Johdatus todennäköisyyslaskentaan

Jatkuvia jakaumia

Jatkuvia jakaumia

Jatkuva tasainen jakauma

Eksponenttijakauma

Normaalijakauma

Keskeinen raja-arvolause

Jatkuvia jakaumia
Mitä opimme? – 1/3

- Tutustumme tässä luvussa seuraaviin **jatkuviin todennäköisyysjakaumiin**:
  - **Jatkuva tasainen jakauma**
  - **Eksponenttijakauma**
  - **Normaalijakauma**

Jatkuvia jakaumia
Mitä opimme? – 2/3

- Tarkastelun kohteena ovat seuraavat jakaumien ominaisuudet:
  (i) **Jakauman määrittely**
  (ii) **Tiheysfunktio** ja **kertymäfunktio**
  (iii) **Odotusarvo**, **varianssi** ja **standardipoikkeama**
  (iv) **Kuvaaja**
- Tarkasteltavien jakaumien **odotusarvot** *johdetaan* suoraan odotusarvon määritelmään nojautuen.
- Todennäköisyysjakauman *momentit* saadaan kuitenkin yleensä kätevimmin johdetuksi käyttämällä hyväksi jakauman **momentit generoivaa funktiota**; ks. lukua **Momenttiemäfunktio ja karakteristinen funktio**.

Jatkuvia jakaumia
Mitä opimme? – 3/3

- Tarkastelemme **normaalijakauman** tapauksessa myös ko. jakaumaa noudattavien *riippumattomien* satunnaismuuttujien **summan** jakaumaa.
- Lisätietoja riippumattomien satunnaismuuttujien summan jakauman määräämisestä: ks. lukua **Satunnaismuuttujien muunnokset ja niiden jakaumat**.
- Huomautus:

  Tarkoitamme *satunnaismuuttujien riippumattomuudella* sitä, että yhdenkään satunnaismuuttujan saamat arvot eivät riipu siitä, mitä arvoja muut satunnaismuuttujat saavat; käsite täsmennetään luvussa **Kaksiulotteiset todennäköisyysjakaumat**.
- Esitämme tässä luvussa myös **keskeisen raja-arvolauseen**, joka on tärkeimpiä perusteluita *normaalijakauman keskeiselle asemalle* tilastotieteessä.

Jatkuvia jakaumia
Esitiedot

- Esitiedot: ks. seuraavia lukuja:

  Satunnaismuuttujat ja todennäköisyysjakaumat

  Kertymäfunktio

  Jakaumien tunnusluvut

Jatkuvia jakaumia
Lisätiedot

- **Todennäköisyysjakaumien momenttien** määräämistä tarkastellaan luvussa

  Momenttiemäfunktio ja karakteristinen funktio

- **Riippumattomien satunnaismuuttujien summan jakauman** määräämistä tarkastellaan luvussa

  Satunnaismuuttujien muunnokset ja niiden jakaumat

- Tilastotieteessä paljon käytettyjä **normaalijakaumasta johdettuja jakaumia** (χ^2-, *F*- ja *t-jakaumia*) käsitellään luvussa

  Normaalijakaumasta johdettuja jakaumia

Jatkuvia jakaumia

>> **Jatkuva tasainen jakauma**
Eksponenttijakauma
Normaalijakauma
Keskeinen raja-arvolause

Jatkuva tasainen jakauma

Avainsanat

Jatkuva tasainen jakauma
Kertymäfunktio
Odotusarvo
Standardipoikkeama
Tiheysfunktio
Todennäköisyyksien määrääminen jatkuvasta tasaisesta jakaumasta
Varianssi

Jatkuva tasainen jakauma ja sen tiheysfunktio 1/3

- Olkoon satunnaismuuttujan X arvoalueena *reaaliakselin äärellinen väli* $[a, b]$.
- Olkoot $[c_1, d_1]$ ja $[c_2, d_2]$ välin $[a, b]$ kaksi *mielivaltaista*, *samanpituista* osaväliä:

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

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

$$d_1 - c_1 = d_2 - c_2$$

- Oletetaan, että väleihin $[c_1, d_1]$ ja $[c_2, d_2]$ liittyvät todennäköisyydet ovat *yhtä suuria*:

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

Jatkuva tasainen jakauma ja sen tiheysfunktio 2/3

- Satunnaismuuttujan X **tiheysfunktio** on

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

- Funktio $f(x)$ kelpaa tiheysfunktioksi, koska

$$\int_a^b f(x)dx = 1$$

Jatkuva tasainen jakauma ja sen tiheysfunktio 3/3

- Sanomme, että satunnaismuuttuja X noudattaa **jatkuvaa tasaista jakaumaa parametreinaan *a* ja *b***.
- Merkintä:

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

Jatkuva tasainen jakauma ja sen kertymäfunktio

- Olkoon $X \sim \text{Uniform}(a, b)$.
- Satunnaismuuttujan X **kertymäfunktio** on

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

Jatkuva tasainen jakauma ja sen kertymäfunktio: Johto

- Olkoon $X \sim \text{Uniform}(a, b)$.
- Satunnaismuuttujan X *tiheysfunktion* lauseke, kun $x \in [a, b]$:

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

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

$$\begin{aligned} F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(t)dt &= \int_{a}^{x} \frac{1}{b-a} dt \\ &= \left[\frac{1}{b-a} \right]_{a}^{x} \\ &= \frac{x-a}{b-a} \end{aligned}$$

Odotusarvo, varianssi ja standardipoikkeama

- Olkoon $X \sim \text{Uniform}(a, b)$.
- **Odotusarvo**:

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

- **Varianssi** ja **standardipoikkeama**:

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

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

Jatkuva tasainen jakauma
Odotusarvon johto

- Olkoon

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

- Tällöin

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

Jatkuva tasainen jakauma
Tiheysfunktion kuvaaja

- Kuva oikealla esittää jatkuvan tasaisen jakauman

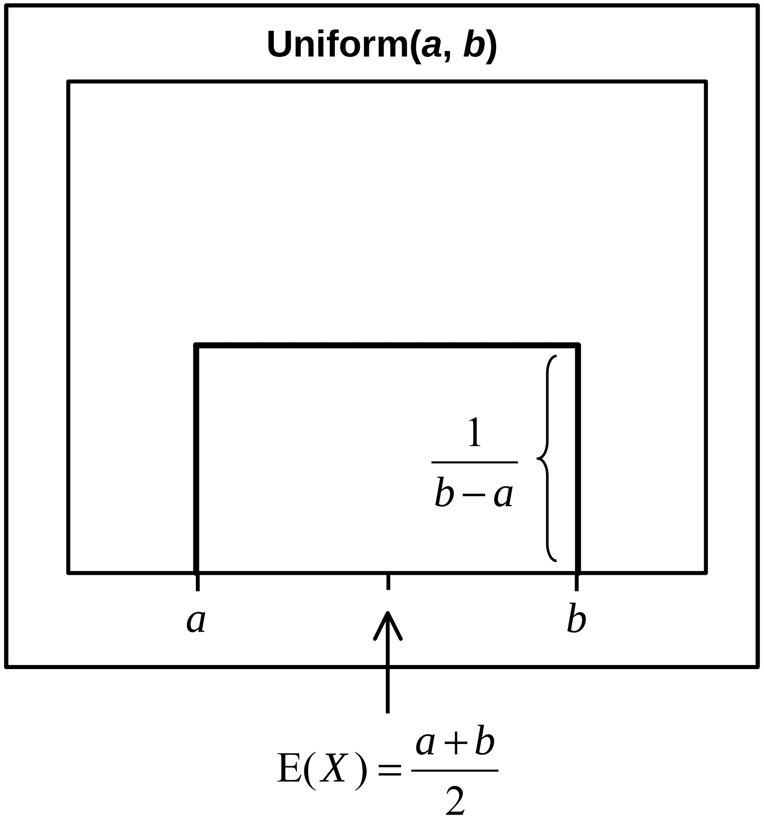
  Uniform(a, b)

  tiheysfunktiota

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

- Jakauman *odotusarvo*:

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



Tiheysfunktion ja sen kuvaajan ominaisuudet

- Jatkuvan tasaisen jakauman tiheysfunktio

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

  saa positiivisen *vakioarvon* $1/(b - a)$ välillä $[a, b]$ ja saa arvon 0 välin $[a, b]$ ulkopuolella.

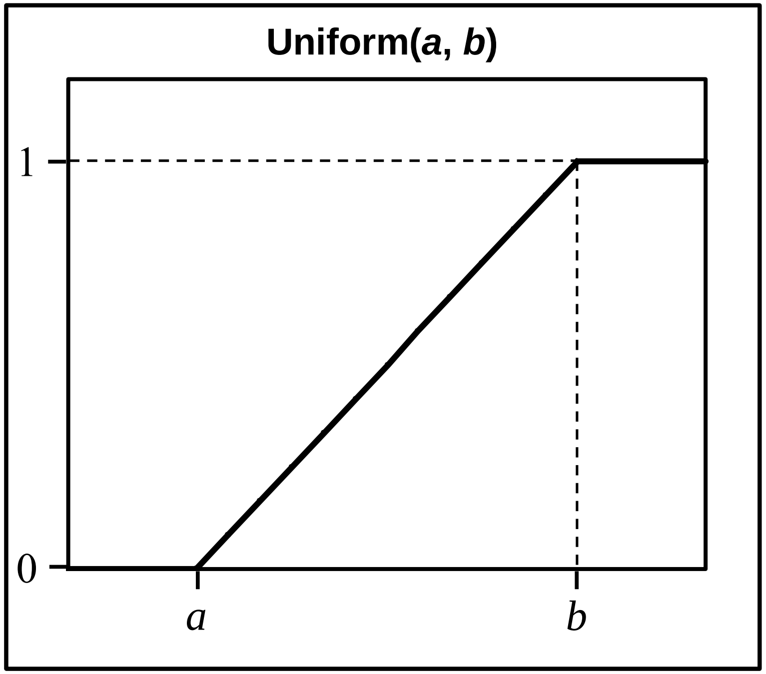
Kertymäfunktion kuvaaja

- Kuva oikealla esittää jatkuvan tasaisen jakauman

  Uniform(a, b)

  kertymäfunktiota

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



Todennäköisyyksien määrääminen jatkuvasta tasaisesta jakaumasta 1/2

- Olkoon $X \sim \text{Uniform}(a, b)$.
- Olkoon

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

jokin välin $[a, b]$ *osaväli*.

