The Lower Bound of Density Estimation for Biased Data in Sobolev Spaces

Jinru Wang[1,a] Yuan Zhou[2,b]

[12]College of Applied Science, Beijing University of Technology, Beijing, China

[a]wangjinru@bjut.edu.cn [b]yuanzhou@emails.bjut.edu.cn

Keywords: Lower Bound, Density Estimation, Biased Data, Sobolev Spaces.

Abstract. In this paper, we consider the density estimation problem from independent and identically distributed (i.i.d.) biased observations. We study the lower bound of convergence rates of density estimation over Sobolev spaces W_r^N （$N \in N^+$） under the L_p risk by using Fano's lemma.

Introduction

In this paper, we consider the problem of estimating the density functions without observing directly the sample $X_1, \dots X_n$. We observe an i.i.d. sample $Y_1, \dots Y_n$ from a biased distribution with the following density function $f^Y(y) = g(y) f^X(y)/\mu$, where g(y) is the so-called weighting or biasing function, $\mu = E\, g(X)$. Our purpose is to estimate the density function f^X from $Y_1, Y_2, \dots Y_n$. Several examples about this biased data can be found in the literature. For instance, in paper [1], it is shown that the distribution of the concentration of alcohol in the blood of intoxicated drivers is of interest, since the drunken driver has a larger chance of being arrested, the collected data are size-biased.

For unbiased data, Kerkyacharian and Picard [2] study a Besov space with matched case. Donoho, Johnstone, Kerkyacharian and Picard [3] consider a Besov space with unmatched case. They show the lower bound by using Korostelev and Assouad lemmas. However, the conditions of those two lemmas are difficult to be verified. In 2011, Huiying Wang [4] give a proof by using Fano's lemma. In reference [3, 4], they show the lower bound of convergence rates over Besov for L_p risk.

So far, wavelet density estimations about the lower bound of convergence rates over Besov spaces have made some achievements. However, there is a few people to study density estimation in Sobolev spaces W_r^N （$N \in N^+$）. In this paper, we study the lower bound of density estimation convergence over Sobolev spaces.

Preliminaries

In this paper, we always assume that scaling wavelet $\varphi(x)$ is orthonormal, compactly supported and $N+1$ regular. We consider the Sobolev balls $\tilde{W}_r^N(A, L)$ which is defined by:

$\tilde{W}_r^N(A, L) := \{ f : f \in W_r^N(R), f$ is a probability density on R with a compact support of length $\le A$, and $\|f^{(N)}\|_r \le L \}$.

Lemma 1 (Fano's lemma, [5]) Let $(\Omega, \mathrm{X}, P_k)$ be probability measurable spaces and $A_k \in \mathrm{X}$, $k = 0,1,\dots,m$. If $A_k \cap A_v = \Phi$ for any $k \ne v$, A_k^c standing for the complement of A_k and $k_m := \inf_{0 \le v \le m} \frac{1}{m} \sum_{k \ne v} K(P_k, P_v)$, we have

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

Lemma 2 (Varshamov-Gilbert lemma, [5]) 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 $M\geq 2^{m/8}$, and $\sum_{k=1}^{m}\left|\varepsilon_k^i-\varepsilon_k^j\right|\geq\frac{m}{8}(0\leq i\neq j\leq M)$.

Main result

Theorem 3 Let φ be a compactly supported, $N+1$ regular and orthonormal scaling function, $f^X\in\tilde{W}_r^N(A,L)$. We assume that there exist two constants g_1 and g_2 such that, for any $x\in R, 0<g_1\leq g(x)\leq g_2<\infty$. If $\hat{f}_n^X(x)$ is any estimator of f^X, then for $\forall 1\leq r,p<\infty, N>1/r$, we have

$$\sup_{f^X\in\tilde{W}_r^N(A,L)}E\left\|\hat{f}_n^X(x)-f^X(x)\right\|_p\succ\max\left\{\left(\frac{\ln n}{n}\right)^{\frac{N-1/r+1/p}{2(N-1/r)+1}},n^{-\frac{N}{2N+1}}\right\}.$$

where $x\succ y$ means $x\geq cy$ with a positive constant c.

