



Aalto-yliopisto

## **Keskeinen raja-arvolause ja sen sovellukset liiketoiminnassa**

<https://github.com/perej1/clt-teaching>

**Jaakko Pere**

**25. maaliskuuta 2025**

## Kohdeyleisö

- Ensimmäisen ja toisen vuoden kandiopiskelijat (lukion matematiikka).

# Kohdeyleisö

- Ensimmäisen ja toisen vuoden kandiopiskelijat (lukion matematiikka).
- Kurssilla on jo käsitelty seuraavat käsitteet:
  - Satunnaismuuttuja
  - Odotusarvo
  - Varianssi
  - Estimaattori
  - Suurten lukujen laki

# **Luennon sisältö**

**Moraali**

**Normaalijakauma**

**Keskeinen raja-arvolause**

**Sovellus: likiarvoinen luottamusväli odotusarvolle**

# Table of Contents

**Moraali**

**Normaalijakauma**

**Keskeinen raja-arvolause**

**Sovellus: likiarvoinen luottamusväli odotusarvolle**

## Esimerkki: luottokorttipetokset

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

## Esimerkki: luottokorttipetokset

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

## Esimerkki: luottokorttipetokset

- Otos sisältää luottokorttitapahtumia Euroopasta syyskuulta 2013. Tapahtumia kirjattiin kahdelta päivältä (Lähde: [Kaggle](#)).
- Luottokorttitapahtumia on yhteensä 284 807, joista 492 luokiteltiin petoksiksi.
- Estimaatti luottokorttipetoksen todennäköisyydelle on  $\hat{p} = 492/284807 \approx 0.0017$ .



## Esimerkki: luottokorttipetokset

- Otos sisältää luottokorttitapahtumia Euroopasta syyskuulta 2013. Tapahtumia kirjattiin kahdelta päivältä (Lähde: [Kaggle](#)).
- Luottokorttitapahtumia on yhteensä 284 807, joista 492 luokiteltiin petoksiksi.
- Estimaatti luottokorttipetoksen todennäköisyydelle on  $\hat{p} = 492/284807 \approx 0.0017$ .
- Kuinka varma voin olla siitä, että saatu piste-estimaatti on lähellä populaatisuuretta?

## Esimerkki: luottokorttipetokset

- Otos sisältää luottokorttitapahtumia Euroopasta syyskuulta 2013. Tapahtumia kirjattiin kahdelta päivältä (Lähde: [Kaggle](#)).
- Luottokorttitapahtumia on yhteensä 284 807, joista 492 luokiteltiin petoksiksi.
- Estimaatti luottokorttipetoksen todennäköisyydelle on  $\hat{p} = 492/284807 \approx 0.0017$ .
- Kuinka varma voin olla siitä, että saatu piste-estimaatti on lähellä populaatisuuretta?
- Voidaanko keskiarvon jakaumasta sanoa jotain edes suurilla otosko'oilla  $n$ ?

# Simulaatio

1. Simuloi otos riippumattomia ja samoin jakauneita havaintoja  $x_1, \dots, x_n$  (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  $x_1, \dots, x_n$  (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

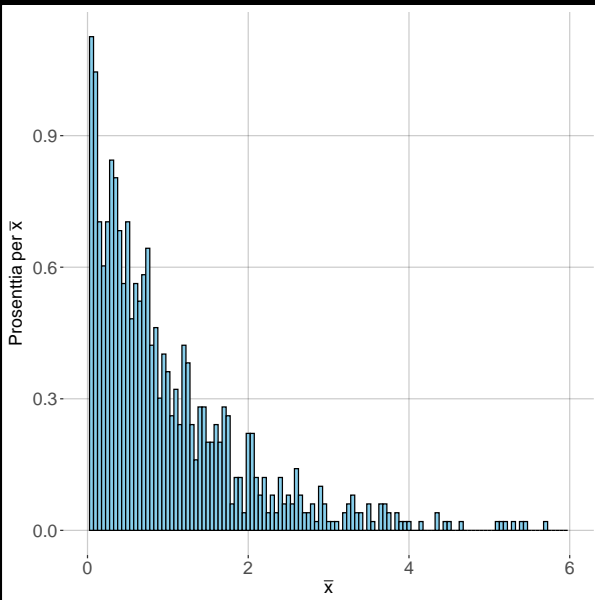
1. Simuloi otos riippumattomia ja samoin jakauneita havaintoja  $x_1, \dots, x_n$  (otoskoko on  $n$ ). Laske keskiarvo

