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 todellista parametria μ?

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 todellista parametria μ?
- Voidaanko keskiarvon jakaumasta yleisesti sanoa jotain suurilla otosko'oilla n?

Simulaatio (1)

1. Simuloi otos riippumattomia ja samoin jakauneita havaintoja $x_1, \ldots, x_n$. Laske keskiarvo

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

Simulaatio (1)

1. **1.** Simuloi otos riippumattomia ja samoin jakauneita havaintoja $x_1, \ldots, x_n$. Laske keskiarvo

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

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

Simulaatio (1)

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

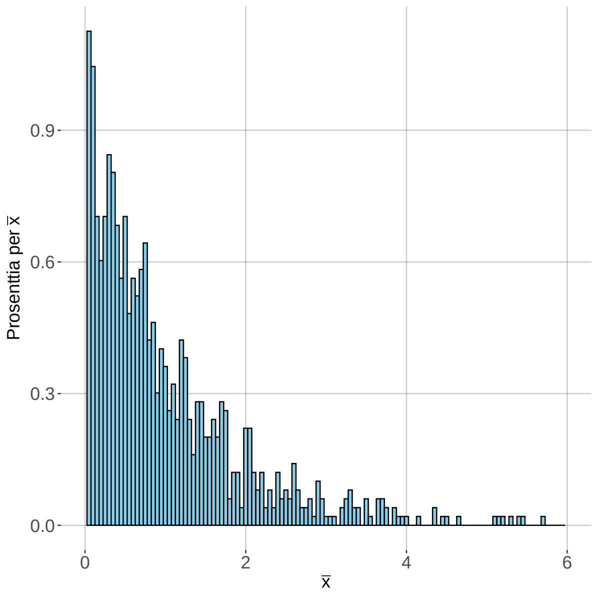
$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

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

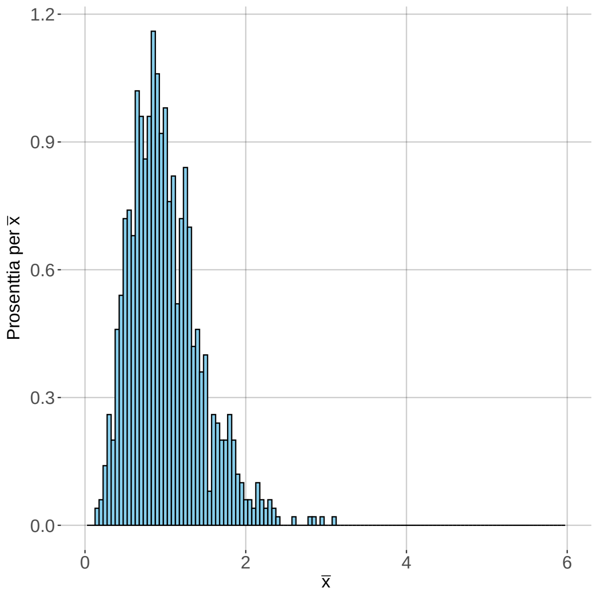
Simulaatio (2)

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

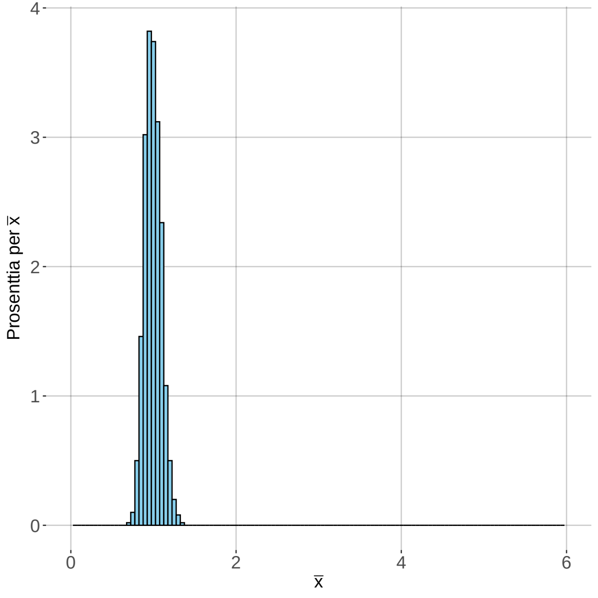
$$f(x) = \begin{cases} e^{-x}, & \text{kun} \quad x \in [0, \infty), \\ 0, & \text{muulloin.} \end{cases}$$



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 $m = 1000$.

