon ja otosvarianssin yhteisjakauma

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

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2. \tag{10.17}$$

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 $X_1, \dots, X_n$ ovat riippumattomia, ja niillä kaikilla on normaalijakauma $N(\mu, \sigma^2)$. Tällöin vektorilla $(X_1, \dots, X_n)$ on multinormaalijakauma $N_n(\mathbf{m}, \mathbf{\Sigma})$ parametreilla

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

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 10.7. *Olkoot $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$, ja määritellään $\bar{X}$ ja S^2 kaavoilla* (10.17). *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 $\mathbf{X} = (X_1, \dots, X_n)$ voidaan lausua muodossa

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

jossa $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^n$ 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

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

josta

$$\bar{X}\,\mathbf{1} = \left(\frac{1}{\sqrt{n}}\mathbf{u}^T\mathbf{X}\right)\sqrt{n}\,\mathbf{u} = \mathbf{u}\mathbf{u}^T\mathbf{X}$$

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

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

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

$$(n-1)S^2 = \mathbf{R}^T\mathbf{R} = \|\mathbf{R}\|^2. \tag{10.18}$$

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 10.4 perusteella $\bar{X}$ ja $\mathbf{R}$ ovat riippumattomia. Koska S^2 saadaan laskettua kaavan (10.18) 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, 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

$$\mathbf{R}^T\mathbf{R} = \left(\mathbf{V}\mathbf{V}^T\mathbf{X}\right)^T (\mathbf{V}\mathbf{V}^T\mathbf{X}) = \mathbf{X}^T\mathbf{V}\mathbf{V}^T\mathbf{X} = \|\mathbf{V}^T\mathbf{X}\|^2. \tag{10.19}$$

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 (10.18) ja (10.19) 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 $ES^2 = \sigma^2$. □

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 $Z \perp\!\!\!\perp Y$.

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 $\mathbf{u}_1, \dots, \mathbf{u}_p$, ja asetetaan nämä vektorit matriisin $\mathbf{U}$ pystyriveiksi, $\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_p \end{bmatrix}$. 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 11

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.

11.1 Suurten lukujen laki

Palautetaan mieleen stokastisen suppenemisen ja melkein varman suppenemisen määritelmät.

Määritelmä 11.1. 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,

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

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

Määritelmä 11.2. 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

$$X_n \xrightarrow{\text{a.s.}} Y \qquad \Rightarrow \qquad X_n \xrightarrow{\mathrm{P}} Y \tag{11.1}$$

Todistimme jaksossa 6.1 ns. heikon suurten lukujen lain (engl. *weak law of large numbers, WLLN*). Jos sitä sovelletaan *i.i.d.*-jonoon eli jonoon riippumattomia ja samoin jakautuneita satunnaismuuttujia, niin heikko suurten lukujen laki toteaa, että tiettyjen momenttiehtojen ollessa voimassa, n ensimmäisen satunnaismuuttujan aritmeettinen keskiarvo suppenee stokastisesti kohti vastaavaa odotusarvoa. Heikon suurten lukujen lain todistus oli suoraviivainen Tšebyševin epäyhtälön sovellus.

Vahvan suurten lukujen lain mukaan keskiarvon suppeneminen on peräti melkein varmaa. Tämän lauseen ainoa momenttiehto on se, että odotusarvon pitää olla olemassa.

Lause 11.1 (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 keskiarvojen*

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

muodostama jono 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.

11.2 Jakaumasuppeneminen

Määritelmä 11.3. 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.

Merkitsemme jakaumasuppenemista seuraavasti,

$$X_n \xrightarrow{\text{d}} Y.$$

Nuolen yläindeksi d tulee sanasta *distribution*. (Myös merkinnät $X_n \xrightarrow{\mathcal{L}} Y$ sekä $X_n \xrightarrow{\text{w}} Y$ ovat yleisiä jakaumasuppenemiselle.)

Jos rajajakauma on jokin tuttu jakauma, kuten $N(0,1)$, niin jakaumasuppenemista voidaan merkitä siten, että satunnaismuuttujan sijasta käytetään rajajakauman tunnusta, seuraavaan tapaan,

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

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.

11.3 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 11.2 (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

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

Keskeisen raja-arvolauseen väite voitaisiin yhtä hyvin muotoilla siten, että

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

mikä auttaa ymmärtämään seuraavaa, moniulotteista versiota keskeisestä raja-arvolauseesta.