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

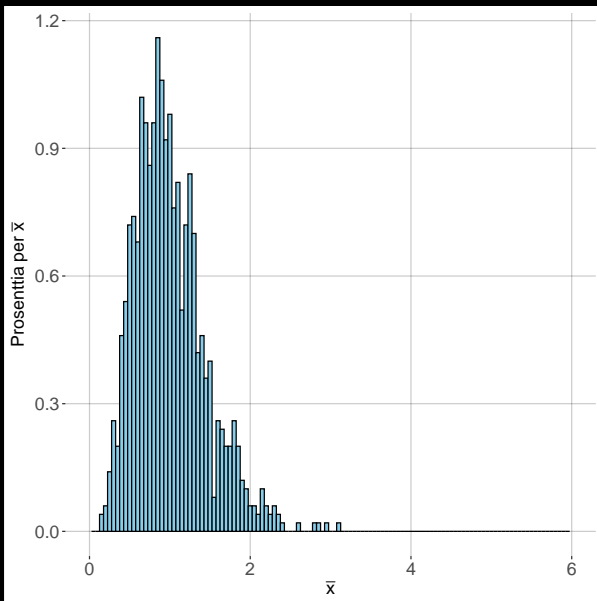
2. Toista yllä oleva toimenpide  $m = 1000$  kertaa. Näin meillä on  $m$  keskiarvoa.
3. Piirrä histogrammi keskiarvoista.

- Simuloimme otoksia eksponenttijakaumasta  $\text{Exp}(1)$  skaalaparametrilla  $\lambda = 1$  (jakauman odotusarvo on 1). Kyseisen eksponenttijakauman tiheysfunktio on

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

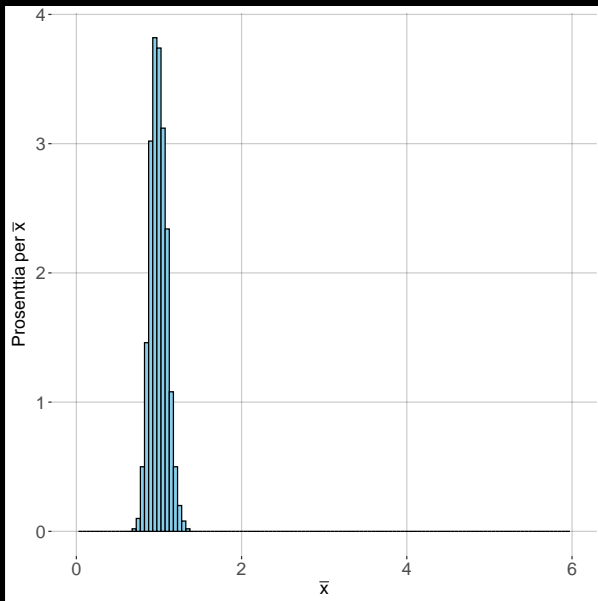


**Kuva:** Eksponenttijakauma  $\text{Exp}(1)$ ,  $n = 1$  ja  $m = 1000$ .



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





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

## Opetustavoite

Luennon jälkeen osaamme muodostaa likiarvoisen luottamusvälin odotusarvolle perustuen keskeiseen raja-arvolauseeseen.

# Table of Contents

Moraali

Normaalijakauma

Keskeinen raja-arvolause

Sovellus: likiarvoinen luottamusväli odotusarvolle

Normaalijakauma on absoluuttisesti jatkuva jakauma.

Merkitsemme normaalijakaumaa parametrein  $\mu \in (-\infty, \infty)$  ja  $\sigma^2 \in (0, \infty)$  notaatiolla  $N(\mu, \sigma^2)$ . Jakauman  $N(\mu, \sigma^2)$  tiheysfunktio on

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

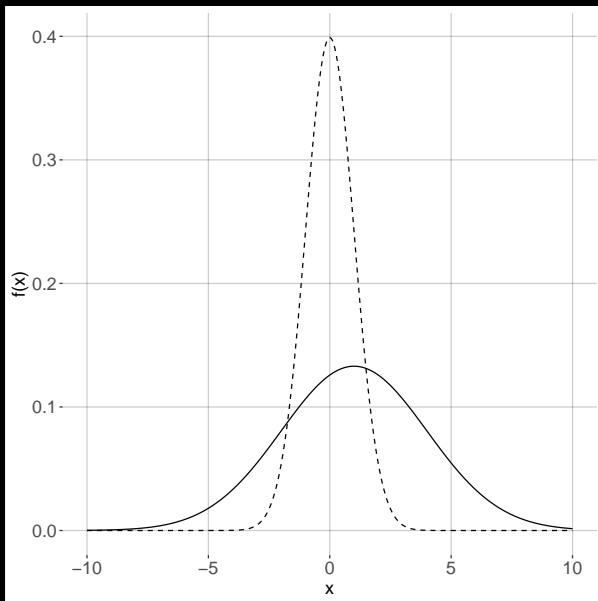
Normaalijakauma on absoluuttisesti jatkuva jakauma.

