7-19-2013

On the Integrated Squared Error of the Linear Wavelet Density Estimator

Lu Lu
University of Connecticut - Storrs, lu.lu@uconn.edu

On the Integrated Squared Error of the Linear Wavelet Density Estimator

Lu Lu, Ph.D.

University of Connecticut, 2013

ABSTRACT

We study the asymptotic properties of two nonparametric density estimators: linear wavelet density estimators (WDE) and Bernstein density estimators. For linear WDE, we use the integrated squared error as the risk measure. We provide the asymptotically exact almost sure rate of concentration about its mean, in fact a law of the iterated logarithm. Regarding the rate in probability, we obtain a polynomial upper bound for the rate of convergence in the central limit theorem of Zhang and Zheng (1999). Bernstein density estimators are used for estimating densities with a compact support, say $[0,1]$. We consider its performance under the sup norm over an interval $[a,b]$, where $0 < a \leq b < 1$ and show the stochastic error is of a smaller order than that over the interval $[0,1]$.

On the Integrated Squared Error of the Linear Wavelet Density Estimator

Lu Lu

B.E., Optical Engineering, Zhejiang University, 2006

M.Sc., Mathematics, University of Connecticut, 2011

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2013

APPROVAL PAGE

Doctor of Philosophy Dissertation

On the Integrated Squared Error of the Linear Wavelet Density Estimator

Presented by

Lu Lu, B.E., M.Sc.

Major Advisor ______________________________
Evarist Giné-Masdéu

Associate Advisor ______________________________
Richard Bass

Associate Advisor ______________________________
Wolodymyr R. Madych

University of Connecticut

2013

Dedicated to my parents.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Evarist Giné for his guidance, patience and understanding throughout my graduate studies. He rekindled my interest in math and made it possible for me to study in the program. He has spent numerous hours with me to discuss and proofread the thesis. Besides teaching me mathematics, he has also taught me rigorous attitude towards research. I could not imagine finishing the thesis without his help.

I am very grateful to my committee members, Richard Bass and Wolodymyr Madych. Prof. Bass taught me several classes and provided some financial support. I would like to thank Prof. Madych for available discussions.

I'm also indebted to Richard Vitale for suggesting the topic of Bernstein density estimator.

Many other professors in the department helped me and gave me great suggestions. Special thanks go to Iddo Ben-Ari, Ron Blei, Maria Gordina and Alexander Teplyaev. I would also like to thank Monique Roy and Tammy Prentice for their service.

Last but not least, I would like to thank Tingting Huan, Hua Ren, Fangjun Xu and Mingfeng Zhao for their encouragement and support.

TABLE OF CONTENTS

Introduction

0.1 Wavelet Density Estimation

Let $X, X_1, X_2, ..., X_n,\ n \in \mathbb{N}$ be i.i.d random variables in $\mathbb{R}$ with common Lebesgue density f. There are several ways to estimate f. Compared to the traditional parametric approach, the nonparametric approach is much more flexible since it does not assume the underlying density has any specific structure. There are several nonparametric estimators in use, among them are the histograms, the kernel density estimator (KDE), the splines, to name a few. The convolution kernel density estimators and orthogonal projection density estimators are perhaps the most common ways. Wavelet density estimators (WDE) fall into this second category. They were first introduced in Doukhan and León (1990) and Kerkyacharian and Picard (1992).

The wavelet theory aims to approximate functions using orthonormal bases consisting of small waves. In particular, the sequence of spaces $\{V_j\}_{j=-\infty}^{\infty}$ form a multiresolution analysis of $\mathbb{R}$ if $V_j \subset V_{j+1}$ and that $\bigcup_j V_j$ is dense in $L_2(\mathbb{R})$. $\phi(x)$ and $\psi(x)$ are its scaling function and the wavelet function, respectively. The translations and dilations, which are defined by $\phi_{0k} := \phi(x-k)$ and $\psi_{jk} := 2^{j/2}\psi(2^j x - k)$, form a complete orthonormal system in $L_2(\mathbb{R})$. Härdle, Kerkyacharian, Picard and Tsybakov (HKPT, 1998) or Daubechies (1992) give good references for the wavelet theory. We will assume in what follows that the functions ϕ and ψ are bounded and have bounded support (e.g., Daubechies wavelets).

Functions $f \in L_p(\mathbb{R})$ have formal expansions

$$f = \sum_k \alpha_{0k}\phi_{0k}(x) + \sum_{j=0}^{\infty}\sum_k \beta_{jk}\psi_{jk}(x). \tag{0.1.1}$$

The orthogonal projection onto the subspace V_{j_n}, $j_n \in \mathbb{N}$, is then

$$\sum_{k\in\mathbb{Z}} \alpha_{0k}\phi_{0k}(x) + \sum_{j=0}^{j_n-1}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(x) = \sum_{k\in\mathbb{Z}} \alpha_{j_nk}\phi_{j_nk}(x).$$

The linear wavelet density estimator an approximation to the above projection. It is defined by

$$\hat{f}_n(x) = \sum_k \hat{\alpha}_{0k}\phi_{0k}(x) + \sum_{j=0}^{j_n-1}\sum_k \hat{\beta}_{jk}\psi_{jk}(x), \tag{0.1.2}$$

where j_n is a sequence of positive integers. $\hat{\alpha}_{jk}$ and $\hat{\beta}_{jk}$ are the empirical coefficients constructed by the plug-in method.

$$\hat{\alpha}_{jk} = P_n(\phi_{jk}) = \frac{1}{n}\sum_{i=1}^{n} 2^{j/2}\phi(2^jX_i - k), \tag{0.1.3}$$

$$\hat{\beta}_{jk} = P_n(\psi_{jk}) = \frac{1}{n}\sum_{i=1}^{n} 2^{j/2}\psi(2^jX_i - k), \tag{0.1.4}$$

where $P_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$ is the empirical measure corresponding to the sample $\{X_i\}_{i=1}^{n}$, $n \in \mathbb{N}$. They are unbiased estimators of α_{jk} and β_{jk}.

It is interesting to compare the kernel density estimator and the linear wavelet density estimator since they enjoy many similarities. In fact, if ϕ is bounded and compactly supported, the estimator $\hat{f}_n(x)$ in (0.1.2) can be written as

$$f_{n,K}(x) := \hat{f}_n(x) = \frac{2^{j_n}}{n} \sum_{i=1}^{n} K(2^{j_n} t, 2^{j_n} X_i), \tag{0.1.5}$$

where the kernel $K(x, y)$ is given by

$$K(x, y) = \sum_{k \in \mathbb{Z}} \phi(x - k)\phi(y - k). \tag{0.1.6}$$

$\{2^{-j_n}\}$ is playing the role of the bandwidth in the classical kernel density estimation. And by Lemma 8.6, HKPT (1998), $K(x, y)$ is majorized by a convolution kernel $\Phi(x - y)$, that is

$$|K(x, y)| \leq \Phi(x - y), \tag{0.1.7}$$

where $\Phi : \mathbb{R} \to \mathbb{R}^+$ is a bounded, compactly supported and symmetric function.

If we apply thresholding to the wavelet coefficients $\hat{\alpha}_{jk}$ and $\hat{\beta}_{jk}$, then we obtain nonlinear wavelet density estimators. Compared to the kernel density estimators, they are spatially adaptive and even accomodates discontinuities (Donoho, Johnstone, Kerkyacharian and Picard, 1995). As a first step to understand their asymptotic properties, we will restrict our attention to the linear ones in this thesis.

Given an estimator $\hat{f}_n$, there are several risk measures to assess the performance of it. For example, we study the pointwise error, e.g. $|\hat{f}_n(x) - f(x)|$ or the sup norm, e.g., $I_\infty = \|\hat{f}_n - f\|_\infty$, uniformly over the whole or part of the domain of f. For the wavelet density estimator described above, Massiani (2002, 2003) proved a pointwise law of the iterated logarithm (LIL) and a LIL for the supremum norm over a compact interval. Varron (2008) extended the latter result to multivariate densities. Giné and Nickl (2009) obtained the almost sure rate of convergence of $\sup_{x\in\mathbb{R}} |\hat{f}_n(x) - f(x)|$ using methods from the theory of empirical processes over general classes of functions.

In this thesis, we study the asymptotic properties under the integrated squared error, e.g. $I_2 := \int |\hat{f}_n - f|^2$. It is a natural measure since the estimator $\hat{f}_n$ should be constructed in such a way that the mean integrated squared error is minimized (Bowman, 1985). Hall (1984) proved a central limit theorem for $I_2 - \mathbb{E}I_2$ in the case of a kernel density estimator. Giné and Mason (2004) established the law of the iterated logarithm for the L_2 error of a kernel density estimator. While the former gives rate of approximation in probability, the latter deals with a.s. rate of convergence. Zhang and Zheng (1999) studied the asymptotic normality of the integrated squared error of a wavelet density estimator and the LIL for WDE is done here. The proof here for the most part follows the same pattern as in Giné and Mason (2004), including U-statistics theory, the Komlós-Major-Tusnády (KMT) approximation and a refined moderate deviations result for weighted chi-squared variables. But the variance computation in the wavelet case is different. It requires nontrivial variance computations based on wavelet approximation. For this purpose, we adapted the proof of Proposition 1 in Zhang and Zheng (1999).

As a complement to Zhang and Zheng's result, we also obtain an upper bound for the rate of convergence in their central limit theorem. Doukhan and León (1993) obtained the rate of convergence in the CLT for generalized density projection estimates with respect to Prohorov's metric. However, the optimal window width was not attained in that case.

More specifically, we will prove a LIL and a bound on the rate of convergence in CLT for the statistic

$$J_n := \|f_{n,K} - f\|_2^2 - \mathbb{E}\|f_{n,K} - f\|_2^2. \tag{0.1.8}$$

The results are stated in Theorems 1.4.1 and 2.0.1, respectively. Both LIL and CLT also appear in the author's published work (Lu, 2013). As was observed by Zhang and Zheng (1999), a difference from the kernel density estimator is that, by orthogonality of the wavelet bases, J_n is equal to $\bar{J}_n$, where

$$\bar{J}_n := \|f_{n,K} - \mathbb{E}f_{n,K}\|_2^2 - \mathbb{E}\|f_{n,K} - \mathbb{E}f_{n,K}\|_2^2. \tag{0.1.9}$$

Although it requires more stringent conditions on f to study the stochastic part, there is no need to study the bias part. Therefore it is not necessary to impose the second order differentiability on the density f, which is required in the kernel case.

It is natural to ask what the order of the mean integrated squared error (MISE) $\mathbb{E}I_2$ is. Hall and Patil (1995) obtained the MISE of the linear wavelet density estimator on a compact interval in $\mathbb{R}$ under some smoothness conditions on f. Theorem 10.1, HKPT (1998) gives a bound on the MISE on $\mathbb{R}$. We will show that, the bound on the stochastic part of MISE cannot be improved for a large class of density functions. That is, it is of the order $2^{j_n}/n$. Thus MISE must be at least of the order $2^{j_n}/n$. This is strictly larger than the order of $J_n = I_2 - \mathbb{E}I_2$, which is $2^{j_n/2}\sqrt{\log\log n}/n$. It makes sense since the MISE gives a first-order approximation of I_2 and the rates in LIL and CLT tell us how well the approximation works.

0.2 Bernstein Density Estimation

Given a sequence of i.i.d. random variables $\{X_i\}_{i=1}^n$ with the Lebesgue density f on $\mathbb{R}$ that has a compact support, we may assume that f lives on $[0, 1]$ without loss of generality. Let F denote the corresponding distribution function. The kernel density estimator treats the density as if it is on $\mathbb{R}$, and thus gives a positive estimate to regions outside $[0, 1]$. This is the well-known boundary effects of the kernel density estimator. It behaves poorly in the boundary region especially when the density is nonzero at the boundary. Several methods of correcting the boundary effects have been proposed. A review of these techniques was given by Jones (1993).

Recently, there have been researches on the Bernstein density estimator, which was originally introduced by Vitale (1975). It is expected to have better boundary behavior than KDE. The definition is motivated by the fact that the Bernstein polynomials of a distribution function F converge to F uniformly on $[0, 1]$. Let $F_n(x) := \frac{1}{n}\sum_{i=1}^n 1(X_i \le x)$ be the empirical distribution function. For $x \in [0, 1]$, $m \in \mathbb{Z}$, $k = 0, ..., m$, let

$$b_{k,m}(x) := \binom{m}{k} x^k (1-x)^{m-k}. \tag{0.2.1}$$

For n observations, the Bernstein estimator of F is given by

$$\hat{F}_{m,n}(x) := \sum_{k=0}^{m} F_n\left(\frac{k}{m}\right) b_{k,m}(x). \tag{0.2.2}$$

Taking derivative of $\hat{F}_{m,n}(x)$ leads to the Bernstein density estimator. That is

$$\hat{f}_{m,n}(x) = \frac{d}{dx}\hat{F}_{m,n}(x) = m\sum_{k=0}^{m-1}\left[F_n\left(\frac{k+1}{m}\right) - F_n\left(\frac{k}{m}\right)\right] b_{k,m-1}(x). \tag{0.2.3}$$

$\hat{f}_{m,n}(x)$ is a polynomial approximation to the density f, and it is itself a genuine density function that is infinitely differentiable. It collects histogram values on the intervals $[k/m, (k+1)/m]$, $k = 0, ..., m-1$. m is a function of n and $m \to \infty$ as $n \to \infty$. Thus it is natural to think of $1/m$ as the bandwidth of the estimator. At a fixed point x, the weights $b_{k,m-1}(x)$ are large for those k's for which x and k/m are close. The number of observations in $[k/m, (k+1)/m]$ for the corresponding k's contribute more to $\hat{f}_{m,n}(x)$ than those in the intervals where the distances between x and k/m are large.

It can also be written as a linear combination of beta densities with random weights. Indeed, let $\beta_{k,m}(x) = \frac{\Gamma(k+m)}{\Gamma(k)\Gamma(m)} x^{k-1}(1-x)^{m-1}$ for $x \in [0,1]$ and 0 otherwise. Then

$$\hat{f}_{m,n}(x) = \sum_{k=0}^{m-1} \left[F_n\left(\frac{k+1}{m}\right) - F_n\left(\frac{k}{m}\right) \right] \beta_{k+1,m-k}(x). \tag{0.2.4}$$

We also note that, through transformations, the estimator can be used to estimate a density on any compact interval in $\mathbb{R}$ and on the whole real line as well.

Several different measures have been used to study the performance of the estimator. Vitale (1975) studied the pointwise mean squared error (MSE) $\mathbb{E}[\hat{f}_{m,n}(x) - f(x)]^2$ and obtained the asymptotic bias and variance. Under certain smoothness conditions on f, for $x \in (0,1)$, $\hat{f}_{m,n}(x)$ is optimal when $m \sim n^{2/5}$. MSE is of the order $n^{-4/5}$. Babu, Canty, Chaubey (2002) considered the supremum norm and showed that $\sup_{x\in[0,1]} |\hat{f}_{m,n}(x) - \mathbb{E}\hat{f}_{m,n}(x)|$ is $O_{a.s.}(\sqrt{m \log n / n})$. The mean integrated squared error has been studied by Leblanc (2010). It has been shown that the optimal bandwidth, in this case, $1/m$, is of the order $n^{-2/5}$. It is different from KDE and WDE since the bias is larger but the variance is smaller.

Tenbusch (1994) and Babu and Chaubey (2006) considered extensions to multivariate densities. There is also some literature on the Bayesian approach, such as Ghosal (2001), Kruijer and van der Vaart(2008), Petrone (1999), Petrone and Wasserman (2002). Lorentz (1986) gives a good reference on the properties of Bernstein polynomials.

We study the convergence rate of the stochastic error under the sup norm over a compact interval $[a, b]$, where $0 < a < b < 1$, i.e., $\sup_{x\in[a,b]} |\hat{f}_{m,n}(x) - \mathbb{E}\hat{f}_{m,n}(x)|$. It is shown in Theorem 3.2.4 that the order is strictly smaller than that on $[0, 1]$. We employ techniques from empirical processes which have also been successful in showing such upper bounds for KDE and WDE (Giné and Guillou, 2002; Giné and Nickl, 2009). Moreover, we also derive a lower bound result on $[a, b]$ in Theorem 3.3.2. It implies that the upper bound result is optimal. This suggests that the Bernstein estimator does experience some boundary effects. It also coincides with the conclusion by Leblanc (2012), where the boundary properties of the Bernstein estimator have been studied under the pointwise error. It would be interesting to extend the upper and lower bounds derived here to $[a_n, b_n]$, where $a_n \to 0$, $b_n \to 1$ and obtain the uniform convergence rate of the bias.

Chapter 1

Law of the Iterated Logarithm for the Wavelet Density Estimation

1.1 Setup and notations

In order to study the the integrated squared error for the wavelet density estimator, we shall impose the following conditions:

(f): $f^2(x)$ is bounded and Riemann integrable on $\mathbb{R}$.

(S1): The scaling function ϕ is bounded and compactly supported.

Then, in (0.1.7), we can assume Φ 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),

$$\operatorname*{ess\,sup}_x \theta_\phi(x) < \infty. \tag{1.1.1}$$

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

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

(B1):

$$j_n \to \infty,\ 2^{-j_n} \asymp n^{-\delta} \quad \text{for some } \delta \in (0, 1/3), \tag{1.1.2}$$

where $a_n \asymp b_n$ means $0 < \liminf a_n/b_n < \limsup a_n/b_n < \infty$.

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

$$\lambda_{k+1}/\lambda_k \to 1,\ \log\log\lambda_k/\log k \to 1,\ \lambda_{k+1}-\lambda_k \to \infty \tag{1.1.3}$$

as $k \to \infty$, such that

$$2^{-j_n} \text{ is constant for } n \in [\lambda_k, \lambda_{k+1}),\ \ k \in \mathbb{N}. \tag{1.1.4}$$

For instance, the sequence $\lambda_k = \exp(k/\log(e+k))$ satisfies these conditions.

Next we set up some notations.

$$K_n(t,x) = K(2^{j_n}t, 2^{j_n}x)\ \ and\ \bar{K}_n(t,x) = K_n(t,x) - \mathbb{E}K_n(t,X).$$

Then

$$\begin{aligned}
\bar{J}_n &= \int |f_{n,K} - \mathbb{E}f_{n,K}|^2 - \mathbb{E}\int |f_{n,K} - \mathbb{E}f_{n,K}|^2 \\
&= \frac{2^{2j_n}}{n^2}\left[\int_{\mathbb{R}}\left(\sum_{i=1}^n \bar{K}(2^{j_n}t, 2^{j_n}X_i)\right)^2 dt - \mathbb{E}\int_{\mathbb{R}}\left(\sum_{i=1}^n \bar{K}(2^{j_n}t, 2^{j_n}X_i)\right)^2 dt\right] \\
&= \frac{2^{2j_n}}{n^2} W_n(\mathbb{R}),
\end{aligned} \tag{1.1.5}$$

where

$$\begin{aligned}
W_n(F) &= \int_F\left(\sum_{i=1}^n \bar{K}(2^{j_n}t, 2^{j_n}X_i)\right)^2 dt - \mathbb{E}\int_F\left(\sum_{i=1}^n \bar{K}(2^{j_n}t, 2^{j_n}X_i)\right)^2 dt \\
&= \int_F\left(\sum_{i=1}^n \bar{K}_n(t, X_i)\right)^2 dt - \mathbb{E}\int_F\left(\sum_{i=1}^n \bar{K}_n(t, X_i)\right)^2 dt \\
&=: U_n(F) + L_n(F),
\end{aligned} \tag{1.1.6}$$

$$U_n(F) := \sum_{1\le i\ne j\le n} \int_F \bar{K}_n(t, X_i)\bar{K}_n(t, X_j)dt \tag{1.1.7}$$

and

$$L_n(F) := \sum_{i=1}^{n} \int_F \left(\bar{K}_n^2(t, X_i) - \mathbb{E}\bar{K}_n^2(t, X)\right) dt. \tag{1.1.8}$$

The measurable set F will be

$$\mathbb{R},\ \ [-M, M]\ \ or\ \ [-M, M]^C,\ \ M > 0,\ \ with\ \int_F f(x)dx > 0. \tag{1.1.9}$$

The main property for F is

$$\lambda(\{x + y : x \in F, |y| < \varepsilon\} \cap F^c) \to 0 \text{ as } \varepsilon \to 0. \tag{1.1.10}$$

To motivate the computations for tail estimations and moderate deviations, let's explain the idea in the proof of the LIL. Since $\bar{J}_n$ can be decomposed into $W_n([-M, M])$ and $W_n([-M, M]^C)$, we need to control statistics of these forms. $W_n([-M, M]^C)$ is the sum of a degenerate U-statistic and a diagonal term. It can be handled with an exponential inequality in Giné, Latała and Zinn (2000). The bound for the diagonal term in $W_n([-M, M]^C)$ is obtained using the Berstein inequality. These tools are very useful in proving the blocking for the upper bound in LIL. $W_n([-M, M])$ is approximated by a Gaussian chaos. The moderate deviation result in Giné and Mason (2004) is used to obtain the bound on the Gaussian chaos. It was originally proved using Pinsky's method (Pinsky, 1966). We also prove a rate of convergence in the CLT of the wavelet density estimator as a complement to Zhang and Zheng's result (1999). In order to get this, we need to assume more conditions on f. $\bar{J}_n$ is composed of $L_n(\mathbb{R})$ and $U_n(\mathbb{R})$. The

Bernstein inequality 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. As was pointed out by Giné and Mason (2004), the methods used here do not apply to L_p norm if $p \neq 2$. This is because $U_n(F)$ can be written as a U-statistic under the L_2 norm whereas it is difficult to handle if $p \neq 2$.

