Tilastollinen päättely, syksy 2014 – kevät 2015
Harjoitus 4 (25. ja 27. 11. 2014)

1. Satunnaiskokeella on toisensa poissulkevat tulosvaihtoehdot A_j, $j = 1, \dots, p$, joiden todennäköisyydet ovat $\theta_j = P(A_j)$ ja jossa $\theta_1 + \cdots + \theta_p = 1$. Koetta toistetaan riippumattomasti n kertaa ja eri tulosvaihtoehtojen esiintymislukumäärät kootaan *frekvenssitaulukkoon*:

Tulos	A_1	A_2	$\cdots$	A_p
Lkm	k_1	k_2	$\cdots$	k_p

Vastaava satunnaisvektori $\boldsymbol{K} = (K_1, \dots, K_p)$ noudattaa *multinomijakaumaa*, jonka parametrit ovat n ja $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$. Sen yhteispistetodennäköisyydet ovat

$$P(K_1 = k_1, \dots, K_p = k_p) = c(k_1, \dots, k_p)\, \theta_1^{k_1} \cdots \theta_p^{k_p},$$

kun $k_1 + \cdots + k_p = n$ (ja $= 0$ muulloin). Tässä $c(k_1, \dots, k_p)$ on ns. multinomikerroin.

Kirjoita aineistoa $(k_1, \dots, k_p)$ vastaava log-uskottavuusfunktio parametrille $\boldsymbol{\theta}$ ja määritä $\boldsymbol{\theta}$:n suurimman uskottavuuden estimaatti. (Toistojen lukumäärä n on tunnettu luku eikä siis osa tarkasteltavan mallin parametria.)

Huom. Opastus kääntöpuolella!

2. (Monisteen teht. 2.14.) Olkoon $f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$ tilastollinen malli, jonka parametri θ on yksiulotteinen. Olkoon $\phi = \phi(\theta)$ kääntäen yksikäsitteinen parametrimuunnos, jonka käänteismuunnos on $\theta = \theta(\phi)$. Tarkastellaan uudelleenparametroitua mallia $f^*_{\boldsymbol{Y}}(\boldsymbol{y}; \phi) = f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta(\phi))$. Näytä, että sen havaittu informaatio ja Fisherin informaatio saadaan alkuperäisen mallin informaatioista kaavoilla

$$j^*(\hat{\phi}; \boldsymbol{y}) = j(\hat{\theta}; \boldsymbol{y})\, \theta'(\hat{\phi})^2, \qquad i^*(\phi) = i(\theta(\phi))\, \theta'(\phi)^2.$$

Oletetaan, että malli täyttää kaikki tarpeelliset säännöllisyysehdot ja että parametrimuunnos on riittävän monta kertaa derivoituva. [*Apu.* $l'(\hat{\theta}; \boldsymbol{y}) = 0$ ja $E[l'(\theta; \boldsymbol{Y})] = 0$.]

3. (Monisteen teht. 3.3.) Oletetaan, että havainnot $Y_1, \dots, Y_n$ ovat riippumattomia ja noudattavat jakaumaa, jolla on odotusarvo μ ja varianssi σ^2. Tarkastellaan parametrin σ^2 estimointia muotoa cV olevilla estimaattoreilla, kun $c > 0$ on vakio ja

$$V = \sum_{i=1}^{n} (Y_i - \overline{Y})^2.$$

Näytä, että cV on harhaton jos ja vain jos $c = 1/(n-1)$. [*Ehdotus.* Käytä harjoituksen 2 tehtävän 2 hajotelmaa valinnalla $a = \mu$.]

4. (Monisteen teht. 3.4.) Havainnoista $Y_1, \dots, Y_n$ oletetaan samoin kuin edellisessä tehtävässä. Keksi jokin harhaton estimaattori odotusarvon neliölle μ^2.

5. Tarkastellaan *Poisson-regressiomallia*: $Y_1, \dots, Y_n \perp\!\!\!\perp$ ja $Y_i \sim P(\beta x_i)$, jossa $\beta > 0$ on positiivinen parametri ja $x_1, \dots, x_n > 0$ ovat tunnettuja lukuja (selittävän muuttujan arvoja). Muodosta log-uskottavuusfunktio ja johda β:n suurimman uskottavuuden estimaattorille lauseke

$$\hat{\beta} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i}.$$

Osoita, että $\hat{\beta}$ on harhaton.

Tässä Y_i voisi olla esimerkiksi liikenneonnettomuuteen vuoden aikana joutuneiden ihmisten lukumäärä (suomalaisessa) kunnassa, jonka väkiluku on x_i. Mikä olisi parametrin β tulkinta tällöin?

Ehdotus tehtävään 1:

Sijoita log-uskottavuusfunktion lausekkeeseen side-ehto $\theta_p = 1 - (\theta_1 + \cdots + \theta_{p-1})$, laske osittaisderivaatat muuttujien $\theta_1, \ldots, \theta_{p-1}$ suhteen ja sijoita θ_p takaisin saatuihin lausekkeisiin. Uskottavuusyhtälöt $\partial l / \partial \theta_j = 0$ $(j = 1, \ldots, p-1)$ voidaan kirjoittaa yhtäpitävästi muotoon $k_1/\theta_1 = k_2/\theta_2 = \cdots = k_p/\theta_p$.