Article

Pointwise Optimality of Wavelet Density Estimation for Negatively Associated Biased Sample

Renyu Ye 1,2, Xinsheng Liu 1,* and Yuncai Yu 1

1 State Key Laboratory of Mechanics and Control of Mechanical Structures, Department of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; yerenyu@nuaa.edu.cn (R.Y.); yuyuncai@nuaa.edu.cn (Y.Y.)
2 School of Mathematics and Computation Science, Anqing Normal University, Anqing 246133, China

Correspondence: xsliu@nuaa.edu.cn

Received: 14 December 2019; Accepted: 28 January 2020; Published: 2 February 2020

Abstract: This paper focuses on the density estimation problem that occurs when the sample is negatively associated and biased. We constructed a block thresholding wavelet estimator to recover the density function from the negatively associated biased sample. The pointwise optimality of this wavelet density estimation is shown as $L^p$ ($1 \leq p < \infty$) risks over Besov space. To validate the effectiveness of the block thresholding wavelet method, we provide some examples and implement the numerical simulations. The results indicate that our block thresholding wavelet density estimator is superior in terms of the mean squared error (MSE) when comparing with the nonlinear wavelet density estimator.

Keywords: pointwise optimality; wavelet; negatively associated; biased sample; density estimation

MSC: 62G07, 62G20, 42C40

1. Introduction

Let $X_1, X_2, \ldots, X_n$ be the unobserved realizations of random variable $X$ with the density function $g$, and $Y_1, Y_2, \ldots, Y_n$ be the recorded observations of random variable $Y$ with the density function:

$$f(y) = \frac{\eta(y)g(y)}{\mu}$$  \hspace{1cm} (1)

where $\eta(y)$ represents a biasing function and its corresponding expectation is $\mu = E(\eta(X)) = \int_0^1 \eta(y)g(y)dy$ ($0 < \mu < \infty$).

To achieve the goal that recovers the density function $g(y)$ in Model (1) from the sample $Y_1, Y_2, \ldots, Y_n$, some statisticians have conducted thorough explorations [1–4]. The wavelet method, which can be adapted to represent the local features of the density function, has been widely used in their researches. For instance, Ramirez and Vidakovic [2] estimated the density from the stratified size-biased sample using a linear wavelet method and proved the consistency of their wavelet estimator on the $L^2$ risk over Besov space. Chesneau [3] considered a pointwise wavelet estimation on the $L^p$ $(1 \leq p < \infty)$ risks [4] and extended this univariate wavelet estimation to the multivariate case. However, the above results rely on the independence assumption of the size-biased sample, which is a serious restriction in practical applications. As the size-biased sample is dependent, many researchers modelled the dependence of the sample as being negatively associated (NA), the definition of which is as follows.
Definition 1. Let $A$ and $B$ be an arbitrary pair of disjoint nonempty subsets of $\{1, 2, \ldots, n\}$. Assume that $f_1$ and $f_2$ are real-valued coordinate-wise nondecreasing functions. For a sequence of random variables $Y_1, Y_2, \ldots, Y_n$, if the covariances of the random variable functions exist and satisfy
\[ \text{Cov} \left( f_1(Y_k, k \in A), f_2(Y_m, m \in B) \right) \leq 0, \tag{2} \]
then $Y_1, Y_2, \ldots, Y_n$ are said to be NA.

The concept of negative association, was first proposed by Alam and Saxena [5] and its basic properties were investigated by Joag-Dev and Proschan [6]. Negative association has been widely applied to multivariate statistical analysis and systems reliability since it contains many multivariate distributions, such as: (a) multinomial, (b) convolution of unlike multinomial, (c) Dirichlet, (d) negatively correlated normal distribution, and (e) random sampling without replacement. Some studies investigated the fundamental and asymptotic properties for an NA sample ([7–12]). Recently, the NA sequence has been introduced into the Model (1), and the wavelet density estimation of NA size-biased sample was studied. For example, Chesneau [13] obtained the optimal convergence rate of the linear wavelet method for a NA size-biased sample on the $L^2$ risk. As the linear wavelet method is not adaptive, Liu and Xu [14] and Guo and Kou [15] considered a nonlinear wavelet method to estimate the density function from a NA (stratified) size-biased sample. This nonlinear wavelet density estimation has been shown to be adaptive, and a pointwise convergence rate over $L^p$ risks was established.

As far as we know, a good density estimation procedure should simultaneously achieve two objectives: computational efficiency and adaptivity [16,17]. Although the nonlinear method in [14,15] is adaptive, the convergence rate of the nonlinear estimation is nearly optimal (up to one logarithmic factor). Cai and Chicken [16] proposed a block thresholding method that can remove the logarithmic factor from the convergence rate of the wavelet density estimation. The block thresholding method provides spatial adaptivity to relatively subtle changes in the underlying density function. More specifically, it considers the wavelet coefficients of blocks of length $L$, but not the term by term; thus, it provides a degree of graduated smoothing, which amounts to choosing an appropriate bandwidth in kernel estimation. In contrast, the hard thresholding method in Liu and Xu [14] applied to density estimation involves using a global bandwidth in kernel estimation.

We aimed to remove the logarithmic factor in the convergence rate of wavelet density estimation and thus enhance computational efficiency. We selected the block thresholding technique [16,18] and allowed the sample to be NA size-biased. We structured an $L^p$ version of the block thresholding wavelet estimator for the density function $g(y)$ in Model (1). This estimator is adaptive and simultaneously achieves the pointwise optimal convergence rate over Besov space. Some examples are provided and the corresponding simulations were conducted using R software. The result indicated that the block thresholding wavelet density estimator is better than the nonlinear wavelet density estimator in terms of the mean squared error (MSE).

