Tilastollinen päättely II, syksy 2015 – kevät 2016
Harjoitus 8 (26. ja 28. 1. 2016), ratkaisut

Tehtävät 1–3 liittyvät su-estimaattorien asymptoottiseen normaalisuuteen, jota käsitellään monisteen jaksossa 3.6. Palauta myös mieleen todennäköisyyslaskennasta keskeinen raja-arvolause.

1. (Vrt. monisteen teht. 3.13.) Olkoot $Y_1, \dots, Y_n \sim P(\mu)$ $\perp\!\!\!\perp$. Mitä normaalijakaumaa su-estimaattori $\hat{\mu} = \overline{Y}$ approksimatiivisesti noudattaa, kun n on suuri? Mihin $\sqrt{n}(\hat{\mu} - \mu)$ toisin sanoen suppenee jakaumaltaan (eli heikon suppenemisen mielessä)?

Ratkaise tämä tehtävä kahdella tavalla: yhtäältä suoraan keskeisen raja-arvolauseen avulla ja toisaalta monisteen kohdan 3.6.5 avulla.

Ratkaisu. Selvitetään estimaattorin jakauma ensin keskeisen raja-arvolauseen avulla. Olkoon $Z \sim N(0, 1)$ ja merkitään

$$Z_n := \sqrt{\frac{n}{\mu}} \left(\overline{Y} - \mu\right).$$

Satunnaismuuttujat $Y_1, \dots, Y_n$ ovat riippumattomia ja samoin jakautuneita odotusarvonaan ja varianssinaan μ, joten keskeisen raja-arvolauseen nojalla $Z_n \xrightarrow{w} Z$, kun $n \to \infty$. Tästä seuraa, että kaikilla tarpeeksi suurilla n pätee likimääräisesti

$$\hat{\mu} = \overline{Y} \sim N\left(\mu, \frac{\mu}{n}\right),$$

jota merkitään

$$\overline{Y} \underset{as}{\sim} N\left(\mu, \frac{\mu}{n}\right).$$

Tarkastellaan seuraavaksi estimaattorin jakaumaa monisteen kohdassa 3.6.5 esitetyn su-estimaattorin asymptoottista jakaumaa koskevan lauseen avulla. Tehtävän Poisson-mallin log-uskottavuusfunktio on

$$\ell(\mu; \mathbf{y}) = -n\mu + n\overline{y}\log(\mu),$$

jonka toinen derivaatta parametrin suhteen on

$$\ell''(\mu; \mathbf{y}) = -\frac{n\overline{y}}{\mu^2}.$$

Tästä saadaan Fisherin informaatio

$$i(\mu) = \mathrm{E}(-\ell''(\mu; \mathbf{Y})) = \frac{nE(\overline{Y})}{\mu^2} = \frac{n}{\mu}.$$

Edellä mainitun lauseen perusteella

$$\hat{\mu} \underset{as}{\sim} N\left(\mu, \frac{\mu}{n}\right).$$

2. Oletetaan, että havaintoja vastaavat satunnaismuuttujat $Y_1, \dots, Y_n$ ovat riippumattomia ja noudattavat kukin jatkuvaa jakaumaa, jonka tiheysfunktio on $f(y; \theta) = \theta y^{\theta - 1}$, kun $0 \le y \le 1$ (ja $= 0$ muulloin).

a) Etsi parametrin θ suurimman uskottavuuden estimaattori $\hat{\theta}$.

b) Mitä normaalijakaumaa $\hat{\theta}$ approksimatiivisesti noudattaa, kun n on suuri?