- *Välin* $[c, d]$ *todennäköisyys* saadaan integroimalla jatkuvan tasaisen jakauman Uniform(a, b) tiheysfunktio

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

välillä $[c, d]$:

$$\Pr(c \le X \le d) = \int_c^d f(x)dx = \frac{d-c}{b-a}$$

Todennäköisyyksien määrääminen jatkuvasta tasaisesta jakaumasta 2/2

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

Jatkuvia jakaumia

Jatkuva tasainen jakauma

>> Eksponenttijakauma

Normaalijakauma

Keskeinen raja-arvolause

Eksponenttijakauma

Avainsanat

Eksponenttijakauma

Kertymäfunktio

Odotusarvo

Poisson-jakauma

Standardipoikkeama

Tiheysfunktio

Todennäköisyyksien määrääminen eksponenttijakaumasta

Varianssi

Eksponenttijakauma ja sen tiheysfunktio

- Olkoon satunnaismuuttujan X **tiheysfunktio**

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

- Funktio $f(x)$ kelpaa tiheysfunktioksi, koska

$$\int_0^{+\infty} f(x)dx = 1$$

- Sanomme, että satunnaismuuttuja X noudattaa **eksponenttijakaumaa parametrinaan** λ.
- Merkintä:

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

Eksponenttijakauma ja sen kertymäfunktio

- Olkoon $X \sim \mathrm{Exp}(\lambda)$.
- Satunnaismuuttujan X **kertymäfunktio** on

$$F(x) = \Pr(X \leq x) = 1 - e^{-\lambda x}\ ,\ \lambda > 0\ ,\ x \geq 0$$

Eksponenttijakauma ja sen kertymäfunktio: Johto

- Olkoon $X \sim \mathrm{Exp}(\lambda)$.
- Satunnaismuuttujan X *tiheysfunktio*:

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

- Siten satunnaismuuttujan X *kertymäfunktioksi* saadaan, kun $x \geq 0$:

$$\begin{aligned} F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} \lambda e^{-\lambda t} dt \\ = \left[-e^{-\lambda t} \right]_0^x \\ = 1 - e^{-\lambda x} \end{aligned}$$

Eksponenttijakauma

Odotusarvo, varianssi ja standardipoikkeama

- Olkoon $X \sim \text{Exp}(\lambda)$.
- **Odotusarvo**:

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

- **Varianssi** ja **standardipoikkeama**:

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

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

Odotusarvon johto

- Olkoon

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

- Tällöin *osittaisintegroinnilla* saadaan:

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

Tiheysfunktion kuvaaja

- Kuva oikealla esittää eksponenttijakauman

  $\mathrm{Exp}(\lambda)$

  tiheysfunktiota

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

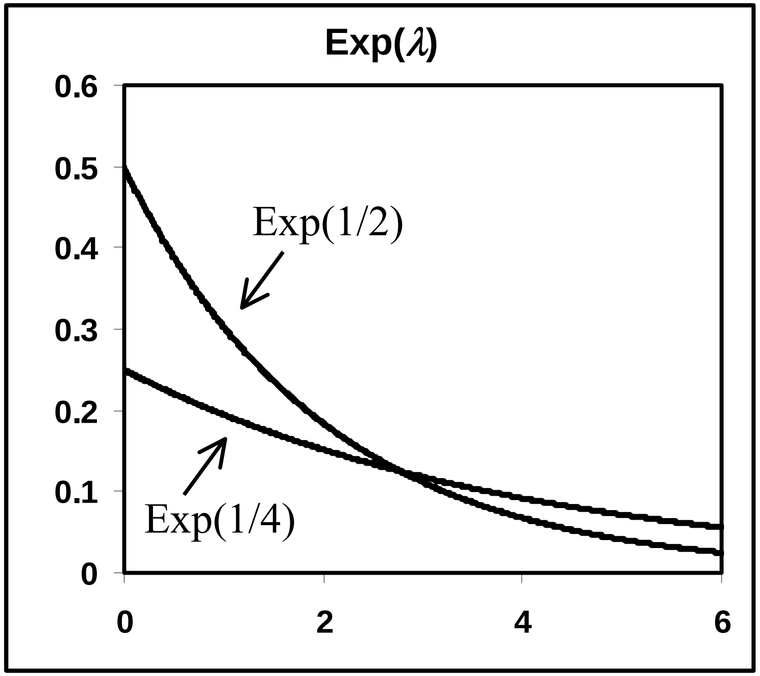
  välillä [0, 6], kun

  (i) $\lambda = 1/2$

  (ii) $\lambda = 1/4$

- Jakauman *odotusarvo*:

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



Tiheysfunktion ja sen kuvaajan ominaisuudet

- Eksponenttijakauman tiheysfunktio

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

  on *positiivinen* kaikille ei-negatiivisille argumentin arvoille:

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

- Tiheysfunktiolla on *maksimi* pisteessä

  $$x = 0$$

- Tiheysfunktio on *monotonisesti laskeva* kaikille $\lambda > 0$.

Eksponenttijakauma
Kertymäfunktion kuvaaja

- Kuva oikealla esittää eksponenttijakauman

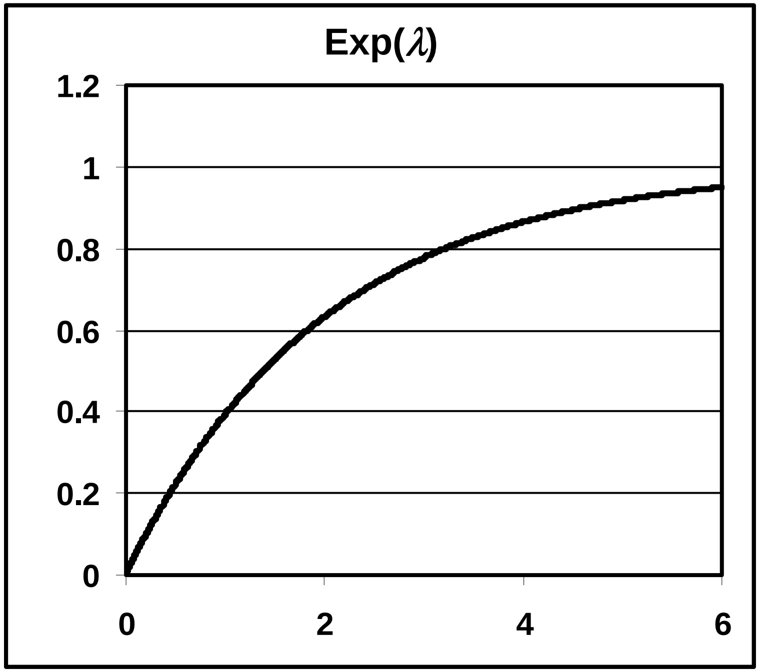
  $\mathrm{Exp}(\lambda)$

  kertymäfunktiota

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

  välillä [0, 6], kun

  $\lambda = 1/2$



Poisson-prosessi

- Tarkastellaan jonkin tapahtuman A sattumista *jatkuvalla* aikavälillä, jonka pituus on t aikayksikköä.
- Määritellään *diskreetti satunnaismuuttuja*

$$Z = \text{Niiden tapahtumien lukumäärä, jotka sattuvat aikavälillä } [0, t]$$

- Sopivin oletuksin satunnaismuuttuja Z noudattaa **Poisson-jakaumaa parametrinaan** $\boldsymbol{\nu t}$:

$$Z \sim \text{Poisson}(\nu t)$$

- Parametri νt kuvaa *tapahtumaintensiteettiä* eli *tapahtumien keskimääräistä lukumäärää aikavälillä, jonka pituus on t aikayksikköä*.

Poisson-prosessi ja eksponenttijakauma

- Olkoon

$$Z \sim \text{Poisson}(\nu t)$$

- Määritellään *jatkuva satunnaismuuttuja*

$$\begin{aligned} X &= \text{Ensimmäisen tapahtuman sattumisaika} \\ &= \text{Tapahtumien väliaika} \end{aligned}$$

- Satunnaismuuttuja X noudattaa **eksponenttijakaumaa parametrina** ν:

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

Eksponenttijakauman tiheysfunktion johto 1/2

- Olkoon

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

- Johdetaan ensin satunnaismuuttujan X *kertymäfunktio* F_X.
- Kertymäfunktion määritelmän ja komplementtitodennäköisyyden kaavan mukaan

$$(*) \quad F_X(x) = \Pr(X \leq x) = 1 - \Pr(X > x)$$

- 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 saadaan:

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

Eksponenttijakauman tiheysfunktion johto 2/2

- Sijoittamalla tämä satunnaismuuttujan X lausekkeeseen (∗) kalvolla 1/2 saadaan

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

josta satunnaismuuttujan X *tiheysfunktioksi* saadaan derivoimalla

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

- Siten

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

Eksponenttijakauman unohtamisominaisuus

- Olkoon $X \sim \mathrm{Exp}(\lambda)$.
- Tällöin

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

- Siten eksponenttijakaumalla on seuraava *unohtamis-ominaisuus*:

  Se, että tapahtuman sattumista on jouduttu odottamaan ajan b, *ei vaikuta* todennäköisyyteen joutua odottamaan ajan a lisää.

- Poisson-prosessin unohtamisominaisuutta on esimerkki stokastisen prosessin *Markov-ominaisuudesta*.

Todennäköisyyksien määrääminen eksponenttijakaumasta 1/2

- Olkoon $X \sim \text{Exp}(\lambda)$.
- Olkoon

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

jokin välin $[0, +\infty)$ *osaväli*.

