On the integrated squared error of the linear wavelet density estimator

Lu Lu*

Department of Mathematics, University of Connecticut, Storrs, CT, USA, 06269

Abstract
Linear wavelet density estimators are wavelet projections of the empirical measure based on independent, identically distributed observations. We study here the law of the iterated logarithm (LIL) and a Berry-Esseen type theorem. These results are proved under different assumptions on the density $f$ that are different from those needed for similar results in the case of convolution kernels (KDE): whereas the smoothness requirements are much less stringent than for the KDE, Riemann integrability assumptions are needed in order to compute the asymptotic variance, which gives the scaling constant in LIL. To study the Berry-Esseen type theorem, a rate of convergence result in the martingale CLT is used.

Keywords: linear wavelet density estimation, law of the iterated logarithm, integrated squared error, Berry-Esseen type theorem

2000 MSC: 62G07, 60F05, 60F15

1. Introduction
Let $X, X_1, X_2, ...$ be i.i.d random variables in $\mathbb{R}$ with common Lebesgue density $f$. Let $\phi \in L_2(\mathbb{R})$ be a scaling function and $\psi$ the corresponding wavelet function. Let $\phi_{0k} := \phi(x - k)$ and $\psi_{jk} := 2^{j/2} \psi(2^j x - k)$. \{\phi_{0k}, \psi_{jk}\} forms an orthonormal system in $L_2(\mathbb{R})$. Every $f \in L_p(\mathbb{R})$ has a formal expansion

$$f(x) = \sum_k \alpha_{0k} \phi_{0k}(x) + \sum_{j=0}^{\infty} \sum_k \beta_{jk} \psi_{jk}(x). \quad (1.1)$$

The linear wavelet density estimator is defined as

$$\hat{f}_n(x) = \sum_k \hat{\alpha}_{0k} \phi_{0k}(x) + \sum_{j=0}^{j_n-1} \sum_k \hat{\beta}_{jk} \psi_{jk}(x), \quad (1.2)$$

where $j_n$ is a sequence of integers. $\hat{\alpha}_{jk}$ and $\hat{\beta}_{jk}$ are constructed by the plug-in method. Let $P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$ be the empirical measure corresponding to the sample $\{X_i\}_{i=1}^{n}, n \in \mathbb{N}$.

*Principal Corresponding Author
Email address: lu.lu@uconn.edu (Lu Lu)

Preprint submitted to Elsevier May 10, 2014
Then
\[ \hat{\alpha}_{jk} = P_n(\phi_{jk}) = \frac{1}{n} \sum_{i=1}^{n} 2^{j/2} \phi(2^j X_i - k), \quad (1.3) \]
\[ \hat{\beta}_{jk} = P_n(\psi_{jk}) = \frac{1}{n} \sum_{i=1}^{n} 2^{j/2} \psi(2^j X_i - k). \quad (1.4) \]
They are unbiased estimators of \( \alpha \) and \( \beta \).

The use of this estimator first appeared in Doukhan and León (1990) and Kerkyacharian and Picard (1992). When \( \phi \) satisfies certain properties, i.e., bounded and compactly supported, one may write \( \hat{f}_n(x) \) in a form similar to that of the classical kernel density estimator:
\[ f_{n,K}(x) := \hat{f}_n(x) = \frac{2^{j_n}}{n} \sum_{i=1}^{n} K(2^{j_n} x, 2^{j_n} X_i), \quad (1.5) \]
where the projection kernel \( K(x, y) \) is given by
\[ K(x, y) = \sum_{k \in \mathbb{Z}} \phi(x - k) \phi(y - k). \quad (1.6) \]
\( \{2^{-j_n}\} \) is playing the role of the bandwidth in the classical kernel density estimation, and the sum is finite for each \( x \) and \( y \). By Lemma 8.6, Härdle, Kerkyacharian, Picard and Tsybakov (HKPT, 1998), \( K(x, y) \) is majorized by a convolution kernel \( \Phi(x - y) \) such that
\[ |K(x, y)| \leq \Phi(x - y), \quad (1.7) \]
where \( \Phi : \mathbb{R} \rightarrow \mathbb{R}^+ \) is a bounded, compactly supported and symmetric function.

A widely accepted measure of performance of an estimator is its mean integrated squared error, which is the expected value of the integrated squared error or \( L_2 \) error defined by \( I_n := \int (f_n(x) - f(x))^2 dx \) (see, e.g., Bowman 1985). The integrated squared error \( I_n \) constitutes in itself a nice global measure of approximation of the density. And it is of interest to obtain the asymptotically exact almost sure rate of approximation, in this measure, of the density by an estimator of interest, often a law of the iterated logarithm. This was done by Giné and Mason (2004) for kernel density estimators, and it is done here for wavelet density estimators. We will refer to several results by Giné and Mason (2004), which will be abbreviated as (GM) in what follows. This type of theorems may be thought of as companion results to central limit theorems: whereas the latter gives rate of approximation in probability, the former deals with a.s. rate of convergence. The central limit theorem for the integrated squared error \( I_n \) was obtained by Hall (1984) for kernel density estimators, and by Zhang and Zheng (1999) for wavelet density estimators. We also prove a Berry-Esseen type theorem as a complement to Zhang and Zheng’s result. Doukhan and León (1993) obtained a bound on the rate of convergence in the CLT for generalized density projection estimates with respect to Prohorov’s metric. However, their bound does not apply to the optimal window width.

To study the integrated square error for the wavelet density estimator, we shall impose the following conditions:
\( (f) \): \( f(x) \) is bounded.
The scaling function $\phi$ is bounded and compactly supported (e.g., Daubechies wavelet). Then, in (1.7), we can assume $\Phi$ is supported on $[-A,A]$ for some $A > 0$. Set $\theta_\phi(x) = \sum_k |\phi(x-k)|$. (S1) also guarantees that (see section 8.5, HKPT, 1998),

$$\text{ess sup}_x \theta_\phi(x) < \infty.$$  \hfill (1.8)

(S2): $\|\phi\|_v < \infty$, where $\| \cdot \|_v$ denotes the total variation norm of $\phi$.

The bandwidth $\{2^{-j_n}\}$ satisfies

(B1): $j_n \to \infty, \ 2^{-j_n} \asymp n^{-\delta}$ for some $\delta \in (0,1/3)$, where $a_n \asymp b_n$ means $0 < \lim \inf a_n/b_n < \lim \sup a_n/b_n < \infty$.

(B2): There exists an increasing sequence of positive constants $\{\lambda_k\}_{k \geq 1}$ satisfying

$$\lambda_{k+1}/\lambda_k \to 1, \ \log \log \lambda_k/\log k \to 1, \ \lambda_{k+1} - \lambda_k \to \infty$$ \hfill (1.9)

as $k \to \infty$, such that $2^{-j_n}$ is constant for $n \in [\lambda_k, \lambda_{k+1}), k \in \mathbb{N}$. For instance, the sequence $\lambda_k = \exp(k/\log(e+k))$ satisfies these conditions.

We will prove the following theorems for the statistic $J_n := \|f_{n,K} - f\|^2_2 - \mathbb{E}\|f_{n,K} - f\|^2_2$.

**Theorem 1.1.** Let $f, \phi$ and $j_n$ satisfy hypotheses (f), (S1), (S2), (B1) and (B2). Set $\sigma^2 := 2\int_{\mathbb{R}} f^2(x)dx$. Then,

$$\lim_{n \to \infty} \sup_{t} \frac{n2^{-j_n/2}}{\sigma \sqrt{2 \log \log n}} J_n = 1, \ a.s.$$ \hfill (1.10)

**Theorem 1.2.** Assume the hypotheses (f), (S1), (B1) and that there exists $L \geq 0$ such that $f$ is Hölder continuous with exponent $0 < \alpha \leq 1$ on $[-L,L]$: $f$ is monotonically increasing on $(-\infty,-L]$ and monotonically decreasing on $[L,\infty)$. Let $Z \sim \mathcal{N}(0,1)$. Then there exists a constant $C$ (depending on $f, \phi$ and $\{j_n\}$), such that

$$\sup_t |\Pr\{n2^{-j_n/2}J_n \leq t\} - \Pr\{\sigma Z \leq t\}| \leq C(n^{-3\delta/16} \vee n^{-\alpha \delta} \sqrt{\log n})$$ \hfill (1.11)

where $\sigma^2 = 2\int_{\mathbb{R}} f^2(x)dx$.

For example, if $2^{-j_n} \asymp n^{-1/5}$, $\sup_t |\Pr\{n2^{-j_n/2}J_n \leq t\} - \Pr\{\sigma Z \leq t\}| \leq C(n^{-3/80} \vee n^{-\alpha /5} \sqrt{\log n})$. No claim of optimality of the rate obtained is made.

Zhang and Zheng (1999) used the fact that $J_n$ coincides with its stochastic part, $\tilde{J}_n$, where

$$\tilde{J}_n := \|f_{n,K} - \mathbb{E}f_{n,K}\|^2_2 - \mathbb{E}\|f_{n,K} - \mathbb{E}f_{n,K}\|^2_2.$$ \hfill (1.13)

This is due to the orthogonality of the wavelet basis. We will include a short proof later for completeness. Thus, there is no need to analyze the bias part and assume more regularity conditions on the density $f$ as is done in the kernel case (e.g., Hall, 1984; GM, 2004).
Next we set up some notations. Let $K$ be the projection kernel associated with the scaling function $\phi$ as in (1.6). Set

$$K_n(t, x) := K(2^{jn}t, 2^{jn}x)$$

and

$$\bar{K}_n(t, x) := K_n(t, x) - \mathbb{E}K_n(t, X).$$

Then by (1.13),

$$\bar{J}_n = \frac{2^{2jn}}{n^2} \left[ \int_{\mathbb{R}} \left( \sum_{i=1}^{n} \bar{K}(2^{jn}t, 2^{jn}X_i) \right)^2 dt - \mathbb{E} \int_{\mathbb{R}} \left( \sum_{i=1}^{n} \bar{K}(2^{jn}t, 2^{jn}X_i) \right)^2 dt \right]$$

where

$$W_n(F) := \int_{F} \left( \sum_{i=1}^{n} \bar{K}(2^{jn}t, 2^{jn}X_i) \right)^2 dt - \mathbb{E} \int_{F} \left( \sum_{i=1}^{n} \bar{K}(2^{jn}t, 2^{jn}X_i) \right)^2 dt = U_n(F) + L_n(F),$$

$$U_n(F) = \sum_{1 \leq i \neq j \leq n} \int_{F} \bar{K}_n(t, X_i)\bar{K}_n(t, X_j)dt, \quad L_n(F) = \sum_{i=1}^{n} \int_{F} (\bar{K}^2_n(t, X_i) - \mathbb{E}\bar{K}^2_n(t, X)) dt. \tag{1.16}$$

The measurable set $F$ will normally be $\mathbb{R}$, $[-M, M]$ or $[-M, M]^C$, $M > 0$, with $\int_{F} f(t)dt > 0$. But in the results below, $F$ can be any set with this property and such that $\lambda(\{x + y : x \in F, |y| < \varepsilon \} \cap F^c) \to 0$ as $\varepsilon \to 0$. \tag{1.17}

The proof of Theorem 1.1 for the most part follows the same pattern in (GM): For some $M$ large enough, $W_n([-M, M]^C)$ is shown to be negligible by using an exponential inequality for degenerate $U$-statistics (Giné, Latała and Zinn, 2000) and Bernstein’s inequality for the diagonal term. Therefore, we may truncate $\bar{J}_n$ and deal with $W_n([-M, M])$. This is approximated by a Gaussian chaos using strong approximations (Komlós-Major-Tusnády inequality) and a moderate deviation is proved for it. Finally, one deals with the usual blocking of laws of the iterated logarithm. Here it can be implemented again because of Bernstein type exponential inequalities for $U$-statistics. However, due to the fact that $K(x, y)$ is not a convolution kernel, the computation of the limiting variance turns out to be a major difficulty, which we surmount using ideas from the proof of CLT in Zhang and Zheng (1999). For this we require $f$ to be (improper) Riemann integrable on $\mathbb{R}$, and this is the purpose of condition (f).

