Luku 5

Jatkuvat jakaumat

Sellaiset suureet kuten esimerkiksi aika, lämpötila, pituus ja paino ajatellaan tavallisesti jatkuviksi muuttujiksi, ts. muuttujiksi, jotka voivat saada mitä tahansa reaaliarvoja annetulla välillä. Esimerkiksi henkilön ikä on *jatkuva satunnaismuuttuja,* joka voi saada positiivisia reaalilukuarvoja. Diskreetin satunnaismuuttujan arvoavaruus on äärellinen tai numeroituva, mutta jatkuvan satunnaismuuttujan arvoavaruus on ylinumeroituva.

5.1 Jatkuvat satunnaismuuttujat

Jokaiseen satunnaismuuttujaan liittyy kertymäfunktio. Satunnaismuuttujan X kertymäfunktio määriteltiin alaluvussa 2.5.2 (Määritelmä 2.4) funktiona

$$F(x) = P(X \leq x), \qquad x \in \mathbb{R}.$$

Diskreetin satunnaismuuttujan kertymäfunktio on porrasfunktio, joka voidaan lausua hyppyfunktioiden summana (4.1.1) [ks. alaluku 4.1].

Lauseen 2.10 mukaan funktio $F(x)$ on kertymäfunktio jos ja vain jos seuraavat kolme ehtoa toteutuvat:

1. $\lim_{x \to -\infty} F(x) = 0$ ja $\lim_{x \to \infty} F(x) = 1$.
2. $F(x)$ on kasvava (ei-vähenevä) funktio.
3. $F(x)$ on oikealta jatkuva eli $\lim_{x \to x_0+} F(x) = F(x_0)$ kaikilla $x_0 \in \mathbb{R}$.

Jos meillä on jokin satunnaismuuttuja X, niin ominaisuudet 1. − 3. voidaan todeta todennäköisyysfunktion $P(X \leq x)$ ominaisuuksien avulla. Jos jokin funktio $F(x)$ toteuttaa ehdot 1. − 3., ei ole aivan helppoa todistaa, että $F(x)$ on todella jonkin satunnaismuuttujan kertymäfunktio. Todistus löytyy vaativista todennäköisyyslaskennan oppikirjoista.

Esimerkki 5.1 Funktio

$$F(x) = \frac{1}{1 + e^{-x}} \tag{5.1.1}$$

on esimerkki jatkuvasta kertymäfunktiosta, joka siis toteuttaa Lauseen 2.10 ehdot 1. – 3. Koska

$$\lim_{x\to-\infty} e^{-x} = \infty, \quad \text{niin} \quad \lim_{x\to-\infty} F(x) = 0$$

ja

$$\lim_{x\to\infty} F(x) = 1, \quad \text{koska} \quad \lim_{x\to\infty} e^{-x} = 0.$$

Funktio $F(x)$ on kasvava, koska sen 1. derivaatta

$$F'(x) = \frac{e^{-x}}{(1+e^{-x})^2} > 0.$$

On myös helppo todeta, että $F(x)$ ei ole ainoastaan oikealta jatkuva vaan *jatkuva*. □

Satunnaismuuttujan jatkuvuus voidaankin määritellä siihen liittyvän kertymäfunktion jatkuvuuden avulla.

Määritelmä 5.1 Satunnaismuuttuja X on jatkuva, jos sen kertymäfunktio $F_X(x)$ on x:n jatkuva funktio. Satunnaismuuttuja X on diskreetti, jos sen kertymäfunktio on x:n porrasfunktio.

Vastaavalla tavalla kuin diskreetin satunnaismuuttujan kertymäfunktio voidaan lausua summana, voidaan jatkuvan satunnaismuuttujan kertymäfunktio lausua integraalina:

$$P(X \le x) = F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t. \tag{5.1.2}$$

Jos $f_X(t)$ on jatkuva, niin integraalilaskennan peruslauseen mukaan

$$F'_X(x) = f_X(x), \tag{5.1.3}$$

missä $F'_X(x)$ on kertymäfunktion $F_X(x)$ derivaatta.

Määritelmä 5.2 Jatkuvan satunnaismuuttujan X tiheysfunktio $f_X(x)$ on funktio, joka toteuttaa yhtälön

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t \quad \text{kaikilla} \quad x \in \mathbb{R}. \tag{5.1.4}$$

Esimerkki 5.2 Olkoon X tiettyyn palvelunumeroon tulevien puheluiden pituus. Oletetaan, että X:n tiheysfunktio on

$$f(x) = \frac{1}{20}\mathrm{e}^{-x/20}, \qquad 0 \le x < \infty.$$

Silloin X noudattaa ns. *eksponenttijakaumaa* keskiarvolla 20. Nyt

$$S = \{\, x \mid 0 \le x < \infty \,\} \quad \text{ja} \quad f(x) > 0 \quad \text{kun} \quad x \in S.$$

Kertymäfunktio on

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} \frac{1}{20} \mathrm{e}^{-t/20} \, \mathrm{d}x = \int_{0}^{x} \frac{1}{20} \mathrm{e}^{-t/20} \, \mathrm{d}x \\ &= \Big/_{\!0}^{\,x} -\mathrm{e}^{-t/20} = 1 - \mathrm{e}^{-x/20}. \end{aligned}$$

Silloin

$$F'(x) = \frac{\mathrm{d}}{\mathrm{d}x}\big(1 - \mathrm{e}^{-x/20}\big) = \frac{1}{20}\mathrm{e}^{-x/20} = f(x), \qquad x \ge 0$$

ja $f(x) = 0$, kun $x < 0$. □

Huomaa, että yksittäisen pisteen $a \in \mathbb{R}$ todennäköisyys $P(X = a)$ on aina nolla, jos X on jatkuva satunnaismuuttuja. Silloin erityisesti kaikilla reaaliluvuilla $b > a$

$$\begin{aligned} F(b) - F(a) &= P(a \le X \le b) = P(a < X \le b) \\ &= P(a \le X < b) = P(a < X < b). \end{aligned}$$

Esimerkki 5.3 Olkoon X jatkuva satunnaismuuttuja, jonka tiheysfunktio on $f(x) = 2x$, kun $0 < x < 1$. Silloin X:n kertymäfunktio on

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

Huomaa, että

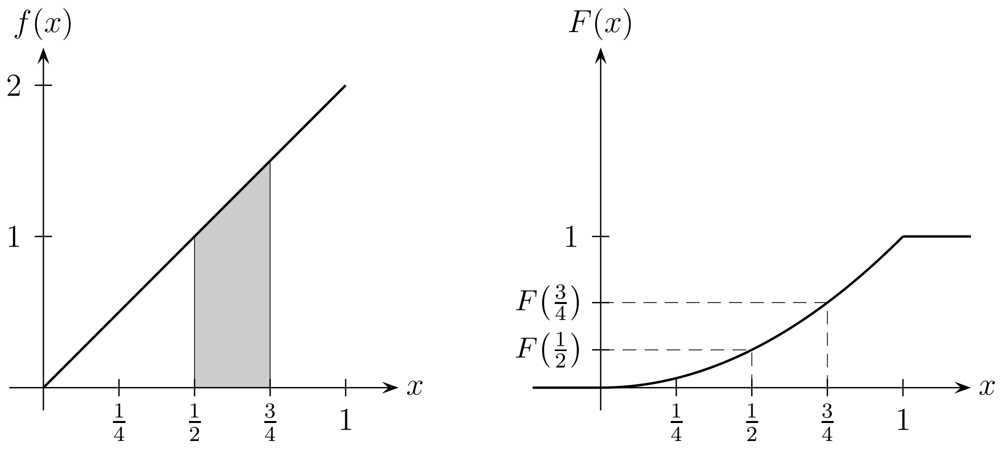
$$F(x) = \int_{0}^{x} 2t \, \mathrm{d}t = x^2, \qquad \text{kun } 0 \le x < 1.$$

Jos kertymäfunktio on annettu, niin tiheysfunktio saadaan derivoimalla kertymäfunktio:

$$F'(x) = \frac{\mathrm{d}}{\mathrm{d}x} x^2 = 2x, \qquad 0 \le x < 1.$$

Kertymäfunktion avulla voidaan laskea todennäköisyyksiä. Esimerkiksi todennäköisyys

$$P\left(\frac{1}{2} < X \le \frac{3}{4}\right) = F\left(\frac{3}{4}\right) - F\left(\frac{1}{2}\right) = \left(\frac{3}{4}\right)^2 - \left(\frac{1}{2}\right)^2 = \frac{5}{16}$$



Kuvio 5.1. Jatkuvan satunnaismuuttujan X tiheysfunktio $f(x) = 2x$ ja kertymäfunktio $F(x) = x^2$.

ja

$$P\left(\frac{3}{4} < X \leq \frac{3}{2}\right) = F\left(\frac{3}{2}\right) - F\left(\frac{3}{4}\right) = 1 - \left(\frac{3}{4}\right)^2 = \frac{7}{16}.$$

Toisaalta tietysti $P\left(\frac{1}{2} \leq X \leq \frac{3}{4}\right)$ voidaan laskea suoran $y = 2x$ ja x-akselin väliin jäävänä pinta-alana:

$$P\left(\frac{1}{2} \leq X \leq \frac{3}{4}\right) = \int_{1/2}^{3/4} 2x \,\mathrm{d}x = \frac{5}{15},$$

joka tietysti voidaan esittää kertymäfunktion avulla. □

Jatkuvan satunnaismuuttujan momentit määritellään vastaavasti kuin diskreetin satunnaismuuttujan tapauksessa, mutta määritelmässä summa korvataan integraalilla. Jatkuvan satunnaismuuttujan *r. momentti* on

$$\alpha_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x) \,\mathrm{d}x,$$

missä $f(x)$ on X:n tiheysfunktio. Satunnaismuuttujan X *r. keskusmomentti* on

$$\mu_r = E[(X - \mu)^r],$$

missä $\mu = E(X) = \alpha_1$ on X:n *odotusarvo.* Satunnaismuuttujan X odotusarvo on siis integraali

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) \,\mathrm{d}x$$

ja X:n *varianssi* σ^2 on 2. keskusmomentti

$$\begin{aligned}\sigma^2 = \mu_2 &= E[(X-\mu)^2] \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,\mathrm{d}x.\end{aligned}$$

Merkitsemme myös $E\big[(X-\mu)^2\big] = \mathrm{Var}(X)$, jolloin X:n *hajonta* on

$$\sigma = \sqrt{\mathrm{Var}(X)}.$$

Momenttifunktio on

$$M(t) = E(\mathrm{e}^{tx}) = \int_{-\infty}^{\infty} \mathrm{e}^{tx} f(x)\,\mathrm{d}x, \tag{5.1.5}$$

jos integraali 5.1.5 on olemassa jollakin avoimella välillä $(-a, a)$, missä $a > 0$. Tietysti esimerkiksi tulokset

$$\begin{aligned}\sigma^2 &= E(X^2) - \mu^2, \\ \mu &= M'(0), \\ \alpha_2 &= E(X^2) = M''(0)\end{aligned}$$

pitävät edelleen paikkansa samalla tavalla kuin diskreettien satunnaismuuttujien tapauksessa.

Esimerkki 5.4 Lasketaan nyt Esimerkissä 5.3 määritellyn satunnaismuuttujan X odotusarvo ja varianssi:

$$\mu = E(X) = \int_0^1 x(2x)\,\mathrm{d}x = \frac{2}{3} \Big/_0^1 x^3 = \frac{2}{3}$$

ja

$$\begin{aligned}\sigma^2 &= E(X^2) - \mu^2 \\ &= \int_0^1 x^2(2x)\,\mathrm{d}x - \left(\frac{2}{3}\right)^2 = \frac{1}{2} \Big/_0^1 x^4 - \frac{4}{9} = \frac{1}{18}.\end{aligned}$$

Kolmas momentti on

$$\alpha_3 = E(X^3) = \int_0^1 x^3(2x)\,\mathrm{d}x = \frac{2}{5} \Big/_0^1 x^5 = \frac{2}{5}$$

ja 3. keskusmomentti on

$$\begin{aligned}
\mu_3 = E\big[(X-\mu)^3\big] &= \int_0^1 (x-\mu)^3(2x)\,\mathrm{d}x \\
&= \int_0^1 (x^3 - 3\mu x^2 + 3\mu^2 x - \mu^3)(2x)\,\mathrm{d}x \\
&= \int_0^1 x^3(2x)\,\mathrm{d}x - 3\mu\int_0^1 x^2(2x)\,\mathrm{d}x + 3\mu^2\int_0^1 x(2x)\,\mathrm{d}x - \mu^3\int_0^1 2x\,\mathrm{d}x \\
&= \alpha_3 - 3\mu\alpha_2 + 3\mu^3 - \mu^3 = \alpha_3 - 3\mu\alpha_2 + 2\mu^3 \\
&= \frac{2}{5} - 3\cdot\frac{2}{3}\cdot\frac{1}{2} + 2\cdot\left(\frac{2}{3}\right)^3 = -\frac{3}{5} + \frac{16}{27} = \frac{1}{15}.
\end{aligned}$$

□

Myös prosenttipisteet ovat tärkeitä jakauman tunnuslukuja. Jakauman $100p$-prosenttipiste π_p määritellään seuraavasti:

$$p = \int_{-\infty}^{\pi_p} f(x)\,\mathrm{d}x = F(\pi_p), \qquad 0 \le p \le 1.$$

Prosenttipistettä $\pi_{0.50}$ kutsutaan *mediaaniksi* ja pistettä $\pi_{0.25}$ ja $\pi_{0.75}$ *alakvartiiliksi* ja *yläkvartiiliksi.* Esimerkissä 5.3 käsitellyn jakauman 36 %:n piste on 0.6, koska

$$F(\pi_{0.36}) = \pi_{0.36}^2 = 0.6^2 = 0.36.$$

Esimerkki 5.5 Olkoon satunnaismuuttujan X kertymäfunktio määritelty seuraavasti

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

Tarkistamme ensin, että F on todella kertymäfunktio. Toteamme helposti, että

1) $\lim_{x\to-\infty} F(x) = 0$ ja $\lim_{x\to\infty} F(x) = 1$,

2) $F(x)$ on x:n kasvava (ei-vähenevä) funktio ja

3) $F(x)$ on oikealta jatkuva, koska se on jatkuva.

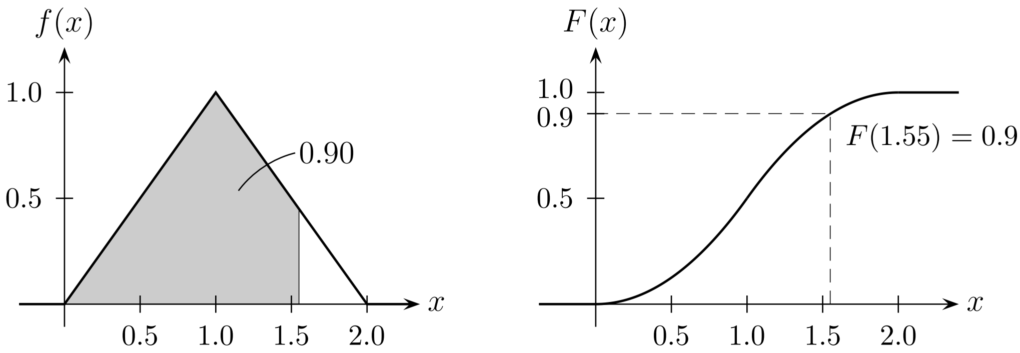
Tiheysfunktio saadaan derivoimalla $F(x)$. Nyt siis $F'(x) = x$ välillä $0 < x \leq 1$ ja $F'(x) = 2 - x$ välillä $1 \leq x \leq 2$. Näin siis tiheysfunktio on

$$f(x) = \begin{cases} x, & 0 < x \leq 1; \\ 2 - x, & 1 \leq x \leq 2; \\ 0 & \text{muualla.} \end{cases}$$

Tiheysfunktio voidaan kirjoittaa lyhyesti muodossa

$$f(x) = 1 - |x - 1|, \qquad 0 \leq x \leq 2.$$

Koska X:n tiheysfunktion kuvaaja on kolmion muotoinen, X:n jakaumaa kutsutaan kolmiojakaumaksi.



Kuvio 5.2. Kolmiojakauman tiheysfunktion ja kertymäfunktion kuvaajat.

Kolmiojakauman odotusarvo on

$$\begin{aligned} \mu = \int_0^2 x f(x)\,\mathrm{d}x &= \int_0^1 x \cdot x\,\mathrm{d}x + \int_1^2 x(2 - x)\,\mathrm{d}x \\ &= \Big/_{\!0}^{1} \frac{x^3}{3} + \Big/_{\!1}^{2} \left(x^2 - \frac{x^3}{3}\right) = \frac{1}{3} + \left(4 - \frac{8}{3}\right) - \left(1 - \frac{1}{3}\right) \\ &= 1. \end{aligned}$$

Koska jakauma on symmetrinen odotusarvon 1 suhteen, on 1 myös jakauman mediaani $\pi_{0.50}$. Se voidaan todeta helposti myös määritelmän perusteella, sillä $F(1) = \frac{1^2}{2} = 0.5$. Jakauman 90 %:n piste $\pi_{0.90}$ saadaan ratkaisemalla yhtälö

$$1 - \frac{(2 - \pi_{0.90})^2}{2} = 0.90.$$

Ratkaisu on $\pi_{0.90} = 2 - \sqrt{0.2} = 1.55$. □

Itse asiassa relaatio (5.1.4) ei välttämättä ole voimassa kaikilla x:n arvoilla, sillä $F(x)$ voi olla jatkuva, mutta ei derivoituva. Jos $f(x)$ on jatkuva, niin silloin tietysti yhtälö (5.1.4) pitää paikkansa. Huomattakoon, että jatkuvan satunnaismuuttujan tiheysfunktio ei välttämättä ole jatkuva, mutta kertymäfunktio on.

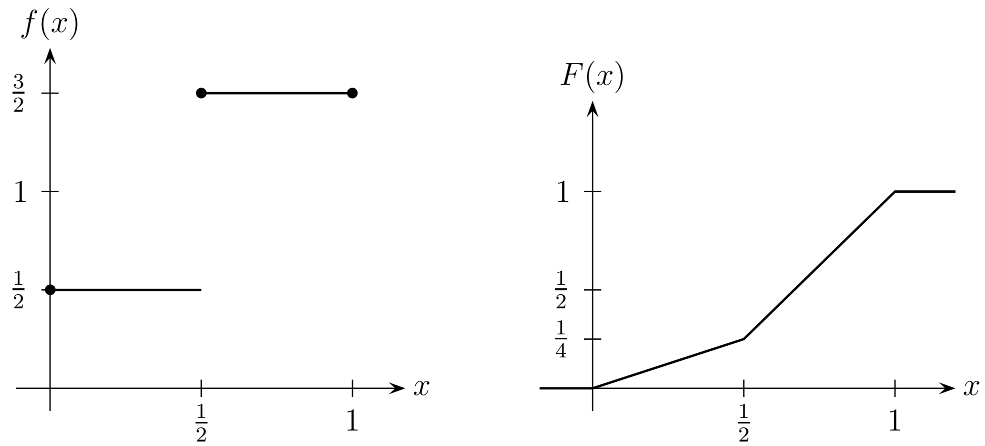
Esimerkki 5.6 Tarkastellaan nyt satunnaismuuttujaa X, jonka tiheysfunktio on

$$f(x) = \begin{cases} \frac{1}{2}, & 0 \leq x < \frac{1}{2}; \\ \frac{3}{2}, & \frac{1}{2} \leq x \leq 1. \end{cases}$$

Vastaavasti X:n kertymäfunktio on

$$F(x) = \begin{cases} 0, & x < 0; \\ \frac{1}{2}x, & 0 \leq x < \frac{1}{2}; \\ \frac{1}{4} + \frac{3}{2}\left(x - \frac{1}{2}\right), & \frac{1}{2} \leq x \leq 1; \\ 1, & 1 \leq x. \end{cases}$$

Havaitsemme nyt, että X:n tiheysfunktio ei ole jatkuva. Nyt myöskään F ei



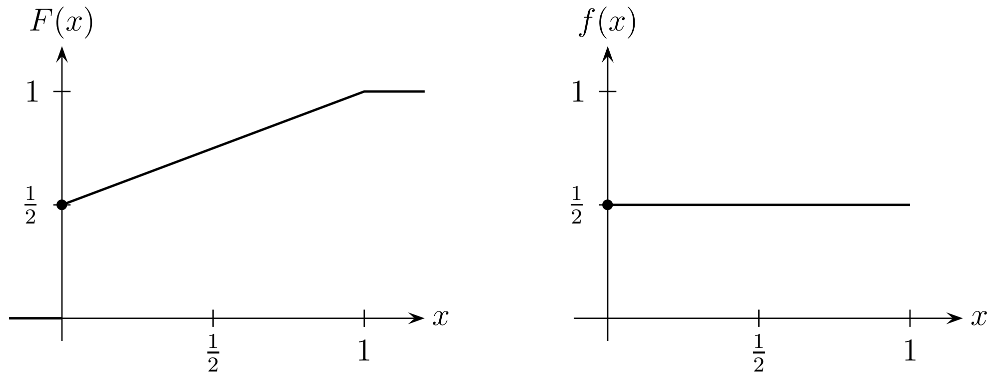
Kuvio 5.3. Satunnaismuuttujan X tiheysfunktion ja kertymäfunktion kuvaajat.

ole derivoituva pisteessä $\frac{1}{2}$. Pisteessä $x = \frac{1}{2}$ *ei ole voimassa,* että $F'(x) = f(x)$. Tässä on esimerkki jatkuvasta satunnaismuuttujasta, jonka tiheysfunktio ei ole jatkuva ja jonka kertymäfunktio ei ole koko määrittelyalueella S derivoituva. $\square$

Jatkuvan satunnaismuuttujan tiheysfunktiolla voi olla äärellinen määrä epäjatkuvuuspisteitä, mutta kertymäfunktio on jatkuva. Esimerkin 5.6 satunnaismuuttujan tiheysfunktiolla on määrittelyalueellaan yksi epäjatkuvuuspiste ja kertymäfunktio on jatkuva. Relaatio (5.1.3) pitää paikkansa vain tiheysfunktion jatkuvuuspisteissä, mutta ei epäjatkuvuuspisteissä.

