Luku 6

Otantajakaumien teoria

6.1 Riippumattomat satunnaismuuttujat

Muistamme edellisistä luvuista, että satunnaismuuttujat X_1 ja X_2 ovat riippumattomat (määritelmät 4.6 ja 5.5), jos

$$f(x_1, x_2) = f_1(x_1)f_2(x_2) \qquad \text{kaikilla } x_1 \in S_1,\ x_2 \in S_2,$$

missä $f(x_1, x_2)$ on X_1:n ja X_2:n yhteisjakauman tiheysfunktio, $f_1(x_1)$ on X_1:n ja $f_2(x_2)$ on X_2:n tiheysfunktio. Määritelmä yleistyy suoraviivaisesti usean satunnaismuuttujan tapaukseen. Satunnaismuuttujat X_1, X_2, …, X_n *ovat riippumattomat,* jos

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n).$$

Jos satunnaismuuttujat X_1, X_2, …, X_n *ovat riippumattomat,* niin myös niiden funktiot $u_1(X_1)$, $u_2(X_2)$, …, $u_n(X_n)$ ovat riippumattomat, mikäli kukin funktio u_i, $i = 1, 2, \ldots, n$ riippuu vain satunnaismuuttujasta X_i eikä siis satunnaismuuttujista X_j, $j \neq i$. Silloin erityisesti Lauseen 3.10 mukaan

$$E[u_1(X_1)u_2(X_2)\cdots u_n(X_n)] = E[u(X_1)]\,E[u_2(X_2)]\cdots E[u_n(X_n)].$$

Jos riippumattomat satunnaismuuttujat X_1, X_2, …, X_n noudattavat samaa jakaumaa (RSJ), jonka kertymäfunktio on $F(x)$, niin sanomme, että $X_1, X_2, \ldots, X_n$ on n:n kokoinen otos jakaumasta F. Kertymäfunktio edustaa populaatiota, josta otos tehdään.

Esimerkki 6.1 Olkoon $X_1, X_2, \ldots, X_n$ otos normaalijakaumasta $\mathrm{N}(\mu, \sigma^2)$. Silloin $X_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ ja X_1, X_2, …, X_n ovat riippumattomat.

Otoksen $X_1, X_2, \dots, X_n$ yhteisjakauman tiheysfunktio on siis

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= \prod_{i=1}^{n} f(x_i) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-[1/(2\sigma^2)](x_i-\mu)^2} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \mathrm{e}^{-[1/(2\sigma^2)]\sum_{i=1}^{n}(x_i-\mu)^2}, \end{aligned}$$

missä

$$f(x_i) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-[1/(2\sigma^2)](x_i-\mu)^2}, \qquad i = 1, 2, \dots, n.$$

□

Lause 6.1 (Apulauseen 5.1 yleistys) *Satunnaisvektorit* $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ *ja* $\boldsymbol{Y} = (Y_1, Y_2, \dots, Y_n)$ *ovat riippumattomat jos ja vain jos on olemassa sellaiset funktiot* $g(\boldsymbol{X})$ *ja* $h(\boldsymbol{Y})$*, että*

$$f(\boldsymbol{x}, \boldsymbol{y}) = g(\boldsymbol{x})h(\boldsymbol{y})$$

kaikilla $\boldsymbol{x}$*:n ja* $\boldsymbol{y}$*:n arvoilla, missä* g *ei riipu* $\boldsymbol{y}$*:stä ja* h *ei riipu* $\boldsymbol{x}$*:stä.*

6.2 Riippumattomien satunnaismuuttujien summan jakauma

Tilastollisissa sovelluksissa tarkastellaan tavallisesti erilaisia satunnaismuuttujien funktioita. Otoksesta $X_1, X_2, \dots, X_n$ laskettua reaali-, tai vektoriarvoista suuretta $T(X_1, X_2, \dots, X_n)$ sanotaan otoksen *tunnusluvuksi* (*statistics*). Kaksi tärkeää otoksen tunnuslukua ovat otoskeskiarvo $\overline{X}$ ja otosvarianssi S^2. Esimerkiksi $T_1(X_1, X_2, \dots, X_n) = \overline{X}$ on reaaliarvoinen ja $T_2(X_1, X_2, \dots, X_n) = (\overline{X}, S^2)$ on vektoriarvoinen.

Lause 6.2 *Olkoon* $X_1, X_2, \dots, X_n$ *otos jakaumasta, jonka tiheysfunktio on* $f(x)$*. Silloin satunnaismuuttujien* X_1, X_2, $\dots$, X_n *yhteisjakuman tiheysfunktio on* $f_1(x_1)f_2(x_2)\cdots f_n(x_n)$*. Jos* $g(y)$ *on satunnaismuuttujan* $Y = u(X_1, X_2, \dots, X_n)$ *tiheysfunktio, niin*

$$\begin{aligned} E(Y) &= \int_{S_y} y g(y)\,\mathrm{d}y \\ &= \int_S\int_S \cdots \int_S u(x_1, x_2, \dots, x_n) f_1(x_1) f_2(x_2) \cdots f_n(x_n)\,\mathrm{d}x_1\,\mathrm{d}x_2 \dots \mathrm{d}x_n, \end{aligned}$$

mikäli odotusarvo on olemassa. Diskreettejä satunnaismuuttujia koskeva vastaava tulos saadaan korvaamalla integraalit summalausekkeilla. Satunnaismuuttujien X_1, X_2, $\dots$, X_n *arvoalue on* S *ja* Y*:n arvoalue on* S_y*.*

Esimerkki 6.2 Heitetään kahta noppaa. Olkoon 1. nopan silmäluku X_1 ja 2. nopan silmäluku X_2. Määritetään nyt silmälukujen summan $Y = X_1 + X_2$ todennäköisyysfunktio $g(y)$. Tarkastellaan ensin yksittäisen arvon, esimerkiksi $y = 4$, todennäköisyyden $g(4)$ laskemista. Tapahtuma $\{Y = 4\}$ voi sattua kolmella toisensa poissulkevalla tavalla: $\{X_1 = 1,\ X_2 = 3\}$, $\{X_1 = 2,\ X_2 = 2\}$ ja $\{X_1 = 3,\ X_2 = 1\}$. Siksi

$$\begin{aligned} g(4) &= P(Y = 4) \\ &= P(X_1 = 1,\ X_2 = 3) + P(X_1 = 2,\ X_2 = 2) + P(X_1 = 3,\ X_2 = 1) \\ &= \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{3}{36}. \end{aligned}$$

Jatkamalla samalla periaatteella saadaan todennäköisyysfunktio $g(y)$:

y	2	3	4	5	6	7	8	9	10	11	12
$g(y)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

□

Yleisesti Esimerkin 6.2 todennäköisyydet voidaan laskea ns. *konvoluutiokaavalla*

$$g(y) = P(Y = y) = \sum_{k=1}^{y-1} f(k) f(y-k),$$

missä

$$f(k) = \frac{1}{6}, \qquad k = 1, 2, 3, 4, 5, 6$$

on nopan silmäluvun todennäköisyysfunktio.