1.2 Tail Estimation

The goal of this section is to obtain exponential inequalities for $W_n(F)$, where F satisfies (1.1.9) and also for $W_n(\mathbb{R}) - W_{n,m}(\mathbb{R})$, where $W_{n,m}(\mathbb{R})$ is defined below. We assume throughout this section that ϕ satisfies $(S1)$, and K is associated with ϕ as in (0.1.6).

Set, for $m < n$,

$$W_{n,m}(\mathbb{R}) := \int_{\mathbb{R}} \left[\left(\sum_{m<i\leq n} \bar{K}_n(t, X_i) \right)^2 - \mathbb{E} \left(\sum_{m<i\leq n} \bar{K}_n(t, X_i) \right)^2 \right] dt \tag{1.2.1}$$

and

$$H_n(x, y) = \int_{\mathbb{R}} \bar{K}_n(t, x)\bar{K}_n(t, y)dt, \tag{1.2.2}$$

$$H_{n,F}(x, y) = \int_F \bar{K}_n(t, x)\bar{K}_n(t, y)dt. \tag{1.2.3}$$

With this notation,

$$U_n(F) = \sum_{1\leq i\neq j\leq n} H_{n,F}(X_i, X_j),\ L_n(F) = \sum_{i=1}^{n} \left(H_{n,F}(X_i, X_i) - \mathbb{E}H_{n,F}(X_i, X_i)\right), \tag{1.2.4}$$

and

$$
\begin{aligned}
&W_n(\mathbb{R}) - W_{n,m}(\mathbb{R}) \\
&= 2\sum_{i=1}^{m}\sum_{j=m+1}^{n} H_n(X_i, X_j) + \sum_{1\le i\ne j\le m} H_n(X_i, X_j) + \sum_{i=1}^{m}\left(H_n(X_i, X_i) - \mathbb{E}H_n(X_i, X_i)\right).
\end{aligned} \tag{1.2.5}
$$

Consider the 3 terms on the right-hand side. Bernstein's inequality (e.g., de la Peña and Giné 1999) says that for centered, i.i.d. variables ξ_i, if $\|\xi_i\|_\infty \le c \le \infty$ and $\sigma^2 = E\xi_i^2$, then

$$
\Pr\left\{\sum_{i=1}^{m} \xi_i > t\right\} \le \exp\left(-\frac{t^2}{2m\sigma^2 + \frac{2}{3}ct}\right). \tag{1.2.6}
$$

Applying it to the 3rd term in (1.2.5), given Corollary A.0.1, and inequality (A.0.5), noting that $c = 8 \cdot 2^{-j_n}\|\Phi\|_2^2$ and $\sigma^2 \le 4 \cdot 2^{-2j_n}\|\Phi\|_2^4$, we get, for all $m \in \mathbb{N}$ and $n \in \mathbb{N}$,

$$
\begin{aligned}
&\Pr\left\{\left|\sum_{i=1}^{m}\left(H_n(X_i, X_i) - \mathbb{E}H_n(X_i, X_i)\right)\right| > \tau n 2^{-\frac{3}{2}j_n}\right\} \\
&\le 2\exp\left(-\frac{\tau^2 n^2 2^{-3j_n}}{8m2^{-2j_n}\|\Phi\|_2^4 + \frac{16}{3}\tau n 2^{-\frac{5}{2}j_n}\|\Phi\|_2^2}\right).
\end{aligned} \tag{1.2.7}
$$

The first two terms in (1.2.5) are of U-statistics type. They can be controlled by using the following exponential inequality for canonical U-statistics. We recall that a U-statistic of order two, $\sum_{1\le i\ne j\le k} h_{i,j}(X_i, X_j)$ is canonical (or degenerate) for the probability law of X if $\mathbb{E}h_{i,j}(X, y) = \mathbb{E}h_{i,j}(x, X) = 0$ for all $i \ne j$ and $x, y \in \mathbb{R}$. Likewise, a decoupled U-statistic $\sum_{1\le i,j\le n} h_{i,j}(X_i^{(1)}, X_j^{(2)})$ where the $2n$ random variables are independent and not necessarily identically distributed, is canonical if for all i, j and $x, y \in \mathbb{R}$, $\mathbb{E}h_{i,j}(X_i^{(1)}, y) = \mathbb{E}h_{i,j}(x, X_j^{(2)}) = 0$.

1.2.1 Theorem. (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 $(X_i^{(1)}, X_j^{(2)})$, $i, j = 1, 2, ..., n$, and if A, B, C, D are as defined below, then

$$\Pr\left\{\left|\sum_{1\le i,j\le n} h_{i,j}(X_i^{(1)}, X_j^{(2)})\right| \ge x\right\} \le L\exp\left[-\frac{1}{L}\min\left(\frac{x^2}{C^2}, \frac{x}{D}, \frac{x^{2/3}}{B^{2/3}}, \frac{x^{1/2}}{A^{1/2}}\right)\right] \tag{1.2.8}$$

for all $x > 0$, where

$$D = \|(h_{i,j})\|_{L^2\to L^2}$$
$$:= \sup\left\{\mathbb{E}\sum_{i,j} h_{i,j}(X_i^{(1)}, X_j^{(2)}) f_i(X_i^{(1)}) g_j(X_j^{(2)}) : \mathbb{E}\sum_i f_i^2(X_i^{(1)}) \le 1, \mathbb{E}\sum_j g_j^2(X_j^{(2)}) \le 1\right\}, \tag{1.2.9}$$

$$C^2 = \sum_{i,j} \mathbb{E}h_{i,j}^2(X_i, X_j), \tag{1.2.10}$$

$$B^2 = \max_{i,j}\left[\left\|\sum_i \mathbb{E}h_{i,j}^2(X_i^{(1)}, y)\right\|_\infty, \left\|\sum_j \mathbb{E}h_{i,j}^2(x, X_j^{(2)})\right\|_\infty\right] \tag{1.2.11}$$

and

$$A = \max_{i,j} \|h_{i,j}\|_\infty. \tag{1.2.12}$$

1.2.2 Remark. Theorem 1.2.1 holds if the decoupled U-statistic $\sum_{1\le i,j\le n} h_{i,j}(X_i^{(1)}, X_j^{(2)})$ is replaced by the undecoupled one $\sum_{i\ne j\le n} h_{i,j}(X_i, X_j)$. This can be proved by setting $h_{i,i} = 0$ so that $\sum_{1\le i\ne j\le n} h_{i,j}(X_i, X_j) = \sum_{1\le i,j\le n} h_{i,j}(X_i, X_j)$, and using the decoupling result in Theorem 3.4.1, de la Peña and Giné (1999).

Due to the remark, to deal with the 2nd term in (1.2.5), we can take $h_{i,j} = H_{n,F,i,j} = H_{n,F}$

and apply Theorem 1.2.1 to $\sum_{1\le i\neq j\le m} H_{n,F}(X_i, X_j)$. Next, we show the calculations on the bounds for A, B, C and D.

Inequality (A.0.6) immediately gives the following two bounds.

$$\begin{aligned} A &= \max_{i,j} \|h_{i,j}\|_\infty = \max_{i,j} \|H_{n,F}\|_\infty \\ &= \max_{i,j} \left\| \int_F \bar{K}_n(t,x)\bar{K}_n(t,y)dt \right\|_\infty \\ &\le 4 \cdot 2^{-j_n} \|\Phi\|_2^2, \end{aligned} \tag{1.2.13}$$

and

$$\begin{aligned} B^2 &= \max_{i,j} \left[\left\| \sum_i \mathbb{E}h_{i,j}^2(X_i^{(1)}, y) \right\|_\infty, \left\| \sum_j \mathbb{E}h_{i,j}^2(x, X_j^{(2)}) \right\|_\infty \right] \\ &\le 16m \cdot 2^{-2j_n} \|\Phi\|_2^4. \end{aligned} \tag{1.2.14}$$

Next we consider

$$C^2 = \sum_{i,j} \mathbb{E}H_{n,F}^2(X_i, X_j) = \sum_{i,j} \mathbb{E}\left(\int_F \bar{K}_n(t, X_1)\bar{K}_n(t, X_2)dt \right)^2. \tag{1.2.15}$$

By (A.0.6),

$$\mathbb{E}\left(\int_F |\bar{K}_n(t, X_1)\bar{K}_n(t, X_2)|dt \right)^2 \le (4 \cdot 2^{-j_n}\|\Phi\|_2^2)^2 < \infty. \tag{1.2.16}$$

So we can apply Fubini's theorem to the effect that

$$\begin{aligned} \mathbb{E}\left(\int_F \bar{K}_n(t, X_1)\bar{K}_n(t, X_2)dt \right)^2 &= \int_{F^2} \mathbb{E}[\bar{K}_n(t, X_1)\bar{K}_n(s, X_1)]\mathbb{E}[\bar{K}_n(t, X_2)\bar{K}_n(s, X_2)]dsdt \\ &= 2^{-2j_n} \int_{F^2} 2^{j_n}\mathbb{E}[\bar{K}_n(t, X_1)\bar{K}_n(s, X_1)]2^{j_n}\mathbb{E}[\bar{K}_n(t, X_2)\bar{K}_n(s, X_2)]dsdt \\ &= 2^{-2j_n} \int_{F^2} (2^{j_n}\mathbb{E}[\bar{K}_n(t, X)\bar{K}_n(s, X)])^2 dsdt \end{aligned}$$

$$= 2^{-2j_n} \int_{F^2} R_n^2(t,s)dsdt, \tag{1.2.17}$$

where

$$R_n(t,s) := 2^{j_n} \int_{\mathbb{R}} \bar{K}_n(t,x)\bar{K}_n(s,x)f(x)dx. \tag{1.2.18}$$

By the above calculations and Lemmas $A.0.2$, $A.0.3$, for n large enough depending on F,

$$\begin{aligned} \mathbb{E}H_{n,F}^2(X_1,X_2) &= 2^{-3j_n}2^{j_n}\int_{F^2} R_n^2(t,s)dsdt \\ &\le 2^{-3j_n}\cdot 2\int_F f^2(x)dx \int_{\mathbb{R}}\left(\int_{\mathbb{R}} \Phi(w+u)\Phi(w)dw\right)^2 du \\ &\le 2^{-3j_n}\cdot 2\|\Phi\|_1^2\|\Phi\|_2^2 \int_F f^2(x)dx. \end{aligned} \tag{1.2.19}$$

Thus we have proved that, for all n large enough depending on F,

$$C^2 \le 2m^2\cdot 2^{-3j_n}\|\Phi\|_1^2\|\Phi\|_2^2\int_F f^2(x)dx. \tag{1.2.20}$$

Finally, we estimate D.

1.2.3 Lemma. If f satisfies condition (f) and ϕ satisfies condition (S1),

$$D \le 4m2^{-2j_n}\|f\|_\infty\|\Phi\|_1^2. \tag{1.2.21}$$

Proof. Since $h_{i,j} = H_{n,F}$ for all i, j, claim that

$$D = m\sup\left\{\mathbb{E}H_{n,F}(X_1,X_2)l(X_1)g(X_2) : \mathbb{E}l^2(X_1)\le 1, \mathbb{E}g^2(X_2)\le 1\right\}. \tag{1.2.22}$$

To see this, if in (1.2.9), we take $f_i = l/\sqrt{m}$ for each i and $g_j = g/\sqrt{m}$ for each j with l

and g satisfying $\mathbb{E}l^2(X_1) \le 1, \mathbb{E}g^2(X_2) \le 1$, it is easy to see that

$$m \sup \left\{ \mathbb{E}H_{n,F}(X_1, X_2) l(X_1) g(X_2) : \mathbb{E}l^2(X_1) \le 1, \mathbb{E}g^2(X_2) \le 1 \right\} \le D.$$

On the other hand, taking $l = \dfrac{1}{\sqrt{m}} \sum_i f_i$ and $g = \dfrac{1}{\sqrt{m}} \sum_j g_j$ with $\mathbb{E} \sum_i f_i^2(X_1) \le 1$ and $\mathbb{E} \sum_j g_j^2(X_2) \le 1$, we see that

$$D \le m \sup \left\{ \mathbb{E}H_{n,F}(X_1, X_2) l(X_1) g(X_2) : \mathbb{E}l^2(X_1) \le 1, \mathbb{E}g^2(X_2) \le 1 \right\}.$$

So (1.2.22) holds. In fact, D can be further simplified. Using the definition of $H_{n,F}$ in (1.2.3), we get

$$\begin{aligned} D &= m \sup \left\{ \mathbb{E}H_{n,F}(X_1, X_2) l(X_1) g(X_2) : \mathbb{E}l^2(X_1) \le 1, \mathbb{E}g^2(X_2) \le 1 \right\} \\ &= m \sup \left\{ \mathbb{E} \int_F \bar{K}_n(t, X_1) \bar{K}_n(t, X_2) l(X_1) g(X_2) dt : \mathbb{E}l^2(X_1) \le 1, \mathbb{E}g^2(X_2) \le 1 \right\} \\ &= m \sup \left\{ \int_F \left[\mathbb{E}(\bar{K}_n(t, X) \varphi(X)) \right]^2 dt : \mathbb{E}\varphi^2(X) \le 1 \right\}. \end{aligned} \tag{1.2.23}$$

By Cauchy-Schwarz,

$$\begin{aligned} \int_{\mathbb{R}} \left[\mathbb{E}|\bar{K}_n(t, X) \varphi(X)| \right]^2 dt &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |\bar{K}_n(t, x) \varphi(x)| f(x) dx \right)^2 dt \\ &\le \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |\bar{K}_n(t, x)| f(x) dx \int_{\mathbb{R}} |\bar{K}_n(t, y)| \varphi^2(y) f(y) dy \right) dt. \end{aligned} \tag{1.2.24}$$

Let $\Phi_n(x) := \Phi(2^{j_n} x)$. Using (0.1.7) and that f is bounded, for all $t \in \mathbb{R}$,

$$\begin{aligned} \int_{\mathbb{R}} |\bar{K}_n(t, x)| f(x) dx &= \int_{\mathbb{R}} |K_n(t, x) - \mathbb{E}K_n(t, X)| f(x) dx \\ &\le \int_{\mathbb{R}} |K_n(t, x)| f(x) dx + \int_{\mathbb{R}} \mathbb{E}|K_n(t, X)| f(x) dx \\ &\le \int_{\mathbb{R}} |\Phi_n(t - x)| f(x) dx + \mathbb{E}|K_n(t, X)| \end{aligned}$$

$$\leq 2\|\Phi_n\|_1\|f\|_\infty.$$

For all ϕ such that $\mathbb{E}\phi^2(X) \leq 1$, we also have

$$\begin{aligned}
&\int_{\mathbb{R}}\int_{\mathbb{R}} |\bar{K}_n(t,y)|\varphi^2(y)f(y)dydt \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} |K_n(t,y) - \mathbb{E}K_n(t,X)|\varphi^2(y)f(y)dydt \\
&\leq \int_{\mathbb{R}}\int_{\mathbb{R}} (|K_n(t,y)| + \mathbb{E}|K_n(t,X)|)\varphi^2(y)f(y)dydt \qquad (1.2.25)\\
&\leq \int_{\mathbb{R}}\int_{\mathbb{R}} \Phi_n(t-y)\varphi^2(y)f(y)dydt \\
&\quad + \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{E}|K_n(t,X)|\varphi^2(y)f(y)dydt \\
&\leq \|\Phi_n\|_1 + \int_{\mathbb{R}} \mathbb{E}\Phi_n(t-X)dt \\
&= 2\|\Phi_n\|_1.
\end{aligned}$$

Then combining (1.2.23) − (1.2.25), we obtain $D \leq 4m\|f\|_\infty\|\Phi_n\|_1^2$. Since $\|\Phi_n\|_1 = \int_{\mathbb{R}} \Phi(2^{j_n}x)dx = 2^{-j_n}\|\Phi\|_1$, we get

$$D \leq 4m2^{-2j_n}\|f\|_\infty\|\Phi\|_1^2. \qquad (1.2.26)$$

□

1.2.4 Proposition. Let X_i be i.i.d. with density f satisfying condition (f). Let F be a measurable subset of $\mathbb{R}$ satisfying condition (1.1.9). ϕ satisfies ($S1$) and K is the projection kernel associated with ϕ. $2^{-j_n} \to 0$. Then there exist constants κ_0 (depending on f and ϕ) and n_0 (depending on F, f, ϕ and the sequence $\{j_n\}$) such that, for all $\tau > 0$

and for all $n \geq n_0$, $0 \leq m < n$,

$$\Pr\left\{\left|\sum_{1\leq i\neq j\leq m} H_{n,F}(X_i, X_j)\right| \geq \tau n 2^{-\frac{3}{2}j_n}\right\}$$
$$\leq \kappa_0 \exp\left(-\frac{1}{\kappa_0}\min\left[\frac{\tau^2 n^2}{m^2\int_F f^2(x)dx}, \frac{\tau n}{m 2^{-j_n/2}}, \frac{\tau^{2/3}n^{2/3}2^{-\frac{j_n}{3}}}{m^{1/3}}, \tau^{1/2}n^{1/2}2^{-\frac{j_n}{4}}\right]\right) \tag{1.2.27}$$

and

$$\Pr\left\{\left|\sum_{i=1}^{m}\sum_{j=m+1}^{n} H_{n,F}(X_i, X_j)\right| \geq \tau n 2^{-\frac{3}{2}j_n}\right\}$$
$$\leq \kappa_0 \exp\left(-\frac{1}{\kappa_0}\min\left[\frac{\tau^2 n^2}{m(n-m)\int_F f^2(x)dx}, \frac{\tau n}{\sqrt{m(n-m)}2^{-j_n/2}}, \right.\right. \tag{1.2.28}$$
$$\left.\left.\frac{\tau^{2/3}n^{2/3}2^{-\frac{j_n}{3}}}{(m\vee(n-m))^{1/3}}, \tau^{1/2}n^{1/2}2^{-\frac{j_n}{4}}\right]\right).$$

Proof. Gathering Theorem 1.2.1, (1.2.13), (1.2.14), (1.2.20) and (1.2.21), we get (1.2.27). (1.2.28) can be obtained in a similar way. □

1.2.5 Proposition. Under the same hypotheses of Proposition 1.2.4 on f, ϕ and $\{j_n\}$, there exist constants κ_0 (depending on ϕ and f) and n_0 (depending on F, f, ϕ and the sequence $\{j_n\}$) such that, for all $\tau > 0$ and for all $n \geq n_0$,

$$\Pr\left\{|W_n(F)| \geq \tau n 2^{-\frac{3}{2}j_n}\right\}$$
$$\leq \kappa_0 \exp\left(-\frac{1}{\kappa_0}\min\left[\frac{\tau^2}{\int_F f^2(x)dx}, 2^{j_n/2}\tau, \tau^{2/3}n^{1/3}2^{-\frac{j_n}{3}}, \tau^{1/2}n^{1/2}2^{-\frac{j_n}{4}}, \tau^2 n 2^{-j_n}, \tau n 2^{-\frac{j_n}{2}}\right]\right). \tag{1.2.29}$$

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 κ_0 and n_0 as above such that

$$\Pr\left\{|W_n(F)| \geq \eta n 2^{-\frac{3}{2}j_n}\sqrt{\log\log n}\right\} \leq \kappa_0 \exp\left(-\frac{\eta^2 \log\log n}{\kappa_0 \int_F f^2(x)dx}\right) \tag{1.2.30}$$

for all $n \geq n_0$.

Proof. Clearly,

$$\Pr\left\{|W_n(F)| \geq 2\tau n 2^{-\frac{3}{2}j_n}\right\} \leq \Pr\left\{|L_n(F)| \geq \tau n 2^{-\frac{3}{2}j_n}\right\} + \Pr\left\{|U_n(F)| \geq \tau n 2^{-\frac{3}{2}j_n}\right\}. \tag{1.2.31}$$

By Bernstein's inequality for $L_n(F)$ (see (1.2.7)), for all the measurable sets F and all n, we have

$$\begin{aligned}\Pr\left\{|L_n(F)| \geq \tau n 2^{-\frac{3}{2}j_n}\right\} &\leq 2\exp\left(-\frac{\tau^2 n^2 2^{-3j_n}}{8n2^{-2j_n}\|\Phi\|_2^4 + \frac{16}{3}\tau n 2^{-\frac{5}{2}j_n}\|\Phi\|_2^2}\right) \\ &\leq C\exp\left(-\frac{1}{C}\frac{\tau^2 n^2 2^{-3j_n}}{n2^{-2j_n} + \tau n 2^{-\frac{5}{2}j_n}}\right) \\ &\leq C\exp\left(-\frac{1}{C}\min\left(\tau^2 n 2^{-j_n}, \tau n 2^{-\frac{1}{2}j_n}\right)\right),\end{aligned} \tag{1.2.32}$$

where C depends on the scaling function ϕ through Φ. By inequality (1.2.27) there is a constant κ_0', such that

$$\begin{aligned}&\Pr\left\{|U_n(F)| \geq \tau n 2^{-\frac{3}{2}j_n}\right\} = \Pr\left\{\left|\sum_{1\leq i\neq j\leq n} H_{n,F}(X_i, X_j)\right| \geq \tau n 2^{-\frac{3}{2}j_n}\right\} \\ &\leq \kappa_0' \exp\left(-\frac{1}{\kappa_0'}\min\left(\frac{\tau^2}{\int_F f^2(x)dx}, 2^{j_n/2}\tau, \tau^{2/3}n^{1/3}2^{-\frac{1}{3}j_n}, \tau^{1/2}n^{1/2}2^{-\frac{1}{4}j_n}\right)\right).\end{aligned} \tag{1.2.33}$$

for all n large enough. These two inequalities give (1.2.29). □

Now the three terms in the decomposition of $W_n(\mathbb{R}) - W_{n,m}(\mathbb{R})$ in (1.2.5) can be bounded. The first two are of the U-statistics type, so Proposition 1.2.4 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 (1.2.7).