2. Notations and Assumptions

Denote the scaling function and its associated wavelet function by $\phi$ and $\psi$, respectively. We assume both $\phi$ and $\psi$ are $r$-regularity and have periodic boundary conditions on $[0, 1]$. Let $\phi_{i_0,j}(y) = 2^{i_0/2} \phi \left( 2^{i_0} y - j \right)$, $\psi_{i,j}(y) = 2^{i/2} \psi \left( 2^i y - j \right)$. We define an orthonormal basis (ONB) of $L^2[0,1]$ as:
\[ \Omega = \left\{ \phi_{i_0,j}, j = 0, 1, \ldots, 2^{i_0-1}, \psi_{i,j}, i \geq i_0 \geq 0, j = 0, 1, \ldots, 2^i-1 \right\}. \]
Then, a function \( g \in L^2[0,1] \) can be reconstructed as:
\[
g(y) = \sum_{j=0}^{2^l-1} \alpha_{ij} \psi_{0ij}(y) + \sum_{i \geq i_0} \sum_{j=0}^{2^l-1} \beta_{ij} \psi_{ij}(y),
\]
where \( \alpha_{ij} = \int_{0}^{1} g(y) \psi_{0ij}(y) \, dy, \beta_{ij} = \int_{0}^{1} g(y) \psi_{ij}(y) \, dy. \)

Besov space \( B^s_{p,q} \) includes the common Sobolev \( (B^s_{2,2}) \) and Hölder \( (B^s_{\infty,\infty}) \) spaces, so they offer a flexible collection of smooth functions. For a function \( g(y) \) that belongs to the Besov space, the Besov sequence norm \( \| \beta \|_{B^s_{p,q}} \) of the wavelet coefficients is bounded [19], which implies that a positive constant \( K \) exists, such that:
\[
\| \beta \|_{B^s_{p,q}} = \| \alpha_{00} \|_{L^p} + \left\{ \sum_{i=i_0}^{\infty} 2^{iq\gamma} \| \beta_i \|_{L^q}^{1/q} \right\}^{1/q} \leq K,
\]
where \( \beta \) and \( \alpha_{00} \) denote the vectors, \( s \) is an index of regularity that satisfies \( 0 < s < r, \theta \) and \( q \) satisfy \( 1 \leq \theta, q \leq \infty, \gamma = s + 1/2 - 1/h, \) and \( s > 1/h. \)

To establish our theorem, we list some assumptions that were necessary in our proofs.

(A1) The function \( g(y) \) is bounded; that is, a positive constant \( c_1 \) exists, such that:
\[
g(y) \leq c_1.
\]

(A2) The biasing function \( \eta(y) \) is non-increasing and for all \( y \in [0,1], \) two positive constants \( c_2 \) and \( c_3 \) exist, i.e.,
\[
c_2 \leq \eta(y) \leq c_3.
\]

**Remark 1.** The assumptions (A1) and (A2) are reasonable to obtain the pointwise adaptivity [3,15]. From the definition of \( g(y) \), the assumption (A2) implies that:
\[
\mu = \int_{0}^{1} \eta(y) f(y) \, dy \geq c_2 \int_{0}^{1} f(y) \, dy = c_2.
\]

### 3. Estimator and Main Result

Let \( L = (\log n)^{\nu/2} \) and divide the set \( \{0,1,\ldots,2^l-1\} \) into consecutive, non-overlapping blocks of length \( L \) at every resolution level \( i \), i.e.,
\[
\Delta_{im} = \{ j : (m-1)L + 1 \leq j \leq mL \} , m \in \mathbb{Z}.
\]

Consider \( \beta_n = \left( \frac{1}{n} \sum_{t=1}^{n} \frac{1}{\eta(Y_t)} \right)^{-1} \) as the estimator of \( \mu \). Define \( \lambda_{0l} = \frac{\beta_n}{n} \sum_{t=1}^{n} \frac{1}{\eta(Y_t)} \psi_{0l}(Y_t) \) and \( \hat{\beta}_{ij} = \frac{\beta_n}{n} \sum_{t=1}^{n} \frac{1}{\eta(Y_t)} \psi_{ij}(Y_t) \). Denote \( i_0 = \frac{\nu/2}{2} \log_2 \left( \log_2 n \right) \) and \( i_1 = \log_2 \left( n/\log_2 n \right) \) and let \( \sum_{(im)} \) represent the summation over \( j \in \Delta_{im} \). Then, the block thresholding estimator is given by:
\[
\hat{g}(y) = \sum_{j=0}^{2^l-1} \lambda_{0j} \psi_{0j}(y) + \sum_{i=i_0}^{i_1-1} \sum_{m} \sum_{(im)} \left( \hat{\beta}_{ij} \psi_{ij}(y) \right) I(\hat{B}_{im} > \lambda),
\]
where \( \hat{B}_{im} = \left( L^{-1} \sum_{(im)} \hat{\beta}_{ij}^p \right)^{1/p} \) and the threshold is set to \( \lambda = \sqrt{48 \left( c_2^{-1} \mu + c_1 \right)} n^{-1}. \)