Esimerkki 5.7 Määritellään satunnaismuuttuja X siten, että sen kertymäfunktio on

$$F(x) = \begin{cases} 0, & x < 0; \\ \frac{1}{2}, & x = 0; \\ \frac{1}{2} + \frac{x}{2}, & 0 < x < 1; \\ 1, & 1 \leq x. \end{cases} \tag{5.1.6}$$



Kuvio 5.4. Satunnaismuuttujan X kertymäfunktion ja 'tiheysfunktion' kuvaajat.

Kertymäfunktio ei ole nyt jatkuva, koska funktio hyppää pisteessä $x = 0$. Kertymäfunktio ei ole myöskään porrasfunktio. Nyt myös yksittäisellä pisteellä $X = 0$ on positiivinen todennäköisyys $P(X = 0) = \frac{1}{2}$, joten $f(x)$ ei ole tiheysfunktio. Itse asiassa kertymäfunktio (5.1.6) voidaan kirjoittaa porrasfunktion (kertymäfunktio) ja jatkuvan kertymäfunktion summana. Alaluvussa 4.1 määriteltiin hyppyfunktio $\varepsilon(x)$ siten, että $\varepsilon(x) = 1$ epänegatiivisilla x:n arvoilla ja $\varepsilon(x) = 0$, kun $x < 0$. Funktio $\varepsilon(x)$ on porrasfunktio ja siis diskreetin satunnaismuuttujan kertymäfunktio. Puoliavoimella välillä $(0, 1]$ tasajakaumaa noudattavan satunnaismuuttujan kertymäfunktio on

$$F_c(x) = \begin{cases} 0, & x \leq 0; \\ x, & 0 < x < 1; \\ 1, & 1 \leq x. \end{cases}$$

Nyt kertymäfunktio (5.1.6) voidaan kirjoittaa muodossa

$$F(x) = \frac{1}{2}\,\varepsilon(x) + \frac{1}{2}F_c(x).$$

Esimerkiksi todennäköisyys

$$\begin{aligned} P\left(X \leq \frac{1}{2}\right) &= \frac{1}{2}\,\varepsilon\left(\frac{1}{2}\right) + \frac{1}{2}F_c\left(\frac{1}{2}\right) \\ &= \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}. \end{aligned}$$

Satunnaismuuttuja X ei ole diskreetti eikä jatkuva. □

Yleisesti jatkuva satunnaismuuttuja voidaan määritellä identiteetin (5.1.4) avulla olettamatta tiheysfunktion $f(x)$ jatkuvuutta. Jos on olemassa sellainen epänegatiivinen funktio $f(x)$ [ts. $f(x) \geq 0$ kaikilla $x \in \mathbb{R}$], että (5.1.4) pitää paikkansa kaikilla $x \in \mathbb{R}$, niin kertymäfunktion $F(x)$ sanotaan olevan *absoluuttisesti jatkuva.* Absoluuttisesti jatkuva funktio on jatkuva. Kaikkien tässä luvussa käsiteltäviät jatkuvien satunnaismuuttujien kertymäfunktiot ovat absoluuttisesti jatkuvia.

5.2 Tasajakauma ja eksponenttijakauma

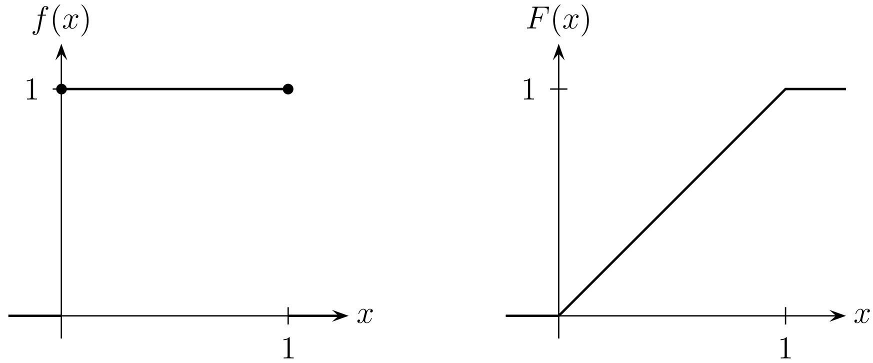
5.2.1 Tasajakauma

Jatkuva satunnaismuuttuja X noudattaa *tasajakaumaa* välillä $[0,1]$, jos sen tiheysfunktio on 1 tällä välillä ja 0 muualla:

$$f(x) = \begin{cases} 1, & \text{kun } x \in [0,1], \\ 0 & \text{muualla.} \end{cases} \tag{5.2.1}$$

Silloin merkitään $X \sim \mathrm{Tas}(0,1)$. On helppo todeta, että $f(x)$ on tiheysfunktio, koska $f(x) \geq 0$ ja

$$\int_0^1 f(x)\,\mathrm{d}x = \int_0^1 \mathrm{d}x = 1.$$



Kuvio 5.5. Tasajakauman $\mathrm{Tas}(0,1)$ tiheysfunktio ja kertymäfunktio.

Tasajakauman keskiarvo ja varianssi ovat:

$$E(X) = \int_0^1 x\,\mathrm{d}x = \frac{1}{2}$$

ja

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \int_0^1 x^2\,\mathrm{d}x - \frac{1}{4} = \frac{1}{12}.$$

Satunnaismuuttujan X momenttifunktio on

$$M_X(t) = \int_0^1 \mathrm{e}^{tx}\,\mathrm{d}x = \Big/_0^1 \frac{1}{t}\mathrm{e}^{tx} = \frac{\mathrm{e}^t - 1}{t}.$$

Huomaa, että $M_X(0) = 1$.

Olkoon $[a, b]$ annettu suljettu väli, $a < b$. Silloin satunnaismuuttuja $U = (b-a)X + a$ noudattaa tasajakaumaa välillä $[a, b]$. Silloin merkitään $U \sim \mathrm{Tas}(a, b)$. Koska $E(U) = (b-a)\,E(X) + a$ ja $\mathrm{Var}(U) = (b-a)^2\,\mathrm{Var}(X)$, niin

$$E(U) = \frac{a+b}{2} \qquad \text{ja} \qquad \mathrm{Var}(U) = \frac{(b-a)^2}{12}.$$

Satunnaismuuttujan U tiheysfunktio on

$$f(u) = \begin{cases} \frac{1}{b-a}, & \text{kun } u \in [a, b]; \\ 0 & \text{muualla} \end{cases} \tag{5.2.2}$$

ja U:n momenttifunktio on

$$M_U(t) = \begin{cases} \dfrac{\mathrm{e}^{tb} - \mathrm{e}^{ab}}{t(b-a)}, & t \neq 0; \\ 1, & t = 0. \end{cases}$$

5.2.2 Eksponenttijakauma

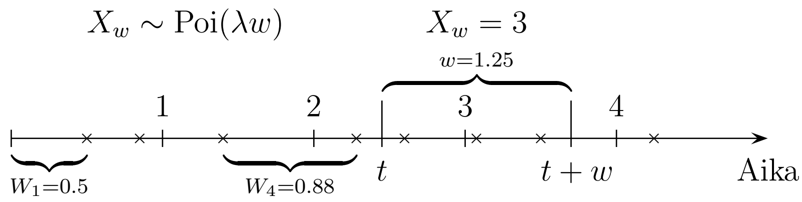
Poissonin prosessissa tarkastellaan, montako tapahtumaa (lisäystä) sattuu jollain aikavälillä. Merkitään w:n pituisella välillä sattuvien tapahtumien lukumäärää satunnaismuuttujalla X_w. Jos Poissonin prosessin intensiteetti on λ, niin Määritelmän 4.3 mukaan todennäköisyys, että w:n pituisella välillä sattuu x tapahtumaa, on

$$P(X_w = x) = \mathrm{e}^{-\lambda w}\frac{(\lambda w)^x}{x!}. \tag{5.2.3}$$

Poissonin prosessilla voidaan mallintaa esimerkiksi asiakkaiden saapumista palvelupisteeseen, puheluiden tuloa vaihteeseen, onnettomuuksien sattumista tarkasteltavalla tieosuudella tai autojen kulkua liikenteen tarkkailupisteen ohi. Tällöin ajatellaan, että yksittäiset tapahtumat sattuvat toisistaan riippumatta täysin satunnaisesti.

Tarkkaillaan nyt Poissonin prosessia, jonka intensiteetti on λ. Olkoon W odotusaika siihen hetkeen, kunnes seuraava tapahtuma sattuu. Odotusaika on jatkuva satunnaismuuttuja. Jos tarkkailemme prosessia hetkestä t hetkeen $t+w$ eli w:n pituisen aikavälin $[t, t+w]$, niin tapahtuma $\{\, W > w \,\}$ sattuu jos ja vain jos Poissonin prosessissa ei satu yhtään tapahtumaa välillä $[t, t+w]$. Siksi identiteetin (5.2.3) mukaan

$$P(W > w) = P(X_w = 0) = \mathrm{e}^{-\lambda w}.$$



Kuvio 5.6. Kaaviokuva esittää Poissonin saapumisprosessia, esimerkiksi autojen kulkemista liikenteen tarkkailupisteen ohi. Esimerkiksi W_1 on 1. auton odotusaika ja W_4 on 3. ja 4. auton välinen aika. Kiinnitetyllä w:n pituisella välillä on kulkenut ohi $X_w = 3$ autoa. Peräkkäiset odotusajat W_1, W_2, W_3, $\ldots$ ovat toisistaan riippumattomat ja noudattavat samaa jakaumaa.

Odotusajan W kertymäfunktio on siis

$$\begin{aligned} F(w) &= P(W \le w) \\ &= 1 - P(W > w) = 1 - P(X_w = 0) \\ &= 1 - \mathrm{e}^{-\lambda w}. \end{aligned}$$

Koska odotusaika W on epänegatiivinen, niin $F(w) = 0$, kun $w < 0$.

Odotusajan W tiheysfunktio on

$$F'(w) = f(w) = \lambda \mathrm{e}^{-\lambda w}$$

derivointisäännön (5.1.3) nojalla. Usein merkitään $\lambda = \frac{1}{\theta}$, missä $\theta > 0$. Sanomme, että W noudattaa *eksponenttijakaumaa* parametrilla θ ja merkitsemme $W \sim \mathrm{Exp}(\theta)$. Parametri θ on jakauman keskiarvo. Eksponenttijakauman tiheysfunktio on silloin muotoa

$$f(w) = \frac{1}{\theta}\mathrm{e}^{-w/\theta}. \tag{5.2.4}$$

Eksponenttijakauman $\mathrm{Exp}(\theta)$ momenttifunktio on

$$\begin{aligned} M(t) &= \int_0^\infty e^{tw}\frac{1}{\theta}\mathrm{e}^{-w/\theta}\,\mathrm{d}w = \Big/_0^\infty -\frac{\mathrm{e}^{-(1-\theta t)w/\theta}}{1-\theta t} \\ &= \frac{1}{1-\theta t}, \qquad t < \frac{1}{\theta}. \end{aligned}$$

Eksponenttijakaumalla on vastaava ”unohtamisominaisuus” kuin geometrisella jakaumalla. Jos $T \sim \mathrm{Exp}(\theta)$, niin

$$P(T > a + b \mid T > a) = P(T > b) \tag{5.2.5}$$

kaikilla epänegatiivisilla a ja b. Tulos voidaan todistaa laskemalla ehdollinen todennäköisyys

$$\begin{aligned} P(T > a + b \mid T > a) &= \frac{P(T > a,\ T > a + b)}{P(T > a)} = \frac{P(T > a + b)}{P(T > a)} \\ &= \frac{\mathrm{e}^{-(a+b)/\theta}}{\mathrm{e}^{-a/\theta}} = \mathrm{e}^{-b/\theta} = P(T > b). \end{aligned}$$

Huomattakoon, että edellä on käytetty tulosta

$$P(T > t) = 1 - P(T \le t) = 1 - F(t) = \mathrm{e}^{-t/\theta}, \qquad t \ge 0.$$

Esimerkki 5.8 Oletetaan, että asiakkaiden saapuminen liikkeeseen noudattaa Poissonin prosessia intensiteetillä 20 asiakasta tunnissa. Mikä on todennäköisyys, että myyjä joutuu odottamaan seuraavaa asiakasta yli 5 minuuttia? Olkoon X odotusaika, kunnes seuraava asiakas saapuu. Silloin prosessissa (5.2.3) $\lambda = 1/3$ asiakasta minuutissa ja $X \sim \mathrm{Exp}(3)$, koska eksponenttijakauman keskiarvo $\theta = 1/\lambda$. Jakauman Exp(3) tiheysfunktio on

$$f(x) = \frac{1}{3}\mathrm{e}^{-x/3}, \qquad 0 \le x < \infty$$

ja

$$P(X > 5) = \int_5^\infty \frac{1}{3}\mathrm{e}^{-x/3}\,\mathrm{d}x = \Big/_{\!5}^{\infty} -\mathrm{e}^{-x/3} = \mathrm{e}^{-5/3} \approx 0.1889.$$

Jatkuvan jakauman *mediaani* m on sellainen piste, että $F(m) = 1/2$. Nyt jakauman Exp(3) mediaanin m tulee toteuttaa ehto $F(m) = 1 - \mathrm{e}^{-m/3} = \frac{1}{2}$, joten

$$m = 3\log(2) \approx 2.0794.$$

□

5.2.3 Elinaikajakauma

Ominaisuuden (5.2.5) perusteella eksponenttijakauma on sopiva elinajan jakauma silloin, kun jäljellä oleva elinaika ei riipu tämänhetkisestä iästä. Olkoon T esimerkiksi jonkin elektronisen komponentin ikä tunteina. Silloin $P(T > b)$ on todennäköisyys, että uusi komponentti kestää ainakin b tuntia, kun taas $P(T > a + b \mid T > a)$ on todennäköisyys, että a tuntia käytössä ollut komponentti kestää vielä b tuntia. Jos elinaika noudattaa eksponenttijakaumaa, niin ominaisuuden (5.2.5) nojalla todennäköisyydet $P(T > b)$ ja $P(T > a + b \mid T > a)$ ovat samat kaikilla a ja b. Todennäköisyys, että komponentti rikkoontuu b:n seuraavan tunnin aikana, ei riipu lainkaan siitä, kuinka kauan komponentti on jo ollut käytössä.

Funktiota $G(t) = P(T > t)$ kutsutaan *eloonjäämisfunktioksi.* Eksponenttijakauma määrittelee eloonjäämisfunktion $G(t) = \mathrm{e}^{-t/\theta}$, jolla on unohtamisominaisuus

$$G(t + s) = G(t)G(s), \qquad t > 0, \quad s > 0. \tag{5.2.6}$$

Määritelmänsä nojalla $G(0) = 1$ ja $G(t) \to 0$, kun t kasvaa. Onko eksponenttifunktion lisäksi muita eloonjäämisfunktioita, joilla on unohtamisominaisuus (5.2.6)? Voidaan osoittaa, että ehdon (5.2.6) toteuttavat eloonjäämisfunktiot ovat aina muotoa $\mathrm{e}^{-\lambda t}$, $\lambda > 0$.

Jos elinaika T noudattaa eksponenttijakaumaa $\mathrm{Exp}(\theta)$, niin vakio $\lambda = \frac{1}{\theta}$ on hetkellinen *kuolleisuusaste* tai *vaaran aste.* Parametri λ säätelee todennäköisyyttä kuolla hetken $T = t$ jälkeisellä yksikön pituisella aikavälillä.

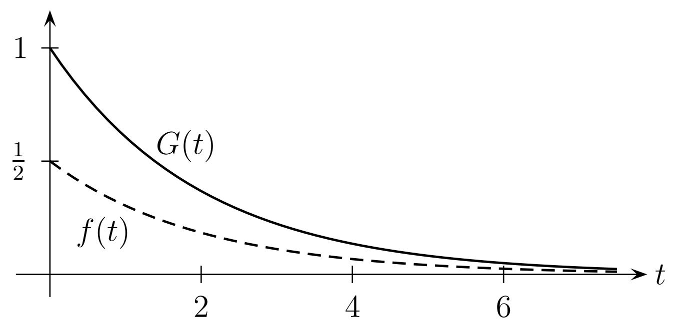
Olkoon Δ tarkasteltavan aikavälin pituus. Määritelläään todennäköisyys

$$
\begin{aligned}
P(T \leq t + \Delta \mid T > t) &= 1 - P(T > t + \Delta \mid T > t) \\
&= 1 - P(T > \Delta) = 1 - \mathrm{e}^{-\lambda\Delta},
\end{aligned}
$$

missä viimeistä edellinen yhtäsuuruus saadaan unohtamisominaisuuden (5.2.6) nojalla. Kun funktiota $\mathrm{e}^{-\lambda\Delta}$ arvioidaan Taylorin polynomin avulla, saadaan

$$
\begin{aligned}
1 - \mathrm{e}^{-\lambda\Delta} &= 1 - (1 - \lambda\Delta + \frac{1}{2}\lambda^2\Delta^2 - \cdots) \\
&= \lambda\Delta - \frac{1}{2}\lambda^2\Delta^2 + \cdots \\
&\approx \lambda\Delta, \qquad \text{kun } \Delta \text{ on pieni.}
\end{aligned}
$$

Arviointivirhe pienenee merkityksettömäksi verrattuna Δ:aan, kun $\Delta \to 0$. Silloin siis $P(T \leq t + \Delta \mid t > t) \approx \lambda\Delta$.



Kuvio 5.7. Eksponenttijakauman $\mathrm{Exp}(2)$ tiheysfunktio $f(t) = \frac{1}{2}\mathrm{e}^{-t/2}$ ja vastaava eloonjäämisfunktio $G(t) = \mathrm{e}^{-t/2}$.

Nyt nähdään, että

$$
\lim_{\Delta \to 0} \frac{P(T \leq t + \Delta \mid T > t)}{\Delta} = \lambda
$$

on riippumaton ajasta t. Eksponentiaalisesti jakautuneen elinajan tapauksessa kuolleisuusaste λ on iästä riippumaton vakio. Yleisesti kuolleisuusaste $\lambda(t)$ on tietysti iän funktio.

5.3 Gammajakauma ja χ^2-jakauma

Gammajakaumajakauma on välillä $[0, \infty)$ määritelty jakauma tai jakaumaperhe, koska parametrien vaihdellessa saadaan hyvinkin erinäköisiä jakaumia, vaikka ne ovat matemaattisesti samaa muotoa. Gammafunktio

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \mathrm{e}^{-x}\, \mathrm{d}x \tag{5.3.1}$$

määriteltiin jo Pykälässä 2.4.7. Jos $\alpha > 0$, niin $\Gamma(\alpha)$ on äärellinen. Jos α on positiivinen kokonaisluku, niin $\Gamma(\alpha)$ voidaan lausua suljetussa muodossa, muutoin ei.

Gammafunktio toteuttaa rekursiivisen relaation

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha),$$

joka voidaan osoittaa osittaisintegroinnilla. Jos $\alpha = n$ on positiivinen kokonaisluku, niin

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)\cdots 2 \cdot 1 \cdot \Gamma(1) = n!\Gamma(1).$$

Koska $\Gamma(1) = 1$, niin

$$\Gamma(n+1) = n!$$

kaikilla positiivisilla kokonaisluvuilla. Myös $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ on tärkeä erikoistapaus.

Funktio

$$f(t) = \frac{t^{\alpha-1}\mathrm{e}^{-t}}{\Gamma(\alpha)}, \qquad 0 < t < \infty \tag{5.3.2}$$

määrittelee tiheysfunktion, sillä gammafunktiossa integroitava on positiivinen välillä $(0, \infty)$. Sanokaamme, että (5.3.2) on satunnaismuuttujan T tiheysfunktio. Kaikkien gammajakaumien perhe saadaan määrittelemällä satunnaismuuttuja $X = \beta T$, missä β on positiivinen vakio. X:n tiheysfunktio voidaan johtaa soveltamalla Lauseen 5.5 muunnostekniikkaa. Merkitsemme $X \sim \mathrm{Gamma}(\alpha, \beta)$ ja sanomme, että X noudattaa gammajakaumaa parametrein α ja β. Jakauman $\mathrm{Gamma}(\alpha, \beta)$ tiheysfunktioksi saadaan

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}\mathrm{e}^{-x/\beta}, \qquad 0 < x < \infty, \quad \alpha > 0, \quad \beta > 0. \tag{5.3.3}$$

Esitämme nyt gammajakauman perusominaisuudet seuraavassa lauseessa.

Lause 5.1 *Oletetaan, että* $X \sim \mathrm{Gamma}(\alpha, \beta)$.

1. *Funktio (5.3.3) määrittelee tiheysfunktion kaikilla* $\alpha > 0$, $\beta > 0$.