- *Välin* $[c, d]$ *todennäköisyys* saadaan integroimalla eksponenttijakauman $\text{Exp}(\lambda)$ tiheysfunktio

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

välillä $[c, d]$:

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

Todennäköisyyksien määrääminen eksponenttijakaumasta 2/2

- *Kaikkien* muiden eksponenttijakaumaan Exp(λ) liittyvien tapahtumien todennäköisyydet saadaan välin [a, b] osavälien todennäköisyyksistä *todennäköisyys-laskennan laskusääntöjen* avulla.

Jatkuvia jakaumia

Jatkuva tasainen jakauma

Eksponenttijakauma

>> Normaalijakauma

Keskeinen raja-arvolause

Normaalijakauma

Avainsanat

- **Kertymäfunktio**
- **Normaalijakauma**
- **Odotusarvo**
- **Riippumattomien normaalijakautuneiden satunnaismuuttujien summan jakauma**
- **Standardipoikkeama**
- **Standardoitu normaalijakauma**
- **Tiheysfunktio**
- **Todennäköisyyksien määrääminen normaalijakaumasta**
- **Varianssi**

Normaalijakauma ja sen tiheysfunktio 1/2

- Olkoon satunnaismuuttujan X **tiheysfunktio**

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

$$-\infty < x < +\infty$$

- Funktio $f(x)$ kelpaa tiheysfunktioksi, koska

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

- Sanomme, että satunnaismuuttuja X noudattaa **normaalijakaumaa parametreinaan** μ **ja** σ^2.
- Merkintä:

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

Normaalijakauma ja sen tiheysfunktio 2/2

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

Normaalijakauma ja sen kertymäfunktio

- Olkoon $X \sim \mathrm{N}(\mu, \sigma^2)$.
- Satunnaismuuttujan X **kertymäfunktio** on

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

- Koska *normaalijakauman tiheysfunktion integraalifunktiota ei tunneta*, niin **normaalijakauman kertymäfunktiolle ei voida antaa eksplisiittistä lauseketta**.
- Siten normaalijakauman kertymäfunktion arvojen määräämiseen on käytettävä *numeerista integrointia*.

Odotusarvo, varianssi ja standardipoikkeama

- Olkoon $X \sim \mathrm{N}(\mu, \sigma^2)$.
- **Odotusarvo**:

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

- **Varianssi** ja **standardipoikkeama**:

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

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

Odotusarvon johto 1/2

- Olkoon

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

- Tällöin

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

- *Sijoituksella*

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

saadaan:

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

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

Odotusarvon johto 2/2

- Nyt

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

- Perustelu:

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

koska integroitava on 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*.

Tiheysfunktion kuvaaja

- 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], kun

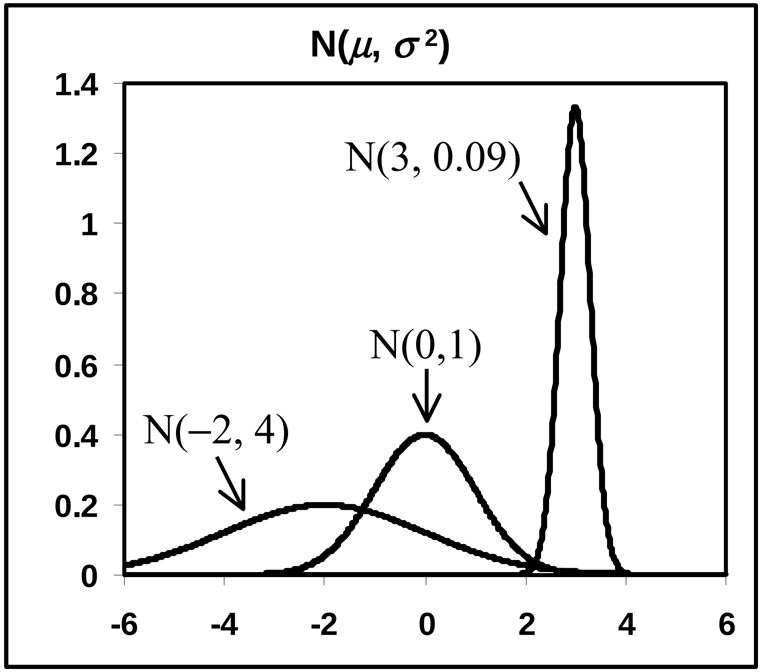
  (i) $\mu = -2 \quad \sigma^2 = 4$

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

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

- Jakauman *odotusarvo*:

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



Tiheysfunktion ja sen kuvaajan ominaisuuksia 1/3

- Normaalijakauman tiheysfunktio
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$
on kaikkialla *positiivinen*:
$$f(x) > 0 \text{ kaikille } x$$
- Tiheysfunktio on *yksihuippuinen*.
- Tiheysfunktio saa *maksimiarvonsa* pisteessä μ.
- Tiheysfunktio on *symmetrinen* suoran $x = \mu$ suhteen:
$$f(\mu - x) = f(\mu + x) \text{ kaikille } x$$

Tiheysfunktion ja sen kuvaajan ominaisuuksia 2/3

- Tiheysfunktiolla on *käännepisteet* pisteissä

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

  ja tiheysfunktio on *kupera ylöspäin* välin $[\mu - \sigma, \mu + \sigma]$ *sisäpuolella* ja *kupera alaspäin* välin $[\mu - \sigma, \mu + \sigma]$ *ulkopuolella*.
- Kaikki normaalijakaumat ovat *samanmuotoisia*, jos ne piirretään *standardoiduissa yksiköissä*

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

Normaalijakauma
Tiheysfunktion ja sen kuvaajan ominaisuuksia 3/3

- Kuva oikealla esittää normaalijakauman

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

  tiheysfunktiota

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

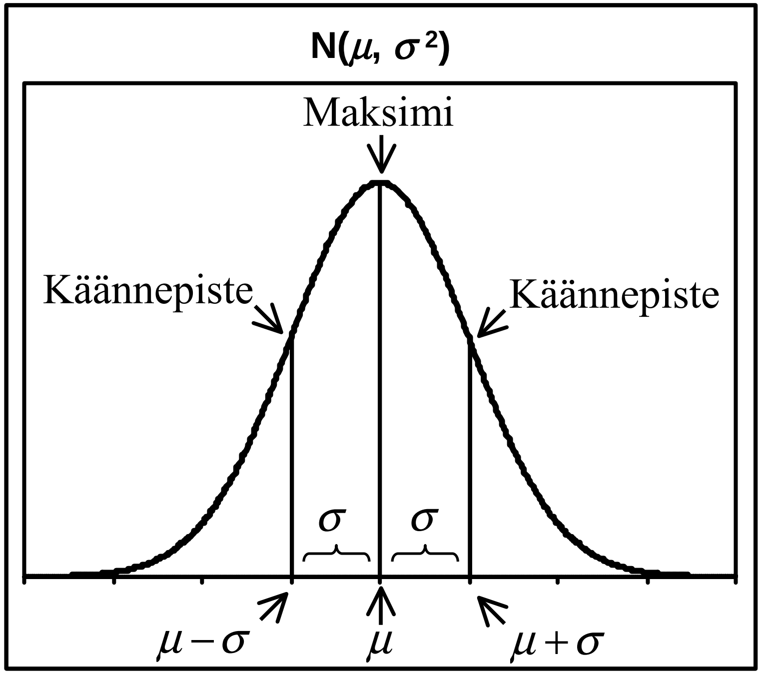
- Tiheysfunktiolla on *maksimi* pisteessä

  $$x = \mu$$

- Tiheysfunktiolla on *käännepisteet* pisteissä

  $$x = \mu - \sigma$$

  $$x = \mu + \sigma$$



68-95-99.7-sääntö

- *Kaikille* normaalijakaumille pätee (likimäärin):

