2.1 Satunnaismuuttuja ja sen jakauma

Satunnaismuuttuja (lyhenne sm, engl. *random variable, rv*; myös *variate*) on satunnaiskokeeseen liittyvä numeerinen muuttuja, jonka arvo määräytyy kokeen lopputuloksesta. Satunnaismuuttuja on siis perusjoukolla määritelty reaaliarvoinen funktio.

Määritelmä (Satunnaismuuttuja)

Olkoon Ω perusjoukko. Kuvaus $X : \Omega \to \mathbb{R}$ on satunnaismuuttuja.

Esimerkkejä satunnaismuuttujista

- Silmäluku nopanheitossa. Jos alkeistapaus $\omega \in \{1, \ldots, 6\}$ kertoo nopan silmäluvun, niin

$$X(\omega) = \omega$$

  on nopan silmäluku.

- Silmälukujen summa kahdessa nopanheitossa. Jos $E = \{1, \ldots, 6\}$, perusjoukko $\Omega = E \times E$, ja alkeistapaukselle $(i, j) \in E \times E$ ensimmäinen koordinaatti i kertoo nopan yksi ja toinen koordinaatti j nopan kaksi silmäluvun, niin niiden summa on

$$X((i, j)) = i + j.$$

Tapahtuman indikaattori (tärkeä esimerkki satunnaismuuttujasta)

Määritelmä (Tapahtuman indikaattori)

Olkoon $A \subset \Omega$ tapahtuma. Sen indikaattori (eli ilmaisin eli osoitinmuuttuja) 1_A on satunnaismuuttuja, jonka määrittelee lauseke

$$1_A(\omega) = \begin{cases} 1, & \text{jos } \omega \in A, \\ 0, & \text{muuten.} \end{cases}$$

Huomautuksia: Tapahtuman A indikaattorille käytetään yleisesti myös merkintää I_A. Jos tapahtuma A esitetään monimutkaisella lausekkeella, niin usein käytetään edellisten sijasta merkintää $1(A)$ tai $I(A)$. Jos on lisäksi tarpeen merkitä argumentti ω näkyviin, niin voidaan käyttää merkintöjä $1(A)(\omega)$ tai $I(A)(\omega)$.

Satunnaismuuttujien laskutoimitukset

- Usein samalla perusjoukolla on määritelty useampi kuin yksi satunnaismuuttuja, esim. X ja Y.
- Tällöin voidaan tarkastella myös *satunnaismuuttujien välisiä laskutoimituksia*, kuten aX (vakiolle $a \in \mathbb{R}$), $X + Y$, XY jne. Myös ne ovat satunnaismuuttujia.
- Esim. $X + Y$ tarkoittaa funktioiden yhteenlaskua, ts. $X + Y$ on se funktio $\Omega \to \mathbb{R}$, jolle

  $$(X + Y)(\omega) = X(\omega) + Y(\omega).$$

  Ts. satunnaismuuttujien laskutoimitukset tulkitaan pisteittäin.
- Jos Y ei saa arvoa nolla, niin myös X/Y on sm.

Merkintöjä (piilota oomega)

- Jos X on sm, niin voidaan kysyä, millä todennäköisyydellä se saa arvon joukosta B, jossa $B \subset \mathbb{R}$.
- Kyseistä tapahtumaa voidaan merkitä $\{\omega \in \Omega : X(\omega) \in B\}$, mutta sitä merkitään tavallisesti lyhyemmin: $\{X \in B\}$.
- Sen todennäköisyyttä voidaan merkitä seuraavasti,

$$\begin{aligned} P(X \in B) &= P\{X \in B\} = P(\{X \in B\}) \\ &= P(\{\omega \in \Omega : X(\omega) \in B\}). \end{aligned} \tag{1}$$

Tavallisesti käytetään lyhyitä merkintöjä, joissa ei esiinny perusjoukkoa Ω eikä sisäkkäisiä sulkuja.

Lisää merkintöjä tapahtumalle $\{X \in B\}$

Huomaa, minkälaisia merkintöjä tapahtumalle $\{X \in B\}$ käytetään seuraavissa tilanteissa.

- B on yksiö $\{x\}$, jossa $x \in \mathbb{R}$. Tällöin kysytään pistetodennäköisyyttä

$$P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\}).$$

- B on väli kuten esim. suljettu väli $[a, b]$, jossa $a < b$,

$$P(a \leq X \leq b) = P(\{\omega \in \Omega : a \leq X(\omega) \leq b\}).$$

  Vastaavasti määritellään $P(a < X \leq b)$, $P(a \leq X < b)$ ja $P(a < X < b)$.

- Tapaus $B = (-\infty, x]$. Tällöin merkitään

$$P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}).$$

Jakauma

Määritelmä (Satunnaismuuttujan jakauma)

Jos X on satunnaismuuttuja, niin sen jakauma on

$$P(X \in B)$$

ymmärrettynä argumentin $B \subset \mathbb{R}$ funktiona.

Huomautus: Itse asiassa tn $P(X \in B)$ on määritelty vain silloin, kun $B \subset \mathbb{R}$ on ns. Borelin joukko, jollaisia ovat esim. kaikki välit ja kaikki joukot, jotka voidaan muodostaa numeroituvasta määrästä välejä soveltamalla numeroituva määrä joukko-operaatiota. Tästä lähtien jätämme tämän seikan vaille huomiota.

Huomautuksia jakaumasta

- On mahdollista osoittaa, että sm:n X jakauma

$$P_X(B) = P(X \in B), \qquad B \subset \mathbb{R}$$

  toteuttaa todennäköisyyden aksioomat, joten jakauma on joukossa $\mathbb{R}$ määritelty tn-mitta.
- Perusjoukko Ω häviää usein kokonaan merkinnöistä. Tilanne voidaan mieltää myös siten, että perusjoukkona onkin $\mathbb{R}$ ja että tapahtumat ovat sen osajoukkoja.
- Jakauma on hyvin abstrakti olio.
- Seuraavaksi tarkastelemme konkreettisempia välineitä, joiden avulla voimme kuvailla ja hallita jakaumia: kertymäfunktio, diskreetin jakauman pistetodennäköisyysfunktio ja jatkuvan jakauman tiheysfunktio.

2.2 Kertymäfunktio

Määritelmä

Jos X on satunnaismuuttuja, niin funktio $F : \mathbb{R} \to \mathbb{R}$

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

on X:n kertymäfunktio (lyhenne kf).

Huomautus: Englannin kielellä

- jakauma on *(probability) distribution* (joskus *law*),
- kertymäfunktio on *distribution function* tai *cumulative distribution function, cdf*.

Tapahtuman indikaattorin kertymäfunktio

- Olkoon A tapahtuma, ja olkoon $X = 1_A$, eli X on tapahtuman A indikaattori.
- Tällöin X:n kf on

$$F(x) = \begin{cases} 0, & \text{jos } x < 0, \\ 1 - P(A), & \text{jos } 0 \le x < 1, \\ 1, & \text{jos } x \ge 1. \end{cases}$$

- Tämä kertymäfunktio on paloittain vakio.
- Huomaa, että kf:lla voi olla hyppyjä ja että se voi olla vakio jollakin välillä.

Monotonisen funktion raja-arvot

- Monotonisella funktiolla $G : \mathbb{R} \to \mathbb{R}$ on jokaisessa pisteessä $x \in \mathbb{R}$ olemassa sekä vasemmanpuoleinen että oikeanpuoleinen raja-arvo.
- Lisäksi sillä on raja-arvo pisteissä $-\infty$ ja ∞ (mutta äärettömässä raja-arvo voi olla joko jokin reaaliluku, tai jompikumpi äärettömistä $-\infty$ tai ∞).
- Näitä raja-arvoja merkitään seuraavasti

$$G(x-) = \lim_{y \to x-} G(y), \qquad G(x+) = \lim_{y \to x+} G(y)$$
$$G(-\infty) = \lim_{y \to -\infty} G(y), \qquad G(\infty) = \lim_{y \to \infty} G(y).$$

Kertymäfunktion ominaisuudet

Lause (Kertymäfunktion ominaisuudet)

Satunnaismuuttujan X kertymäfunktiolla $F : \mathbb{R} \to \mathbb{R}$ on seuraavat ominaisuudet.

(a) F on kasvava funktio.

(b) F on oikealta jatkuva funktio, eli

$$F(x+) = F(x), \qquad \text{kaikilla } x \in \mathbb{R}.$$

(c) $F(-\infty) = 0$ ja $F(\infty) = 1$.

Kääntäen, jos funktiolla $F : \mathbb{R} \to \mathbb{R}$ on nämä ominaisuudet, niin on olemassa sm X siten, että F on X:n kf.

