

Mat-1.2620 Sovellettu todennäköisyysslaskenta B**9. harjoitukset / Ratkaisut**

Aiheet: Estimointi
 Estimointimenetelmät
 Väliestimointi

Avainsanat:

Aritmeettinen keskiarvo, Bernoulli-jakauma, Bernoulli-koe, Estimaatti, Estimaattori, Frekvenssi, Frekvenssitulkinta, Harhaton estimaattori, Keskeinen raja-arvolause, Keskihajonta, χ^2 -jakauma, Logaritminen uskottavuusfunktio, Luottamuskerroin, Luottamustaso, Luottamusväli, Maksimointi, Momentti, Momenttiestimaattori, Momenttimenetelmä, Normaalijakauma, Odotusarvo, Otos, Otosjakauma, Otoskoko, Otosvarianssi, Riippumattomuus, Standardoitu normaali-jakauma, Suhteellinen frekvenssi, Suhteellinen osuus, Suurimman uskottavuuden estimaattori, Suurimman uskottavuuden menetelmä, t -jakauma, Todennäköisyys, Uskottavuusfunktio, Varianssi, Yksinkertainen satunnaisotos

Estimointi**Yksinkertainen satunnaisotos**

Olkoon

$$X_i, i = 1, 2, \dots, n$$

yksinkertainen satunnaisotos jakaumasta, jonka *pistetodennäköisyys-* tai *tiheysfunktio* $f(x; \theta)$ riippuu *parametrista* θ . Tällöin havainnot $X_i, i = 1, 2, \dots, n$ ovat *riippumattomia, identtisesti jakautuneita satunnaismuuttujia*, joilla on *sama pistetodennäköisyys-* tai *tiheysfunktio* $f(x; \theta)$:

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

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

Estimaattori ja estimaatti

Oletetaan, että todennäköisyysjakauman $f(x; \theta)$ *parametri* θ on *tuntematon* ja sen *estimoimiseen* käytetään havaintojen $X_i, i = 1, 2, \dots, n$ funktiota eli (*otos-*) *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**.

Otosjakauma

Oletetaan, että havainnot

$$X_i, i = 1, 2, \dots, n$$

muodostavat *yksinkertaisen satunnaisotoksen* jakaumasta $f(x; \theta)$ ja olkoon

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

jokin *parametrin θ estimaattori*.

Koska estimaattori T on *satunnaismuuttuja*, sillä on *todennäköisyysjakauma*, jota kutsutaan *estimaattorin T otosjakaumaksi*. Estimaattorin T otosjakauma muodostaa *tilastollisen mallin* eli *todennäköisyysmallin estimaattorin T arvojen satunnaisvaihtelulle otoksesta toiseen*.

Harhattomuus

Estimaattoria T sanotaan parametrin θ **harhattomaksi** *estimaattoriksi*, jos

$$E(T) = \theta$$

Tehokkuus

Estimaattori T_1 on **tehokkaampi** kuin estimaattori T_2 , jos

$$\text{Var}(T_1) < \text{Var}(T_2)$$

Täystehokkuus (minimivarianssisuus)

Estimaattori T sanotaan **täystehokkaaksi**, jos sen varianssi

$$\text{Var}(T)$$

on pienempi kuin minkä tahansa muun estimaattorin.

Estimointimenetelmät

Suurimman uskottavuuden menetelmä

Yksinkertainen satunnaisotos

Olkoon

$$X_i, i = 1, 2, \dots, n$$

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

Tällöin havainnot $X_i, i = 1, 2, \dots, n$ ovat riippumattomia, identtisesti jakautuneita satunnaismuuttujia, joilla on sama pistetodennäköisyys- tai tiheysfunktio $f(x; \theta)$:

$$\begin{aligned} X_1, X_2, \dots, X_n &\perp \\ X_i &\sim f(x; \theta), i = 1, 2, \dots, n \end{aligned}$$

Uskottavuusfunktio

Koska havainnot $X_i, i = 1, 2, \dots, n$ muodostavat yksinkertaisen satunnaisotoksen jakaumasta $f(x; \theta)$, otoksen yhteisjakauman tiheysfunktio on

$$f(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \times f(x_2; \theta) \times \dots \times f(x_n; \theta)$$

jossa

$$f(x_i; \theta), i = 1, 2, \dots, n$$

on yksittäiseen havaintoon X_i liittyvä pistetodennäköisyys- tai tiheysfunktio.

Otoksen $X_i, i = 1, 2, \dots, n$ uskottavuusfunktio

$$L(\theta; x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n; \theta)$$

on havaintojen $X_i, i = 1, 2, \dots, n$ yhteisjakauman pistetodennäköisyys- tai tiheysfunktion f arvo pisteessä

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

tulkittuna parametrin θ arvojen funktioksi. Uskottavuusfunktio L sisältää kaiken informaation otoksesta.

Suurimman uskottavuuden estimaattori

Olkoon

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

parametrin θ arvo, joka maksimoi otoksen $X_i, i = 1, 2, \dots, n$ uskottavuusfunktion

$$L(\theta; x_1, x_2, \dots, x_n)$$

parametrin θ suhteen. Huomaa, että uskottavuusfunktion L maksimin antava parametrin θ arvo t on muuttujien (havaintoarvojen) $x_1, x_2, \dots, x_n$ funktio.

Sijoittamalla uskottavuusfunktion L maksimin parametrin θ suhteen antavassa lausekkeessa

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

muuttujien

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

paikalle satunnaismuuttujat (havainnot)

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

saadaan parametrin θ suurimman uskottavuuden estimaattori eli **SU-estimaattori**

$$\hat{\theta} = g(X_1, X_2, \dots, X_n)$$

Parametrin θ suurimman uskottavuuden estimaattori $\hat{\theta}$ tuottaa parametrille θ arvon, joka *maksimoi poimitun otoksen eli saatujen havaintoarvojen uskottavuuden (todennäköisyyden)*. Siten parametrin θ suurimman uskottavuuden estimaattorin $\hat{\theta}$ otoskohtainen arvo *maksimoi todennäköisyyden saada juuri se otos, joka on saatu*.

Suurimman uskottavuuden estimaattorin määrittäminen

Parametrin θ suurimman uskottavuuden estimaattori määrätään *maksimoimalla uskottavuusfunktio*

$$L(\theta; x_1, x_2, \dots, x_n)$$

parametrin θ suhteen.

Kaikissa säännöllisissä tapauksissa maksimi löydetään merkitsemällä uskottavuusfunktion $L(\theta)$ *derivaatta*

$$L'(\theta)$$

nollaksi ja ratkaisemalla θ saadusta *normaaliyhtälöstä*

$$L'(\theta) = 0$$

Logaritminen uskottavuusfunktio

Uskottavuusfunktion maksimi kannattaa tavallisesti etsiä maksimoimalla uskottavuusfunktion sijasta *logaritmista uskottavuusfunktioita* (uskottavuusfunktion logaritmia)

$$l(\theta; x_1, x_2, \dots, x_n) = \log L(\theta; x_1, x_2, \dots, x_n)$$

Koska havainnot $X_1, X_2, \dots, X_n$ on oletettu tässä *riippumattomiksi*, *logaritminen uskottavuusfunktio* voidaan kirjoittaa seuraavaan muotoon:

$$\begin{aligned} l(\theta) &= \log L(\theta) \\ &= \log (f(x_1; \theta) \times f(x_2; \theta) \times \dots \times f(x_n; \theta)) \\ &= \log f(x_1; \theta) + \log f(x_2; \theta) + \dots + \log f(x_n; \theta) \\ &= l(\theta; x_1) + l(\theta; x_2) + \dots + l(\theta; x_n) \end{aligned}$$

jossa

$$l(\theta; x_i) = \log f(x_i; \theta), \quad i = 1, 2, \dots, n$$

on havaintoarvoon x_i liittyvä logaritminen uskottavuusfunktio.

Jos parametrin θ SU-estimaattori $\hat{\theta}$ ei täytä mitään hyvän estimaattorin kriteeriä *äärellisillä havaintojen lukumäärillä*, SU-estimaattorin $\hat{\theta}$ käyttöä parametrin θ estimaattorina voidaan kuitenkin usein perustella SU-estimaattorin *yleisillä asymptoottisilla ominaisuuksilla*:

Hyvin yleisin ehdoin pätee:

- (i) SU-estimaattori $\hat{\theta}$ on **tarkentuva** eli

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

- (ii) SU-estimaattori $\hat{\theta}$ on **asymptoottisesti normaalin**.

Normaalijakauman parametrin suurimman uskottavuuden estimaattorit

Satunnaismuuttuja X noudattaa **normaalijakaumaa**, jos sen *tiheysfunktio* on muotoa

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

Normaalijakauman *parametreina* ovat jakauman *odotusarvo*

$$E(X) = \mu$$

ja *varianssi*

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

Normaalijakauman $N(\mu, \sigma^2)$ odotusarvon μ ja varianssin σ^2 **SU-estimaattorit** ovat havaintojen $X_1, X_2, \dots, X_n$ *aritmeettinen keskiarvo*

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

ja *otosvarianssi*

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Normaalijakauman $N(\mu, \sigma^2)$ odotusarvon μ SU-estimaattorilla $\bar{X}$ on seuraavat ominaisuudet:

- (i) $\bar{X}$ on *harhaton*.
- (ii) $\bar{X}$ ja $\hat{\sigma}^2$ ovat yhdessä *tyhjentäviä* parametreille μ ja σ^2 .
- (iii) $\bar{X}$ on *tehokas* eli minimivarianssinen estimaattori.
- (iv) $\bar{X}$ on *tarkentuva*.
- (v) $\bar{X}$ noudattaa *normaalijakaumaa*:

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

Normaalijakauman $N(\mu, \sigma^2)$ varianssin σ^2 SU-estimaattorilla $\hat{\sigma}^2$ on seuraavat ominaisuudet:

(i) $\hat{\sigma}^2$ on *harhainen*, mutta estimaattori

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} \hat{\sigma}^2$$

on *harhaton*.

(ii) $\bar{X}$ ja $\hat{\sigma}^2$ ovat yhdessä *tyhjentäviä* parametreille μ ja σ^2 .

(iii) $\hat{\sigma}^2$ ei ole *tehokas* eli minimivarianssinen estimaattori.

(iv) $\hat{\sigma}^2$ on *tarkentuva*.

(v) $(n-1)s^2/\sigma^2$ noudattaa χ^2 -jakaumaa:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

Eksponenttijakauman parametrin suurimman uskottavuuden estimaattori

Satunnaismuuttuja X noudattaa **eksponenttijakaumaa** $\text{Exp}(\lambda)$, jos sen *tiheysfunktio* on

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

Eksponenttijakauman ainoana *parametrina* on

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

Eksponenttijakauman $\text{Exp}(\lambda)$ parametrin λ **SU-estimaattori** on

$$\hat{\lambda} = \frac{1}{\bar{X}}$$

jossa

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

on havaintojen $X_1, X_2, \dots, X_n$ *aritmeettinen keskiarvo*.

Bernoulli-jakauman odotusarvoparametrin suurimman uskottavuuden estimaattori

Olkoon A *tapahtuma*, jonka todennäköisyys on p :

$$\Pr(A) = p$$

Määritellään satunnaismuuttuja X seuraavasti:

$$X = \begin{cases} 1, & \text{jos } A \text{ tapahtuu} \\ 0, & \text{jos } A \text{ ei tapahdu} \end{cases}$$

Satunnaismuuttuja X noudattaa **Bernoulli-jakaumaa** parametrilla p :

$$X \sim \text{Ber}(p)$$

jossa

$$\Pr(A) = p = E(X)$$

Bernoulli-jakauman $\text{Ber}(p)$ odotusarvoparametrin p **SU-estimaattori** on havaintojen $X_1, X_2, \dots, X_n$ *aritmeettinen keskiarvo*

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

Huomaa, että

$$\bar{X} = \frac{f}{n}$$

jossa f on kiinnostuksen kohteena olevan tapahtuman A *suhteellinen frekvenssi* otoksessa.