(i) 68% jakauman todennäköisyysmassasta on välillä

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

(ii) 95% jakauman todennäköisyysmassasta on välillä

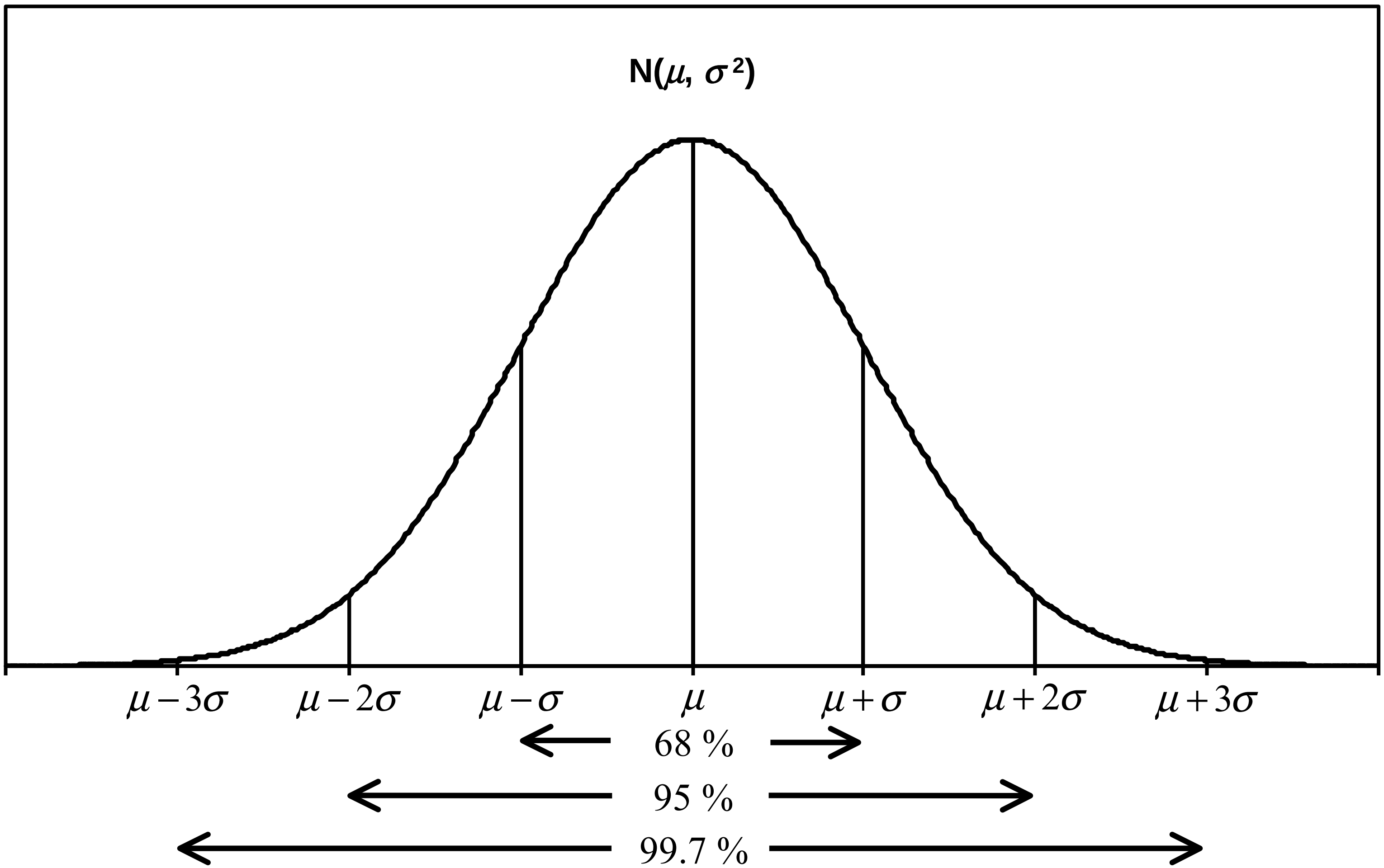
$$[\mu - 2\sigma, \mu + 2\sigma]$$

(iii) 99.7% jakauman todennäköisyysmassasta on välillä

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

Normaalijakauma

68-95-99.7-sääntö: Havainnollistus



Standardoitu normaalijakauma

- Olkoon

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

jolloin siis

$$\mathrm{E}(X) = 0$$

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

- Tällöin sanomme, että satunnaismuuttuja X noudattaa **standardoitua normaalijakaumaa**.

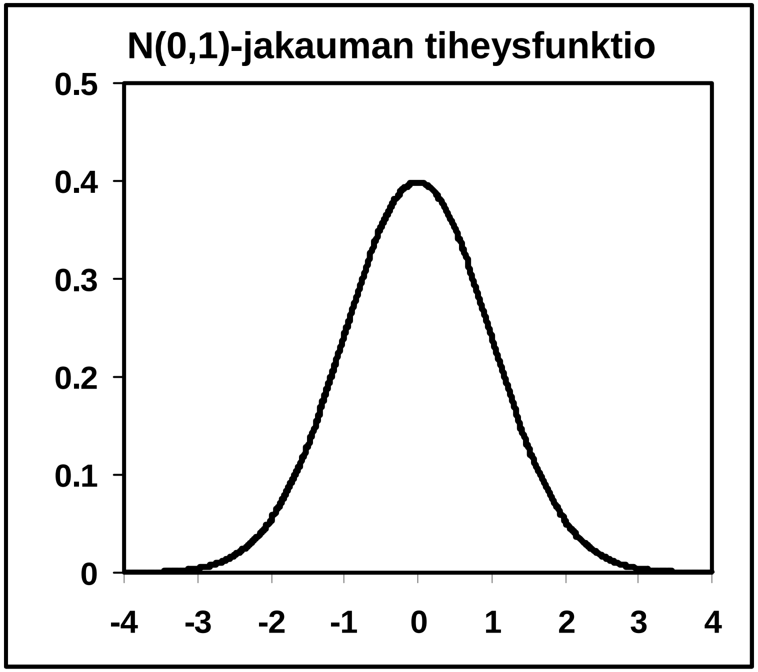
Standardoitu normaalijakauma: Tiheysfunktion kuvaaja

- Kuva oikealla esittää standardoidun normaali-jakauman

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

  tiheysfunktiota

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



Standardoitu normaalijakauma: Kertymäfunktion kuvaaja

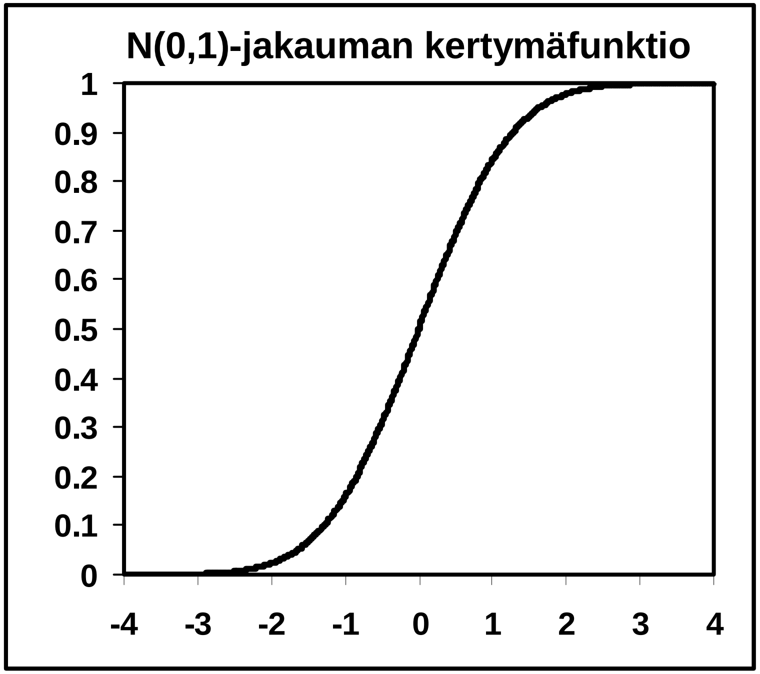
- Kuva oikealla esittää standardoidun normaali-jakauman

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

  kertymäfunktiota.
- Standardoidun normaali-jakauman kertymäfunktion $F(x)$ määrittelee kaava

  $$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

  jossa $f(x)$ on standardoidun normaalijakauman tiheysfunktio.



Lineaarimuunnoksen jakauma

- Olkoon $X \sim \mathrm{N}(\mu, \sigma^2)$.
- Määritellään satunnaismuuttuja

$$Y = a + bX$$

jossa a ja b ovat (ei-satunnaisia) *vakioita*.

- Tällöin Y on *normaalinen*:

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

- Perustelu:

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

Standardointi

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

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

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

- *Standardoidaan* satunnaismuuttuja X:

$$Z = \frac{X - \mu}{\sigma}$$

- Standardoitu satunnaismuuttuja Z noudattaa *standardoitua normaalijakaumaa* N(0,1):

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

Normaalijakauma ja standardoitu normaalijakauma 1/2

- *Kaikki* normaalijakaumat $\mathrm{N}(\mu, \sigma^2)$ ovat *samanmuotoisia standardoiduissa yksiköissä*

$$Z = \frac{X - \mu}{\sigma}$$

- Siten todennäköisyydet *mielivaltaisesta normaalijakaumasta* $\mathrm{N}(\mu, \sigma^2)$ voidaan aina määrätä *standardoidun normaalijakauman* $\mathrm{N}(0,1)$ avulla.

Normaalijakauma ja standardoitu normaalijakauma 2/2

- Olkoon siis

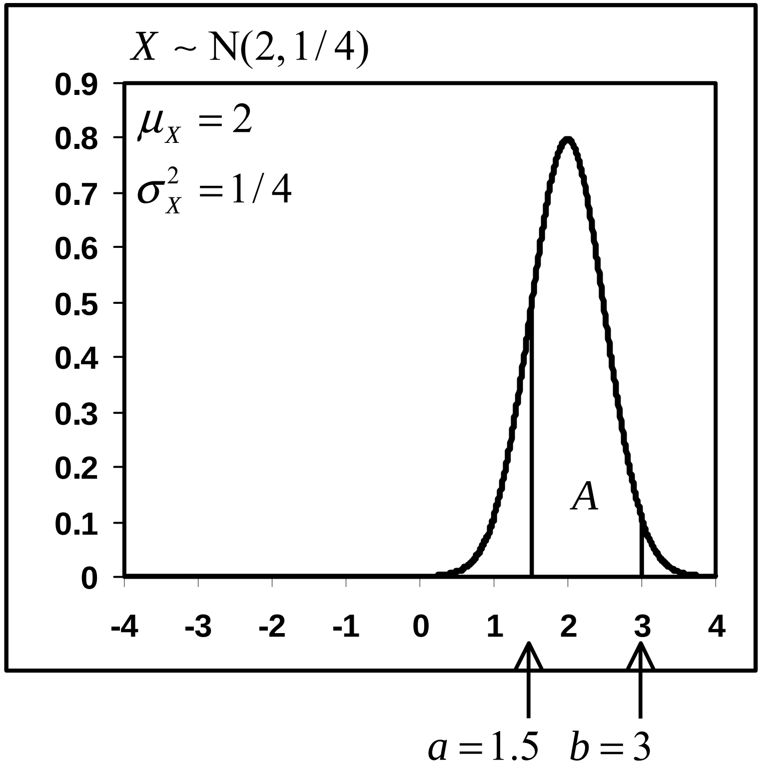
$$X \sim \mathrm{N}(\mu, \sigma^2)$$

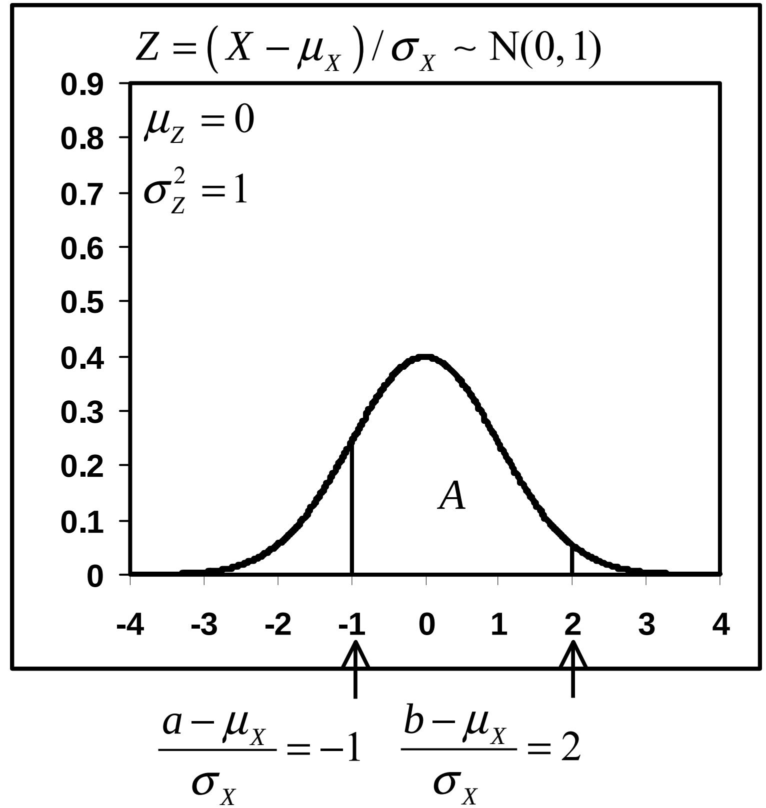
$$Z \sim \mathrm{N}(0,1)$$

- Tällöin

$$\begin{aligned}
&\Pr(a \le X \le b) \\
&= \Pr\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) \\
&= \Pr\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)
\end{aligned}$$