We outline the pointwise convergence rate of the wavelet density estimator \( \hat{g}(y) \) in the following theorem.
Theorem 1. Let \( Y_1, Y_2, \cdots, Y_n \) be an identically distributed NA sequence in Model (1). Suppose \( g \in B_{p,q}^s \) and \( s > \max \{1/h, 1/2\} \). For the block thresholding wavelet estimator \( \hat{g}(y) \) defined by (5), if the conditions (A1) and (A2) are satisfied, then for \( 1 \leq p < \infty \) and \( h \geq p/(2s+1) \), a positive constant \( C \) exists, such that:

\[
\sup_{g \in B_{p,q}^s} E \int (\hat{g}(y) - g(y))^2 dy \leq \begin{cases} 
C n^{-ps/(2s+1)}, & \text{for } h \geq p, \\
C n^{-ps/(2s+1)}(\log n)^{ps/(2s+1)}, & \text{for } p/(2s+1) \leq h < p.
\end{cases}
\]

Remark 2. The block thresholding wavelet estimator is adaptive since \( \delta \) and \( \eta \) do not rely on \( s, h, q \). If \( p = 2 \) and \( \eta(y) = 1 \), Model (1) reduces to the standard density estimation model and the convergence rate in Theorem 1 is the same as the results of Cai and Chicken [16]. If the NA bias reduces to the independent bias, the results of the Theorem 1 become Theorem (4.1) in Chesneau [3].

Remark 3. With the presence of NA bias, Liu and Xu [14] and Guo and Kou [15] established the near optimal convergence rates (up to a logarithmic term) of a nonlinear wavelet estimator for the density function in Model (1). Note that the logarithmic term in the convergence rate has been removed in Theorem 1, which improves the convergence rates in [14,15].

4. Simulation Study

We used the method in Alam and Saxena [5] and Liu and Xu [14] to generate NA samples, firstly drawing \( n \) samples \( Z_1, Z_2, \cdots, Z_n \) from the distribution \( \text{Beta}(2, 2n - 2) \), then choosing \( Y_1 = Z_1/\delta > 0 \) for \( p \) and \( \eta(y) = 1 \). Thus, \( Y_1, Y_2, \cdots, Y_n \) are NA (refer to Liu and Xu [14]). Throughout the simulations, we took \( \delta = 1/2 \).

Now, we provide three examples including the linear and nonlinear density functions, and their derivatives are continuous and discontinuous.

Example 1. Let \( \eta(y) = y \) and \( g(y) = 3y^3 + \frac{y}{2}, y \in [0, 1] \), then \( g(y) \) is nonlinear density function and its derivative is continuous, \( \mu = \frac{23}{30} \) and:

\[
f(y) = \frac{30}{23} \left( 3y^4 + \frac{y^2}{2} \right), \quad 0 \leq y \leq 1.
\]

Example 2. Let \( \eta(y) = y + 1 \) and:

\[
g(y) = \begin{cases} 
2y + \frac{1}{2}, & 0 \leq y \leq \frac{1}{2}, \\
-2y + \frac{5}{2}, & \frac{1}{2} < y \leq 1.
\end{cases}
\]

\( g(y) \) is a linear continuous function but its derivative is discontinuous at \( y = \frac{1}{2} \). We have \( \mu = \frac{13}{12} \) and:

\[
f(y) = \begin{cases} 
\frac{12}{13} \left( 2y^2 + \frac{2}{3}y + \frac{1}{3} \right), & 0 \leq y \leq \frac{1}{2}, \\
\frac{12}{13} \left( -2y^2 + \frac{2}{3}y + \frac{5}{3} \right), & \frac{1}{2} < y \leq 1.
\end{cases}
\]

Example 3. Let \( \eta(y) = y + 1 \) and

\[
g(y) = \begin{cases} 
3y^2 + \frac{3}{4}, & 0 \leq y \leq \frac{1}{2}, \\
3y^2 - 6y + \frac{15}{4}, & \frac{1}{2} < y \leq 1.
\end{cases}
\]

\( g(y) \) is a nonlinear continuous function but its derivative is discontinuous at \( y = \frac{1}{2} \). We have \( \mu = \frac{3}{2} \) and:

\[
f(y) = \begin{cases} 
\frac{2}{9} \left( 3y^3 + 3y^2 + \frac{3}{2}y + \frac{3}{4} \right), & 0 \leq y \leq \frac{1}{2}, \\
\frac{2}{9} \left( 3y^3 - 3y^2 - \frac{3}{2}y + \frac{15}{4} \right), & \frac{1}{2} < y \leq 1.
\end{cases}
\]
For the block wavelet threshold, we set $c_2 = \mu$ and $c_1 = \max_{1 \leq t \leq n} \{g(Y_t)\} + 1$. To evaluate the performance of the density estimators, we considered the MSE, which is defined as:

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^{n} (\hat{g}(Y_t) - g(Y_t))^2.$$

Figures 1–3 display the recovery of the density functions by wavelet density estimators for NA samples in the above examples. Table 1 lists the MSEs of these estimators. The results indicated both the nonlinear wavelet method and the block thresholding wavelet method performed well in the density estimation problem even though the samples were NA size-biased. The block wavelet density estimator performed better than the nonlinear wavelet density estimator in terms of the MSE; the estimations were increasingly accurate with increasing sample size.

| $n$  | Example 1 Block | Example 1 Nonlinear | Example 2 Block | Example 2 Nonlinear | Example 3 Block | Example 3 Nonlinear |
|------|----------------|---------------------|----------------|---------------------|----------------|---------------------|
| 512  | 0.0047         | 0.0072              | 0.0064         | 0.0102              | 0.0067         | 0.0114              |
| 1024 | 0.0043         | 0.0065              | 0.0056         | 0.0098              | 0.0054         | 0.0101              |
| 2048 | 0.0039         | 0.0058              | 0.0053         | 0.0096              | 0.0048         | 0.0093              |

Figure 1. Recovery of the density function for NA samples in Example 1 ($n = 2048$).
Figure 2. Recovery of the density function for NA samples in Example 2 \((n = 2048)\).