In order to get the convergence rate in CLT, we need to assume more conditions on $f$. $\bar{J}_n$ is composed of $L_n(\mathbb{R})$ and $U_n(\mathbb{R})$. The exponential inequality for $U$-statistics is used to show $L_n(\mathbb{R})$ is negligible. Then $U_n(\mathbb{R})$ is approximated by a martingale and the rate of convergence was obtained using Erickson, Quine and Weber (1979)’s result. The $U$-statistics method and the application of the martingale limit theory can be traced back to Hall (1984). It makes the study of $L_2$ error easier, but it does not apply to $L_p$ error if $p \neq 2$. \[4\]
The article is organized as follows. In section 2 we collect the variance computation results. In section 3 we state results of tail estimation. In section 4, we obtain a moderate deviation result for $W_n([-M,M])$. In section 5, we complete the proofs of Theorem 1.1 and Theorem 1.2. In the appendix, we give proofs to some lemmas stated in section 2. $C$ is a universal constant which might differ from line to line.

2. Variance Computations

We present here some inequalities and variance computations used throughout the paper. Only the exact limits present problems and must be treated differently than in the case of convolution kernels, but upper bounds can be dealt with essentially as in the convolution kernel case because of the majorization property (1.7). We will state these results without giving detailed proofs. They can be verified by replacing, in the corresponding proofs by (GM), the bandwidth $h_n$ by $2^{-j_n}$ and the projection kernel $K(x,y)$ by a convolution kernel $\Phi(x-y)$ that is given by (1.7). More specifically, if the kernel $K(x,y)$ satisfies (1.7), we have the following estimates: For all $x$ and $y$, and all measurable sets $F$,

$$\int_F \bar{K}_n^2(t,x) dt \leq 4 \cdot 2^{-j_n} \|\Phi\|_2^2,$$

(2.1)

$$\left| \int_F \bar{K}_n^2(t,x) dt - \mathbb{E} \int_F \bar{K}_n^2(t,X) dt \right| \leq 8 \cdot 2^{-j_n} \|\Phi\|_2^2,$$

(2.2)

and by Cauchy-Schwarz,

$$\int_F |\bar{K}_n(t,x)\bar{K}_n(t,y)| dt \leq 4 \cdot 2^{-j_n} \|\Phi\|_2^2.$$

(2.3)

We have an analogue to Corollary 2.7, (GM).

**Corollary 2.1.** Assume (f), (S1) and (B1) hold, and that $F$ satisfies condition (1.17). Then there exists $n_0 = n_0(F)$ such that, for all $n \geq n_0$,

$$\text{Var} \int_F \bar{K}_n^2(t,X) dt \leq 8 \cdot 2^{-2j_n} \|\Phi\|_2^4 \int_F f(x) dx.$$

(2.4)

And for all $n$,

$$\text{Var} \int_F \bar{K}_n^2(t,X) dt \leq 4 \cdot 2^{-2j_n} \|\Phi\|_2^4.$$

(2.5)

Set

$$C_n(t,s) := 2^{j_n} \int_\mathbb{R} K_n(t,x)K_n(s,x)f(x)dx, \quad R_n(t,s) := 2^{j_n} \int_\mathbb{R} \bar{K}_n(t,x)\bar{K}_n(s,x)f(x)dx.$$

(2.6)

Define the operator $\mathcal{R}_{n,F}$ for $\varphi \in L_2(F)$,

$$\mathcal{R}_{n,F}\varphi(s) = \int_F R_n(s,t)\varphi(t) dt.$$

(2.7)

The next three lemmas are similar to Lemmas 2.3, 2.4 and 2.5, (GM).
Lemma 2.2. Under the hypotheses of Corollary 2.1, for the operator $\mathcal{R}_{n,F}$, we have

$$\sup\{\|\mathcal{R}_{n,F}\varphi\|_2^2 : \|\varphi\|_2 = 1, \varphi \in L_2(F)\} \leq 2^{-2n}C(\Phi, f), \quad (2.8)$$

where

$$C(\Phi, f) = 2\|\Phi\|_1^4 (\|f\|_\infty^2 + \|f\|_2^4). \quad (2.9)$$

Lemma 2.3. Under the hypotheses of Corollary 2.1,

$$\limsup_{n \to \infty} 2^{2n} \int_{F^2} C^2_n(s,t)dsdt \leq \int_F f^2(x)dx \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \Phi(w+u)\Phi(w)dw\right)^2 du \leq \int_F f^2(x)dx \|\Phi\|_1^2 \|\Phi\|_2^2. \quad (2.10)$$

Lemma 2.4. Under the hypotheses of Corollary 2.1,

$$2^{2n} \int_{F^2} (C_n(s,t) - R_n(s,t))^2dsdt \leq 2^{-2n}\|\Phi\|_1^4 \|f\|_2^4 \to 0 \quad \text{as} \quad n \to \infty. \quad (2.11)$$

Note that in Lemma 2.3, we can only get an upper bound instead of the limit using the same method from the convolution kernel case. Calculation of the exact limit of $2^{2n} \int_{[-M,M]^2} R_n^2(s,t)dsdt$ is the key to obtaining the scaling constant in LIL. By Lemma 2.4, we shall approximate it by $2^{2n} \int_{[-M,M]^2} C^2_n(s,t)dsdt$ and calculate the limit of this quantity.

Lemma 2.5. Assume (f) and (B1) holds, and the scaling function $\phi$ satisfies (S1) such that the kernel $K$ associated with $\phi$ is dominated by $\Phi$ whose support is contained in $[-A,A]$, where $A$ is an integer. Then for any $M > 0$,

$$\lim_{n \to \infty} 2^j \int_{[-M,M]^2} C^2_n(s,t)dsdt = \int_{-M}^M f^2(y)dy. \quad (2.12)$$

In order to prove Theorem 1.2, we need to estimate how fast $2^{2n} \int_{\mathbb{R}^2} C^2_n(s,t)dsdt$ converges to $\int_{\mathbb{R}} f^2(y)dy$. This can be done by imposing more regularity conditions on $f$.

Lemma 2.6. Under the hypotheses of Lemma 2.5, and assume that, in addition, $f$ is Hölder continuous with exponent $0 < \alpha \leq 1$ on $[-L,L]$, and monotone on tails $(-\infty, -L] \cup [L, \infty)$, where $L \geq 0$. Then for all $n$, there exists a constant $C$ (depending on $f$, $\phi$ and $\{j_n\}$), such that

$$\left|2^{2n} \int_{\mathbb{R}^2} C^2_n(s,t)dsdt - \int_{\mathbb{R}} f^2(y)dy\right| \leq Cn^{-\delta} \quad (2.13)$$

where $\delta \in (0, 1/3)$ is the same as in (B1).

Together with Lemma 2.3, we obtain

Corollary 2.7. Assume the same conditions in Lemma 2.6, for all $n$ sufficiently large depending on $f$ and $\{j_n\}$,

$$\left|2^{2n} \int_{\mathbb{R}^2} R^2_n(s,t)dsdt - \int_{\mathbb{R}} f^2(y)dy\right| \leq C(n^{-\delta/2} + n^{-\delta}), \quad (2.14)$$

where the constant $C$ depends on $f$, $\phi$ and $\{j_n\}$.

The proofs of Lemmas 2.5 and 2.6 are provided in the appendix.
3. Tail Estimation

The goal of this section is to obtain exponential inequalities for $W_n(F)$, where $F$ satisfies (1.17) and also for $W_n(\mathbb{R}) - W_{n,m}(\mathbb{R})$. We assume throughout this section that $\phi$ satisfies (S1), and $K$ is associated with $\phi$ given by (1.6).

Set, for $m < n$,

$$W_{n,m}(\mathbb{R}) := \int_{\mathbb{R}} \left[ \left( \sum_{m<i\leq n} \tilde{K}_n(t, X_i) \right)^2 - \mathbb{E} \left( \sum_{m<i\leq n} \tilde{K}_n(t, X_i) \right)^2 \right] dt$$

(3.1)

and

$$H_n(x, y) := \int_{\mathbb{R}} \tilde{K}_n(t, x) \tilde{K}_n(t, y) dt, \quad H_{n,F}(x, y) = \int_{F} \tilde{K}_n(t, x) \tilde{K}_n(t, y) dt.$$  

(3.2)

With this notation,

$$U_n(F) = \sum_{1 \leq i \neq j \leq n} H_{n,F}(X_i, X_j), \quad L_n(F) = \sum_{i=1}^{n} \left( H_{n,F}(X_i, X_i) - \mathbb{E} H_{n,F}(X_i, X_i) \right),$$

(3.3)

and

$$W_n(\mathbb{R}) - W_{n,m}(\mathbb{R}) = 2 \sum_{i=1}^{m} \sum_{j=m+1}^{n} H_n(X_i, X_j) + \sum_{1 \leq i \neq j \leq m} H_n(X_i, X_j) + \sum_{i=1}^{m} \left( H_n(X_i, X_i) - \mathbb{E} H_n(X_i, X_i) \right).$$

(3.4)

Bernstein’s inequality (e.g., de la Peña and Giné, 1999) says that for centered, i.i.d. random variables $\xi_i$, if $\|\xi_i\|_{\infty} \leq c < \infty$ and $\sigma^2 = \mathbb{E} \xi_i^2$, then

$$\Pr \left\{ \sum_{i=1}^{m} \xi_i > t \right\} \leq \exp \left( -\frac{t^2}{2m\sigma^2 + 2ct/3} \right).$$

(3.5)

Applying it to the 3rd term in the above equation, given Corollary 2.1 and inequality (2.2), we obtain

$$\Pr \left\{ \left| \sum_{i=1}^{m} \left( H_n(X_i, X_i) - \mathbb{E} H_n(X_i, X_i) \right) \right| > \tau n 2^{-\frac{j_n}{2}} \right\} \leq 2 \exp \left( -\frac{\tau^2 n 2^{2-3j_n}}{8m2^{-2j_n}\|\Phi\|_2^4 + \frac{16}{3} \tau n 2^{-\frac{5}{2}j_n}\|\Phi\|_2^2} \right).$$

(3.6)

