Oulun yliopisto
Matemaattisten tieteiden laitos
Funktioiden estimointi
7. harjoitus, viikko 11, 2011

1. Olkoon φ normaalijakauman $N(0,1)$ tiheysfunktio. Osoita, että

$$K(x) = \frac{3}{2}\varphi(x) + \frac{1}{2}x\varphi'(x),$$

on kertalukua neljä oleva ydin ja piirrä sen kuvaaja.

2. Oletetaan, että $K_{[k]}$ on jatkuvasti derivoituva kertalukua k oleva ydin, joka toteuttaa ehdon

$$x^{k+3}K_{[k]}(x) \to 0 \qquad \text{kun } x \to \infty \text{ ja kun } x \to -\infty.$$

Määrää vakiot α_k ja β_k siten, että kaava

$$K_{[k+2]}(x) = \alpha_k K_{[k]}(x) + \beta_k x K'_{[k]}(x),$$

määrittelee kertalukua $k+2$ olevan ytimen.

3. Olkoon $X_1, \ldots, X_n$ satunnaisotos tiheysfunktiosta f. Tarkastelemme f:n estimointia ydinestimaattorilla käyttämällä kahta toisen kertaluvun ydintä $K^{(1)}$ ja $K^{(2)}$. Jos tiedämme, että h_1 on hyvä silotusparametri ytimelle $K^{(1)}$, niin voimmeko laskea yksinkertaisella kaavalla silotusparametrin h_2, joka olisi "yhtä hyvä" ytimelle $K^{(2)}$? Eräs lähestymistapa on määrätä vakio B siten, että

$$h_n^{(2)*} = Bh_n^{(1)*},$$

missä $h_n^{(i)*}$ tarkoittaa AMISE-optimaalista silotusparametria ytimellä $K^{(i)}$. Tämän jälkeen voimme sopia, että siirrymme ytimen $K^{(1)}$ silotusparametrista h_1 ytimen $K^{(2)}$ silotusparametriin h_2 yksinkertaisesti kaavalla $h_2 = Bh_1$.

Määrää vakiolle B numeerinen arvo, kun $K^{(1)}$ on Gaussin ydin ja $K^{(2)}$ on Epanechnikovin ydin.

4. (Jatkoa edelliseen tehtävään.) Olkoon $A_i(h)$ ytimellä $K^{(i)}$ ja silotusparametrilla h saatava AMISE,

$$A_i(h) = \frac{1}{4}h^4\mu_2(K^{(i)})^2R(f'') + \frac{R(K^{(i)})}{nh}.$$

Osoita, että jos vakio B määrätään kuten edellisessä tehtävässä, niin pätee

$$\frac{A_2(Bh)}{A_1(h)} = C,$$

jossa vakio C ei riipu n:stä, h:sta eikä tiheydestä f.

Mikä numeerinen arvo tulee vakiolle C, jos $K^{(1)}$ on Gaussin ydin ja $K^{(2)}$ on Epanechnikovin ydin?

5. On tilanteita, joissa satunnaismuuttujan X tiheysfunktion f_X estimointi ydinestimaattorilla satunnaisotoksen $X_1, \ldots, X_n$ perusteella ei tuota hyvää tulosta. Tällöin sattaa löytyä aidosti monotoninen funktio t siten, että satunnaismuuttujan $Y = t(X)$ tiheysfunktion f_Y estimointi muunnetun otoksen $Y_1, \ldots, Y_n$ perusteella on helpompaa (jossa $Y_i = t(X_i)$). Tällaisessa tapauksessa voidaan menetellä seuraavasti.

- Muodostetaan ydinestimaattori $\hat{f}_Y$ otoksen $Y_1, \ldots, Y_n$ perusteella.
- Ilmaistaan $f_X(x)$ arvon $f_Y(t(x))$ avulla.
- Sijoitetaan tulokseen funktion f_Y:n tilalle sen estimaattori $\hat{f}_Y$.

(a) Kirjoita edellisen kuvauksen perusteella tällaisen muunnosydinestimaattorin (engl. *transformation kernel (density) estimator*) kaava.

Seuraavaksi kokeilemme sekä tavallista että muunnosydinestimaattoria, kun otos $X_1, \ldots, X_n$ on saatu generoimalla otos $Z_1, \ldots, Z_n$ normaalijakaumasta $N(0, 1)$ ja asettamalla $X_i = \exp(Z_i)$. (Tällaista jakaumaa kutsutaan log-normaalijakaumaksi.)

(b) Muodosta ensin ydinestimaatti alkuperäisen otoksen perusteella. Käytä otoskokoa $n = 100$ ja Gaussin ydintä sekä valitse silotusparametri h ns. normaaliskaalamenetelmällä:

$$h = (4/3)^{1/5} s n^{-1/5},$$

jossa s on otoksen $X_1, \ldots, X_n$ otoskeskihajonta. Vertaa tulosta todellisen tiheysfunktion kanssa.

(c) Muodosta sitten muunnosydinestimaatti, kun muunnoksena t käytetään logaritmifunktiota ja silotusparametri f_Y:n estimaatissa valitaan normaaliskaalamenetelmällä siten, että s:n tilalle sijoitetaan otoksen $Y_1, \ldots, Y_n$ keskihajonta. Vertaa tulosta todellisen tiheysfunktion kanssa.