Bernoulli-jakauman $\text{Ber}(p)$ odotusarvoparametrin p SU-estimaattorilla $\hat{p}$ on seuraavat ominaisuudet:

- (i) $\hat{p}$ on *harhaton*.
- (ii) $\hat{p}$ on *tyhjentävä*.
- (iii) $\hat{p}$ on *tehokas* eli minimivarianssinen estimaattori.
- (iv) $\hat{p}$ on *tarkentuva*.
- (v) $\hat{p}$ noudattaa *asymptoottisesti normaalijakaumaa*:

$$\hat{p} \sim_a N\left(p, \frac{pq}{n}\right)$$

Momenttimenetelmä

Olkoon

$$X_i, i = 1, 2, \dots, n$$

yksinkertainen satunnaisotos jakaumasta, jonka *pistetodennäköisyys-* tai *tiheysfunktio* $f(x; \theta)$, jonka *parametrina* on p -vektori

$$\theta = (\theta_1, \theta_2, \dots, \theta_p)$$

Tällöin havainnot $X_i, i = 1, 2, \dots, n$ ovat *riippumattomia, identtisesti jakautuneita satunnaismuuttujia*, joilla on *sama pistetodennäköisyys-* tai *tiheysfunktio* $f(x; \theta)$:

$$\begin{aligned} X_1, X_2, \dots, X_n &\perp \\ X_i &\sim f(x; \theta), i = 1, 2, \dots, n \end{aligned}$$

Oletetaan, että jakaumalla $f(x; \theta)$ on kaikki (*origo-*) *momentit* kertalukuun p saakka:

$$E(X_i^k) = \alpha_k, k = 1, 2, \dots, p, i = 1, 2, \dots, n$$

Oletetaan, että momenttien

$$\alpha_1, \alpha_2, \dots, \alpha_p$$

ja parametrien

$$\theta_1, \theta_2, \dots, \theta_p$$

välillä on jatkuva *bijektio* eli kääntäen yksikäsitteinen kuvaus:

$$(1) \quad \begin{cases} \alpha_1 = g_1(\theta_1, \theta_2, \dots, \theta_p) \\ \alpha_2 = g_2(\theta_1, \theta_2, \dots, \theta_p) \\ \vdots \\ \alpha_p = g_p(\theta_1, \theta_2, \dots, \theta_p) \end{cases}$$

Tällöin parametrit

$$\theta_1, \theta_2, \dots, \theta_p$$

voidaan esittää momenttien

$$\alpha_1, \alpha_2, \dots, \alpha_p$$

funktioina:

$$(2) \quad \begin{cases} \theta_1 = h_1(\alpha_1, \alpha_2, \dots, \alpha_p) \\ \theta_2 = h_2(\alpha_1, \alpha_2, \dots, \alpha_p) \\ \vdots \\ \theta_p = h_p(\alpha_1, \alpha_2, \dots, \alpha_p) \end{cases}$$

Estimoidaan momentit $\alpha_1, \alpha_2, \dots, \alpha_p$ vastaavilla *otosmomenteilla*:

$$a_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k=1, 2, \dots, p$$

Sijoittamalla estimaattorit $a_1, a_2, \dots, a_p$ momenttien $\alpha_1, \alpha_2, \dots, \alpha_p$ paikalle yhtälöihin (2), saadaan parametrien $\theta_1, \theta_2, \dots, \theta_p$ **momenttiestimaattorit** eli **MM-estimaattorit**

$$\begin{cases} \hat{\theta}_1 = h_1(a_1, a_2, \dots, a_p) \\ \hat{\theta}_2 = h_2(a_1, a_2, \dots, a_p) \\ \vdots \\ \hat{\theta}_p = h_p(a_1, a_2, \dots, a_p) \end{cases}$$

Normaalijakauman parametrien momenttiestimaattorit

Normaalijakauman $N(\mu, \sigma^2)$ odotusarvon μ ja varianssin σ^2 **MM-estimaattorit** ovat havaintojen $X_1, X_2, \dots, X_n$ *aritmeettinen keskiarvo*

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

ja *otosvarianssi*

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Eksponenttijakauman parametrin momenttiestimaattori

Eksponenttijakauman $\text{Exp}(\lambda)$ parametrin λ **MM-estimaattori** on

$$\hat{\lambda} = \frac{1}{\bar{X}}$$

jossa

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

on havaintojen $X_1, X_2, \dots, X_n$ *aritmeettinen keskiarvo*.

Bernoulli-jakauman odotusarvoparametrin momenttiestimaattori

Bernoulli-jakauman $\text{Ber}(p)$ odotusarvoparametrin p **MM-estimaattori** on havaintojen $X_1, X_2, \dots, X_n$ *aritmeettinen keskiarvo*

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

Huomaa, että

$$\bar{X} = \frac{f}{n}$$

jossa f on kiinnostuksen kohteena olevan tapahtuman A *suhteellinen frekvenssi* otoksessa.

Väliestimointi

Normaalijakauman odotusarvon luottamusväli, kun jakauman varianssi on tunnettu

Otos normaalijakaumasta

Olkoon

$$X_i, i = 1, 2, \dots, n$$

yksinkertainen satunnaisotos normaalijakaumasta $N(\mu, \sigma^2)$. Tällöin satunnaismuuttujat $X_i, i = 1, 2, \dots, n$ ovat riippumattomia ja noudattavat samaa normaalijakaumaa $N(\mu, \sigma^2)$:

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

$$X_i \sim N(\mu, \sigma^2), i = 1, 2, \dots, n$$

Normaalijakauman parametrien estimointi

Oletetaan, että normaalijakauman $N(\mu, \sigma^2)$ varianssi σ^2 on tunnettu ja estimoidaan odotusarvoparametri $E(X) = \mu$ sen *harhattomalla estimaattorilla*:

Havaintojen aritmeettinen keskiarvo

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

on odotusarvoparametrin $E(X) = \mu$ harhaton estimaattori.

Normaalijakauman odotusarvon luottamusväli, kun jakauman varianssi on tunnettu

Valitaan *luottamustasoksi*

$$1 - \alpha$$

Luottamustaso kiinnittää todennäköisyyden, jolla konstruoitava luottamusväli peittää normaalijakauman odotusarvon μ todellisen arvon.

Määrätään *luottamuskertoimet* $-z_{\alpha/2}$ ja $+z_{\alpha/2}$ siten, että

$$\Pr(z \leq -z_{\alpha/2}) = \frac{\alpha}{2}$$

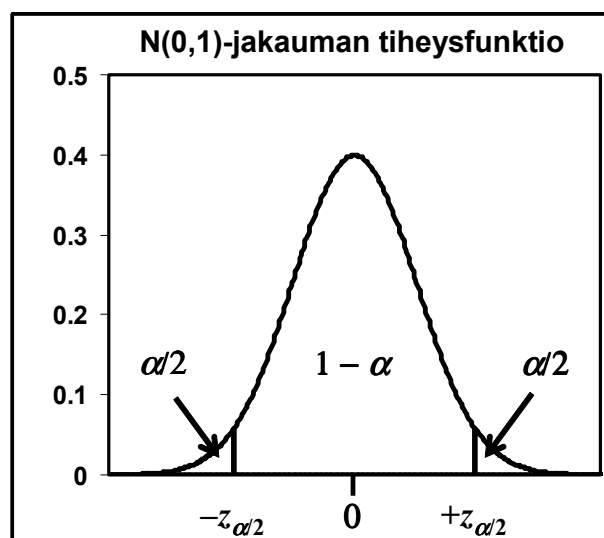
$$\Pr(z \geq +z_{\alpha/2}) = \frac{\alpha}{2}$$

jossa satunnaismuuttuja z noudattaa *standardoitua normaalijakaumaa*:

$$z \sim N(0,1)$$

Siten luottamuskertoimet $-z_{\alpha/2}$ ja $+z_{\alpha/2}$ toteuttavat ehdon

$$\Pr(-z_{\alpha/2} \leq z \leq +z_{\alpha/2}) = 1 - \alpha$$



Normaalijakauman *odotusarvoparametrin* μ *luottamusväli luottamustasolla* $(1 - \alpha)$ on tunnetun varianssin σ^2 tapauksessa muotoa

$$\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

jossa

$\bar{X}$ = havaintojen *aritmeettinen keskiarvo* otoksessa

σ^2 = jakauman *varianssi*

n = havaintojen *lukumäärä*

$-z_{\alpha/2}$ ja $+z_{\alpha/2}$ = luottamustasoon $(1 - \alpha)$ liittyvät *luottamuskertoimet* normaalijakaumasta $N(0,1)$

Luottamusvälin konstruktio perustuu siihen, että

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Koska luottamusväli on *symmetrinen* keskipisteensä $\bar{X}$ suhteen, luottamusväli esitetään usein muodossa

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Luottamusvälin *pituus* on

$$2 \times z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Luottamusvälin konstruktioista seuraa, että

$$\Pr \left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

Siten luottamusväli *peittää* parametrin μ todellisen arvon todennäköisyydellä $(1 - \alpha)$ ja se *ei peitä* parametrin μ todellista arvoa todennäköisyydellä α .

Luottamusvälin ominaisuudet

- (i) Normaalijakauman odotusarvon μ luottamusvälin *keskipiste* $\bar{X}$ vaihtelee otoksesta toiseen.
- (ii) Luottamusvälin *pituus ei vaihtelee* otoksesta toiseen.
- (iii) Luottamusvälin *pituus* riippuu valitusta luottamustasosta $(1 - \alpha)$, havaintojen lukumäärästä n ja jakauman varianssista σ^2 .
- (iv) Luottamusväli *lyhenee (pitenee)*, jos *luottamustasoa* $(1 - \alpha)$ *pienennetään (kasvatetaan)*.
- (v) Luottamusväli *lyhenee (pitenee)*, jos *havaintojen lukumäärää* n *kasvatetaan (pienennetään)*.
- (vi) Luottamusväli *lyhenee (pitenee)*, jos jakauman *varianssi* σ^2 *pienenee (kasvaa)*.

Luottamusvälin frekvenssitulkinta

Normaalijakauman odotusarvon μ luottamusvälillä on seuraava *frekvenssitulkinta*:

- (i) Jos otantaa jakaumasta $N(\mu, \sigma^2)$ toistetaan, *keskimäärin*

$$100 \times (1 - \alpha) \%$$

otoksista konstruoiduista luottamusväleistä *peittää* parametrin μ todellisen arvon.

- (ii) Jos otantaa jakaumasta $N(\mu, \sigma^2)$ toistetaan, *keskimäärin*

$$100 \times \alpha \%$$

otoksista konstruoiduista luottamusväleistä *ei peitä* parametrin μ todellista arvoa.

Johtopäätökset luottamusvälistä

Oletetaan, että olemme tehneet *johtopäätöksen*, että konstruoitu luottamusväli peittää odotusarvoparametrin μ todellisen arvon:

- (i) Luottamusvälin konstruktioista seuraa, että tehty johtopäätös *on oikea*

$$100 \times (1 - \alpha) \% \text{:ssa}$$

tapauksia.

- (ii) Luottamusvälin konstruktioista seuraa, että tehty johtopäätös *on väärä*

$$100 \times \alpha \% \text{:ssa}$$

tapauksia.

Virheellisen johtopäätöksen mahdollisuutta *ei saada häviämään*, ellei luottamusväliä tehdä *äärettömän leveäksi*, jolloin väli *ei enää sisällä informaatiota* odotusarvoparametrin μ todellisesta arvosta.

Vaatimukset luottamusvälille

Olisi toivottavaa pystyä konstruoimaan parametrille μ mahdollisimman *lyhyt* luottamusväli, johon liittyvä luottamustaso olisi samanaikaisesti mahdollisimman *korkea*.

Molempien vaatimusten samanaikainen täyttäminen *ei ole* kuitenkaan mahdollista, *jos otoskoko pidetään kiinteänä*:

- (i) *Luottamustason kasvattaminen pidentää luottamusväliä*, jolloin tieto parametrin μ todellisen arvon sijainnista tulee *epätarkemmaksi*.
- (ii) *Luottamusvälin lyhentäminen pienentää luottamustasoa*, jolloin tieto parametrin μ todellisen arvon sijainnista tulee *epävarmemmaksi*.

Otoskoon määrääminen

Oletetaan, että normaalijakauman odotusarvoparametrille μ halutaan konstruoida luottamusväli, jonka *toivottu pituus* on $2A$.