The first two terms in (3.4) are of U-statistics type. They can be controlled by the following exponential inequality for canonical U-statistics.
Theorem 3.1. (Giné, Latała, Zinn, 2000) There exists a universal constant $L < \infty$ such that, if $h_{i,j}$ are bounded canonical kernels of two variables for the independent random variables $(X^{(1)}_i, X^{(2)}_j)$, $i, j = 1, 2, ..., n$, and if $A, B, C, D$ are as defined below, then

$$
\Pr \left\{ \left| \sum_{1 \leq i,j \leq n} h_{i,j}(X^{(1)}_i, X^{(2)}_j) \right| \geq x \right\} \leq L \exp \left[ -\frac{1}{L} \min \left( \frac{x^2}{C^2}, \frac{x^{2/3}}{D^{2/3}}, \frac{x^{1/2}}{A^{1/2}} \right) \right]
$$

(3.7)
for all $x > 0$, where

$$
D = \| (h_{i,j}) \|_{L^2 \rightarrow L^2}
$$

and

$$
A = \max_{i,j} \| h_{i,j} \|_\infty.
$$

(3.11)

Theorem 3.1 also holds if the decoupled U-statistic $\sum_{1 \leq i,j \leq n} h_{i,j}(X^{(1)}_i, X^{(2)}_j)$ is replaced by the undecoupled U-statistic $\sum_{1 \leq i \neq j \leq n} h_{i,j}(X_i, X_j)$. We will take $h_{i,j} = H_{n,F,i,j} = H_{n,F}$, calculate the constants $A, B, C, D$ in Theorem 3.1 and apply it to $\sum_{1 \leq i \neq j \leq m} H_{n,F}(X_i, X_j)$. Theorem 3.1 gives

$$
A \leq 4 \cdot 2^{-jn} \| \Phi \|_2^2, \quad B^2 \leq 16m \cdot 2^{-2jn} \| \Phi \|_2^4.
$$

(3.12)

By Lemmas 2.3 and 2.4 for $n$ large enough depending on $F$,

$$
C^2 \leq 2m^2 \cdot 2^{-3jn} \| \Phi \|_2^2 \| \Phi \|_2^2 \int_F f^2(x)dx.
$$

(3.13)

If $f$ satisfies condition (f) and $\phi$ satisfies condition (S1), the bound on $D$ can be calculated by following the proof in the kernel case and making obvious modifications there.

$$
D \leq 4m2^{-2jn} \| f \|_\infty \| \Phi \|_1^2.
$$

(3.14)

Proposition 3.2. Let $X_i$ be i.i.d. with density $f$ satisfying condition (f). Let $F$ be a measurable subset of $\mathbb{R}$ satisfying condition (1.17). $\phi$ satisfies (S1) and $K$ is the projection kernel associated with $\phi$. $2^{-jn} \rightarrow 0$. Then there exist constants $\kappa_0$ (depending on $f$ and $\phi$) and $n_0$ (depending on $F, f, \phi$ and the sequence $\{j_n\}$) such that, for all $\tau > 0$ and for all $n \geq n_0$, $0 \leq m < n$,

$$
\Pr \left\{ \left| \sum_{1 \leq i \neq j \leq m} H_{n,F}(X_i, X_j) \right| \geq \tau n2^{-jn} \right\} \leq \kappa_0 \exp \left( -\frac{1}{\kappa_0} \min \left( \frac{\tau^2 n^2}{m^2 \int_F f^2(x)dx}, \frac{\tau n}{m2^{-jn/2}}, \frac{\tau^{2/3}n^{2/3}2^{-jn/3}}{m^{1/3}}, \frac{\tau^{1/2}n^{1/2}2^{-jn/4}}{m} \right) \right)
$$

(3.15)
and
\[
\Pr \left\{ \left| \sum_{i=1}^{m} \sum_{j=m+1}^{n} H_{n,F}(X_i, X_j) \right| \geq \tau n 2^{-\frac{3}{2}j_n} \right\} 
\leq \kappa_0 \exp \left( -\frac{1}{\kappa_0} \min \left[ \frac{\tau^2 n^2}{m(n-m) \int_f f^2(x) dx}, \frac{\tau n}{\sqrt{m(n-m)2^{-j_n/2}}}, \frac{\tau^2 / 3 2^{-j_n/3} n^{1/3} 2^{-j_n / 4}}{(m \vee (n-m))^{1/3}}, \frac{\tau^{1/2} n^{1/2} 2^{-j_n / 4}}{\tau n^{2^{-j_n / 2}}}, \frac{\tau^{1/2} n^{1/2} 2^{-j_n / 4}}{\tau n^{2^{-j_n / 2}}} \right] \right) .
\] (3.16)

**Proof.** Gathering Theorem 3.1, (3.12), (3.13) and (3.14), we get (3.15). (3.16) can be obtained in a similar way.

Using this and (3.6) for the diagonal \( L_n(F) \), we also have

**Proposition 3.3.** Under the same hypotheses of Proposition 3.2 on \( f, \phi \) and \( \{j_n\} \), there exist constants \( \kappa_0 \) (depending on \( \phi \) and \( f \)) and \( n_0 \) (depending on \( F, f, \phi \) and the sequence \( \{j_n\} \)) such that, for all \( \tau > 0 \) and for all \( n \geq n_0 \),
\[
\Pr \left\{ |W_n(F)| \geq \tau n 2^{-\frac{3}{2}j_n} \right\} 
\leq \kappa_0 \exp \left( -\frac{1}{\kappa_0} \min \left[ \frac{\tau^2}{\int_f f^2(x) dx}, \frac{\tau n}{2^{j_n / 2} n^{1/3} 2^{-j_n / 4}}, \frac{\tau n}{\sqrt{2 \log \log n}}, \frac{\tau n}{\sqrt{2 \log \log n}} \right] \right) .
\] (3.17)

In particular, if the sequence \( 2^{j_n} \) satisfies condition (B1) and \( \tau = \eta \sqrt{\log \log n} \), the first term dominates. For every \( \eta > 0 \) there exist \( \kappa_0 \) and \( n_0 \) as above such that
\[
\Pr \left\{ |W_n(F)| \geq \eta n 2^{-\frac{3}{2}j_n} \sqrt{\log \log n} \right\} \leq \kappa_0 \exp \left( -\frac{\eta^2 \log \log n}{\kappa_0 \int_f f^2(x) dx} \right) .
\] (3.18)

for all \( n \geq n_0 \).

Now the three terms in the decomposition of \( W_n(F) - W_{n,m}(F) \) in (3.4) can be bounded. The first two are of the U-statistics type, so Proposition 3.2 is used to obtain the estimation. The last one is a sum of mean zero i.i.d. r.v.’s and can be dealt with by (3.6).

**Lemma 3.4.** Under the same hypotheses of Proposition 3.2 on \( f, \phi \) and \( \{j_n\} \), there exist a constant \( \kappa_0 \) (depending on \( f \) and \( \phi \)) and \( \eta > 0 \) such that, for all \( \epsilon > 0, \sigma > 0, \) if \( n \) is large enough (depending on \( f, \phi \) and \( \{j_n\} \)), and \( m \) fixed is such that \( 0 \leq m < n \),
\[
\Pr \left\{ \left| W_n(F) - W_{n,m}(F) \right| \geq \epsilon \sigma n 2^{-j_n / 3} \sqrt{2 \log \log n} \right\} \leq \kappa_0 \exp \left( -\frac{\epsilon^2 n^2}{\kappa_0 \sigma} \right) .
\] (3.19)
4. Moderate Deviations

In this section, we’ll prove a moderate deviation result for $W_n([-M, M])$. This statistic can be approximated by a Gaussian chaos due to the Komlós-Major-Tusnády (KMT) theorem and the Dvoretzky-Kiefer-Wolfowitz (DKW) inequalities. Then a moderate deviation result in (GM) is used for the Gaussian chaos. \( \phi \) satisfies both (S1) and (S2).

Let \( F_n(t) := \frac{1}{n} \sum_{i=1}^{n} 1(X_i \leq t) \) and \( B_n \) be a sequence of Brownian bridges. For all \( x \in \mathbb{R} \), set

\[
E_n(x) := \sqrt{n2^{-jn}}[f_{n,K}(x) - \mathbb{E}f_{n,K}(x)] = \sqrt{\frac{2jn}{n}} \sum_{i=1}^{n} [K(2^{jn}x, 2^{jn}X_i) - \mathbb{E}K(2^{jn}x, 2^{jn}X)] \\
= \sqrt{n2^{jn}} \int_{\mathbb{R}} K(2^{jn}x, 2^{jn}t) d[f_n(t) - F(t)].
\]

Let \( K_{n,x}(t) := K(2^{jn}x, 2^{jn}t) \) and \( \mu_{K_{n,x}}(t) \) be the Borel measure associated with \( K_{n,x}(t) \). Define the Gaussian process

\[
\Gamma_n(x) := 2^{jn/2} \int_{\mathbb{R}} [B_n(F(x)) - B_n(F(t))] d\mu_{K_{n,x}}(t).
\]

We want to approximate

\[
\frac{2^{3jn/2}}{n} W_n([-M, M]) = 2^{jn/2} \int_{-M}^{M} [(E_n(t))^2 - \mathbb{E}(E_n(t))^2] \ dt
\]

by a Gaussian chaos:

\[
2^{jn/2} \int_{-M}^{M} [(\Gamma_n(t))^2 - \mathbb{E}((\Gamma_n(t))^2)] \ dt.
\]

In order to apply the KMT theorem, we need an integration by parts formula for \( E_n(x) \). This requires us to check two conditions: (i) \( F_n(t) - F(t) \) and \( K_{n,x}(t) \) are in the space \( NBV \), where \( NBV \) is defined by

\[
NBV = \{ G \text{ is of bounded variation, } G \text{ is right continuous and } G(-\infty) = 0 \}.
\]

(ii) Almost surely, for fixed \( N \), there are no points in \([-N, N]\) where \( F_n(t) - F(t) \) and \( K_{n,x}(t) \) are both discontinuous.

For any \( m \in \mathbb{N} \), let \( \{-\infty < t_0 < ... < t_m = t\} \) be a partition over \((-\infty, t)\). Then

\[
\sum_{l=1}^{m} |K_{n,x}(t_l) - K_{n,x}(t_{l-1})| \leq \sum_{k} |\phi(2^{jn}x - k)| \sum_{l=1}^{m} |\phi(2^{jn}t_l - k) - \phi(2^{jn}t_{l-1} - k)| \\
\leq \sum_{k} |\phi(2^{jn}x - k)||\phi(2^{jn} \cdot -k)||_v.
\]

Since \( \phi \) satisfies (1.8) and (S2), we have, for almost every \( x \),

\[
||K_{n,x}||_v \leq \sum_{k} |\phi(2^{jn}x - k)||\phi||_v := C_{\phi},
\]
where $C_\phi$ is a constant that depends only on the scaling function $\phi$. The other conditions in (i) are obvious. To verify (ii), we note that $K_{n,x}(t)$ could only have discontinuities at dyadic points whereas $F_n(t) - F(t)$ could only have discontinuities at $X_i$, $1 \leq i \leq n$.

