Todennäköisyyslaskenta: Todennäköisyysjakaumia

Sisällys

16. Diskreettejä jakaumia

Määrittelemme tässä luvussa seuraavat *diskreetit todennäköisyysjakaumat*:

- **Diskreetti tasainen jakauma**
- **Bernoulli-jakauma**
- **Binomijakauma**
- **Geometrinen jakauma**
- **Negatiivinen binomijakauma**
- **Hypergeometrinen jakauma**
- **Poisson-jakauma**

Johdamme jokaisen jakauman **pistetodennäköisyysfunktion**. Lisäksi havainnollistamme pistetodennäköisyysfunktioita **graafisesti** ja johdamme jakaumien **odotusarvot** ja **varianssit** sekä (hypergeometrista jakaumaa lukuun ottamatta) myös niiden **momenttiemäfunktiot**. Tarkastelemme myös jakaumien yhteyksiä muihin jakaumiin.

Avainsanat:

Bernoulli-jakauma, Bernoulli-koe, Binomijakauma, Diskreetti tasainen jakauma, Eksponenttijakauma, Geometrinen jakauma, Hypergeometrinen jakauma, Kertymäfunktio, Momenttiemäfunktio, Negatiivinen binomijakauma, Odotusarvo, Otanta, Otanta palauttaen, Otanta palauttamatta, Otantasuhde, Pistetodennäköisyysfunktio, Poisson-jakauma, Standardipoikkeama, Varianssi

16.1. Diskreetti tasainen jakauma

Diskreettiä tasaista jakaumaa voidaan käyttää sellaisten satunnaisilmiöiden mallintamiseen, joissa on äärellinen määrä *symmetrisiä* eli *yhtä todennäköisiä* alkeistapahtumia.

Olkoon X *diskreetti satunnaismuuttuja*, jonka mahdolliset arvot ovat

$$x_1, x_2, \ldots , x_n$$

Oletetaan, että satunnaismuuttujan X mahdollisiin arvoihin $x_1, x_2, \ldots , x_n$ liittyvät todennäköisyydet ovat *yhtä suuria*:

$$\Pr(X = x_i) = \frac{1}{n}, \; i = 1, 2, \ldots , n$$

Tällöin satunnaismuuttuja X noudattaa **diskreettiä tasaista jakaumaa** ja sen *pistetodennäköisyysfunktio* on

$$f(x_i) = \Pr(X = x_i) = p_i = \frac{1}{n}, \; i = 1, 2, \ldots , n$$

Funktio $f(x_i)$, $i = 1, 2, \ldots , n$ määrittelee todennäköisyysjakauman, koska

$$f(x_i) > 0, \; i = 1, 2, \ldots , n$$

ja

$$\sum_{i=1}^{n} f(x_i) = n \cdot \frac{1}{n} = 1$$

Diskreetin tasaisen jakauman tunnusluvut

Oletetaan, että satunnaismuuttuja X noudattaa *diskreettiä tasaista jakaumaa*, jonka *pistetodennäköisyysfunktio* on

$$f(x_i) = \Pr(X = x_i) = p_i = \frac{1}{n}, \; i = 1, 2, \ldots , n$$

Odotusarvo:

$$\mathrm{E}(X) = \mu_X = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

Standardipoikkeama:

$$\mathrm{D}(X) = \sigma_X = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2}$$

Perustelu:

Suoraan diskreetin jakauman odotusarvon määritelmän mukaan

$$\mathrm{E}(X) = \mu_X = \sum_{i=1}^{n} x_i f(x_i) = \sum_{i=1}^{n} x_i \frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

Suoraan diskreetin jakauman varianssin määritelmän mukaan

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = \sum_{i=1}^{n} (x_i - \mu_X)^2 f(x_i) = \sum_{i=1}^{n} (x_i - \overline{x})^2 \frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2$$

■

Diskreetin tasaisen jakauman odotusarvon ja varianssin ominaisuudet

Diskreetin tasaisen jakauman

$$f(x_i) = \Pr(X = x_i) = p_i = \frac{1}{n}, \; i = 1, 2, \ldots, n$$

odotusarvo

$$\mathrm{E}(X) = \mu_X = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

on lukujen $x_1, x_2, \ldots, x_n$ *aritmeettinen keskiarvo* ja jakauman varianssi

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2$$

on lukujen $x_1, x_2, \ldots, x_n$ *otosvarianssi*, jossa jakajana on käytetty havaintojen lukumäärää n. Lisätietoja *havaintoarvojen jakaumaa kuvaavista otostunnusluvuista*: ks. monisteen **Tilastolliset menetelmät** lukua **Tilastollisten aineistojen kuvaaminen**.

Diskreetin tasaisen jakauman varianssi voidaan kirjoittaa myös muotoon

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} x_i\right)^2$$

jossa

$$\mathrm{E}(X^2) = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

on satunnaismuuttujan X 2. *momentti* (= lukujen $x_1, x_2, \ldots, x_n$ 2. *otosmomentti*).

Perustelu:

$$\begin{aligned}\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2 &= \frac{1}{n}\sum_{i=1}^{n} (x_i - 2\cdot\overline{x}\cdot x_i + \overline{x}^2) \\ &= \frac{1}{n}\left(\sum_{i=1}^{n} x_i - 2\cdot\overline{x}\cdot\sum_{i=1}^{n} x_i + n\overline{x}^2\right)\end{aligned}$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n} x_i - 2 \cdot \overline{x} \cdot (n\overline{x}) + n\overline{x}^2\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2$$
$$= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$$

■

Diskreetin tasaisen jakauman pistetodennäköisyysfunktio

Heitetään *virheetöntä* noppaa. Olkoon satunnaismuuttuja

X = Nopanheiton tulos

Satunnaismuuttujan X *pistetodennäköisyysfunktio* on

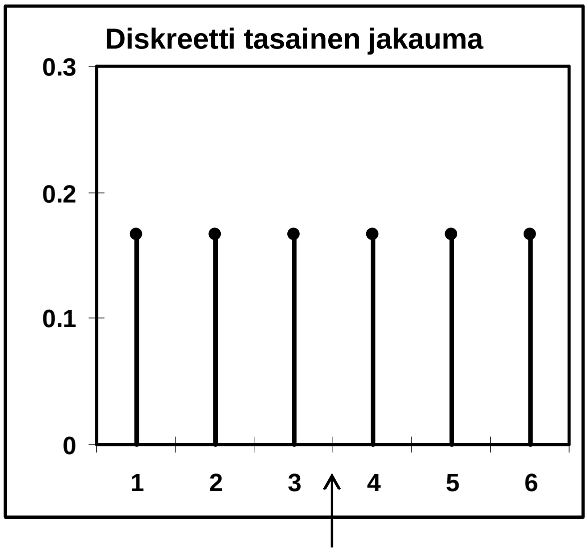
$f(i) = \Pr(X = i) = p_i = 1/6$

$i = 1, 2, 3, 4, 5, 6$

Kuva oikealla esittää jakauman pistetodennäköisyysfunktion *kuvaajaa*.

Kuvaan on merkitty myös jakauman *odotusarvo*

$\mathrm{E}(X) = 3.5$



E(X) = 3.5

Jakauman *odotusarvo* saadaan seuraavalla laskutoimituksella:

$$\mathrm{E}(X) = \sum_{i=1}^{6} if(i) = \frac{1}{6}\sum_{i=1}^{6} i = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2} = 3.5$$

Jakauman *varianssi* saadaan seuraavalla laskutoimituksella:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{D}^2(X) \\
&= \sum_{i=1}^{6} (i - \mathrm{E}(X))^2 f(i) \\
&= \frac{1}{6}\sum_{i=1}^{6} (i - \mathrm{E}(X))^2 \\
&= \frac{1}{6}\left[\left(1-\frac{7}{2}\right)^2 + \left(2-\frac{7}{2}\right)^2 + \left(3-\frac{7}{2}\right)^2 + \left(4-\frac{7}{2}\right)^2 + \left(5-\frac{7}{2}\right)^2 + \left(6-\frac{7}{2}\right)^2\right] \\
&= \frac{1}{6}\left[\left(-\frac{5}{2}\right)^2 + \left(-\frac{3}{2}\right)^2 + \left(-\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{3}{2}\right)^2 + \left(\frac{5}{2}\right)^2\right] \\
&= \frac{35}{12} \approx 2.917
\end{aligned}$$

Siten jakauman standardipoikkeamaksi saadaan:

$$\mathrm{D}(X) = \frac{1}{2}\sqrt{\frac{35}{3}} \approx \sqrt{2.917} \approx 1.708$$

Diskreetin tasaisen jakauman momenttiemäfunktio

Diskreetin tasaisen jakauman

$$f(x_i) = \Pr(X = x_i) = p_i = \frac{1}{n}\,,\ i = 1, 2, \ldots, n$$

momenttiemäfunktio (ks. lukua **Momenttiemäfunktio ja karakteristinen funktio**) on

$$m_X(t) = \mathrm{E}(e^{tX}) = \frac{1}{n}\sum_{k=1}^{n} e^{tx_k}$$

Perustelu:

Suoraan diskreetin jakauman momenttiemäfunktion määritelmän mukaan

$$m_X(t) = \mathrm{E}(e^{tX}) = \sum_x e^{tx} f(x) = \sum_{i=1}^{n} e^{tx_i} f(x_i) = \sum_{i=1}^{n} e^{tx_i}\frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n} e^{tx_i}$$

■

Diskreetin tasaisen jakauman momenttiemäfunktio ja jakauman tunnusluvut

Johdetaan diskreetin tasaisen jakauman *odotusarvo* ja *varianssi* jakauman momenttiemäfunktion avulla.

Odotusarvo:

$$\mathrm{E}(X) = \mu_X = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Perustelu:

Diskreetin tasaisen jakauman *momenttiemäfunktio* on

$$m_X(t) = \frac{1}{n}\sum_{i=1}^{n} e^{tx_i}$$

Momenttiemäfunktion 1. derivaatta pisteessä $t = 0$:

$$\left.\frac{dm_X(t)}{dt}\right|_{t=0} = \left.\frac{1}{n}\sum_{i=1}^{n} x_i e^{tx_i}\right|_{t=0} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Momenttiemäfunktion 2. derivaatta pisteessä $t = 0$:

$$\left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 e^{tx_i}\Big|_{t=0} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Siten diskreetin tasaisen jakauman *odotusarvo* μ_X, 2. *momentti* α_2 ja *varianssi* σ_X^2 saadaan seuraavilla kaavoilla:

$$\mu_X = \mathrm{E}(X) = \alpha_1 = \left.\frac{dm_X(t)}{dt}\right|_{t=0} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

$$\alpha_2 = \mathrm{E}(X^2) = \left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$$\sigma_X^2 = \mathrm{Var}(X) = \alpha_2 - \alpha_1^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

■

16.2. Bernoulli-jakauma

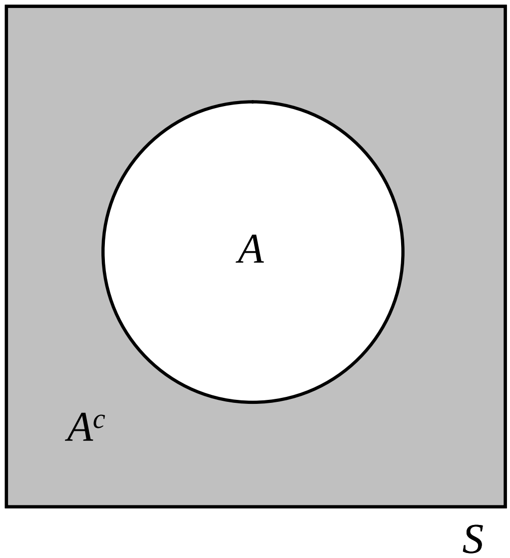
Olkoon A jokin otosavaruuden S *tapahtuma* ja olkoon

$$\Pr(A) = p$$

Olkoon tapahtuman A *komplementtitapahtuman* (tapahtuma A *ei satu*) A^c todennäköisyys

$$\Pr(A^c) = 1 - \mathrm{P}(A) = 1 - p = q$$

Ks. Venn-diagrammia oikealla.



Määritellään *diskreetti satunnaismuuttuja* X seuraavalla tavalla:

$$X = \begin{cases} 1\text{ , jos tapahtuma } A \textit{ sattuu} \\ 0\text{ , jos tapahtuma } A \textit{ ei satu} \end{cases}$$

Satunnaismuuttujan X *todennäköisyysjakauma* on

$$\Pr(X = 1) = p$$
$$\Pr(X = 0) = 1 - p = q$$

ja sen *pistetodennäköisyysfunktio* voidaan esittää seuraavassa muodossa:

$$f(x) = p^x q^{1-x}\text{ , } 0 < p < 1\text{ , } q = 1 - p\text{ , } x = 0,1$$

Sanomme, että satunnaismuuttuja X noudattaa **Bernoulli-jakaumaa parametrilla *p*** ja käytämme tästä merkintää:

$$X \quad \text{Bernoulli}(p)$$

Funktio $f(x)$ määrittelee todennäköisyysjakauman, koska

$$f(0) > 0$$
$$f(1) > 0$$

ja

$$f(0)+f(1)=q+p=(1-p)+p=1$$

Bernoulli-kokeet

Oletetaan, että olemme kiinnostuneita satunnaisilmiössä vain siitä sattuuko jokin ilmiöön liittyvä tapahtuma A vai ei – toisin sanoen kiinnostuksen kohteena on vain se sattuuko tapahtuma A vai tapahtuman A komplementti A^c. Kutsumme tällaista satunnaisilmiötä **Bernoulli-kokeeksi**.
Yllä esitetyn mukaan Bernoulli-kokeita voidaan mallintaa *Bernoulli-jakaumalla*.

Bernoulli-jakauman tunnusluvut

Olkoon

$$X \sim \text{Bernoulli}(p)$$

Odotusarvo:

$$\mathrm{E}(X)=p$$

Varianssi:

$$\mathrm{Var}(X)=\mathrm{D}^2(X)=\sigma_X^2=pq$$

Standardipoikkeama:

$$\mathrm{D}(X)=\sigma_X=\sqrt{pq}$$

Perustelu:

Suoraan diskreetin jakauman odotusarvon määritelmän mukaan

$$\mathrm{E}(X)=0\times\Pr(X=0)+1\times\Pr(X=1)=0\times q+1\times p=p$$

Määrätään satunnaismuuttujan X varianssi käyttäen kaavaa

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

jossa $\mathrm{E}(X^2)$ on satunnaismuuttujan X 2. momentti. Suoraan diskreetin jakauman 2. momentin määritelmän mukaan

$$\mathrm{E}(X^2)=0^2\times\Pr(X=0)+1^2\times\Pr(X=1)=0\times q+1\times p=p$$

Siten

$$\mathrm{Var}(X)=\mathrm{E}(X^2)-[\mathrm{E}(X)]^2=p-p^2=p(1-p)=pq$$

■

Bernoulli-jakauman odotusarvon ja varianssin ominaisuudet

Olkoon

$$X \sim \text{Bernoulli}(p)$$

Bernoulli-jakauman odotusarvo

$$\mathrm{E}(X)=p$$

yhtyy kiinnostuksen kohteena olevan tapahtuman A todennäköisyyteen $\Pr(A)=p$.

Bernoulli-jakauman varianssi

$$\mathrm{Var}(X) = pq = p(1 - p) = p - p^2$$

saavuttaa *maksiminsa* 1/4, kun

$$p = q = 1/2$$

Bernoulli-jakauman pistetodennäköisyysfunktio

Kuva oikealla esittää Bernoulli-jakauman

Bernoulli(0.8)

pistetodennäköisyysfunktiota

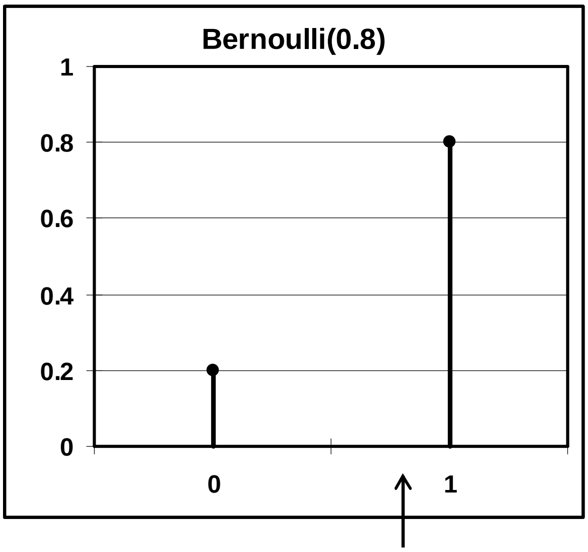
$$f(x) = p^x q^{1-x}\ ,\ x = 0,1$$

jossa

$$p = 0.8\ ,\ q = 1 - p$$

Kuvaan on merkitty myös jakauman *odotusarvo*

$$\mathrm{E}(X) = p = 0.8$$



E(*X*) = 0.8

Bernoulli-jakauman momenttiemäfunktio

Bernoulli-jakauman

$$f(x) = \Pr(X = x) = p^x q^{1-x}$$
$$0 < p < 1\ ,\ q = 1 - p\ ,\ x = 0,1$$

momenttiemäfunktio (ks. lukua **Momenttiemäfunktio ja karakteristinen funktio**) on

$$m_X(t) = \mathrm{E}(e^{tX}) = q + pe^t$$

Perustelu:

Suoraan diskreetin jakauman momenttiemäfunktion määritelmän mukaan

$$m_X(t) = \mathrm{E}(e^{tX}) = \sum_x e^{tx} f(x) = e^{t\times 0} \times \Pr(X = 0) + e^{t\times 1} \times \Pr(X = 1) = q + pe^t$$

■

Bernoulli-jakauman momenttiemäfunktio ja jakauman tunnusluvut

Johdetaan Bernoulli-jakauman Bernoulli(p) *odotusarvo* ja *varianssi* jakauman momentti-emäfunktion avulla.

Odotusarvo:

$$\mathrm{E}(X) = p$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = pq$$

Perustelu:

Bernoulli-jakauman Bernoulli(p) *momenttiemäfunktio* on

$$m_X(t) = q + pe^t$$

Momenttiemäfunktion 1. derivaatta pisteessä $t = 0$:

$$\left.\frac{dm_X(t)}{dt}\right|_{t=0} = \left.pe^t\right|_{t=0} = p$$

Momenttiemäfunktion 2. derivaatta pisteessä $t = 0$:

$$\left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \left.pe^t\right|_{t=0} = p$$

Siten Bernoulli-jakauman Bernoulli(p) *odotusarvo* μ_X, 2. *momentti* α_2 ja *varianssi* σ_X^2 saadaan seuraavilla kaavoilla:

$$\mu_X = \mathrm{E}(X) = \alpha_1 = \left.\frac{dm_X(t)}{dt}\right|_{t=0} = p$$

$$\alpha_2 = \mathrm{E}(X^2) = \left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = p$$

$$\sigma_X^2 = \mathrm{Var}(X) = \alpha_2 - \alpha_1^2 = p - p^2 = pq$$

■

Bernoulli-jakaumaa noudattavien satunnaismuuttujien summan jakauma

Olkoot

$$X_1, X_2, \ldots, X_n$$

riippumattomia satunnaismuuttujia, jotka noudattavat *samaa Bernoulli-jakaumaa* parametrilla p:

$$X_1, X_2, \ldots, X_n \perp$$

$$X_i \sim \text{Bernoulli}(p),\ i = 1, 2, \ldots, n$$

Tällöin satunnaismuuttujien $X_1, X_2, \ldots, X_n$ summa

$$Y = X_1 + X_2 + \cdots + X_n$$

on diskreetti satunnaismuuttuja, joka noudattaa **binomijakaumaa** (lisätietoja: ks. kappaletta **Binomijakauma**) parametrein (n, p):

$$Y \sim \text{Bin}(n, p)$$

Perustelu:

Todistus perustuu siihen, että *riippumattomien satunnaismuuttujien summan momenttiemäfunktio on ko. satunnaismuuttujien momenttiemäfunktioiden tulo*.

Oletetaan, että

$X_1, X_2, \ldots , X_n \perp$

$X_i \sim \text{Bernoulli}(p),\ i = 1, 2, \ldots , n$

Satunnaismuuttujien $X_1, X_2, \ldots , X_n$ *momenttiemäfunktiot* ovat muotoa

$$m_i(t) = q + pe^t \ , \ i = 1, 2, \ldots , n$$

Muodostetaan satunnaismuuttujien $X_1, X_2, \ldots , X_n$ *summa*:

$$Y = X_1 + X_2 + \cdots + X_n$$

Summamuuttujan Y momenttiemäfunktio on

$$m_Y(t) = m_1(t)m_2(t) \cdots m_n(t) = (q + pe^t)(q + pe^t) \cdots (q + pe^t) = (q + pe^t)^n$$

Funktio $m_Y(t)$ on *binomijakauman*

$$\text{Bin}(n, p)$$

momenttiemäfunktio (lisätietoja: ks. kappaletta **Binomijakauma**). Lisäksi funktio $m_Y(t)$ on jatkuva pisteen

$$t = 0$$

ympäristössä. Koska momenttiemäfunktio $m_Y(t)$ on tällöin *yksikäsitteinen*, summamuuttuja Y noudattaa *binomijakaumaa* parametrein n ja p:

$$Y = X_1 + X_2 + \cdots + X_n \sim \text{Bin}(n, p)$$

■

Huomautus:

- Kaikilla Bernoulli-jakaumilla on oltava sama tapahtuman A todennäköisyyttä kuvaava parametri p.

Bernoulli-kokeet diskreettien todennäköisyysjakaumien perustana

Bernoulli-kokeet muodostavat perustan monille diskreeteille todennäköisyysjakaumille. Oletetaan, että toistamme *samaa* Bernoulli-koetta niin, että toistot ovat toisistaan riippumattomia ja olkoon A se kiinnostuksen kohteena oleva tapahtuma, jonka sattumista toistojen aikana seurataan.

(i) **Binomijakauma** saadaan määräämällä todennäköisyys sille, että tapahtuma A sattuu x kertaa, kun koetta toistetaan n kertaa; ks. kappaletta **Binomijakauma**.

(ii) **Geometrinen jakauma** saadaan määräämällä todennäköisyys sille, että tapahtuma A sattuu ensimmäisen kerran x. koetoistossa; ks. kappaletta **Geometrinen jakauma**.

(iii) **Negatiivinen binomijakauma** saadaan määräämällä todennäköisyys sille, että tapahtuma A sattuu r. kerran x. koetoistossa; ks. kappaletta **Negatiivinen binomijakauma**.

(iv) **Poisson-jakauma** voidaan johtaa binomijakauman raja-arvona, kun koetoistojen lukumäärän annetaan tiettyjen ehtojen vallitessa kasvaa rajatta. Poisson-jakauma kuvaa *harvinaisten* tapahtumien todennäköisyyksiä *pitkissä* toistokoesarjoissa; ks. kappaletta **Poisson-jakauma**.

16.3. Binomijakauma

Toistetaan *samaa* Bernoulli-koetta niin, että toistot ovat toisistaan riippumattomia ja seurataan tapahtuman A sattumista toistojen aikana. Olkoon

$$\Pr(A) = p$$

jolloin tapahtuman A *komplementtitapahtuman* (= tapahtuma A *ei satu*) A^c todennäköisyys on

$$\Pr(A^c) = 1 - p = q$$

Oletetaan, että teemme koetoistoja n kappaletta, jossa n on kiinteä eli ei-satunnainen, etukäteen valittu luku, ja olkoon X *diskreetti satunnaismuuttuja*, joka kuvaa tapahtuman A esiintymiskertojen lukumäärää koetoistojen joukossa.

Satunnaismuuttuja X noudattaa **binomijakaumaa** parametrein n ja p:

$$X \quad \mathrm{Bin}(n, p)$$

Binomijakauman *pistetodennäköisyysfunktio* on

$$f(x) = \Pr(X = x) = \binom{n}{x} p^x q^{n-x}\ ,\ 0 < p < 1\ ,\ q = 1 - p\ ,\ x = 0, 1, 2, \ldots, n$$

Binomijakauman johto

Toistetaan *samaa* Bernoulli-koetta n kertaa niin, että toistot ovat toisistaan riippumattomia ja seurataan tapahtuman A sattumista toistojen aikana. Olkoon

$$\Pr(A) = p$$

ja

$$\Pr(A^c) = 1 - p = q$$

Määrätään todennäköisyys sille, että saadaan tapahtumajono, joka toteuttaa seuraavan ehdon:

($*$) Jonossa on x kappaletta tapahtumia A ja $(n - x)$ kappaletta tapahtumia A^c.

Olkoon

$$\underbrace{A\,A\,A^c\,A\,A^c \ldots A}_{n\ \text{kappaletta}}$$

mielivaltainen tapahtumajono, joka toteuttaa ehdon ($*$). Tämän jonon todennäköisyys on *riippumattomien tapahtumien tulosäännön* nojalla

$$ppqpq \ldots p = p^x q^{n-x}$$

Sama todennäköisyys on *jokaisella* tapahtumajonolla, joka toteuttaa ehdon ($*$).

Erilaiset järjestykset, joihin voimme asettaa x kappaletta tapahtumia A ja $(n - x)$ kappaletta tapahtumia A^c, *ovat tapahtumajonoina toisensa poissulkevia*. Siten todennäköisyys saada jono, joka toteuttaa ehdon ($*$), saadaan *toisensa poissulkevien tapahtumien yhteenlaskusäännön* mukaan laskemalla *kaikkien* ehdon ($*$) toteuttavien jonojen todennäköisyydet

$$p^x q^{n-x}$$

yhteen. Siten meidän on määrättävä *kaikkien* sellaisten jonojen lukumäärä, jotka toteuttavat ehdon (∗). Tämä lukumäärä on sama kuin niiden *järjestysten* lukumäärä, joihin voimme asettaa x kappaletta tapahtumia A ja $(n - x)$ kappaletta tapahtumia A^c.

Oletetaan, että käytössämme on kahdenlaisia *objekteja*, A ja A^c. *Lokeromallin* mukaan kysytty lukumäärä saadaan selville laskemalla kaikkien niiden tapojen lukumäärä, joilla x kappaletta objekteja A voidaan asettaa lokerikkoon, jossa on n lokeroa. Huomaa, että objektien A^c paikat on määrätty heti, kun objektit A on asetettu lokerikkoon.

Kysytyn lukumäärän antaa *binomikerroin*

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Perustelu: ks. lukua **Klassinen todennäköisyys ja kombinatoriikka**.

Siten todennäköisyys saada tapahtumajono, jossa on x kappaletta tapahtumia A ja $(n - x)$ kappaletta tapahtumia A^c, on

$$\binom{n}{x} p^x q^{n-x}$$

Olkoon X diskreetti satunnaismuuttuja, joka kuvaa tapahtuman A esiintymiskertojen lukumäärää koetoistojen joukossa. Yllä esitetyn mukaan

$$f(x) = \Pr(X = x) = \binom{n}{x} p^x q^{n-x},\ q = 1 - p,\ x = 0, 1, 2, \ldots, n$$

Funktio $f(x)$ määrittelee todennäköisyysjakauman, koska

$$f(x) > 0,\ x = 0, 1, 2, \ldots, n$$

ja *binomikaavan* mukaan

$$\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p+q)^n = 1^n = 1$$

Binomijakauman pistetodennäköisyydet p_x, $x = 0, 1, 2, \ldots, n$ toteuttavat siis yhtälön

$$\sum_{x=0}^{n} p_x = \sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = 1$$

Binomijakauman tunnusluvut

Olkoon

$$X \sim \mathrm{Bin}(n, p)$$

Odotusarvo:

$$\mathrm{E}(X) = \mu_X = np$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = npq$$

Standardipoikkeama:

$$\mathrm{D}(X) = \sigma_X = \sqrt{npq}$$

Perustelu:

Johdamme tässä binomijakauman odotusarvon käyttäen diskreetin jakauman odotusarvon määritelmää. Johdamme jakauman odotusarvon ja varianssin jakauman *momenttiemäfunktion* avulla kohdassa **Binomijakauman momenttiemäfunktio ja jakauman tunnusluvut**. Lisäksi jakauman odotusarvo ja varianssi johdetaan käyttäen hyväksi *binomijakauman ja Bernoulli-jakauman yhteyttä* kohdassa **Binomijakauma ja Bernoulli-jakauma**.

Suoraan diskreetin jakauman odotusarvon määritelmän mukaan

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{x=0}^{n} xf(x) \\
&= \sum_{x=0}^{n} x\binom{n}{x} p^x(1-p)^{n-x} \\
&= \sum_{x=0}^{n} x\cdot\frac{n!}{x!(n-x)!} p^x(1-p)^{n-x} \\
&= \sum_{x=1}^{n} x\cdot\frac{n!}{x!(n-x)!} p^x(1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x(1-p)^{n-x} \\
&= np\sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1}(1-p)^{n-x} \\
&= np
\end{aligned}$$

Yhtälöketjun viimeinen yhtälö perustuu siihen, että summassa

$$\begin{aligned}
\sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1}(1-p)^{n-x} &= \sum_{x=0}^{n-1} \frac{(n-1)!}{x!(n-1-x)!} p^x(1-p)^{n-1-x} \\
&= \sum_{x=0}^{n-1} \binom{n-1}{x} p^x(1-p)^{n-1-x} \\
&= 1
\end{aligned}$$

on laskettu yhteen *kaikki* binomijakauman $\mathrm{Bin}(n-1, p)$ pistetodennäköisyydet.

■

Odotusarvon ominaisuudet

Olkoon

$$X \quad \mathrm{Bin}(n, p)$$

Satunnaismuuttujan X odotusarvo

$$\mathrm{E}(X) = np$$

on *suoraan verrannollinen* tehtävien koetoistojen lukumäärään n että toistojen aikana seurattavan tapahtuman A todennäköisyyteen p.

Binomijakauman pistetodennäköisyysfunktio

Kuva oikealla esittää binomijakauman

$$\text{Bin}(12,1/3)$$

pistetodennäköisyysfunktiota

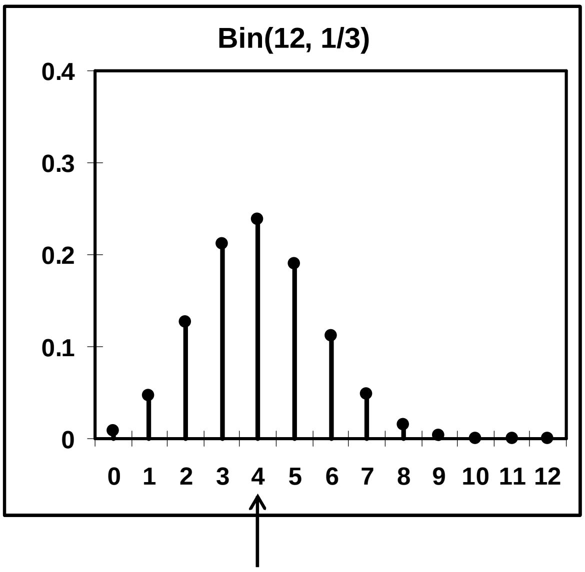
$$f(x) = \binom{n}{x} p^x q^{n-x} \text{ , } x = 0,1,2,\ldots,n$$

jossa

$$n = 12 \text{ , } p = 1/3 \text{ , } q = 1 - p$$

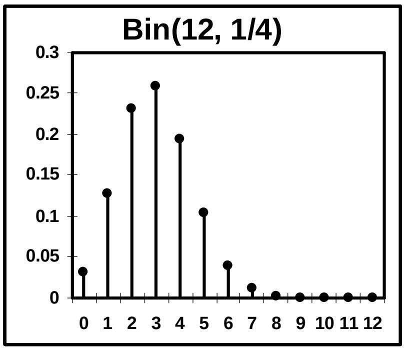
Kuvaan on merkitty myös jakauman *odotusarvo*

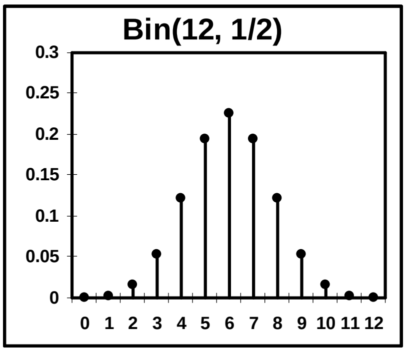
$$\text{E}(X) = np = 4$$

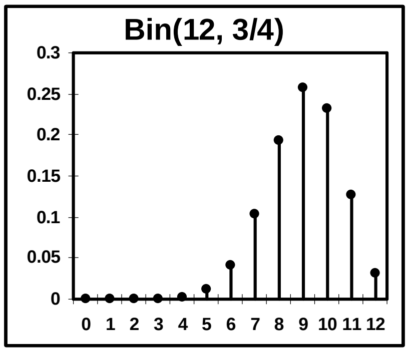


E(X) = 4

Alla oleva kuvasarja havainnollistaa sitä, miten parametri p vaikuttaa binomijakauman *vinouteen*:







$p < 1/2$: Binomijakauma on *positiivisesti vino* eli *vino oikealle*.

$p = 1/2$: Binomijakauma on *symmetrinen*.

$p > 1/2$: Binomijakauma on *negatiivisesti vino* eli *vino vasemmalle*.

Binomijakauman momenttiemäfunktio

Binomijakauman

$$f(x) = \Pr(X = x) = \binom{n}{x} p^x q^{n-x} \text{ , } 0 < p < 1 \text{ , } q = 1 - p \text{ , } x = 0,1,2,\ldots,n$$

momenttiemäfunktio (ks. lukua **Momenttiemäfunktio ja karakteristinen funktio**) on

$$m_X(t) = \text{E}(e^{tX}) = (q + pe^t)^n$$

Perustelu:

Suoraan diskreetin jakauman momenttiemäfunktion määritelmän mukaan

$$m_X(t) = \mathrm{E}(e^{tX}) = \sum_x e^{tx} f(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} = (q + pe^t)^n$$

■

Binomijakauman momenttiemäfunktio ja jakauman tunnusluvut

Johdetaan binomijakauman Bin(n, p) *odotusarvo* ja *varianssi* jakauman momenttiemäfunktion avulla.

Odotusarvo:

$$\mathrm{E}(X) = \mu_X = np$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = npq$$

Perustelu:

Binomijakauman Bin(n, p) *momenttiemäfunktio* on

$$m_X(t) = (q + pe^t)^n$$

Momenttiemäfunktion 1. derivaatta pisteessä $t = 0$:

$$\left.\frac{dm_X(t)}{dt}\right|_{t=0} = n(q + pe^t)^{n-1} pe^t \Big|_{t=0} = np$$

Momenttiemäfunktion 2. derivaatta pisteessä $t = 0$:

$$\begin{aligned}\left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} &= \left[n(n-1)(q + pe^t)^{n-2} pe^t pe^t + n(q + pe^t)^{n-1} pe^t \right]\Big|_{t=0} \\ &= npe^t (q + pe^t)^{n-2} \left[(n-1)pe^t + (q + pe^t) \right]\Big|_{t=0} \\ &= np + n(n-1)p^2\end{aligned}$$

Siten binomijakauman Bin(n, p) *odotusarvo* μ_X, 2. *momentti* α_2 ja *varianssi* σ_X^2 saadaan seuraavilla kaavoilla:

$$\mu_X = \mathrm{E}(X) = \alpha_1 = \left.\frac{dm_X(t)}{dt}\right|_{t=0} = np$$

$$\alpha_2 = \mathrm{E}(X^2) = \left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = np + n(n-1)p^2$$

$$\sigma_X^2 = \mathrm{Var}(X) = \alpha_2 - \alpha_1^2 = np + n(n-1)p^2 - n^2 p^2 = npq$$

■

Binomijakaumaa noudattavien satunnaismuuttujien summan jakauma

Olkoot

$$X_1, X_2, \ldots , X_k$$

riippumattomia satunnaismuuttujia, jotka noudattavat *binomijakaumaa* parametrein (n_1, p), (n_2, p), … , (n_k, p):

$$X_1, X_2, \ldots, X_k \perp$$
$$X_i \sim \mathrm{Bin}(n_i, p) \,,\; i = 1, 2, \ldots, k$$

Tällöin satunnaismuuttujien $X_1, X_2, \ldots , X_k$ *summa*

$$Y = X_1 + X_2 + \cdots + X_k$$

on diskreetti satunnaismuuttuja, joka noudattaa *binomijakaumaa* parametrein $n = n_1 + n_2 + \cdots + n_k$ ja p:

$$Y \sim \mathrm{Bin}(n, p)$$

Perustelu:

Todistus perustuu siihen, että *riippumattomien satunnaismuuttujien summan momentti-emäfunktio on ko. satunnaismuuttujien momenttiemäfunktioiden tulo*.

Oletetaan, että

$$X_1, X_2, \ldots, X_k \perp$$
$$X_i \sim \mathrm{Bin}(n_i, p) \,,\; i = 1, 2, \ldots, k$$

Satunnaismuuttujien $X_1, X_2, \ldots , X_k$ *momenttiemäfunktiot* ovat muotoa

$$m_i(t) = (q + pe^t)^{n_i} \,,\; i = 1, 2, \ldots, k$$

Muodostetaan satunnaismuuttujien $X_1, X_2, \ldots , X_k$ *summa*:

$$Y = X_1 + X_2 + \cdots + X_k$$

Summamuuttujan Y momenttiemäfunktio on

$$\begin{aligned} m_Y(t) &= m_1(t) m_2(t) \cdots m_k(t) \\ &= (q + pe^t)^{n_1} (q + pe^t)^{n_2} \cdots (q + pe^t)^{n_k} \\ &= (q + pe^t)^{n_1 + n_2 + \cdots + n_k} \\ &= (q + pe^t)^n \end{aligned}$$

jossa

$$n = n_1 + n_2 + \cdots + n_k$$

Funktio $m_Y(t)$ on *binomijakauman*

$$\mathrm{Bin}(n, p)$$

momenttiemäfunktio, jossa

$$n = n_1 + n_2 + \cdots + n_k$$

Lisäksi funktio $m_Y(t)$ on jatkuva pisteen

$$t = 0$$

ympäristössä. Koska momenttiemäfunktio $m_Y(t)$ on tällöin *yksikäsitteinen*, summamuuttuja Y noudattaa *binomijakaumaa* parametrein $n = n_1 + n_2 + \cdots + n_k$ ja p:

$$Y = X_1 + X_2 + \cdots + X_k \sim \mathrm{Bin}(n, p)$$

Huomautus:

- Kaikilla binomijakaumilla on oltava *sama* tapahtuman A todennäköisyyttä kuvaava parametri p, mutta sen sijaan toistokokeiden lukumäärää kuvaava parametri saa vaihdella jakaumasta toiseen.

Binomijakauma ja Bernoulli-jakauma

Toistetaan *samaa* Bernoulli-koetta niin, että toistot ovat toisistaan riippumattomia n kertaa ja seurataan tapahtuman A sattumista toistojen aikana.

Olkoon

$$\Pr(A) = p$$

ja

$$\Pr(A^c) = 1 - p = q$$

Määritellään *diskreetit satunnaismuuttujat*

$$X_i\ ,\ i = 1, 2, \ldots , n$$

seuraavalla tavalla:

$$X_i = \begin{cases} 1 \text{ , jos tapahtuma } A \textit{ sattuu} \text{ kokeessa } i \\ 0 \text{ , jos tapahtuma } A \textit{ ei satu} \text{ kokeessa } i \end{cases}$$

Tällöin

$$X_i \sim \mathrm{Bernoulli}(p),\ i = 1, 2, \ldots , n$$

Määritellään *diskreetti satunnaismuuttuja* X seuraavalla tavalla:

$$X = \text{Tapahtuman } A \text{ esiintymiskertojen lukumäärä}$$

Tällöin

$$X \quad \mathrm{Bin}(n, p)$$

Selvästi

$$X = \sum_{i=1}^{n} X_i$$

koska luku 1 esiintyy summassa $\sum X_i$ *täsmälleen* yhtä monta kertaa kuin tapahtuma A sattuu n:n koetoiston aikana.

Tämä merkitsee sitä, että *jokainen binomijakautunut satunnaismuuttuja voidaan esittää riippumattomien, samaa Bernoulli-jakaumaa noudattavien satunnaismuuttujien summana.*

Bernoulli-jakauma ja binomijakauman odotusarvo ja varianssi

Binomijakauman odotusarvoa ja varianssia johdettaessa voidaan käyttää hyväksi sitä, että jokainen binomijakautunut satunnaismuuttuja voidaan esittää riippumattomien, samaa Bernoulli-jakaumaa noudattavien satunnaismuuttujien summana.

Olkoot

$$X_i\,,\ i = 1,2,\ \ldots\ ,\ n$$

riippumattomia satunnaismuuttujia, jotka noudattavat *samaa Bernoulli-jakaumaa* parametrilla p:

$$X_1, X_2,\ \ldots\ , X_n \perp$$

$$X_i \sim \mathrm{Bernoulli}(p),\ i = 1,2,\ \ldots\ ,\ n$$

Olkoon

$$X = \sum_{i=1}^{n} X_i$$

Tällöin

$$X \quad \mathrm{Bin}(n, p)$$

Satunnaismuuttujan $X = \sum X_i$ *odotusarvo* on

$$\mathrm{E}(X) = \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \sum_{i=1}^{n} p = np$$

koska satunnaismuuttujien summan odotusarvo on *aina* satunnaismuuttujien odotusarvojen summa.

Satunnaismuuttujan $X = \sum X_i$ *varianssi* on

$$\mathrm{D}^2(X) = \mathrm{D}^2\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{D}^2(X_i) = \sum_{i=1}^{n} pq = npq$$

koska *riippumattomille satunnaismuuttujille pätee* se, että satunnaismuuttujien summan varianssi on satunnaismuuttujien varianssien summa.

Binomijakauma ja otanta palauttaen

Olkoon perusjoukon S alkioiden lukumäärä

$$n(S) = N$$

Muodostetaan perusjoukon S alkioiden osajoukko B, jossa on

$$n(B) = n$$

alkiota käyttämällä alkioiden poiminnassa **otantaa takaisinpanolla** eli **palauttaen**.

Otanta palauttaen:

(i) Poimitaan alkiot perusjoukosta S osajoukkoon B *satunnaisesti* yksi kerrallaan.

(ii) **Palautetaan** jokainen poimittu alkio ennen seuraavan alkion poimimista perusjoukkoon S.