Tarvittava *otoskoko* saadaan kaavasta

$$n = \left(\frac{z_{\alpha/2} \sigma}{A} \right)^2$$

Normaalijakauman odotusarvon ja varianssin luottamusvälit, kun jakauman varianssi ei ole tunnettu

Otos normaalijakaumasta

Olkoon

$$X_i, i = 1, 2, \dots, n$$

yksinkertainen satunnaisotos normaalijakaumasta $N(\mu, \sigma^2)$. Tällöin satunnaismuuttujat $X_i, i = 1, 2, \dots, n$ ovat riippumattomia ja noudattavat samaa normaalijakaumaa $N(\mu, \sigma^2)$:

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

$$X_i \sim N(\mu, \sigma^2), i = 1, 2, \dots, n$$

Normaalijakauman parametrien estimointi

Estimoidaan normaalijakauman $N(\mu, \sigma^2)$ parametrit μ ja σ^2 niiden harhattomilla estimaattoreilla:

Havaintojen aritmeettinen keskiarvo

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

on odotusarvoparametrin $E(X) = \mu$ harhaton estimaattori.

Havaintojen otosvariassi

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

on varianssiparametrin $\text{Var}(X) = \sigma^2$ harhaton estimaattori.

Normaalijakauman odotusarvon luottamusväli, kun jakauman varianssi ei ole tunnettu

Valitaan luottamustasoksi

$$1 - \alpha$$

Luottamustaso kiinnittää todennäköisyyden, jolla konstruoitava luottamusväli peittää normaalijakauman odotusarvon μ todellisen arvon.

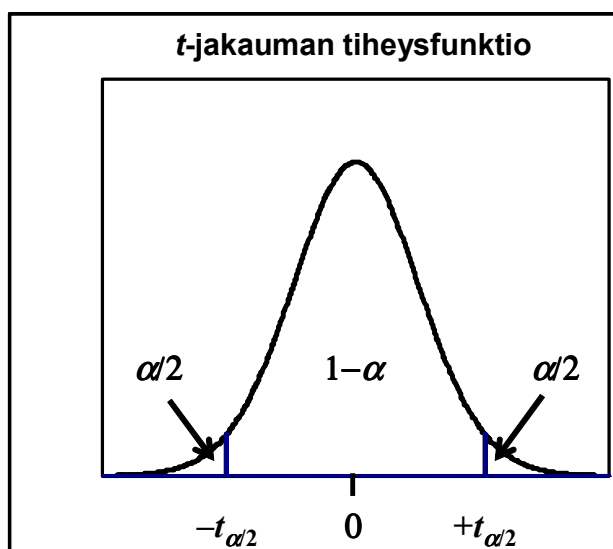
Määrätään luottamuskertoimet $-t_{\alpha/2}$ ja $+t_{\alpha/2}$ siten, että

$$\Pr(t \leq -t_{\alpha/2}) = \frac{\alpha}{2}$$

$$\Pr(t \geq +t_{\alpha/2}) = \frac{\alpha}{2}$$

jossa satunnaismuuttuja t noudattaa t -jakaumaa vapausastein $(n-1)$:

$$t \sim t(n-1)$$



Siten luottamuskertoimet $-t_{\alpha/2}$ ja $+t_{\alpha/2}$ toteuttavat ehdon

$$\Pr(-t_{\alpha/2} \leq t \leq +t_{\alpha/2}) = 1 - \alpha$$

Normaalijakauman odotusarvoparametrin μ luottamusväli luottamustasolla $(1 - \alpha)$ on tuntemattoman varianssin σ^2 tapauksessa muotoa

$$\left(\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right)$$

jossa

$\bar{X}$ = havaintojen aritmeettinen keskiarvo otoksessa

s^2 = otosvarianssi

n = havaintojen lukumäärä

$-t_{\alpha/2}$ ja $+t_{\alpha/2}$ = luottamustasoon $(1 - \alpha)$ liittyvät luottamuskertoimet t -jakaumasta $t(n - 1)$

Luottamusvälin konstruktio perustuu siihen, että

$$\frac{\bar{X} - \mu}{s / \sqrt{n}} \sim t(n - 1)$$

Koska luottamusväli on *symmetrinen* keskipisteensä $\bar{X}$ suhteen, luottamusväli esitetään usein muodossa

$$\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Luottamusvälin *pituus* on

$$2 \times t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Luottamusvälin konstruktioista seuraa, että

$$\Pr\left(\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

Siten luottamusväli *peittää* parametrin μ todellisen arvon todennäköisyydellä $(1 - \alpha)$ ja se *ei peitä* parametrin μ todellista arvoa todennäköisyydellä α .

Odotusarvon luottamusvälin ominaisuudet

- (i) Normaalijakauman odotusarvon μ luottamusvälin *keskipiste* $\bar{X}$ vaihtelee otoksesta toiseen.
- (ii) Luottamusvälin *pituus vaihtelee* otoksesta toiseen.
- (iii) Luottamusvälin *pituus* riippuu valitusta luottamustasosta $(1 - \alpha)$, havaintojen lukumäärästä n ja otosvarianssista s^2 .
- (iv) Luottamusväli *lyhenee (pitenee)*, jos *luottamustasoa* $(1 - \alpha)$ *pienennetään (kasvatetaan)*.
- (v) Luottamusväli *lyhenee (pitenee)*, jos *havaintojen lukumäärää* n *kasvatetaan (pienennetään)*.
- (vi) Luottamusväli *lyhenee (pitenee)*, jos *otosvarianssi* s^2 *pienenee (kasvaa)*.

Odotusarvon luottamusvälin frekvenssitulkinta

Normaalijakauman odotusarvon μ luottamusvälillä on seuraava *frekvenssitulkinta*:

- (i) Jos otantaa jakaumasta $N(\mu, \sigma^2)$ toistetaan, *keskimäärin*

$$100 \times (1 - \alpha) \%$$

otoksista konstruoiduista luottamusväleistä *peittää* parametrin μ todellisen arvon.

- (ii) Jos otantaa jakaumasta $N(\mu, \sigma^2)$ toistetaan, *keskimäärin*

$$100 \times \alpha \%$$

otoksista konstruoiduista luottamusväleistä *ei peitä* parametrin μ todellista arvoa.

Johtopäätökset odotusarvon luottamusvälistä

Oletetaan, että olemme tehneet *johtopäätöksen*, että konstruoitu luottamusväli peittää odotusarvoparametrin μ todellisen arvon:

- (i) Luottamusvälin konstruktioista seuraa, että tehty johtopäätös *on oikea*

$$100 \times (1 - \alpha) \% \text{:ssa}$$

tapauksia.

- (ii) Luottamusvälin konstruktioista seuraa, että tehty johtopäätös *on väärä*

$$100 \times \alpha \% \text{:ssa}$$

tapauksia.

Virheellisen johtopäätöksen mahdollisuutta *ei saada häviämään*, ellei luottamusväliä tehdä *äärettömän leveäksi*, jolloin väli *ei enää sisällä informaatiota* odotusarvoparametrin μ todellisesta arvosta.

Vaatimukset odotusarvon luottamusvälille

Olisi toivottavaa pystyä konstruoimaan odotusarvoparametrille μ mahdollisimman *lyhyt* luottamusväli, johon liittyvä luottamustaso olisi samanaikaisesti mahdollisimman *korkea*.

Molempien vaatimusten samanaikainen täyttäminen *ei ole* kuitenkaan mahdollista, jos otoskoko pidetään kiinteänä:

- (i) Luottamustason kasvattaminen pidentää luottamusväliä, jolloin tieto parametrin μ todellisen arvon sijainnista tulee epätarkemmaksi.
- (ii) Luottamusvälin lyhentäminen pienentää luottamustasoa, jolloin tieto parametrin μ todellisen arvon sijainnista tulee epävarmemmaksi.

Otoskoon määrittäminen

Oletetaan, että normaalijakauman odotusarvoparametrille μ halutaan konstruoida luottamusväli, jonka *toivottu pituus* on $2A$.

Tarvittava otoskoko saadaan kaavasta

$$n = \left(\frac{z_{\alpha/2} \sigma}{A} \right)^2$$

jossa

$z_{\alpha/2}$ = luottamustasoon $(1 - \alpha)$ liittyvä luottamuserroin normaalijakaumasta

Normaalijakauman varianssin luottamusväli

Valitaan luottamustasoksi

$$1 - \alpha$$

Luottamustaso kiinnittää todennäköisyyden, jolla konstruoitava luottamusväli peittää normaalijakauman varianssin σ^2 todellisen arvon.

Määritään luottamuskertoimet $\chi^2_{1-\alpha/2}$ ja $\chi^2_{\alpha/2}$ siten, että

$$\Pr(\chi^2 \leq \chi^2_{1-\alpha/2}) = \frac{\alpha}{2}$$

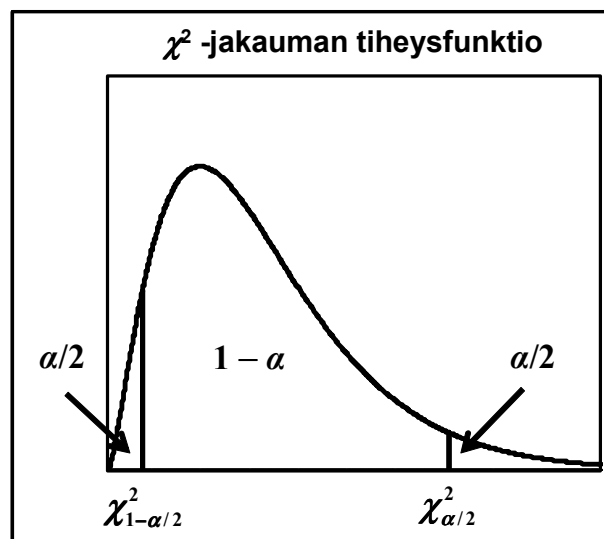
$$\Pr(\chi^2 \geq \chi^2_{\alpha/2}) = \frac{\alpha}{2}$$

jossa satunnaismuuttuja χ^2 noudattaa χ^2 -jakaumaa vapausastein $(n - 1)$:

$$\chi^2 \sim \chi^2(n-1)$$

Siten luottamuskertoimet $\chi^2_{1-\alpha/2}$ ja $\chi^2_{\alpha/2}$ toteuttavat ehdon

$$\Pr(\chi^2_{1-\alpha/2} \leq \chi^2 \leq \chi^2_{\alpha/2}) = 1 - \alpha$$



Normaalijakauman *varianssiparametrin* σ^2 *luottamusväli luottamustasolla* $(1 - \alpha)$ on muotoa

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \right)$$

jossa

$$s^2 = \text{otosvarianssi}$$

$$n = \text{havaintojen lukumäärä}$$

$$\chi^2_{1-\alpha/2} \text{ ja } \chi^2_{\alpha/2} = \text{luottamustasoon } (1 - \alpha) \text{ liittyvät luottamuskertoimet } \chi^2\text{-jakaumasta vapausastein } (n - 1)$$

Luottamusvälin konstruktio perustuu siihen, että

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

Luottamusvälin *pituus* on

$$(n-1)s^2 \left(\frac{1}{\chi^2_{1-\alpha/2}} - \frac{1}{\chi^2_{\alpha/2}} \right)$$

Luottamusvälin konstruktioista seuraa, että

$$\Pr \left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \right) = 1 - \alpha$$

Siten konstruoitu luottamusväli *peittää* parametrin σ^2 todellisen arvon todennäköisyydellä $(1 - \alpha)$ ja se *ei peitä* parametrin σ^2 todellista arvoa todennäköisyydellä α .

Varianssin luottamusvälin ominaisuudet

- (i) Normaalijakauman varianssin σ^2 luottamusvälin *pituus* vaihtelee otoksesta toiseen.
- (ii) Luottamusvälin *pituus* riippuu valitusta luottamustasosta $(1 - \alpha)$, havaintojen lukumäärästä n ja otosvarianssista s^2 .
- (iii) Luottamusväli *lyhenee (pitenee)*, jos *luottamustasoa* $(1 - \alpha)$ *pienennetään (kasvatetaan)*.
- (iv) Luottamusväli *lyhenee (pitenee)*, jos *havaintojen lukumäärää* n *pienennetään (kasvatetaan)*.
- (v) Luottamusväli *lyhenee (pitenee)*, jos *otosvarianssi* s^2 *pienenee (kasvaa)*.

Varianssin luottamusvälin frekvenssitulkinta

Normaalijakauman odotusarvon σ^2 luottamusvälillä on seuraava *frekvenssitulkinta*:

- (i) Jos otantaa jakaumasta $N(\mu, \sigma^2)$ toistetaan, *keskimäärin*

$$100 \times (1 - \alpha) \%$$

otoksista konstruoiduista luottamusväleistä *peittää* parametrin σ^2 todellisen arvon.

