Johdatus tilastolliseen päättelyyn, kevät 2015
Harjoitus 2 (24.–27. 3.)

1. Jatkoa harjoituksen 1 tehtäviin 1, 2 ja 5. Yhden havainnon sijasta tehdäänkin ko. jakaumasta kaksi riippumatonta havaintoa, satunnaismuuttujina siis $Y_1 \perp\!\!\!\perp Y_2$, joilla kummallakin on sama pistetodennäköisyysfunktio $f(y; \theta)$.

Oletetaan, että havainnot ovat $y_1 = 4$ ja $y_2 = 3$. Laske vastaavan uskottavuusfunktion $L(\theta; y_1, y_2)$ arvot ja määritä suurimman uskottavuuden estimaatti $\hat{\theta}$.

2. Jatkoa harjoituksen 1 tehtävään 6.

a) Määritä parametrin θ suurimman uskottavuuden estimaatti $\hat{\theta}$. Ohje: Huomaat törmääväsi ongelmaan, mikäli käytit $\mathrm{Tas}(0, \theta)$-jakauman tiheysfunktiosta versiota, joka on $= 1/\theta$ vain avoimella välillä $]0, \theta[$. Kannattaakin valita tiheysfunktiolle versio, joka on $= 1/\theta$ vastaavalla suljetulla välillä. Se määrittelee saman jakauman mutta toimii suurimman uskottavuuden estimoinnissa paremmin.

b) Mitkä ovat $\hat{\theta}$:n arvot harjoituksen 1 tehtävän 4 kohdissa a–c?

3. Tehdas valmistaa tuotetta, jonka kestoaika on jakaumaltaan $\mathrm{Exp}(\lambda)$, jossa $\lambda > 0$ on meille tuntematon (vrt. JTN-kurssin harjoituksen 5 tehtävä 9). Poimimme satunnaisotoksen (n kpl) kyseisiä tuotteita ja mittaamme otoksen tuoteyksilöiden kestoajat $y_1, \ldots, y_n$. Tilastollinen mallimme on siis lyhyesti ilmaistuna $Y_1, \ldots, Y_n \sim \mathrm{Exp}(\lambda) \perp\!\!\!\perp$, jossa $\mathrm{Exp}(\lambda)$-jakauman tf on $g(x; \lambda) = \lambda e^{-\lambda x}$, kun $x > 0$.

Muodosta tilastollisen mallin ja uskottavuusfunktion lausekkeet ja osoita uskottavuusfunktion logaritmia tutkimalla, että $\hat{\lambda} = 1/\bar{y}$, jossa $\bar{y} = (y_1 + \cdots + y_n)/n$ on otoskeskiarvo.

Mikä ongelma mahtaa liittyä tällaisen satunnaiskokeen suorittamiseen käytännössä?

4. Mallina on satunnaisotos normaalijakaumasta $N(\mu, \sigma^2)$, ts. $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2) \perp\!\!\!\perp$. Luennoilla on näytetty, että odotusarvoparametrin μ suurimman uskottavuuden estimaatti on $\hat{\mu} = \bar{y}$, kun havainnot ovat $y_1, \ldots, y_n$ (ks. muistiinpanojen jakso 4.4).

Tarkastellaan vastaavaa satunnaismuuttujaa eli *estimaattoria* $\hat{\mu} = \overline{Y} = (Y_1 + \cdots + Y_n)/n$ (merkiten sitä siis samalla symbolilla kuin estimaattia).

a) Totea, että $\hat{\mu}$ on *harhaton* eli $E(\hat{\mu}) = \mu$. Mitä tämä tulkinnallisesti merkitsee ”toistetun aineistonkeruun” ajatuksen näkökulmasta?

b) Laske varianssi $\mathrm{var}(\hat{\mu})$ ja totea, että se lähestyy nollaa, kun havaintojen lukumäärä n kasvaa rajatta. Mitä tämä tulos voisi merkitä?

Tämän tehtävän aihepiireistä puhutaan muistiinpanojen luvussa 3 (ks. erityisesti jaksot 3.4 ja 3.5). Emme kuitenkaan syvenny niihin tällä kurssilla laajemmin.

KÄÄNNÄ!

5. a) Olkoot $y_1, \ldots, y_n$ reaalilukuja ja $\bar{y} = \sum_{i=1}^n y_i/n$ niiden aritmeettinen keskiarvo (otoskeskiarvo) kuten edellä. Näytä, että jos μ on reaaliluku, niin

$$\sum_{i=1}^{n} (y_i - \mu)^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2.$$

Luennoilla tätä hajotelmaa käytettiin normaalijakauman su-estimoinnin yhteydessä (ks. muistiinpanojen sivu 29). Ehdotus: Kirjoita aluksi $y_i - \mu = (y_i - \bar{y}) + (\bar{y} - \mu)$.

b) Olkoot $Y_1, \ldots, Y_n$ riippumattomia satunnaismuuttujia, joille $E(Y_i) = \mu$ ja $\mathrm{var}(Y_i) = \sigma^2$ (esimerkiksi $Y_i \sim N(\mu, \sigma^2)$). Osoita a-kohdan avulla, että tavanomaiselle otosvarianssille

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

pätee $E(S^2) = \sigma^2$. Tämä kertoo, että S^2 on *harhaton* estimaattori σ^2:lle.