Tilastollinen päättely II, syksy 2015 – kevät 2016
Harjoitus 6 (8. ja 10. 12. 2015)

Tehtävät 1–4 liittyvät harhattomuuteen ja tehokkuuteen. Monisteen jaksot 3.2 ja 3.4.

1. Jatkoa harjoituksen 5 tehtävään 4. Tarkastellaan edelleen mallia $Y_1, \ldots, Y_n \sim \text{Tas}(0, \theta) \perp\!\!\!\perp$ ja θ:n estimointia.

a) Totea, että estimaattori $\check{\theta} = \dfrac{n+1}{n}\hat{\theta}$ on harhaton.

b) Totea, että myös estimaattori $\tilde{\theta} = 2\overline{Y}$ on harhaton. (Tämä on momenttimenetelmän antama; ks. monisteen esim. 3.3.3.)

Ratkaisu

a) Harjoitusten 5 tehtävän 4 ratkaisuissa todetaan, että $\mathrm{E}(\hat{\theta}) = n\theta/(n+1)$, joten

$$\mathrm{E}(\check{\theta}) = \mathrm{E}\left(\frac{n+1}{n}\hat{\theta}\right) = \frac{n+1}{n}\mathrm{E}(\hat{\theta}) = \frac{n+1}{n}\frac{n}{n+1}\theta = \theta.$$

b) Tasajakauman ominaisuuksien nojalla $\mathrm{E}(Y_i) = \theta/2$, $i = 1, \ldots, n$. Tästä saadaan

$$\mathrm{E}(\tilde{\theta}) = \mathrm{E}(2\overline{Y}) = \mathrm{E}\left(\frac{2}{n}\sum_{i=1}^{n} Y_i\right) = \frac{2}{n}\sum_{i=1}^{n}\mathrm{E}\left(Y_i\right) = \frac{2}{n}\cdot n \cdot \frac{\theta}{2} = \theta.$$

2. Jatkoa edelliseen tehtävään. Laske estimaattorien $\hat{\theta}$, $\check{\theta}$ ja $\tilde{\theta}$ keskineliövirheet ja vertaa niitä toisiinsa (kun $n > 2$). Kiinnitä erityisesti huomiota siihen, miten ne käyttäytyvät suurilla n:n arvoilla. (Vihjeitä kääntöpuolella.)

Ratkaisu Estimaattorin $\hat{\theta}$ harha on $b(\theta) = n/(n+1)\theta - \theta = \theta/(n+1)$. Toinen origomomentti on (ks. harj 5, teht 4)

$$\mathrm{E}(\hat{\theta}^2) = \int_0^\theta y^2\theta^{-n}ny^{n-1}dy = \frac{n}{\theta^n}\int_0^\theta y^{n+1}dy = \frac{n}{\theta^n}\cdot\frac{\theta^{n+2}}{n+2} = \frac{n}{n+2}\theta^2,$$

joten estimaattorin $\hat{\theta}$ varianssiksi saadaan

$$\mathrm{var}(\hat{\theta}) = \mathrm{E}(\hat{\theta}^2) - (\mathrm{E}(\hat{\theta}))^2 = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 = \frac{n}{(n+1)^2(n+2)}\theta^2.$$

Yhdistämällä tämä estimaattorin $\hat{\theta}$ harhaan

$$b(\theta) = \mathrm{E}(\hat{\theta}) - \theta = \frac{\theta^2}{(n+1)^2},$$

saadaan kurssimonisteen luvun 3.4 määritelmän mukaan SU-estimaattorin keskineliövirheeksi

$$\mathrm{E}[(\hat{\theta} - \theta)^2] = \frac{n}{(n+1)^2(n+2)}\theta^2 + \frac{\theta^2}{(n+1)^2} = \frac{2n+2}{(n+1)^2(n+2)}\theta^2 = \frac{2}{(n+1)(n+2)}\theta^2.$$

SU-estimaattorista skaalaamalla saadulle harhattomalle estimaattorille $\check{\theta}$ pätee harhattomuuden nojalla

$$\mathrm{E}[(\check{\theta} - \theta)^2] = \mathrm{var}(\check{\theta}) = \frac{(n+1)^2}{n^2}\mathrm{var}(\hat{\theta}) = \frac{1}{n(n+2)}\theta^2.$$