- (ii) Jos otantaa jakaumasta $N(\mu, \sigma^2)$ toistetaan, *keskimäärin*

$$100 \times \alpha \%$$

otoksista konstruoiduista luottamusväleistä *ei peitä* parametrin σ^2 todellista arvoa.

Johtopäätökset varianssin luottamusvälistä

Oletetaan, että olemme tehneet *johtopäätöksen*, että konstruoitu luottamusväli peittää varianssi-parametrin σ^2 todellisen arvon:

- (i) Luottamusvälin konstruktiosta seuraa, että tehty johtopäätös *on oikea*

$$100 \times (1 - \alpha) \% \text{:ssa}$$

tapauksia.

- (ii) Luottamusvälin konstruktiosta seuraa, että tehty johtopäätös *on väärä*

$$100 \times \alpha \% \text{:ssa}$$

tapauksia.

Vaatimukset varianssin luottamusvälille

Olisi toivottavaa pystyä konstruoimaan varianssiparametrille σ^2 mahdollisimman *lyhyt* luottamusväli, johon liittyvä luottamustaso olisi samanaikaisesti mahdollisimman *korkea*.

Vaatimusten samanaikainen täyttäminen *ei ole* kuitenkaan mahdollista:

- (i) *Luottamustason kasvattaminen pidentää luottamusväliä*, jolloin tieto parametrin σ^2 todellisen arvon sijainnista tulee *epätarkemmaksi*.
- (ii) *Luottamusvälin lyhentäminen pienentää luottamustasoa*, jolloin tieto parametrin σ^2 todellisen arvon sijainnista tulee *epävarmemmaksi*.

Bernoulli-jakauman odotusarvon luottamusväli

Bernoulli-jakauma

Olkoon A on jokin *tapahtuma* ja olkoon

$$\Pr(A) = p$$

$$\Pr(A^c) = 1 - p = q$$

Määritellään satunnaismuuttuja

$$X = \begin{cases} 1, & \text{jos } A \text{ tapahtuu} \\ 0, & \text{jos } A \text{ ei tapahdu} \end{cases}$$

Tällöin satunnaismuuttuja X noudattaa **Bernoulli-jakaumaa** parametrinaan

$$p = \Pr(A) = E(X)$$

Merkitään:

$$X \sim \text{Ber}(p)$$

Bernoulli-jakauman *pistetodennäköisyysfunktio* on

$$f(x; p) = p^x (1 - p)^{1-x}, \quad x = 0, 1; 0 < p < 1$$

Otos Bernoulli-jakaumasta

Olkoon

$$X_i, i = 1, 2, \dots, n$$

yksinkertainen satunnaisotos Bernoulli-jakaumasta $\text{Ber}(p)$. Tällöin satunnaismuuttujat $X_i, i = 1, 2, \dots, n$ ovat *riippumattomia* ja noudattavat *samaa Bernoulli-jakaumaa* $\text{Ber}(p)$:

$$\begin{aligned} X_1, X_2, \dots, X_n &\perp \\ X_i &\sim \text{Ber}(p), i = 1, 2, \dots, n \end{aligned}$$

Bernoulli-jakauman odotusarvoparametrin estimointi

Estimoidaan Bernoulli-jakauman $\text{Ber}(p)$ odotusarvoparametri p sen *harhattomalla estimaattorilla*:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$$

Koska

$$X_i = \begin{cases} 1, & \text{jos } A \text{ tapahtuu} \\ 0, & \text{jos } A \text{ ei tapahdu} \end{cases}, \quad i = 1, 2, \dots, n$$

niin

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{f}{n}$$

jossa f on tapahtuman A *frekvenssi* otoksessa. Siten Bernoulli-jakauman odotusarvoparametrin p *estimaattori* $\hat{p}$ on tapahtuman A *suhteellinen frekvenssi* otoksessa.

Huomaa, että

$$f \sim \text{Bin}(n, p)$$

Bernoulli-jakauman odotusarvoparametrin luottamusväli

Valitaan *luottamustasoksi*

$$1 - \alpha$$

Luottamustaso kiinnittää todennäköisyyden, jolla konstruoitava luottamusväli peittää Bernoulli-jakauman odotusarvoparametrin p todellisen arvon.

Määrätään *luottamuskertoimet* $-z_{\alpha/2}$ ja $+z_{\alpha/2}$ siten, että

$$\Pr(z \leq -z_{\alpha/2}) = \frac{\alpha}{2}$$

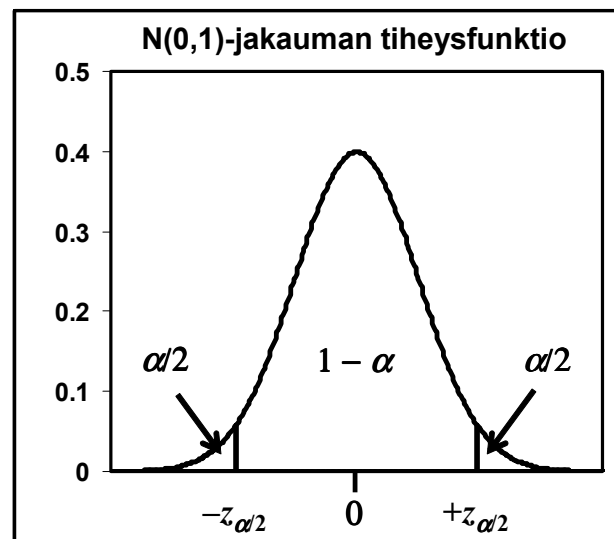
$$\Pr(z \geq +z_{\alpha/2}) = \frac{\alpha}{2}$$

jossa satunnaismuuttuja z noudattaa *standardoitua normaalijakaumaa*:

$$z \sim N(0,1)$$

Siten luottamuskertoimet $-z_{\alpha/2}$ ja $+z_{\alpha/2}$ toteuttavat ehdon

$$\Pr(-z_{\alpha/2} \leq z \leq +z_{\alpha/2}) = 1 - \alpha$$



Bernoulli-jakauman *odotusarvoparametrin* p *approksimatiivinen luottamusväli luottamustasolla* $(1 - \alpha)$ on muotoa

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

jossa

$\hat{p}$ = odotusarvoparametrin p *harhaton estimaattori*

n = havaintojen *lukumäärä*

$-z_{\alpha/2}$ ja $+z_{\alpha/2}$ = luottamustasoon $(1 - \alpha)$ liittyvät *luottamuskertoimet* normaalijakaumasta $N(0,1)$

Luottamusvälin konstruktio perustuu siihen, että *keskeisen raja-arvolauseen* mukaan

$$\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \sim_a N(0,1)$$

Koska luottamusväli on *symmetrinen* keskipisteensä $\hat{p}$ suhteen, luottamusväli esitetään usein muodossa

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Luottamusvälin *pituus* on

$$2 \times z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Luottamusvälin konstruktioista seuraa, että

$$\Pr \left(\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right) =_a 1 - \alpha$$

Siten luottamusväli *peittää* parametrin p todellisen arvon approksimatiivisesti todennäköisyydellä $(1 - \alpha)$ ja se *ei peitä* parametrin p todellista arvoa approksimatiivisesti todennäköisyydellä α .

Luottamusvälin ominaisuudet

- (i) Bernoulli-jakauman odotusarvoparametrin p luottamusvälin *keskipiste* $\hat{p}$ vaihtelee otoksesta toiseen.
- (ii) Luottamusvälin *pituus vaihtelee* otoksesta toiseen.
- (iii) Luottamusvälin *pituus* riippuu valitusta luottamustasosta $(1 - \alpha)$, havaintojen lukumäärästä n ja estimaattorista $\hat{p}$.
- (iv) Luottamusväli *lyhenee (pitenee)*, jos *luottamustasoa* $(1 - \alpha)$ *pienennetään (kasvatetaan)*.
- (v) Luottamusväli *lyhenee (pitenee)*, jos *havaintojen lukumäärää* n *kasvatetaan (pienennetään)*.
- (vi) Luottamusväli on *lyhimmillään*, kun

$$\hat{p} \approx 0 \text{ tai } 1$$

- (vii) Luottamusväli on *pisimmillään*, kun

$$\hat{p} = \frac{1}{2}$$

Luottamusvälin frekvenssitulkinta

Bernoulli-jakauman odotusarvoparametrin p approksimatiivisella luottamusvälillä on seuraava *frekvenssitulkinta*:

- (i) Jos otantaa jakaumasta $\text{Ber}(p)$ toistetaan, *keskimäärin*

$$100 \times (1 - \alpha) \%$$

otoksista konstruoiduista luottamusväleistä *peittää* parametrin p todellisen arvon.

- (ii) Jos otantaa jakaumasta $\text{Ber}(p)$ toistetaan, *keskimäärin*

$$100 \times \alpha \%$$

otoksista konstruoiduista luottamusväleistä *ei peitä* parametrin p todellista arvoa.

Johtopäätökset luottamusvälistä

Oletetaan, että olemme tehneet *johtopäätöksen*, että luottamusväli peittää odotusarvoparametrin p todellisen arvon:

- (i) Luottamusvälin konstruktiosta seuraa, että tehty johtopäätös *on oikea*

$$100 \times (1 - \alpha) \% \text{:ssa}$$

tapauksia.

- (ii) Luottamusvälin konstruktiosta seuraa, että tehty johtopäätös *on väärä*

$$100 \times \alpha \% \text{:ssa}$$

tapauksia.

Virheellisen johtopäätöksen mahdollisuutta *ei saada häviämään*, ellei luottamusväliä tehdä *äärettömän leveäksi*, jolloin väli *ei enää sisällä informaatiota* odotusarvoparametrin p todellisesta arvosta.

Vaatimukset luottamusvälille

Olisi toivottavaa pystyä konstruoimaan parametrille p mahdollisimman *lyhyt* luottamusväli, johon liittyvä luottamustaso olisi samanaikaisesti mahdollisimman *korkea*.

Molempien vaatimusten samanaikainen täyttäminen *ei ole* kuitenkaan mahdollista, *jos otoskoko pidetään kiinteänä*:

- (i) *Luottamustason kasvattaminen pidentää luottamusväliä*, jolloin tieto parametrin p todellisen arvon sijainnista tulee *epätarkemmaksi*.
- (ii) *Luottamusvälin lyhentäminen pienentää luottamustasoa*, jolloin tieto parametrin p todellisen arvon sijainnista tulee *epävarmemmaksi*.

Otoskoon määrääminen

Oletetaan, että Bernoulli-jakauman odotusarvoparametrille p halutaan konstruoida luottamusväli, jonka *toivottu pituus* on

$$2A$$

Tarvittava *otoskoko* saadaan kaavasta

$$n = \left(\frac{z_{\alpha/2} \sqrt{p(1-p)}}{A} \right)^2$$

Tarvittava otoskoko *saavuttaa maksiminsa*

$$n = \left(\frac{z_{\alpha/2}}{2A} \right)^2$$

kun

$$p = \frac{1}{2}$$

Tehtävä 9.1.

Kolme tutkijaa A, B ja C ovat määrittäneet erään teollisuuslaitoksen jätevesistä pH-arvoja tavoitteenaan estimoida jätevesien keskimääräinen pH-arvo μ havaintojen perusteella. Määritykset tehtiin keräämällä useita toisistaan riippumattomia vesinäytteitä ja määrämällä näytekohtaisten pH-arvojen keskiarvot.

Tutkijoiden saamat tulokset:

Tutkija	Näytteiden lukumäärä	pH-lukujen aritmeettinen keskiarvo
A	10	7.4
B	15	7.7
C	200	6.2

- (a) Näytä, että estimaattorit

$$\bar{X}_A, \bar{X}_B, \bar{X}_C \text{ ja } \bar{X}_{ABC} = \frac{\bar{X}_A + \bar{X}_B + \bar{X}_C}{3}$$

ovat *harhattomia* keskimääräiselle pH-arvolle μ .

- (b) Mikä estimaattoreista on luotettavin siinä mielessä, että sen varianssi on *pienin*?
- (c) Näytä, että vielä pienempi varianssi kuin yhdelläkään ym. estimaattorilla on sellaisella estimaattorilla, joka saadaan laskemalla näytteiden lukumäärillä painotettu aritmeettinen keskiarvo (ts. aritmeettinen keskiarvo, joka saadaan yhdistämällä tutkijoiden aineistot ja laskemalla yhdistetyn aineiston pH-lukujen aritmeettinen keskiarvo).