Sattumistodennäköisyydet väleille

- Lasketaan tn $P(a < X \leq b)$, jossa $a < b$. Koska

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\},$$

jossa osat ovat erillisiä, on

$$P(a < X \leq b) = F(b) - F(a).$$

- Seuraavan lauseen jälkeen pystymme toistamaan vastaavan laskun myös väleille (a, b), $[a, b)$ ja $[a, b]$.

Lause

Jos F on sm:n X kf, niin kaikilla x

$$P(X < x) = F(x-).$$

Kertymäfunktion hypyt

- Kertymäfunktiolla voi olla hyppyjä.
- Koska
$$\{X \le x\} = \{X < x\} \cup \{X = x\},$$
jossa osat ovat erillisiä, niin
$$P(X = x) = P(X \le x) - P(X < x) = F(x) - F(x-).$$
- Pistetodennäköisyys $P(X = x)$ on yhtä kuin kertymäfunktion hyppy pisteessä x.

Satunnaismuuttuja voidaan ilmoittaa funktion alaindeksillä

- Voidaan tarkastella useampaa kuin yhtä satunnaismuuttujaa, vaikkapa X ja Y.
- Tällöin usein käytetään sellaista merkintätapaa, jossa satunnaismuuttuja kirjoitetaan kertymäfunktion alaindeksiksi: esim. F_X on X:n kf ja F_Y on Y:n kf.
- Tällaista merkintätapaa voidaan käyttää muidenkin jakauman esitysten kuin kertymäfunktioiden kohdalla.

Sama jakauma

Määritelmä

Satunnaismuuttujilla X ja Y on sama jakauma eli ne ovat samoin jakautuneet, jos

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

Tämä seikka voidaan ilmaista merkinnällä

$$X \stackrel{\mathrm{d}}{=} Y.$$

Kertymäfunktiot ja jakaumat vastaavat toisiaan

Seuraava lause toteaa, että kertymäfunktiot ja jakaumat vastaavat toisiaan yksikäsitteisesti. Tämän takia voidaan puhua tietyn jakauman kertymäfunktiosta, eikä aina tarvitse puhua sm:n kertymäfunktiosta.

Lause (Kertymäfunktio määrää jakauman yksikäsitteisesti)

Seuraavat asiat ovat yhtäpitäviä.

(a) X ja Y ovat samoin jakautuneet, eli $X \stackrel{\mathrm{d}}{=} Y$.

(b) $F_X = F_Y$ (ts. X:n ja Y:n kertymäfunktiot ovat samat).

2.3 Diskreetti jakauma

Määritelmä

Satunnaismuuttujan X jakauma on diskreetti, jos sen arvojoukko $\{x_1, x_2, \ldots\}$ on äärellinen tai korkeintaan numeroituvasti ääretön.

Asia voidaan ilmaista myös jollakin seuraavista tavoista

- X on diskreetti sm,
- sm X on diskreetisti jakautunut
- X:llä on diskreetti jakauma.

Laajennettu määritelmä diskreetille satunnaismuuttujalle

- Joskus määritelmää laajennetaan: vaaditaan vain, että X saa todennäköisyydellä yksi arvon korkeintaan numeroituvasti äärettömästä joukosta S.
- Tällöin diskreetti sm X voi saada arvoja joukon S ulkopuolelta, mutta tämän poikkeustapahtuman todennäköisyys on nolla.
- Yksinkertaisuuden vuoksi kutsumme myös tässä tapauksessa X:n arvojoukoksi jotakin sellaista korkeintaan numeroituvaa joukkoa S, jolle $P(X \in S) = 1$.

Pistetodennäköisyysfunktio, ptnf

Määritelmä

Satunnaismuuttujan X pistetodennäköisyysfunktio (lyh. ptnf) on funktio $f : \mathbb{R} \to \mathbb{R}$, jolle

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

- Pistetodennäköisyysfunktio $x \mapsto P(X = x)$ (engl. *probability (mass) function, pmf*) on määritelty koko reaaliakselilla kaikenlaisille satunnaismuuttujille X.
- Käsite on käyttökelpoinen lähinnä vain diskreeteille satunnaismuuttujille.

Merkintäsopimuksia

- Tyypillisesti pistetodennäköisyydet ilmoitetaan vain niille x_i, jotka kuuluvat diskreetin sm:n määritelmässä esiintyvään, korkeintaan numeroituvasti äärettömään joukkoon, joka voi esimerkiksi olla jokin kokonaislukujen osajoukko. Tällöin asiayhteydestä pitää ymmärtää, että muilla $x \in \mathbb{R}$ on $P(X = x) = 0$.
- Diskreetin satunnaismuuttujan ptnf:lle käytetään muuten samaa symbolia kuin sen kf:lle, mutta ptnf:n symboli kirjoitetaan pienellä kirjaimella (esim. f, g tai f_X) ja kf:n symboli vastaavalla isolla kirjaimella (esim. F, G tai F_X).

Diskreetin jakauman ptnf:n ominaisuuksia

Lause

Olkoon X diskreetti sm, f sen ptnf, ja olkoon S sen arvojoukko. Tällöin

(a) $0 \leq f(x) \leq 1$ kaikilla x, ja $f(x) = 0$, kun $x \notin S$.

(b) f määrää X:n jakauman kaavalla

$$P(X \in B) = \sum_{x \in B} f(x)$$

Huomautus: Kohdassa (b) lasketaan yhteen korkeintaan numeroituva määrä ei-negatiivisia termejä $f(x_i)$, jossa $x_i \in S \cap B$, minkä takia kyseinen summamerkintä on mielekäs. (Summausjärjestyksellä ei ole väliä.)

Ptnf:stä diskreetti jakauma

Lause

Olkoon $f : \mathbb{R} \to \mathbb{R}$ ei-negatiivinen funktio, joka saa positiivisia arvoja korkeintaan numeroituvassa $\mathbb{R}$:n osajoukossa siten, että

$$\sum_x f(x) = 1$$

Tällöin se on jonkin diskreetin satunnaismuuttujan ptnf.

2.4 Esimerkkejä diskreeteistä jakaumista

- diskreetti tasajakauma
- Bernoullin jakauma
- binomijakauma
- hypergeometrinen jakauma, negatiivinen binomijakauma, geometrinen jakauma, Poissonin jakauma (luvussa 5).

Diskreetti tasajakauma

- Satunnaismuuttujalla X on diskreetti tasajakauma joukossa $\{1, \dots, N\}$, mikäli sen ptnf on
$$f(x) = \frac{1}{N}, \qquad \text{kun } x = 1, 2, \dots, N.$$
- Esimerkiksi, jos X on silmäluku nopanheitossa, niin X:llä on diskreetti tasajakauma joukossa $1, \dots, 6$.

Bernoullin jakauma

- Olkoon A on tapahtuma, ja merkitään $p = P(A)$, jolloin $0 \leq p \leq 1$.
- Indikaattorilla $X = 1_A$ on Bernoullin jakauma parametrilla p (eli X noudattaa Bernoullin jakaumaa parametrillä p). Tämä asia voidaan merkitä $X \sim \mathrm{Bernoulli}(p)$.
- Tavallisesti sanotaan, että satunnaiskoe onnistuu silloin, kun $X = 1$ ja epäonnistuu muuten. Tällöin p:stä käytetään nimeä onnistumistodennäköisyys.

Bernoullin jakauman ptnf

- X:n arvojoukko on $\{0, 1\}$, ja sen ptnf on

$$f(x) = f(x \mid p) = \begin{cases} 1 - p, & \text{kun } x = 0, \\ p, & \text{kun } x = 1, \\ 0, & \text{muuten.} \end{cases}$$

- Merkintäsopimus: ptnf:n arvot tarvitsee mainita vain ko. sm:n arvojoukossa, muualla ymmärretään sen olevan nolla.
- Jakauman $\mathrm{Bernoulli}(p)$ ptnf voidaan kirjoittaa lyhyesti

$$f(x \mid p) = p^x (1 - p)^{1-x}, \qquad \text{kun } x = 0, 1,$$

Tässä tulkitaan $0^0 = 1$, mikäli $p = 0$ tai $p = 1$.

Jakaumien merkinnöistä

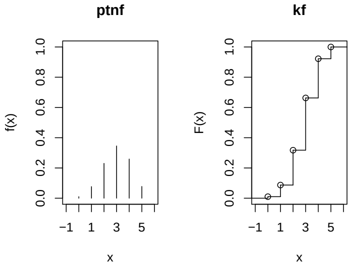
- Useimmat mielenkiintoiset jakaumat riippuvat yhden tai useamman parametrin arvosta. *Oikeastaan tarkastellaan saman tien perhettä jakaumia.*
- Pitääksemme lukua parametrin arvosta, merkitsemme sen tarvittaessa pystyviivan oikealle puolelle. Tätä merkintää voidaan käyttää kertymäfunktioille, pistetodennäköisyysfunktiolle ynnä muille funktioille.
- Jos taas sekaantumisen vaaraa ei ole, parametri voidaan jättää pois merkinnöistä.