Proof of Theorem 3: Using the idea of reference[4], firstly, we prove

$$\sup_{f^X\in\tilde{W}_r^N(A,L)}E\left\|\hat{f}_n^X(x)-f^X(x)\right\|_p\succ\left(\frac{\ln n}{n}\right)^{\frac{N-1/r+1/p}{2(N-1/r)+1}}.$$

Now we construct $h_k(x)$, such that $h_k(x)\in\tilde{W}_r^N(A,L)$ and

$$\sup_k E\left\|\hat{f}_n^X(x)-h_k(x)\right\|_p\succ\left(\frac{\ln n}{n}\right)^{\frac{N-1/r+1/p}{2(N-1/r)+1}}.$$

Let φ be a compactly supported, $N+1$ regular and orthonormal scaling function, ψ be the corresponding wavelet, and $\operatorname{supp}\psi\in[0,l)$, $l<A$. Then there exists a compactly supported density function $h(x)$ satisfying $h(x)\in W_r^N(R)$, and $h(x)\big|_{[0,l)}=C_0>0$.

Denote $\Delta_j=\{0,l,2l,...,(2^j-1)l,2^jl\}$, then the number of elements in Δ_j is 2^j+1. We define $a_j=2^{-j(N+1/2-1/r)}$, $h_k(x)=h(x)+a_j\psi_{j,k}(x)I\{k\neq 2^jl\}, k\in\Delta_j$. Obviously we have $h_{2^jl}(x)=h(x)$, $h_k(x)\in\tilde{W}_r^N(A,L)$. Let $h_k^Y(x)$ be the density function of $Y_1,Y_2,\cdots,Y_n$, then $h_k^Y(x)=\frac{g(x)h_k(x)}{\mu}$. For any $k\neq k'$, we get

$$\left\|h_k-h_{k'}\right\|_p=a_j\left\|\psi_{j,k}(x)-\psi_{j,k'}(x)\right\|_p\geq a_j\left\|\psi_{j,k}\right\|_p=2^{-j(N+1/p-1/r)}\left\|\psi\right\|_p:=\eta_j.$$

If denote $A_k=\left\{\left\|\hat{f}_n^X-h_k\right\|_p<\frac{\eta_j}{2}\right\}$, then using Fano's Lemma, we have

$$\sup_{k\in\Delta_j}P_{h_k^Y}^n\left(A_k^c\right)\geq\min\left\{\frac{1}{2},\sqrt{2^j}\exp\left(-3e^{-1}-k_{2^j}\right)\right\},$$

where P_f^n stands for the probability measure corresponding to the density function $f^n(x)=f(x_1)f(x_2)\cdots f(x_n)$. Since

$$E\left\|\hat{f}_n^X(x)-h_k(x)\right\|_p\geq\frac{\eta_j}{2}P_{h_k^Y}^n\left(\left\|\hat{f}_n^X-h_k\right\|_p\geq\frac{\eta_j}{2}\right)=\frac{\eta_j}{2}P_{h_k^Y}^n\left(A_k^c\right),$$

then $\sup_{k\in\Delta_j}E\left\|\hat{f}_n^X(x)-h_k(x)\right\|_p\geq\sup_{k\in\Delta_j}\frac{\eta_j}{2}P_{h_k^Y}^n\left(A_k^c\right)\geq\frac{\eta_j}{2}\min\left\{\frac{1}{2},\sqrt{2^j}\exp\left(-3e^{-1}-k_{2^j}\right)\right\}$.

where $k_{2^j}:=\inf_{v\in\Delta_j}\frac{1}{2^j}\sum_{k\neq v}K\left(P_{h_k^Y}^n,P_{h_v^Y}^n\right)$. Next, we shows $k_{2^j}\leq g_2C_0^{-1}na_j^2/g_1$. From

$$K\left(P_{f_1}^n, P_{f_2}^n\right)=\int_{f_1^n \cdot f_2^n>0} f_1^n(x)\ln\frac{f_1^n(x)}{f_2^n(x)}dx=\sum_{j=1}^{n}\int f_1(x_j)\ln\frac{f_1(x_j)}{f_2(x_j)}dx_j = nK\left(P_{f_1}^1, P_{f_2}^1\right),$$