Then we apply an integration by parts formula (Ex. 3.34, Folland 1999) to the integral $\int_{[-N,N]} K_{n,x}(t)d[F_n(t) - F(t)]$ and let $N \to \infty$. By dominated convergence, this gives

$$\int_{\mathbb{R}} K_{n,x}(t)d[F_n(t) - F(t)] + \int_{\mathbb{R}} (F_n(t) - F(t))d\mu_{K_{n,x}}(t) = 0. \quad (4.8)$$

Moreover, since $\int_{\mathbb{R}} d\mu_{K_{n,x}}(t) = 0$,

$$E_n(x) = \sqrt{n}2^{j_n} \int_{\mathbb{R}} [F(t) - F_n(t) - (F(x) - F_n(x))]d\mu_{K_{n,x}}(t). \quad (4.9)$$

Now we are able to bound the difference between (4.3) and (4.4). We set $\alpha_n(t) := \sqrt{n}[F_n(t) - F(t)]$ and $D_n := \sup_{-\infty < t < \infty} |\alpha_n(t) - B_n(F(t))|$. We have

$$D_n(M) := \left| \frac{2^{3j_n/2}}{n} W_n([-M, M]) - 2^{j_n/2} \int_{-M}^{M} ((\Gamma_n(t))^2 - E((\Gamma_n(t))^2)) dt \right| \leq 2^{j_n} \cdot 4MD_nC_\phi \text{ ess sup}_{x} (|E_n(x)| + |\Gamma_n(x)|) \quad (4.10)
$$

$$\leq 2^{3j_n/2}8MD_n(\|\alpha_n\|_\infty + \|B_n\|_\infty)C_\phi^2.$$

We use the KMT theorem for $D_n$ and the DKW inequalities for $\|\alpha_n\|_\infty$ and $\|B_n\|_\infty$.

**Theorem 4.1.** (Komlós, Major, Tusnády, 1975) There exists a probability space $(\Omega, \mathcal{A}, P)$ with i.i.d random variables $X_1, X_2, \ldots$, with density $f$ and a sequence of Brownian bridges $B_1, B_2, \ldots$, such that, for all $n \geq 1$ and $x \in \mathbb{R}$,

$$\Pr\left\{ D_n \geq n^{-1/2}(a \log n + x) \right\} \leq b \exp(-cx), \quad (4.11)$$

where $a, b$ and $c$ are positive constants that do not depend on $n$, $x$ or $f$.

The DKW inequalities (Dvoretzky, Kiefer, Wolfowitz, 1956; or see Shorack and Wellner, 1986) give that, for every $z > 0$,

$$\Pr\left\{ \|\alpha_n\|_\infty > z \right\} \leq 2 \exp(-2z^2), \quad \Pr\left\{ \|B_n\|_\infty > z \right\} \leq 2 \exp(-2z^2). \quad (4.12)$$

We arrive at the following proposition.

**Proposition 4.2.** Assuming the scaling function $\phi$ satisfies (S1), (S2) and $j_n$ satisfies (B1), for any $\gamma > 0$ there exists $C_{M,\phi} > 0$ such that

$$\Pr\left\{ D_n(M) \geq \frac{C_{M,\phi}(\log n)^2}{2^{-3j_n/2}\sqrt{n}} \right\} \leq n^{-\gamma} \quad (4.13)$$

for all $n > n_0(\gamma)$. 

11
Proof. For $\gamma > 0$, take $x = 2\gamma \log n / c$ in (4.11). If $n$ is sufficiently large depending on $\gamma$,

$$
\Pr \left\{ D_n \geq \frac{1}{\sqrt{n}} \left( a + \frac{2\gamma}{c} \right) \log n \right\} \leq b \exp (-2\gamma \log n) \leq \frac{1}{2} n^{-\gamma}.
$$

(4.14)

From DKW inequalities (4.12), it is easy to see that for $n$ large enough,

$$
\Pr \left\{ \|\alpha_n\|_\infty + \|B_n\|_\infty > \frac{\log n}{a + 2\gamma/c} \right\} \leq \frac{1}{2} n^{-\gamma}.
$$

(4.15)

Combining these with (4.10), we get

$$
\Pr \left\{ D_n \left( M \right) \geq 8MC^2_\phi (\log n)^2 \right\} \leq \Pr \left\{ D_n \geq \frac{1}{\sqrt{n}} \left( a + \frac{2\gamma}{c} \right) \log n \right\} + \Pr \left\{ \|\alpha_n\|_\infty + \|B_n\|_\infty > \frac{\log n}{a + 2\gamma/c} \right\}
$$

(4.16)

$$
\leq n^{-\gamma}.
$$

Setting $C_{M,\phi} = 8MC^2_\phi$ yields (4.13).

It is easier to obtain a moderate deviation result for $2^{3j_n/2}W_n([-M, M])n$ than for $2^{3j_n/2}W_n([-M, M])n$. For the former we can adapt the method in (GM) where they obtain a moderate deviation result for similar random variables by adapting a method of Pinsky (Pinsky, 1966) to prove the LIL for sums of random variables with finite moments higher than 2. It is a well-known fact that $\int_{-M}^{M} \left( (\Gamma_n(t))^2 - E(\Gamma_n(t))^2 \right) dt$ can be written as a sum of weighted, centered chi-squared random variables (e.g., Proposition 4.3, GM, 2004). Recall the operator $R_{n,F}$ defined in (2.7). Let $\lambda_{n,1} \geq \lambda_{n,2} \geq \ldots \geq 0$ be the eigenvalues of the operator $R_{n,F}$ with $F = [-M, M]$. $Z_k$ are i.i.d $\mathcal{N}(0, 1)$. We then have

$$
\int_{F} \left[ (\Gamma_n(t))^2 - \mathbb{E}(\Gamma_n(t))^2 \right] dt = \sum_{k=1}^{\infty} \lambda_{n,k} (Z_k^2 - 1).
$$

(4.17)

The limiting variance is calculated using Lemmas 2.4, 2.5

$$
\lim_{n \to \infty} 2^{j_n} \mathbb{E} \left[ \int_{-M}^{M} \left( (\Gamma_n(t))^2 - E(\Gamma_n(t))^2 \right) dt \right]^2 = \lim_{n \to \infty} 2 \cdot 2^{j_n} \sum_{k=1}^{\infty} \lambda_{n,k}^2
$$

$$
= \lim_{n \to \infty} 2 \cdot 2^{j_n} \int_{-M}^{M} \int_{-M}^{M} R_n^2(s,t) ds dt
$$

$$
= 2 \int_{-M}^{M} f^2(x) dx =: \sigma^2(M).
$$

(4.18)

Set $b_n := \left( \lambda_{n,1} / \sqrt{\sum_{k=1}^{\infty} \lambda_{n,k}^2} \right)^{\eta}$ for some $0 < \eta \leq 1$ and

$$
V_n(M) := \frac{2^{3j_n/2}}{\sigma(M)} \int_{-M}^{M} \left( (\Gamma_n(t))^2 - E(\Gamma_n(t))^2 \right) dt.
$$

(4.19)
Using (4.17) and a modification of Pinsky’s method, we have a moderate deviation for $V_n(M)$, which is parallel to (4.15), (GM). For any sequence $a_n$ converging to infinity at the rate $a_n^2 + \log b_n \to -\infty$ and for all $0 < \epsilon < 1$,

$$\exp \left( -\frac{a_n^2(1 + \epsilon)}{2} \right) \leq \Pr \{ \pm V_n(M) \geq a_n \} \leq \exp \left( -\frac{a_n^2(1 - \epsilon)}{2} \right)$$

if $n$ is large enough depending on $\epsilon$.

We can use this result, the triangle inequality and Proposition 4.2 to obtain:

**Proposition 4.3.** Let $a_n = C\sqrt{2 \log \log n}$, $0 < C < \infty$. Under the hypotheses of Proposition 4.2 and further assuming that $f$ satisfies condition (f) and that $\int_M^M f^2(x)dx > 0$, then we have a two-sided inequality,

$$\exp \left( -\frac{a_n^2(1 + \epsilon)}{2} \right) - \frac{1}{n^2} \leq \Pr \left\{ \pm \frac{2^{3j_n/2}}{\sigma(M)n} W_n([-M, M]) \geq a_n \right\} \leq \exp \left( -\frac{a_n^2(1 - \epsilon)}{2} \right) + \frac{1}{n^2}$$

for all $0 < \epsilon < 1$ and $n$ large enough (depending on $M$ and $\epsilon$).

### 5. Main Proofs

#### 5.1. Theorem 1.1

**Proof.** We show that $J_n = \bar{J}_n$, where $\bar{J}_n$ is defined in (1.13). Since we have,

$$J_n = \int_{\mathbb{R}} f^2_{n,K} - \mathbb{E}f^2_{n,K} - 2f f_{n,K} + 2f \mathbb{E}f_{n,K},$$

and

$$\bar{J}_n = \int_{\mathbb{R}} f^2_{n,K} - 2f f_{n,K} \mathbb{E}f_{n,K} - \mathbb{E}f^2_{n,K} + 2(\mathbb{E}f_{n,K})^2.$$

It remains to show that the difference

$$J_n - \bar{J}_n = 2 \int_{\mathbb{R}} (f - \mathbb{E}f_{n,K})(\mathbb{E}f_{n,K} - f_{n,K}) = 0.$$

$\mathbb{E}f_{n,K} - f_{n,K}$ is a linear combination of $\{\phi_{0k}\}$ and $\{\psi_{jk}\}$, $0 \leq j \leq j_n - 1$, whereas $f - \mathbb{E}f_{n,K}$ is a linear combination of $\{\psi_{jk}\}$, $j \geq j_n$. By orthogonality of $\{\phi_{0k}, \psi_{jk}\}$, we have $J_n - \bar{J}_n = 0$. Thus the proof of Theorem 1.1 reduces to proving that

$$\limsup_{n \to \infty} \frac{n^{2 - j_n/2}}{\sigma \sqrt{2 \log \log n}} \bar{J}_n = 1, \quad a.s.$$ (5.4)

By (1.14), this is equivalent to

$$\limsup_{n \to \infty} \frac{2^{3j_n/2} W_n(\mathbb{R})}{n \sigma \sqrt{2 \log \log n}} = 1.$$ (5.5)

Since we have analogous variance computation, tail estimation and moderate deviation results to those for the kernel density estimator, the proof is the same as in Theorem 5.1, (GM). We give an outline of the proof but readers should refer to (GM) for details.
Thus, satisfies the same properties as $\lambda$. By Borel-Cantelli, there exists $c < 1$ such that
\[ \limsup_{k} \frac{W_{r_k}(\mathbb{R})}{\sigma r_k 2^{-3j r_k / 2} \sqrt{2 \log \log r_k}} = c \quad a.s. \] (5.6)
The proof of Lemma 3.4 also applies to $W_{r_k}(\mathbb{R}) - W_{r_k r_{k-1}}(\mathbb{R})$ since $r_k / r_{k-1} \geq k$. And we have
\[ \frac{|W_{r_k}(\mathbb{R}) - W_{r_k r_{k-1}}(\mathbb{R})|}{r_k \sigma 2^{-3j r_k / 2} \sqrt{2 \log \log r_k}} \to 0 \quad a.s. \] (5.7)
Thus
\[ \limsup_{k} \frac{W_{r_k r_{k-1}}(\mathbb{R})}{\sigma r_k 2^{-3j r_k / 2} \sqrt{2 \log \log r_k}} = c \quad a.s. \] (5.8)
By Borel-Cantelli, there exists $c'$ satisfying $c < c' < 1$, s.t.
\[ \sum_k \Pr \left\{ \frac{W_{r_k r_{k-1}}(\mathbb{R})}{\sigma r_k 2^{-3j r_k / 2} \sqrt{2 \log \log r_k}} \geq c' \right\} < \infty. \] (5.9)
Set $m_k := r_k - r_{k-1}$ and define
\[ W_{m_k}^{'}(\mathbb{R}) := \int_{\mathbb{R}} \left( \sum_{i=1}^{r_k - r_{k-1}} K(2^{jr_k} t, 2^{jr_k} X_i) \right)^2 dt - E \int_{\mathbb{R}} \left( \sum_{i=1}^{r_k - r_{k-1}} K(2^{jr_k} t, 2^{jr_k} X_i) \right)^2 dt. \] (5.10)
Since $W_{m_k}^{'}(\mathbb{R})$ and $W_{r_k r_{k-1}}(\mathbb{R})$ have the same distribution, (5.9) holds with $W_{r_k r_{k-1}}(\mathbb{R})$ replaced by $W_{m_k}^{'}(\mathbb{R})$. This and $m_k / r_k \to 1$ imply that there exists $c''$ satisfying $c' < c'' < 1$, s.t.
\[ \sum_k \Pr \left\{ W_{m_k}^{'}(\mathbb{R}) \geq c'' \sigma m_k 2^{-3j r_k / 2} \sqrt{2 \log \log m_k} \right\} < \infty. \] (5.11)
We choose $M$ large enough so that $\int_{[-M,M]} f^2(x)dx < (\delta c'' / \kappa_0)$, where $\kappa_0$ is the constant in (3.18). $W_{m_k}^{'}(\mathbb{R})$ can be split into $W_{m_k}^{'}([-M,M])$ and $W_{m_k}^{'}([-M,M]^c)$. (3.18) is used for $W_{m_k}^{'}([-M,M]^c)$ and Proposition 4.3 for $W_{m_k}^{'}([-M,M])$. Then we would reach a contradiction to (5.11) and thus prove the lower bound.