2.
$$E(X) = \alpha\beta, \qquad \mathrm{Var}(X) = \alpha\beta^2$$

   ja

$$M(t) = E(\mathrm{e}^{tX}) = \left(\frac{1}{1-\beta t}\right)^\alpha, \qquad t < \frac{1}{\beta}.$$

3.

$$E(X^c) = \frac{\Gamma(\alpha + c)\beta^c}{\Gamma(\alpha)}$$

kaikilla $c > -\alpha$.

4. Olkoon $U = bX$, $b > 0$. *Silloin* $U \sim \mathrm{Gamma}(\alpha, b\beta)$.

Eksponettijakauma on gammajakauman erikoistapaus. Kun sijoitetaan tiheysfunktioon (5.3.3) $\alpha = 1$, saadaan

$$f(x;\beta) = \frac{1}{\beta}\mathrm{e}^{-x/\beta}, \qquad x > 0.$$

Havaitaan siis, että $\mathrm{Gamma}(1, \beta) = \mathrm{Exp}(\beta)$.

χ^2-jakauma

Toinen tärkeä gammajakauman erikoistapaus on χ^2-jakauma. Jos valitaan $\alpha = \frac{r}{2}$, missä r on positiivinen kokonaisluku, ja $\beta = 2$, tulee tiheysfunktio (5.3.3) muotoon

$$f(x) = \frac{1}{\Gamma\left(\frac{r}{2}\right)2^{r/2}} x^{(r/2)-1}\mathrm{e}^{-x/2}, \qquad 0 < x < \infty, \tag{5.3.4}$$

mikä on χ^2-jakauman tiheysfunktio vapausastein r. Jos X noudattaa χ^2-jakaumaa vapausastein r, merkitään $X \sim \mathrm{Khi2}(r)$. χ^2-jakauman keskiarvo, varianssi ja momenttifunktio saadaan nyt suoraan gammajakauman avulla. Jos $X \sim \mathrm{Khi2}(r)$, niin

$$E(X) = r, \qquad \mathrm{Var}(X) = 2r$$

ja

$$M(t) = (1 - 2t)^{-r/2}, \qquad t < \frac{1}{2}.$$

Odotusaika Poissonin prosessissa

Seuraavan tapahtuman odotusaika Poissonin prosessissa noudattaa eksponettijakaumaa. Olkoon W nyt odotusaika, kunnes sattuu α tapahtumaa, missä α on siis positiivinen kokonaisluku. Jos Poissonin prosessin intensiteetti on λ, niin todennäköisyys, että w:n pituisella aikavälillä sattuu x tapahtumaa, saadaan kaavalla (5.2.3):

$$P(X_w = x) = \mathrm{e}^{-\lambda w}\frac{(\lambda w)^x}{x!}.$$

Odotusajan W kertymäfunktio, kun $W \geq 0$, on

$$\begin{aligned} F(w) &= P(W \leq w) = 1 - P(W > w) \\ &= 1 - P(\text{vähemmän kuin } \alpha \text{ tapahtumaa välillä } [t, t+w]) \\ &= 1 - \sum_{x=0}^{\alpha-1} \mathrm{e}^{-\lambda w} \frac{(\lambda w)^x}{x!}, \end{aligned}$$

koska tapahtumien lukumäärä aikavälillä $[t, t+w]$ noudattaa Poissonin jakaumaa keskiarvolla λw [ks. (5.2.3)]. Laskemalla derivaatta $F'(w) = f(w)$ saadaan tiheysfunktio

$$f(w) = \frac{\lambda(\lambda w)^{\alpha-1}}{(\alpha-1)!} \mathrm{e}^{-\lambda w}.$$

Jos $w < 0$, niin $F(w) = 0$ ja $f(w) = 0$. Nyt huomaamme, että

$$W \sim \mathrm{Gamma}\Big(\alpha, \frac{1}{\lambda}\Big).$$

5.4 Normaalijakauma

5.4.1 Standardimuotoinen normaalijakauma

Tarkastelemme nyt todennäköisyysteorian ja tilastotieteen tärkeintä jakaumaa, normaalijakaumaa. Olkoon Z jatkuva satunnaismuuttuja, jonka tiheysfunktio on

$$(5.4.1) \qquad f(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2} \qquad -\infty < z < \infty.$$

Silloin Z noudattaa standardimuotoista normaalijakaumaa. Käytetään myös sanontaa "Z noudattaa standardoitua normaalijakaumaa".

Tarkistamme nyt, että (5.4.1) on todellakin tiheysfunktio. Koska $f(z) > 0$, pitää vain osoittaa, että

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-z^2/2} \, \mathrm{d}z = 1.$$

Osoitamme siis, että

$$(5.4.2) \qquad \int_{-\infty}^{\infty} \mathrm{e}^{-z^2/2} \, \mathrm{d}z = \sqrt{2\pi}.$$

Emme pysty suoraan integroimaan funktiota $\mathrm{e}^{-z^2/2}$, koska sen integraalifunktio ei ole lausuttavissa suljetussa muodossa. Osoittautuu kuitenkin, että integraalin (5.4.2) neliö on helppo laskea.

Integraalin arvo ei muutu, jos integrointimuuttuja nimetään uudelleen, joten

$$I = \int_{-\infty}^{\infty} \mathrm{e}^{-z^2/2}\,\mathrm{d}z = \int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2}\,\mathrm{d}x = \int_{-\infty}^{\infty} \mathrm{e}^{-y^2/2}\,\mathrm{d}y.$$

Riittää osoittaa, että $I^2 = 2\pi$. Nyt

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2}\,\mathrm{d}x\right)\left(\int_{-\infty}^{\infty} \mathrm{e}^{-y^2/2}\,\mathrm{d}y\right) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y = \left(\int_0^{\infty} r\mathrm{e}^{-r^2/2}\,\mathrm{d}r\right)\left(\int_0^{2\pi}\mathrm{d}\theta\right) \\ &= 2\pi\int_0^{\infty} \mathrm{e}^{-u}\,\mathrm{d}u = 2\pi. \end{aligned}$$

Näin siis tulos (5.4.2) pitää paikkansa. Edellä kolmas yhtäsuuruus saadaan siirtymällä napakoordinaatteihin:

$$x = r\cos\theta \qquad \text{ja} \qquad y = r\sin\theta.$$

Silloin $x^2 + y^2 = r^2$, $\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}\theta\,\mathrm{d}r$ ja integrointirajat ovat $0 < r < \infty$, $0 < \theta < 2\pi$.

Integraalilla (5.4.2) on myös läheinen yhteys gammafunktioon. Koska integraalissa (5.4.2) integroitava on symmetrinen nollan suhteen, niin integraalit yli välien $(-\infty, 0)$ ja $(0, \infty)$ ovat yhtä suuret. Siksi

$$\int_0^{\infty} \mathrm{e}^{-z^2/2}\,\mathrm{d}z = \sqrt{\frac{\pi}{2}}. \tag{5.4.3}$$

Tekemällä sijoitus $x = \frac{1}{2}z^2$ integraaliin (5.4.3) saadaan integraali, joka on $\Gamma\left(\frac{1}{2}\right)$. Silloin

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} x^{-1/2}\mathrm{e}^{-x}\,\mathrm{d}x = \sqrt{\pi}. \tag{5.4.4}$$

Lause 5.2 *Oletetaan, että Z noudattaa standardoitua normaalijakaumaa. Silloin*

1. *Z:n momenttifunktio on*

$$M(t) = \mathrm{e}^{t^2/2}, \qquad -\infty < t < \infty.$$

2. $E(Z) = 0$ *ja* $\mathrm{Var}(Z) = 1$.

Todistus. *1.* Määritelmän mukaan

$$M(t) = \int_{-\infty}^{\infty} \mathrm{e}^{tz} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}\,\mathrm{d}z.$$

Tehdään sijoitus $x = z - t$. Silloin $\mathrm{d}z = \mathrm{d}x$ ja $\mathrm{e}^{tz}\mathrm{e}^{-z^2/2} = \mathrm{e}^{(t^2-x^2)/2}$, joten

$$M(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{(t^2-x^2)/2}\,\mathrm{d}x = \mathrm{e}^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}\,\mathrm{d}x = \mathrm{e}^{t^2/2}.$$

Viimeinen yhtäsuuruus seuraa siitä, että integraali yli normaalijakauman tiheysfunktion $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ on 1.

2. Koska $M(t) = \mathrm{e}^{t^2/2}$, niin $M'(t) = t\mathrm{e}^{t^2/2}$ ja $M''(t) = \mathrm{e}^{t^2/2} + t^2\mathrm{e}^{t^2/2}$. Silloin $M'(0) = 0$, $M''(0) = 1$ ja $\mathrm{Var}(Z) = M''(0) - [M'(0)]^2 = 1$. □

Merkitään $Z \sim \mathrm{N}(0,1)$, missä siis $E(Z) = 0$ ja $\mathrm{Var}(Z) = 1$. Seuraavassa pykälässä määritellään normaalijakauma, jonka keskiarvo on μ ja varianssi σ^2.

5.4.2 Yleinen normaalijakauma

Satunnaismuuttuja X *noudattaa normaalijakaumaa* keskiarvolla μ ja varianssilla $\sigma^2 > 0$, jos se voidaan esittää muodossa

$$X = \mu + \sigma Z,$$

missä $Z \sim \mathrm{N}(0,1)$. Silloin merkitään $X \sim \mathrm{N}(\mu, \sigma^2)$. Jos $X \sim \mathrm{N}(\mu, \sigma^2)$, niin vastaavasti

$$Z = \frac{X - \mu}{\sigma} \sim \mathrm{N}(0,1).$$

Seuraavassa lauseessa esitetään jakaumaa koskevat perustulokset.

Lause 5.3 *Jos $X \sim \mathrm{N}(\mu, \sigma^2)$, niin*

1. $E(X) = \mu, \quad \mathrm{Var}(X) = \sigma^2$ *ja*

2. $$M_X(t) = E\big(\mathrm{e}^{tX}\big) = \mathrm{e}^{\mu t + \sigma^2 t^2/2}, \qquad -\infty < t < \infty.$$

3. *X:n tiheysfunktio on*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty.$$

Todistus. *1.* Koska $X \sim \mathrm{N}(\mu, \sigma^2)$, niin $X = \mu + \sigma Z$, missä $Z \sim \mathrm{N}(0,1)$. Silloin

$$E(X) = E(\mu + \sigma Z) = \mu + \sigma\, E(Z) = \mu$$

ja

$$\mathrm{Var}(X) = \mathrm{Var}(\mu + \sigma Z) = \sigma^2\,\mathrm{Var}(Z) = \sigma^2.$$

2. Määritelmän mukaan (ks. myös Lause 3.14)

$$\begin{aligned} M_X(t) &= E\left(\mathrm{e}^{tX}\right) = E\big[\mathrm{e}^{t(\mu+\sigma Z)}\big] = \mathrm{e}^{t\mu}\, E\big(\mathrm{e}^{t\sigma Z}\big) \\ &= \mathrm{e}^{t\mu} M_Z(t\sigma) = \mathrm{e}^{t\mu}\mathrm{e}^{t^2\sigma^2/2} = \mathrm{e}^{t\mu + t^2\sigma^2/2}. \end{aligned}$$

3. Tehdään muunnos $x = h(z) = \mu + \sigma z$. Silloin h:lla on käänteisfunktio g ja $z = g(x) = \frac{x-\mu}{\sigma}$ sekä $g'(x) = \frac{1}{\sigma}$. Alaluvussa 5.5 esitettävän muunnostekniikan avulla saadaan X:n tiheysfunktioksi

$$\begin{aligned} f_X(x) &= f_Z\left(\frac{x-\mu}{|\sigma|}\right)\frac{1}{|\sigma|} \\ &= \frac{1}{\sqrt{2\pi}|\sigma|}\mathrm{e}^{-(x-\mu)^2/2\sigma^2}. \end{aligned} \tag{5.4.5}$$

Tavallisesti tiheysfunktio kirjoitetaan muodossa

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-\mu)^2/2\sigma^2},$$

missä

$$\sigma = +\sqrt{\mathrm{Var}(X)} = +\sqrt{\sigma^2}$$

on X:n hajonta. Todistuksessa ei oletettu, että $\sigma > 0$. □

Esimerkki 5.9 Jos X:n tiheysfunktio on

$$f(x) = \frac{1}{\sqrt{32\pi}}\mathrm{e}^{-(x+7)^2/32}, \qquad -\infty < x < \infty,$$

niin $X \sim \mathrm{N}(-7, 16)$ ja

$$M_X(t) = \mathrm{e}^{-7t+8t^2}.$$

□

Esimerkki 5.10 Jos X:n momenttifunktio on

$$M_X(t) = \mathrm{e}^{5t+12t^2},$$

niin $X \sim \mathrm{N}(5, 24)$ ja X:n tiheysfunktio on

$$f(x) = \frac{1}{\sqrt{48\pi}}\mathrm{e}^{-(x-5)^2/48}, \qquad -\infty < x < \infty.$$

□

Jos $X \sim \mathrm{N}(\mu, \sigma^2)$, niin X:n tiheysfunktio saavuttaa maksimin pisteessä $x = \mu$ ja käänteispisteet ovat $x = \mu \pm \sigma$. Todennäköisyysmassa on jakautunut siten, että

$$\begin{aligned} P(|X-\mu| \le \sigma) &= P(|Z| \le 1) = 0.6826, \\ P(|X-\mu| \le 2\sigma) &= P(|Z| \le 2) = 0.9544, \\ P(|X-\mu| \le 3\sigma) &= P(|Z| \le 3) = 0.9974, \end{aligned}$$

missä $Z \sim \mathrm{N}(0,1)$. Esimerkiksi

$$\begin{aligned} P(|X-\mu| \le \sigma) &= P(|Z| \le 1) = P(-1 \le Z \le 1) \\ &= \Phi(1) - \Phi(-1) = 0.8413447 - 0.1586553 = 0.6826895, \end{aligned}$$

missä

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-v^2/2}\,\mathrm{d}v$$

on standardimuotoisen normaalijakauman kertymäfunktio. Sen arvot on taulukoitu ja se saadaan laskettua useilla ohjelmistoilla. Edellä esitettyjen todennäköisyyksien kahden numeron likiarvoina käytetään tavallisesti lukuja 0.68, 0.95 ja 0.99, jotka eivät ole pyöristettyjä vaan katkaistuja arvoja. Myös yllä esitetyt neljän numeron likiarvot ovat katkaistuja arvoja.

Lause 5.4

1. *Olkoon $X \sim \mathrm{N}(\mu, \sigma^2)$ ja $U = aX + b$, missä $a \neq 0$ ja b ovat annettuja vakioita. Silloin*
$$U \sim \mathrm{N}(a\mu + b, a^2\sigma^2).$$

2. *Olkoot $X_1, X_2, \ldots, X_n$ riippumattomat, $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$ ja $a_1, a_2, \ldots, a_n$, b ovat annetut vakiot, joista ainakin yksi a_i poikkeaa nollasta. Silloin $Y = \sum_{i=1}^n a_i X_i + b$ noudattaa normaalijakaumaa*
$$Y \sim \mathrm{N}\left(\sum_{i=1}^n a_i\mu_i + b,\ \sum_{i=1}^n a_i^2\sigma_i^2\right).$$

Esimerkki 5.11 Riippumattomat satunnaismuuttujat X_1, X_2, X_3 noudattavat normaalijakaumaa siten, että $X_i \sim \mathrm{N}(2^i, i^i)$, $i = 1, 2, 3$. Silloin $Y = X_1 + X_2 + X_3 \sim \mathrm{N}(14, 32)$, sillä

$$E(Y) = 2 + 2^2 + 2^3 = 14 \qquad \text{ja} \qquad \mathrm{Var}(Y) = 1 + 2^2 + 3^3 = 32$$

ja Lauseen 5.4 mukaan Y noudattaa normaalijakaumaa. Satunnaismuuttuja $Y = X_1 + 2X_2 + 3X_3 \sim \mathrm{N}(34, 260)$, koska

$$E(Y) = 2 + 2 \cdot 2^2 + 3 \cdot 2^3 = 34$$

ja

$$\mathrm{Var}(Y) = 1 + 2^2 \cdot 2^2 + 3^2 \cdot 3^3 = 260.$$

□

5.5 Muuttujien vaihto

Oletetaan, että X on jatkuva satunnaismuuttuja, jonka kertymäfunktio on $F(x)$. Lukuisissa sovelluksissa tarvitaan satunnaismuuttujan X jonkin funktion $Y = h(X)$ jakaumaa, kun X:n jakauma tunnetaan. Tehtävänämme on nyt siis määrittää satunnaismuuttujan $Y = h(X)$ jakauma, missä $h(x)$ on x:n reaaliarvoinen funktio.

5.5.1 Muunnos kertymäfunktio avulla

Voimme pyrkiä johtamaan Y:n kertymäfunktion

$$G(y) = P(Y \leq y)$$

suoraan X:n kertymäfunktion $F(x)$ avulla. Y:n tiheysfunktio $g(y)$ voidaan määrittää sitten identiteetin (5.1.3) avulla, kun $G(y)$ on derivoituva.

Esimerkki 5.12 Olkoon X jatkuva satunnaismuuttuja, jonka tiheysfunktio on

$$f(x) = \frac{3x^2}{2}, \qquad -1 \leq x \leq 1.$$

Tarkastellaan satunnaismuuttujan $Y = X^2$ jakaumaa. Silloin Y:n arvoavaruus on $S_Y = [0, 1]$ ja Y:n kertymäfunktio on

$$\begin{aligned} G(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \geq X \geq \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{3x^2}{2}\,\mathrm{d}x = \Big/_{-\sqrt{y}}^{\sqrt{y}} \frac{x^3}{2} = y^{3/2}, \qquad 0 \leq y \leq 1. \end{aligned}$$

Derivoimalla saadaan Y:n tiheysfunktioksi

$$g(y) = G'(y) = \frac{3y^{1/2}}{2}, \qquad 0 \leq y \leq 1.$$

□

Esimerkki 5.13 Olkoon X jatkuva satunnaismuuttuja, jonka kertymäfunktio on

$$F(x) = 1 - (1 + x)\mathrm{e}^{-x}, \qquad x > 0.$$

Johdetaan satunnaismuuttujan $Y = \mathrm{e}^{-X}$ jakauma. Merkitään Y:n kertymäfunktiota G:llä. Silloin

$$\begin{aligned} G(y) &= P(Y \leq y) = P\big(\mathrm{e}^{-X} \leq y\big) = P[-X \leq \log(y)] \\ &= P[X \geq -\log(y)] = 1 - P[X < -\log(y)] \\ &= 1 - F[-\log(y)], \end{aligned}$$

missä $F(x)$ on X:n kertymäfunktio. Sijoittamalla $x = -\log(y)$ X:n kertymäfunktioon saadaan

$$G(y) = [1 - \log(y)]\mathrm{e}^{\log(y)} = [1 - \log(y)]y.$$

Koska $S_X = (0, \infty)$, niin $S_Y = (0, 1)$. Y on jatkuva satunnaismuuttuja, koska $G(y)$ on jatkuva ja sillä on jatkuva derivaatta muualla paitsi pisteessä $y = 0$. Y:n tiheysfunktio on

$$g(y) = G'(y) = \begin{cases} -\log(y), & \text{kun } 0 < y < 1; \\ 0 & \text{muualla.} \end{cases}$$

Huomaa, että $-\log(y) > 0$, kun $0 < y < 1$. Nyt siis $g(y) \geq 0$ kaikilla $y \in S_Y = (0, 1)$. □

5.5.2 Muunnos tiheysfunktion avulla

Seuraavaksi esitetään yleinen menetelmä, jonka avulla voidaan johtaa satunnaismuuttujan X funktion $Y = h(X)$ tiheysfunktio suoraan X:n tiheysfunktion $f_X(x)$ avulla. Menetelmän edellyttää kuitenkin, että funktiolla $h(x)$ on tarkasteltavalla välillä *käänteisfunktio.* Esimerkiksi funktion $y = \mathrm{e}^x$ käänteisfunktio on $x = \log(y)$. Myös funktio $y = x^2$ on *kääntyvä,* kun $x > 0$, sillä silloin $x = \sqrt{y}$. Funktio $y = x^2$ *ei ole kääntyvä* koko reaaliakselilla, koska silloin $x = \pm\sqrt{y}$, joka ei ole funktio. Huomattakoon, että jatkuva funktio $h(x)$ on *kääntyvä,* jos ja vain jos se on joko aidosti kasvava tai aidosti vähenevä.

Lineaarinen munnos

Tarkastellaan ensin yksinkertaista lineaarista muunnosta $Y = aX + b$, missä a ja b ovat annettuja vakioita. Nyt siis $h(X) = aX + b$. Funktion $y = h(x)$ derivaatta on

$$\frac{\mathrm{d}y}{\mathrm{d}x} = h'(x) = a.$$

Funktiolla $h(x)$ on käänteisfunktio

$$g(y) = \frac{y - b}{a}, \qquad a \neq 0$$

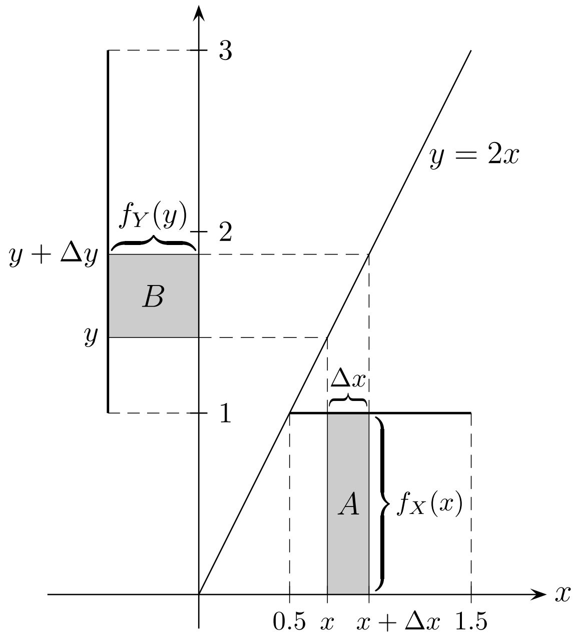
ja

$$\frac{\mathrm{d}y}{\mathrm{d}x} = g'(y) = \frac{1}{a}.$$

Esimerkki 5.14 Oletetaan, että $X \sim \mathrm{Tas}(0.5, 1.5)$ ja $Y = 2X$. Mitä jakaumaa Y noudattaa?