Binomijakauma

- Olkoon $n \geq 1$ kokonaisluku ja $0 \leq p \leq 1$.
- Sm X noudattaa binomijakaumaa parametreilla n ja p, jos sen arvojoukko on $\{0, 1, \dots, n\}$ ja ptnf on

$$f(x) = f(x \mid n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad \text{kun } x = 0, 1, \dots, n.$$

- Tämä voidaan merkitä $X \sim \mathrm{Bin}(n, p)$.
- *Kuva:* Binomijakauman $\mathrm{Bin}(5, 0.6)$ ptfn ja kf.



Binomijakauman toistokoetulkinta

- Binomijakaumaa $\mathrm{Bin}(n, p)$ noudattava sm syntyy toistokokeessa, jossa toistetaan riippumattomasti n kertaa sellaista (Bernoullin) koetta, jossa yhdessä kokeessa onnistumistodennäköisyys on p.
- Jos X on onnistumisten lukumäärä n toistossa, niin $X \sim \mathrm{Bin}(n, p)$.
- Tulos perustellaan myöhemmin.
- Bernoullin jakauma $\mathrm{Bernoulli}(p)$ on tietenkin sama kuin binomijakauma $\mathrm{Bin}(1, p)$.

2.5 Integraalilaskennan kertausta

- Matemaattisessa analyysissä on määritelty erilaisia integrointiprosesseja, joista sinulle (tässä vaiheessa opintoja) ainoastaan Riemannin integraali on tuttu.
- Todennäköisyyslaskennan teoriassa Riemannin integraalin käyttö johtaisi tiettyihin asian kannalta epäoleellisiin hankaluuksiin, minkä takia matemaatikot käyttävät sen sijasta ns. Lebesguen integraalia.
- Tämän kurssin puitteissa näiden kahden integraalikäsitteen eroilla ei ole merkitystä, vaan kurssin tehtävät saadaan ratkaistua käyttämällä tuttua Riemannin integraalia, mutta usein tarvitaan ns. epäoleellista Riemannin integraalia.

Antiderivaatta eli integraalifunktio

Määritelmä

Olkoon f välillä I määritelty reaalifunktio. Tällöin sanomme, että F on funktion f antiderivaatta eli integraalifunktio, jos

$$F'(x) = f(x), \qquad \text{kaikilla } x \in I. \tag{2}$$

(Välin päätepisteissä derivaatta tässä tarvittaessa määritellään toispuoleisena derivaattana.)

- Jos F on f:n antiderivaatta ja C on vakio, niin myös $F + C$ on f:n antiderivaatta. Antiderivaattoja on siis monta!
- Jokaisella suljetulla äärellisellä välillä $I = [a, b]$ määritellyllä jatkuvalla funktiolla f on olemassa antiderivatta välillä I.

Riemannin integraalin määritelmästä

- Funktion f Riemannin interaali välin $[a, b]$ yli määritellään eräänlaisena summien raja-arvona, mikäli ko. raja-arvo on olemassa. Tällöin sanotaan, että f on Riemann-integroituva välin $[a, b]$ yli.
- Määritelmä toimii ainoastaan äärellisille integrointiväleille $[a, b]$, ja lisäksi integroitavan funktion eli integrandin f tulee on rajoitettu integrointivälillä.
- Integraalien arvoja lasketaan hyvin harvoin suoraan määritelmän perusteella, vaan tämän sijasta integraali tyypillisesti lasketaan antiderivaatan avulla.

Integraalin laskeminen antiderivaatan avulla

- Jos f on äärellisellä välillä $[a, b]$ määritelty jatkuva funktio, niin sen määrätty integraali (engl. *definite integral*) välin $[a, b]$ yli saadaan laskettua seuraavasti, jos tunnetaan jokin f:n antiderivaatta F:

$$\int_a^b f(x)\,\mathrm{d}x = \Big/_{a}^{b} F(x) = F(b) - F(a). \tag{3}$$

- Integrandin muuttaminen äärellisessä määrässä pisteitä ei muuta integraalin arvoa. Jos f on Riemann-integroituva välin $[a, b]$ yli ja $g = f$ kaikkialla muualla paitsi äärellisessä määrässä pisteitä, niin g on Riemann-integroituva, ja

$$\int_a^b g(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x. \tag{4}$$

Osittaisintegrointi

- Olkoot u ja v välillä $[a, b]$ määriteltyjä jatkuvasti derivoituvia funktioita.
- Kun otetaan huomioon tulon derivointikaava

$$\frac{\mathrm{d}}{\mathrm{d}x}[u(x)\,v(x)] = u'(x)\,v(x) + u(x)\,v'(x),$$

saadaan tulos

$$\Big/_{a}^{b} u(x)\,v(x) = \int_a^b (u'(x)\,v(x) + u(x)\,v'(x))\,\mathrm{d}x,$$

josta järjestelmällä saadaan osittaisintegrointikaava (engl. *integration by parts*)

$$\int_a^b u(x)\,v'(x)\,\mathrm{d}x = \Big/_{a}^{b} u(x)\,v(x) - \int_a^b u'(x)\,v(x)\,\mathrm{d}x. \qquad (5)$$

Sijoitus määrätyssä integraalissa

- Olkoon f jatkuva funktio välillä I, joka sisältää integrointivälin $[a, b]$, ja olkoon $g : [c, d] \to I$ jatkuvasti derivoituva funktio siten, että $g(c) = a$ ja $g(d) = b$. Olkoon F funktion f integraalifunktio.
- Tällöin

$$\int_a^b f(x)\,\mathrm{d}x = \Big/_{\!a}^{\,b} F(x) = F(g(d)) - F(g(c)) = \Big/_{\!c}^{\,d} F(g(t)).$$

- Toisaalta

$$\frac{\mathrm{d}}{\mathrm{d}t} F(g(t)) = f(g(t))\,g'(t), \qquad \text{kaikilla } t \in [c, d]$$

jota käyttämällä saadaan sijoituskeinon kaava

$$\int_a^b f(x)\,\mathrm{d}x = \int_c^d f(g(t))\,g'(t)\,\mathrm{d}t. \tag{6}$$

Sijoituskeino formaalisti

Formaalisti kaavassa

$$\int_a^b f(x)\,\mathrm{d}x = \int_c^d f(g(t))\,g'(t)\,\mathrm{d}t.$$

sijoitetaan integaaliin

- $x = g(t)$
- $\mathrm{d}x = g'(t)\,\mathrm{d}t.$
- Lisäksi alkuperäiset x-rajat (a, b) vaihdetaan vastaaviksi t-rajoiksi (c, d), joille $a = g(c)$ ja $b = g(d)$.

Sijoituskeinon käyttämisestä

Sijoituskaava

$$\int_a^b f(x)\,\mathrm{d}x = \int_c^d f(g(t))\,g'(t)\,\mathrm{d}t$$

on yhtälö, joten sitä voi käyttää kumpaan suuntaan tahansa; usein käytetään oikealta vasemmalle.

- *Ideana* on tunnistaa monimutkaisesta integroitavasta lausekkeesta jokin osalauseke $g(t)$, jonka merkitseminen uudeksi muuttujaksi u yksinkertaistaa lauseketta.
- *Vaikeutena* on, että "jostain" pitää löytyä (tekijänä) kyseisen osalausekkeen derivaatta $g'(t)$.
- Aina voi kokeilla monenlaisia sijoituksia, mutta vaatii taitoa ja kokemusta löytää toimiva sijoitus joka todella yksinkertaistaa integraalin johonkin helpompaan muotoon.

Sijoituskeino: Helppo esimerkki

Muistetaan, että $\int \sin x \, \mathrm{d}x = -\cos x + C$. Yritetään laskea

$$\int_0^{\pi/2} (\sin 2x) \, \mathrm{d}x.$$

- Toimisiko sijoitus $u = g(x) = 2x$?
- Nyt $\mathrm{d}u = g'(x)\mathrm{d}x = 2 \, \mathrm{d}x$. Vakiotekijä 2 on helppo järjestää.

$$\begin{aligned} \int_0^{\pi/2} (\sin 2x) \, \mathrm{d}x &= \frac{1}{2} \int_{x=0}^{x=\pi/2} (\sin 2x) \, 2 \, \mathrm{d}x \\ &= \frac{1}{2} \int_{u=0}^{u=\pi} (\sin u) \, \mathrm{d}u = \frac{1}{2}(-\cos \pi + \cos 0) = 1. \end{aligned}$$

Huomaa integrointirajat! Havainnollisuuden vuoksi tässä on merkitty, minkä muuttujan rajoista on kysymys. Piirrä kuva!

