Tilastollisen päättelyn jatkokurssi
3. harjoitus (19. 11. 2013)

1. Tarkastellaan muistiinpanojen sivujen 14–16 mallin yleistystä
$$Y_j = Z_j\beta_j + \varepsilon_j, \qquad j = 1, \dots, N,$$
jossa Z_j on kiinteä (selittävistä muuttujista koostuva) $n_j \times p$ -matriisi ja $\varepsilon_1, \dots, \varepsilon_N \perp\!\!\!\perp$ sekä $\varepsilon_j \sim \mathsf{N}_{n_j}(0, \sigma^2 I_{n_j})$. Oletetaan, että tässä kerroinvektorit β_j $(p \times 1)$ ovat satunnaisia ja toteuttavat
$$\beta_j \sim \mathsf{N}_p(\beta, \Omega), \qquad \beta_1, \dots, \beta_N \perp\!\!\!\perp, \qquad (\beta_1, \dots, \beta_N) \perp\!\!\!\perp (\varepsilon_1, \dots, \varepsilon_N).$$
Johda tilastollisen mallin lauseke eli koko satunnaisvektorin $\boldsymbol{Y} = (Y_1, \dots, Y_N)$ yhteistiheysfunktio.

2. Tarkastellaan muistiinpanojen sivuilla 13 ja 17 käsiteltyä autoregressiivistä mallia. Johda tarkasti perustellen tämän mallin yhteistiheysfunktion lauseke
$$f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \phi y_{i-1})^2\right\}$$
jossa $\boldsymbol{y} = (y_1, \dots, y_n)$ ja parametri on $\theta = (\phi, \sigma^2)$.

   Ehdotus. Merkitään $\boldsymbol{Y}_n = (Y_1, \dots, Y_n)$. Etene induktiolla n:n suhteen. Jos $f_{\boldsymbol{Y}_{n-1}}$ on jo löydetty, esitä $\boldsymbol{Y}_n = (\boldsymbol{Y}_{n-1}, Y_n)$ lineaarisena muunnoksena sv:sta $(\boldsymbol{Y}_{n-1}, \varepsilon_n)$, jossa $\boldsymbol{Y}_{n-1} \perp\!\!\!\perp \varepsilon_n$ (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 $\mu = \mathsf{E}(X_1)$, $\sigma^2 = \mathsf{Var}(X_1)$ ja $\mu_4 = \mathsf{E}[(X_1 - \mu)^4]$ (neljäs keskusmomentti) sekä
$$\tilde{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2, \quad \hat{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2, \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,$$
jossa $\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i$.

   a) Osoita, että $\tilde{S}_n^2 \xrightarrow{p} \sigma^2$ ja $\sqrt{n}(\tilde{S}_n^2 - \sigma^2) \xrightarrow{d} \mathsf{N}(0, \nu)$, jossa $\nu = \mathsf{Var}[(X_1 - \mu)^2] = \mu_4 - \sigma^4$.

   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 $\sum_i (X_i - \mu)^2 = \sum_i (X_i - \overline{X}_n)^2 + n(\overline{X}_n - \mu)^2$, 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 $Y_i \sim P(\mu_i(\alpha, \beta))$, kun $\mu_i(\alpha, \beta) = \exp(\alpha + \beta x_i)$ ja $x_1, \dots, x_n$ ovat kiinteitä tunnettuja lukuja (selittävän muuttujan arvoja). Johda tilastollisen mallin lauseke $f_{\boldsymbol{Y}}(\boldsymbol{y}; \alpha, \beta)$, vastaava log-uskottavuusfunktio, pistemääräfunktio ja havaittu informaatiomatriisi. (Palauta mieleen näiden käsitteiden määritelmät aineopintojen päättelyn kurssilta; ks. myös muistiinpanojen s. 22 ja 23.)

   Muista. $P(\mu)$ tarkoittaa Poisson-jakaumaa, jonka odotusarvo on μ.