Tehtävä 9.1. – Mitä opimme?

Tehtävässä vertaillaan todennäköisyysjakauman parametrin erilaisia *harhattomia estimaattoreita, estimaattoreiden odotusarvoja ja variansseja*.

Tehtävä 9.1. – Ratkaisu:

Jos

$$X_i, i = 1, 2, \dots, n$$

on yksinkertainen satunnaisotos jakaumasta, jonka odotusarvo on μ ja varianssi on σ^2 , niin havainnot $X_i, i = 1, 2, \dots, n$ ovat *riippumattomia* satunnaismuuttujia, joilla on sama odotusarvo ja varianssi:

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

$$E(X_i) = \mu, i = 1, 2, \dots, n$$

$$\text{Var}(X_i) = \sigma^2, i = 1, 2, \dots, n$$

Olkoon

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

havaintojen X_i , $i = 1, 2, \dots, n$ *aritmeettinen keskiarvo*.

Tällöin

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

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

(a)&(b)

Estimaattorit $\bar{X}_A$, $\bar{X}_B$, $\bar{X}_C$ ja $\bar{X}_{ABC}$ ovat *harhattomia*, koska

$$E(\bar{X}_A) = E(\bar{X}_B) = E(\bar{X}_C) = \mu$$

ja

$$\begin{aligned} E(\bar{X}_{ABC}) &= E\left[\frac{1}{3}(\bar{X}_A + \bar{X}_B + \bar{X}_C)\right] \\ &= \frac{1}{3}[E(\bar{X}_A) + E(\bar{X}_B) + E(\bar{X}_C)] \\ &= \frac{1}{3}(\mu + \mu + \mu) = \mu \end{aligned}$$

Estimaattoreiden $\bar{X}_A$, $\bar{X}_B$, $\bar{X}_C$ varianssit ovat

$$\text{Var}(\bar{X}_A) = \frac{\sigma^2}{10} = 0.1\sigma^2$$

$$\text{Var}(\bar{X}_B) = \frac{\sigma^2}{15} = 0.067\sigma^2$$

$$\text{Var}(\bar{X}_C) = \frac{\sigma^2}{200} = 0.005\sigma^2$$

Estimaattoreista $\bar{X}_A$, $\bar{X}_B$, $\bar{X}_C$ *luotettavin* on estimaattori $\bar{X}_C$, koska sen varianssi on *pienin*, mikä johtuu siitä, että se perustuu *suurimpaan* näytteiden lukumäärään.

Koska estimaattorit $\bar{X}_A$, $\bar{X}_B$, $\bar{X}_C$ ovat *riippumattomia*, niin

$$\begin{aligned} \text{Var}(\bar{X}_{ABC}) &= \text{Var}\left[\frac{1}{3}(\bar{X}_A + \bar{X}_B + \bar{X}_C)\right] \\ &= \frac{1}{9}[\text{Var}(\bar{X}_A) + \text{Var}(\bar{X}_B) + \text{Var}(\bar{X}_C)] \\ &= \frac{1}{9}\left(\frac{\sigma^2}{10} + \frac{\sigma^2}{15} + \frac{\sigma^2}{200}\right) = 0.019\sigma^2 \end{aligned}$$

Siten estimaattorin $\bar{X}_{ABC}$ varianssi on *suurempi* kuin estimaattorin $\bar{X}_C$, mikä johtuu siitä, että luotettavinta estimaattoria $\bar{X}_C$ epäluotettavimmat estimaattorit $\bar{X}_A$ ja $\bar{X}_B$ saavat estimaattorissa $\bar{X}_{ABC}$ yhtä suuren painon kuin estimaattori $\bar{X}_C$.

- (c) Määritellään näytteiden lukumäärillä *painotettu* aritmeettinen keskiarvo $\bar{X}$ kaavalla

$$\bar{X} = \frac{10\bar{X}_A + 15\bar{X}_B + 200\bar{X}_C}{10 + 15 + 200}$$

Estimaattori $\bar{X}$ on *harhaton* parametrille μ , koska

$$\begin{aligned} E(\bar{X}) &= \frac{10}{225} E(\bar{X}_A) + \frac{15}{225} E(\bar{X}_B) + \frac{200}{225} E(\bar{X}_C) \\ &= \frac{10}{225} \mu + \frac{15}{225} \mu + \frac{200}{225} \mu \\ &= \mu \end{aligned}$$

Estimaattorin $\bar{X}$ varianssiksi saadaan aritmeettisten keskiarvojen $\bar{X}_A, \bar{X}_B, \bar{X}_C$ *riippumattomuuden* takia:

$$\begin{aligned} \text{Var}(\bar{X}) &= \frac{1}{225^2} [10^2 \text{Var}(\bar{X}_A) + 15^2 \text{Var}(\bar{X}_B) + 200^2 \text{Var}(\bar{X}_C)] \\ &= \frac{1}{225^2} \left(\frac{10^2}{10} + \frac{15^2}{15} + \frac{200^2}{200} \right) \sigma^2 \\ &= \frac{\sigma^2}{225} = 0.0044\sigma^2 \end{aligned}$$

Siten estimaattorin $\bar{X}$ varianssi on *pienempi* kuin estimaattoreiden $\bar{X}_A, \bar{X}_B, \bar{X}_C$ ja $\bar{X}_{ABC}$, mikä johtuu siitä, että estimaattorissa $\bar{X}$ estimaattorit $\bar{X}_A, \bar{X}_B, \bar{X}_C$ ovat saaneet painoikseen niihin liittyvien havaintojen lukumäärät.

Huomautuksia:

- (i) Todennäköisyysjakauman parametreilla on aina useita erilaisia *harhattomia* estimaattoreita; kaikilla niillä on siis yhtä suuret odotusarvot, mutta niiden varianssit eivät välttämättä ole yhtä suuria.
- (ii) Harhattomista estimaattoreita *parhaimpana* voidaan pitää sitä, jonka *varianssi on pienin*. Tätä vaatimusta on tapana kutsua (*täys-*) *tehokkuus*- tai *minimivarianssisuus-kriteeriksi*.

Tehtävä 9.2.

Satunnaismuuttujan X tiheysfunktio on

$$f(x) = (1 + \theta)x^\theta, \quad 0 < x < 1$$

Kysymys: Miksi parametrin θ pitää toteuttaa ehto $\theta > -1$?

Oletetaan, että satunnaismuuttujasta X on saatu havainnot

$$0.5, 0.3, 0.1, 0.1, 0.2$$

- Hahmottele tiheysfunktion kuvaaja parametrin θ arvoilla $-0.5, 0, 1, 2$ ja arvioi mikä arvoista sopisi parhaiten havaintoihin.
- Estimoi parametri θ *momenttimenetelmällä*.
- Estimoi parametri θ *suurimman uskottavuuden menetelmällä*.
- Vertaa parametrin θ momenttiestimaattoria ja suurimman uskottavuuden estimaattoria toisiinsa.

Tehtävä 9.2. – Mitä opimme?

Tehtävässä näytetään, että *suurimman uskottavuuden menetelmä* ja *momenttimenetelmä* eivät välttämättä tuota samaa estimaattoria todennäköisyysjakauman parametreille.

Tehtävä 9.2. – Ratkaisu:

Koska $f(x)$ on *tiheysfunktio*, sen pitää toteuttaa ehto

$$f(x) \geq 0, \quad 0 < x < 1$$

mikä toteutuu, jos

$$1 + \theta \geq 0$$

Parametrin arvo $\theta = -1$ ei kuitenkaan käy, koska tällöin

$$f(x) \equiv 0, \quad 0 < x < 1$$

Siten parametrin θ on toteuttava ehto

$$\theta > -1$$

Tämä ehto myös riittävä, koska tällöin

$$\int_{-\infty}^{+\infty} f(x) dx = \int_0^1 (1 + \theta)x^\theta dx = [x^{1+\theta}]_0^1 \equiv 1$$

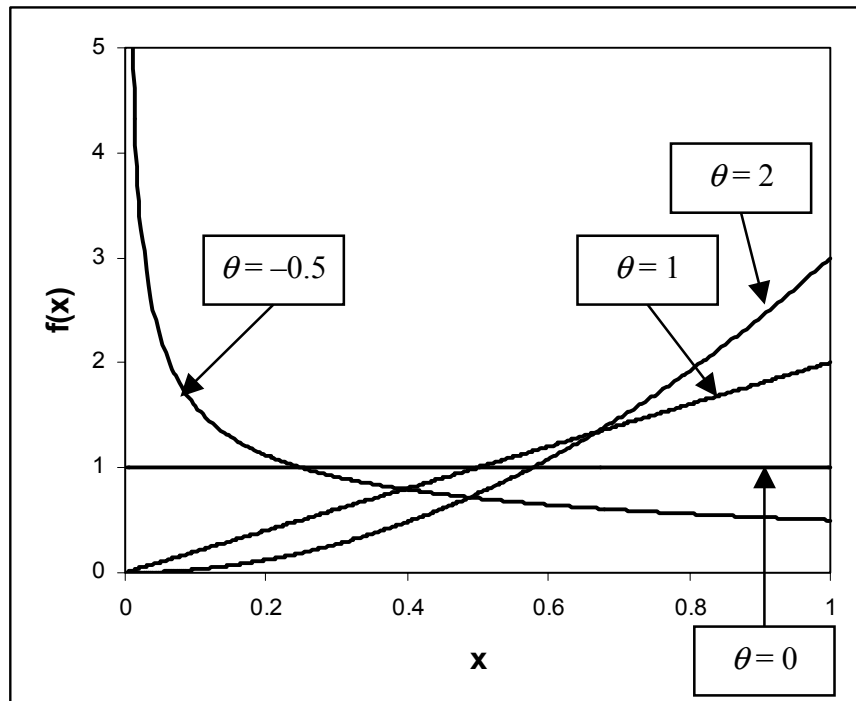
- Kuvio alla esittää tiheysfunktia

$$f(x) = (1 + \theta)x^\theta, \quad 0 < x < 1$$

kun

$$\theta = -0.5, 0, 1, 2$$

Kuviosta nähdään, että havainnot sopivat parhaiten jakaumaan, jossa $\theta < 0$, koska tällöin suuri osa jakauman todennäköisyysmassasta keskittyy välin $(0,1)$ vasemmanpuoleiseen päähän kuten havainnot.



(b) Estimoidaan parametri θ momenttimenetelmällä.

Määritetään ensin satunnaismuuttujan X odotusarvo:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 x(1+\theta)x^\theta dx = \left[\frac{1+\theta}{2+\theta} x^{2+\theta} \right]_0^1 = \frac{1+\theta}{2+\theta}$$

Parametrin θ momenttiestimaattori $\hat{\theta}_{MM}$ toteuttaa yhtälön

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

jossa

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

on havaintojen aritmeettinen keskiarvo.

Siten

$$\bar{X} = \frac{1+\hat{\theta}_{MM}}{2+\hat{\theta}_{MM}}$$

Koska tarkastellussa otoksessa

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{5} (0.5 + 0.3 + 0.1 + 0.1 + 0.2) = 0.24$$

niin saadaan yhtälö

$$0.24 = \frac{1 + \hat{\theta}_{MM}}{2 + \hat{\theta}_{MM}}$$

josta edelleen

$$\hat{\theta}_{MM} = -0.684$$

- (c) Estimoidaan parametri θ suurimman uskottavuuden menetelmällä.

Riippumattoman otoksen $X_i, i = 1, 2, \dots, n$ uskottavuusfunktio on

$$\begin{aligned} L(\theta; x_1, x_2, \dots, x_n) &= f(x_1, x_2, \dots, x_n; \theta) \\ &= (1 + \theta)x_1^\theta \times (1 + \theta)x_2^\theta \times \dots \times (1 + \theta)x_n^\theta \\ &= (1 + \theta)^n u^\theta \end{aligned}$$

missä

$$u = x_1 x_2 \cdots x_n$$

Parametrin θ suurimman uskottavuuden estimaattori $\hat{\theta}_{ML}$ saadaan maksimoimalla uskottavuusfunktio $L(\theta)$ parametrin θ suhteen. Tämä tapahtuu etsimällä funktion L derivaatan nollakohdat:

$$L'(\theta) = 0 \Leftrightarrow (1 + \theta)^{n-1} u^\theta (n + (1 + \theta) \log u) = 0 \Leftrightarrow n + (1 + \theta) \log u = 0$$

Siten

$$\hat{\theta}_{ML} = -1 - \frac{n}{\log u}$$

josta saadaan

$$\hat{\theta}_{ML} = -0.384$$

koska $n = 5$ ja $u = 0.0003$.

