Tilastollisen päättelyn jatkokurssi, sl 2011 HT 2, viikko 38

1. Oletetaan, että parametrivektorille θ_0 $(d \times 1)$ on käytettävissä asymptoottisesti normaalinen estimaattori $\hat{\theta}_n$ eli

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathsf{N}_d\left(0, \Sigma\left(\theta_0\right)\right),$$

jossa kovarianssimatriisi on positiivisesti definiitti (ja siten epäsingulaarinen eli kääntyvä). Oletetaan lisäksi, että funktio $\theta \longmapsto \Sigma(\theta)$ on jatkuva pisteessä θ_0.
(i) Osoita yksityiskohtaisesti perustellen, että

$$n(\hat{\theta}_n - \theta_0)'\Sigma(\hat{\theta}_n)^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \chi^2_d.$$

(ii) Johda edellisen kohdan tulokseen perustuva testi hypoteesille $\theta_0 = 0$.

Vihje: Kohdassa (i) monisteen Lause 1.1 ja Seuraus 1.2 (tai Lauseet 1.3 ja 1.4) sekä lineaarisen mallin kurssilla esitetty tulos $Z'\Sigma^{-1}Z \sim \chi^2_k$, kun $Z \sim \mathsf{N}_k(0, \Sigma)$ ja Σ $(k \times k)$ on positiivisesti definiitti. Kohdassa (ii) voit tarvittaessa katsoa Tilastollisen päättelyn kurssilla esitettyä Waldin testiä vektoriparametrin tapauksessa.

2. Olkoon $(Y_1, X_1), ..., (Y_n, X_n)$ riippumaton otos kaksiulotteisesta normaalijakaumasta ja $\rho = \mathsf{Cor}(Y_1, X_1)$ $(|\rho| < 1)$ havaintojen välinen teoreettinen korrelaatiokerroin. Tiedetään, että otoskorrelaatiokertoimelle (eli ρ:n suurimman uskottavuuden estimaattorille)

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

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

$$Z_n = \frac{1}{2}\log\left(\frac{1 + R_n}{1 - R_n}\right) \quad \text{ja} \quad \zeta = \frac{1}{2}\log\left(\frac{1 + \rho}{1 - \rho}\right).$$

(i) Osoita deltamenetelmää soveltaen, että $\sqrt{n}\left(Z_n - \zeta\right) \xrightarrow{d} \mathsf{N}(0, 1)$ (alaindeksi 0 on jätetty yksinkertaisuuden vuoksi pois ρ:sta ja ζ:sta).
(ii) Johda edellisen kohdan perusteella (eli muuttujaan Z_n perustuva) approksimatiivinen testi hypoteesille $\rho = 0$ vaihtoehtoa $\rho \neq 0$ vastaan.

Huom.: Koska Z_n:n asymptoottisen jakauman varianssia ei tarvitse estimoida, toimii sen normaalisuusapproksimaatio testejä ja luottamusvälejä muodostettaessa paremmin kuin R_n:n.

3. Olkoon $Y_1, ..., Y_n$ riippumaton otos jakaumasta, jolla on äärellinen neljäs momentti. Merkitään $\mathsf{E}(Y_1) = \mu$, $\mathsf{Var}(Y_1) = \sigma^2$ ja $\mu_4 = \mathsf{E}[(Y_1 - \mu)^4]$ sekä

$$\tilde{S}^2_n = \frac{1}{n}\sum_{i=1}^n (Y_i - \mu)^2 \quad \text{ja} \quad \hat{S}^2_n = \frac{1}{n}\sum_{i=1}^n \left(Y_i - \bar{Y}_n\right)^2.$$

(i) Osoita, että $\tilde{S}_n^2 \xrightarrow{p} \sigma^2$ ja $\sqrt{n}(\tilde{S}_n^2 - \sigma^2) \xrightarrow{d} \mathsf{N}\,(0, \upsilon)$, jossa $\upsilon = \mathsf{Var}[(Y_1 - \mu)^2] = \mu_4 - \sigma^4$.
(ii) Osoita, että $\sqrt{n}(\tilde{S}_n^2 - \hat{S}_n^2) \xrightarrow{p} 0$ ja totea tämän ja edellisen tehtävän avulla, että $\hat{S}_n^2 \xrightarrow{p} \sigma^2$ ja $\sqrt{n}(\hat{S}_n^2 - \sigma^2) \xrightarrow{d} \mathsf{N}\,(0, \upsilon)$.