(iii) Oletetaan, että *jokaisella perusjoukon S alkiolla* on jokaisessa poiminnassa *sama todennäköisyys*

$$1/N$$

tulla poimituksi osajoukkoon B.

Tällöin sanomme, että osajoukko B muodostaa **yksinkertaisen satunnaisotoksen** perusjoukosta S.

Otannassa takaisinpanolla alkioiden poiminta voidaan toteuttaa käyttämällä seuraavaa *arpomis-menettelyä*:

(1) Pannaan *uurnaan* jokaista perusjoukon S alkiota vastaava arpalippu.

(2) *Sekoitetaan* uurnan sisältö huolellisesti.

(3) *Nostetaan* uurnasta arpalippu, jota vastaava alkio valitaan otokseen B.

(4) **Palautetaan** nostettu arpalippu uurnaan.

(5) Palataan vaiheeseen (2), kunnes haluttu otoskoko n on saavutettu.

Huomautuksia:

- Perusjoukon S *alkioidenn todennäköisyys* tulla poimituksi otokseen *säilyy samana* poiminnan aikana.
- Jokaisella perusjoukon S samankokoisella *osajoukolla* on *sama* todennäköisyys tulla poimituksi otokseen.
- Sama perusjoukon S alkio voi tulla poimituksi otokseen *useita kertoja*.

Olkoon A perusjoukon S osajoukko, jonka alkioiden lukumäärä on

$$n(A) = r$$

Tällöin todennäköisyys poimia alkio joukosta A säilyy poiminnan kaikissa vaiheissa samana:

$$\Pr(A) = p = \frac{r}{N}$$

Otannassa takaisinpanolla otokseen B poimittujen A-tyyppisten alkioiden lukumäärä X on diskreetti satunnaismuuttuja, joka noudattaa *binomijakaumaa parametrein n* ja p:

$$X \quad \mathrm{Bin}(n, p)$$

Huomautus:

- **Otantaa ilman takaisinpanoa** eli **palauttamatta** tarkastellaan kappaleessa **Hyper-geometrinen jakauma**.

16.4. Geometrinen jakauma

Toistetaan *samaa* Bernoulli-koetta niin, että toistot ovat toisistaan riippumattomia ja seurataan tapahtuman A sattumista toistojen aikana. Olkoon

$$\Pr(A) = p$$

ja

$$\Pr(A^c) = 1 - p = q$$

Oletetaan, että teemme koetoistoja *kunnes tapahtuma A sattuu* 1. *kerran* ja olkoon *X diskreetti satunnaismuuttuja*, joka kuvaa tehtyjen koetoistojen lukumäärää, kun tapahtuma A sattuu 1. kerran.

Satunnaismuuttuja X noudattaa **geometrista jakaumaa** parametrilla p:

$$X \sim \mathrm{Geom}(p)$$

ja sen *pistetodennäköisyysfunktio* on

$$f(x) = \Pr(X = x) = q^{x-1}p\,,\ 0 < p < 1\,,\ q = 1 - p\,,\ x = 1, 2, 3, \ldots$$

Geometrisen jakauman johto

Toistetaan *samaa* Bernoulli-koetta niin, että toistot ovat toisistaan riippumattomia kunnes kiinnostuksen kohteena oleva tapahtuma A sattuu 1. kerran. Olkoon

$$\Pr(A) = p$$

ja

$$\Pr(A^c) = 1 - p = q$$

Oletetaan, että tapahtuma A sattuu 1. kerran x. kokeessa. Tällöin toistokokeiden tuloksena on saatu tapahtumajono

$$\underbrace{A^c A^c A^c \ldots A^c A}_{x \text{ kappaletta}}$$

jossa on *ensin* $(x - 1)$ kappaletta tapahtumia A^c ja tapahtuma A on jonossa *viimeisenä. Tämän* tapahtumajonon todennäköisyys on *riippumattomien tapahtumien tulosäännön* nojalla

$$qqq \cdots qp = q^{x-1}p\,,\ x = 1, 2, 3, \ldots$$

Olkoon X diskreetti satunnaismuuttuja, joka kuvaa tehtyjen koetoistojen lukumäärää, kun tapahtuma A sattuu 1. kerran. Yllä esitetyn mukaan

$$f(x) = \Pr(X = x) = q^{x-1}p\,,\ q = 1 - p\,,\ x = 1, 2, 3, \ldots$$

Funktio $f(x)$ määrittelee todennäköisyysjakauman, koska

$$f(x) > 0\,,\ x = 1, 2, 3, \ldots$$

ja *geometrisen sarjan summan kaavan* mukaan

$$\sum_{x=1}^{\infty} f(x) = \sum_{x=1}^{\infty} q^{x-1}p = p\sum_{x=0}^{\infty} q^x = p\frac{1}{1-q} = p\frac{1}{p} = 1$$

Geometrisen jakauman kertymäfunktio

Olkoon

$$X \sim \mathrm{Geom}(p)$$

On helppo nähdä (esimerkiksi täydellisellä induktiolla), että satunnaismuuttujan X *kertymäfunktio* on

$$F(x) = \Pr(X \le x) = 1 - (1 - p)^{[x]}$$

jossa

$$[x] = \text{suurin kokonaisluku, joka} \le x$$

Kertymäfunktion kaavasta seuraa *komplementtitapahtuman todennäköisyyden kaavan* nojalla

$$\Pr(X > x) = 1 - \Pr(X \le x) = 1 - F(x) = (1-p)^{[x]}$$

Geometrisen jakauman tunnusluvut

Olkoon

$$X \quad \mathrm{Geom}(p)$$

Odotusarvo:

$$\mathrm{E}(X) = \mu_X = \frac{1}{p}$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = \frac{q}{p^2}$$

Standardipoikkeama:

$$\mathrm{D}(X) = \sigma_X = \frac{\sqrt{q}}{p}$$

Perustelu:

Johdamme tässä geometrisen jakauman odotusarvon käyttäen diskreetin jakauman odotusarvon määritelmää. Jakauman odotusarvo ja varianssi johdetaan jakauman *momenttiemäfunktion* avulla kohdassa **geometrisen jakauman momenttiemäfunktio ja jakauman tunnusluvut**.

Suoraan diskreetin jakauman odotusarvon määritelmän mukaan

$$\mathrm{E}(X) = \sum_{x=1}^{\infty} xf(x) = \sum_{x=1}^{\infty} x(1-p)^{x-1} p$$

Tästä seuraa, että

$$(1-p)\,\mathrm{E}(X) = \sum_{x=1}^{\infty} x(1-p)^x p = \sum_{x=0}^{\infty} x(1-p)^x p = \sum_{x=1}^{\infty} (x-1)(1-p)^{x-1} p$$

Siten

$$\begin{aligned}
p\,\mathrm{E}(X) &= \mathrm{E}(X) - (1-p)\,\mathrm{E}(X) \\
&= \sum_{x=1}^{\infty} x(1-p)^{x-1} p - \sum_{x=1}^{\infty} (x-1)(1-p)^{x-1} p \\
&= \sum_{x=1}^{\infty} [x-(x-1)](1-p)^{x-1} p \\
&= \sum_{x=1}^{\infty} (1-p)^{x-1} p \\
&= 1
\end{aligned}$$

Yhtälöketjun viimeinen yhtälö perustuu siihen, että summassa

$$\sum_{x=1}^{\infty}(1-p)^{x-1}p$$

on laskettu yhteen *kaikki* geometrisen jakauman Geom(p) pistetodennäköisyydet.

Siten olemme saaneet yhtälön

$$p\,\mathrm{E}(X)=1$$

odotusarvon E(X) ratkaisemiseksi, joten

$$\mathrm{E}(X)=\frac{1}{p}$$

■

Odotusarvon ominaisuudet

Olkoon

$$X \quad \mathrm{Geom}(p)$$

Satunnaismuuttujan X odotusarvo

$$\mathrm{E}(X)=\frac{1}{p}$$

on *kääntäen verrannollinen* tapahtuman A todennäköisyyteen p. Siten tapahtumaa A saa odottaa keskimäärin sitä *kauemmin* mitä *pienempi* on tapahtuman A todennäköisyys.

Geometrisen jakauman pistetodennäköisyysfunktio

Kuva oikealla esittää geometrisen jakauman

$$\mathrm{Geom}(1/3)$$

pistetodennäköisyysfunktiota

$$f(x)=q^{x-1}p$$

pisteissä

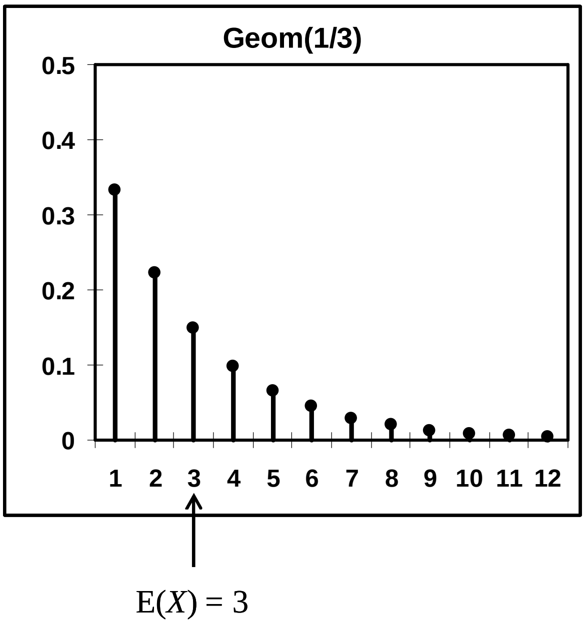
$$x = 1, 2, \ldots , 12$$

kun

$$p=1/3\,,\; q=1-p$$

Kuvaan on merkitty myös jakauman *odotusarvo*

$$\mathrm{E}(X)=\frac{1}{p}=3$$



Geometrisen jakauman momenttiemäfunktio

Geometrisen jakauman

$$f(x) = \Pr(X = x) = q^{x-1}p$$
$$0 < p < 1,\ q = 1 - p$$
$$x = 1, 2, 3, \ldots$$

momenttiemäfunktio (ks. lukua **Momenttiemäfunktio ja karakteristinen funktio**) on

$$m_X(t) = \mathrm{E}(e^{tX}) = \frac{pe^t}{1 - qe^t}$$

Perustelu:

Suoraan diskreetin jakauman momenttiemäfunktion määritelmän mukaan

$$\begin{aligned} m_X(t) = \mathrm{E}(e^{tX}) &= \sum_x e^{tx} f(x) \\ &= \sum_{x=1}^{\infty} e^{tx} pq^{x-1} \\ &= pe^t \sum_{x=1}^{\infty} e^{tx-t} q^{x-1} \\ &= pe^t \sum_{x=1}^{\infty} (qe^t)^{x-1} \\ &= \frac{pe^t}{1 - qe^t} \end{aligned}$$

■

Geometrisen jakauman momenttiemäfunktio ja jakauman tunnusluvut

Johdetaan geometrisen jakauman Geom(p) *odotusarvo* ja *varianssi* jakauman momenttiemäfunktion avulla.

Odotusarvo:

$$\mathrm{E}(X) = \mu_X = \frac{1}{p}$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = \frac{q}{p^2}$$

Perustelu:

Geometrisen jakauman Geom(p) *momenttiemäfunktio* on

$$m_X(t) = \frac{pe^t}{1 - qe^t}$$

Momenttiemäfunktion 1. derivaatta pisteessä t = 0:

$$\left.\frac{dm_X(t)}{dt}\right|_{t=0} = \left.\frac{pe^t(1-qe^t) - pe^t(-qe^t)}{(1-qe^t)^2}\right|_{t=0} = \left.\frac{pe^t}{(1-qe^t)^2}\right|_{t=0} = \frac{1}{p}$$

Momenttiemäfunktion 2. derivaatta pisteessä t = 0:

$$\left.\frac{d^2m_X(t)}{dt^2}\right|_{t=0} = \left.\frac{pe^t(1-qe^t)^2 - pe^t \cdot 2(1-qe^t)(-qe^t)}{(1-qe^t)^4}\right|_{t=0} = \left.\frac{pe^t(1+qe^t)}{(1-qe^t)^3}\right|_{t=0} = \frac{1+q}{p^2}$$

Siten geometrisen jakauman Geom(p) *odotusarvo* μ_X , 2. *momentti* α_2 ja *varianssi* σ_X^2 saadaan seuraavilla kaavoilla:

$$\mu_X = \mathrm{E}(X) = \alpha_1 = \left.\frac{dm_X(t)}{dt}\right|_{t=0} = \frac{1}{p}$$

$$\alpha_2 = \mathrm{E}(X^2) = \left.\frac{d^2m_X(t)}{dt^2}\right|_{t=0} = \frac{1+q}{p^2}$$

$$\sigma_X^2 = \mathrm{Var}(X) = \alpha_2 - \alpha_1^2 = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}$$

■

Geometrisen jakauman unohtamisominaisuus

Olkoon

$$X \sim \mathrm{Geom}(p)$$

Tällöin

$$\Pr(X \geq a + b \mid X \geq a) = \Pr(X \geq b + 1)$$

jossa $a, b \in$.

Perustelu:

Todetaan ensin, että

$$\Pr(X \geq c) = 1 - \Pr(X < c) = 1 - \Pr(X \leq c - 1) = 1 - F(c - 1) = (1 - p)^{[c - 1]}$$

jossa $c \in$, F on geometrisen jakauman *kertymäfunktio* ja

$[x]$ = suurin kokonaisluku, joka $\leq x$

Koska

$$\{X \geq a + b\} \subset \{X \geq a\}$$

niin

$$\Pr(X \geq a + b \text{ ja } X \geq a) = \Pr(X \geq a + b)$$

Siten

$$\Pr(X \geq a + b \mid X \geq a) = \frac{\Pr(X \geq a + b \text{ ja } X \geq a)}{\Pr(X \geq a)}$$

$$
\begin{aligned}
&= \frac{\Pr(X \geq a+b)}{\Pr(X \geq a)} \\
&= \frac{(1-p)^{[a+b-1]}}{(1-p)^{[a-1]}} \\
&= (1-p)^{[b]} \\
&= 1 - F(b) \\
&= 1 - \Pr(X \leq b) \\
&= \Pr(X > b) \\
&= \Pr(X \geq b+1)
\end{aligned}
$$

■

Siten geometrisella jakaumalla on ns. *unohtamisominaisuus*: Se, että tapahtuman A sattumista on jouduttu odottamaan a koetoistoa, *ei vaikuta* todennäköisyyteen joutua odottamaan b koetoistoa *lisää*. Tällaista unohtamisominaisuutta kutsutaan *stokastisten prosessien* teoriassa **Markov-ominaisuudeksi**.

Samanlainen unohtamisominaisuus on **eksponenttijakaumalla**, jota voidaan pitää geometrisen jakauman *jatkuvana* vastineena; lisätietoja eksponenttijakaumasta: ks. lukua **Jatkuvia jakaumia**.

16.5. Negatiivinen binomijakauma

Toistetaan *samaa* Bernoulli-koetta niin, että toistot ovat toisistaan riippumattomia ja seurataan tapahtuman A sattumista toistojen aikana. Olkoon

$$\Pr(A) = p$$

ja

$$\Pr(A^c) = 1 - p = q$$

Oletetaan, että teemme koetoistoja *kunnes tapahtuma A sattuu r. kerran* ja olkoon *X diskreetti satunnaismuuttuja*, joka kuvaa tehtyjen koetoistojen lukumäärää, kun tapahtuma A sattuu r. kerran.

Satunnaismuuttuja X noudattaa **negatiivista binomijakaumaa** parametrein r ja p:

$$X \quad \mathrm{NegBin}(r, p)$$

ja sen *pistetodennäköisyysfunktio* on

$$f(x) = \Pr(X = x) = \binom{x-1}{r-1} q^{x-r} p^r \text{ , } 0 < p < 1 \text{ , } q = 1 - p$$
$$r = 1, 2, 3, \ldots \text{ ; } x = r, r+1, r+2, \ldots$$

Negatiivisen binomijakauman johto

Toistetaan *samaa* Bernoulli-koetta niin, että toistot ovat toisistaan riippumattomia kunnes kiinnostuksen kohteena oleva tapahtuman A sattuu r. kerran. Olkoon

$$\Pr(A) = p$$

ja

$$\Pr(A^c) = 1 - p = q$$

Määrätään todennäköisyys sille, että saadaan tapahtumajono joka toteuttaa seuraavan ehdon:

$(*)$ Jonossa on r kappaletta tapahtumia A ja $(x - r)$ kappaletta tapahtumia A^c ja tapahtuma A on jonossa *viimeisenä*.

Olkoon

$$\underbrace{A A A^c A A^c \ldots A}_{n \text{ kappaletta}}$$

mielivaltainen tapahtumajono, joka toteuttaa ehdon $(*)$. Tämän jonon todennäköisyys on *riippumattomien tapahtumien tulosäännön* nojalla

$$ppqpq \ldots p = p^r q^{x-r}$$

Sama todennäköisyys on *jokaisella* tapahtumajonolla, joka toteuttaa ehdon $(*)$.

Erilaiset järjestykset, joihin voimme asettaa r kappaletta tapahtumia A ja $(x - r)$ kappaletta tapahtumia A^c ja joissa tapahtuma A on *viimeisenä*, *ovat tapahtumajonoina toisensa poissulkevia*. Siten todennäköisyys saada jono, joka toteuttaa ehdon $(*)$, saadaan *toisensa poissulkevien tapahtumien yhteenlaskusäännön* mukaan laskemalla *kaikkien* ehdon $(*)$ toteuttavien jonojen todennäköisyydet

$$p^r q^{x-r}$$

yhteen. Siten meidän on määrättävä *kaikkien* sellaisten jonojen lukumäärä, jotka toteuttavat $(*)$. Koska tapahtuma A pitää olla jonossa viimeisenä, tämä lukumäärä on sama kuin niiden *järjestysten* lukumäärä, joihin voimme asettaa $(r - 1)$ kappaletta tapahtumia A ja $((x - 1) - (r - 1)) = (x - r)$ kappaletta tapahtumia A^c.

Oletetaan, että meillä on käytössä kahdenlaisia *objekteja*, A ja A^c. *Lokeromallin* mukaan kysytty lukumäärä saadaan selville laskemalla kaikkien niiden tapojen lukumäärä, joilla $(r - 1)$ kappaletta objekteja A voidaan asettaa lokerikkoon, jossa on $(x - 1)$ lokeroa. Huomaa, että objektien A^c paikat on määrätty heti, kun objektit A on asetettu lokerikkoon.

Kysytyn lukumäärän antaa *binomikerroin*

$$\binom{x-1}{r-1} = \frac{(x-1)!}{(r-1)!((x-1)-(r-1))!} = \frac{(x-1)!}{(r-1)!(x-r)!}$$

Perustelu: ks. lukua **Klassinen todennäköisyys ja kombinatoriikka**.

Siten todennäköisyys saada tapahtumajono, jossa on r kappaletta tapahtumia A ja $(x - r)$ kappaletta tapahtumia A^c ja jossa tapahtuma A on *viimeisenä*, on

$$\binom{x-1}{r-1} p^r q^{x-r}$$

Olkoon X *diskreetti satunnaismuuttuja*, joka kuvaa tehtyjen koetoistojen lukumäärää, kun tapahtuma A sattuu r. kerran. Yllä esitetyn mukaan

$$f(x) = \binom{x-1}{r-1} q^{x-r} p^r \ , \ q = 1 - p$$

$$r = 1, 2, 3, \ldots \ ; \ x = r, r+1, r+2 \ldots$$

Voidaan osoittaa, että negatiivisen binomijakauman pistetodennäköisyyksien summa = 1. Tulos seuraa *binomikaavan yleistyksestä negatiivisille eksponenteille* (todistus sivuutetaan).

Negatiivisen binomijakauman tunnusluvut

Olkoon

$$X \quad \mathrm{NegBin}(r, p)$$

Odotusarvo:

$$\mathrm{E}(X) = \mu_X = \frac{r}{p}$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = \frac{rq}{p^2}$$

Standardipoikkeama:

$$\mathrm{D}(X) = \sigma_X = \frac{\sqrt{rq}}{p}$$

Perustelu:

Johdamme tässä negatiivisen binomijakauman odotusarvon käyttäen diskreetin jakauman odotusarvon määritelmää. Jakauman odotusarvo ja varianssi johdetaan jakauman *momentti-emäfunktion* avulla kohdassa **Negatiivisen binomijakauman momenttiemäfunktio ja jakauman tunnusluvut**.

Suoraan diskreetin jakauman odotusarvon määritelmän mukaan

$$\begin{aligned}
\mathrm{E}(X) = \sum_{x=r}^{\infty} x f(x) &= \sum_{x=r}^{\infty} x \binom{x-1}{r-1} q^{x-r} p^r \\
&= \sum_{x=r}^{\infty} x \frac{(x-1)!}{(r-1)!(x-r)!} q^{x-r} p^r \\
&= \frac{r}{p} \sum_{x=r}^{\infty} \frac{x!}{r!(x-r)!} q^{x-r} p^{r+1} \\
&= \frac{r}{p} \sum_{x=r}^{\infty} \binom{x}{r} q^{x-r} p^{r+1} = \frac{r}{p}
\end{aligned}$$

Yhtälöketjun viimeinen yhtälö perustuu siihen, että summassa

$$\sum_{x=r}^{\infty} \binom{x}{r} q^{x-r} p^{r+1} = \sum_{x=r+1}^{\infty} \binom{x-1}{r} q^{x-(r+1)} p^{r+1} = 1$$

on laskettu yhteen *kaikki* negatiivisen binomijakauman $\mathrm{NegBin}(r+1, p)$ piste-todennäköisyydet.

■

Odotusarvon ominaisuudet

Olkoon

$$X \quad \mathrm{NegBin}(r, p)$$

Satunnaismuuttujan X odotusarvo

$$\mathrm{E}(X) = \frac{r}{p}$$

on *suoraan verrannollinen* tapahtuman A esiintymisten odotettuun lukumäärään r ja *kääntäen verrannollinen* tapahtuman A todennäköisyyteen p. Siten tapahtuman A r. esiintymistä saa odottaa keskimäärin sitä *kauemmin* mitä *useampia* tapahtuman A esiintymisiä odotetaan ja mitä *pienempi* on tapahtuman A todennäköisyys.

Negatiivisen binomijakauman pistetodennäköisyysfunktio

Kuva oikealla esittää negatiivisen binomijakauman

$$\mathrm{NegBin}(3,1/3)$$

pistetodennäköisyysfunktiota

$$f(x) = \binom{x-1}{r-1} q^{x-r} p^r$$

pisteissä

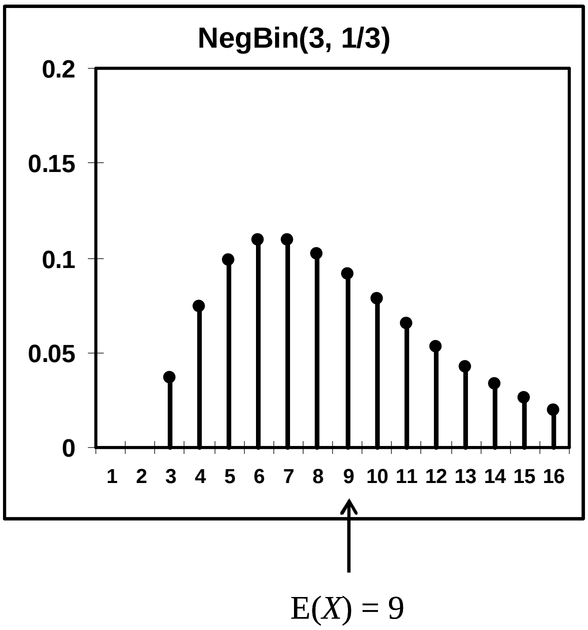
$$x = 3, 4, \ldots , 16$$

kun

$$r = 3\,,\ p = 1/3\,,\ q = 1 - p$$

Kuvaan on merkitty myös jakauman *odotusarvo*

$$\mathrm{E}(X) = \frac{r}{p} = 9$$



Negatiivisen binomijakauman momenttiemäfunktio

Negatiivisen binomijakauman

$$f(x) = \Pr(X = x) = \binom{x-1}{r-1} q^{x-r} p^r \ ,\ 0 < p < 1\,,\ q = 1 - p$$

$$r = 1, 2, 3, \ldots\ ;\ x = r, r+1, r+2, \ldots$$

momenttiemäfunktio (ks. lukua **Momenttiemäfunktio ja karakteristinen funktio**) on

$$m_X(t) = \mathrm{E}(e^{tX}) = \frac{(pe^t)^r}{(1-qe^t)^r}$$

Perustelu:

Suoraan diskreetin jakauman momenttiemäfunktion määritelmän mukaan

$$\begin{aligned} m_X(t) = \mathrm{E}(e^{tX}) &= \sum_x e^{tx} f(x) \\ &= \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} q^{x-r} p^r \\ &= (pe^t)^r \sum_{x=0}^{\infty} e^{tx} \binom{r+x-1}{r-1} q^x \\ &= (pe^t)^r (1-qe^t)^{-r} \\ &= \frac{(pe^t)^r}{(1-qe^t)^r} \end{aligned}$$

■

Negatiivisen binomijakauman momenttiemäfunktio ja jakauman tunnusluvut

Johdetaan negatiivisen binomijakauman NegBin(r, p) *odotusarvo* ja *varianssi* jakauman momentti-emäfunktion avulla.

Odotusarvo:

$$\mathrm{E}(X) = \mu_X = \frac{r}{p}$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = \frac{rq}{p^2}$$

Perustelu:

Negatiivisen binomijakauman NegBin(r, p) *momenttiemäfunktio* on

$$m_X(t) = \frac{(pe^t)^r}{(1-qe^t)^r}$$

Momenttiemäfunktion 1. derivaatta pisteessä t = 0:

$$\begin{aligned} \left.\frac{dm_X(t)}{dt}\right|_{t=0} &= \left.\frac{r(pe^t)^{r-1} pe^t (1-qe^t)^r - (pe^t)^r r(1-qe^t)^{r-1}(-qe^t)}{(1-qe^t)^{2r}}\right|_{t=0} \\ &= \left.\frac{r(pe^t)^r}{(1-qe^t)^{r+1}}\right|_{t=0} = \frac{r}{p} \end{aligned}$$

Momenttiemäfunktion 2. derivaatta pisteessä t = 0:

$$\begin{aligned} \left.\frac{d^2m_X(t)}{dt^2}\right|_{t=0} &= \left.\frac{r^2(pe^t)^{r-1} pe^t (1-qe^t)^{r+1} - r(pe^t)^r (r+1)(1-qe^t)^r(-qe^t)}{(1-qe^t)^{2r+2}}\right|_{t=0} \\ &= \left.\frac{r(pe^t)^r (r+qe^t)}{(1-qe^t)^{r+2}}\right|_{t=0} = \frac{r^2+rq}{p^2} \end{aligned}$$

Siten negatiivisen binomijakauman NegBin(r, p) *odotusarvo* μ_X , 2. *momentti* α_2 ja *varianssi* σ_X^2 saadaan seuraavilla kaavoilla:

$$\mu_X = \mathrm{E}(X) = \alpha_1 = \left.\frac{dm_X(t)}{dt}\right|_{t=0} = \frac{r}{p}$$

$$\alpha_2 = \mathrm{E}(X^2) = \left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \frac{r^2 + rq}{p^2}$$

$$\sigma_X^2 = \mathrm{Var}(X) = \alpha_2 - \alpha_1^2 = \frac{r^2 + rq}{p^2} - \frac{r^2}{p^2} = \frac{rq}{p^2}$$

■

Negatiivinen binomijakauma ja geometrinen jakauma

Olkoon

X NegBin(r, p)

jossa

$r = 1$

Tällöin X noudattaa *geometrista jakaumaa* parametrilla p:

X Geom(p)

16.6. Hypergeometrinen jakauma

Oletetaan, että perusjoukossa S

$n(S) = N$

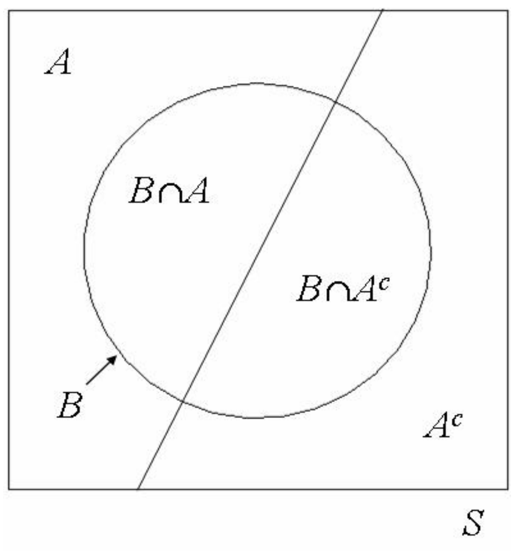
alkioita. Olkoon A perusjoukon S jokin *osajoukko*:

$A \subset S$

Tällöin A ja sen *komplementti* A^c muodostavat perusjoukon S *osituksen*:

$A \cup A^c = S$

$A \cap A^c = \emptyset$



Oletetaan, että joukossa A on

$n(A) = r$

alkiota ja joukossa A^c on

$n(A^c) = N - r$

alkiota.

Poimitaan perusjoukosta S *satunnaisesti* osajoukko B, jossa on

$n(B) = n$

alkiota. I

Perusjoukon S ositus joukoiksi A ja A^c *indusoi* joukon B osituksen joukoiksi $B \cap A$ ja $B \cap A^c$; ks. kuvaa edellä.

Olkoon X *diskreetti satunnaismuuttuja*, joka kuvaa joukkoon B poimittujen perusjoukon S osajoukon A (eli joukon $B \cap A$) alkioiden lukumäärää.

Satunnaismuuttuja X noudattaa **hypergeometrista jakaumaa** parametrein N, r ja n:

$$X \quad \mathrm{HyperGeom}(N, r, n)$$

ja sen *pistetodennäköisyysfunktio* on

$$f(x) = \Pr(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, \max[0, n-(N-r)] \le x \le \min(n, r)$$

Hypergeometrisen jakauman johto

Oletetaan, että otosavaruudessa S on

$$n(S) = N$$

alkiota.

Olkoon A jokin otosavaruuden S osajoukko:

$$A \subset S$$

ja olkoon A^c joukon A komplementti. Tällöin joukot A ja A^c muodostavat otosavaruuden S *osituksen*:

$$A \cup A^c = S$$

$$A \cap A^c = \varnothing$$

Oletetaan, että

$$n(A) = r$$

$$n(A^c) = N - r$$

Olkoon edelleen

$$B \subset S$$

ja

$$n(B) = n$$

Otosavaruuden S ositus joukkoihin A ja A^c *indusoi* joukon B osituksen joukoiksi $B \cap A$ ja $B \cap A^c$:

$$(B \cap A) \cup (B \cap A^c) = B$$

$$(B \cap A) \cap (B \cap A^c) = \varnothing$$

Ks. kuvaa edellä.

Olkoon

$$n(B \cap A) = x$$

$$n(B \cap A^c) = n - x$$

Joukosta S (jossa on siis N alkiota) voidaan poimia n alkiota osajoukkoon B

$$\binom{N}{n}$$

eri tavalla. Joukosta A (jossa on siis r alkiota) voidaan poimia x alkiota

$$\binom{r}{x}$$

eri tavalla. Joukosta A^c (jossa on siis $(N - r)$ alkiota) voidaan poimia $(n - x)$ alkiota

$$\binom{N-r}{n-x}$$

eri tavalla.

Joukosta A (jossa on siis r alkiota) voidaan poimia x alkiota *riippumatta* siitä, mitkä $(n - x)$ alkiota poimitaan joukosta A^c (jossa on siis $(N - r)$ alkiota). Siten *kombinatoriikan kertolaskuperiaatteesta* seuraa, että niiden tapojen lukumäärä, joilla voidaan poimia n alkiota joukosta S siten, että saadaan r alkiota joukosta A ja $(N - r)$ alkiota joukosta A^c on

$$\binom{r}{x}\binom{N-r}{n-x}$$

Olkoon X *diskreetti satunnaismuuttuja*, joka kuvaa joukkoon B poimittujen perusjoukon S osajoukon A (eli joukon $B \cap A$) alkioiden lukumäärää. Soveltamalla *klassisen todennäköisyyden* määritelmää saadaan:

$$f(x) = \Pr(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Voidaan osoittaa, että hypergeometrisen jakauman pistetodennäköisyyksien summa = 1 (todistus sivuutetaan).

Hypergeometrisen jakauman tunnusluvut

Olkoon

$$X \quad \mathrm{HyperGeom}(N, r, n)$$

Odotusarvo:

$$\mathrm{E}(X) = \mu_X = n\frac{r}{N}$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = n\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Standardipoikkeama:

$$\mathrm{D}(X) = \sigma_X = \sqrt{n\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)}$$

Perustelu:

Johdamme tässä hypergeometrisen jakauman odotusarvon käyttäen diskreetin jakauman odotusarvon määritelmää. Jakauman varianssin johtaminen sivuutetaan.

Koska

$$x\binom{r}{x} = x\frac{r!}{x!(r-x)!} = r\frac{(r-1)!}{(x-1)!(r-x)!} = r\binom{r-1}{x-1}$$

niin

$$\mathrm{E}(X) = \sum_{x=0}^{n} xf(x) = \sum_{x=0}^{n} x\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = r\sum_{x=1}^{n}\frac{\binom{r-1}{x-1}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Koska

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N}{n}\cdot\frac{(N-1)!}{(n-1)!(N-n)!} = \frac{N}{n}\binom{N-1}{n-1}$$

niin

$$\mathrm{E}(X) = r\sum_{x=1}^{n}\frac{\binom{r-1}{x-1}\binom{N-r}{n-x}}{\frac{N}{n}\binom{N-1}{n-1}} = \frac{nr}{N}\sum_{x=1}^{n}\frac{\binom{r-1}{x-1}\binom{N-r}{n-x}}{\binom{N-1}{n-1}} = \frac{nr}{N}$$

Yhtälöketjun viimeinen yhtälö perustuu siihen, että summassa

$$\sum_{x=1}^{n}\frac{\binom{r-1}{x-1}\binom{N-r}{n-x}}{\binom{N-1}{n-1}} = 1$$

on laskettu yhteen *kaikki* hypergeometrisen jakauman $\mathrm{HyperGeom}(N-1, r-1, n-1)$ pistetodennäköisyydet.

■

Odotusarvon ominaisuudet

Olkoon

$$X \quad \mathrm{HyperGeom}(N,r,n)$$

Satunnaismuuttujan X odotusarvo

$$\mathrm{E}(X)=n\frac{r}{N}$$

on *suoraan verrannollinen* sekä perusjoukosta S poimittavan *joukon B alkioiden lukumäärään n* että perusjoukon S *osajoukon A alkioiden lukumäärään r* ja *kääntäen verrannollinen perusjoukon S alkioiden lukumäärään N*.

Hypergeometrisen jakauman pistetodennäköisyysfunktio

Kuva oikealla esittää hypergeometrisen jakauman

$$\mathrm{HyperGeom}(100,12,20)$$

pistetodennäköisyysfunktiota

$$f(x)=\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

pisteissä

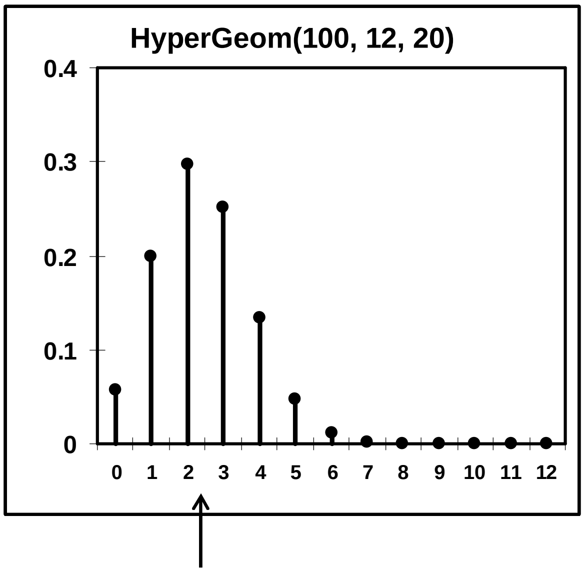
$$x = 0, 1, 2, \ldots , 12$$

kun

$$N=100\,,\ r=12\,,\ n=20$$

Kuvaan on merkitty myös jakauman *odotusarvo*

$$\mathrm{E}(X)=n\frac{r}{N}=2.4$$



E(X) = 2.4

Hypergeometrinen jakauma ja binomijakauma

Hypergeometrista jakaumaa voidaan *approksimoida* binomijakaumalla, jos ns. *otantasuhde*

$$\frac{n}{N}$$

jossa

$N = n(S)$ = perusjoukon S koko

ja

$n = n(B)$ = perusjoukosta S poimitun osajoukon B koko

on kyllin pieni. Näin on käytännössä, jos

$$\frac{n}{N}<0.05$$

Olkoon

$$X \quad \mathrm{HyperGeom}(N, r, n)$$

ja merkitään

$$\frac{r}{N} = p$$

Siten

$$r = Np$$

ja voimme kirjoittaa:

$$X \quad \mathrm{HyperGeom}(N, Np, n)$$

Huomaa, että p on todennäköisyys poimia joukkoon B alkio perusjoukon S osajoukosta A, jonka koko on

$$r = n(A)$$

Annetaan nyt

$$N \to +\infty$$

ja

$$r \to +\infty$$

niin, että

$$\frac{r}{N} = p$$

Tällöin hypergeometrisen jakauman $\mathrm{HyperGeom}(N, Np, n)$ pistetodennäköisyydet *konvergoivat* kohden binomijakauman $\mathrm{Bin}(n, p)$ pistetodennäköisyyksiä:

$$\lim_{N \to +\infty} f_{\mathrm{HyperGeom}(N, Np, n)}(x) = f_{\mathrm{Bin}(n,p)}(x)\,,\ x = 0, 1, 2, \ldots, n$$

Perustelu:

Olkoon

$$X \quad \mathrm{HyperGeom}(N, r, n)$$

Tällöin

$$\begin{aligned}
&\Pr(X = x) \\
&= \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \\
&= \frac{r!}{x!(r-x)!} \times \frac{(N-r)!}{(n-x)!(N-r-n+x)!} \times \frac{n!(N-n)!}{N!} \\
&= \frac{n!}{x!(n-x)!} \times \frac{r!}{(r-x)!} \times \frac{(N-r)!}{(N-r-n+x)!} \times \frac{(N-n)!}{N!}
\end{aligned}$$

$$=\binom{n}{x}\times[r(r-1)\ldots(r-x+1)]\times[(N-r)(N-r-1)\ldots(N-r-n+x+1)]$$

$$\times\frac{1}{N(N-1)\ldots(N-n+1))}$$

$$=\binom{n}{x}\times\frac{r(r-1)\ldots(r-x+1)}{N(N-1)\ldots(N-x+1)}\times\frac{(N-r)(N-r-1)\ldots(N-r-n+x+1)}{(N-x)(N-x-1)\ldots(N-x-(n-x)+1))}$$

$$=\binom{n}{x}\times\frac{r}{N}\times\frac{\frac{r-1}{N}\ldots\frac{r-(x-1)}{N}}{\frac{N-1}{N}\ldots\frac{N-(x-1)}{N}}\times\frac{\frac{N-r}{N}\cdot\frac{N-r-1}{N}\ldots\frac{N-r-(n-x-1)}{N}}{\frac{N-x}{N}\cdot\frac{N-x-1}{N}\ldots\frac{N-x-(n-x-1)}{N}}$$

Oletetaan, että

$$N\to+\infty$$

ja

$$r\to+\infty$$

niin, että

$$\frac{r}{N}=p$$

Tällöin

$$\frac{r-i}{N}\to p\,,\; i=1,2,\ldots,x-1$$

$$\frac{N-i}{N}\to 1\,,\; i=1,2,\ldots,x-1$$

$$\frac{N-r-i}{N}\to 1-p\,,\; i=0,1,2,\ldots,n-x-1$$

$$\frac{N-x-i}{N}\to 1\,,\; i=0,1,2,\ldots,n-x-1$$

Siten

$$\Pr(X=x)=\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}\quad\to\quad\binom{n}{x}p^x(1-p)^{n-x}$$

mikä on binomijakauman Bin(n, p) pistetodennäköisyys pisteessä x.

■

Hypergeometrisen jakauman $\mathrm{HyperGeom}(N, Np, n)$ ja binomijakauman $\mathrm{Bin}(n, p)$ yhteys näkyy myös siinä, että jakaumilla on *sama* odotusarvo:

$$\mathrm{E}(X_{\mathrm{HyperGeom}(N,Np,n)}) = \mathrm{E}(X_{\mathrm{Bin}(n,p)}) = np$$

ja niiden varianssit

$$\mathrm{Var}(X_{\mathrm{HyperGeom}(N,Np,n)}) = np(1-p)\frac{N-n}{N-1}$$
$$\mathrm{Var}(X_{\mathrm{Bin}(n,p)}) = np(1-p)$$

eroavat vain multiplikatiivisella tekijällä

$$\frac{N-n}{N-1}$$

jota sanotaan *äärellisen perusjoukon korjaustekijäksi*. Korjaustekijä vaikuttaa hypergeometrisen jakauman varianssiin sitä *vähemmän* mitä *pienempi* on otantasuhde n/N, sillä

$$\frac{N-n}{N-1} \approx 1$$

jos

$$\frac{n}{N} \approx 0$$

Hypergeometrinen jakauma ja otanta palauttamatta

Olkoon perusjoukon S alkioiden lukumäärä

$$n(S) = N$$

Muodostetaan perusjoukon S alkiosta osajoukko B, jonka alkioiden lukumäärä on

$$n(B) = n$$

käyttämällä alkioiden poiminnassa **otantaa ilman takaisinpanoa** eli **palauttamatta**.

Otanta palauttamatta:

(i) Poimitaan alkiot perusjoukosta S osajoukkoon B *satunnaisesti* yksi kerrallaan.

(ii) **Ei palauteta** poimittua alkiota ennen seuraavan alkion poimimista perusjoukkoon S.

(iii) Oletetaan, että *jokaisella perusjoukon S jäljellä olevalla alkiolla* on k. alkiota poimittaessa *sama todennäköisyys*

$$1/(N - k + 1)$$

tulla poimituksi osajoukkoon B.

Tällöin sanomme, että osajoukko B muodostaa **yksinkertaisen satunnaisotoksen** perusjoukosta S.