Sijoituskeino: Muita esimerkkejä

Samankaltainen vakiotekijän eliminointi tulee kohta vastaan ns. eksponenttijakauman tiheysfunktion integraalissa, joka on muotoa

$$\int \lambda e^{-\lambda x}\,\mathrm{d}x.$$

Tämä sujuu helposti sijoituksella $u = -\lambda x$.

Entäpä jos integraali onkin tällainen?

$$\int_0^\infty x\, e^{-x^2}\,\mathrm{d}x$$

Integraali sievenee huomattavasti sijoituksella $u = x^2$. Huomaa differentiaali $\mathrm{d}u = 2x\,\mathrm{d}x$, johon tarvittava derivaatta $2x$ onnekkaasti löytyy integroitavasta (paitsi vakiotekijää 2, joka ei tuota ongelmia).

Mitä ehtoja tarvitaan, kun integraali lasketaan antiderivaatan avulla?

- Kaavan (3) kohdalla vaadittiin, että integrandin f tulee olla jatkuva funktio kyseessä olevalla äärellisellä suljetulla välillä ennen kuin integraalin saa laskettua antiderivaatan F avulla.
- Jos f on muuten jatkuva, mutta sillä on hyppy pisteessä $a < c < b$, niin tällöin integrointialue voidaan pilkkoa jakopisteellä c väleihin $[a, c]$ ja $[c, b]$ ja yrittää laskea integraali summana

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x.$$

  Suotuisassa tapauksessa kumpikin integraaleista saadaan laskettua antiderivaatan avulla.

Milloin tarvitaan epäoleellista integraalia?

- Jos (a) integrointivälin päätepiste on äärettömyydessä tai jos (b) funktio f kasvaa rajatta integrointivälin päätepisteessä, niin Riemannin integraalin perusmääritelmä ei tuota mitään tulosta.
- Tällöin integraali joudutaan laskemaan ns. epäoleellisena (engl. *improper*) integraalina.
- Tämä tarkoittaa käytännössä raja-arvon laskemista sellaisesta integraalista, jossa ongelmallinen päätepiste on korvattu hyvin käyttäytyvällä pisteellä, jonka sitten annetaan lähestyä alkuperäistä ongelmallista päätepistettä.

Esimerkki, jossa integrandi kasvaa rajatta

- Laskemme integraalin $\int_0^1 x^{-a}\,\mathrm{d}x$, jossa $0 < a < 1$.
- Tämä integraali on epäoleellinen, sillä funktio x^{-a} kasvaa rajatta, kun $x \to 0+$.
- Jos $\epsilon > 0$, niin integraali välin $[\epsilon, 1]$ yli saadaan laskettua antiderivaatan avulla, sillä

$$\int_\epsilon^1 x^{-a}\,\mathrm{d}x = \Big/_{\epsilon}^{1} \frac{x^{-a+1}}{-a+1} = \frac{1}{1-a} - \frac{\epsilon^{1-a}}{1-a}.$$

  Tässä $\epsilon^{1-a} \to 0$, kun $\epsilon \to 0+$ sen takia, että $1 - a > 0$.
- Koska raja-arvo on olemassa, niin alkuperäinen integraali saadaan laskettua epäoleellisena integraalina, ja

$$\int_0^1 x^{-a}\,\mathrm{d}x = \lim_{\epsilon \to 0+} \int_\epsilon^1 x^{-a}\,\mathrm{d}x = \frac{1}{1-a}.$$

Esimerkki, jossa toinen integrointiraja on äärettömyydessä

- Laskemme integraalin $\int_1^\infty x^{-a}\,\mathrm{d}x$, kun $a > 1$.
- Integraali on epäoleellinen sen takia, että ylempi päätepiste on ääretön.

$$\int_1^\infty x^{-a}\,\mathrm{d}x = \lim_{M\to\infty}\int_1^M x^{-a}\,\mathrm{d}x = \lim_{M\to\infty} \Big/_1^M -\frac{x^{1-a}}{a-1}$$
$$= \lim_{M\to\infty}\left(-\frac{M^{1-a}}{a-1}+\frac{1}{a-1}\right) = \frac{1}{a-1}.$$

- Lausekkeessa M^{1-a} potenssi on negatiivinen, joten tämä lauseke lähestyy nollaa, kun $M \to \infty$.

2.6 Jatkuva jakauma

Määritelmä (Jatkuva jakauma, tiheysfunktio)

Satunnaismuuttujalla X on jatkuva jakauma tiheysfunktiolla f (lyh. tf), jos f on ei-negatiivinen ja jos kaikilla $a < b$

$$P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x$$

Tiheysfunktio on englanniksi (probability) density function, pdf. Toisinaan puhutaan tiheysfunktion sijasta lyhyesti tiheydestä. Sen sijaan, että sanotaan että sm:lla X on jatkuva jakauma voidaan sanoa että

- X:n jakauma on jatkuva,
- X on jatkuvasti jakautunut,
- X on jatkuva satunnaismuuttuja.

Merkintäsopimuksia

- Jatkuvan jakauman tiheysfunktiolle käytetään muuten samaa symbolia kuin sen kf:lle, mutta tiheysfunktion symboli kirjoitetaan pienellä kirjaimella (esim. f, g tai f_X) ja kf:n symboli vastaavalla isolla kirjaimella (esim. F, G tai F_X).
- Joskus tiheysfunktio häviää jonkin joukon $B \subset \mathbb{R}$ ulkopuolella. Tällöin tiheysfunktion kaava voidaan ilmoittaa vain joukossa B, ja asiayhteydestä pitää ymmärtää, että $f(x) = 0$, kun $x \notin B$.

Mikä integraalikäsite?

- Jatkuvan jakauman määritelmässä esiintyvä integraali tarkoittaa itseasiassa Lebesguen integraalia, mutta sovelluksissa esiintyvät tiheysfunktiot ovat jatkuvia funktioita kaikkialla muualla paitsi äärellisessä määrässä pisteitä, ja niiden integraalit voidaan käytännössä laskea Riemannin integraalien avulla.
- Integrointialue ei välttämättä ole äärellinen väli.
- Tiheysfunktio voi olla yhdessä tai useammassa pisteessä singulaarinen siten, että ainakin toinen sen toispuoleisista raja-arvoista on ääretön.
- Integrointialue joudutaan usein ensin pilkkomaan paloihin, minkä jälkeen tiheysfunktion integraali saadaan palautettua summaksi (mahdollisesti epäoleellisia) Riemannin integraaleja.

Tiheysfunktion (epä)yksikäsitteisyydestä

- Jakauman tiheysfunktio ei ole yksikäsitteinen, mutta melkein yksikäsitteinen.
- Jos f on tietyn jakauman tf ja g on yhtä kuin f muualla paitsi äärellisessä määrässä pisteitä, niin

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^b g(x)\,\mathrm{d}x$$

  kaikilla $a < b$. Tämän takia myös g on kyseisen jakauman tf.
- Ei-negatiiviset funktiot f ja g ovat saman jakauman tiheysfunktioita, jos ja vain jos ne yhtyvät melkein kaikkialla, eli joukko jossa $f \neq g$ on nollamittainen (mikä käsite määritellään reaalianalyysin kursseilla).

Välien sattumistodennäköisyydet

- Jos X:n jakauma on jatkuva ja sen tf on f, niin

$$P(X = x) = P(x \le X \le x) = \int_x^x f(u)\,\mathrm{d}u = 0$$

kaikilla x, joten pistetodennäköisyys on aina nolla.

- Ptnf ei ole hyödyllinen käsite jatkuville jakaumille.
- Koska kaikki pistetodennäköisyydet ovat nollia, niin

$$\begin{aligned} P(a < X < b) = P(a \le X < b) &= P(a < X \le b) \\ &= P(a \le X \le b) = \int_a^b f(x)\,\mathrm{d}x, \end{aligned}$$

kun $a < b$.

Toinen tapa karakterisoida jatkuva jakauma

- Rajankäynnin jälkeen huomataan, että puoliäärettömien välien todennäköisyydet saadaan myöskin tf:a integroimalla. Esim.

$$P(X \le x) = P(X < x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u.$$

- Itse asiassa pätee

$$P(X \in B) = \int_B f(x)\,\mathrm{d}x, \qquad \text{kaikille } B \subset \mathbb{R}, \tag{7}$$

jossa integraalin alaindeksi B tarkoittaa joukon B yli laskettua integraalia, ts.

$$\int_B f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} 1_B(x)\, f(x)\,\mathrm{d}x.$$

Kertymäfunktion ja tiheysfunktion yhteys

Lause (Kertymäfunktion ja tiheysfunktion yhteys)

Olkoon X:llä jatkuva jakauma.