(ii) Proof of the upper bound: We shall first use conditions (B1) and (B2) to introduce a blocking and reduce $W_n(\mathbb{R})$ to $W_{n_k}(\mathbb{R})$ for the sequence $n_k := \min \{ n \in \mathbb{N} : n \geq \lambda_k \}$. $n_k$ satisfies the same properties as $\lambda_k$ does. $I_k$ is the block defined by $I_k := [n_k, n_{k+1}) \cap \mathbb{N}$. $I_k$ is nonempty for $k \geq k_0$.

By Borel-Cantelli, it suffices to show that, for every $\delta > 0$,
\[ \sum_{k \geq k_0} \Pr \left\{ \max_{n \in I_k} |W_n(\mathbb{R})| > (1 + \delta) \sigma n_k 2^{-3j n_k / 2} \sqrt{2 \log \log n_k} \right\} < \infty. \] (5.12)
We will prove that for every $\tau > 0$,
\[ \sum_{k \geq k_0} \Pr \left\{ \max_{n \in I_k} |W_n(\mathbb{R}) - W_{n_k}(\mathbb{R})| > \tau \sigma n_k 2^{-3j n_k / 2} \sqrt{2 \log \log n_k} \right\} < \infty. \] (5.13)
For $n \in I_k$, similar to (3.4), we have
\[
W_n(\mathbb{R}) - W_{n_k}(\mathbb{R}) = 2 \sum_{i=1}^{n_k} \sum_{j=n_k+1}^{n} H_{n_k}(X_i, X_j) + \sum_{n_k < i \neq j \leq n} H_{n_k}(X_i, X_j) + \sum_{i=n_k+1}^{n} \left( H_{n_k}(X_i, X_i) - \mathbb{E}H_{n_k}(X_i, X_i) \right).
\]  
(5.14)

$H_n$ is replaced by $H_{n_k}$ since $\{2^{-j_n}\}$ is constant for $n \in I_k$ by hypothesis. We will apply Montgomery-Smith maximal inequality (Montgomery-Smith, 1993) to the first and the last summands directly: If $X_i$ are i.i.d. r.v.’s taking values in a Banach space and $\| \cdot \|$ is a norm in the Banach space, then
\[
\Pr \left\{ \max_{1 \leq k \leq n_k} \left\| \sum_{i=1}^{k} X_i \right\| > t \right\} \leq 9 \Pr \left\{ \left\| \sum_{i=1}^{n_k} X_i \right\| > \frac{t}{30} \right\}.
\]  
(5.15)

However, the second summand is not a sum of i.i.d random variables. A decoupling inequality (e.g., de la Peña and Giné, 1999, Theorem 3.4.1) is used to transform it into independent variables, i.e., $\sum_{n_k < i \neq j \leq n} H_{n_k}(X_i^{(1)}, X_j^{(2)})$, where $X_i^{(1)}$ and $X_j^{(2)}$, $i, j \in \mathbb{N}$ are i.i.d. copies of $X_1$. Then we add the diagonal, apply Montgomery-Smith inequality twice and subtract the diagonal at last. We will be able to reduce (5.13) to proving that, for every $\tau > 0$,
\[
\sum_{k \geq k_0} \Pr \left\{ \left| \sum_{j=n_k+1}^{n_{k+1}-1} \sum_{i=1}^{k} H_{n_k}(X_i, X_j) \right| > \tau \sigma n_k 2^{-3j_n}/\sqrt{2\log \log n_k} \right\} < \infty,
\]  
(5.16)

\[
\sum_{k \geq k_0} \Pr \left\{ \left| \sum_{i=n_k+1}^{k+1} \left( H_{n_k}(X_i, X_i) - \mathbb{E}H_{n_k}(X_i, X_i) \right) \right| > \tau \sigma n_k 2^{-3j_n}/\sqrt{2\log \log n_k} \right\} < \infty,
\]  
(5.17)

and
\[
\sum_{k \geq k_0} \Pr \left\{ \left| \sum_{j=n_k+1}^{n_{k+1}-1} \sum_{i=1}^{k} H_{n_k}(X_i^{(1)}, X_j^{(2)}) \right| > \tau \sigma n_k 2^{-3j_n}/\sqrt{2\log \log n_k} \right\} < \infty,
\]  
(5.18)

\[
\sum_{k \geq k_0} \Pr \left\{ \left| \sum_{i=n_k+1}^{k+1} H_{n_k}(X_i^{(1)}, X_i^{(2)}) \right| > \tau \sigma n_k 2^{-3j_n}/\sqrt{2\log \log n_k} \right\} < \infty.
\]  
(5.19)

(5.16), (5.17) come from the first and last summands in (5.14) whereas (5.18), (5.19) come from the second summand. We apply Bernstein’s inequality to (5.17) and (5.19). Proposition 3.2 will take care of (5.16) and (5.18). Therefore, (5.13) is proved. Thus (5.12) is reduced to showing that for every $\delta > 0$,
\[
\sum_{k \geq k_0} \Pr \left\{ |W_{n_k}(\mathbb{R})| > (1 + \delta)\sigma n_k 2^{-3j_n}/\sqrt{2\log \log n_k} \right\} < \infty.
\]  
(5.20)
The second step is to reduce $W_{nk}(R)$ to $W_{nk}([-M,M])$ for some $M$ large enough. Given $\delta > 0$, there exists $M < \infty$ such that $\int_{[-M,M]} f^2(x)dx < \delta^2 \sigma^2/(4\kappa_0)$, where $\kappa_0$ is the constant in inequality (3.18). Application of (3.18) gives that, from some $k_0$ on,

$$
\Pr \left\{ \left| W_{nk}([-M,M]) \right| > \delta \sigma n_k 2^{-3\delta n_k/2} \sqrt{2 \log \log n_k} \right\} \leq \kappa_0 \exp (-2 \log \log n_k), \quad (5.21)
$$

where the right hand side is the general term of a convergent series. Let $\epsilon$ be so small that $(1 + \delta/2)^2 (1 - \epsilon) > 1$. Now we use (4.21) to obtain that, for $n_k$ large enough,

$$
\Pr \left\{ \left| W_{nk}([-M,M]) \right| > (1 + \delta/2) \sigma n_k 2^{-3\delta n_k/2} \sqrt{2 \log \log n_k} \right\} \leq \exp \left( -(1 + \delta/2)^2 (1 - \epsilon) \log \log n_k \right) + \frac{1}{n_k^2},
$$

which is also the general term of a convergent series. Hence the series (5.20) converges for every $\delta > 0$.

$$
\text{5.2. Theorem 1.2}
$$

**Proof.** Without loss of generality, we will assume that, for all $n$, there exist constants $C_1$ and $C_2$, such that $C_1 n^\delta \leq 2^{j_n} \leq C_2 n^\delta$. Proving Theorem 1.2 is equivalent to proving that

$$
\sup_t \left| \Pr\{n2^{-j_n/2} \bar{J}_n \leq t\} - \Pr\{\sigma Z \leq t\} \right| \leq C(n^{-3\delta/16} + n^{-\alpha \delta} \sqrt{\log n}). \quad (5.23)
$$

By (1.14) and (1.15), we have that

$$
n2^{-j_n/2} \bar{J}_n/\sigma = \frac{2^{3j_n/2}}{n\sigma} W_n(\mathbb{R}) = \frac{2^{3j_n/2}}{n\sigma} U_n(\mathbb{R}) + \frac{2^{3j_n/2}}{n\sigma} L_n(\mathbb{R}). \quad (5.24)
$$

Using the triangle inequality, we can obtain an upper bound and a lower bound for this statistic. For an arbitrary positive sequence $\epsilon_{1,n}$,

$$
\sup_t \left| \Pr\{n2^{-j_n/2} \bar{J}_n/\sigma \leq t\} - \Pr\{Z \leq t\} \right| \\
\leq \sup_t \left| \Pr\left\{ \frac{2^{3j_n/2}}{n\sigma} U_n(\mathbb{R}) \leq t \right\} - \Pr\{Z \leq t\} \right| \\
+ \Pr\left\{ \frac{2^{3j_n/2}}{n\sigma} |L_n(\mathbb{R})| > \epsilon_{1,n} \right\} + \sup_t \Pr\{t - \epsilon_{1,n} < Z \leq t + \epsilon_{1,n}\}. \quad (5.25)
$$

It’s easy to bound the last term:

$$
\sup_t \Pr\{t - \epsilon_{1,n} < Z \leq t + \epsilon_{1,n}\} < \epsilon_{1,n}. \quad (5.26)
$$
By (3.6), for $0 < \epsilon_{1,n} \leq 1$ so that $\epsilon_{1,n}^2 \leq \epsilon_{1,n}$,

$$
\Pr \left\{ |L_n(\mathbb{R})| > \sigma \epsilon_{1,n} n^{2-3\delta/2} \right\} \leq C \exp \left( -\frac{1}{C} \min \left( \sigma^2 \epsilon_{1,n}^2 n^{2-\delta}, \sigma \epsilon_{1,n} n^{2-\delta/2} \right) \right)
\leq C \exp \left( -\frac{1}{C} \epsilon_{1,n}^2 n^{1-\delta} \right),
$$

(5.27)

where $C$ depends on both $\phi$ and $f$, $\delta \in (0, 1/3)$. We may take $\epsilon_{1,n} = n^{-1/3}$ to obtain

$$
\Pr \left\{ |L_n(\mathbb{R})| > \sigma \epsilon_{1,n} n^{2-3\delta/2} \right\} \leq C \exp ( - \log n ) = C n^{-1}
$$

(5.28)

when $n$ is large enough. Using (5.26), we get

$$
\sup_t \Pr \{ t - \epsilon_{1,n} < Z \leq t + \epsilon_{1,n} \} \leq n^{-1/3}.
$$

