

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 $X : \Omega \rightarrow \mathbb{R}$ on **satunnaismuuttuja**.

Esimerkkejä satunnaismuuttujista

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

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

on tuo silmäluku.

- Silmälukujen summa kahdessa nopanheitossa. Jos $E = \{1, \dots, 6\}$, 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 summan ilmoittaa

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

Tapahtuman indikaattori (tärkeä esimerkki satunnaismuuttujasta)

Määritelmä (Tapahtuman indikaattori)

Olkoon $A \subset \Omega$ tapahtuma. Sen **indikaattori** (eli ilmaisoin 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 $\Omega \rightarrow \mathbb{R}$, 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

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

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

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

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 $F : \mathbb{R} \rightarrow \mathbb{R}$

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

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

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

- 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 $G : \mathbb{R} \rightarrow \mathbb{R}$ 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, $-\infty$ tai ∞).
- Näitä raja-arvoja merkitään seuraavasti

$$G(x-) = \lim_{y \rightarrow x-} G(y),$$

$$G(x+) = \lim_{y \rightarrow x+} G(y)$$

$$G(-\infty) = \lim_{y \rightarrow -\infty} G(y),$$

$$G(\infty) = \lim_{y \rightarrow \infty} G(y).$$

Kertymäfunktion ominaisuudet

Lause (Kertymäfunktion ominaisuudet)

Satunnaismuuttujan X kertymäfunktioilla $F : \mathbb{R} \rightarrow \mathbb{R}$ on seuraavat ominaisuudet.

- (a) F on kasvava funktio.
- (b) F on oikealta jatkuva funktio, eli

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

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

Kääntäen, jos funktiolla $F : \mathbb{R} \rightarrow \mathbb{R}$ 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 hyppyt

- Kertymäfunktiolla voi olla hyppyjä.
- Koska

$$\{X \leq x\} = \{X < x\} \cup \{X = x\},$$

jossa osat ovat erillisiä, niin

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

- 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

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

Tämä seikka voidaan ilmaista merkinnällä

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

Kertymäfunktiot ja jakaumat vastaavat toisiaan

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

Lause (Kertymäfunktio määrää jakauman yksikäsitteisesti)

Seuraavat asiat ovat yhtäpitäviä.

- (a) X ja Y ovat samoin jakautuneet.
- (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 $\{x_1, x_2, \dots\}$ 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 $P(X \in S) = 1$.

Määritelmä

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

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

Pistetodennäköisyysfunktio $x \mapsto P(X = x)$ (engl. *probability (mass) function, pmf*) on periaattessa määritelty koko reaaliakselilla kaikenlaisille satunnaismuuttujille X . Tämä käsite on mielenkiintoinen lähes yksinomaan diskreeteille satunnaismuuttujille.

- 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 $f : \mathbb{R} \rightarrow \mathbb{R}$ 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

$$f(x) = \frac{1}{N}, \quad \text{kun } x = 1, 2, \dots, N.$$

- 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 parametrilla p). Tämä asia voidaan merkitä $X \sim \text{Bernoulli}(p)$.
- Tavallisesti ajatellaan niin, että kun sattuu $A = \{X = 1\}$, niin satunnaiskoe onnistuu ja muussa tapauksessa epäonnistuu. Tällöin p on yhtä kuin kokeen onnistumistodennäköisyys.

Bernoullin jakauman ptmf

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

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

- Merkintäsopimus: ptmf:n arvot tarvitsee kertoa vain ko. sm:n arvojoukossa.
- Jakauman Bernoulli(p) ptmf voidaan kirjoittaa lyhyesti