(a) Jos X:n tiheysfunktio on f, niin sen kertymäfunktio F on

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

(b) Jos X:n kertymäfunktio on F, niin F on derivoituva melkein kaikkialla, ja X:n tiheysfunktioksi voidaan valita F'.

Huomautus: Kun sanotaan, että tiheysfunktioksi voidaan valita $f = F'$, niin tarkoitetaan sitä, että tiheysfunktioksi valitaan jokin funktio f, joka yhtyy funktion F derivaattaan niissä pisteissä, joissa derivaatta on määritelty, ja muissa pisteissä f saadaan määritellä mielivaltaisesti.

Eräs tapa tunnistaa kertymäfunktiosta, että jakauma on jatkuva

Lause

Olkoon F sellainen kertymäfunktio, että

(1) F on jatkuva koko reaaliakselilla,

(2) F on derivoituva kaikkialla muualla paitsi mahdollisesti äärellisessä määrässä pisteitä,

(3) ja lisäksi derivaatta $f = F'$ on jatkuva kaikkialla muualla paitsi mahdollisesti äärellisessä määrässä pisteitä.

Tällöin F on jatkuvan jakauman kertymäfunktio, ja jakauman tiheysfunktioksi voidaan valita sen derivaatta $f = F'$.

Todistuksen perusidea

Jos $[a, b]$ on sellainen äärellinen väli, joka ei sisällä yhtään pistettä, jossa F:n derivaatta ei ole jatkuva, niin

$$F(b) - F(a) = \int_a^b F'(u)\,\mathrm{d}u.$$

Lisäksi otetaan raja-arvoja ja käytetään kertymäfunktion ominaisuuksia.

Huomautuksia

- Jos X:llä on jatkuva jakauma, niin sen kf on jatkuva. Sen sijaan tf voi olla hyvin epäsäännöllinen funktio. Tilastollisissa sovelluksissa esiintyvät tiheysfunktiot ovat varsin säännöllisiä funktioita, sillä ne ovat jatkuvia kaikkialla paitsi äärellisessä määrässä pisteitä. Tf:n ei tarvitse olla rajoitettu funktio.
- Täydellinen karakterisointi: kf F on jatkuvan jakauman kf täsmälleen silloin, kuin F on absoluuttisesti jatkuva, mutta tämän asian määritteleminen jätetään reaalianalyysin kurssien huoleksi.
- Täsmällisempi nimitys jatkuvalle jakaumalle olisi *absoluuttisesti jatkuva jakauma*.

Kuinka tunnistan, että funktio on tiheysfunktio?

Jos X:llä on tiheys f, niin

$$1 = P(\Omega) = P(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x,$$

joten tiheysfunktion integraali koko reaaliakselin yli on yksi.

Lause

Jos $f : \mathbb{R} \to \mathbb{R}$ on ei-negatiivinen, ja sen integraali koko reaaliakselin yli on yksi, niin se on jonkin satunnaismuuttujan tiheysfunktio.

Normalisoimaton tiheysfunktio

Lause

Jos $g \geq 0$ on sellainen reaaliakselilla määritelty funktio, jonka integraali $\int_{-\infty}^{\infty} g(x)\,\mathrm{d}x$ on äärellinen ja aidosti positiivinen, niin tällöin on olemassa vakio k siten, että $f = k\,g$ on tiheysfunktio.

Todistus Valitaan $1/k = \int_{-\infty}^{\infty} g(x)\,\mathrm{d}x$, jonka jälkeen

$$\int f(x)\,\mathrm{d}x = k \int g(x)\,\mathrm{d}x = \frac{k}{k} = 1.$$

Huomautus. Tällä tavalla jokainen normalisoimaton tiheysfunktio g määrää yksikäsitteisellä tavalla jatkuvan jakauman, jonka tiheysfunktio on edellä konstruoitu f.

Tiheysfunktion arvioiminen aineistosta

- Olkoon sm:lla X tiheysfunktio f, jota emme tunne.
- Oletetaan, että meillä on aineisto $x_1, x_2, \ldots, x_N$, jota voidaan pitää N riippumattoman sm:n X:n jakaumaa noudattavan sm:n havaittuina arvoina.
- Toisin sanoen

$$x_1 = X_1(\omega^{\mathrm{act}}), x_2 = X_2(\omega^{\mathrm{act}}), \ldots x_N = X_N(\omega^{\mathrm{act}}),$$

  jossa ω^{act} on aktualisoitunut alkeistapaus, kullakin X_i on sama jakauma kuin X:llä ja lisäksi $X_1, \ldots, X_N$ ovat riippumattomia (mikä määritellään myöhemmin).
- Kuinka tällaisessa tilanteessa voidaan arvioida tiheysfunktion arvoa pisteessä u? Oletetaan, että u on f:n jatkuvuuspiste.

Eräs ratkaisu

- Valitaan jokin pieni luku $h > 0$ ja lasketaan, kuinka moni otospisteistä x_i osuu h:n pituiselle välille
$$[u - \frac{h}{2}, u + \frac{h}{2}].$$
Olkoon tämä lukumäärä $N_{u,h}$.
- Tiheysfunktion määritelmän ja frekvenssitulkinnan nojalla
$$P(u - \frac{h}{2} \le X \le u + \frac{h}{2}) = \int_{u-h/2}^{u+h/2} f(x)\,\mathrm{d}x \approx \frac{N_{u,h}}{N}.$$
- Toisaalta, integraalilaskennan väliarvolauseen nojalla
$$\int_{u-h/2}^{u+h/2} f(x)\,\mathrm{d}x \approx h\,f(u).$$
- Yhdistämällä saadaan arvio $f(u) \approx \frac{N_{u,h}}{h\,N}$. Tässä voidaan käyttää myös muita h:n pituisia välejä, jotka sisältävät pisteen u kuin tämä.

2.7 Esimerkkejä jatkuvista jakaumista

- tasajakauma
- eksponenttijakauma
- gammajakauma, beetajakauma, normaalijakauma, khiin neliön jakauma, t-jakauma, F-jakauma (luvussa 5).

Tasajakauma

- Olkoon $a, b \in \mathbb{R}$ ja $a < b$. Satunnaismuuttujalla X on välin (a, b) tasajakauma, $X \sim \mathrm{U}(a, b)$, jos sillä on jatkuva jakauma tiheysfunktiona f, jossa

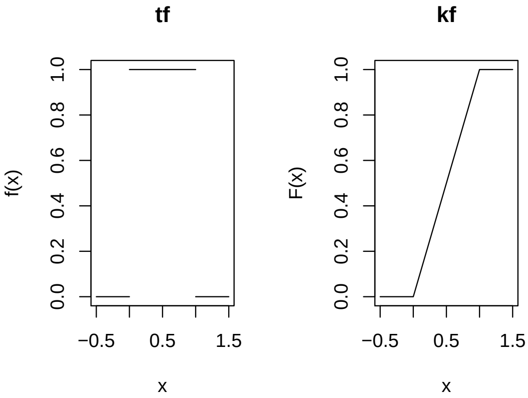
$$f(x) = f(x \mid a, b) = \frac{1}{b - a}, \qquad \text{kun } a < x < b.$$

  (Tästä kaavasta pitää ymmärtää, että $f(x) = 0$, kun $x \leq a$ tai $x \geq b$.) Tiheys on siis *vakio* koko välillä (a, b).

- Kertymäfunktio on

$$F(x) = \int_{-\infty}^{x} f(u) \, \mathrm{d}u = \begin{cases} 0, & \text{kun } x \leq a, \\ \dfrac{x - a}{b - a}, & \text{kun } a < x < b, \\ 1, & \text{kun } x \geq b. \end{cases}$$

Tasajakauman $\mathrm{U}(0,1)$ tiheys- ja kertymäfunktio



Eksponenttijakauma (engl. *exponential distribution*)

- Satunnaismuuttujalla X on eksponenttijakauma parametrilla $\lambda > 0$, $X \sim \mathrm{Exp}(\lambda)$, jos X:llä on tf

$$f(x) = f(x \mid \lambda) = \lambda\, \mathrm{e}^{-\lambda x}, \qquad \text{kun } x > 0.$$

- Kun $x > 0$, niin eksponenttijakauman kf:lla on arvo

$$F(x) = \Big/_{0}^{x} (-\mathrm{e}^{-\lambda u}) = 1 - \mathrm{e}^{-\lambda x},$$

ja $F(x) = 0$, kun $x \leq 0$.