Lause 11.3 (Keskeinen raja-arvolause, moniulotteinen versio)**.** *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*

$$\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{\text{d}} N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{11.2}$$

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

11.4 Normaaliapproksimaatio

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

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

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

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

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

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

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 11.1. (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

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

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

$$P(\sum_{1}^{n} X_i \leq 13) \approx 0.268.$$

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,

$$\{\sum_{1}^{n} X_i \leq 13\} = \{\sum_{1}^{n} X_i \leq 13 + 0.5\}.$$

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

$$P(\sum_{1}^{n} X_i \leq 13) \approx P(Z \leq \frac{13.5-15}{\sqrt{6}}) \approx 0.270.$$

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

Huomaa, että edellisessä esimerkissä käytettiin jatkuvuuskorjausta. Kun satunnaismuuttujat X_i ovat kokonaislukuarvoisia, niin myös niiden summa $S = \sum_{i=1}^{n} X_i$ on kokonaislukuarvoinen, joten sen yhteydessä on mielekästä tarkastella häntätodennäköisyyksiä kokonaislukuarvoisissa pisteissä. Jos x ja y ovat kokonaislukuja, niin

$$\{S \geq x\} = \{S \geq x - \frac{1}{2}\}$$
$$\{S \leq y\} = \{S \leq y + \frac{1}{2}\}$$
$$\{x \leq S \leq y\} = \{x - \frac{1}{2} \leq S \leq y + \frac{1}{2}\}$$

Yleisesti ottaen tässä tapauksessa normaaliapproksimaatiossa saavutetaan parempi tarkkuus, mikäli tällaiset kokonaislukuarvoisen satunnaismuuttujan kokonaislukuarvoiset rajat korvataan tähän tapaan kokonaislukujen puoliväleissä olevilla arvoilla. Tämä johtaa summalle S approksimaatioihin

$$\begin{aligned} P(S \geq x) &= P(S \geq x - \frac{1}{2}) \approx 1 - \Phi\left(\frac{x - \frac{1}{2} - n\mu}{\sqrt{n}\,\sigma}\right) \\ P(S \leq y) &= P(S \leq y + \frac{1}{2}) \approx \Phi\left(\frac{y + \frac{1}{2} - n\mu}{\sqrt{n}\,\sigma}\right) \\ P(x \leq S \leq y) &= P(x - \frac{1}{2} \leq S \leq y + \frac{1}{2}) \approx \Phi\left(\frac{y + \frac{1}{2} - n\mu}{\sqrt{n}\,\sigma}\right) - \Phi\left(\frac{x - \frac{1}{2} - n\mu}{\sqrt{n}\,\sigma}\right) \end{aligned}$$

11.5 Deltamenetelmä

Seuraavassa esitetään kaksi yleisesti käytettyä versiota ns. deltamenetelmästä (engl. *delta method*). Sen avulla voidaan normaaliapproksimaatiota käyttää myös sileille funktioille otoskeskiarvoista.

Lause 11.4 (Deltamenetelmä). *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

$$\sqrt{n}(g(X_n) - g(a)) \xrightarrow{\mathrm{d}} N(0, \sigma^2 (g'(a))^2). \tag{11.3}$$

Heuristinen perustelu. Oletuksesta seuraa, että sm:n X_n jakauma on voimakkaasti keskittynyt arvon a lähelle, sillä

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

Tässä jono $\frac{1}{\sqrt{n}}$ suppenee kohti nollaa ja jono $\sqrt{n}(X_n - a)$ on siinä mielessä stabiili, että sillä on rajajakauma. Tämän takia suurilla n tuntuu järkevältä approksimoida satunnaismuuttujaa $g(X_n)$ pisteessä a kehitetyllä ensimmäisen asteen Taylorin polynomilla

$$g(X_n) \approx g(a) + g'(a)(X_n - a)$$

Tämän jälkeen

$$\begin{aligned} \sqrt{n}(g(X_n) - g(a)) &\approx g'(a)\,\sqrt{n}(X_n - a) \\ &\overset{\mathrm{d}}{\approx} g'(a)\,N(0, \sigma^2) \\ &\overset{\mathrm{d}}{=} N(0, \sigma^2\,(g'(a))^2). \end{aligned}$$

Deltamenetelmää sovelletaan usein siten, että äärellisellä otoskoolla n approksimoidaan

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

Esimerkki 11.2. Tarkastellaan riippumattomia muuttujia $X_i \sim \mathrm{Bernoulli}(p)$, $i = 1, \ldots, n$, 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 helppojen laskujen avulla approksimaatio
$$\sqrt{n}(\hat{\theta}_n - \theta) \overset{\mathrm{d}}{\approx} N(0, \frac{1}{p(1 - p)}).$$
Tällä perusteella on esim. mahdollista johtaa asymptoottinen luottamusväli parametrille θ. △

Deltamenetelmän heuristisen perustelun saa helposti yleistettyä kattamaan seuraavan moniulotteisen version.

