Aalto-yliopisto

Keskeinen raja-arvolause ja sen sovellukset liiketoiminnassa

25. maaliskuuta 2025

Agenda of the presentation

Lyhyt kertaus: Suurten lukujen laki

Normaalijakauma

Keskeinen raja-arvolause

Todennäköisyyksien approksimointi

Sovellus: Likiarvoinen luottamusväli

Kohdeyleisö

- Ensimmäisen ja toisen vuoden kandiopiskelijat

Moraali

- Suurten lukujen laki

Opetustavoitteet

Luennon jälkeen

1. osaamme approksimoida todennäköisyyksiä klassisen keskeisen raja-arvo lauseen avulla,

Table of Contents

Table of Contents

Normaalijakauma on absoluuttisesti jatkuva jakauma. Merkitsemme normaalijakaumaa parametrein $\mu \in (-\infty, \infty)$ ja $\sigma^2 \in (0, \infty)$ notaatiolla $\mathrm{N}\left(\mu, \sigma^2\right)$. Jakauman $\mathrm{N}\left(\mu, \sigma^2\right)$ tiheysfunktio on

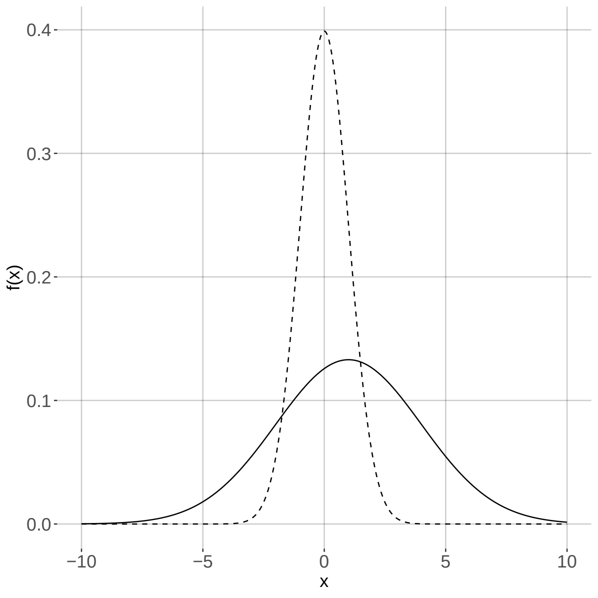
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Normaalijakauma on absoluuttisesti jatkuva jakauma. Merkitsemme normaalijakaumaa parametrein $\mu \in (-\infty, \infty)$ ja $\sigma^2 \in (0, \infty)$ notaatiolla $\mathrm{N}\left(\mu, \sigma^2\right)$. Jakauman $\mathrm{N}\left(\mu, \sigma^2\right)$ tiheysfunktio on

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

Tällöin normaalijakauman kertymäfunktio voidaan esittää tiheysfunktion avulla

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



Kuva: Normaalijakaumat $\mathrm{N}\,(0, 1)$ (katkoviiva) ja $\mathrm{N}\,(1, 9)$ (jatkuva viiva).

Table of Contents

Simulaatio

1. Simuloi otos riippumattomia ja samoin jakauneita havaintoja $x_1, \ldots, x_n$ (otoskoko on n). Laske keskiarvo

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Simulaatio

1. Simuloi otos riippumattomia ja samoin jakauneita havaintoja $x_1, \ldots, x_n$ (otoskoko on n). Laske keskiarvo

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

2. Toista yllä oleva toimenpide $m = 1000$ kertaa. Näin meillä on m keskiarvoa.

Simulaatio

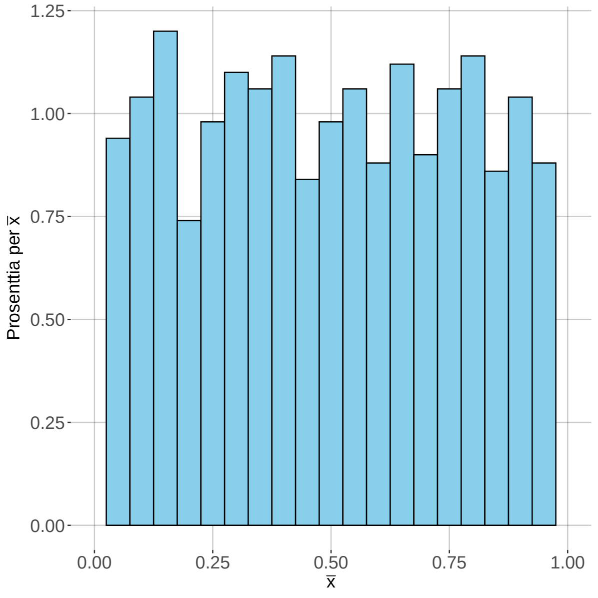
1. Simuloi otos riippumattomia ja samoin jakauneita havaintoja $x_1, \ldots, x_n$ (otoskoko on n). Laske keskiarvo