2.8 Satunnaismuuttujan muunnos

- Jos X on sm, ja $g : \mathbb{R} \to \mathbb{R}$ on funktio, niin myös $Y = g(X)$ on sm.
- Merkintä $Y = g(X)$ tarkoittaa sitä, että

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

  ts. Y on yhdistetty funktio $g \circ X$.
- Mitä voimme sanoa Y:n jakaumasta?

Alkukuva

Jos $g : \mathbb{R} \to \mathbb{R}$ on funktio ja $A \subset \mathbb{R}$, niin merkintä $g^{-1}(A)$ tarkoittaa joukon A alkukuvaa kuvauksessa g, eli

$$g^{-1}(A) = \{x \in \mathbb{R} : g(x) \in A\}.$$

Huomaa, että

$$g(x) \in A \quad \iff \quad x \in g^{-1}(A).$$

Muunnoksen $g(X)$ jakauma diskreetissä tapauksessa

Lause

Olkoon X diskreetti sm ptnf:lla f_X, ja olkoon $g : \mathbb{R} \to \mathbb{R}$ funktio. Tällöin sm $Y = g(X)$ diskreetti, ja sen ptnf f_Y on

$$f_Y(y) = \sum_{x \in g^{-1}(\{y\})} f_X(x).$$

Todistus: Y voi saada korkeintaan numeroituvasti äärettömän määrän eri arvoja, joten Y on diskreetti. Mille tahansa $y \in \mathbb{R}$ on

$$\begin{aligned} P(Y = y) &= P(g(X) = y) = P(g(X) \in \{y\}) = P(X \in g^{-1}(\{y\})) \\ &= \sum_{x \in g^{-1}(\{y\})} f_X(x), \end{aligned}$$

Erikoistapaus

- Jos yhtälöllä $y = g(x)$ on yksikäsitteinen ratkaisu $x = h(y)$ kaikilla y, jotka kuuluvat sm:n Y arvojoukkoon S_Y, niin

$$f_Y(y) = f_X(h(y)), \qquad y \in S_Y. \tag{8}$$

- Jatkuvan jakauman tilanteessa saadaan erilainen kaava.
- Jos taas ratkaisuja on useita, niin jokainen erisuuri ratkaisu antaa ptnf:n arvoon pisteessä y yhden tällaisen termin lisää.

Jatkuvasti jakautuneen satunnaismuuttujan muunnos

- Jos X:llä on jatkuva jakauma, niin sen muunnoksen $Y = g(X)$ jakauma voi funktion g luonteesta riippuen joko diskreetti, jatkuva tai ei kumpaakaan näistä.
- Esim. jos g on porrasfunktio, niin $Y = g(X)$ on diskreetti satunnaismuuttuja.
- Myöhemmin tarkastelemme tapausta, jossa satunnaismuuttujalla $Y = g(X)$ on jatkuva jakauma.
- Luentomonisteessa on esimerkki, jossa muunnoksella $Y = g(X)$ ei ole diskreetti eikä jatkuva jakauma, vaan ns. sekatyypin jakauma.

2.9 Kvantiilifunktio ja sen käyttö simuloinnissa

- Tarkastelemme sellaista jatkuvaa jakaumaa, jonka tiheysfunktio f on aidosti positiivinen välillä (a, b) ja nolla muualla.
- Päätepisteille sallitaan reaalisten arvojen lisäksi arvot $a = -\infty$ tai $b = \infty$.
- Tällaisen jakauman kertymäfunktio F, on aidosti kasvava välillä (a, b), ja lisäksi $F(a) = 0$ ja $F(b) = 1$.

Kvantiilifunktio on kertymäfunktion käänteisfunktio

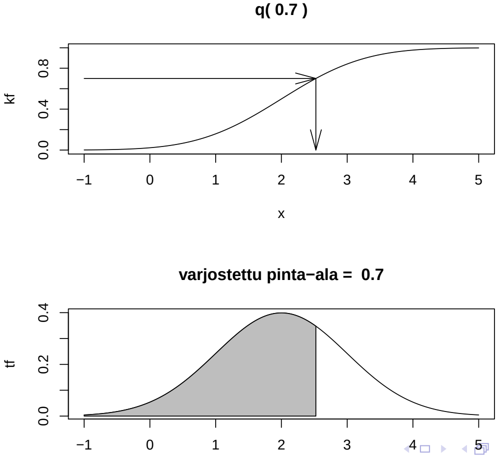
- Oletusten ansiosta funktiolla F on olemassa käänteisfunktio $F^{-1} : (0,1) \to (a,b)$. (Tarkemmin sanoen F:n rajoittumalla joukkon (a,b) on käänteisfunktio F^{-1}.)
- Merkintä F^{-1} ei tarkoita funktiota $1/F$, vaan kertymäfunktion käänteisfunktiota.
- Käytännössä $F^{-1}(u)$ määritetään ratkaisemalla x yhtälöstä

$$F(x) = u, \qquad 0 < u < 1.$$

- Tällaista (avoimella välillä $(0,1)$ määriteltyä) kertymäfunktion käänteisfunktiota F^{-1} kutsutaan kyseisen jakauman kvantiilifunktioksi (engl. *quantile function*). (Myös nimitys fraktiilifunktio on käytössä.)

Kvantiilifunktion määritelmää havainnollistava kuva

Erään jatkuvan jakauman kvantiilifunktion $q = F^{-1}$ arvo pisteessä $u = 0.7$. Ylempänä on kertymäfunktio F ja alempana tiheysfunktio f.



Kvantiilifunktion merkitys

- Osuus u jakauman todennäköisyysmassasta jää pisteen $F^{-1}(u)$ vasemmalle puolelle eli vasemmanpuoleiseen häntään: jos sm:lla X on kf F, niin

$$P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u, \qquad \text{mielivaltaiselle } 0 < u < 1.$$

- Ts. tiheysfunktion f alle jää pisteen $F^{-1}(u)$ vasemmalle puolelle alue, jonka pinta-ala on u, sillä

$$u = P(X \leq F^{-1}(u)) = \int_{-\infty}^{F^{-1}(u)} f(x)\,\mathrm{d}x.$$

- Koska jakauma on jatkuva, niin pisteen $F^{-1}(u)$ oikealle puolelle jää jakauman todennäköisyysmassasta osuus $1 - u$, eli tiheysfunktion alle jäävän oikeanpuoleisen häntäalueen pinta-ala on $1 - u$.

Alakvantiilit ja yläkvantiilit

- Kvantiilifunktio määriteltiin tarkastelemalla jakauman vasemmanpuoleista häntää.
- Tilastollisessa päättelyssä määritetään testien kriittisiä pisteitä usein jakauman oikeanpuoleisen hännän avulla.
- Selvyyden vuoksi voimme puhua jakauman *alakvantiileista* ja *yläkvantiileista*.
- Jos $0 < u < 1$, niin jakauman u-alakvantiili (eli u-kvantiili tai u-fraktiili) on sellainen piste, josta vasemmalle jää jakaumasta osuus u. Ts. u-alakvantiili on $F^{-1}(u)$.
- Jakauman u-yläkvantiili taas on sellainen piste, josta oikealle jää jakaumasta osuus u, joten u-yläkvantiili on $F^{-1}(1-u)$, sillä

$$P(X \geq F^{-1}(1-u)) = 1 - P(X < F^{-1}(1-u)) = 1 - (1-u) = u.$$

Erikoisnimityksiä tietyille kvantiileille

- mediaani on $\frac{1}{2}$-kvantiili,
- alakvartiili on $\frac{1}{4}$-kvantiili,
- yläkvartiili on $\frac{3}{4}$-kvantiili.
- Näiden lisäksi puhutaan myös kvintiileistä, desiileistä ja persentiileistä.

Huom! Tässä kvantiilit määriteltiin pisteinä, joiden alapuolelle jää tietty osuus jakaumasta. Tilastotieteessä esim. kvartiililla tai desiilillä saatetaan tarkoittaa myös jakauman osaa, esim. "tulonjakotilaston alakvartiili" $=$ tuloiltaan alin neljäsosa väestöstä; "III desiili" $=$ se osa joka on kvantiilipisteiden $\frac{2}{10}$ ja $\frac{3}{10}$ välissä.

Kertymäfunktion käänteisfunktion ominaisuuksia

- Funktio F^{-1} toteuttaa identiteetit

$$F^{-1}(F(x)) = x, \quad \text{kaikilla} \quad a < x < b,$$

  ja

$$F(F^{-1}(u)) = u, \quad \text{kaikilla} \quad 0 < u < 1.$$

- Koska F on aidosti kasvava funktio, niin myös F^{-1} on aidosti kasvava funktio.
- Jos jompaa kumpaa sovelletaan epäyhtälön molemmille puolille, niin epäyhtälön ratkaisujoukko säilyy muuttumattomana.

