6. LUOTTAMUSVÄLIT NORMAALIJAKAUMALLE

[vrt. Arjas-Siren, luku 3]

Tark. malli $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$.

Tavoite: Muodostetaan luottamusväli odotusarvolle μ annetulla luottamustasolla $1-\alpha$ (esim. 0.95)

Pal. mieleen, että μ:n (piste-)estimaattina käytetään

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

Vastaavasti σ^2:n estimaattina (useimmiten) käytetään

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 .$$

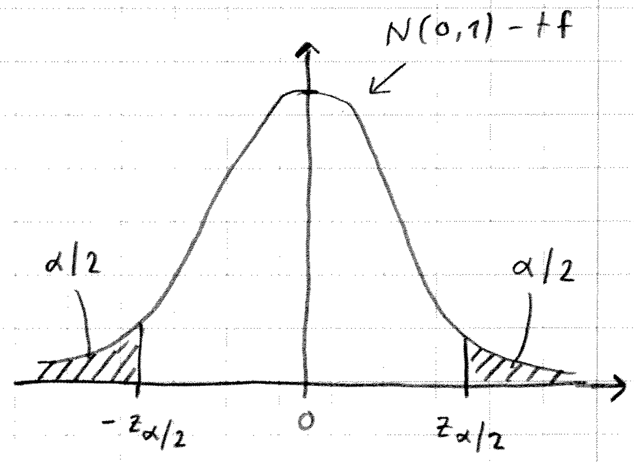
(A) Olet. että $\sigma^2 > 0$ on tunnettu luku (ei siis parametri)

Sivulla 16 todettiin, että sm:lle $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ pätee

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \qquad \text{[vrt. Tuominen, TN I, Esim. 3.7.9]}$$

eli

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



Valitaan $N(0,1)$-jakaumasta kohta $z_{\alpha/2}$, josta oikealle "häntätn" $= \alpha/2$

Symmetria $\Rightarrow$ $-z_{\alpha/2}$:sta vasemmalle "häntätn" on samoin $\alpha/2$

Välille $(-z_{\alpha/2}, z_{\alpha/2})$ jäävä tn-massa $= 1-\alpha$.

Siten

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

$\Longrightarrow$

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \qquad (**)$$

Siis (satunnaisten) pisteiden $\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ ja $\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ rajoittama väli sisältää μ:n tn:llä $1-\alpha$.

Tämä merkitsee:

> Kun $\underset{\sim}{x} = (x_1, \ldots, x_n)$ on havaittu aineisto ja $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, niin väli
>
> $$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \qquad (\#)$$
>
> on <u>μ:n luottamusväli luottamustasolla $1-\alpha$</u>.

<u>Esim.</u> Kun $\alpha = 0.05$, niin

$$z_{\alpha/2} = z_{0.025} \approx 1.96 \quad (\approx 2)$$

joten <u>95 %:n luottamusväli</u> μ:lle on

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

[<u>Huom.</u> Lukujen $z_{\alpha/2}$ arvoja on normaalijakauman taulukoissa!]

(B) Oletetaan, että myös σ^2 on <u>(tuntematon) parametri</u>

Ongelma: väli (#) yllä ei käytettävissä, koska σ ei tunnettu $\Longrightarrow$ joudumme estimoimaan sen aineistosta.

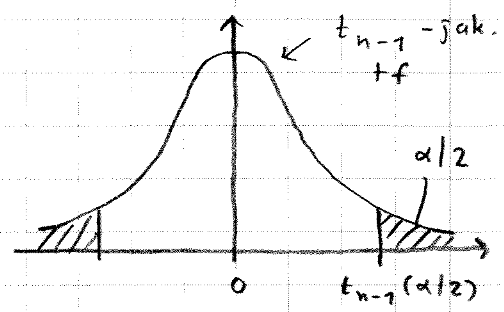
Estimaattina käytämme

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \left(\frac{\text{otos-}}{\underline{\text{keskihajonta}}}\right)$$

Olkoon S vastaava satunnaismuuttuja.
Tällöin pätee (vrt. A-tapauksen jakaumatulos (*) !)