Ratkaisu. a) Mallin tiheysfunktio on

$$f(\mathbf{y};\theta) = \theta^n \prod_{i=1}^{n} y_i^{\theta-1} = \theta^n \prod_{i=1}^{n} y_i^{\theta} \prod_{i=1}^{n} y_i^{-1},$$

josta (kertomalla parametrista riippumaton osa pois) saadaan log-uskottavuusfunktioksi

$$\ell(\theta;\mathbf{y}) = n\log(\theta) + \theta \sum_{i=1}^{n} \log(y_i).$$

Asettamalla ensimmäinen derivaatta nollaksi saadaan

$$\ell'(\theta;\mathbf{y}) = \frac{n}{\theta} + \sum_{i=1}^{n} \log(y_i) = 0 \Leftrightarrow \theta = -\frac{n}{\sum_{i=1}^{n} \log(y_i)}.$$

Toinen derivaatta $\ell''(\theta;\mathbf{y})) = -n/\theta^2$ on selvästi aina negatiivinen, joten tämä on myös suurimman uskottavuuden estimaatti.

b) Edellä todetun perusteella Fisherin informaatio on $i(\theta) = n/\theta^2$. Kurssimonisteen kohdan 3.6.5. lauseen nojalla

$$\hat{\theta} \underset{as}{\sim} N\left(\theta, \frac{\theta^2}{n}\right).$$

3. (Monisteen teht. 3.15.) Olkoot $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$. Parametrin (μ, σ^2) su-estimaattori $(\hat{\mu}, \hat{\sigma}^2)$ on laskettu monisteen kohdassa 2.2.6.

a) Palauta todennäköisyyslaskennasta mieleen, mikä on satunnaismuuttujaparin $(\hat{\mu}, \hat{\sigma}^2)$ tarkka yhteisjakauma (reunajakaumat ja komponenttien välinen riippuvuus/riippumattomuus).

b) Mitkä ovat sen kaksiulotteisen normaalijakauman parametrit (odotusarvovektori ja kovarianssimatriisi), jota $(\hat{\mu}, \hat{\sigma}^2)$ approksimatiivisesti noudattaa, kun n on suuri?

[*Apuja.* Syksyn 2015 Todennäköisyyslaskenta II, luentokalvojen kohta 10.9. Tarkasteltavan mallin informaatiomatriisi on laskettu aikaisemmin luennoilla ja monisteen luvussa 2.]

Ratkaisu. a) Odotusarvon estimaattori $\hat{\mu}$ on normaalijakautuneiden satunnaismuuttujien keskiarvona normaalijakautunut odotusarvolla μ ja varianssilla σ^2/n. Varianssin estimaattorille $\hat{\sigma}^2$ pätee

$$\frac{n\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \sim \chi^2_{n-1},$$

mikä voidaan ilmaista myös muodossa

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{n} \chi^2_{n-1}.$$

Lisäksi estimaattorit ovat keskenään riippumattomat.

b) Normaalimallissa parametrit μ ja σ^2 ovat ortogonaaliset, joten Fisherin informaatiomatriisin alkiot diagonaalin ulkopuolella ovat nollia. Kurssimonisteen kohdassa 3.6.8 esitetyn lauseen perusteella

$$(\hat{\mu}, \hat{\sigma}^2) \underset{as}{\sim} N((\mu, \sigma^2), \boldsymbol{i}^{-1}(\mu, \sigma^2)),$$

missä Fisherin informaatiomatriisin käänteismatriisi on

$$\boldsymbol{i}^{-1}(\mu, \sigma^2) = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{bmatrix}.$$

Tehtävä 4 liittyy tyhjentävyyden käsitteeseen. Opiskele sitä varten ainakin alustavasti monisteen lukua 4.

4. (Monisteen teht. 4.1.) Olkoot $Y_1, \ldots, Y_n \sim \text{Exp}(\lambda) \perp\!\!\!\perp$. Osoita faktorointikriteerin avulla, että otoskeskiarvo $\overline{Y}$ tai yhtä hyvin summa $\sum_{i=1}^n Y_i$ on parametrin λ tyhjentävä tunnusluku.

Ratkaisu. Mallin yhteistiheysfunktio on

$$f_{\mathbf{Y}}(\mathbf{y}; \lambda) = \prod_{i=1}^{n} \lambda \exp\{-\lambda y_i\} = \lambda^n \exp\{-\lambda n \overline{y}\}.$$

Yhteistiheysfunktio riippuu aineistosta vain $\bar{y}$:n tai yhtäpitävästi summan $\sum_{i=1}^n y_i$ kautta, joten molemmat näistä ovat tyhjentäviä tunnuslukuja.