Otannassa ilman takaisinpanoa alkioiden poiminta voidaan toteuttaa käyttämällä seuraavaa *arpomismenettelyä*:

(1) Pannaan *uurnaan* jokaista perusjoukon S alkiota vastaava arpalippu.

(2) *Sekoitetaan* uurnan sisältö huolellisesti.

(3) *Nostetaan* uurnasta arpalippu, jota vastaava alkio valitaan otokseen B.

(4) **Ei palauteta** nostettua arpalippu uurnaan.

(5) Palataan vaiheeseen (3), kunnes haluttu otoskoko n on saavutettu.

Huomautuksia:

- Perusjoukon S *alkioiden todennäköisyys* tulla poimituksi otokseen *muuttuu* poiminnan aikana.
- Jokaisella perusjoukon S samankokoisella *osajoukolla* on *sama* todennäköisyys tulla poimituksi otokseen.
- Sama perusjoukon S alkio voi tulla poimituksi otokseen *vain kerran*.

Olkoon A perusjoukon S osajoukko, jonka alkioiden lukumäärä on

$$n(A) = r$$

Tällöin todennäköisyys poimia *yksi* alkio joukosta A on

$$\Pr(A) = p = \frac{r}{N}$$

Otannassa ilman takaisinpanoa otokseen B poimittujen A-tyyppisten alkioiden lukumäärä X on diskreetti satunnaismuuttuja, joka noudattaa *hypergeometrista jakaumaa parametrein* N, n ja r:

$$X \quad \mathrm{HyperGeom}(N, r, n)$$

Huomautus:

- **Otantaa takaisinpanolla** eli **palauttaen** tarkastellaan kappaleessa **Binomijakauma**.

Otanta palauttaen *vs* otanta palauttamatta

Poimitaan perusjoukosta *satunnaisesti otos* (osajoukko) arpomalla alkiot perusjoukosta otokseen yksi kerrallaan.

Otoksen poiminta voidaan toteuttaa joko **palauttaen** (*takaisinpanolla*) tai **palauttamatta** (*ilman takaisinpanoa*):

(i) *Otannassa palauttaen* perusjoukon alkiot *arvotaan* otokseen yksi kerrallaan niin, että jokainen alkio *palautetaan* välittömästi arpomisen jälkeen takaisin perusjoukkoon, jolloin sama alkio voi tulla poimituksi otokseen *useita kertoja*.

(ii) *Otannassa palauttamatta* alkiot *arvotaan* otokseen yksi kerrallaan niin, että alkiota *ei palauteta* arpomisen jälkeen takaisin perusjoukkoon, jolloin sama alkio voi tulla poimituksi otokseen *vain kerran*.

Olkoon perusjoukon S koko

$$n(S) = N$$

Tarkastellaan perusjoukon S osajoukkoa A, jonka koko on

$$n(A) = r$$

Poimitaan perusjoukosta S satunnaisesti osajoukko B, jonka koko on

$$n(B) = n$$

Määritellään *diskreetti satunnaismuuttuja*

X = ”A-tyyppisten alkioiden lukumäärä otoksessa B”

Jos otos poimitaan perusjoukosta *palauttaen*, satunnaismuuttuja X noudattaa *binomijakaumaa* parametrein n ja $p = r/N$:

$$X \quad \mathrm{Bin}(n, p)$$

Jos otos poimitaan perusjoukosta *palauttamatta* eli *ilman takaisinpanoa*, satunnaismuuttuja X noudattaa *hypergeometrista jakaumaa* parametrein N, r ja n:

$$X \quad \mathrm{HyperGeom}(N, r, n)$$

Huomautus:

- Koska hypergeometrinen jakauma konvergoi kohden binomijakaumaa, kun perusjoukon koon N annetaan kasvaa rajatta, ero otannan palauttaen ja otannan palauttamatta välillä on merkityksetön, jos otoskoko n on *pieni* perusjoukon kokoon N verrattuna ja ero häviää kokonaan, jos perusjoukko on *ääretön*.

16.7. Poisson-jakauma

Toistetaan *samaa satunnaiskoetta* ja oletetaan, että toistot ovat toisistaan *riippumattomia*. Tarkastellaan jonkin *tapahtuman A* sattumista toistojen aikana. Oletetaan, että tapahtuman A *tapahtumaintensiteetti* eli tapahtuman A esiintymisten *keskimääräinen lukumäärä aikayksikköä kohden* on λ.

Määritellään *diskreetti satunnaismuuttuja X*:

X = Tapahtuman A esiintymisten *lukumäärä ajanjaksona, jonka kesto on s aikayksikköä*

Tietyin oletuksin satunnaismuuttuja X noudattaa **Poisson-jakaumaa** parametrilla λs:

$$X \sim \mathrm{Poisson}(\lambda s)$$

jossa

s = ajanjakson pituus aikayksiköissä

λ = tapahtuman A esiintymisten *keskimääräinen lukumäärä aikayksikköä* kohden

ja sen *pistetodennäköisyysfunktio* on

$$f(x) = \Pr(X = x) = \frac{e^{-\lambda s}(\lambda s)^x}{x!}\ ,\ \lambda > 0\ ,\ x = 0, 1, 2, \ldots$$

Huomautus:

- Poisson-jakauma syntyy myös sellaisessa tilanteessa, jossa tarkastellaan tapahtuman A esiintymistä *avaruudessa*. Tällöin parametri λ kuvaa tapahtumien A esiintymisten *keskimääräistä lukumäärää tilavuusyksikköä kohden* ja satunnaismuuttuja X kuvaa tapahtuman A esiintymisten *lukumäärää avaruuden osa-alueessa, jonka koko on s tilavuusyksikköä*.

Poisson-jakauman johto

Luonnostelemme seuraavassa vain pääkohdat Poisson-jakauman johdosta.

Tarkastellaan jonkin tapahtuman A sattumista *saman* satunnaiskokeen toistojen aikana.

Oletukset:

(1) Toistot ovat toisistaan *riippumattomia*.

(2) $\Pr(\text{Yksi tapahtuma } A \textit{ lyhyellä aikavälillä } ds) = \lambda ds$

(3) Aikaväli on ds on *niin lyhyt*, että todennäköisyys

$$\Pr(k \text{ kappaletta tapahtumia } A \text{ aikavälillä } ds,\ k > 1)$$

on häviävän pieni eli kertaluokkaa $o(s)$.

Olkoon

$$f(x; s) = \Pr(x \text{ kappaletta tapahtumia } A \text{ aikavälillä } [0, s])$$

Oletusten (1)-(3) pätiessä aikavälillä

$$[0, s + ds]$$

voi sattua x kappaletta tapahtumia A kahdella *toisensa poissulkevalla tavalla* (tn = todennäköisyys):

Tapaus 1: x kappaletta tapahtumia A ajanhetkeen s mennessä: $tn = f(x; s)$

Ei tapahtumia A aikavälillä ds: $tn = 1 - \lambda ds$

Lisäksi nämä ovat tapahtumina toisistaan *riippumattomia*.

Tapaus 2: $(x - 1)$ kappaletta tapahtumia A ajanhetkeen s mennessä: $tn = f(x - 1; s)$

Yksi tapahtuma A aikavälillä ds: $tn = \lambda ds$

Lisäksi nämä ovat tapahtumina toisistaan *riippumattomia*.

Riippumattomien tapahtumien tulosäännön ja *toisensa poissulkevien tapahtumien yhteenlasku-säännön* mukaan

$$f(x; s + ds) = f(x; s)(1 - \lambda ds) + f(x - 1; s)\lambda ds$$

josta saamme *erotusosamäärän*

$$\frac{f(x; s + ds) - f(x; s)}{ds} = \lambda[f(x - 1; s) - f(x; s)]$$

Antamalla

$$ds \to 0$$

saamme (x:n suhteen) *differenssiyhtälön*

$$\frac{df(x; s)}{ds} = \lambda[f(x - 1; s) - f(x; s)]$$

On helppo nähdä, että

$$f(x) = \frac{e^{-\lambda s}(\lambda s)^x}{x!},\ x = 0, 1, 2, \ldots$$

on tämän differenssiyhtälön *ratkaisu*.

Funktio $f(x)$ määrittelee todennäköisyysjakauman, koska

$$f(x) > 0\,,\ x = 0, 1, 2, \ldots$$

ja *eksponenttifunktion määritelmän* mukaan

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda s}(\lambda s)^x}{x!} = e^{-\lambda s}\sum_{x=0}^{\infty}\frac{(\lambda s)^x}{x!} = e^{-\lambda s}e^{\lambda s} = 1$$

Poisson-jakauman tunnusluvut

Olkoon

$$X \sim \text{Poisson}(\lambda s)$$

Odotusarvo:

$$\text{E}(X) = \mu_X = \lambda s$$

Varianssi:

$$\text{Var}(X) = \text{D}^2(X) = \sigma_X^2 = \lambda s$$

Standardipoikkeama:

$$\text{D}(X) = \sigma_X = \sqrt{\lambda s}$$

Perustelu:

Johdamme tässä Poisson-jakauman odotusarvon käyttäen diskreetin jakauman odotusarvon määritelmää. Jakauman odotusarvo ja varianssi johdetaan jakauman *momenttiemäfunktion* avulla kohdassa **Poisson-jakauman momenttiemäfunktio ja jakauman tunnusluvut**.

Suoraan diskreetin jakauman odotusarvon määritelmän mukaan ja eksponenttifunktion ominaisuuksien mukaan

$$\begin{aligned}
\text{E}(X) = \sum_{x=0}^{\infty} x f(x) &= \sum_{x=0}^{\infty} x\frac{e^{-\lambda s}(\lambda s)^x}{x!} \\
&= e^{-\lambda s}\sum_{x=1}^{\infty} x\frac{(\lambda s)^x}{x!} \\
&= \lambda s e^{-\lambda s}\sum_{x=1}^{\infty}\frac{(\lambda s)^{x-1}}{(x-1)!} \\
&= \lambda s e^{-\lambda s}e^{\lambda s} = \lambda s
\end{aligned}$$

■

Odotusarvon ominaisuudet

Poisson-jakauman odotusarvo λ kuvaa *tapahtumaintensiteettiä*: Mitä suurempi on λ, sitä enemmän tapahtumia on odotettavissa aikayksikköä kohden.

Poisson-jakauman pistetodennäköisyysfunktio

Kuva okealla esittää Poisson-jakauman

Poisson(5)

pistetodennäköisyysfunktiota

$$f(x) = \frac{e^{-\lambda s}(\lambda s)^x}{x!}$$

pisteissä

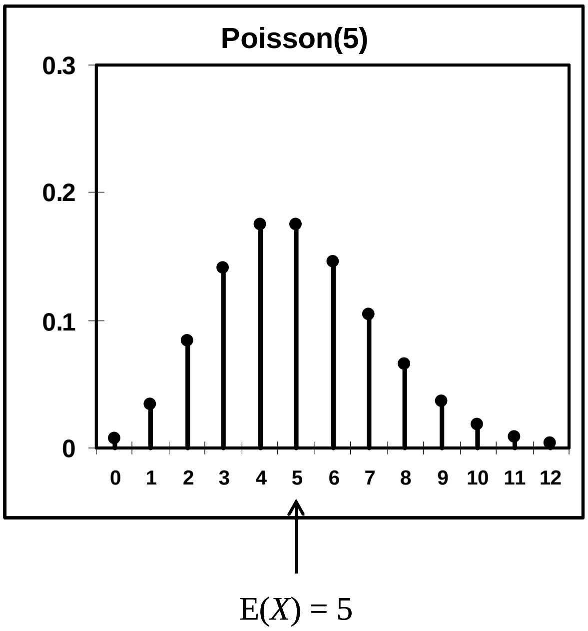
x = 0, 1, 2, … , 12

kun

$$\lambda s = 5$$

Kuvaan on merkitty myös jakauman *odotusarvo*

$$\mathrm{E}(x) = \lambda s = 5$$



Esimerkki 1. Radioaktiivinen hajoaminen.

Tarkastellaan *radioaktiivista hajoamista* ja olkoon

X = aikavälillä [0, s] hajoavien atomien lukumäärä

Tällöin

X ~ Poisson(λs)

jossa λ on *alkuainekohtainen parametri*, joka kuvaa *keskimäärin aikayksikköä kohden hajoavien atomien lukumäärää.*

Poisson-jakauman momenttiemäfunktio

Poisson-jakauman

$$f(x) = \Pr(X = x) = \frac{e^{-\lambda s}(\lambda s)^x}{x!}\ ,\ \lambda > 0\ ,\ x = 0,1,2,\ldots$$

momenttiemäfunktio (ks. lukua **Momenttiemäfunktio ja karakteristinen funktio**) on

$$m_X(t) = \mathrm{E}(e^{tX}) = e^{\lambda s(e^t - 1)}$$

Perustelu:

Suoraan diskreetin jakauman momenttiemäfunktion määritelmän mukaan

$$m_X(t) = \mathrm{E}(e^{tX}) = \sum_x e^{tx} f(x) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda s}(\lambda s)^x}{x!} = e^{-\lambda s} \sum_{x=0}^{\infty} \frac{(\lambda s e^t)^x}{x!} = e^{-\lambda s} e^{\lambda s e^t} = e^{\lambda s(e^t - 1)}$$

■

Poisson-jakauman momenttiemäfunktio ja jakauman tunnusluvut

Johdetaan Poisson-jakauman Poisson(λs) *odotusarvo* ja *varianssi* jakauman momenttiemäfunktion avulla.

Odotusarvo:

$$\mathrm{E}(X) = \mu_X = \lambda s$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2 = \lambda s$$

Perustelu:

Poisson-jakauman Poisson(λ) *momenttiemäfunktio* on

$$m_X(t) = e^{\lambda s(e^t - 1)}$$

Momenttiemäfunktion 1. derivaatta pisteessä $t = 0$:

$$\left.\frac{dm_X(t)}{dt}\right|_{t=0} = e^{\lambda s(e^t-1)}\lambda s e^t\Big|_{t=0} = \lambda s\, e^{t+\lambda s(e^t-1)}\Big|_{t=0} = \lambda s$$

Momenttiemäfunktion 2. derivaatta pisteessä $t = 0$:

$$\left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \lambda s\, e^{t+\lambda s(e^t-1)}(1+\lambda s e^t)\Big|_{t=0} = \lambda s + (\lambda s)^2$$

Siten Poisson-jakauman Poisson(λ) *odotusarvo* μ_X, 2. *momentti* α_2 ja *varianssi* σ_X^2 saadaan seuraavilla kaavoilla:

$$\mu_X = \mathrm{E}(X) = \alpha_1 = \left.\frac{dm_X(t)}{dt}\right|_{t=0} = \lambda s$$

$$\alpha_2 = \mathrm{E}(X^2) = \left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \lambda s + (\lambda s)^2$$

$$\sigma_X^2 = \mathrm{Var}(X) = \alpha_2 - \alpha_1^2 = \lambda s + (\lambda s)^2 - (\lambda s)^2 \quad = \lambda s$$

■

Poisson-jakaumaa noudattavien satunnaismuuttujien summan jakauma

Olkoot

$$X_1, X_2, \ldots, X_k$$

riippumattomia satunnaismuuttujia, jotka noudattavat *Poisson-jakaumia* parametrein $\lambda_1 s$, $\lambda_2 s$, … ja $\lambda_k s$:

$$X_1, X_2, \ldots, X_k \perp$$

$$X_i \quad \mathrm{Poisson}(\lambda_i s),\ i = 1, 2, \ldots, k$$

Tällöin satunnaismuuttujien $X_1, X_2, \ldots, X_n$ *summa*

$$Y = X_1 + X_2 + \cdots + X_k$$

on diskreetti satunnaismuuttuja, joka noudattaa *Poisson-jakaumaa* parametrilla $\lambda s = \lambda_1 s + \lambda_2 s + \cdots + \lambda_k s$:

$$X \sim \text{Poisson}(\lambda s)$$

Perustelu:

Todistus perustuu siihen, että *riippumattomien satunnaismuuttujien summan momentti-emäfunktio ko. satunnaismuuttujien momenttiemäfunktioiden tulo*.

Oletetaan, että

$$X_1, X_2, \ldots, X_k \perp$$
$$X_i \sim \text{Poisson}(\lambda_i)\,,\ i = 1, 2, \ldots, k$$

Satunnaismuuttujien X_1, X_2, … , X_k *momenttiemäfunktiot* ovat muotoa

$$m_i(t) = e^{\lambda_i s(e^t - 1)}\,,\ i = 1, 2, \ldots, k$$

Muodostetaan satunnaismuuttujien X_1, X_2, … , X_k *summa*:

$$Y = X_1 + X_2 + \cdots + X_k$$

Summamuuttujan Y momenttiemäfunktio on

$$m_Y(t) = m_1(t) m_2(t) \cdots m_k(t) = e^{\lambda_1 s(e^t-1)} e^{\lambda_2 s(e^t-1)} \cdots e^{\lambda_k s(e^t-1)} = e^{(\lambda_1 + \lambda_2 + \cdots + \lambda_1) s(e^t-1)} = e^{\lambda s(e^t-1)}$$

jossa

$$\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_k$$

Funktio $m_Y(t)$ on *Poisson-jakauman*

$$\text{Poisson}(\lambda s)$$

momenttiemäfunktio, jossa

$$\lambda s = \lambda_1 s + \lambda_2 s + \cdots + \lambda_k s$$

Lisäksi funktio $m_Y(t)$ on selvästi jatkuva pisteen

$$t = 0$$

ympäristössä. Koska momenttiemäfunktio $m_Y(t)$ on tällöin *yksikäsitteinen*, summamuuttuja Y noudattaa *Poisson-jakaumaa* parametrilla $\lambda s = \lambda_1 s + \lambda_2 s + \cdots + \lambda_k s$:

$$Y = X_1 + X_2 + \cdots + X_k \sim \text{Poisson}(\lambda s)$$

■

Huomautus:

- Jokaisella Poisson-jakaumalla saa olla *eri* parametri.

Poisson-jakauma ja binomijakauma

Olkoon

$$X_n \sim \text{Bin}(n, p)$$

Annetaan

$$n \to \infty$$

ja

$$p \to 0$$

siten, että

$$np = \lambda s$$

Tällöin *satunnaismuuttujan* X_n *jakauma lähestyy Poisson-jakaumaa parametrilla* λ:

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda s}} f_{\mathrm{Bin}(n,p)}(x) = f_{\mathrm{Poisson}(\lambda s)}(x)\,,\ x = 0,1,2,\ldots$$

Perustelu:

Olkoon

$$X \sim \mathrm{Bin}(n, p)$$

Tällöin

$$f_X(x) = \binom{n}{x} p^x q^{n-x}\,,\ q = 1-p\,,\ x = 0,1,2,\ldots,n$$

Oletetaan, että

$$n \to +\infty$$

ja samaan aikaan

$$p \to 0$$

niin, että

$$np = \lambda s$$

jossa $\lambda > 0$ on ei-satunnainen *vakio*. Ottamalla huomioon, että $np = \lambda s$ ja $q = 1 - p$, voimme kirjoittaa, että

$$\begin{aligned}
f_X(x) &= \binom{n}{x} p^x q^{n-x} \\
&= \frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x x!}(np)^x(1-p)^{n-x} \\
&= \frac{n(n-1)(n-2)\cdots(n-x+1)}{n^x}\cdot\frac{(\lambda s)^x}{x!}\cdot\frac{(1-p)^n}{(1-p)^x} \\
&= 1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\cdot\frac{(\lambda s)^x}{x!}\cdot\frac{(1-p)^n}{(1-p)^x}
\end{aligned}$$

Edelleen voimme kirjoittaa, että

$$(1-p)^n = [(1-p)^{-1/p}]^{-np} = [(1-p)^{-1/p}]^{-\lambda}$$

Eksponenttifunktio ominaisuuksien perusteella

(1) $$\lim_{p\to 0}[(1-p)^{-1/p}]^{-\lambda} = e^{-\lambda}$$

Lisäksi pätee:

(2) $$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)=1$$

(3) $$\lim_{p\to 0}(1-p)^x = 1$$

Yhdistämällä kaavat (1), (2) ja (3) saadaan haluttu lopputulos:

$$\lim_{\substack{n\to\infty\\ p\to 0\\ np=\lambda s}} f_X(x) = \frac{e^{-\lambda s}(\lambda s)^x}{x!}$$

■

Poisson-jakauman ja binomijakauman välinen yhteys tulee esille myös siinä, että

$$\mathrm{E}(X) = \lambda s \approx np$$

jos n on *suuri* ja p on *pieni*. Lisäksi myös jakaumien varianssit ovat tällöin lähellä toisiaan:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \lambda s \approx np \approx np(1-p)$$

koska

$$p \approx 0 \Rightarrow 1 - p \approx 1$$

Poisson-jakauma ja eksponenttijakauma

Tarkastellaan jonkin tapahtuman A sattumista *saman* satunnaiskokeen toistojen aikana ja olkoon

X = tapahtumien lukumäärä aikayksikköä kohden

Oletetaan, että

$X \sim \mathrm{Poisson}(\lambda)$

jolloin

$\mathrm{E}(X) = \lambda$

Parametri λ kuvaa tarkastelun kohteena olevan tapahtuman *tapahtumaintensiteettiä* eli tarkastelun kohteena olevan tapahtuman esiintymisten keskimääräistä lukumäärää aikayksikköä kohden.

Olkoon

Y = odotusaika 1. tapahtumalle (tai tapahtumien väliaika)

Tällöin Y on *jatkuva satunnaismuuttuja*, joka noudattaa **eksponenttijakaumaa** parametrilla λ:

$Y \sim \mathrm{Exp}(\lambda)$

jolloin

$\mathrm{E}(Y) = 1/\lambda$

Lisätietoja eksponenttijakaumasta: ks. lukua **Jatkuvia jakaumia**.

Voidaan osoittaa, että jakaumien välinen yhteys toimii myös toiseen suuntaan: Olkoon siis

Y = odotusaika 1. tapahtumalle (tai tapahtumien väliaika) ja oletetaan, että

$$Y \sim \mathrm{Exp}(\lambda)$$

Olkoon

X = tapahtumien lukumäärä aikayksikköä kohden

Tällöin X on *diskreetti satunnaismuuttuja*, joka noudattaa **Poisson-jakaumaa** parametrilla λ:

$$X \sim \text{Poisson}(\lambda)$$

17. Jatkuvia jakaumia

Määrittelemme tässä luvussa seuraavat *jatkuvat todennäköisyysjakaumat*:

- **Jatkuva tasainen jakauma**
- **Eksponenttijakauma**
- **Normaalijakauma**
- **Log-normaalijakauma**
- **Cauchy-jakauma**
- **Gamma-jakauma**
- **Beta-jakauma**
- **Weibull-jakauma**

Tarkastelemme jakaumien ominaisuuksia havainnollistamalla jakaumien **tiheysfunktioita graafisesti** ja johtamalla jakaumien **odotusarvot** ja **varianssit** sekä (useimmissa tapauksissa) jakaumien **momenttiemäfunktiot**. Tarkastelemme myös jakaumien yhteyksiä muihin jakaumiin.

Kohdassa **Keskeinen raja-arvolause** esitämme perustelun sille, miksi normaalijakauma on ”normaali”.

Avainsanat:

Binomijakauma, Beta-jakauma, Cauchy-jakauma, Eksponenttijakauma, Gamma-jakauma, Jatkuva tasainen jakauma, Kertymäfunktio, Keskeinen raja-arvolause, Log-normaalijakauma, Momenttiemäfunktio, Normaaliapproksimaatio, Normaalijakauma, Odotusarvo, Poisson-jakauma, Standardipoikkeama, Standardointi, Taulukot, Tiheysfunktio, Varianssi, Weibull-jakauma

17.1. Jatkuva tasainen jakauma

Jatkuva tasainen jakauma on *diskreetin tasaisen jakauman* (ks. lukua **Diskreettejä jakaumia**) *jatkuva vastine*.

Olkoon *jatkuvan satunnaismuuttujan X tiheysfunktio*

$$f(x) = \begin{cases} 0 & , x < a \\ \dfrac{1}{b-a} & , a \leq x \leq b \\ 0 & , x > a \end{cases}$$

Tällöin satunnaismuuttuja X noudattaa **jatkuvaa tasaista jakaumaa** parametrein a ja b ja merkitsemme:

$$X \quad \mathrm{Uniform}(a,b)$$

Funktio $f(x)$ kelpaa tiheysfunktioksi, koska

$$f(x) > 0 \,,\; x \in [a,b]$$

ja

$$\int_{-\infty}^{+\infty} f(x)dx = \int_a^b \frac{1}{b-a}dx = \left[\frac{x}{b-a}\right]_a^b = \frac{b-a}{b-a} = 1$$

Jakauman *nimi* johtuu siitä, että jatkuvalla tasaisella jakaumalla on seuraava ominaisuus:

Olkoot $[c_1,d_1]$ ja $[c_2,d_2]$ jatkuvan tasaisen jakauman $\mathrm{Uniform}(a,b)$ määrittelyvälin $[a, b]$ kaksi *mielivaltaista*, *samanpituista* osaväliä:

$$[c_1,d_1] \subset [a, b]$$

$$[c_2,d_2] \subset [a, b]$$

$$d_1 - c_1 = d_2 - c_2$$

Tällöin väleihin $[c_1,d_1]$ ja $[c_2,d_2]$ liittyvät todennäköisyydet ovat *yhtä suuria*.

Perustelu:

Todetaan ensin, että

$$\Pr(X \in [c_1,d_1]) = \int_{c_1}^{d_1} f(x)dx = \int_{c_1}^{d_1} \frac{1}{b-a}dx = \left[\frac{x}{b-a}\right]_{c_1}^{d_1} = \frac{d_1-c_1}{b-a}$$

ja vastaavasti

$$\Pr(X \in [c_2,d_2]) = \int_{c_2}^{d_2} f(x)dx = \int_{c_2}^{d_2} \frac{1}{b-a}dx = \left[\frac{x}{b-a}\right]_{c_2}^{d_2} = \frac{d_2-c_2}{b-a}$$

Koska oletuksen mukaan

$$d_1 - c_1 = d_2 - c_2$$

niin

$$\Pr(X \in [c_1, d_1]) = \Pr(X \in [c_2, d_2])$$

■

Jatkuvan tasaisen jakauman tunnusluvut

Olkoon

$$X \quad \text{Uniform}(a,b)$$

Odotusarvo:

$$\mathrm{E}(X) = \frac{a+b}{2}$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \frac{(b-a)^2}{12}$$

Standardipoikkeama:

$$\mathrm{D}(X) = \frac{b-a}{2\sqrt{3}}$$

Perustelu:

Johdamme tässä jatkuvan tasaisen jakauman odotusarvon käyttäen jatkuvan jakauman odotusarvon määritelmää. Jakauman odotusarvo ja varianssi johdetaan jakauman *momentti-emäfunktion* avulla kohdassa **Jatkuvan tasaisen jakauman momenttiemäfunktio ja jakauman tunnusluvut**.

Suoraan jatkuvan jakauman odotusarvon määritelmästä saadaan, että

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}\left[\frac{1}{2}x^2\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

■

Jatkuvan tasaisen jakauman tiheysfunktio

Kuva oikealla esittää jatkuvan tasaisen jakauman

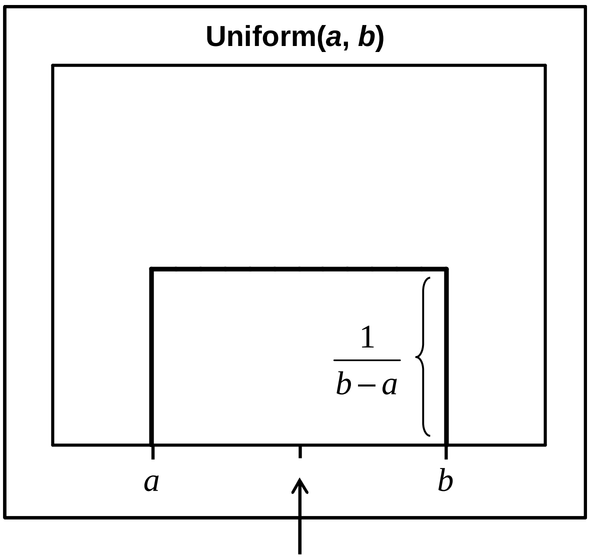
Uniform(a, b)

tiheysfunktiota

$$f(x) = \begin{cases} 0 & , x < a \\ \dfrac{1}{b-a} & , a \le x \le b \\ 0 & , x > a \end{cases}$$

Kuvaan on merkitty myös jakauman *odotusarvo*:

$$\mathrm{E}(X) = \frac{a+b}{2}$$



$$\mathrm{E}(X) = \frac{a+b}{2}$$

Jatkuvan tasaisen jakauman tiheysfunktiolla $f(x)$ on seuraavat ominaisuudet:

(i) Tiheysfunktio $f(x)$ saa positiivisen *vakioarvon*

$$1/(b-a)$$

välillä $[a, b]$ ja arvon 0 välin $[a, b]$ ulkopuolella.

(ii) Tiheysfunktio on *symmetrinen* pisteen

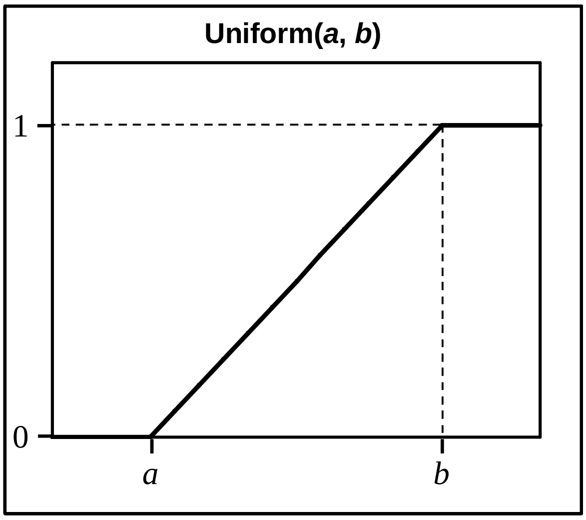
$$\mathrm{E}(X) = (a+b)/2$$

suhteen.

Jatkuvan tasaisen jakauman kertymäfunktio

Jatkuvan tasaisen jakauman *kertymäfunktio* on

$$F(x) = \Pr(X \le x) = \begin{cases} 0 & , x \le a \\ \dfrac{x-a}{b-a} & , a \le x \le b \\ 1 & , x \ge b \end{cases}$$



Ks. kuvaa oikealla

Perustelu:

Olkoon

$$X \quad \mathrm{Uniform}(a,b)$$

Satunnaismuuttujan X *tiheysfunktion* lauseke on välillä $[a, b]$:

$$f(x) = \frac{1}{b-a}$$

Siten satunnaismuuttujan X *kertymäfunktion* lausekkeeksi saadaan, kun $x \in [a, b]$:

$$F(x) = \Pr(X \le x) = \int_{-\infty}^{x} f(t)dt = \int_{a}^{x} \frac{1}{b-a}dt = \left[\frac{t}{b-a}\right]_a^x = \frac{x-a}{b-a}$$

■

Jatkuvan tasaisen jakauman momenttiemäfunktio

Jatkuvan tasaisen jakauman

$$f(x) = \frac{1}{b-a}, a \le x \le b$$

momenttiemäfunktio (ks. lukua **Momenttiemäfunktio ja karakteristinen funktio**) on

$$m_X(t) = \mathrm{E}(e^{tX}) = \frac{e^{bt} - e^{at}}{t(b-a)}$$

Perustelu:

Suoraan jatkuvan jakauman momenttiemäfunktion määritelmän mukaan

$$m_X(t) = \mathrm{E}(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} f(x)dx = \int_a^b e^{tx} \frac{1}{b-a} dx = \frac{1}{b-a}\left[\frac{e^{tx}}{t}\right]_a^b = \frac{e^{bt}-e^{at}}{t(b-a)}$$

■

Jatkuvan tasaisen jakauman momenttiemäfunktio ja jakauman tunnusluvut

Johdetaan jatkuvan tasaisen jakauman Uniform(a, b) *odotusarvo* ja *varianssi* jakauman momenttiemäfunktion avulla.

Odotusarvo:

$$\mathrm{E}(X) = \frac{a+b}{2}$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \frac{(b-a)^2}{12}$$

Perustelu:

Jatkuvan tasaisen jakauman Uniform(a, b) *momenttiemäfunktio* on

$$m_X(t) = \frac{e^{bt}-e^{at}}{t(b-a)}$$

Momenttiemäfunktion 1. derivaatta pisteessä t = 0:

$$\begin{aligned}
\left.\frac{dm_X(t)}{dt}\right|_{t=0} &= \left.\frac{(be^{bt}-ae^{at})t(b-a)-(e^{bt}-e^{at})(b-a)}{t^2(b-a)^2}\right|_{t=0} \\
&= \left.\frac{(be^{bt}-ae^{at})t-(e^{bt}-e^{at})}{t^2(b-a)}\right|_{t=0} \\
&= \frac{a+b}{2}
\end{aligned}$$

Momenttiemäfunktion 2. derivaatta pisteessä t = 0:

$$\begin{aligned}
\left.\frac{d^2m_X(t)}{dt^2}\right|_{t=0} &= \left\{\frac{[(b^2e^{bt}-a^2e^{at})t+(be^{bt}-ae^{at})-(be^{bt}-ae^{at})]t^2(b-a)}{t^4(b-a)^2}\right. \\
&\qquad \left.\left. -\frac{[(be^{bt}-ae^{at})t-(e^{bt}-e^{at})]2t(b-a)}{t^4(b-a)^2}\right\}\right|_{t=0} \\
&= \left.\frac{(b^2e^{bt}-a^2e^{at})t^2-2(be^{bt}-ae^{at})t+2(e^{bt}-e^{at})}{t^3(b-a)}\right|_{t=0} \\
&= \frac{a^2+ab+b^2}{3}
\end{aligned}$$

Siten jatkuvan tasaisen jakauman Uniform(a, b) *odotusarvo* μ_X , 2. *momentti* α_2 ja *varianssi* σ_X^2 saadaan seuraavilla kaavoilla:

$$\mu_X = \mathrm{E}(X) = \alpha_1 = \left.\frac{dm_X(t)}{dt}\right|_{t=0} = \frac{a+b}{2}$$

$$\alpha_2 = \mathrm{E}(X^2) = \left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \frac{a^2+ab+b^2}{3}$$

$$\sigma_X^2 = \mathrm{Var}(X) = \alpha_2 - \alpha_1^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

■

Jatkuvan tasaisen jakauman todennäköisyydet

Olkoon

$$X \quad \mathrm{Uniform}(a,b)$$

ja olkoon

$$[c, d] \subset [a, b]$$

jokin jakauman määrittelyvälin [a, b] *osaväli*. *Välin* [c, d] *todennäköisyys* saadaan integroimalla tiheysfunktio

$$f(x) = \begin{cases} 0 & , x < a \\ \dfrac{1}{b-a} & , a \le x \le b \\ 0 & , x > a \end{cases}$$

välillä [c, d]:

$$\Pr(c \le X \le d) = \int_c^d f(x)dx = \int_c^d \frac{1}{b-a}dx = \left[\frac{x}{b-a}\right]_c^d = \frac{d-c}{b-a}$$

Kaikkien muiden jatkuvaan tasaiseen jakaumaan Uniform(a, b) liittyvien tapahtumien todennäköisyydet saadaan välin [a, b] osavälien todennäköisyyksistä *todennäköisyyslaskennan laskusääntöjen* avulla.

17.2. Eksponenttijakauma

Eksponenttijakauma on *geometrisen jakauman* (ks. lukua **Diskreettejä jakaumia**) *jatkuva vastine* ja sitä käytetään usein *elinajan* mallintamiseen. Tietyt ehdot täyttävissä *jonotapahtumissa* 1. *tapahtuman odotuaika* ja *tapahtumien väliaika* noudattaa eksponenttijakaumaa.

Olkoon *jatkuvan satunnaismuuttujan X tiheysfunktio*

$$f(x) = \lambda\exp(-\lambda x)\ ,\ \lambda > 0\ ,\ x \ge 0$$

Tällöin satunnaismuuttuja X noudattaa **eksponenttijakaumaa** parametrilla λ ja merkitsemme:

$$X \quad \mathrm{Exp}(\lambda)$$

Johdamme eksponenttijakauman tiheysfunktion kohdassa **Eksponenttijakauma ja Poisson-jakauma**.

Funktio $f(x)$ kelpaa tiheysfunktioksi, koska

$$f(x) > 0\,,\; x \geq 0$$

ja

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{0}^{+\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^{+\infty} = 1$$

Eksponenttijakauman tunnusluvut

Olkoon

$$X \quad \mathrm{Exp}(\lambda)$$

Odotusarvo:

$$\mathrm{E}(X) = \frac{1}{\lambda}$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \frac{1}{\lambda^2}$$

Standardipoikkeama:

$$\mathrm{D}(X) = \frac{1}{\lambda}$$

Perustelu:

Johdamme tässä eksponenttijakauman odotusarvon käyttäen jatkuvan jakauman odotusarvon määritelmää. Jakauman odotusarvo ja varianssi johdetaan jakauman *momenttiemäfunktion* avulla kohdassa **Eksponenttijakauman momenttiemäfunktio ja jakauman tunnusluvut**.

Jatkuvan jakauman odotusarvon määritelmästä saadaan soveltamalla *osittaisintegrointia*, että

$$\mathrm{E}(X) = \int_0^{+\infty} xf(x)dx = \int_0^{+\infty} x\lambda e^{-\lambda x} dx = \left[-xe^{-\lambda x}\right]_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx = \left[-\frac{1}{\lambda} e^{-\lambda x}\right]_0^{+\infty} = \frac{1}{\lambda}$$

■

Eksponenttijakauman tiheysfunktio

Kuva alla esittää eksponenttijakauman

$$\mathrm{Exp}(\lambda)$$

tiheysfunktiota

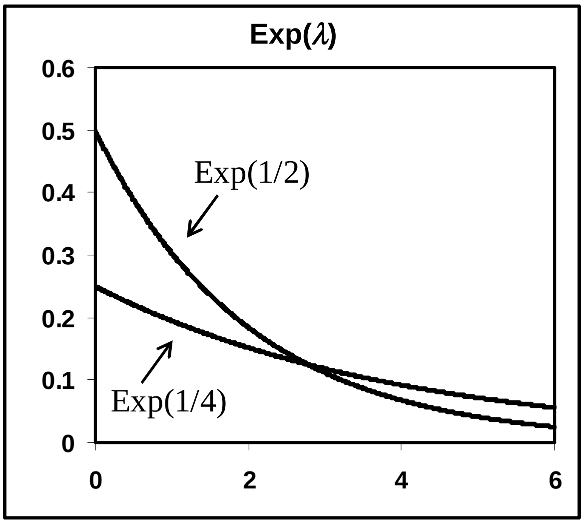
$$f(x) = \lambda \exp(-\lambda x)$$

$$\lambda > 0\,,\; x \geq 0$$

välillä [0, 6] seuraavissa tapauksissa:

(i) $\lambda = 1/2$

(ii) $\lambda = 1/4$



Eksponenttijakauman *odotusarvo*:

$$\mathrm{E}(X) = \frac{1}{\lambda}$$

Eksponenttijakauman tiheysfunktiolla on seuraavat ominaisuudet:

(i) Eksponenttijakauman tiheysfunktio $f(x)$ saa *positiivisia* arvoja kaikille ei-negatiivisille argumentin x arvoille:

$$f(x) > 0,\ x \geq 0$$

(ii) Tiheysfunktio *monotonisesti laskeva* ja sillä on *maksimi* pisteessä $x = 0$.

(iii) Tiheysfunktio on *vino oikealle*.

Eksponenttijakauman kertymäfunktio

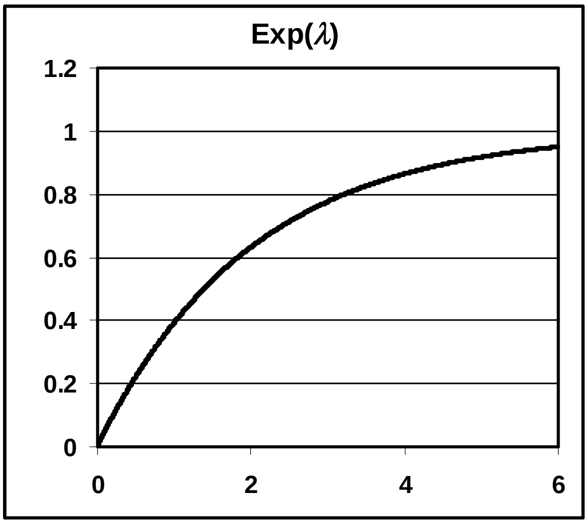
Eksponenttijakauman *kertymäfunktioksi* saadaan

$$F(x) = \int_0^x f(t)dt = \int_0^x \lambda \exp(-\lambda t)dt = \left[-\exp(-\lambda t)\right]_0^x = 1 - \exp(-\lambda x)$$

$$\lambda > 0\,,\ x \geq 0$$

Komplementtitapahtuman todennäköisyyden kaavan mukaan

$$\begin{aligned}\Pr(X > x) &= 1 - \mathrm{P}(X \leq x)\\ &= 1 - F(x)\\ &= \exp(-\lambda x)\end{aligned}$$



Kuva oikealla esittää eksponenttijakauman

$$\mathrm{Exp}(\lambda)$$

kertymäfunktiota

$$F(x) = 1 - \exp(-\lambda x)$$

$$\lambda > 0\,,\ x \geq 0$$

välillä [0, 6], kun

$$\lambda = 1/2$$

Eksponenttijakauman momenttiemäfunktio

Eksponenttijakauman

$$f(x) = \lambda e^{-\lambda x}\,,\ \lambda > 0\,,\ x \geq 0$$

momenttiemäfunktio (ks. lukua **Momenttiemäfunktio ja karakteristinen funktio**) on

$$m_X(t) = \mathrm{E}(e^{tX}) = \frac{\lambda}{\lambda - t}$$

Perustelu:

Suoraan jatkuvan jakauman momenttiemäfunktion määritelmän mukaan

$$\begin{aligned} m_X(t) = \mathrm{E}(e^{tX}) &= \int_{-\infty}^{+\infty} e^{tx} f(x)dx \\ &= \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^{\infty} e^{(t-\lambda)x} dx \\ &= \lambda \left[\frac{e^{(t-\lambda)x}}{t-\lambda} \right]_0^{\infty} \\ &= \frac{\lambda}{\lambda - t} \end{aligned}$$

■

Eksponenttijakauman momenttiemäfunktio ja jakauman tunnusluvut

Johdetaan eksponenttijakauman Exp(λ) *odotusarvo* ja *varianssi* jakauman momenttiemäfunktion avulla.

