Tilastollinen päättely II, syksy 2015 – kevät 2016
Harjoitus 7 (19. ja 21. 1. 2016)
Malliratkaisut

1. Perustehtävä tarkentuvuudesta (monisteen jakso 3.5). Jatkoa harjoituksen 5 tehtävään 2 ja harjoituksen 6 tehtävään 3. Perustele, että estimaattorit $\hat{\beta}$ ja T ovat tarkentuvia, jos $x_i \geq c$ kaikilla i ja $c > 0$ on vakio.

Ratkaisu

Harjoitusten 5 tehtävässä 2 on osoitettu molempien estimaattorien harhattomuus. Nyt kurssimonisteen 3.5.3 perusteella riittää, että osoitetaan estimaattorien varianssien suppenevan kohti nollaa otoskoon kasvaessa.

Harjoitusten 6 tehtävässä 3 on laskettu molemmille estimaattoreille varianssit.

$$var(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}$$

ja

$$var(T) = \frac{n\sigma^2}{(\sum_{i=1}^n x_i)^2}.$$

Oletetaan, että $x_i \geq c > 0$ kaikilla i. Nyt

$$0 \leq \lim_{n\to\infty} var(\hat{\beta}) = \lim_{n\to\infty} \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \leq \lim_{n\to\infty} \frac{\sigma^2}{nc^2} = 0$$

ja

$$0 \leq \lim_{n\to\infty} var(T) = \lim_{n\to\infty} \frac{n\sigma^2}{(\sum_{i=1}^n x_i)^2} \leq \lim_{n\to\infty} \frac{n\sigma^2}{(nc)^2} = \lim_{n\to\infty} \frac{\sigma^2}{nc^2} = 0.$$

Siis estimaattorit $\hat{\beta}$ ja T ovat tarkentuvia, kun $x_i \geq c > 0$ kaikilla i.

Tehtävät 2–4 ovat ”täydentävää ainesta” ja liittyvät ns. eksponenttiperheen jakaumiin. Näillä jakaumilla on monia todennäköisyyslaskennan ja tilastollisen päättelyn kannalta miellyttäviä ominaisuuksia, joihin on viitattu esim. monisteen kohdissa 2.5.2 ja 4.2.5.

2. (Monisteen tehtävä 2.20.) Parametrilla $\theta \in \Omega \subset \mathbb{R}^d$ indeksoitu jakaumaperhe on d-ulotteinen *eksponenttiperhe*, jos sen yptf/ytf voidaan kirjoittaa muotoon

$$f_{\mathbf{Y}}(\mathbf{y};\theta) = c(\theta)h(\mathbf{y}) \exp\left\{\sum_{j=1}^d \phi_j(\theta)t_j(\mathbf{y})\right\},$$

jossa $c(\theta)$ sekä $\phi_j(\theta)$:t ovat reaaliarvoisia ja riippuvat vain parametrista θ ja vastaavasti $h(\mathbf{y})$ sekä $t_j(\mathbf{y})$:t ovat reaaliarvoisia vain aineistosta $\mathbf{y}$ riippuvia tunnuslukuja. Vektoria $\phi = \big(\phi_1(\theta), \dots, \phi_d(\theta)\big)$ kutsutaan perheen *luonnolliseksi parametriksi.*

a) Osoita, että mallia $K \sim \mathrm{Bin}(n, \theta)$ vastaava jakaumaperhe $f_K(k;\theta)$ (ks. monisteen esim. 2.1.5) on yksiulotteinen eksponenttiperhe, luonnollisena parametrina θ:n logit-muunnos $\log\{\theta/(1-\theta)\}$.

b) Osoita, että mallia $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2) \perp\!\!\!\perp$ vastaava jakaumaperhe $f_{\mathbf{Y}}(\mathbf{y}; \mu, \sigma^2)$ (ks. esim. monisteen kohta 1.2.3) on kaksiulotteinen eksponenttiperhe, luonnollisena parametrina $(\mu/\sigma^2, 1/\sigma^2)$. [*Ehdotus.* Aloita kirjoittamalla $(y_i - \mu)^2 = y_i^2 - 2\mu y_i + \mu^2$.]

Ratkaisu

a) Mallin pistetodennäköisyysfunktio voidaan esittää muodossa

$$
\begin{aligned}
f_K(k;\theta) &= \binom{n}{k}\theta^k(1-\theta)^{n-k} \\
&= \binom{n}{k}\exp\{\log(\theta^k(1-\theta)^{n-k})\} \\
&= \binom{n}{k}\exp\{k\log(\theta) + (n-k)\log(1-\theta)\} \\
&= \binom{n}{k}\exp\{k\log(\theta) + n\log(1-\theta) - k\log(1-\theta)\} \\
&= \binom{n}{k}\exp\{n\log(1-\theta)\}\exp\{k(\log(\theta) - \log(1-\theta))\} \\
&= \binom{n}{k}\exp\{n\log(1-\theta)\}\exp\{k\log(\theta/(1-\theta))\} \\
&= c(\theta)h(k)\exp\{\phi(\theta)t(k)\},
\end{aligned}
$$

missä $c(\theta) = \exp\{n\log(1-\theta)\}$, $h(k) = \binom{n}{k}$, $\phi(\theta) = \log(\theta/(1-\theta))$ ja $t(k) = k$. Kyseessä on siis yksiulotteinen eksponenttiperhe.

