Luku 6

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 tietyllä välillä positiivisia reaalilukuarvoja. Diskreetin satunnaismuuttujan arvojoukko on äärellinen tai numeroituva, mutta jatkuvan satunnaismuuttujan arvojoukko on ylinumeroituva. Satunnaismuuttuja X määriteltiin alaluvussa 2.6.1 jatkuvaksi, jos sen kertymäfunktio F_X on jatkuva (Määritelmä 2.7).

6.1 Jatkuvat satunnaismuuttujat

Jokaiseen satunnaismuuttujaan liittyy kertymäfunktio. Satunnaismuuttujan X kertymäfunktio määriteltiin alaluvussa 2.6.1 (Määritelmä 2.6 (ii)) pistefunktiona

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

Diskreetin satunnaismuuttujan kertymäfunktio on porrasfunktio, joka voidaan lausua hyppyfunktioiden summana (5.1.1) [ks. alaluku 5.1].

Lauseen 2.9 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 6.1 Funktio

$$F(x) = \frac{1}{1+e^{-x}} \tag{6.1.1}$$

on esimerkki jatkuvasta kertymäfunktiosta, joka siis toteuttaa Lauseen 2.9 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*. □

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

$$P(X \le x) = F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t. \tag{6.1.2}$$

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

$$F'_X(x) = f_X(x), \tag{6.1.3}$$

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

Määritelmä 6.1 Jatkuvan satunnaismuuttujan X tiheysfunktio $f_X(x)$ on funktio, joka toteuttaa yhtälön

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t \quad \text{kaikilla} \quad x \in \mathbb{R}. \tag{6.1.4}$$

Esimerkki 6.2 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

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

Silloin

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

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 $b > a$

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

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

$$M(t) = E(\mathrm{e}^{tx}) = \int_{-\infty}^{\infty} \mathrm{e}^{tx} f(x)\,\mathrm{d}x, \tag{6.1.5}$$

jos integraali 6.1.5 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 6.3 Lasketaan nyt Esimerkissä 2.12 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 2.12 käsitellyn jakauman 36 %:n piste on 0.6, koska

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

Esimerkki 6.4 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) $\lim\limits_{x\to-\infty} F(x) = 0$ ja $\lim\limits_{x\to\infty} F(x) = 1$,

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 $0 < x \le 1$ ja $F'(x) = 2 - x$ välillä $1 \le x \le 2$. Näin siis tiheysfunktio on

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

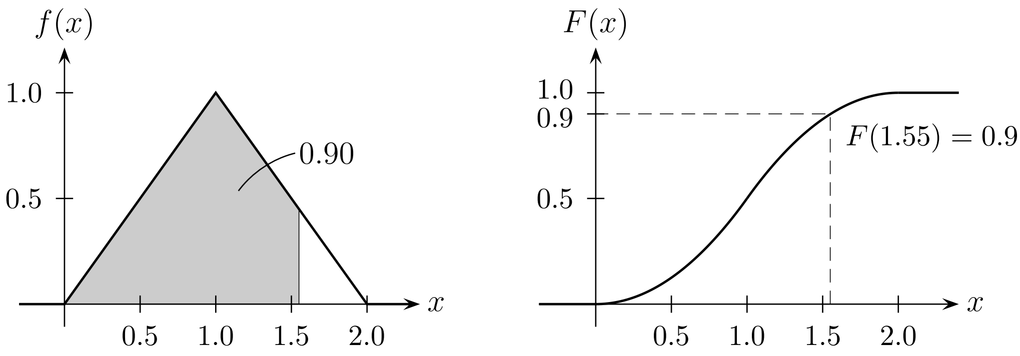
Tiheysfunktio voidaan kirjoittaa lyhyesti muodossa

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

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

Kolmiojakauman odotusarvo on

$$\begin{aligned} \mu = \int_0^2 x f(x)\,\mathrm{d}x &= \int_0^1 x \cdot x\,\mathrm{d}x + \int_1^2 x(2-x)\,\mathrm{d}x \\ &= \Big/_0^1 \frac{x^3}{3} + \Big/_1^2 \left(x^2 - \frac{x^3}{3}\right) = \frac{1}{3} + \left(4 - \frac{8}{3}\right) - \left(1 - \frac{1}{3}\right) \\ &= 1. \end{aligned}$$



Kuvio 6.1. Kolmiojakauman tiheysfunktion ja kertymäfunktion kuvaajat.

