Rolf Nevanlinna -instituutti
Funktioiden estimointi
Loppukoe 8.8.2002

1. Tarkastelemme parametrista perhettä

$$f(x;\theta) = \frac{1}{\theta} 1_{[0,\theta]}(x), \qquad x \in \mathbb{R}, \quad \theta > 0.$$

(Tasainen jakauma välillä $[0,\theta]$.) Olkoon $X_1, \ldots, X_n \sim f(\cdot;\theta)$ iid-otos.

- Osoita, että parametrin θ suurimman uskottavuuden estimaattori on $\hat{\theta}_n = \max(X_1, \ldots, X_n)$.
- Laske estimaattorin $\hat{\theta}_n$ tiheysfunktio (vihje: kertymäfunktio on helppo laskea).
- Johda lauseke
$$\mathbb{E}(\hat{\theta}_n - \theta)^2 = \frac{2\theta^2}{(n+1)(n+2)}.$$
- Selitä lyhyesti mikä on parametrisessa estimoinnissa esiintyvä Cramerin-Raon alaraja ja pohdi sitä estimaattorin $\hat{\theta}_n$ tapauksessa.

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. Parametrittomassa funktion estimoinnissa esiintyy usein "silotusparametri", jonka avulla säädetään harhan ja varianssin suhteellista osuutta estimointivirheessä. Valitse jokin tiheysfunktion tai regressiofunktion estimointimenetelmä ja pohdi sen kohdalla harhan ja varianssin tasapainoittamista silotusparametrin säädön avulla.

4. Kun tiheysfunktiota f estimoidaan ei-negatiivisesta ytimestä K saadulla ydinestimaattorilla, on asymptoottisesti optimaalinen keskimääräinen integroitu neliövirhe muotoa

$$\frac{5}{4} C(K) R(f'')^{1/5} n^{-4/5},$$

missä $C(K) = [\mu_2(K)^2 R(K)^4]^{1/5}$, $R(g) = \int_{-\infty}^{\infty} g(x)^2 dx$ ja $\mu_2(K) = \int_{-\infty}^{\infty} x^2 K(x) dx$. Osoita, että Epanechnikovin ydin $K^*(x) = (3/4)(1 - x^2)_+$ on optimaalinen tämän asymptoottisen virheen mielessä.

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.

Matematiikan ja tilastotieteen laitos
Funktioiden estimointi
Loppukoe 16.3.2004

1. Tarkastelemme parametrista perhettä

$$f(x;\theta) = \frac{1}{\theta} 1_{[0,\theta]}(x), \qquad x \in \mathbb{R}, \quad \theta > 0.$$

(Tasainen jakauma välillä $[0,\theta]$.) Olkoon $X_1, \ldots, X_n \sim f(\cdot;\theta)$ iid-otos.

- Osoita, että parametrin θ suurimman uskottavuuden estimaattori on $\hat{\theta}_n = \max(X_1, \ldots, X_n)$.
- Laske estimaattorin $\hat{\theta}_n$ tiheysfunktio (vihje: kertymäfunktio on helppo laskea).
- Johda lauseke
$$\mathbb{E}(\hat{\theta}_n - \theta)^2 = \frac{2\theta^2}{(n+1)(n+2)}.$$
- Selitä lyhyesti mikä on parametrisessa estimoinnissa esiintyvä Cramerin-Raon alaraja ja pohdi sitä estimaattorin $\hat{\theta}_n$ tapauksessa.

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

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

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

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

eräällä $C > 0$.

3. Parametrittomassa funktion estimoinnissa esiintyy usein "silotusparametri", jonka avulla säädetään harhan ja varianssin suhteellista osuutta estimointivirheessä. Valitse jokin tiheysfunktion tai regressiofunktion estimointimenetelmä ja pohdi sen kohdalla harhan ja varianssin tasapainoittamista silotusparametrin säädön avulla.

4. Kun tiheysfunktiota f estimoidaan ei-negatiivisesta ytimestä K saadulla ydinestimaattorilla, on asymptoottisesti optimaalinen keskimääräinen integroitu neliövirhe muotoa

$$\frac{5}{4} C(K) R(f'')^{1/5} n^{-4/5},$$

missä $C(K) = [\mu_2(K)^2 R(K)^4]^{1/5}$, $R(g) = \int_{-\infty}^{\infty} g(x)^2 dx$ ja $\mu_2(K) = \int_{-\infty}^{\infty} x^2 K(x) dx$. Osoita, että Epanechnikovin ydin $K^*(x) = (3/4)(1-x^2)_+$ on optimaalinen tämän asymptoottisen virheen mielessä.

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.