Odotusarvo:

$$\mathrm{E}(X) = \frac{1}{\lambda}$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \frac{1}{\lambda^2}$$

Perustelu:

Eksponenttijakauman Exp(λ) momenttiemäfunktio on

$$m_X(t) = \frac{\lambda}{\lambda - t}$$

Momenttiemäfunktion 1. derivaatta pisteessä $t = 0$:

$$\left.\frac{dm_X(t)}{dt}\right|_{t=0} = \left.\frac{\lambda}{(\lambda - t)^2}\right|_{t=0} = \frac{1}{\lambda}$$

Momenttiemäfunktion 2. derivaatta pisteessä $t = 0$:

$$\left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \left.\frac{2\lambda}{(\lambda - t)^3}\right|_{t=0} = \frac{2}{\lambda^2}$$

Siten eksponenttijakauman Exp(λ) *odotusarvo* μ_X, 2. *momentti* α_2 ja *varianssi* σ_X^2 saadaan seuraavilla kaavoilla:

$$\mu_X = \mathrm{E}(X) = \alpha_1 = \left.\frac{dm_X(t)}{dt}\right|_{t=0} = \frac{1}{\lambda}$$

$$\alpha_2 = \mathrm{E}(X^2) = \left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \frac{2}{\lambda^2}$$

$$\sigma_X^2 = \mathrm{Var}(X) = \alpha_2 - \alpha_1^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

■

Eksponenttijakauma ja Poisson-jakauma

Tarkastellaan jonkin tapahtuman sattumista *saman* satunnaiskokeen toistojen aikana ja olkoon

X = odotusaika 1. tapahtumalle (tai tapahtumien väliaika)

Oletetaan, että

$X \sim \mathrm{Exp}(\lambda)$

jolloin

$\mathrm{E}(X) = 1/\lambda$

Olkoon

Z = tapahtumien lukumäärä aikayksikköä kohden

Tällöin Z on *diskreetti satunnaismuuttuja*, joka noudattaa **Poisson-jakaumaa** parametrilla λ:

$Z \sim \mathrm{Poisson}(\lambda)$

jolloin siis

$\mathrm{E}(Z) = \lambda$

Parametri λ kuvaa tarkastelun kohteena olevan tapahtuman *tapahtumaintensiteettiä* eli tarkastelun kohteena olevan tapahtuman esiintymisten keskimääräistä lukumäärää aikayksikköä kohden. Lisätietoja Poisson-jakaumasta; ks. lukua **Diskreettejä jakaumia**.

Voidaan osoittaa, että jakaumien välinen yhteys toimii myös toiseen suuntaan: Olkoon siis

Z = tapahtumien lukumäärä aikayksikköä kohden

ja oletetaan, että

$Z \sim \mathrm{Poisson}(\lambda)$

Olkoon

X = odotusaika 1. tapahtumalle (tai tapahtumien väliaika)

Tällöin X on *jatkuva satunnaismuuttuja*, joka noudattaa **eksponenttijakaumaa** parametrilla λ:

$X \sim \mathrm{Exp}(\lambda)$

Perustelu:

Olkoon

X = odotusaika 1. tapahtumalle

Johdetaan satunnaismuuttujan X *kertymäfunktio* F_X.

Kertymäfunktion määritelmän *ja komplementtitapahtuman todennäköisyyden* kaavan mukaan

(∗) $$F_X(x) = \Pr(X \le x) = 1 - \Pr(X > x)$$

Ensimmäinen tapahtuma sattuu ajanhetken x jälkeen, jos ja vain jos aikavälillä $[0, x]$ *ei ole sattunut yhtään tapahtumaa*. Siten

$$\Pr(X > x) = \Pr(Z = 0)$$

jossa

$$Z \sim \text{Poisson}(\lambda x)$$

Poisson-jakauman *pistetodennäköisyysfunktion* kaavasta

$$f(z) = \Pr(Z = z) = \frac{e^{-\lambda x}(\lambda x)^z}{z!}, \ z = 0, 1, 2, \ldots$$

saadaan:

$$\Pr(X > x) = \Pr(Z = 0) = \exp(-\lambda x)$$

Sijoittamalla tämä kaavaan (∗) saamme satunnaismuuttujan X kertymäfunktioksi

$$F_X(x) = \Pr(X \le x) = 1 - \Pr(X > x) = 1 - \exp(-\lambda x)$$

Satunnaismuuttujan X *tiheysfunktio* saadaan derivoimalla kertymäfunktio $F_X(x)$ argumenttinsa x suhteen:

$$f_X(x) = \frac{d}{dx} F_X(x) = \lambda \exp(-\lambda x)$$

Siten

$$X \sim \text{Exp}(\lambda)$$

■

Eksponenttijakauman unohtamisominaisuus

Olkoon

$$X \sim \text{Exp}(\lambda)$$

Tällöin

$$\Pr(X \ge a + b \mid X \ge a) = \Pr(X \ge b)$$

jossa $a \ge 0$, $b \ge 0$.

Perustelu:

Todetaan ensin, että

$$\Pr(X \ge c) = 1 - \Pr(X \le c) = 1 - F(c) = \exp(-\lambda c)$$

jossa $c \geq 0$ ja F on eksponenttijakauman *kertymäfunktio*.

Koska

$$\{X \geq a+b\} \subset \{X \geq a\}$$

niin

$$\Pr(X \geq a+b \text{ ja } X \geq a) = \Pr(X \geq a+b)$$

Siten

$$\begin{aligned}\Pr(X \geq a+b \mid X \geq a) &= \frac{\Pr(X \geq a+b \text{ ja } X \geq a)}{\Pr(X \geq a)} \\ &= \frac{\Pr(X \geq a+b)}{\Pr(X \geq a)} \\ &= \frac{\exp(-\lambda(a+b))}{\exp(-\lambda a)} \\ &= \exp(-\lambda b) \\ &= 1 - F(b) \\ &= 1 - \Pr(X \leq b) \\ &= \Pr(X \geq b)\end{aligned}$$

■

Siten eksponenttijakaumalla ns. on *unohtamisominaisuus*: Se, että tapahtuman sattumista on jouduttu odottamaan ajan *a*, *ei vaikuta* todennäköisyyteen joutua odottamaan ajan *b lisää*. Tällaista unohtamisominaisuutta kutsutaan *stokastisten prosessien* teoriassa **Markov-ominaisuudeksi**.

Samanlainen unohtamisominaisuus on **geometrisella jakaumalla**, jota voidaan pitää eksponenttijakauman *diskreettinä* vastineena; lisätietoja geometrisesta jakaumasta: ks. lukua **Diskreettejä jakaumia**.

Eksponenttijakauman todennäköisyydet

Olkoon

$$X \sim \text{Exp}(\lambda)$$

ja olkoon

$$[c, d] \subset [0, +\infty)$$

jokin jakauman määrittelyvälin $[0, +\infty)$ *osaväli*. *Välin* $[c, d]$ *todennäköisyys* saadaan integroimalla eksponenttijakauman $\text{Exp}(\lambda)$ tiheysfunktio

$$f(x) = \lambda\exp(-\lambda x) \,,\ \lambda > 0 \,,\ x \geq 0$$

välillä $[c, d]$:

$$\Pr(c \leq X \leq d) = \int_c^d f(x)dx = \int_c^d \lambda e^{-\lambda x}dx = \left[-e^{-\lambda x}\right]_c^d = e^{-\lambda c} - e^{-\lambda d}$$

Kaikkien muiden eksponenttijakaumaan $\text{Exp}(\lambda)$ liittyvien tapahtumien todennäköisyydet saadaan välin $[0, +\infty)$ osavälien todennäköisyyksistä *todennäköisyyslaskennan laskusääntöjen* avulla.

17.3. Normaalijakauma

Normaalijakauma on sekä *teoreettisen* että *soveltavan tilastotieteen tärkein jakauma*. Tämä johtuu seuraavista seikoista:

(i) **Empiirisen kokemuksen mukaan monia satunnaisilmiöitä koskevat *havainnot* noudattavat ainakin approksimatiivisesti normaalijakaumaa**.

(ii) **Monet *satunnaismuuttujat* noudattavat approksimatiivisesti normaalijakaumaa**.

(iii) **Monet tärkeät *todennäköisyysjakaumat* saadaan sopivilla muunnoksilla normaalijakaumasta**.

(iv) ***Keskeisen raja-arvolauseen* mukaan usean satunnaismuuttujan *summat* noudattavat ainakin approksimatiivisesti normaalijakaumaa**, **jos summassa on riittävästi tekijöitä**.

Olkoon *jatkuvan satunnaismuuttujan X tiheysfunktio*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, -\infty < \mu < +\infty\,, \sigma > 0\,, -\infty < x < +\infty$$

Tällöin satunnaismuuttuja X noudattaa **normaalijakaumaa** parametrein μ ja σ^2 ja merkitsemme:

$$X \quad \mathrm{N}(\mu, \sigma^2)$$

Funktio $f(x)$ kelpaa tiheysfunktioksi, koska

$$f(x) > 0\,, -\infty < x < +\infty$$

ja

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Perustelu:

Sen perusteleminen, että

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

on siinä mielessä hankalaa, että *normaalijakauman tiheysfunktio kuuluu sellaisten funktioiden joukkoon, joiden integraalifunktiota ei voida esittää suljetussa muodossa* (so. alkeisfunktioiden avulla). Perusteleminen onnistuu kuitenkin alla käytettävän ”tempun” avulla.

Tarkastellaan yksinkertaisuuden vuoksi ns. *standardoitua normaalijakaumaa*, jossa

$$\mu = 0\,, \sigma^2 = 1$$

Tämä ei ole rajoittava oletus, koska *jokainen* yleistä normaalijakaumaa N(μ, σ^2) noudattava satunnaismuuttuja saadaan lineaarimuunnoksella standardoitua normaalijakaumaa N(0,1) noudattavasta satunnaismuuttujasta; ks. kohtaa **Normaalijakautuneen satunnaismuuttujan lineaarimuunnoksen jakauma**.

Jos

$$\mu = 0 \,, \sigma^2 = 1$$

niin normaalijakauman tiheysfunktio on muotoa

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Koska $f(x)$ on *symmetrinen*, väite

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

tulee todistetuksi, jos näytämme, että

$$\int_0^{\infty} e^{-x^2/2} dx = \frac{\sqrt{2\pi}}{2} = \sqrt{\frac{\pi}{2}}$$

Tarkastellaan integraalia

$$\left(\int_0^{\infty} e^{-x^2/2} dx \right)^2 = \left(\int_0^{\infty} e^{-u^2/2} du \right) \left(\int_0^{\infty} e^{-v^2/2} dv \right) = \int_0^{\infty} \int_0^{\infty} e^{-(u^2+v^2)/2} du\, dv$$

Siirrytään tässä integraalissa *napakoordinaatteihin* muunnoksella

$$u = r\cos(\theta)$$
$$v = r\sin(\theta)$$

Tällöin

$$u^2 + v^2 = r^2\cos^2(\theta) + r^2\cos^2(\theta) = r^2[\cos^2(\theta) + \cos^2(\theta)] = r^2$$

$$du\, dv = r\, d\theta\, dr$$

ja integroimisalueeksi tulee

$$0 < r < \infty$$
$$0 < \theta < \pi/2$$

Siten

$$\int_0^{\infty} \int_0^{\infty} e^{-(u^2+v^2)/2} dudv = \int_0^{\infty} \int_0^{\pi/2} re^{-r^2/2} d\theta dr = \frac{\pi}{2} \int_0^{\infty} re^{-r^2/2} dr = \frac{\pi}{2} \left[-e^{-r^2/2} \right]_0^{\infty} = \frac{\pi}{2}$$

joten olemme todistaneet, että

$$\left(\int_0^{\infty} e^{-x^2/2} dx \right)^2 = \frac{\pi}{2}$$

josta väite seuraa.

■

Normaalijakaumaa kutsutaan kehittäjänsä mukaan usein **Gaussin jakaumaksi** ja sen tiheysfunktion kuvaajaa kutsutaan tavallisesti joko **Gaussin käyräksi** tai **kellokäyräksi** (engl. *bell curve*).

Normaalijakauman tunnusluvut

Olkoon

$$X \sim \mathrm{N}(\mu, \sigma^2)$$

Odotusarvo:

$$\mathrm{E}(X) = \mu$$

Varianssi:

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

Standardipoikkeama:

$$\mathrm{D}(X) = \sigma$$

Perustelu:

Johdamme tässä normaalijakauman odotusarvon käyttäen jatkuvan jakauman odotusarvon määritelmää. Jakauman odotusarvo ja varianssi johdetaan jakauman *momenttiemäfunktion* avulla kohdassa **Normaalijakauman momenttiemäfunktio ja jakauman tunnusluvut**.

Suoraan jatkuvan jakauman odotusarvon määritelmästä saadaan, että

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Sijoituksella

$$z = \frac{x-\mu}{\sigma}$$

saadaan:

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} (\mu + \sigma z) e^{-\frac{1}{2}z^2} dz = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \mu e^{-\frac{1}{2}z^2} dz + \int_{-\infty}^{+\infty} \frac{\sigma}{\sqrt{2\pi}} z e^{-\frac{1}{2}z^2} dz$$

Nyt

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \mu e^{-\frac{1}{2}z^2} dz + \int_{-\infty}^{+\infty} \frac{\sigma}{\sqrt{2\pi}} z e^{-\frac{1}{2}z^2} dz = \mu$$

Tämä seuraa siitä, että

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \mu e^{-\frac{1}{2}z^2} dz = \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \mu \cdot 1 = \mu$$

koska integroitava on ns. *standardoidun normaalijakauman* N(0,1) tiheysfunktio ja

$$\int_{-\infty}^{+\infty} \frac{\sigma}{\sqrt{2\pi}} z e^{-\frac{1}{2}z^2} dz = 0$$

koska integroitava on muuttujan *z pariton funktio*.

■

Normaalijakauman tiheysfunktio

Kuva oikealla esittää normaalijakauman

$$\mathrm{N}(\mu,\sigma^2)$$

tiheysfunktiota

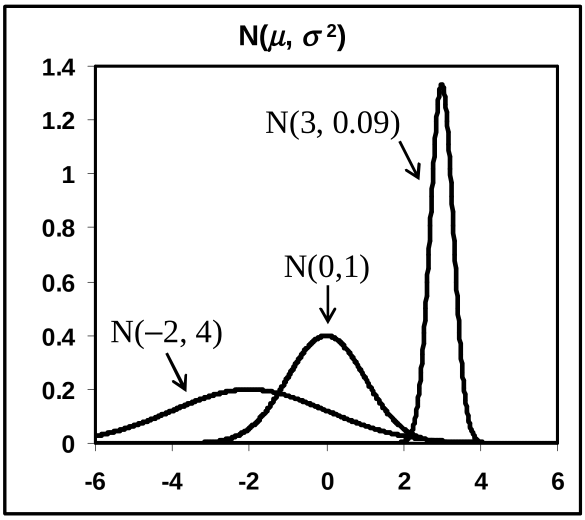
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

välillä [–6,+6] seuraavissa tapauksissa:

(i) $\mu=-2\ ,\sigma^2=4$

(ii) $\mu=0\ \ ,\sigma^2=1$

(iii) $\mu=+3\ ,\sigma^2=0.09$



Normaalijakauman *odotusarvo*:

$$\mathrm{E}(X)=\mu$$

Normaalijakauman *varianssi*:

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

Normaalijakauman tiheysfunktion ominaisuudet

Kaikille normaalijakaumille pätee:

(i) Normaalijakauman tiheysfunktio $f(x)$ on kaikkialla *positiivinen*:

$$f(x)>0,\ -\infty<x<+\infty$$

(ii) Tiheysfunktio on *yksihuippuinen*.

(iii) Tiheysfunktiolla on *maksimi* pisteessä

$$\mathrm{E}(X)=\mu$$

(iv) Tiheysfunktio on *symmetrinen* pisteen $x=\mu$ suhteen:

$$f(\mu-x)=f(\mu+x)\ ,\ -\infty<x<+\infty$$

(v) Tiheysfunktiolla on *käännepisteet* pisteissä

$$\mu-\sigma \text{ ja } \mu+\sigma$$

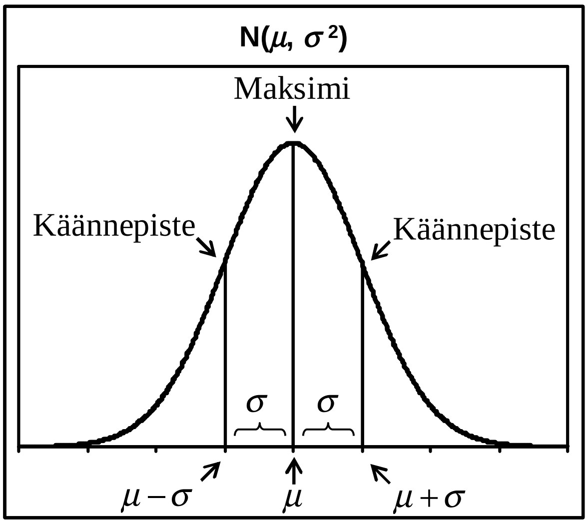
(vi) Tiheysfunktio on *kupera ylöspäin* välin

$$[\mu-\sigma,\mu+\sigma]$$

sisäpuolella ja *kupera alaspäin* välin

$$[\mu-\sigma,\mu+\sigma]$$

ulkopuolella.



(vii) Kaikkien normaalijakaumien tiheysfunktiot ovat *samanmuotoisia*, jos ne piirretään *standardoiduissa yksiköissä*

$$z = \frac{x-\mu}{\sigma}$$

68-95-99.7-sääntö

Kaikille normaalijakaumille pätee:

(i) Noin 68% jakauman todennäköisyysmassasta on välillä

$$[\mu - \sigma, \mu + \sigma]$$

(ii) Noin 95% jakauman todennäköisyysmassasta on välillä

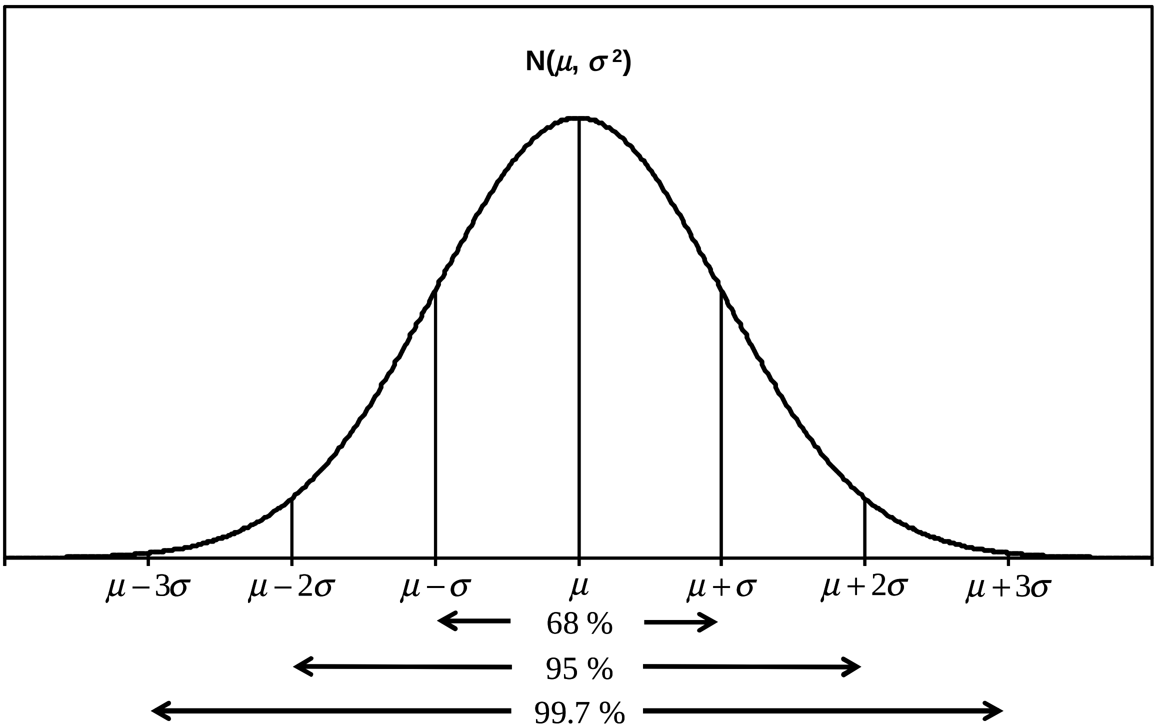
$$[\mu - 2\sigma, \mu + 2\sigma]$$

(iii) Noin 99.7% jakauman todennäköisyysmassasta on välillä

$$[\mu - 3\sigma, \mu + 3\sigma]$$

68-95-99.7-sääntö nähdään helposti oikeaksi vetoamalla normaalijakauman tiheysfunktion ominaisuuteen (vii) ja *standardoidun normaalijakauman taulukoihin*; lisätietoja normaalijakauman taulukoiden käytöstä: ks. kohtia **Todennäköisyyksien määrääminen standardoidusta normaalijakaumasta** ja **Todennäköisyyksien määrääminen mielivaltaisesta normaalijakaumasta**.

Kuva alla havainnollistaa 68-95-99.7-sääntöä.



Normaalijakauman kertymäfunktio

Olkoon

$$X \sim \mathrm{N}(\mu,\ \sigma^2)$$

Satunnaismuuttujan X **kertymäfunktio** on

$$F(x) = \Pr(X \le x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

Huomautuksia:

- **Normaalijakauman kertymäfunktiolle ei voida antaa eksplisiittistä lauseketta**, koska normaalijakauman tiheysfunktio kuuluu sellaisten funktioiden joukkoon, joiden integraalifunktiota ei voida esittää suljetussa muodossa (so. alkeisfunktioiden avulla).
- Normaalijakauman kertymäfunktion arvojen määräämiseen on käytettävä **numeerista integrointia** ja siksi tilastotieteen oppikirjoihin on tavallisesti liitetty ns. standardoidun normaalijakauman kertymäfunktion arvoja sisältävä **taulukko**.
- Standardoidun normaalijakauman kertymäfunktion arvoja sisältävän taulukon avulla voidaan määrätä todennäköisyydet reaaliakselin väleille mielivaltaisesta normaalijakaumasta.

Lisätietoja: ks. kohtia **Todennäköisyyksien määrääminen standardoidusta normaali-jakaumasta** ja **Todennäköisyyksien määrääminen mielivaltaisesta normaalijakaumasta**.

Normaalijakauman momenttiemäfunktio

Normaalijakauman

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, -\infty < \mu < +\infty, \sigma > 0, -\infty < x < +\infty$$

momenttiemäfunktio (ks. lukua **Momenttiemäfunktio ja karakteristinen funktio**) on

$$m_X(t) = \mathrm{E}(e^{tX}) = \exp(\mu t + \tfrac{1}{2}\sigma^2 t^2)$$

Perustelu:

Jatkuvan jakauman momenttiemäfunktion määritelmän mukaan

$$\begin{aligned}
m_X(t) = \mathrm{E}(e^{tX}) &= \int_{-\infty}^{+\infty} e^{tx} f(x)dx \\
&= \int_{-\infty}^{+\infty} \exp(tx)\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}dx \\
&= \exp(\mu t + \tfrac{1}{2}\sigma^2 t^2)\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}[x - 2(\mu+\sigma^2 t)]^2\right\}dx \\
&= \exp(\mu t + \tfrac{1}{2}\sigma^2 t^2)
\end{aligned}$$

■

Normaalijakauman momenttiemäfunktio ja jakauman tunnusluvut

Johdetaan normaalijakauman $\mathrm{N}(\mu, \sigma^2)$ *odotusarvo* ja *varianssi* jakauman momenttiemäfunktion avulla.

Odotusarvo:

$$\mathrm{E}(X) = \mu$$

Varianssi:

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

Perustelu:

Normaalijakauman $\mathrm{N}(\mu, \sigma^2)$ *momenttiemäfunktio* on

$$m_X(t) = \exp(\mu t + \tfrac{1}{2}\sigma^2 t^2)$$

Momenttiemäfunktion 1. derivaatta pisteessä $t = 0$:

$$\left.\frac{dm_X(t)}{dt}\right|_{t=0} = \left. e^{\mu t + \frac{1}{2}\sigma^2 t^2}(\mu + \sigma^2 t)\right|_{t=0} = \mu$$

Momenttiemäfunktion 2. derivaatta pisteessä $t = 0$:

$$\left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \left.\left[e^{\mu t + \frac{1}{2}\sigma^2 t^2}(\mu + \sigma^2 t)^2 + e^{\mu t + \frac{1}{2}\sigma^2 t^2}\sigma^2 \right]\right|_{t=0} = \mu^2 + \sigma^2$$

Siten normaalijakauman $\mathrm{N}(\mu, \sigma^2)$ *odotusarvo* μ_X, 2. *momentti* α_2 ja *varianssi* σ_X^2 saadaan seuraavilla kaavoilla:

$$\mu_X = \mathrm{E}(X) = \alpha_1 = \left.\frac{dm_X(t)}{dt}\right|_{t=0} = \mu$$

$$\alpha_2 = \mathrm{E}(X^2) = \left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \mu^2 + \sigma^2$$

$$\sigma_X^2 = \mathrm{Var}(X) = \alpha_2 - \alpha_1^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

■

Standardoitu normaalijakauma

Jos

$$Z \quad \mathrm{N}(0,1)$$

jolloin siis

$$\mathrm{E}(Z) = \mu = 0$$

$$\mathrm{Var}(Z) = \mathrm{D}^2(Z) = \sigma^2 = 1$$

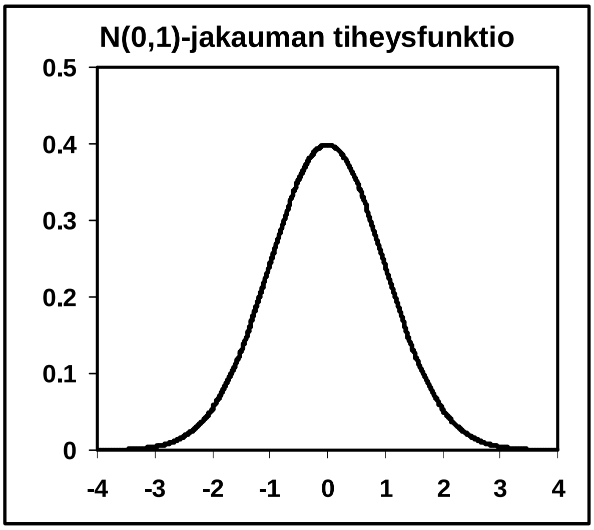
niin sanomme, että satunnaismuuttuja X noudattaa **standardoitua normaalijakaumaa**.

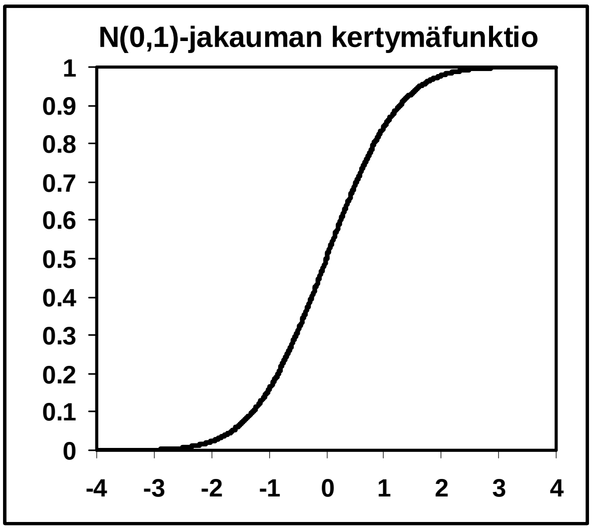
Alla olevista kuvista vasemmanpuoleinen esittää standardoidun normaalijakauman $\mathrm{N}(0,1)$ *tiheysfunktiota*

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

ja kuvista oikeanpuoleinen esittää standardoidun normaalijakauman N(0,1) *kertymäfunktiota*

$$\Phi(z) = \Pr(Z \le z) = \int_{-\infty}^{z} f(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2} dt$$





Kuten edellä on todettu, **normaalijakauman kertymäfunktiolle ei voida antaa eksplisiittistä lauseketta**, koska normaalijakauman tiheysfunktio kuuluu sellaisten funktioiden joukkoon, joiden integraalifunktiota ei voida esittää suljetussa muodossa (so. alkeisfunktioiden avulla).

Normaalijakautuneen satunnaismuuttujan lineaarimuunnoksen jakauma

Olkoon

$$X \quad \mathrm{N}(\mu, \sigma^2)$$

Määritellään satunnaismuuttuja

$$Y = a + bX$$

jossa a ja b ovat ei-satunnaisia *vakioita*. Tällöin satunnaismuuttuja Y on **normaalinen**:

$$Y \sim \mathrm{N}(a + b\mu, b^2\sigma^2)$$

Perustelu:

Ks. lukua **Satunnaismuuttujien muunnokset ja niiden jakaumat**.

Tästä tuloksesta seuraa erikoistapauksena seuraava tulos: Jos

$$X \quad \mathrm{N}(0,1)$$

ja

$$Y = \mu + \sigma X$$

jossa μ ja σ ovat ei-satunnaisia *vakioita*, niin

$$Y \sim \mathrm{N}(\mu, \sigma^2)$$

Siten **kaikki normaalijakautuneet satunnaismuuttujat saadaan lineaarimuunnoksella standardoitua normaalijakaumaa** N(0,1) **noudattavista satunnaismuuttujista**.

Huomautus:

- Kaikille satunnaismuuttujille (joilla on odotusarvo ja varianssi) pätee seuraava:

Oletetaan, että

$$\mathrm{E}(X) = \mu$$
$$\mathrm{Var}(X) = \sigma^2$$

ja olkoon

$$Y = a + bX$$

Tällöin

$$\mathrm{E}(Y) = a + b\mu$$
$$\mathrm{Var}(Y) = b^2\sigma^2$$

Ks. lukua **Jakaumien tunnusluvut**.

Edellä esitetyn tuloksen merkitys on siinä, että sen mukaan *satunnaismuuttujan X normaalisuudesta seuraa sen lineaarimuunnoksen* $Y = a + bX$ *normaalisuus*.

Standardointi

Olkoon

$$X \quad \mathrm{N}(\mu, \sigma^2)$$

jolloin siis

$$\mathrm{E}(X) = \mu$$
$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma^2$$

Tällöin **standardoitu satunnaismuuttuja**

$$Z = \frac{X - \mathrm{E}(X)}{\mathrm{D}(X)} = \frac{X - \mu}{\sigma}$$

noudattaa *standardoitua normaalijakaumaa*:

$$Z \quad \mathrm{N}(0,1)$$

Perustelu:

Tulos seuraa edellä esitetystä normaalijakautuneen satunnaismuuttujan lineaarimuunnoksen jakaumaa koskevasta tuloksesta, koska

$$Z = \frac{X - \mu}{\sigma} = -\frac{\mu}{\sigma} + \frac{1}{\sigma} X$$

■

Todennäköisyyksien määrääminen standardoidusta normaalijakaumasta

Todennäköisyydet *standardoidusta normaalijakaumasta* N(0,1) voidaan määrätä jakauman *kertymäfunktion* avulla.

Oletetaan, että

$$Z \quad N(0,1)$$

ja olkoon

$$\Phi(z) = \Pr(Z \le z)$$

satunnaismuuttujan Z *kertymäfunktio*.

Kaikkien standardoituun normaalijakaumaan liittyvien tapahtumien todennäköisyydet saadaan todennäköisyyksistä

$$\Phi(z) = \Pr(Z \le z)$$

todennäköisyyslaskennan laskusääntöjen avulla. Perusoperaationa on tällöin muotoa

$$[a, b]$$

olevien reaaliakselin *välien* todennäköisyyksien määrääminen. Jatkuvan jakauman kertymäfunktioiden yleisten ominaisuuksien perusteella

$$\Pr(a \le Z \le b) = \Phi(b) - \Phi(a)$$

Kuten edellä on todettu, **normaalijakauman kertymäfunktiolle ei voida antaa eksplisiittistä lauseketta**, koska normaalijakauman tiheysfunktio kuuluu sellaisten funktioiden joukkoon, joiden integraalifunktiota ei voida esittää suljetussa muodossa (so. alkeisfunktioiden avulla). Siksi todennäköisyyksien määräämisessä standardoidusta normaalijakaumasta on turvauduttava joko **taulukoihin** tai **tietokoneohjelmiin**; ks. seuraavia kappaleita.

Todennäköisyyksien määrääminen standardoidusta normaalijakaumasta ja normaalijakauman taulukot

Standardoidun normaalijakauman *taulukot* sisältävät standardoidun normaalijakauman N(0,1) *kertymäfunktion* $\Phi(z)$ *arvoja* taulukoituna usealle eri argumentin z arvolle.

Tähän monisteeseen liittyvässä taulukkokokoelmassa on taulukoituna standardoidun normaalijakauman N(0,1) kertymäfunktion $\Phi(z)$ arvoja, kun z vaihtelee välillä [–3.59, +3.59] askelin, joiden pituus on 0.01:

$$z = -3.59\ (0.01)\ +3.59$$

Taulukot mahdollistavat seuraavien tehtävien ratkaisemisen:

(i) Määrää todennäköisyys

$$\Pr(Z \le z) = \Phi(z)$$

kun z on annettu.

(ii) Määrää z, kun todennäköisyys

$$\Pr(Z \le z) = \Phi(z)$$

on annettu.

Monissa standardoidun normaalijakauman taulukoissa on taulukoituna todennäköisyyksiä

$$\Pr(Z \le z) = \Phi(z)$$

vain, kun $z \ge 0$. Tällöin todennäköisyydet

$$\Pr(Z \le -z) = \Phi(-z)$$

saadaan nojaamalla standardoidun normaalijakauman tiheysfunktion *symmetrisyyteen* pisteen

$$z = 0$$

suhteen:

$$\Phi(-z) = \Pr(Z \le -z) = 1 - \Pr(Z \ge -z) = 1 - \Pr(Z \le z) = 1 - \Phi(z)$$

Esimerkki 1. Todennäköisyyksien määrääminen standardoidun normaalijakauman taulukoista.

Olkoon

$$Z \quad \mathrm{N}(0,1)$$

Tällöin

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

on satunnaismuuttujan Z *tiheysfunktio* ja

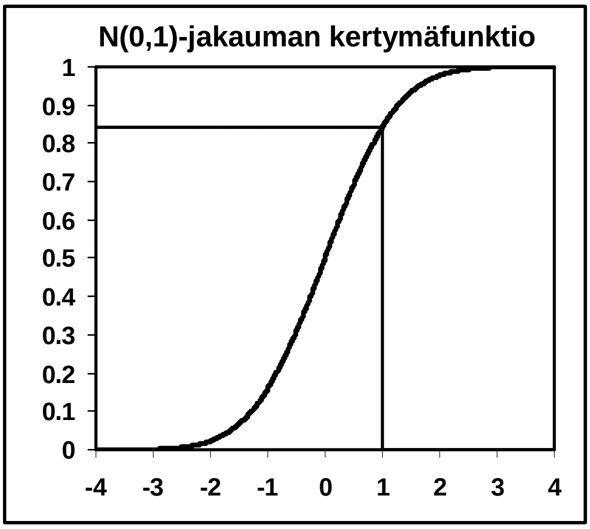
$$\Phi(z) = \Pr(Z \le z) = \int_{-\infty}^{z} f(t)dt$$

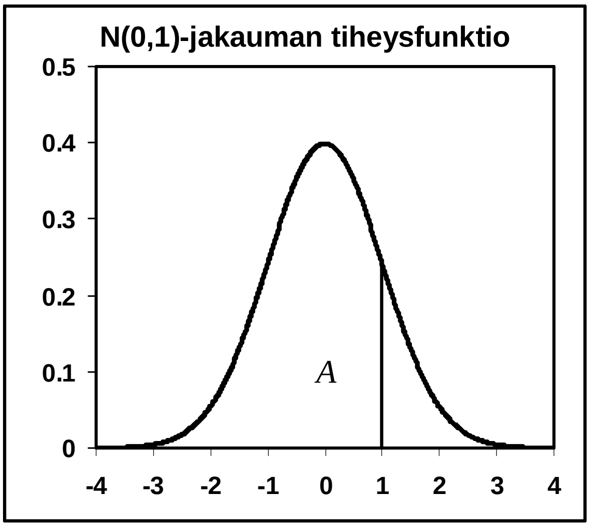
on satunnaismuuttujan Z *kertymäfunktio*.

Standardoidun normaalijakauman N(0,1) taulukoista saadaan:

$$\Phi(1) = \Pr(Z \le 1) = 0.8413$$

Ks. vasemmanpuoleista kuvaa alla.





Oikeanpuoleinen kuva yllä havainnollistaa sitä, että

$$\text{Alueen } A \text{ pinta-ala} = \int_{-\infty}^{1} f_Z(z)dz = \Pr(Z \le 1) = \Phi(1) = 0.8413$$

Todennäköisyyksien määrääminen mielivaltaisesta normaalijakaumasta

Kuten edellä on todettu, *kaikki* normaalijakaumat $\mathrm{N}(\mu,\sigma^2)$ ovat *samanmuotoisia standardoiduissa yksiköissä*

$$Z = \frac{X-\mu}{\sigma}$$

Tämä merkitsee sitä, että todennäköisyydet *mielivaltaisesta normaalijakaumasta* $\mathrm{N}(\mu,\sigma^2)$ voidaan määrätä *standardoidun normaalijakauman* N(0,1) todennäköisyyksien avulla.

Oletetaan, että

$$X \quad \mathrm{N}(\mu,\sigma^2)$$

jolloin

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

Tällöin siis

$$\Pr(a \le X \le b) = \Pr\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) = \Pr\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$$

Esimerkki 1: Todennäköisyyksien määrääminen mielivaltaisesta normaalijakaumasta.

Olkoon

$$X \quad \mathrm{N}(2, 1/4)$$

jolloin

$$\mathrm{E}(X) = \mu = 2$$

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma^2 = \frac{1}{4}$$

ja

$$\mathrm{D}(X) = \sigma = \frac{1}{2}$$

Haluamme määrätä todennäköisyyden

$$\Pr(1.5 \le X \le 3)$$

Käytämme tehtävän ratkaisemisessa *standardoidun normaalijakauman* N(0,1) *taulukoita*.

Olkoon

$$X \quad \mathrm{N}(\mu,\sigma^2)$$

jossa

$$\mu = 2$$
$$\sigma^2 = 1/4$$

Tällöin *standardoitu satunnaismuuttuja*

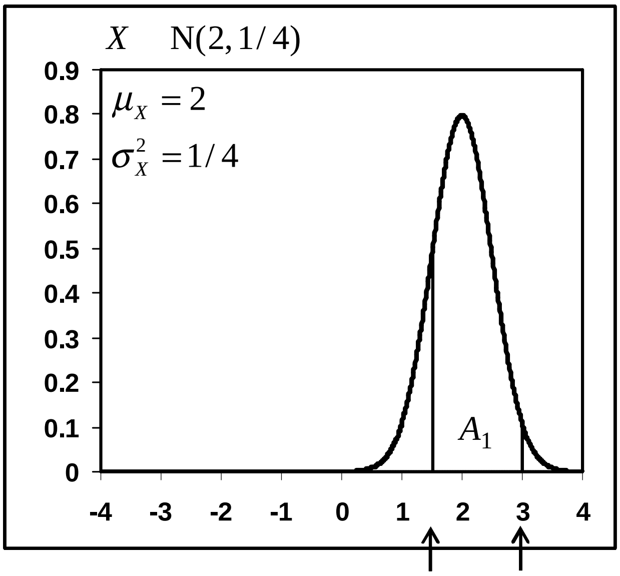
$$Z = \frac{X-\mu}{\sigma} = \frac{X-2}{1/2} \quad \mathrm{N}(0,1)$$

Standardoidun normaalijakauman N(0,1) taulukoista saadaan:

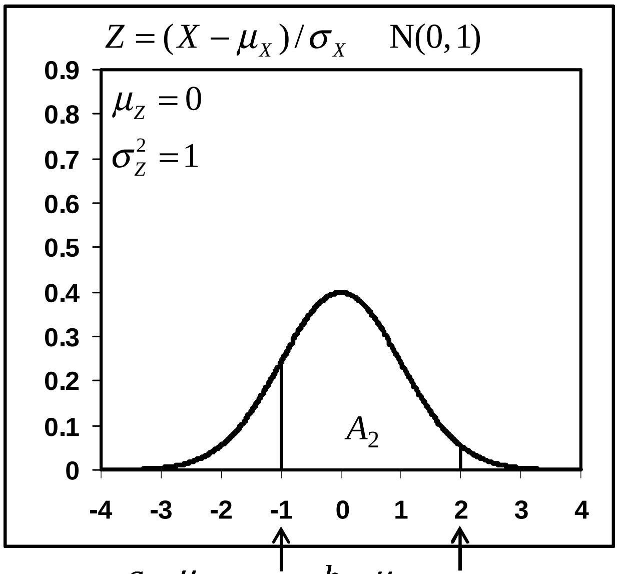
$$\begin{aligned}
\Pr(1.5 \le X \le 3) &= \Pr\left(\frac{1.5-2}{1/2} \le \frac{X-2}{1/2} \le \frac{3-2}{1/2}\right) \\
&= \Pr(-1 \le Z \le 2) \qquad\qquad | \; Z \quad \mathrm{N}(0,1) \\
&= \Pr(Z \le 2) - \Pr(Z \le -1) \\
&= 0.9772 - 0.1587 \\
&= 0.8185
\end{aligned}$$

Ks. alla olevia kuvia, joissa siis

$$\begin{aligned}
\text{Alueen } A_1 \text{ pinta-ala} &= \Pr(1.5 \le X \le 3) = \Pr(-1 \le Z \le 2) = \text{Alueen } A_2 \text{ pinta-ala} \\
&= 0.8185
\end{aligned}$$



$a = 1.5 \quad b = 3$



$$\frac{a-\mu_X}{\sigma_X} = -1 \qquad \frac{b-\mu_X}{\sigma_X} = 2$$

Todennäköisyyksien määrääminen normaalijakaumasta ja tietokoneohjelmat

Olkoon

$$X \quad \mathrm{N}(\mu, \sigma^2)$$

Monet *tietokoneohjelmat* mahdollistavat seuraavien tehtävien ratkaisemisen mielivaltaisille parametrien μ ja σ^2 arvoille:

(i) Määrää todennäköisyys

$$\Pr(X \le x)$$

kun x on annettu.

