Tilastollisen päättelyn jatkokurssi
7. harjoitus (joulukuu 2014)

Lisätehtäviä eri aihepiireistä. Niitä ei käsitellä harjoituksissa eikä niistä saa lisäpisteitä, mutta ratkaisut ainakin osaan tehtävistä tulevat verkkoon.

1. Olkoon $\hat{\theta}_n$ asymptoottisesti normaalinen estimaattori kuten harjoituksen 1 tehtävässä 4:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathsf{N}(0, \sigma^2(\theta_0)).$$

Funktio $h(\theta) = \theta^2$ ei toteuta delta-menetelmän oletusta $h'(\theta_0) \neq 0$ pisteessä $\theta_0 = 0$. Osoita, että nyt pätee

a) $\sqrt{n}h(\hat{\theta}_n) = \sqrt{n}\hat{\theta}_n^2 \xrightarrow{p} 0$

b) $nh(\hat{\theta}_n) = n\hat{\theta}_n^2 \xrightarrow{d} \sigma^2(0)\chi_1^2$.

Opetus. Oletus $h'(\theta_0) \neq 0$ on tärkeä delta-menetelmän mielekkyyden kannalta. Tässä tapauksessa muuttujan $h(\hat{\theta}_n)$ asymptoottisen jakauman muoto saatiin kuitenkin näkyviin valitsemalla "lihotuskertoimeksi" $\sqrt{n}$:n sijasta n.

2. Olkoon (X_i, Y_i), $i = 1, \ldots, n$, riippumaton otos kaksiulotteisesta multinormaalijakaumasta, jonka korrelaatiokerroin on $\rho = \mathsf{Cor}(X_1, Y_1) \in \,]-1, 1[$. Voidaan osoittaa, että otoskorrelaatiokertoimelle (eli ρ:n su-estimaattorille)

$$R_n = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sqrt{\sum_{i=1}^n (X_i - \overline{X}_n)^2}\sqrt{\sum_{i=1}^n (Y_i - \overline{Y}_n)^2}}$$

pätee $\sqrt{n}(R_n - \rho) \xrightarrow{d} \mathsf{N}(0, (1-\rho^2)^2)$. Tarkastellaan *Fisherin z-muunnosta*

$$Z_n = \frac{1}{2}\log\frac{1+R_n}{1-R_n}, \qquad \zeta = \frac{1}{2}\log\frac{1+\rho}{1-\rho}.$$

a) Osoita delta-menetelmän avulla, että $\sqrt{n}(Z_n - \zeta) \xrightarrow{d} \mathsf{N}(0, 1)$.

b) Johda edellisen perusteella muuttujaan Z_n perustuva (approksimatiivinen) testi hypoteesille $\mathsf{H}_0\colon \rho = 0$ (kun vastahypoteesi on kaksisuuntainen $\rho \neq 0$).

Huom. Z_n:llä on se hyvä piirre, että sen asymptoottisen jakauman varianssi ei riipu parametreista eikä sitä näin ollen tarvitse estimoida. Fisherin z-muunnos toimii siis tässä *varianssin stabiloivana muunnoksena.* Sen normaaliapproksimaatio toimii testejä ja luottamusvälejä muodostettaessa paremmin kuin R_n:n.

3. Tarkastellaan muistiinpanojen sivuilla 21–22 esitetyn lineaarisen mallin epälineaarista yleistystä, jossa yhtälön (2.16) sijasta pätee

$$Y_i = g(Z_i; \beta) + \varepsilon_i, \qquad i = 1, \ldots, n.$$

Tässä Z_i on k-ulotteinen vektorin $(\boldsymbol{W}_{i-1}, X_i)$ osavektori, β on p-ulotteinen parametri ja $g(z; \beta)$ on tunnettua muotoa oleva (kyllin siisti) funktio. Muilta osin oletetaan, että ehdollisen mallin yhteydessä esitetyt yleiset oletukset ovat voimassa ja virhetermit ε_i ovat kuten sivulla 22.

a) Johda parametrin $\theta = (\beta, \sigma^2)$ ehdollinen uskottavuusfunktio $L^{(c)}(\theta; \boldsymbol{w})$ ja log-uskottavuusfunktio $l^{(c)}(\theta; \boldsymbol{w})$.

