

Oulun yliopisto
 Matemaattisten tieteiden laitos
 Tilastollinen hahmontunnistus
 11. harjoitus, viikko 13, 2010

1. Jatkamme tiheysfunktion histogrammiestimaattorin $\hat{f}_n$ tarkastelua viime viikon tehtävän 4 pohjalta. Osoita, että kun $x \in B_k$, niin histogrammin pisteessä x lasketulle keskimääräiselle neliölliselle virheellemä pätee

$$\mathbb{E}(\hat{f}_n(x) - f(x))^2 = \left(\frac{p_k}{h} - f(x)\right)^2 + \frac{p_k(1-p_k)}{nh^2}.$$

2. Luentojen lauseen 6.2 mukaan pätee ydintestimaattorille $\hat{f}_n(\cdot; h)$

$$\mathbb{E} \int_{-\infty}^{\infty} (\hat{f}_n(x; h) - f(x))^2 dx \approx \frac{1}{4} h^4 \sigma_K^4 \|f''\|_2^2 + \frac{1}{nh} \|K\|_2^2,$$

kun h on pieni. Merkitään tämän kaavan oikeaa puolta $\text{AMISE}(h)$:lla (=Asymptotic Mean Integrated Squared Error). Osoita, että silotusparametri $h = \bar{h}_n$, joka annetulla n minimoi $\text{AMISE}(h)$:n on

$$\bar{h}_n = \left[\frac{\|K\|_2^2}{\sigma_K^4 \|f''\|_2^2} \right]^{1/5} \cdot n^{-1/5}.$$

Osoita sitten, että tämä pienin $\text{AMISE}(h)$:n arvo on

$$\text{AMISE}(\bar{h}_n) = \frac{5}{4} (\sigma_K^4 \|f''\|_2^2)^{1/5} (\|K\|_2^2)^{4/5} \cdot n^{-4/5}.$$

3. Osoita, että kun estimoitavana on $N(0, \sigma^2)$:n tiheysfunktio f ja käytetään Gaussin ydintä, niin

$$\bar{h}_n = \left(\frac{4}{3}\right)^{\frac{1}{5}} \sigma n^{-\frac{1}{5}}.$$

4. Olkoon $\hat{f}_n(\cdot; k)$ tiheysfunktion 1-ulotteinen k -lähinaapuriestimaattori. Osoita, että

$$\int_{-\infty}^{\infty} \hat{f}_n(x; k) dx = \infty.$$

Osoita myös, että $\hat{f}_n(\cdot; k)$:n ei ole derivoituva.

5. Olkoon X satunnaisvektori, jonka kovarianssimatriisi on Σ_X , ja jolla on tiheysfunktio f_X , ja olkoon $X_1, \dots, X_n$ otos X :n jakaumasta. Olkoon A valkai-sumatriisi, ts. satunnaisvektorin $Y = AX$ kovarianssimatriisi Σ_Y on yksikkö-matriisi I_d . Määritellään $Y_i = AX_i, i = 1, \dots, n$ ja olkoon $\hat{f}_{Y,n}(\cdot; k)$ otokseen $Y_1, \dots, Y_n$ perustuva k -kähinaapuriestimaattori Y :n tiheysfunktiolle f_Y . Muodostetaan nyt $\hat{f}_{Y,n}(\cdot; k)$:n avulla estimaattori

$$\hat{f}_{X,n}(x; k) = |\det A| \hat{f}_{Y,n}(Ax; k), \quad x \in \mathbb{R}^d,$$

X :n tiheysfunktiolle f_X . Osoita, että $\hat{f}_{X,n}(\cdot; k)$ on itse asiassa otokseen $X_1, \dots, X_n$ perustuva k -kähinaapuriestimaattori, kun $\mathbb{R}^d$:n normina on painotettu euklidi-nen normi $\|\cdot\|_{\Sigma_X^{-1}}$ (vrt. harjoitus 3, tehtävä 2).