Applied Mathematics, **2011, 2, 1258-1262**
doi:10.4236/am.2011.210175 Published Online October 2011 (http://www.SciRP.org/journal/am)

Convergence Rates of Density Estimation in Besov Spaces

Huiying Wang
Department of Applied Mathematics, Beijing University of Technology, Beijing, China
E-mail: *b200806005@emails.bjut.edu.cn*
Received July 29, 2011; *revised August* 23, 2011; *accepted August* 30, 2011

Abstract

The optimality of a density estimation on Besov spaces $B^s_{r,q}(\boldsymbol{R})$ for the L_p risk was established by Donoho, Johnstone, Kerkyacharian and Picard ("Density estimation by wavelet thresholding," The Annals of Statistics, Vol. 24, No. 2, 1996, pp. 508-539). To show the lower bound of optimal rates of convergence $R_n\left(B^s_{r,q}, p\right)$, they use Korostelev and Assouad lemmas. However, the conditions of those two lemmas are difficult to be verified. This paper aims to give another proof for that bound by using Fano's Lemma, which looks a little simpler. In addition, our method can be used in many other statistical models for lower bounds of estimations.

Keywords: Optimal Rate of Convergence, Density Estimation, Besov Spaces, Wavelets

1. Introduction

Wavelet analysis has many applications, one of which is to estimate an unknown density function based on independent and identically distributed (i.i.d.) random samples. Let $(\Omega, \aleph, P)$ be a probability measurable space and $X_1, \cdots, X_n$ be i.i.d. random variables with an unknown density function *f*. We use $E(X)$ to denote the expectation of a random variable *X*. The sequence $R_n(V,p) := \inf_{f_n} \sup_{f\in V} E\left(\|f_n - f\|_p\right)$ is called optimal rate of convergence on the functional class *V* for the L_p risk. Here, f_n is an arbitrary estimator of *f* with *n* i.i.d. random samples. Kerkyacharian and Picard [1] study $R_n(V,p)$ when *V* is a Besov space with matched case. Donoho, Johnstone, Kerkyacharian and Picard [2] consider unmatched cases. In fact, they show the optimal convergence rates for L_r Besov class $\tilde{B}^s_{r,q}$ and L_p risk

$$R_n\left(\tilde{B}^s_{r,q}, p\right) \sim \begin{cases} (\ln n/n)^{\frac{s-1/r+1/p}{2(s-1/r)+1}}, & r \le \dfrac{p}{2s+1}, \\ n^{-\frac{s}{2s+1}}, & r > \dfrac{p}{2s+1}. \end{cases} \quad (1.1)$$

To show the lower bound of (1.1), authors of [2,3] use Korostelev and Assouad lemmas. However, the conditions of those two lemmas are difficult to be verified. In this small paper, we give another proof for the lower bound of (1.1) by using Fano's lemma [4]. It should be pointed out that Fano's lemma can be used to a variety of statistical models, see [5-7].

As usual, $L_p(R)\,(p \ge 1)$ denotes the classical Lebesgue space on the real line *R*. In particular, $L_2(R)$ stands for the Hilbert space, which consists of all square integrable functions. As a subspace of $L_p(R)$, the Sobolev space with an integer exponent *k* means

$$W^k_p(R) := \left\{ f,\ f^{(m)} \in L_p(R),\ m = 0,1,\cdots,k \right\} (p \ge 1).$$

The corresponding norm

$$\|f\|_{W^k_p} := \|f\|_p + \left\|f^{(k)}\right\|_p.$$

Moreover, the Besov space $B^s_{p,q}(R)$ [3] ($1 \le p,\ q \le \infty$, $s = n + \alpha$ and $\alpha \in (0,1]$) can be defined by

$$B^s_{p,q}(R) = \left\{ f \in W^n_p(R), \left(2^{j\alpha}\omega^2_p\left(f^{(n)}, 2^{-j}\right)\right)_{j\in Z} \in l_q \right\}$$

with the associated norm

$$\|f\|_{B^s_{p,q}} := \|f\|_{W^n_p} + \left\|\left\{2^{j\alpha}\omega^2_p\left(f^{(n)}, 2^{-j}\right)\right\}\right\|_{l_q(Z)},$$

where

$$\omega^2_p(f,t) := \sup_{|h|\le t} \|f(x+2h) - 2f(x+h) + f(x)\|_p.$$

In general, it can be shown that compactly supported and *n* times differentiable functions belong to $B^s_{p,q}(R)$ for

$0 < s < n$ and $1 \le p,\ q \le \infty$.