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1} \qquad \left(\begin{array}{l}\text{Studentin } \underline{t\text{-jakauma}},\\ n-1 \ \underline{\text{vapausostetta}}\end{array}\right)$$

Menetellään kuten A-kohdassa: Valitaan piste $t_{n-1}(\alpha/2)$, josta oikealle t_{n-1}-jakauman häntätn $=\alpha/2$. Päädytään (**):n sijasta tulokseen



$$P\left(\overline{X}-t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}} < \mu < \overline{X}+t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}}\right)=1-\alpha$$

Saatu:

> Kun $\underset{\sim}{x}=(x_1,\dots,x_n)$ on havaittu aineisto ja $\overline{x}$ = otoskeskiarvo sekä s = otoskeskihajonta (ks. edell. sivu), niin väli
>
> $$\left(\overline{x}-t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}},\ \overline{x}+t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}\right) \qquad (\#\#)$$
>
> on $\underline{\mu\text{:n luottamusväli luottamustasolla } 1-\alpha}$.

Huom. * Verrattuna lv:iin (#) tässä lasketaan aineistosta $\overline{x}$:n lisäksi myös keskihajonta s. Lisäksi kerroin $t_{n-1}(\alpha/2)$ riippuu otoskoosta n.
* t_{n-1}-jakaumalla on "paksummat hännät" kuin $N(0,1)$-jakaumalla $\Rightarrow$ kerroin $t_{n-1}(\alpha/2)$ on suurempi kuin $z_{\alpha/2}$. Käytännössä ero on merkityksellinen vain pienillä n:n arvoilla ($\sim$ alle 50), sillä t_{n-1} lähestyy $N(0,1)$-jakaumaa kun $n\to\infty$.
* Lukuja $t_{n-1}(\alpha/2)$ löytyy taulukoista ja tietokoneohjelmista, joten t-jakaumien tiheysfunktioita ei tarvitse tietää.

Esim. Tutkitaan erään bussityypin keskikulutusta (litraa / 100 km) kaupunkiliikenteessä. Oletetaan, että kulutus vaihtelee likimain normaalisti bussiyksilön ja ajoreitin mukaan. Tehtiin 20 kulutusmittausta satunnaisesti valituilla busseilla ja reiteillä: niiden keskiarvo oli 36 ja keskihajonta 3.5 (litraa / 100 km). Muodosta keskikulutukselle 95 %:n ja 99 % lv:t.

Ratkaisu: Olkoon μ keskikulutus, jolloin kyseessä on $n = 20$ riippumatonta havaintoa $N(\mu, \sigma^2)$ - jakaumasta (jossa μ:n lisäksi myös σ^2 tuntematon).

Aineistosta laskettu: $\bar{x} = 36$, $s = 3.5$

95 %:n luottamustasolla ($\alpha = 0.05$) on

$$t_{19}(0.025) \overset{\text{taulukosta}}{\approx} 2.093, \quad t_{19}(0.025)\frac{s}{\sqrt{n}} \approx 1.64 \approx 1.6$$

$$\Rightarrow \text{lv on } (36 - 1.6,\ 36 + 1.6) = (34.4,\ 37.6)$$

99 %:n luottamustasolla ($\alpha = 0.01$) on

$$t_{19}(0.005) \approx 2.861, \quad t_{19}(0.005)\frac{s}{\sqrt{n}} \approx 2.24 \approx 2.2$$

$$\Rightarrow \text{lv on } (36 - 2.2,\ 36 + 2.2) = (33.8,\ 38.2)$$

Huom. Luottamusvälin tulkinnosta:

* Yhtälössä $P\left(\bar{X} - t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}}\right) = 1 - \alpha$ tn:n käsite liittyy aineistoon (sm:ien $\bar{X}$ ja S kautta): pisteiden $\bar{X} \pm t_{n-1}(\alpha/2)\, S/\sqrt{n}$ rajoittama satunnainen väli peittää μ:n (= kiinteä mutta tuntematon piste) tn:llä $1 - \alpha$.