Figure 3. Recovery of the density function for NA samples in the Example 3 \((n = 2048)\).

5. Proof of Theorem 1

The proof of Theorem 1 is similar to that of Theorem (4.2) in Chesneau [20] and Theorem (4.1) in Chesneau [3]. The difference is that we considered that the samples were not only biased, but also NA. Hence, we had to overcome some non-trivial technical difficulties. Before elaborating the detailed proof, we introduce some basic properties and inequalities of the NA sequence in the following lemmas.

Lemma 1. [13]. For a sequence of NA random variables \(Z_1, Z_2, \cdots, Z_n\) and the non-empty subsets \(B_1, B_2, \cdots, B_k\) of \(\{1, 2, \cdots, n\}\), if \(B_1, B_2, \cdots, B_k\) are pairwise disjoint and \(f_1, f_2, \cdots, f_k\) are \(k\) coordinate-wise non-decreasing Borel functions, then \(f_1(Z_t, t \in B_1), f_2(Z_t, t \in B_2), \cdots, f_k(Z_t, t \in B_k)\) are still NA.
Lemma 2. [20]. Let \( \{Z_t, t \geq 1\} \) be NA random variables and \( \{Z^*_t, t \geq 1\} \) be independent random variables with the same marginal distribution as \( Z_t \). If \( f \) is a non-decreasing function, then:

\[
E \left[ f \left( \max_{1 \leq k \leq n} \left| \sum_{i=1}^{k} Z_t \right| \right) \right] \leq E \left[ f \left( \max_{1 \leq k \leq n} \left| \sum_{i=1}^{k} Z^*_t \right| \right) \right].
\]

Lemma 3. [20]. If \( \{Z_t\} \) and \( \{Z^*_t\} \) are NA random variables with \( EZ_t = 0 \) and \( EZ_t^2 < \infty \), then we have:

1. Rosenthal-type inequality: If \( E|Z_t|^p < \infty \) for some \( p \geq 2 \), then a constant \( C_p \) (only depends on \( p \)) exists, such that:

\[
E \left( \max_{1 \leq k \leq n} \left| \sum_{i=1}^{k} Z_t \right|^p \right) \leq C_p \left\{ \sum_{i=1}^{n} E|Z_t|^p + \left( \sum_{i=1}^{n} EZ_t^2 \right)^{p/2} \right\}.
\]

2. Kolmogorov-type inequality: Denote \( b_n^2 = \sum_{i=1}^{n} EZ_t^2 \), we have, for all \( 0 < b < 1 \) and \( \tau > 0 \):

\[
P \left( \max_{1 \leq k \leq n} \left| \sum_{i=1}^{k} Z_t \right| \geq z \right) \leq 2P \left( \max_{1 \leq k \leq n} |Z_k| > z \right) + \frac{1}{1-b} \exp \left\{ -\frac{\tau^2 b}{2(\tau z + b_n^2)} \right\}.
\]

Proof of Theorem 1. According to the proof of Theorem (4.1) in Chesneau [3], the proof of Theorem 1 will be completed if we can show that the moment inequality

\[
E|\hat{\beta}_{ij} - \beta_{ij}|^{2p} \leq Cn^{-p},
\]

and the large deviation inequality

\[
P \left( \left( \sum_{i \in B_m} \left| \hat{\beta}_{ij} - \beta_{ij} \right|^p \right)^{1/p} \geq \frac{\lambda}{2} (\log n)^{-1/2} \right) \leq n^{-p},
\]

hold at resolution level \( i \) (the case of primary resolution level \( i_0 \) can be treated similarly). Therefore, the remainder of the proof is composed of the following two parts.

**Part one:** Moment inequality. By the proof of Proposition 4.1 in Chesneau [3], we have that the inequality (6) is equivalent to

\[
E|\hat{\beta}_{ij} - \beta_{ij}|^{2p} \leq C_p \left( G_{ij} + H_{ij} \right),
\]

where:

\[
G_{ij} = E \left( \left| \mu n^{-1} \sum_{i=1}^{n} \eta^{-1} (Y_i) \psi_{ij} - \beta_{ij} \right|^{2p} \right),
\]

\[
H_{ij} = E \left( \left| n^{-1} \sum_{i=1}^{n} \eta^{-1} (Y_i) - \mu^{-1} \right|^{2p} \right).
\]

We first consider the term \( G_{ij} \). Write:

\[
n^{-1} \sum_{i=1}^{n} \mu \eta^{-1} (Y_i) \psi_{ij} - \beta_{ij} = n^{-1} \sum_{i=1}^{n} \left( \mu \eta^{-1} (Y_i) \psi_{ij} - \beta_{ij} \right) = n^{-1} \sum_{i=1}^{n} \xi_{ij}.
\]

Since \( \psi \) is a bounded variation function, then two bounded nonnegative nondecreasing functions exist, \( \bar{\psi} \) and \( \tilde{\psi} \), such that \( \psi = \bar{\psi} - \tilde{\psi} \). Denote:

\[
\xi_{ij} = \mu \eta^{-1} (Y_i) \tilde{\psi}_{ij} (Y_i) - \bar{\psi}_{ij}, \quad \xi_{ij} = \mu \eta^{-1} (Y_i) \bar{\psi}_{ij} (Y_i) - \tilde{\psi}_{ij},
\]

