

Oulun yliopisto
 Matemaattisten tieteiden laitos
 Funktioiden estimointi
 10. harjoitus, viikko 14, 2011

1. Muodostamme d -ulotteisen satunnaisvektorin X tiheysfunktion f ydinestimääntöriin $\hat{f}_n(\cdot; H)$ siten, että otos on X_1, \dots, X_n , ydin on K ja $H = h^2 A$, jossa $h > 0$ ja A on positiivisesti definiitti $d \times d$ matriisi.

Osoita, että seuraavalla tavalla saadaan sama estimaatti.

- (i) Määritellään satunnaisvektori $Y = A^{-1/2} X$ ja sen jakaumasta otos $Y_i = A^{-1/2} X_i$.
 - (ii) Ilmaistaan X :n tiheysfunktio f satunnaisvektorin Y tiheysfunktion f_Y avulla.
 - (iii) Sijoitetaan tulokseen funktion f_Y tilalle sen ydinestimääntöri \hat{f}_Y , joka perustuu otokseen Y_1, \dots, Y_n ja silotusparametrimatriisiin $h^2 I_d$.
2. Osoita seuraavat tiheysfunktion k -lähinaapuriestimaattoriin $\hat{f}_n(\cdot; k)$ ominaisuudet.

- (a) Kun $k = 1$, niin estimaatilla $\hat{f}_n(\cdot; k)$ on napa jokaisessa otospisteessä (ts. $|\hat{f}_n(x; k)| \rightarrow \infty$, kun $x \rightarrow X_i$).
- (b) Jos $k > 1$, niin $\hat{f}_n(\cdot; k)$ on jatkuva, mutta derivaatta $\hat{f}'_n(\cdot; k)$ ei ole jatkuva. (Oletetaan, että otospisteet X_1, \dots, X_n ovat kaikki erisuuria; näin tapahtuu todennäköisyydellä yksi, jos kyseessä on i.i.d. otos jostakin tiheydestä f .)
- (c) Integraali $\int \hat{f}_n(x; k) dx$ on ääretön kaikilla k .

3. Olkoon satunnaismuuttujaparilla (X, Y) (ts. vektorilla $[X \ Y]^T$) kaksiulotteinen normaalijakauma $N(\mu, \Sigma)$.

- (a) Perustele, miksi kovarianssimatriisi voidaan aina kirjoittaa muotoon

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \rho \sigma_X \sigma_Y \\ \rho \sigma_X \sigma_Y & \sigma_Y^2 \end{bmatrix},$$

jossa $-1 \leq \rho \leq 1$.

- (b) Tarkista, että Y voidaan kirjoittaa homoskedastisen mallin muotoon

$$Y = m(X) + \sigma \varepsilon,$$

jossa ε ja X ovat riippumattomia, $\mathbb{E}(\varepsilon) = 0$ ja $\text{Var}(\varepsilon) = 1$, ja jossa

$$m(x) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X),$$

kun kirjoitetaan $\mu_X = \mathbb{E}X$ ja $\mu_Y = \mathbb{E}Y$. Määrää σ suureiden ρ , σ_X ja σ_Y avulla. (Oletamme, että Σ on positiivisesti definiitti.)

4. Osoita s. 123 laskuharjoitukseksi jätetty tosiasia: kun X :n ja Y :n yhteisjakauksen tiheysfunktio estimoidaan ydinestimaattorilla $\hat{f}_n(x, y; h^2 I_2)$, jossa ytimenä on tuloydin $L(x, y) = K(x)K(y)$ ja K on kertaluvun kaksi ydin, niin lauseke

$$\frac{\int y \hat{f}_n(x, y; h^2 I_2) dy}{\int \hat{f}_n(x, y; h^2 I_2) dy}$$

on Nadarayan-Watsonin estimaattori pisteessä x .

5. Kokeile regressiofunktion estimointia `motorcycle`-aineistolla, jonka saat käyttöösi kurssin kotisivun <http://cc.oulu.fi/~11h/FE2007> kautta (tiedosto `mcyc.m`). Aineisto sisältää moottoripyörän törmäyskokeesta mitattuja testinuken pään kiihtyvyyksiä (y) ajan (x) funktiona.

- (a) Kokeile regressiofunktion estimointia Nadarayan-Watsonin estimaattorilla eri h :n arvoilla. Käytä Gaussin ydintä.
- (b) Piirrä arvot $(X_i, W_h(x - X_i))$, kun x on lähellä aineiston pienintä ajanhetkeä, ja toisaalta, kun x on aineiston keskivaiheilla jollakin hyväksi katsomallasi h :n arvolla. (Tässä $W_h(x - X_i)$ määriteltiin luennoilla siten, että Nadarayan-Watsonin estimaattori voidaan kirjoittaa arvojen Y_i painotettuna keskiarvona, $\hat{m}_n(x; h) = \sum_i W_h(x - X_i) Y_i$).