Momenttiestimaattorin keskineliövirhe saadaan laskettua varianssin laskusääntöjen ja harhattomuuden avulla

$$\mathrm{E}[(\tilde{\theta}-\theta)^2] = \frac{4}{n^2}\mathrm{var}\left(\sum_{i=1}^n Y_i\right) = \frac{4}{n^2}\cdot n\cdot\frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

Estimaattorien keskineliövirheet voidaan asettaa suuruusjärjestykseen

$$\frac{\theta^2}{n(n+2)} < \frac{2\theta^2}{(n+1)(n+2)} < \frac{\theta^2}{3n}$$

kaikilla $n > 2$. Siten keskineliövirheen mielessä tarkastelluista estimaattoreista paras on $\check{\theta}$ ja momentti estimaattori $\tilde{\theta}$ huonoin.

3. Jatkoa harjoituksen 5 tehtävään 2. Laske estimaattorien $\hat{\beta}$ ja T varianssit. Osoita, että $\hat{\beta}$ on tehokkaampi kuin T. (Apu kääntöpuolella.)

Ratkaisu Satunnaismuuttujien $Y_i, i = 1, \ldots, n$ varianssi on σ^2, joten

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

ja

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

Käyttäen annettua vihjettä saadaan

$$\mathrm{var}(T) = \frac{n\sigma^2}{\left(\sum_{i=1}^n x_i\right)^2} \geq \frac{n\sigma^2}{n\sum_{i=1}^n x_i^2} = \frac{\sigma^2}{\sum_{i=1}^n x_i^2} = \mathrm{var}(\hat{\beta}).$$

Siis SU-estimaattorin $\hat{\beta}$ varianssi on pienempi ja siten se on tehokkaampi.

4. (Monisteen teht. 3.5.)

a) Palautetaan todennäköisyyslaskennasta mieleen, että jos $X_1, \ldots, X_k \sim N(0,1) \perp\!\!\!\perp$, niin $X_1^2 + \cdots + X_k^2$ noudattaa khii-toiseen-jakaumaa χ_k^2. Tarkista tämän avulla, että χ_k^2-jakauman odotusarvo on k, varianssi $2k$ ja toinen momentti $k^2 + 2k$. [*Muista.* Standardinormaalijakauman neljäs origomomentti on 3.]

b) Olkoot $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2) \perp\!\!\!\perp$, ja määritellään V kuten harjoituksen 4 tehtävässä 3: $V = \sum_i (Y_i - \overline{Y})^2$. Todennäköisyyslaskennassa on opittu, että $V/\sigma^2 \sim \chi_{n-1}^2$. Näytä tämän tiedon ja a-kohdan perusteella, että σ^2:n estimaattorin cV keskineliövirhe on

$$E[(cV - \sigma^2)^2] = [(n^2-1)c^2 - 2(n-1)c + 1]\,\sigma^4.$$

Näytä sitten derivaattatarkastelun avulla, että se saa pienimmän arvonsa (c:n funktiona) täsmälleen pisteessä $c = 1/(n+1)$.

Opetus. Varianssin σ^2 estimaattoriksi on normaalijakaumamallissa siis ainakin kolme muotoa cV olevaa perusteltua kandidaattia: $V/(n-1)$ (harhaton), V/n (su) ja $V/(n+1)$ (pienin keskineliövirhe). Ei ole mahdollista yksiselitteisesti sanoa, mikä niistä on ”paras”, mutta $V/(n-1)$ on varmasti käytetyin.

Ratkaisu

a) Odotusarvo on

$$\mathrm{E}\left(\sum_{i=1}^k X_i^2\right) = \sum_{i=1}^k \mathrm{E}(X_i^2) = k.$$