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 $\pi_{0.90} = 2 - \sqrt{0.2} = 1.55$. $\square$

Itse asiassa relaatio (6.1.4) 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 (6.1.4) pitää paikkansa. Huomattakoon, että jatkuvan satunnaismuuttujan tiheysfunktio ei välttämättä ole jatkuva, mutta kertymäfunktio on.

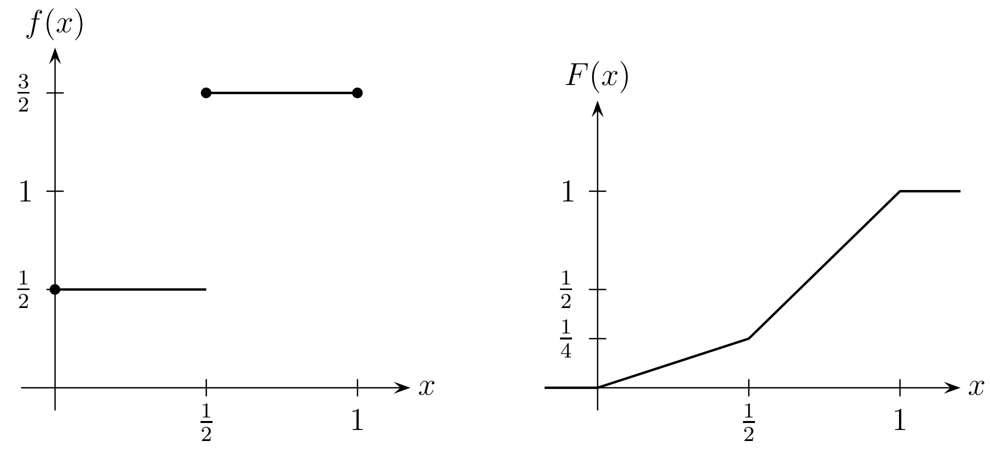
Esimerkki 6.5 Tarkastellaan nyt satunnaismuuttujaa X, jonka tiheysfunktio on

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

Vastaavasti X:n kertymäfunktio on

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

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

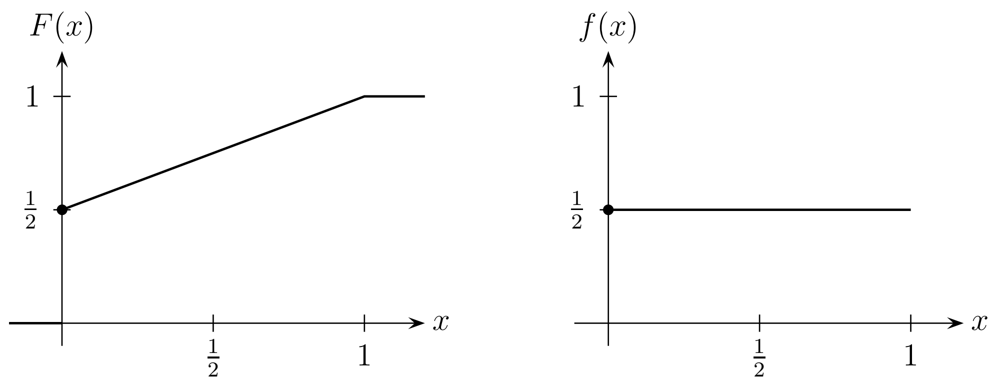


Kuvio 6.2. Satunnaismuuttujan X tiheysfunktion ja kertymäfunktion kuvaajat.