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

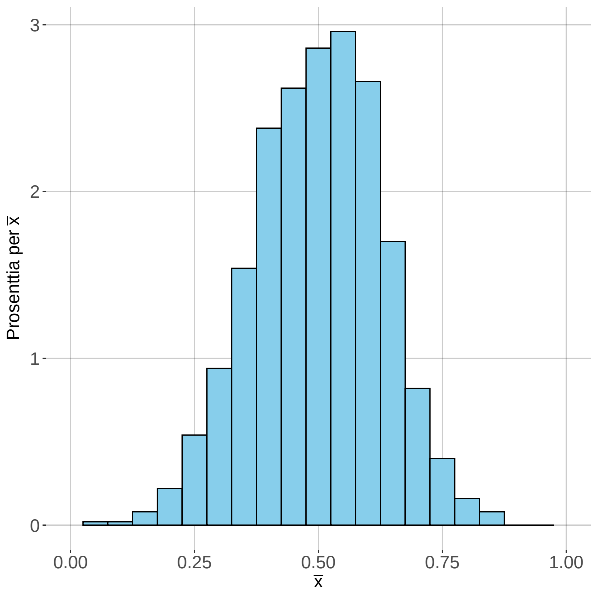
2. Toista yllä oleva toimenpide $m = 1000$ kertaa. Näin meillä on m keskiarvoa.
3. Piirrä histogrammi keskiarvoista.

- Ensin simuloimme otoksia tasajakaumasta $U[0,1]$ välillä 0-1 (jakauman odotusarvo on 0.5). Kyseisen tasajakauman tiheysfunktio on

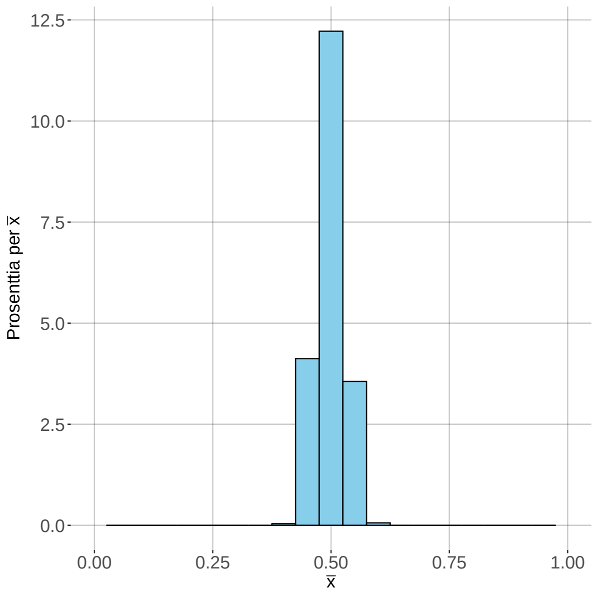
$$f(x) = \begin{cases} 1, & \text{kun} \quad x \in [a,b], \\ 0, & \text{muulloin.} \end{cases}$$



Kuva: Tasajakauma $U[0, 1]$, $n = 1$ ja $m = 1000$.



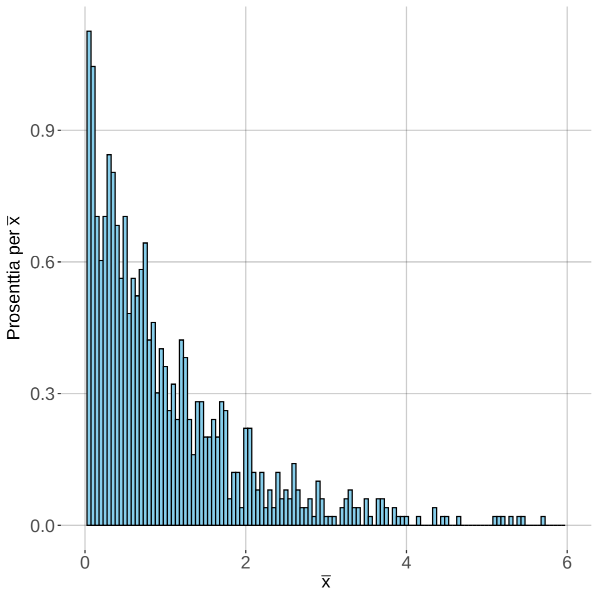
Kuva: Tasajakauma $U[0, 1]$, $n = 5$ ja $m = 1000$.



