**Tilastollinen päättely, syksy 2013 - kevät 2014**

Harjoitus 4

1. Oletetaan, että havainnot $Y_1, \ldots, Y_n$ ovat riippumattomia ja noudattavat jakaumaa, jonka tiheysfunktio on

$$f(y; \alpha, \beta) = \begin{cases} \left(\frac{\alpha}{\beta}\right)\left(\frac{y}{\beta}\right)^{\alpha-1} & \text{jos } y \in [0, \beta], \\ 0 & \text{jos } y \notin [0, \beta]. \end{cases}$$

   jossa $\alpha > 0$ ja $\beta > 0$. Johda suurimman uskottavuuden estimaatit parametreille $\alpha$ ja $\beta$.

2. Hernekasvit luokitellaan niiden tuottamien herneiden mukaan yhtäältä muodon puolesta (pyöreä tai särmikäs) ja toisaalta värin puolesta (keltainen tai vihreä). Näin hernekasvit jakautuvat neljään luokkaan: PK, PV, SK ja SV. Niiden esiintymistodennäköisyydet ovat genetiikan mukaan vastaavasti $\alpha\beta$, $\alpha(1-\beta)$, $(1-\alpha)\beta$ ja $(1-\alpha)(1-\beta)$, jossa $0 < \alpha < 1$ ja $0 < \beta < 1$.

   Parametrien $\alpha$ ja $\beta$ estimoimiseksi kasvatettiin sata (toisistaan riippumatonta) kasvia, jolloin havaittiin, että ne jakautuivat em. luokkiin seuraavasti:

| Luokka: | PK | PV | SK | SV |
|---|---|---|---|---|
| Havaintoja: | 52 | 21 | 17 | 10 |

   a) Ilmoita näitä havaintoja vastaava uskottavuusfunktio ja log-uskottavuusfunktio sekä suurimman uskottavuuden estimaatti $(\hat{\alpha}, \hat{\beta})$.

   b) Ovatko tarkasteltavan mallin parametrit $\alpha$ ja $\beta$ ortogonaaliset? [Vrt. monisteen kohta 2.4.6]

3. Tarkastellaan toistokoemallia $Y_1, \ldots, Y_n \sim B(\theta) \perp\!\!\!\perp$. Totea, että suurimman uskottavuuden estimaattori $\hat{\theta} = (Y_1 + \cdots + Y_n)/n$ on harhaton: $E_\theta(\hat{\theta}) = \theta$ kaikilla $\theta \in [0, 1]$, ja laske sen varianssi $\mathrm{var}(\hat{\theta})$. Palauta mieleen tämän mallin Fisherin informaatio $i(\theta)$. Mikä yhteys sillä on em. varianssiin?

4. Jatkoa edellisen harjoituksen tehtävälle 4. Olkoot $Y_1, \ldots, Y_n \perp\!\!\!\perp$ ja ja $Y_i \sim N(\beta x_i, \sigma_0^2)$, jossa $x_1, \ldots, x_n$ ovat tunnettuja lukuja ja $\sigma_0^2 > 0$ on tunnettu.

   (a) Osoita, että parametrin $\beta$ suurimman uskottavuuden estimaattori $\hat{\beta}$ on harhaton $\beta$:n estimaattori ja laske sen varianssi. Mikä yhteys $\hat{\beta}$:n varianssilla on tämän mallin Fisherin informaatioon $i(\beta)$?

   (b) Osoita, että estimaattorit $T_1 = \sum_{i=1}^n Y_i / \sum_{i=1}^n x_i$ ja $T_2 = [\sum_{i=1}^n (Y_i/x_i)]/n$ ovat myös harhattomia $\beta$:n estimaattoreita ja laske niiden varianssit. Vertaa saatuja variansseja suurimman uskottavuuden estimaattorin varianssiin.

5. Mallissa $Y_1, \ldots, Y_n \sim \text{Tas}(0, \theta) \perp\!\!\!\perp$ on su-estimaattoriksi saatu $\hat{\theta} = \max(Y_1, \ldots, Y_n)$ (ks. luentojen kohta 2.2.8).

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

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

   ja derivoi siitä tiheysfunktio $f = F'$.

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

   c) Laske $\hat{\theta}$:n varianssi ja keskineliövirhe $E[(\hat{\theta} - \theta)^2]$ ja vertaa jälkimmäistä momenttimenetelmän antaman harhattoman estimaattorin $\tilde{\theta} = 2\overline{Y}$ (luentojen kohta 3.3.3) varianssiin. Kumpi estimaattori on parempi?

   d) Olisiko $\check{\theta} = [(n+1)/n]\hat{\theta}$ hyvä estimaattori?

   [Monisteen harjoitustehtävä 3.10]