(5.29)

To control the first term in (5.25), we will approximate $2^{3\delta/2} U_n(\mathbb{R})/(n \sigma)$ by $S_{nn}$, which is defined below. We set

$$
U_{nn} := \sum_{i=2}^{n} \sum_{j=1}^{i-1} H_n(X_i, X_j), \quad s_n^2 := \mathbb{E}(U_{nn}^2),
$$

(5.30)

and

$$
X_{ni} := \frac{H_n(X_i, X_j)}{s_n}, \quad S_{nk} := \sum_{i=2}^{k} X_{ni},
$$

(5.31)

then

$$
S_{nn} = \sum_{i=2}^{n} \sum_{j=1}^{i-1} \frac{H_n(X_i, X_j)}{s_n}.
$$

(5.32)

Analogous to (5.25), for any positive sequence $\epsilon_{2,n}$,

$$
\sup_t \left| \Pr \left\{ \frac{2^{3\delta/2}}{n \sigma} U_n(\mathbb{R}) \leq t \right\} - \Pr \{ Z \leq t \} \right| 
\leq \sup_t \left| \Pr \{ S_{nn} \leq t \} - \Pr \{ Z \leq t \} \right| 
+ \Pr \left\{ \left| \frac{2^{3\delta/2}}{n \sigma} U_n(\mathbb{R}) - S_{nn} \right| > \epsilon_{2,n} \right\} + \sup_t \Pr \{ t - \epsilon_{2,n} < Z \leq t + \epsilon_{2,n} \}.
$$

(5.33)

By (5.3),

$$
\Pr \left\{ \left| \frac{2^{3\delta/2}}{n \sigma} U_n(\mathbb{R}) - S_{nn} \right| > \epsilon_{2,n} \right\} = \Pr \left\{ \left| \sum_{1 \leq i \neq j \leq n} H_n(X_i, X_j) \right| > \frac{\epsilon_{2,n}}{d_n} \right\},
$$

(5.34)

where $d_n = \left| \frac{2^{3\delta/2}}{n \sigma} - \frac{1}{2s_n} \right|$. We then estimate the order of $d_n$. Recall the definition of $R_n(s, t)$ in (2.7) and set $e_n := (2^{n} \int_{\mathbb{R}^2} R_n^2(s, t) ds dt)^{1/2}$. Using the definition of $s_n^2$ and Fubini’s theorem, we get

$$
s_n^2 = \sum_{i=2}^{n} \sum_{j=1}^{i-1} \mathbb{E}H_n^2(X_i, X_j) = \frac{n(n-1)}{2} 2^{-3\delta/2} \epsilon_n^2.
$$

(5.35)
Plugging it into $d_n$ and using a triangle inequality, we then have

$$d_n \leq C^2 \frac{3j_n}{2} \left| \frac{1}{\sqrt{n} \int f^2(x) dx} - \frac{1}{\sqrt{n(n - 1)} \int f^2(x) dx} \right| + C^2 \frac{3j_n}{2} \left| \frac{1}{\sqrt{n} \int f^2(x) dx} - \frac{1}{\sqrt{n(n - 1)} e_n} \right|. \quad (5.36)$$

Since $2^{j_n} \leq Cn^\delta$ for some $\delta \in (0, 1/3)$ and $1/\sqrt{n(n - 1)} - 1/n \leq n^{-2}$ when $n \geq 2$, the first term is bounded by $Cn^{3\delta/2 - 2}$. Corollary 2.1 gives that $|\epsilon_n^2 - \int f^2(x) dx| \leq C(n^{-\delta/2} + n^{-\alpha\delta})$. The second term is bounded by $Cn^{3\delta/2 - 1}(n^{-\delta/2} + n^{-\alpha\delta})$ when $n$ is large enough. Combining the two terms, $d_n \leq Cn^{3\delta/2 - 1}(n^{-\delta/2} + n^{-\alpha\delta})$, where $C$ depends on $f$, $\{j_n\}$ and $\phi$. Taking $\epsilon_{2n} = n^{-\delta(1/\alpha + \delta)} \sqrt{\log n}$ and using (5.34), Proposition 3.2, we obtain

$$\Pr \left\{ \left| \frac{2^{j_n}}{n\sigma^2} U_n(\mathbb{R}) - S_{nn} \right| > \epsilon_{2n} \right\} \leq \kappa_0 \exp (-\log n) = Cn^{-1} \quad (5.37)$$

when $n$ is large enough. Consequently,

$$\sup_t \Pr \{ t - \epsilon_{2n} < Z \leq t + \epsilon_{2n} \} \leq n^{-\delta(1/\alpha + \delta)} \sqrt{\log n}. \quad (5.38)$$

We then deal with $\sup_t |\Pr \{ S_{nn} \leq t \} - \Pr \{ Z \leq t \}|$. Let $F_i$ be the $\sigma$-field generated by $\{X_1, X_2, ..., X_i\}$ for $i = 1, 2, ....$. We first observe that, by the definitions in (5.30)-(5.32),

$$\mu_{ni} := \mathbb{E}(X_{ni}|\mathcal{F}_{i-1}) = 0, \quad (5.39)$$

and thus $S_{nk}$ is a martingale with respect to $\mathcal{F}_k$. We will use the result of Erickson, Quine and Weber (1979) to derive a bound for $\sup_t |\Pr \{ S_{nn} \leq t \} - \Pr \{ Z \leq t \}|$. For $i \geq 2$, let $X_{ni}' := X_{ni} - \mu_{ni}$; $\sigma_{ni}^2 := \mathbb{E}(X_{ni}'^2|\mathcal{F}_{i-1})$ and $\sigma_n^2 := \sum_{i=2}^n \sigma_{ni}^2$. Also define $Y_{ni} := \sum_{j=1}^{i-1} H_n(X_i, X_j)$ and $V_n := \sum_{i=2}^n \mathbb{E}(Y_{ni}^2|\mathcal{F}_{i-1})$.

**Theorem 5.1** (Erickson, Quine, Weber, 1979). Given $X = \{X_{ni}, i = 2, ..., n; n = 1, 2, ...\}$ and $\mathcal{F} = \{\mathcal{F}_i, i = 1, 2, ...\}$, let $S_{nn} := \sum_{i=2}^n X_{ni}$. If $\mu_{ni} = 0$ for all $n, i$, then for $\eta \in (0, 1]$, there exists a constant $C$,

$$\sup_t |\Pr \{ S_{nn} \leq t \} - \Pr \{ Z \leq t \}| \leq C \left\{ \sum_{i=2}^n \mathbb{E}|X_{ni}|^{2+\eta} + \mathbb{E}|1 - \sigma_n^2|^{1+\eta/2} \right\}^{1/(3+\eta)}. \quad (5.40)$$

Consider the second term:

$$\mathbb{E} \left| 1 - \sigma_n^2 \right|^2 = s_n^{-4} \mathbb{E} \left| s_n^2 - V_n^2 \right|^2 \leq s_n^{-4} \mathbb{E}(V_n^4). \quad (5.41)$$

Set $G_n(x, y) = \mathbb{E}(H_n(X_1, x)H_n(X_1, y))$, then by the proof of Theorem 1, Hall (1984),

$$\mathbb{E}(V_n^4) \leq C \left( n^4 \mathbb{E}G_n^2(X_1, X_2) + n^3 \mathbb{E}C_n^2(X_1) \right) \leq C \left( n^4 \mathbb{E}G_n^2(X_1, X_2) + n^3 \mathbb{E}H_n^2(X_1, X_2) \right). \quad (5.42)$$
By (5.35) and Corollary 2.7, $s_n^4 \sim n^{4-\delta}$. The calculations in Theorem 1, Zhang and Zheng (1999) can be applied here directly. $H_n(x,y)$ defined in (3.2) is off by a scaling constant $2^{-2jn}n^2$ from their definition.

$$\mathbb{E}H_n^4(X_1, X_2) = (2^{-2jn}n^2)^4 O(2^{3jn}/n^8) = O(2^{-5jn}) = O(n^{-5\delta}), \quad (5.43)$$

and

$$\mathbb{E}G_n^2(X_1, X_2) = (2^{-2jn}n^2)^4 O(2^{jn}/n^8) = O(2^{-7jn}) = O(n^{-7\delta}). \quad (5.44)$$

Combining these estimates and using Hölder inequality, we see

$$\mathbb{E}|1 - \sigma_n^2|^{1+\eta/2} \leq Cn^{-\delta(2+\eta)/4}. \quad (5.45)$$

For the first term in (5.40), we observe that

$$\sum_{i=2}^{n} \mathbb{E}|X_{ni}|^{2+\eta} \leq \sum_{i=2}^{n} \frac{1}{S_n^{2+\eta}} \left( \mathbb{E} \left| \sum_{j=1}^{i-1} H_n(X_i, X_j) \right|^{3} \right)^{(2+\eta)/3}. \quad (5.46)$$

Let $\mathbb{E}_i$ denote the expectation with respect to $X_i$ and $\mathbb{E}_{i'}$ denote the expectation with respect to $X_1, \ldots, X_{i-1}$. We can apply a Hoffmann-Jorgensen type inequality with respect to $\mathbb{E}_{i'}$ (Theorem 1.5.13, de la Peña and Giné, 1999),

$$\mathbb{E} \left| \sum_{j=1}^{i-1} H_n(X_i, X_j) \right|^3 \leq C \mathbb{E}_i \left( \mathbb{E}_{i'} \max_{1 \leq j \leq i-1} |H_n(X_i, X_j)|^3 + \left( \mathbb{E}_{i'} \left( \sum_{j=1}^{i-1} H_n(X_i, X_j) \right)^2 \right)^{3/2} \right). \quad (5.47)$$

The first term can be bounded using (2.3). For the second one, we use Jensen’s inequality, Hölder inequality and (5.43) to get

$$\mathbb{E}_i \left( \mathbb{E}_{i'} \left( \sum_{j=1}^{i-1} H_n(X_i, X_j) \right)^2 \right)^{3/2} = \mathbb{E}_i \left( (i-1) \mathbb{E}_1 H_n^2(X_1, X_2) \right)^{3/2} \leq C(i-1)^{3/2}n^{-15\delta/4}. \quad (5.48)$$

These inequalities and $\sum_{i=2}^{n} i^{(2+\eta)/2} \leq Cn^{2+\eta/2}$ lead to

$$\sum_{i=2}^{n} \mathbb{E}|X_{ni}|^{2+\eta} \leq Cn^{(\delta/2-1)(2+\eta)}n^{-\delta(2+\eta)} \sum_{i=2}^{n} \max(1, i^{3/2}n^{-3\delta/4})^{(2+\eta)/3} \leq Cn^{\delta/2+\eta/4-\eta/2}. \quad (5.49)$$

Gathering (5.40), (5.45) and (5.49) and noting that the bound is minimized when $\eta = 1$, we arrive at

$$\sup_t |\Pr \{S_{nn} \leq t\} - \Pr \{Z \leq t\}| \leq C \max(n^{3\delta/16-1/8}, n^{-3\delta/16}) \leq C n^{-3\delta/16}. \quad (5.50)$$