Kuviossa 5.8 on alueen A pinta-ala

$$P[X \in (x, x + \Delta x)] = f_X(x) \cdot \Delta x = \Delta x$$



Kuvio 5.8. Tasajakaumaa Tas$(0.5, 1.5)$ noudattavan satunnaismuuttujan X lineaarinen muunnos.

ja alueen B pinta-ala

$$P[Y \in (y, y + \Delta y)] = f_Y(y) \cdot \Delta y.$$

Tapahtumat $X \in (x, x + \Delta x)$ ja $Y \in (y, y + \Delta y)$ sattuvat täsmälleen samanaikaisesti, joten

$$P[X \in (x, x + \Delta x)] = P[Y \in (y, y + \Delta y)]. \tag{5.5.1}$$

Koska $y = 2x$ ja $y + \Delta y = 2(x + \Delta x)$, niin $\Delta y = 2\Delta x$ ja identiteetistä (5.5.1) seuraa, että $f_Y(y) = \frac{1}{2}$. Koska $0.5 < x < 1.5$, niin $1 < y < 3$. Näin siis $Y \sim \text{Tas}(1, 3)$:

$$f_Y(y) = \begin{cases} \frac{1}{2}, & 1 < y < 3; \\ 0, & \text{muualla.} \end{cases}$$

□

Olkoon X satunnaismuuttuja, jonka arvoavaruus on S_X. Silloin satunnaismuuttujan $Y = h(X)$ arvoavaruus S_Y määräytyy siten, että

$$X \in S_X \Leftrightarrow Y \in S_Y.$$

Seuraavassa lauseessa esitettävässä menetelmässä oletetaan, että funktio $y = h(x)$ on tarkasteltavalla arvoalueella kääntyvä. Silloin on olemassa sellainen funktio $x = g(y)$, että

$$y = h(x) \Leftrightarrow x = g(y).$$

Lause 5.5 *Olkoon X jatkuva satunnaismuuttuja, jonka tiheysfunktio on $f_X(x)$ ja arvoavaruus S_X. Olkoon $Y = h(X)$ sellainen funktio, että sillä on käänteisfunktio $x = g(y)$ ja käänteisfunktion derivaatta $g'(y)$ on olemassa kaikilla $y \in S_Y$, missä S_Y on Y:n arvoavaruus. Silloin Y:n tiheysfunktio on*

$$f_Y(y) = f_X\big(g(y)\big)|g'(y)|, \qquad y \in S_Y.$$

Todistus. Oletuksen mukaan $g(y)$ on derivoituva, joten se on jatkuva. Koska h ja g ovat kääntyviä, niin h ja g ovat molemmat joko kasvavia tai väheneviä. Oletetaan h ja g ovat väheneviä. Silloin

$$F_Y(y) = P(Y \le y) = P(h(X) \le y) = P(X \ge g(y)) = 1 - F_X\big(g(y)\big).$$

Derivoidaan $1 - F_X[g(y)]$ ketjusäännön avulla, jolloin saadaan

$$\begin{aligned} f_Y(y) &= F_Y'(y) = -F_X'\big(g(y)\big)g'(y) \\ &= -f_X\big(g(y)\big)g'(y) = f_X\big(g(y)\big)|g'(y)|. \end{aligned}$$

Viimeinen yhtäsuuruus seuraa siitä, että $g'(y)$ on negatiivinen, koska g on vähevä.

Jos h ja g ovat kasvavia, niin todistus on melkein samanlainen ja se jätetään harjoitustehtäväksi. □

Esimerkki 5.15 Olkoon X jatkuva satunnaismuuttuja, jonka tiheysfunktio on $f_X(x) = \mathrm{e}^{-x}$ ja $S_X = \{\, x \mid x > 0 \,\}$. Olkoon $Y = X^{1/2}$, joten $X = Y^2 = g(Y)$ ja $S_Y = S_X$. Koska $g'(y) = 2y$, niin

$$f_Y(y) = f_X(y^2)|2y| = 2y\mathrm{e}^{-y^2}, \qquad y > 0.$$

Tarkastellaan vielä satunnaismuuttujaa $V = \mathrm{e}^{-X}$. Silloin $X = -\log(V)$. Merkitään nyt $-\log(V) = \tilde{g}(V)$. Silloin $S_V = [0, 1]$ ja $\tilde{g}'(v) = -1/v$. Siksi

$$f_V(v) = f_X[-\log(v)]|\frac{-1}{v}| = \frac{v}{v} = 1,$$

joten V noudattaa tasajakaumaa välillä $[0, 1]$. □

Mikäli muunnosfunktiolla h ei ole käänteisfunktiota X:n arvoavaruudessa S_X, niin Lauseen 5.5 muunnosmenetelmää ei voi suoraan soveltaa. Jos kuitenkin on olemassa sellainen S_X:n ositus yhteispisteettömiin osaväleihin A_1, A_2, …, A_m, että

(5.5.2) $$S_X = A_1 \cup A_2 \cup \cdots \cup A_m$$

ja h on kääntyvä jokaisella osavälillä, voidaan muunnos tehdä jokaisella osavälillä erikseen. Sitä varten määritellään funktiot

$$h(x) = \begin{cases} h_i(x), & \text{kun } x \in A_i; \\ 0 & \text{muualla.} \end{cases}$$

Silloin $h(x)$ voidaan kirjoittaa muodossa $h(x) = \sum_{i=1}^m h_i(x)$, missä jokainen $h_i(x)$ on kääntyvä välillä A_i. Olkoot funktioiden h_i käänteisfunktiot vastaavasti g_i, $i = 1, 2, \ldots, m$. Satunnaismuuttujan $Y = h(X)$ tiheysfunktio voidaan nyt esittää Lauseen 5.5 avulla muodossa

$$(5.5.3) \qquad f_Y(y) = \sum_{i=1}^{m} f_X\big(g_i(y)\big)|g_i'(y)|. \qquad y \in S_Y.$$

Huomattakoon, että joskus tarvitaan äärellisen osituksen (5.5.2) sijasta ositus, jossa jakovälejä A_1, A_2, $\ldots$ on ääretön määrä ($m = \infty$).

5.5.3 Normaalimuuttujan muunnokset

Jos $X \sim \mathrm{N}(0, 1)$, niin X:n tiheysfunktio on

$$f(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}, \qquad -\infty < x < \infty,$$

joka on standardimuotoisen normaalijakauman tiheysfunktio. Johdetaan nyt satunnaismuuttujan $U = X^2$ jakauma. Muunnosfunktio $u = h(x) = x^2$ ei ole kääntyvä, koska $x = \pm\sqrt{u}$ ei ole funktio. Siksi esitämme arvoavaruuden $S_X = \{\, -\infty < x < \infty \,\}$ ositettuna muodossa

$$S_X = (-\infty, 0] \cup (0, \infty).$$

Silloin funktiolla $h(x)$ on välillä $(-\infty, 0]$ käänteisfunktio $g_1(u) = -\sqrt{u}$ ja välillä $(0, \infty)$ käänteisfunktio $g_2(u) = \sqrt{u}$. Nyt siis kaavan (5.5.3) mukaan U:n tiheysfunktio on

$$(5.5.4) \qquad f_U(u) = f_X(-\sqrt{u})\frac{1}{2\sqrt{u}} + f_X(\sqrt{u})\frac{1}{2\sqrt{u}} = \frac{1}{\sqrt{2\pi u}} \mathrm{e}^{-u/2},$$

kun $u \in (-\infty, \infty)$. U noudattaa χ^2-jakaumaa vapausastein 1. Käsittelemme tilastotieteessä tärkeää χ^2-jakaumaa vielä jatkossa tarkemmin.

Lause 5.6 *Jos* $X \sim \mathrm{N}(\mu, \sigma^2)$, $\sigma^2 > 0$, *niin silloin*

$$\frac{(X-\mu)^2}{\sigma^2} \sim \mathrm{Khi2}(1).$$

Todistus. Koska $X \sim \mathrm{N}(\mu, \sigma^2)$, niin määritelmän mukaan $\frac{X-\mu}{\sigma} = Z \sim \mathrm{N}(0, 1)$. Edellä näytettiin, että $Z^2 \sim \mathrm{Khi2}(1)$. Näin on lause todistettu. □

Lause 5.7 *Jos Z_i:t ovat riippumattomat ja* $Z_i \sim \mathrm{N}(0, 1)$, $i = 1, 2, \ldots, n$, *niin*

$$Z_1^2 + Z_2^2 + \cdots Z_n^2 \sim \mathrm{Khi2}(n).$$

Jos tehdään otos normaalijakaumasta $\mathrm{N}(0,1)$, niin Lauseen 5.7 mukaan havaintojen neliösumma noudattaa Khi2-jakaumaa vapausastein n, missä n on otoskoko.

Seuraus 5.1 *Jos X_i:t ovat riippumattomat ja $X_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$, niin*

$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \sim \mathrm{Khi2}(n).$$

Jos vastaavasti tehdään n:n suuruinen otos normaalijakaumasta $\mathrm{N}(\mu, \sigma^2)$, niin Seurauslauseen 5.1 mukaan standardoitujen havaintojen neliösumma noudattaa Khi2-jakaumaa vapausastein n.

Lause 5.8 *Olkoot X_1 ja X_2 riippumattomat ja $X_i \sim \mathrm{Khi2}(n_i)$, $i = 1, 2$. Silloin*

$$X_1 + X_2 \sim \mathrm{Khi2}(n_1 + n_2).$$

5.6 Satunnaismuuttujan funktion odotusarvo

Olkoon X jatkuva satunnaismuuttuja, jonka tiheysfunktio $f(x)$ on määritelty arvoavaruudessa S. Olkoon $h(X)$ satunnaismuuttujan X reaaliarvoinen funktio, joka siis määrittelee uuden satunnaimuuttujan.

Määritelmä 5.3 Jos X on jatkuva satunnaismuuttuja, niin satunnaismuuttujan $h(X)$ odotusarvo on

$$E[h(X)] = \int_S h(x) f(x)\,\mathrm{d}x, \tag{5.6.1}$$

mikäli $E\big(|h(X)|\big) < \infty$. Jos $E\big(|h(X)|\big) = \infty$, niin sanomme, että $E[h(X)]$ ei ole olemassa.

Huomautus 5.1 Odotusarvon $E[h(X)]$ olemassaolo tarkoittaa siis sitä, että funktion $|h(X)|$ odotusarvo on äärellinen. Jos X noudattaa esimerkiksi eksponenttijakaumaa keskiarvolla 1, niin $f(x) = \mathrm{e}^{-x}$ ja $S = [0, \infty)$. Silloin X:n odotusarvo on

$$\begin{aligned} E(X) &= \int_0^\infty x\mathrm{e}^{-x}\,\mathrm{d}x \\ &= \Big/_0^\infty (-x\mathrm{e}^{-x}) + \int_0^\infty \mathrm{e}^{-x}\,\mathrm{d}x \qquad \text{(osittaisintegrointi)} \\ &= \int_0^\infty \mathrm{e}^{-x}\,\mathrm{d}x = 1, \end{aligned}$$

joten odotusarvo on olemassa. Hyvin usein odotusarvot ovat epäoleellisia integraaleja, niin kuin tässäkin esimerkissä.

Jos $h(X)$ integroituu itseisesti, eli

$$\int_S |h(x)|$$

on äärellisenä olemassa, niin $E[h(X)]$ on olemassa. Funktio $V = h(X)$ on satunnaismuuttuja, jonka tiheysfunktio $g(v)$ on määritelty arvoavaruudessa $S_V = \{\, v \mid v = h(x),\ x \in S \,\}$. Silloin

$$E[h(X)] = E(V) = \int_{S_V} vg(v).$$

Esimerkki 5.16 Tarkastellaan nyt *Cauchyn jakaumaa* noudattavaa satunnaismuuttujaa X, jonka tiheysfunktio on

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad -\infty < x < \infty. \tag{5.6.2}$$

Kaava (5.6.2) todellakin määrittelee tiheysfunktion, koska

$$\int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)}\,\mathrm{d}x = \frac{2}{\pi} \Big/_{0}^{\infty} \arctan(x) = \frac{2}{\pi} \cdot \frac{\pi}{2} = 1.$$

Osoitamme nyt, että $E(|X|) = \infty$, mistä seuraa, että *Cauchyn jakaumalla ei ole keskiarvoa.* Symmetrian nojalla voidaan kirjoittaa

$$E(|X|) = \int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)}\,\mathrm{d}x = \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2}\,\mathrm{d}x.$$

Jokaista reaalilukua $M > 0$ kohti saadaan

$$\int_0^M \frac{x}{(1+x^2)}\,\mathrm{d}x = \Big/_{0}^{M} \frac{\log(1+x^2)}{2} = \frac{\log(1+M^2)}{2}.$$

Tästä seuraa, että

$$E(|X|) = \lim_{M\to\infty} \frac{2}{\pi} \int_0^M \frac{x}{1+x^2}\,\mathrm{d}x = \frac{1}{\pi} \lim_{M\to\infty} \log(1+M^2) = \infty,$$

joten $E(X)$ ei ole olemassa. □

Taulukko 5.1. Tärkeitä odotusarvoja.

$h(x)$	$E[h(X)]$	Merkintä	Nimitys
x	$E(X)$	μ	odotusarvo
x^r	$E(X^r)$	α_r	r. momentti
$x^{(r)}$	$E[X^{(r)}]$	g_r	r. tekijämomentti
$(x-\mu)^2$	$E[(X-\mu)^2]$	σ^2	varianssi
$(x-\mu)^r$	$E[(X-\mu)^r]$	μ_r	r. keskusmomentti

5.6.1 Momentifunktio ja momentit

Kun $h(X) = X^r$, niin $E[h(X)] = E(X^r)$ on X:n r. momentti. Jatkuvien satunnaismuuttujien momentit määritellään vastaavasti kuin diskreettien satunnaismuuttujien momentit. Summalausekkeet vain korvataan integraaleilla. Taulukossa 5.1 esitetään yhteenveto eri momenteista

Momenttifunktio määriteltiin 3. luvussa (Määritelmä 3.12) ja jatkuville satunnaismuuttujille alaluvussa 5.1 [ks. identiteetti (5.1.5)]. Jatkuvan satunnaismuuttujan X momentifunktio on

$$M(t) = E(\mathrm{e}^{tX}) = \int\limits_S \mathrm{e}^{tx} f(x)\,\mathrm{d}x, \qquad t \in A,$$

missä $f(x)$ on X:n tiheysfunktio ja A sellainen t:n arvojen joukko, että $M(t)$ on äärellinen kaikilla $t \in A$. Koska $M(0) = 1$, niin $0 \in A$. Sanomme, että $M(t)$ on olemassa, jos $(-a, a) \subset A$ jollakin $a > 0$. Momenttifunktion perusominaisuudet esitettiin Pykälässä 3.5.2.

Esimerkki 5.17 Huomautuksessa 5.1 laskettiin odotusarvo $E(X)$, kun $X \sim$Exp(1). Silloin X:n tiheysfunktio on $f(x) = \mathrm{e}^{-x} \geq 0$ välillä $S = [0, \infty)$ ja $f(x) = 0$ muualla. Kaikki momentit $E(X^r)$ voidaan määrittää osittaisintegroinnilla, mutta käytetämme nyt momenttifunktiota, joka on

$$M(t) = E(\mathrm{e}^{tX}) = \int\limits_0^\infty \mathrm{e}^{tx}\mathrm{e}^{-x}\,\mathrm{d}x = \frac{1}{1-t}, \qquad t < 1.$$

Derivoimalla $M(t)$ toistuvasti r kertaa saadaan $M^{(r)}(t) = \frac{r!}{(1-t)^{k+1}}$. Siksi

$$E(X^r) = M^{(r)}(0) = r!,$$

joten

$$\mu = E(X) = 1, \qquad E(X^2) = 2, \qquad \sigma^2 = E(X^2) - \mu^2 = 1.$$

□

Erityisesti keskiarvo μ, varianssi σ^2 ja hajonta $\sigma = \sqrt{\mathrm{Var}(X)}$ ovat tavallisimmat tunnusluvut, joilla jakaumaa luonnehditaan. Jakauman yksityiskohtaisemmassa tarkastelussa voidaan käyttää myös korkeampia momentteja, mikäli ne ovat olemassa.

Vinous ja huipukkuus

Satunnaismuuttujan 1. momentti μ määrittää jakauman sijainnin. Keskistetyn muuttujan $X-\mu$ toinen momentti (keskusmomentti) on varianssi σ^2 ja se mittaa todennäköisyysmassan hajaantumista. Normeeratun muuttujan $(X-\mu)/\sigma$ kolmas ja neljäs momentti luonnehtivat jakauman muotoa.

Jakauman *vinouskerroin,* josta käytetään merkintää γ_1, määritellään seuraavasti:

$$\gamma_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3}, \tag{5.6.3}$$

missä μ_3 on jakauman 3. keskusmomentti ja $\sigma = \sqrt{\mathrm{Var}(X)}$ on hajonta. Olkoon X:n tiheysfunktio $f(x)$. Silloin X:n jakauma on *symmetrinen pisteen* a suhteen, jos

$$f(a-x) = f[-(a-x)]$$

kaikilla x:n arvoilla. Jos $E(X)$ on olemassa, niin silloin $E(X) = a$. Symmetrisen jakauman vinouskerroin on nolla. Jos jakaumalla on pitkä häntä oikealle, kuten Poissonin jakaumalla ja geometrisella jakaumalla, niin jakauma on positiivisesti vino ja $\gamma_1 > 0$. Jos jakaumalla on pitkä häntä vasemmalle, niin $\gamma_1 < 0$. Jakaumalla on tietysti oltava 3. momentti, jotta vinouskerroin voidaan laskea. Huomaa, että Cauchyn jakauma, jonka tiheysfunktio on

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad -\infty < x < \infty,$$

on symmetrinen pisteen $a = 0$ suhteen, mutta 0 ei ole jakauman keskiarvo, koska jakaumalla ei ole keskiarvoa (ks. Esimerkki 5.16). Cauchyn jakauman vinouskerrointa ei voida laskea, vaikka määritelmän nojalla voimme todeta jakauman olevan symmetrinen.

Huipukkuuskerrointa merkitään γ_2 ja se määritellään 4. keskusmomentin avulla seuraavasti:

$$\gamma_2 = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4}, \tag{5.6.4}$$

missä μ_4 on X:n 4. keskusmomentti. Standardimuotoisen normaalijakauman $\mathrm{N}(0,1)$ huipukkuus on 3. Jos jakaumalla on paksummat hännät kuin normaalijakaumalla $\mathrm{N}(0,1)$, niin silloin $\gamma_2 > 3$. Jos hännät ovat ohuemmat kuin normaalijakaumalla $\mathrm{N}(0,1)$, niin $\gamma_2 < 3$. Usein huipukkuuden mittana käytetäänkin poikkeamaa normaalijakauman $\mathrm{N}(0,1)$ huipukkuudesta: $\frac{\mu_4}{\sigma^4} - 3$.

5.7 Kaksiulotteiset jakaumat

Tarkastellaan nyt kahden jatkuvan satunnaismuuttujan yhteisjakaumaa. Yleistys usean muuttujan tapaukseen on sen jälkeen suoraviivainen.

Määritelmä 5.4 Olkoot X ja Y samassa otosavaruudessa määritellyt jatkuvat satunnaismuuttujat. Olkoon kaksiulotteisen jatkuvan satunnaismuuttujan (X, Y) arvoavaruus S. Funktio $f(x, y)$ on (X, Y):n tiheysfunktio (X:n ja Y:n yhteisjakauman tiheysfunktio), jos sillä on seuraavat ominaisuudet:

1. $f(x, y) \geq 0 \qquad \text{kaikilla} \qquad (x, y) \in \mathbb{R}^2,$
2. $\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f(x, y) \, \mathrm{d}x \, \mathrm{d}y = 1 \qquad \text{ja}$
3. $$P[(X, Y) \in A] = \iint\limits_{(x,y) \in A} f(x, y) \, \mathrm{d}x \, \mathrm{d}y,$$
   missä $(X, Y) \in A$ on tasossa määritelty tapahtuma.