1.2.6 Lemma. Under the same hypotheses of Proposition 1.2.4 on f, ϕ and $\{j_n\}$, there exist a constant κ_0 (depending on f and ϕ) and $\eta > 0$ such that, for all $\epsilon > 0$, $\sigma > 0$, if n is large enough (depending on f, ϕ and $\{j_n\}$), and m is fixed such that $0 \le m < n$,

$$Pr\left\{|W_n(\mathbb{R}) - W_{n,m}(\mathbb{R})| \ge \epsilon\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}\right\} \le \kappa_0 \exp\left(-\frac{\epsilon^2 n^\eta}{\kappa_0}\right). \tag{1.2.34}$$

Proof. By the decomposition (1.2.5), $W_n(\mathbb{R}) - W_{n,m}(\mathbb{R})$ is split into three terms.

$$\begin{aligned}
&\Pr\left\{|W_n(\mathbb{R}) - W_{n,m}(\mathbb{R})| \ge \epsilon\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}\right\}\\
&= \Pr\left\{\left|2\sum_{i=1}^{m}\sum_{j=m+1}^{n} H_n(X_i, X_j) + \sum_{1\le i\ne j\le m} H_n(X_i, X_j) + \sum_{i=1}^{m}(H_n(X_i, X_i) - \mathbb{E}H_n(X_i, X_i))\right|\right.\\
&\qquad \left.\ge \epsilon\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}\right\}\\
&\le \Pr\left\{\left|2\sum_{i=1}^{m}\sum_{j=m+1}^{n} H_n(X_i, X_j)\right| \ge \frac{1}{3}\epsilon\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}\right\}\\
&\quad + \Pr\left\{\left|\sum_{1\le i\ne j\le m} H_n(X_i, X_j)\right| \ge \frac{1}{3}\epsilon\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}\right\}\\
&\quad + \Pr\left\{\left|\sum_{i=1}^{m}(H_n(X_i, X_i) - \mathbb{E}H_n(X_i, X_i))\right| \ge \frac{1}{3}\epsilon\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}\right\}.
\end{aligned} \tag{1.2.35}$$

By (1.2.7), for all n and fixed $m < n$,

$$\begin{aligned}
&\Pr\left\{\left|\sum_{i=1}^{m}(H_n(X_i, X_i) - \mathbb{E}H_n(X_i, X_i))\right| \ge \frac{1}{3}\epsilon\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}\right\}\\
&\le 2\exp\left(-\frac{\frac{2}{9}\epsilon^2\sigma^2 n^2 2^{-3j_n}\log\log n}{8m2^{-2j_n}\|\Phi\|_2^4 + \frac{16}{3}\cdot\frac{1}{3}\epsilon\sigma n 2^{-5j_n/2}\|\Phi\|_2^2\sqrt{2\log\log n}}\right)\\
&\le c_1\exp\left(-\frac{1}{c_1}\min\left(\epsilon^2 n^2 2^{-j_n}\log\log n, \epsilon n 2^{-j_n/2}\sqrt{\log\log n}\right)\right),
\end{aligned} \tag{1.2.36}$$

where c_1 is a constant that depends on m, σ and ϕ. Then by condition (B1), there exists $\eta_1 > 0$ such that the last term is dominated by $c_1' \exp\left(-\frac{1}{c_1'}\epsilon^2 n^{\eta_1}\right)$.

Take $\tau = \frac{1}{3}\epsilon\sigma\sqrt{2\log\log n}$ and $F = \mathbb{R}$ in (1.2.27), for all n large enough depending on f, ϕ and $\{j_n\}$,

$$\begin{aligned}
&\Pr\left\{\left|\sum_{1\le i\neq j\le m} H_n(X_i, X_j)\right| \ge \frac{1}{3}\epsilon\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}\right\} \\
&\le c_2 \exp\left(-\frac{1}{c_2}\min\left[\frac{\frac{2}{9}\epsilon^2\sigma^2 n^2\log\log n}{m^2\int_{\mathbb{R}} f^2(x)dx}, \frac{\frac{1}{3}\epsilon\sigma n 2^{j_n/2}\sqrt{2\log\log n}}{m},\right.\right. \\
&\quad \left.\left.\frac{(\frac{2}{9}\epsilon^2\sigma^2 n^2 2^{-j_n}\log\log n)^{1/3}}{m^{1/3}}, \left(\frac{1}{3}\epsilon\sigma n 2^{-j_n/2}\sqrt{2\log\log n}\right)^{1/2}\right]\right) \\
&= c_2' \exp\left(-\frac{1}{c_2'}\min\left[\frac{\epsilon^2 n^2\log\log n}{m^2}, \frac{\epsilon n 2^{j_n/2}\sqrt{\log\log n}}{m}, \frac{(\epsilon^2 n^2 2^{-j_n}\log\log n)^{1/3}}{m^{1/3}},\right.\right. \\
&\quad \left.\left.\epsilon^{1/2} n^{1/2} 2^{-j_n/4}(\log\log n)^{1/4}\right]\right),
\end{aligned} \tag{1.2.37}$$

where c_2 and c_2' depends on f, ϕ and σ. Since $2^{-j_n} \asymp n^{-\delta}$, for some $\delta \in (0, 1/3)$, and for every fixed m such that $0 \le m < n$, this is

$$\begin{aligned}
&\le c_2'' \exp\left(-\frac{1}{c_2''}\min\left[\epsilon^2 n^2\log\log n, \epsilon n^{1+\delta/2}\sqrt{\log\log n}, \epsilon^{2/3} n^{2/3-\delta/3}(\log\log n)^{1/3},\right.\right. \\
&\quad \left.\left.\epsilon^{1/2} n^{1/2-\delta/4}(\log\log n)^{1/4}\right]\right) \\
&\le c_2'' \exp\left(-\frac{1}{c_2''}\epsilon^2 n^{\eta_2}\right)
\end{aligned} \tag{1.2.38}$$

for some $\eta_2 > 0$. For the first summand in (1.2.35), by (1.2.28), through similar calculations, we have

$$\Pr\left\{\left|\sum_{i=1}^{m}\sum_{j=m+1}^{n} H_n(X_i, X_j)\right| \ge \frac{1}{6}\epsilon\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}\right\} \le c_3 \exp\left(-\frac{1}{c_3'}\epsilon^2 n^{\eta_3}\right) \tag{1.2.39}$$

for some $\eta_3 > 0$. Gathering the bounds for the three summands in (1.2.35), we conclude that there exists $\eta > 0$ and $\kappa_0 < \infty$, such that for all $0 < \epsilon < 1$ and all n large enough,

$$Pr\left\{|W_n(\mathbb{R}) - W_{n,m}(\mathbb{R})| \geq \epsilon\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}\right\} \leq \kappa_0 \exp\left(-\frac{\epsilon^2 n^\eta}{\kappa_0}\right). \tag{1.2.40}$$

□

1.3 Moderate Deviations

In this section, we'll prove a moderate deviation result for $W_n([-M, M])$, $M > 0$. 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 Giné and Mason (2004) is used for the Gaussian chaos. We assume throughout this section that ϕ satisfies both (S1) and (S2).

Given a sequence of i.i.d. random variables $X, X_1, ..., X_n$, $n \in \mathbb{N}$ with Lebesgue density f. Let $F(t)$ be the distribution function. Let $F_n(t)$ be the empirical distribution function, that is

$$F_n(t) := \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq t) \tag{1.3.1}$$

For all $x \in \mathbb{R}$, set

$$\begin{aligned} E_n(x) :&= \sqrt{n2^{-j_n}}[f_{n,K}(x) - \mathbb{E}f_{n,K}(x)] = \sqrt{\frac{2^{j_n}}{n}}\sum_{i=1}^{n}[K(2^{j_n}x, 2^{j_n}X_i) - \mathbb{E}K(2^{j_n}x, 2^{j_n}X)] \\ &= \sqrt{n2^{j_n}}\int_{\mathbb{R}} K(2^{j_n}x, 2^{j_n}t)d[F_n(t) - F(t)]. \end{aligned} \tag{1.3.2}$$

Define $K_{n,x}(t) := K(2^{j_n}x, 2^{j_n}t)$ and let $\mu_{K_{n,x}}(t)$ be the Borel measure associated with $K_{n,x}(t)$.

Define the Gaussian process

$$\Gamma_n(x) := 2^{j_n/2} \int_{\mathbb{R}} [B_n(F(x)) - B_n(F(t))] d\mu_{K_{n,x}}(t), \tag{1.3.3}$$

where B_n is a sequence of Brownian bridges. By definition of $W_n([-M, M])$, we may write

$$\frac{2^{3j_n/2}}{n} W_n([-M, M]) = 2^{j_n/2} \int_{-M}^{M} \left[(E_n(t))^2 - \mathbb{E}(E_n(t))^2)\right] dt. \tag{1.3.4}$$

We want to approximate it by the Gaussian chaos:

$$2^{j_n/2} \int_{-M}^{M} \left[(\Gamma_n(t))^2 - \mathbb{E}((\Gamma_n(t))^2)\right] dt. \tag{1.3.5}$$

In order to apply the KMT theorem (to be stated below) to control the difference of the two statistics, we need an integration by parts formula for $E_n(x)$. We check two conditions first (Exercise 3.34, Folland, 1999): (i)$F_n(t) - F(t)$ and $K_{n,x}(t)$ are in the space NBV, where NBV is defined by

$$NBV = \{G \ is \ of \ bounded \ variation, \ G \ is \ right \ continuous \ and \ G(-\infty) = 0\}. \tag{1.3.6}$$

(ii) For fixed $N > 0$ and all $t \in [-N, N]$, the probability that $F_n(t) - F(t)$ and $K_{n,x}(t)$ are both discontinuous is 0.

It's obvious that, with probability one, $F_n(t) - F(t)$ is right continuous and $\lim_{t \to -\infty}(F_n(t) - F(t)) = 0$. $F_n(t)$ and $F(t)$ are both bounded and increasing, so $F_n(t) - F(t)$ is of bounded variation. $K_{n,x}(t)$ is also right continuous.

$$\lim_{t\to-\infty} K_{n,x}(t) = \lim_{t\to-\infty} \sum_k \phi\left(2^{j_n}x-k\right)\phi\left(2^{j_n}t-k\right). \tag{1.3.7}$$

For fixed x, this is a finite sum of functions with compact support. So,

$$\lim_{t\to-\infty} K_{n,x}(t) = 0. \tag{1.3.8}$$

Similarly,

$$\lim_{t\to\infty} K_{n,x}(t) = 0. \tag{1.3.9}$$

To see that $K_{n,x}(t)$ is of bounded variation, let $T_{K,n,x}(t) := \sup\{\sum_{l=1}^{m} |K_{n,x}(t_l) - K_{n,x}(t_{l-1})| : m \in \mathbb{N}, -\infty < t_0 < ... < t_m = t\}$. For any $m \in \mathbb{N}$ and any partition over $(-\infty, t)$,

$$\begin{aligned}\sum_{l=1}^{m} |K_{n,x}(t_l) - K_{n,x}(t_{l-1})| &= \sum_{l=1}^{m} \left| \sum_k \phi(2^{j_n}x-k)\phi(2^{j_n}t_l-k) - \sum_k \phi(2^{j_n}x-k)\phi(2^{j_n}t_{l-1}-k) \right| \\ &\le \sum_k |\phi(2^{j_n}x-k)| \sum_{l=1}^{m} |\phi(2^{j_n}t_l-k) - \phi(2^{j_n}t_{l-1}-k)| \\ &\le \sum_k |\phi(2^{j_n}x-k)| \|\phi(2^{j_n}\cdot-k)\|_v. \end{aligned} \tag{1.3.10}$$

Now if $t_0 < ... < t_m$ is a partition, so is $2^{j_n}t_0 - k < ... < 2^{j_n}t_m - k$. Also recall ϕ satisfies (1.1.1) and (S2), then we have, for almost every x,

$$\|K_{n,x}\|_v = \lim_{t\to\infty} T_{K,n,x}(t) \le \sum_k |\phi(2^{j_n}x-k)| \|\phi\|_v := C_\phi, \tag{1.3.11}$$

where C_ϕ is a constant that depends only on the scaling function ϕ.

Now we check (ii). $K_{n,x}(t)$ could only have discontinuities at countably many points, say, $\{e_j\}$, $j \in \mathbb{N}$. $F_n(t) - F(t)$ could only have discontinuities at $X_i, i = 1, ..., n$.

$$\begin{aligned}
&\Pr\left\{F_n(t) - F(t) \text{ is discontinuous at } e_j, j \in \mathbb{N}\right\} \\
&= \Pr\left\{X_i = e_j, 1 \le i \le n, j \in \mathbb{N}\right\} \\
&\le n\Pr\left\{X = e_j, j \in \mathbb{N}\right\} = 0.
\end{aligned} \tag{1.3.12}$$

Thus with probability one, for fixed $N > 0$, there are no points in $[-N, N]$ where $F_n(t) - F(t)$ and $K_{n,x}(t)$ are both discontinuous.

Let $\mu_{F'}$ be the signed Borel measure associated with $F_n(t) - F(t)$. We then apply the integration by parts formula (Exercise 3.34, Folland, 1999) to the integral $\int_{-N}^{N} K(2^{j_n}x, 2^{j_n}t)d[F_n(t) - F(t)]$. With probability one,

$$\begin{aligned}
&\int_{[-N,N]} K_{n,x}(t)d[F_n(t) - F(t)] + \int_{[-N,N]} (F_n(t) - F(t))d\mu_{K_{n,x}}(t) \\
&= K_{n,x}(N)\left(F_n(N) - F(N)\right) - K_{n,x}\left((-N)-\right)\left(F_n((-N)-) - F((-N)-)\right).
\end{aligned} \tag{1.3.13}$$

Due to (1.3.8) and (1.3.9), the right hand side of (1.3.13) approaches zero as $N \to \infty$. For the left hand side, we notice

$$\left|K_{n,x}(t)1_{[-N,N]}\right| \le \Phi(2^{j_n}(x - t)), \tag{1.3.14}$$

$$\int \Phi(2^{j_n}(x - t))d|\mu_{F'}| \le C\|F_n - F\|_v < \infty, \tag{1.3.15}$$

and

$$|F_n(t) - F(t)|1_{[-N,N]} \le 2. \tag{1.3.16}$$

Applying Lebesgue dominated convergence,

$$\int 2d|\mu_{K_{n,x}}|(t) = 2\|K_{n,x}\|_v < \infty. \tag{1.3.17}$$

The left hand side converges to $\int_{\mathbb{R}} K_{n,x}(t)d[F_n(t)-F(t)] + \int_{\mathbb{R}}(F_n(t)-F(t))d\mu_{K_{n,x}}(t)$. Letting $N \to \infty$ on both sides of (1.3.13), we have shown that

$$\int_{\mathbb{R}} K_{n,x}(t)d[F_n(t) - F(t)] = -\int_{\mathbb{R}}(F_n(t) - F(t))d\mu_{K_{n,x}}(t). \tag{1.3.18}$$

Thus in (1.3.2), with probability one,

$$\begin{aligned} E_n(x) &= \sqrt{n2^{j_n}} \int_{\mathbb{R}} [F(t) - F_n(t)]d\mu_{K_{n,x}}(t) \\ &= \sqrt{n2^{j_n}} \int_{\mathbb{R}} [F(t) - F_n(t) - (F(x) - F_n(x))]d\mu_{K_{n,x}}(t). \end{aligned} \tag{1.3.19}$$

The last equality follows since by (1.3.8) and (1.3.9),

$$\begin{aligned} \int_{\mathbb{R}} [F(x) - F_n(x)]d\mu_{K_{n,x}}(t) &= (F(x) - F_n(x)) \int_{\mathbb{R}} d\mu_{K_{n,x}}(t) \\ &= (F(x) - F_n(x)) \lim_{N\to\infty} |K_{n,x}(N) - K_{n,x}(-N)| \\ &= 0. \end{aligned} \tag{1.3.20}$$

Let $\alpha_n(t) := \sqrt{n}\,[F_n(t) - F(t)]$ and $D_n := \sup_{-\infty<t<\infty} |\alpha_n(t) - B_n(F(t))|$. By the definition of $\Gamma_n(x)$, (1.3.19) and (1.3.11),

$$\begin{aligned} &|E_n(x) - \Gamma_n(x)| \\ &= \left| 2^{j_n/2} \int_{\mathbb{R}} [\alpha_n(x) - \alpha_n(t)]d\mu_{K_{n,x}}(t) - 2^{j_n/2} \int_{\mathbb{R}} [B_n(F(x)) - B_n(F(t))]d\mu_{K_{n,x}}(t) \right| \\ &\leq 2^{j_n/2} 2D_n \int_{\mathbb{R}} d|\mu_{K_{n,x}}| \leq 2^{j_n/2} 2D_n C_\phi < \infty. \end{aligned} \tag{1.3.21}$$

Similar to the above,

$$\operatorname*{ess\,sup}_x(|E_n(x)| + |\Gamma_n(x)|) \leq 2^{j_n/2} \cdot 2C_\phi(\|\alpha_n\|_\infty + \|B_n\|_\infty). \tag{1.3.22}$$

Next we'll prove a bound on the size of the difference between (1.3.4) and (1.3.5). With probability one,

$$\begin{aligned} D_n(M) :&= \left| \frac{2^{3j_n/2}}{n} W_n([-M, M]) - 2^{j_n/2} \int_{-M}^{M} \Big((\Gamma_n(t))^2 - \mathbb{E}((\Gamma_n(t))^2) \Big)\, dt \right| \\ &= 2^{j_n/2} \left| \int_{-M}^{M} \Big((E_n(t))^2 - \mathbb{E}((E_n(t))^2) \Big)\, dt - \int_{-M}^{M} \Big((\Gamma_n(t))^2 - \mathbb{E}((\Gamma_n(t))^2) \Big)\, dt \right| \\ &= 2^{j_n/2} \left| \int_{-M}^{M} \Big[(E_n(t))^2 - (\Gamma_n(t))^2 \Big]\, dt \right| \\ &\leq 2^{j_n/2} \operatorname*{ess\,sup}_x(|E_n(x)| + |\Gamma_n(x)|) \int_{-M}^{M} |E_n(t) - \Gamma_n(t)| dt. \end{aligned} \tag{1.3.23}$$

Using the bounds in (1.3.21) and (1.3.22), we then have

$$\begin{aligned} D_n(M) &\leq 2^{j_n/2} \operatorname*{ess\,sup}_x(|E_n(x)| + |\Gamma_n(x)|) 2^{j_n/2} \cdot 2D_n C_\phi 2M \\ &\leq 2^{3j_n/2} 8MD_n(\|\alpha_n\|_\infty + \|B_n\|_\infty) C_\phi^2. \end{aligned} \tag{1.3.24}$$

We use the KMT theorem for D_n and the Dvoretzky-Kiefer-Wolfowitz (DKW) inequalities for $\|\alpha_n\|_\infty$ and $\|B_n\|_\infty$.

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

$$\Pr\left\{ D_n \geq n^{-1/2}(a \log n + x) \right\} \leq b \exp(-cx), \tag{1.3.25}$$

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

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

$$Pr\{\|\alpha_n\|_\infty > z\} \leq 2\exp(-2z^2), \quad Pr\{\|B_n\|_\infty > z\} \leq 2\exp(-2z^2). \tag{1.3.26}$$

These inequalities lead to the following proposition.