The Besov space can be discretized by the sequence norm of wavelet coefficients. Many useful wavelets are generated by scaling functions. More precisely, if φ is a scaling function with

$$\varphi(x) = \sum_k h_k \sqrt{2}\varphi(2x-k),$$

then $\psi(x) := \sum_k (-1)^k h_{1-k} \sqrt{2}\varphi(2x-k)$ defines a wavelet [3]. Clearly, when φ is compactly supported and continuous, the corresponding wavelet ψ has the same properties. An orthonormal wavelet basis of $L_2(R)$ is generated from dilation and translation of a scaling function and its corresponding wavelet, *i.e.*

$$\left\{\begin{array}{l}\varphi_{j_0k}(x) := 2^{\frac{j_0}{2}}\varphi\left(2^{j_0}x-k\right),\\ \left.\psi_{jk}(x) := 2^{\frac{j}{2}}\psi\left(2^{j}x-k\right)\right\}_{j\ge j_0,k\in Z}\end{array}\right. .$$

Although wavelet basis are constructed for $L_2(R)$, most of them constitute unconditional bases for $L_p(R)$. A scaling function φ is called t regular, if φ has continuous derivatives of order t and its corresponding wavelet ψ has vanishing moments of order t, *i.e.*

$$\int x^k \psi(x)\mathrm{d}x = 0,\ k = 0,1,\cdots,t-1.$$

The following lemma [3] plays important roles in this paper.

Lemma 1.1. Let φ be a compactly supported, t regular orthonormal scaling function with the corresponding wavelet ψ and $0 < s < t$. If $f \in L_p(R)$, $s_{0k} := \langle f, \varphi_{0k}\rangle$, $d_{jk} := \langle f, \psi_{jk}\rangle$ and $1 \le p,\ q \le \infty$, then the following two conditions are equivalent:

1) $f \in B^s_{p,q}(R)$;

2) $\|s_{0\cdot}\|_p + \left\| \left\{ 2^{j\left(s+\frac{1}{2}-\frac{1}{p}\right)} \|d_{j\cdot}\|_p \right\}_{j\ge 0} \right\|_q < \infty$.

Furthermore,

$$\|f\|_{B^s_{p,q}} \sim \|s_{0\cdot}\|_p + \left\| \left\{ 2^{j\left(s+\frac{1}{2}-\frac{1}{p}\right)} \|d_{j\cdot}\|_p \right\}_{j\ge 0} \right\|_q .$$

Before introducing Fano's Lemma, we need the notation of Kullback-Leilber distance [4]. Let P and Q with P being absolutely continuous with respect to Q (denoted by $P \ll Q$). Then the Kullback-Leilber distance is defined by

$$K(P,Q) := \int_{p\cdot q>0} p(x)\ln\frac{p(x)}{q(x)}\mathrm{d}x,$$

Where p and q are density functions of P, Q respectively.

Lemma 1.2. (Fano's Lemma, [4]) Let $(\Omega, \aleph, P_k)$ be probability measurable spaces and $A_k \in \aleph$, $k = 0,1,\cdots,m$. If $A_k \cap A_v = \phi$ for $k \ne v$, then with A^c standing for the complement of A and $\kappa_m := \inf_{0\le v\le m} \frac{1}{m}\sum_{k\ne v} K(P_k, P_v)$,

$$\sup_{0\le k\le m} P_k\left(A_k^c\right) \ge \min\left\{\frac{1}{2}, \sqrt{m}\exp\left(-3e^{-1}-\kappa_m\right)\right\}.$$

By Lemma 1.1 and 1.2, we can show the following result:

Theorem 1.1. Let $f \in \tilde{B}^s_{r,q}(R,L)$ with $1 \le r,\ q \le \infty$, $1 \le p < \infty$ and $sr > 1$. If f_n is an estimator of f with n i.i.d. random samples, then

$$\sup_{f\in\tilde{B}^s_{r,q}(R,L)} E\left(\|f_n - f\|_p\right) \succ \max\left\{\left(\frac{\ln n}{n}\right)^{\frac{s-1/r+1/p}{2(s-1/r)+1}}, n^{-\frac{s}{2s+1}}\right\},$$

where $\tilde{B}^s_{r,q}(R,L) := \left\{ f \in B^s_{r,q}(R), \|f\|_{B^s_{r,q}} \le L \right\}$ and f has compact support}; The notation $x \succ y$ means $x \ge Cy$ with a constant C.

