Estimation of the smoothness of density

Karol Dziedziul
Magdalena Kucharska
Barbara Wolnik

9-13 August 2010 MCI Gothenburg

Plan of presentation

- **General concept of Besov spaces** $B^s_{p,q}$, $0 < s$
- **The smoothness of the functions - Besov spaces** $B^s_{2,\infty}$, $0 < s < k$
- **Estimator of the smoothness (of the parameter** s**) of density**

Gagliardo interpolation spaces

Let V be a topological vector space and A_1 i A_2 two Banach spaces, with norms $\|\cdot\|_1$, $\|\cdot\|_2$. We can convert $A_1 \cap A_2$ and

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

into Banach spaces. An intermediate space $(A_1, A_2)_{\theta,q}$ between A_1 and A_2 satisfy

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

Procedure of creating such spaces is connected with Peetre's K-functional , $0 < t < \infty$

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

Then for $0 < \theta < 1$, $0 < q \leq \infty$

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

Sobolev and Besov spaces

Defintion

A Sobolev space $W_p^k = W_p^k(R)$, $1 \leq p < \infty$ *and* $k \in N$ *is a closure of the smooth functions with compact supports* $f \in C_c^\infty(R)$ *in the norm*

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

For $p = \infty$ *we have some modification.*

Defintion

A Besov space $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ *and* $s > 0$

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

where the smoothness parameter

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

Examples

A Zygmund space $B^1_{\infty,\infty}$.

The Höldera spaces $B^s_{\infty,\infty}$ for $0 < s < 1$,

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

Ciesielski, Kerkyacharian, Roynette 1993, Sjögren 1982. Let X^{α} be a fractional brownian motion. If $0 < 1/p < \alpha/2$, $p < \infty$, $0 < \alpha < 2$ then

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

If $p = \infty$, then for all $0 < \beta < \alpha/2$

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

Properties of Biesov spaces

Theorem

Let $1 \leq p_1, p_2 \leq \infty$, $1 \leq q_1, q_2 \leq \infty$ *i* $s_1, s_2 > 0$. *Then*

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

where

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

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

We have continuous embedding

$$B^s_{pq} \subset B^{s_1}_{pq_1}, \tag{1}$$

where $s > s_1$ *or* $s = s_1$, $q_1 > q$.

Smoothness of function in terms of Besov spaces

Let us consider a function $f \in L^2(R)$, $p = 2$. Since $B^s_{2,q} \subset B^s_{2,\infty}$ we consider only $B^s_{2,\infty}$. From the following continuous embedding

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

we have: f belongs to each $B^s_{2,\infty}$ or f belongs to none of $B^s_{2,\infty}$ or **there is the parameter s^* such that**

$$\text{for all} \quad s < s^*, \quad f \in B^s_{2,\infty}$$

and

$$\text{for all} \quad s > s^*, \quad f \notin B^s_{2,\infty}.$$

In the last case we say that s^* is the smoothness parameter of the function f.

Multiresolution analysis

Let us define the orthogonal projection on piecewise constant function

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

where the kernel is given by $K_h(x,y) = \frac{1}{h}K\left(\frac{x}{h}, \frac{y}{h}\right)$

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

($\mathbb{I}_{[a,b)}$ is the characteristic functions of the interval $[a,b)$). From multiresolution analysis point of view introduce

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

The Besov spaces $B^s_{2,\infty}$ are characterized by operators Q_j only for $0 < s < 1/2$, B.I. Golubov 1972, for full scale of Besov spaces in S. Ropella. 1976.

Determination of smoothness of function

Theorem

Let $0 < s^* < 1/2$ *be such that for all* $s < s^*$ $f \in B^s_{2,\infty}$ *and for* $s > s^*$ $f \notin B^s_{2,\infty}$. *Then*

$$\liminf_{j \to \infty} \frac{-\log_2^+ \|Q_j f\|_2}{j} = s^*, \tag{2}$$

where $\log_2^+$ *means that we take only nonzero arguments.*

Histogram

Let $X_1, X_2, \ldots$ be a sequence of iid rvs with unknown density $f \in L^2(R)$. The histogram is defined by

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

where recall the kernel is given by $K_h(x,y) = \frac{1}{h} K\left(\frac{x}{h}, \frac{y}{h}\right)$

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

Now $Ef_{h,n} = P_{(h)}f$, where $P_{(h)}$ is the orthogonal projection. This projection can written by the Haar basis for $h = 2^{-j}$.

Estimation of smoothness of density

Let us denote the density-histogram for parameters $h = 2^{-j}$ and $n = 2^{2j}$ by f_j, i.e.

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

Theorem

Let $X_1, X_2, \ldots$ be a sequence of iid random variables with density $f \in L^2(\mathbb{R})$. Moreover let $0 < s^* < 1/2$ be such that for all $s < s^*$ $f \in B^s_{2,\infty}$ and for all $s > s^*$ $f \notin B^s_{2,\infty}$. Then

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

Proof = references

P. Hall, *Central limit theorem for square error of multivariate nonparametric density estimators*, J. Multivar. Anal. (1984) 14, pp. 1-16.
K. Dziedziul, *Central limit for square error of multivariate nonparametric box spline density estimators*, Appl. Math. 28,3 (2001) pp. 437-456.

We apply theorem into financial and biological signals.

This research is supported in part by the grant "The exploitation of white cabbage for phytoremediation and biofumigation of soils (AGROBIOKAP)"
co-financed by European Union from European Regional Development Fund in a framework of the Innovative Economy Operational Programme 2007-2013

Thank you

FAQ

1. We think that one can use the Franklin system, wavelets or positive operator. One can use also Rosenblatt Parzen estimator since we have characterization of Besov spaces by the convolution operators.

2. By the Korovkin's theorem the rate of approximation by positive operators is not greater then two. We think that we can extend the smoothness from $0 < s^* < 1/2$ to $0 < s^* < 2$.

Bandwidth and parameter of smoothness

Let $X_1, X_2, \ldots$ be a sequence of iid random variables with density $f \in L^2(\mathbb{R})$. Moreover let $0 < s^* < 1/2$ be such that for all $s < s^*$ $f \in B^s_{2,\infty}$ and for all $s > s^*$ $f \notin B^s_{2,\infty}$. Then

$$MISE(f,h) = E\left[\int_{\mathbb{R}^2} [f_{h,n} - f]^2\right]$$

$$= E\left[\int_{\mathbb{R}^2} [f_{h,n} - P_h f]^2\right] + \int_{\mathbb{R}^2} [P_h f - f]^2$$

$$\approx \frac{1}{nh} + h^{2s} \|f\|^2_{B^s_{2,\infty}}.$$