and for any $u>0$, $\ln u \le u-1$, we have

$$K\left(P_{f_1}^n, P_{f_2}^n\right)=nK\left(P_{f_1}^1, P_{f_2}^1\right)\le n\int f_1(x)\left(\frac{f_1(x)}{f_2(x)}-1\right)dx = n\int f_2^{-1}(x)\left(f_1(x)-f_2(x)\right)^2 dx,$$

and

$$k_{2^j} \le 2^{-j}\sum_{k\in\Delta_j, k\ne 2^j l}\int\left(\frac{g(x)h_{2^j l}(x)}{\mu}\right)^{-1}\left(\frac{g(x)h_k(x)}{\mu}-\frac{g(x)h_{2^j l}(x)}{\mu}\right)^2 dx$$

$$\le \frac{g_2}{g_1}2^{-j}n\sum_{k\in\Delta_j, k\ne 2^j l} C_0^{-1}\int a_j^2\left|\psi_{j,k}\right|^2 dx = g_2 n C_0^{-1}a_j^2/g_1,$$

Taking $2^j \sim \left(\frac{n}{\ln n}\right)^{\frac{1}{2(N-1/r)+1}}$, then $na_j^2 = n2^{-2j(N+1/2-1/r)}\sim n\left(\frac{n}{\ln n}\right)^{-1}=\ln n$.

We choose C such that $na_j^2 \le C\ln n$, $Cg_2\left[4(N-1/r)+2\right]<C_0 g_1$, then

$$\sqrt{2^j}e^{-k_{2^j}} \ge \sqrt{2^j}e^{-g_2C_0^{-1}na_j^2/g_1} \ge \sqrt{2^j}n^{-g_2C_0^{-1}C/g_1}\ge 1.$$

Hence, $\sup_{k\in\Delta_j}E\left\|\hat f_n^X(x)-h_k(x)\right\|_p \ge \frac{\eta_j}{2}\min\left\{\frac{1}{2},\sqrt{2^j}\exp\left(-3e^{-1}-k_{2^j}\right)\right\}\ge C\eta_j$. Since

$\eta_j = 2^{-j(N+1/p-1/r)}\|\psi\|_p \sim C\left(\frac{\ln n}{n}\right)^{\frac{N+1/p-1/r}{2(N-1/r)+1}}$, then

$$\sup_{k\in\Delta_j}E\left\|\hat f_n^X(x)-h_k(x)\right\|_p \succ \left(\frac{\ln n}{n}\right)^{\frac{N+1/p-1/r}{2(N-1/r)+1}}.$$

Therefore, $\sup_{f^X\in\tilde W_r^N(A,L)}E\left\|\hat f_n^X(x)-f^X(x)\right\|_p \succ \left(\frac{\ln n}{n}\right)^{\frac{N-1/r+1/p}{2(N-1/r)+1}}$.

Next, we prove $\sup_{f^X\in\tilde W_r^N(A,L)}E\left\|\hat f_n^X(x)-f^X(x)\right\|_p \succ n^{-\frac{N}{2N+1}}$. Defining

$$a_j = 2^{-j(N+1/2)},\quad h_{\varepsilon^i}(x)=h(x)+a_j\sum_{k\in\Delta_j}\varepsilon_k^i\psi_{j,k}(x), i=0,1,\cdots,M.$$

with $\varepsilon^i=\left(\varepsilon_k^i\right)_{k\in\Delta_j}, \varepsilon_k^i\in\{0,1\}$. Obviously $h_{\varepsilon^0}(x)=h(x)$. Since $h_{\varepsilon^i}(x)=h(x)+a_j\sum_{k\in\Delta_j}\varepsilon_k^i\psi_{j,k}(x)\ge C_0-a_j\sum_{k\in\Delta_j}\left\|\psi_{j,k}\right\|_\infty \ge C_0-2^{-j(N-1)}\|\psi\|_\infty \ge 0$ (for large j), so $h_{\varepsilon^i}(x)$ is a probability density function. By the assumptions of φ, the wavelet ψ is compactly supported and $N+1$ times differentiable. Therefore $\psi\in\tilde W_r^N$, and we get

$$\left\|h_{\varepsilon^i}^{(N)}(x)\right\|_r \le \left\|h^{(N)}(x)\right\|_r + \left\|a_j\sum_{k\in\Delta_j}\varepsilon_k^i\psi_{j,k}^{(N)}(x)\right\|_r.$$

