

Tilastollisen päättelyn jatkokurssi 3. harjoitus (20. 11. 2012)

1. Tarkastellaan muistiinpanojen sivujen 14–15 mallin yleistystä

$$Y_j = Z_j \beta_j + \varepsilon_j, \quad j = 1, \dots, N,$$

jossa Z_j on kiinteä (selittävistä muuttujista koostuva) $n_j \times p$ -matriisi ja $\varepsilon_1, \dots, \varepsilon_N \perp\!\!\!\perp$ sekä $\varepsilon_j \sim \mathbf{N}_{n_j}(0, \sigma^2 I_{n_j})$. Oletetaan, että tässä kerroinvektorit β_j ($p \times 1$) ovat satunnaisia ja toteuttavat

$$\beta_j \sim \mathbf{N}_p(\beta, \Omega), \quad \beta_1, \dots, \beta_N \perp\!\!\!\perp, \quad (\beta_1, \dots, \beta_N) \perp\!\!\!\perp (\varepsilon_1, \dots, \varepsilon_N).$$

Johda tilastollisen mallin lauseke eli koko satunnaisvektorin $\mathbf{Y} = (Y_1, \dots, Y_N)$ yhteistiheysfunktio.

2. Tarkastellaan muistiinpanojen sivuilla 13 ja 17 käsiteltyä autoregressiivistä mallia. Johda tarkasti perustellen tämän mallin yhteistiheysfunktion lauseke

$$f_{\mathbf{Y}}(\mathbf{y}; \theta) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \phi y_{i-1})^2\right\}$$

jossa $\mathbf{y} = (y_1, \dots, y_n)$ ja parametri on $\theta = (\phi, \sigma^2)$.

Ehdotus. Merkitään $\mathbf{Y}_n = (Y_1, \dots, Y_n)$. Etene induktiolla n :n suhteen. Jos $f_{\mathbf{Y}_{n-1}}$ on jo löydetty, esitä $\mathbf{Y}_n = (\mathbf{Y}_{n-1}, Y_n)$ lineaarisena muunnoksena sv:sta $(\mathbf{Y}_{n-1}, \varepsilon_n)$, jossa $\mathbf{Y}_{n-1} \perp\!\!\!\perp \varepsilon_n$ (tämän muunnoksen Jacobin determinantti on 1). Palauta mieleen todennäköisyyslaskennasta ytf:n muunnoskaava.

3. Olkoon X_1, \dots, X_n riippumaton otos jakaumasta, jolla on äärellinen neljäs momentti. Merkitään $\mu = \mathbf{E}(X_1)$, $\sigma^2 = \mathbf{Var}(X_1)$ ja $\mu_4 = \mathbf{E}[(X_1 - \mu)^4]$ (neljäs keskusmomentti) sekä

$$\tilde{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2, \quad \hat{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

jossa $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$.

- a) Osoita, että $\tilde{S}_n^2 \xrightarrow{p} \sigma^2$ ja $\sqrt{n}(\tilde{S}_n^2 - \sigma^2) \xrightarrow{d} \mathbf{N}(0, \nu)$, jossa $\nu = \mathbf{Var}[(X_1 - \mu)^2] = \mu_4 - \sigma^4$.
 b) Osoita, että $\sqrt{n}(\tilde{S}_n^2 - \hat{S}_n^2) \xrightarrow{p} 0$, ja päättele, että a-kohdan tulokset pätevät myös \hat{S}_n^2 :lle.
 c) Totea, että a-kohdan tulokset pätevät niin ikään S_n^2 :lle.

Vihje. Käytä b-kohdassa identiteettiä $\sum_i (X_i - \mu)^2 = \sum_i (X_i - \bar{X}_n)^2 + n(\bar{X}_n - \mu)^2$, joka on mainittu myös aineopintojen päättelyn kurssilla. Muista stokastisen suppenemisen ja jakaumasuppenemisen väliset implikaatiot.

4. Tarkastellaan mallia, jossa $Y_1, \dots, Y_n \perp\!\!\!\perp$ ja $Y_i \sim \text{Exp}(\alpha + \beta x_i)$, kun x_1, \dots, x_n ovat kiinteitä tunnettuja lukuja (selittävän muuttujan arvoja). Johda tilastollisen mallin lauseke $f_{\mathbf{Y}}(\mathbf{y}; \alpha, \beta)$, vastaava log-uskottavuusfunktio, pistemääräfunktio ja havaittu informaatiomatriisi. (Palauta mieleen näiden käsitteiden määritelmät aineopintojen päättelyn kurssilta; ks. myös muistiinpanojen s. 22.)

Muista. $\text{Exp}(\lambda)$ tarkoittaa eksponenttijakaumaa, jonka tiheysfunktio on $\lambda e^{-\lambda y}$, kun $y > 0$.