Tilastollisen päättelyn jatkokurssi – yleistentti 24. 1. 2013

1. Olkoot $Y_1, \dots, Y_n \sim \mathsf{N}(0, \sigma^2) \perp\!\!\!\perp$, jossa $\sigma^2 > 0$. Tarkastellaan suurimman uskottavuuden estimaattoria $\hat{\sigma}_n^2 = \sum_{i=1}^n Y_i^2/n$.

   a) Osoita, että $\hat{\sigma}_n^2$ on harhaton, ja laske sen varianssi.
   b) Mikä on $\hat{\sigma}_n^2$:n asymptoottinen jakauma eli mihin muuttuja $\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2)$ suppenee jakaumaltaan?
   c) Mikä on muunnoksen $\log(\hat{\sigma}_n^2)$ asymptoottinen jakauma eli mihin $\sqrt{n}[\log(\hat{\sigma}_n^2) - \log(\sigma^2)]$ suppenee jakaumaltaan?

   Perustele vastauksesi.

2. a) Kuvaile yksityiskohtaisesti tilastollisen mallin yleinen tuloesitys, kun aineistoa vastaava satunnaisvektori on $\boldsymbol{Y}_n = (Y_1, \dots, Y_n)$.
   b) Tarkastellaan autoregressiivistä aikasarjamallia
   $$Y_i = \phi Y_{i-1} + \varepsilon_i, \qquad i = 1, \dots, n,$$
   jossa $Y_0 = y_0$ on tunnettu reaalinen vakio, $\varepsilon_1, \dots, \varepsilon_n \sim \mathsf{N}(0, \sigma^2) \perp\!\!\!\perp$ ja $\phi \in \mathbb{R}$ sekä $\sigma^2 > 0$. Johda tuloesityksen avulla tämän tilastollisen mallin lauseke.

3. Olkoon $f_{\boldsymbol{Y}_n}(\boldsymbol{y}_n; \theta)$ jatkuva tilastollinen malli, jossa parametri $\theta \in \Theta$ on yksiulotteinen ja merkitään $\boldsymbol{Y}_n = (Y_1, \dots, Y_n)$ sekä $\boldsymbol{y}_n = (y_1, \dots, y_n)$.

   a) Miten määritellään mallin $f_{\boldsymbol{Y}_n}(\boldsymbol{y}_n; \theta)$ säännöllisyys?
   b) Jos $f_{\boldsymbol{Y}_n}(\boldsymbol{y}_n; \theta)$ on säännöllinen, sen pistemäärällä on ns. martingaaliominaisuus. Mitä tämä tarkoittaa? Esitä eksplisiittisesti tähän liittyvät ehdot. (Todistusta ei tarvita.)
   c) Selosta lyhyesti, miksi pistemäärän martingaaliominaisuus on tärkeä uskottavuuspäättelyn asymptotiikassa.

4. Tarkastellaan tilastollista mallia $f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta)$, jonka parametriavaruus on $\Theta \subset \mathbb{R}^d$ ja joka toteuttaa tavanomaiset pistemäärän ja su-estimaattorin asymptotiikkaan liittyvät oletukset. Asetetaan hypoteesi $\mathsf{H}_0\colon A\theta_0 = c$, jossa θ_0 viittaa parametrin θ todelliseen arvoon, A on tunnettu $q \times d$ -matriisi astetta q ja $c \in \mathbb{R}^q$ on tunnettu vektori.

   Johda huolellisesti perustellen Waldin testi hypoteesille H_0 vaihtoehtoa $A\theta_0 \neq c$ vastaan. Vastauksesta tulee käydä ilmi testisuureen lauseke ja sen asymptoottinen jakauma H_0:n pätiessä sekä miten siihen päädytään ja lisäksi testisuureen arvojen tulkinta (ts. miten p-arvo lasketaan).

Huom: Tehtävissä 2 ja 3 pyri erottelemaan lihavoidut symbolit $\boldsymbol{y}$ ja $\boldsymbol{Y}$ lihavoimattomista symboleista y ja Y esimerkiksi käyttämällä alleviivausta lihavoinnin merkkinä.

Muistin tueksi:
Jos $X \sim \mathsf{N}(\mu, \sigma^2)$, niin X:n tiheysfunktio on $f(x) = (2\pi\sigma^2)^{-1/2} \exp\{-(x-\mu)^2/2\sigma^2\}$ ja $\mathsf{E}[(X-\mu)^4] = 3\sigma^4$.
Jos $X \sim \mathsf{N}_k(\mu, \Sigma)$ ja $Y = BX + \nu$, jossa B on ei-satunnainen $q \times k$ -matriisi ja ν ei-satunnainen q-ulotteinen vektori, niin $Y \sim \mathsf{N}_q(B\mu + \nu, B\Sigma B')$. Jos lisäksi Σ on positiivisesti definiitti, niin $(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi_k^2$.