Putting together the last inequality with (5.25), (5.28), (5.29), (5.33), (5.37) and (5.38), we conclude that when $n$ is large enough (depending on $f$ and $\phi$), there exists a constant $C$
(depending on $f$, $\phi$ and $\{j_n\}$),
\[
\sup_t \Pr\{n^{-j_n/2} \tilde{J}_n / \sigma \leq t\} - \Pr\{Z \leq t\} \leq C \left(n^{-\delta(1/2+\alpha)} \sqrt{\log n + n^{-3\delta/16}}\right) \\
\leq C(n^{-3\delta/16} \vee n^{-\alpha \delta} \sqrt{\log n}).
\]
Taking $C$ sufficiently large so that (1.12) is true for all $n$. 

\[\square\]

Appendix

Proof of Lemma 2.5. By the definition of $C_n(s,t)$,
\[
2^{j_n} \int_{[-M,M]^2} C_n^2(s,t) ds dt = 2^{3j_n} \int_{[-M,M]^2} \left\{ \int_{\mathbb{R}^2} K(2^{j_n}t, 2^{j_n}x) K(2^{j_n}s, 2^{j_n}x) \\
K(2^{j_n}t, 2^{j_n}y) K(2^{j_n}s, 2^{j_n}y) f(x)f(y) dx dy \right\} ds dt
\]
By change of variables $y = x - 2^{-j_n} u, t = 2^{-j_n} w + x, s = 2^{-j_n} z + x$ and the compactness of $\Phi$, this integral is equal to
\[
\int_{-A}^A \int_{-A}^A \int_{-2A}^{2A} \int_{\mathbb{R}} K(2^{j_n}x + z, 2^{j_n}x) K(2^{j_n}x + w, 2^{j_n}x) K(2^{j_n}x + u, 2^{j_n}x) K(2^{j_n}x + t, 2^{j_n}x + u) f(x) f(x - 2^{-j_n}u) 1(2^{-j_n} z + x \in [-M, M]) 1(2^{-j_n} w + x \in [-M, M]) dx du dz dw
\]
\[
= \int_{-A}^A \int_{-A}^A \int_{-2A}^{2A} \sum_{i=-\infty}^\infty \int_0^{2^{-j_n}} K(2^{j_n}x + z + i, 2^{j_n}x + i) K(2^{j_n}x + w + i, 2^{j_n}x + i) K(2^{j_n}x + u + i, 2^{j_n}x + i) f(x + 2^{-j_n}i) f(x + 2^{-j_n}i - 2^{-j_n}u) 1(2^{-j_n} z + x + 2^{-j_n}i \in [-M, M]) 1(2^{-j_n} w + x + 2^{-j_n}i \in [-M, M]) dx du dz dw. 
\]
Using $K(x+1,y+1) = K(x,y)$ and change of variables, it is in turn equal to
\[
\int_{-A}^A \int_{-A}^A \int_{-2A}^{2A} \sum_{i=-\infty}^\infty \int_0^1 2^{-j_n} K(x+z,x) K(x+w,x) K(x+z,x-u) K(x+w,x-u) f(2^{-j_n}(x+i)) f(2^{-j_n}(x+u)) 1(2^{-j_n}(z+x+i) \in [-M, M]) \\
1(2^{-j_n}(w+x+i) \in [-M, M]) dx du dz dw. 
\]
To continue, it is convenient to write
\[
\sum_{i=-\infty}^\infty 2^{-j_n} f(2^{-j_n}(x+i)) f(2^{-j_n}(x+i-u)) 1(2^{-j_n}(z+x+i) \in [-M, M]) \\
1(2^{-j_n}(w+x+i) \in [-M, M]) \\
= \left\{ \sum_{i=2A}^\infty + \sum_{i=-2A}^{-2A+1} + \sum_{i=-2A+1}^{-2A} \right\} 2^{-j_n} f(2^{-j_n}(x+i)) f(2^{-j_n}(x+i-u)) \\
1(2^{-j_n}(z+x+i) \in [-M, M]) 1(2^{-j_n}(w+x+i) \in [-M, M]) =: \text{I}_1(j_n) + \text{I}_2(j_n) + \text{I}_3(j_n) = \text{I}(j_n).
\]
The next lemma proves the convergence of $I(j_n)$.

**Lemma A.1.** Assume that $f$ is bounded. For fixed $M > 0$,

$$I(j_n) \to \int_{-M}^{M} f^2(y)dy$$

uniformly for $x \in [0, 1], u \in [-2A, 2A], z \in [-A, A], w \in [-A, A]$ as $n \to \infty$.

**Proof.** To simplify the notation, let $u' = x - u, z' = x + z, w' = x + w$. Then $u' \in [-2A, 2A + 1], z' \in [-A, A + 1], w' \in [-A, A + 1]$. Consider $I_1(j_n)$. The general summand of $I_1(j_n)$ is zero if $2^{-j_n}(-A + i) > M$.

$$I_1(j_n) = \left( \sum_{i=2A}^{[2^{j_n}M] - 2A - 1} + \sum_{[2^{j_n}M] - 2A}^{[2^{j_n}M] + A} \right) 2^{-j_n} f(2^{-j_n}(x + i)) f(2^{-j_n}(u' + i))$$

$$1(2^{-j_n}(z' + i) \in [0, M]) 1(2^{-j_n}(w' + i) \in [0, M])$$

where $[2^{j_n}M]$ is the largest integer less than or equal to $2^{j_n}M$.

$I_5(j_n)$ is a finite sum with each summand bounded by a constant times $2^{-j_n}$. So $I_5(j_n) \to 0$ uniformly for $x \in [0, 1], u \in [-2A, 2A], z \in [-A, A], w \in [-A, A]$.

Setting $\triangle y = 2^{-j_n}(4A + 1)$, we can simplify $I_4(j_n)$ since the indicator function in the general summand of $I_4(j_n)$ must be 1.

$$I_4(j_n) = \frac{1}{4A + 1} \sum_{i=2A}^{[2^{j_n}M] - 2A - 1} \sum_{j=0}^{N_i} \triangle y f(2^{-j_n}(x + i) + j\triangle y) f(2^{-j_n}(u' + i) + j\triangle y),$$

where $N_i$ is the largest $j$ such that for fixed $i, i + j(4A + 1) \leq [2^{j_n}M] - 2A - 1$. $N_i = \lfloor M / \triangle y - 1 \rfloor$ or $\lfloor M / \triangle y - 2 \rfloor$ depending on $i$.

For each $2A \leq i \leq 6A$, consider the partition of $[0, M]$:

$$P_{i,n} = \{0, 2^{-j_n}(i - 2A), 2^{-j_n}(i - 2A) + \triangle y, ..., 2^{-j_n}(i - 2A) + (N_i + 1)\triangle y, M\}.$$ 

There are at most $N_i + 3$ subintervals. Except for the first and the last subintervals, whose lengths we denote respectively by $\triangle y_i,1$ and $\triangle y_i,N_i+3$, all the subintervals in this partition have length $\triangle y = 2^{-j_n}(4A + 1)$. We also have $0 \leq \triangle y_{i,1} \leq \triangle y$ and $0 \leq \triangle y_{i,N_i+3} \leq \triangle y$.

Setting

$$S_{i,n} := f^2(0) \triangle y_{i,1} + \sum_{j=0}^{N_i} \triangle y f(2^{-j_n}(x + i) + j\triangle y) f(2^{-j_n}(u' + i) + j\triangle y) + f^2(M) \triangle y_{i,N_i+3},$$

(A.7)
we see that

\[ S_{i,n} \leq f^2(0)\Delta y_{i,1} + \sum_{j=0}^{N_i} M_{i,j}^2\Delta y + f^2(M)\Delta y_{i,N_i+3} \quad (A.8) \]

and

\[ S_{i,n} \geq f^2(0)\Delta y_{i,1} + \sum_{j=0}^{N_i} m_{i,j}^2\Delta y + f^2(M)\Delta y_{i,N_i+3}, \quad (A.9) \]

where \( M_{i,j} \) and \( m_{i,j} \) denote respectively the supremum and the infimum of \( f \) on the partition \([2^{-j_n}(i - 2A) + j\Delta y, 2^{-j_n}(i - 2A) + (j + 1)\Delta y]\). As \( n \to \infty \), the mesh of \( P_{i,n} \) tends to zero. Obviously, \( f^2(0)\Delta y_{i,1} + f^2(M)\Delta y_{i,N_i+3} \to 0 \). \( f \in L_1 \) and boundness of \( f \) implies that \( f^2 \) is Riemann integrable on \([0, M]\) for any \( M > 0 \). It follows that \( S_{i,n} \to \int_0^M f^2(y)dy \) for \( 2A \leq i \leq 6A \) and by \((A.6)\),

\[ I_4(j_n) \to \int_0^M f^2(y)dy. \quad (A.10) \]

Note that this convergence is uniform for \( x \in [0,1] \) and \( u' \in [-2A, 2A + 1] \), therefore, it is uniform for \( x \in [0,1], u \in [-2A, 2A], z \in [-A, A], w \in [-A, A] \). We have thus proved that \( \lim_{n \to \infty} I_1(j_n) = \int_0^M f^2(y)dy \) uniformly for \( x, u, z, w \) in the corresponding intervals. By analogy, \( I_3(j_n) \to \int_{-M}^0 f^2(y)dy \) uniformly for \( x, u, z, w \) in the same intervals.

Since \( f \) is bounded, \( I_2(j_n) \to 0 \) as \( n \to \infty \). \( (A.5) \) is proved when collecting the results for \( I_1(j_n), I_2(j_n) \) and \( I_3(j_n) \).

**Lemma A.2.** Assume the scaling function \( \phi \) satisfies \((S1)\) such that the kernel \( K \) associated with \( \phi \) is dominated by \( \Phi \) whose support is contained in \([-A, A]\), where \( A \) is an integer. Then

\[ \int_{-A}^{A} \int_{-A}^{A} \int_{-2A}^{2A} \int_{0}^{1} K(x + z, x)K(x + w, x)K(x + z, x - u)K(x + w, x - u)dxdudzdw = 1. \quad (A.11) \]

**Proof.** Since \( K(x + z, x)K(x + w, x)K(x + z, x - u)K(x + w, x - u) \) is absolutely integrable, by Fubini’s theorem,

\[ \int_{-A}^{A} \int_{-A}^{A} \int_{-2A}^{2A} \int_{0}^{1} K(x + z, x)K(x + w, x)K(x + z, x - u)K(x + w, x - u)dxdudzdw \]

\[ = \int_{0}^{1} \int_{\mathbb{R}} \int_{\mathbb{R}} K(x + z, x)K(x + z, x - u)dz \int_{\mathbb{R}} K(x + w, x)K(x + w, x - u)dwdudx. \quad (A.12) \]
We make the following observation: For any \( y \) and \( z \), by orthogonality of \( \phi \),

\[
\int K(x, y)K(x, z)dx = \int \sum_{k} \phi^2(x - k)\phi(y - k)\phi(z - k)dx + \int \sum_{k \neq l} \phi(x - k)\phi(y - k)\phi(x - l)\phi(z - l)dx
\]

\[
= \sum_{k \in \mathbb{Z}} \phi(y - k)\phi(z - k) \int \phi^2(x - k)dx + \sum_{k \neq l} \phi(y - k)\phi(z - l) \int \phi(x - k)\phi(x - l)dx
\]

\[
= \sum_{k \in \mathbb{Z}} \phi(y - k)\phi(z - k) = K(y, z). \tag{A.13}
\]