Jatkuvan satunnaismuuttujan tiheysfunktiolla voi olla äärellinen määrä epäjatkuvuuspisteitä, mutta kertymäfunktio on jatkuva. Esimerkin 6.5 satunnaismuuttujan tiheysfunktiolla on määrittelyalueellaan yksi epäjatkuvuuspiste ja kertymäfunktio on jatkuva. Relaatio (6.1.3) pitää paikkansa vain tiheysfunktion jatkuvuuspisteissä, mutta ei epäjatkuvuuspisteissä.

Esimerkki 6.6 Määritellään satunnaismuuttuja X siten, että sen kertymäfunktio on

$$F(x) = \begin{cases} 0, & x < 0; \\ \frac{1}{2}, & x = 0; \\ \frac{1}{2} + \frac{x}{2}, & 0 < x < 1; \\ 1, & 1 \leq x. \end{cases} \tag{6.1.6}$$



Kuvio 6.3. 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 (6.1.6) voidaan kirjoittaa porrasfunktion (kertymäfunktio) ja jatkuvan kertymäfunktion summana. Alaluvussa 5.1 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

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

Nyt kertymäfunktio (6.1.6) voidaan kirjoittaa muodossa

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

Esimerkiksi todennäköisyys

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

Satunnaismuuttuja X ei ole diskreetti eikä jatkuva. □

Yleisesti jatkuva satunnaismuuttuja voidaan määritellä identiteetin (6.1.4) avulla olettamatta tiheysfunktion $f(x)$ jatkuvuutta. Jos on olemassa sellainen epänegatiivinen funktio $f(x)$ [ts. $f(x) \ge 0$ kaikilla $x \in \mathbb{R}$], että (6.1.4) 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.

6.2 Tasajakauma ja eksponenttijakauma

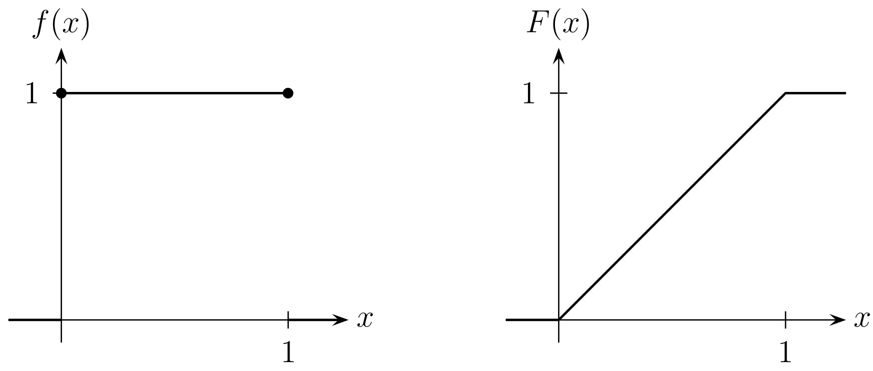
6.2.1 Tasajakauma

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

$$f(x) = \begin{cases} 1, & \text{kun } x \in [0, 1], \\ 0 & \text{muualla.} \end{cases} \tag{6.2.1}$$

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

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



Kuvio 6.4. Tasajakauman Tas(0, 1) tiheysfunktio ja kertymäfunktio.

Tasajakauman keskiarvo ja varianssi ovat:

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

ja

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

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 $U \sim \operatorname{Tas}(a, b)$. Koska $E(U) = (b-a)\,E(X) + a$ ja $\operatorname{Var}(U) = (b-a)^2 \operatorname{Var}(X)$, niin