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

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

Jakaumien merkinnöistä

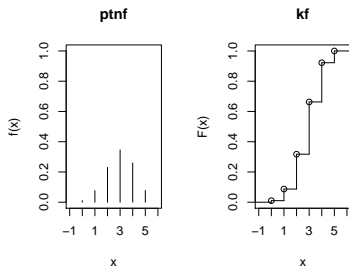
- Useimmat mielenkiintoiset jakaumat riippuvat yhden tai useamman parametrin arvosta. *Oikeastaan tarkastellaan saman tien perhettä jakaumia.*
- Pitääksemme lukua parametrin arvosta, merkitsemme sen tarvittaessa pystyviivan oikealle puolelle. Tätä merkintää voidaan käyttää kertymäfunktioille, pistetodennäköisyysfunktioille 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 parametreillä n ja p , jos sen arvojoukko on $\{0, 1, \dots, n\}$ ja ptnfn on

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

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



Binomijakauman toistokoetulkinta

- Binomijakaumaa $\text{Bin}(n, p)$ 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 $X \sim \text{Bin}(n, p)$.
- Tulos perustellaan myöhemmin.
- Bernoullin jakauma $\text{Bernoulli}(p)$ on tietenkin sama kuin binomijakauma $\text{Bin}(1, p)$.

2.5 Jatkuva jakauma

Määritelmä (Jatkuva jakauma, tiheysfunktio)

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

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

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.

- Jatkuvan jakauman tiheysfunktioille 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 vähintään paloittain jatkuvia funktioita, 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

$$\int_a^b f(x) dx = \int_a^b g(x) dx$$

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

$$P(X = x) = P(x \leq X \leq x) = \int_x^x f(u) du = 0$$

kaikilla x , joten pistetodennäköisyys on aina nolla.

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

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

kun $a < b$.

Toinen tapa karakterisoida jatkuva jakauma

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

$$P(X \leq x) = P(X < x) = \int_{-\infty}^x f(u) du.$$

- Itse asiassa pätee

$$P(X \in B) = \int_B f(x) dx, \quad \text{kaikille } B \subset \mathbb{R}, \quad (2)$$

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

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

- Ominaisuus (2) oltaisiin voitu alunperin valita tiheysfunktion (ja jatkuvan jakauman) määritelmäksi.

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

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

Jos x on tiheysfunktion f jatkuvuus piste, niin $F'(x) = f(x)$.

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

Huomautus: Kun sanotaan, että tiheysfunktiksi voidaan valita $f = F'$, niin tarkoitetaan sitä, että tiheysfunktiksi 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 kertymäfunktioista, että jakauma on jatkuva

Lause

Olkoon F sellainen kertymäfunktio, joka on

- jatkuva koko reaaliakselilla ja
- jatkuvasti derivoituva kaikkialla muualla, paitsi 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 ei ole jatkuvasti derivoituva, niin **analyysin peruslauseen** mukaan

$$F(b) - F(a) = \int_a^b F'(u) \, du.$$

Lisäksi otetaan raja-arvoja ja käytetään kertymäfunktion ominaisuuksia.

- Jos X :llä on jatkuva jakauma, niin sen k_f on jatkuva. Sen sijaan t_f :n ei tarvitse olla jatkuva eikä edes paloittain jatkuva funktio (sovelluksissa t_f on paloittain jatkuva). T_f :n ei tarvitse olla rajoitettu funktio.
- Täydellinen karakterisointi: k_f F on jatkuvan jakauman k_f täsmälleen silloin, kuin F on **absoluuttisesti jatkuva**, mutta tämän asian määrittäminen 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

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

joten tiheysfunktion integraali koko reaaliakselin yli on yksi.

Lause

Jos $f : \mathbb{R} \rightarrow \mathbb{R}$ 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 $\int_{-\infty}^{\infty} g(x) dx$ on äärellinen ja aidosti positiivinen, niin tällöin on olemassa vakio k siten, että $f = k g$ on tiheysfunktio.

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

$$\int f(x) dx = k \int g(x) dx = \frac{k}{k} = 1.$$

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

Tiheysfunktion frekvenssitulkinta

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

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

jossa ω^{act} on aktualisoitunut alkeistapaus, kullakin X_i on sama jakauma kuin X :llä ja lisäksi X_1, \dots, X_N ovat riippumattomia (mikä määritellään myöhemmin).