Toinen tapa johtaa $g(y)$, on käyttää momenttifunktiota. Nopan silmäluvun momenttifunktio on

$$M_X(t) = E\big(\mathrm{e}^{tX}\big) = \frac{1}{6}\mathrm{e}^{t} + \frac{1}{6}\mathrm{e}^{2t} + \frac{1}{6}\mathrm{e}^{3t} + \frac{1}{6}\mathrm{e}^{4t} + \frac{1}{6}\mathrm{e}^{5t} + \frac{1}{6}\mathrm{e}^{6t}.$$

Koska silmäluvut ovat riippumattomat, niin Y:n momenttifunktio on

$$M_Y(t) = M_{X_1}(t) M_{X_2}(t) = [M_X(t)]^2.$$

Koska e^{kt}:n kerroin $M_Y(t)$:n lausekkeessa on todennäköisyys $P(Y = k)$, $k = 2, 3, \ldots, 12$, ne muodostavat Y:n todennäköisyysfunktion.

Lause 6.3 *Olkoot riippumattomien satunnaismuuttujien $X_1,\ X_2,\ \ldots,\ X_n$ odotusarvot $\mu_1,\ \mu_2,\ \ldots,\ \mu_n$ ja varianssit $\sigma_1^2,\ \sigma_2^2,\ \ldots,\ \sigma_n^2$. Silloin satunnaismuuttujan $Y = \sum_{i=1}^{n} a_i X_i$ odotusarvo ja varianssi ovat*

$$\mu_Y = \sum_{i=1}^{n} a_i \mu_i \qquad \textit{ja} \qquad \sigma_Y^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2,$$

missä $a_1,\ a_2,\ \ldots,\ a_n$ ovat annettuja vakioita.

Todistus. Koska odotusarvo on lineaarinen operaattori, niin

$$E(Y) = E\Big(\sum_{i=1}^{n} a_i X_i\Big) = \sum_{i=1}^{n} E(a_i X_i)$$
$$= \sum_{i=1}^{n} a_i\, E(X_i) = \sum_{i=1}^{n} a_i \mu_i.$$

Vastaavasti

$$\sigma^2 = E[(Y-\mu_Y)^2] = E\bigg[\Big(\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i \mu_i\Big)^2\bigg]$$
$$= E\Big[\sum_{i=1}^{n} a_i (X_i - \mu_i)\Big]^2 = E\Big[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (X_i-\mu_i)(X_j-\mu_j)\Big]$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j\, E[(X_i-\mu_i)(X_j-\mu_j)] = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma_{ij},$$

missä

$$\sigma_{ij} = E[(X_i-\mu_i)(X_j-\mu_j)].$$

Koska X_i ja X_j ovat riippumattomat, niin $\sigma_{ij} = 0$, kun $i \neq j$. Tästä seuraa, että

$$\sigma^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2.$$

□

Esimerkki 6.3 Olkoot X_1 ja X_2 riippumattomat satunnaismuuttujat, joiden odotusarvot ovat $\mu_1 = -4$ ja $\mu_2 = 3$ sekä varianssit vastaavasti $\sigma_1^2 = 4$ ja $\sigma_2^2 = 9$. Silloin satunnaismuuttujan $Y = 3X_1 - 2X_2$ odotusarvo ja varianssi ovat

$$\mu_Y = 3 \cdot (-4) + (-2) \cdot 3 = -18$$

ja

$$\sigma_Y^2 = 3^2 \cdot 4 + (-2)^2 \cdot 9 = 72.$$

□

Esimerkki 6.4 Olkoon $X_1, X_2, \ldots, X_n$ otos jakaumasta, jonka odotusarvo on μ ja varianssi σ^2. Silloin otoskeskiarvon

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

odotusarvo ja varianssi ovat

$$\mu_{\overline{X}} = \sum_{i=1}^{n} \left(\frac{1}{n}\right)\mu = \mu \qquad \text{ja} \qquad \sigma_{\overline{X}}^2 = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 \sigma^2 = \frac{\sigma^2}{n}.$$

Otosvarianssi on muotoa

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\Big(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\Big),$$

joten

$$E(S^2) = \frac{1}{n-1}\Big[\sum_{i=1}^{n} E(X_i^2) - n\,E(\overline{X}^2)\Big].$$

Koska $E(X_i^2) = \sigma^2 + \mu^2$ ja $E(\overline{X}^2) = \sigma^2/n + \mu^2$, niin laskemalla on helppo todeta, että $E(S^2) = \sigma^2$. Olemme siis osoittaneet, että $\overline{X}$ on μ:n ja S^2 on σ^2:n harhaton estimaattori. □

Lause 6.4 *Jos X_1, X_2, ..., X_n ovat riippumattomat satunnaismuuttujat, joiden momenttifunktiot ovat $M_{X_i}(t)$, $i = 1, 2, \ldots, n$, niin satunnaismuuttujan $Y = \sum_{i=1}^{n} a_i X_i$ momenttifunktio on*

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t).$$

Todistus. Satunnaismuuttujan Y momenttifunktio on

$$\begin{aligned} M_Y(t) &= E\big(\mathrm{e}^{tY}\big) = E\big(\mathrm{e}^{t(a_1X_1 + a_2X_2 + \cdots + a_nX_n)}\big) \\ &= E\big(\mathrm{e}^{a_1tX_1} \cdot \mathrm{e}^{a_2tX_2} \cdots \mathrm{e}^{a_ntX_n}\big) \\ &= E\big(\mathrm{e}^{a_1tX_1}\big)\, E\big(\mathrm{e}^{a_2tX_2}\big) \cdots E\big(\mathrm{e}^{a_ntX_n}\big), \end{aligned}$$

koska satunnaismuuttujat $\mathrm{e}^{a_itX_i}$ ovat keskenään riippumattomat. Momenttifunktion määritelmän mukaan

$$E\big(\mathrm{e}^{tX_i}\big) = M_{X_i}(t),$$

joten

$$E\big(\mathrm{e}^{a_itX_i}\big) = M_{X_i}(a_i t).$$

Siksi

$$M_Y(t) = M_{X_1}(a_1t)M_{X_2}(a_2t)\cdots M_{X_n}(a_nt) = \prod_{i=1}^{n} M_{X_i}(a_i t).$$

□

Esimerkki 6.5 Olkoon $X_1, X_2, \ldots, X_n$ otos Bernoullin jakaumasta $\mathrm{Ber}\big(\frac{1}{3}\big)$. Silloin

$$M(t) = \tfrac{2}{3} + \tfrac{1}{3}\mathrm{e}^t.$$

Jos $Y = X_1 + X_2 + \cdots + X_n$, niin

$$M_Y(t) = \prod_{i=1}^{n}\big(\tfrac{2}{3} + \tfrac{1}{3}\mathrm{e}^t\big) = \big(\tfrac{2}{3} + \tfrac{1}{3}\mathrm{e}^t\big)^n.$$

Tästä näemme, että $Y \sim \mathrm{Bin}\big(n, \frac{1}{3}\big)$. □

Seuraus 6.1 *Jos $X_1, X_2, \ldots, X_n$ on otos jakaumasta, jonka momenttifunktio on $M(t)$, niin*

1. *satunnaismuuttujan $Y = \sum_{i=1}^n X_i$ momenttifunktio on*

$$M_Y(t) = \prod_{i=1}^n M(t) = [M(t)]^n.$$

2. *otoskeskiarvon $\overline{X} = \sum_{i=1}^n (1/n)X_i$ momenttifunktio on*

$$M_{\overline{X}}(t) = \prod_{i=1}^n M\left(\frac{t}{n}\right) = \left[M\left(\frac{t}{n}\right)\right]^n.$$