b)Merkitään mallin parametria $\theta = (\mu, \sigma^2)$. Mallin ytf voidaan esittää muodossa

$$
\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y};\theta) &= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(y_i-\mu)^2}{2\sigma^2}\right\} \\
&= (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i^2 - 2\mu y_i + \mu^2)\right\} \\
&= (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{n\mu^2}{2\sigma^2}\right\}\exp\left\{\frac{\mu\sum_{i=1}^{n}y_i}{\sigma^2} - \frac{\sum_{i=1}^{n}y_i^2}{2\sigma^2}\right\} \\
&= c(\theta)h(\mathbf{y})\exp\left\{\sum_{j=1}^{2}\phi_j(\theta)t_j(\mathbf{y})\right\},
\end{aligned}
$$

missä

$$
\begin{aligned}
&c(\theta) = (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{n\mu^2}{2\sigma^2}\right\},\ h(\mathbf{y}) = 1,\ \phi_1(\theta) = \frac{\mu}{\sigma^2},\ \phi_2(\theta) = \frac{1}{\sigma^2}, \\
&t_1(\mathbf{y}) = \sum_{i=1}^{n}y_i \quad \text{ja} \quad t_2(\mathbf{y}) = -\frac{1}{2}\sum_{i=1}^{n}y_i^2,
\end{aligned}
$$

eli kyseessä on kaksiulotteinen eksponenttiperhe

3. Jatkoa edelliseen tehtävään. Perustele, että eksponenttiperheen määritelmässä vakio $c(\theta)$ riippuu θ:sta vain ϕ:n kautta. Siten perhe voidaan aina uudelleenparametroida luonnollisen parametrinsa avulla. [*Muista.* Funktiot $f_{\mathbf{Y}}(\mathbf{y};\theta)$ ovat yhteispistetodennäköisyysfunktiota tai -tiheysfunktioita!]

Ratkaisu

$f_{\mathbf{Y}}(\mathbf{y};\theta)$ on yptf/ytf, joten sen täytyy summautua/integroitua ykköseksi yli kaikkien mahdollisten $\mathbf{y}$:n arvojen. Siten kun ϕ on kiinnitetty, täytyy jatkuvan jakauman tapauksessa olla

$$\begin{aligned} 1 &= \int_{\mathbb{R}^n} c(\theta)h(\mathbf{y})\exp\left\{\sum_{j=1}^{d}\phi_j(\theta)t_j(\mathbf{y})\right\}\mathrm{d}\mathbf{y} \\ \Leftrightarrow c(\theta) &= \frac{1}{\int_{\mathbb{R}^n} h(\mathbf{y})\exp\left\{\sum_{j=1}^{d}\phi_j(\theta)t_j(\mathbf{y})\right\}\mathrm{d}\mathbf{y}}, \end{aligned}$$

eli vakio $c(\theta)$ riippuu θ:sta vain ϕ:n kautta. Diskreetin jakauman tapauksessa integraali korvataan summalla.

4. Perustele, että jakaumaperhe $\mathrm{Tas}(0,\theta)$, $\theta > 0$, ei ole (yksiulotteinen) eksponenttiperhe.

Ratkaisu

Eksponenttiperheeseen kuuluvan jakauman alusta on muotoa

$$\begin{aligned} A &= \{\mathbf{y} : f_{\mathbf{Y}}(\mathbf{y};\theta) > 0\} \\ &= \left\{\mathbf{y} : c(\theta)h(\mathbf{y})\exp\left\{\sum_{j=1}^{d}\phi_j(\theta)t_j(\mathbf{y})\right\} > 0\right\} \end{aligned}$$

Havaitaan, että aina $c(\theta) \neq 0$. Lisäksi $\exp\left\{\sum_{j=1}^{d}\phi_j(\theta)t_j(\mathbf{y})\right\} > 0$. Näiden perusteella huomataan, että

$$A = \{\mathbf{y} : f_{\mathbf{Y}}(\mathbf{y};\theta) > 0\} = \{\mathbf{y} : f_{\mathbf{Y}}(\mathbf{y};\theta) \neq 0\}$$

on sama joukko kuin $\{\mathbf{y} : h(\mathbf{y}) \neq 0\}$. Eksponenttiperheeseen kuuluvan jakauman alusta on siis riippumaton parametrista θ.

Tehtävän mallin tf:n alusta

$$A = \{y : f_Y(y;\theta) > 0\} = \{y : 0 < y < \theta\}$$

kuitenkin riippuu parametrista θ, joten malli ei voi kuulua eksponenttiperheeseen.

Moniulotteisessa tapauksessa mallin ytf:n alusta olisi muotoa

$$A = \{\mathbf{y} : f_{\mathbf{Y}}(\mathbf{y};\theta) > 0\} = \{\mathbf{y} : 0 < y_i < \theta \quad \forall i = 1,\ldots,n\},$$

joka myöskin riippuu parametrista θ.