Normaalijakauma ja standardoitu normaalijakauma: Esimerkki





- Standardoidun normaalijakauman N(0,1) taulukoista saadaan:

$$\text{Alueen } A \text{ pinta-ala} = \Pr\left(1.5 \leq X \leq 3\right) = \Pr\left(-1 \leq Z \leq 2\right) = 0.8185$$

Todennäköisyyksien määrääminen standardoidusta normaalijakaumasta 1/2

- Todennäköisyydet *standardoidusta normaalijakaumasta* N(0,1) voidaan määrätä jakauman *kertymäfunktion* avulla.
- Olkoon $Z \sim \mathrm{N}(0,1)$.
- Olkoon satunnaismuuttujan Z *kertymäfunktio*

$$\Phi(z) = \Pr(Z \leq z)$$

- Huomautus:

Koska *normaalijakauman tiheysfunktion integraalifunktiota ei tunneta*, normaalijakauman kertymäfunktion määräämiseen on käytettävä jotakin *numeerista menetelmää*.

Siksi useimmissa tilastotieteen ja todennäköisyyslaskennan oppikirjoissa on valmis *taulukko*, jossa on taulukoituna normaalijakauman kertymäfunktion arvoja ja niihin liittyviä todennäköisyyksiä.

Todennäköisyyksien määrääminen standardoidusta normaalijakaumasta 2/2

- *Kaikkien* standardoituun normaalijakaumaan liittyvien tapahtumien todennäköisyydet saadaan todennäköisyyksistä

$$\Pr(Z \leq z) = \Phi(z)$$

todennäköisyyslaskennan laskusääntöjen avulla.

- Esimerkiksi

$$\Pr(a \leq Z \leq b) = \Phi(b) - \Phi(a)$$

Todennäköisyyksien määrääminen standardoidusta normaalijakaumasta: Taulukot 1/2

- Standardoidun normaalijakauman *taulukot* sisältävät standardoidun normaalijakauman N(0,1) *kertymäfunktion* $\Phi(z)$ *arvoja* taulukoituna usealle eri argumentin z arvolle.
- Siten 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.

Todennäköisyyksien määrääminen standardoidusta normaalijakaumasta: Taulukot 2/2

- Monissa normaalijakauman taulukoissa on taulukoitu todennäköisyyksiä

  $$\Pr(Z \leq z) = \Phi(z)$$

  vain, kun $z \geq 0$.

- Tällöin todennäköisyydet $\Pr(Z \leq -z) = \Phi(-z)$ saadaan soveltamalla standardoidun normaalijakauman tiheysfunktion *symmetrisyyttä* suoran $z = 0$ suhteen:

  $$\begin{aligned}\Phi(-z) &= \Pr(Z \leq -z) \\ &= 1 - \Pr(Z \geq -z) \\ &= 1 - \Pr(Z \leq z) \\ &= 1 - \Phi(z)\end{aligned}$$

Todennäköisyyksien määrääminen standardoidusta normaalijakaumasta: Esimerkki 1/2

- Olkoon

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

  ja olkoon

  $$f_Z(z)$$

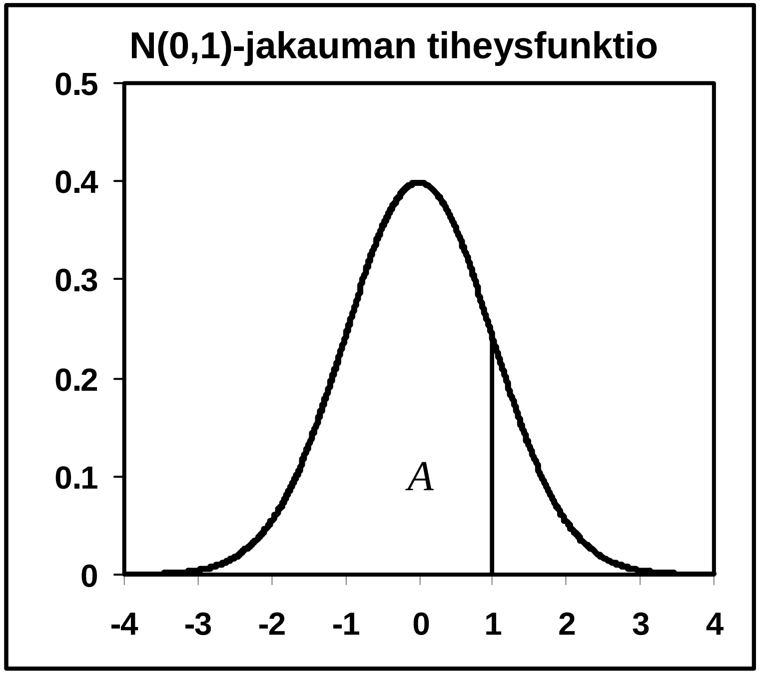
  satunnaismuuttujan Z *tiheysfunktio*.
- Standardoidun normaalijakauman N(0,1) taulukoista saadaan:

  Alueen A pinta-ala

  $$= \int_{-\infty}^{1} f_Z(z)\,dz$$

  $$= \Pr(Z \leq 1)$$

  $$= 0.8413$$



Todennäköisyyksien määrääminen standardoidusta normaalijakaumasta: Esimerkki 2/2

- Olkoon

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

ja olkoon

$$\Phi(z)$$

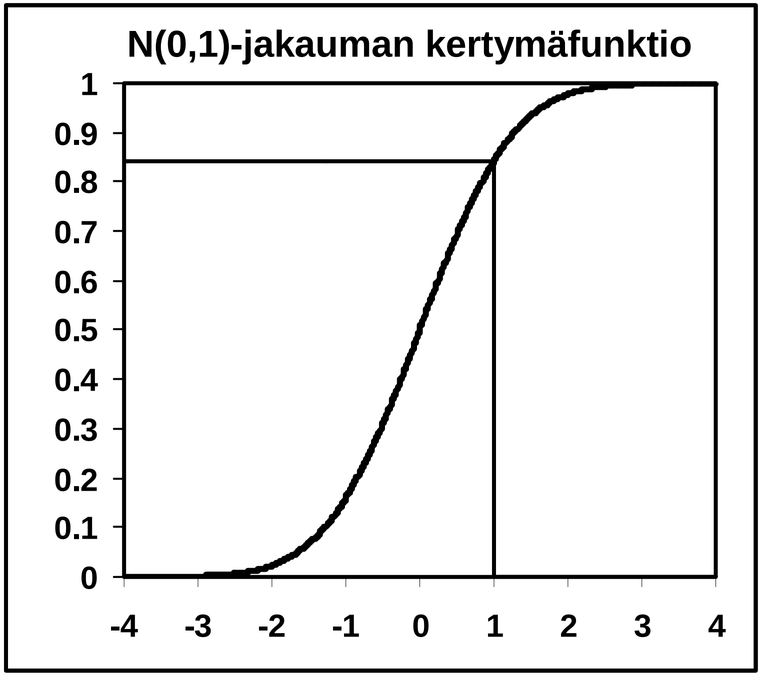
satunnaismuuttujan Z *kertymäfunktio*.

- Standardoidun normaalijakauman N(0,1) taulukoista saadaan:

$$\Phi(1)$$

$$= \Pr(Z \leq 1)$$

$$= 0.8413$$



Todennäköisyyksien määrääminen normaalijakaumasta: Ohjelmat

- Olkoon $X \sim \mathrm{N}(\mu, \sigma^2)$.
- Monet *tietokoneohjelmat* mahdollistavat seuraavien tehtävien ratkaisemisen mielivaltaisille parametrien μ, σ^2 arvoille:

(i) Määrää todennäköisyys

$$\Pr(X \leq x)$$

kun x on annettu.

(ii) Määrää x, kun todennäköisyys

$$\Pr(X \leq x)$$

kun on x annettu.

Kahden normaalijakautuneen satunnaismuuttujan summan jakauma

- Olkoon

$$X \sim \mathrm{N}(\mu_X,\ \sigma_X{}^2)$$

$$Y \sim \mathrm{N}(\mu_Y,\ \sigma_Y{}^2)$$

ja olkoot X ja Y lisäksi *riippumattomia*.

- Määritellään satunnaismuuttuja

$$W = X + Y$$

- Tällöin *summa* $W = X + Y$ on *normaalinen*:

$$W \sim \mathrm{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

- Perustelu:

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

Normaalijakautuneiden satunnaismuuttujien summan jakauma 1/2

- Olkoon X_i , $i = 1, 2, \ldots , n$ jono *riippumattomia normaalijakautuneita satunnaismuuttujia*.
- Siten

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

$$X_i \sim \mathrm{N}(\mu_i, \sigma_i^2) \text{ , } i = 1, 2, \ldots, n$$

- Olkoon

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

satunnaismuuttujien X_i , $i = 1, 2, \ldots , n$ *summa*.