Kertymäfunktiomuunnos

- Olkoon X sm, jolla on kertymäfunktio F. Tarkastellaan satunnaismuuttujaa $Y = F(X)$.
- Jos $0 < u < 1$, niin

$$\begin{aligned} P[Y \leq u] = P[F^{-1}(Y) \leq F^{-1}(u)] &= P[X \leq F^{-1}(u)] \\ &= F(F^{-1}(u)) = u. \end{aligned}$$

- Tämä tarkoittaa sitä, että

$$Y = F(X) \sim \mathrm{U}(0,1). \tag{9}$$

- Satunnaismuuttujaan $F(X)$, ja tulokseen $F(X) \sim \mathrm{U}(0,1)$ viitataan toisinaan nimellä kertymäfunktiomuunnos (engl. *probability integral transform, PIT*).

Käänteisfunktiomenetelmä

- Olkoon $U \sim \mathrm{U}(0,1)$, ja tarkastellaan satunnaismuuttujaa $Z = F^{-1}(U)$.
- Jos $a < x < b$, niin

$$
\begin{aligned}
P(Z \leq x) = P[F^{-1}(U) \leq x] &= P[F(F^{-1}(U)) \leq F(x)] \\
&= P[U \leq F(x)] = F(x)
\end{aligned}
$$

- Siis

$$
U \sim \mathrm{U}(0,1) \quad \Rightarrow \quad \text{sm:n } F^{-1}(U) \text{ kertymäfunktio on } F.
$$

Tällöin sanotaan, että sm $Z = F^{-1}(U)$ on saatu käänteisfunktiomenetelmällä (engl. *inverse transform method*).

Käänteisfunktiomenetelmä simuloinnissa

- Matemaattisissa ohjelmistoissa on yleensä satunnaislukugeneraattori, jonka tuottamia arvoja voidaan pitää riippumattomina jakaumaa $\mathrm{U}(0,1)$ noudattavien satunnaismuuttujien arvoina.
- Jos jakauman kvantiilifunktiolla on yksinkertainen lauseke, niin käänteisfunktiomenetelmä antaa kätevän keinon simuloida kyseistä jakaumaa:

$$\text{generoi } u \text{ jakaumasta } \mathrm{U}(0,1)\text{, ja laske } x = F^{-1}(u).$$

Esimerkki: eksponenttijakauman $\mathrm{Exp}(1)$ simulointi

- Kf on

$$F(x) = \begin{cases} 1 - \mathrm{e}^{-x}, & \text{jos } x \geq 0 \\ 0, & \text{muuten,} \end{cases}$$

- Kvantiilifunktion arvo $F^{-1}(u)$ pisteessä $0 < u < 1$ saadaan ratkaisemalla x yhtälöstä

$$F(x) = u \quad \Leftrightarrow \quad x = -\ln(1-u).$$

- Tämän takia jakaumaa $\mathrm{Exp}(1)$ voidaan simuloida asettamalla $X = -\ln(1-U)$, kun U on ensin generoitu jakaumasta $\mathrm{U}(0,1)$.

Kvantiilifunktion käyttö empiiriseen tarkasteluun

- Aiemmin todettiin: Jos X noudattaa tiettyä jakaumaa, jonka kertymäfunktio on F (voidaan merkitä: $X \sim F$), niin muunnoksen $Y = F(X)$ jakauma on tasajakauma $U(0, 1)$.
- Olkoon käytettävissä otos $x_1, x_2, \ldots x_n$ ja hypoteesi (arvelu), että ne ovat peräisin jakaumasta, jonka kertymäfunktio on F.
- Lasketaan muunnokset $y_1 = F(x_1)$ jne. Nyt saadaan n lukua, joiden pitäisi oletuksen ja kertymäfunktiomuunnoksen mukaan olla otos tasajakaumasta $U(0, 1)$.
- Esim. kuvallisesti voidaan tutkia, "näyttääkö" otos $y_1, y_2, \ldots, y_n$ sellaiselta otokselta. Jos ei, on ehkä oletus $X \sim F$ väärä! Vrt. tilastotieteen *P-P plot*.

Huomautuksia

- Kvantiilifunktio voidaan yleistää mielivaltaiselle jakaumalle, ks. luentomoniste.
- Myös käänteisfunktiomenetelmä yleistyy, ja tämän jälkeen reseptiä $F^{-1}(U)$ voidaan pitää tapana konstruoida sm, jolla on haluttu kf F, mikäli ensin todistetaan että tasajakauma $\mathrm{U}(0,1)$ on olemassa.

2.10 Muunnetun satunnaismuuttujan tiheys

- Olkoon X:llä jatkuva jakauma, ja $g : \mathbb{R} \to \mathbb{R}$.
- Satunnaismuuttujalla $Y = g(X)$ ei tällöin välttämättä ole jatkuva jakauma.
- Jos kuitenkin g on riittävän sileä ja aidosti monotoninen, tai jos sen määrittelyjoukko $\mathbb{R}$ voidaan osittaa paloihin, joissa g on riittävän sileä ja aidosti monotoninen, niin osoittautuu, että tällöin Y:llä on jatkuva jakauma.

Kertymäfunktiotekniikka

- Jos muunnoksella $Y = g(X)$ on jatkuva jakauma, niin tiheysfunktio on mahdollista johtaa laskemalla ensin kertymäfunktio ja sitten kertymäfunktion derivaatta.
- Tämän lisäksi pitää jollakin tavalla tarkistaa, että Y:n jakauma todellakin on jatkuva.
- Tarkistus voidaan tehdä esim. suoraan tiheysfunktion määritelmän nojalla.
- Vaihtoehtona on käyttää lausetta 2.7 (moniste s. 27)

Kertymäfunktiotekniikka, $Y = \mathrm{e}^X$

- Olkoon X:llä jokin jatkuva jakauma, jolla on jatkuva tiheysfunktio f_X, ja tarkastellaan muunnosta $Y = \mathrm{e}^X$.
- Selvästi $Y > 0$, joten $F_Y(y) = 0$, kun $y < 0$. Kun $y > 0$, on

$$F_Y(y) = P(Y \le y) = P(\mathrm{e}^X \le y) = P(X \le \ln(y)) = F_X(\ln(y)),$$

- Lasketaan kf:n derivaatta:

$$f_Y(y) = F_Y'(y) = f_X(\ln(y))/y, \qquad \text{kun } y > 0,$$

  ja $f_Y(y) = 0$, kun $y \le 0$.
- Lauseen 2.7 avulla voidaan tarkistaa, että Y:n jakauma on jatkuva. Tämän jälkeen tiedetään, että F:n derivaatta kelpaa Y:n tiheysfunktioksi.

$Y = \mathrm{e}^X$: muuttujanvaihtotekniikka

- Tarkistetaan, että tn $P(a \le Y \le b)$ saadaan laskettua integroimalla tiettyä funktiota välin (a, b) yli, kun $a < b$ ovat mielivaltaisia. Koska $Y > 0$, voidaan olettaa, että $a > 0$.

$$
\begin{aligned}
P(a \le Y \le b) &= P(a \le \mathrm{e}^X \le b) \\
&= P(\ln a \le X \le \ln b) \\
&= \int_{\ln a}^{\ln b} f_X(x)\,\mathrm{d}x \qquad (X\text{:n tf on } f_X) \\
&= \int_a^b f_X(\ln(y))\frac{1}{y}\,\mathrm{d}y \quad (\text{Sij. } y = \mathrm{e}^x \Leftrightarrow x = \ln(y))
\end{aligned}
$$

- Esimerkin tekniikka yleistetään kohta lauseeksi.

Diffeomorfismi

Määritelmä (Diffeomorfismi)

Kuvaus $g : A \to B$, jossa A ja B ovat d-ulotteisen avaruuden $\mathbb{R}^d$ avoimia osajoukkoja, on diffeomorfismi, jos

(a) g on bijektio joukkojen A ja B välillä

(b) sekä g että sen käänteisfunktio $g^{-1} : B \to A$ ovat jatkuvasti derivoituvia.