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

$$Pr\left\{D_n(M) \geq \frac{C_{M,\phi}(\log n)^2}{2^{-3j_n/2}\sqrt{n}}\right\} \leq n^{-\gamma} \tag{1.3.27}$$

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

Proof. For $\gamma > 0$, take $x = 2\gamma \log n/c$ in (1.3.25), where c is the constant in (1.3.25). For n large enough depending on γ,

$$2\gamma \log n \geq \gamma \log n + \log 2b. \tag{1.3.28}$$

Then by (1.3.25) in the KMT theorem,

$$Pr\left\{D_n \geq \frac{1}{\sqrt{n}}\left(a + \frac{2\gamma}{c}\right)\log n\right\} \leq b\exp\left(-2\gamma \log n\right) \leq \frac{1}{2}n^{-\gamma}. \tag{1.3.29}$$

From DKW inequalities (1.3.26), it is easy to see that for n large enough,

$$\begin{aligned} Pr\left\{\|\alpha_n\|_\infty + \|B_n\|_\infty > \frac{\log n}{a + 2\gamma/c}\right\} &\leq 4\exp\left(-\frac{(\log n)^2}{2\left(a + \frac{2\gamma}{c}\right)^2}\right) \\ &\leq \frac{1}{2}n^{-\gamma}. \end{aligned} \tag{1.3.30}$$

Using (1.3.24),

$$\begin{aligned} &\Pr\left\{D_n(M) \geq \frac{8MC_\phi^2(\log n)^2}{2^{-3j_n/2}\sqrt{n}}\right\} \\ &\leq \Pr\left\{D_n\left(\|\alpha_n\|_\infty + \|B_n\|_\infty\right) \geq \frac{(\log n)^2}{\sqrt{n}}\right\} \\ &\leq \Pr\left\{D_n \geq \frac{1}{\sqrt{n}}\left(a + \frac{2\gamma}{c}\right)\log n\right\} + \Pr\left\{\|\alpha_n\|_\infty + \|B_n\|_\infty > \frac{\log n}{a + 2\gamma/c}\right\} \\ &\leq n^{-\gamma}. \end{aligned} \tag{1.3.31}$$

Now take $C_{M,\phi} = 8MC_\phi^2$ to prove (1.3.27). □

Next we derive a moderate deviation result for the Gaussian chaos $2^{j_n/2}\int_{-M}^{M}[(\Gamma_n(t))^2 - \mathbb{E}((\Gamma_n(t))^2))]dt$. It is easier to obtain a moderate deviation result for this than for $2^{3j_n/2}W_n([-M,M])/n$. We first justify that it can be written as a sum of independent random variables. We have,

$$\mathbb{E}\Gamma_n(x) = 2^{j_n/2}\int_{\mathbb{R}} \mathbb{E}[B_n(F(x)) - B_n(F(t))]d\mu_{K_{n,x}}(t) = 0. \tag{1.3.32}$$

Recall the definition of R_n in (1.2.18). Using (1.3.2), we can verify that

$$\mathbb{E}\left[E_n(s)E_n(t)\right] = R_n(s,t). \tag{1.3.33}$$

Comparing (1.3.3) and (1.3.19), and using

$$n\mathbb{E}\left[F(t) - F_n(t)\right]\left[F(s) - F_n(s)\right] = \mathbb{E}[B_n(F(t))B_n(F(s))] = F(t \wedge s) - F(s)F(t), \tag{1.3.34}$$

we get

$$\mathbb{E}\left[\Gamma_n(s)\Gamma_n(t)\right] = \mathbb{E}\left[E_n(s)E_n(t)\right] = R_n(s,t). \tag{1.3.35}$$

By Fubini and (A.0.4), for any subset F satisfying conditions (1.1.9),

$$\mathbb{E}\left(\int_F \Gamma_n^2(s)ds\right) = \int_F R_n(s,s)ds \leq 2^{j_n} \int_{\mathbb{R}} \int_{\mathbb{R}} \bar{K}_n^2(s,u) f(u) du ds \\ \leq 4\|\Phi\|_2^2. \tag{1.3.36}$$

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

$$\mathcal{R}_{n,F}\varphi(s) = \int_F R_n(s,t)\varphi(t)dt. \tag{1.3.37}$$

The following well-known proposition shows that the Gaussian chaos can be written as a sum of weighted, centered chi-squared random variables, e.g.,

1.3.3 Proposition. (Giné and Mason, 2004) A centered non-degenerate Gaussian process $\{\Gamma(t),\, t \in F\}$, for F a Borel subset of $\mathbb{R}$, with covariance function

$$R(s,t) = \mathbb{E}(\Gamma(t)\Gamma(s)), s, t \in F \tag{1.3.38}$$

has a version with all of its sample paths in $L_2(F)$ if and only if

$$0 < \int_F R(s,s)ds < \infty. \tag{1.3.39}$$

If this is the case, then

$$0 < \int_{F^2} R^2(s,t)dsdt < \infty, \tag{1.3.40}$$

and the spectrum of the operator

$$\mathcal{R}\varphi(s) = \int_F R(s,t)\varphi(t)dt, \quad \varphi \in L_2(F) \tag{1.3.41}$$

consists of a sequence of non-negative eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq 0$ in ℓ_1, corresponding to eigenvectors $e_1, e_2, ...$, that can be taken to be orthonormal, in which case $R(s,t) = \sum_i \lambda_i e_i(s) e_i(t)$ in the $L_2(F \times F)$ sense; moreover, for this sequence of eigenvalues and eigenvectors,

$$\Gamma(t) \stackrel{d}{=} \sum_{k=1}^{\infty} \lambda_k^{1/2} e_k(t) Z_k, \tag{1.3.42}$$

$$\int_F \left[(\Gamma(t))^2 - \mathbb{E}(\Gamma(t))^2 \right] dt = \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1), \tag{1.3.43}$$

where $Z_1, Z_2, ...,$ are i.i.d $N(0,1)$ random variables and

$$\sum_{k=1}^{\infty} \lambda_k = \int_F R(s,s) ds, \tag{1.3.44}$$

and

$$\sum_{k=1}^{\infty} \lambda_k^2 = \int_{F^2} R^2(s,t) ds dt. \tag{1.3.45}$$

According to Proposition 1.3.3, $\{\Gamma_n(t), t \in [-M, M]\}$ has a version with all of its sample paths in $L_2([-M, M])$. It also follows from the proposition that,

$$\int_{-M}^{M} \left[(\Gamma_n(t))^2 - \mathbb{E}(\Gamma_n(t))^2) \right] dt = \sum_{k=1}^{\infty} \lambda_{n,k} (Z_k^2 - 1), \tag{1.3.46}$$

where $\lambda_{n,1} \geq \lambda_{n,2} \geq \ldots \geq 0$ are the eigenvalues of the operator $\mathcal{R}_{n,F}$ where $F = [-M, M]$. For some $0 < \eta \leq 1$, $\mathbb{E}|Z_k^2 - 1|^{2+\eta} \leq 1$. We calculate the limiting variance of the Gaussian chaos by Lemmas $A.0.3$ and $A.0.4$,

$$
\begin{aligned}
&\lim_{n\to\infty} 2^{j_n}\mathbb{E}\left[\int_{-M}^{M}\left((\Gamma_n(t))^2 - E(\Gamma_n(t))^2\right)dt\right]^2 \\
&= \lim_{n\to\infty} 2^{j_n}\mathbb{E}\left[\sum_{k=1}^{\infty}\lambda_{n,k}(Z_k^2-1)\right]^2 \\
&= \lim_{n\to\infty} 2\cdot 2^{j_n}\sum_{k=1}^{\infty}\lambda_{n,k}^2 \quad \text{as the partial sums are Cauchy in } L^2(P) \\
&= \lim_{n\to\infty} 2\cdot 2^{j_n}\int_{-M}^{M}\int_{-M}^{M} R_n^2(s,t)dsdt \\
&= 2\int_{-M}^{M} f^2(x)dx \\
&=: \sigma^2(M).
\end{aligned}
\tag{1.3.47}
$$

Set $b_n := (\lambda_{n,1}/\Lambda_n)^{\eta}$, where $\Lambda_n^2 := \sum_{k=1}^{\infty}\lambda_{n,k}^2$. Also define

$$
V_n := \frac{1}{\sqrt{2}\Lambda_n}\sum_{k=1}^{\infty}\lambda_{n,k}(Z_k^2-1). \tag{1.3.48}
$$

Corollary 4.6 in Giné and Mason (2004) gives a moderate deviation result for V_n. It was proved using Pinsky's method (Pinsky, 1966). For any sequence a_n converging to infinity at the rate $a_n^2 + \log b_n \to -\infty$, for all $0 < \epsilon < 1$,

$$
\exp\left(-\frac{a_n^2}{2}(1+\epsilon)\right) \le \Pr\{\pm V_n \ge a_n\} \le \exp\left(-\frac{a_n^2}{2}(1-\epsilon)\right) \tag{1.3.49}
$$

if n is large enough depending on ϵ. In order to apply the Corollary, we first check the condition $b_n \to 0$. By Lemma $A.0.9$,

$$
\begin{aligned}
\lambda_{n,1} &\le \sup\{\|\mathcal{R}_{n,[-M,M]}\varphi\|_2 : \|\varphi\|_2 = 1, \varphi \in L_2([-M,M])\} \\
&\le 2^{-j_n}C^{1/2}(\Phi, f).
\end{aligned}
\tag{1.3.50}
$$

And by (1.3.47),

$$\frac{\lambda_{n,1}}{\sqrt{2\sum_{k=1}^{\infty}\lambda_{n,k}^2}} \le \frac{2^{-j_n}C^{1/2}(\Phi,f)}{\sqrt{2\sum_{k=1}^{\infty}\lambda_{n,k}^2}} \asymp \frac{2^{-j_n}C^{1/2}(\Phi,f)}{2^{-j_n/2}\sigma(M)} \asymp 2^{-j_n/2} \to 0. \tag{1.3.51}$$

This gives that $b_n \to 0$. The above calculation also estimates the order of b_n, which determines the size of the moderate deviation a_n. Set

$$V_n(M) := \frac{2^{j_n/2}}{\sigma(M)}\int_{-M}^{M}\Big((\Gamma_n(t))^2 - E(\Gamma_n(t))^2\Big)dt. \tag{1.3.52}$$

Comparing the definitions of V_n and $V_n(M)$ in (1.3.46) and (1.3.48), we see that

$$V_n = c_{n,M}V_n(M), \tag{1.3.53}$$

where $c_{n,M} = \dfrac{2^{-j_n/2}\sigma(M)}{\sqrt{2}\Lambda_n}$, and by the calculation in (1.3.47), $c_{n,M} \to 1$ as $n \to \infty$.

$$\Pr\{\pm V_n(M) \ge a_n\} = \Pr\{\pm V_n \ge a_n c_{n,M}\}. \tag{1.3.54}$$

Take $a_n = C\sqrt{2\log\log n}$, with $0 < C < \infty$. Then $a_n^2 c_{n,M}^2 = C(1+o(1))\log\log n$. By condition (B1), b_n is dominated by $Cn^{-\delta\eta/2}$. So

$$a_n^2 c_{n,M}^2 + \log b_n \le (C+o(1))\log\log n + \log Cn^{-\delta\eta/2}. \tag{1.3.55}$$

Since $(\log n)^{C+o(1)}/n^{\delta\eta/2} \to 0$, the sequence $a_n c_{n,M}$ satisfies that $a_n^2 c_{n,M}^2 + \log b_n \to -\infty$.

Take $a_n c_{n,M}$ instead of a_n in (1.3.49), we have

$$\exp\left(-\frac{a_n^2 c_{n,M}^2}{2}(1+\epsilon)\right) \le \Pr\{\pm V_n \ge a_n c_{n,M}\} \le \exp\left(-\frac{a_n^2 c_{n,M}^2}{2}(1-\epsilon)\right) \tag{1.3.56}$$

for all $0 < \epsilon < 1$ if n is sufficiently large depending on ϵ. Now if given $0 < \epsilon < 1$, choose N so that $c_{n,M} > \sqrt{1-\epsilon}$ for all $n \ge N$. Then there exists some $0 < \epsilon' < 1$ satisfying $c_{n,M}^2(1-\epsilon') \ge 1-\epsilon$. Then choosing $n \ge N$ large enough and using (1.3.50) and (1.3.53), we get

$$\Pr\{\pm V_n(M) \ge a_n\} \le \exp\left(-\frac{a_n^2 c_{n,M}^2}{2}(1-\epsilon')\right) \le \exp\left(-\frac{a_n^2}{2}(1-\epsilon)\right). \tag{1.3.57}$$

Similarly, we also have the inequality in the other direction. So we may replace V_n by $V_n(M)$ in (1.3.49) and obtain, for $a_n = C\sqrt{2\log\log n}$, $0 < C < \infty$,

$$\exp\left(-\frac{a_n^2(1+\epsilon)}{2}\right) \le Pr\left\{\pm V_n(M) \ge a_n\right\} \le \exp\left(-\frac{a_n^2(1-\epsilon)}{2}\right) \tag{1.3.58}$$

for all n large enough depending on ϵ.

We can use this result, the triangle inequality and Proposition 1.3.2 to obtain the following proposition.

1.3.4 Proposition. Let $a_n = C\sqrt{2\log\log n}$, $0 < C < \infty$. Under the hypotheses of Proposition 1.3.2, and further assuming that f satisfies condition (f) and that $\int_{-M}^{M} f^2(x)dx > 0$, then

$$\exp\left(-\frac{a_n^2(1+\epsilon)}{2}\right) - \frac{1}{n^2} \le \Pr\left\{\pm\frac{2^{3j_n/2}}{\sigma(M)n}W_n([-M,M]) \ge a_n\right\} \le \exp\left(-\frac{a_n^2(1-\epsilon)}{2}\right) + \frac{1}{n^2} \tag{1.3.59}$$

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

Proof. Using the definition of $D_n(M)$ in (1.3.23), we write

$$D_n(M) = \left| \frac{2^{3j_n}}{n} W_n\left([-M, M]\right) - V_n(M)\sigma(M) \right|. \tag{1.3.60}$$

Consider the right-side inequality in (1.3.59) first. By Proposition 1.3.2, there exists a constant $C_{M,\phi} > 0$ so that, for n sufficiently large,

$$\Pr\left\{ \frac{1}{\sigma(M)} D_n(M) \geq \frac{C_{M,\phi}(\log n)^2}{\sigma(M) 2^{-3j_n/2}\sqrt{n}} \right\} \leq \frac{1}{n^2}. \tag{1.3.61}$$

Let $d_{n,M} =: \dfrac{C_{M,\phi}(\log n)^2}{\sigma(M) 2^{-3j_n/2}\sqrt{n}}$,

$$\begin{aligned} &\Pr\left\{ \pm \frac{2^{3j_n/2}}{\sigma(M) n} W_n([-M, M]) \geq a_n \right\} \\ &\leq \Pr\left\{ \frac{1}{\sigma(M)} D_n(M) \geq d_{n,M} \right\} + \Pr\left\{ \pm V_n(M) \geq a_n - d_{n,M} \right\}. \end{aligned} \tag{1.3.62}$$

For $\delta \in (0, 1/3)$,

$$d_{n,M} \asymp \frac{C_{M,\phi}(\log n)^2}{\sigma(M) n^{(1-3\delta)/2}} \to 0, \text{ and } \frac{d_{n,M}}{a_n} \asymp \frac{C_{M,\phi} n^{3\delta/2}(\log n)^2}{\sigma(M)\sqrt{n \log\log n}} \to 0. \tag{1.3.63}$$

Thus given $\epsilon > 0$, we can choose n large enough so that $0 < d_{n,M}/a_n < 1 - \sqrt{1-\epsilon}$. And there exists $0 < \epsilon' < 1$ with

$$\left(\frac{a_n - d_{n,M}}{a_n} \right)^2 (1 - \epsilon') \geq 1 - \epsilon. \tag{1.3.64}$$

Since $(a_n - d_{n,M})^2 = a_n^2(1 + o(1))$, $(a_n - d_{n,M})^2 - \log b_n \to -\infty$. Thus (1.3.58) gives,

$$\Pr\left\{ \pm V_n(M) \geq a_n - d_{n,M} \right\} \leq \exp\left(-\frac{(a_n - d_{n,M})^2(1-\epsilon')}{2} \right) \leq \exp\left(-\frac{a_n^2(1-\epsilon)}{2} \right). \tag{1.3.65}$$

(1.3.61), (1.3.62) and (1.3.65) yield that, for n sufficiently large depending on ϵ and M,

$$\Pr\left\{\pm\frac{2^{3j_n/2}}{\sigma(M)n}W_n([-M,M]) \geq a_n\right\} \leq \exp\left(-\frac{a_n^2(1-\epsilon)}{2}\right) + \frac{1}{n^2}. \tag{1.3.66}$$

The left-side inequality in (1.3.59) can be obtained in a similar way. □

1.4 LIL for the wavelet density estimator

In this section, we prove the law of the iterated logarithm for the wavelet density estimator using the tools established in the previous sections. Recall the definition of J_n in (0.1.8). Set $\sigma^2 =: \sigma^2(\infty) = 2\int_{\mathbb{R}} f^2(x)dx$.

1.4.1 Theorem. Let f, ϕ and $\{j_n\}$ satisfy hypotheses (f), (S1), (S2), (B1) and (B2). Then,

$$\limsup_{n\to\infty} \pm\frac{n2^{-j_n/2}}{\sigma\sqrt{2\log\log n}}J_n = 1, \quad a.s. \tag{1.4.1}$$

1.4.2 Remark. Theorem 1.4.1 includes the case when $2^{-j_n} \asymp n^{-1/5}$, where the mean integrated squared error attains its minimum under certain conditions. This is in contrast to kernel density estimators, where it has been proved that (see section 6, Giné and Mason, 2004),

$$\limsup_{n\to\infty} \pm\frac{n2^{-j_n/2}}{\sigma\sqrt{2\log\log n}}J_n = C, \ \ a.s. \tag{1.4.2}$$

when the bandwidth is of the order $n^{-1/5}$. But the constant C has not been determined.

Proof. We will first show that $J_n = \bar{J}_n$, where

$$\bar{J}_n = \|f_{n,K} - \mathbb{E}f_{n,K}\|_2^2 - \mathbb{E}\|f_{n,K} - \mathbb{E}f_{n,K}\|_2^2. \tag{1.4.3}$$

Since we have that,

$$J_n = \int_{\mathbb{R}} f_{n,K}^2 - \mathbb{E} f_{n,K}^2 - 2 f_{n,K} f + 2 f \mathbb{E} f_{n,K}, \tag{1.4.4}$$

and

$$\bar{J}_n = \int_{\mathbb{R}} f_{n,K}^2 - 2 f_{n,K} \mathbb{E} f_{n,K} - \mathbb{E} f_{n,K}^2 + 2 \left(\mathbb{E} f_{n,K}\right)^2. \tag{1.4.5}$$

It remains to show that the difference

$$J_n - \bar{J}_n = 2 \int_{\mathbb{R}} (f - \mathbb{E} f_{n,K})(\mathbb{E} f_{n,K} - f_{n,K}) = 0. \tag{1.4.6}$$

$\mathbb{E} f_{n,K} - f_{n,K}$ is a linear combination of $\{\phi_{0k}\}$ and $\{\psi_{jk}\}$, $0 \le j \le j_n - 1$, whereas $f - \mathbb{E} f_{n,K}$ is a linear combination of $\{\psi_{jk}\}$, $j \ge j_n$. By orthogonality of $\{\phi_{0k}, \psi_{jk}\}$, we have $J_n - \bar{J}_n = 0$. Thus it is equivalent to proving that

$$\limsup_{n\to\infty} \pm \frac{n 2^{-j_n/2}}{\sigma \sqrt{2 \log\log n}} \bar{J}_n = 1, \quad a.s. \tag{1.4.7}$$

i) lower bound

By (1.1.5), it suffices to prove that

$$\limsup_{n\to\infty} \pm \frac{2^{3j_n/2} W_n(\mathbb{R})}{n\sigma \sqrt{2\log\log n}} \ge 1 \;\; a.s. \tag{1.4.8}$$

By Lemma 1.2.6 and Borel-Cantelli, for every $0 \le m < n$,

$$\frac{|W_n(\mathbb{R}) - W_{n,m}(\mathbb{R})|}{n\sigma 2^{-3j_n/2} \sqrt{2\log\log n}} \to 0 \;\; a.s. \tag{1.4.9}$$

as $n \to \infty$.