KÄÄNNÄ!

b) Osoita, että parametrin β su-estimaatti $\hat{\beta}$ voidaan määrittää minimoimalla jäännösneliösummafunktio

$$S(\beta) = \sum_{i=1}^{n} \big(y_i - g(z_i;\beta)\big)^2$$

ja että $\hat{\sigma}^2$ saadaan kaavasta $\hat{\sigma}^2 = S(\hat{\beta})/n$ (olettaen, että $\hat{\beta}$ on olemassa).

4. Jatkoa harjoituksen 5 tehtävään 1. Oletetaan aikaisempien oletusten lisäksi, että $\hat{\beta}$ on asymptoottisesti normaalinen (ks. muistiinpanojen sivu 31):

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathsf{N}_p(0, \sigma^2 Q^{-1}).$$

Asetetaan hypoteesi $\mathsf{H}_0\colon A\beta = c$, jossa $c \in \mathbb{R}^q$ ja A on $q \times p$ -matriisi, jonka aste on q.

a) Perustele, että $\sqrt{n}(A\hat{\beta} - c) \xrightarrow{d} \mathsf{N}_q(0, \sigma^2 AQ^{-1}A')$, kun H_0 pätee.

b) Tarkastellaan lineaarisen mallin kurssilla esitettyä F-testisuuretta

$$F = (A\hat{\beta} - c)' \left[A \Big(\sum_{i=1}^{n} Z_i Z_i' \Big)^{-1} A' \right]^{-1} (A\hat{\beta} - c) \Big/ qS^2,$$

jossa $S^2 = n\hat{\sigma}^2/(n-p)$. Osoita, että $qF \xrightarrow{d} \chi^2_q$, kun H_0 pätee. (Vrt. Saikkonen: *Lineaarinen malli* (kevät 2007), s. 18.)

5. Olkoot $Y_1, \ldots, Y_n \perp\!\!\!\perp$ ja $Y_i \sim \mathsf{N}(\mu, \sigma_i^2(\delta))$, jossa varianssi $\sigma_i^2(\delta) = e^{\delta v_i}$ riippuu ei-satunnaisesta selittäjästä v_i ja tuntemattomasta parametrista δ. Johda tilastollisen mallin lauseke (ytf), log-uskottavuusfunktio, pistemäärä ja havaittu informaatio, kun parametrina on $\theta = (\mu, \delta) \in \mathbb{R}^2$. Totea, että μ ja δ ovat ortogonaaliset.

6. Jatkoa edelliseen tehtävään. Muodosta Raon pistemäärätestisuure hypoteesille $\mathsf{H}_0\colon \delta = 0$ (ts. kaikkien havaintojen varianssi on 1). Voit käyttää testisuureen versiota ortogonaalisille parametreille (ks. sivu 41).

7. Jatkoa kahteen edelliseen tehtävään. Muodosta Waldin testisuure hypoteesille $\mathsf{H}_0\colon \delta = 0$. Voit käyttää sen versiota ortogonaalisille parametreille (ks. sivu 38).

Vihjeitä:

1. Muista, että $Z^2 \sim \chi^2_1$, kun $Z \sim \mathsf{N}(0,1)$.

4. Tarvinnet ainakin lausetta 1.3 ja seurausta 1.2 aikaisemman tehtävän lisäksi. Muista myös lineaarisen mallin kurssilta neliömuoto $(Z-\mu)'\Sigma^{-1}(Z-\mu)$, kun $Z \sim \mathsf{N}(\mu, \Sigma)$. Kurssin kotisivulla on linkki lineaarisen mallin luentomuistiinpanoihin.