- (d) Momenttimenetelmä ja suurimman uskottavuuden menetelmä tuottavat tässä eri tuloksen.

Huomautuksia:

- (i) Momenttimenetelmä ja suurimman uskottavuuden menetelmä saattavat tuottaa todennäköisyysjakauman parametreille *eri estimaattorit*.
- (ii) *Hyvän estimaattorin* valinta on vaikea ongelma; juuri sen takia *estimaattoreiden vertailuun* käytetään sellaisia *hyvyysominaisuuksia* kuten *harhattomuus*, *tehokkuus*, *tyhjentävyys* ja *tarkentuvuus*.
- (iii) *Suurimman uskottavuuden estimaattorille* voidaan hyvin yleisin ehdoin todistaa *tyhjentävyys* ja *tarkentuvuus* sekä *asymptoottinen normaalisuus*.

Tehtävä 9.3.

Olkoot $X_i, i = 1, 2, \dots, n$ riippumattomia, samaa eksponenttijakaumaa noudattavia satunnaismuuttujia, joiden odotusarvo $E(X_i) = \beta$, ts. satunnaismuuttujat X_i muodostavat yksinkertaisen satunnaisotoksen eksponenttijakaumasta, jonka parametri on $1/\beta$. Määää parametrin β suurimman uskottavuuden estimaattori.

Tehtävä 9.3. – Mitä opimme?

Tehtävässä tarkastellaan eksponenttijakauman parametrin *suurimman uskottavuuden estimointia*.

Tehtävä 9.3. – Ratkaisu:

Oletetaan, että $X_i, i = 1, 2, \dots, n$ on yksinkertainen satunnaisotos eksponenttijakaumasta, jonka parametrina on $1/\beta$. Siten

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

$$X_i \sim \text{Exp}(1/\beta), i = 1, 2, \dots, n$$

Otoksen $X_1, X_2, \dots, X_n$ uskottavuusfunktio on

$$L(x_1, x_2, \dots, x_n; \beta) = f(x_1; \beta) \times f(x_2; \beta) \times \dots \times f(x_n; \beta) = \frac{1}{\beta^n} \exp\left(-\frac{1}{\beta} \sum_{i=1}^n x_i\right)$$

jossa

$$f(x_i; \beta) = \frac{1}{\beta} \exp\left(-\frac{1}{\beta} x_i\right), i = 1, 2, \dots, n$$

on havaintoon X_i liittyvä tiheysfunktio.

Vastaava *logaritminen uskottavuusfunktio* on

$$l(x_1, x_2, \dots, x_n; \beta) = \log L(x_1, x_2, \dots, x_n; \beta) = -\frac{1}{\beta} \sum_{i=1}^n x_i - n \log(\beta)$$

Suurimman uskottavuuden estimaattori parametrille θ löydetään maksimoimalla logaritmisen uskottavuusfunktio l parametrin θ suhteen.

Tämä tapahtuu derivoimalla logaritminen uskottavuusfunktio l parametrin θ suhteen, merkitsemällä derivaatta nolaksi ja *ratkaisemalla saatu normaaliyhtälö* parametrin θ suhteen:

$$\frac{\partial l(x_1, x_2, \dots, x_n; \beta)}{\partial \beta} = \frac{1}{\beta^2} \sum_{i=1}^n x_i - n \frac{1}{\beta} = 0$$

Ratkaisuksi saadaan

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

Saatu ratkaisu antaa logaritmisen uskottavuusfunktion maksimin, mikä nähdään esim. sijoittamalla saatu ratkaisu logaritmisen uskottavuusfunktion 2. derivaatan lausekkeeseen.

Tehtävä 9.4.

Tehdas väittää, että sen valmistamista tuotteista korkeintaan 5 % on viallisia. Asiakas poimii tuotteiden joukosta yksinkertaisen satunnaisotoksen, jonka koko on 150 ja löytää 15 viallista tuotetta. Voidaanko tehtaan väitettä viallisten suhteellisesta osuudesta pitää oikeutettuna?

Ohje: Määrää otoksesta 95 %:n ja 99 %:n *luottamusväli* tehtaan väittämälle viallisten suhteelliselle osuudelle ja tee johtopäätös niiden perusteella.

Lisäkysymys: Miten valittu luottamustaso vaikuttaa luottamusvälin pituuteen?

Tehtävä 9.4. – Mitä opimme?

Tehtävässä tarkastellaan *Bernoulli-jakauman odotusarvon luottamusvälin* määrittämistä.

Tehtävä 9.4. – Ratkaisu:

Tarkastellaan tapahtumaa

$$A = \{\text{Satunnaisesti valittu tuote on viallinen}\}$$

ja olkoon tapahtuman A todennäköisyys

$$\Pr(A) = p$$

$$\Pr(A^c) = 1 - p = q$$

Määritellään satunnaismuuttuja

$$X = \begin{cases} 1, & \text{jos satunnaisesti valittu tuote on viallinen} \\ 0, & \text{jos satunnaisesti valittu tuote ei ole viallinen} \end{cases}$$

Tällöin satunnaismuuttuja X noudattaa *Bernoulli-jakaumaa* parametrinaan

$$p = \Pr(A) = E(X)$$

Valmistettujen tuotteiden joukosta poimittiin yksinkertainen satunnaisotos, jonka koko oli

$$n = 150$$

ja otoksessa havaittiin 15 viallista tuotetta.

Konstruoidaan otoksesta saatujen tietojen perusteella $(1 - \alpha) \%$:n *luottamusväli* odotusarvoparametrille p . Luottamusväli on muotoa

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

jossa

$$\hat{p} = \text{odotusarvoparametrin } p \text{ harhaton estimaattori}$$

$$n = \text{havaintojen lukumäärä}$$

$$-z_{\alpha/2} \text{ ja } +z_{\alpha/2} = \text{luottamustasoon } (1 - \alpha) \text{ liittyvät luottamuskertoimet normaali-jakaumasta } N(0,1)$$

Parametrin p *estimaatiksi* saadaan

$$\hat{p} = \frac{f}{n} = \frac{15}{150} = 0.1$$

Valitaan *luottamustasoksi*

$$1 - \alpha = 0.95$$

Koska

$$\alpha = 0.05$$

luottamustasoa 0.95 vastaavat *luottamuskertoimet* ovat

$$-z_{\alpha/2} = -z_{0.025}$$

$$+z_{\alpha/2} = +z_{0.025}$$

Luottamuskertoimet $-z_{\alpha/2} = -z_{0.025}$ ja $+z_{\alpha/2} = +z_{0.025}$ toteuttavat yhtälöt

$$\Pr(z \leq -z_{\alpha/2}) = \Pr(z \leq -z_{0.025}) = \frac{\alpha}{2} = 0.025$$

$$\Pr(z \geq +z_{\alpha/2}) = \Pr(z \geq +z_{0.025}) = \frac{\alpha}{2} = 0.025$$

jossa satunnaismuuttuja z noudattaa *standardoitua normaalijakaumaa*:

$$z \sim N(0,1)$$

Siten

$$\Pr(-z_{\alpha/2} \leq z \leq +z_{\alpha/2}) = \Pr(-z_{0.025} \leq z \leq +z_{0.025}) = 1 - \alpha = 0.95$$

Standardoidun normaalijakauman $N(0,1)$ taulukoiden mukaan

$$-z_{0.025} = -1.96$$

$$+z_{0.025} = +1.96$$

Siten 95 %:n *luottamusväli* Bernoulli-jakauman parametrille p on muotoa

$$\begin{aligned} \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.1 \pm 1.96 \sqrt{\frac{0.1 \times (1-0.1)}{150}} \\ &= 0.1 \pm 1.96 \times 0.024 \\ &= 0.1 \pm 0.048 \\ &= (0.052, 0.148) \end{aligned}$$

Väli *ei peitä* parametrin p oletettua arvoa 0.05.

Valitaan *luottamustasoksi*

$$1 - \alpha = 0.99$$

Koska

$$\alpha = 0.01$$

luottamustasoa $1 - \alpha = 0.99$ vastaavat *luottamuskertoimet* ovat

$$-z_{\alpha/2} = -z_{0.005}$$

$$+z_{\alpha/2} = +z_{0.005}$$

Luottamuskertoimet $-z_{\alpha/2} = -z_{0.005}$ ja $+z_{\alpha/2} = +z_{0.005}$ toteuttavat yhtälöt

$$\Pr(z \leq -z_{\alpha/2}) = \Pr(z \leq -z_{0.005}) = \frac{\alpha}{2} = 0.005$$

$$\Pr(z \geq +z_{\alpha/2}) = \Pr(z \geq +z_{0.005}) = \frac{\alpha}{2} = 0.005$$

jossa satunnaismuuttuja z noudattaa *standardoitua normaalijakaumaa*:

$$z \sim N(0,1)$$

Siten

$$\Pr(-z_{\alpha/2} \leq z \leq +z_{\alpha/2}) = \Pr(-z_{0.005} \leq z \leq +z_{0.005}) = 1 - \alpha = 0.99$$

Standardoidun normaalijakauman $N(0,1)$ taulukoiden mukaan

$$-z_{0.005} = +2.58$$

$$+z_{0.005} = -2.58$$

Siten 99 %:n *luottamusväli* Bernoulli-jakauman parametrille p on muotoa

$$\begin{aligned} \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.1 \pm 2.58 \sqrt{\frac{0.1 \times (1-0.1)}{150}} \\ &= 0.1 \pm 2.58 \times 0.024 \\ &= 0.1 \pm 0.063 \\ &= (0.037, 0.163) \end{aligned}$$

Väli *peittää* parametrin p oletetun arvon 0.05.

Siten otoksesta saatu evidenssi viittaa siihen suuntaan, että valmistajan väitteeseen voidaan kohdistaa jonkin verran epäilyjä. Asiaa tarkastellaan lisää *tilastollisen testauksen* yhteydessä.

Huomautuksia:

- (i) Luottamusväli *levenee*, jos luottamustasoa *kasvatetaan*, jolloin siitä tulee *epäinformatiivisempi*.
- (ii) Luottamusväli *kapenee*, jos otoskokoa *kasvatetaan*.
- (iii) Jos luottamusvälin pituus halutaan *puolittaa*, pitää havaintojen lukumäärä *nelinkertaistaa*.

Tehtävä 9.5.

Tehdas valmistaa ruuveja. Ruuvien paino vaihtelee satunnaisesti noudattaen normaali-jakaumaa. Ruuvien joukosta poimittiin yksinkertainen satunnaisotos. Otoskeskiarvoksi saatiin tällöin 25 g. Tehdään (epärealistinen) oletus, että normaalijakauman varianssi 0.25 g^2 on tunnettu.

Määrittää 99 %:n *luottamusvälit* ruuvien painon odotusarvolle, jos otoskokona oli

- (a) 1
- (b) 100
- (c) 10000

Vertaa saatujen luottamusvälien pituuksia toisiinsa. Miten luottamusvälin pituus käyttäytyy otoskoon funktiona?

Tehtävä 9.5. – Mitä opimme?

Tehtävässä tarkastellaan *normaalijakauman odotusarvon luottamusvälin määräämistä* (epärealistisessa) tilanteessa, jossa *jakauman varianssi oletetaan tunnetuksi*.

Normaalijakauman odotusarvon luottamusvälin määräämistä esimerkitapauksessa, jossa jakauman varianssia ei oleteta tunnetuksi käsitellään tehtävässä 9.6.

Tehtävä 9.5. – Ratkaisu:

Tehdas valmistaa ruuveja. Ruuvien paino vaihtelee satunnaisesti noudattaen normaali-jakaumaa. Ruuvien joukosta poimittiin yksinkertainen satunnaisotos, jonka koko oli n .

Määritellään satunnaismuuttujat

$$X_i = \text{Ruuvin } i \text{ paino otoksessa, } i = 1, 2, \dots, n$$

Oletuksien mukaan

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

$$X_i \sim N(\mu, \sigma^2), i = 1, 2, \dots, n$$

jossa varianssi

$$\sigma^2 = 0.25 \text{ g}$$

on tunnettu.