Esimerkki 6.6 Olkoon X_1, X_2, X_3 otos eksponenttijakaumasta, jonka odotusarvo on θ. Eksponenttijakauman momenttifunktio on $M(t) = 1/(1 - \theta t)$, $t < 1/\theta$. Silloin summan $Y = X_1 + X_2 + X_3$ momenttifunktio on

$$M_Y(t) = [1/(1 - \theta t)]^3 = (1 - \theta t)^{-3}, \qquad t < \frac{1}{\theta},$$

mikä on gammajakauman $\text{Gamma}(3, \theta)$ momenttifunktio, joten $Y \sim \text{Gamma}(3, \theta)$. Toisaalta $\overline{X}$:n momenttifunktio on

$$M_{\overline{X}}(t) = \left[\left(1 - \frac{\theta t}{3}\right)^{-1}\right]^3 = \left(1 - \frac{\theta t}{3}\right)^{-3}, \qquad \text{kun } t < \frac{3}{t}.$$

Otoskeskiarvo $\overline{X}$ noudattaa siis gammajakaumaa $\text{Gamma}(3, \theta/3)$. $\square$

6.3 Normaalijakaumaan liittyvät jakaumat

Lause 6.5 *Jos $X_1, X_2, \ldots, X_n$ on otos normaalijakaumasta $\mathrm{N}(\mu, \sigma^2)$, niin otoskeskiarvon $\overline{X} = \sum_{i=1}^n (1/n)X_i$ jakauma on $\mathrm{N}(\mu, \sigma^2/n)$.*

Todistus. Koska $X_i \sim \mathrm{N}(\mu, \sigma^2)$, niin

$$M_{X_i}(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right).$$

Seurauslauseen 6.1 mukaan

$$\begin{aligned} M_{\overline{X}}(t) &= \left[\exp\left(\mu \cdot \frac{t}{n} + \frac{\sigma^2 (t/n)^2}{2}\right)\right]^n \\ &= \exp\left(\mu t + \frac{(\sigma^2/n)t^2}{2}\right), \end{aligned}$$

joka on normaalijakauman $\mathrm{N}(\mu, \sigma^2/n)$ momenttifunktio. Koska momenttifunktio määrittää yksikäsitteisesti satunnaismuuttujan jakauman, niin $\overline{X} \sim \mathrm{N}(\mu, \sigma^2/n)$. $\square$

Lause 6.6 *Olkoot $X_i \sim \chi^2(r_i)$, $i = 1, 2, \ldots, n$. Jos X_1, X_2, $\ldots$, X_n ovat riippumattomat, niin satunnaismuuttujan $Y = X_1 + X_2 + \cdots + X_n$ jakauma on $\chi^2(r_1 + r_2 + \cdots + r_k)$.*

Todistus. Satunnaismuuttujan Y momenttifunktio voidaan kirjoittaa muodossa

$$\begin{aligned} M_Y(t) &= E\left[\mathrm{e}^{t(X_1+X_2+\cdots+X_n)}\right] \\ &= E\left(\mathrm{e}^{tX_1}\right) \cdot E\left(\mathrm{e}^{tX_2}\right) \cdots E\left(\mathrm{e}^{tX_n}\right) \\ &= \prod_{i=1}^{n} M_{X_i}(t). \end{aligned}$$

Koska

$$M_{X_i}(t) = (1-2t)^{-r_i/2}; \qquad t < \frac{1}{2},$$

niin

$$M_Y(t) = \prod_{i=1}^{n} (1-2t)^{-r_i/2} = (1-2t)^{-(r_1+r_2+\cdots+r_n)/2}, \qquad t < \frac{1}{2}$$

on χ^2-jakauman $\chi^2(r_1 + r_2 + \cdots + r_n)$ momenttifunktio. Tästä seuraa, että $Y \sim \chi^2(r_1 + r_2 + \cdots + r_n)$. □

Lause 6.7 *Olkoon $Z_1, Z_2, \ldots, Z_n$ otos standardimuotoisesta normaalijakaumasta* $\mathrm{N}(0, 1)$. *Silloin $W = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ noudattaa jakaumaa $\chi^2(n)$.*

Todistus. Koska $Z_i \sim \chi^2(1)$, $i = 1, 2, \ldots, n$ ja Z_1^2, Z_2^2, $\ldots$, Z_n^2 ovat keskenään riippumattomat, niin tulos seuraa Lauseesta 6.6. □

Seuraus 6.2 *Olkoot X_1, X_2, $\ldots$, X_n riippumattomat ja $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$. Silloin satunnaismuuttuja*

$$W = \sum_{i=1}^{n} \frac{(X_i - \mu_i)^2}{\sigma_i^2}$$

noudattaa jakaumaa $\chi^2(n)$.

Lause 6.8 *Olkoon $X_1, X_2, \ldots, X_n$ otos normaalijakaumasta* $\mathrm{N}(\mu, \sigma^2)$, *$\overline{X} = \sum_{i=1}^{n} (1/n) X_i$ on otoskeskiarvo ja $S^2 = [1/(n-1)] \sum_{i=1}^{n} (X_i - \overline{X})^2$ on otosvarianssi. Silloin*

1. *$\overline{X}$ ja S^2 ovat riippumattomat satunnaismuuttujat,*
2. *$\overline{X}$ noudattaa normaalijakaumaa* $\mathrm{N}(\mu, \sigma^2/n)$,
3. *$(n-1)S^2/\sigma^2$ noudattaa χ^2-jakaumaa $\chi^2(n-1)$.*

Todistus. Kohdan 1 todistus sivuutetaan tässä yhteydessä. Kohta 2 on lause 6.5. Todistetaan nyt väite, että

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

Laskemalla voidaan todeta, että

$$\begin{aligned}
\sum_{i=1}^{n}\Big(\frac{X_i-\mu}{\sigma}\Big)^2 &= \sum_{i=1}^{n}\Big[\frac{(X_i-\overline{X})+(\overline{X}-\mu)}{\sigma}\Big]^2 \\
&= \sum_{i=1}^{n}\Big(\frac{X_i-\overline{X}}{\sigma}\Big)^2 + \frac{n(\overline{X}-\mu)^2}{\sigma^2} \\
&= \frac{(n-1)S^2}{\sigma^2} + \Big(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\Big)^2.
\end{aligned}$$

Koska $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$, niin $Z^2 \sim \chi^2(1)$. Vastaavasti Seurauslauseen 6.2 mukaan $W = \sum_{i=1}^{n}\big(\frac{X_i-\mu}{\sigma}\big)^2 \sim \chi^2(n)$.