X_i:en neliöiden varianssit saadaan laskettua standardinormaalijakauman toisen ja neljännen momentin avulla:

$$\mathrm{var}(X_i^2) = \mathrm{E}(X_i^4) - \mathrm{E}(X_i^2)^2 = 3 - 1^2 = 2.$$

Tästä saadaan χ_k^2:n jakauman varianssi

$$\mathrm{var}\left[\sum_{i=1}^{k} X_i^2\right] = \sum_{i=1}^{k} \mathrm{var}(X_i^2) = 2k$$

ja toinen momentti

$$\mathrm{E}\left[\left(\sum_{i=1}^{k} X_i^2\right)^2\right] = \left(\mathrm{E}\left[\sum_{i=1}^{k} X_i^2\right]\right)^2 + \mathrm{var}\left[\sum_{i=1}^{k} X_i^2\right] = k^2 + 2k.$$

b) Annetun jakaumatuloksen ja a-kohdan nojalla

$$\mathrm{E}\left(\frac{V}{\sigma^2}\right) = n - 1$$

ja

$$\frac{\mathrm{E}(V^2)}{\sigma^4} = \mathrm{E}\left[\left(\frac{V}{\sigma^2}\right)^2\right] = (n-1)^2 + 2(n-1) = n^2 - 1.$$

Siten cV:n keskineliövirheeksi saadaan

$$\begin{aligned}
\mathrm{E}[(cV - \sigma^2)^2] &= c^2\mathrm{E}(V^2) - 2c\sigma^2\mathrm{E}(V) + \sigma^4 \\
&= \left(c^2\frac{\mathrm{E}(V^2)}{\sigma^4} - 2c\frac{\mathrm{E}(V)}{\sigma^2} + 1\right)\sigma^4 \\
&= \left[(n^2-1)c^2 - 2(n-1)c + 1\right]\sigma^4.
\end{aligned}$$

Etsitään tämän minimikohta derivoimalla ja ratkaisemalla derivaatan nollakohdat:

$$\begin{aligned}
\frac{\partial}{\partial c}E[(cV - \sigma^2)^2] = 2(n^2-1)c - 2(n-1) &= 0 \\
c &= \frac{n-1}{n^2-1} \\
c &= \frac{n-1}{(n-1)(n+1)} \\
c &= \frac{1}{n+1}.
\end{aligned}$$

Keskineliövirheen lauseke on c:n funktiona tarkasteltuna ylöspäin aukeava paraabeli kun $n \geq 2$, joten sen globaali minimikohta on sen derivaatan nollakohdassa $c = 1/(n+1)$.

Tehtävää 5 varten opiskele momenttimenetelmää koskeva monisteen jakso 3.3. Emme käsittele momenttimenetelmää varsinaisesti luennoilla, mutta aiheeseen liittyviä kysymyksiä voi esittää esim. ti 8. 12. ja to 10. 12.

5. Estimoi momenttimenetelmällä

a) parametri λ, kun $Y_1, \ldots, Y_n \sim \mathrm{Exp}(\lambda) \perp\!\!\!\perp$,

b) parametri (α, β), kun $Y_1, \ldots, Y_n \sim G(\alpha, 1/\beta) \perp\!\!\!\perp$. (Kyseessä on gammajakauma, ks. monisteen liite. Vrt. myös harjoituksen 2 tehtävään 5a.)

Ratkaisu

a) Satunnaismuuttujat Y_i ovat eksponenttijakautuneita, joten $\mathrm{E}(Y_i) = 1/\lambda$. Momenttiestimaattori $\tilde{\lambda}$ saadaan ratkaisemalla parametri λ yhtälöstä

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{\lambda},$$

eli

$$\tilde{\lambda} = \frac{n}{\sum_{i=1}^{n} Y_i}.$$

Tämä on sama kuin parametrin λ SU-estimaattori.

b) Satunnaismuuttujat Y_i, noudattavat $G(\alpha, 1/\beta)$-jakaumaa, joten $\mathrm{E}(Y_i) = \alpha\beta$ ja $\mathrm{Var}(Y_i) = \alpha\beta^2$, $i = 1, \ldots, n$. Nyt estimoitavana on kaksiulotteinen parametrivektori, joten momenttiestimaattori saadaan ratkaisemalla parametrit yhtälöryhmästä

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} Y_i = \alpha\beta \\ \frac{1}{n}\sum_{i=1}^{n} Y_i^2 = \alpha\beta^2 + (\alpha\beta)^2 . \end{cases}$$

Tulokseksi saadaan

$$\begin{cases} \tilde{\alpha} = \frac{\overline{Y}^2}{\tilde{\sigma}^2}, \\ \tilde{\beta} = \frac{\tilde{\sigma}^2}{\overline{Y}}, \end{cases}$$

jossa $\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2$.