
Ilkka Mellin
Tilastolliset menetelmät

Osa 2: Otokset, otosjakaumat ja estimointi
Estimointi

Estimointi

- >> Todennäköisyysjakaumien parametrit ja niiden estimointi**
Hyvän estimaattorin ominaisuudet

Todennäköisyysjakaumat tilastollisten aineistojen kuvaajina

- **Tilastollinen aineisto** koostuu tutkimuksen kohteita kuvaavien muuttujien *havaituista arvoista*.
- Tilastollisissa tutkimusasetelmissä havaintoarvoihin liittyy aina *epävarmuutta ja satunnaisuutta*.
- Tilastollisissa tutkimusasetelmissä tutkimuksen kohteita kuvaavat muuttujat tulkitaan *satunnaismuuttujiksi*, jotka *generoivat* muuttujien havaitut arvot.
- *Havaintoarvot generoineiden satunnaismuuttujien todennäköisyysjakauma* muodostaa **tilastollisen mallin** sille satunnaisilmiölle, jota havainnot koskevat.

Todennäköisyysjakaumien parametrit 1/2

- Tarkastellaan jotakin tutkimuksen kaikkien mahdollisten kohteiden muodostaman perusjoukon S alkioiden ominaisuutta kuvaavaa *satunnaismuuttujaa* X .
- Oletetaan, että satunnaismuuttuja X noudattaa *todennäköisyysjakaumaa*, jonka *pistetodennäköisyys-* tai *tiheysfunktio*

$$f(x ; \theta)$$

riippuu **parametrista** θ .

- Merkintä:

$$X \sim f(x ; \theta)$$

Todennäköisyysjakaumien parametrit 2/2

- Satunnaismuuttujan X pistetodennäköisyys- tai tiheysfunktio

$$f(x ; \theta)$$

kuvaa satunnaismuuttujan X *todennäköisyysjakaumaa* ja parametri θ kuvaa jotakin jakauman *karakteristista ominaisuutta*.

- Koska parametrin θ arvoa *ei yleensä tunneta*, tilastollisen tutkimuksen tärkeimpiä osatehtäviä on **estimoida** eli *arvioida* parametrin θ tuntematon arvo jakaumasta $f(x ; \theta)$ *poimitun otoksen perusteella*.

Todennäköisyysjakaumien parametrit ja niiden estimointi

Yksinkertainen satunnaisotos

- Olkoon

$$X_1, X_2, \dots, X_n$$

(**yksinkertainen**) **satunnaisotos** jakaumasta, jonka pistetodennäköisyys- tai tiheysfunktio $f(x; \theta)$ riippuu parametrilla θ .

- Tällöin havainnot X_1, X_2, \dots, X_n ovat *riippumattomia, identtisesti jakautuneita satunnaismuuttujia*, joilla on *sama pistetodennäköisyys- tai tiheysfunktio* $f(x; \theta)$:

$$X_1, X_2, \dots, X_n \perp$$

$$X_i \sim f(x; \theta), i = 1, 2, \dots, n$$

Todennäköisyysjakaumien parametrit ja niiden estimointi

Havainnot ja havaintoarvot

- Oletetaan, että satunnaismuuttujat (havainnot)

$$X_1, X_2, \dots, X_n$$

saavat *poimitussa otoksessa* **havaituiksi arvoikseen** luvut

$$x_1, x_2, \dots, x_n$$

- **Havaintoarvot**

$$x_1, x_2, \dots, x_n$$

vaihtelevat satunnaisesti otoksesta toiseen jakaumasta

$$f(x; \theta)$$

saatavin todennäköisyyksin.

Todennäköisyysjakaumien parametrit ja niiden estimointi

Estimaattorit ja estimaatit 1/2

- Oletetaan, että todennäköisyysjakauman $f(x; \theta)$ parametrin θ estimointiin käytetään satunnaismuuttujien X_1, X_2, \dots, X_n funktiota eli *tunnuslukua*

$$T = g(X_1, X_2, \dots, X_n)$$

- Tällöin funktiota $T = g(X_1, X_2, \dots, X_n)$ kutsutaan parametrin θ **estimaattoriksi** ja funktion g *havaintoarvoista*

$$x_1, x_2, \dots, x_n$$

laskettua arvoa

$$t = g(x_1, x_2, \dots, x_n)$$

kutsutaan parametrin θ **estimaatiksi**.

Todennäköisyysjakaumien parametrit ja niiden estimointi

Estimaattorit ja estimaatit 2/2

- Olkoon

$$T = g(X_1, X_2, \dots, X_n)$$

jakauman $f(x; \theta)$ parametrin θ *estimaattori*.

- Tällöin estimaattorin T havaintoarvoista

$$x_1, x_2, \dots, x_n$$

laskettu arvo eli *estimaatti*

$$t = g(x_1, x_2, \dots, x_n)$$

on satunnaismuuttujan T arvon realisaatio otoksessa.

Estimaattorit ja estimaatit:

Kommentti

- Todennäköisyysjakauman $f(x; \theta)$ parametrin θ *estimaattorilla*

$$T = g(X_1, X_2, \dots, X_n)$$

tarkoitetaan siis sellaista jakaumaa $f(x; \theta)$ noudattavien *satunnaismuuttujien*

$$X_1, X_2, \dots, X_n$$

funktiota, joka generoi muuttujien X_1, X_2, \dots, X_n havaittuihin arvoihin x_1, x_2, \dots, x_n sovellettuna *estimaatteja eli arvioita*

$$t = g(x_1, x_2, \dots, x_n)$$

parametrille θ .