Esimerkki 5.18 Olkoon X:n ja Y:n yhteisjakauman tiheysfunktio

$$f(x, y) = \frac{3}{2} x^2 (1 - |y|), \qquad -1 < x < 1, \quad -1 < y < 1.$$

Määritellään $A = \{\, (x, y) \mid 0 < x < 1,\ 0 < y < x \,\}$. Todennäköisyys, että $(X, Y) \in A$, on

$$\begin{aligned} P[(X, Y) \in A] &= \int\limits_0^1 \int\limits_0^x \frac{3}{2} x^2 (1 - y) \, \mathrm{d}y \, \mathrm{d}x = \int\limits_0^1 \frac{3}{2} x^2 \Big/_{\!0}^{x} \left(y - \frac{y^2}{2} \right) \mathrm{d}x \\ &= \int\limits_0^1 \frac{3}{2} \left(x^3 - \frac{x^4}{2} \right) \mathrm{d}x = \frac{3}{2} \Big/_{\!0}^{1} \left(\frac{x^4}{4} - \frac{x^5}{10} \right) = \frac{9}{40}. \end{aligned}$$

□

5.7.1 Reunajakauma ja ehdollinen jakauma

Kaksiulotteista satunnaismuuttujaa (X, Y) kutsutaan kaksiulotteiseksi satunnaisvektoriksi. Silloin X ja Y ovat tietysti (yksiulotteisia) satunnaismuuttujia. X:n *reunajakauman tiheysfunktio,* jota merkitään $f_X(x)$, on pelkästään X:n tiheysfunktio, jossa Y:tä ei oteta huomioon. Satunnaismuuttujan X *ehdollinen tiheysfunktio* ehdolla $Y = y$ on on X:n tiheysfunktio, kun Y:n arvo tunnetaan. X:n ehdollista tiheysfunktiota ehdolla $Y = y$ merkitään $f_X(x \mid Y = y)$ tai lyhyesti $f_X(x \mid y)$.

Määritelmä 5.5 Olkoon $f(x, y)$ jatkuvan satunnaisvektorin (X, Y) tiheysfunktio ja S sen arvoavaruus. Silloin satunnaismuuttujat X ja Y ovat jatkuvia ja niiden reunajakaumien tiheysfunktiot ovat

$$f_X(x) = \int\limits_{-\infty}^{\infty} f(x, y) \, \mathrm{d}y, \quad x \in S_X; \qquad f_Y(y) = \int\limits_{-\infty}^{\infty} f(x, y) \, \mathrm{d}x, \quad y \in S_Y,$$

missä S_X on X:n ja S_Y on Y:n arvoavaruus. Satunnaismuuttujat *X ja Y ovat riippumattomat* jos ja vain jos

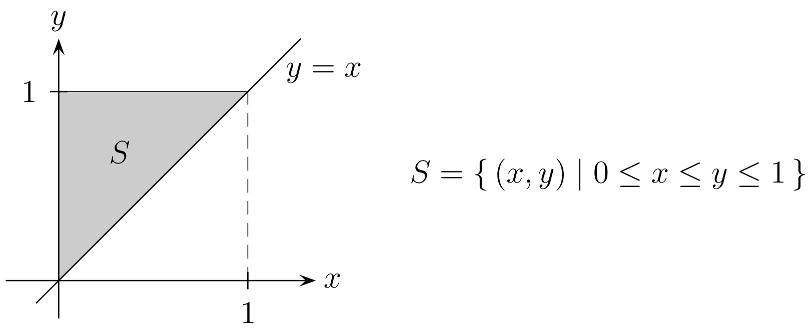
(5.7.1) $$f(x,y) = f_X(x)f_Y(y) \qquad \text{kaikilla } x \in S_X \text{ ja } y \in S_Y;$$

muutoin *X ja Y riippuvat toisistaan.*

Esimerkki 5.19 Olkoon X:n ja Y:n yhteisjakauman tiheysfunktio

$$f(x,y) = 2, \qquad 0 \le x \le y \le 1,$$

muualla $f(x,y) = 0$. Satunnaisvektorin (X,Y) arvoavaruus on $S = \{\,(x,y) \mid 0 \le x \le y \le 1\,\}$.



Kuvio 5.9. Tasajakauman $f(x,y) = 2$ määrittelyalue S.

Silloin esimerkiksi todennäköisyys

$$\begin{aligned} P\left(0 \le X \le \frac{1}{2},\ 0 \le Y \le \frac{1}{2}\right) &= P\left(0 \le X \le Y,\ 0 \le Y \le \frac{1}{2}\right) \\ &= \int_0^{1/2}\int_0^{y} 2\,\mathrm{d}y\,\mathrm{d}x = \int_0^{1/2} 2y\,\mathrm{d}y = \frac{1}{4}. \end{aligned}$$

Reunajakaumien tiheysfunktiot ovat

$$f_X(x) = \int_x^1 2\,\mathrm{d}y = 2(1-x), \qquad 0 \le x \le 1,$$

ja

$$f_Y(y) = \int_0^y 2\,\mathrm{d}x = 2y, \qquad 0 \le y \le 1.$$

Lasketaan vielä X:n ja Y:n odotusarvot sekä Y:n 2. momentti.

$$E(X) = \int_0^1 \int_x^1 2x \,\mathrm{d}y \,\mathrm{d}x = \int_0^1 2x(1-x) \,\mathrm{d}x = \frac{1}{3},$$

$$E(Y) = \int_0^1 \int_0^y 2y \,\mathrm{d}x \,\mathrm{d}y = \int_0^1 2y^2 \,\mathrm{d}y = \frac{2}{3},$$

$$E(Y^2) = \int_0^1 \int_0^y 2y^2 \,\mathrm{d}x \,\mathrm{d}y = \int_0^1 2y^3 \,\mathrm{d}y = \frac{1}{2}.$$

Odotusarvot $E(X)$, $E(Y)$ ja $E(Y)$ voidaan laskea joko suoraan reunajakaumasta tai sitten yhteisjakaumasta. □

Nähdään helposti, että Esimerkissä 5.19 satunnaismuuttujat X ja Y eivät ole riippumattomat, koska

$$f_X(x)f_Y(y) = 2(1-x)2y \neq f(x,y) = 2, \qquad (x,y) \in S.$$

Sen sijaan voidaan osoittaa, että Esimerkissä 5.18 satunnaismuuttujat X ja Y ovat riippumattomat.

Jatkuvan satunnaismuuttujan ehdollinen tiheysfunktio määritellään seuraavasti:

Määritelmä 5.6 Jos jatkuvan satunnaisvektorin (X, Y) tiheysfunktio on $f(x, y)$ ja arvoavaruus S, niin *X:n ehdollinen tiheysfunktio ehdolla $Y = y$* on

$$f_X(x \mid y) = \frac{f(x,y)}{f_Y(y)}, \qquad (x,y) \in S$$

ja *Y:n ehdollinen tiheysfunktio ehdolla $X = x$* on

$$f_Y(y \mid x) = \frac{f(x,y)}{f_X(x)}, \qquad (x,y) \in S.$$

Huomattakoon, että Määritelmässä 5.6 oletetaan, että $f_Y(y) > 0$ ja $f_X(x) > 0$.

Esimerkki 5.20 Olkoot satunnaismuuttujat X ja Y samat kuin Esimerkissä 5.19 Silloin

$$\begin{aligned} f(x,y) &= 2, \qquad 0 \le x \le y \le 1, \\ f_X(x) &= 2(1-x), \qquad 0 \le x \le 1, \\ f_Y(y) &= 2y, \qquad 0 \le y \le 1. \end{aligned}$$

Määritetään nyt Y:n ehdollisen jakauman tiheysfunktio, kun $X = x$ on annettu. Määritelmän 5.6 mukaan

$$f(y \mid x) = \frac{f(x,y)}{f_X(x)} = \frac{2}{2(1-x)} = \frac{1}{1-x}, \qquad x \le y \le 1, \quad 0 \le x \le 1.$$

Y:n ehdollinen odotusarvo ehdolla $X = x$ on

$$E(Y \mid x) = \int_x^1 y \frac{1}{1-x} \, \mathrm{d}y = \Big/_x^1 \frac{y^2}{2(1-x)} = \frac{1+x}{2}, \qquad 0 \le x \le 1.$$

Samalla tavalla voidaan osoittaa, että

$$E(X \mid y) = \frac{y}{2}, \qquad 0 \le y \le 1.$$

Suoraan määritelmän perusteella Y:n ehdollinen varianssi ehdolla $X = x$ on

$$\begin{aligned} E\big([Y - E(Y \mid x)]^2 \mid x\big) &= \int_x^1 \left(y - \frac{1+x}{2}\right)^2 \frac{1}{1-x} \, \mathrm{d}y \\ &= \Big/_x^1 \frac{1}{3(1-x)} \left(y - \frac{1+x}{2}\right)^3 \\ &= \frac{(1-x)^2}{12}. \end{aligned}$$

Jos $U \sim \mathrm{Tas}(a, b)$, niin $E(U) = \frac{a+b}{2}$ ja $\mathrm{Var}(U) = \frac{(b-a)^2}{12}$. Koska Y:n ehdollinen jakauma ehdolla $X = x$ on $\mathrm{Tas}(x, 1)$, niin olisimme voineet tasajakauman ominaisuuksien perusteella suoraan todeta, että

$$E(Y \mid x) = \frac{x+1}{2} \qquad \text{ja} \qquad \mathrm{Var}(Y \mid x) = \frac{(1-x)^2}{12}.$$

Lasketaan vielä ehdollinen todennäköisyys

$$P(3/4 \le Y \le 7/8 \mid X = 1/4) = \int_{3/4}^{7/8} f(y \mid 1/4) \, \mathrm{d}y = \int_{3/4}^{7/8} \frac{1}{3/4} \, \mathrm{d}y = \frac{1}{6}.$$

□

Havaitsimme edellisessä esimerkissä, että Y:n ehdollinen odotusarvo on x:n lineaarinen funktio:

$$E(Y \mid x) = \frac{1}{2} + \frac{1}{2}x, \qquad 0 \le x \le 1.$$

Jos $E(Y \mid x)$ on lineaarinen, niin pitää yleisesti paikkansa, että

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

missä $\rho = \mathrm{Cor}(X, Y)$ on X:n ja Y:n välinen korrelaatio, σ_X on X:n hajonta ja σ_Y on Y:n hajonta. Jos $E(X \mid y)$ on lineaarinen, niin

$$E(X \mid y) = \mu_X + \rho \frac{\sigma_X}{\sigma_Y}(y - \mu_Y).$$

Ehdollisten odotusarvojen $E(Y \mid x)$ ja $E(X \mid y)$ yhtälöissä kertoimien $\rho\frac{\sigma_Y}{\sigma_X}$ ja $\rho\frac{\sigma_X}{\sigma_Y}$ tulo on ρ^2. Esimerkissä 5.20 näiden kertoimien tulo on $\rho^2 = \frac{1}{4}$. Siksi $\rho = \frac{1}{2}$, koska molemmat kertoimet ovat positiiviset. Näiden kertoimien suhde on σ_Y^2/σ_X^2 ja esimerkissä tämä suhde on 1. Tästä voimme päätellä, että Esimerkissä 5.20 $\sigma_X^2 = \sigma_Y^2$.

Satunnaismuuttujien X ja Y riippumattomuuden tarkistaminen suoraan relaation (5.7.1) perusteella edellyttää reunajakaumien tiheysfunktioiden $f_X(x)$ ja $f_Y(y)$ tuntemista. Seuraava apulause tekee riippumattomuuden tarkistamisen jonkin verran helpommaksi, koska siinä ei edellytetä reunajakaumien tuntemista.

Apulause 5.1 *Olkoon (X, Y) kaksiulotteinen satunnaisvektori, jonka yhteisjakauman tiheysfunktio on $f(x, y)$. Silloin satunnaismuuttujat X ja Y ovat riippumattomat, jos ja vain jos on olemassa sellaiset funktiot $g(x)$ ja $h(y)$, että*

$$f(x, y) = g(x)h(y) \qquad \textit{kaikilla } x \in \mathbb{R} \textit{ ja kaikilla } y \in \mathbb{R},$$

missä g riippuu vain x:stä ja h vain y:stä.

Kertymäfunktio

Kaksiulotteinen jakauma voidaan täydellisesti luonnehtia kertymäfunktionsa avulla. Satunnaisvektorin (X, Y) yhteisjakauman kertymäfunktio $F(x, y)$ määritellään relaatiolla

$$F(x, y) = P(X \leq x,\ Y \leq y)$$

missä $(x, y) \in \mathbb{R}^2$. Tiheysfunktion avulla lausuttuna kertymäfunktio on

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t)\, \mathrm{d}s\, \mathrm{d}t.$$

Integraalilaskennan peruslause kahden muuttujan tapauksessa sanoo, että

$$\frac{\partial^2 F(x, y)}{\partial x\, \partial y} = f(x, y) \tag{5.7.2}$$

kaikissa $f(x, y)$:n jatkuvuuspisteissä. Relaatio (5.7.2) on hyödyllinen silloin, kun kertymäfunktio tunnetaan ja halutaan johtaa tiheysfunktio. Silloin tiheysfunktio $f(x, y)$ saadaan derivoimalla $F(x, y)$ sekä x:n että y:n suhteen eli laskemalla osittaisderivaatta $\frac{\partial^2 F(x,y)}{\partial x\, \partial y}$.

Esimerkki 5.21 Olkoon X:n ja Y:n yhteisjakauman kertymäfunktio

$$F(x, y) = \begin{cases} xy, & 0 \leq x \leq 1 \text{ ja } 0 \leq y \leq 1; \\ y, & x > 1,\ 0 \leq y \leq 1; \\ x, & y > 1,\ 0 \leq x \leq 1; \\ 1, & x > 1 \text{ ja } y > 1; \\ 0, & x < 0 \text{ tai } y < 0. \end{cases}$$

Laskemalla osittaisderivaatta $\frac{\partial^2 F(x,y)}{\partial x\, \partial y}$ saadaan

$$f(x,y) = \begin{cases} 1, & 0 \le x \le 1,\ 0 \le y \le 1; \\ 0 & \text{muualla.} \end{cases}$$

Satunnaisvektori (X,Y) noudattaa siis kaksiulotteista tasajakaumaa $\mathrm{Tas}[(0,1)\times(0,1)]$. Todennäköisyys voidaan lausua kertymäfunktion avulla seuraavasti:

$$\begin{aligned} &P(x_1 \le X \le x_2,\ y_1 \le Y \le y_2) \\ &\qquad = \int_{x_1}^{x_2}\int_{y_1}^{y_2} \mathrm{d}y\, \mathrm{d}x = \Big/_{x_1}^{x_2} \Big/_{y_1}^{y_2} xy = (x_2 - x_1)(y_2 - y_1) \\ &\qquad = x_2 y_2 - x_2 y_1 - x_1 y_2 + x_1 y_1 \\ &\qquad = F(x_2,y_2) - F(x_2,y_1) - F(x_1,y_2) + F(x_1,y_1). \end{aligned}$$

Yleisesti pitää paikkansa, että

$$\begin{aligned} &P(x_1 \le X \le x_2,\ y_1 \le Y \le y_2) \\ &\qquad = F(x_2,y_2) - F(x_2,y_1) - F(x_1,y_2) + F(x_1,y_1). \end{aligned}$$

Kahden muuttujan tasajakauman $\mathrm{Tas}[(0,1)\times(0,1)]$ tapauksessa todennäköisyys $P\big(\frac{1}{4} \le X \le \frac{1}{2},\ \frac{1}{2} \le Y \le \frac{3}{4}\big)$ on

$$\begin{aligned} &F\left(\frac{1}{2}, \frac{3}{4}\right) - F\left(\frac{1}{2}, \frac{1}{2}\right) - F\left(\frac{1}{4}, \frac{3}{4}\right) + F\left(\frac{1}{4}, \frac{1}{2}\right) \\ &\qquad = \frac{1}{2}\cdot\frac{3}{4} - \frac{1}{2}\cdot\frac{1}{2} - \frac{1}{4}\cdot\frac{3}{4} + \frac{1}{4}\cdot\frac{1}{2} = \frac{1}{16}. \end{aligned}$$

□

5.7.2 Yhteisjakauman momenttifunktio

Kaksiulotteisen diskreetin satunnaisvektorin momenttifunktio määriteltiin alaluvussa 4.7.2. Jatkuvien satunnaismuuttujien X_1 ja X_2 yhteisjakauman eli jatkuvan satunnaisvektorin (X_1, X_2) jakauman momenttifunktio määritellään samalla tavalla kuin diskreetissä tapauksessa. Olkoon (X_1, X_2) jatkuva satunnaisvektori ja $t_1 X_1 + t_2 X_2$ satunnaismuuttujien X_1 ja X_2 lineaarinen yhdiste, missä $t_1, t_2 \in \mathbb{R}$. Satunnaisvektorin (X_1, X_2) jakauman momenttifunktio on

$$M(t_1, t_2) = E\big(\mathrm{e}^{t_1 X_1 + t_2 X_2}\big).$$

Jatkuvien satunnaismuuttujien tapauksessa odotusarvon lauseke on muotoa

$$E\big(\mathrm{e}^{t_1 X + t_2 X_2}\big) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{t_1 x_1 + t_2 x_2} f(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2.$$

Merkitään

$$
\begin{aligned}
M_i(t_1,t_2) &= \frac{\partial M(t_1,t_2)}{\partial t_i},\\
M_{ii}(t_1,t_2) &= \frac{\partial^2 M(t_1,t_2)}{\partial t_i^2},\\
M_{ij}(t_1,t_2) &= \frac{\partial^2 M(t_1,t_2)}{\partial t_i\,\partial t_j},
\end{aligned}
$$

missä $M_i(t_1,t_2)$ on M:n osittaisderivaatta t_i:n suhteen, $M_{ii}(t_1,t_2)$ on M:n 2. osittaisderivaatta t_i:n suhteen ja $M_{ij}(t_1,t_2)$ on osittaisderivaatta t_i:n ja t_j:n suhteen ($i=1,2$; $j=1,2$). Esitämme nyt seuraavassa lauseessa, miten momenttifunktio generoi satunnaisvektorin momentit.

Lause 5.9 *Oletetaan, että satunnaisvektorilla* (X_1,X_2) *on momenttifunktio. Silloin* $E(X_i)$, $E(X_i^2)$ *ja* $E(X_iX_j)$ *ovat äärelliset ja*

$$
E(X_i) = M_i(0,0), \qquad E(X_i^2) = M_{ii}(0,0), \qquad E(X_iX_j) = M_{ij}(0,0)
$$

kaikilla $i=1,2$ *ja* $j=1,2$.

Esimerkiksi X_1:n odotusarvo saadaan derivoimalla ensin momenttifunktio t_1:n suhteen ja sijoittamalla sitten derivaatan lausekkeeseen $t_1=0$ ja $t_2=0$. Sekamomentti $E(X_1X_2)$ saadaan määrittämällä toisen kertaluvun osittaisderivaatta $M_{12}(t_1,t_2)$ (derivoidaan momenttifunktio t_2:n ja t_1:n suhteen) ja laskemalla osottaisderivaatan arvo $M_{ij}(0,0)$ pisteessä $(t_1,t_2)=(0,0)$.

Esimerkki 5.22 Jos Z_1 ja Z_2 ovat riippumattomat ja noudattavat standardimuotoista normaalijakaumaa, niin (Z_1,Z_2) noudattaa kaksiulotteista standardimuotoista normaalijakaumaa. (Z_1,Z_2):n momenttifunktio on

$$
\begin{aligned}
M(t_1,t_2) &= E\big(\mathrm{e}^{t_1Z_1+t_2Z_2}\big) = E(\mathrm{e}^{t_1Z_1})\,E(\mathrm{e}^{t_2Z_2})\\
&= \mathrm{e}^{t_1^2/2}\mathrm{e}^{t_2^2/2} = \mathrm{e}^{(t_1^2+t_2^2)/2}.
\end{aligned}
$$

Tässä tapauksessa $M_1(t_1,t_2) = t_1\mathrm{e}^{(t_1^2+t_2^2)/2}$, joten $E(X_1) = M_1(0,0) = 0$. Vastaavasti $M_{11} = \mathrm{e}^{(t_1^2+t_2^2)/2} + t_1\mathrm{e}^{(t_1^2+t_2^2)/2}$ ja $E(X_1^2) = M_{11}(0,0) = 1$. □

Huomattakoon, että myös satunnaisvektoreiden tapauksessa pätee momenttifunktioden yksikäsitteisyyttä koskeva lause (vrt. Lause 3.12). Jos siis satunnaisvektoreilla (X_1,X_2) ja (Y_1,Y_2) on sama momenttifunktio, niin niillä on sama jakauma. Reunajakaumien momenttifunktiot saadaan kätevästi yhteisjakuman momenttifunktiosta.

Lause 5.10 *Oletetaan, että satunnaisvektorin* (X,Y) *momenttifunktio on* $M(s,t)$ *sekä* X*:n ja* Y*:n momenttifunktiot vastaavasti* $M_X(s)$ *ja* $M_Y(t)$.

