

Estymacja parametru gładkości gęstości

Karol Dzedziul
Magdalena Kucharska
Barbara Wolnik

6-10 Grudzień 2010 Wisła

- Przestrzeń Biesowa $B_{p,q}^s$, $0 < s$ jako przestrzeń interpolacyjna
- Własności przestrzeni Biesowa i wybór $B_{2,\infty}^s$, $0 < s < k$
- Estymator parametru gładkości s nieznanego gęstości
- Przykład zastosowań.

Przestrzenie interpolacyjne, konstrukcja Gagliardo, interpolacja rzeczywista

Niech V będzie przestrzenią topologiczną i wektorową oraz niech A_1 i A_2 będą dwoma przestrzeniami Banacha o normach odpowiednio $\|\cdot\|_1$, $\|\cdot\|_2$. W przestrzeniach $A_1 \cap A_2$ oraz

$$A_1 + A_2 := \{a_1 + a_2 : a_1 \in A_1, a_2 \in A_2\}$$

w naturalny sposób definiujemy normy tak aby były przestrzeniami Banacha. Przestrzenią interpolacyjną $(A_1, A_2)_{\theta, q}$ nazywamy przestrzeń Banacha

$$A_1 \cap A_2 \subset (A_1, A_2)_{\theta, q} \subset A_1 + A_2.$$

Procedura tworzenia takich przestrzeni jest oparta o K-funkcjonal Peetre, $0 < t < \infty$

$$K(t, a, A_1, A_2) := \inf_{a=a_1+a_2} (\|a_1\|_{A_1} + t\|a_2\|_{A_2}).$$

Wówczas $0 < \theta < 1$, $0 < q \leq \infty$

$$a \in (A_1, A_2)_{\theta, q} \iff \int_0^\infty \left(\frac{K(t, a)}{t^\theta} \right)^q \frac{dt}{t} < \infty.$$

Przestrzeń Biesowa jako przestrzeń interpolacyjna dla przestrzeni Sobolewa

Definicja

Przestrzenie Sobolewa $W_p^k = W_p^k(\mathbb{R})$, $1 \leq p < \infty$ i $k \in \mathbb{N}$ nazywamy domknięcie przestrzeni funkcji gładkich o zwatym nośniku $f \in C_c^\infty(\mathbb{R})$ w normie

$$\|f\|_{p,k} = \|f\|_{L^p} + \|D^k f\|_{L^p},$$

$$\|f\|_{L^p} = \left(\int_{\mathbb{R}} |f(x)|^p dx \right)^{\frac{1}{p}}.$$

Dla $p = \infty$ odpowiednio modyfikujemy.

Definicja

Przestrzeń Biesowa $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ i $s > 0$

$$B_{pq}^s = (W_p^{k_1}, W_p^{k_2})_{\theta,q},$$

$$s = (1 - \theta)k_1 + \theta k_2.$$

Przykłady

Przestrzeń Zygmunda $B_{\infty, \infty}^1$. Przestrzenie Höldera $B_{\infty, \infty}^s$ dla $0 < s < 1$,

$$(C[0, 1], C^1[0, 1])_{\theta, \infty} = B_{\infty, \infty}^{\theta}$$

Ciesielski, Kerkyacharian, Roynette 1993, Sjögren 1982. Niech X^α będzie ułamkowym procesem Wienera (ułamkowym ruchem Browna). Wówczas jeśli $0 < 1/p < \alpha/2$, $p < \infty$, $0 < \alpha < 2$, to

$$P(X^\alpha \in B_{p, \infty}^{\alpha/2}([0, 1])) = 1$$

Jeśli $p = \infty$, to dla wszystkich $0 < \beta < \alpha/2$

$$P(X^\alpha \in B_{p, \infty}^{\beta}([0, 1])) = 1$$

Powyższa charakteryzacja pozwala zdefiniować całkę stochastyczną po trajektoriach (podejście Younga)

$$\int_0^1 f(s) dX_s^\alpha.$$

Twierdzenie

Niech $1 \leq p_1, p_2 \leq \infty$, $1 \leq q_1, q_2 \leq \infty$ i $s_1, s_2 > 0$. Wówczas

$$(B_{p_1 q_1}^{s_1}, B_{p_2 q_2}^{s_2})_{\theta, q} = B_{pq}^s,$$

gdzie

$$\frac{1}{p} = \frac{1-\theta}{p_1} + \frac{\theta}{p_2}, \quad \frac{1}{q} = \frac{1-\theta}{q_1} + \frac{\theta}{q_2},$$

$$s = (1-\theta)s_1 + \theta s_2.$$

Ponadto zachodzi (ciągłe włożenie)

$$B_{pq}^s \subset B_{p q_1}^{s_1}, \tag{1}$$

gdzie $s > s_1$ lub $s = s_1$, $q_1 > q$.

Gładkość funkcji w terminach przestrzeni Biesowa

Dla prostoty ustalamy $p = 2$, tak, że zakładamy $f \in L^2(R)$. Ponieważ $B_{2,q}^s \subset B_{2,\infty}^s$ rozważamy $B_{2,\infty}^s$. Z twierdzenia o ciągłym włożeniu wynika, że

$$B_{2,\infty}^{s_1} \subset B_{2,\infty}^{s_2} \text{ for } s_1 > s_2$$

Zatem albo f należy do wszystkich przestrzeni $B_{2,\infty}^s$ albo f nie należy do żadnej $B_{2,\infty}^s$ albo **istnieje parametr s^* taki, że**

$$\text{dla wszystkich } s < s^*, \quad f \in B_{2,\infty}^s$$

oraz

$$\text{dla wszystkich } s > s^*, \quad f \notin B_{2,\infty}^s.$$

W ostatnim przypadku mówimy, że s^* jest parametrem gładkości funkcji f .