(9)
where $\bar{\beta}_{ij} = \int_0^1 g(y) \bar{\psi}_{ij}(y) \, dy$ and $\tilde{\beta}_{ij} = \int_0^1 g(y) \tilde{\psi}_{ij}(y) \, dy$ and, hence, $\xi_i = \bar{\xi}_i - \tilde{\xi}_i$. As Lemma 1 and $E(\tilde{\xi}_1) = \int_0^1 \mu \eta^{-1}(y) (g(y) \eta(y) \mu^{-1}) \, dy - \beta_{ij} = 0$, $\xi_i$ and $\tilde{\xi}_i$ ($t=1,2,\cdots,n$) are zero NNA random variables according to the monotonicity of $\eta(y)$ in assumption (A2).

Note that $|\beta_{ij}| \leq \left( \int_0^1 g^2(y) \, dy \right)^{1/2} \left( \int_0^1 \psi^2_{ij}(y) \, dy \right)^{1/2} \leq c_1$ and $|\psi_{ij}(y)| \leq \sup_{y \in [0,1]} |\psi_{ij}(y)| \leq 2^{1/2}$ \sup_{y \in [0,1]} |\psi(y)| by assumption (A2) and Model (4). It follows that:

$$
E\left( |\tilde{\xi}_1|^p \right) = 2^{p-1}\left( E\left( |\mu \eta^{-1}(Y_1) \bar{\psi}_{ij}(Y_1)|^p \right) + |\tilde{\beta}_{ij}|^p \right) 
\leq 2^{p-1}\left( \mu^p c_2^{-\frac{1}{2}} 2^{(p-2)/2} \left( \sup_{y \in [0,1]} |\psi(y)| \right)^{p-2} E\left( \eta^{-1}(Y_1) \bar{\psi}_{ij}^2(Y_1) \right) + c_1^p \right).$$

For any $y \in [0,1]$, we have:

$$
E\left( \eta^{-1}(Y_1) \bar{\psi}_{ij}^2(Y_1) \right) = \int_0^1 \eta^{-1}(y) \bar{\psi}_{ij}^2(y) (\mu^{-1} \eta(y) g(y)) \, dy 
\leq \mu^{-1} c_1 \int_0^1 \bar{\psi}_{ij}^2(y) \, dy \leq \mu^{-1} c_1.
$$

Hence, for all resolution levels $i_0 \leq i \leq i_1$:

$$
E\left( |\tilde{\xi}_1|^p \right) \leq 2^{p-1}\left( c_1 \mu^{p-1} c_2^{-\frac{1}{2}} 2^{(p-2)/2} \left( \sup_{y \in [0,1]} |\psi(y)| \right)^{p-2} + c_1^p \right) 
\leq 2^{p-1}\left( C n^{(p-2)/2} + c_1^p \right) = O\left( n^{(p-2)/2} \right).
$$

Resorting to the Rosenthal-type inequality in Lemma 3, we obtain:

$$
E\left( \left| n^{-1} \sum_{i=1}^n \tilde{\xi}_i \right|^{2p} \right) \leq C \left( n^{1-2p} E\left( \tilde{\xi}_1^{2p} \right) + n^{-p} \left( E\left( \tilde{\xi}_1^{2p} \right) \right)^p \right) = O\left( n^{-p} \right).
$$

Similarly, for $\bar{\xi}_i$, we have:

$$
E\left( \left| n^{-1} \sum_{i=1}^n \bar{\xi}_i \right|^{2p} \right) = O\left( n^{-p} \right).
$$

Recall that $\tilde{\xi}_i = \xi_i - \bar{\xi}_i$, then we obtain:

$$
G_{jk} = E\left( \left| n^{-1} \sum_{i=1}^n \tilde{\xi}_i \right|^{2p} \right) \leq \frac{1}{n^{2p}} \left[ E\left| \sum_{i=1}^n \bar{\xi}_i \right|^{2p} + E\left| \sum_{i=1}^n \xi_i \right|^{2p} \right] = O\left( n^{-p} \right).
$$

(10)

Now, we consider the term $H_{ij}$. Denote $\zeta_i = \frac{1}{\eta(Y_i)} - \frac{1}{\eta}$. It is easy to check that $E\left( |\xi_1|^p \right) \leq 2^{p-1} (E(\eta^{-p}(Y_1)) + \mu^{-p}) \leq c_1$. Resorting to the Rosenthal-type inequality again, we have:

$$
H_{ij} = E\left( \left| n^{-1} \sum_{i=1}^n \xi_i \right|^{2p} \right) 
\leq C \left( n^{1-2p} E\left( \xi_1^{2p} \right) + n^{-p} \left( E\left( \xi_1^{2p} \right) \right)^p \right) 
\leq C \left( n^{1-2p} + n^{-p} \right) = O\left( n^{-p} \right).
$$

(11)

