Tilastollisen päättelyn jatkokurssi
7. harjoitus (11. 12. 2012)

Tehtäviä eri aihepiireistä. Niitä ei käsitellä harjoituksissa eikä niistä saa lisäpisteitä, mutta ratkaisut tulevat verkkoon lähipäivinä.

1. Olkoon (X_i, Y_i), $i = 1, \dots, n$, 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. (Vrt. myös harj. 2, teht. 3.)

2. Olkoot X ja Z k-ulotteisia satunnaisvektoreita siten, että $X \perp\!\!\!\perp Z$, X on jatkuvasti jakautunut ja $Z \sim \mathsf{N}_k(\mu, \Sigma)$. Osoita, että satunnaisvektorin $Y = X + Z$ ehdollinen jakauma ehdolla $X = x$ on $\mathsf{N}_k(x + \mu, \Sigma)$ eli symbolisesti

$$Y \mid (X = x) \sim \mathsf{N}_k(x + \mu, \Sigma).$$

Huomaa, että kyseessä on muuttujan $x + Z$ jakauma.

Tätä (kenties intuitiivisesti selvää) tulosta käytettiin mm. sivulla 17 autoregressiivisen mallin yhteydessä ja sivulla 21 pääteltäessä yhtälöstä (2.16) tulos (2.15). Mieti miten.

Ohje. Parin (X, Z) ytf on helppo. Johda muunnoskaavan avulla parin (X, Y) ytf.

3. Jatkoa harjoituksen 4 tehtäviin 4 ja 5. Johda parametrin $\theta = (\beta, \sigma^2)$ Fisherin informaatio olettaen tarvittavien momenttien äärellisyys. Totea, että β ja σ^2 ovat ortogonaaliset.

4. Olkoot $Y_1, \dots, Y_n \perp\!\!\!\perp$ 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 $\theta = (\mu, \delta) \in \mathbb{R}^2$.

5. Jatkoa edelliseen tehtävään.

a) Totea, että μ ja δ ovat ortogonaaliset.

b) Johda Raon pistemäärätesti hypoteesille $\mathsf{H}_0\colon \delta = 0$ (ts. kaikkien havaintojen varianssi on 1). Voit käyttää testisuureen versiota ortogonaalisille parametreille (ks. sivun 41 alussa).