Koska S^2 ja Z^2 ovat kohdan 1 mukaan riippumattomat, niin

$$\begin{aligned}
E\big(\mathrm{e}^{tW}\big) &= E\big(\mathrm{e}^{t[(n-1)S^2/\sigma^2+Z^2]}\big) = E\big(\mathrm{e}^{t(n-1)S^2/\sigma^2}\cdot \mathrm{e}^{tZ^2}\big) \\
&= E\big(\mathrm{e}^{t(n-1)S^2/\sigma^2}\big)\,E\big(\mathrm{e}^{tZ^2}\big).
\end{aligned}$$

Koska $W \sim \chi^2(n)$ ja $Z^2 \sim \mathrm{N}(0,1)$, niin

$$(1-2t)^{-n/2} = E\big[\mathrm{e}^{t(n-1)S/\sigma^2}\big]\cdot(1-2t)^{-1/2}.$$

Tästä seuraa, että

$$E\big[\mathrm{e}^{t(n-1)S/\sigma^2}\big] = (1-2t)^{-(n-1)/2}; \qquad t < \frac{1}{2},$$

joka on jakauman $\chi^2(n-1)$ momenttifunktio. Näin on lauseen väite 3 todistettu. □

Lause 6.9 *Olkoot X_1, X_2, ..., X_n keskenään riippumattomat normaalijakaumaa noudattavat satunnaismuuttujat, joiden odotusarvot ovat μ_1, μ_2, ..., μ_n ja varianssit σ_1^2, σ_2^2, ..., σ_n^2. Silloin lineaarikombinaatio*

$$Y = \sum_{i=1}^{n} a_i X_i$$

noudattaa normaalijakaumaa

$$\mathrm{N}\Big(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\Big).$$

Todistus. Tulos saadaan soveltamalla Lausetta 6.4 normaalijakaumaan. □

6.4 Järjestyssuureet

Otoksen suurin ja pienin arvo sekä keskimääräinen arvo, mediaani, ovat tärkeitä otossuureiden arvojen järjestykseen perustuvia tunnuslukuja. Olkoon $X_1, X_2, \ldots, X_n$ otos. Merkitään otoksen pienintä arvoa $X_{(1)}$ seuraavaksi pienintä $X_{(2)}$ ja niin edelleen, joten

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}.$$

Tämä indeksointi tarkoittaa sitä, että otosarvot pannaan kasvavaan järjestykseen. Jos otos on esimerkiksi 5.0, 3.1, 2.7, 6.1, 5.3, niin järjestetty otos on 2.7, 3.1, 5.0, 5.3, 6.1. Nyt siis esimerkiksi $X_1 = 5.0$, $X_{(1)} = 2.7$ ja $X_{(3)} = 5.0$ on mediaani ja $X_3 = 2.7$. Nyt siis

$$X_{(1)} = \min(X_1, \ldots, X_n)$$

ja

$$X_{(n)} = \max(X_1, \ldots, X_n).$$

Tunnusluku $X_{(k)}$ on otoksen k. järjestystunnusluku.

6.4.1 Maksimi ja minimi

Olkoon $X_1, X_2, \ldots, X_n$ otos jakaumasta, jonka kertymäfunktio on $F(x)$. Maksimin kertymäfunktio on

$$\begin{aligned} F_{(n)}(x) &= P(X_{(n)} \le x) \\ &= P(X_1 \le x,\ X_2 \le x,\ \ldots,\ X_n \le x) \\ &= P(X_1 \le x)P(X_2 \le x)\cdots P(X_n \le x), \end{aligned}$$

koska X_1, X_2, $\ldots$, X_n ovat riippumattomat. Kertymäfunktion määritelmän mukaan $P(X_i \le x) = F(x)$, joten

$$F_{(n)}(x) = [F(x)]^n.$$

Minimin kertymäfunktio on

$$\begin{aligned} F_{(1)}(x) &= P(X_{(1)} \le x) \\ &= 1 - P(X_{(1)} > x) \\ &= 1 - P(X_1 > x,\ X_2 > x,\ \ldots,\ X_n > x) \\ &= 1 - P(X_1 > x)P(X_2 > x)\cdots P(X_n > x) \\ &= 1 - [1 - F(x)]^n. \end{aligned}$$

Esimerkki 6.7 Olkoon $X_1, X_2, \ldots, X_n$ otos eksponenttijakaumasta $\mathrm{Exp}(\lambda)$. Määritetään minimin $X_{(1)}$ jakauma. Eksponenttijakauman $\mathrm{Exp}(\lambda)$ kertymäfunktio on

$$F(x) = \begin{cases} 0, & \text{kun } x < 0; \\ 1 - \mathrm{e}^{\lambda x}, & \text{kun } x \ge 0. \end{cases}$$

Silloin minimin kertymäfunktio on

$$F_{(1)}(x) = \begin{cases} 0, & \text{kun } x < 0; \\ 1 - \mathrm{e}^{-n\lambda x}, & \text{kun } x \geq 0. \end{cases}$$

Minimi noudattaa siis eksponenttijakaumaa $\mathrm{Exp}(n\lambda)$. □

Jos otos on jatkuvasta jakaumasta, jonka tiheysfunktio on $f(x)$, saadaan $X_{(1)}$:n ja $X_{(n)}$:n jakaumien tiheysfunktiot derivoimalla kertymäfunktiot $F_{(n)}(x)$ ja $F_{(1)}(x)$. Nyt siis maksimin tiheysfunktio on

$$f_{(n)}(x) = \frac{\mathrm{d}}{\mathrm{d}x}[F(x)]^n = n[F(x)]^{n-1} f(x)$$

ja minimin tiheysfunktio on

$$f_{(1)}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\big(1 - [1 - F(x)]^n\big) = n[1 - F(x)]^{n-1} f(x).$$

6.4.2 Järjestyssuureen $X_{(k)}$ jakauma

Olkoon $X_1, X_2, \ldots, X_n$ otos jakaumasta, jonka kertymäfunktio on $F(x)$. Johdetaan nyt järjestystunnusluvun $X_{(k)}$, $1 < k < n$, jakauma. Jos $\{X_{(k)} \leq x\}$, niin silloin ainakin k otosarvoa on pienempiä tai korkeintaan yhtä suuria kuin x. Tarkastellaan yksinkertaisuuden vuoksi tapausta $n = 3$. Johdetaan mediaanin $X_{(2)}$ jakauma. Tapahtuma $\{X_{(2)} \leq x\}$ toteutuu täsmälleen silloin, kun $\{X_1 \leq x,\ X_2 \leq x\}$ tai $\{X_1 \leq x,\ X_3 \leq x\}$ tai $\{X_2 \leq x,\ X_3 \leq x\}$ tai $\{X_1 \leq x,\ X_2 \leq x,\ X_3 \leq x\}$. Koska

$$P(X_i \leq x,\ X_j \leq x) = [F(x)]^2[1 - F(x)], \qquad i \neq j$$

ja

$$P(X_1 \leq x,\ X_2 \leq x,\ X_3 \leq x) = [F(x)]^3,$$

niin

$$\begin{aligned} F_{(2)}(x) = P(X_{(2)} \leq x) &= 3[F(x)]^2[1 - F(x)] + [F(x)]^3 \\ &= \sum_{i=2}^{3} \binom{3}{i} [F(x)]^i [1 - F(x)]^{3-i}. \end{aligned} \tag{6.4.1}$$

Yleisessä tapauksessa vastaava kaava voidaan johtaa samalla periaatteella. Emme kuitenkaan käsittele yleisen kaavan johtoa sen tarkemmin, toteamme vain, että $X_{(k)}$:n kertymäfunktio on

$$F_{(k)}(x) = P(X_{(k)} \leq x) = \sum_{i=k}^{n} \binom{n}{i} [F(x)]^i [1 - F(x)]^{n-i}. \tag{6.4.2}$$