(ii) Määrää x, kun todennäköisyys

$$\Pr(X \le x)$$

on annettu.

Normaalijakaumaa noudattavien satunnaismuuttujien summan jakauma

Olkoot

$$X_1,\ X_2,\ \ldots\ ,\ X_n$$

riippumattomia satunnaismuuttujia, jotka noudattavat *normaalijakaumaa* parametrein (μ_1,σ_1^2), (μ_2,σ_2^2), … , (μ_n,σ_n^2):

$$X_1, X_2, \ldots, X_n \perp$$
$$X_i \sim \mathrm{N}(\mu_i,\sigma_i^2),\ i=1,2,\ldots,n$$

Tällöin satunnaismuuttujien $X_1,\ X_2,\ \ldots\ ,\ X_n$ *summa*

$$Y = X_1 + X_2 + \cdots + X_n$$

noudattaa *normaalijakaumaa* parametrein $\mu = \mu_1 + \mu_2 + \cdots + \mu_n$ ja $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$:

$$Y \sim \mathrm{N}(\mu,\sigma^2)$$

Perustelu:

Todistus perustuu siihen, että *riippumattomien satunnaismuuttujien summan momenttiemäfunktio ko. satunnaismuuttujien momenttiemäfunktioiden tulo*.

Oletetaan, että

$$X_1, X_2, \ldots, X_n \perp$$
$$X_i \sim \mathrm{N}(\mu_i,\sigma_i^2),\ i=1,2,\ldots,n$$

Satunnaismuuttujien $X_1,\ X_2,\ \ldots\ ,\ X_n$ *momenttiemäfunktiot* ovat muotoa

$$m_i(t) = \exp(\mu_i t + \tfrac{1}{2}\sigma_i^2 t^2),\ i=1,2,\ldots,n$$

Muodostetaan satunnaismuuttujien $X_1,\ X_2,\ \ldots\ ,\ X_n$ *summa*:

$$Y = X_1 + X_2 + \cdots + X_n$$

Summamuuttujan Y momenttiemäfunktio on

$$\begin{aligned} m_Y(t) &= m_1(t)m_2(t)\cdots m_n(t) \\ &= \exp(\mu_1 t + \tfrac{1}{2}\sigma_1^2 t^2)\exp(\mu_2 t + \tfrac{1}{2}\sigma_2^2 t^2)\cdots\exp(\mu_n t + \tfrac{1}{2}\sigma_n^2 t^2) \\ &= \exp((\mu_1 + \mu_2 + \cdots + \mu_n)t + \tfrac{1}{2}(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)t^2) \\ &= \exp(\mu t + \tfrac{1}{2}\sigma^2 t^2) \end{aligned}$$

jossa

$$\mu = \mu_1 + \mu_2 + \cdots + \mu_n$$
$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$$

Funktio $m_Y(t)$ on *normaalijakauman*

$$\mathrm{N}(\mu,\sigma^2)$$

momenttiemäfunktio. Lisäksi funktio $m_Y(t)$ on selvästi jatkuva pisteen

$$t = 0$$

ympäristössä. Koska momenttiemäfunktio $m_Y(t)$ on tällöin *yksikäsitteinen*, summamuuttuja Y noudattaa *normaalijakaumaa* parametrein

$$\mu = \mu_1 + \mu_2 + \cdots + \mu_n$$

ja

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$$

eli

$$Y = X_1 + X_2 + \cdots + X_n \sim \mathrm{N}(\mu,\sigma^2)$$

■

Normaalijakaumaa noudattavien satunnaismuuttujien lineaarikombinaation jakauma

Yhdistämällä riippumattomien normaalijakaumaa noudattavien satunnaismuuttujien summan jakaumaa koskeva tulos ja normaalijakauman lineaarimuunnoksen jakaumaa koskeva tulos saadaan seuraava, *riippumattomien normaalijakautuneiden satunnaismuuttujien lineaarikombinaation jakaumaa* koskeva tulos:

Olkoot

$$X_1,\ X_2,\ \ldots,\ X_n$$

riippumattomia satunnaismuuttujia, jotka noudattavat *normaalijakaumaa* parametrein (μ_1,σ_1^2), (μ_2,σ_2^2), … , (μ_n,σ_n^2):

$$X_1, X_2, \ldots, X_n \perp$$
$$X_i \sim \mathrm{N}(\mu_i,\sigma_i^2),\ i = 1, 2, \ldots, n$$

Olkoot lisäksi

$$a_1,\ a_2,\ \ldots,\ a_n$$

ei satunnaisia *vakioita*. Tällöin satunnaismuuttujien $X_1,\ X_2,\ \ldots,\ X_n$ *lineaarikombinaatio*

$$Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

noudattaa *normaalijakaumaa* parametrein

$$\mu = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$$

ja

$$\sigma^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2$$

eli

$$Y \sim \mathrm{N}(\mu,\sigma^2)$$

Samaa normaalijakaumaa noudattavien satunnaismuuttujien summan jakauma

Riippumattomien normaalijakaumaa noudattavien satunnaismuuttujien summan jakaumaa koskevan tuloksen seurauksena saadaan *riippumattomien, samaa normaalijakaumaa noudattavien satunnaismuuttujien summan jakaumaa* koskeva tulos:

Olkoot

$$X_1,\ X_2,\ \ldots\ ,\ X_n$$

riippumattomia satunnaismuuttujia, jotka noudattavat *samaa normaalijakaumaa* parametrein μ ja σ^2:

$$X_1, X_2, \ldots, X_n \perp$$
$$X_i \sim \mathrm{N}(\mu, \sigma^2)\ ,\ i = 1, 2, \ldots, n$$

Tällöin satunnaismuuttujien $X_1,\ X_2,\ \ldots\ ,\ X_n$ *summa*

$$Y = X_1 + X_2 + \cdots + X_n$$

noudattaa *normaalijakaumaa* parametrein $n\mu$ ja $n\sigma^2$:

$$Y \sim \mathrm{N}(n\mu, n\sigma^2)$$

Normaalijakaumaa noudattavien satunnaismuuttujien aritmeettisen keskiarvon jakauma

Olkoot

$$X_1,\ X_2,\ \ldots\ ,\ X_n$$

riippumattomia satunnaismuuttujia, jotka noudattavat *samaa normaalijakaumaa* parametrein μ ja σ^2:

$$X_1, X_2, \ldots, X_n \perp$$
$$X_i \sim \mathrm{N}(\mu, \sigma^2)\ ,\ i = 1, 2, \ldots, n$$

Olkoon

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

satunnaismuuttujien $X_1,\ X_2,\ \ldots\ ,\ X_n$ *aritmeettinen keskiarvo*. Tällöin *riippumattomien, samaa normaalijakaumaa noudattavien satunnaismuuttujien summan jakaumaa* ja *normaalijakautuneen satunnaismuuttujan lineaarimuunnoksen jakaumaa* koskevista tuloksista seuraa, että

$$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

17.4. Keskeinen raja-arvolause

Se, että **normaalijakauma** (ks. kappaletta **Normaalijakauma**) **on sekä *teoreettisen* että *soveltavan tilastotieteen* tärkein jakauma** perustuu hyvin pitkälti siihen *teoreettiseen* ja *empiiriseen* havaintoon, että moniin satunnaisilmiöihin liittyvät satunnaismuuttujat noudattavat ainakin *approksimatiivisesti* normaalijakaumaa.

Mikä on selitys tälle havainnolle? Selityksenä on **keskeinen raja-arvolause**.

Tarkastellaan jonoa

$$X_i\ ,\ i = 1, 2, \ldots , n$$

riippumattomia ja *samaa normaalijakaumaa noudattavia satunnaismuuttujia*:

$$X_1, X_2, \ldots, X_n \perp$$
$$X_i \sim \mathrm{N}(\mu, \sigma^2)\ ,\ i = 1, 2, \ldots, n$$

Edellä esitettyjen tulosten mukaan *satunnaismuuttujien* X_i , $i = 1, 2, \ldots , n$ *summa*

$$Y_n = \sum_{i=1}^{n} X_i$$

noudattaa normaalijakaumaa:

$$Y_n \sim \mathrm{N}(n\mu, n\sigma^2)$$

Kysymys: Mitä voidaan sanoa *riippumattomien, samaa jakaumaa noudattavien satunnaismuuttujien summan jakaumasta*, jos ko. satunnaismuuttujat *eivät noudata normaalijakaumaa*?

Vastaus: *Ei-normaalisten* satunnaismuuttujien summa *ei yleensä ole* normaalinen.

Mutta: Jos yhteenlaskettavia on ”tarpeeksi paljon”, *satunnaismuuttujien summa on* (hyvin yleisin ehdoin) **approksimatiivisesti normaalinen**. Tämä on **keskeisen raja-arvolauseen** olennainen sisältö.

Koska monia satunnaismuuttujia voidaan pitää *usean riippumattoman tekijän summana*, antaa keskeinen raja-arvolause selityksen *empiiriselle havainnolle* niiden normaalisuudesta.

Olkoon siis

$$X_i\ ,\ i = 1, 2, \ldots , n$$

jono *riippumattomia* ja *samoin jakautuneita satunnaismuuttujia,* joiden odotusarvo ja varianssi ovat

$$\mathrm{E}(X_i) = \mu\ ,\ i = 1, 2, \ldots, n$$
$$\mathrm{Var}(X_i) = \mathrm{D}^2(X_i) = \sigma^2\ ,\ i = 1, 2, \ldots, n$$

Olkoon

$$Y_n = \sum_{i=1}^{n} X_i$$

satunnaismuuttujien X_i , $i = 1, 2, \ldots , n$ *summa*. Tällöin

$$\mathrm{E}(Y_n) = n\mu$$
$$\mathrm{Var}(Y_n) = \mathrm{D}^2(Y_n) = n\sigma^2$$

Standardoidaan satunnaismuuttuja Y_n:

$$Z_n = \frac{Y_n - \mathrm{E}(Y_n)}{\mathrm{D}(Y_n)} = \frac{Y_n - n\mu}{\sigma\sqrt{n}}$$

Annetaan

$$n \to +\infty$$

Keskeisen raja-arvolauseen mukaan *satunnaismuuttujan Z_n jakauma lähestyy tällöin rajatta standardoitua normaalijakaumaa* N(0,1):

$$\lim_{n\to+\infty} \Pr\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le z \right) = \Phi(z)$$

jossa $\Phi(z)$ on *standardoidun normaalijakauman* N(0,1) *kertymäfunktio*.

Merkintä:

$$Z_n \overset{a}{\sim} \mathrm{N}(0,1)$$

jossa

a = *asymptoottisesti* (tai approksimatiivisesti suurille n)

Perustelu:

Ks. luvun **Stokastiikan konvergenssikäsitteet ja raja-arvolauseet** kappaletta **Keskeinen raja-arvolause**.

Keskeisen raja-arvolauseen mukaan *usean satunnaismuuttujan summa on* (tietyin ehdoin) *approksimatiivisesti normaalinen* (lähes) *riippumatta yhteenlaskettavien jakaumasta*.

Huomautus:

- Yhteenlaskettavien ei tarvitse olla edes *jatkuvia*, vaan ne voivat olla jopa *diskreettejä*.

Approksimaation *hyvyys* riippuu yhteenlaskettavien satunnaismuuttujien lukumäärästä ja niiden jakaumasta. Approksimaation *hyvyys paranee, kun yhteenlaskettavien satunnaismuuttujien lukumäärän annetaan kasvaa*.

Jos yhteenlaskettavien satunnaismuuttujien jakauma on *symmetrinen*, approksimaatio on hyvä jo suhteellisen pienillä yhteenlaskettavien lukumäärillä. Sen sijaan, jos yhteenlaskettavien satunnaismuuttujien jakauma on *epäsymmetrinen*, hyvä approksimaatio vaatii enemmän yhteenlaskettavia.

Keskeisen raja-arvolauseen tulos esitetään usein seuraavassa muodossa: Riippumattomien samaa jakaumaa noudattavien satunnaismuuttujien X_i , $i = 1, 2, \ldots , n$ **aritmeettinen keskiarvo**

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

on suurille *n approksimatiivisesti normaalinen parametrein* μ ja σ^2/n:

$$\bar{X}_n \overset{a}{\sim} \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Keskeinen raja-arvolause koskee satunnaismuuttujien **asymptoottista käyttäytymistä** samaan tapaan kuin luvussa **Jakaumien tunnusluvut** esitetty **suurten lukujen laki**. Keskeisessä raja-arvolauseessa esiintyvä *rajakäyttäytymisen muoto* on esimerkki ns. **jakaumakonvergenssista** eli **heikosta konvergenssista**; lisätietoja: ks. lukua **Stokastiikan konvergenssikäsitteet ja raja-arvolauseet**.

Keskeisestä raja-arvolauseesta on olemassa *yleisempiä muotoja*, joissa voidaan lieventää *samoinjakautuneisuus*- ja *riippumattomuusoletuksia*. Tällä on suuri merkitys esimerkiksi **stokastisten prosessien** ja **aikasarja-analyysin** teoriassa.

Binomijakauman approksimointi normaalijakaumalla: De Moivren ja Laplacen raja-arvolause

Oletetaan, että satunnaismuuttuja X noudattaa **binomijakaumaa** parametrein n ja p:

$$X \quad \mathrm{Bin}(n, p)$$

Tällöin

$$\mathrm{E}(X) = np$$
$$\mathrm{Var}(X) = npq$$

jossa

$$q = 1 - p$$

Lisätietoja binomijakaumasta: ks. lukua **Diskreettejä jakaumia**.

Koska jokainen binomijakautunut satunnaismuuttuja voidaan esittää *riippumattomien ja samaa* **Bernoulli-jakaumaa** Bernoulli(p) *noudattavien satunnaismuuttujien summana* (lisätietoja: ks. lukua **Diskreettejä jakaumia**), niin *keskeisestä raja-arvolauseesta* seuraa, että

$$\lim_{n \to +\infty} \Pr\left(\frac{X - np}{\sqrt{npq}} \le z \right) = \Phi(z)$$

jossa Φ on *standardoidun normaalijakauman* N(0,1) *kertymäfunktio*.

Perustelu:

Olkoon

$$X \quad \mathrm{Bin}(n, p)$$

Tällöin satunnaismuuttuja X voidaan esittää *riippumattomien* ja *samaa* **Bernoulli-jakaumaa** Bernoulli(p) *noudattavien satunnaismuuttujien summana*:

$$X = \sum_{i=1}^{n} X_i$$

jossa

$$X_i \quad \mathrm{Bernoulli}(p)\,,\ i = 1, 2, \mathrm{K}\,, n$$

Koska

$$\mathrm{E}(X_i) = p\,,\ i = 1, 2, \mathrm{K}\,, n$$
$$\mathrm{Var}(X_i) = pq\,,\ q = 1 - p\,,\ i = 1, 2, \mathrm{K}\,, n$$

niin

$$\mathrm{E}(X) = \sum_{i=1}^{n} \mathrm{E}(X_i) = np$$
$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = npq$$

Siten keskeisestä raja-arvolauseesta seuraa, että

$$\lim_{n\to+\infty} \Pr\left(\frac{X-np}{\sqrt{npq}} \leq z \right) = \Phi(z)$$

■

Tätä keskeisen raja-arvolauseen erikoistapausta kutsutaan tavallisesti **De Moivren ja Laplacen raja-arvolauseeksi**.

Binomijakauman approksimointi normaalijakaumalla onnistuu kohtuullisen hyvin jo melko pienille toistokokeiden lukumäärille n, jos $p \approx 1/2$. Sen sijaan, jos $p \approx 0$ tai $p \approx 1$, vaatii hyvä approksimaatio selvästi suuremman toistokokeiden lukumäärän n.

De Moivren ja Laplacen raja-arvolauseen havainnollistus

Alla esitetty neljän kuvan sarja näyttää miten satunnaismuuttujien

$$X \quad \mathrm{Bin}(n, p)$$
$$Z \quad \mathrm{N}(np, npq)\,,\; q = 1 - p$$

jakaumat alkavat muistuttaa yhä enemmän toisiaan, kun toistokokeiden lukumäärän n annetaan kasvaa.

Kuvasarjan kaikissa kuvissa

$$p = 0.1$$

mutta n vaihtelee saaden arvot

$$n = 1, 10, 30, 100$$

Kuviin piirretyn normaalijakauman $\mathrm{N}(\mu, \sigma^2)$ odotusarvo ja varianssi on asetettu samoiksi kuin binomijakauman Bin(n, p) odotusarvo ja varianssi, ts.

$$\mu = np$$
$$\sigma^2 = npq\,,\; q = 1 - p$$

Kuva 1:

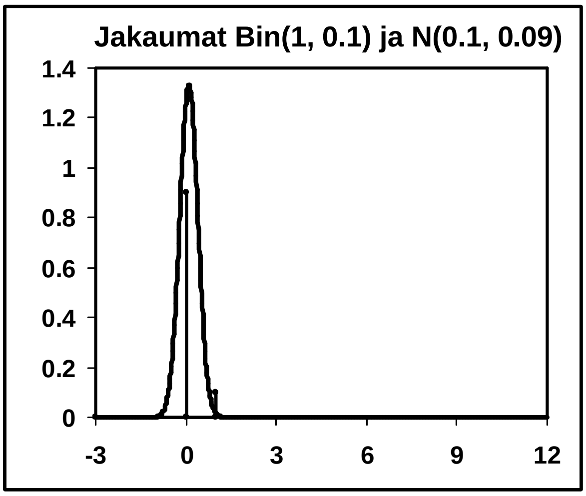
Olkoon $X \quad \mathrm{Bin}(n, p)$, jossa

$$n = 1$$
$$p = 0.1$$

ja olkoon $Z \quad \mathrm{N}(\mu, \sigma^2)$, jossa

$$\mu = np = 0.1$$
$$\sigma^2 = np(1-p) = 0.09$$

Kuva 1 esittää satunnaismuuttujan X *pistetodennäköisyysfunktiota* ja satunnaismuuttujan Z *tiheysfunktiota* välillä [–3, 12].



Kuva 2:

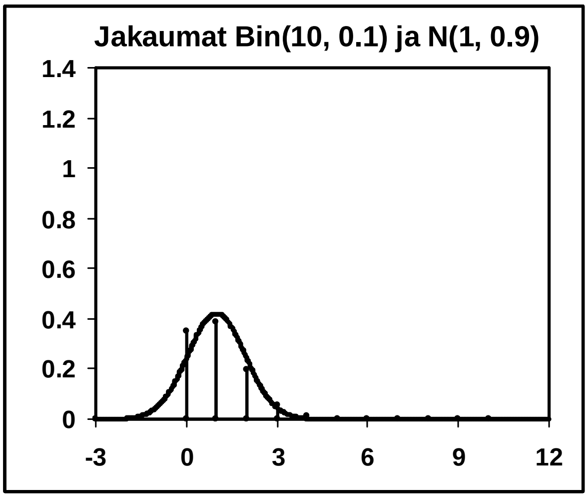
Olkoon $X \quad \mathrm{Bin}(n, p)$, jossa

$$n = 10$$
$$p = 0.1$$

ja olkoon $Z \quad \mathrm{N}(\mu, \sigma^2)$, jossa

$$\mu = np = 1$$
$$\sigma^2 = np(1 - p) = 0.9$$

Kuva 2 esittää satunnaismuuttujan X *pistetodennäköisyysfunktiota* ja satunnaismuuttujan Z *tiheysfunktiota* välillä [–3, 12].



Kuva 3:

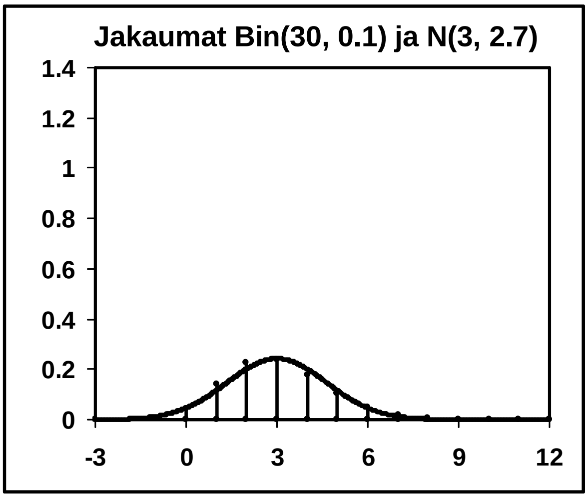
Olkoon $X \quad \mathrm{Bin}(n, p)$, jossa

$$n = 30$$
$$p = 0.1$$

ja olkoon $Z \quad \mathrm{N}(\mu, \sigma^2)$, jossa

$$\mu = np = 3$$
$$\sigma^2 = np(1 - p) = 2.7$$

Kuva 3 esittää satunnaismuuttujan X *pistetodennäköisyysfunktiota* ja satunnaismuuttujan Z *tiheysfunktiota* välillä [–3, 12].



Kuva 4:

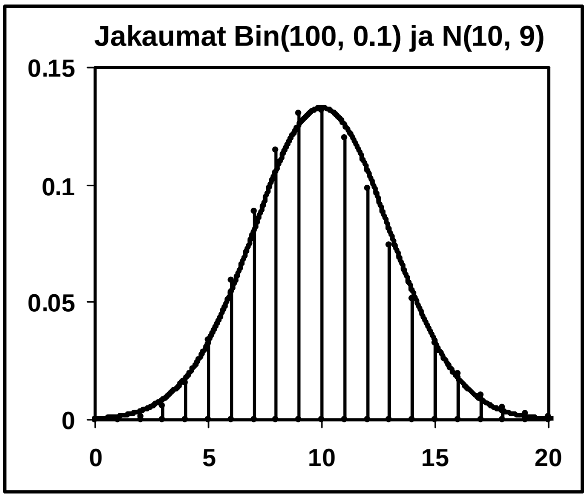
Olkoon $X \quad \mathrm{Bin}(n, p)$, jossa

$$n = 100$$
$$p = 0.1$$

ja olkoon $Z \quad \mathrm{N}(\mu, \sigma^2)$, jossa

$$\mu = np = 10$$
$$\sigma^2 = np(1 - p) = 9$$

Kuva 4 esittää satunnaismuuttujan X *pistetodennäköisyysfunktiota* ja satunnaismuuttujan Z *tiheysfunktiota* välillä [0, 20].



Huomaa, että kuvan 4 koordinaattiakseleilla on *eri* skaala kuin kuvien 1, 2 ja 3 koordinaattiakseleilla.

Binomitodennäköisyyksien approksimointi normaalijakauman avulla

De Moivren ja Laplacen raja-arvolauseen mukaan *binomijakaumaa*

$$\mathrm{Bin}(n, p)$$

voidaan suurille n *approksimoida normaalijakaumalla*

$$\mathrm{N}(\mu, \sigma^2)$$

jossa

$$\mu = np$$

$$\sigma^2 = npq\ , \ q = 1 - p$$

Jos siis

$$X \quad \mathrm{Bin}(n, p)$$

niin *De Moivren ja Laplacen raja-arvolauseen* mukaan suurille n

$$\Pr(a < X \le b) \approx \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right)$$

jossa Φ on standardoidun normaalijakauman N(0,1) kertymäfunktio.

Jos a ja b ovat *kokonaislukuja*, approksimaatio *on* hieman *parempi*, jos käytetään kaavaa

$$\Pr(a < X \le b) \approx \Phi\left(\frac{b + 1/2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - 1/2 - np}{\sqrt{npq}}\right)$$

Korjaustekijä 1/2 perustuu siihen, että *diskreettiä* binomijakaumaa approksimoidaan *jatkuvalla* normaalijakaumalla; ks. seuraavaa esimerkkiä.

Jos annetaan $a \to -\infty$, saadaan approksimaatiotulos

$$\Pr(X \le b) = F_X(b) \approx \Phi\left(\frac{b + 1/2 - np}{\sqrt{npq}}\right)$$

jossa $F_X(\cdot)$ on *binomijakauman kertymäfunktio*.

Jos $a = b$, saadaan approksimaatiotulos

$$\Pr(X = a) = f_X(a) \approx \Phi\left(\frac{a + 1/2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - 1/2 - np}{\sqrt{npq}}\right)$$

jossa $f_X(\cdot)$ on *binomijakauman pistetodennäköisyysfunktio*.

Esimerkki 2. Binomitodennäköisyyksien approksimointi normaalijakauman avulla.

Kuva alla esittää jakauman

$$\mathrm{Bin}(100,0.1)$$

pistetodennäköisyysfunktiota ja jakauman

$$\mathrm{N}(10,9)$$

tiheysfunktiota välillä [6, 12].

De Moivren ja Laplacen raja-arvolauseen mukaan binomitodennäköisyyttä pisteessä

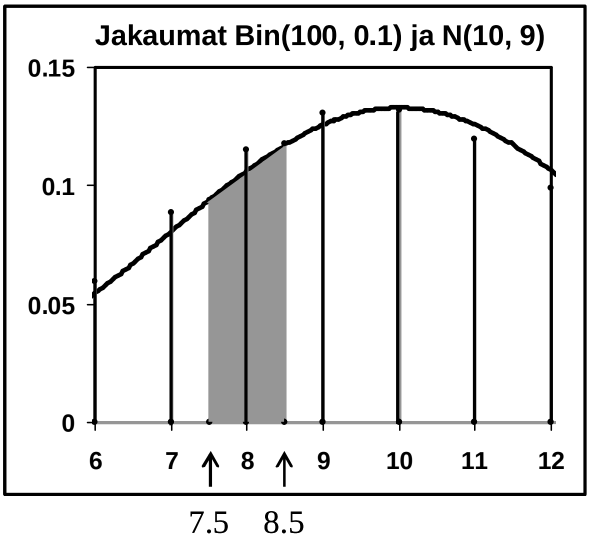
$$x = 8$$

voidaan *approksimoida* varjostetun alueen pinta-alalla.

Olkoon siis $X \quad \mathrm{Bin}(n, p)$, jossa

$$n = 100$$

$$p = 0.1$$



Tällöin

$$f_X(8) = 0.1148$$

jossa

$$f_X(x) = \binom{n}{p} p^x (1-p)^{n-x}$$

on binomijakauman Bin(100,0.1) *pistetodennäköisyysfunktio*.

Asetetaan

$$\mu = np = 10$$

$$\sigma^2 = np(1-p) = 9$$

Tällöin

$$\begin{aligned}\Phi\left(\frac{8+1/2-\mu}{\sigma}\right) - \Phi\left(\frac{8-1/2-\mu}{\sigma}\right) &= \Phi(-0.5) - \Phi(-0.8333) \\ &= 0.3085 - 0.2033 \\ &= 0.1052\end{aligned}$$

jossa Φ on standardoidun normaalijakauman N(0,1) *kertymäfunktio*.

Todennäköisyydet ovat kahden desimaalin tarkkuudella samat!

Hypergeometrisen jakauman approksimointi normaalijakauman avulla

Hypergeometrinen jakauma

$$\mathrm{HyperGeom}(N, r, n)$$

lähestyy perusjoukon koon N kasvaessa rajatta *binomijakaumaa*

$$\mathrm{Bin}(n, p)$$

jos samanaikaisesti myös r kasvaa rajatta ja

$$r / N = p$$

Lisätietoja binomijakaumasta ja hypergeometrisesta jakaumasta: ks. lukua **Diskreettejä jakaumia**.

Siten *De Moivren ja Laplacen raja-arvolauseesta* seuraa, että *hypergeometrista jakaumaa*

$$\mathrm{HyperGeom}(N, r, n)$$

voidaan suurille N *approksimoida normaalijakaumalla*

$$\mathrm{N}(\mu, \sigma^2)$$

jossa

$$\mu = n\frac{r}{N}$$
$$\sigma^2 = n\frac{r}{N}\left(1-\frac{r}{N}\right)$$

Poisson-jakauman approksimointi normaalijakaumalla

Oletetaan, että satunnaismuuttuja X noudattaa **Poisson-jakaumaa** parametrilla λ:

$$X \quad \mathrm{Poisson}(\lambda)$$

Tällöin

$$\mathrm{E}(X) = \lambda$$
$$\mathrm{Var}(X) = \lambda$$

Parametri λ kuvaa tarkastelun kohteena olevan tapahtuman *tapahtumaintensiteettiä* eli tarkastelun kohteena olevan tapahtuman esiintymisten keskimääräistä lukumäärää aikayksikköä kohden; lisätietoja Poisson-jakaumasta: ks. lukua **Diskreettejä jakaumia**.

Poisson-jakaumaa koskevassa kappaleessa (ks. lukua **Diskreettejä jakaumia**) todetaan, että Poisson-jakauma saadaan raja-prosessin kautta binomijakaumasta. Koska *De Moivren ja Laplacen raja-arvolauseen* mukaan binomijakaumaa voidaan approksimoida normaalijakaumalla, on perusteltua olettaa, että myös Poisson-jakaumaa voidaan approksimoida normaalijakaumalla.

On helppoa todistaa momenttiemäfunktioita hyväksi käyttäen (ks. lukua **Momenttiemäfunktio ja karakteristinen funktio**), että

$$\lim_{\lambda \to +\infty} \Pr\left(\frac{X-\lambda}{\sqrt{\lambda}} \le z\right) = \Phi(z)$$

jossa Φ on *standardoidun normaalijakauman* $\mathrm{N}(0,1)$ *kertymäfunktio*.

Perustelu:

Olkoon

$$X \quad \mathrm{Poisson}(\lambda)$$

Satunnaismuuttujan X *momenttiemäfunktio* on

$$m_X(t) = \mathrm{E}(e^{tX}) = \exp[\lambda(e^t - 1)]$$

Tarkastellaan *standardoitua muuttujaa*

$$Z = \frac{X-\lambda}{\sqrt{\lambda}}$$

Satunnaismuuttujan *lineaarimuunnoksen* momenttiemäfunktiota koskevan tuloksen perusteella satunnaismuuttujan *Z momenttiemäfunktio* on

$$m_Z(t) = \mathrm{E}(e^{tZ}) = \exp[-\sqrt{\lambda}t]\exp[\lambda(e^{t/\sqrt{\lambda}} - 1)] = \exp[\lambda e^{t/\sqrt{\lambda}} - \lambda - \sqrt{\lambda}t]$$

Kehittämällä funktio

$$e^{t/\sqrt{\lambda}}$$

Taylorin sarjaksi saadaan lauseke

$$\begin{aligned} m_Z(t) &= \exp[\lambda e^{t/\sqrt{\lambda}} - \lambda - \sqrt{\lambda}t] \\ &= \exp\left[\lambda\left(1 + \frac{t}{\lambda^{1/2}} + \frac{t^2}{2\lambda} + \frac{t^3}{3!\lambda^{3/2}} + \cdots\right) - \lambda - \sqrt{\lambda}t\right] \\ &= \exp\left[\lambda + \lambda^{1/2}t + \frac{t^2}{2} + \frac{t^3}{3!\lambda^{1/2}} + \cdots - \lambda - \sqrt{\lambda}t\right] \\ &= \exp\left[\frac{t^2}{2} + \frac{t^3}{3!\lambda^{1/2}} + \cdots\right] \end{aligned}$$

Jos

$$\lambda \to \infty$$

niin

$$\exp\left[\frac{t^2}{2} + \frac{t^3}{3!\lambda^{1/2}} + \cdots\right] \to \exp\left[\frac{t^2}{2}\right]$$

joka on *standardoidun normaalijakauman* N(0,1) momenttiemäfunktio.

Siten satunnaismuuttujan

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

jakauma konvergoi standardoitua normaalijakauma kohti, kun $\lambda \to \infty$.

■

Poisson-todennäköisyyksien approksimointi normaalijakauman avulla

Edellä esitetyn mukaan *Poisson-jakaumaa*

$$\mathrm{Poisson}(\lambda)$$

voidaan approksimoida normaalijakaumalla

$$\mathrm{N}(\mu, \sigma^2)$$

jossa

$$\mu = \lambda$$
$$\sigma^2 = \lambda$$

Jos siis

$$X \quad \text{Poisson}(\lambda)$$

niin

$$\Pr(a < X \le b) \approx \Phi\left(\frac{b-\lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a-\lambda}{\sqrt{\lambda}}\right)$$

jossa Φ on standardoidun normaalijakauman N(0,1) kertymäfunktio.

Jos *a* ja *b* ovat *kokonaislukuja*, approksimaatio *on* hieman *parempi*, jos käytetään kaavaa

$$\Pr(a < X \le b) \approx \Phi\left(\frac{b+1/2-\lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a-1/2-\lambda}{\sqrt{\lambda}}\right)$$

Korjaustekijä 1/2 perustuu siihen, että *diskreettiä* Poisson-jakaumaa approksimoidaan *jatkuvalla* normaalijakaumalla.

Jos annetaan $a \to -\infty$, saadaan approksimaatiotulos

$$\Pr(X \le b) = F_X(b) \approx \Phi\left(\frac{b+1/2-\lambda}{\sqrt{\lambda}}\right)$$

jossa $F_X(\cdot)$ on *Poisson-jakauman kertymäfunktio*.

Jos $a = b$, saadaan approksimaatiotulos

$$\Pr(X = a) = f_X(a) \approx \Phi\left(\frac{a+1/2-\lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a-1/2-\lambda}{\sqrt{\lambda}}\right)$$

jossa $f_X(\cdot)$ on *Poisson-jakauman pistetodennäköisyysfunktio*.

17.5. Log-normaalijakauma

Log-normaalijakaumaa voidaan käyttää mallintamaan sellaisten satunnaismuuttujien jakaumia, jotka *eivät voi saada negatiivisia arvoja*, mutta *voivat saada miten suuria arvoja tahansa*. Tyypillinen esimerkki tällaisesta jakaumasta on *tulojakauma*.

Olkoon *jatkuvan satunnaismuuttujan X tiheysfunktio*

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \frac{1}{x} \exp\left\{-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right\}, -\infty < \mu < +\infty \,,\, \sigma > 0 \,,\, 0 < x < +\infty$$

Tällöin satunnaismuuttuja *X* noudattaa **log-normaalijakaumaa** parametrein μ ja σ^2 ja merkitsemme:

$$X \quad \text{LogN}(\mu, \sigma^2)$$

Log-normaalijakauman johto

Olkoon

$$Y = \log X \quad \text{N}(\mu, \sigma^2)$$

Tällöin

$$X = \exp(Y) \quad \mathrm{LogN}(\mu, \sigma^2)$$

Perustelu:

Olkoon

$$Y \quad \mathrm{N}(\mu, \sigma^2)$$

Tällöin satunnaismuuttujan *Y* tiheysfunktio on

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2}\left(\frac{y-\mu}{\sigma} \right)^2 \right\}, -\infty < \mu < +\infty \,,\, \sigma > 0 \,,\, -\infty < y < +\infty$$

Tehdään muuttujan vaihto

$$X = \exp(Y)$$

Funktio

$$x = h(y) = \exp(y)$$

on *aidosti kasvava ja derivoituva* ja sen käänteismuunnoksen

$$y = h^{-1}(x) = \log(x)$$

derivaatta on

$$\frac{dh^{-1}(x)}{dx} = \frac{d\log(x)}{dx} = \frac{1}{x}$$

Siten (ks. luvun **Satunnaismuuttujien muunnosten jakaumat** kappaletta **Satunnaismuuttujan monotonisen muunnoksen jakauma**)

$$f_X(x) = f_Y(h^{-1}(x)) \left| \frac{dh^{-1}(x)}{dx} \right| = \frac{1}{\sqrt{2\pi}\sigma} \cdot \frac{1}{x} \exp\left\{ -\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma} \right)^2 \right\}$$

■

Log-normaalijakauman johdossa käytetystä menetelmästä seuraa automaattisesti, että funktio $f_X(x)$ on tiheysfunktio.

Log-normaalijakauman tunnusluvut

Olkoon

$$X \quad \mathrm{LogN}(\mu, \sigma^2)$$

Odotusarvo:

$$\mathrm{E}(X) = e^{\mu + \sigma^2/2}$$

Varianssi:

$$\mathrm{D}^2(X) = \mathrm{Var}(X) = e^{2(\mu+\sigma^2)} - e^{2(\mu+\sigma^2/2)}$$

Standardipoikkeama:

$$\mathrm{D}(X) = \sqrt{e^{\mu+\sigma^2} - e^{\mu+\sigma^2/2}}$$

Perustelu:

Johdamme tässä log-normaalijakauman odotusarvon.

Käytetään hyväksi sitä, että jos

$$Y = \log X \quad \mathrm{N}(\mu, \sigma^2)$$

niin

$$X = e^Y \quad \mathrm{LogN}(\mu, \sigma^2)$$

Koska

$$Y \quad \mathrm{N}(\mu, \sigma^2)$$

niin satunnaismuuttujan Y momenttiemäfunktio on

$$m_Y(t) = \mathrm{E}(e^{tY}) = \exp(\mu t + \tfrac{1}{2}\sigma^2 t^2)$$

Tällöin

$$\mathrm{E}(X) = \mathrm{E}(e^{\log(X)}) = \mathrm{E}(e^Y) = m_Y(1) = \exp(\mu + \sigma^2/2)$$

■

Log-normaalijakauman tiheysfunktio

Kuva oikealla esittää log-normaalijakauman

$$\mathrm{LogN}(\mu, \sigma^2)$$

tiheysfunktiota

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \frac{1}{x} \cdot \exp\left\{-\frac{1}{2}Q(x)\right\}$$

$$-\infty < \mu < +\infty \,, \sigma > 0$$

$$0 < x < +\infty$$

jossa

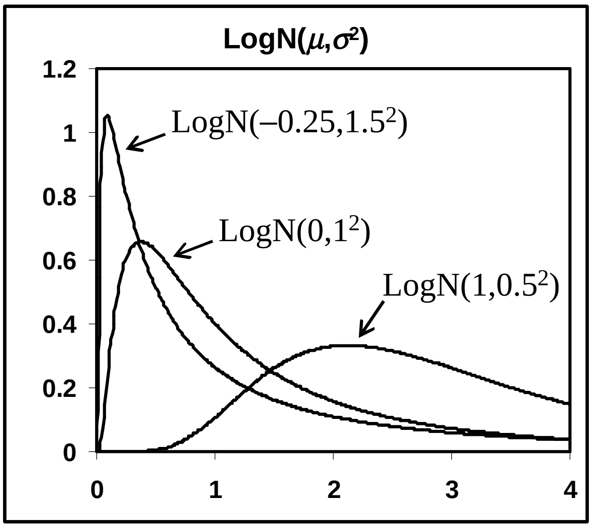
$$Q(x) = \left(\frac{\log x - \mu}{\sigma}\right)^2$$

välillä [0, 4] seuraavissa tapauksissa:

(i) $\mu = -0.25 \,, \sigma^2 = 1.5^2$

(ii) $\mu = 0 \,, \sigma^2 = 1^2$

(iii) $\mu = +1 \,, \sigma^2 = 0.5^2$



Log-normaalijakauman *odotusarvo*:

$$\mathrm{E}(X) = e^{\mu+\sigma^2/2}$$

Log-normaalijakauman tiheysfunktion ominaisuudet

(i) Log-normaalijakauman tiheysfunktio $f(x)$ on *positiivinen* kaikille positiivisille argumentin x arvoille:

$$f(x) > 0,\ x > 0$$

(ii) Tiheysfunktio on *yksihuippuinen* ja sillä on *maksimi* pisteessä $x = e^{\mu-\sigma^2}$.

(iii) Tiheysfunktio on *vino oikealle*.

Log-normaalijakauman kertymäfunktio

Log-normaalijakauma kuuluu niiden jakaumien joukkoon, joiden kertymäfunktiolle ei voida antaa eksplisiittistä lauseketta, koska *log-normaalijakauman tiheysfunktio kuuluu sellaisten funktioiden joukkoon, joiden integraalifunktiota ei voida esittää suljetussa muodossa* (so. alkeisfunktioiden avulla).

17.6. Cauchy-jakauma

Cauchy-jakauma on esimerkki jakaumasta, *jolla ei ole momentteja*.

Olkoon *jatkuvan satunnaismuuttujan X tiheysfunktio*

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

Tällöin satunnaismuuttuja X noudattaa **Cauchy-jakaumaa** parametrilla θ ja merkitsemme:

$$X \quad \mathrm{Cauchy}(\theta)$$

Funktio $f(x)$ kelpaa tiheysfunktioksi, koska

$$f(x) > 0,\ -\infty < x < +\infty$$

ja

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{+\infty} \frac{1}{\pi}\cdot\frac{1}{1+(x-\theta)^2}dx = \frac{1}{\pi}\left[\arctan(x-\theta)\right]_{-\infty}^{+\infty} = \frac{1}{\pi}\cdot\pi = 1$$

Cauchy-jakauman johto

Olkoon

$$Y \sim \mathrm{Uniform}\,(-\pi/2, \pi/2)$$

Tällöin

$$X = \tan(Y) \sim \mathrm{Cauchy}(0)$$

Perustelu: ks. luvun **Satunnaismuuttujien muunnosten jakaumat** kappaletta **Satunnaismuuttujan monotonisen muunnoksen jakauma**.

Yleinen Cauchy-jakauma parametrilla θ saadaan *siirrolla* Cauchy-jakaumasta, jonka parametrina on 0.

Cauchy-jakauman tunnusluvut

Olkoon

$$X \quad \text{Cauchy}(\theta)$$

Cauchy-jakaumalla *ei ole momentteja*.

Perustelu:

Olkoon

$$X \quad \text{Cauchy}(\theta)$$

Näytetään, että

$$\mathrm{E}(|X|) = \infty$$

jolloin *satunnaismuuttujalla X ei ole odotusarvoa*, mistä seuraa, että *satunnaismuuttujalla X ei ole myöskään korkeampia momentteja*.

Voimme yleisyyden kärsimättä olettaa, että $\theta = 0$. Kirjoitetaan

$$\mathrm{E}(|X|) = \int_{-\infty}^{+\infty} |x| f(x)dx = \int_{-\infty}^{+\infty} \frac{|x|}{\pi} \cdot \frac{1}{1+x^2} dx = \frac{2}{\pi} \int_0^{+\infty} \frac{x}{1+x^2} dx$$

Olkoon M mielivaltainen *positiivinen* luku. Tällöin voimme kirjoittaa

$$\int_0^M \frac{x}{1+x^2} dx = \left[\frac{1}{2} \log(1+x^2) \right]_0^M = \frac{1}{2} \log(1+M^2)$$

js siten

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

■

On helppo nähdä, että parametri θ on Cauchy-jakauman **mediaani** ja **moodi**.