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

Satunnaismuuttujan U tiheysfunktio on

$$f(u) = \begin{cases} \frac{1}{b-a}, & \text{kun } u \in [a, b]; \\ 0 & \text{muualla} \end{cases} \tag{6.2.2}$$

ja U:n momenttifunktio on

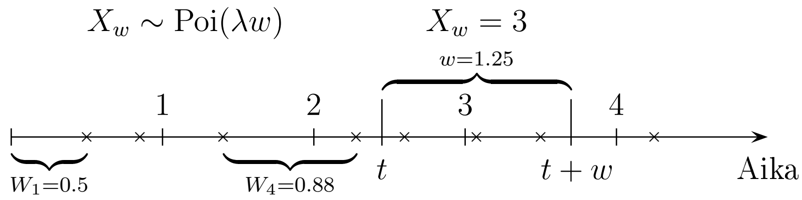
$$M_U(t) = \begin{cases} \dfrac{\mathrm{e}^{tb} - \mathrm{e}^{ab}}{t(b-a)}, & t \neq 0; \\ 1, & t = 0. \end{cases}$$

6.2.2 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 5.2 mukaan todennäköisyys, että w:n pituisella välillä sattuu x tapahtumaa, on

$$P(X_w = x) = \mathrm{e}^{-\lambda w} \frac{(\lambda w)^x}{x!}. \tag{6.2.3}$$

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.



Kuvio 6.5. 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, ... ovat toisistaan riippumattomat ja noudattavat samaa jakaumaa.

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 (6.2.3) mukaan

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

Odotusajan W kertymäfunktio on siis

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

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 (6.1.3) 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

$$f(w) = \frac{1}{\theta}\mathrm{e}^{-w/\theta}. \tag{6.2.4}$$

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

$$P(T > a + b \mid T > a) = P(T > b) \tag{6.2.5}$$

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 6.7 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 (6.2.3) $\lambda = 1/3$ asiakasta minuutissa ja $X \sim \mathrm{Exp}(3)$, koska eksponenttijakauman keskiarvo $\theta = 1/\lambda$. Jakauman $\mathrm{Exp}(3)$ 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 $\mathrm{Exp}(3)$ mediaanin m tulee toteuttaa ehto $F(m) = 1 - \mathrm{e}^{-m/3} = \frac{1}{2}$, joten

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

□

6.2.3 Elinaikajakauma

Ominaisuuden (6.2.5) 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 (6.2.5) 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

$$G(t + s) = G(t)G(s), \qquad t > 0, \quad s > 0. \tag{6.2.6}$$

Määritelmänsä nojalla $G(0) = 1$ ja $G(t) \to 0$, kun t kasvaa. Onko eksponenttifunktion lisäksi muita eloonjäämisfunktioita, joilla on unohtamisominaisuus (6.2.6)? Voidaan osoittaa, että ehdon (6.2.6) 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. Määritellään todennäköisyys

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

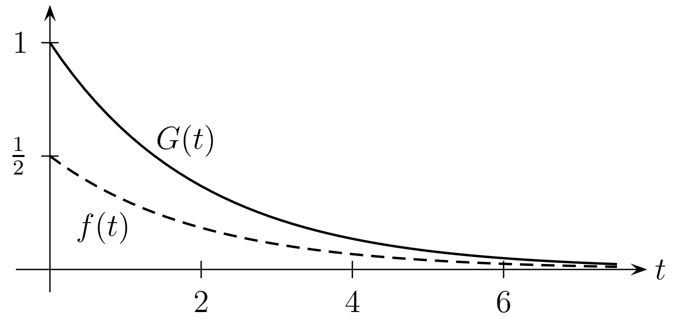
missä viimeistä edellinen yhtäsuuruus saadaan unohtamisominaisuuden (6.2.6) nojalla. Kun funktiota $\mathrm{e}^{-\lambda\Delta}$ arvioidaan Taylorin polynomin avulla, saadaan

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

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

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

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



Kuvio 6.6. Eksponenttijakauman Exp(2) tiheysfunktio $f(t) = \frac{1}{2}\mathrm{e}^{-t/2}$ ja vastaava eloonjäämisfunktio $G(t) = \mathrm{e}^{-t/2}$.