Jos otos on jatkuvasta jakaumasta, saadaan vastaava tiheysfunktio derivoimalla kertymäfunktio. Esitetään ensin $X_{(2)}$:n tiheysfunktio, kun $n = 3$. Kun kertymäfunktio (6.4.1) derivoidaan, saadaan

$$\begin{aligned} f_{(2)}(x) = F'_{(2)}(x) &= 3 \cdot 2F(x)f(x)[1 - F(x)] - 3[F(x)]^2 f(x) + 3[F(x)]^2 f(x) \\ &= 3!F(x)[1 - F(x)]f(x). \end{aligned}$$

Derivoimalla lauseke (6.4.2) saadaan satunnaismuuttujan $X_{(k)}$ tiheysfunktio yleisessä tapauksessa $(1 \le k \le n)$:

$$f_{(k)}(x) = \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1 - F(x)]^{n-k} f(x). \tag{6.4.3}$$

Esimerkki 6.8 Olkoon X_1, X_2, X_3, X_4, X_5 otos jakaumasta, jonka tiheysfunktio on $f(x) = 2x$, $0 < x < 1$. Jakauman kertymäfunktio on

$$F(x) = \begin{cases} 0, & x < 0; \\ x^2, & 0 \le x \le 1; \\ 1, & x > 1. \end{cases}$$

Silloin mediaanin tiheysfunktio on lausekkeen (6.4.3) nojalla

$$f_{(3)}(x) = \frac{5!}{2!2!} x^4 (1 - x^2)^2 \cdot 2x = 60x^5(1 - x^2)^2, \qquad 0 < x < 1.$$

Vastaavasti minimin tiheysfunktio on

$$f_{(1)}(x) = 10x(1 - x^2)^4, \qquad 0 < x < 1$$

ja maksimin tiheysfunktio on

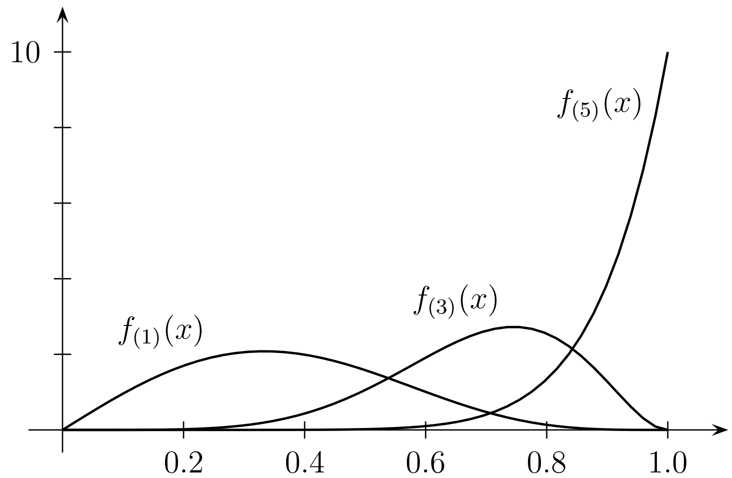
$$f_{(5)}(x) = 10x^9, \qquad 0 < x < 1.$$

□

6.5 Keskeinen rajaväittämä

Olemme havainneet, että otossuureen jakauma riippuu tavallisesti otoskoosta n. Jos $X_1, X_2, \ldots, X_n$ on otos Bernoullin jakaumasta $\mathrm{Ber}(p)$, niin $X = X_1 + X_2 + \cdots + X_n \sim \mathrm{Bin}(n, p)$. Satunnaismuuttujan X jakauma riippuu siis otoskoosta n. Jos $X_1, X_2, \ldots, X_n$ on otos normaalijakaumasta $\mathrm{N}(\mu, \sigma^2)$, niin otoskeskiarvon $\overline{X}$ jakauma $\mathrm{N}(\mu, \sigma^2/n)$ riippuu n:stä.

Olkoon $(X_i; i \ge 1) = X_1, X_2, X_3, \ldots$ satunnaismuuttujien jono, missä satunnaismuuttujien X_n, $n = 1, 2, 3, \ldots$ jakauma riippuu n:stä. Merkitään satunnaismuuttujan X_n kertymäfunktiota $F_n(x)$, joka siis riippuu n:stä. Seuraavassa määritellään satunnaismuuttujien jonon $(X_i; i \ge 1)$ *suppeneminen jakaumamielessä*.



Kuvio 6.1. Minimin, maksimin ja mediaanin tiheysfunktio, kun otos on jakaumasta $f(x) = 2x$, $0 < x < 1$.

Määritelmä 6.1 Satunnaismuuttujien X_1, X_2, $X_3, \ldots$ jono *suppenee jakaumaltaan* kohti satunnaismuuttujaa X, jos $\lim_{n\to\infty} F_n(x) = F(x)$ kaikissa pisteissä x, joissa $F(x)$ on jatkuva.

Kun jono $\{X_n\}$ suppenee jakaumaltaan kohti satunnaismuuttujaa X, merkitään $X_n \xrightarrow{d} X$. Momenttifunktioiden yhteydessä esitettiin momenttifunktion ja ja jakauman (kertymäfunktion) yksikäsitteistä vastaavuutta koskeva Lause 3.12. Samassa yhteydessä esitettiin myös momenttifunktioiden suppenemista koskeva Lause 3.15, jota voidaan soveltaa raja-jakaumien määrittämiseen.

Merkitään satunnaismuuttujien X_1, X_2, $\ldots$, X_n summaa ja keskiarvoa seuraavasti:

$$S_n = \sum_{i=1}^{n} X_i \qquad \text{ja} \qquad \overline{X}_n = \frac{S_n}{n}$$

Lause 6.10 (Keskeinen rajaväittämä) *Olkoon $X_1, X_2, \ldots, X_n$ otos jakaumasta, jonka keskiarvo on μ ja varianssi σ^2. Merkitään*

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{S_n - n\mu}{\sqrt{n}\sigma}.$$

Silloin Z_n:n jakauma lähenee normaalijakaumaa $\mathrm{N}(0,1)$*, kun $n \to \infty$.*

Keskeisen rajaväittämän mukaan riippumattomien satunnaismuuttujien summa noudattaa likimain normaalijakaumaa, kun n on suuri. Merkitsemme

$$Z_n \simeq \mathrm{N}(0,1),$$

kun n on suuri. Merkki $\simeq$ tarkoittaa ”noudattaa likimain jakaumaa”. Käytännössä keskeisen rajaväittämän avulla voidaan arvioida Z_n:n jakaumaa, kun

n on riittävän suuri. Silloin

$$P(Z_n \leq z) \approx \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}\, \mathrm{d}x = \Phi(z),$$

missä $\Phi(z)$ on normitetun normaalijakauman kertymäfunktio. Voimme merkitä saman asian myös seuraavasti:

$$P(Z_n \leq z) \to \Phi(z),$$

kun $n \to \infty$.