Merkitsemme normaalijakaumaa parametrein  $\mu \in (-\infty, \infty)$  ja  $\sigma^2 \in (0, \infty)$  notaatiolla  $N(\mu, \sigma^2)$ . Jakauman  $N(\mu, \sigma^2)$  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

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



**Kuva:** Normaalijakaumat  $N(0, 1)$  (katkoviiva) ja  $N(1, 9)$  (jatkuva viiva).

# Table of Contents

Moraali

Normaalijakauma

Keskeinen raja-arvolause

Sovellus: likiarvoinen luottamusväli odotusarvolle

## Lause

*Olkoon  $X_1, X_2, \dots, X_n$  riippumattomia ja samoin jakautuneita satunnaismuuttujia niin, että  $0 < \text{Var}(X_1) < \infty$ . Merkitsemme  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ ,  $\mu = \mathbb{E}(X_1)$  ja  $\sigma = \sqrt{\text{Var}(X_1)}$ . Tällöin, kun otoskoko  $n$  on suuri, niin*

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

*noudattaa likimain standardinormaalijakaumaa  $N(0, 1)$ ,*



## Lause

*Olkoon  $X_1, X_2, \dots, X_n$  riippumattomia ja samoin jakautuneita satunnaismuuttujia niin, että  $0 < \text{Var}(X_1) < \infty$ . Merkitsemme  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ ,  $\mu = \mathbb{E}(X_1)$  ja  $\sigma = \sqrt{\text{Var}(X_1)}$ . Tällöin, kun otoskoko  $n$  on suuri, niin*

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

*noudattaa likimain standardinormaalijakaumaa  $N(0, 1)$ ,*

$$\mathbb{P}(\tilde{X} \leq x) \approx \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt =: \Phi(x).$$

# Intuitio

Kun otoskoko  $n$  on suuri, niin melkeinpä  $\bar{X} \sim \mathbf{N}\left(\mu, \frac{\sigma^2}{n}\right)$ .

# Intuitio

Kun otoskoko  $n$  on suuri, niin melkeinpä  $\bar{X} \sim \text{N}\left(\mu, \frac{\sigma^2}{n}\right)$ .

Huom! Dian väite ei pidä täysin paikkaansa vaan antaa vain intuition aiemmille simulaatioille.

# Table of Contents

Moraali

Normaalijakauma

Keskeinen raja-arvolause

**Sovellus: likiarvoinen luottamusväli odotusarvolle**

## Luottamusvälin määritelmä odotusarvolle

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

## Luottamusvälin määritelmä odotusarvolle

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

$$\mathbb{P}(\mu \in [\ell(\mathbf{X}), u(\mathbf{X})]) \geq 1 - \alpha.$$

## Luottamusvälin approksimointi (1)

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

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

## Luottamusvälin approksimointi (1)

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

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

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

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

Yllä  $F$  on jakauman  $N(0, 1)$  kertymäfunktio.



## Luottamusvälin approksimointi (2)

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

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

## Luottamusvälin approksimointi (2)

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

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

Käytännössä yleensä (tuntematon) keskihajonta korvataan vielä otoskeskihajonnalla  $\hat{\sigma}$ , sillä suurten lukujen lain perusteella  $\hat{\sigma} \approx \sigma$ . Päädymme luottamusväliin

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

## 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.
- Havainnot  $x_1, \dots, x_{284807}$  ovat binäärisiä (0 tai 1). Oletamme, että havainnot ovat riippumattomia ja samoin jakautuneita.

## Ratkaisu

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

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

## 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

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



## 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

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

Luottamusvälin laskemiseen liittyvät yksityiskohdat löytyvät skriptistä `credit.R`.

## Tiivistelmä

Sovelsimme keskeistä raja-arvolauseetta luottokorttipetokseen

## Tiivistelmä

Sovelsimme keskeistä raja-arvolausetta luottokorttipetokseen

- laskemalla likiarvoistettuja luottamusvälejä petoksen todennäköisyydelle.

## Tiivistelmä

Sovelsimme keskeistä raja-arvolausetta luottokorttipetokseen

- laskemalla likiarvoistettuja luottamusvälejä petoksen todennäköisyydelle.

Kaikki materiaalit löytyvät osoitteesta

<https://github.com/perej1/clt-teaching>.