Since $\forall k\ne k'$, $\operatorname{supp}\psi_{jk}\cap\operatorname{supp}\psi_{jk'}=\Phi$, then we get

$$\left\|a_j\sum_{k\in\Delta_j}\varepsilon_k^i\psi_{j,k}^{(N)}(x)\right\|_r = a_j\left(\int\left|\sum_{k\in\Delta_j}\varepsilon_k^i\psi_{j,k}^{(N)}(x)\right|^r dx\right)^{1/r} = a_j 2^{j(1/2-1/r)}2^{jN}\left(\sum_{k\in\Delta_j}\left|\varepsilon_k^i\right|^r\right)^{1/r}\left\|\psi^{(N)}\right\|_r.$$

From $a_j=2^{-j(N+1/2)}$, and $\sum_{k\in\Delta_j}\left|\varepsilon_k^i\right|^r\le 2^j$, we have

$$\left\| a_j \sum_{k\in\Delta_j} \varepsilon_k^i \psi_{j,k}^{(N)}(x) \right\|_r \le \left\| \psi^{(N)} \right\|_r ,$$

hence, $\left\| h_{\varepsilon^i}^{(N)}(x) \right\|_r \le L$. Therefore $h_{\varepsilon^i}(x) \in \tilde{W}_r^N(A,L)$.

Let $h_{\varepsilon^i}^Y(x)$ be the density function of $Y_1, Y_2, \cdots, Y_n$, then $h_{\varepsilon^i}^Y(x) = \dfrac{g(x) h_{\varepsilon^i}(x)}{\mu}$. By Varshamov-Gilbert lemma, there exists $\{\varepsilon^0, \cdots, \varepsilon^M\}$ with $\varepsilon^0 = (0,\cdots,0)$, such that $M \ge 2^{2^j/8}$ and $\sum_{k=1}^m \left| \varepsilon_k^i - \varepsilon_k^l \right| \ge \dfrac{2^j}{8} (0 \le i \ne l \le M)$. For any $i \ne l$,

$$\left\| h_{\varepsilon^i} - h_{\varepsilon^l} \right\|_p^p = \left\| a_j \sum_{k\in\Delta_j} (\varepsilon_k^l - \varepsilon_k^i) \psi_{j,k} \right\|_p^p \ge \|\psi\|_p^p 2^{-(pN+1)j} 2^{j-3} = \|\psi\|_p^p 2^{-pjN} 2^{-3},$$

we have $\left\| h_{\varepsilon^i} - h_{\varepsilon^l} \right\|_p \ge \|\psi\|_p 2^{-Nj} 8^{-1/p} := \eta_j$. If we let

$$A_{\varepsilon^i} = \left\{ \left\| \hat{f}_n^X - h_{\varepsilon^i} \right\|_p < \frac{\eta_j}{2} \right\}, \quad i = 0,1,\ldots,M,$$

then $\forall i \ne l$, we get $A_{\varepsilon^i} \cap A_{\varepsilon^l} = \Phi$. Using Fano's Lemma, we have

$$\sup_i P_{h_{\varepsilon^i}^Y}^n \left(A_{\varepsilon^i}^c \right) \ge \min\left\{ \frac{1}{2}, \sqrt{M} \exp\left(-3e^{-1} - k_M\right) \right\}.$$

and $E\left\| \hat{f}_n^X(x) - h_{\varepsilon^i}(x) \right\|_p = \dfrac{\eta_j}{2} P_{h_{\varepsilon^i}^Y}^n \left(A_k^c \right)$. Moreover,