Esimerkki 6.9 Olkoon $X_1, X_2, \ldots, X_{15}$ otos jakaumasta, jonka tiheysfunktio on $f(x) = \left(\frac{3}{2}\right)x^2$, $-1 < x < 1$. Jakauman odotusarvo $\mu = 0$ ja varianssi $\sigma^2 = 3/5$. Esimerkiksi todennäköisyys $P(\overline{X} \leq 0.15)$ voidaan laskea johtamalla ensin $\overline{X}$:n jakauma ja määrittämällä siitä kysytty todennäköisyys. Keskeisen rajaväittämän avulla saadaan tämän todennäköisyyden tarkka arvio ilman tietoa $\overline{X}$:n tarkasta jakaumasta:

$$\begin{aligned} P(\overline{X} \leq 0.15) &= P\left(\frac{\overline{X} - 0}{\sqrt{3/5}\,/\sqrt{15}} \leq \frac{0.15 - 0}{\sqrt{3/5}\,/\sqrt{15}} \right) \\ &= P(Z_{15} \leq 0.75) \\ &\approx \Phi(0.75) = 0.7734. \end{aligned}$$

Arvion tarkkuudesta keskeinen rajaväittämä ei kuitenkaan anna käsitystä. □

6.6 Jakaumien likiarvot normaalijakauman avulla

Olkoon $X_1, X_2, \ldots, X_n$ otos Bernoullin jakaumasta $\mathrm{Ber}(p)$. Silloin $S_n = X_1 + X_2 + \cdots + X_n$ noudattaa binomijakaumaa $\mathrm{Bin}(n, p)$. Keskeisen rajaväittämän mukaan

$$Z_n = \frac{\overline{X}_n - p}{\sqrt{p(1-p)/n}} = \frac{S_n - np}{\sqrt{np(1-p)}}$$

noudattaa likimain normaalijakaumaa $\mathrm{N}(0, 1)$, kun n on suuri. Tuloksen mukaan binomijakauma lähenee normaalijakaumaa, kun n kasvaa. Peukalosääntönä voidaan pitää, että n on riittävän suuri, kun $np \geq 5$ ja $n(1-p) \geq 5$. Mitä enemmän p poikkeaa 0.5:stä, sitä suurempi n tarvitaan.

Esimerkki 6.10 Oletetaan, että $X \sim \mathrm{Bin}(10, 0.5)$. Lasketaan todennäköisyys $P(3 \leq X < 6)$. Voidaan kirjoittaa

$$P(3 \leq X < 6) = P(2.5 \leq X \leq 5.5).$$

Arvioidaan nyt jälkimmäistä todennäköisyyttä keskeisen rajaväittämän nojalla normaalijakauman avulla. Silloin

$$\begin{aligned} P(2.5 \leq X \leq 5.5) &= P\left(\frac{2.5-5}{\sqrt{10/4}} \leq \frac{X-5}{\sqrt{10/4}} \leq \frac{5.5-5}{\sqrt{10/4}}\right) \\ &\approx \Phi(0.316) - \Phi(-1.581) = 0.5670. \end{aligned}$$

Tarkka todennäköisyys binomijakauman avulla on $P(3 \leq X < 6) = 0.5683$. □

6.7 t-jakauma ja F-jakauma

Oletetaan, että $X_1, X_2, \ldots, X_n$ on otos jakaumasta $\mathrm{N}(\mu, \sigma^2)$, jonka varianssi σ^2 tunnetaan. Tarkastellaan lauseketta

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}},$$

joka tunnetaan t-testisuureena. Tiedämme, että

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$$

ja Lauseen 6.8 mukaan

$$U = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

Lisäksi Lauseen 6.8 mukaan Z ja U ovat riippumattomat. Tällainen satunnaismuuttuja noudattaa t-jakaumaa vapausastein $r = n-1$. Alaluvussa 5.9.2 esitettiin t-jakauman tiheysfunktio.

Usein halutaan verrata kahden normaalijakauman $\mathrm{N}(\mu_1, \sigma_1^2)$ ja $\mathrm{N}(\mu_2, \sigma_2^2)$ varriansseja. Teemme n_1:n kokoisen otoksen jakaumasta $\mathrm{N}(\mu_1, \sigma_1^2)$ ja n_2:n kokoisen otoksen jakaumasta $\mathrm{N}(\mu_2, \sigma_2^2)$. Oletetaan, että otokset ovat toisistaan riippumattomat. Olkoot S_1^2 ja S_2^2 näistä eri otoksista lasketut otosvarianssit. Lauseen 6.8 mukaan

$$U = (n_1 - 1)\frac{S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1) \qquad \text{ja} \qquad V = (n_2 - 1)\frac{S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1).$$

Koska otokset ovat keskenään riippumattomat, niin satunnaismuuttujat U ja V ovat riippumattomat. Varianssien yhtäsuuruutta voidaan testatata tarkastelemalla suhdetta

$$F = \frac{U/r_1}{V/r_2}, \tag{6.7.1}$$

missä $r_1 = n_1 - 1$ ja $r_2 = n_2 - 1$. Alaluvussa 5.9.2 osoitettiin, että suhde (6.7.1) noudattaa F-jakaumaa vapausastein r_1 ja r_2.

6.8 Momenttifunktion rajafunktiot

Tarkastelemme nyt satunnaismuuttujan (usein otoksen tunnusluku) jakauman riippuvuutta otoskoosta n. Otoskeskiarvo ja otosvarianssi ovat tavallisimmat otoksesta lasketut tunnusluvut. Oletetaan esimerkiksi, että $X_n \sim \mathrm{Bin}(n, p)$. Eri n:n arvoilla saamme eri binomijakauman. Miten jakauma muuttuu n:n kasvaessa? Olemme keskeisen rajaväittämän avulla jo osoittaneet, että $\mathrm{Bin}(n, p)$ lähenee normaalijakaumaa, kun n kasvaa.

Voimme tutkia $\mathrm{Bin}(n, p)$:n rajajakaumaa myös ehdolla, että jakauman odotusarvo np pidetään vakiona λ. Jos $np = \lambda$ on vakio ja $n \to \infty$, niin $p \to 0$. Satunnaismuuttujan $X_n \sim \mathrm{Bin}(n, p)$ momenttifunktio on

$$M_n(t) = (1 - p + p\mathrm{e}^t)^n.$$

Koska $p = \lambda/n$, niin

$$M_n(t) = \left[1 - \frac{\lambda}{n} + \frac{\lambda}{n}\mathrm{e}^t\right]^n = \left[1 + \frac{\lambda(\mathrm{e}^t - 1)}{n}\right]^n.$$

Käyttäen hyväksi analyysin tulosta

$$\lim_{n\to\infty}\left(1 + \frac{a}{n}\right)^n = \mathrm{e}^a,$$

saadaan

$$\lim_{n\to\infty} M_n(t) = \mathrm{e}^{\lambda(\mathrm{e}^t-1)} = M(t),$$

joka on olemassa kaikilla $t \in \mathbb{R}$. Koska

$$M(t) = \mathrm{e}^{\lambda(\mathrm{e}^t-1)}$$

on Poissonin jakauman $\mathrm{Poi}(\lambda)$ momenttifunktio, niin Lauseen 3.15 mukaan X_n:n jakauma lähestyy siis Poissonin jakaumaa $\mathrm{Poi}(\lambda)$, kun $n \to \infty$.