Consequently, the inequality (6) is followed by (8), (10), and (11).
Part two: Large deviation inequality. From the proof of Proposition 4.2 in Chesneau [3], we know that the inequality (7) is bounded by:

\[
P\left(\left(\sum_{j \in \Delta_m} |\hat{\beta}_{ij} - \beta_{ij}|^p\right)^{1/p} \geq \sqrt{32\left(c_2^{-1}\mu + c_1\right)^2 n^{-1/2}(\log_2 n)^{1/2}}\right) \leq P\left(\frac{1}{2} \sum_{t=1}^n \left(\frac{1}{\eta(Y_t)} - \frac{1}{2}\right) \right) \geq \frac{8\left(c_2^{-1}\mu + c_1\right)^2 n^{-1/2}(\log_2 n)^{1/2}}{1/2} + P\left(\left(\sum_{j \in \Delta_m} \left|\frac{\hat{\mu}}{\eta(Y_t)} \psi_j(Y_t) - \beta_{ij}\right|^p\right)^{1/p} \geq \frac{8\left(c_2^{-1}\mu + c_1\right)^2 n^{-1/2}(\log_2 n)^{1/2}}{1/2}\right)
\]

Now, let us consider the upper bounds of \(S_1\) and \(S_2\), separately.

Bound for \(S_1\): Recall that \(z_t = \frac{1}{\eta(Y)} - \frac{1}{\mu} \) and \(t = 1, \cdots, n\), are NA random variables with zero mean and finite variance \(\sigma^2\). Hence, the triangular inequality yields:

\[|z_t| \leq \mu^{-1} + \eta^{-1}(Y_t) \leq 2c_2^{-1}.
\]

Taking \(\tau = n^{1/2}(\log_2 n)^{1/2}, b = \frac{1}{2}\) and \(z = 2c_2^{-1}\), using the Kolmogorov-type inequality in Lemma 3, we have:

\[
S_1 \leq P\left(\max_{1 \leq t \leq n} \left|\sum_{t=1}^n \hat{\mu}_t \right| \geq C n^{1/2}(\log_2 n)^{1/2}\right) \leq 4 \exp\left\{-\frac{\delta_1^2\mu + c_1^2}{n(\log_2 n)}\right\}
\]

Bound for \(S_2\): Recall that \(L = (\log_2 n)^{(p-2)},\) we set \(\lambda_s = \sqrt{8\left(c_2^{-1}\mu + c_1\right)^2 n^{-1/2}(\log_2 n)^{1/2}}\), then \(S_2 \leq O(n^{-p})\) implies that:

\[
P\left\{\left(\sum_{j \in \Delta_m} \left|\frac{\hat{\mu}}{\eta(Y_t)} \psi_j(Y_t) - \beta_{ij}\right|^p\right)^{1/p} \geq \lambda_s\right\} \leq C \exp\left\{-\frac{n \lambda_s^2}{8\left(c_2^{-1}\mu + c_1\right)^2}\right\}.
\]

Now, we will check the inequality in (13). Let \(d = \frac{p}{p-1}\) and define a set \(\Theta = \left\{\theta = (\theta_1, \theta_2, \cdots, \theta_L) \in \mathbb{R}^L : \sum_{l=1}^L \theta_l^d \leq 1\right\}\). By the elementary inequality and the argument of duality, we show that for all integers \(l, m\):

\[
\left(\sum_{j \in \Delta_m} \left|\frac{n^{-1} \sum_{t=1}^n \frac{\hat{\mu}}{\eta(Y_t)} \psi_j(Y_t) - \beta_{ij}\right|^p\right)^{1/p} \leq \left(\sum_{j \in \Delta_m} \left|\frac{n^{-1} \sum_{t=1}^n \frac{\mu}{\eta(Y_t)} \psi_j(Y_t) - \beta_{ij}\right|^p\right)^{1/2}\right)^{1/2} = \sup_{\theta \in \Theta} \sum_{j=1}^L \theta_j \left(\sum_{j \in \Delta_m} \left|\frac{n^{-1} \sum_{t=1}^n \frac{\mu}{\eta(Y_t)} \psi_j(Y_t) - \beta_{ij}\right|^p\right)^{1/2}.
\]

Therefore, substituting Equation (14) into Equation (13) produces:

\[
P\left\{\sup_{\theta \in \Theta} \sum_{j=1}^L \theta_j \left(\sum_{j \in \Delta_m} \left|\frac{n^{-1} \sum_{t=1}^n \frac{\mu}{\eta(Y_t)} \psi_j(Y_t) - \beta_{ij}\right|^p\right)^{1/2} \geq \lambda_s\right\} \leq C \exp\left\{-\frac{n \lambda_s^2}{8\left(c_2^{-1}\mu + c_1\right)^2}\right\}.
\]
To prove Equation (15), we define a stochastic process \( \{ T_n (\theta), \theta \in \Theta \} \) with

\[
T_n (\theta) = \sum_{j=1}^{L} \theta_j \left( n^{-1} \sum_{t=1}^{n} \frac{\mu}{\eta (Y_t)} \psi_{ij} (Y_t) - \beta_{ij} \right).
\]

Note that \( \beta_{ij} = \int_{0}^{1} g (y) \psi_{ij} (y) dy = \frac{1}{n} \sum_{t=1}^{n} g (Y_t) \psi_{ij} (Y_t) \). Denote \( e_t = \frac{\mu}{\eta (Y_t)} - g (Y_t) \), then \( T_n (\theta) \) can be rewritten as:

\[
T_n (\theta) = \frac{1}{n} \sum_{t=1}^{n} e_t \sum_{j=1}^{L} \theta_j \psi_{ij} (Y_t).
\]