Otokseen poimittujen ruuvien painojen *aritmeettinen keskiarvo* oli

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = 25 \text{ g}$$

Konstruoidaan otoksesta saatujen tietojen perusteella $(1 - \alpha) \%$:n *luottamusväli* odotusarvoparametrille μ . Koska varianssi σ^2 oletettiin *tunnetuksi*, luottamusväli on muotoa

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

jossa

$\bar{X}$ = havaintojen *aritmeettinen keskiarvo* otoksessa

σ^2 = jakauman *varianssi*

n = havaintojen *lukumäärä*

$-z_{\alpha/2}$ ja $+z_{\alpha/2}$ = luottamustasoon $(1 - \alpha)$ liittyvät *luottamuskertoimet* normaalijakaumasta $N(0,1)$

Valitaan *luottamustasoksi*

$$1 - \alpha = 0.99$$

Koska

$$\alpha = 0.01$$

luottamustasoa $1 - \alpha = 0.99$ vastaavat *luottamuskertoimet* ovat

$$-z_{\alpha/2} = -z_{0.005}$$

$$+z_{\alpha/2} = +z_{0.005}$$

Luottamuskertoimet $-z_{\alpha/2} = -z_{0.005}$ ja $+z_{\alpha/2} = +z_{0.005}$ toteuttavat yhtälöt

$$\Pr(z \leq -z_{\alpha/2}) = \Pr(z \leq -z_{0.005}) = \frac{\alpha}{2} = 0.005$$

$$\Pr(z \geq +z_{\alpha/2}) = \Pr(z \geq +z_{0.005}) = \frac{\alpha}{2} = 0.005$$

jossa satunnaismuuttuja z noudattaa *standardoitua normaalijakaumaa*:

$$z \sim N(0,1)$$

Siten

$$\Pr(-z_{\alpha/2} \leq z \leq +z_{\alpha/2}) = \Pr(-z_{0.005} \leq z \leq +z_{0.005}) = 1 - \alpha = 0.99$$

Standardoidun normaalijakauman $N(0,1)$ taulukoiden mukaan

$$-z_{0.005} = -2.58$$

$$+z_{0.005} = +2.58$$

Siten 99 %:n *luottamusväli* normaalijakauman odotusarvoparametrille μ on muotoa

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 25 \pm 2.58 \times \frac{0.5}{\sqrt{n}}$$

- (a)
- $n = 1$
- :

Luottamusväliksi saadaan

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 25 \pm 2.58 \times \frac{0.5}{\sqrt{1}} = 25 \pm 1.29 = (23.71, 26.29)$$

- (b)
- $n = 100$
- :

Luottamusväliksi saadaan

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 25 \pm 2.58 \times \frac{0.5}{\sqrt{100}} = 25 \pm 0.129 = (24.871, 25.129)$$

- (c)
- $n = 10000$
- :

Luottamusväliksi saadaan

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 25 \pm 2.58 \times \frac{0.5}{\sqrt{10000}} = 25 \pm 0.0129 = (24.9871, 25.0129)$$

Jos otantaa toistetaan, niin *luottamustason frekvenssitulkinnan* mukaan otoksista konstruoidut luottamusvälit peittävät (keskimäärin)

99 %:ssa

otoksia parametrin μ tuntemattoman arvon ja (keskimäärin)

1 %:ssa

otoksia ei sitä tee.

Huomautuksia:

- (i) Luottamusväli *levenee*, jos luottamustasoa *kasvatetaan*, jolloin siitä tulee *epäinformatiivisempi*.
- (ii) Luottamusväli *kapenee*, jos otoskokoa *kasvatetaan*.
- (iii) Jos luottamusvälin pituus halutaan *puolittaa*, pitää havaintojen lukumäärä *nelinkertaistaa*.
- (iv) **Luottamuskertoimet pitää valita normaalijakauman sijasta *t-jakaumasta*, jos varianssi σ^2 ei ole tunnettu ja se joudutaan *estimoimaan* otoksesta.**

Näin saatava *estimoituun varianssiin σ^2 perustuva luottamusväli on *leveämpi* kuin tässä konstruoitu *tunnettuun varianssiin σ^2 perustuva luottamusväli*; ks. tehtävää 9.6.*

Jos havaintojen lukumäärän annetaan kasvaa rajatta, *estimoituun varianssiin σ^2 perustuva luottamusväli lähestyy tunnettuun varianssiin σ^2 perustuvaa luottamusväliä*.

Tehtävä 9.6.

Tehdas valmistaa nauloja. Naulojen pituus vaihtelee satunnaisesti noudattaen normaali-jakaumaa. Naulojen joukosta poimittiin yksinkertainen satunnaisotos, jonka koko oli 30. Otokeskiarvoksi saatiin 9.99 cm ja otosvarianssiksi 0.01 cm.

- (a) Määrää 95 %:n luottamusväli naulojen pituuden odotusarvolle.
- (b) Määrää 90 %:n luottamusväli naulojen pituuden varianssille.

Tehtävä 9.6. – Mitä opimme?

Tehtävässä tarkastellaan *normaalijakauman odotusarvon ja varianssin luottamusvälin määräämistä tilanteessa, jossa jakauman varianssi ei ole tunnettu.*

Normaalijakauman odotusarvon luottamusvälin määräämistä esimerkkitapauksessa, jossa jakauman varianssi on tunnettu käsitellään tehtävässä 9.5.

Tehtävä 9.6. – Ratkaisu:

- (a) Tehdas valmistaa nauloja. Naulojen pituus vaihtelee satunnaisesti noudattaen normaalijakaumaa. Naulojen joukosta poimittiin yksinkertainen satunnaisotos, jonka koko oli 30.

Määritellään satunnaismuuttujat

$$X_i = \text{Naulan } i \text{ pituus otoksessa, } i = 1, 2, \dots, n = 30$$

Oletuksien mukaan

$$X_1, X_2, \dots, X_{30} \perp$$

$$X_i \sim N(\mu, \sigma^2), i = 1, 2, \dots, 30$$

Otokseen poimittujen ruuvien painojen *aritmeettinen keskiarvo* oli

$$\bar{X} = \frac{1}{30} \sum_{i=1}^{30} X_i = 9.99 \text{ cm}$$

ja *otosvarianssi* oli

$$s^2 = \frac{1}{30-1} \sum_{i=1}^{30} (X_i - \bar{X})^2 = 0.01 \text{ cm}$$

Konstruoidaan otoksesta saatujen tietojen perusteella $(1 - \alpha) \%$:n *luottamusväli* odotusarvoparametrille μ . Koska varianssi σ^2 oletettiin *tuntemattomaksi*, luottamusväli on muotoa

$$\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

jossa

$\bar{X}$ = havaintojen *aritmeettinen keskiarvo* otoksessa

s^2 = *otosvarienssi*

n = havaintojen *lukumäärä*

$-t_{\alpha/2}$ ja $+t_{\alpha/2}$ = luottamustasoon $(1 - \alpha)$ liittyvät *luottamuskertoimet* *t-jakaumasta* $t(n - 1)$

Valitaan *luottamustasoksi*

$$1 - \alpha = 0.95$$

Koska

$$\alpha = 0.05$$

luottamustasoa $1 - \alpha = 0.95$ vastaavat *luottamuskertoimet* ovat

$$-t_{\alpha/2} = -t_{0.025}$$

$$+t_{\alpha/2} = +t_{0.025}$$

Luottamuskertoimet $-t_{\alpha/2} = -t_{0.025}$ ja $+t_{\alpha/2} = +t_{0.025}$ toteuttavat yhtälöt

$$\Pr(t \leq -t_{\alpha/2}) = \Pr(t \leq -t_{0.025}) = \frac{\alpha}{2} = 0.025$$

$$\Pr(t \geq +t_{\alpha/2}) = \Pr(t \geq +t_{0.025}) = \frac{\alpha}{2} = 0.025$$

jossa satunnaismuuttuja t noudattaa *t-jakaumaa vapausastein* $(n - 1) = 29$:

$$t \sim t(n - 1) = t(29)$$

Siten

$$\Pr(-t_{\alpha/2} \leq t \leq +t_{\alpha/2}) = \Pr(-t_{0.025} \leq t \leq +t_{0.025}) = 1 - \alpha = 0.95$$

t-jakauman $t(29)$ taulukoiden mukaan

$$-t_{0.025} = -2.045$$

$$+t_{0.025} = +2.045$$

Siten 95 %:n *luottamusväli* normaalijakauman odotusarvoparametrille μ on muotoa

$$\begin{aligned} \bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} &= 9.99 \pm 2.045 \times \frac{0.1}{\sqrt{30}} \\ &= 9.99 \pm 0.037 = (9.953, 10.027) \end{aligned}$$

Jos otantaa toistetaan, niin *luottamustason frekvenssitulkinnan* mukaan otoksista konstruoidut luottamusvälit peittävät (keskimäärin)

95 %:ssa

otoksia parametrin μ tuntemattoman arvon ja (keskimäärin)

5 %:ssa

otoksia ei sitä tee.

Huomautuksia:

- (i) Luottamusväli *levenee*, jos luottamustasoa *kasvatetaan*, jolloin siitä tulee *epäinformatiivisempi*.
- (ii) Luottamusväli *kapenee*, jos otoskoko *kasvatetaan*.
- (iii) Jos luottamusvälin pituus halutaan *puolittaa*, pitää havaintojen lukumäärä *nelinkertaistaa*.
- (iv) **Luottamuskertoimet voidaan valita t -jakauman sijasta *normaalijakaumasta*, jos varianssi σ^2 on tunnettu.**

Näin saatava *tunnettuun varianssiin σ^2 perustuva luottamusväli on kapeampi* kuin tässä konstruoitu *estimoituun varianssiin σ^2 perustuva luottamusväli*; ks. tehtävää 9.5.

Jos havaintojen lukumäärän annetaan kasvaa rajatta, *estimoituun varianssiin σ^2 perustuvan luottamusväli lähestyy tunnettuun varianssiin σ^2 perustuvaa luottamusväliä*.

- (b) Tehdas valmistaa nauloja. Naulojen pituus vaihtelee satunnaisesti noudattaen normaalijakaumaa. Naulojen joukosta poimittiin yksinkertainen satunnaisotos, jonka koko oli 30.

Määritellään satunnaismuuttujat

$$X_i = \text{Naulan } i \text{ pituus otoksessa, } i = 1, 2, \dots, n = 30$$

Oletuksien mukaan

$$X_1, X_2, \dots, X_{30} \perp$$

$$X_i \sim N(\mu, \sigma^2), i = 1, 2, \dots, 30$$

Otokseen poimittujen ruuvien painojen *aritmeettinen keskiarvo* oli

$$\bar{X} = \frac{1}{30} \sum_{i=1}^{30} X_i = 9.99 \text{ cm}$$

ja *otosvarianssi* oli

$$s^2 = \frac{1}{30-1} \sum_{i=1}^{30} (X_i - \bar{X})^2 = 0.01 \text{ cm}$$

Konstruoidaan otoksesta saatujen tietojen perusteella $(1 - \alpha)$ %:n *luottamusväli* varianssiparametrille σ^2 .