6.3 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 on

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\mathrm{e}^{-x}\,\mathrm{d}x. \tag{6.3.1}$$

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

$$f(t) = \frac{t^{\alpha-1}\mathrm{e}^{-t}}{\Gamma(\alpha)}, \qquad 0 < t < \infty \tag{6.3.2}$$

määrittelee tiheysfunktion, sillä gammafunktiossa integroitava on positiivinen välillä $(0, \infty)$. Sanokaamme, että (6.3.2) 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 6.5 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

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \mathrm{e}^{-x/\beta}, \qquad 0 < x < \infty, \quad \alpha > 0, \quad \beta > 0. \tag{6.3.3}$$

Esitämme nyt gammajakauman perusominaisuudet seuraavassa lauseessa.

Lause 6.1 *Oletetaan, että* $X \sim \mathrm{Gamma}(\alpha, \beta)$.

1. *Funktio (6.3.3) 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 (6.3.3) $\alpha = 1$, 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 (6.3.3) muotoon

$$f(x) = \frac{1}{\Gamma\left(\frac{r}{2}\right) 2^{r/2}} x^{(r/2)-1} \mathrm{e}^{-x/2}, \qquad 0 < x < \infty, \tag{6.3.4}$$

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 (6.2.3):

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

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. (6.2.3)]. 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

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

6.4 Normaalijakauma

6.4.1 Standardimuotoinen normaalijakauma

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

$$f(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2} \qquad -\infty < z < \infty. \tag{6.4.1}$$

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

Tarkistamme nyt, että (6.4.1) 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

$$\int_{-\infty}^{\infty} \mathrm{e}^{-z^2/2}\,\mathrm{d}z = \sqrt{2\pi}. \tag{6.4.2}$$

Emme pysty suoraan integroimaan funktiota $\mathrm{e}^{-z^2/2}$, koska sen integraalifunktio ei ole lausuttavissa suljetussa muodossa. Osoittautuu kuitenkin, että integraalin (6.4.2) 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 (6.4.2) 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 (6.4.2) on myös läheinen yhteys gammafunktioon. Koska integraalissa (6.4.2) integroitava on symmetrinen nollan suhteen, niin integraalit yli välien $(-\infty, 0)$ ja $(0, \infty)$ ovat yhtä suuret. Siksi

$$\int_{0}^{\infty} \mathrm{e}^{-z^2/2}\,\mathrm{d}z = \sqrt{\frac{\pi}{2}}. \tag{6.4.3}$$

Tekemällä sijoitus $x = \frac{1}{2}z^2$ integraaliin (6.4.3) saadaan integraali, joka on $\Gamma\big(\frac{1}{2}\big)$. Silloin

$$\Gamma\left(\frac{1}{2}\right) = \int_{0}^{\infty} x^{-1/2}\mathrm{e}^{-x}\,\mathrm{d}x = \sqrt{\pi}. \tag{6.4.4}$$

Lause 6.2 *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.

6.4.2 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 6.3 *Jos $X \sim \mathrm{N}(\mu, \sigma^2)$, niin*

1. *$E(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$ 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 4.12)

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

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 6.5 esitettävän muunnostekniikan avulla saadaan X:n tiheysfunktioksi

$$\begin{aligned} f_X(x) &= f_Z\left(\frac{x-\mu}{|\sigma|}\right)\frac{1}{|\sigma|} \\ &= \frac{1}{\sqrt{2\pi}|\sigma|}\mathrm{e}^{-(x-\mu)^2/2\sigma^2}. \end{aligned} \tag{6.4.5}$$

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 6.8 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 6.9 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

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

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

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

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 6.4

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, ..., X_n riippumattomat, $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$ ja a_1, a_2, ..., 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 6.10 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 6.4 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

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

□

6.5 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.

6.5.1 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 (6.1.3) avulla, kun $G(y)$ on derivoituva.