From the definition of $W_{n,m}(\mathbb{R})$ in (1.2.1), we observe that it does not depend on $X_1, ..., X_m$, so $\limsup_n \frac{W_n(\mathbb{R})}{\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}}$ is measurable with respect to the tail σ-algebra of $\{X_i\}$. Assume the lower bound (1.4.8) is not true. Then there exists $c < 1$ such that,

$$\limsup_n \frac{W_n(\mathbb{R})}{\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}} = c \quad a.s., \tag{1.4.10}$$

and for $r_k = k^k$,

$$\limsup_k \frac{W_{r_k}(\mathbb{R})}{\sigma r_k 2^{-3j_{r_k}/2}\sqrt{2\log\log r_k}} = c' \leq c \quad a.s. \tag{1.4.11}$$

Although $r_{k-1} \to \infty$ as $k \to \infty$, the argument in Lemma 1.2.6 still applies to $W_{r_k}(\mathbb{R}) - W_{r_k,r_{k-1}}(\mathbb{R})$ due to the fact that $r_k/r_{k-1} \geq k$, and by Borel-Cantelli, we have

$$\frac{|W_{r_k}(\mathbb{R}) - W_{r_k,r_{k-1}}(\mathbb{R})|}{r_k \sigma 2^{-3j_{r_k}/2}\sqrt{2\log\log r_k}} \to 0 \quad a.s. \tag{1.4.12}$$

Thus

$$\limsup_k \frac{W_{r_k,r_{k-1}}(\mathbb{R})}{\sigma r_k 2^{-3j_{r_k}/2}\sqrt{2\log\log r_k}} = c' \leq c \quad a.s. \tag{1.4.13}$$

Let $c'' \in (c', 1)$, and set

$$A_k = \left\{ \frac{W_{r_k,r_{k-1}}(\mathbb{R})}{\sigma r_k 2^{-3j_{r_k}/2}\sqrt{2\log\log r_k}} \geq c'' \right\}, \tag{1.4.14}$$

then $\Pr(A_k \;\; i.o.) = 0$. Since the variables $W_{r_k,r_{k-1}}(\mathbb{R})$ are independent from each other for all k, by Borel-Cantelli,

$$\sum_k \Pr\left\{\frac{W_{r_k,r_{k-1}}(\mathbb{R})}{\sigma r_k 2^{-3j_{r_k}/2}\sqrt{2\log\log r_k}} \geq c''\right\} < \infty. \tag{1.4.15}$$

Set $m_k := r_k - r_{k-1}$ and define

$$W'_{m_k}(F) = \int_F \left(\sum_{i=1}^{r_k-r_{k-1}} \bar{K}(2^{j_{r_k}}t, 2^{j_{r_k}}X_i)\right)^2 dt - \mathbb{E}\int_F \left(\sum_{i=1}^{r_k-r_{k-1}} \bar{K}(2^{j_{r_k}}t, 2^{j_{r_k}}X_i)\right)^2 dt. \tag{1.4.16}$$

Comparing $W'_{m_k}(F)$ to $W_{r_k,r_{k-1}}(F)$, where

$$W_{r_k,r_{k-1}}(F) := \int_F \left[\left(\sum_{r_{k-1}<i\leq r_k} \bar{K}(2^{j_{r_k}}t, 2^{j_{r_k}}X_i)\right)^2 - \mathbb{E}\left(\sum_{r_{k-1}<i\leq r_k} \bar{K}(2^{j_{r_k}}t, 2^{j_{r_k}}X_i)\right)^2\right] dt, \tag{1.4.17}$$

we see that they have the same distribution. Hence (1.4.15) is true with $W_{r_k,r_{k-1}}(\mathbb{R})$ replaced by $W'_{m_k}(\mathbb{R})$. Since $m_k/r_k \to 1$, we may find $c''' \in (c'', 1)$, so that

$$\sum_k \Pr\left\{W'_{m_k}(\mathbb{R}) \geq c'''\sigma m_k 2^{-3j_{r_k}/2}\sqrt{2\log\log m_k}\right\} < \infty. \tag{1.4.18}$$

For any $\varsigma > 0$ satisfying $0 < c'''(1+\varsigma) < 1$, choose M_0 so that $\int_{[-M_0,M_0]^C} f^2(x)dx < \varsigma^2 c'''^2\sigma^2/\kappa_0$, where κ_0 is the constant in (1.2.30). Since

$$W'_{m_k}([-M_0, M_0]) \leq W'_{m_k}(\mathbb{R}) + |W'_{m_k}([-M_0, M_0]^C)|, \tag{1.4.19}$$

we have

$$\begin{aligned} &\Pr\left\{W'_{m_k}(\mathbb{R}) \geq c'''\sigma m_k 2^{-3j_{r_k}/2}\sqrt{2\log\log m_k}\right\} \\ &\geq \Pr\left\{W'_{m_k}([-M_0, M_0]) \geq c'''(1+\varsigma)\sigma m_k 2^{-3j_{r_k}/2}\sqrt{2\log\log m_k}\right\} \\ &\quad - \Pr\left\{|W'_{m_k}\left([-M_0, M_0]^C\right)| \geq c'''\varsigma\sigma m_k 2^{-3j_{r_k}/2}\sqrt{2\log\log m_k}\right\}. \end{aligned} \tag{1.4.20}$$

By the tail estimation result (1.2.30), we obtain

$$\begin{aligned}&\sum_k \Pr\left\{W'_{m_k}([-M_0, M_0]^C) \geq c''' \varsigma \sigma m_k 2^{-3j_{r_k}/2} \sqrt{2\log\log m_k}\right\} \\ &\leq \sum_k \kappa_0 \exp\left(-\frac{2c'''^2 \varsigma^2 \sigma^2 \log\log m_k}{\kappa_0 \int_{[-M_0,M_0]^C} f^2(x)dx}\right) < \infty.\end{aligned} \tag{1.4.21}$$

Fix $\epsilon \in (0,1)$, so that $c'''^2(1+\varsigma)^2(1+\epsilon)^3 < 1$. Now let M_0 be large enough so that $\sigma/\sigma(M_0) < 1+\epsilon$. Letting $a_n = c'''(1+\varsigma)(1+\epsilon)\sqrt{2\log\log m_k}$ in the moderate deviation result (1.3.59), we have

$$\begin{aligned}&\sum_k \Pr\left\{W'_{m_k}([-M_0, M_0]) \geq c'''(1+\varsigma)\sigma m_k 2^{-3j_{r_k}/2} \sqrt{2\log\log m_k}\right\} \\ &= \sum_k \Pr\left\{W'_{m_k}([-M_0, M_0]) \geq c'''(1+\varsigma)\sigma(M_0)\frac{\sigma}{\sigma(M_0)} m_k 2^{-3j_{r_k}/2} \sqrt{2\log\log m_k}\right\} \\ &\geq \sum_k \Pr\left\{W'_{m_k}([-M_0, M_0]) \geq c'''(1+\varsigma)(1+\epsilon)\sigma(M_0) m_k 2^{-3j_{r_k}/2} \sqrt{2\log\log m_k}\right\} \\ &\geq \sum_k \left(\exp(-c'''^2(1+\varsigma)^2(1+\epsilon)^3 \log\log m_k) - \frac{1}{m_k^2}\right).\end{aligned} \tag{1.4.22}$$

But this is a divergent series, so

$$\sum_k \Pr\left\{W'_{m_k}([-M_0, M_0]) \geq c'''(1+\varsigma)\sigma m_k 2^{-3j_{r_k}/2} \sqrt{2\log\log m_k}\right\} = \infty. \tag{1.4.23}$$

Combining (1.4.20), (1.4.21) and (1.4.23), we obtain

$$\sum_k \Pr\left\{W'_{m_k}(\mathbb{R}) \geq c''' \sigma m_k 2^{-3j_{r_k}/2} \sqrt{2\log\log m_k}\right\} = \infty, \tag{1.4.24}$$

which is a contradiction to (1.4.18). Thus the lower bound for the LIL, (1.4.8) is proved.

(ii) Upper bound

Let λ_k be an increasing sequence of positive constants satisfying condition (B2). We shall use (B1) and (B2) to introduce a blocking and reduce $W_n(\mathbb{R})$ to $W_{n_k}(\mathbb{R})$ for the sequence $n_k := \min\{n \in \mathbb{N} : n \geq \lambda_k\}$, then $\{n_k\}$ also satisfies the same properties: $n_{k+1}/n_k \to 1, \log\log n_k/\log k \to 1, n_{k+1} - n_k \to \infty$, and 2^{-j_n} is constant for $n \in [n_k, n_{k+1})$, $k \in \mathbb{N}$. For each $k \in \mathbb{N}$, let I_k be the blocks $I_k = [\lambda_k, \lambda_{k+1}) \cap \mathbb{N} = [n_k, n_{k+1}) \cap \mathbb{N}$. So 2^{-j_n} is constant on I_k for all k. And $I_k \neq \emptyset$ for k large enough, say k_0. Now we'll prove the upper bound: using $J_n = \bar{J}_n$ and the definition of $\bar{J}_n$,

$$\limsup_n \frac{|W_n(\mathbb{R})|}{\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}} \leq 1 \ \ a.s. \tag{1.4.25}$$

It suffices to prove that for every $\varsigma > 0$,

$$\limsup_n \frac{|W_n(\mathbb{R})|}{\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}} \leq 1 + \varsigma \ \ a.s. \tag{1.4.26}$$

Suppose we have

$$\sum_{k \geq k_0} \Pr\left\{\max_{n \in I_k} |W_n(\mathbb{R})| > (1+\varsigma)\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\} < \infty. \tag{1.4.27}$$

Then by Borel-Cantelli, with probability 1,

$$\max_{n \in I_k} |W_n(\mathbb{R})| \leq (1+\varsigma)\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k} \tag{1.4.28}$$

when k is sufficiently large. Thus for $n \in I_k$, we have

$$|W_n(\mathbb{R})| \leq (1+\varsigma)\sigma n 2^{-3j_n/2}\sqrt{2\log\log n}, \tag{1.4.29}$$

which would imply (1.4.26). So, it suffices to prove (1.4.27) for every $\varsigma > 0$. We will first prove the following lemma.

1.4.3 Lemma. Under the hypotheses of Theorem 1.4.1,

$$\sum_{k\geq k_0} \Pr\left\{\max_{n\in I_k} |W_n(\mathbb{R}) - W_{n_k}(\mathbb{R})| > \tau\sigma n_k 2^{-3j_{n_k}} \sqrt{2\log\log n_k}\right\} < \infty \tag{1.4.30}$$

for every $\tau > 0$.

If Lemma 1.4.3 holds, then in order to prove (1.4.27) for every $\varsigma > 0$, it suffices to prove

$$\sum_{k\geq k_0} \Pr\left\{|W_{n_k}(\mathbb{R})| > (1+\varsigma)\sigma n_k 2^{-3j_{n_k}/2} \sqrt{2\log\log n_k}\right\} < \infty. \tag{1.4.31}$$

For every $\varsigma > 0$: Observe that, for $0 < \tau < \varsigma$,

$$\begin{aligned}
&\sum_{k\geq k_0} \Pr\left\{\max_{n\in I_k} |W_n(\mathbb{R})| > (1+\varsigma)\sigma n_k 2^{-3j_{n_k}/2} \sqrt{2\log\log n_k}\right\} \qquad (1.4.32)\\
&\leq \sum_{k\geq k_0} \Pr\left\{|W_{n_k}(\mathbb{R})| + \max_{n\in I_k} |W_n(\mathbb{R}) - W_{n_k}(\mathbb{R})| > (1+\varsigma)\sigma n_k 2^{-3j_{n_k}/2} \sqrt{2\log\log n_k}\right\}\\
&\leq \sum_{k\geq k_0} \Pr\left\{\max_{n\in I_k} |W_n(\mathbb{R}) - W_{n_k}(\mathbb{R})| > \tau\sigma n_k 2^{-3j_{n_k}/2} \sqrt{2\log\log n_k}\right\} +\\
&\sum_{k\geq k_0} \Pr\left\{|W_{n_k}(\mathbb{R})| > (1+\varsigma-\tau)\sigma n_k 2^{-3j_{n_k}/2} \sqrt{2\log\log n_k}\right\}.
\end{aligned}$$

So Lemma 1.4.3 reduces (1.4.27) to (1.4.30) and (1.4.31).

Proof. For $n \in I_k$, $k \geq k_0$, using the fact that 2^{-j_n} is constant in each block I_k, we observe

$$H_n(x,y) = H_{n_k}(x,y). \tag{1.4.33}$$

By the definition of $W_n(\mathbb{R})$ in (1.1.6),

$$\begin{aligned} W_{n_k}(\mathbb{R}) &= U_{n_k}(\mathbb{R}) + L_{n_k}(\mathbb{R}) \\ &= \sum_{1\le i\ne j\le n_k} H_{n_k}(X_i, X_j) + \sum_{i=1}^{n_k}(H_{n_k}(X_i, X_i) - \mathbb{E}H_{n_k}(X_i, X_i)) \end{aligned} \tag{1.4.34}$$

By (1.2.5) and the above two equations, we have

$$\begin{aligned} W_n(\mathbb{R}) - W_{n,n_k}(\mathbb{R}) &= 2\sum_{i=1}^{n_k}\sum_{j=n_k+1}^{n} H_{n_k}(X_i, X_j) + \sum_{1\le i\ne j\le n_k} H_{n_k}(X_i, X_j) \\ &\quad + \sum_{i=1}^{n_k}(H_{n_k}(X_i, X_i) - \mathbb{E}H_{n_k}(X_i, X_i)) \\ &= 2\sum_{i=1}^{n_k}\sum_{j=n_k+1}^{n} H_{n_k}(X_i, X_j) + W_{n_k}(\mathbb{R}). \end{aligned} \tag{1.4.35}$$

Rearranging the equation, we get

$$W_n(\mathbb{R}) - W_{n_k}(\mathbb{R}) = 2\sum_{i=1}^{n_k}\sum_{j=n_k+1}^{n} H_{n_k}(X_i, X_j) + W_{n,n_k}(\mathbb{R}). \tag{1.4.36}$$

Recall Montgomery-Smith maximal inequality (Montgomery-Smith, 1993): If V_i are i.i.d. r.v.'s taking values in a Banach space and $\|\cdot\|$ denotes a norm in a Banach space, then

$$\mathrm{P}r\left\{\max_{1\le k\le n}\left\|\sum_{i=1}^{k} V_i\right\| > t\right\} \le 9\mathrm{P}r\left\{\left\|\sum_{i=1}^{n} V_i\right\| > \frac{t}{30}\right\}. \tag{1.4.37}$$

Let $\mathrm{P}r_{\{>n_k\}}$ be the conditional probability given $X_1, ..., X_{n_k}$, and $\mathrm{P}r_{\{\le n_k\}}$ be the conditional probability with respect to $X_{n_k+1}, ..., X_n$. Application of the Montgomery-Smith inequality to the first summand on the right hand side of (1.4.36) gives,

$$\begin{aligned}
&\mathrm{P}r\left\{\max_{n\in I_k}\left|\sum_{j=n_k+1}^{n}\sum_{i=1}^{n_k}H_{n_k}(X_i,X_j)\right|>t\right\}\\
&=\mathrm{P}r_{\{\le n_k\}}\mathrm{P}r_{\{>n_k\}}\left\{\max_{n\in I_k}\left|\sum_{j=n_k+1}^{n}\sum_{i=1}^{n_k}H_{n_k}(X_i,X_j)\right|>t\right\}\\
&\le 9\mathrm{P}r_{\{\le n_k\}}\mathrm{P}r_{\{>n_k\}}\left\{\left|\sum_{j=n_k+1}^{n_{k+1}-1}\sum_{i=1}^{n_k}H_{n_k}(X_i,X_j)\right|>\frac{t}{30}\right\}\\
&=9\mathrm{P}r\left\{\left|\sum_{j=n_k+1}^{n_{k+1}-1}\sum_{i=1}^{n_k}H_{n_k}(X_i,X_j)\right|>\frac{t}{30}\right\}
\end{aligned}\tag{1.4.38}$$

for $t>0$. By definition of $W_{n,n_k}(\mathbb{R})$ in (1.2.1), we split the second summand in (1.4.36) into two terms.

$$\begin{aligned}
W_{n,n_k}(\mathbb{R}) &= \sum_{n_k<i\ne j\le n}H_{n_k}(X_i,X_j)+\sum_{i=n_k+1}^{n}\left(H_{n_k}(X_i,X_i)-\mathbb{E}H_{n_k}(X_i,X_i)\right)\\
&=: U_{n,n_k}+L_{n,n_k}.
\end{aligned}\tag{1.4.39}$$

To control the size of $\mathrm{P}r\{\max_{n\in I_k}|U_{n,n_k}|>t\}$, we need to cancel the maximum over $n\in I_k$. However, since U_{n,n_k} is not a sum of i.i.d. variables, we cannot apply (1.4.37) directly. In order to achieve this, we use a decoupling here to add more independence. We first justify that $\max_{n\in I_k}|U_{n,n_k}|$ is a norm in a Banach space. Let ℓ_m^∞ denote $\mathbb{R}^m$ with the norm $\|(a_1,...,a_m)\|=\max_{1\le i\le m}|a_i|$. Define $\ell^\infty_{n_{k+1}-1}-$valued function: For $i\ne j$,

$$\tilde{H}_{n_k,i,j}(x,y):=\{\underbrace{0,...0}_{i\vee j-1},\underbrace{H_{n_k},...,H_{n_k}}_{n_{k+1}-i\vee j}\}\in\ell^\infty_{n_{k+1}-1},\tag{1.4.40}$$

and

$$\tilde{H}_{n_k,i,i}(x,y):=0.\tag{1.4.41}$$

For $i \neq j$, the $n-$th coordinate of $\tilde{H}_{n_k,i,j}$ is not zero if and only if $i \vee j \leq n$. So the $n-$th coordinate of $\sum_{n_k<i,j\leq n_{k+1}-1} \tilde{H}_{n_k,i,j}$ is $\sum_{n_k<i\neq j\leq n} H_{n_k}(X_i, X_j)$. Then,

$$\max_{n\in I_k} |U_{n,n_k}| = \max_{n\in I_k} \left| \sum_{n_k<i\neq j\leq n} H_{n_k}(X_i, X_j) \right| = \left\| \sum_{n_k<i,j\leq n_{k+1}-1} \tilde{H}_{n_k,i,j} \right\|. \tag{1.4.42}$$

So we can apply the de la Peña-Montgomery-Smith decoupling inequality (e.g., de la Peña and Giné, 1999, Theorem 3.4.1) to obtain that there exists a constant $C > 0$, such that

$$\Pr\left\{ \left\| \sum_{n_k<i,j\leq n_{k+1}-1} \tilde{H}_{n_k,i,j}(X_i, X_j) \right\| > t \right\} \leq C\Pr\left\{ \left\| \sum_{n_k<i\neq j\leq n_{k+1}-1} \tilde{H}_{n_k,i,j}(X_i^{(1)}, X_j^{(2)}) \right\| > \frac{t}{C} \right\}, \tag{1.4.43}$$

where $X_i^{(1)}$ and $X_j^{(2)}$, $i, j \in \mathbb{N}$ are i.i.d. copies of X_1. It follows that

$$\Pr\left\{ \max_{n\in I_k} |U_{n,n_k}| > t \right\} \leq C\Pr\left\{ \max_{n\in I_k} |U_{n,n_k}^{dec}| > \frac{t}{C} \right\}, \tag{1.4.44}$$

where $U_{n,n_k}^{dec} = \sum_{n_k<i\neq j\leq n} H_{n_k}(X_i^{(1)}, X_j^{(2)})$. But since U_{n,n_k}^{dec} is not a sum of i.i.d. random variables, in order to apply Montgomery-Smith inequality, we need to add the diagonal first. Let $\Pr^{(1)}$ and $\Pr^{(2)}$ denote the conditional probabilities given $X_i^{(1)}$ and $X_j^{(2)}$. We have

$$\begin{aligned}
&\Pr\left\{ \max_{n\in I_k} \left| \sum_{n_k<i,j\leq n} H_{n_k}(X_i^{(1)}, X_j^{(2)}) \right| > t \right\} \\
&= \Pr^{(2)}\Pr^{(1)}\left\{ \max_{n\in I_k} \left| \sum_{i=n_k+1}^{n} \sum_{j=n_k+1}^{n} H_{n_k}(X_i^{(1)}, X_j^{(2)}) \right| > t \right\} \\
&\leq \Pr^{(2)}\Pr^{(1)}\left\{ \max_{n\in I_k} \max_{m\in I_k} \left| \sum_{i=n_k+1}^{n} \sum_{j=n_k+1}^{m} H_{n_k}(X_i^{(1)}, X_j^{(2)}) \right| > t \right\}
\end{aligned}$$

$$\leq 9\Pr^{(2)}\Pr^{(1)}\left\{\max_{m\in I_k}\left|\sum_{i=n_k+1}^{n_{k+1}-1}\sum_{j=n_k+1}^{m} H_{n_k}(X_i^{(1)}, X_j^{(2)})\right| > \frac{t}{30}\right\}$$

$$= 9\Pr^{(2)}\Pr^{(1)}\left\{\max_{m\in I_k}\left|\sum_{j=n_k+1}^{m}\sum_{i=n_k+1}^{n_{k+1}-1} H_{n_k}(X_i^{(1)}, X_j^{(2)})\right| > \frac{t}{30}\right\}$$

$$\leq 81\Pr\left\{\left|\sum_{j=n_k+1}^{n_{k+1}-1}\sum_{i=n_k+1}^{n_{k+1}-1} H_{n_k}(X_i^{(1)}, X_j^{(2)})\right| > \frac{t}{900}\right\}. \tag{1.4.45}$$

For all $t > 0$, by the decoupling inequality (1.4.43) and Montgometry-Smith inequality for L_{n,n_k}, we have

$$\Pr\left\{\max_{n\in I_k}|W_{n,n_k}(\mathbb{R})| > t\right\}$$

$$= \Pr\left\{\max_{n\in I_k}|U_{n,n_k}(\mathbb{R}) + L_{n,n_k}(\mathbb{R})| > t\right\} \tag{1.4.46}$$

$$\leq \Pr\left\{\max_{n\in I_k}|U_{n,n_k}(\mathbb{R})| > \frac{t}{2}\right\} + \Pr\left\{\max_{n\in I_k}|L_{n,n_k}(\mathbb{R})| > \frac{t}{2}\right\}$$

$$\leq C\Pr\left\{\max_{n\in I_k}|U_{n,n_k}^{dec}(\mathbb{R})| > \frac{t}{2C}\right\} + \Pr\left\{\max_{n\in I_k}|L_{n,n_k}(\mathbb{R})| > \frac{t}{2}\right\}$$

$$\leq C\Pr\left\{\max_{n\in I_k}\left(\left|\sum_{n_k<i,j\leq n} H_{n_k}(X_i^{(1)}, X_j^{(2)})\right| + \left|\sum_{n_k<i\leq n} H_{n_k}(X_i^{(1)}, X_i^{(2)})\right|\right) > \frac{t}{2C}\right\}$$

$$+ \Pr\left\{\max_{n\in I_k}|L_{n,n_k}(\mathbb{R})| > \frac{t}{2}\right\}$$

$$\leq C\Pr\left\{\max_{n\in I_k}\left|\sum_{n_k<i,j\leq n} H_{n_k}(X_i^{(1)}, X_j^{(2)})\right| > \frac{t}{4C}\right\} + C\Pr\left\{\max_{n\in I_k}\left|\sum_{n_k<i\leq n} H_{n_k}(X_i^{(1)}, X_i^{(2)})\right| > \frac{t}{4C}\right\}$$

$$+ \Pr\left\{\max_{n\in I_k}\left|\sum_{i=n_k+1}^{n}(H_{n_k}(X_i, X_i) - \mathbb{E}H_{n_k}(X_i, X_i))\right| > \frac{t}{2}\right\}$$

$$\leq 81C\Pr\left\{\left|\sum_{j=n_k+1}^{n_{k+1}-1}\sum_{i=n_k+1}^{n_{k+1}-1} H_{n_k}(X_i^{(1)}, X_j^{(2)})\right| > \frac{t}{3600C}\right\} + 9C\Pr\left\{\left|\sum_{n_k<i\leq n_{k+1}-1} H_{n_k}(X_i^{(1)}, X_i^{(2)})\right|\right.$$

$$\left. > \frac{t}{120C}\right\} + 9\Pr\left\{\left|\sum_{i=n_k+1}^{n_{k+1}-1}(H_{n_k}(X_i, X_i) - \mathbb{E}H_{n_k}(X_i, X_i))\right| > \frac{t}{60}\right\}.$$