1. *Silloin*

$$
M_X(s) = M(s,0) \qquad \text{ja} \qquad M_Y(t) = M(0,t).
$$

2. X *ja* Y *ovat riippumattomat jos ja vain jos*

$$
M(s,t) = M_X(s)M_Y(t).
$$

5.8 Kahden muuttujan normaalijakauma

5.8.1 Standardimuoto

Oletetaan, että satunnaismuuttujat Z ja V ovat riippumattomat ja noudattavat standardimuotoista normaalijakaumaa. Silloin Z:n ja V:n riippumattomuuden nojalla satunnaisvektorin (Z, V) yhteisjakauman tiheysfunktio on

$$f_{Z,V}(z, v) = f(z)f(v) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-v^2/2} = \frac{1}{2\pi}\mathrm{e}^{-(z^2+v^2)/2}. \tag{5.8.1}$$

Sanomme, että satunnaisvektori (Z, V) noudattaa kaksiulotteista standardimuotoista normaalijakaumaa ja funktio (5.8.1) on tämän jakauman tiheysfunktio. Merkitään $(Z, V) \sim \mathrm{N}_2(\mathbf{0}, \boldsymbol{I})$, missä $\mathbf{0}$ on 2×1-nollavektori eli $\mathbf{0} = (0, 0)^\mathsf{T}$. Merkintä $(0, 0)^\mathsf{T}$ tarkoittaa vektorin $(0, 0)$ transponointia, joka muuntaa vaakavektorin $(0, 0)$ pystytoriksi. Matriisi

$$\boldsymbol{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

on 2×2-identiteettimatriisi. Yhteisjakauman reunajakaumien keskiarvot ovat $E(Z) = E(V) = 0$ ja varianssit $\mathrm{Var}(Z) = \mathrm{Var}(V) = 1$ sekä $\mathrm{Cov}(Z, V) = 0$. Satunnaisvektorin (Z, V) odotusarvovektori on $[E(Z), E(V)]^\mathsf{T} = \mathbf{0}$ ja kovarianssimatriisi

$$\begin{pmatrix} \mathrm{Var}(Z) & \mathrm{Cov}(Z, V) \\ \mathrm{Cov}(V, Z) & \mathrm{Var}(V) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Huomaa, että aina $\mathrm{Cov}(Z, V) = \mathrm{Cov}(V, Z)$, joten kovarianssimatriisi on symmetrinen. Voidaan merkitä myös $(Z, V) \sim \mathrm{N}_2(0, 0; 1, 1, 0)$, missä odotusarvot, varianssit ja korrelaatio on annettu sulkeissa.

5.8.2 Korreloivat muuttujat

Oletetaan, että $X \sim \mathrm{N}(0, 1)$ ja $Z \sim \mathrm{N}(0, 1)$ ovat riippumattomat. Niiden avulla voidaan konstruoida normaalijakaumaa noudattava satunnaismuuttuja Y siten, että X ja Y korreloivat. Kiertämällä x-akselia kulman θ verran vastapäivään saadaan y-akseli (Kuvio 5.10). Projisoidaan satunnaispiste (X, Z) y-akselille ja merkitään tätä projektiota Y:llä. On helppo todeta geometrisen päättelyn avulla (Kuvio 5.10), että

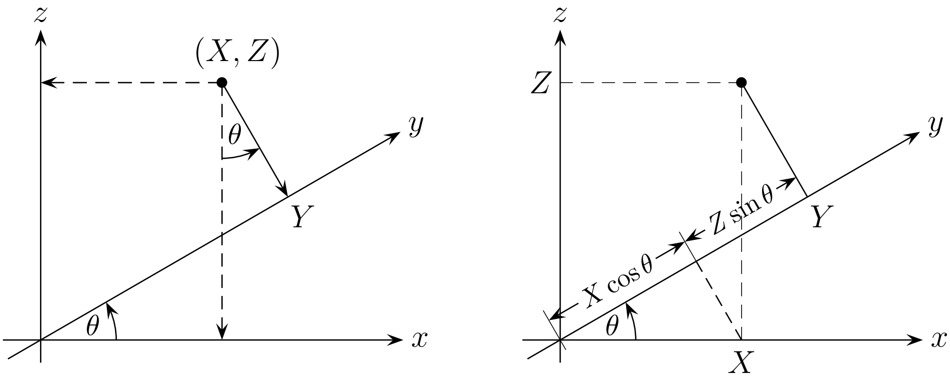
$$Y = X \cos\theta + Z \sin\theta.$$

Satunnaismuuttuja Y saadaan siis X:n ja Z:n lineaarisena muunoksena. Tästä seuraa, että

$$E(Y) = \cos\theta \cdot E(X) + \sin\theta \cdot E(Z) = 0$$

ja

$$\mathrm{Var}(Y) = \cos^2\theta \cdot \mathrm{Var}(X) + \sin^2\theta \cdot \mathrm{Var}(Z) = 1,$$



Kuvio 5.10.

koska $E(X) = E(Z) = 0$ ja $\mathrm{Var}(X) = \mathrm{Var}(Z) = 1$. Lauseen 5.4 mukaan $Y \sim \mathrm{N}(0,1)$. Satunnaisuuttujien X ja Y 2. kertaluvun sekamomentti on

$$\begin{aligned} E(XY) &= E[X(X\cos\theta + Z\sin\theta)] \\ &= \cos\theta \cdot E(X^2) + \sin\theta \cdot E(XZ) \\ &= \cos\theta. \end{aligned}$$

Viimeinen yhtäsuuruus seuraa siitä, että $E(X^2) = 1$ ja $E(XZ) = E(X)\,E(Z) = 0$. Satunnaisuuttujien X ja Y välinen korrelaatio $\mathrm{Cor}(X,Y) = E(X,Y)$, koska $E(X) = E(Y) = 0$ ja $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$.

5.9 Satunnaisvektoreiden muunnokset

Oletetaan, että jatkuvien satunnaismuuttujien X ja Y yhteisjakauman tiheysfunktio on f. Olkoon

(5.9.1) $$U = h_1(X,Y); \qquad V = h_2(X,Y)$$

sellainen satunnaisvektorin (X,Y) muunnos, että sillä on käänteismuunnos. Silloin mitä tahansa satunnaisvektorin (U,V) arvoa $(u,v) \in \mathbb{R}^2$ vastaa yksikäsitteinen satunnaisvektorin (X,Y) arvo $(x,y) \in \mathbb{R}^2$. Voimme silloin määritellä käänteiskuvauksen

$$x = g_1(u,v); \qquad y = g_2(u,v).$$

Vektoreiden $(u,v) \in \mathbb{R}^2$ ja $(x,y) \in \mathbb{R}^2$ välillä on yksi-yksinen vastaavuus. Oletetamme lisäksi, että funktioilla g_1 ja g_2 on jatkuvat osittaisderivaatat. Yksiulotteisen muunnoksen tapauksessa laskettavaa derivaattaa g' vastaa satunnaisvektorien muunnoksen *Jacobin determinantti,* joka on funktioiden g_1 ja g_2 osittaisderivaattojen matriisin determinantti. Jacobin determinanttia kutsutaan muunnoksen *Jakobiaaniksi.*

Muunnoksen (5.9.1) Jakobiaani on

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}, \tag{5.9.2}$$

missä

$$\frac{\partial x}{\partial u} = \frac{\partial g_1(u,v)}{\partial u}, \qquad \frac{\partial x}{\partial v} = \frac{\partial g_1(u,v)}{\partial v},$$
$$\frac{\partial y}{\partial u} = \frac{\partial g_2(u,v)}{\partial u}, \qquad \frac{\partial y}{\partial v} = \frac{\partial g_2(u,v)}{\partial v}.$$

Jacobin determinanttia merkitään $J = \frac{\partial(x,y)}{\partial(u,v)}$. Oletetaan, että $J \neq 0$, kun $f(x,y) > 0$. Satunnaisvektorin (U,V) yhteisjakauman tiheysfunktio on

$$f_{U,V}(u,v) = f\big(g_1(u,v), g_2(u,v)\big)|J|. \tag{5.9.3}$$

Satunnaisvektorin (U,V) arvoavaruus $S_{U,V}$ saadaan tarkastelemalla kuvausta (5.9.1), joka kuvaa satunnaisvektorin (X,Y) arvojoukon $S_{X,Y}$ kuvajoukoksi $S_{U,V}$.

Esimerkki 5.23 Olkoot X ja Y jatkuvat satunnaismuuttujat, joiden yhteisjakauman tiheysfunktio on f. Määritellään satunnaismuuttujat U ja V siten, että

$$U = X + Y; \qquad V = X - Y. \tag{5.9.4}$$

Johdetaan nyt (U,V):n jakauman tiheysfunktio. Muunnoksen (5.9.4) käänteismuunnos on

$$x = g_1(u,v) = \frac{u+v}{2}, \qquad y = g_2(u,v) = \frac{u-v}{2},$$

ja muunnoksen Jakobiaani on

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}.$$

Satunnaisvektorin (U,V) yhteisjakauman tiheysfunktio on yhtälön (5.9.3) nojalla

$$f_{U,V}(u,v) = \frac{1}{2} f\left(\frac{u+v}{2}, \frac{u-v}{2}\right). \tag{5.9.5}$$

Jos esimerkiksi X ja Y ovat riippumattomat ja noudattavat tasajakaumaa Tas$(0,1)$, niin (X,Y):n yhteisjakauman tiheysfunktio $f(x,y) = 1$, kun $x \in [0,1]$ ja $y \in [0,1]$. Silloin (U,V):n tiheysfunktio on

$$f_{U,V}(u,v) = \begin{cases} \frac{1}{2}, & 0 \le u+v \le 2,\ 0 \le u-v \le 2. \\ 0 & \text{muualla.} \end{cases}$$

□

Yleinen muunnos

Tarkasteltavalla muunnoksella ei tietenkään aina ole käänteismuunnosta. Jos muunnos (5.9.1) ei ole yksi-yksinen eli bijektio, niin sillä ei ole käänteismuunnosta. Jos kuitenkin on olemassa sellainen arvoavaruuden $S_{X,Y}$ ositus yhteispisteettömiin (x, y)-tason osaväleihin A_1, A_2, ..., A_m, että

$$S_{X,Y} = A_1 \cup A_2 \cup \cdots \cup A_m \tag{5.9.6}$$

ja muunnoksella

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

on käänteismuunnos

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

jokaisella osavälillä A_i, $i = 1, 2, \ldots, m$, niin kaavaa (5.9.3) voidaan soveltaa kullakin osavälillä erikseen vastaavalla tavalla kuin yhden muuttujan tapauksessa. Määritellään funktiot

$$h_{ki}(x, y) = \begin{cases} h_k(x, y), & \text{kun } (x, y) \in A_i; \\ 0 & \text{muualla,} \end{cases}$$

kun $k = 1, 2$. Silloin $h_1(x, y) = \sum_{i=1}^m h_{1i}(x, y)$ ja $h_2(x, y) = \sum_{i=1}^m h_{2i}(x, y)$. Jokaisella muunnoksella

$$u = h_{1i}(x, y), \qquad v = h_{2i}(x, y)$$

on käänteismuunnos välillä A_i, $i = 1, 2, \ldots, m$. Merkitään näitä käänteismuunnoksia

$$x = g_{1i}(u, v), \qquad y = g_{2i}(u, v).$$

Satunnaisvektorin (U, V) tiheysfunktio voidaan silloin esittää kaavan (5.9.3) avulla seuraavasti

$$f_{U,V}(u, v) = \sum_{i=1}^{m} f_{X,Y}\big(g_{1i}(u, v), g_{2i}(u, v)\big)|J_i|, \tag{5.9.7}$$

missä J_i on muunnoksen $x = g_{1i}(u, v)$, $y = g_{2i}(u, v)$ Jakobiaani.

Satunnaismuuttujien funktion jakauma

Usein tarkasteltavana on vain yksi satunnaismuuttujien X ja Y funktio $U = h_1(X, Y)$. Funktio $h_1(X, Y)$ voi olla esimerkiksi muotoa $X + Y$, XY, $X^2 + Y^2$ jne. Esitettyä muunnostekniikkaa voidaan edelleen käyttää, jos löydetään sellainen apumuuttuja $V = h_2(X, Y)$, että muunnoksella

$$U = h_1(X, Y), \qquad V = h_2(X, Y)$$

on käänteismuunnos. Muunnoskaavan (5.9.3) avulla saadaan sitten satunnaisvektorin (U,V) yhteisjakauman tiheysfunktio, josta voidaan määrittää U:n reunajakauman tiheysfunktio. Jos esimerkiksi $U = h_1(X,Y) = X + Y$, niin voidaan valita apumuuttuja $V = h_2(X,Y) = X - Y$. Silloin muunnoksella $u = x + y$, $v = x - y$ on käänteismuunnos

$$x = g_1(u,v) = \frac{u+v}{2}, \qquad y = g_2(u,v) = \frac{u-v}{2}$$

ja (U,V):n tiheysfunktio saadaan kaavalla (5.9.3), niinkuin Esimerkissä 5.23 osoitettiin. Kun tästä (U,V):n tiheysfunktiosta ”integroidaan v pois”, saadaan U:n tiheysfunktio. Jos ollaan kiinnostuneita esimerkiksi satunnaismuuttujan $U = XY$ tiheysfunktiosta, voidaan valita apumuuttuja $V = X$, sillä muunnoksella $u = xy$, $v = x$ on käänteismuunnos. Sitten sovelletaan jälleen edellä kuvattua tekniikkaa. Huomaa, että apumuuttujan valinta ei ole yksikäsitteinen, vaan useilla eri valinnoilla voidaan päästä haluttuun tulokseen.

Tässä yhteydessä on syytä palauttaa mieleen Lause 3.6. Siinä osoitettiin riippumattomille satunnaismuuttujille X ja Y seuraava tulos: Jos $g(X)$ ei riipu Y:stä ja $h(Y)$ ei riipu X:stä, niin silloin satunnaismuuttujat $g(X)$ ja $h(Y)$ ovat riippumattomat. Lause todistettiin diskreettien satunnaismuuttujien tapauksessa, mutta se pitää paikkansa myös jatkuville muuttujille.

Esimerkki 5.24 Jos X ja Y ovat riippumattomat ja noudattavat standardimuotoista normaalijakumaa, niin mitä jakaumaa noudattaa $X + Y$? Määritelään ensin apumuuttuja $V = X - Y$. Kuten esimerkissä 5.23 osoitettiin, muunnoksen $u = x + y$, $v = x - y$ käänteismuunnos on

$$x = g_1(u,v) = \frac{u+v}{2}, \qquad y = g_2(u,v) = \frac{u-v}{2},$$

ja muunnoksen Jakobiaani $J = -\frac{1}{2}$. Yhtälön (5.9.5) perusteella (U,V):n tiheysfunktio on

$$\begin{aligned} f_{U,V}(u,v) &= \frac{1}{4\pi}\mathrm{e}^{-[(u+v)^2/8+(u-v)^2/8]} \\ &= \frac{1}{4\pi}\mathrm{e}^{-(u^2+v^2)/4} = \frac{1}{\sqrt{4\pi}}\mathrm{e}^{-u^2/4}\frac{1}{\sqrt{4\pi}}\mathrm{e}^{-v^2/4} \\ &= f_U(u)f_V(v). \end{aligned} \tag{5.9.8}$$

Näemme, että $f_U(u)$ ja $f_V(v)$ ovat kumpikin normaalijakauman $\mathrm{N}(0,2)$ tiheysfunktioita. Siksi

$$\int\limits_{-\infty}^{\infty} f_{U,V}(u,v)\,\mathrm{d}v = f_U(u)\int\limits_{-\infty}^{\infty} f_V(v)\,\mathrm{d}v = f_U(u),$$

joten $U = X + Y \sim \mathrm{N}(0,2)$. Identiteetistä (5.9.8) seuraa, että $X + Y$ ja $X - Y$ ovat riippumattomat. Havaitsemme myös, että $X - Y \sim \mathrm{N}(0,2)$. Itse asiassa voidaan todistaa seuraava tulos: Jos X ja Y ovat riippumattomat ja noudattavat samaa jakaumaa F, niin $X+Y$ ja $X-Y$ ovat riippumattomat jos ja vain jos F on normaalijakauma. □

Esimerkki 5.25 Olkoot X ja Y jatkuvat satunnaismuuttujat, joiden yhteisjakauman tiheysfunktio on f. Määritellään lineaarinen muunnos

$$\begin{aligned} u &= ax + by, \\ v &= cx + dy. \end{aligned} \tag{5.9.9}$$

Ratkaisemalla yhtälöryhmästä (5.9.9) x ja y saadaan käänteismuunnos

$$\begin{aligned} x &= \frac{du - bv}{D}, \\ y &= \frac{av - cu}{D}, \end{aligned}$$

missä $D = ad - bc$. Käänteismuunnos on olemassa, jos $D \neq 0$. Muunnoksen Jakobiaani on

$$J = \begin{vmatrix} \frac{d}{D} & \frac{-b}{D} \\ \frac{-c}{D} & \frac{a}{D} \end{vmatrix} = \frac{ad - bc}{D^2} = \frac{1}{D}.$$

Satunnaisvektorin (U, V) yhteisjakauman tiheysfunktio on yhtälön (5.9.3) nojalla

$$f_{U,V}(u, v) = \frac{1}{|D|} f[(du - bv)/D, (av - cu)/D]. \tag{5.9.10}$$

Esimerkin 5.23 yhtälö (5.9.5) on yhtälön (5.9.10) erikoistapaus. Kun $a = b = c = 1$ ja $d = -1$ sijoitetaan yhtälöön (5.9.10), saadaan yhtälö (5.9.5). Lineaarinen muunnos

$$\begin{aligned} u &= ax + by + e, \\ v &= cx + dy + f. \end{aligned}$$

voidaan palauttaa muunnokseen (5.9.9) merkitsemällä $u^* = u - e$ ja $v^* = v - f$, jolloin

$$\begin{aligned} u^* &= ax + by, \\ v^* &= cx + dy. \end{aligned}$$

Yhtälöstä (5.9.10) saadaan sitten (U^*, V^*):n tiheysfunktio. □

5.9.1 Yleinen kahden muuttujan normaalijakauma

Standardimuotoinen kaksiulotteinen normaalijakauma määriteltiin alaluvussa 5.8.1. Yleinen kaksiulotteinen normaalijakauma voidaan määritellä standardimuotoisen normaalijakauman avulla vastaavasti kuin yhden muuttujan tapauksessa. Olkoon (X, Y) sellainen satunnaisvektori, että $E(X) = \mu_1$,

$E(Y) = \mu_2$, $\mathrm{Var}(X) = \sigma_1^2$, $\mathrm{Var}(Y) = \sigma_2^2$ ja $\mathrm{Cov}(X,Y) = \sigma_{12}$. Silloin satunnaisvektorin (X,Y) keskiarvovektori on $\boldsymbol{\mu} = (\mu_1, \mu_2)^\mathsf{T}$ ja kovarianssimatriisi

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

Satunnaismuuttujien X ja Y välinen korrelaatiokerroin on $\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}$.

Määritelmä 5.7 Satunnaisvektori (X,Y) noudattaa normaalijakaumaa, jonka keskiarvovektori on $\boldsymbol{\mu}$ ja kovarianssimatriisi $\boldsymbol{\Sigma}$, jos se voidaan lausua muodossa

$$\begin{aligned} X - \mu_1 &= \sigma_1\sqrt{1-\rho^2}Z_1 + \rho\sigma_1 Z_2, \\ Y - \mu_2 &= \sigma_2 Z_2, \end{aligned} \tag{5.9.11}$$

missä Z_1 ja Z_2 ovat riippumattomat ja noudattavat standardimuotoista normaalijakaumaa $\mathrm{N}(0,1)$ sekä $\sigma_1 > 0$, $\sigma_2 > 0$ ja $|\rho| \le 1$.

Merkitsemme $(X,Y) \sim \mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, kun (X,Y) noudattaa normaalijakaumaa, jonka keskiarvovektori on $\boldsymbol{\mu}$ ja kovarianssimatriisi $\boldsymbol{\Sigma}$. Normaalijakaumaa $\mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ noudattava satunnaisvektori (X,Y) saadaan aina riippumattomista standardoiduista normaalimuuttujista lineaarisella muunnoksella (5.9.11)

Yleisen normaalijakauman tiheysfunktio saadaan suoraan Esimerkissä 5.25 esitetyllä tekniikalla. Merkitään $x - \mu_1 = u$ ja $y - \mu_2 = v$. Koska nyt lineaarisessa muunnoksessa (5.9.9) $c = 0$ ja $D = ad$, saa yhtälö (5.9.10) muodon

$$f_{U,V}(u,v) = \frac{1}{ad} f\left(\frac{du - bv}{ad}, \frac{v}{d}\right).$$

Koska f on kahden muuttujan standardimuotoisen normaalijakauman tiheysfunktio ja muunnoksessa (5.9.11) $a = \sigma_1\sqrt{1-\rho^2}$, $b = \rho\sigma_1$ ja $d = \sigma_2$, niin

$$\begin{aligned} f_{U,V}(u,v) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2\sigma_1^2\sigma_2^2(1-\rho^2)}(\sigma_2 u - \rho\sigma_1 v)^2 + \frac{v^2}{\sigma_2^2}\right] \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{u^2}{\sigma_1^2} - 2\rho\frac{u}{\sigma_1}\frac{v}{\sigma_2} + \frac{v^2}{\sigma_2^2}\right)\right]. \end{aligned}$$

Kun edellä johdettuun tiheysfunktioon sijoitetaan $u = x - \mu_1$ ja $v = y - \mu_2$, saadaan (X,Y):n tiheysfunktio

$$\begin{aligned} f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Bigg[-\frac{1}{2(1-\rho^2)}\Bigg(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 \\ - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\Bigg)\Bigg]. \end{aligned} \tag{5.9.12}$$

Seuraavassa lauseessa esitetään kaksiulotteisen normaalijakauman keskeiset ominaisuudet.