Luottamusväli on muotoa

$$\left(\frac{(n-1)s^2}{\chi_{\alpha/2}^2}, \frac{(n-1)s^2}{\chi_{1-\alpha/2}^2} \right)$$

jossa

$$s^2 = \text{otosvarianssi}$$

$$n = \text{havaintojen lukumäärä}$$

$$\chi_{1-\alpha/2}^2 \text{ ja } \chi_{\alpha/2}^2 = \text{luottamustasoon } (1 - \alpha) \text{ liittyvät luottamuskertoimet } \chi^2\text{-jakaumasta vapausastein } (n - 1)$$

Valitaan *luottamustasoksi*

$$1 - \alpha = 0.90$$

Koska

$$\alpha = 0.10$$

luottamustasoa $1 - \alpha = 0.90$ vastaavat *luottamuskertoimet* ovat

$$\chi_{1-\alpha/2}^2 = \chi_{0.95}^2$$

$$\chi_{\alpha/2}^2 = \chi_{0.05}^2$$

Luottamuskertoimet $\chi_{1-\alpha/2}^2 = \chi_{0.95}^2$ ja $\chi_{\alpha/2}^2 = \chi_{0.05}^2$ toteuttavat yhtälöt

$$\Pr(\chi^2 \leq \chi_{1-\alpha/2}^2) = \Pr(\chi^2 \leq \chi_{0.95}^2) = \frac{\alpha}{2} = 0.05$$

$$\Pr(\chi^2 \geq \chi_{\alpha/2}^2) = \Pr(\chi^2 \geq \chi_{0.05}^2) = \frac{\alpha}{2} = 0.05$$

jossa satunnaismuuttuja χ^2 noudattaa χ^2 -jakaumaa vapausastein $(n - 1) = 29$:

$$\chi^2 \sim \chi^2(n-1) = \chi^2(29)$$

Siten

$$\Pr(\chi_{1-\alpha/2}^2 \leq \chi^2 \leq \chi_{\alpha/2}^2) = \Pr(\chi_{0.95}^2 \leq \chi^2 \leq \chi_{0.05}^2) = 1 - \alpha = 0.90$$

χ^2 -jakauman $\chi^2(29)$ taulukoiden mukaan

$$\chi_{1-\alpha/2}^2 = 17.708$$

$$\chi_{\alpha/2}^2 = 42.557$$

Siten 90 %:n *luottamusväli* normaalijakauman varianssiparametrille σ^2 on muotoa

$$\left(\frac{(n-1)s^2}{\chi_{\alpha/2}^2}, \frac{(n-1)s^2}{\chi_{1-\alpha/2}^2} \right) = \left(\frac{(30-1) \times 0.01}{42.557}, \frac{(30-1) \times 0.01}{17.708} \right) = (0.0068, 0.0164)$$

Jos otantaa toistetaan, niin *luottamustason frekvenssitulkinnan* mukaan otoksista konstruoidut luottamusvälit peittävät (keskimäärin)

90 %:ssa

otoksia parametrin σ^2 tuntemattoman arvon ja (keskimäärin)

10 %:ssa

otoksia ei sitä tee.

Huomautuksia:

- (i) Luottamusväli *levenee*, jos luottamustasoa *kasvatetaan*, jolloin siitä tulee *epäinformatiivisempi*.
- (ii) Luottamusväli pituus *pysyy suunnilleen samanmittaisena*, jos otoskokoa *kasvatetaan*.

Tehtävä 9.7.

Kuntaan suunnitellaan ydinvoimalaa. Kunnan asukkaiden mielipiteet halutaan selvittää yksinkertaiseen satunnaisotantaan perustuvalla kyselytutkimuksella.

Kuinka suuri otos kuntalaisten joukosta on poimittava, jotta saataisiin 99 %:n varmuus siitä, että otoksesta laskettu voimalan rakentamisen kannattajien suhteellinen osuus ei poikkea enempää kuin 0.5 %-yksikköä voimalan rakentamisen kannattajien todellisesta suhteellisesta osuudesta?

Tehtävä 9.7. – Mitä opimme?

Tehtävässä tarkastellaan *tarvittavan otoskoon määrittämistä*, kun Bernoulli-jakauman odotusarvoparametrille halutaan saada tietyn mittainen luottamusväli.

Tehtävä 9.7. – Ratkaisu:

Tarkastellaan tapahtumaa

$$A = \{\text{Satunnaisesti valittu kunnan asukas kannattaa voimalan rakentamista}\}$$

ja olkoon tapahtuman A todennäköisyys

$$\Pr(A) = p$$

$$\Pr(A^c) = 1 - p = q$$

Määritellään satunnaismuuttuja

$$X = \begin{cases} 1, & \text{jos satunnaisesti valittu kunnan asukas kannattaa voimalan rakentamista} \\ 0, & \text{jos satunnaisesti valittu kunnan asukas ei kannata voimalan rakentamista} \end{cases}$$

Oletetaan, että satunnaismuuttuja X noudattaa *Bernoulli-jakaumaa* parametrinaan

$$p = \Pr(A) = E(X)$$

Jos kunnassa on N asukasta,

$$pN = F$$

on voimalan rakentajien lukumäärä *kaikkien kunnan asukkaiden joukossa*. Siten odotusarvoparametri

$$p = \frac{F}{N}$$

voidaan tulkita voimalan rakentamisen kannattajien *suhteelliseksi osuudeksi kaikkien kunnan asukkaiden joukossa*.

Oletetaan, että kunnan asukkaiden joukosta poimitaan *yksinkertainen satunnaisotos*, jonka koko n . Jos voimalan rakentamisen kannattajien *lukumäärä otoksessa* on f , niin kannattajien *suhteellinen osuus otoksessa*

$$\hat{p} = \frac{f}{n}$$

on odotusarvoparametrin p harhaton estimaattori.

Bernoulli-jakauman odotusarvoparametrin p luottamusväli on muotoa

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

jossa

$\hat{p}$ = odotusarvoparametrin p harhaton estimaattori

n = havaintojen lukumäärä

$-z_{\alpha/2}$ ja $+z_{\alpha/2}$ = luottamustasoon $(1 - \alpha)$ liittyvät luottamuskertoimet normaalijakaumasta $N(0,1)$

Siten Bernoulli-jakauman parametrin p luottamusväli on muotoa

$$\hat{p} \pm a$$

jossa $\hat{p}$ on otoksesta laskettu suhteellinen osuus. Jos haluamme parametrille p luottamusvälin, jonka pituus on $2a$, voimme muodostaa otoskoon n ratkaisemiseksi yhtälön

$$z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = a$$

Ratkaisuksi saadaan

$$n = \left(\frac{z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})}}{a} \right)^2$$

Jos otoskooksi n valitaan pienin kaavan antamaa lukuarvoa suuremmista kokonaisluvuista, saadaan haluttu varmuus siitä, että voimalan rakentamisen kannattajien *suhteellinen osuus kaikkien kuntalaisten joukossa* on otoksesta konstruoidun luottamusvälin sisällä.

Koska $\hat{p}$:n arvoa ei tunneta ($\hat{p}$:n arvo saadaan tietysti selville vasta otoksen poimimisen jälkeen), on järkevää korvata $\hat{p}$ luvulla, joka *maksimoi* tarvittavan otoskoon n .

Funktio

$$f(\hat{p}) = \hat{p}(1-\hat{p})$$

on alaspäin aukeava paraabeli, joka saavuttaa maksiminsa pisteessä

$$\hat{p} = \frac{1}{2}$$

Siten (maksimaalinen) tarvittava otoskoko n saadaan kaavasta

$$n = \left(\frac{z_{\alpha/2}}{2a} \right)^2$$

Tehtävässä

$$a = 0.005$$

$$1 - \alpha = 0.99$$

$$z_{\alpha/2} = 2.58$$

Siten tarvittava otoskoko on

$$n = \left(\frac{z_{\alpha/2}}{2a} \right)^2 = \left(\frac{2.58}{2 \times 0.005} \right)^2 = 66564$$

Kommentteja:

- (i) Jos haluamme näin kapean luottamisvälin ($\pm 0.5\%$) näin korkealla luottamustasolla (99 %) voimalan rakentajien suhteelliselle osuudelle kaikkien kunnan asukkaiden joukossa, tarvitsemme otoksen, joka on paljon suurempi (> 60000 asukasta) kuin kyselytutkimuksissa tavallisesti käytetyt otokset (n. 500-2000).

Olemme siis saaneet paradoksaalisen tuloksen, että tarvittava otoskoko ylittäisi useimmissa Suomen kunnissa kunnan asukasluvun! Tämä on se hinta, jonka joudumme maksamaan otoksen poimintaan liittyvästä sattumasta ja asetusta tarkkuus-vaatimuksesta.

- (ii) Tavallisesti kyselytutkimuksissa käytetään otoskokoja 500-2000 ja 95 %:n luottamustasoa. Suhteellisen osuuden luottamusvälin pituus on

$$2a = 2 \times z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Jos tässäkin korvaamme (ennen otoksen poimintaa tuntemattoman) suhteellisen osuuden $\hat{p}$ lukuarvolla, joka maksimoi funktion

$$f(\hat{p}) = \hat{p}(1-\hat{p})$$

saamme luottamusvälin maksimipituudeksi

$$2a = 2 \times z_{\alpha/2} \sqrt{\frac{1}{4n}}$$

Jos luottamustasona on

$$1 - \alpha = 0.95$$

niin $\alpha/2 = 0.025$ ja normaalijakauman taulukoiden mukaan

$$z_{\alpha/2} = 1.96$$

Jos siis otoskoko n vaihtelee välillä $[500, 2000]$, niin 95 %:n luottamusvälin pituus vaihtelee välillä

$$(2 \times 0.022, 2 \times 0.044)$$

Tästä nähdään se, että tyypillisissä vaalikyselyissä puolueiden todellisiin kannatus-osuuksiin liittyvä epävarmuus on huomattavan suurta.

Tehtävä 9.8.

Tölkitetyn tuoremehun C-vitamiinipitoisuus (mg/dl) vaihtelee jonkin verran valmistuserästä toiseen noudattaen normaalijakaumaa. Laboratorio haluaa selvittää erään tuoremehumerkin keskimääräisen C-vitamiinipitoisuuden mittaamalla pitoisuudet myynnissä olevien tuoremehutölkkien joukosta poimitusta yksinkertaisesta satunnaisotoksesta.

Laboratorio haluaa niin tarkan arvion C-vitamiinipitoisuudesta, että voidaan 95 %:n varmuudella tehdä johtopäätös, että otoksesta laskettu keskimääräinen C-vitamiinipitoisuus ei poikkea todellisesta keskimääräisestä C-vitamiinipitoisuudesta enempää kuin 0.5 mg.

Määrää tarvittava otoskoko, kun aikaisempien tutkimusten perusteella tiedetään, että C-vitamiinipitoisuuden otoskeskihajonta on tavallisesti n. 2 mg.

Tehtävä 9.8. – Mitä opimme?

Tehtävässä tarkastellaan *tarvittavan otoskoon määrittämistä*, kun normaalijakauman odotusarvoparametrille halutaan saada tietyn mittainen luottamusväli.

Tehtävä 9.8. – Ratkaisu:

Oletetaan, että tölkitetyn tuoremehun C-vitamiinipitoisuus vaihtelee satunnaisesti valmistuserästä toiseen noudattaen normaalijakaumaa.

Määritellään satunnaismuuttujat

$$X = \text{Tuoremehun C-vitamiinipitoisuus (mg/dl)}$$

Oletamme, että

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

Oletetaan, että tölkkien joukosta poimitaan yksinkertainen satunnaisotos, jona koko on n .

Konstruoidaan otoksesta saatujen tietojen perusteella $(1 - \alpha) \%$:n *luottamusväli* odotusarvoparametrille μ . Koska varianssi σ^2 on oletettu *tunnetuksi*, luottamusväli on muotoa

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

jossa

$$\bar{X} = \text{havaintojen aritmeettinen keskiarvo otoksessa}$$

$$\sigma^2 = \text{varianssi}$$

$$n = \text{havaintojen lukumäärä}$$

$$-z_{\alpha/2} \text{ ja } +z_{\alpha/2} = \text{luottamustasoon } (1 - \alpha) \text{ liittyvät luottamuskertoimet standardoidusta normaalijakaumasta } N(0,1)$$

Siten odotusarvon μ luottamusväli on muotoa

$$M \pm a$$

jossa M on havaintojen aritmeettinen keskiarvo otoksessa. Jos haluamme parametrille μ luottamusvälin, jonka pituus on $2a$, voimme muodostaa otoskoon n ratkaisemiseksi yhtälön

$$a = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Ratkaisuksi saadaan

$$n = \left(\frac{z_{\alpha/2} \sigma}{a} \right)^2$$

Jos *otoskooksi* n valitaan pienin kaavan antamaa lukuarvoa suuremmista kokonaisluvuista, saadaan haluttu varmuus siitä, että todellinen C-vitamiinipitoisuus on otoksesta konstruoidun luottamusvälin sisällä.

Tehtävässä

$$a = 0.5$$

$$1 - \alpha = 0.95$$

$$z_{\alpha/2} = 1.96$$

$$\sigma = 2$$

Siten tarvittavan otoskoon kaava antaa tulokseksi luvun

$$n = \left(\frac{z_{\alpha/2} \sigma}{a} \right)^2 = \left(\frac{1.96 \times 2}{0.5} \right)^2 \approx 61.5$$

joten otoskooksi pitää valita vähintään 62.