Cauchy-jakauman tiheysfunktio

Kuva oikealla esittää Cauchy-jakauman

$$\text{Cauchy}(\theta)$$

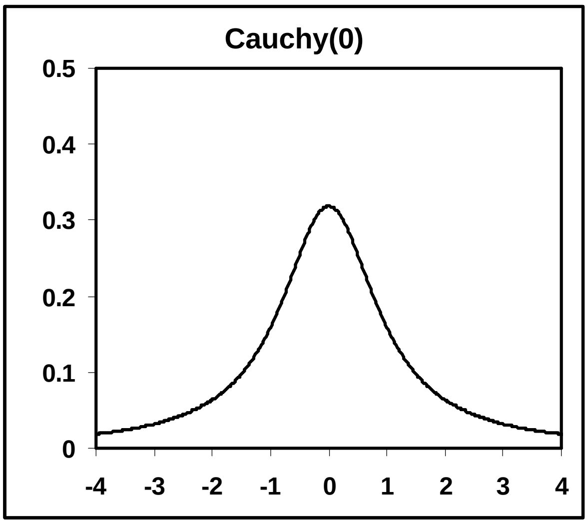
tiheysfunktiota

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+(x-\theta)^2}$$

$$-\infty < \theta < +\infty \,, -\infty < x < +\infty$$

välillä [–4,+4], kun

$$\theta = 0$$



Kuten edellä todettiin, Cauchy-jakaumalla *ei ole momentteja*, mutta parametri θ on siis jakauman *mediaani* ja *moodi*.

Cauchy-jakauman tiheysfunktion ominaisuudet

(i) Cauchy-jakauman tiheysfunktio $f(x)$ on kaikkialla *positiivinen*:

$$f(x) > 0,\ -\infty < x < +\infty$$

(ii) Tiheysfunktio on *yksihuippuinen* ja sillä on *maksimi* pisteessä $x = \theta$.

(iii) Tiheysfunktio on *symmetrinen* pisteen $x = \theta$ suhteen.

Cauchy-jakauma ja *t*-jakauma

Voidaan osoittaa, että ns. ***t*-jakauma yhdellä vapausasteella** on sama kuin Cauchy-jakauma; lisätietoja *t*-jakaumasta: ks. lukua **Normaalijakaumasta johdettuja jakaumia**.

17.7. Gamma-jakauma

Gamma-jakauma on joustava jakauma, jolla voidaan mallintaa monia erilaisia satunnaisilmiöitä. Sitä käytetään mm. *elinajan* mallintamiseen kuten erikoistapaustaan *eksponenttijakaumaa*.

Olkoon *jatkuvan satunnaismuuttujan X tiheysfunktio*

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \ , \alpha > 0 \ , \beta > 0 \ , x \ge 0$$

jossa

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$$

on ns. *gamma-funktio*. Tällöin satunnaismuuttuja X noudattaa **gamma-jakaumaa** parametrein α ja β ja merkitsemme:

$$X \quad \text{Gamma}(\alpha, \beta)$$

Funktio $f(x)$ kelpaa tiheysfunktioksi, koska

$$f(x) > 0 \ , x \ge 0$$

ja gamma-funktion määritelmän perusteella

$$\int_{-\infty}^{+\infty} f(x)dx = \int_0^{+\infty} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} dx = \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} t^{\alpha-1} e^{-t} dt = \frac{1}{\Gamma(\alpha)} \cdot \Gamma(\alpha) = 1$$

Gamma-funktio

Gamma-funktio määritellään kaavalla

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$$

Voidaan osoittaa, että gamma-funktiolla on seuraavat ominaisuudet:

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

(ii) $\Gamma(n) = (n-1)!\,,\ n \in$

(iii) $\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$

Perustelu:

(i) Saamme osittaisintegroinnilla rekursiokaavan

$$\Gamma(\alpha+1) = \int_0^\infty t^\alpha e^{-t} dt = \left[-t^\alpha e^{-t}\right]_0^\infty + \alpha \int_0^\infty t^{\alpha-1} e^{-t} dt = \alpha \int_0^\infty t^{\alpha-1} e^{-t} dt = \alpha\Gamma(\alpha)$$

(ii) Todetaan ensin, että

$$\Gamma(1) = \int_0^\infty e^{-t} dt = \left[-e^{-t}\right]_0^\infty = 1$$

Soveltamalla kaavan (i) rekursiota ja sitä, että $\Gamma(1) = 1$, saamme

$$\begin{aligned}
\Gamma(n) &= (n-1)\Gamma(n-1) \\
&= (n-1)(n-2)\Gamma(n-2) \\
&= (n-1)(n-2)(n-3)\Gamma(n-3) \\
&\ \cdots \\
&= (n-1)(n-2)(n-3) \cdots (n-(n-2))(n-(n-1))\Gamma(n-(n-1)) \\
&= (n-1)(n-2)(n-3) \cdots 2 \cdot 1 \cdot \Gamma(1) \\
&= (n-1)!
\end{aligned}$$

(iii) Kaava

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$

voidaan todistaa vetoamalla kappaleessa **Normaalijakauma** esitettyyn todistukseen sille, että normaalijakauman tiheysfunktio $f(x)$ toteuttaa ehdon

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

Tällöin todettiin, että

$$\int_0^\infty e^{-z^2/2} dz = \sqrt{\tfrac{1}{2}\pi}$$

Tehdään tässä integraalissa muuttujan vaihto

$$z = \sqrt{2t}$$

jolloin

$$dz = (2t)^{-1/2} dt$$

Saamme siten yhtälön

$$\int_0^\infty 2^{-1/2} t^{-1/2} e^{-t}\, dw = \sqrt{\tfrac{1}{2}\pi}$$

Tästä yhtälöstä seuraa välittömästi, että

$$\int_0^{\infty} t^{-1/2} e^{-t}\, dw = \sqrt{\pi}$$

Koska toisaalta

$$\Gamma(\tfrac{1}{2}) = \int_0^{\infty} t^{-1/2} e^{-t} dt$$

saamme lopulta haluamamme yhtälön

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$

■

Gamma-jakauman tunnusluvut

Olkoon

$$X \quad \text{Gamma}(\alpha, \beta)$$

Odotusarvo:

$$\mathrm{E}(X) = \alpha\beta$$

Varianssi:

$$\mathrm{D}^2(X) = \mathrm{Var}(X) = \alpha\beta^2$$

Standardipoikkeama:

$$\mathrm{D}(X) = \beta\sqrt{\alpha}$$

Perustelu:

Johdamme tässä gamma-jakauman odotusarvon. Jakauman odotusarvo ja varianssi johdetaan jakauman *momenttiemäfunktion* avulla kohdassa **Gamma-jakauman momenttiemäfunktio ja jakauman tunnusluvut**.

Suoraan jatkuva jakauman odotusarvon määritelmän mukaan

$$\mathrm{E}(X) = \int_0^{+\infty} x f(x) dx = \int_0^{+\infty} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x x^{\alpha-1} e^{-x/\beta} dx = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{+\infty} x^{\alpha} e^{-x/\beta} dx$$

Koska integroitava

$$x^{\alpha} e^{-x/\beta}$$

on gamma-jakauman $\text{Gamma}(\alpha+1, \beta)$ tiheysfunktion *ydin* (funktion ydin on funktion *pääosa* eli se osa funktiosta, joka jää jäljelle, kun integroimiseen vaikuttamattomat vakiot jätetään funktiosta pois), niin

$$\int_0^{+\infty} x^{\alpha} e^{-x/\beta} dx = \Gamma(\alpha+1)\beta^{\alpha+1}$$

Siten

$$\mathrm{E}(X) = \frac{\Gamma(\alpha+1)\beta^{\alpha+1}}{\Gamma(\alpha)\beta^{\alpha}} = \frac{\alpha\Gamma(\alpha)\beta^{\alpha+1}}{\Gamma(\alpha)\beta^{\alpha}} = \alpha\beta$$

jossa olemme käyttäneet hyväksi sitä gamma-funktion ominaisuutta, että

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

Gamma-jakauman tiheysfunktio

Kuva oikealla esittää gamma-jakauman

$$\mathrm{Gamma}(\alpha, \beta)$$

tiheysfunktiota

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}$$

$$\alpha > 0 \,,\ \beta > 0 \,,\ x \geq 0$$

välillä [0, 8] seuraavissa tapauksissa:

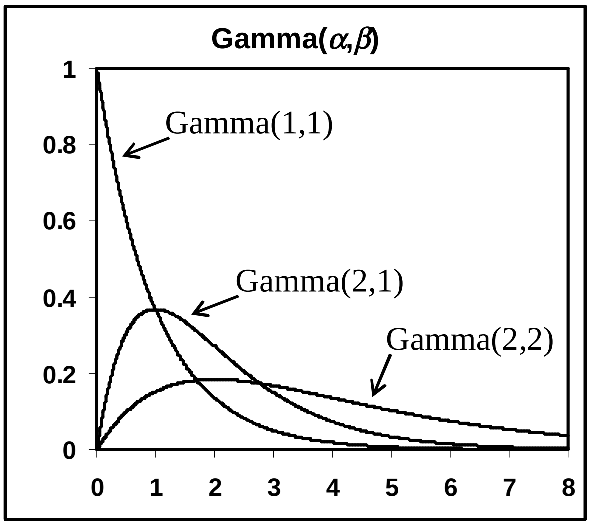
(i) $\alpha = 1 \,,\ \beta = 1$

(ii) $\alpha = 2 \,,\ \beta = 1$

(iii) $\alpha = 2 \,,\ \beta = 2$

Gamma-jakauman *odotusarvo*:

$$\mathrm{E}(X) = \alpha\beta$$



Gamma-jakauman tiheysfunktion ominaisuudet

(i) Gamma-jakauman tiheysfunktio on *positiivinen* kaikille positiivisille argumentin *x* arvoille:

$$f(x) > 0,\ x > 0$$

(ii) Jos $\alpha \leq 1$, niin tiheysfunktio on *monotonisesti laskeva* ja sillä on *maksimi* pisteessä $x = 0$.

(iii) Jos $\alpha > 1$, niin tiheysfunktio on *yksihuippuinen* ja sillä on *maksimi* pisteessä $x = (\alpha - 1)\beta$.

(iv) Tiheysfunktio on *vino oikealle*.

Gamma-jakauman kertymäfunktio

Gamma-jakauma kuuluu niiden jakaumien joukkoon, joiden kertymäfunktiolle ei voida antaa eksplisiittistä lauseketta, koska *gamma-jakauman tiheysfunktio kuuluu sellaisten funktioiden joukkoon, joiden integraalifunktiota ei voida esittää suljetussa muodossa* (so. alkeisfunktioiden avulla).

Gamma-jakauman momenttiemäfunktio

Gamma-jakauman

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \,,\ \alpha > 0 \,,\ \beta > 0 \,,\ x \geq 0$$

momenttiemäfunktio (ks. lukua **Momenttiemäfunktio ja karakteristinen funktio**) on

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

Perustelu:

Jatkuvan jakauman momenttiemäfunktion määritelmän mukaan

$$\begin{aligned}
m_X(t) = \mathrm{E}(e^{tX}) &= \int_{-\infty}^{+\infty} e^{tx} f(x)dx \\
&= \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} dx \\
&= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_{-\infty}^{+\infty} x^{\alpha-1} e^{-\left(\frac{1}{\beta}-t\right)x} dx \\
&= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_{-\infty}^{+\infty} x^{\alpha-1} e^{-x/\left(\frac{\beta}{1-\beta t}\right)} dx
\end{aligned}$$

Koska integroitava

$$x^{\alpha-1} e^{-x/\left(\frac{\beta}{1-\beta t}\right)}$$

on gamma-jakauman $\mathrm{Gamma}(\alpha, \beta/(1-\beta t))$ tiheysfunktion *ydin* (funktion ydin on funktion *pääosa* eli se osa funktiosta, joka jää jäljelle, kun integroimisen kannalta vakiot jätetään funktiosta pois), niin

$$\int_{0}^{+\infty} x^{\alpha-1} e^{-x/\left(\frac{\beta}{1-\beta t}\right)} dx = \Gamma(\alpha)\left(\frac{\beta}{1-\beta t}\right)^{\alpha}$$

Siten

$$m_X(t) = \mathrm{E}(e^{tX}) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \Gamma(\alpha)\left(\frac{\beta}{1-\beta t}\right)^{\alpha} = \left(\frac{1}{1-\beta t}\right)^{\alpha}$$

joka pätee, kun

$$t < \frac{1}{\beta}$$

koska lausekkeen $\beta/(1-\beta t)$ on oltava gamma-jakauman parametrina *positiivinen*.

■

Gamma-jakauman momenttiemäfunktio ja jakauman tunnusluvut

Johdetaan gamma-jakauman $\mathrm{Gamma}(\alpha, \beta)$ *odostusarvo* ja *varianssi* jakauman momentti-emäfunktion avulla.

Odotusarvo:

$$\mathrm{E}(X) = \alpha\beta$$

Varianssi:

$$\mathrm{D}^2(X) = \mathrm{Var}(X) = \alpha\beta^2$$

Perustelu:

Gamma-jakauman $\mathrm{Gamma}(\alpha, \beta)$ momenttiemäfunktio on

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

Momenttiemäfunktion 1. derivaatta pisteessä $t = 0$:

$$\left.\frac{dm_X(t)}{dt}\right|_{t=0} = \left.\frac{\alpha\beta}{(1-\beta t)^{\alpha+1}}\right|_{t=0} = \alpha\beta$$

Momenttiemäfunktion 2. derivaatta pisteessä $t = 0$:

$$\left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \left.\frac{\alpha(\alpha+1)\beta^2}{(1-\beta t)^{\alpha+2}}\right|_{t=0} = \alpha(\alpha+1)\beta^2$$

Siten eksponenttijakauman $\mathrm{Exp}(\lambda)$ *odotusarvo* μ_X, 2. *momentti* α_2 ja *varianssi* σ_X^2 saadaan seuraavilla kaavoilla:

$$\mu_X = \mathrm{E}(X) = \alpha_1 = \left.\frac{dm_X(t)}{dt}\right|_{t=0} = \alpha\beta$$

$$\alpha_2 = \mathrm{E}(X^2) = \left.\frac{d^2 m_X(t)}{dt^2}\right|_{t=0} = \alpha(\alpha+1)\beta^2$$

$$\sigma_X^2 = \mathrm{Var}(X) = \alpha_2 - \alpha_1^2 = \alpha(\alpha+1)\beta^2 - \alpha^2\beta^2 = \alpha\beta^2$$

■

Gamma-jakauma ja Poisson-jakauma

Olkoon

$$X \quad \mathrm{Gamma}(\alpha, \beta)$$

jossa α on *kokonaisluku* eli

$$\alpha \in$$

Voidaan osoittaa, että tällöin

$$\Pr(X \le x) = \Pr(Y \ge \alpha)$$

jossa satunnaismuuttuja Y noudattaa **Poisson-jakaumaa** parametrilla x/β:

$$Y \quad \mathrm{Poisson}(x/\beta)$$

Lisätietoja Poisson-jakaumasta: ks. lukua **Diskreettejä jakaumia**.

Gamma-jakauma ja eksponenttijakauma

Olkoon

$$X \sim \text{Gamma}(\alpha,\beta)$$

jossa

$$\alpha = 1$$

Tällöin X noudattaa **eksponenttijakaumaa** parametrilla $1/\beta$:

$$X \sim \text{Exp}(1/\beta)$$

Lisätietoja eksponenttijakaumasta: ks. kappaletta **Eksponenttijakauma**.

Gamma-jakauma ja χ^2-jakauma

Olkoon

$$X \sim \text{Gamma}(\alpha,\beta)$$

jossa

$$\alpha = \frac{p}{2}$$

$$\beta = 2$$

Voidaan osoittaa, että tällöin X noudattaa **χ^2-jakaumaa *p*:llä vapausasteella**:

$$X \sim \chi^2(p)$$

Lisätietoja χ^2-jakaumasta: ks. lukua **Normaalijakaumasta johdettuja jakaumia**.

17.8. Beta-jakauma

Beta-jakaumaa käytetään (*bayeslaisessa tilastotieteessa*) *suhteellisen osuuden* (priori-) jakaumana.

Olkoon *jatkuvan satunnaismuuttujan X tiheysfunktio*

$$f(x) = \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,,\ \alpha > 0\,,\ \beta > 0\,,\ 0 \le x \le 1$$

jossa

$$\mathrm{B}(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt$$

on ns. *beta-funktio*. Tällöin satunnaismuuttuja X noudattaa **beta-jakaumaa** parametrein α ja β ja merkitsemme:

$$X \sim \text{Beta}(\alpha,\beta)$$

Funktio $f(x)$ kelpaa tiheysfunktioksi, koska

$$f(x) > 0\,,\ x \in (0,1)$$

ja beta-funktion määritelmän perusteella

$$\int_{-\infty}^{+\infty} f(x)dx = \int_0^1 \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx$$
$$= \frac{1}{\mathrm{B}(\alpha,\beta)} \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dt = \frac{1}{\mathrm{B}(\alpha,\beta)} \cdot \mathrm{B}(\alpha,\beta) = 1$$

Beta-funktio

Beta-funktio määritellään kaavalla

$$\mathrm{B}(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$$

Beta-funktiolla ja *gamma-funktiolla* on seuraava yhteys:

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

jossa $\Gamma(\cdot)$ ns. *gamma-funktio*; lisätietoja gamma-funktiosta: ks. kappaletta **Gamma-jakauma**.

Beta-jakauman tunnusluvut

Olkoon

$$X \quad \mathrm{Beta}(\alpha,\beta)$$

Odotusarvo:

$$\mathrm{E}(X) = \frac{\alpha}{\alpha+\beta}$$

Varianssi:

$$\mathrm{D}^2(X) = \mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Standardipoikkeama:

$$\mathrm{D}(X) = \frac{1}{\alpha+\beta}\sqrt{\frac{\alpha\beta}{(\alpha+\beta+1)}}$$

Perustelu:

Johdamme tässä beta-jakauman odotusarvon.

Suoraan jatkuva jakauman odotusarvon määritelmän mukaan

$$\mathrm{E}(X) = \int_0^1 xf(x)dx = \int_0^1 \frac{1}{\mathrm{B}(\alpha,\beta)} xx^{\alpha-1}(1-x)^{\beta-1}dx = \frac{1}{\mathrm{B}(\alpha,\beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1}dx$$

Koska integroitava

$$x^{\alpha}(1-x)^{\beta-1}$$

on beta-jakauman $\mathrm{Beta}(\alpha+1,\beta)$ tiheysfunktion *ydin* (funktion ydin on funktion *pääosa* eli se osa funktiosta, joka jää jäljelle, kun integroimiseen vaikuttamattomat vakiot jätetään funktiosta pois), niin

$$\int_0^1 x^{\alpha}(1-x)^{\beta-1}dx = \mathrm{Beta}(\alpha+1,\beta)$$

Siten

$$\mathrm{E}(X) = \frac{\mathrm{B}(\alpha+1,\beta)}{\mathrm{B}(\alpha,\beta)} = \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\alpha}{\alpha+\beta}$$

jossa olemme käyttäneet hyväksi sekä beta-funktion ja gamma-funktion yhteyttä, että sitä gamma-funktion ominaisuutta, että

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

■

Beta-jakauman tiheysfunktio

Ylempi kuva oikealla esittää beta-jakauman

$$\mathrm{Beta}(\alpha,\beta)$$

tiheysfunktiota

$$f(x) = \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

$$\alpha > 0\,,\ \beta > 0\,,\ 0 \le x \le 1$$

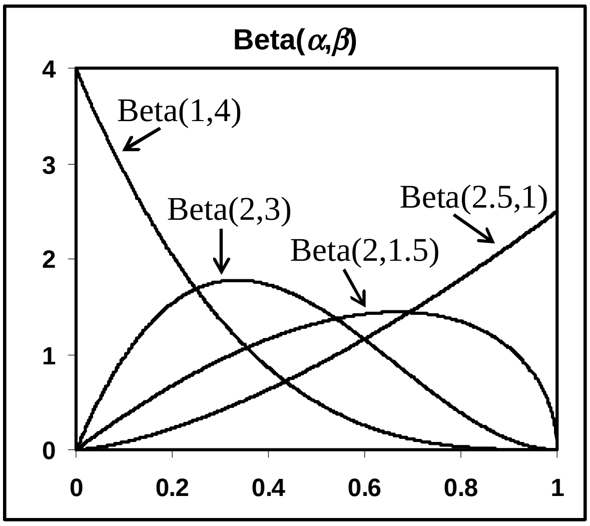
välillä [0, 1] seuraavissa tapauksissa:

(i) $\alpha = 1 \quad , \beta = 4$

(ii) $\alpha = 2 \quad , \beta = 3$

(iii) $\alpha = 2 \quad , \beta = 1.5$

(iv) $\alpha = 2.5 \,, \beta = 1$



Alempi kuva oikealla esittää beta-jakauman

$$\mathrm{Beta}(\alpha,\beta)$$

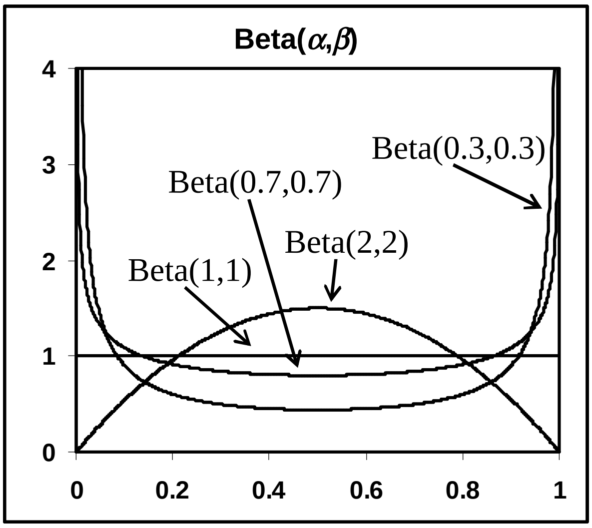
tiheysfunktiota välillä [0, 1] seuraavissa tapauksissa:

(i) $\alpha = 0.3 \,, \beta = 0.3$

(ii) $\alpha = 0.7 \,, \beta = 0.7$

(iii) $\alpha = 1 \quad , \beta = 1$

(iv) $\alpha = 2 \quad , \beta = 1$



Beta-jakauma *odotusarvo*:

$$\mathrm{E}(X) = \frac{\alpha}{\alpha+\beta}$$

Beta-jakauman tiheysfunktion ominaisuudet

(i) Beta-jakauman tiheysfunktio $f(x)$ on *positiivinen* välillä (0,1):

$$f(x) > 0,\ x \in (0,1)$$

(ii) Jos $\alpha > 1$ ja $\beta = 1$, niin tiheysfunktio on *aidosti kasvava* ja sillä on *maksimi* pisteessä $x = 1$.

(iii) Jos $\alpha = 1$ ja $\beta > 1$, niin tiheysfunktio on *aidosti vähenevä* ja sillä on *maksimi* pisteessä $x = 0$.

(iv) Jos $\alpha > 1$ ja $\beta > 1$, niin tiheysfunktio on *yksihuippuinen* ja sillä on *maksimi* pisteessä $x = (\alpha - 1)/(\alpha + \beta - 2)$.

(v) Jos $\alpha < 1$ ja $\beta < 1$, niin tiheysfunktio on *U:n muotoinen* ja sillä on *lokaalit maksimit* pisteissä $x = 0$ ja $x = 1$.

(vi) Tiheysfunktio on *symmetrinen*, jos $\alpha = \beta$.

Beta-jakauman kertymäfunktio

Beta-jakauma kuuluu niiden jakaumien joukkoon, joiden kertymäfunktiolle ei voida antaa eksplisiittistä lauseketta, koska *beta-jakauman tiheysfunktio kuuluu sellaisten funktioiden joukkoon, joiden integraalifunktiota ei voida esittää suljetussa muodossa* (so. alkeisfunktioiden avulla).

Beta-jakauma ja jatkuva tasainen jakauma

Olkoon

$$X \quad \mathrm{Beta}(\alpha,\beta)$$

jossa

$$\alpha = \beta = 1$$

Helposti nähdään, että tällöin X noudattaa **jatkuvaa tasaista jakaumaa** välillä (0,1):

$$X \quad \mathrm{Uniform}(0,1)$$

Lisätietoja jatkuvasta tasaisesta jakaumasta: ks. kappaletta **Jatkuva tasainen jakauma**.

17.9. Weibull-jakauma

Weibull-jakaumaa käytetään *luotettavuustekniikassa* komponentin *elinän* jakaumana tilanteessa, jossa komponenttia ei huolleta ja todennäköisyys sille, että komponentti vikaantuu saattaa muuttua käytön aikana.

Olkoon *jatkuvan satunnaismuuttujan X tiheysfunktio*

$$f(x) = \frac{\gamma}{\beta} x^{\gamma-1} e^{-x^{\gamma}/\beta}\ ,\ \gamma > 0\ ,\ \beta > 0\ ,\ 0 \le x < \infty$$

Tällöin satunnaismuuttuja X noudattaa **Weibull-jakaumaa** parametrein γ ja β ja merkitsemme:

$$X \quad \mathrm{Weibull}(\gamma,\beta)$$

Funktio $f(x)$ kelpaa tiheysfunktioksi, koska

$$f(x) > 0\,,\; x > 0$$

ja

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Perustelu:

Todistetaan, että

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{0}^{+\infty} \frac{\gamma}{\beta} x^{\gamma-1} e^{-x^{\gamma}/\beta} dx = \frac{\gamma}{\beta} \int_{0}^{+\infty} x^{\gamma-1} e^{-x^{\gamma}/\beta} dx = 1$$

Tarkastellaan siksi integraalia

$$\int_{0}^{+\infty} x^{\gamma-1} e^{-x^{\gamma}/\beta} dx$$

Tehdään tässä integraalissa muuttujan vaihto

$$x^{\gamma} / \beta = t$$

jolloin

$$x = \beta^{1/\gamma} t^{1/\gamma}$$

ja

$$dx = \frac{\beta^{1/\gamma}}{\gamma} \cdot \frac{t^{1/\gamma}}{t} dt$$

Siten

$$\int_{0}^{+\infty} x^{\gamma-1} e^{-x^{\gamma}/\beta} dx = \int_{0}^{+\infty} \beta^{(\gamma-1)/\gamma} t^{(\gamma-1)/\gamma} e^{-t} \frac{\beta^{1/\gamma}}{\gamma} \cdot \frac{t^{1/\gamma}}{t} dt = \frac{\beta}{\gamma} \int_{0}^{+\infty} e^{-t} dt = \frac{\beta}{\gamma}$$

joten

$$\int_{-\infty}^{+\infty} f(x)dx = \frac{\gamma}{\beta} \int_{0}^{+\infty} x^{\gamma-1} e^{-x^{\gamma}/\beta} dx = \frac{\gamma}{\beta} \cdot \frac{\beta}{\gamma} = 1$$

■

Weibull-jakauman tunnusluvut

Olkoon

$$X \quad \text{Weibull}(\gamma, \beta)$$

Odotusarvo:

$$\mathrm{E}(X) = \beta^{1/\gamma} \Gamma(1 + 1/\gamma)$$

jossa $\Gamma(\cdot)$ ns. *gamma-funktio*; lisätietoja gamma-funktiosta: ks. tämän luvun kappaletta **Gamma-jakauma**.

Varianssi:

$$\mathrm{D}^2(X) = \mathrm{Var}(X) = \beta^{2/\gamma}\Gamma(1+2/\gamma) - [\mathrm{E}(X)]^2$$

Standardipoikkeama:

$$\mathrm{D}(X) = \sqrt{\mathrm{Var}(X)}$$

Perustelu:

Johdamme tässä Weibull-jakauman odotusarvon.

Suoraan jatkuva jakauman odotusarvon määritelmän mukaan

$$\mathrm{E}(X) = \int_0^{+\infty} xf(x)dx = \int_0^{+\infty} \frac{\gamma}{\beta} xx^{\gamma-1}e^{-x^\gamma/\beta}dx = \frac{\gamma}{\beta}\int_0^{+\infty} x^\gamma e^{-x^\gamma/\beta}dx$$

Tarkastellaan integraalia

$$\int_0^{+\infty} x^\gamma e^{-x^\gamma/\beta}dx$$

Tehdään tässä integraalissa muuttujan vaihto

$$x^\gamma/\beta = t$$

jolloin

$$x = \beta^{1/\gamma}t^{1/\gamma}$$

ja

$$dx = \frac{\beta^{1/\gamma}}{\gamma}\cdot\frac{t^{1/\gamma}}{t}dt$$

Siten

$$\int_0^{+\infty} x^\gamma e^{-x^\gamma/\beta}dx = \int_0^{+\infty} \beta t e^{-t}\frac{\beta^{1/\gamma}}{\gamma}\cdot\frac{t^{1/\gamma}}{t}dt = \frac{\beta^{1/\gamma+1}}{\gamma}\int_0^{+\infty} t^{1/\gamma}e^{-t}dt$$

Koska integroitava

$$t^{1/\gamma}e^{-t}$$

on gamma-jakauman $\mathrm{Gamma}(1+1/\gamma,1)$ tiheysfunktion *ydin* (funktion ydin on funktion *pääosa* eli se osa funktiosta, joka jää jäljelle, kun integroimisen kannalta vakiot jätetään funktiosta pois), niin

$$\int_0^{+\infty} t^{1/\gamma}e^{-t}dt = \Gamma(1+1/\gamma)$$

Siten

$$\mathrm{E}(X) = \frac{\gamma}{\beta}\cdot\frac{\beta^{1/\gamma+1}}{\gamma}\Gamma(1+1/\gamma) = \beta^{1/\gamma}\Gamma(1+1/\gamma)$$

■

Weibull-jakauman tiheysfunktio

Kuva oikealla esittää Weibull-jakauman

$$\mathrm{Weibull}(\gamma, \beta)$$

tiheysfunktiota

$$f(x) = \frac{\gamma}{\beta} x^{\gamma-1} e^{-x^{\gamma}/\beta}$$

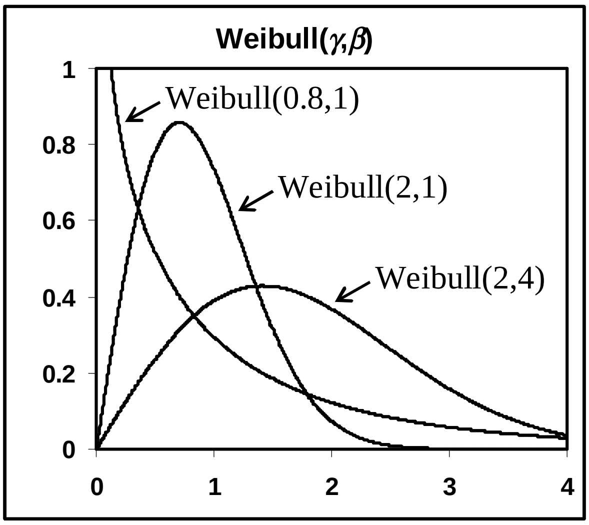
$$\gamma > 0 \, , \, \beta > 0 \, , \, 0 \le x < \infty$$

välillä [0, 4] seuraavissa tapauksissa:

(i) $\gamma = 0.8 \; , \; \beta = 1$

(ii) $\gamma = 2 \quad , \; \beta = 1$

(iii) $\gamma = 2 \quad , \; \beta = 4$



Weibull-jakauman *odotusarvo*:

$$\mathrm{E}(X) = \beta^{1/\gamma}\Gamma(1 + 1/\gamma)$$

Weibull-jakauman tiheysfunktion ominaisuudet

(i) Weibull-jakauman tiheysfunktio on *positiivinen* kaikille positiivisille argumentin x arvoille:

$$f(x) > 0, \; x > 0$$

(ii) Jos $\gamma \le 1$, niin tiheysfunktio on *monotonisesti laskeva* ja sillä on *maksimi* pisteessä $x = 0$.

(iii) Jos $\gamma > 1$, niin tiheysfunktio on *yksihuippuinen* ja sillä on *maksimi* pisteessä $x = [\beta(1 - 1/\gamma)]^{1/\gamma}$.

(iv) Tiheysfunktio on *vino oikealle*.

Weibull-jakauman kertymäfunktio

Weibull-jakauma kuuluu niiden jakaumien joukkoon, joiden kertymäfunktiolle ei voida antaa eksplisiittistä lauseketta, koska *Weibull-jakauman tiheysfunktio kuuluu sellaisten funktioiden joukkoon, joiden integraalifunktiota ei voida esittää suljetussa muodossa* (so. alkeisfunktioiden avulla).

Weibull-jakauma ja eksponenttijakauma

Olkoon

$$X \quad \mathrm{Exp}(\beta)$$

Voidaan osoittaa, että tällöin satunnaismuuttuja

$$Y = X^{1/\gamma}$$

noudattaa *Weibullin jakaumaa* parametrein γ ja β:

$$Y \quad \mathrm{Weibull}(\gamma, \beta)$$

Lisätietoja eksponenttijakaumasta: ks. kappaletta **Eksponenttijakauma**.

Kääntäen: Jos

$$Y \quad \mathrm{Weibull}(1, \beta)$$

niin

$$Y \quad \mathrm{Exp}(\beta)$$

18. Normaalijakaumasta johdettuja jakaumia

Tässä luvussa määritellään seuraavat *normaalijakaumasta johdetut todennäköisyysjakaumat*:

(i) χ^2**-jakauma.**

(ii) ***F*-jakauma.**

(iii) ***t*-jakauma.**

Näillä jakaumilla on jakaumilla on normaalijakauman itsensä rinnalla keskeinen rooli **tilastotieteessä**; ks. tästä esimerkkejä seuraavissa monisteen **Tilastolliset menetelmät** luvuissa:

Otos ja otosjakaumat

Väliestimointi

Testit suhdeasteikollisille muuttujille

Yhteensopivuuden, homogeenisuuden ja riippumattomuuden testaaminen

Yhden selittäjän lineaarinen regressiomalli

Yleinen lineaarinen malli

Avainsanat:

χ^2-jakauma, *F*-jakauma, Kertymäfunktio. Normaalijakauma, Normaalijakaumasta johdetutut jakaumat, Odotusarvo, Standardipoikkeama, *t*-jakauma, Taulukot, Tiheysfunktio, Vapausasteet, Varianssi

18.1. χ^2-jakauma

Olkoot

$$X_i\ ,\ i = 1, 2, \ldots , n$$

riippumattomia ja *standardoitua normaalijakaumaa* N(0,1) *noudattavia satunnaismuuttujia*:

$$X_1, X_2, \ldots, X_n \perp$$
$$X_i \sim \mathrm{N}(0,1)\ ,\ i = 1, 2, \ldots, n$$

Tällöin satunnaismuuttuja

$$X = \sum_{i=1}^{n} X_i^2$$

noudattaa **χ^2-jakaumaa** vapausasteella (eli *parametrilla*) n:

$$X \quad \chi^2(n)$$

χ^2-jakauman *tiheysfunktio* on muotoa

$$f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}\ ,\ 0 \le x < \infty$$

jossa $\Gamma(\cdot)$ ns. *gamma-funktio*; lisätietoja gamma-funktiosta: ks. luvun **Jatkuvia jakaumia** kappaletta **Gamma-jakauma**. χ^2-jakauman tiheysfunktio on johdettu luvussa **Satunnais-muuttujien muunnosten jakaumat**.

χ^2-jakauman tunnusluvut

Olkoon

$$X \quad \chi^2(n)$$

Odotusarvo:

$$\mathrm{E}(X) = n$$

Varianssi:

$$\mathrm{Var}(X) = \mathrm{D}^2(X) = 2n$$

Standardipoikkeama:

$$\mathrm{D}(X) = \sqrt{2n}$$

Perustelu:

Johdamme tässä χ^2-jakauman odotusarvon. Nojaamme johdossa jakauman määritelmään.

Olkoon

$$X_1, X_2, \ldots, X_n \perp$$
$$X_i \sim \mathrm{N}(0,1)\ ,\ i = 1, 2, \ldots, n$$

Tällöin

$$X = \sum_{i=1}^{n} X_i^2 \quad \chi^2(n)$$

Oletuksista seuraa, että

$$\mathrm{E}(X_i) = 0\,,\ i = 1, 2, \ldots, n$$
$$\mathrm{Var}(X_i) = \mathrm{E}(X_i^2) - [\mathrm{E}(X_i)]^2 = \mathrm{E}(X_i^2) = 1\,,\ i = 1, 2, \ldots, n$$

Siten

$$\mathrm{E}(X) = \mathrm{E}\left(\sum_{i=1}^{n} X_i^2\right) = \sum_{i=1}^{n} \mathrm{E}(X_i^2) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = \underbrace{1 + 1 + \cdots + 1}_{n \text{ kpl}} = n$$

koska odotusarvo on operaattorina *lineaarinen*.

■

χ^2-jakauman tiheysfunktio

Kuva oikealla esittää χ^2-jakauman

$$\chi^2(n)$$

tiheysfunktiota välillä [0, 10], kun vapausasteiden lukumäärällä n on seuraavat arvot:

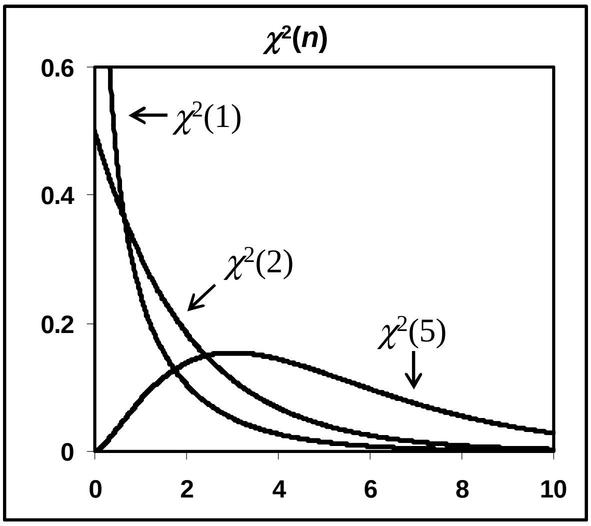
$n = 1$

$n = 2$

$n = 5$

Jakauman *odotusarvo*:

$$\mathrm{E}(X) = n$$



χ^2-jakauman tiheysfunktion ominaisuudet

(i) χ^2-jakauman tiheysfunktio $f(x)$ on *positiivinen* kaikille *positiivisille* argumentin x arvoille:

$$f(x) > 0\,,\ x > 0$$

(ii) Jos vapausasteiden lukumäärä

$$n = 1, 2$$

niin tiheysfunktio on *monotonisesti laskeva* kaikille $x \geq 0$.

(iii) Jos vapausasteiden lukumäärä

$$n \geq 3$$

niin tiheysfunktiolla on *maksimi* jossakin pisteessä $x > 0$.

Todennäköisyyksien määrääminen χ^2-jakaumasta

Todennäköisyydet χ^2*-jakaumasta* voidaan määrätä jakauman *kertymäfunktion* avulla. Olkoon

$$X \quad \chi^2(n)$$

ja olkoon satunnaismuuttujan X *kertymäfunktio*

$$F_{Chi}(x;n) = \Pr(X \le x)$$

Kaikkien χ^2-jakaumaan liittyvien tapahtumien todennäköisyydet saadaan todennäköisyyksistä

$$\Pr(X \le x) = F_{Chi}(x;n)$$

todennäköisyyslaskennan laskusääntöjen avulla. Siten esimerkiksi

$$\Pr(a \le X \le b) = F_{Chi}(b) - F_{Chi}(a)$$

Koska **χ^2-jakauman tiheysfunktion integraalifunktiota ei pystytä esittämään suljetussa muodossa** (so. alkeisfunktioiden avulla), todennäköisyyksien määräämisessä χ^2-jakaumasta on käytettävä jotakin *numeerista menetelmää*. Tämä voi tapahtua tietokoneohjelmien tai *taulukoiden* avulla.

Todennäköisyyksien määrääminen χ^2-jakaumasta ja χ^2-jakauman taulukot

Olkoon

$$X \quad \chi^2(n)$$

Tähän kirjaan liittyvässä taulukkokokoelmassa on taulukoituna usealle eri vapausasteiden $df = n$ arvolle pisteet x, jotka toteuttavat ehdon

$$p_x = \Pr(X \ge x)$$

kun

$$p_x = 0.999, 0.99, 0.95, 0.9, 0.1, 0.05, 0.01, 0.001$$

Siten piste x erottaa χ^2-jakauman *oikealle* hännälle todennäköisyysmassan, jonka koko on

$$p_x = \Pr(X \ge x)$$

Huomaa, että piste x erottaa χ^2-jakauman *vasemmalle* hännälle todennäköisyysmassan, jonka koko on

$$1 - p_x = \Pr(X \le x)$$

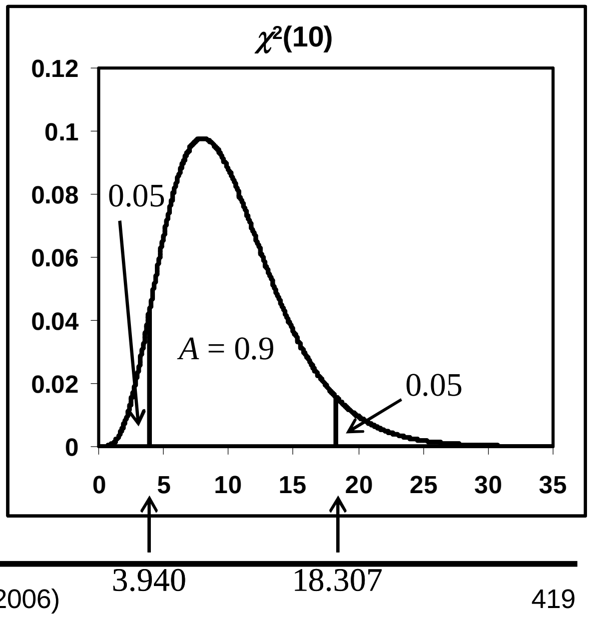
Koska eräs taulukoiden tärkeimmistä käyttöalueista on **tilastollinen testaus** (ks. monistetta **Tilastolliset menetelmät**), todennäköisyyttä p_x kutsutaan *merkitsevyystasoksi* ja pistettä x merkitsevyystasoa p_x vastaavaksi *kriittiseksi arvoksi* tai *rajaksi*.