Similar to the definition of \( \bar{\psi} \) and \( \bar{\psi} \) in Equation (9), we set \( \bar{\psi}_{ij} (Y_t) = \frac{1}{L} \sum_{j=1}^{L} \theta_j \tilde{\psi}_{ij} (Y_t) \). Let \( \epsilon > 0 \), if \( |\epsilon T_n (\theta)| \leq 1 \), then:

\[
E \sup_{\theta \in \Theta} \exp \left\{ \epsilon |T_n (\theta)| \right\} \leq 1 + \sum_{k=2}^{\infty} E \sup_{\theta \in \Theta} \left( |\epsilon T_n (\theta)|^k \right) \leq 1 + \epsilon^2 E \sup_{\theta \in \Theta} |T_n (\theta)|^2 \left\{ \frac{1}{p} + \frac{1}{p} + \cdots \right\} \leq \exp \left( \epsilon^2 E \sup_{\theta \in \Theta} |T_n (\theta)|^2 \right).
\]

Recall that \( \sum_{j=1}^{L} \theta_j^d \leq 1 \) and \( \frac{1}{d} + \frac{1}{p} = 1 \), \( |e_t| \leq \left| \frac{\mu}{\eta (Y_t)} \right| + |g (Y_t)| \leq c_2^{-1} \mu + c_1 \), invoking the Hölder’s inequality produces:

\[
\exp \left\{ E \sup_{\theta \in \Theta} |\epsilon T_n (\theta)|^2 \right\} \leq \sup_{\theta \in \Theta} \prod_{t=1}^{n} \left( \exp \left\{ E \frac{\mu^2}{\eta^2} \sum_{j=1}^{L} |\theta_j \tilde{\psi}_{ij} (Y_t) |^2 (e_t)^2 \right\} \right) \leq \sup_{\theta \in \Theta} \prod_{t=1}^{n} \left( \exp \left\{ \left( c_2^{-1} \mu + c_1 \right)^2 \frac{\mu^2}{\eta^2} \left( \sum_{j=1}^{L} \theta_j^d \right)^{2/4} \left( \sum_{j=1}^{L} \tilde{\psi}_{ij}^p (Y_t) \right)^{2/p} \right) \right) \leq \exp \left\{ \left( c_2^{-1} \mu + c_1 \right)^2 \frac{\mu^2}{\eta^2} \prod_{t=1}^{n} \left( \sum_{j=1}^{L} |\tilde{\psi}_{ij} (Y_t)|^p \right)^{2/p} \right\}.
\]

Note that when \( t = j \), \( |\tilde{\psi}_{ij} (Y_t)|^p = 1 \), and for all \( t \neq j \), \( |\tilde{\psi}_{ij} (Y_t)|^p = 0 \). Hence, for every \( j (j = 1, 2, \ldots, L) \), only a counterpart \( t = j \) exists, such that \( |\tilde{\psi}_{ij} (Y_t)|^p = 1 \); that is,

\[
\sum_{t=1}^{n} \left( \sum_{j=1}^{L} |\tilde{\psi}_{ij} (Y_t)|^p \right)^{2/p} = L^{2/p},
\]

which leads to:

\[
\exp \left\{ \epsilon^2 E \sup_{\theta \in \Theta} |T_n (\theta)|^2 \right\} \leq C \exp \left\{ \frac{\left( c_2^{-1} \mu + c_1 \right)^2 \epsilon^2 L^{2/p}}{n^2} \right\}.
\]

(17)
Now we consider the case $|ε\hat{T}_n(θ)| > 1$. By Lemmas 1 and 2, for $ε > 0$, we obtain:

\[
E \sup_{θ ∈ Θ} \exp \left\{ |ε\hat{T}_n(θ)| \right\} \leq E \exp \left\{ E \sup_{θ ∈ Θ} |ε\hat{T}_n(θ)|^2 \right\} \\
\leq \sup_{θ ∈ Θ} \prod_{i=1}^{n} E \left( \exp \left\{ \frac{c_1^2(θ^2 + 1)^2}{n^2} \sum_{j=1}^{n} \left| \theta_{ij} Y_j(θ) \right|^2 \right\} \right)^{2/p} \\
\leq C \exp \left\{ \frac{\left( c_1^2 1 + c_1 \right)^2}{n^2} \right\}.
\]  

(18)

Thus, using Equations (16)–(18) together, we derive that:

\[
E \sup_{θ ∈ Θ} \exp \left\{ |ε\hat{T}_n(θ)| \right\} \leq C \exp \left\{ \frac{\left( c_1^2 1 + c_1 \right)^2}{n^2} \right\}.
\]

Likewise, for the term $\hat{T}_n(θ)$:

\[
E \sup_{θ ∈ Θ} \exp \left\{ |ε\hat{T}_n(θ)| \right\} \leq C \exp \left\{ \frac{\left( c_1^2 1 + c_1 \right)^2}{n^2} \right\}.
\]

Applying Markov’s inequality, all $λ_s > 0$ and $ε > 0$, we have:

\[
P \left\{ \sup_{θ ∈ Θ} |\hat{T}_n(θ)| ≥ λ_s \right\} \leq \frac{E \exp \left\{ \sup_{θ ∈ Θ} \hat{T}_n(θ) \right\}}{\exp \{ελ_s\}} \leq \exp \left\{ -ελ_s + \left( c_1^2 1 + c_1 \right)^2 ε^2 / n \right\},
\]  

(19)

\[
P \left\{ \sup_{θ ∈ Θ} |\hat{T}_n(θ)| ≥ λ_s \right\} \leq \frac{E \exp \left\{ \sup_{θ ∈ Θ} \hat{T}_n(θ) \right\}}{\exp \{ελ_s\}} \leq \exp \left\{ -ελ_s + \left( c_1^2 1 + c_1 \right)^2 ε^2 / n \right\}.
\]  