- Kuinka tällaisessa tilanteessa voidaan arvioida tiheysfunktion arvoa pisteessä u ? Oletetaan, että u on f :n jatkuvuus piste.

Eräs ratkaisu

- Valitaan jokin pieni luku $h > 0$ ja lasketaan, kuinka moni otospisteistä x_i osuu h :n pituiselle välille

$$\left[u - \frac{h}{2}, u + \frac{h}{2}\right].$$

Olkoon tämä lukumäärä $N_{u,h}$.

- Tiheysfunktion määritelmän ja frekvenssitulkinnan nojalla

$$P\left(u - \frac{h}{2} \leq X \leq u + \frac{h}{2}\right) = \int_{u-h/2}^{u+h/2} f(x) dx \approx \frac{N_{u,h}}{N}.$$

- Toisaalta, integraalilaskennan väliarvolauseen nojalla

$$\int_{u-h/2}^{u+h/2} f(x) dx \approx hf(u).$$

- Yhdistämällä saadaan arvio $f(u) \approx \frac{N_{u,h}}{hN}$. Tässä voidaan käyttää myös muita h :n pituisia välejä, jotka sisältävät pisteen u kuin tämä.

2.6 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, $X \sim U(a, b)$, jos sillä on jatkuva jakauma tiheysfunktiona f , jossa

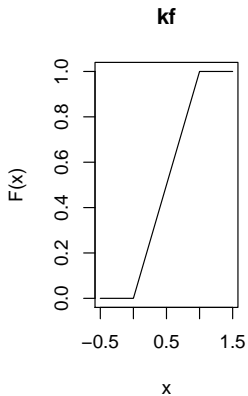
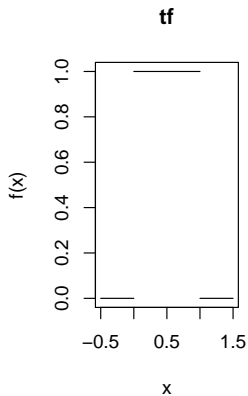
$$f(x) = f(x \mid a, b) = \frac{1}{b-a}, \quad \text{kun } a < x < b.$$

(Tästä kaavasta pitää ymmärtää, että $f(x) = 0$, kun $x \leq a$ tai $x \geq b$.)

- Kertymäfunktio on

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

Tasajakauman $U(0, 1)$ tiheys- ja kertymäfunktio



Eksponenttijakauma (engl. *exponential distribution*)

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

$$f(x) = f(x | \lambda) = \lambda e^{-\lambda x}, \quad \text{kun } x > 0.$$

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

$$F(x) = \int_0^x (-e^{-\lambda u}) = 1 - e^{-\lambda x},$$

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

2.7 Satunnaismuuttujan muunnos

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

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

ts. Y on yhdistetty funktio $g \circ X$.

- Mitä voimme sanoa Y :n jakaumasta?

Jos $g : \mathbb{R} \rightarrow \mathbb{R}$ 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

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

Muunoksen $g(X)$ jakauma diskreetissä tapauksessa

Lause

Olkoon X diskreetti sm ptnf:lla f_X , ja olkoon $g : \mathbb{R} \rightarrow \mathbb{R}$ 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

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

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

$$f_Y(y) = f_X(h(y)), \quad y \in S_Y. \quad (3)$$

- 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 muunoksen $Y = g(X)$ jakauma voi funktion g luonteesta riippuen joko diskreetti, jatkuva tai ei kumpaakaan näistä.
- Esim. jos g on paloittain vakio funktio, 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.

