Tilastollinen päättely, syksy 2014 – kevät 2015
Harjoitus 5 (2. ja 4. 12. 2014)

1. (Monisteen teht. 3.1.) Johda funktion $g(\boldsymbol{\theta})$ estimaattorin T keskineliövirheelle monisteen kohdassa 3.4.1 mainittu hajotelma

$$E_{\boldsymbol{\theta}}[(T - g(\boldsymbol{\theta}))^2] = \mathrm{var}_{\boldsymbol{\theta}}(T) + b(\boldsymbol{\theta})^2,$$

   jossa $b(\boldsymbol{\theta})$ on estimaattorin harha.

2. (Monisteen teht. 3.2.) Tarkastellaan Poisson-mallia $Y_1, \ldots, Y_n \sim P(\mu) \perp\!\!\!\perp$. Varmista, että su-estimaattori $\hat{\mu} = \overline{Y} = (Y_1 + \cdots + Y_n)/n$ on harhaton, ja laske sen varianssi. Onko $\hat{\mu}$ täystehokas?

3. (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.

4. (Vrt. monisteen teht. 3.10.) Mallissa $Y_1, \ldots, Y_n \sim \mathit{Tas}(0, \theta) \perp\!\!\!\perp$ on su-estimaattoriksi saatu $\hat{\theta} = \max(Y_1, \ldots, Y_n)$ (ks. kohta 2.2.8).

   a) Muodosta $\hat{\theta}$:n kertymäfunktio F lähtien havainnosta

$$P\{\hat{\theta} \le t\} = P\{Y_1 \le t\} \cdots P\{Y_n \le t\}$$

   ja derivoi siitä tiheysfunktio $f = F'$.

   b) Laske $\hat{\theta}$:n odotusarvo ja totea, että $\hat{\theta}$ on harhainen mutta asymptoottisesti harhaton.

5. Jatkoa edelliseen tehtävään.

   a) Totea, että estimaattorit $\tilde{\theta} = 2\overline{Y}$ ja $\check{\theta} = [(n+1)/n]\hat{\theta}$ ovat harhattomia. (Myöhemmin näemme, että $\tilde{\theta}$ on momenttimenetelmän antama.)

   b) Laske estimaattorien $\hat{\theta}$, $\tilde{\theta}$ ja $\check{\theta}$ keskineliövirheet ja vertaa niitä toisiinsa (kun $n > 2$). Kiinnitä myös huomiota siihen, mitä ”vauhtia” keskineliövirheet lähestyvät nollaa, kun $n \to \infty$. Mikä on keskineliövirheen mielessä paras näistä kolmesta estimaattorista? [*Ohje.* Aloita laskemalla $E(\hat{\theta}^2)$.]