

Oulun yliopisto
 Matemaattisten tieteiden laitos
 Funktioiden estimointi
 8. harjoitus, 14. 3. 2007

1. Kun ydinestimäattorin silotusparametria valitaan ristiinvalidoinnilla, minimoidaan suuretta

$$\text{LSCV}(h) = \int [\hat{f}_n(x; h)]^2 dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{-i,n}(X_i; h).$$

Johda $\text{LSCV}(h)$:lle muoto

$$\text{LSCV}(h) = \frac{1}{2nh\sqrt{\pi}} + \frac{1}{nh\sqrt{\pi}} \sum_{i<j} \left[\frac{1}{n} \exp\left(-\frac{1}{4}\Delta_{ij}^2\right) - \frac{\sqrt{8}}{n-1} \exp\left(-\frac{1}{2}\Delta_{ij}^2\right) \right]$$

siinä tapauksessa, kun käytetään Gaussin ydintä ja kun merkitään

$$\Delta_{ij} = \frac{X_i - X_j}{h}.$$

2. Harhaisessa ristiinvalidoinnissa täytyy laskea suure $R(\hat{f}_n''(\cdot; h)) - R(K'')/(nh^5)$, jossa $\hat{f}_n(\cdot; h)$ on ydinestimäattori silotusparametrilla h . Osoita, että

$$R(\hat{f}_n''(\cdot; h)) - \frac{R(K'')}{nh^5} = \frac{2}{n^2h^5} \sum_{i<j} \tau(\Delta_{ij}),$$

jossa Δ_{ij} määritellään samoin kuin edellisessä tehtävässä ja jossa

$$\tau(\Delta) = \int K''(u)K''(u + \Delta) du.$$

(Oletamme, että $\tau : \mathbb{R} \rightarrow \mathbb{R}$ on hyvin määritelty.)

3. Kun K on Gaussin ydin, johda edellisessä tehtävässä määritellylle funktiolle τ esitys

$$\tau(\Delta) = \frac{\exp(-\frac{1}{4}\Delta^2)(\Delta^4 - 12\Delta^2 + 12)}{32\sqrt{\pi}}.$$

Johda edelleen Gaussin ytimelle kaava

$$\text{BCV}(h) = \frac{1}{2nh\sqrt{\pi}} + \frac{1}{64n^2h\sqrt{\pi}} \sum_{i<j} \exp(-\Delta_{ij}^2/4)(\Delta_{ij}^4 - 12\Delta_{ij}^2 + 12).$$

4. Tässä tehtävässä kokeilemme silotusparametrin valintaa LSCV- ja BCV-menetelmillä. Estimoitava tiheys on (Wandin ja Jonesin esimerkitiheys)

$$f(x) = \frac{3}{4}f(x; 0, 1) + \frac{1}{4}f(x; \frac{3}{2}, (\frac{1}{3})^2),$$

jossa $f(x; \mu, \sigma^2)$ on $N(\mu, \sigma^2)$ -normaalijakauman tiheysfunktio pisteessä x . (Tästä jakaumasta saa generoitua satunnaismuuttujan generoimalla todennäköisyydellä $\frac{3}{4}$ muuttujan jakaumasta $N(0, 1)$ ja muussa tapauksessa muuttujan jakaumasta $N(\frac{3}{2}, (\frac{1}{3})^2)$.)

- (a) Implementoi Gaussin ytimelle tehtävissä 1 ja 3 johdetut laskentakaavat suureille LSCV(h) ja BCV(h). Generoi kokoa $n = 100$ oleva otos tiheydestä f . Tarkastele, onko käyrillä LSCV(h) ja BCV(h), $h > 0$ yksikäsitteinen globaali minimikohta. Toista koe muutamia kertoja.
- (b) Valitse nyt silotusparametrille arvo etsimällä sekä LSCV- että BCV-funktion minimikohta. Voit esimerkiksi laskea funktioiden arvot sopivalla välillä $[\varepsilon, 2]$ asetetulla hilalla ($\varepsilon > 0$) ja valita silotusparametriksi \hat{h}_n^{LSCV} sen hilapisteen, jossa LSCV(h):lle tuli pienin arvo. Silotusparametri \hat{h}_n^{BCV} valitaan vastaavasti. Toista koe 500 kertaa. Wand ja Jones totesivat, että satunnaismuuttujalla \hat{h}_n^{LSCV} on suurempi varianssi mutta pienempi harha kuin satunnaismuuttujalla \hat{h}_n^{BCV} . Päädytkö samanlaiseen tulokseen? Kuten luennolla mainittiin, on teoreettisesti optimaalinen silotusparametrin arvo on 0.318 (vrt. luennot).

5. Toteuta luennoissa esitetty silotusparametrin valinta kaksivaiheisella DPI-menetelmällä, $\hat{h}_n^{\text{DPI},2}$ Gaussin ytimelle. Vertaa saatavaa silotusparametrien jakaumaa edellisen tehtävän tapauksessa ristiinvalidoinnilla saataviin silotusparametrien jakaumiin.

Tässä tehtävässä tarvitaan lausekkeet Gaussin ytimen K kertaluvun 4, ja 6 derivaatoille. Ne ovat

$$K^{(4)}(x) = (x^4 - 6x^2 + 3)K(x)$$

$$K^{(6)}(x) = (x^6 - 15x^4 + 45x^2 - 15)K(x).$$