2.8 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

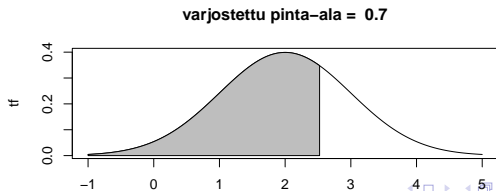
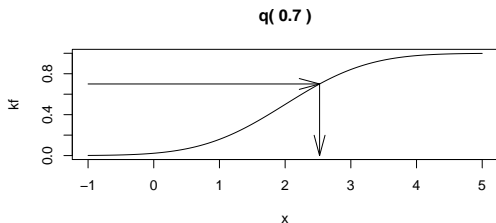
- Oletusten ansiosta funktiolla F on olemassa käänteisfunktio $F^{-1} : (0, 1) \rightarrow (a, b)$. (Tarkemmin sanoen F :n rajoittumalla joukon (a, b) on käänteisfunktio F^{-1} .)
- Merkintä F^{-1} **ei tarkoita** funktiota $1/F$, vaan kertymäfunktion **käänteisfunktioita**.
- Käytännössä $F^{-1}(u)$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

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

- Tällaista (avoimella välillä $(0, 1)$ määriteltyä) kertymäfunktion käänteisfunktioita 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

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

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

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

- 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.

Kertymäfunktion käänteisfunktion ominaisuuksia

- Funktio F^{-1} toteuttaa identiteetit

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

ja

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

- 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.

Kertymäfunktio muunnos

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

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

- Tämä tarkoittaa sitä, että

$$F(X) \sim U(0, 1). \tag{4}$$

- Satunnaismuuttujaan $F(X)$, ja tulokseen $F(X) \sim U(0, 1)$ viitataan toisinaan nimellä **kertymäfunktio muunnos** (engl. *probability integral transform*).

Käänteisfunktio menetelmä

- Olkoon $U \sim U(0, 1)$, 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

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

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

Käänteisfunktio menetelmä simuloinnissa

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

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

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

- Kf on

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

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

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

2.9 Muunnetun satunnaismuuttujan tiheys

- Olkoon X :llä jatkuva jakauma, ja $g : \mathbb{R} \rightarrow \mathbb{R}$.
- 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.

- Jos muunnoksella $Y = g(X)$ on jatkuva jakauma, niin tiheysfunktion lauseke on helpointa 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 7

Esimerkki kertymäfunktio-tekniikasta, $Y = e^X$

- Olkoon X :llä jatkuva jakauma, jolla on jatkuva tiheysfunktio f_X , ja $Y = e^X$.
- Selvästi $Y > 0$, joten $F_Y(y) = 0$, kun $y < 0$. Kun $y > 0$, on
$$F_Y(y) = P(Y \leq y) = P(e^X \leq y) = P(X \leq \ln(y)) = F_X(\ln(y)),$$
- Laskemalla kf:n derivaatan, arvaamme että tiheysfunktioksi kelpaa

$$f_Y(y) = F'_Y(y) = f_X(\ln(y))/y, \quad \text{kun } y > 0,$$

ja $f_Y(y) = 0$, kun $y \leq 0$.

Esimerkki jatkuu: tarkistetaan että jakauma on jatkuva

- Tarkistetaan, että tn $P(a \leq Y \leq b)$ saadaan laskettua integroimalla tiheysfunktioehdokasta välin (a, b) yli, kun $0 < a < b$ (vrt. tiheysfunktion määritelmä).
- Tarkistaminen on helppoa, sillä

$$\begin{aligned} P(a \leq Y \leq b) &= P(a \leq e^X \leq b) \\ &= P(\ln a \leq X \leq \ln b) \\ &= \int_{\ln a}^{\ln b} f_X(x) dx && (X\text{:n tf on } f_X) \\ &= \int_a^b f_X(\ln(y)) \frac{1}{y} dy && (\text{Sij. } y = e^x \Leftrightarrow x = \ln(y)) \end{aligned}$$

- Esimerkin tekniikka yleistetään kohta lauseeksi.

Määritelmä (Diffeomorfismi)

Kuvaus $g : A \rightarrow B$, 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 $g^{-1} : B \rightarrow A$ ovat jatkuvasti derivoituvia.