Remark 1.1. Note that

$$\max\left\{\left(\frac{\ln n}{n}\right)^{\frac{s-1/r+1/p}{2(s-1/r)+1}}, n^{-\frac{s}{2s+1}}\right\} = \left(\frac{\ln n}{n}\right)^{\frac{s-1/r+1/p}{2(s-1/r)+1}}$$

for $r \le \frac{p}{2s+1}$ and for $r > \frac{p}{2s+1}$

$$\max\left\{\left(\frac{\ln n}{n}\right)^{\frac{s-1/r+1/p}{2(s-1/r)+1}}, n^{-\frac{s}{2s+1}}\right\} = n^{-\frac{s}{2s+1}}.$$

Then theorem 1.1 is a reformulate of the lower bound in (1.1). By using the idea of reference [5], we show this theorem in the next two sections.

2. Proof of Theorem 1.1

Firstly, we prove

$$\sup_{f\in\tilde{B}^s_{r,q}(R,L)} E\left(\|f_n - f\|_p\right) \succ \left(\frac{\ln n}{n}\right)^{\frac{s-1/r+1/p}{2(s-1/r)+1}}.$$

One need construct g_k such that $g_k \in \tilde{B}^s_{r,q}(R,L)$ and

$$\sup_k E\left(\|f_n - g_k\|_p\right) \succ \left(\frac{\ln n}{n}\right)^{\frac{s-1/r+1/p}{2(s-1/r)+1}}.$$

Let φ be a compactly supported, $t(t > s)$ regular and orthonormal scaling function, ψ be the corresponding wavelet with $\operatorname{supp}\psi \subseteq [0,l)$, $l \in N^+$. Here and after,

N^{+} denotes the set of positive integers. Then there exists a compactly supported density function g (*i.e.* $g(x)\ge 0$ and $\int g(x)\mathrm{d}x=1$) satisfying

$$g(x)\in B_{r,q}^{s}(R) \text{ and } g(x)\big|_{[0,l]}=c_0>0.$$

Let $\Delta_j:=\left\{0,l,2l,\cdots,\left(2^j-1\right)l,2^jl\right\}$. Then the number of elements in Δ_j is 2^j+1, denoted by $\#\Delta_j=2^j+1$. Motivated by [5], one defines $a_j:=2^{-j(s+1/2-1/r)}$ and

$$g_k(x):=g(x)+a_j\psi_{jk}(x)I_{\{k\ne 2^jl\}},\ k\in\Delta_j$$

with $I_{\{k\ne 2^jl\}}:=1$ if $k\ne 2^jl$, else $I_{\{k\ne 2^jl\}}:=0$. Obviously, $g_{2^jl}=g$, $\int g(x)\mathrm{d}x=1$ and $g_k(x)\ge c_0-2^{-j(s-1/r)}\|\psi\|_\infty\ge 0$ for large j, which implies that g_k is a density function for each k.

By the assumptions of φ, the wavelet ψ is compactly supported and t times differentiable. Therefore, $\psi\in B_{r,q}^{s}(R)(t>s)$ and $g_k\in\tilde{B}_{r,q}^{s}(R)$. Because $a_j2^{j(s+1/2-1/r)}=1$, $\|a_j\psi_{jk}\|_{B_{r,q}^s}\le C$ and so is $\|g_k\|_{B_{r,q}^s}$ due to Lemma 1.1. Hence, $g_k(x)\in\tilde{B}_{r,q}^{s}(R,L)$. Clearly,

$$\begin{aligned}\|g_k-g_{k'}\|_p&\ge\left\|g_k-g_{2^jl}\right\|_p=\|a_j\psi_{jk}\|_p\\&=\|\psi\|_p2^{-j(s+1/p-1/r)}:=\delta_j\end{aligned}\tag{2.1}$$

For $k\ne k'\in\Delta_j/\{2^jl\}$ due to $a_j:=2^{-j(s+1/2-1/r)}$. Furthermore, $A_k:=\left\{\|f_n-g_k\|_p<\frac{\delta_j}{2}\right\}$ satisfies $A_k\cap A_{k'}=\phi$ for $k\ne k'$. Recall that $\#\Delta_j=2^j+1$. By Lemma 1.2, $\sup_{k\in\Delta_j}P_{g_k}^n\left(A_k^c\right)\ge\min\left\{\frac12,\sqrt{2^j}\exp\left(-\kappa_{2^j}-\frac3e\right)\right\}$. Here and after, P_f^n stands for the probability measure corresponding to the density function $f^n(x):=f(x_1)f(x_2)\cdots f(x_n)$. It is easy to see that $P_{g_k}^n\ll P_{g_0}^n$ from the constructions of g_k. Since f_n is an estimator of density with n i.i.d. random samples,

$$E\left(\|f_n-g_k\|_p\right)\ge\frac{\delta_j}{2}P_{g_{\varepsilon^i}}^n\left(\|f_n-g_k\|_p\ge\frac{\delta_j}{2}\right)=\frac{\delta_j}{2}P_{g_k}^n\left(A_k^c\right).$$