Normaalijakautuneiden satunnaismuuttujien summan jakauma 2/2

- Tällöin summa Y_n on *normaalinen*:

$$Y_n \sim \mathrm{N}(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)$$

- Sanoin:

  Riippumattomien, normaalijakautuneiden satunnais-muuttujien summa on normaalinen ja parametrit saadaan yhteenlaskettavien satunnaismuuttujien vastaavien parametrien summina.

- Perustelu:

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

Normaalijakautuneiden satunnaismuuttujien summan jakauma 1/2

- Olkoon X_i , i = 1, 2, … , n jono *riippumattomia*, *samaa normaalijakaumaa noudattavia satunnaismuuttujia*.
- Siten

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

$$X_i \sim \mathrm{N}(\mu, \sigma^2)\ ,\ i = 1, 2, \ldots, n$$

- Olkoon

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

satunnaismuuttujien X_i , i = 1, 2, … , n *summa*.

Normaalijakautuneiden satunnaismuuttujien summan jakauma 2/2

- Tällöin summa Y_n on *normaalinen*:

$$Y_n = \sum_{i=1}^{n} X_i \sim \mathrm{N}(n\mu, n\sigma^2)$$

- Siten *riippumattomien, samaa normaalijakaumaa noudattavien satunnaismuuttujien summa on normaalinen* ja parametrit saadaan yhteenlaskettavien satunnaismuuttujien vastaavien parametrien summina.
- Huomautus:

  Tulos on *erikoistapaus* riippumattomien, normaalijakautuneiden satunnaismuuttujien summaa koskevasta yleisestä jakaumatuloksesta.

Normaalijakautuneiden satunnaismuuttujien aritmeettisen keskiarvon jakauma 1/2

- Olkoon X_i , i = 1, 2, … , n jono *riippumattomia*, *samaa normaalijakaumaa noudattavia satunnaismuuttujia*.
- Siten

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

$$X_i \sim \mathrm{N}(\mu, \sigma^2)\ ,\ i = 1, 2, \ldots, n$$

- Olkoon

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

satunnaismuuttujien X_i , i = 1, 2, … , n *aritmeettinen keskiarvo*.

Normaalijakautuneiden satunnaismuuttujien aritmeettisen keskiarvon jakauma 2/2

- Tällöin aritmeettinen keskiarvo $\bar{X}$ on *normaalinen*:

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

- Siten *riippumattomien, samaa normaalijakaumaa noudattavien satunnaismuuttujien aritmeettinen keskiarvo on normaalinen*.
- Huomautus:

Ilman normaalisuusoletustakin pätee:

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

$$\mathrm{D}^2(\bar{X}) = \frac{\sigma^2}{n}$$

Miksi normaalijakauma on ”normaali”?

- **Normaalijakauma on sekä *teoreettisen* että *soveltavan tilastotieteen* tärkein jakauma**.
- Normaalijakauman keskeinen asema tilastotieteessä perustuu siihen *teoreettiseen* ja *empiiriseen* tosiseikkaan, että moniin satunnaisilmiöihin liittyvät satunnaismuuttujat noudattavat ainakin *approksimatiivisesti* normaalijakaumaa.
- Mikä on tämän tosiseikan selitys?
- Selityksenä on **keskeinen raja-arvolause**; ks. seuraavaa kappaletta.

Jatkuvia jakaumia

Jatkuva tasainen jakauma

Eksponenttijakauma

Normaalijakauma

>> Keskeinen raja-arvolause

Keskeinen raja-arvolause

Avainsanat

Approksimointi
Asymptoottinen
Aritmeettinen keskiarvo
Binomijakauma
De Moivren ja Laplacen raja-arvolause
Hypergeometrinen jakauma
Kertymäfunktio
Keskeinen raja-arvolause
Normaalijakauma
Poisson-jakauma
Riippumattomien normaalijakautuneiden satunnaismuuttujien summan jakauma
Standardointi
Standardoitu normaalijakauma
Tiheysfunktio

Johdanto 1/2

- Olkoon X_i , $i = 1, 2, \dots, n$ jono *riippumattomia*, *samaa normaalijakaumaa* $\mathrm{N}(\mu, \sigma^2)$ *noudattavia satunnais-muuttujia*.
- Tällöin *satunnaismuuttujien* X_i *summa* Y_n on normaalinen:

$$Y_n = \sum_{i=1}^{n} X_i \sim \mathrm{N}(n\mu, n\sigma^2)$$

- Kysymys:

  Mitä voidaan sanoa *riippumattomien*, *samaa jakaumaa noudattavien satunnaismuuttujien summan jakaumasta*, jos ko. satunnaismuuttujat *eivät noudata normaali-jakaumaa*?

Johdanto 2/2

- *Ei-normaalisten* satunnaismuuttujien summa *ei yleensä ole* normaalinen.
- Kuitenkin, 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.

Keskeisen raja-arvolauseen formulointi 1/3

- Olkoon X_i , i = 1, 2, … jono *riippumattomia*, *samoin jakautuneita satunnaismuuttujia*, joiden odotusarvo ja varianssi ovat

$$\mathrm{E}(X_i) = \mu \ , \ i = 1, 2, \ldots$$

$$\mathrm{D}^2(X_i) = \sigma^2 \ , \ i = 1, 2, \ldots$$

- Olkoon

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

satunnaismuuttujien X_i , i = 1, 2, … , n *summa*.

Keskeisen raja-arvolauseen formulointi 2/3

- Summan Y_n odotusarvo ja varianssi ovat

$$\mathrm{E}(Y_n) = n\mu$$

$$\mathrm{D}^2(Y_n) = n\sigma^2$$

- *Standardoidaan* summa Y_n :

$$Z_n = \frac{Y_n - n\mu}{\sigma\sqrt{n}}$$

- Annetaan $n \to +\infty$
- Tällöin satunnaismuuttujan Z_n *jakauma lähestyy standardoitua normaalijakaumaa* N(0,1).

Keskeisen raja-arvolauseen formulointi 3/3

- Siten **keskeinen raja-arvolause** sanoo, että

$$\lim_{n\to+\infty} \Pr\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \leq z \right) = \Phi(z)$$

jossa Φ on *standardoidun normaalijakauman* N(0,1) *kertymäfunktio*.

- Merkintä:

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \sim_a \mathrm{N}(0,1)$$

Kommentteja 1/3

- Keskeiselle raja-arvolauseelle esitetään *todistus* luvussa
  Konvergenssikäsitteet ja raja-arvolauseet.
- Keskeisen raja-arvolauseen mukaan usean satunnais-muuttujan *summa on* (*tietyin ehdoin*) *approksimatiivisesti normaalinen* (*lähes*) *riippumatta yhteenlaskettavien jakaumasta*.
- Huomautus:

  Yhteenlaskettavien ei tarvitse olla edes *jatkuvia*, vaan ne voivat olla jopa *diskreettejä*.

Kommentteja 2/3

- Approksimaation *hyvyys riippuu* yhteenlaskettavien satunnaismuuttujien lukumäärästä, niiden jakaumasta ja erityisesti niiden jakauman *vinoudesta*.
- Approksimaation *hyvyys paranee*, kun yhteenlaskettavien satunnaismuuttujien *lukumäärä kasvaa*.
- Jos yhteenlaskettavien satunnaismuuttujien jakauma on *symmetrinen*, approksimaatio on hyvä jo suhteellisen pienillä yhteenlaskettavien lukumäärillä.
- Jos yhteenlaskettavien satunnaismuuttujien jakauma on *epäsymmetrinen*, hyvä approksimaatio vaatii enemmän yhteenlaskettavia.

Kommentteja 3/3

- 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. *jakauma-konvergenssista* eli *heikosta konvergenssista*.
- Keskeisestä raja-arvolauseesta on olemassa *yleisempiä muotoja*, joissa lievennetään *samoinjakautuneisuus*- ja *riippumattomuusoletuksia*.

Keskeinen raja-arvolause

Aritmeettisen keskiarvon approksimatiivinen jakauma

- Keskeisestä raja-arvolauseesta seuraa:

  Riippumattomien samoin jakautuneiden satunnaismuuttujien X_i , $i = 1, 2, \ldots , n$ **aritmeettinen keskiarvo**

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

  on suurille (mutta äärellisille) n *approksimatiivisesti normaalinen parametreinaan* μ *ja* σ^2/n:

$$\bar{X}_n \sim_a \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Keskeisen raja-arvolauseen seurauksia 1/3

- Keskeisellä raja-arvolauseesta seuraa erikoistapauksina monet yksittäisiä jakaumia koskevat asymptoottiset tulokset.
- Käsittelemme seuraavia erikoistapauksia:

(i) **Binomijakauma lähestyy normaalijakaumaa**, kun toistokokeiden lukumäärän n annetaan kasvaa.

(ii) **Poisson-jakauma lähestyy normaalijakaumaa**, kun jakauman intensiteettiparametrin λ arvon annetaan kasvaa.

Keskeisen raja-arvolauseen seurauksia 2/3

- Sitä, että *binomijakauma lähestyy* toistokokeiden lukumäärän *n* kasvaessa *normaalijakaumaa*, kutsutaan tavallisesti **De Moivren ja Laplacen raja-arvolauseeksi**.
- De Moivren ja Laplacen raja-arvolauseen mukaan *binomitodennäköisyyksiä voidaan approksimoida normaalijakaumasta määrätyillä todennäköisyyksillä*, jos toistokokeiden lukumäärä on kyllin suuri.
- Koska hypergeometrinen jakauma muistuttaa tietyin ehdoin binomijakaumaa, myös *hypergeometrisen jakauman todennäköisyyksiä voidaan approksimoida normaalijakaumasta määrätyillä todennäköisyyksillä*.

Keskeisen raja-arvolauseen seurauksia 3/3

- Poisson-jakaumaa koskevan keskeisen raja-arvolauseen muodon mukaan *Poisson-jakauman todennäköisyyksiä voidaan approksimoida normaalijakaumasta määrätyillä todennäköisyyksillä*.