(20)

Taking $ε = \frac{λ_s}{2(c_1^2 1 + c_1)^2}$ in Equations (19) and (20), we obtain:

\[
P \left\{ \sup_{θ ∈ Θ} |\hat{T}_n(θ)| ≥ λ_s \right\} \leq C \exp \left\{ -λ_s^2 n / 4 \left( c_1^2 1 + c_1 \right)^2 \right\},
\]

\[
P \left\{ \sup_{θ ∈ Θ} |\hat{T}_n(θ)| ≥ λ_s \right\} \leq C \exp \left\{ -λ_s^2 n / 4 \left( c_1^2 1 + c_1 \right)^2 \right\}.
\]

Note that $\sup_{θ ∈ Θ} |\hat{T}_n(θ)| ≤ \sup_{θ ∈ Θ} ( |\hat{T}_n(θ) - T_n(θ)| ) ≤ \sup_{θ ∈ Θ} |\hat{T}_n(θ)| + \sup_{θ ∈ Θ} |T_n(θ)|$; hence:

\[
P \left\{ \sup_{θ ∈ Θ} T_n(θ) ≥ λ_s \right\} \leq P \left\{ \sup_{θ ∈ Θ} |\hat{T}_n(θ)| ≥ λ_s / 2 \right\} + P \left\{ \sup_{θ ∈ Θ} |\hat{T}_n(θ)| ≥ λ_s / 2 \right\}
\]

\[
\leq C \exp \left\{ -λ_s^2 n / 8 \left( c_1^2 1 + c_1 \right)^2 \right\}.
\]

This is the result of Equation (15) and thus concludes the proof of Theorem 1. □
Author Contributions: R.Y. compiled the manuscript with support from X.L. and Y.Y. All authors read and approved the final manuscript.

Funding: Xinsheng Liu was supported by National Natural Science Foundation of China (No. 61374183). Renyu Ye was supported by the Key Project Foundation of Natural Science Research in Universities of Anhui Province in China (No. KJ2019A0557) and the Natural Science Foundation of Anhui Province of China (No.1908085MA01). Yuncai Yu was supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX19_0149).

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Efromovich, S. Density estimation for biased data. *Ann. Statist.* 2004, 32, 1137–1161.
2. Ramirez, P.; Vidakovic, B. Wavelet density estimation for stratified size-biased sample. *J. Stat. Plan. Inference* 2010, 140, 419–432.
3. Chesneau, C. Wavelet block thresholding for density estimation in the presence of bias. *J. Korean Stat. Soc.* 2010, 39, 43–53.
4. Guo, H.; Kou, J. Pointwise density estimation for biased sample. *J. Comput. Appl. Math.* 2019, 361, 444–458.
5. Alam, K.; Saxena, K.L. Positive dependence in multivariate distributions. *Commun. Stat. Theory Method.* 1981, 10, 1183–1196.
6. Joag-Dev, K.; Proschan, F. Negative association of random variables with applications. *Ann. Stat.* 1983, 11, 286–295.
7. Liang, H.Y. Complete convergence for weighted sums of negatively associated random variables. *Statist. Probab. Lett.* 2000, 48, 317–325.
8. Matula, P. A note on the almost sure convergence of sums of negatively dependent random variables. *Statist. Probab. Lett.* 1992, 5, 209–212.
9. Roussas, G.G. Asymptotic normality of random fields of positively or negatively associated processes. *J. Multivariate Anal.* 1994, 50, 152–173.
10. Chen, P.; Sung, S.H. On the strong convergence for weighted sums of negatively associated random variables. *Statist. Probab. Lett.* 2014, 92, 45–52.
11. Miao, Y.; Xu, W.; Chen, S. Some limit theorems for negatively associated random variables. *Pro. Math. Sci.* 2014, 124, 447–456.
12. Wu, Y. On complete moment convergence for arrays of rowwise negatively associated random variables. *RACSAM* 2014, 108, 669–681.
13. Chesneau, C.; Dewan, I.; Doosti, H. Wavelet linear density estimation for associated stratified size-biased sample. *J. Nonparametr. Stat.* 2012, 2, 429–445.
14. Liu, Y.M.; Xu, J.L. Wavelet density estimation for negatively associated stratified size-biased sample. *J. Nonparametr. Stat.* 2014, 26, 537–554.
15. Guo, H.J.; Kou, J.K. Pointwise density estimation based on negatively associated data. *J. Inequal. Appl.* 2019, 206, 1–16.
16. Cai, T.; Chicken, E. Block thresholding for density estimation: Local and global adaptivity. *J. Multivariate Anal.* 2005, 95, 76–106.
17. Brown, L.; Cai, T.; Zhang, R.; Zhao, L. The root-unroot algorithm for density estimation as implemented via wavelet block thresholding. *Probab. Theory Related Fields* 2010, 146, 401–433.
18. Chesneau, C. Wavelet estimation via block thresholding: A minimax study under the $L^p$ risk. *Statist. Sinica.* 2008, 18, 1007–1024.
19. Donoho, D.L.; Johnstone, I.M. Minimax estimation via wavelet shrinkage. *Ann. Stat.* 1998, 26, 879–921.
20. Shao, Q.M. A comparison theorem on maximum inequalities between negatively associated and independent random variables. *J. Theoret. Probab.* 2000, 13, 343–356.