Lause 5.11 *Oleteaan, että satunnaisvektori* (X, Y) *noudattaa kaksiulotteista normaalijakaumaa* $\mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *missä* $\boldsymbol{\mu} = (\mu_1, \mu_2)^\mathsf{T} = [E(X), E(Y)]^\mathsf{T}$ *ja*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X, Y) \\ \mathrm{Cov}(Y, X) & \mathrm{Var}(Y) \end{pmatrix}$$

ja $\rho = \mathrm{Cor}(X, Y) = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$. *Silloin pitävät paikkansa seuraavat ominaisuudet:*

1. $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ *ja* $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$,

2. X *ja* Y *ovat riippumattomat jos ja vain jos* $\rho = 0$.

3. X*:n ja* Y*:n ehdolliset jakaumat:*

$$X \mid y \sim \mathrm{N}\bigg(\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(y - \mu_2),\ \sigma_1^2(1 - \rho^2)\bigg),$$

eli X *noudattaa normaalijakaumaa ehdolla, että* $Y = y$ *on annettu. Vastaavasti*

$$Y \mid x \sim \mathrm{N}\bigg(\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x - \mu_1),\ \sigma_2^2(1 - \rho^2)\bigg).$$

4. *Satunnaisvektorin* (X, Y) *momenttifunktio on*

$$M(s, t) = \exp\bigg(\mu_1 s + \mu_2 t + \frac{\sigma_1^2 s^2 + \sigma_2^2 t^2 + \rho\sigma_1\sigma_2 st}{2}\bigg).$$

Lause 5.12 *Satunnaisvektori* (X, Y) *noudattaa kaksiulotteista normaalijakaumaa, jos ja vain jos satunnaismuuttujien* X *ja* Y *lineaarikombinaatiot* $aX+bY$ *noudattavat yksiulotteista normaalijakaumaa kaikilla* $a \in \mathbb{R}$ *ja* $b \in \mathbb{R}$.

5.9.2 Studentin t-jakauma, F-jakauma ja beta-jakauma

Oletetaan, että satunnaismuuttujat Z ja U ovat riippumattomat, $Z \sim \mathrm{N}(0, 1)$ ja U noudattaa χ^2-jakaumaa vapausastein r eli $U \sim \mathrm{Khi2}(r)$. Tarkastellaan nyt satunnaismuuttujan

$$T = \frac{Z}{\sqrt{U/r}} \tag{5.9.13}$$

jakaumaa. Tätä muotoa olevalla satunnaismuuttujalla on erittäin keskeinen rooli tilastollisessa päättelyssä. Satunnaismuuttuja (5.9.13) noudattaa ns. *Studentin t-jakaumaa* (tai lyhyesti t-jakaumaa) vapausastein r. Jakauma on nimetty englantilaisen tilastotieteilijän W. S. Gossetin mukaan. Gosset esitti tämän jakauman *Biometrikassa* vuonna 1908 nimimerkillä "Student". Kun T noudattaa Studentin t-jakaumaa vapausastein r, merkitään $T \sim \mathrm{t}(r)$.

Olkoon $X_1, X_2, \ldots, X_n$ on otos normaalijakaumasta $\mathrm{N}(\mu, \sigma^2)$. Silloin siis satunnaismuuttujat $X_1, X_2, \ldots, X_n$ ovat toisistaan riippumattomat ja $X_i \sim\mathrm{N}(\mu, \sigma^2)$, $1 \le i \le n$. Otoksesta laskettu suure

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{(\overline{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{S^2/\sigma^2}} \tag{5.9.14}$$

eli otossuure noudattaa t-jakaumaa vapausastein $n - 1$, missä $\overline{X}$ on otoskeskiarvo ja

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

on otosvarianssi. Lausekkeen (5.9.14) osoittaja $(\overline{X} - \mu)/(\sigma/\sqrt{n})$ noudattaa normaalijakaumaa $\mathrm{N}(0, 1)$ ja satunnaismuuttuja S^2/σ^2 noudattaa jakaumaa $\mathrm{Khi2}(n - 1)$. Koska $\overline{X}$ ja S^2 ovat toisistaan riippumattomat, niin lausekkeen (5.9.14) osoittaja ja nimittäjä ovat toisistaan riippumattomat.

Satunnaismuuttujan (5.9.13) eli Studentin t-jakauman tiheysfunktio vapausastein r on

$$f_T(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)} \frac{1}{\sqrt{r\pi}} \frac{1}{(1 + t^2/r)^{(r+1)/2}}, \qquad -\infty < t < \infty. \tag{5.9.15}$$

Huomaa, että vapausastein $r = 1$ tiheysfunktiosta (5.9.15) tulee Cauchyn jakauman tiheysfunktio.

Tiheysfunktio (5.9.15) saadaan suoraviivaisesti edellä esitetyllä muunnostekniikalla. Lähdetään liikkeelle satunnaisvektorin (Z, U) yhteisjakaumasta. Koska Z ja U ovat riippumattomat, niin

$$f_{Z,U}(z, u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \frac{1}{\Gamma(r/2) 2^{r/2}} u^{r/2-1} e^{-u/2}, \ -\infty < z < \infty, \ 0 < v < \infty.$$

Tehdään muunnos

$$t = \frac{z}{\sqrt{u/r}}, \qquad w = u,$$

jonka käänteismuunnos on

$$z = t\sqrt{w/r}, \qquad u = w.$$

Muunnoksen jakobiaani on $\sqrt{w/r}$. Sen jälkeen voidaan soveltaa muunnoskaavaa (5.9.3). T:n tiheysfunktio saadaan (T, W):n yhteisjakauman tiheysfunktiosta integroimalla w:n yli:

$$f_T(t) = \int_0^\infty f_{Z,U}\big[t(w/r)^{1/2}, w\big] (w/r)^{1/2} \,\mathrm{d}w.$$

Laskennan lopputuloksena on t-jakauman tiheysfunktio (5.9.15). Yksityiskohdat jätetään lukijan tehtäväksi.

Studentin t-jakaumalla ei ole momenttifunktiota, koska sillä ei ole kaikkien kertalukujen momentteja. Jos $T_r \sim \mathrm{t}(r)$, niin silloin T_r:llä on ainoastaan $r-1$ ensimmäistä momenttia. Esimerkiksi jakaumalla $\mathrm{t}(1)$ ei ole keskiarvoa ja jakaumalla $\mathrm{t}(2)$ ei ole varianssia. Voidaan laskemalla osoittaa, että

$$E(T_r) = 0, \quad \text{jos } r > 1, \qquad \mathrm{Var}(T_r) = \frac{r}{r-2}, \quad \text{jos } r > 2. \tag{5.9.16}$$

Toinen tilastollisessa päättelyssä keskeinen jakauma, *Snedecorin F-jakauma* tai vain lyhyesti F-jakauma, voidaan johtaa t-jakauman tapaan. Jos $X_1, X_2, \ldots, X_n$ on otos normaalijakaumasta $\mathrm{N}(\mu_1, \sigma_1^2)$ ja $X_1, X_2, \ldots, X_m$ otos normaalijakaumasta $\mathrm{N}(\mu_2, \sigma_2^2)$, niin silloin satunnaismuuttuja

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \tag{5.9.17}$$

noudattaa F-jakaumaa vapausastein $n-1$ ja $m-1$. Lausekkeessa (5.9.17) S_1^2 ja S_2^2 ovat otosvarianssit. S_1^2/σ_1^2 ja S_2^2/σ_2^2 ovat riippumattomat ja

$$(n-1)S_1^2/\sigma_1^2 \sim \mathrm{Khi2}(n-1), \qquad (m-1)S_2^2/\sigma_2^2 \sim \mathrm{Khi2}(m-1).$$

F-jakauma määritellään seuraavasti: Jos $X \sim \mathrm{Khi2}(r)$ ja $Y \sim \mathrm{Khi2}(s)$ ovat riippumattomat, niin satunnaismuuttuja

$$F = \frac{X/r}{Y/s} \tag{5.9.18}$$

noudattaa F-jakaumaa vapausastein r ja s. Silloin merkitään $F \sim \mathrm{F}(r, s)$. F-jakauman tiheysfunktio voidaan johtaa vastaavalla tavalla kuin t-jakauman tiheysfunktio. Määritellään muunnos

$$U = X + Y, \qquad V = \frac{X/r}{Y/s},$$

missä F-suuretta (5.9.18) on merkitty kirjaimella V. Muunnoksen käänteismuunnos on

$$X = \frac{r}{s}\Big(1 + \frac{r}{s}V\Big)^{-1} UV, \qquad Y = \Big(1 + \frac{r}{s}V\Big)^{-1} U.$$

Koska X ja Y ovat riippumattomat, niin

$$f_{X,Y}(X, Y) = k_r k_s x^{(r/2)-1} y^{(s/2)-1} \mathrm{e}^{-(x+y)/2}, \qquad 0 < x < \infty, \quad 0 < y < \infty,$$

missä $k_r = 1/\Gamma\big(\frac{r}{2}\big)2^{r/2}$ ja $k_s = 1/\Gamma\big(\frac{s}{2}\big)2^{s/2}$. Kun lasketaan Jakobiaani ja tehdään sijoitukset, saadaan satunnaisvektorin (U, V) tiheysfunktio $f_{U,V}(u, v)$ kaavan (5.9.3) mukaisesti. V:n reunajakauman tiheysfuntio on sitten F-jakauman tiheysfuntio:

$$f_F(v) = \frac{\Gamma\big(\frac{r+s}{2}\big)}{\Gamma\big(\frac{r}{2}\big)\Gamma\big(\frac{s}{2}\big)} \Big(\frac{r}{s}\Big)^{r/2} \frac{v^{(r/2)-1}}{\big[1 + \frac{r}{s}v\big]^{(r+s)/2}}, \qquad 0 < v < \infty. \tag{5.9.19}$$

F-jakauman keskiarvo ja varianssi ovat

$$E(F) = \frac{s}{s-2}, \qquad s > 2 \tag{5.9.20}$$

$$\operatorname{Var}(F) = \frac{2s^2(r+s-2)}{r(s-2)^2(s-4)}, \qquad s > 4. \tag{5.9.21}$$

Beta-jakauma

Olkoot $X \sim \text{Gamma}(\alpha, \theta)$ ja $Y \sim \text{Gamma}(\beta, \theta)$ riippumattomat gammajakaumaa noudattavat satunnaismuuttujat. Silloin satunnaisvektorin (X, Y) tiheysfunktio on

$$f(x,y) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)\theta^{\alpha+\beta}} x^{\alpha-1} y^{\beta-1} \exp\Big(-\frac{x+y}{\theta}\Big), \qquad 0 < x, \quad y < \infty.$$

Tarkastellaan muunnosta

$$U = \frac{X}{X+Y}, \qquad V = X + Y,$$

jonka käänteismuunnos on

$$X = UV, \qquad Y = V - UV.$$

Muunnoksen Jakobiaani on

$$\begin{vmatrix} v & u \\ -v & 1-u \end{vmatrix} = v(1-u) + uv = v.$$

Satunnaisvektorin (U, V) tiheysfunktio on kaavan (5.9.3) mukaan

$$f_{U,V}(u,v) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)} (uv)^{\alpha-1} (v - uv)^{\beta-1} \mathrm{e}^{-v\theta} v,$$

missä $0 < u < 1$ ja $0 < v < \infty$. Kun tästä (U,V):n yhteisjakauman tiheysfunktiosta määritetään U:n reunajakuman tiheysfunktio, saadaan

$$f_U(u) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1} (1-u)^{\beta-1}, \qquad 0 < u < 1. \tag{5.9.22}$$

Sanomme, että U noudattaa betajakaumaa parametrein α ja β. Silloin merkitsemme $U \sim \text{Beta}(\alpha, \beta)$. Koska (5.9.22) on tiheysfunktio, niin

$$\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,\mathrm{d}u.$$

Funktiota

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{5.9.23}$$

kutsutaan *betafunktioksi.*

Betajakauma on yksi niitä harvoja nimettyjä jakaumia, joiden koko todennäköisyysmassa on äärellisellä valillä, eli välillä $(0,1)$. Betajakauman momentit on helppo laskea tiheysfunktion erityisominaisuuksien avulla. Kun $n > -\alpha$, niin

$$E(X^n) = \frac{B(\alpha+n,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+n)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)\Gamma(\alpha)}. \tag{5.9.24}$$

Sijoittamalla $n = 1$ ja $n = 2$ kaavaan (5.9.24) saadaan 1. ja 2. momentti ja niiden avulla varianssi:

$$E(X) = \frac{\alpha}{\alpha+\beta} \qquad \text{ja} \qquad \mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{5.9.25}$$

Esimerkki 5.26 Oletetaan, että on suoritettavana $n+m$ työtä. Töiden suorittamiseen tarvittavat ajat noudattavat toisistaan riippumatta eksponettijakaumaa keskiarvolla $\theta > 0$ (ts. gammajakaumaa parametrein $\alpha = 1$ ja θ.) Oletetaan, että kaksi eri työntekijää tekee nämä työt siten, että työntekijä A tekee työt $1, 2, \ldots, n$ ja B tekee työt $n+1, n+2, \ldots, n+m$. Jos merkitään X:llä työntekijän A käyttämää aikaa ja Y:llä työntekijän B käyttämää aikaa, niin toisistaan riippumatta $X \sim \mathrm{Gamma}(n,\theta)$ ja $Y \sim \mathrm{Gamma}(m,\theta)$. Silloin $(n+m)$:n työn vaatima kokonaisaika $X+Y$ noudattaa gammajakaumaa $\mathrm{Gamma}(n+m,\theta)$. Työntekijän A käyttämä suhteellinen osuus $X/(X+Y)$ kokonaisajasta noudattaa betajakaumaa $\mathrm{Beta}(n,m)$. □

5.9.3 Hierarkkiset mallit ja yhdistetyt jakaumat

Tarkastellaan aluksi esimerkkinä tietyn laitteen, esimerkiksi kopiokoneen, rikkoontumista. Oletetaan, että rikkoontumisten lukumäärä X vuodessa noudattaa Poissonin jakaumaa parametrilla λ. Kun laite on rikkoontunut,sen korjaamiseen tarvittava aika noudattaa eksponettijakaumaa keskiarvolla $\theta > 0$. Olkoon Y_i aika, joka tarvitaan i. rikkoontumisen korjaamiseen. Oletetaan lisäksi, että korjausajat eri kerroilla ovat riippumattomat. Jos vuodessa sattuu $X = x$ rikkoontumista, niin kokonaiskorjausaika on

$$Y = Y_1 + Y_2 + \cdots + Y_x.$$

Silloin

$$E(Y \mid x) = E(Y_1) + E(Y_2) + \cdots + E(Y_x) = x\theta$$

ja

$$\mathrm{Var}(Y \mid x) = \mathrm{Var}(Y_1) + \mathrm{Var}(Y_2) + \cdots + \mathrm{Var}(Y_x) = x\theta^2.$$

Edellä lasketut keskiarvo ja varianssi ovat siis ehdollisia ehdolla $X = x$. Ehdollinen keskiarvo ja varianssi

$$\mu(x) = E(Y \mid x) = x\theta \qquad \text{ja} \qquad \sigma^2(x) = \mathrm{Var}(Y \mid x) = x\theta^2$$

ovat x:n funktioita. Koska $X \sim \mathrm{Poi}(\lambda)$ on satunnaismuuttuja, niin myös

$$\mu(X) = E(Y \mid X) = X\theta \qquad \text{ja} \qquad \sigma^2(X) = \mathrm{Var}(Y \mid X) = X\theta^2$$

ovat satunnaismuuttujia. Silloin

$$\begin{aligned} E[\mu(X)] &= E[E(Y \mid X)] = \theta\, E(X) = \theta\lambda, \\ E[\sigma^2(X)] &= E[\mathrm{Var}(Y \mid X)] = \theta^2\, E(X) = \theta^2\lambda, \\ \mathrm{Var}[\mu(X)] &= \mathrm{Var}[E(Y \mid X)] = \theta\, \mathrm{Var}(X) = \theta\lambda. \end{aligned}$$

Määritimme siis edellä kaksinkertaisen odotusarvon $E[E(Y \mid X)]$, missä sisimmäinen odotusarvo on otettu Y:n suhteen ja ulompi X:n suhteen.

Esitämme nyt kaksinkertaisia odotusarvoja koskevan lauseen, joka usein helpottaa huomattavasti odotusarvojen laskemista.

Lause 5.13 *Olkoot X ja Y mitkä tahansa kaksi satunnaismuuttujaa. Silloin*

$$E(Y) = E[E(Y \mid X)], \tag{5.9.26}$$

jos odotusarvot ovat olemassa.

Todistus. Olkoon $f(x, y)$ satunnaisvektorin (X, Y) tiheysfunktio. Määritelmän mukaan

$$E(Y) = \iint y f(x, y)\,\mathrm{d}y\,\mathrm{d}x = \int \left[\int y f(y \mid x)\,\mathrm{d}y \right] f_X(x)\,\mathrm{d}x, \tag{5.9.27}$$

missä $f(y \mid x)$ on Y:n ehdollisen jakauman tiheysfunktio ehdolla $X = x$ ja $f_X(x)$ on X:n reunajakauman tiheysfunktio. Integrointi tehdään yli (X, Y):n arvoavaruuden. Koska sisempi integraali luasekkeessa (5.9.27) on ehdollinen odotusarvo $E(Y \mid x)$, niin odotusarvo (5.9.27) voidaan kirjoittaa muodossa

$$E(Y) = \int E(Y \mid x) f_X(x)\,\mathrm{d}x = E[E(Y \mid X)],$$

niin kuin lauseessa väitetään. □

Voimme soveltaa nyt Lausetta 5.13 laitteen rikkoontumisten vaatimaa korjausaikaa koskevaan esimerkkiin. Kokonaiskorjausaika vuodessa on Y ja sen keskiarvo on Lauseen 5.13 mukaan

$$\begin{aligned} E(Y) &= E[E(Y \mid X)] \\ &= E(\theta X) = \theta\, E(X) = \theta\lambda. \end{aligned}$$

Tälläinen sovellus on selvintä ajatella kaksitasoisena hierarkisena mallina, missä 1. vaiheessa sattuvat rikkoontumiset Poissonin jakauman mukaan ja sitten 2. vaiheessa tarvittavat korjausajat jakaantuvat eksponenttijakauman mukaan. Huomaa, että kokonaiskorjausajan Y keskiarvoparametri on nyt satunnaismuuttuja θX, missä X noudattaa Poissonin jakaumaa. Siksi Y:n jakaumaa on perusteltua kutsua yhdistetyksi jakaumaksi, koska siinä yhdistyvät parametrin Poissonin jakauma ja korjausajan eksponenttijakauama. Odotusarvoa $E(Y)$ laskettaessa on yksinkertaisinta laskea ensin ehdollinen odotusarvo $E(Y \mid X)$ ja sitten näiden ehdollisten odotusarvojen odotusarvo. Silloin laskennassa ei tarvita Y:n reunajakaummaa. Lopputulos on kuitenkin Lauseen 5.13 mukaan $E(Y)$.

Myös varianssi voidaan laskea ehdollisen jakauman varianssin avulla.

Lause 5.14 *Olkoot X ja Y mitkä tahansa kaksi satunnaismuuttujaa. Silloin*

$$\mathrm{Var}(Y) = E[\mathrm{Var}(Y \mid X)] + \mathrm{Var}[E(Y \mid X)], \tag{5.9.28}$$

jos odotusarvot ovat olemassa.

Todistus. Varianssin määritelmän mukaan

$$\mathrm{Var}(Y) = E[Y - E(Y)]^2 = E\big([Y - E(Y \mid X) + E(Y \mid X) - E(Y)]^2\big),$$

missä viimeinen yhtäsuuruus on saatu lisäämällä ja vähentämällä $E(Y \mid X)$. Korottamalla lauseke neliöön ja ottamalla odotusarvot saadaan

$$\begin{aligned} \mathrm{Var}(Y) = E\big([Y - E(Y \mid X)]^2\big) &+ E\big([(Y \mid X) - E(Y)]^2\big) \\ &+ 2\,E\big([Y - E(Y \mid X)]\,E[(Y \mid X) - E(Y)]\big). \end{aligned} \tag{5.9.29}$$

Lausekkeen (5.9.29) viimeinen termi on nolla, mikä voidaan osoittaa laskemalla merkitytyt odotusarvot. Yhtälön (5.9.29) oikean puolen ensimmäinen termi voidaan kirjoittaa muodossa:

$$\begin{aligned} E\big([Y - E(Y \mid X)]^2\big) &= E\big(E[Y - E(Y \mid X)]^2 \mid X\big) \\ &= E[\mathrm{Var}(Y \mid X)] \end{aligned}$$

ja toinen termi muodossa

$$E\big([E(Y \mid X) - E(Y)]^2\big) = \mathrm{Var}[E(Y \mid X)],$$

joten identiteetti (5.9.28) pitää paikkansa. □

Jatkuvat jakaumat: Yhteenveto

- Satunnaismuuttuja X on (absoluuttisesti) jatkuva, jos X:llä on tiheysfunktio $f(x) \geq 0$, joka toteuttaa seuraavat ehdot:

  1. $f(x) > 0$, kun $x \in S$,
  2. $\int_S f(x)\,\mathrm{d}x = 1$,
  3. $P(X \in A) = \int_A f(x)\,\mathrm{d}x$ on tapahtuman $\{X \in A\}$ todennäköisyys.