Or, with $C' = \max(C, 1)$,

$$\Pr\left\{\max_{n\in I_k}|W_{n,n_k}(\mathbb{R})| > t\right\} \leq 81C'\Pr\left\{\left|\sum_{j=n_k+1}^{n_{k+1}-1}\sum_{i=n_k+1}^{n_{k+1}-1} H_{n_k}(X_i^{(1)}, X_j^{(2)})\right| > \frac{t}{3600C'}\right\} + 9C'\Pr\left\{\left|\sum_{n_k<i\leq n_{k+1}-1} H_{n_k}(X_i^{(1)}, X_i^{(2)})\right| > \frac{t}{120C'}\right\} + 9C'\Pr\left\{\left|\sum_{i=n_k+1}^{n_{k+1}-1}(H_{n_k}(X_i, X_i) - \mathbb{E}H_{n_k}(X_i, X_i))\right| > \frac{t}{60C'}\right\}. \quad (1.4.47)$$

By (1.4.36),

$$\begin{aligned}
&\sum_{k\geq k_0}\Pr\left\{\max_{n\in I_k}|W_n(\mathbb{R}) - W_{n_k}(\mathbb{R})| > \tau\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\} \quad (1.4.48)\\
&\leq \sum_{k\geq k_0}\Pr\left\{\max_{n\in I_k}\left|2\sum_{i=1}^{n_k}\sum_{j=n_k+1}^{n} H_{n_k}(X_i, X_j)\right| + \max_{n\in I_k}\left|W_{n,n_k}(\mathbb{R})\right| > \tau\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\}\\
&\leq \sum_{k\geq k_0}\Pr\left\{\max_{n\in I_k}\left|\sum_{i=1}^{n_k}\sum_{j=n_k+1}^{n} H_{n_k}(X_i, X_j)\right| > \frac{1}{4}\tau\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\}\\
&\quad + \sum_{k\geq k_0}\Pr\left\{\max_{n\in I_k}|W_{n,n_k}(\mathbb{R})| > \frac{1}{2}\tau\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\}.
\end{aligned}$$

Using (1.4.38) and (1.4.47) to bound these two terms, it follows that the above expression is less than or equal to

$$\begin{aligned}
&\sum_{k\geq k_0} 9\Pr\left\{\left|\sum_{j=n_k+1}^{n_{k+1}-1}\sum_{i=1}^{n_k} H_{n_k}(X_i, X_j)\right| > \frac{1}{120}\tau\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\}\\
&+ \sum_{k\geq k_0} 81C'\Pr\left\{\left|\sum_{j=n_k+1}^{n_{k+1}-1}\sum_{i=n_k+1}^{n_{k+1}-1} H_{n_k}(X_i^{(1)}, X_j^{(2)})\right| > \frac{1}{7200C'}\tau\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\}\\
&+ \sum_{k\geq k_0} 9C'\Pr\left\{\left|\sum_{n_k<i\leq n_{k+1}-1} H_{n_k}(X_i^{(1)}, X_i^{(2)})\right| > \frac{1}{240C'}\tau\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\}
\end{aligned}$$

$$+\sum_{k\geq k_0} 9C'\Pr\left\{\left|\sum_{i=n_k+1}^{n_{k+1}-1}(H_{n_k}(X_i,X_i)-\mathbb{E}H_{n_k}(X_i,X_i))\right| > \frac{1}{120C'}\tau\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\}. \tag{1.4.49}$$

So Lemma 1.4.3 reduces to showing that, for every $\tau > 0$, we have

$$\sum_{k\geq k_0}\Pr\left\{\left|\sum_{j=n_k+1}^{n_{k+1}-1}\sum_{i=1}^{n_k} H_{n_k}(X_i,X_j)\right| > \tau\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\} < \infty, \tag{1.4.50}$$

$$\sum_{k\geq k_0}\Pr\left\{\left|\sum_{j=n_k+1}^{n_{k+1}-1}\sum_{i=n_k+1}^{n_{k+1}-1} H_{n_k}(X_i^{(1)},X_j^{(2)})\right| > \tau\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\} < \infty, \tag{1.4.51}$$

$$\sum_{k\geq k_0}\Pr\left\{\left|\sum_{i=n_k+1}^{n_{k+1}-1} H_{n_k}(X_i^{(1)},X_i^{(2)})\right| > \tau\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\} < \infty, \tag{1.4.52}$$

and

$$\sum_{k\geq k_0}\Pr\left\{\left|\sum_{i=n_k+1}^{n_{k+1}-1}(H_{n_k}(X_i,X_i)-\mathbb{E}H_{n_k}(X_i,X_i))\right| > \tau\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\} < \infty. \tag{1.4.53}$$

Applying (1.2.7) to (1.4.53), we obtain

$$\begin{aligned}
&\sum_{k\geq k_0}\Pr\left\{\left|\sum_{i=n_k+1}^{n_{k+1}-1}(H_{n_k}(X_i,X_i)-\mathbb{E}H_{n_k}(X_i,X_i))\right| > \tau\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\}\\
&\leq \sum_{k\geq k_0} 2\exp\left(-\frac{\tau^2\sigma^2 {n_k}^2 2^{-3j_{n_k}} 2\log\log n_k}{8(n_{k+1}-n_k-1)2^{-2j_{n_k}}\|\Phi\|_2^4 + \frac{16}{3}\tau\sigma n_k 2^{-\frac{5}{2}j_{n_k}}\|\Phi\|_2^2\sqrt{2\log\log n_k}}\right)\\
&\leq 2\sum_{k\geq k_0}\exp\left(-c(\tau,\sigma)\frac{n_k^2 2^{-j_{n_k}}\log\log n_k}{(n_{k+1}-n_k-1)+n_k 2^{-\frac{1}{2}j_{n_k}}\sqrt{\log\log n_k}}\right)\\
&\leq 2\sum_{k\geq k_0}\exp\left(-c'(\tau,\sigma)n_k 2^{-j_{n_k}}\min\left(\frac{n_k\log\log n_k}{n_{k+1}-n_k-1}, 2^{j_{n_k}/2}\sqrt{\log\log n_k}\right)\right),
\end{aligned} \tag{1.4.54}$$

where $c(\tau,\sigma)$ and $c'(\tau,\sigma)$ are constants depending on τ and σ. Due to the assumptions on n_k,

$$2^{j_{n_k}/2}\sqrt{\log\log n_k} \to \infty \text{ as } k \to \infty, \tag{1.4.55}$$

and

$$\frac{n_k}{n_{k+1}-n_k-1} = \frac{1}{\frac{n_{k+1}-n_k-1}{n_k}} \to \infty \text{ as } k \to \infty. \tag{1.4.56}$$

So the k-th summand is dominated, from some k on, by $\exp(-c'(\tau,\sigma)n_k 2^{-j_{n_k}})$ for some $c'(\tau,\sigma) > 0$. Since for some $\epsilon > 0$, $2^{-j_{n_k}} \geq \epsilon n_k^{-\delta}$ when k is large enough, then

$$\exp\left(-c'(\tau,\sigma)n_k 2^{-j_{n_k}}\right) \leq \exp\left(-c'(\tau,\sigma)\epsilon n_k^{1-\delta}\right), \tag{1.4.57}$$

which is the general term of a convergent series.

For (1.4.52), we apply Bernstein's inequality. We check that $\mathbb{E}H_{n_k}(X_i^{(1)}, X_i^{(2)}) = 0$. By (1.2.19),

$$\mathbb{E}H_{n_k}^2(X_i^{(1)}, X_i^{(2)}) \leq 2 \cdot 2^{-3j_{n_k}}\|\Phi\|_1^2\|\Phi\|_2^2 \int_{\mathbb{R}} f^2(x)dx, \tag{1.4.58}$$

and by (A.0.6),

$$\|H_{n_k}(X_i^{(1)}, X_i^{(2)})\|_\infty \leq 4 \cdot 2^{-j_{n_k}}\|\Phi\|_2^2. \tag{1.4.59}$$

So (1.2.6) gives that

$$Pr\left\{\left|\sum_{i=n_k+1}^{n_{k+1}-1} H_{n_k}(X_i^{(1)}, X_i^{(2)})\right| > \tau\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\}$$

$$\leq \exp\left(-c(\tau,\sigma)\frac{n_k^2 2^{-3j_{n_k}}\log\log n_k}{(n_{k+1}-n_k-1)2^{-3j_{n_k}}+2^{-j_{n_k}}n_k 2^{-3j_{n_k}/2}\sqrt{\log\log n_k}}\right) \tag{1.4.60}$$
$$\leq \exp\left(-c'(\tau,\sigma)\min\left(\frac{n_k^2\log\log n_k}{n_{k+1}-n_k-1}, n_k 2^{-j_{n_k}/2}\sqrt{\log\log n_k}\right)\right)$$
$$\leq \exp\left(-c'(\tau,\sigma)n_k 2^{-j_{n_k}/2}\min\left(\frac{n_k 2^{j_{n_k}/2}\log\log n_k}{n_{k+1}-n_k-1}, \sqrt{\log\log n_k}\right)\right),$$

and the two terms in the minimum function approach infinity as $k\to\infty$.

So the last expression is dominated by $\exp(-c'(\tau,\sigma)n_k^{1-\delta/2})$ from some k on, which is also the general term of a convergent series. We use the tail estimation result, Proposition 1.2.4, for (1.4.50) and (1.4.51). For (1.4.50), four terms in the exponent in (1.2.28) are

$$\frac{\tau^2\sigma^2 n_k^2 2\log\log n_k}{n_k(n_{k+1}-n_k-1)\int_{\mathbb{R}} f^2(x)dx}, \quad \frac{\tau\sigma n_k 2^{j_{n_k}/2}\sqrt{2\log\log n_k}}{\sqrt{n_k(n_{k+1}-n_k-1)}}, \tag{1.4.61}$$
$$\left(\frac{\tau^2\sigma^2 n_k^2 2^{-j_{n_k}} 2\log\log n_k}{n_k\vee(n_{k+1}-n_k-1)}\right)^{1/3}, \quad \left(\tau\sigma n_k 2^{-j_{n_k}/2}\sqrt{2\log\log n_k}\right)^{1/2}.$$

The first term is of the order of

$$\frac{n_k^2}{n_k(n_{k+1}-n_k-1)}\log\log n_k \asymp \frac{n_k}{n_{k+1}-n_k}\log\log n_k \asymp \frac{n_k}{n_{k+1}-n_k}\log k, \tag{1.4.62}$$

and

$$M_k := \frac{n_k}{n_{k+1}-n_k} = \frac{1}{n_{k+1}/n_k-1}\to\infty. \tag{1.4.63}$$

So if the first term dominates, then from some k on, it is greater than $2\log k$, and we have

$$\Pr\left\{\left|\sum_{j=n_k+1}^{n_{k+1}-1}\sum_{i=1}^{n_k} H_{n_k}(X_i,X_j)\right| > \tau\sigma n_k 2^{-3j_{n_k}}\sqrt{2\log\log n_k}\right\} \leq \frac{c}{k^2}, \tag{1.4.64}$$

where the right side of the inequality is the general term of a convergent series.

The second term is of the order of

$$n_k^{\delta/2}\sqrt{\log\log n_k}\sqrt{\frac{n_k}{n_{k+1}-n_k}} \asymp n_k^{\delta/2}\sqrt{M_k \log k}. \tag{1.4.65}$$

The third term is of the order of

$$(n_k^{1-\delta}\log\log n_k)^{1/3} \asymp (\log k)^{1/3} n_k^{\frac{1-\delta}{3}}. \tag{1.4.66}$$

The fourth term is of the order of

$$\left(n_k^{1-\delta/2}\sqrt{\log\log n_k}\right)^{1/2} \asymp \left(n_k^{1-\delta/2}\sqrt{\log k}\right)^{1/2} \asymp (\log k)^{1/4} n_k^{1/2-\delta/4}. \tag{1.4.67}$$

So the last three terms are at least as large as positive powers of n_k. If one of these dominates, then we also have a convergent series.

For (1.4.51), we do similar calculations for the four terms in the exponent in (1.2.27). They are:

$$\begin{gathered} \frac{\tau^2\sigma^2 n_k^2 2\log\log n_k}{(n_{k+1}-n_k-1)^2}, \quad \frac{\tau\sigma 2^{j_{n_k}/2} n_k\sqrt{2\log\log n_k}}{n_{k+1}-n_k-1}, \\ \left(\frac{\tau^2\sigma^2 n_k^2 2^{-j_{n_k}} 2\log\log n_k}{n_{k+1}-n_k-1}\right)^{1/3}, \quad \left(\tau\sigma n_k 2^{-j_{n_k}/2}\sqrt{2\log\log n_k}\right)^{1/2}. \end{gathered} \tag{1.4.68}$$

The 1st term is of the order of

$$\left(\frac{n_k}{n_{k+1}-n_k}\right)^2 \log\log n_k \asymp M_k^2 \log k. \tag{1.4.69}$$

The 2nd term is of the order of

$$\frac{n_k}{n_{k+1}-n_k} n_k^{\delta/2}\sqrt{\log k} \asymp M_k n_k^{\delta/2}\sqrt{\log k}. \tag{1.4.70}$$

The 3rd term is of the order of

$$\left(\frac{n_k^{2-\delta}\log k}{n_{k+1}-n_k}\right)^{1/3} \asymp \left(\frac{n_k}{n_{k+1}-n_k} n_k^{1-\delta}\log k\right)^{1/3} \asymp \left(M_k n_k^{1-\delta}\log k\right)^{1/3}. \tag{1.4.71}$$

The 4th term is of the order of

$$(\log k)^{1/4} n_k^{1/2-\delta/4}. \tag{1.4.72}$$

Hence, these are the general terms of convergent series and (1.4.51) also holds. This concludes the proof of Lemma 1.4.3. □

By Lemma 1.4.3, it only remains to prove that (1.4.31) converges for every $\varsigma > 0$. Note that

$$\begin{aligned}
&\sum_{k\geq k_0} \Pr\left\{|W_{n_k}(\mathbb{R})| > (1+\varsigma)\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\} \\
\leq &\sum_{k\geq k_0} \Pr\left\{|W_{n_k}([-M,M])| + \left|W_{n_k}([-M,M]^C)\right| > (1+\varsigma)\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\} \\
\leq &\sum_{k\geq k_0} \Pr\left\{|W_{n_k}([-M,M])| > (1+\frac{\varsigma}{2})\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\} \\
&+\sum_{k\geq k_0} \Pr\left\{\left|W_{n_k}([-M,M]^C)\right| > \frac{\varsigma}{2}\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\}.
\end{aligned} \tag{1.4.73}$$

By condition (f), $\|f\|_2 < \infty$. Hence, given $\varsigma > 0$, there exists $M_1 < \infty$ such that $\int_{[-M_1,M_1]^C} f^2(x)dx < \varsigma^2\sigma^2/(4\kappa_0)$, where κ_0 is the constant in inequality (1.2.30). Taking

$\eta = \sqrt{2}\varsigma\sigma/2$ in (1.2.30), from some k on, we have

$$\begin{aligned}
&\Pr\left\{|W_{n_k}([-M_1, M_1]^C)| > \frac{\varsigma}{2}\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\} \\
&\leq \kappa_0 \exp\left(-\frac{\varsigma^2\sigma^2 \log\log n_k/2}{\kappa_0 \int_{[-M_1,M_1]^C} f^2(x)dx}\right) \\
&\leq \kappa_0 \exp\left(-\frac{\varsigma^2\sigma^2 \log\log n_k/2}{\varsigma^2\sigma^2/4}\right) \\
&= \kappa_0 \exp\left(-2\log\log n_k\right).
\end{aligned} \tag{1.4.74}$$

For all $1/2 < \tau < 1$, there exists $K(\tau)$, such that $\log\log n_k/\log k \geq 1 - \tau$ for all $k \geq K(\tau)$, hence

$$\exp(-2\log\log n_k) \leq \exp(-2(1-\tau)\log k) = k^{-2(1-\tau)}, \tag{1.4.75}$$

which is the general term of a convergent series. Let ϵ be so small that $(1+\varsigma/2)^2(1-\epsilon) > 1$. Using the moderate deviation (1.3.59) on the first summand in (1.4.73), we obtain, for k large enough,

$$\begin{aligned}
&Pr\left\{|W_{n_k}([-M_1, M_1])| > (1+\frac{\varsigma}{2})\sigma n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\} \\
&\leq Pr\left\{|W_{n_k}([-M_1, M_1])| > (1+\frac{\varsigma}{2})\sigma(M_1) n_k 2^{-3j_{n_k}/2}\sqrt{2\log\log n_k}\right\} \\
&\leq \exp\left(-(1+\varsigma/2)^2(1-\epsilon)\log\log n_k\right) + \frac{1}{n_k^2},
\end{aligned} \tag{1.4.76}$$

which is also the general term of a convergent series. Hence the series (1.4.73) converges for every $\varsigma > 0$. □

1.5 Asymptotic stochastic error

In this section, we study the asymptotic stochastic error of the estimator $f_{n,K}$, that is,

$\mathbb{E}\|f_{n,K} - \mathbb{E}f_{n,K}\|_2^2$. We obtain the order of this quantity as $n \to \infty$.

1.5.1 Theorem. Assume that f is bounded and Riemann integrable on $\mathbb{R}$ and there exists $M > 0$ such that f is monotonically increasing on $(-\infty, -M]$ and monotonically decreasing on $[M, \infty)$. Also assume that ϕ satisfies (S1). Then

$$\mathbb{E}\|f_{n,K} - \mathbb{E}f_{n,K}\|_2^2 = \frac{2^{j_n}}{n}\|\phi\|_2^4 + o(2^{j_n}/n). \tag{1.5.1}$$

Proof. By definition of $f_{n,K}$ in (0.1.5),

$$\begin{aligned}
\mathbb{E}\|f_{n,K} - \mathbb{E}f_{n,K}\|_2^2 &= \frac{2^{2j_n}}{n^2}\mathbb{E}\int_{\mathbb{R}}\left[\sum_{i=1}^{n}(K_n(x,X_i) - \mathbb{E}K_n(x,X_i))\right]^2 dx \\
=&\frac{2^{2j_n}}{n^2}\int_{\mathbb{R}}\sum_{i=1}^{n}\mathbb{E}\left[K_n(x,X_i) - \mathbb{E}K_n(x,X_i)\right]^2 dx \\
=&\frac{2^{2j_n}}{n}\int_{\mathbb{R}}\mathbb{E}\left[K_n(x,X) - \mathbb{E}K_n(x,X)\right]^2 dx \\
=&\frac{2^{2j_n}}{n}\int_{\mathbb{R}}\left[\mathbb{E}K^2(2^{j_n}x, 2^{j_n}X) - \left(\mathbb{E}K(2^{j_n}x, 2^{j_n}X)\right)^2\right] dx
\end{aligned} \tag{1.5.2}$$

by independence of $\{X_i\}$, $i = 1, ..., n$. Then we consider $2^{j_n}\int \mathbb{E}K^2(2^{j_n}x, 2^{j_n}X)dx$. By a change of variable $2^{j_n}x = 2^{j_n}y + w$ and periodicity of K, it is equal to

$$\begin{aligned}
2^{j_n}\int\int K^2(2^{j_n}x, 2^{j_n}y)f(y)dydx &= \int\int K^2(2^{j_n}y + w, 2^{j_n}y)f(y)dydw \\
&= \int\sum_{i=-\infty}^{\infty}\int_{2^{-j_n}i}^{2^{-j_n}(i+1)} K^2(2^{j_n}y + w, 2^{j_n}y)f(y)dydw \\
&= \int\sum_{i=-\infty}^{\infty}\int_{0}^{2^{-j_n}} K^2(2^{j_n}y + i + w, 2^{j_n}y + i)f(y + 2^{-j_n}i)dydw \\
&= \int\sum_{i=-\infty}^{\infty}\int_{0}^{2^{-j_n}} K^2(2^{j_n}y + w, 2^{j_n}y)f(y + 2^{-j_n}i)dydw \\
&= 2^{-j_n}\int\sum_{i=-\infty}^{\infty}\int_{0}^{1} K^2(y + w, y)f(2^{-j_n}y + 2^{-j_n}i)dydw.
\end{aligned} \tag{1.5.3}$$

Let M be an integer such that f is monotone on $[M, \infty)$. For fixed $y \in [0,1]$, set

$$\begin{aligned}
&\sum_{i=-\infty}^{\infty} 2^{-j_n} f(2^{-j_n} y + 2^{-j_n} i) \\
=&\left(\sum_{i=0}^{2^{j_n}M} + \sum_{i=2^{j_n}M+1}^{\infty} + \sum_{i=-\infty}^{-1}\right) 2^{-j_n} f(2^{-j_n} y + 2^{-j_n} i) \qquad (1.5.4)\\
=:&I_1(j_n) + I_2(j_n) + I_3(j_n).
\end{aligned}$$

By the hypothesis that f is Riemann integrable, $I_1(j_n) \to \int_0^M f(y)dy$. Using that f is monotone on tails, we have

$$\int_{M+2\cdot 2^{-j_n}}^{\infty} f(y)dy \le I_2(j_n) \le \int_M^{\infty} f(y)dy. \qquad (1.5.5)$$