Esimerkki 1. Todennäköisyyksien määrääminen χ^2-jakauman taulukoista.

Kuva oikealla esittää χ^2-jakauman

$$\chi^2(10)$$

tiheysfunktiota välillä [0, 35].



χ^2-jakauman taulukoiden mukaan:

$$\begin{aligned}
&\text{Alueen } A \text{ pinta-ala}\\
&\quad = \Pr(3.940 \le X \le 18.307)\\
&\quad = F_{Chi}(18.307;10) - F_{Chi}(3.940;10)\\
&\quad = 0.95 - 0.05 = 0.9
\end{aligned}$$

Todennäköisyyksien määrääminen χ^2-jakaumasta ja tietokoneohjelmat

Olkoon

$$X \quad \chi^2(n)$$

Monet *tietokoneohjelmat* mahdollistavat seuraavien tehtävien ratkaisemisen ilman χ^2-jakauman taulukoiden asettamia rajoituksia:

(i) Määrää todennäköisyys

$$\Pr(X \le x) = F_{Chi}(x\,;n)$$

kun x on annettu.

(ii) Määrää x, kun todennäköisyys

$$\Pr(X \le x) = F_{Chi}(x\,;n)$$

on annettu.

18.2. *F*-jakauma

Olkoot Y_i, i = 1, 2, … , m ja X_i, i = 1, 2, … , n *riippumattomia, standardoitua normaalijakaumaa* N(0,1) *noudattavia satunnaismuuttujia*:

$$\begin{aligned}
&Y_i \sim \mathrm{N}(0,1)\,,\ i = 1,2,\ldots,m\,;\ X_i \sim \mathrm{N}(0,1)\,,\ i = 1,2,\ldots,n\\
&Y_1, Y_2, \ldots, Y_m, X_1, X_2, \ldots, X_n \perp
\end{aligned}$$

Tällöin satunnaismuuttujat

$$Y = \sum_{i=1}^{m} Y_i^2\ ;\ X = \sum_{i=1}^{n} X_i^2$$

ovat *riippumattomia* ja noudattavat χ^2*-jakaumia* vapausastein m ja n:

$$\begin{aligned}
&Y \quad \chi^2(m)\,;\ X \quad \chi^2(n)\\
&Y \perp X
\end{aligned}$$

Määritellään satunnaismuuttuja

$$F = \frac{\frac{1}{m}Y}{\frac{1}{n}X} = \frac{n}{m}\cdot\frac{Y}{X}$$

Satunnaismuuttujat F noudattaa ***F*-jakaumaa** vapausastein (eli *parametrein*) m ja n:

$$F \quad F(m,n)$$

F-jakauman *tiheysfunktio* on muotoa

$$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}}, 0 \le x < \infty$$

jossa $\Gamma(\cdot)$ ns. *gamma-funktio*; lisätietoja gamma-funktiosta: ks. luvun **Jatkuvia jakaumia** kappaletta **Gamma-jakauma**. *F*-jakauman tiheysfunktio on johdettu luvussa **Satunnaismuuttujien muunnosten jakaumat**.

Suoraan *F*-jakauman määritelmästä nähdään, että *F*-jakaumalla on seuraava ominaisuus: Jos satunnaismuuttuja *F* noudattaa *F*-jakaumaa vapausastein *m* ja *n*:

$$F \quad F(m,n)$$

niin satunnaismuuttuja 1/*F* noudattaa *F*-jakaumaa vapausastein *n* ja *m*:

$$\frac{1}{F} \quad F(n,m)$$

F-jakauman tunnusluvut

Olkoon

$$F \quad F(m,n)$$

Odotusarvo:

$$\mathrm{E}(X) = \frac{n}{n-2}, n > 2$$

F-jakaumalla *ei ole* odotusarvoa, kun *n* = 1, 2. Huomaa myös, että *F*-jakauman odotusarvo *ei riipu* parametrista *m* (= yhteenlaskettavien lukumäärä *F*-jakauman määritelmän *osoittajassa* olevassa χ^2-satunnaismuuttujassa $Y = \sum Y_i^2$).

Varianssi:

$$\mathrm{D}^2(F) = \mathrm{Var}(F) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, n > 4$$

F-jakaumalla *ei ole* varianssia, kun *n* = 1, 2, 3, 4.

Standardipoikkeama:

$$\mathrm{D}(F) = \sqrt{\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}}, n > 4$$

F-jakaumalla *ei ole* standardipoikkeamaa, kun *n* = 1, 2, 3, 4.

F-jakauman tiheysfunktio

Kuva alla esittää *F*-jakauman

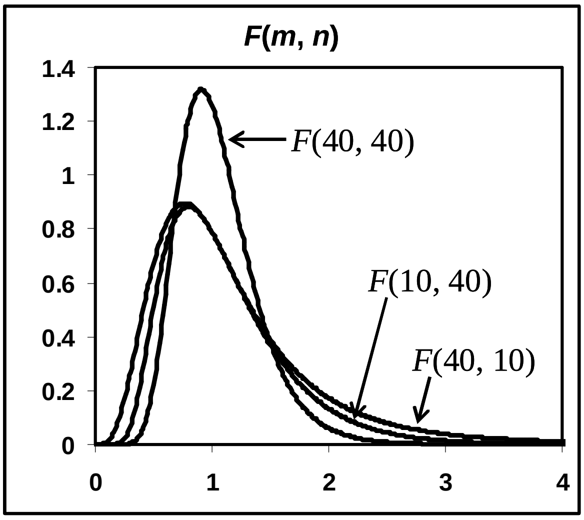
$$F(m, n)$$

tiheysfunktiota välillä [0, 5], kun vapausasteiden lukumäärillä *m* ja *n* on seuraavat arvot:

$m = 10,\ n = 40$

$m = 40,\ n = 10$

$m = 40,\ n = 40$



Jakauman *odotusarvo*:

$$\mathrm{E}(X) = \frac{n}{n-2},\ n > 2$$

F-jakauman tiheysfunktion ominaisuudet

(i) *F*-jakauman tiheysfunktio $f(x)$ on *positiivinen* kaikille *positiivisille* argumentin x arvoille:

$$f(x) > 0\ ,\ x > 0$$

(ii) Jos *osoittajan* vapausasteiden lukumäärä

$$m = 1, 2$$

niin tiheysfunktio on *monotonisesti laskeva* kaikille $x \ge 0$.

(iii) Jos *osoittajan* vapausasteiden lukumäärä

$$m \ge 3$$

niin tiheysfunktiolla on *maksimi* jossakin pisteessä $x > 0$.

Todennäköisyyksien määrääminen *F*-jakaumasta

Todennäköisyydet *F-jakaumasta* voidaan määrätä jakauman *kertymäfunktion* avulla. Olkoon

$$F \quad F(m,n)$$

ja olkoon satunnaismuuttujan *F kertymäfunktio*

$$F_F(x;m,n) = \Pr(F \le x)$$

Kaikkien F-jakaumaan liittyvien tapahtumien todennäköisyydet saadaan todennäköisyyksistä

$$\Pr(F \le x) = F_F(x;m,n)$$

todennäköisyyslaskennan laskusääntöjen avulla. Siten esimerkiksi

$$\Pr(a \le F \le b) = F_F(b) - F_F(a)$$

Koska ***F*-jakauman tiheysfunktion integraalifunktiota ei pystytä esittämään suljetussa muodossa** (so. alkeisfunktioiden avulla), todennäköisyyksien määräämisessä *F*-jakaumasta on käytettävä jotakin *numeerista menetelmää*. Tämä voi tapahtua tietokoneohjelmien tai *taulukoiden* avulla.

Todennäköisyyksien määrääminen *F*-jakaumasta ja *F*-jakauman taulukot

Olkoon

$$F \quad F(m,n)$$

Tähän kirjaan liittyvässä taulukkokokoelmassa on taulukoituna usealle eri vapausasteiden $df_1 = m$ ja $df_2 = n$ arvoille pisteet x, jotka toteuttavat ehdon

$$p_x = \Pr(F \geq x)$$

kun

$$p_x = 0.05,\ 0.025,\ 0.01,\ 0.005$$

Siten piste x erottaa F-jakauman *oikealle* hännälle todennäköisyysmassan, jonka koko on

$$p_x = \Pr(F \geq x)$$

Koska eräs taulukoiden tärkeimmistä käyttöalueista on **tilastollinen testaus** (ks. monistetta **Tilastolliset menetelmät**), kutsutaan todennäköisyyttä p_x taulukoissa *merkitsevyystasoksi* ja pistettä x merkitsevyystasoa p_x vastaavaksi *kriittiseksi arvoksi* tai *rajaksi*.

F-jakauman *vasemman* hännän todennäköisyydet saadaan käyttämällä hyväksi sitä, että satunnaismuuttuja $1/F$ noudattaa F-jakaumaa vapausastein n ja m, jos satunnaismuuttuja F noudattaa F-jakaumaa vapausastein m ja n. Jos siis

$$F \quad F(m,n)$$

niin

$$G = \frac{1}{F} \quad F(n,m)$$

Oletetaan nyt, että haluamme määrätä pisteen x siten, että

$$\Pr(F \leq x) = p$$

jossa

$$F \quad F(m,n)$$

Tällöin voimme toimia seuraavasti: Määrätään F-jakauman taulukoista piste y siten, että

$$\Pr(G \geq y) = p$$

jossa

$$G \quad F(n,m)$$

Yllä esitetyn mukaan

$$\Pr(G \geq y) = \Pr\left(\frac{1}{G} \leq \frac{1}{y}\right) = \Pr\left(F \leq \frac{1}{y}\right)$$

jossa

$$F \quad F(m,n)$$

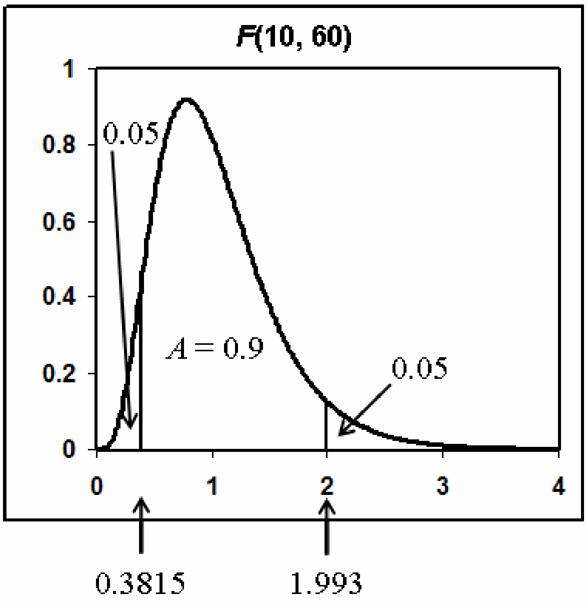
Siten piste

$$x = 1/y$$

toteuttaa ehdon

$$\Pr(F \leq x) = p$$

Esimerkki 1. Todennäköisyyksien määrääminen *F*-jakauman taulukoista.



Kuva oikealla esittää *F*-jakauman

$F(10, 60)$

tiheysfunktiota välillä [0, 4].

F-jakauman taulukoiden mukaan:

$$\begin{aligned}
&\text{Alueen } A \text{ pinta-ala} \\
&\quad = \Pr(0.3815 \le F \le 1.993) \\
&\quad = F_F(1.993; 10, 60) \\
&\qquad - F_F(0.3815; 10, 60) \\
&\quad = 0.95 - 0.05 = 0.9
\end{aligned}$$

Piste

$b = 1.993$

joka erottaa *F*-jakauman $F(10,60)$ *oikealle* hännälle todennäköisyysmassan 0.05, saadaan suoraan *F*-jakauman taulukoista. Piste

$a = 0.3815$

joka erottaa *F*-jakauman $F(10,60)$ *vasemmalle* hännälle todennäköisyysmassan 0.05, saadaan taulukoista seuraavalla tavalla: Olkoon

$$G \quad F(60,10)$$

F-jakauman taulukoiden mukaan

$$\Pr(G \ge 2.621) = 0.05$$

Siten piste

$$a = \frac{1}{2.621} \approx 0.3815$$

toteuttaa ehdon

$$\Pr(F \le 0.3815) = 0.05$$

jossa

$$F \quad F(10,60)$$

Todennäköisyyksien määrääminen *F*-jakaumasta ja tietokoneohjelmat

Olkoon

$$F \quad F(m,n)$$

Monet *tietokoneohjelmat* mahdollistavat seuraavien tehtävien ratkaisemisen ilman *F*-jakauman taulukoiden asettamia rajoituksia:

(i) Määrää todennäköisyys

$$\Pr(F \le x) = F_F(x\,;\, m, n)$$

kun x on annettu.

(ii) Määrää x, kun todennäköisyys

$$\Pr(F \leq x) = F_F(x\,;\, m, n)$$

on annettu.

18.3. *t*-jakauma

Olkoot Y ja X_i , i = 1, 2, … , n *riippumattomia, standardoitua normaalijakaumaa* N(0,1) *noudattavia satunnaismuuttujia*:

$$Y \sim \mathrm{N}(0,1)\,;\; X_i \sim \mathrm{N}(0,1)\,,\; i = 1, 2, \ldots, n$$
$$Y, X_1, X_2, \ldots, X_n \perp$$

Tällöin satunnaismuuttuja

$$X = \sum_{i=1}^{n} X_i^2$$

noudattaa χ^2*-jakaumaa* vapausastein n:

$$X \quad \chi^2(n)$$

Lisäksi satunnaismuuttujat Y ja X ovat *riippumattomia*:

$$Y \perp X$$

Määritellään satunnaismuuttuja

$$t = \frac{Y}{\sqrt{\dfrac{1}{n}X}}$$

Satunnaismuuttuja t noudattaa (**Studentin**) ***t*-jakaumaa** vapausastein (eli *parametrilla*) n:

$$t \quad t(n)$$

t-jakauman *tiheysfunktio* on muotoa

$$f(x) = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)} \left(1 + \frac{1}{n}x^2\right)^{-\frac{n+1}{2}} \,,\; -\infty < x < +\infty$$

jossa $\Gamma(\cdot)$ ns. *gamma-funktio*; lisätietoja gamma-funktiosta: ks. luvun **Jatkuvia jakaumia** kappaletta **Gamma-jakauma**. t-jakauman tiheysfunktio on johdettu luvussa **Satunnaismuuttujien muunnosten jakaumat**.

t-jakauman tunnusluvut

Olkoon

$$t \quad t(n)$$

Odotusarvo:

$$\mathrm{E}(X) = 0 \text{ , } n > 1$$

t-jakaumalla *ei ole* odotusarvoa, kun $n = 1$.

Varianssi:

$$\mathrm{Var}(t) = \mathrm{D}^2(t) = \frac{n}{n-2} \text{ , } n > 2$$

t-jakaumalla *ei ole* varianssia, kun $n = 1$ tai 2.

Standardipoikkeama:

$$\mathrm{D}(t) = \sqrt{\frac{n}{n-2}} \text{ , } n > 2$$

t-jakaumalla *ei ole* standardipoikkeamaa, kun $n = 1$ tai 2.

t-jakauman tiheysfunktio

Kuva oikealla esittää *t*-jakauman

$$t(n)$$

tiheysfunktiota välillä [–4,+4], kun vapausasteiden lukumäärällä n on seuraavat arvot:

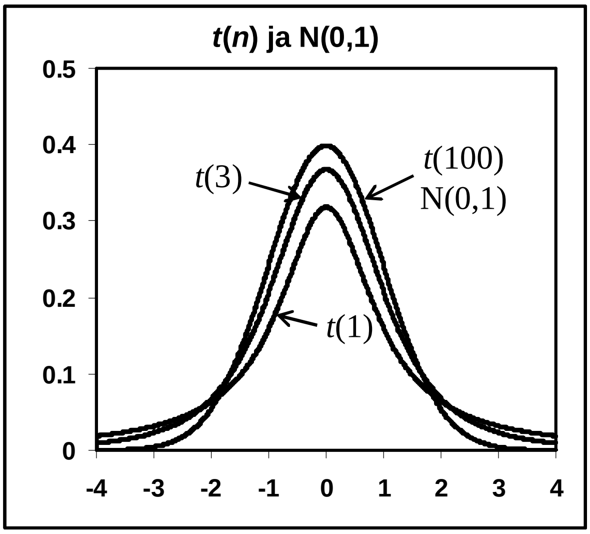
$$n = 1$$

$$n = 3$$

$$n = 100$$

Jakauman *odotusarvo*:

$$\mathrm{E}(X) = 0 \text{ , } n > 1$$



Kuvaan on piirretty myös standardoidun normaalijakauman N(0,1) tiheysfunktion kuvaaja. Huomaa, että standardoidun normaalijakauman N(0,1) tiheysfunktion kuvaaja ei erotu *t*-jakauman kuvaajasta, kun vapausteiden lukumäärä $n = 100$; ks. kohtaa ***t*-jakauma ja normaalijakauma**.

t-jakauman tiheysfunktion ominaisuudet

(i) *t*-jakauman tiheysfunktio $f(x)$ on *positiivinen* kaikille argumentin x arvoille:

$$f(x) > 0 \text{ , } -\infty < x < +\infty$$

(ii) Tiheysfunktio on *yksihuippuinen* ja sillä on *maksimi* pisteessä $x = 0$.

(iii) Tiheysfunktio on symmetrinen pisteen 0 suhteen:

$$f(-x) = f(+x) \text{ , } -\infty < x < +\infty$$

Todennäköisyyksien määrääminen *t*-jakaumasta

Todennäköisyydet *t-jakaumasta* voidaan määrätä jakauman *kertymäfunktion* avulla. Olkoon

$$t \quad t(n)$$

ja olkoon satunnaismuuttujan *t kertymäfunktio*

$$F_t(x;n) = \Pr(t \le x)$$

Kaikkien t-jakaumaan liittyvien tapahtumien todennäköisyydet saadaan todennäköisyyksistä

$$\Pr(t \le x) = F_t(x;n)$$

todennäköisyyslaskennan laskusääntöjen avulla. Siten esimerkiksi

$$\Pr(a \le t \le b) = F_t(b) - F_t(a)$$

Koska ***t*-jakauman tiheysfunktion integraalifunktiota ei pystytä esittämään suljetussa muodossa** (so. alkeisfunktioiden avulla), todennäköisyyksien määräämisessä *t*-jakaumasta on käytettävä jotakin *numeerista menetelmää*. Tämä voi tapahtua tietokoneohjelmien tai *taulukoiden* avulla.

Todennäköisyyksien määrääminen *t*-jakaumasta ja *t*-jakauman taulukot

Olkoon

$$t \quad t(n)$$

Tähän kirjaan liittyvässä taulukkokokoelmassa on taulukoituna usealle eri vapausasteiden *df* = *n* arvoille pisteet *x*, jotka toteuttavat ehdon

$$p_x = \Pr(t \ge x)$$

kun

$$p_x = 0.4,\ 0.4,\ 0.2,\ 0.1,\ 0.05,\ 0.025,\ 0.01,\ 0.005,\ 0.001,\ 0.0005$$

Siten piste *x* erottaa *t*-jakauman *oikealle* hännälle todennäköisyysmassan, jonka koko on

$$p_x = \Pr(t \ge x)$$

Koska eräs taulukoiden tärkeimmistä käyttöalueista on **tilastollinen testaus** (ks. monistetta **Tilastolliset menetelmät**), kutsutaan todennäköisyyttä p_x taulukoissa *merkitsevyystasoksi* ja pistettä *x* merkitsevyystasoa p_x vastaavaksi *kriittiseksi arvoksi* tai *rajaksi*.

Pisteet *x*, jotka erottavat todennäköisyysmassat

$$0.4,\ 0.3,\ 0.2,\ 0.1,\ 0.05,\ 0.025,\ 0.01,\ 0.005,\ 0.001,\ 0.0005$$

t-jakauman *vasemmalle* hännälle saadaan käyttämällä hyväksi sitä, että *t*-jakauma on *symmetrinen* origon suhteen.

Siten

$$p_x = \Pr(t \ge x) = \Pr(t \le -x)$$

Esimerkki 1. Todennäköisyyksien määrääminen *t*-jakauman taulukoista.

Kuva oikealla esittää *t*-jakauman

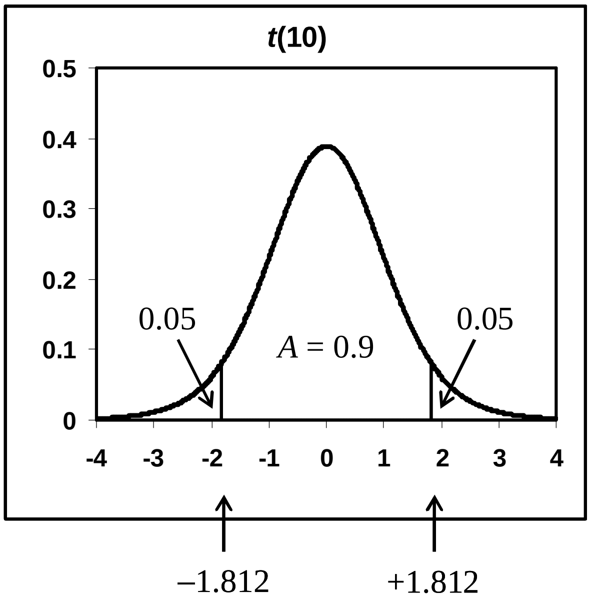
$$t(10)$$

tiheysfunktiota välillä [0, 4].

t-jakauman taulukoista saadaan:

Alueen A pinta-ala

$$\begin{aligned} &= \Pr(-1.812 \le t \le +1.812) \\ &= F_t(+1.812;10) - F_t(-1.812;10) \\ &= 0.95 - 0.05 \\ &= 0.9 \end{aligned}$$



Todennäköisyyksien määrääminen *t*-jakaumasta ja tietokoneohjelmat

Olkoon

$$t \quad t(n)$$

Monet *tietokoneohjelmat* mahdollistavat seuraavien tehtävien ratkaisemisen ilman *t*-jakauman taulukoiden asettamia rajoituksia:

(i) Määrää todennäköisyys

$$\Pr(t \le x) = F_t(x\,;n)$$

kun x on annettu.

(ii) Määrää x, kun todennäköisyys

$$\Pr(t \le x) = F_t(x\,;n)$$

on annettu.

t-jakauma ja *F*-jakauma

Suoraan *t*-jakauman ja *F*-jakauman määritelmistä nähdään seuraavaa:

Jos

$$t \quad t(n)$$

niin

$$t^2 \quad F(1,n)$$

Kääntäen, jos

$$F \quad F(1,n)$$

niin

$$\sqrt{F} \quad t(n)$$

t-jakauma ja Cauchy-jakauma

Olkoon

$$t \quad t(n)$$

jossa

$$n = 1$$

Tällöin t noudattaa **Cauchy-jakaumaa** parametrilla 0:

$$t \quad \text{Cauchy}(0)$$

Lisätietoja Cauchy-jakaumasta: ks. luvun **Jatkuvia jakaumia** kappaletta **Cauchy-jakauma**.

t-jakauma ja normaalijakauma

t-jakauma lähestyy standardoitua normaalijakaumaa, kun *vapausasteiden* lukumäärä n kasvaa. Olkoon

$$t \sim t(n)$$

Tällöin

$$\lim_{n \to +\infty} \Pr(t \le z) = \Phi(z)$$

missä Φ on *standardoidun normaalijakauman* N(0,1) *kertymäfunktio*.

Koska t-jakauma lähestyy vapausasteiden lukumäärän n kasvaessa standardoitua normaalijakaumaa N(0,1), voidaan t-jakaumaan liittyvät todennäköisyydet määrätä suurilla vapausasteiden luvuilla n standardoidun normaalijakauman avulla. *Normaalijakauma-approksimaatio t-jakaumalle* on kohtuullisen hyvä jo, kun $n = 30$, ja riittää useimpiin tarkoituksiin, kun $n > 100$.

Esimerkki 2. Normaalijakauma-approksimaation hyvyys.

Kohdan ***t*-jakauman tiheysfunktio** kuvassa jakaumien $t(100)$ ja N(0,1) tiheysfunktioiden kuvaajia ei pysty erottamaan toisistaan.

19. Moniulotteisia jakaumia

Tässä luvussa määritellään seuraavat *moniulotteiset todennäköisyysjakaumat*:

(i) **Multinomijakauma.**

(ii) **Kaksiulotteinen normaalijakauma.**

Multinomijakauma on **binomijakauman** *yleistys useamman kuin kahden toisensa poissulkevan tapahtuman tilanteeseen*.

Normaalijakauman *yleistystä* useampiulotteiseen avaruuteen kutsutaan **multinormaalijakaumaksi** tai **moniulotteiseksi normaalijakaumaksi**. Multinormaalijakaumalla on keskeinen rooli *useamman muuttujan tilastollisten aineistojen* analyysissa käytettävien tilastollisten menetelmien kuten **regressioanalyysin** tai **monimuuttujamenetelmien** *teoriassa*.

Rajoitumme tässä esityksessä tarkastelemaan multinormaalijakauman kaksiulotteista erikoistapausta **kaksiulotteista normaalijakaumaa**, mutta esitetty teoria on (suhteellisen) helppo yleistää useampiulotteisen normaalijakauman tapaukseen.

Avainsanat:

Binomijakauma, Diskreetti jakauma, Ehdollinen jakauma, Ehdollinen odotusarvo, Ehdollinen varianssi, Jatkuva jakauma, Kaksiulotteinen normaalijakauma, Korrelaatio, Korreloimattomuus, Korreloituneisuus, Kovarianssi, Kulmakerroin, Multinomijakauma, Odotusarvo, Painopiste, Pistetodennäköisyysfunktio, Regressiofunktio, Regressiosuora, Reunajakauma, Riippumattomuus, Riippuvuus, Suora, Tiheysfunktio, Todennäköisyysmassa, Varianssi, Yhteisjakauma, Yhteiskorrelaatiokerroin

19.1. Multinomijakauma

Multinomijakauma on **binomijakauman** *yleistys* useamman kuin kahden toisensa poissulkevan tapahtuman tilanteeseen; lisätietoja binomijakaumasta: ks. lukua **Diskreettejä jakaumia**.

Olkoon

$$A_1, A_2, \ldots , A_k$$

otosavaruuden *S ositus*. Tällöin

$$S = A_1 \cup A_2 \cup \cdots \cup A_k$$

$$A_i \cap A_j = \emptyset \text{ , } i \neq j$$

Oletetaan, että tapahtumien $A_1, A_2, \ldots , A_k$ *todennäköisyydet*

$$\Pr(A_i) = p_i \text{ , } i = 1, 2, \ldots , k$$

toteuttavat ehdon

$$p_1 + p_2 + \cdots + p_k = 1$$

Olkoon satunnaismuuttuja

$$X_i = \text{Tapahtuman } A_i \text{ esiintymisten lukumäärä } n\text{-kertaisessa toistokokeessa}$$

$$i = 1, 2, \ldots , k$$

Tällöin *binomijakauman* määritelmästä seuraa, että

$$X_i \sim \mathrm{Bin}(n,\ p_i) \text{ , } i = 1, 2, \ldots , k$$

jossa

$$p_i = \Pr(A_i) \text{ , } i = 1, 2, \ldots , k$$

Lisäksi satunnaismuuttujat $X_1, X_2, \ldots , X_k$ toteuttavat ehdon

$$X_1 + X_2 + \cdots + X_k = n$$

Satunnaismuuttujien

$$X_1, X_2, \ldots , X_k$$

yhteisjakaumaa kutsutaan **multinomijakaumaksi**.

Huomautus:

Satunnaismuuttuja X_i *eivät ole riippumattomia*, koska niitä sitoo toisiinsa ehto

$$X_1 + X_2 + \cdots + X_k = n$$

jossa toistokokeiden lukumäärä n on *kiinteä, ei-satunnainen* luku.

Voidaan osoittaa, että satunnaismuuttujat $X_1, X_2, \ldots , X_k$ noudattavat $(k - 1)$-ulotteista **multinomijakaumaa**, jos niiden *yhteisjakauman pistetodennäköisyysfunktio* on muotoa

$$\Pr(X_1 = n_1 \text{ ja } X_2 = n_2 \text{ ja } \cdots \text{ ja } X_k = n_k) = \frac{n!}{n_1!n_2!\cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

jossa

$$p_1 + p_2 + \cdots + p_k = 1$$
$$n_1 + n_2 + \cdots + n_k = n$$

Merkintä:

$$(X_1, X_2, \ldots , X_k) \sim \text{Multinom}(p_1, p_2 , \ldots , p_k ; n)$$

Multinomijakauman ominaisuudet

(i) Jos $k = 2$, niin multinomijakauma yhtyy *binomijakaumaan*:

$$\Pr_{Multinom}(X_1 = n_1 \text{ ja } X_2 = n - n_1) = \Pr_{Bin}(X_1 = n_1)$$

(ii) Kaikki multinomijakauman *yksiulotteiset reunajakaumat* ovat *binomijakaumia*.

(iii) *Multinomitodennäköisyydet* saadaan korottamalla *multinomi*

$$(p_1 + p_2 + \cdots + p_k)$$

potenssiin n:

$$(p_1 + p_2 + \cdots + p_k)^n = \sum \frac{n!}{n_1!n_2!\cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

jossa summa lasketaan yli kaikkien lukujen $n_1, n_2, \ldots , n_k$, jotka toteuttavat ehdon

$$n_1 + n_2 + \cdots + n_k = n$$

19.2. Kaksiulotteinen normaalijakauma

Kaksiulotteinen normaalijakauma on *normaalijakauman* (ks. lukua **Jatkuvia jakaumia**) *kaksiulotteinen yleistys*. Normaalijakauman yleistystä *p-ulotteiseen avaruuteen* ($p > 1$) kutsutaan joko ***p*-ulotteiseksi normaalijakaumaksi** tai **multinormaalijakaumaksi**. Rajoitumme tässä esityksessä pelkästään kaksiulotteisen normaalijakauman käsittelyyn.

On syytä huomata, että multinormaalijakaumalla on keskeinen asema **lineaaristen mallien** ja **monimuuttujamenetelmien** teoriassa.

Satunnaismuuttujien X ja Y muodostama pari (X, Y) noudattaa **2-ulotteista normaalijakaumaa** parametrein $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$ ja ρ_{XY}, jos sen *tiheysfunktio* on

(1)
$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} \exp\left\{-\frac{1}{2(1-\rho_{XY}^2)}Q(x,y)\right\}$$

jossa

$$Q(x,y) = \left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2$$

ja

$$-\infty < \mu_X < +\infty \qquad \sigma_X > 0$$
$$-\infty < \mu_Y < +\infty \qquad \sigma_Y > 0$$
$$-1 \le \rho_{XY} \le +1$$

Muuttujat x ja y voivat saada kaavassa (1) kaikki reaalilukuarvot:

$$-\infty < x < +\infty$$
$$-\infty < y < +\infty$$

Merkitään:

$$(X,Y) \quad \mathrm{N}_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$$

Kaksiulotteisen normaalijakauman johto

Näytämme, että **kaksiulotteinen normaalijakauma saadaan lineaarimuunnoksella kahdesta riippumattomasta standardoitua normaalijakaumaa noudattavasta satunnaismuuttujasta**.

Olkoot Z_1 ja Z_2 kaksi *riippumatonta* satunnaismuuttujaa, jotka noudattavat *standardoitua normaalijakaumaa* N(0,1):

$$Z_1 \perp Z_2$$
$$Z_1 \quad \mathrm{N}(0,1) \qquad Z_2 \quad \mathrm{N}(0,1)$$

Muodostetaan uudet satunnaismuuttujat X ja Y seuraavilla kaavoilla:

$$X = \mu_X + \sigma_X Z_1$$
$$Y = \mu_X + \sigma_Y \left[\rho_{XY} Z_1 + (1-\rho_{XY}^2)^{1/2} Z_2\right]$$

Oletetaan, että ei-satunnaiset *vakiot* $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ ja ρ_{XY} toteuttavat seuraavat ehdot:

$$-\infty < \mu_X < +\infty \qquad \sigma_X > 0$$
$$-\infty < \mu_Y < +\infty \qquad \sigma_Y > 0$$
$$-1 \le \rho_{XY} \le +1$$

Satunnaismuuttujilla X ja Y on seuraavat **odotusarvot**, **varianssit**, **kovarianssi** ja **korrelaatio**:

$$\mathrm{E}(X) = \mu_X \qquad \mathrm{E}(Y) = \mu_Y$$

$$\mathrm{Var}(X) = \sigma_X^2 \qquad \mathrm{Var}(Y) = \sigma_Y^2$$

$$\mathrm{Cov}(X,Y) = \rho_{XY}\sigma_X\sigma_Y$$

$$\mathrm{Cor}(X,Y) = \rho_{XY}$$

Lisäksi satunnaismuuttujat X ja Y ovat **normaalijakautuneita** ja satunnaismuuttujien X ja Y yhteisjakauma on **kaksiulotteinen normaalijakauma** parametrein $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$ ja ρ_{XY} :

$$(X,Y) \quad \mathrm{N}_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$$

Perustelu:

Olkoot

$$Z_1 \perp Z_2$$
$$Z_1 \quad \mathrm{N}(0,1) \qquad Z_2 \quad \mathrm{N}(0,1)$$

Tällöin

$$\mathrm{E}(Z_1) = \mathrm{E}(Z_2) = 0$$
$$\mathrm{Var}(Z_1) = \mathrm{E}(Z_1^2) = \mathrm{E}(Z_2^2) = \mathrm{Var}(Z_2) = 1$$

Määritellään satunnaismuuttujat X ja Y kaavoilla

$$X = \mu_X + \sigma_X Z_1$$
$$Y = \mu_X + \sigma_Y[\rho_{XY}Z_1 + (1-\rho_{XY}^2)^{1/2} Z_2]$$

jossa

$$-\infty < \mu_X < +\infty \qquad \sigma_X > 0$$
$$-\infty < \mu_Y < +\infty \qquad \sigma_Y > 0$$
$$-1 \le \rho_{XY} \le +1$$

(i) Johdetaan ensin satunnaismuuttujien X ja Y *tunnusluvut*.

Satunnaismuuttujien X ja Y *odotusarvot* ovat

$$\mathrm{E}(X) = \mathrm{E}\{\mu_X + \sigma_X Z_1\} = \mu_X + \sigma_X\,\mathrm{E}(Z_1) = \mu_X + \sigma_X \times 0 = \mu_X$$

ja

$$\begin{aligned}
\mathrm{E}(Y) &= \mathrm{E}\left\{\mu_Y + \sigma_Y[\rho Z_1 + (1-\rho_{XY}^2)^{1/2} Z_2]\right\} \\
&= \mu_Y + \sigma_Y[\rho\,\mathrm{E}(Z_1) + (1-\rho_{XY}^2)^{1/2}\,\mathrm{E}(Z_2)] \\
&= \mu_Y + \sigma_Y[\rho \times 0 + (1-\rho_{XY}^2)^{1/2} \times 0] = \mu_Y
\end{aligned}$$

Satunnaismuuttujien X ja Y *varianssit* ovat

$$\mathrm{Var}(X) = \mathrm{E}\left\{(X - \mathrm{E}(X))^2\right\} = \mathrm{E}\left\{\sigma_X^2 Z_1^2\right\} = \sigma_X^2\,\mathrm{E}(Z_1^2) = \sigma_X^2 \cdot 1 = \sigma_X^2$$

ja

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{E}\left\{(Y - \mathrm{E}(Y))^2\right\} \\
&= \mathrm{E}\left\{\sigma_Y^2[\rho Z_1 + (1-\rho^2)^{1/2} Z_2]^2\right\} \\
&= \sigma_Y^2\,\mathrm{E}\left\{\rho^2 Z_1^2 + (1-\rho^2) Z_2^2 + 2\rho(1-\rho^2)^{1/2} Z_1 Z_2\right\} \\
&= \sigma_Y^2\left\{\rho^2\,\mathrm{E}(Z_1^2) + (1-\rho^2)\,\mathrm{E}(Z_2^2) + 2\rho(1-\rho^2)^{1/2}\,\mathrm{E}(Z_1 Z_2)\right\} \\
&= \sigma_Y^2\left\{\rho^2 \times 1 + (1-\rho^2) \times 1 + 2\rho(1-\rho^2)^{1/2} \times 0\right\} \\
&= \sigma_Y^2
\end{aligned}$$

Satunnaismuuttujien X ja Y *kovarianssi* on

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathrm{E}\left\{(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right\} \\
&= \mathrm{E}\left\{\sigma_X Z_1 \times \sigma_Y[\rho Z_1 + (1-\rho^2)^{1/2} Z_2]\right\} \\
&= \sigma_X \sigma_Y\,\mathrm{E}\left\{\rho Z_1^2 + (1-\rho^2)^{1/2} Z_1 Z_2\right\} \\
&= \sigma_X \sigma_Y\left\{\rho\,\mathrm{E}(Z_1^2) + (1-\rho^2)^{1/2}\,\mathrm{E}(Z_1 Z_2)\right\} \\
&= \sigma_X \sigma_Y\left\{\rho \times 1 + (1-\rho^2)^{1/2} \times 0\right\} \\
&= \sigma_X \sigma_Y \rho
\end{aligned}$$

Satunnaismuuttujien X ja Y *korrelaatio* on

$$\mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_X} = \frac{\sigma_X \sigma_X \rho}{\sigma_X \sigma_X} = \rho$$

(ii) Määrätään seuraavaksi satunnaismuuttujien X ja Y *jakaumat*.

Satunnaismuuttuja X noudattaa *normaalijakaumaa*, koska se on standardoitua normaalijakaumaa noudattavan satunnaismuuttujan Z_1 *lineaarimuunnos*. Satunnaismuuttuja Y noudattaa *normaalijakaumaa*, koska se on riippumattomien standardoitua normaalijakaumaa noudattavien satunnaismuuttujien Z_1 ja Z_2 *lineaarikombinaatio*.

Siten kohdassa (i) saatujen tulosten perusteella:

$$X \quad \mathrm{N}(\mu_X, \sigma_X^2)$$

$$Y \quad \mathrm{N}(\mu_Y, \sigma_Y^2)$$

(iii) Johdetaan lopuksi satunnaismuuttujien X ja Y *yhteisjakauman* tiheysfunktio.

Koska satunnaismuuttujat Z_1 ja Z_2 ovat *riippumattomia*, satunnaismuuttujien Z_1 ja Z_2 yhteisjakauman tiheysfunktio on satunnaismuuttujien Z_1 ja Z_2 tiheysfunktioiden tulo:

$$f_{Z_1Z_2}(z_1,z_2) = f_{Z_1}(z_1)f_{Z_2}(z_2)$$
$$= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}z_1^2\right\}\times\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}z_2^2\right\} = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}(z_1^2+z_2^2)\right\}$$

Olkoon

$$\begin{cases} x = \mu_X + \sigma_X z_1 \\ y = \mu_X + \sigma_Y[\rho_{XY}z_1 + (1-\rho_{XY}^2)^{1/2}z_2] \end{cases}$$

Transformaatio

$$(z_1, z_2) \to (x, y)$$

on *bijektio*:

$$\begin{cases} z_1 = \dfrac{x-\mu_X}{\sigma_X} \\ z_2 = \dfrac{1}{(1-\rho_{XY}^2)^{1/2}}\left[\dfrac{y-\mu_Y}{\sigma_Y} - \rho_{XY}\dfrac{x-\mu_X}{\sigma_X}\right] \end{cases}$$

Muunnoksen

$$(z_1, z_2) \to (x, y)$$

Jacobin determinantti on

$$\frac{\partial(x,y)}{\partial(z_1,z_2)} = \frac{\partial x}{\partial z_1}\cdot\frac{\partial y}{\partial z_2} - \frac{\partial x}{\partial z_2}\cdot\frac{\partial y}{\partial z_1}$$
$$= \sigma_X \times \sigma_Y(1-\rho_{XY}^2)^{1/2} - 0\times\sigma_2\rho_{XY}$$
$$= \sigma_X\sigma_Y(1-\rho_{XY}^2)^{1/2}$$

Jos

$$\rho_{XY} = \pm 1$$

niin Jacobin determinantti

$$\frac{\partial(x,y)}{\partial(z_1,z_2)} = 0$$

Tällöin muunnoksen $(Z_1, Z_2) \to (X, Y)$ tuloksena saadaan ns. *singulaarinen kaksiulotteinen normaalijakauma*, jonka jätämme tässä vain maininnan varaan.