$$\sup_{k\in\Delta_j} E\left\| \hat{f}_n^X(x) - h_{\varepsilon^i}(x) \right\|_p \ge \sup_{k\in\Delta_j} \frac{\eta_j}{2} P_{h_{\varepsilon^i}^Y}^n \left(A_k^c \right) \ge \frac{\eta_j}{2} \min\left\{ \frac{1}{2}, \sqrt{M} \exp\left(-3e^{-1} - k_M\right) \right\}.$$

where $k_M := \inf_{0\le v\le M} \dfrac{1}{M} \sum_{i\ne v} K\left(P_{h_{\varepsilon^i}^Y}^n, P_{h_{\varepsilon^v}^Y}^n \right)$. Next, we shows $k_M \le g_2 C_0^{-1} n a_j^2 2^j / g_1$. We compute that

$$\begin{aligned} k_M &= \inf_{0\le v\le M} \frac{1}{M} \sum_{i\ne v} K\left(P_{h_{\varepsilon^i}^Y}^n, P_{h_{\varepsilon^v}^Y}^n \right) \le \frac{1}{M} \sum_{0<i\le M} K\left(P_{h_{\varepsilon^i}^Y}^n, P_{h_{\varepsilon^0}^Y}^n \right) \\ &\le \frac{n}{M} \sum_{0<i\le M} \int \left(\frac{g(x)h(x)}{\mu} \right)^{-1} \left(\frac{g(x) h_{\varepsilon^i}(x)}{\mu} - \frac{g(x)h(x)}{\mu} \right)^2 dx \\ &\le \frac{g_2}{g_1} \frac{n}{M} \sum_{0<i\le M} C_0^{-1} a_j^2 \sum_{k\in\Delta_j} \int \left| \psi_{j,k}(x) \right|^2 dx = g_2 n C_0^{-1} a_j^2 2^j / g_1 \end{aligned}$$

Taking $2^j \sim n^{\frac{1}{2N+1}}$, then $n a_j^2 = n 2^{-j(2N+1)} \sim 1$. We choose C such that $n a_j^2 \le C$, and $C < C_0 g_1 / 32 g_2$. From $\sqrt{M} \ge \sqrt{2^{2^{j-3}}}$, we get

$$\sqrt{M} e^{-k_M} \ge 2^{2^{j-4}} e^{-g_2 C_0^{-1} n a_j^2 2^j / g_1} \ge 2^{2^{j-4}(1 - 32 C g_2 C_0^{-1} g_1^{-1})} \ge 1.$$

So $\sup_{k\in\Delta_j} E\left\| \hat{f}_n^X(x) - h_{\varepsilon^i}(x) \right\|_p \ge \sup_{k\in\Delta_j} \dfrac{\eta_j}{2} P_{h_{\varepsilon^i}^Y}^n \left(A_k^c \right) \ge c\eta_j$, and $\eta_j = \|\psi\|_p 2^{-Nj} 8^{-1/p} \sim n^{-\frac{N}{2N+1}}$, therefore $\sup_{f^X \in \tilde{W}_r^N(A,L)} E\left\| \hat{f}_n^X(x) - f^X(x) \right\|_p \succ n^{-\frac{N}{2N+1}}$.

Acknowledgements

This paper is supported by Beijing Educational Committee Foundation (No.PHR201008022) and

Fundamental Research Foundation of Beijing University of Technology.

References

[1] S. Efromovich, Nonparametric curve estimation. Methods, Theory, and Applications, Springer Series in Statistics, 1999.
[2] G. Kerkyacharian, D. Picard. Density estimation in Besov spaces. Statist. Proab. Lett., 13 (1992) 15-24.
[3] D. L. Donoho, I. M.Johnstone, G. Kerkyacharian, D. Picard. Density estimation by wavelet thresholding. Ann. Statist., 24 (1996) 508-539.
[4] H. Y. Wang. Convergence rates of density estimation in Besov spaces. Applied Mathematics, 2 (2011) 1258-1262.
[5] A. B. Tsybakov. Introduction to Nonparametric Estimation. Springer-Verlag, Berlin, 2009.
[6] S. Efromovich, Density estimation for biased data. The Annals of Statistics, 32 (2004) 1137-1161.
[7] C. Chesnean. Wavelet block thresholding for density estimation in the presence of bias. Journal of the Korean Statistical Society, 39 (2010) 43-53.