As $n \to \infty$, $I_2(j_n) \to \int_M^{\infty} f(y)dy$. By analogy, $I_3(j_n) \to \int_{\infty}^{0} f(y)dy$. Thus we have

$$\sum_{i=-\infty}^{\infty} 2^{-j_n} f(2^{-j_n} y + 2^{-j_n} i) \to \int f(y)dy = 1 \qquad (1.5.6)$$

as $n \to \infty$. The convergence is uniform for $y \in [0,1]$ by the definition of Riemann integrability. To continue with (1.5.3), next we consider

$$\begin{aligned}
&\int_{\mathbb{R}} \int_0^1 K^2(y+w, y)dydw \\
=&\int_{\mathbb{R}} \int_0^1 \left[\sum_k \phi(y+w-k)\phi(y-k)\right]^2 dydw \qquad (1.5.7)\\
=&\int_{\mathbb{R}} \int_0^1 \sum_k \phi^2(y+w-k)\phi^2(y-k)dydw \\
&+ \int_{\mathbb{R}} \int_0^1 \sum_{k \ne l} \phi(y+w-k)\phi(y-k)\phi(y+w-l)\phi(y-l)dydw.
\end{aligned}$$

For the first summand,

$$\begin{aligned}&\int_{\mathbb{R}}\int_0^1\sum_k \phi^2(y+w-k)\phi^2(y-k)dydw\\ =&\sum_k\int_0^1\phi^2(y-k)\int_{\mathbb{R}}\phi^2(y+w-k)dwdy=\|\phi\|_2^4.\end{aligned} \tag{1.5.8}$$

Then by orthogonality of $\{\phi(\cdot-k)\}, k\in\mathbb{Z}$ and Fubini's theorem, the second summand

$$\int_{\mathbb{R}}\int_0^1\sum_{k\neq l}\phi(y+w-k)\phi(y-k)\phi(y+w-l)\phi(y-l)dydw=0.$$

Therefore,

$$\int_{\mathbb{R}}\int_0^1 K^2(y+w,y)dydw=\|\phi\|_2^4. \tag{1.5.9}$$

Now applying Fubini's theorem to the last equation in (1.5.3) and using (1.5.6), (1.5.9),

$$\begin{aligned}&2^{j_n}\int\int K^2(2^{j_n}x,2^{j_n}y)f(y)dydx\\ =&\int_{\mathbb{R}}\int_0^1 K^2(y+w,y)\sum_{i=-\infty}^{\infty}2^{-j_n}f(2^{-j_n}y+2^{-j_n}i)dydw\\ \to&\|\phi\|_2^4.\end{aligned} \tag{1.5.10}$$

It follows that

$$\frac{2^{2j_n}}{n}\int\mathbb{E}K^2(2^{j_n}x,2^{j_n}X)dx=\frac{2^{j_n}}{n}\|\phi\|_2^4+o(2^{j_n}/n). \tag{1.5.11}$$

Now we consider $2^{2j_n}\int(\mathbb{E}K(2^{j_n}x,2^{j_n}X))^2dx$.

$$\begin{aligned}&2^{2j_n}\int\left(\int K(2^{j_n}x,2^{j_n}y)f(y)dy\right)^2dx\\ =&2^{2j_n}\int\int\int K(2^{j_n}x,2^{j_n}y)K(2^{j_n}x,2^{j_n}z)f(y)f(z)dydzdx\end{aligned}$$

$$
\begin{aligned}
&= \int\int\int K(2^{j_n}x, 2^{j_n}x - s)K(2^{j_n}x, 2^{j_n}x - t)f(x - 2^{-j_n}s)f(x - 2^{-j_n}t)dsdtdx \qquad (1.5.12)\\
&\leq \int\int\int \Phi(s)\Phi(t)f(x - 2^{-j_n}s)f(x - 2^{-j_n}t)dsdtdx\\
&\leq C\int\int \Phi(s)\Phi(t)\int f(x - 2^{-j_n}s)dxdsdt\\
&\leq C\|\Phi\|_1^2.
\end{aligned}
$$

Combining these with (1.5.2), we get (1.5.1). □

Chapter 2

Rate of Convergence in the Central Limit Theorem

In this chapter, we show that the integrated squared error of the linear wavelet density estimator converges to $\mathcal{N}(0, \sigma^2)$ at polynomial rate. No claim of optimality of the rate obtained is made, but it is reassuring that the rate is not logarithmic. C is a universal constant which may vary from line to line. Without loss of generality, we will assume that, for all n, there exist constants C_1 and C_2, such that $C_1 n^\delta \le 2^{j_n} \le C_2 n^\delta$.

2.0.1 Theorem. Assume the hypotheses (f), (S1), (B1) and that there exists $L \ge 0$ such that f is Hölder continuous with exponent $0 < \alpha \le 1$ 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, ϕ and $\{j_n\}$), such that, for all n,

$$\sup_{t\in\mathbb{R}} |\Pr\{n2^{-j_n/2} J_n \le t\} - \Pr\{\sigma Z \le t\}| \le C(n^{-3\delta/16} \vee n^{-\alpha\delta}\sqrt{\log n}), \tag{2.0.1}$$

where $\sigma^2 = 2\int_{\mathbb{R}} f^2(x)dx$. δ is the same as in condition (B1).

For example, if $2^{-j_n} \asymp n^{-1/5}$, $\sup_t |\Pr\{n2^{-j_n/2} J_n \le t\} - \Pr\{\sigma Z \le t\}| \le C(n^{-3/80} \vee n^{-\alpha/5}\sqrt{\log n})$.

Proof. By the proof of Theorem 1.4.1, it is equivalent to showing that

$$\sup_{t} |\Pr\{n2^{-j_n/2} \bar{J}_n \le t\} - \Pr\{\sigma Z \le t\}| \le C(n^{-3\delta/16} \vee n^{-\alpha\delta}\sqrt{\log n}). \tag{2.0.2}$$

By (1.1.5) and (1.1.6), we have that

$$n2^{-j_n/2}\bar{J}_n/\sigma = \frac{2^{3j_n/2}}{n\sigma}W_n(\mathbb{R}) = \frac{2^{3j_n/2}}{n\sigma}U_n(\mathbb{R}) + \frac{2^{3j_n/2}}{n\sigma}L_n(\mathbb{R}). \tag{2.0.3}$$

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

$$\begin{aligned}
&\Pr\{n2^{-j_n/2}\bar{J}_n/\sigma \le t\} - \Pr\{Z \le t\} \\
&\le \Pr\left\{\frac{2^{3j_n/2}}{n\sigma}U_n(\mathbb{R}) \le t + \epsilon_{1,n}\right\} + \Pr\left\{\frac{2^{3j_n/2}}{n\sigma}L_n(\mathbb{R}) < -\epsilon_{1,n}\right\} - \Pr\{Z \le t\} \\
&= \Pr\left\{\frac{2^{3j_n/2}}{n\sigma}U_n(\mathbb{R}) \le t + \epsilon_{1,n}\right\} - \Pr\{Z \le t + \epsilon_{1,n}\} \\
&\quad + \Pr\left\{\frac{2^{3j_n/2}}{n\sigma}L_n(\mathbb{R}) < -\epsilon_{1,n}\right\} + \Pr\{t < Z \le t + \epsilon_{1,n}\},
\end{aligned} \tag{2.0.4}$$

and

$$\begin{aligned}
&\Pr\{n2^{-j_n/2}\bar{J}_n/\sigma \le t\} - \Pr\{Z \le t\} \\
&\ge \Pr\left\{\frac{2^{3j_n/2}}{n\sigma}U_n(\mathbb{R}) \le t - \epsilon_{1,n}\right\} - \Pr\left\{-\frac{2^{3j_n/2}}{n\sigma}L_n(\mathbb{R}) < -\epsilon_{1,n}\right\} - \Pr\{Z \le t\} \\
&= \Pr\left\{\frac{2^{3j_n/2}}{n\sigma}U_n(\mathbb{R}) \le t - \epsilon_{1,n}\right\} - \Pr\{Z \le t - \epsilon_{1,n}\} \\
&\quad - \left(\Pr\left\{-\frac{2^{3j_n/2}}{n\sigma}L_n(\mathbb{R}) < -\epsilon_{1,n}\right\} + \Pr\{t - \epsilon_{1,n} < Z \le t\}\right).
\end{aligned} \tag{2.0.5}$$

So

$$\begin{aligned}
&\sup_{t\in\mathbb{R}}\left|\Pr\{n2^{-j_n/2}\bar{J}_n/\sigma \le t\} - \Pr\{Z \le t\}\right| \\
&\le \sup_{t\in\mathbb{R}}\left|\Pr\left\{\frac{2^{3j_n/2}}{n\sigma}U_n(\mathbb{R}) \le t\right\} - \Pr\{Z \le t\}\right| \\
&\quad + \Pr\left\{\frac{2^{3j_n/2}}{n\sigma}\left|L_n(\mathbb{R})\right| > \epsilon_{1,n}\right\} + \sup_{t\in\mathbb{R}}\Pr\left\{t - \epsilon_{1,n} < Z \le t + \epsilon_{1,n}\right\}.
\end{aligned} \tag{2.0.6}$$

It's easy to bound the last term as follows:

$$\begin{aligned}
&\sup_{t\in\mathbb{R}} \Pr\left\{t-\epsilon_{1,n} < Z \le t+\epsilon_{1,n}\right\} \\
&= \sup_t \frac{1}{\sqrt{2\pi}} \int_{t-\epsilon_{1,n}}^{t+\epsilon_{1,n}} \exp(-x^2/2)dx \\
&\le \sup_t \frac{1}{\sqrt{2\pi}} \int_{t-\epsilon_{1,n}}^{t+\epsilon_{1,n}} 1dx \\
&< \epsilon_{1,n}.
\end{aligned} \tag{2.0.7}$$

By (1.2.32), for $0 < \epsilon_{1,n} \le 1$ so that $\epsilon_{1,n}^2 \le \epsilon_{1,n}$,

$$\begin{aligned}
\Pr\left\{|L_n(\mathbb{R}))| > \sigma\epsilon_{1,n} n 2^{-3j_n/2}\right\} &\le C\exp\left(-\frac{1}{C}\min\left(\sigma^2\epsilon_{1,n}^2 n2^{-j_n}, \sigma\epsilon_{1,n} n 2^{-j_n/2}\right)\right) \\
&\le C\exp\left(-\frac{1}{C}\epsilon_{1,n}^2 n 2^{-j_n}\right) \\
&\le C\exp\left(-\frac{1}{C}\epsilon_{1,n}^2 n^{1-\delta}\right),
\end{aligned} \tag{2.0.8}$$

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

$$\Pr\left\{|L_n(\mathbb{R})| > \sigma\epsilon_{1,n} n 2^{-3j_n/2}\right\} \le C\exp(-\log n) = Cn^{-1} \tag{2.0.9}$$

when n is large enough. Using (2.0.7), we get

$$\sup_t \Pr\left\{t-\epsilon_{1,n} < Z \le t+\epsilon_{1,n}\right\} \le n^{-1/3}. \tag{2.0.10}$$

To control the first term in (2.0.6), we will approximate $2^{3j_n/2}U_n(\mathbb{R})/(n\sigma)$ by S_{nn}, which is defined below. We set

$$U_{nn} := \sum_{i=2}^{n}\sum_{j=1}^{i-1} H_n(X_i, X_j),\ s_n^2 := \mathbb{E}(U_{nn}^2), \tag{2.0.11}$$

and

$$X_{ni} := \sum_{j=1}^{i-1} \frac{H_n(X_i, X_j)}{s_n}, \; S_{nk} := \sum_{i=2}^{k} X_{ni}. \tag{2.0.12}$$

Then according to these notations,

$$S_{nn} = \sum_{i=2}^{n} \sum_{j=1}^{i-1} \frac{H_n(X_i, X_j)}{s_n}. \tag{2.0.13}$$

Analogous to (2.0.6), for any positive sequence $\epsilon_{2,n}$,

$$\begin{aligned}
&\sup_t \left| \Pr\left\{ \frac{2^{3j_n/2}}{n\sigma} U_n(\mathbb{R}) \le t \right\} - \Pr\{Z \le t\} \right| \\
&= \sup_t \left| \Pr\left\{ \frac{2^{3j_n/2}}{n\sigma} U_n(\mathbb{R}) - S_{nn} + S_{nn} \le t \right\} - \Pr\{Z \le t\} \right| \\
&\le \sup_t \left| \Pr\left\{ S_{nn} \le t \right\} - \Pr\{Z \le t\} \right| \\
&\quad + \Pr\left\{ \left| \frac{2^{3j_n/2}}{n\sigma} U_n(\mathbb{R}) - S_{nn} \right| > \epsilon_{2,n} \right\} + \sup_t \Pr\left\{ t - \epsilon_{2,n} < Z \le t + \epsilon_{2,n} \right\}.
\end{aligned} \tag{2.0.14}$$

By the definition of $U_n(\mathbb{R})$ in (1.2.4),

$$\begin{aligned}
&\Pr\left\{ \left| \frac{2^{3j_n/2}}{n\sigma} U_n(\mathbb{R}) - S_{nn} \right| > \epsilon_{2,n} \right\} \\
&= \Pr\left\{ \left| \frac{2^{3j_n/2}}{n\sigma} \sum_{1 \le i \ne j \le n} H_n(X_i, X_j) - \frac{1}{2s_n} \sum_{1 \le i \ne j \le n} H_n(X_i, X_j) \right| > \epsilon_{2,n} \right\} \\
&= \Pr\left\{ \left| \sum_{1 \le i \ne j \le n} H_n(X_i, X_j) \right| > \frac{\epsilon_{2,n}}{d_n} \right\},
\end{aligned} \tag{2.0.15}$$

where $d_n = \left| \frac{2^{3j_n/2}}{n\sigma} - \frac{1}{2s_n} \right|$. We then estimate the order of d_n. Let $Y_{ni} := \sum_{j=1}^{i-1} H_n(X_i, X_j)$. Since

$$\mathbb{E} H_n(X_1, X_2) H_n(X_3, X_4) = 0 \tag{2.0.16}$$

and

$$\mathbb{E}H_n(X_1, X_2)H_n(X_1, X_3) = \mathbb{E}\left[H_n(X_1, X_2)\mathbb{E}\left(H_n(X_1, X_3)|X_1, X_2\right)\right] = 0, \tag{2.0.17}$$

we have $\mathbb{E}(Y_{ni}Y_{nj}) = 0$ for $i \neq j$. Thus, by definition of s_n^2 and (1.2.17),

$$\begin{aligned} s_n^2 &= \mathbb{E}\left(\sum_{i=2}^{n} Y_{ni}\right)^2 = \mathbb{E}\sum_{i=2}^{n} Y_{ni}^2 \\ &= \sum_{i=2}^{n}\sum_{j=1}^{i-1} \mathbb{E}H_n^2(X_i, X_j) \\ &= \frac{n(n-1)}{2}\mathbb{E}H_n^2(X_1, X_2) \\ &= \frac{n(n-1)}{2}2^{-3j_n}2^{j_n}\int_{\mathbb{R}^2} R_n^2(s,t)dsdt. \end{aligned} \tag{2.0.18}$$

Set $e_n := \left(2^{j_n}\int_{\mathbb{R}^2} R_n^2(s,t)dsdt\right)^{1/2}$. Then

$$\begin{aligned} d_n &= \left|\frac{2^{3j_n/2}}{n\sqrt{2\int f^2(x)dx}} - \frac{2^{3j_n/2}}{\sqrt{2n(n-1)}e_n}\right| \\ &\leq C2^{3j_n/2}\left|\frac{1}{n\sqrt{\int f^2(x)dx}} - \frac{1}{\sqrt{n(n-1)\int f^2(x)dx}}\right| \\ &\quad + C2^{3j_n/2}\left|\frac{1}{\sqrt{n(n-1)\int f^2(x)dx}} - \frac{1}{\sqrt{n(n-1)}e_n}\right|. \end{aligned} \tag{2.0.19}$$

By condition (B1), $2^{j_n} \leq Cn^{\delta}$ for some $\delta \in (0, 1/3)$. Since $1/\sqrt{n(n-1)} - 1/n \leq n^{-2}$, the first term is bounded by $Cn^{3\delta/2-2}$ for $n \geq 2$. The constant C here depends on f and $\{j_n\}$. Corollary $A.0.8$ gives that $\left|e_n^2 - \int_{\mathbb{R}} f^2(x)dx\right| \leq C(n^{-\delta/2} + n^{-\delta\alpha})$. From this, we get a bound on the second term. When n is sufficiently large depending on f,

$$\begin{aligned}
&\frac{C2^{3j_n/2}}{\sqrt{n(n-1)}}\left|\frac{1}{e_n}-\frac{1}{\sqrt{\int f^2(x)dx}}\right| \\
\leq& Cn^{3\delta/2-1}\frac{\left|\sqrt{\int f^2(x)dx}-e_n\right|}{e_n\sqrt{\int f^2(x)dx}} \\
\leq& Cn^{3\delta/2-1}\frac{\left|e_n^2-\int_{\mathbb{R}} f^2(x)dx\right|}{\left(e_n+\sqrt{\int f^2(x)dx}\right)e_n\sqrt{\int f^2(x)dx}} \\
\leq& Cn^{3\delta/2-1}\left(n^{-\delta/2}+n^{-\delta\alpha}\right),
\end{aligned} \tag{2.0.20}$$

where C depends on $f, \{j_n\}$ and ϕ. The last inequality follows since $e_n \to \sqrt{\int f^2(x)dx}$, so e_n is bounded away from zero if n is large. Combining the two terms, $d_n \leq Cn^{3\delta/2-1}\left(n^{-\delta/2}+n^{-\delta\alpha}\right)$.

By Proposition 1.2.4 with $\tau = 2^{3j_n/2}\epsilon_{2,n}/(nd_n)$, there exist constants κ_0 (depending on ϕ and f) and n_0 such that for all $n \geq n_0$ (depending on f, ϕ and $\{j_n\}$),

$$\begin{aligned}
&\Pr\left\{\left|\sum_{1\leq i\neq j\leq n} H_n(X_i,X_j)\right| > \frac{\epsilon_{2,n}}{d_n}\right\} \\
\leq& \kappa_0\exp\left(-\frac{1}{\kappa_0}\min\left[\frac{2^{3j_n}\epsilon_{2,n}^2}{n^2d_n^2\int f^2(x)dx}, \frac{2^{2j_n}\epsilon_{2,n}}{nd_n}, \frac{\epsilon_{2,n}^{2/3}2^{2j_n/3}}{d_n^{2/3}n^{1/3}}, \frac{\epsilon_{2,n}^{1/2}2^{j_n/2}}{d_n^{1/2}}\right]\right).
\end{aligned} \tag{2.0.21}$$

The four terms in the exponent are respectively bounded below by:

$Cn^{\delta\min(1,2\alpha)}\epsilon_{2,n}^2, Cn^{\delta/2}n^{\delta\min(1/2,\alpha)}\epsilon_{2,n}, Cn^{1/3-\delta/3}n^{\delta\min(1,2\alpha)/3}\epsilon_{2,n}^{2/3}, n^{1/2-\delta/4}n^{\delta\min(1/2,\alpha)/2}\epsilon_{2,n}^{1/2}$.

If $0 \leq \epsilon_{2,n} \leq 1$, then all of them are bounded below by $C\epsilon_{2,n}^2 n^{\min(\delta,2\delta\alpha)}$. Taking $\epsilon_{2,n} = \kappa_0 n^{-\delta(\frac{1}{2}\wedge\alpha)}\sqrt{\log n}/C$ and using (2.0.15), we obtain

$$\Pr\left\{\left|\frac{2^{3j_n/2}}{n\sigma}U_n(\mathbb{R})-S_{nn}\right| > \epsilon_{2,n}\right\} \leq \kappa_0\exp\left(-\log n\right) = \kappa_0 n^{-1} \tag{2.0.22}$$

when n is large enough. Consequently,

$$\sup_t \Pr\{t - \epsilon_{2,n} < Z \le t + \epsilon_{2,n}\} \le n^{-\delta(\frac{1}{2}\wedge\alpha)}\sqrt{\log n}. \tag{2.0.23}$$

Let $\mathcal{F}_i$ be the σ-field generated by $\{X_1, X_2, ..., X_i\}$ for i=1,2,.... To deal with $\sup_t |\Pr\{S_{nn} \le t\} - \Pr\{Z \le t\}|$, we first observe that, by the definitions in (2.0.11)-(2.0.13),

$$\mu_{ni} := \mathbb{E}(X_{ni}|\mathcal{F}_{i-1}) = 0. \tag{2.0.24}$$

Thus S_{nk} is a martingale with respect to $\mathcal{F}_k$. Central limit theorems for martingales have been studied by several authors (e.g., Hall and Heyde, 1980) and rates of convergence to normality are also available. We will use the result of Erickson, Quine and Weber (1979) to derive a bound for $\sup_t |\Pr\{S_{nn} \le t\} - \Pr\{Z \le t\}|$. For $i \ge 2$, let $X'_{ni} := X_{ni} - \mu_{ni}$, $\sigma^2_{ni} := \mathbb{E}\left({X'_{ni}}^2|\mathcal{F}_{i-1}\right)$ and $\sigma^2_n := \sum_{i=2}^n \sigma^2_{ni}$.