Opetustavoite

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

Table of Contents

Moraali

Normaalijakauma

Keskeinen raja-arvolause

Sovellus: likiarvoinen luottamusväli odotusarvolle

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

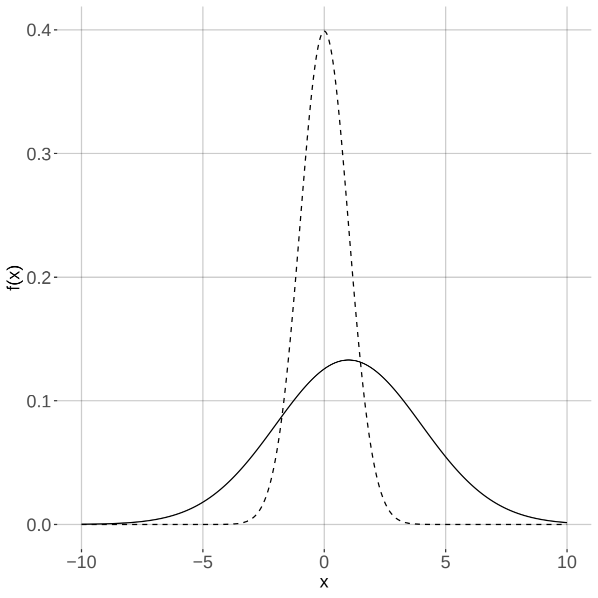
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

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

Moraali

Normaalijakauma

Keskeinen raja-arvolause

Sovellus: likiarvoinen luottamusväli odotusarvolle

Lause

Olkoon $X_1, X_2, \ldots, X_n$ *riippumattomia ja samoin jakautuneita satunnaismuuttujia niin, että* $0 < \mathrm{Var}\left(X_1\right) < \infty$. *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, niin*

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

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* $0 < \mathrm{Var}\left(X_1\right) < \infty$. *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, niin*

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

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

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

Intuitio

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

Intuitio

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

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 $\boldsymbol{X} = (X_1, \ldots, X_n)$ satunnaismuuttujan X havaintoja.

Luottamusvälin määritelmä odotusarvolle

- Oletetaan, että X on satunnaismuuttuja 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

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

Luottamusvälin approksimointi (1)

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

$$\mathbb{P}\left(z^{(\ell)} \le \sqrt{n}\frac{\mu - \bar{X}}{\sigma} \le z^{(u)}\right) \approx 1 - \alpha,$$

Luottamusvälin approksimointi (1)

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

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

jossa väli $[z^{(\ell)}, z^{(u)}]$ valitaan niin, että

$$\Phi(z^{(u)}) - \Phi(z^{(\ell)}) = 1 - \alpha.$$

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

Luottamusvälin approksimointi (2)

Olkoon $z_{1-\alpha/2}$ standardinormaalijakauman $(1-\alpha/2)$-kvantiili. Valitsemalla esimerkiksi $z^{(u)} = z_{1-\alpha/2}$ ja $z^{(\ell)} = -z_{1-\alpha/2}$ 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)

Olkoon $z_{1-\alpha/2}$ standardinormaalijakauman $(1-\alpha/2)$-kvantiili. Valitsemalla esimerkiksi $z^{(u)} = z_{1-\alpha/2}$ ja $z^{(\ell)} = -z_{1-\alpha/2}$ päädymme seuraavaan luottamusväliin

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

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

Ratkaisu

- Voimme laskea $\mathbb{E}(X) = p$ ja $\mathrm{Var}(X) = p(1-p)$.
- 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-arvolausetta luottokorttipetoksiin

Tiivistelmä

Sovelsimme keskeistä raja-arvolausetta luottokorttipetoksiin

- laskemalla likiarvoistettuja luottamusvälejä petoksen todennäköisyydelle.

Tiivistelmä

Sovelsimme keskeistä raja-arvolausetta luottokorttipetoksiin

- laskemalla likiarvoistettuja luottamusvälejä petoksen todennäköisyydelle.

Kaikki materiaalit löytyvät osoitteesta `https://github.com/perej1/clt-teaching`.