Vihje: Kirjoita kohdassa (ii) $\left(Y_i - \bar{Y}_n\right)^2 = \left(Y_i - \mu + \mu - \bar{Y}_n\right)^2$, korota neliöön, summaa ja käytä "sopivia" monisteen lauseita (huomaa myös monisteen s. 5 loppupuolella mainittu tulos, jonka mukaan jakaumakonvergenssista kohti vakiota seuraa stokastinen konvergenssi).

4. Olkoon $\varepsilon_1, \varepsilon_2, ...$ jono $\perp\!\!\!\perp$ $\mathsf{N}\,(0,1)$–jakautuneita sm:ia ja $Y_i = \sum_{j=1}^{i} \varepsilon_j$, $i = 1, 2, ...$

(i) Osoita, että $\mathsf{Cov}\,(Y_i, Y_{i+h}) = \mathsf{E}\,(Y_i Y_{i+h}) = i$ kaikilla $h \geq 0$ ja tätä käyttäen edelleen, että $\mathsf{Cor}\,(Y_i, Y_{i+h}) = 1/\sqrt{1 + h/i}$ $(h \geq 0,\, i \geq 1)$. Miten käy korrelaatiokertoimelle $\mathsf{Cor}\,(Y_i, Y_{i+h})$, kun $i \to \infty$ ja h on kiinteä?

(ii) Totea identiteetti

$$X_n := \frac{1}{n} \sum_{i=1}^{n} Y_{i-1}\varepsilon_i = \frac{1}{2n} \left(Y_n^2 - \sum_{i=1}^{n} \varepsilon_i^2 \right),$$

jossa $Y_0 = 0$ ja merkintä $:=$ määrittelee sm:n X_n.

Vihje: Yksi mahdollisuus kohdassa (ii) on käyttää Y_i:n määritelmästä seuraavaa identiteettiä $\varepsilon_i = Y_i - Y_{i-1}$ ja johtaa summalle $\sum_{i=1}^{n} \varepsilon_i^2$ lauseke, johon saadaan pienen muokkauksen jälkeen summa $\sum_{i=1}^{n} Y_{i-1}\varepsilon_i$. Kysytty identiteetti saadaan ratkaisemalla näin syntyvä yhtälö.

5. (Jatkoa edelliselle) Totea, että $\mathsf{E}\,(Y_{i-1}\varepsilon_i) = 0$ kaikilla $i \geq 1$ ja että $X_n \xrightarrow{d} \frac{1}{2}\left(\chi_1^2 - 1\right)$. Päättele tästä edelleen, että suurten lukujen laki (SLL) ja keskeinen raja-arvolause (KRL) eivät päde muuttujien $Y_{i-1}\varepsilon_i$, $i = 1, ..., n$ keskiarvolle. Viimeisessä kohdassa ei välttämättä vaadita täsmällistä matemaattista perustelua.

Huom.: Tämä ja edellinen tehtävä osoittavat, että tavanomainen SLL ja KRL eivät päde, jos havaintojen riippuvuus on "liiallista" (muuttujat $Y_{i-1}\varepsilon_i$, $i = 1, ..., n$, ovat vahvasti riippuvia, vaikkakin korreloimattomia). Huomaa myös, että Y_i saadaan ratkaisuna autoregressiivisestä mallista $Y_i = \phi Y_{i-1} + \varepsilon_i$ $(i = 1, ..., n)$, kun $\phi = 1$ ja $Y_0 = 0$.