Esimerkki 6.11 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 6.12 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 $y \in S_Y = (0, 1)$. □

6.5.2 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

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

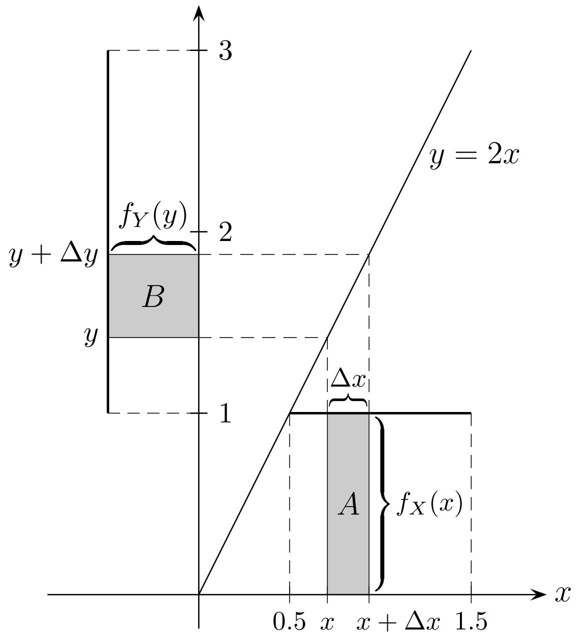
ja

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

Esimerkki 6.13 Oletetaan, että $X \sim \mathrm{Tas}(0.5, 1.5)$ ja $Y = 2X$. Mitä jakaumaa Y noudattaa?

Kuviossa 6.7 on alueen A pinta-ala

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



Kuvio 6.7. 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

$$P[X \in (x, x + \Delta x)] = P[Y \in (y, y + \Delta y)]. \tag{6.5.1}$$

Koska $y = 2x$ ja $y + \Delta y = 2(x + \Delta x)$, niin $\Delta y = 2\Delta x$ ja identiteetistä (6.5.1) 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 6.5 *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

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

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 6.14 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 6.5 muunnosmenetelmää ei voi suoraan soveltaa. Jos kuitenkin on olemassa sellainen S_X:n ositus yhteispisteettömiin osaväleihin $A_1, A_2, \ldots, A_m$, että

$$S_X = A_1 \cup A_2 \cup \cdots \cup A_m \tag{6.5.2}$$

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 6.5 avulla muodossa

$$f_Y(y) = \sum_{i=1}^{m} f_X\bigl(g_i(y)\bigr)|g_i'(y)|. \qquad y \in S_Y. \tag{6.5.3}$$

Huomattakoon, että joskus tarvitaan äärellisen osituksen (6.5.2) sijasta ositus, jossa jakovälejä A_1, A_2, $\ldots$ on ääretön määrä ($m = \infty$).

6.5.3 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 (6.5.3) mukaan U:n tiheysfunktio on

$$f_U(u) = f_X(-\sqrt{u}) \frac{1}{2\sqrt{u}} + f_X(\sqrt{u}) \frac{1}{2\sqrt{u}} = \frac{1}{\sqrt{2\pi u}} \mathrm{e}^{-u/2}, \tag{6.5.4}$$

kun $u \in (0, \infty)$. U noudattaa χ^2-jakaumaa vapausastein 1. Käsittelemme tilastotieteessä tärkeää χ^2-jakaumaa vielä jatkossa tarkemmin.

Lause 6.6 *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. $\square$

Lause 6.7 *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 6.7 mukaan havaintojen neliösumma noudattaa Khi2-jakaumaa vapausastein n, missä n on otoskoko.

Seuraus 6.1 *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 6.1 mukaan standardoitujen havaintojen neliösumma noudattaa Khi2-jakaumaa vapausastein n.

Lause 6.8 *Olkoot X_1 ja X_2 riippumattomat ja $X_i \sim \mathrm{Khi2}(n_i)$, $i = 1, 2$. Silloin*

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

6.6 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 6.2 Jos X on jatkuva satunnaismuuttuja, niin satunnaismuuttujan $h(X)$ odotusarvo on

$$E[h(X)] = \int_S h(x) f(x) \, \mathrm{d}x, \tag{6.6.1}$$

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

Huomautus 6.1 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

$$\int\limits_S |h(x)|$$

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

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

Esimerkki 6.15 Tarkastellaan nyt *Cauchyn jakaumaa* noudattavaa satunnaismuuttujaa X, jonka tiheysfunktio on

