Oulun yliopisto
Matemaattisten tieteiden laitos
Funktioiden estimointi
Loppukoe 23.5.2011

1. Olkoon $\lambda > 0$ ja

$$f(x;\lambda) = \begin{cases} \frac{1}{\lambda} e^{-x/\lambda}, & x > 0, \\ 0, & \text{muulloin.} \end{cases}$$

(i) Osoita, että $f(\cdot;\lambda)$ on tiheysfunktio.

(ii) Olkoon $\lambda_0 > 0$ ja $X_1, \ldots, X_n$ i.i.d. otos tiheysfunktion $f(\cdot;\lambda_0)$ määräämästä jakaumasta. Olkoon edelleen $\hat{\lambda}_n$ parametrin λ_0 otokseen $X_1, \ldots, X_n$ perustuva suurimman uskottavuuden estimaattori. Osoita, että

$$\hat{\lambda}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

- Osoita, että estimaattori $\hat{\lambda}_n$ on harhaton ja että

$$\mathbb{E}_{\lambda_0} (\hat{\lambda}_n - \lambda_0)^2 = \frac{\lambda_0^2}{n}.$$

2. Olkoon $\varphi_k : [0,1] \to \mathbb{R}$, $k = 1, \ldots, m$, ortonormaali joukko avaruudessa $L^2([0,1])$, eli

$$\int_0^1 \varphi_k(x)\varphi_\ell(x)dx = \begin{cases} 0, & \text{kun } k \neq \ell, \\ 1, & \text{kun } k = \ell. \end{cases}$$

Olkoon $a_1, \ldots, a_m \in \mathbb{R}$ ja $f = \sum_{k=1}^{m} a_k \varphi_k$ tiheysfunktio. Kun $X_1, \ldots, X_n \sim f$ on iid-otos, estimoimme funktiota f estimaattorilla

$$\hat{f}_n = \sum_{k=1}^{m} \hat{a}_{kn} \varphi_k,$$

missä $\hat{a}_{kn} = (1/n) \sum_{i=1}^{n} \varphi_k(X_i)$. Osoita, että

$$\mathbb{E} \int_0^1 [\hat{f}_n(x) - f(x)]^2 dx = C/n$$

eräällä $C > 0$.

3. Vertaile parametrista ja parametritonta funktion estimointia. Miten nämä kaksi funktion estimointitapaa eroavat toisistaan? Mitkä ovat niiden heikkoudet ja vahvuudet?

4. Olkoon f rajoitettu tiheysfunktio, $X_1, \dots, X_n \sim f$ i.i.d. otos, $K \in L^2(\mathbb{R})$ ja

$$\hat{f}(x;h) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i), \qquad x \in \mathbb{R},$$

f:n ydinestimaattori. Kuten tavallista, $h > 0$ on silotusparametri ja $K_h(x) = (1/h)K(x/h)$. Osoita, että pisteittäiselle neliölliselle virheelle pätee hajoitelma

$$\mathbb{E}[\hat{f}(x;h) - f(x)]^2 = [(f * K_h)(x) - f(x)]^2 + \frac{1}{nh}(f * (K^2)_h)(x) - \frac{1}{n}[(f * K_h)(x)]^2.$$

5. Määrittele Nadaryan-Watsonin ydinregressioestimaattori ja johda se kahdesta eri periaatteesta lähtien:

- Regressionfunktion yleisen kaavan

$$m(x) = \mathbb{E}(Y \mid X = x) = \frac{\int y f(x,y)dy}{\int f(x,y)dy}$$

  ja tiheysfunktion f ydinestimattorin avulla.

- Vakiofunktioon perustuvan lokaalin regression avulla.