Muuttujanvaihtokaava tiheysfunktioille

Lause (Muuttujanvaihtokaava tiheysfunktioille)

Olkoon sm:lla X jatkuva jakauma tf:lla f_X . Olkoon $g : A \rightarrow B$ 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 tiheysfunktioilla

$$f_Y(y) = \begin{cases} f_X(h(y)) |h'(y)|, & \text{kun } y \in B \\ 0, & \text{muuten.} \end{cases} \quad (5)$$

- Tulos (5) voidaan ilmaista lyhyemmin kaavalla

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

- 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 (5) vastaavan diskreetin tapauksen tulokseen (3). Vain jatkuvan jakauman tapauksessa tarvitaan käänteisfunktion derivaatan itseisarvoa.

Muunnetun satunnaismuuttujan tiheysfunktio voidaan johtaa kahdella erilaisella tekniikalla:

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

Huomautuksia kertymäfunktioitekniikasta

- Kertymäfunktioitekniikassa 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 paloittain jatkuvasti derivoituva (mutta derivaatan ei välttämättä tarvitse olla olemassa äärellisessä määrässä pisteitä).

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

$$f_X(x) |dx| = f_Y(y) |dy| \quad (6)$$

pitää säilyä bijektiivisessä muuttujanvaihdossa

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

- Kun tiedosta (6) ratkaistaan $f_Y(y)$, saadaan

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right| = f_X(h(y)) |h'(y)|,$$

mutta tietenkin tämä kaava pätee vain g :n kuvajoukossa B .

- Yksinkertaisissa tapauksissa muunnoksen tf on tällä tavalla menetellen mahdollista johtaa päässälaskulla.

Muistisääntö on robusti

- Muistisäännöstä (6) voidaan ratkaista $f_Y(y)$ myös toisella tavalla, nimittäin seuraavasti,

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

- Myös tämä kaava pitää paikkansa joukossa B , sillä kaava

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}$$

kertoo oikean yhteyden funktion ja sen käänteisfunktion derivaattojen välillä.

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

- 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 $Y = X^2$, ja lasketaan sen kf.
- Selvästi $Y \geq 0$, joten $F_Y(y) = 0$, kun $y < 0$.
- Kun $y > 0$, on

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

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

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

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

Esimerkki $Y = X^2$: tarkistus

- Tarkistetaan, että tn $P(a \leq Y \leq b)$ todella saadaan laskettua integroimalla tätä tiheysfunktiota välin (a, b) yli, kun $0 < a < b$.

$$\begin{aligned} P(a \leq Y \leq b) &= P(a \leq X^2 \leq b) = P(\sqrt{a} \leq |X| \leq \sqrt{b}) \\ &= P(-\sqrt{b} \leq X \leq -\sqrt{a}) + P(\sqrt{a} \leq X \leq \sqrt{b}) \\ &= \int_{-\sqrt{b}}^{-\sqrt{a}} f_X(x) dx + \int_{\sqrt{a}}^{\sqrt{b}} f_X(x) dx \end{aligned}$$

Muuttujanvaihdot integraaleissa

- Nyt ensimmäisessä integraalissa tehdään muuttujanvaihto

$$y = x^2 \iff x = -\sqrt{y} \quad (\text{jossa } x < 0 \text{ ja } y > 0)$$

ja toisessa

$$y = x^2 \iff x = \sqrt{y} \quad (\text{jossa } x, y > 0).$$

- Tällöin saadaan

$$P(a \leq Y \leq b) = \int_a^b \left(f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} \right) dy$$

Tekniikka on muotoiltu lauseeksi

- Jos edellisen esimerkin tekniikka yleistetään, saadaan muotoiltua luentomonisteen lause 2.13.
- 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 kertymäfunktio tekniikalla: johdetaan ensin sm:n Y kertymäfunktio F_Y ja sitten derivoidaan tämä funktio.