Wielopoziomowa aproksymacja

Niech $P_{(h)}$ oznacza rzut ortogonalny

$$P_{(h)} : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$$

dany wzorem

$$P_{(h)}f(x) = \int_{\mathbb{R}} K_h(x, y)f(y)dy,$$

gdzie jądro $K_h(x, y) = \frac{1}{h}K\left(\frac{x}{h}, \frac{y}{h}\right)$

$$K(x, y) = \sum_{k \in \mathbb{Z}} \mathbb{I}_{[0,1)}(y - k)\mathbb{I}_{[0,1)}(x - k)$$

($\mathbb{I}_{[a,b)}$ jest funkcją charakterystyczną odcinka $[a, b)$). Wprowadzamy następujące oznaczenia

$$P_j = P_{(2^{-j})}, \quad Q_j = P_j - P_{j-1}.$$

Przestrzeń Biesowa $B_{2,\infty}^s$ można zcharakteryzować za pomocą operatorów Q_j dla $0 < s < 1/2$, B.I. Golubov 1972, (S. Ropella 1976.)

Twierdzenie deterministyczne, jak wyznaczyć parametr gładkości

Twierdzenie

Niech $0 < s^* < 1/2$ tak, że $s < s^* \implies f \in B_{2,\infty}^s$ oraz $s > s^* \implies f \notin B_{2,\infty}^s$.
Wówczas

$$\liminf_{j \rightarrow \infty} \frac{-\log_2^+ \|Q_j f\|_2}{j} = s^*, \quad (2)$$

gdzie \log_2^+ oznacza że bierzemy tylko logarytm dla niezerowych argumentów.

Uwaga: jeśliby istniała stała N , że dla wszystkich $j > N \implies \|Q_j f\|_2 = 0$ wówczas funkcja f należałaby do wszystkich przestrzeni Biesowa.

Niech X_1, X_2, \dots będzie ciągiem zmiennych losowych niezależnych o jednakowym rozkładzie z nieznaną gęstością $f \in L^2(\mathbb{R})$. Histogram dla powyższej próby z parametrem grupowania $h > 0$ (parametr okna)

$$f_{h,n}(x) = \frac{1}{n} \sum_{j=1}^n K_h(x, X_j),$$

gdzie przypomnijmy $K_h(x, y) = \frac{1}{h} K\left(\frac{x}{h}, \frac{y}{h}\right)$,

$$K(x, y) = \sum_{k \in \mathbb{Z}} \mathbb{I}_{[0,1)}(y - k) \mathbb{I}_{[0,1)}(x - k).$$

Zauważmy, że $Ef_{h,n} = P_{(h)}f$, gdzie $P_{(h)}$ jest rzutem ortogonalnym.

Histogram odpowiadający parametrom $h = 2^{-j}$ oraz $n = 2^{2j}$ oznaczmy przez f_j , czyli

$$f_j = f_{2^{-j}, 2^{2j}}.$$

Wówczas

Twierdzenie

Niech X_1, X_2, \dots będzie ciągiem niezależnych zmiennych losowych o wspólnej gęstości $f \in L^2(\mathbb{R})$. Ponadto niech $0 < s^* < 1/2$ będzie parametrem gładkości f , czyli $s < s^* \implies f \in B_{2,\infty}^s$ oraz $s > s^* \implies f \notin B_{2,\infty}^s$. Wówczas

$$\liminf_{j \rightarrow \infty} \frac{-\log_2^+ \|f_j - f_{j-1}\|_2}{j} = s^* \quad \text{a.e.} \quad (3)$$

P. Hall, Central limit theorem for square error of multivariate nonparametric density estimators, J. Multivar. Anal. (1984) 14, pp. 1-16.

K. Dzedziul, Central limit for square error of multivariate nonparametric box spline density estimators, Appl. Math. 28,3 (2001) pp. 437-456.

Przykładowe EKG, baza PhysioBank ATM

Niech X_1, X_2, \dots będzie ciągiem niezależnych zmiennych losowych o gęstości $f \in L^2(\mathbb{R})$. Ponadto niech $0 < s^* < 1/2$ będzie parametrem gładkości, czyli dla $s < s^*$ $f \in B_{2,\infty}^s$ oraz dla $s > s^*$ $f \notin B_{2,\infty}^s$. Wówczas dla $\varepsilon > 0$

$$\begin{aligned} MISE(f, h) &= E \left[\int_{\mathbb{R}^2} [f_{h,n} - f]^2 \right] \\ &= E \left[\int_{\mathbb{R}^2} [f_{h,n} - P_h f]^2 \right] + \int_{\mathbb{R}^2} [P_h f - f]^2 \\ &\approx \frac{1}{nh} + h^{2(s-\varepsilon)} \|f\|_{B^{(s-\varepsilon)}_{2,\infty}}^2. \end{aligned}$$

Dziękuję