6.9 Suppenemiskäsitteet

Olemme edellä jo useaan otteeseen tutustuneet suppenemiseen jakaumamielessä. Käsite määriteltiin alaluvussa 4.3.4 (Määritelmä 4.2). Satunnaismuuttujien jono $(X_n, n \geq 1) = (X_1, X_2, \ldots)$ suppenee jakaumaltaan kohti satunnaismuuttujaa X ($X_n \xrightarrow{d} X$, kun $n \to \infty$), jos

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

kaikssa pisteissä x, joissa $F_X(x)$ on jatkuva. Tässä yhteydessä on myös syytä muistaa, että momenttifunktioiden jonon suppenemisesta seuraa vastaavien jakaumien suppeneminen jakaumamielessä.

Esimerkki 6.11 Olkoon $\{X_n\}$ sellainen satunnaismuuttujien jono, että

$$p_n(x) = P(X_n = x) = \begin{cases} 1, & \text{kun } x = 2 + \frac{1}{n}; \\ 0, & \text{kun } x \neq 2 + \frac{1}{n}. \end{cases}$$

Huomaa, että $p_n(2) = 0$ kaikilla n. Tästä seuraa, että $p_n(x) \to p(x)$, missä $p(x) = 0$ kaikilla x. Satunnaismuuttujan X_n kertymäfunktio on muotoa

$$F_n(x) = \begin{cases} 0, & \text{kun } x < 2 + \frac{1}{n}; \\ 1, & \text{kun } x \geq 2 + \frac{1}{n}. \end{cases}$$

Kun $n \to \infty$, niin $F_n(x) \to F(x)$, missä

$$F(x) = \begin{cases} 0, & x < 2; \\ 1, & x \geq 2. \end{cases}$$

$F(x)$ on pisteeseen $x = 2$ degeneroituneen jakauman kertymäfunktio eli $P(X = 2) = 1$. Todennäköisyysfunktioiden $p_n(x)$ jono ei kuitenkaan suppene kohti tämän jakauman todennäköisyysfunktiota. □

Olkoon $\{X_n\}$ jono satunnaismuuttujia, joiden odotusarvo on μ ja varianssi σ^2. Silloin keskeisen rajaväittämän mukaan

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z,$$

missä $Z \sim \mathrm{N}(0, 1)$. Huomattakoon, että Z_n:n jakaumat ovat usein diskreettejä, mutta silti rajajakauma on normaalijakauma. Kun n on riittävän suuri, niin

$$P\left(a \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) \approx \Phi(b) - \Phi(a).$$

Jos esimerkiksi $X_n \sim \mathrm{Bin}(n, p)$, niin silloin keskeisen rajaväittämän mukaan

$$\frac{\sqrt{n}(\overline{X}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{d} Z,$$

missä $Z \sim \mathrm{N}(0, 1)$. Tätä tulosta kutsutaan *De Moivren ja Laplacen* lauseeksi.

Osoitimme alavuvussa 3.4 Tšebyševin epäyhtälön avulla, että otoskeskiarvo $\overline{X}$ on hyvä populaation keskiarvon tunnusluku. Tarkastelu ei perustunut suppenemiseen jakaumamielessä vaan ns. stokastiseen suppenemiseen.

Määritelmä 6.2 Satunnaismuuttujien jono $\{X_n\}$ suppenee stokastisesti kohti satunnaismuuttujaa X, jos kaikilla $\varepsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$$

tai yhtäpitävästi

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1.$$

Stokastista suppenemista sanotaan myös suppenemiseksi todennäköisyyden mielessä ja merkitään $X_n \xrightarrow{P} X$. Usein tarkastellaan tilannetta, että satunnaismuuttuja, jota lähestytään, on vakio. Tällainen tilanne on heikossa suurten lukujen laissa (Lause 3.11, HSLL). Esitetty heikon suurten lukujan lain todistus oli sillä tavalla yleinen, että se on pätevä myös jatkuville satunnaismuuttujille. HSLL sanoo, että otoskeskiarvo suppenee stokastisesti kohti populaation keskiarvoa, kun otoskoko kasvaa.

Olkoon $\{X_n\}$ sellaisten satunnaismuuttujien jono, että $E(X_n) = \mu$ ja $\mathrm{Var}(X_n) = \sigma^2$. Heikon suurten lukujen lain mukaan

$$\overline{X}_n \xrightarrow{P} \mu,$$

missä $\overline{X}_n = (X_1 + X_2 + \cdots + X_n)/n$. Lause todistettiin Tšebyševin epäyhtälön avulla.

Esimerkki 6.12 Olkoon $\{X_n\}$ jono sellaisia diskreettejä satunnaismuuttujia, että

$$P(X_n = 1) = \frac{1}{n} \qquad \text{ja} \qquad P(X_n = 0) = 1 - \frac{1}{n}.$$

Silloin

$$P(|X_n| > \varepsilon) = \begin{cases} P(X_n = 1) = \dfrac{1}{n}, & \text{kun } 0 < \varepsilon < 1; \\ 0, & \text{kun } \varepsilon \geq 1. \end{cases}$$

Tästä nähdään, että $P(|X_n| > \varepsilon) \to 0$, kun $n \to \infty$. Voimme siis sanoa, että $X_n \xrightarrow{P} 0$. □

Esimerkki 6.13 (Otosvarianssin tarkentuvuus) Olkoon $\{X_n\}$ sellainen satunnaismuuttujien jono, että $E(X_n) = \mu$ ja $\mathrm{Var}(X_n) = \sigma^2 < \infty$. Otosvarianssi on

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Tiedämme, että $E(S_n^2) = \sigma^2$. Tšebyševin epäyhtälön mukaan

$$P(|S_n^2 - \sigma^2| \geq \varepsilon) \leq \frac{E(S_n^2 - \sigma^2)}{\varepsilon^2} = \frac{\mathrm{Var}(S_n^2)}{\varepsilon^2}.$$

Jos nyt $\mathrm{Var}(S_n^2) \to 0$, kun $n \to \infty$, niin $\lim_{n\to\infty} P(|S_n^2 - \sigma^2| \geq \varepsilon) = 0$ ja $(S_n^2, n \geq 1)$ suppenee stokastisesti kohti populaation varianssia. □

Tässä yhteydessä on tietysti luonnollista kysyä, miten stokastinen suppeneminen ja suppeneminen jakaumamielessä suhteutuvat toisiinsa. Voidaan osoittaa, että stokastinen suppeneminen implikoi suppenemisen jakaumamielessä. Jos siis $X_n \xrightarrow{P} X$, niin $X_n \xrightarrow{d} X$. Jos jono $\{X_n\}$ suppenee kohti vakiota μ, niin silloin $X_n \xrightarrow{P} \mu$ jos ja vain jos $X_n \xrightarrow{d} \mu$.

Rajoitumme tässä esityksessä kahteen edellä esitettyyn suppenemiskäsitteeseen: stokastiseen suppenemiseen ja suppenemiseen jakaumamielessä. Esitämme kuitenkin vielä ns. melkein varman (m.v.) suppenemisen.

Määritelmä 6.3 Jono $\{X_n\}$ suppenee melkein varmasti kohti satunnaismuuttujaa X, jos

$$P\big(\lim_{n\to\infty} |X_n - X| < \varepsilon\big) = 1.$$

Näennäisesti määritelmä muistuttaa stokastisen suppenemisen määritelmää, vaikka käsitteet ovat sisällöllisesti erilaisia.