De Moivren ja Laplacen raja-arvolause

- Olkoon $X \sim \mathrm{Bin}(n, p)$ ja $q = 1 - p$.
- Siten

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

$$\mathrm{Var}(X) = npq$$

- Tällöin

$$\lim_{n \to +\infty} \Pr\left(\frac{X - np}{\sqrt{npq}} \le z \right) = \Phi(z)$$

jossa Φ on *standardoidun normaalijakauman* N(0,1) *kertymäfunktio*.

De Moivren ja Laplacen raja-arvolause: Perustelu 1/2

- Olkoon $X \sim \mathrm{Bin}(n, p)$.
- Tällöin satunnaismuuttuja X voidaan esittää *riippumattomien*, *samaa Bernoulli-jakaumaa* noudattavien satunnaismuuttujien X_i summana:

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

jossa

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

- Koska

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

$$\mathrm{Var}(X_i) = npq\,,\ q = 1 - p$$

niin

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

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = npq$$

De Moivren ja Laplacen raja-arvolause: Perustelu 2/2

- Tällöin *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*.

De Moivren ja Laplacen raja-arvolause: Havainnollistus 1/5

- Kalvoilla 2/5-5/5 oleva kuvasarja havainnollistaa De Moivren ja Laplacen raja-arvolausetta.
- Kuvasarja näyttää miten satunnaismuuttujien

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

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

  jakaumat alkavat muistuttaa yhä enemmän toisiaan, kun toistokokeiden lukumäärän n annetaan kasvaa.
- Kuvasarjassa

  $$p = 0.1$$

  $$n = 1, 10, 30, 100$$

  $$\mu = np$$

  $$\sigma^2 = np(1-p)$$

De Moivren ja Laplacen raja-arvolause: Havainnollistus 2/5

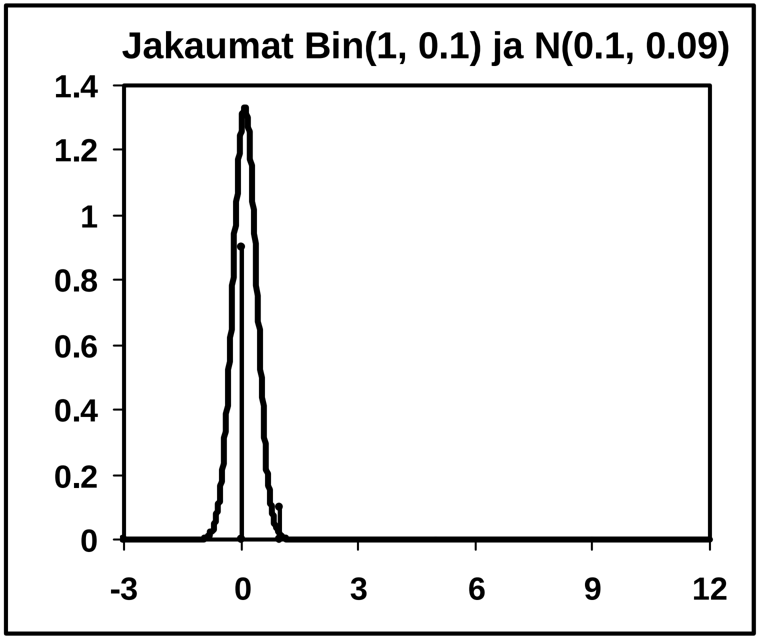
- Olkoon

$$X \sim \mathrm{Bin}(n, p)$$
$$n = 1$$
$$p = 0.1$$

ja

$$Z \sim \mathrm{N}(\mu, \sigma^2)$$
$$\mu = np = 0.1$$
$$\sigma^2 = np(1-p) = 0.09$$

- Kuva oikealla esittää satunnaismuuttujan X *pistetodennäköisyysfunktiota* ja satunnaismuuttujan Z *tiheysfunktiota* välillä [–3, 12].



De Moivren ja Laplacen raja-arvolause: Havainnollistus 3/5

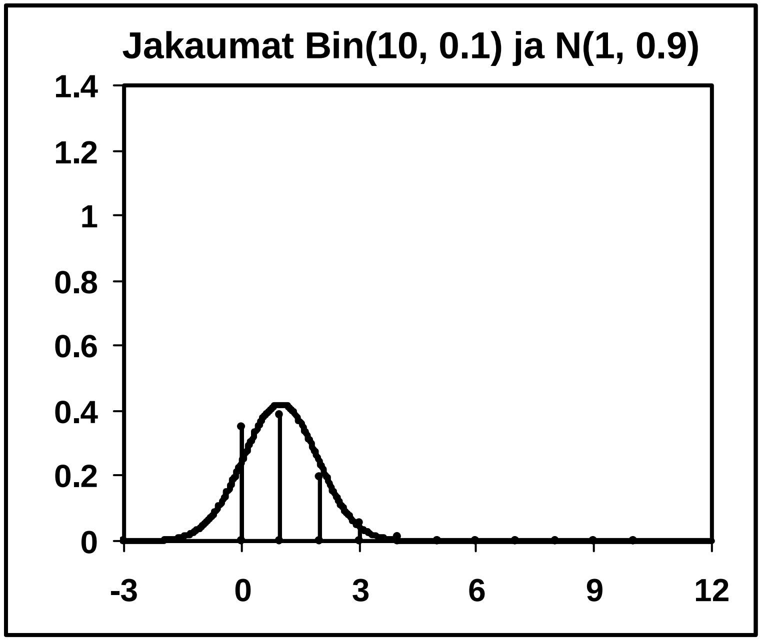
- Olkoon

$$X \sim \mathrm{Bin}(n, p)$$
$$n = 10$$
$$p = 0.1$$

ja

$$Z \sim \mathrm{N}(\mu, \sigma^2)$$
$$\mu = np = 1$$
$$\sigma^2 = np(1-p) = 0.9$$

- Kuva oikealla esittää satunnaismuuttujan X *pistetodennäköisyysfunktiota* ja satunnaismuuttujan Z *tiheysfunktiota* välillä [–3, 12].



De Moivren ja Laplacen raja-arvolause: Havainnollistus 4/5

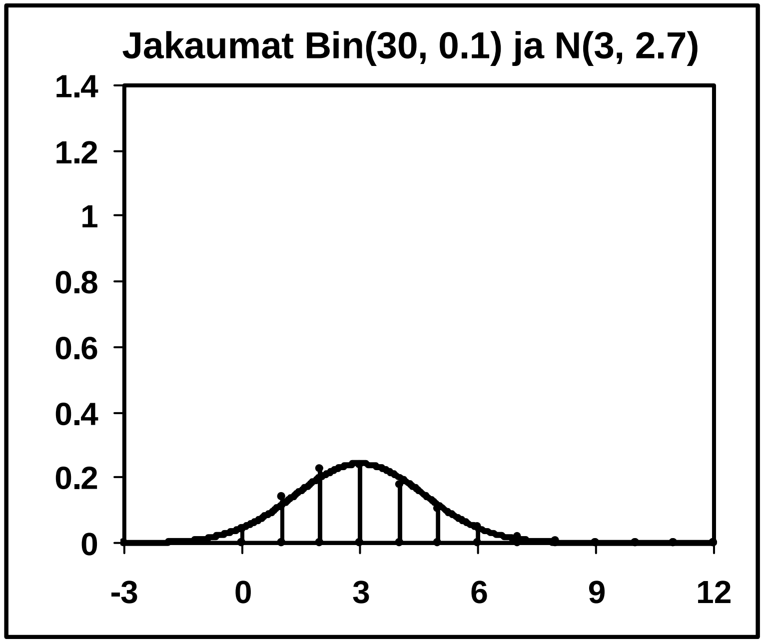
- Olkoon

$$X \sim \mathrm{Bin}(n, p)$$
$$n = 30$$
$$p = 0.1$$

ja

$$Z \sim \mathrm{N}(\mu, \sigma^2)$$
$$\mu = np = 3$$
$$\sigma^2 = np(1-p) = 2.7$$

- Kuva oikealla esittää satunnaismuuttujan X *pistetodennäköisyysfunktiota* ja satunnaismuuttujan Z *tiheysfunktiota* välillä $[-3, 12]$.



De Moivren ja Laplacen raja-arvolause: Havainnollistus 5/5

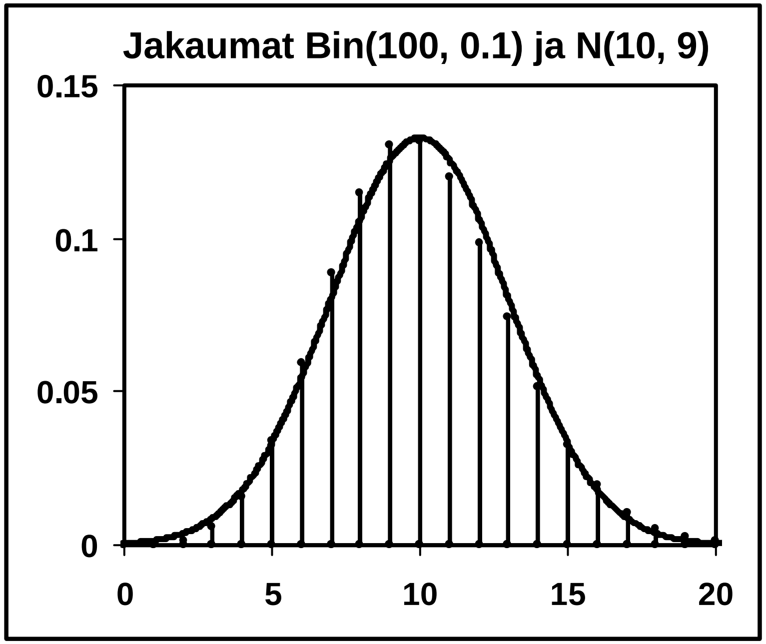
- Olkoon

$$X \sim \mathrm{Bin}(n, p)$$
$$n = 100$$
$$p = 0.1$$

ja

$$Z \sim \mathrm{N}(\mu, \sigma^2)$$
$$\mu = np = 10$$
$$\sigma^2 = np(1-p) = 9$$

- Kuva oikealla esittää satunnaismuuttujan X *pistetodennäköisyysfunktiota* ja satunnaismuuttujan Z *tiheysfunktiota* välillä [0, 20].