Sen sijaan jos

$$-1 < \rho_{XY} < +1$$

niin

$$\frac{\partial(x,y)}{\partial(z_1,z_2)} = \sigma_X\sigma_Y(1-\rho_{XY}^2)^{1/2} \neq 0$$

ja *satunnaismuuttujien X* ja *Y tiheysfunktio* saadaan kaavasta

$$f_{XY}(x,y) = f_{Z_1Z_2}(z_1,z_2)\left|\frac{\partial(x,y)}{\partial(z_1,z_2)}\right|^{-1}$$

Edellä on todettu, että

$$f_{Z_1Z_2}(z_1,z_2) = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}(z_1^2+z_2^2)\right\}$$

Tarkastellaan lauseketta

$$z_1^2+z_2^2$$

Sijoitetaan tähän lausekkeeseen z_1 ja z_2 lausuttuina muuttujien x ja y funktioina eli

$$\begin{cases} z_1 = \dfrac{x-\mu_X}{\sigma_X} \\ z_2 = \dfrac{1}{(1-\rho_{XY}^2)^{1/2}}\left[\dfrac{y-\mu_Y}{\sigma_Y} - \rho_{XY}\dfrac{x-\mu_X}{\sigma_X}\right] \end{cases}$$

Siten

$$\begin{aligned} z_1^2+z_2^2 &= \left(\frac{x-\mu_X}{\sigma_X}\right)^2 \\ &\quad + \frac{1}{1-\rho_{XY}^2}\left[\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho_{XY}\left(\frac{y-\mu_Y}{\sigma_Y}\right)\left(\frac{x-\mu_X}{\sigma_X}\right) + \rho_{XY}^2\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right] \\ &= \frac{1}{1-\rho_{XY}^2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{y-\mu_Y}{\sigma_Y}\right)\left(\frac{x-\mu_X}{\sigma_X}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right] \end{aligned}$$

Lisäksi

$$\left|\frac{\partial(x,y)}{\partial(z_1,z_2)}\right|^{-1} = \left|\sigma_X\sigma_Y(1-\rho_{XY}^2)^{1/2}\right|^{-1} = \frac{1}{\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}}$$

Yhdistämällä saadut tulokset saamme *satunnaismuuttujien X* ja *Y yhteisjakauman tiheysfunktion* lausekkeeksi *kaksiulotteisen normaalijakauman* $\mathrm{N}_2(\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2,\rho_{XY})$ *tiheysfunktion* lausekkeen:

$$f_{XY}(x,y) = f_{Z_1Z_2}(z_1,z_2)\left|\frac{\partial(x,y)}{\partial(z_1,z_2)}\right|^{-1} = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}}\exp\left\{-\frac{1}{2(1-\rho_{XY}^2)}Q(x,y)\right\}$$

jossa

$$Q(x,y) = \left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{y-\mu_Y}{\sigma_Y}\right)\left(\frac{x-\mu_X}{\sigma_X}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2$$

■

Kaksiulotteisen normaalijakauman tunnusluvut

Olkoon

$$(X,Y) \quad \mathrm{N}_2(\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2,\rho_{XY})$$

Odotusarvot:

$$\mathrm{E}(X)=\mu_X$$
$$\mathrm{E}(Y)=\mu_Y$$

Varianssit:

$$\mathrm{Var}(X)=\mathrm{D}^2(X)=\sigma_X^2=\mathrm{E}[(X-\mu_X)^2]$$
$$\mathrm{Var}(Y)=\mathrm{D}^2(Y)=\sigma_Y^2=\mathrm{E}[(Y-\mu_Y)^2]$$

Korrelaatio:

$$\mathrm{Cor}(X,Y)=\rho_{XY}$$

Kovarianssi:

$$\mathrm{Cov}(X,Y)=\sigma_{XY}=\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]=\rho_{XY}\sigma_X\sigma_Y$$

2-ulotteisen normaalijakauman tiheysfunktio ja sen ominaisuudet

Kaksiulotteisen normaalijakauman tiheysfunktio $f_{XY}(x,y)$ muodostaa *pinnan*

$$z=f_{XY}(x,y)$$

kolmiulotteisessa avaruudessa ^{3}. Pinnalla $z=f_{XY}(x,y)$ on *maksimi* satunnaismuuttujien X ja Y odotusarvojen μ_X ja μ_Y määräämässä pisteessä

$$(\mu_X,\mu_Y)$$

Piste (μ_X,μ_Y) määrää kaksiulotteisen normaalijakauman *todennäköisyysmassan* **painopisteen**.

2-ulotteisen normaalijakauman tiheysfunktion määrittelevässä yhtälössä (1) lauseke

$$Q(x,y)=\left(\frac{x-\mu_X}{\sigma_X}\right)^2-2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2$$

määrää pinnan $z=f_{XY}(x,y)$ muodon määräämällä sen *tasa-arvokäyrät*. Tasa-arvokäyrät ovat **ellipsejä**, joiden yhtälöt voidaan esittää muodossa

$$Q(x,y)=c^2$$

missä c on vakio. Kaksiulotteisen normaalijakauman tiheysfunktion muodostaman pinnan muodon määräävillä *tasa-arvoellipseillä* on seuraavat *ominaisuudet*:

(i) Ellipsien *keskipisteenä* on jakauman todennäköisyysmassan *painopiste*

$$(\mu_X,\mu_Y)$$

(ii) Ellipsien *eksentrisyys* on sekä korrelaatiokertoimen ρ_{XY} että standardipoikkeamien σ_X ja σ_Y funktio.

(iii) Ellipsi on *sitä eksentrisempi mitä voimakkaammin satunnaismuuttujat X ja Y ovat korreloituneita* eli mitä suurempi on

$$|\rho_{XY}|$$

(iv) Jos

$$\rho_{XY} = 0$$

ellipsien *pääakselit* ovat koordinaattiakseleiden suuntaiset.

(v) Jos

$$\rho_{XY} = 0$$

ja lisäksi

$$\sigma_X = \sigma_Y$$

niin ellipsit ovat *ympyröitä*.

(vi) Jos

$$\rho_{XY} = \pm 1$$

niin ellipsit surkastuvat *janoiksi*.

2-ulotteisen normaalijakauman odotusarvovektori ja kovarianssimatriisi

Olkoon

$$(X,Y) \quad \mathrm{N}_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$$

Määritellään 2-ulotteista normaalijakaumaa noudattavan satunnaismuuttujien *X* ja *Y* muodostaman parin (*X*,*Y*) **odotusarvovektori** $\boldsymbol{\mu}$ ja **kovarianssimatriisi** $\boldsymbol{\Sigma}$ kaavoilla

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

Huomaa, että kovarianssimatriisi $\boldsymbol{\Sigma}$ on *symmetrinen*:

$$\boldsymbol{\Sigma}' = \boldsymbol{\Sigma}$$

Odotusarvovektori $\boldsymbol{\mu} = (\mu_X, \mu_Y)$ määrää 2-ulotteisen normaalijakauman tiheysfunktion muodostaman pinnan muodon määräävien *tasa-arvoellipsien*

$$Q(x,y) = \left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 = c^2$$

keskipisteen.

2-ulotteisen normaalijakauman tiheysfunktion muodostaman pinnan muodon määräävien *tasa-arvoellipsien pääakseleiden pituudet* (oikeammin: pituuksien suhteet) ja *suunnat* saadaan määräämällä *kovarianssimatriisin*

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

ominaisarvot ja *ominaisvektorit*.

Olkoon

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

2×2-*yksikkömatriisi*.

Kovarianssimatriisin $\mathbf{\Sigma}$ **ominaisarvot** saadaan määräämällä *determinanttiyhtälön*

$$(1) \qquad |\mathbf{\Sigma} - \lambda\mathbf{I}| = \begin{vmatrix} \sigma_X^2 - \lambda & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 - \lambda \end{vmatrix} = \lambda^2 - (\sigma_X^2 + \sigma_Y^2)\lambda + \sigma_X^2\sigma_Y^2 - \sigma_{XY}^2 = 0$$

ratkaisut muuttujan λ suhteen. Ominaisarvot ovat *reaalisia*, koska kovarianssimatriisi $\mathbf{\Sigma}$ on *symmetrinen*. 2. asteen yhtälön (1) juuret saadaan kaavasta

$$\lambda = \frac{\sigma_X^2 + \sigma_Y^2 \pm \sqrt{(\sigma_X^2 + \sigma_Y^2)^2 - 4(\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2)}}{2}$$

Olkoot juuret

$$(2) \qquad \begin{aligned} \lambda_1 &= \frac{1}{2}\left[\sigma_X^2 + \sigma_Y^2 + \sqrt{(\sigma_X^2 + \sigma_Y^2)^2 - 4(\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2)}\right] \\ \lambda_2 &= \frac{1}{2}\left[\sigma_X^2 + \sigma_Y^2 - \sqrt{(\sigma_X^2 + \sigma_Y^2)^2 - 4(\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2)}\right] \end{aligned}$$

Näemme myös kaavoista (2), että ominaisarvot λ_1 ja λ_2 ovat reaalisia. Tämä johtuu siitä, että yhtälön (1) *diskriminantti*

$$D = (\sigma_X^2 + \sigma_Y^2)^2 - 4(\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2)$$

on aina *ei-negatiivinen*:

$$D = (\sigma_X^2 + \sigma_Y^2)^2 - 4(\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2) = (\sigma_X^2 - \sigma_Y^2)^2 + 4\sigma_{XY}^2 \geq 0$$

Lisäksi yhtälön (1) juurien kaavoista (2) nähdään, että

$$\lambda_1 \geq \lambda_2$$

Ominaisarvot λ_1 ja λ_2 ovat aina *ei-negatiivisia*. On helppo nähdä, että ominaisarvo λ_1 on itse asiassa aina *positiivinen*. Ominaisarvon λ_2 *ei-negatiivisuus* nähdään seuraavalla tavalla:

Koska

$$\rho_{XY}^2 = \frac{\sigma_{XY}^2}{\sigma_X^2\sigma_Y^2} \leq 1$$

niin

$$D = (\sigma_X^2 - \sigma_Y^2)^2 + 4\sigma_{XY}^2 \leq (\sigma_X^2 - \sigma_Y^2)^2 + 4\sigma_X^2\sigma_Y^2 = (\sigma_X^2 + \sigma_Y^2)^2$$

jolloin

$$\lambda_2 = \frac{1}{2}\left[\sigma_X^2 + \sigma_Y^2 - \sqrt{(\sigma_X^2 - \sigma_Y^2)^2 + 4\sigma_{XY}^2}\right] \geq \frac{1}{2}\left[\sigma_X^2 + \sigma_Y^2 - \sqrt{(\sigma_X^2 + \sigma_Y^2)^2}\right] = 0$$

Lisäksi

$$\lambda_2 = 0$$

täsmälleen silloin, kun

$$\rho_{XY}^2 = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} = 1$$

Matriisin $\mathbf{\Sigma}$ ominaisarvoja λ_1 ja λ_2 vastaavat **ominaisvektorit**

$$\mathbf{q}_1 = (q_{11}, q_{21})$$
$$\mathbf{q}_2 = (q_{12}, q_{22})$$

saadaan yhtälöistä

$$\mathbf{\Sigma}\mathbf{q}_i = \lambda_i \mathbf{q}_i \ , \ i = 1, 2$$

ottamalla huomioon ehdot

$$\mathbf{q}_i'\mathbf{q}_i = q_{1i}^2 + q_{2i}^2 = 1 \ , \ i = 1, 2$$

Siten ominaisvektori $\mathbf{q}_i$, $i = 1, 2$ saadaan ratkaistuksi yhtälöryhmästä

$$\begin{cases} (\sigma_X^2 - \lambda_i)q_{1i} + \sigma_{XY}q_{2i} = 0 \\ \sigma_{XY}q_{1i} + (\sigma_Y^2 - \lambda)q_{2i} = 0 \\ q_{1i}^2 + q_{2i}^2 = 1 \end{cases}$$

Muodostetaan matriisin $\mathbf{\Sigma}$ ominaisarvoista λ_1 ja λ_2 *diagonaalimatriisi*

$$\mathbf{L} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

ja matriisin $\mathbf{\Sigma}$ ominaisvektoreiden $\mathbf{q}_1 = (q_{11}, q_{21})$ ja $\mathbf{q}_2 = (q_{12}, q_{22})$ komponenteista matriisi

$$\mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$$

Symmetristen matriisien ominaisarvojen ja -vektoreiden yleisten ominaisuuksien perusteella tiedetään, että matriisi $\mathbf{Q}$ on *ortogonaalinen*, joten

$$\mathbf{Q}'\mathbf{Q} = \mathbf{Q}\mathbf{Q}' = \mathbf{I}$$

Matriisit $\mathbf{\Sigma}$, $\mathbf{L}$ ja $\mathbf{Q}$ toteuttavat matriisiyhtälöt

$$\mathbf{\Sigma} = \mathbf{Q}\mathbf{L}\mathbf{Q}'$$

ja

$$\mathbf{L} = \mathbf{Q}'\mathbf{\Sigma}\mathbf{Q}$$

Yhtälöä $\mathbf{\Sigma} = \mathbf{Q}\mathbf{L}\mathbf{Q}'$ kutsutaan tavallisesti matriisin $\mathbf{\Sigma}$ **pääakselihajotelmaksi**.

Tasa-arvoellipsien

$$Q(x,y) = c^2$$

pääakseleiden *pituudet* suhtautuvat toisiinsa kuten *ominaisarvojen* λ_1 ja λ_2 *neliöjuuret* ja pää-akseleiden *suunnat* yhtyvät vastaavien *ominaisvektoreiden suuntiin*.

2-ulotteisen normaalijakauman reunajakaumat

Olkoon

$$(X,Y) \quad \mathrm{N}_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$$

Satunnaismuuttujien X ja Y **reunajakaumat** ovat 1-*ulotteisia normaalijakaumia*:

$$X \quad \mathrm{N}(\mu_X, \sigma_X^2)$$
$$Y \quad \mathrm{N}(\mu_Y, \sigma_Y^2)$$

2-ulotteinen normaalijakauma, korreloimattomuus ja riippumattomuus

Olkoon

$$(X,Y) \quad \mathrm{N}_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$$

Jos

$$\rho_{XY} = 0$$

niin satunnaismuuttujat X ja Y ovat *riippumattomia*.

Perustelu:

Kaksiulotteisen normaalijakauman *tiheysfunktio* on

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} \exp\left\{-\frac{1}{2(1-\rho_{XY}^2)} Q(x,y)\right\}$$

jossa

$$Q(x,y) = \left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2$$

Oletetaan, että

$$\rho_{XY} = 0$$

Tällöin satunnaismuuttujien X ja Y yhteisjakauman tiheysfunktio hajoaa satunnaismuuttujien reunajakaumien tiheysfunktioiden tuloksi:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left\{-\frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right\} \cdot \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\}$$

$$= f_X(x)\, f_Y(y)$$

jossa $f_X(x)$ on satunnaismuuttujan X reunajakauman tiheysfunktio ja $f_Y(y)$ on satunnaismuuttujan Y reunajakauman tiheysfunktio.

Koska satunnaismuuttujien X ja Y yhteisjakauman tiheysfunktio $f_{XY}(x, y)$ on voitu esittää satunnaismuuttujien X ja Y reunajakaumien tiheysfunktioiden tulona, satunnaismuuttujat X ja Y ovat *riippumattomia*.

■

On syytä huomata, että **vaikka 2-ulotteisen normaalijakauman tapauksessa satunnaismuuttujien X ja Y korreloimattomuudesta seuraa niiden riippumattomuus**, **tämä ei ole yleisesti totta**. Sen sijaan aina pätee, että **satunnaismuuttujien riippumattomuudesta seuraa niiden korreloimattomuus**.

2-ulotteisen normaalijakauman ehdolliset jakaumat

Olkoon

$$(X, Y) \quad \mathrm{N}_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$$

2-ulotteisen normaalijakauman **ehdolliset jakaumat** ovat 1-*ulotteisia normaalijakaumia*:

$$(X \mid Y) \quad \mathrm{N}(\mu_X + \beta_{XY}(y - \mu_Y)\,,\, \sigma_X^2(1-\rho_{XY}^2))\,,\ \beta_{XY} = \rho_{XY}\frac{\sigma_X}{\sigma_Y}$$

$$(Y \mid X) \quad \mathrm{N}(\mu_Y + \beta_{YX}(x - \mu_X)\,,\, \sigma_Y^2(1-\rho_{XY}^2))\,,\ \beta_{YX} = \rho_{XY}\frac{\sigma_Y}{\sigma_X}$$

Perustelu:

Esitetään perustelu *kaksiulotteisen normaalijakauman ehdollisten jakaumien normaalisuudelle* tarkastelemalla satunnaismuuttujan X *ehdollista jakaumaa* satunnaismuuttujan Y suhteen (ehdolla $Y = y$).

Olkoon

$f_{XY}(x, y)$ = Satunnaismuuttujien X ja Y *yhteisjakauman* tiheysfunktio

$f_{X|Y}(x \mid y)$ = Satunnaismuuttujan X *ehdollisen jakauman* tiheysfunktio satunnaismuuttujan Y suhteen

$f_Y(y)$ = Satunnaismuuttujan Y *reunajakauman* tiheysfunktio

Ehdollisen jakauman tiheysfunktion määritelmän mukaan

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

Kaksiulotteisen normaalijakauman *tiheysfunktio* on muotoa

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}}\exp\left\{-\frac{1}{2(1-\rho_{XY}^2)}Q(x, y)\right\}$$

jossa

$$Q(x,y)=\left(\frac{x-\mu_X}{\sigma_X}\right)^2-2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2$$

Satunnaismuuttujan Y reunajakauman *tiheysfunktio* on muotoa

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_Y}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\}$$

Melko helposti nähdään, että

$$f_{X|Y}(x\,|\,y)=\frac{f_{XY}(x,y)}{f_Y(y)}=\frac{1}{\sqrt{2\pi\sigma_X^2(1-\rho_{XY}^2)}}\exp\left\{-\frac{1}{2\sigma_X^2(1-\rho_{XY}^2)}Q(x\,|\,y)\right\}$$

jossa

$$Q(x\,|\,y)=\left[x-\mu_X-\rho_{XY}\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right]^2$$

Siten satunnaismuuttujan X *ehdollinen jakauma* satunnaismuuttujan Y suhteen (ehdolla $Y = y$) on *normaalinen*:

$$(X\,|\,Y=y)\sim \mathrm{N}(\mu_{X|Y},\sigma_{X|Y}^2)$$

jossa

$$\mu_{X|Y}=\mathrm{E}(X\,|\,Y=y)=\mu_X+\rho_{XY}\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)$$

$$\sigma_{X|Y}^2=\mathrm{Var}(X\,|\,Y=y)=(1-\rho_{XY}^2)\sigma_X^2$$

■

2-ulotteisen normaalijakauman ehdolliset odotusarvot

Olkoon

$$(X,Y)\quad \mathrm{N}_2(\mu_X,\mu_Y,\sigma_X^2,\sigma_Y^2,\rho_{XY})$$

Satunnaismuuttujan X **ehdollinen odotusarvo** satunnaismuuttujan Y suhteen on muotoa

$$\mathrm{E}(X\,|\,Y)=\mu_X+\beta_{XY}(y-\mu_Y)$$

Satunnaismuuttujan X ehdollinen odotusarvo satunnaismuuttujan Y suhteen muodostaa satunnaismuuttujan Y arvojen y funktiona satunnaismuuttujan X **regressiofunktion** satunnaismuuttujan Y arvojen suhteen. 2-ulotteisen normaalijakauman tapauksessa satunnaismuuttujan X regressiofunktio satunnaismuuttujan Y suhteen määrää *suoran*

$$x=\mu_X+\beta_{XY}(y-\mu_Y)$$

Kutsumme tätä suoraa muuttujan x **regressiosuoraksi** muuttujan y suhteen. Suoran *kulmakerroin* on

$$\beta_{XY}=\rho_{XY}\frac{\sigma_X}{\sigma_Y}$$

ja se kulkee satunnaismuuttujien X ja Y odotusarvojen määräämän pisteen (μ_X, μ_Y) kautta.

Satunnaismuuttujan X **ehdollinen odotusarvo** satunnaismuuttujan Y suhteen on

$$\mathrm{E}(X \mid Y) = \mu_X + \beta_{XY}(y - \mu_Y)$$

Satunnaismuuttujan Y ehdollinen odotusarvo satunnaismuuttujan X suhteen muodostaa satunnaismuuttujan X arvojen x funktiona satunnaismuuttujan Y **regressiofunktion** satunnaismuuttujan X arvojen suhteen. 2-ulotteisen normaalijakauman tapauksessa satunnaismuuttujan Y regressiofunktio satunnaismuuttujan X suhteen määrää *suoran*

$$y = \mu_Y + \beta_{YX}(x - \mu_X)$$

Kutsumme tätä suoraa muuttujan y **regressiosuoraksi** muuttujan x suhteen. Suoran *kulmakerroin* on

$$\beta_{YX} = \rho_{XY}\frac{\sigma_Y}{\sigma_X}$$

ja se kulkee satunnaismuuttujien X ja Y odotusarvojen määräämän pisteen (μ_X, μ_Y) kautta.

Huomaa, että regressiosuorien kulmakertoimet β_{YX} ja β_{XY} toteuttavat yhtälön

$$\beta_{YX}\beta_{XY} = \rho_{XY}^2$$

Lisäksi *molemmat* regressiosuorat kulkevat 2-ulotteisen normaalijakauman todennäköisyysmassan painopisteen (μ_X, μ_Y) kautta.

Regressiosuorien ominaisuudet

Kirjoitetaan molempien regressiosuorien yhtälöt muuttujan x funktioina.

Muuttujan x regressiosuora muuttujan y suhteen:

$$y = \mu_Y + \frac{1}{\beta_{XY}}(x - \mu_X) = \mu_Y + \frac{1}{\rho_{XY}} \cdot \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

Muuttujan y regressiosuora muuttujan x suhteen:

$$y = \mu_Y + \beta_{YX}(x - \mu_X) = \mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

Regressiosuorilla on seuraavat ominaisuudet:

(i) Molemmat regressiosuorat kulkevat 2-ulotteisen normaalijakauman todennäköisyysmassan painopisteen (μ_X, μ_Y) kautta.

(ii) Molempien regressiosuorien kulmakertoimilla β_{YX} ja β_{XY} sekä satunnaismuuttujien X ja Y korrelaatiokertoimella ρ_{XY} on *aina sama merkki*:

Suorat ovat *nousevia*, jos

$$\rho_{XY} > 0$$

Suorat ovat *laskevia*, jos

$$\rho_{XY} < 0$$

(iii) Muuttujan *y* regressiosuora muuttujan *x* suhteen on *aina loivempi* kuin muuttujan *x* regressiosuora muuttujan *y* suhteen, koska

$$\rho_{XY}^2 \le 1$$

(iv) Muuttujan *x* regressiosuora muuttujan *y* suhteen on *sitä loivempi mitä voimakkaammin satunnaismuuttujat X ja Y ovat korreloituneita* eli mitä suurempi on

$$|\,\rho_{XY}\,|$$

(v) Muuttujan *y* regressiosuora muuttujan *x* suhteen on *sitä jyrkempi mitä voimakkaammin satunnaismuuttujat X ja Y ovat korreloituneita* eli mitä suurempi on

$$|\,\rho_{XY}\,|$$

(vi) Molemmat regressiosuorat ovat *sitä jyrkempiä mitä pienempi on satunnaismuuttujan X varianssi*

$$\sigma_X^2$$

(vii) Molemmat regressiosuorat ovat *sitä jyrkempiä mitä suurempi on satunnaismuuttujan Y varianssi*

$$\sigma_Y^2$$

(vii) Regressiosuorat *yhtyvät* täsmälleen silloin, kun

$$\rho_{XY} = \pm 1$$

(ix) Jos

$$\rho_{XY} = 0$$

niin regressiosuorat ovat *kohtisuorassa toisiaan vastaan* ja muuttujan *x* regressiosuora muuttujan *y* suhteen on

$$x = \mu_X$$

ja muuttujan *y* regressiosuora muuttujan *x* suhteen on

$$y = \mu_Y$$

Tällöin satunnaismuuttujan *X* saamat arvot *eivät riipu* satunnaismuuttujan *Y* saamista arvoista ja satunnaismuuttujan *Y* saamat arvot *eivät riipu* satunnaismuuttujan *X* saamista arvoista.

Regressiosuorien yhtälöt ja standardointi

Regressiosuorien yhtälöt voidaan kirjoittaa **standardoitujen muuttujien**

$$x' = \frac{x - \mu_X}{\sigma_X}$$

$$y' = \frac{y - \mu_Y}{\sigma_Y}$$

funktioina seuraaviin muotoihin: Muuttujan *x* regressiosuora muuttujan *y* suhteen on standardoiduissa muuttujissa *x*´ ja *y*´ muotoa

$$y' = \frac{1}{\rho_{XY}} x'$$

ja muuttujan *y* regressiosuora muuttujan *x* suhteen on standardoiduissa muuttujissa *x´* ja *y´* muotoa

$$y' = \rho_{XY} x'$$

Huomaa, että *standardoitujen muuttujien* välisten regressiosuorien kulmakertoimet ovat toistensa *käänteislukuja*.

Yhteiskorrelaatiokerroin

2-ulotteisen normaalijakauman tapauksessa satunnaismuuttujien *X* ja *Y* **yhteiskorrelaatiokerroin** on

$$R_{X|Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \rho_{XY} = \frac{\sigma_{YX}}{\sigma_Y \sigma_X} = R_{Y|X}$$

2-ulotteisen normaalijakauman ehdolliset varianssit

Satunnaismuuttujan *X* **ehdollinen varianssi** satunnaismuuttujan *Y* suhteen on

$$\sigma_{X|Y}^2 = \sigma_X^2 (1 - \rho_{XY}^2)$$

Satunnaismuuttujan *X* **ehdollinen varianssi** satunnaismuuttujan *Y* suhteen on

$$\sigma_{Y|X}^2 = \sigma_Y^2 (1 - \rho_{XY}^2)$$

Huomaa, että kumpikaan ehdollinen varianssi *ei riipu* ehtomuuttujansa arvosta.

Koska

$$|\rho_{XY}| \le 1$$

niin ehdollisten varianssien kaavoista nähdään, että

$$\sigma_{X|Y}^2 \le \sigma_X^2$$

ja

$$\sigma_{Y|X}^2 \le \sigma_Y^2$$

Esimerkki 2-ulotteisesta normaalijakaumasta

Olkoon

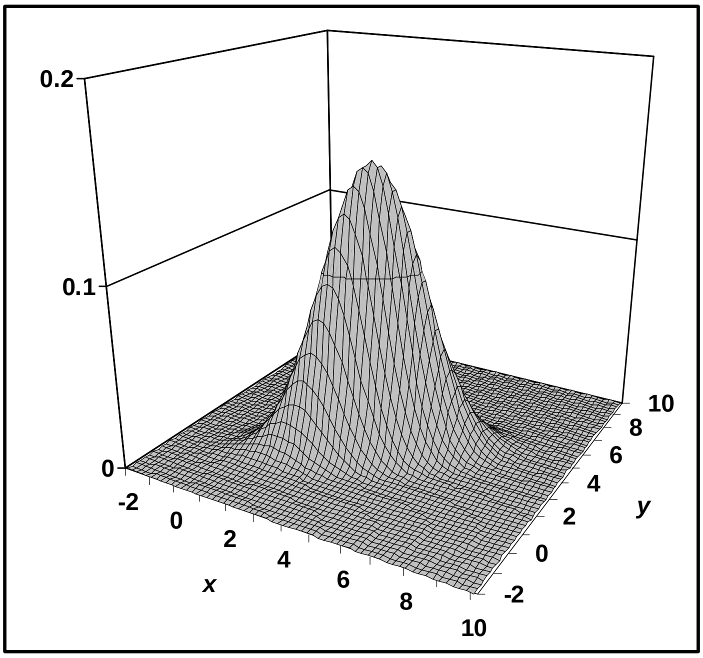
$$(X, Y) \sim \mathrm{N}_2(4, 3, 2, 1, 0.7)$$

Siten jakauman *parametrit* ovat

$$\begin{array}{ll} \mathrm{E}(X) = \mu_X = 4 & \mathrm{Var}(X) = \sigma_X^2 = 2 \\ \mathrm{E}(Y) = \mu_Y = 3 & \mathrm{Var}(Y) = \sigma_Y^2 = 1 \\ \mathrm{Cor}(X, Y) = \rho_{XY} = 0.7 & \end{array}$$

joten

$$\mathrm{Cov}(X, Y) = \rho_{XY} \sigma_X \sigma_Y = 0.7 \times \sqrt{2} \times \sqrt{1} \approx 0.9899$$



Kuva oikealla esittää esimerkin 2-ulotteisen normaalijakauman *tiheysfunktiota*

$$f_{XY}(x,y)$$

Jakauman todennäköisyysmassan *paino-pisteenä* on piste

$$(\mu_X, \mu_Y) = (4,3)$$

Odotusarvovektori ja kovarianssimatriisi

Jakauman *odotusarvovektori* on

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

Jakauman *kovarianssimatriisi* on

$$\begin{aligned}\boldsymbol{\Sigma} &= \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0.7\times\sqrt{2}\times\sqrt{1} \\ 0.7\times\sqrt{2}\times\sqrt{1} & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0.9899 \\ 0.9899 & 1 \end{bmatrix}\end{aligned}$$

Tasa-arvoellipsit

Tiheysfunktion $f_{XY}(x,y)$ määräävän pinnan muodon määräävät *tasa-arvoellipsit*

$$Q(x,y) = \left(\frac{x-4}{\sqrt{2}}\right)^2 - 2\times 0.7\left(\frac{x-4}{\sqrt{2}}\right)\left(\frac{y-3}{\sqrt{1}}\right) + \left(\frac{y-3}{\sqrt{1}}\right)^2 = c^2$$

Ellipsien *keskipisteenä* on jakauman todennäköisyysmassan *painopiste*

$$(\mu_X, \mu_Y) = (4,3)$$

Olkoon

$$\boldsymbol{\Sigma} = \mathbf{QLQ}'$$

kovarianssimatriisin $\boldsymbol{\Sigma}$ *pääakselihajotelma*, jossa **L** on matriisin $\boldsymbol{\Sigma}$ *ominaisarvojen* muodostama *diagonaalimatriisi* ja **Q** on vastaavien *ominaisvektoreiden* muodostama *ortogonaalinen matriisi*, jossa ominaisvektorit ovat sarakkeina.

Olkoot

$$\lambda_1 \geq \lambda_2$$

matriisin $\boldsymbol{\Sigma}$ *ominaisarvot* ja

$$\mathbf{q}_1 = (q_{11}, q_{21})$$
$$\mathbf{q}_2 = (q_{12}, q_{22})$$

vastaavat *ominaisvektorit*. Tällöin

$$\mathbf{L} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$$

Matriisit $\mathbf{\Sigma}$, $\mathbf{L}$ ja $\mathbf{Q}$ toteuttavat matriisiyhtälöt

$$\mathbf{\Sigma} = \mathbf{QLQ}'$$

ja

$$\mathbf{L} = \mathbf{Q}'\mathbf{\Sigma Q}$$

Olkoon λ kovarianssimatriisin $\mathbf{\Sigma}$ *ominaisarvo*. Tällöin λ toteuttaa 2. asteen yhtälön

$$\det(\mathbf{\Sigma} - \lambda \mathbf{I}_2) = \det \begin{bmatrix} \sigma_X^2 - \lambda & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 - \lambda \end{bmatrix} = \lambda^2 - (\sigma_X^2 + \sigma_Y^2)\lambda + \sigma_X^2\sigma_Y^2 - \sigma_{XY}^2 = 0$$

Tämän 2. asteen yhtälön ratkaisut saadaan kaavasta

$$\lambda = \frac{\sigma_X^2 + \sigma_Y^2 \pm \sqrt{(\sigma_X^2 - \sigma_Y^2)^2 + 4\sigma_{XY}^2}}{2}$$

Ratkaisuiksi saadaan

$$\begin{aligned} \lambda_1 &= \left[\sigma_X^2 + \sigma_Y^2 + \sqrt{(\sigma_X^2 - \sigma_Y^2)^2 + 4\sigma_{XY}^2}\right] \\ &= \frac{1}{2}\left[2 + 1 + \sqrt{(2-1)^2 + 4 \times (0.7 \times \sqrt{2})^2}\right] \\ &= \frac{1}{2}\left[3 + \sqrt{4.92}\right] \approx 2.6091 \end{aligned}$$

$$\begin{aligned} \lambda_2 &= \left[\sigma_X^2 + \sigma_Y^2 - \sqrt{(\sigma_X^2 - \sigma_Y^2)^2 + 4\sigma_{XY}^2}\right] \\ &= \frac{1}{2}\left[2 + 1 - \sqrt{(2-1)^2 + 4 \times (0.7 \times \sqrt{2})^2}\right] \\ &= \frac{1}{2}\left[3 - \sqrt{4.92}\right] \approx 0.3909 \end{aligned}$$

Olkoon $\mathbf{q} = (q_1, q_2)$ kovarianssimatriisin $\mathbf{\Sigma}$ ominaisarvoa λ vastaava *ominaisvektori*. Tällöin $\mathbf{q}$ toteuttaa *matriisiyhtälön*

$$\mathbf{\Sigma q} = \lambda \mathbf{q}$$

Koska vaadimme, että

$$\mathbf{q}'\mathbf{q} = q_1^2 + q_2^2 = 1$$

niin vektori $\mathbf{q} = (q_1, q_2)$ saadaan ratkaistuksi yhtälöryhmästä

$$\begin{cases}(\sigma_X^2-\lambda)q_1+\sigma_{XY}q_2=0\\ \sigma_{XY}q_1+(\sigma_Y^2-\lambda)q_2=0\\ q_1^2+q_2^2=1\end{cases}$$

Ominaisarvoa

$$\lambda_1 = 2.6091$$

vastaavaksi ominaisvektoriksi saadaan

$$\mathbf{q}_1 = (q_{11}, q_{21}) = (0.8517, 0.5240)$$

Ominaisarvoa

$$\lambda_2 = 0.3909$$

vastaavaksi ominaisvektoriksi saadaan

$$\mathbf{q}_2 = (q_{12}, q_{22}) = (-0.5240, 0.8517)$$

Siten kovarianssimatriisin

$$\mathbf{\Sigma} = \begin{bmatrix}\sigma_X^2 & \sigma_{XY}\\ \sigma_{XY} & \sigma_Y^2\end{bmatrix} = \begin{bmatrix}2 & 0.7\sqrt{2}\\ 0.7\sqrt{2} & 1\end{bmatrix} = \begin{bmatrix}2 & 0.9899\\ 0.9899 & 1\end{bmatrix}$$

pääakselihajotelman

$$\mathbf{\Sigma} = \mathbf{QLQ}'$$

matriiseiksi **L** ja **Q** saadaan

$$\mathbf{L} = \begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix} = \begin{bmatrix}2.6091 & 0\\ 0 & 0.3909\end{bmatrix}$$

$$\mathbf{Q} = \begin{bmatrix}q_{11} & q_{12}\\ q_{21} & q_{22}\end{bmatrix} = \begin{bmatrix}0.8517 & -0.5240\\ 0.5240 & 0.8517\end{bmatrix}$$

jossa **L** on matriisin $\mathbf{\Sigma}$ *ominaisarvojen* muodostama *diagonaalimatriisi* ja **Q** on vastaavien *ominaisvektoreiden* muodostama *ortogonaalinen matriisi*, jossa matriisin $\mathbf{\Sigma}$ vastaavat ominaisvektorit ovat sarakkeina.

Yllä esitetyn mukaan jakauman tiheysfunktion määräämän pinnan muodon määräävien tasa-arvoellipsien *pääakselit leikkaavat* jakauman todennäköisyysmassan *painopisteessä*

$$(\mu_X, \mu_Y) = (4, 3)$$

Tasa-arvoellipsien *pääakseleiden pituudet* suhtautuvat toisiinsa kuten kovarianssimatriisin $\mathbf{\Sigma}$ ominaisarvojen

$$\lambda_1 = 2.6091$$
$$\lambda_2 = 0.3909$$

neliöjuuret ja vastaavat ominaisvektorit $\mathbf{q}_1$ ja $\mathbf{q}_2$ määräävät *pääakseleiden suunnat*.

Tasa-arvoellipsien *pääakseleiden suuntaisten suorien yhtälöt* ovat

$$y = a_1 + b_1 x$$
$$y = a_2 + b_2 x$$

jossa

$$b_1 = \frac{q_{21}}{q_{11}} = \frac{0.5240}{0.8517} = 0.6152$$
$$a_1 = \mu_Y - b_1\mu_X = 3 - b_1 \times 4 = 0.5390$$

ovat suurempaa ominaisarvoa λ_1 = 2.6091 vastaavan, pitempään pääakseliin liittyvän suoran kertoimet ja

$$b_2 = \frac{q_{22}}{q_{12}} = -\frac{0.8517}{0.5240} = -1.6254$$
$$a_2 = \mu_Y - b_2\mu_X = 3 - b_2 \times 4 = 9.5015$$

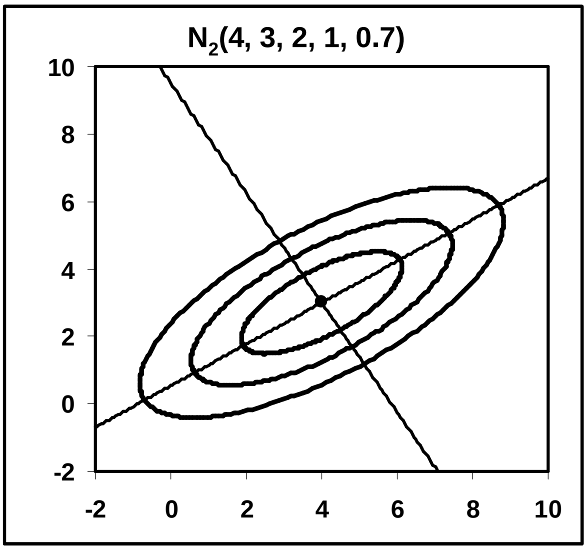
ovat pienempää ominaisarvoa λ_2 = 0.3909 vastaavan, lyhyempään pääakseliin liittyvän suoran kertoimet.

Kuva oikealla esittää esimerkin 2-ulotteisen normaalijakauman tiheysfunktion *tasa-arvoellipsejä*, jotka vastaavat (likimäärin) *todennäköisyyksiä* 68 %, 95 % ja 99.7 %.

Esimerkiksi *uloimman* ellipsin sisään jää n. 99.7 % jakauman todennäköisyysmassasta.

Kuvaan on lisäksi piirretty tasa-arvoellipsien *pääakselien suuntaiset suorat*

$$y = 0.5390 + 0.6152 \times x$$
$$y = 9.5015 - 1.6254 \times x$$

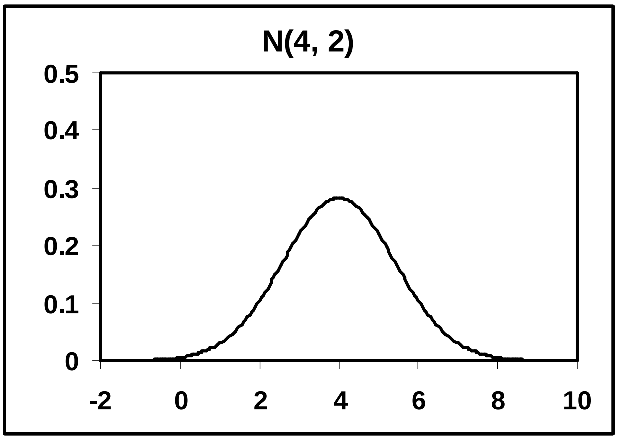


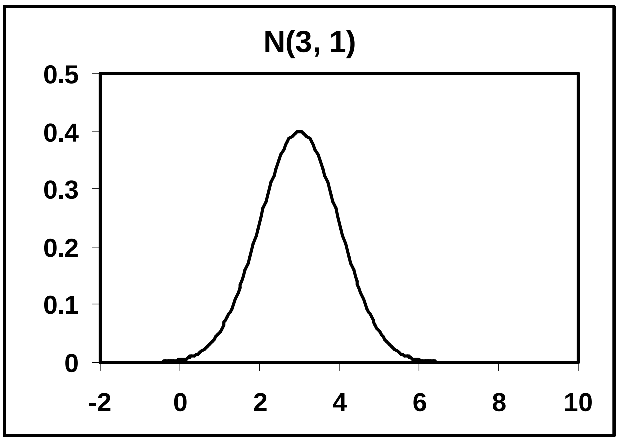
Reunajakaumat

Esimerkin 2-ulotteisen normaalijakauman *reunajakaumat* ovat seuraavat:

$X \quad \mathrm{N}(4,2) \qquad\qquad Y \quad \mathrm{N}(3,1)$

Kuvat alla esittävät reunajakaumien *tiheysfunktioita*.





Regressiosuorat

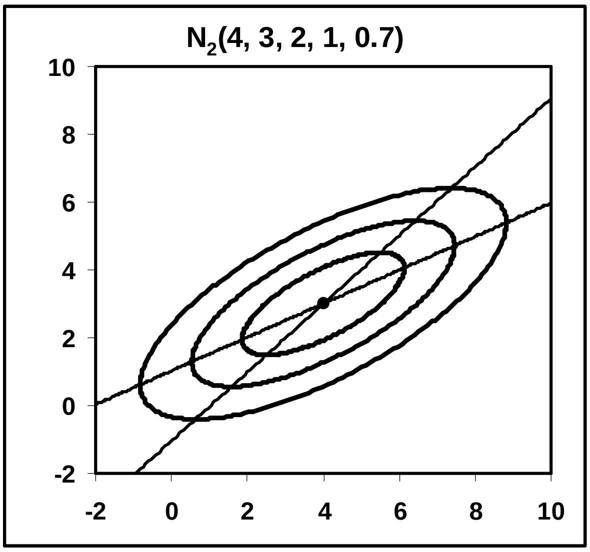
Esimerkin tapauksessa muuttujan *x regressiosuora* muuttujan *y* suhteen on

$$
\begin{aligned}
y &= \mu_Y + \frac{1}{\rho_{XY}} \times \frac{\sigma_Y}{\sigma_X}(x-\mu_X) \\
&= 3 + \frac{1}{0.7} \times \frac{1}{\sqrt{2}}(x-4) \\
&= -1.0406 + 1.0101x
\end{aligned}
$$

ja muuttujan *y regressiosuora* muuttujan *x* suhteen on

$$
\begin{aligned}
y &= \mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(x-\mu_X) \\
&= 3 + 0.7\frac{1}{\sqrt{2}}(x-4) \\
&= 1.0201 + 0.4950x
\end{aligned}
$$

Kuva oikealla esittää esimerkin 2-ulotteisen normaalijakauman tiheysfunktion *tasa-arvoellipsejä*, jotka vastaavat (likimäärin) *todennäköisyyksiä* 68 %, 95 % ja 99.7 %.



Esimerkiksi *uloimman* ellipsin sisään jää n. 99.7 % jakauman todennäköisyysmassasta.

Kuvaan on lisäksi piirretty sekä muuttujan *x* regressiosuora muuttujan *y* suhteen että muuttujan *y* regressiosuora muuttujan *x* suhteen.

Suorista *jyrkempi* eli

$$y = -1.0406 + 1.0101 \times x$$

on *x*:*n regressiosuora y*:*n suhteen* ja suorista *loivempi* eli

$$y = 1.0201 + 0.4950 \times x$$

on *y*:*n regressiosuora x*:*n suhteen*.

Ehdolliset varianssit

Esimerkin tapauksessa satunnaismuuttujan *X ehdollinen varianssi* satunnaismuuttujan *Y* suhteen on

$$0 \le \sigma^2_{X|Y} = (1-\rho^2_{XY})\sigma^2_X = (1-0.7^2) \times 2 = 1.02 \le 2 = \sigma^2_X$$

ja satunnaismuuttujan *Y ehdollinen varianssi* satunnaismuuttujan *X* suhteen on

$$0 \le \sigma^2_{Y|X} = (1-\rho^2_{XY})\sigma^2_Y = (1-0.7^2) \times 1 = 0.51 \le 1 = \sigma^2_Y$$