6.10 Estimaattorit

6.10.1 Estimaattoreiden ominaisuuksia

Olkoon $X_1, X_2, \ldots, X_n$ otos jakaumasta, jonka tiheysfunktio on $f(x, \theta)$. Jos haluamme estimoida jakauman tunnuslukua θ jollakin otoksen tunnusluvulla, merkitsemme usein tätä otoksen tunnuslukua $\hat{\theta}$. On siis muistettava, että $\hat{\theta}$ on otoksen funktio ja täydellisempi merkintä olisi $\hat{\theta}(X_1, X_2, \ldots, X_n)$. Havaitusta otoksesta $x_1, x_2, \ldots, x_n$ laskettua estimaattorin arvoa $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$ sanotaan estimaatiksi. Olemme edellä jo tarkastelleet useita estimaattoreita. Tavanomaisia odotusarvon μ ja varianssin σ^2 estimaattoreita ovat otoskeskiarvo $\overline{X}$ ja otosvarianssi S^2, eli $\hat{\mu} = \overline{X}$ ja $\hat{\sigma}^2 = S^2$.

Määritelmä 6.4 Estimaattori $\hat{\theta}$ on parametrin *harhaton estimaattori*, jos $E(\hat{\theta}) = \theta$ kaikilla θ:n arvoilla. Muutoin $\hat{\theta}$ on harhainen ja $\hat{\theta}$:n harha on $\text{harha}(\hat{\theta}) = E(\hat{\theta}) - \theta$.

Olemme jo aikaisemmin osoittaneet, että $\hat{\mu} = \overline{X}$ ja $\hat{\sigma}^2 = S^2$ ovat harhattomia estimaattoreita.

Eräs intuitiivisesti hyväksyttävä estimaattorille asetettava vaatimus on, että se antaa 'tarkempia' estimaatteja kun otoskoko kasvaa. Tarkan estimaattorin arvot osuvat suurella todennäköisyydellä lähelle parametrin θ oikeata arvoa. *Tarkentvuvuus* sisältää tämän ajatuksen.

Määritelmä 6.5 Tunnusluku $\hat{\theta}$ on parametrin θ *tarkentuva estimaattori*, jos $\hat{\theta} \xrightarrow{P} \theta$, kun otoskoko n kasvaa rajatta.

Selvempi olisi merkitä $\hat{\theta}_n = \hat{\theta}(X_1, X_2, \ldots, X_n)$, missä $(\hat{\theta}_n; n \geq 1)$ on satunnaismuuttujien jono. Jos $\hat{\theta}_n$ on θ:n tarkentuva estimaattori, niin jono $(\hat{\theta}_n; n \geq 1)$ suppenee stokastisesti kohti parametrin arvoa θ.

Määritelmä 6.6 Estimaattorin $\hat{\theta}$ keskineliövirhe (MSE = Mean Square Error) on

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2].$$

Määritelmästä seuraa suoraviivaisesti, että

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + [\text{harha}(\hat{\theta})]^2.$$

Voidaan osoittaa, että $\hat{\theta}$ on θ:n tarkentuva estimaattori, jos $\text{MSE}(\hat{\theta}) \to 0$ otoskoon n kasvaessa rajatta.

6.10.2 Delta-menetelmä

Määritelmä 6.7 Funktion $g(x)$ r. asteen *Taylorin polynomi* pisteessä a on

(6.10.1) $$T_r(x) = g(a) + g'(a)(x-a) + \frac{g''(a)}{2!}(x-a)^2 + \cdots + \frac{g^{(r)}(a)}{r!}(x-a)^r,$$

missä $g^{(r)}(x) = \frac{\mathrm{d}^r}{\mathrm{d}x^r} g(x)$ on funktion $g(x)$ r. derivaatta.

Taylorin lauseen mukaan

(6.10.2) $$\lim_{x \to a} \frac{g(x) - T_r(x)}{(x-a)^r} = 0,$$

jos $g^{(r)}(a)$ on olemassa. Funktio $g(x)$ voidaan lausua pisteen $x = a$ ympäristössä muodossa

$$g(x) = T_r(x) + R_{r+1}(x),$$

missä $R_{r+1}(x) = g(x) - T_r(x)$ on jäännöstermi, joka siis toteuttaa ehdon 6.10.2.

Oletetaan, että X on satunnaismuuttuja, jonka odotusarvo on $E(X) = \mu \neq 0$. Jos estimoidaan funktiota $g(\mu)$, niin sen Taylorin polynomiin perustuva 1. kertaluvun likiarvo pisteessä μ on

(6.10.3) $$g(X) = g(\mu) + g'(\mu)(X - \mu).$$

Jos käytetään $g(\mu)$:n estimaattorina funktiota $g(X)$, niin

$$E[g(X)] \approx g(\mu)$$

ja

$$\mathrm{Var}[g(X)] = [g'(\mu)]^2 \,\mathrm{Var}(X).$$

Esimerkki 6.14 Tarkastellaan odotusarvon $E(X) = \mu \neq 0$ funktion $g(\mu) = 1/\mu$ estimointia. Olkoon estimaattorina $1/X$. Silloin edellisen mukaan

$$E\left(\frac{1}{X}\right) \approx \frac{1}{\mu}$$

ja

$$\mathrm{Var}\left(\frac{1}{X}\right) \approx \left(\frac{1}{\mu}\right)^4 \mathrm{Var}(X).$$

□

Lause 6.11 (Delta-menetelmä) *Olkoon* $\{X_n\}$ *sellainen satunnaismuuttujien jono, että* $\sqrt{n}(X_n - \theta)$ *lähenee jakaumamielessä normaalijakaumaa* $\mathrm{N}(0, \sigma^2)$. *Oletetaan, että annetulla funktiolla* g *on määrätyllä arvolla* θ *derivaatta* $g'(\theta) \neq 0$. *Silloin*

$$\sqrt{n}\,[g(X_n) - g(\theta)] \to \mathrm{N}\big(0, \sigma^2 [g'(\theta)]^2\big)$$

jakaumamielessä.

Esimerkki 6.15 Olkoon $X_1, \ldots, X_n$ otos jakaumasta $\mathrm{Ber}(p)$. Onnistumisen todennäköisyyden p estimaattori on tavallisesti $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Onnistumisen mahdollisuus (odds) $p/(1-p)$ on vedonlyönnissä ja biostatiikassa tavanomainen parametri. Voimme käyttää $p/(1-p)$:n estimaattorina $\hat{p}$:n funktiota $\hat{p}/(1-\hat{p})$. Mitä voimme sanoa tämän estimaattorin ominaisuuksista? Nyt estimoidaan siis funktiota $g(p) = p/(1-p)$. Koska $g'(p) = 1/(1-p)^2$, niin lausekkeen 6.10.3 mukaan

$$\begin{aligned} \mathrm{Var}\left(\frac{\hat{p}}{1-\hat{p}}\right) &\approx [g'(p)]^2 \,\mathrm{Var}(\hat{p}) \\ &= \left[\frac{1}{(1-p)^2}\right]^2 \frac{p(1-p)}{n} = \frac{p}{n(1-p)^3}. \end{aligned}$$

□