Tilastollinen päättely II, syksy 2015 – kevät 2016
Harjoitus 2 (10. ja 12. 11. 2015)

Näissä tehtävissä käsitellään uskottavuusfunktiota ja suurimman uskottavuuden estimointia. Luentomonisteen jaksot 2.1–2.3. Tehtävissä esiintyvien jakaumien tiheysfunktioiden lausekkeet löytyvät monisteen lopussa olevasta liitteestä.

1. (Monisteen teht. 2.2.) Olkoot $y_1, \ldots, y_n$ ja a mielivaltaisia reaalilukuja. Varmista, että

$$\sum_{i=1}^{n}(y_i - a)^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2 + n(\overline{y} - a)^2,$$

kun $\overline{y} = \sum_{i=1}^{n} y_i/n$. [Tulosta käytettiin normaalijakaumamallin uskottavuusfunktion muodostamisessa, ks. monisteen kohta 2.1.4.]

2. (Monisteen teht. 2.3.) Olkoot $Y_1, \ldots, Y_n \sim \mathrm{Exp}(\lambda) \perp\!\!\!\perp$. Kirjoita vastaava tilastollisen mallin lauseke (yhteistiheysfunktio). Muodosta sitten aineistoa $\mathbf{y} = (y_1, \ldots, y_n)$ vastaavat uskottavuus- ja log-uskottavuusfunktio sekä määritä huolellisesti perustellen parametrin λ suurimman uskottavuuden estimaatti. Hahmottele log-uskottavuusfunktion kuvaajaa.

3. (Monisteen teht. 2.6.) Olkoon mallina $Y_1, \ldots, Y_n \sim \mathrm{Tas}(\theta, \theta + 1) \perp\!\!\!\perp$. Johda aineistoa $\mathbf{y} = (y_1, \ldots, y_n)$ vastaava uskottavuusfunktio ja totea, että se saa suurimman arvonsa jokaisessa välin $\big(y_{(n)} - 1, y_{(1)}\big)$ pisteessä, kun merkitään $y_{(1)} = \min(y_1, \ldots, y_n)$ ja $y_{(n)} = \max(y_1, \ldots, y_n)$. Siten su-estimaatti $\hat{\theta}(\mathbf{y})$ ei ole yksikäsitteinen (todennäköisyydellä yksi).

Mitä mahtaa tapahtua välin $\big(y_{(n)} - 1, y_{(1)}\big)$ pituudelle, kun havaintojen lukumäärä $n \to \infty$?

4. (Vrt. monisteen teht. 2.10.) Ladislaus Bortkiewicz julkaisi vuonna 1898 kirjan *The Law of Small Numbers*, joka käsitteli Poissonin jakaumaa. Kuuluisin kirjassa esitetty aineisto oli hevosenpotkuun kuolleiden miesten vuosittaiset lukumäärät neljässätoista Preussin armeijan yksikössä kahdenkymmenen vuoden (1875–1894) ajalta, yhteensä siis 280 havaintoa. Yhteenveto aineistosta on alla.

Kuolleita	0	1	2	3	4	≥ 5
Havaintoja	144	91	32	11	2	0

Bortkiewicz havaitsi, että Poissonin jakauma sopii kuvaamaan hevosenpotkuun kuolleiden miesten määrää kussakin yksikössä vuoden aikana. Olkoon μ kyseisen Poissonin jakauman odotusarvo, ja oletetaan, että kuolleiden lukumäärät eri vuosina ja eri yksiköissä ovat toisistaan riippumattomat. Tällöin aineistoa voidaan mallintaa satunnaismuuttujilla $Y_1, \ldots, Y_{280} \sim P(\mu) \perp\!\!\!\perp$.

Muodosta aineistoa vastaava log-uskottavuusfunktio ja määritä μ:n suurimman uskottavuuden estimaatti.

[Tästä aineistosta on netissä paljon tarinaa (hakusanat esim. ”horse kick data”). Koko aineisto löytyy esim. osoitteesta http://www.math.uah.edu/stat/data/HorseKicks.html.]

KÄÄNNÄ!

5. (Monisteen teht. 2.8.) Tarkastellaan riippumatonta otosta gammajakaumasta: $Y_1, \ldots, Y_n \sim G(\alpha, 1/\beta) \perp\!\!\!\perp$, jossa $\alpha, \beta > 0$.

a) Parametrina on (α, β). Totea, että uskottavuusyhtälöt voidaan saattaa muotoon

$$\begin{cases} \log \alpha - \psi(\alpha) = \log \overline{y} - \frac{1}{n} \sum_{i=1}^{n} \log y_i, \\ \beta = \overline{y}/\alpha, \end{cases}$$

jossa $\psi(\alpha) = \Gamma'(\alpha)/\Gamma(\alpha)$ on ns. *digammafunktio.* Yhtälöitä ei voi ratkaista suljetussa muodossa.

b) Parametrina on vain β, ja α on tunnettu luku. Mikä on β:n su-estimaatti?

Muista: Presemo-keskusteluhuone osoitteessa http://presemo.helsinki.fi/pni/.