Then,

$$\begin{aligned}\sup_{k\in\Delta_j}E\left(\|f_n-g_k\|_p\right)&\ge\sup_{k\in\Delta_j}\frac{\delta_j}{2}P_{g_k}^n\left(A_k^c\right)\\&\ge\frac{\delta_j}{2}\min\left\{\frac12,\sqrt{2^j}\exp\left(-\kappa_{2^j}-\frac3e\right)\right\}.\end{aligned}\tag{2.2}$$

Next, one shows $\kappa_{2^j}\le c_0^{-1}na_j^2$: Recall that

$K\left(P_1^n,P_2^n\right):=\int_{f_1^n\cdot f_2^n>0}f_1^n(x)\ln\frac{f_1^n(x)}{f_2^n(x)}\mathrm{d}x$, $f_1^n(x)=\prod_{j=1}^n f_1(x_j)$ and $f_2^n(x)=\prod_{j=1}^n f_2(x_j)$. Then

$$K\left(P_1^n,P_2^n\right)=\sum_{i=1}^n\int f_1(x_i)\ln\frac{f_1(x_i)}{f_2(x_i)}\mathrm{d}x_i=nK\left(P_1^1,P_2^1\right).$$

Note that $K\left(P_1^1,P_2^1\right):=\int f_1(x)\ln\frac{f_1(x)}{f_2(x)}\mathrm{d}x$ and $\ln u\le u-1$ for $u>0$. Then

$$\begin{aligned}K\left(P_1^n,P_2^n\right)&=n\int f_1(x)\ln\frac{f_1(x)}{f_2(x)}\mathrm{d}x\\&\le n\int f_1(x)\left[\frac{f_1(x)}{f_2(x)}-1\right]\mathrm{d}x\\&=n\int\left|f_2(x)\right|^{-1}\left|f_1(x)-f_2(x)\right|^2\mathrm{d}x.\end{aligned}$$

Hence,

$$\kappa_{2^j}:=\inf_{v\in\Delta_j}\sum_{k\ne v}2^{-j}K\left(P_{g_k}^n,P_{g_v}^n\right)\le 2^{-j}\sum_{k\in\Delta_j}K\left(P_{g_k}^n,P_{g_{2^jl}}^n\right).$$

Moreover,

$$\kappa_{2^j}\le 2^{-j}n\sum_{k\in\Delta_j}\int\left|g(x)\right|^{-1}\left|g_k(x)-g(x)\right|^2\mathrm{d}x.\tag{2.3}$$

According to the definition of g_k, $\operatorname{supp}(g_k-g)\subseteq[0,l]$ and $g(x)=c_0$ on $[0,l]$. Thus, $\int\left|g(x)\right|^{-1}\left|g_k(x)-g(x)\right|^2 dx=c_0^{-1}\int\left|a_j\psi_{jk}(x)\right|^2\mathrm{d}x=c_0^{-1}a_j^2\left\|\psi_{jk}(x)\right\|_2^2=c_0^{-1}a_j^2$ by the orthonormality of ψ_{jk}. Then (2.3) reduces to

$$\kappa_{2^j}\le c_0^{-1}na_j^2.\tag{2.4}$$

Take $2^j\sim\left(\frac{n}{\ln n}\right)^{\frac{1}{2(s-1/r)+1}}$. Then $na_j^2=n2^{-2j\left(s+\frac12-\frac1r\right)}\sim\ln n$. Now, one can choose $C>0$ such that $na_j^2\le C\ln n$ and $C\left[4(s-1/r)+2\right]<c_0$. Therefore,

$$\sqrt{2^j}e^{-\kappa_{2^j}}\ge\sqrt{2^j}e^{-c_0^{-1}na_j^2}\succ\left(\frac{n}{\ln n}\right)^{\left[4(s-1/r)+2\right]^{-1}}n^{-Cc_0^{-1}}\ge 1$$

and (2.2) reduces to $\sup_{k\in\Delta_j}E\left(\|f_n-g_k\|_p\right)\ge C\delta_j$. Then the desired follows from $\delta_j=\|\psi\|_p2^{-j(s+1/p-1/r)}$ by (2.1) and $2^j\sim\left(\frac{n}{\ln n}\right)^{\frac{1}{2(s-1/r)+1}}$.