* Toistetun aineistonkeruun kannalta: jos aineistonkeruu toistetaan uudelleen ja uudelleen ja jokaisesta aineistosta lasketaan luottamusväli (##), saaduista väleistä keskimäärin osuus $1 - \alpha$ sisältää μ:n ja osuus α "menee ohi".

Lisätietoa: Mistä t-jakauma tulee? (aineopintojen tn-laskennan kurssi!)

Olet. $Z \sim N(0,1)$ ja $Y \sim \chi^2_m$ (khii-toiseen-jakauma, m vapausastetta) sekä $Z \perp\!\!\!\perp Y$.

Määritelmä: Sm:n $T = \dfrac{Z}{\sqrt{Y/m}}$ noudattama jakauma on t-jakauma, jolla m vapausastetta; merh. $T \sim t_m$.

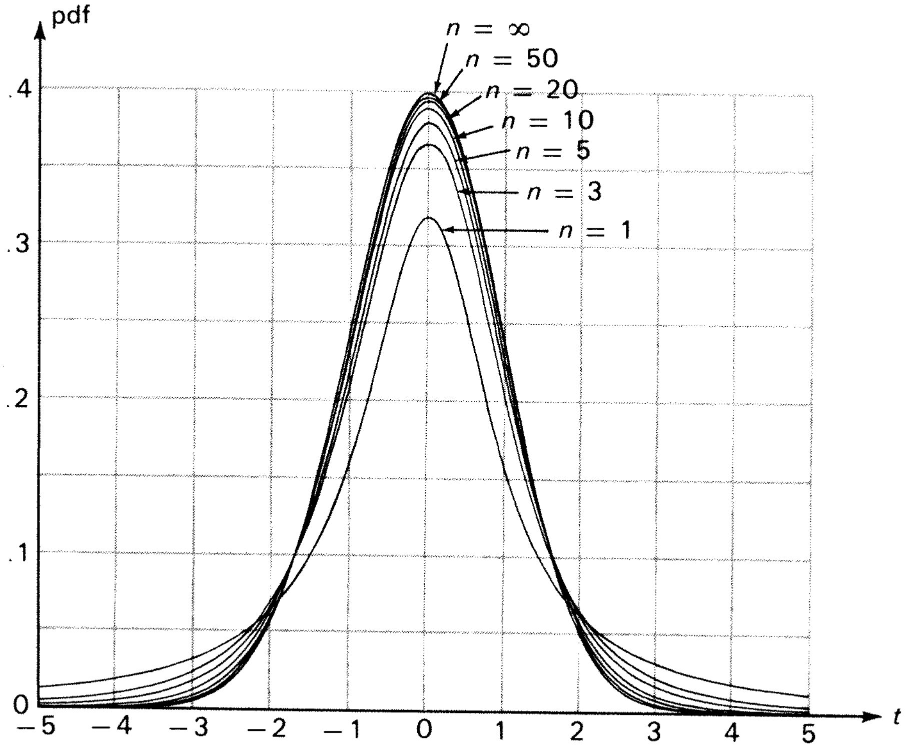
Tark. mallia $X_1, \dots, X_n \sim N(\mu, \sigma^2) \perp\!\!\!\perp$. Olkoot

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}, \quad Y = \frac{(n-1)S^2}{\sigma^2}$$

Sivulla 16 on todettu, että $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ eli $Z \sim N(0,1)$ ja $Y \sim \chi^2_{n-1}$ sekä $Z \perp\!\!\!\perp Y$. Siispä yo. määritelmän nojalla

$$T = \frac{Z}{\sqrt{Y/(n-1)}} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

(kuten sivulla 22 väitettiin!)



KUVA

t_n-jakauman tiheysfunktioita eri n:n arvoilla

$n = \infty$ tark. $N(0,1)$-jakaumaa