Muuttujanvaihtokaava tiheysfunktioille

Lause (Muuttujanvaihtokaava tiheysfunktioille)

Olkoon sm:lla X jatkuva jakauma tf:lla f_X. Olkoon $g : A \to B$ diffeomorfismi, jossa $A, B \subset \mathbb{R}$ ovat avoimia välejä, ja olkoon

$$P(X \in A) = 1.$$

Määritellään $h(y) = g^{-1}(y)$, kun $y \in B$. Tällöin sm:lla $Y = g(X)$ on jatkuva jakauma tiheysfunktiolla

$$f_Y(y) = \begin{cases} f_X(h(y))\,|h'(y)|, & \text{kun } y \in B \\ 0, & \text{muuten.} \end{cases} \tag{10}$$

Huomautuksia lauseesta

- Tulos (10) voidaan ilmaista lyhyemmin kaavalla

$$f_Y(y) = f_X(h(y))\,|h'(y)|, \qquad \text{kun } y \in B.$$

- Kun muunnoksen $Y = g(X)$ tiheysfunktio ilmoitetaan, on tärkeää paitsi kertoa kaava $f_X(h(y))\,|h'(y)|$ myös ilmoittaa selkeästi kaavan pätevyysalue, eli se joukko B, jossa kyseinen kaava pätee.
- Vertaa jatkuvan jakauman tapauksen tulosta (10) vastaavan diskreetin tapauksen tulokseen (8). Vain jatkuvan jakauman tapauksessa tarvitaan käänteisfunktion derivaatan itseisarvoa.

Kaksi erilaista tekniikkaa

Muunnetun satunnaismuuttujan tiheysfunktio voidaan johtaa kahdella erilaisella tekniikalla:

- johdetaan kertymäfunktio ja lasketaan tiheysfunktio kertymäfunktion derivaattana,
- käytetään muuttujanvaihtokaavaa.

Huomautuksia kertymäfunktiotekniikasta

- Kertymäfunktiotekniikassa pitää erikseen tarkistaa, että saatu kertymäfunktio on jatkuvan jakauman kertymäfunktio.
- Ehdoton edellytys tälle asialle on se, että kertymäfunktion pitää olla jatkuva funktio koko reaaliakselilla.
- Lauseen 7 mukaan riittävää jakauman jatkuvuudelle on se, jos kertymäfunktio on lisäksi jatkuvasti derivoituva muualla paitsi äärellisessä määrässä poikkeuspisteitä. (Poikkeuspisteissä derivaatan ei välttämättä tarvitse olla olemassa.)

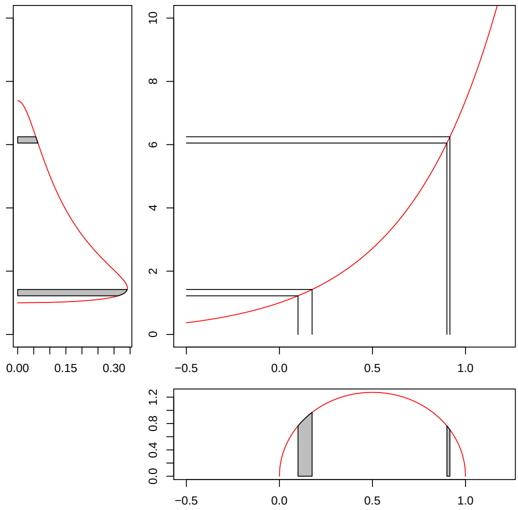
Huomautuksia muuttujanvaihtokaavasta

- Muuttujanvaihtokaavaa käytettäessä pitää miettiä, mitkä ovat avoimet joukot A ja B.
- Lisäksi käänteisfunktio h pitää pystyä ratkaisemaan yhtälöstä $g(x) = y$ (ja tämä ei onnistu ellei kyseessä ole bijektio).
- Tämän jälkeen on tyypillisesti helppo tarkistaa, onko lauseke $g(x)$ jatkuvasti derivoituva joukossa A ja onko lauseke $h(y)$ jatkuvasti derivoituva joukossa B.
- Jos nämä ehdot täyttyvät, niin kyseessä on diffeomorfismi.

Muistisääntö

- Jos g on diffeomorfismi, niin muuttujanvaihtokaavan muistamista helpottaa seuraava muistisääntö.
- Yhtälön
$$f_X(x)\,|\mathrm{d}x| = f_Y(y)\,|\mathrm{d}y| \tag{11}$$
pitää säilyä bijektiivisessä muuttujanvaihdossa
$$y = g(x) \quad \Longleftrightarrow \quad x = h(y).$$
- Kun tiedosta (11) ratkaistaan $f_Y(y)$, saadaan
$$f_Y(y) = f_X(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = f_X(h(y))\,|h'(y)|,$$
mutta tietenkin tämä kaava pätee vain g:n kuvajoukossa B.

Muistisääntö $f_X(x)\,|\mathrm{d}x| = f_Y(y)\,|\mathrm{d}y|$ infinitesimaalisen järkeilyn avulla



Muistisääntö on robusti

- Muistisäännöstä (11) voidaan ratkaista $f_Y(y)$ myös toisella tavalla, nimittäin seuraavasti,

$$f_Y(y) = f_X(x)\frac{1}{\left|\frac{\mathrm{d}y}{\mathrm{d}x}\right|} = \frac{f_X(x)}{|g'(x)|} = \frac{f_X(h(y))}{|g'(h(y))|}.$$

- Myös tämä kaava pitää paikkansa joukossa B, sillä kaava

$$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{1}{\frac{\mathrm{d}y}{\mathrm{d}x}}$$

  kertoo oikean yhteyden funktion ja sen käänteisfunktion derivaattojen välillä.

- Aidosta monotonisuudesta seuraa se, että $|g'| > 0$ joukossa A ja $|h'| > 0$ joukossa B.

Esimerkki $Y = \mathrm{e}^X$ muistisäännön avulla

- Tarkastelemme bijektiivistä vastaavuutta

$$y = \mathrm{e}^x \quad \Leftrightarrow \quad x = \ln(y)$$

pisteiden $x \in \mathbb{R}$ ja $y > 0$ välillä. Koska molemmat lausekkeet ovat jatkuvasti derivoituvia, on kyseessä diffeomorfismi.

- Ratkaisemme muistisäännöstä kaavan

$$f_Y(y) = f_X(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right|$$

- Koska tahdomme laskea f_Y:n arvon pisteessä y, pitää kaavan oikealla puolella x esittää y:n avulla kaavalla $x = \ln(y)$. Kaavassa tarvitaan myös tämän lausekkeen derivaattaa $1/y$. Kun teemme nämä sijoitukset, saamme tuloksen

$$f_Y(y) = f_X(\ln(y)) \, \frac{1}{y}, \qquad y > 0.$$

Yleistys

- Joskus tiheysfunktion muuttujanvaihtokaavaa tarvitaan myös sellaisessa tilanteessa, jossa g ei ole monotoninen, mutta jossa sen määrittelyjoukko voidaan pilkkoa väleiksi, joille rajoitettuna g on aidosti monotoninen.
- Tällaiset tilanteet saadaan hoidettua samaan tapaan kuin seuraavassa esimerkissä.

Esimerkki $Y = X^2$

- Olkoon X:llä jatkuva jakauma, jolla on jatkuva tiheysfunktio f_X. Tarkastellaan satunnaismuuttujaa $Y = X^2$.
- Selvästi $Y \geq 0$, joten $F_Y(y) = 0$, kun $y < 0$.
- Kun $y > 0$, on

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

- Tarkistetaan lauseen 7 avulla, että Y:n jakauma on jatkuva.
- Funktio F_X on jatkuvan jakauman kertymäfunktiona jatkuva, joten ainoa ongelmallinen piste on $y = 0$. Tässä pisteessä pitää tarkistaa F_Y:n jatkuvuus, mutta tarkistus on helppoa, koska $F_X(x)$ on jatkuva pisteessä $x = 0$.
- $F_Y(y)$ on derivoituva muualla paitsi mahdollisesti pisteessä $y = 0$.

Tiheysfunktio kertymäfunktiota derivoimalla

- Derivaatan laskemisen jälkeen näemme, että Y:n tf on

$$f_Y(y) = f_X(\sqrt{y})\,\frac{1}{2\sqrt{y}} + f_X(-\sqrt{y})\,\frac{1}{2\sqrt{y}}, \qquad \text{kun } y > 0,$$

  ja $f_Y(y) = 0$, kun $y \leq 0$.
- Kumpaakin yhtälön $y = x^2$ ratkaisua vastaa tässä yksi termi.

Muuttujanvaihtotekniikka muunnokselle $Y = g(X)$, kun g ei ole monotoninen

- Luentomonisteessa on muotoiltu kohtuullisen yleinen lause 2.13.
- Tiheysfunktioon $f_Y(y)$ saadaan summa termejä, yksi termi kutakin yhtälön $y = g(x)$ ratkaisua kohti. Kukin termeistä on samaa muotoa kuin diffeomorfismin tapauksessa.
- Jos lauseen 2.13 oletukset ovat voimassa, niin muunnoksen $Y = g(X)$ tiheysfunktio saadaan laskea myös kertymäfunktiotekniikalla: johdetaan ensin sm:n Y kertymäfunktio F_Y ja sitten derivoidaan tämä funktio.