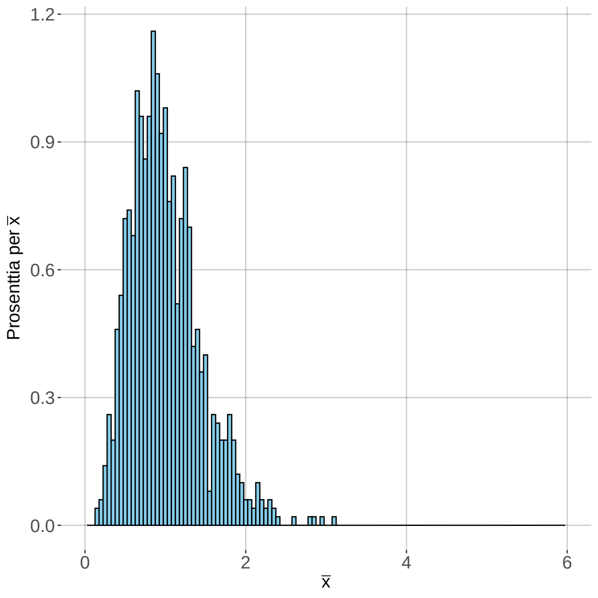
Kuva: Tasajakauma $U[0, 1]$, $n = 100$ ja $m = 1000$.

- Seuraavaksi simuloimme otoksia eksponenttijakaumasta $\mathrm{Exp}(1)$ skaalaparametrilla $\lambda = 1$ (jakauman odotusarvo on 1). Kyseisen eksponentijakauman tiheysfunktio on

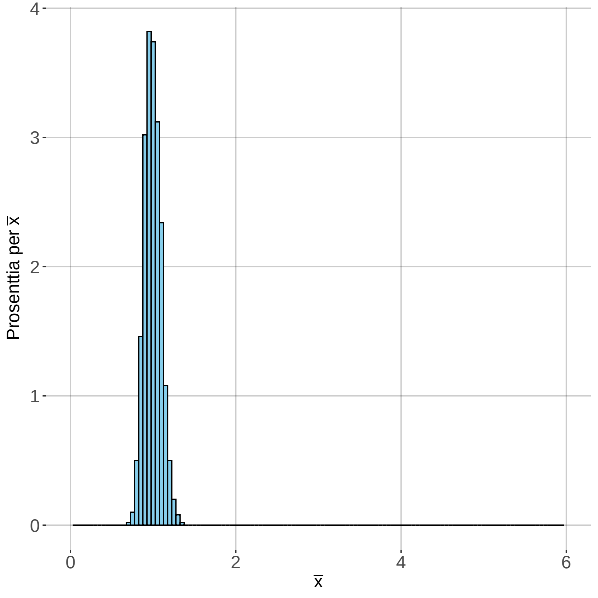
$$f(x) = \begin{cases} e^{-x}, & \text{kun} \quad x \in [0, \infty), \\ 0, & \text{muulloin.} \end{cases}$$



Kuva: Eksponenttijakauma $\mathrm{Exp}(1)$, $n = 1$ ja $m = 1000$.



Kuva: Eksponenttijakauma $\mathrm{Exp}\,(1)$, $n = 5$ ja $m = 1000$.



Kuva: Eksponenttijakauma $\mathrm{Exp}\,(1)$, $n = 100$ ja $m = 1000$.

Lause

Olkoon $X_1, X_2, \ldots, X_n$ *riippumattomia ja samoin jakautuneita satunnaismuuttujia niin, että* $\mathbb{E}\left(|X_1|^2\right) < \infty$. *Merkitsemme* $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}\left(X_1\right)$ *ja* $\sigma = \sqrt{\mathrm{Var}\left(X_1\right)}$. *Tällöin, kun otoskoko n on suuri,*

$$\tilde{X} = \sqrt{n}\frac{\bar{X} - \mu}{\sigma}$$

noudattaa likimain standardinormaalijakaumaa $\mathrm{N}\left(0, 1\right)$,

Lause

Olkoon $X_1, X_2, \ldots, X_n$ riippumattomia ja samoin jakautuneita satunnaismuuttujia niin, että $\mathbb{E}\left(|X_1|^2\right) < \infty$. Merkitsemme $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}\left(X_1\right)$ ja $\sigma = \sqrt{\mathrm{Var}\left(X_1\right)}$. Tällöin, kun otoskoko n on suuri,

$$\tilde{X} = \sqrt{n}\frac{\bar{X} - \mu}{\sigma}$$

noudattaa likimain standardinormaalijakaumaa $\mathrm{N}\left(0, 1\right)$,

$$\mathbb{P}\left(\tilde{X} \leq x\right) \approx \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,\mathrm{d}t.$$

Intuitio

Toisin sanoen kun otoskoko n on suuri, niin $\bar{X}$ noudattaa likimain jakaumaa

$$\mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Kurssin ulkopuolista asiaa

Table of Contents

Table of Contents

Luottamusvälin määritelmä odotusarvolle

- Oletetaan, että X on satunnismuuttuja odotusarvolla $\mu = \mathbb{E}(X) \in (-\infty, \infty)$. Olkoon $\boldsymbol{X} = (X_1, \ldots, X_n)$ satunnaismuuttujan X havaintoja.