Lause 11.5. *(Deltamenetelmä, moniulotteinen versio) Jos k-ulotteisten satunnaisvektorien jono $(\mathbf{X}_n)$ toteuttaa ehdon*
$$\sqrt{n}(\mathbf{X}_n - \mathbf{a}) \xrightarrow{\mathrm{d}} N(0, \mathbf{\Sigma}),$$
jossa $\mathbf{a} \in \mathbb{R}^k$ on vakio, ja funktio $g : U \subset \mathbb{R}^k \to \mathbb{R}$ on derivoituva pisteessä a, niin
$$\sqrt{n}(g(\mathbf{X}_n) - g(\mathbf{a})) \xrightarrow{\mathrm{d}} N(\mathbf{0}, \nabla g(\mathbf{a})^T \, \mathbf{\Sigma} \, \nabla g(\mathbf{a})). \tag{11.4}$$

Heuristinen perustelu. Taas suurilla n approksimoida satunnaismuuttujaa $g(\mathbf{X}_n)$ pisteessä $\mathbf{a}$ kehitetyllä ensimmäisen asteen Taylorin polynomilla
$$g(\mathbf{X}_n) \approx g(\mathbf{a}) + \nabla g(\mathbf{a})^T (\mathbf{X}_n - \mathbf{a})$$
Tästä saadaan
$$\begin{aligned}\sqrt{n}(g(\mathbf{X}_n) - g(\mathbf{a})) &\approx \nabla g(\mathbf{a})^T \sqrt{n}(\mathbf{X}_n - \mathbf{a}) \\ &\overset{\mathrm{d}}{\approx} \nabla g(\mathbf{a})^T N(\mathbf{0}, \mathbf{\Sigma}) \\ &\overset{\mathrm{d}}{=} N(0, \nabla g(\mathbf{a})^T \, \mathbf{\Sigma} \, \nabla g(\mathbf{a})).\end{aligned}$$

11.6 Deltamenetelmän todistus

Edellisessä jaksossa esitetyt perustelut jättivät monia detaljeja huomioimatta. Tässä jaksossa esitellään työkaluja, joilla perustelujen aukot saadaan tilkittyä. Tätä tarkoitusta varten esitellään ilman todistuksia tiettyjä asymptoottisissa tarkasteluissa tärkeitä tuloksia.

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.

Voidaan osoittaa, että seuraavat implikaatiot pitävät paikkansa:

$$\mathbf{X}_n \xrightarrow{\text{a.s.}} \mathbf{Y} \quad \Rightarrow \quad \mathbf{X}_n \xrightarrow{\text{p}} \mathbf{Y} \quad \Rightarrow \quad \mathbf{X}_n \xrightarrow{\text{d}} \mathbf{Y}. \tag{11.5}$$

Yksikään implikaatioista ei päde yleisesti käänteiseen suuntaan. Jos rajamuuttuja $\mathbf{Y}$ on deterministinen eli vakio $\mathbf{c}$ eli jos $\mathbf{Y}$:n jakauma on degeneroitunut, niin tällöin osoittautuu, että jakaumasuppenemisesta seuraa stokastinen suppeneminen, eli

$$\mathbf{X}_n \xrightarrow{\text{d}} \mathbf{c} \quad (\mathbf{c} \text{ vakio}) \quad \Rightarrow \quad \mathbf{X}_n \xrightarrow{\text{p}} \mathbf{c}. \tag{11.6}$$

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

Lause 11.6 (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 $\mathbf{g} : U \subset \mathbb{R}^k \to \mathbb{R}^d$ on jatkuva sellaisessa joukossa $A \subset \mathbb{R}^k$, jossa rajasatunnaisvektorin $\mathbf{Y}$ arvot ovat melkein varmasti, ts. $P(\mathbf{Y} \in A) = 1$*

Otetaan käyttöön merkintä satunnaisvektoreiden komponenteille,

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

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

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

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 11.7 (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. □

Nyt voimme vihdoin esittää todistuksen yksiulotteiselle versiolle deltamenetelmästä, eli lauseelle 11.4

Todistus. Sovelletaan ensimmäisen kertaluvun Taylorin kehitelmää

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

jossa jäännöstermissä esiintyvä funktio r on

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

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) $X_n - a \xrightarrow{\mathrm{d}} 0$, ja koska rajamuuttuja on vakio, niin suppeneminen on peräti stokastista (ks. kaava (11.6)). Siis $X_n - a \xrightarrow{\mathrm{P}} 0$, eli $X_n \xrightarrow{\mathrm{P}} a$. Lisäksi

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

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

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

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 $g'(a)Y \sim N(0, (g'(a))^2\sigma^2)$. □