Now, we prove $\sup_{f\in\tilde{B}_{r,q}^s(R,L)}E\left(\|f_n-f\|_p\right)\succ n^{-\frac{s}{2s+1}}$. Our proof depends on another lemma [4].

Lemma 2.1. (Varshamov-Gilbert) Let $\Theta := \{\varepsilon = (\varepsilon_1, \cdots, \varepsilon_m)\}$, $\varepsilon_i \in \{0,1\}$. Then there exists a subset $\{\varepsilon^0, \cdots, \varepsilon^M\}$ of Θ with $\varepsilon^0 = (0, \cdots, 0)$ such that $M \ge 2^{m/8}$ and $\sum_{k=1}^{m} |\varepsilon_k^i - \varepsilon_k^j| \ge \frac{m}{8} (0 \le i \ne j \le M)$.

It is sufficient to construct g_{ε^i} $(i = 0,1,\cdots,M)$ such that $g_{\varepsilon^i} \in \tilde{B}^s_{r,q}(R,L)$ and

$$\sup_i E\left(\left\| f_n - g_{\varepsilon^i} \right\|_p\right) \succ n^{-\frac{s}{2s+1}}. \tag{2.5}$$

As proved above, let φ be a compactly supported, $t(t > s)$ regular and orthonormal scaling function, ψ be the corresponding wavelet with $\operatorname{supp}\psi \subseteq [0,l)$, $l \in N^+$. Assumes $g \in \tilde{B}^s_{r,q}(R,L)$ and $g|_{[0,l]} = c_0 > 0$. Define $a_j := 2^{-j(s+1/2)}$, $\Delta_j := \{0, l, 2l, \cdots, (2^j - 1)l\}$ and

$$g_\varepsilon(x) := g(x) + a_j \sum_{k \in \Delta_j} \varepsilon_k \psi_{jk}(x)$$

with $\varepsilon = (\varepsilon_k)_{k \in \Delta_j} \in \{0,1\}^{2^j}$ (note that $g_0 = g$). Since $\varepsilon_k \in \{0,1\}$, one knows that $\sum_{k \in \Delta_j} |\varepsilon_k|^r \le 2^j$ and

$$2^{j(s+1/2-1/r)} a_j \left(\sum_{k \in \Delta_j} |\varepsilon_k|^r\right)^{1/r} \le 1.$$

By Lemma 1.1, $\left\| a_j \sum_{k \in \Delta_j} \varepsilon_k \psi_{jk} \right\|_{B^s_{r,q}} \le C$, and so is $\|g_\varepsilon\|_{B^s_{r,q}}$. Hence $g_\varepsilon \in \tilde{B}^s_{r,q}(R,L)$.

Note that the supports of ψ_{jk} for $k \in \Delta_j$ are mutually disjoint. Then

$$g_\varepsilon(x) \ge c_0 - a_j \|\psi_{jk}\|_\infty \ge c_0 - 2^{-js} \|\psi\|_\infty \ge 0$$

for big j. This with $\int g_\varepsilon(x)\mathrm{d}x = \int g(x)\mathrm{d}x = 1$ implies that g_ε is a density function for each $\varepsilon \in \{0,1\}^{2^j}$. According to Lemma 2.1, there exists $\{\varepsilon^0, \varepsilon^1, \cdots, \varepsilon^M\}$ such that $M \ge 2^{2^{j-3}}$ and

$$\sum_{k \in \Delta_j} |\varepsilon_k^l - \varepsilon_k^i| \ge 2^{j-3}. \tag{2.6}$$

Because $\operatorname{supp}\psi_{jk} \cap \operatorname{supp}\psi_{jk'} = \phi$ for $k \ne k' \in \Delta_j$, one knows that

$$\begin{aligned} \left\| g_{\varepsilon^l} - g_{\varepsilon^i} \right\|_p^p &= \sum_{k \in \Delta_j} a_j^p |\varepsilon_k^l - \varepsilon_k^i|^p \|\psi_{jk}\|_p^p \\ &= 2^{-(sp+1)j} \|\psi\|_p^p \sum_{k \in \Delta_j} |\varepsilon_k^l - \varepsilon_k^i|^p \end{aligned}$$