Luottamusvälin määritelmä odotusarvolle

- Oletetaan, että X on satunnismuuttuja odotusarvolla $\mu = \mathbb{E}(X) \in (-\infty, \infty)$. Olkoon $\boldsymbol{X} = (X_1, \ldots, X_n)$ satunnaismuuttujan X havaintoja.
- Luottamustason $1 - \alpha$ luottamusväli on satunnaisväli $[\ell(\boldsymbol{X}), u(\boldsymbol{X})]$, jolle pätee

$$\mathbb{P}(\mu \in [\ell(\boldsymbol{X}), u(\boldsymbol{X})]) = 1 - \alpha$$

Luottamusvälin approksimointi (1)

Oletetaan, että satunnaismuuttujan X varianssi $0 < \sigma^2 < \infty$ on tiedossa. Tällöin keskeisen raja-arvolauseen mukaan

$$\mathbb{P}\left(z_\ell \le \frac{\mu - \bar{X}}{\sigma} \le z_u\right) \approx 1 - \alpha,$$

Luottamusvälin approksimointi (1)

Oletetaan, että satunnaismuuttujan X varianssi $0 < \sigma^2 < \infty$ on tiedossa. Tällöin keskeisen raja-arvolauseen mukaan

$$\mathbb{P}\left(z_\ell \le \frac{\mu - \bar{X}}{\sigma} \le z_u\right) \approx 1 - \alpha,$$

jossa väli $[z_\ell, z_u]$ valitaan niin, että

$$F(z_u) - F(z_\ell) = 1 - \alpha.$$

Yllä F on jakauman $\mathrm{N}\,(0, 1)$ kertymäfunktio.

Luottamusvälin approksimointi (2)

Valitsemalla esimerkiksi $z_u = z_{1-\alpha/2}$ ja $z_\ell = -z_u$ päädymme seuraavaan luottamusväliin

$$\left[\bar{X} - \sigma z_{1-\alpha/2}, \bar{X} + \sigma z_{1-\alpha/2}\right].$$

Esimerkki: luottokorttipetokset

- Otos sisältää luottokorttitapahtumia syyskuulta 2013. Tapahtumia kirjattiin kahdelta päivältä (Lähde: Kaggle).

Esimerkki: luottokorttipetokset

- Otos sisältää luottokorttitapahtumia syyskuulta 2013. Tapahtumia kirjattiin kahdelta päivältä (Lähde: Kaggle).
- Luottokorttitapahtumia on yhteensä 284 807, joista 492 luokiteltiin petoksiksi.

Esimerkki: luottokorttipetokset

- Otos sisältää luottokorttitapahtumia syyskuulta 2013. Tapahtumia kirjattiin kahdelta päivältä (Lähde: Kaggle).
- Luottokorttitapahtumia on yhteensä 284 807, joista 492 luokiteltiin petoksiksi.
- Tavoite: Estimoidaan 95% luottamusväli petoksien osuudelle luottokorttitapahtumista.

Malli

- Satunnaismuuttuja X noudattaa Bernoullijakaumaa tuntemattomalla parametrilla p, jossa p on petoksen todennäköisyys.
- $\mathbb{P}(X = 1) = p$ ja $\mathbb{P}(X = 0) = 1 - p$, jossa tapahtuma $\{X = 1\}$ indikoi petosta ja $\{X = 0\}$ vastaa normaalia luottokorttitapahtumaa.
- Otos $x_1, \ldots x_{284807}$ on 0-1 muotoista. Oletamme, että havainnot ovat riippumattomia ja samoin jakautuneita.

Ratkaisu

- Voimme laskea $\mathbb{E}\left(X\right) = p$ ja $\mathrm{Var}\left(X\right) = p(1-p)$.
- Suurten lukujen lain mukaan $p \approx \hat{p} = \frac{1}{n}\sum_{i=1}^{284807} x_i$.
- Joten voimme approksimoida 95% luottamusvälin seuraavasti

$$\left[\hat{p} - z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})}, \hat{p} + z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})}\right]$$

Ratkaisu (2)

Sijoittamalla lukuarvot

- $\hat{p} \approx 0.0017$, (R komento `mean(data)`) ja
- $z_{1-\alpha/2} \approx 1.96$ (R komento `qnorm(1 - 0.05 / 2)`)

saamme luottamusväliksi

$$\approx [0.0016, 0.0019] .$$

Tiivistelmä