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad -\infty < x < \infty. \tag{6.6.2}$$

Kaava (6.6.2) todellakin määrittelee tiheysfunktion, koska

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

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

$$E(|X|) = \int\limits_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)}\,\mathrm{d}x = \frac{2}{\pi} \int\limits_{0}^{\infty} \frac{x}{1+x^2}\,\mathrm{d}x.$$

Jokaista reaalilukua $M > 0$ kohti saadaan

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

Tästä seuraa, että

$$E(|X|) = \lim_{M\to\infty} \frac{2}{\pi} \int\limits_{0}^{M} \frac{x}{1+x^2}\,\mathrm{d}x = \frac{1}{\pi} \lim_{M\to\infty} \log(1+M^2) = \infty,$$

joten $E(X)$ ei ole olemassa. □

Taulukko 6.1. 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

6.6.1 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 6.1 esitetään yhteenveto eri momenteista

Momenttifunktio määriteltiin 3. luvussa (Määritelmä 4.8) ja jatkuville satunnaismuuttujille alaluvussa 6.1 [ks. identiteetti (6.1.5)]. Jatkuvan satunnaismuuttujan X momentifunktio on

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

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 4.6.2.

Esimerkki 6.16 Huomautuksessa 6.1 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

$$M(t) = E(\mathrm{e}^{tX}) = \int_0^\infty \mathrm{e}^{tx}\mathrm{e}^{-x}\,\mathrm{d}x = \frac{1}{1-t}, \qquad t < 1.$$

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:

$$\gamma_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3}, \tag{6.6.3}$$

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 6.15). 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:

$$\gamma_2 = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4}, \tag{6.6.4}$$

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: $\frac{\mu_4}{\sigma^4} - 3$.

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. $\int\limits_S f(x)\,\mathrm{d}x = 1$,
  3. $P(X \in A) = \int\limits_A f(x)\,\mathrm{d}x$ on tapahtuman $\{X \in A\}$ todennäköisyys.

  Satunnaismuuttujan X kertymäfunktio on

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

  ja momenttifunktio

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

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

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

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

  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 $X \sim \text{Gamma}(\alpha, \beta)$, $\alpha > 0$ 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
$$E(X) = \alpha\beta \qquad \text{ja} \qquad \text{Var}(X) = \alpha\beta^2.$$

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

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

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

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

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

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty,$$
$$E(X) = \mu, \qquad \text{Var}(X) = \sigma^2 \qquad \text{ja} \qquad M(t) = \mathrm{e}^{\mu t + (\sigma^2 t^2)/2}.$$

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

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

Harjoituksia

1. Olkoon X jatkuva satunnaismuuttuja, jonka tiheysfunktio

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

(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 $0 \leq x \leq 1$ ja $f(x) = 0$ muualla.

(a) Piirrä X:n tiheysfunktio.

(b) Määritä ja piirrä X:n kertymäfunktio.

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

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 $F(x) = P(X \leq x)$ 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 $X \sim \text{Khi2}(12)$. Määritä vakiot a ja b siten, että

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

6. Olkoon $X \sim \text{Khi2}(23)$.

(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 $P[15.364 \leq (X-7)^2 \leq 20.096]$ (Vihje: Lause 6.6).

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

(a) Laske

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

(b) Määritä b, c ja d siten, että

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

(Vihje: Lause 6.8)

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

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

(Vihje: Voit olettaa, että $\bar{Z}$ noudattaa normaalijakaumaa. Ks. myös Lause 6.7).

11. 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.

12. 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.)

13. Olkoon $X \sim \text{Khi2}(23)$.

(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.

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

15. 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.

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

17. 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 $\text{Tas}(0, 1)$.

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

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

20. 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?

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

(a) $E(X)$,

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

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

22. Oletetaan, että $X \sim \text{Gamma}(3, 1.5)$. 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 6.1.)

23. 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

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

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

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