This with (2.6) and ε_k^l, $\varepsilon_k^i \in \{0,1\}$ leads to $\left\| g_{\varepsilon^l} - g_{\varepsilon^i} \right\|_p^p \ge \|\psi\|_p^p 2^{-spj} 2^{-3}$ and

$$\left\| g_{\varepsilon^l} - g_{\varepsilon^i} \right\|_p \ge 8^{-1/p} \|\psi\|_p 2^{-sj} := \delta_j. \tag{2.7}$$

Clearly, the sets $A_{\varepsilon^i} = \left\{ \left\| f_n - g_{\varepsilon^i} \right\| < \frac{\delta_j}{2} \right\} (i = 0,1,\cdots,M)$ satisfy $A_{\varepsilon^l} \cap A_{\varepsilon^i} = \phi$ for $i \ne l$. Then Fano's Lemma yields

$$\sup_{0 \le i \le M} P^n_{g_{\varepsilon^i}}\left(A^c_{\varepsilon^i}\right) \ge \min\left\{\frac{1}{2}, \sqrt{M} \exp\left(-\kappa_M - 3e^{-1}\right)\right\}. \tag{2.8}$$

On the other hand, it follows $\kappa_M \le c_0^{-1} n a_j^2 2^j$ from the similar arguments to the proof of (2.4). Take $2^j \sim n^{\frac{1}{2s+1}}$. Then $na_j^2 = a2^{-(2s+1)j} \sim 1$. Hence, one can choose a constant $C > 0$ such that

$$\sqrt{M} e^{-\kappa_M} \ge 2^{2^{j-4}} \mathrm{e}^{-c_0^{-1} n a_j^2 2^j} \ge 2^{2^{j-4}} \mathrm{e}^{-c_0^{-1} C 2^j} \ge 1.$$

Therefore, (2.8) reduces to $\sup_{0 \le i \le M} P^n_{g_{\varepsilon^i}}\left(A^c_{\varepsilon^i}\right) \ge C > 0$ and

$$\begin{aligned} &\sup_{0 \le i \le M} E\left(\left\| f_n - g_{\varepsilon^i} \right\|_p\right) \\ &\ge \sup_{0 \le i \le M} \frac{\delta_j}{2} P^n_{g_{\varepsilon^i}}\left(\left\| f_n - g_{\varepsilon^i} \right\|_p \ge \frac{\delta_j}{2}\right) \ge C\delta_j. \end{aligned}$$

This with (2.7) and $2^j \sim n^{\frac{1}{2s+1}}$ yield (2.5).

3. Acknowledgements

The author Huiying Wang is grateful to the referees for their valuable comments and thanks her advisor, Professor Youming Liu, for his helpful guidance. This work is supported by the National Natural Science Foundation of China (No. 10871012) and Natural Science Foundation of Beijing (No. 1082003).

4. References

[1] G. Kerkyacharian and D. Picard, "Density Estimation in Besov Spaces," *Statistics & Probability Letters*, Vol. 13, No. 1, 1992, pp. 15-24. doi:10.1016/0167-7152(92)90231-S

[2] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian and D. Picard, "Density Estimation by Wavelet Thresholding," *The Annals of Statistics*, Vol. 24, No. 2, 1996, pp. 508-539. doi:10.1214/aos/1032894451

[3] W. Härdle, G. Kerkyacharian, D. Picard and A. B. Tsybakov, "Wavelets, Approximation and Statistical Applications," Springer-Verlag, New York, 1997.

[4] A. B. Tsybakov, "Introduction to Nonparametric Estimation," (English) Revised and Extended from the 2004 French Original, Translated by Vladimir Zaiats, Springer Series in Statistics, Springer, New York, 2009.

[5] P. Baldi, G. Kerkyacharian, D. Marinucci and D. Picard, "Adaptive Density Estimation for Directional Data Using Needlets," *The Annals of Statistics*, Vol. 37, No. 6A,

2009, pp. 3362-3395. doi:10.1214/09-AOS682

[6] C. Christophe, "Regression with Random Design: A Minimax Study," *Statistics & Probability Letters*, Vol. 77, No. 1, 2007, pp. 40-53. doi:10.1016/j.spl.2006.05.010

[7] A. B. Tsybakov, "Optimal Rates of Aggregation," *COLT/ Kernel* 2003 *Lecture Notes in Artificial Intelligence* 2777, Springer, Heidelberg, 2003, pp. 303-313.