Binomitodennäköisyydet ja normaalijakauma 1/4

- De Moivren ja Laplacen raja-arvolauseen mukaan *binomijakaumaa*

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

  voidaan suurille n *approksimoida normaalijakaumalla*

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

  jossa

  $$\mu = np$$

  $$\sigma^2 = npq \text{ , } q = 1 - p$$

Binomitodennäköisyydet ja normaalijakauma 2/4

- Jos siis

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

niin De Moivren ja Laplacen raja-arvolauseen mukaan suurille n

$$\Pr(a < X \leq b) \approx \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right)$$

jossa Φ on standardoidun normaalijakauman N(0,1) kertymäfunktio.

Binomitodennäköisyydet ja normaalijakauma 3/4

- Jos *a* ja *b* ovat kokonaislukuja, approksimaatio *on* hieman *parempi*, jos käytetään kaavaa

$$\Pr(a < X \leq b) \approx \Phi\left(\frac{b + 1/2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - 1/2 - np}{\sqrt{npq}}\right)$$

- *Korjaustekijä* 1/2 perustuu siihen, että *diskreettiä* binomijakaumaa approksimoidaan *jatkuvalla* normaalijakaumalla.

Binomitodennäköisyydet ja normaalijakauma 4/4

- 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 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 on *binomijakauman pistetodennäköisyysfunktio*.

Binomitodennäköisyydet ja normaalijakauma: Esimerkki 1/3

- Kuva oikealla esittää jakauman

  Bin(100, 0.1)

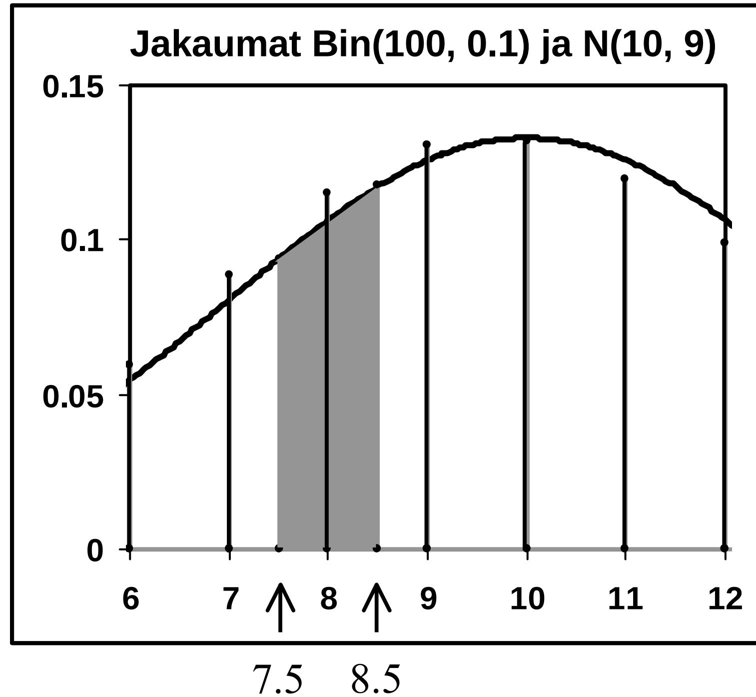
  pistetodennäköisyysfunktiota ja jakauman

  N(10, 9)

  tiheysfunktiota välillä [6, 12].
- De Moivren ja Laplacen raja-arvolauseen mukaan binomitodennäköisyyttä pisteessä

  $x = 8$

  voidaan *approksimoida* varjostetun alueen pinta-alalla; ks. kalvoja 2/3-3/3.



Binomitodennäköisyydet ja normaalijakauma: Esimerkki 2/3

- Olkoon $X \sim \text{Bin}(n, p)$, jossa

  $n = 100$
  $p = 0.1$

- Tällöin

  $f_X(8) = 0.1148$

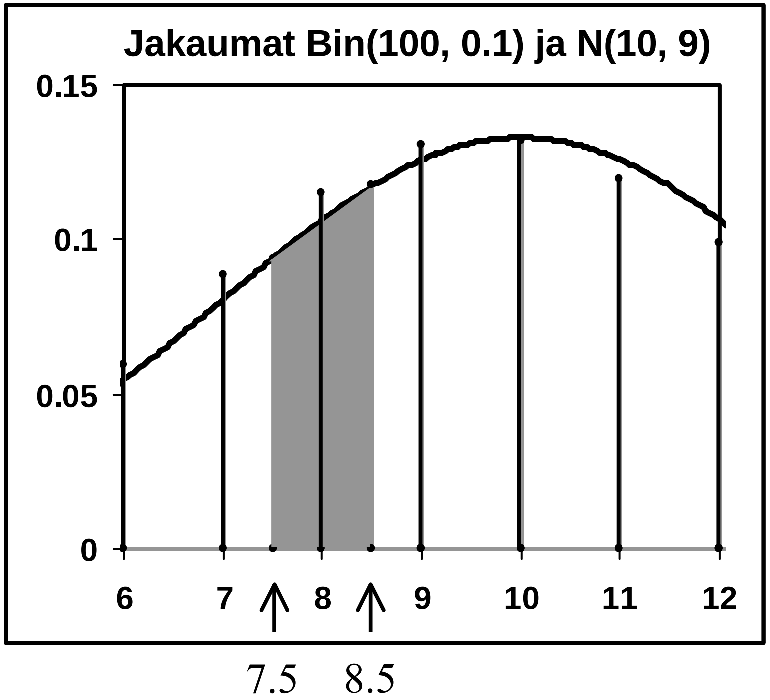
  jossa

  $f_X(x)$

  on binomijakauman

  Bin(100, 0.1)

  pistetodennäköisyysfunktio.



Binomitodennäköisyydet ja normaalijakauma: Esimerkki 3/3

- Olkoot

$$\mu = np = 10$$
$$\sigma^2 = np(1-p) = 9$$

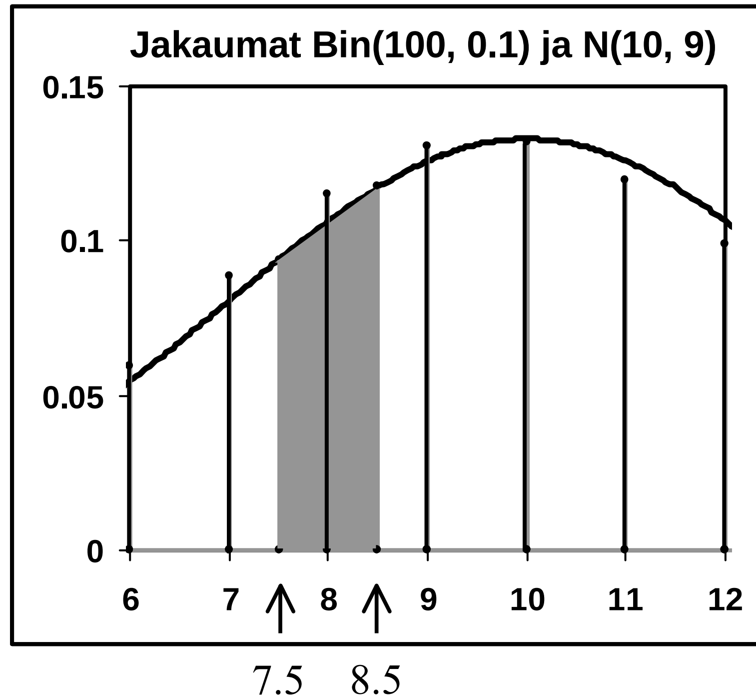
jossa

$$n = 100$$
$$p = 0.1$$

- Tällöin

$$\Phi\left(\frac{8+1/2-\mu}{\sigma}\right) - \Phi\left(\frac{8-1/2-\mu}{\sigma}\right) = 0.1052$$

jossa Φ on standardoidun normaalijakauman N(0,1) *kertymäfunktio*.



Hypergeometrisen jakauman todennäköisyydet ja normaalijakauma 1/2

- *Hypergeometrinen jakauma*

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

  lähestyy perusjoukon koon N kasvaessa rajatta *binomi-jakaumaa*

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

  jossa

  $$p = r/N$$

Hypergeometrisen jakauman todennäköisyydet ja normaalijakauma 2/2

- Siten *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-jakauma ja normaalijakauma

- Olkoon $X \sim \text{Poisson}(\lambda)$.
- Siten

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

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

- Tällöin

$$\lim_{\lambda \to +\infty} \Pr\left(\frac{X - \lambda}{\sqrt{\lambda}} \le z \right) = \Phi(z)$$

jossa Φ on *standardoidun normaalijakauman* N(0,1) *kertymäfunktio*.

Poisson-jakauman todennäköisyydet ja normaalijakauma 1/4

- Poisson-jakaumaa koskevan raja-arvolauseen mukaan *Poisson-jakaumaa*

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

  voidaan suurille λ *approksimoida normaalijakaumalla*

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

  jossa

  $$\mu = \lambda$$

  $$\sigma^2 = \lambda$$

Poisson-jakauman todennäköisyydet ja normaalijakauma 2/4

- Jos siis

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

niin Poisson-jakaumaa koskevan raja-arvolauseen mukaan suurille λ

$$\Pr(a < X \leq b) \approx \Phi\left(\frac{b-\lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a-\lambda}{\sqrt{\lambda}}\right)$$

jossa Φ on *standardoidun normaalijakauman* N(0,1) *kertymäfunktio*.

Poisson-jakauman todennäköisyydet ja normaalijakauma 3/4

- 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.

Poisson-jakauman todennäköisyydet ja normaalijakauma 4/4

- 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 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 on *Poisson-jakauman pistetodennäköisyysfunktio*.