

1. Palaamme tarkastelemaan parametrin  $\theta$  estimointia, kun  $f(x; \theta)$  on välin  $[0, \theta]$  tasaisen jakauman tiheysfunktio, ja  $\theta > 0$ . Edellisissä harjoituksissa johdettiin suurimman uskottavuuden estimaattori sekä sen tiheysfunktio. Tällä kertaa sinun pitää laskea

$$\text{MISE}[f(\cdot; \hat{\theta}_n)] = \mathbb{E}_\theta \int_{-\infty}^{\infty} [f(x; \hat{\theta}_n) - f(x; \theta)]^2 dx$$

(Oletetaan, että  $n \geq 2$ . Vastaukseksi pitäisi saada  $\frac{1}{\theta} \frac{1}{n-1}$ .)

2. Tässä tehtävässä tarkastelemme parametrin estimoinnin käyttöä tilanteessa, jossa malli on väärin spesifioitu.

Sovitamme havaintoihin parametrissa mallia

$$f(x; \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}, \quad x \in \mathbb{R}, \quad \theta \in \mathbb{R},$$

(normaalijakauma  $N(\theta, 1)$ ) jolloin suurimman uskottavuuden estimaattori on  $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Tosiasiassa  $X_1, \dots, X_n$  on kuitenkin satunnaisotos jakaumasta  $N(0, \sigma^2)$ , jossa  $\sigma \neq 1$ . Käytämme tuttuun tapaan tiheysfunktioestimaattoria  $\hat{f}_n = f(\cdot; \hat{\theta}_n)$ .

Onko estimaattori missään pisteessä  $x$  tarkentuva, eli päteekö missään

$$\lim_{n \rightarrow \infty} f(x; \hat{\theta}_n) = f(x),$$

kun  $f$  on jakauman  $N(0, \sigma^2)$  tiheysfunktio? (Vihje: suurten lukujen laki.)

3. Luennoilla tarkastellaan lineaarista regressiomallia

$$Y_i = a_0 x_i + b_0 + \varepsilon_i, \quad i = 1, \dots, n,$$

jossa  $x_i$ :t ovat kiinteitä lukuja,  $\sum_{i=1}^n x_i = 0$ ,  $a_0$  ja  $b_0$  ovat tuntemattomia vakioita ja satunnaismuuttujat  $\varepsilon_i$  ovat riippumattomia ja niillä on kullakin jakauma  $N(0, \sigma^2)$ . Näillä oletuksilla luennoilla johdetaan parametrien  $a_0$  ja  $b_0$  suurimman uskottavuuden estimaattorit

$$\hat{a}_n = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}, \quad \hat{b}_n = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Tehtävänä on todistaa, että  $\text{Cov}_{\theta_0}(\hat{a}_n, \hat{b}_n) = 0$ .

4. Olkoon  $f(x; \mu, \sigma^2)$  jakauman  $N(\mu, \sigma^2)$  tiheysfunktion arvo pisteessä  $x$

(a) Johda (ainakin seuraavissa harjoituksissa hyödyllinen) kaava

$$f(x; \mu_1, \sigma_1^2) f(x; \mu_2, \sigma_2^2) = f(0; \mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2) f(x; \frac{\mu_1 \sigma_2^2 + \mu_2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}).$$

(b) Laske (kaavaan nojautuen) integraali

$$\int_{-\infty}^{\infty} f(x; \mu_1, \sigma_1^2) f(x; \mu_2, \sigma_2^2) dx.$$

5. Tarkastelemme väliä  $D = [0, 1]$ , sekä tiheysfunktiota  $f \in L^2(D)$ , jota estimoidaan sarjakehitelmän avulla, kun  $L^2(D)$ :ssä käytetään ortonormaalialueita kantaa

$$\varphi_k(x) = \begin{cases} 1, & \text{kun } k = 1 \\ \sqrt{2} \cos(2\pi\ell x), & \text{kun } k = 2\ell \text{ ja } \ell = 1, 2, \dots \\ \sqrt{2} \sin(2\pi\ell x), & \text{kun } k = 2\ell + 1 \text{ ja } \ell = 1, 2, \dots \end{cases}$$

Tällöin  $f$  voidaan esittää sarjana  $\sum_{k=1}^{\infty} a_{k0} \varphi_k$ , jossa kertoimien  $a_{k0}$  harhattomat, i.i.d. otokseen  $X_1, \dots, X_n \sim f$  perustuvat estimaattorit ovat

$$\hat{a}_{kn} = \frac{1}{n} \sum_{i=1}^n \varphi_k(X_i).$$

Eräs mieleentuleva ehdotus  $f$ :n estimaattoriksi on

$$\tilde{f}_n = \sum_{k=1}^{\infty} \hat{a}_{kn} \varphi_k.$$

Tämän tehtävän tarkoituksena on perustella, miksi kyseinen ehdotus ei ole mielekäs.

Jotta olisi  $\text{MISE}[\tilde{f}_n] < \infty$ , eli  $\mathbb{E}_{\theta_0} \int_0^1 (\sum_k \hat{a}_{kn} \varphi_k(x) - f(x))^2 dx < \infty$ , niin tällöin täytyisi välttämättä olla  $\mathbb{E}_{\theta_0} \int_0^1 (\sum_k \hat{a}_{kn} \varphi_k(x))^2 dx < \infty$ . Erityisesti tässä integraali olisi äärellinen melkein varmasti, jonka ansiosta se voitaisiin kirjoittaa (kannan  $(\varphi_k)$  ortonormaalisuutta hyväksi käyttäen) melkein varmasti muotoon

$$\int_0^1 (\sum_k \hat{a}_{kn} \varphi_k(x))^2 dx = \sum_k \hat{a}_{kn}^2.$$

Koska vielä  $\mathbb{E}_{\theta_0} \sum_k \hat{a}_{kn}^2 = \sum_k \mathbb{E}_{\theta_0} \hat{a}_{kn}^2$ , niin päädyimme ehdosta  $\text{MISE}[\tilde{f}_n] < \infty$  ehtoon

$$\sum_k \mathbb{E}_{\theta_0} \hat{a}_{kn}^2 < \infty.$$

Tämä on ekvivalenttia sen kanssa, että

$$\sum_k \text{Var}_{\theta_0}(\hat{a}_{kn}) < \infty,$$

sillä  $\mathbb{E}_{\theta_0} \hat{a}_{kn} = a_{k0}$  ja  $\sum_k a_{k0}^2 < \infty$  (koska  $f \in L^2(D)$ ).

Oletamme nyt, että  $f$  toteuttaa ehdon

$$f(x) \geq \varepsilon, \quad \forall x \in [x_0 - \delta, x_0 + \delta],$$

missä  $\varepsilon > 0$  ja  $0 < x_0 < 1$ . Tehtävänäsi on osoittaa, että tällöin

$$\sum_{k=1}^{\infty} \text{Var}_{\theta_0}(\hat{a}_{kn}) = \infty.$$

Tällöin tulee siis osoitettua, että ehdotettu  $\tilde{f}_n$  ei voi toteuttaa ehtoa  $\text{MISE}[\tilde{f}_n] < \infty$ .