  Satunnaismuuttujan X kertymäfunktio on

  $$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

  ja momenttifunktio

  $$M(t) = E(\mathrm{e}^{tX}) = \int_{-\infty}^{\infty} \mathrm{e}^{tx} f(x)\,\mathrm{d}x.$$

- Tasajakauma $X \sim \mathrm{Tas}(a, b)$. Tiheysfunktio on

  $$f(x) = \begin{cases} \frac{1}{b-a}, & \text{kun } x \in [a, b]; \\ 0 & \text{muualla} \end{cases} \quad ja$$

  $$E(X) = \frac{a+b}{2} \qquad \text{ja} \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

  Momenttifunktio

  $$M(t) = \begin{cases} \dfrac{\mathrm{e}^{tb} - \mathrm{e}^{ta}}{t(b-a)}, & t \neq 0; \\ 1, & t = 0. \end{cases}$$

- Eksponenttijakauma $X \sim \mathrm{Exp}(\theta)$, $\theta > 0$ ja $x > 0$,

  $$f(x) = \frac{1}{\theta}\mathrm{e}^{-x/\theta} \qquad \text{ja} \qquad F(x) = 1 - \mathrm{e}^{-x/\theta}.$$

  Silloin

  $$E(X) = \theta \qquad \text{ja} \qquad \mathrm{Var}(X) = \theta^2.$$

  Momenttifunktio

  $$M(t) = \frac{1}{1 - \theta t}, \qquad t < \frac{1}{\theta}.$$

- Gammajakauma $X \sim \mathrm{Gamma}(\alpha, \beta)$, $\alpha > 0$ ja $\beta > 0$. Silloin

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \mathrm{e}^{-x/\beta}, \qquad 0 < x < \infty,$$

$$E(X^c) = \frac{\Gamma(\alpha + c)\beta^c}{\Gamma(\alpha)} \qquad \text{kaikilla } c > -\alpha.$$

Erityisesti

$$E(X) = \alpha\beta \qquad \text{ja} \qquad \mathrm{Var}(X) = \alpha\beta^2.$$

Momenttifunktio

$$M(t) = \left(\frac{1}{1 - \beta t}\right)^\alpha, \qquad t < \frac{1}{\beta}.$$

- χ^2-jakauma, $X \sim \mathrm{Khi2}(r)$. χ^2-jakauma saadaan, kun gammajakaumassa valitaan $\alpha = \frac{r}{2}$ ja $\beta = 2$, missä r on positiivinen kokonaisluku. Silloin

$$E(X) = r, \qquad \mathrm{Var}(X) = 2r \qquad \text{ja} \qquad M(t) = (1 - 2t)^{-r/2}, \quad t < \frac{1}{2}.$$

Jos $X_i \sim \mathrm{Khi2}(r_i)$, $i = 1, 2$, ovat riippumattomat, niin $X_1 + X_2 \sim \mathrm{Khi2}(r_1 + r_2)$.

- Normaalijakauma $X \sim \mathrm{N}(\mu, \sigma^2)$. Silloin

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty,$$

$$E(X) = \mu, \qquad \mathrm{Var}(X) = \sigma^2 \qquad \text{ja} \qquad M(t) = \mathrm{e}^{\mu t + (\sigma^2 t^2)/2}.$$

- Jos $Z \sim \mathrm{N}(0, 1)$, niin $Z^2 \sim \mathrm{Khi2}(1)$.

Jos $Z_i \sim \mathrm{N}(0, 1)$, $i = 1, 2$, ovat riippumattomat, niin $Z_1^2 + Z_2^2 \sim \mathrm{Khi2}(2)$.

- Studentin t-jakauma, $T \sim t(r)$.

$$T = \frac{Z}{\sqrt{U/r}} \sim t(r),$$

jos $Z \sim \mathrm{N}(0, 1)$ ja $U \sim \mathrm{Khi2}(r)$ ovat riippumattomat. Silloin

$$E(T) = 0, \quad r > 1 \qquad \text{ja} \qquad \mathrm{Var}(T) = \frac{r}{r - 2}, \quad r > 2.$$

- F-jakauma, $F \sim \mathrm{F}(r, s)$.

$$F = \frac{X/r}{Y/s} \sim F(r, s),$$

jos $X \sim \mathrm{Khi2}(r)$ ja $Y \sim \mathrm{Khi2}(s)$ ovat riippumattomat. Silloin

$$E(F) = \frac{s}{s - 2}, \quad s > 2; \qquad \mathrm{Var}(F) = \frac{2s^2(r + s - 2)}{r(s - 2)^2(s - 4)}, \quad s > 4.$$

- Beta-jakauma, $X \sim \text{Beta}(\alpha, \beta)$; $\alpha > 0$, $\beta > 0$.

  Tiheysfunktio

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 < x < 1.$$

$$E(X) = \frac{\alpha}{\alpha+\beta} \qquad \text{ja} \qquad \text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Muuttujien vaihto

Satunnaismuuttujan $Y = h(X)$ tiheysfunktio, kun X:n tiheysfunktio on $f_X(x)$:

$$f_Y(y) = f_X\big[g(y)\big]|g'(y)|, \qquad y \in S_Y,$$

missä $g(y)$ on $h(x)$:n käänteisfunktio.

Kaksiulotteiset jakaumat

- Jatkuvien satunnaismuuttujien X ja Y muodostaman satunnaisvektorin (X, Y) tiheysfunktio toteuttaa ehdot

  1. $f(x, y) \geq 0$ kaikilla (x, y),
  2. $\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1$,
  3. $P[(X, Y) \in A] = \iint\limits_{(x,y)\in A} f(x, y)\,\mathrm{d}x\,\mathrm{d}y$.

  Reunajakaumien tiheysfunktiot

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}y, \quad x \in S_X; \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}x, \quad y \in S_Y$$

  Ehdolliset tiheysfunktiot

$$f_X(x \mid y) = \frac{f(x, y)}{f_Y(y)} \qquad \text{ja} \qquad f_Y(y \mid x) = \frac{f(x, y)}{f_X(x)},$$

  missä $f_X(x) > 0$ ja $f_Y(y) > 0$.

  Kertymäfunktio

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t)\,\mathrm{d}s\,\mathrm{d}t.$$

- Normaalijakauma $(X, Y) \sim \mathrm{N}_2\big(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\big)$.
  Tiheysfunktio
$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Big[-\frac{1}{2(1-\rho^2)} Q(x, y)\Big],$$
  missä
$$Q(x, y) = \Big(\frac{x-\mu_1}{\sigma_1}\Big)^2 - 2\rho\Big(\frac{x-\mu_1}{\sigma_1}\Big)\Big(\frac{y-\mu_2}{\sigma_2}\Big) + \Big(\frac{y-\mu_2}{\sigma_2}\Big)^2.$$
  Reunajakaumat:
$$X \sim \mathrm{N}(\mu_1, \sigma_1^2) \qquad \text{ja} \qquad Y \sim \mathrm{N}(\mu_2, \sigma_2^2).$$
  Ehdolliset jakaumat:

  X:n jakauma ehdolla $Y = y$ on
$$X \mid y \sim \mathrm{N}\Big[\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(y - \mu_2),\ \sigma_1^2(1-\rho^2)\Big],$$
  Y:n jakauma ehdolla $X = x$ on
$$Y \mid x \sim \mathrm{N}\Big[\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x - \mu_1),\ \sigma_2^2(1-\rho^2)\Big].$$

Muuttujien vaihto

Olkoon
$$U = h_1(X, Y); \qquad V = h_2(X, Y)$$
muunnos, jolla on käänteismuunnos $x = g_1(u, v)$, $\quad y = g_2(u, v)$. Satunnaisvektorin (U, V) yhteisjakauman tiheysfunktio on
$$f_{U,V}(u, v) = f\big(g_1(u, v), g_2(u, v)\big)|J|,$$
missä $f(x, y)$ on satunnaismuuttujien X ja Y yhteisjakauman tiheysfunktio ja
$$J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}.$$

Ehdolliset odotusarvot ja varianssit

- Ehdollinen odotusarvo
$$E(Y) = E[E(Y \mid X)]$$
- Ehdollinen varianssi
$$\mathrm{Var}(Y) = E[\mathrm{Var}(Y \mid X)] + \mathrm{Var}[E(Y \mid X)].$$

Harjoituksia

1. Olkoon X jatkuva satunnaismuuttuja, jonka tiheysfunktio

$$f(x) = \begin{cases} \frac{1}{2}, & 0 \le x < \frac{1}{3}; \\ 1, & \frac{1}{3} \le x < \frac{2}{3}; \\ c, & \frac{2}{3} \le x \le 1; \\ 0 & \text{muualla.} \end{cases}$$

(a) Laske c.

(b) Määritä X:n kertymäfunktio.

(c) Piirrä X:n tiheysfunktio ja kertymäfunktio.

2. Olkoon jatkuvan satunnaismuuttujan X tiheysfunktio $f(x) = 2(1 - x)$, kun $0 \le x \le 1$ ja $f(x) = 0$ muualla.

(a) Piirrä X:n tiheysfunktio.

(b) Määritä ja piirrä X:n kertymäfunktio.

(c) Laske (i) $P(0 \le X \le 1/2)$, (ii) $P(1/4 \le X \le 3/4)$, (iii) $P(X = 3/4)$ ja (iv) $P(X \ge 3/4)$.

3. Olkoon $f(x)$ jatkuvan satunnaismuuttujan X tiheysfunktio. (i) Määritä jokaisesta alla määritellystä funktiosta $f(x)$ vakio c siten, että $f(x)$ on tiheysfunktio. (ii) Määritä kertymäfunktio $F(x) = P(X \le x)$ ja (iii) hahmottele tiheysfunktion $f(x)$ ja kertymäfunktion $F(x)$ kuvaajat.

(a) $f(x) = x^3/4, \quad 0 < x < c;$

(b) $f(x) = (3/16)x^2, \quad -c < x < c;$

(c) $f(x) = c/\sqrt{x}, \quad 0 < x < 1.$ Onko $f(x)$ rajoitettu?

4. Olkoon X:n tiheysfunktio $f(x) = c/x^2$, $1 < x < \infty$.

(a) Määritä c:n arvo siten, että $f(x)$ on tiheysfunktio.

(b) Osoita, että $E(X)$ ei ole äärellinen.

5. Olkoon $X \sim$ Khi2(12). Määritä vakiot a ja b siten, että

$$P(a < X < b) = 0.90 \qquad \text{ja} \qquad P(X < a) = 0.05.$$

6. Olkoon $X \sim$ Khi2(23).

(a) Laske $P(14.85 < X < 32.01)$.

(b) Määritä a ja b siten, että $P(a < X < b) = 0.95$ ja $P(X < a) = 0.025$.

(c) X:n keskiarvo ja varianssi.

7. Olkoon X:n momenttifunktio $M(t) = (1-2t)^{-12}$, $t < 1/2$. Laske

(a) $E(X)$,

(b) $\mathrm{Var}(X)$ ja

(c) $P(15.66 < X < 42.98)$.

8. Olkoon $X \sim \mathrm{N}(7, 4)$. Laske todennäköisyys $P[15.364 \le (X-7)^2 \le 20.096]$ (Vihje: Lause 5.6).

9. Oletetaan, että X ja Y ovat riippumattomat ja $X \sim \mathrm{Khi2}(8)$ ja $Y \sim \mathrm{Khi2}(12)$.

(a) Laske

$$P(1.646 < X \le 20.09), \qquad P(Y > 6.304), \qquad P(X+Y = 19.34).$$

(b) Määritä b, c ja d siten, että

$$P(X \le b) = 0.9, \qquad P(Y > c) = 0.9, \qquad P(X+Y > d) = 0.05.$$

(Vihje: Lause 5.8)

10. Olkoot Z_1, Z_2 ja Z_3 riippumattomat ja ne noudattavat $\mathrm{N}(0,1)$-jakaumaa. Määritellään satunnaismuuttujat

$$\bar{Z} = \frac{1}{3}(Z_1 + Z_2 + Z_3) \qquad \text{ja} \qquad U = Z_1^2 + Z_2^2 + Z_3^2.$$

Määritä vakiot a, b siten, että

$$P(|\bar{Z}| \le a) = 0.95; \qquad P(U > b) = 0.025.$$

(Vihje: Voit olettaa, että $\bar{Z}$ noudattaa normaalijakaumaa. Ks. myös Lause 5.7).

11. Määritä Esimerkissä 5.13 X:n ja Y:n jakaumien (reunajakaumien) tiheysfunktiot. Totea, että X ja Y ovat riippumattomat.

12. Totea laskemalla, että Esimerkissä 5.20 $E(X \mid y) = \frac{y}{2}$, $0 \le y \le 1$.

13. Olkoon jatkuvien satunnaismuuttujien X ja Y yhteisjakauman kertymäfunktio

$$F(x,y) = \begin{cases} kxy(x+y), & \text{kun } 0 < x < 1 \text{ ja } 0 < y < 1; \\ 0 & \text{muualla.} \end{cases}$$

(a) Laske vakion k arvo ja määritä X:n ja Y:n yhteisjakauman tiheysfunktio.

(b) Määritä X:n reunajakauman ja ehdollisen jakauman tiheysfunktio.

(c) Laske todennäköisyydet

$$P(X < 0.5,\ Y < 0.5); \qquad P(X < 0.5); \qquad P(X < 0.5 \mid Y < 0.5).$$

14. Määritä a, b, c, d siten, että

(a) $P(a \leq F_{8,12} \leq b) = 0.8$; $\quad P(c \leq F_{6,6} \leq d) = 0.98$.

(b) $P(|t_9| \geq a) = 0.05$; $\quad P(|t_{20}| > b) = 0.95$.

(c) $P(F_{1,12} \geq b) = 0.05$; $\quad P(F_{1,12} \leq c) = 0.2$.

15. Olkoot Y ja Z sellaiset riippumattomat satunnaismuuttujat, että $Z \sim \mathrm{N}(0,1)$ ja $Y \sim \chi^2_4$. Määritä a, b siten, että

$$P(Z \leq a\sqrt{Y}) = 0.975; \qquad P(Y + Z^2 \leq b) = 0.975.$$

16. Oletetaan, että $X \sim t(\nu)$. Osoita muuttujien vaihtotekniikalla, että $X^2 \sim \mathrm{F}_{1,\nu}$. (Huomaa, että muunnos ei ole monotoninen).

17. Olkoon $X \sim \mathrm{Khi2}(12)$.

(a) Määritä vakiot a ja b siten, että

$$P(a < X < b) = 0.90 \qquad \text{ja} \qquad P(X < a) = 0.05.$$

(b) Olkoon $X \sim \mathrm{N}(1,4)$ ja $P(a < X < b) = 0.50$. Määritä a ja b siten, että $b - a$ on mahdollisimman pieni.

18. Olkoot Z_1, Z_2 ja Z_3 riippumattomat ja $Z_i \sim \mathrm{N}(0,1)$-jakaumaa, $i = 1,2,3$. Määritellään satunnaismuuttujat $X_i = iZ_i + i$ ja

$$\bar{X} = \frac{1}{3}(X_1 + X_2 + X_3).$$

Määritä vakio a siten, että $P(|\bar{X}| \leq a) = 0.95$. (Vihje: Voit olettaa, että $\bar{X}$ noudattaa normaalijakaumaa.)

19. Olkoon $X \sim \mathrm{Khi2}(23)$.

(a) Laske $P(14.85 < X < 32.01)$.

(b) Määritä a ja b siten, että $P(a < X < b) = 0.95$ ja $P(X < a) = 0.025$.

(c) Laske X:n keskiarvo ja varianssi.

20. Oletetaan, että $X \sim \mathrm{Gamma}(\alpha, \beta)$. Osoita gammajakauman momenttifunktion avulla, että $E(X) = \alpha\beta$ ja $\mathrm{Var}(X) = \alpha\beta^2$.

21. Sillalle saapuu autoja Poissonin prosessin mukaan keskimäärin 15 autoa 10:ssä minuutissa. Laske todennäköisyys, että siltamaksujen kerääjä joutuu odottamaan annetusta hetkestä alkaen kahdeksaa autoa (eli kahdeksatta autoa) ainakin puoli tuntia.

22. Oletetaan, että X noudattaa gammajakaumaa Gamma$(3, 2)$. Määritä satunnaismuuttujan $Y = \sqrt{X}$ tiheysfunktio.

23. Logistista jakaumaa noudattavan satunnaismuuttujan X tiheysfunktio on

$$f(x) = \frac{\mathrm{e}^{-x}}{(1+\mathrm{e}^{-x})^2}, \qquad -\infty < x < \infty.$$

Osoita, että satunnaismuuttuja $Y = \frac{1}{1+\mathrm{e}^{-X}}$ noudattaa tasajakaumaa Tas$(0, 1)$.

24. Oletetaan, että $X \sim$ Tas$(-1, 3)$. Määritä satunnaismuuttujan $Y = X^2$ jakauma.

25. Olkoon momenttifunktio $M(t) = (1 - 2t)^{-12}$, $t < \frac{1}{2}$. Määritä $E(X)$, $\mathrm{Var}(X)$ ja $P(15 < X \leq 42)$.

26. Oletetaan, että matkustusaika kotoa töihin noudattaa normaalijakaumaa, jonka keskiarvo on 40 minuuttia ja hajonta 7 minuuttia. Jos haluat 95 %:n todelläköisyydellä olla työpaikalla klo 8:00, niin milloin viimeistään on lähdettävä kotoa?

27. Olkoon X:n momenttifunktio $M(t) = (1 - 2t)^{-12}$, $t < 1/2$. Laske

(a) $E(X)$,
(b) $\mathrm{Var}(X)$ ja
(c) $P(15.66 < X < 42.98)$.

28. Oletetaan, että $X \sim$ Gamma$(3, 1.5)$. Laske

(a) $P(X > 5)$,
(b) jakauman moodi (tiheysfunktion maksimi) sekä
(c) $E(Y)$ ja $\mathrm{Var}(Y)$, kun $Y = \frac{1}{X}$. (Ks. Lause 5.1.)

29. Tarkastellaan Poissonin prosessia, jonka intensiteetti on λ. Olkoon W odotusaika, kunnes α tapahtumaa sattuu. Silloin W:n kertymäfunktio on

$$F(w) = 1 - \sum_{x=0}^{\alpha-1} \mathrm{e}^{-\lambda w} \frac{(\lambda w)^x}{x!}.$$

Osoita, että tiheysfunktio $f(w)$ on

$$f(w) = \frac{\lambda(\lambda w)^{\alpha-1}}{(\alpha-1)!}\,\mathrm{e}^{-\lambda w}.$$

30. Satunnaismuuttujan Z tiheysfunktio on $f_Z(z)$. Olkoon $Y = aZ+b$, missä a ja b ovat annettuja vakioita.

(a) Osoita, että Y:n tiheysfunktio on

$$f_Y(y) = \frac{1}{|a|} f_Z(\frac{y-b}{a}).$$

(b) Esitä Y:n tiheysfunktio, kun $a = 2$, $b = 1$ ja $Z \sim \mathrm{N}(0, 1)$.

31. Olkoon X:n ja Y:n yhteisjakauman tiheysfunktio (Esimerkki 5.18)

$$f(x, y) = \frac{3}{2}x^2(1 - |y|), \qquad -1 < x < 1, \quad -1 < y < 1.$$

Määritä X:n ja Y:n jakaumien (reunajakaumien) tiheysfunktiot. Totea, että X ja Y ovat riippumattomat.

32. Oletetaan, että riippumattomat satunnaismuuttujat X ja Y noudattavat tasajakaumaa $\mathrm{Tas}(0, 1)$. Laske todennäköisyydet

(a) $P(|X - Y| \leq \frac{1}{2})$ ja

(b) $P(|\frac{X}{Y} - 1| \leq \frac{1}{2})$.

33. Olkoot X ja Y riippumattomat standardoidut normaalimuuttujat. Määritellään muuttujat

$$U = \frac{X + Y}{\sqrt{2}}; \qquad V = \frac{X - Y}{\sqrt{2}}.$$

Kirjoita U:n ja V:n yhteisjakauman tiheysfunktio. Näytä, että U ja V ovat riippumattomat.

34. Satunnaisvektori (X, Y) noudattaa kaksiulotteista normaalijakaumaa parametrein $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = 4$, $\sigma_2^2 = 9$ ja $\rho = 0.8$.

(a) Kirjoita (X, Y):n yhteisjakauman tiheysfunktion lauseke,

(b) $E(Y \mid x)$ ja

(c) $\mathrm{Var}(Y \mid x)$.

(Ks. Lause 5.11.)

35. Erääseen naisten kunto-ohjelmaan osallistuneilta mitattiin kehon rasvaprosentti X ennen ohjelman alkua ja rasvaprosenttin muutos Y ohjelman lopussa. Oletetaan, että muuttujien yhteisjakauma on normaalinen ja $\mu_X = 24.5$, $\sigma_X = 4.8$, $\mu_Y = -0.2$, $\sigma_Y = 3.0$ ja $\rho_{XY} = -0.32$. Laske

(a) $P(1.3 \leq Y \leq 5.8)$,

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

(c) $\mathrm{Var}(Y \mid X = x)$,

(d) $P(1.3 \leq Y \leq 5.8 \mid X = 18)$.

(Ks. Lause 5.11.)

36. Oletetaan, että satunnaisvektori (X, Y) noudattaa standardimuotoista normaalijakaumaa, missä $\mathrm{Cor}(X, Y) = 0.6$. Laske $P(X - Y < 1,\ X + Y > 2)$.

37. Olkoot X ja Y riippumattomat satunnaismuuttujat, joiden tiheysfunktiot ovat

$$f(x) = \mathrm{e}^{-x}, \qquad f(y) = \mathrm{e}^{-y}, \qquad 0 < x < \infty, \quad 0 < y < \infty.$$

Esitä satunnaisvektorin (U, V) yhteisjakauman tiheysfunktio, kun $U = X - Y$ ja $V = X + Y$.