For fixed \( x \in [0, 1] \), by repeated applications of the above equation,

\[
\int_\mathbb{R} \int_\mathbb{R} K(x + z, x)K(x + z, x - u)dz \int_\mathbb{R} K(x + w, x)K(x + w, x - u)dwdu
\]

\[
= K(x, x) = \sum_{k \in \mathbb{Z}} \phi^2(x - k). \tag{A.14}
\]

Finally we consider

\[
\int_0^1 \sum_{k \in \mathbb{Z}} \phi^2(x - k)dx = \sum_{k \in \mathbb{Z}} \int_0^1 \phi^2(x - k)dx = \int \phi^2(x)dx = 1. \tag{A.15}
\]

We now continue with the proof of Lemma 2.5. Since in Lemma A.1, the convergence is uniform for \( x \in [0, 1], u \in [-2A, 2A], z \in [-A, A], w \in [-A, A] \), then if \( n \) is sufficiently large, for fixed \( M > 0 \),

\[
|I(j_n)| \leq 2 \int_{-M}^{M} f^2(t)dt. \tag{A.16}
\]

The quantity in (A.3) is bounded in absolute value by

\[
\|\Phi\|_\infty^4 \int_A A \int_A A \int_0^{2A} I(j_n)dx 
\]

\[
\leq 2 \|\Phi\|_\infty^4 \int_A A \int_A A \int_0^{2A} f^2(t)dx 
\]

\[
\int_{-M}^{M} \int_{-M}^{M} f^2(t)dt < \infty \tag{A.17}
\]

for \( n \) large. So, by Fubini, (A.3) is equal to

\[
\int_{-A}^{A} \int_{-A}^{A} \int_{-2A}^{2A} \int_0^1 K(x + z, x)K(x + w, x)K(x + z, x - u)K(x + w, x - u)I(j_n)dx 
\]

\[
\int_{-M}^{M} f^2(t)dy. \tag{A.18}
\]

By dominated convergence and Lemmas A.1 A.2, it converges to \( \int_{-M}^{M} f^2(y)dy. \]
Proof of Lemma 2.6. Choosing $M$ to be an integer such that $M \geq L + 2^{-j_n}(4A + 1)$, we divide the plane $\mathbb{R}^2$ into four regions: $[-M, M]^2$, $[-M, M]^C \times [-M, M]^C$, $[-M, M] \times [-M, M]^C$ and $[-M, M]^C \times [-M, M]$. To get the rate at which $2^{j_n} \int_{[-M,M]^2} C_n^2(s,t)ds dt$ tends to $\int_{-M}^M f^2(y)dy$, we estimate $\left| I(j_n) - \int_{-M}^M f^2(y)dy \right|$. 

$I_1(j_n)$, which was defined in (A.4), can be decomposed into 4 terms as follows.

$$I_1(j_n) = \left( \sum_{i=2A}^{2^{j_n}L} + \sum_{i=2^{j_n}L}^{2^{j_n}L+2A} \sum_{i=2^{j_n}L+2A}^{2^{j_n}M-2A} + \sum_{i=2^{j_n}M-2A}^{2^{j_n}M} \right) 2^{-j_n} f(2^{-j_n}(x + i)) f(2^{-j_n}(u' + i)) 1(2^{-j_n}(z' + i) \in [0, M]) 1(2^{-j_n}(w' + i) \in [0, M]) =: I'_1(j_n) + I'_2(j_n) + I'_3(j_n) + I'_4(j_n).$$

$I'_1(j_n)$ is essentially the same as $I_4(j_n)$ in (A.9). We follow the argument from (A.6) to (A.9) but consider the interval $[0, L]$ instead. Due to the hypothesis of Hölder continuity, there exists $C$ depending on $f$ and $\{j_n\}$, such that

$$|M^2_i - m^2_{ij}| \leq |M_{ij} + m_{ij}||M_{ij} - m_{ij}| \leq C(\Delta y)^\alpha \leq Cn^{-\delta \alpha}.$$ (A.20)

So we obtain

$$\left| S_{i,n} - \int_0^L f^2(y)dy \right| \leq C L n^{-\delta \alpha}. $$ (A.21)

Obviously, $f^2(0)\Delta y_{i,1}$ and $f^2(L)\Delta y_{i,N_i+3}$ are both bounded by $C n^{-\delta}$. From (A.6), for all $x \in [0, 1]$, $u' \in [-2A, 2A + 1]$, $z' \in [-A, A + 1]$, $w' \in [-A, A + 1]$,

$$\left| I'_1(j_n) - \int_0^L f^2(y)dy \right| \leq \frac{1}{4A + 1} \sum_{i=2A}^{2^{j_n}L} |S_{i,n} - f^2(0)\Delta y_{i,1} - f^2(L)\Delta y_{i,N_i+3} - \int_0^L f^2(y)dy| \leq C n^{-\delta \alpha}.$$ (A.22)

for $n$ large enough depending on $\{j_n\}$. $C$ depends on $f$ and $\{j_n\}$.

Next we will look at $I'_2(j_n)$ and consider a partition $P_{i,n}$ on $[L, M]$. Let $\xi_{ij} := 2^{-j_n}(x + i) + j \Delta y$, $\xi'_{ij} := 2^{-j_n}(u' + i) + j \Delta y$. Similar to (A.6), but for a different $N_i$, we write,

$$I'_2(j_n) = \frac{1}{4A + 1} \sum_{i=2^{j_n}L+2A}^{2^{j_n}L+6A} \sum_{j=0}^{N_i} \Delta y f(\xi_{ij}) f(\xi'_{ij}).$$ (A.23)

Since $f$ is bounded and monotonically decreasing on $[L, \infty)$, it follows that

$$\int_{L+\Delta y_{i,1}+\Delta y}^{M-\Delta y_{i,N_i+3}} f^2(y)dy \leq \sum_{j=0}^{N_i} \Delta y f(\xi_{ij}) f(\xi'_{ij}) \leq C \Delta y + \int_{L+\Delta y_{i,1}}^{M-\Delta y_{i,N_i+3}-\Delta y} f^2(y)dy.$$ (A.24)

Thus when $M \geq L + 2^{-j_n}(4A + 1)$, for all $x \in [0, 1]$, $u' \in [-2A, 2A + 1]$, $z' \in [-A, A + 1]$, $w' \in [-A, A + 1]$ and $n$ large enough depending on $\{j_n\}$,

$$\left| I'_2(j_n) - \int_L^M f^2(y)dy \right| \leq C n^{-\delta},$$ (A.25)
where \( C \) depends on \( f \) and \( \{ j_n \} \). We also have \( |I'_5(j_n)| \leq Cn^{-\delta} \) and \( |I'_7(j_n)| \leq Cn^{-\delta} \).

Collecting these bounds,

\[
I_1(j_n) - \int_0^M f^2(y)dy \leq C(n^{-\delta\alpha} + n^{-\delta}) \leq Cn^{-\delta\alpha}.
\]

(A.26)

Now it’s easy to see \( \left| I(j_n) - \int_M^M f^2(y)dy \right| \leq Cn^{-\delta\alpha} \). By (A.3), (A.4) and Lemma A.2, we get

\[
\left| 2^{j_n} \int_{[-M,M]^2} C_n^2(s,t)dsdt - \int_{-M}^M f^2(y)dy \right| \leq Cn^{-\delta\alpha}.
\]

(A.27)

The derivation of a bound on \( \left| 2^{j_n} \int_{[-M,M]^C} \int_{[-M,M]^C} C_n^2(s,t)dsdt - \int_{-M,M}^M f^2(y)dy \right| \) is similar. The analysis of the key component is analogous to \( I'_6(j_n) \), where the monotonicity of the tail of \( f \) is used.

\[
\left| 2^{j_n} \int_{[-M,M]^C} \int_{[-M,M]^C} C_n^2(s,t)dsdt - \int_{[-M,M]^C} f^2(y)dy \right| \leq Cn^{-\delta}.
\]

(A.28)

It’s easier to analyze the integral on the regions \([ -M, M ] \times [ -M, M ]^C \) and \([ -M, M ]^C \times [ -M, M ] \). Both are bounded by \( Cn^{-\delta} \) since there are at most finitely many summands that are not zero. (2.13) follows by collecting the bounds on the four regions and taking \( C \) sufficiently large so that it is true for all \( n \).

\[ \Box \]

**Acknowledgement**

I would like to express my sincere gratitude to my advisor Prof. Evarist Giné for his constant support during the dissertation. I appreciate his patience, numerous hours of insightful discussions and careful proofreading of the manuscript. It would have been almost impossible for me to write this article without his help.

**References**

[1] Bowman, A.W. (1985) *A comparative study of some kernel-based nonparametric density estimators*. J. Stat. Comput. Simul. 21, 313-327.

[2] de la Peña, V. and Giné E. (1999) *Decoupling: From Dependence to Independence*, Springer, New York.

[3] Doukhan, P. and León, J.R. (1990) *Déviation quadratique d’estimateurs de densité par projections orthogonales*. C.R. Acad. Sci. Paris Sr. I 310, 425-430.

[4] Doukhan, P. and León, J.R. (1993) *Quadratic deviation of projection density estimates*. Rebrape 7, 37-63.

[5] Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956) *Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator*. Ann. Math. Statist. 27, 642-669.
[6] Erickson, R.V., Quine M.P. and Weber N.C. (1979) *Explicit bounds for the departure from normality of sums of dependent random variables*. Acta Math. Acad. Sci. H. **34**, 27-32.

[7] Folland G.B. (1999) *Real Analysis: Modern Techniques and Their Applications*, 2nd ed., Wiley, New York.

[8] Giné, E., Latala, R. and Zinn, J. (2000) *Exponential and moment inequalities for U-statistics*. In: Giné, E., Mason, D.M., Wellner, J.A. (eds.): High dimensional probability II, Progr. Probab. **47**, 13-38.

[9] Giné, E. and Mason D.M. (2004) *The law of the iterated logarithm for the integrated squared deviation of a kernel density estimator*. Bernoulli, **10**, 721-752.

[10] Hall, P. (1984) *Central limit theorem for integrated square error of multivariate nonparametric density estimators*. J. Multivariate Anal., **14**, 1-16.

[11] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998) *Wavelets, approximation, and statistical applications*. Lecture Notes in Statistics **129**, Springer, New York.

[12] Kerkyacharian, G. and Picard, D. (1992) *Density estimation in Besov spaces*. Statist. Probab. Lett. **13**, 15-24.

[13] Komlós, J., Major, P. and Tusnády, G. (1975) *An approximation of partial sums of independent rv’s and the sample distribution function*, I. Z. Wahrscheinlichkeitstheorie Verw. Geb., **32**, 111-131.

[14] Montgomery-Smith, S.J. (1993) *Comparison of sums of independent identically distributed random vectors*, Probab. Math. Statist. **14**, 281-285.

[15] Pinsky, M. (1966) *An elementary derivation of Khintchine’s estimate for large deviations*, Probab. Math. Statist. **14**, 281-285.

[16] Shorack, G. and Wellner, J. (1986) *Empirical Processes with Applications to Statistics*, Wiley, New York.

[17] Zhang, S. and Zheng Z. (1999) *On the asymptotic normality for L₂-error of wavelet density estimator with application*, Comm. Statist. Theory Methods. **28**, 1093-1104.