Todennäköisyysjakaumien parametrit ja niiden estimointi

Estimaattorin otosjakauma

- Estimaattorin

$$T = g(X_1, X_2, \dots, X_n)$$

havaintoarvoista

$$x_1, x_2, \dots, x_n$$

lasketut arvot eli estimaatit

$$t = g(x_1, x_2, \dots, x_n)$$

vaihtelevat satunnaisesti otoksesta toiseen.

- Estimaattorin T arvojen satunnaista vaihtelua otoksesta toiseen voidaan kuvata *estimaattorin T otosjakaumalla*.

Todennäköisyysjakaumien parametrit ja niiden estimointi

Estimaattoreiden johtaminen

- *Hyvien estimaattoreiden johtaminen todennäköisyysjakaumien tuntemattomille parametreille on teoreettisen tilastotieteen keskeisiä ongelmia.*
 - Tärkeimmät estimaattoreiden johtamiseen käytettävät menetelmät:
 - **Suurimman uskottavuuden menetelmä**
 - **Momenttimenetelmä**
- Ks. lukua Estimointimenetelmät.

Todennäköisyysjakaumien parametrit ja niiden estimointi

Piste-estimointi ja väliestimointi

- Todennäköisyysjakauman *parametrin arvon estimointia* kutsutaan usein **piste-estimoinniksi**.
- Parametrin estimaattiin on aina syytä liittää **luottamusväliksi** kutsuttu *väli, joka sisältää estimoidun parametrin todellisen, mutta tuntemattoman arvon tietyllä, soveltajan valittavissa olevalla todennäköisyydellä*.
- Luottamusvälin määräämistä kutsutaan **wäliestimoinniksi**.
Ks. lukua Väliestimointi.

Estimointi

Todennäköisyysjakaumien parametrit ja niiden estimointi

>> Hyvän estimaattorin ominaisuudet

Hyvä estimaattori

- Todennäköisyysjakauman parametreille on tavallisesti tarjolla useita *vaihtoehtoisia estimaattoreita*.
- Estimaattorin valintaa ohjaavat **hyvyyskriteerit**, joilla pyritään takamaan se, että valittu estimaattori tuottaa järkeviä arvoja estimoitavalle parametrille.
- Estimaattoreiden hyvyyskriteereitä:
 - **Tyhjentyvyys**
 - **Harhattomuus**
 - **Tehokkuus**
 - **Tarkentuvuus**

Tyhjentävyys

- Olkoon $\hat{\theta}$ parametrin θ estimaattori
- Estimaattori $\hat{\theta}$ on **tyhjentävä** parametrille θ , jos se *käyttää* parametrin arvon estimoimiseen *kaiken otoksessa olevan informaation*.

Hyvän estimaattorin ominaisuudet

Harhattomuus ja harha

- Olkoon $\hat{\theta}$ parametrin θ estimaattori.
- Estimaattori $\hat{\theta}$ on **harhaton** parametrille θ , jos sen odotusarvo yhtyy parametrin θ arvoon:

$$E(\hat{\theta}) = \theta$$

- Estimaattorin $\hat{\theta}$ **harha** on

$$\text{Bias}(\hat{\theta}) = \theta - E(\hat{\theta})$$

- Jos estimaattori $\hat{\theta}$ on harhaton parametrille θ , niin

$$\text{Bias}(\hat{\theta}) = 0$$

Estimaattorin keskineliövirhe ja tarkkuus

- Parametrin θ estimaattorin $\hat{\theta}$ **keskineliövirhe** on

$$\begin{aligned}\text{MSE}(\hat{\theta}) &= \text{E}\left[(\hat{\theta} - \theta)^2\right] \\ &= \text{Var}(\hat{\theta}) + \left[\text{Bias}(\hat{\theta})\right]^2\end{aligned}$$

- Jos $\hat{\theta}$ on parametrin θ *harhaton* estimaattori eli

$$\text{Bias}(\hat{\theta}) = \theta - \text{E}(\hat{\theta}) = 0$$

niin

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta})$$

- Estimaattoria sanotaan **tarkaksi**, jos se on *harhaton* ja sen *varianssi* on *pieni*.

Hyvän estimaattorin ominaisuudet

Tehokkuus

- Olkoot $\hat{\theta}_1$ ja $\hat{\theta}_2$ kaksi parametrin θ estimaattoria.
- Estimaattori $\hat{\theta}_1$ on **tehokkaampi** kuin estimaattori $\hat{\theta}_2$, jos
$$\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$$
- Parametrin θ estimaattori $\hat{\theta}$ **täystehokas**, jos sen varianssi on *pienempi kuin minkä tahansa muun parametrin θ estimaattorin*.

Hyvän estimaattorin ominaisuudet

Tarkentuvuus

- Olkoon $\hat{\theta}$ parametrin θ estimaattori.
- Estimaattori $\hat{\theta}$ on θ **tarkentuva** parametrille θ , jos se *konvergoi* melkein varmasti *kohti parametrin oikeata arvoa*, kun otoskoon n annetaan kasvaa rajatta:

$$\Pr(T_n \rightarrow \theta) = 1, \text{ kun } n \rightarrow +\infty$$

Ks. monisteen **Todennäköisyyslaskenta** lukua **Stokastiikan konvergenssikäsitteet ja raja-arvolauseet**.