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

$$\sup_t |\Pr\{S_{nn} \le t\} - \Pr\{Z \le t\}| \le C\left\{\sum_{i=2}^n \mathbb{E}|X_{ni}|^{2+\eta} + \mathbb{E}|1 - \sigma^2_n|^{1+\eta/2}\right\}^{1/(3+\eta)}. \tag{2.0.25}$$

By definition of σ^2_n and Y_{ni},

$$\begin{aligned}
&\mathbb{E}\left|1 - \sigma^2_n\right|^2 \\
=&\mathbb{E}\left|1 - \frac{1}{s^2_n}\sum_{i=2}^n \mathbb{E}\left(Y^2_{ni}|\mathcal{F}_{i-1}\right)\right|^2 \\
=&s_n^{-4}\mathbb{E}\left|s^2_n - \sum_{i=2}^n \mathbb{E}\left(Y^2_{ni}|\mathcal{F}_{i-1}\right)\right|^2 \\
=&s_n^{-4}\mathbb{E}\left|s^2_n - V^2_n\right|^2,
\end{aligned} \tag{2.0.26}$$

where $V_n^2 := \sum_{i=2}^n \mathbb{E}\left(Y_{ni}^2|\mathcal{F}_{i-1}\right)$. By (2.0.18), $\mathbb{E}V_n^2 = \sum_{i=2}^n \mathbb{E}Y_{ni}^2 = s_n^2$, $\mathbb{E}\left|s_n^2 - V_n^2\right|^2 \leq \mathbb{E}(V_n^4)$. Set $G_n(x,y) := \mathbb{E}\left(H_n(X_1,x)H_n(X_1,y)\right)$, then by the proof of Theorem 1, Hall (1984),

$$\mathbb{E}(V_n^4) \leq C\left(n^4\mathbb{E}G_n^2(X_1,X_2) + n^3\mathbb{E}G_n^2(X_1,X_1)\right) \leq C\left(n^4\mathbb{E}G_n^2(X_1,X_2) + n^3\mathbb{E}H_n^4(X_1,X_2)\right). \tag{2.0.27}$$

By (2.0.18) and Corollary $A.0.8$, $s_n^4 \asymp n^{4-6\delta}(2^{j_n}\int_{\mathbb{R}^2} R_n^2(s,t)dsdt)^2 \asymp n^{4-6\delta}$. The calculations in Theorem 1, Zhang and Zheng (1999) can be applied directly here. $H_n(x,y)$ defined in (1.2.2) is off by a scaling constant $2^{-2j_n}n^2$ from their definition.

$$\mathbb{E}H_n^4(X_1,X_2) = \left(2^{-2j_n}n^2\right)^4 O(2^{3j_n}/n^8) = O(2^{-5j_n}) = O(n^{-5\delta}), \tag{2.0.28}$$

and

$$\mathbb{E}G_n^2(X_1,X_2) = \left(2^{-2j_n}n^2\right)^4 O(2^{j_n}/n^8) = O(2^{-7j_n}) = O(n^{-7\delta}). \tag{2.0.29}$$

Combining these estimates and using Hölder inequality, we see

$$\mathbb{E}|1-\sigma_n^2|^{1+\eta/2} \leq (\mathbb{E}|1-\sigma_n^2|^2)^{(2+\eta)/4} \leq C\left(\frac{n^{4-7\delta}+n^{3-5\delta}}{n^{4-6\delta}}\right)^{(2+\eta)/4} \leq Cn^{-\delta(2+\eta)/4}. \tag{2.0.30}$$

For the first term in (2.0.25), we observe that

$$\sum_{i=2}^n \mathbb{E}|X_{ni}|^{2+\eta} = \sum_{i=2}^n \frac{1}{s_n^{2+\eta}}\mathbb{E}\left|\sum_{j=1}^{i-1} H_n(X_i,X_j)\right|^{2+\eta} \leq \sum_{i=2}^n \frac{1}{s_n^{2+\eta}}\left(\mathbb{E}\left|\sum_{j=1}^{i-1} H_n(X_i,X_j)\right|^3\right)^{(2+\eta)/3}. \tag{2.0.31}$$

Let $\mathbb{E}_i$ denote the expectation with respect to X_i and $\mathbb{E}_{i'}$ denote the expectation with respect to $X_1, ..., X_{i-1}$. Then by independence of $\{X_i\}$,

$$\mathbb{E}\left|\sum_{j=1}^{i-1} H_n(X_i, X_j)\right|^3 = \mathbb{E}_i\mathbb{E}_{i'}\left|\sum_{j=1}^{i-1} H_n(X_i, X_j)\right|^3. \tag{2.0.32}$$

We can apply a Hoffmann-Jorgensen type inequality with respect to $\mathbb{E}_{i'}$ (Theorem 1.5.13, de la Peña and Giné, 1999). Since for fixed i, $H_n(X_i, X_j)$, $j = 1, ..., i-1$ are centered independent real random variables with respect to the inner expectation,

$$\mathbb{E}_i\mathbb{E}_{i'}\left|\sum_{j=1}^{i-1} H_n(X_i, X_j)\right|^3 \leq C\mathbb{E}_i\left\{\mathbb{E}_{i'}\max_{1\leq j\leq i-1}\left|H_n(X_i, X_j)\right|^3 + \left(\mathbb{E}_{i'}\left(\sum_{j=1}^{i-1} H_n(X_i, X_j)\right)^2\right)^{3/2}\right\}. \tag{2.0.33}$$

By (A.0.6), $\left|H_n(X_i, X_j)\right| \leq Cn^{-\delta}$ for all i and j. So

$$\mathbb{E}_i\mathbb{E}_{i'}\max_{1\leq j\leq i-1}\left|H_n(X_i, X_j)\right|^3 \leq Cn^{-3\delta}. \tag{2.0.34}$$

Using Jensen's inequality, Hölder inequality and (2.0.28), we get

$$\begin{aligned}
&\mathbb{E}_i\left(\mathbb{E}_{i'}\left(\sum_{j=1}^{i-1} H_n(X_i, X_j)\right)^2\right)^{3/2} \\
=&\mathbb{E}_i\left(\sum_{j=1}^{i-1}\mathbb{E}_{i'}H_n^2(X_i, X_j) + \sum_{1\leq j\neq l\leq i-1}\mathbb{E}_{i'}H_n(X_i, X_j)\mathbb{E}_{i'}H_n(X_i, X_l)\right)^{3/2} \\
=&\mathbb{E}_i\left((i-1)\mathbb{E}_1 H_n^2(X_1, X_2)\right)^{3/2} \\
\leq&(i-1)^{3/2}\mathbb{E}\left|H_n(X_1, X_2)\right|^3 \\
\leq&(i-1)^{3/2}\left(\mathbb{E}\left|H_n(X_1, X_2)\right|^4\right)^{3/4} \\
\leq&C(i-1)^{3/2}n^{-15\delta/4}.
\end{aligned} \tag{2.0.35}$$

Since $s_n^2 \asymp n^{2-3\delta}$ and $\sum_{i=2}^n i^{(2+\eta)/2} \le \int_2^{n+1} x^{(2+\eta)/2} dx \le Cn^{2+\eta/2}$,

$$\begin{aligned}
\sum_{i=2}^{n} \mathbb{E}|X_{ni}|^{2+\eta} \le & \frac{1}{s_n^{2+\eta}} C \sum_{i=2}^{n} \left(n^{-3\delta} + (i-1)^{3/2} n^{-15\delta/4}\right)^{(2+\eta)/3} \\
\le & Cn^{(3\delta/2-1)(2+\eta)} n^{-\delta(2+\eta)} \sum_{i=2}^{n} \max\left(1, i^{3/2} n^{-3\delta/4}\right)^{(2+\eta)/3} \\
\le & Cn^{(\delta/2-1)(2+\eta)} \max\left(n, \sum_{i=2}^{n} i^{(2+\eta)/2} n^{-\delta(2+\eta)/4}\right) \\
\le & Cn^{(\delta/2-1)(2+\eta)} \max\left(n, n^{(1/2-\delta/4)(2+\eta)+1}\right) \\
\le & Cn^{(\delta/2-1)(2+\eta)} n^{(1/2-\delta/4)(2+\eta)+1} \\
\le & Cn^{\delta/2+\eta\delta/4-\eta/2}.
\end{aligned} \tag{2.0.36}$$

Gathering (2.0.25), (2.0.30) and (2.0.36), we obtain that, for $\eta \in (0, 1]$,

$$\sup_t |\Pr\{S_{nn} \le t\} - \Pr\{Z \le t\}| \le C \max\left(n^{\frac{\delta/2+\eta\delta/4-\eta/2}{3+\eta}}, n^{-\frac{\delta(2+\eta)}{4(3+\eta)}}\right). \tag{2.0.37}$$

The quantity on the right hand side is minimized when $\eta = 1$. So

$$\sup_t |\Pr\{S_{nn} \le t\} - \Pr\{Z \le t\}| \le C \max\left(n^{3\delta/16-1/8}, n^{-3\delta/16}\right) \le Cn^{-3\delta/16}. \tag{2.0.38}$$

Putting together the last inequality with (2.0.6), (2.0.9), (2.0.10), (2.0.14), (2.0.22) and (2.0.23), we conclude that there exists a constant C (depending on f, ϕ and $\{j_n\}$),

$$\begin{aligned}
\sup_t |\Pr\{n2^{-j_n/2}\bar{J}_n/\sigma \le t\} - \Pr\{Z \le t\}| \le & C\left(n^{-\delta(\frac{1}{2}\wedge\alpha)} \sqrt{\log n} + n^{-3\delta/16}\right) \\
\le & C(n^{-3\delta/16} \vee n^{-\alpha\delta} \sqrt{\log n}).
\end{aligned} \tag{2.0.39}$$

Taking C sufficiently large so it is true for all n. (2.0.1) then follows. □

Chapter 3

Bernstein Density Estimation

3.1 Preliminaries: An exponential inequality for VC classes of functions

Let $\mathcal{G}$ be a class of real measurable functions on a measurable space $(S, \mathcal{S})$, uniformly bounded by a constant. $\{X_i\}$, $i \in \mathbb{N}$ are the coordinate functions $S^{\mathbb{N}} \to S$. X_i's have a common probability law P. Set $\|\Phi\|_{\mathcal{G}} = \sup_{g \in \mathcal{G}} |\Phi(g)|$. We have the maximal version of Talagrand's inequality (Talagrand, 1996; in this form, see Giné and Nickl, 2009): For any measurable, uniformly bounded classes of functions $\mathcal{G}$, there exists a universal constant L, such that for all $t > 0$,

$$\Pr\left\{\max_{k \le n} \left\| \sum_{i=1}^{k} g(X_i) \right\|_{\mathcal{G}} \ge E + t \right\} \le L \exp\left\{ -\frac{t}{LU} \log\left(1 + \frac{tU}{V}\right) \right\}, \tag{3.1.1}$$

where $E := \mathbb{E}\| \sum_{i=1}^{n} g(X_i)\|_{\mathcal{G}}$, $V := \mathbb{E}\| \sum_{i=1}^{n} g^2(X_i)\|_{\mathcal{G}}$, $U \ge \sup_{g \in \mathcal{G}} \|g\|_{\infty}$.

We say that a class $\mathcal{G}$ of measurable functions is VC (Vapnik-Červonenkis) with respect to the constant envelope U if there exist $A, v < \infty$, such that, for all probability measures Q,

$$N(\mathcal{G}, L_2(Q), \epsilon) \le \left(\frac{AU}{\epsilon}\right)^{v}. \tag{3.1.2}$$

Here $N(\mathcal{G}, L_2(Q), \epsilon)$ is the covering number: It is the minimum number of balls of radius ϵ needed to cover $\mathcal{G}$ with $L_2(Q)$ metric. Let $\mathcal{G}$ be a P-centered ($\int g dP = 0 \ \ \forall g \in \mathcal{G}$)

VC-class of functions with $A \geq 3\sqrt{e}$, $v \geq 1$. Set $\sigma^2 := \sup_{g \in \mathcal{G}} \mathbb{E}_P g^2$. Then by Proposition 2.1, Giné and Guillou (2001), there exists a universal constant C, such that

$$E \leq C\left(vU \log \frac{AU}{\sigma} + \sigma\sqrt{nv}\sqrt{\log \frac{AU}{\sigma}}\right). \tag{3.1.3}$$

From inequality (2.6) in Giné and Guillou (2001), one then has

$$V \leq \left(\sqrt{n}\sigma + C\sqrt{v}U\sqrt{\log \frac{AU}{\sigma}}\right)^2. \tag{3.1.4}$$

The last inequality combined with (3.1.1) gives that: there exists a constant K depending on v, such that for all $t > 0$,

$$\begin{aligned} &\Pr\left\{\max_{k \leq n}\left\|\sum_{i=1}^{k} g(X_i)\right\|_{\mathcal{G}} \geq E + t\right\} \\ \leq &K \exp\left\{-\frac{t}{KU}\log\left(1 + \frac{tU}{K\left(\sqrt{n}\sigma + U\sqrt{\log(AU/\sigma)}\right)^2}\right)\right\}. \end{aligned} \tag{3.1.5}$$

Setting $t' := E + t$ and using the bound for E in (3.1.3), for all $t' > C(vU \log \frac{AU}{\sigma} + \sigma\sqrt{nv}\sqrt{\log \frac{AU}{\sigma}})$, we have

$$\begin{aligned} &\Pr\left\{\max_{k \leq n}\left\|\sum_{i=1}^{k} g(X_i)\right\|_{\mathcal{G}} \geq t'\right\} \\ \leq &K \exp\left\{-\frac{t' - E}{KU}\log\left(1 + \frac{(t' - E)U}{K\left(\sqrt{n}\sigma + U\sqrt{\log(AU/\sigma)}\right)^2}\right)\right\}. \end{aligned} \tag{3.1.6}$$

If for some $0 < d < 1$, we take $t' > \frac{C}{1-d}(vU \log \frac{AU}{\sigma} + \sigma\sqrt{nv}\sqrt{\log \frac{AU}{\sigma}})$, then $t' - E > t'd$.

And we have

$$\Pr\left\{\max_{k\le n}\left\|\sum_{i=1}^{k} g(X_i)\right\|_{\mathcal{G}} \ge t'\right\} \le K\exp\left\{-\frac{t'd}{KU}\log\left(1+\frac{(t'd)U}{K\left(\sqrt{n}\sigma + U\sqrt{\log(AU/\sigma)}\right)^2}\right)\right\}. \tag{3.1.7}$$

We arrive at the following version of Proposition 2.2, Giné and Guillou (2001).

3.1.1 Proposition. Let $\mathcal{G}$ be a measurable uniformly bounded VC-class of functions with $A \ge 3\sqrt{e}$, $v \ge 1$. $\mathcal{G}$ is P-centered. There exists K and C depending only on the VC characteristic v of the class $\mathcal{G}$, such that

$$\Pr\left\{\max_{k\le n}\left\|\sum_{i=1}^{k} g(X_i)\right\|_{\mathcal{G}} \ge t\right\} \le K\exp\left\{-\frac{t}{KU}\log\left(1+\frac{tU}{K\left(\sqrt{n}\sigma + U\sqrt{\log(AU/\sigma)}\right)^2}\right)\right\}. \tag{3.1.8}$$

valid for all $t \ge C(vU\log(AU/\sigma) + \sigma\sqrt{nv}\sqrt{\log(AU/\sigma)})$.

We are mainly interested in the range of 't' when the tail probability is of Gaussian type. This is given below, and it is a version of Corollary 2.2, Giné and Guillou (2002).

3.1.2 Corollary. Under the assumptions of Proposition 2.1, and further assuming that for some $\lambda > 0$,

$$\sqrt{n}\sigma \ge \lambda U\sqrt{\log\frac{AU}{\sigma}}, \tag{3.1.9}$$

there exist positive constants K and C_1 depending only on v such that for all $C_2 \ge C_1\lambda^{-1}$ and for all t satisfying

$$C_1\sqrt{n}\sigma\sqrt{\log\frac{AU}{\sigma}} \le t \le C_2\frac{n\sigma^2}{U}, \tag{3.1.10}$$

we have

$$\Pr\left\{\max_{k\le n}\left\|\sum_{i=1}^{k} g(X_i)\right\|_{\mathcal{G}} \ge t\right\} \le K\exp\left\{-C_3\frac{t^2}{n\sigma^2}\right\}, \tag{3.1.11}$$

where $C_3 = \log(1 + \frac{C_2}{K(1+\lambda^{-1})^2})/(KC_2)$.

Proof. Using assumption (3.1.9) and Proposition 3.1.1,

$$\begin{aligned}\Pr\left\{\max_{k\le n}\left\|\sum_{i=1}^{k} g(X_i)\right\|_{\mathcal{G}} \ge t\right\} \le& K\exp\left\{-\frac{t}{KU}\log\left(1+\frac{tU}{K(\sqrt{n}\sigma+\lambda^{-1}\sqrt{n}\sigma)^2}\right)\right\}\\ =& K\exp\left\{-\frac{t}{KU}\log\left(1+\frac{tU}{K(1+\lambda^{-1})^2n\sigma^2}\right)\right\}\end{aligned} \tag{3.1.12}$$

For t satisfying (3.1.10),

$$\frac{tU}{K(1+\lambda^{-1})^2n\sigma^2} \le \frac{C_2n\sigma^2}{K(1+\lambda^{-1})^2n\sigma^2} = \frac{C_2}{K(1+\lambda^{-1})^2}. \tag{3.1.13}$$

Since $\log(1+x)/x$ is decreasing for $x > 0$,

$$\begin{aligned}\log\left(1+\frac{tU}{K(1+\lambda^{-1})^2n\sigma^2}\right) \ge& \frac{\log\left(1+\frac{C_2}{K(1+\lambda^{-1})^2}\right)}{\frac{C_2}{K(1+\lambda^{-1})^2}}\frac{tU}{K(1+\lambda^{-1})^2n\sigma^2}\\ =& \frac{\log\left(1+\frac{C_2}{K(1+\lambda^{-1})^2}\right)tU}{C_2n\sigma^2}.\end{aligned} \tag{3.1.14}$$

Combining it with (3.1.12) yields (3.1.11). □

3.2 Upper Bound

The main idea to prove the upper bound is to apply Corollary 3.1.2 to certain classes related to $\hat{f}_{m,n}(x)$. We will first show that these are of VC type by calculating their covering numbers. From (0.2.3), we may write

$$\hat{f}_{m,n}(x) = \frac{m}{n}\sum_{k=0}^{m-1}\sum_{i=1}^{n}\left[I\left(\frac{k}{m} < X_i \le \frac{k+1}{m}\right)\right] b_{k,m-1}(x). \tag{3.2.1}$$

For fixed $m \in \mathbb{Z}^+$, define

$$\mathcal{F}_m := \left\{\sum_{k=0}^{m-1} I_{(k/m,(k+1)/m]}(\cdot) b_{k,m-1}(x) : x \in (0,1)\right\}. \tag{3.2.2}$$

$$\mathcal{F}_m^* := \left\{\sum_{k=0}^{m-1}\left[I_{(k/m,(k+1)/m]}(\cdot) - \left(F\left(\frac{k+1}{m}\right) - F\left(\frac{k}{m}\right)\right)\right] b_{k,m-1}(x) : x \in [a,b]\right\}. \tag{3.2.3}$$

Set $n_l := 2^l$.

$$\mathcal{H}_l := \bigcup_{m=m_{n_{l-1}}+1}^{m_{n_l}} \mathcal{F}_m^*. \tag{3.2.4}$$

3.2.1 Lemma. For $\epsilon \le 1/3$, we have the entropy bound for all Borel probability measures Q on $\mathbb{R}$,

$$N\left(\mathcal{F}_m, L_2(Q), \epsilon\right) \le \frac{3m}{\epsilon}, \tag{3.2.5}$$

and

$$N(\mathcal{H}_l, L_2(Q), \epsilon) \le (\sqrt{6} m_{n_l}/\epsilon)^2. \tag{3.2.6}$$

Proof. It's easy to see that $0 \le b_{k,m-1}(x) \le 1$ and the maximum of $b_{k,m-1}(x)$ occurs at $x = k/(m-1)$. Let $z_{m,k}$ denote this maximum value. Given $0 < \epsilon < 1$, for fixed m and

$0 \leq k \leq m-1$, take all points $x \in [0,1]$, such that

$$b_{k,m-1}(x) = 0, \epsilon, 2\epsilon, ..., \lfloor \frac{z_{m,k}}{\epsilon} \rfloor \epsilon, z_{m,k} \tag{3.2.7}$$

if they exist. Since $b_{k,m-1}(x)$ is monotonically increasing on $[0, k/(m-1)]$ and monotonically decreasing on $[k/(m-1), 1]$, on each level k, the number $N_{m,k}(\epsilon)$ of such points satisfies $N_{m,k}(\epsilon) \leq 2(\lfloor 1/\epsilon \rfloor + 1) + 1 \leq 3/\epsilon := N(\epsilon)$ such points for $\epsilon \leq 1/3$. Again by monotonicity property of $b_{k,m-1}(\cdot)$, if $x_k^1, x_k^2, ..., x_k^{N_{m,k}(\epsilon)}$ is the set of these points in increasing order, then for all i, k, and all $x, y \in [x_k^i, x_k^{i+1}]$,

$$|b_{k,m-1}(x) - b_{k,m-1}(y)| \leq \epsilon. \tag{3.2.8}$$

Claim that $\{\sum_{k=0}^{m-1} I_{(k/m,(k+1)/m]}(\cdot) b_{k,m-1}(x_l^i), l = 0, ..., m-1; i = 1, ..., N_{m,l}(\epsilon)\}$ are the centers of a covering of $\mathcal{F}_m$ with radius ϵ. To see this, we order all the x_l^i, $0 \leq l \leq m-1$ in increasing order and note that any interval between two consecutive such points is contained in or coincide with an interval of two consecutive points from the level l for all $0 \leq l \leq m-1$. Hence by (3.2.8), if x falls in one of these intervals with e.g. left end point x_l^i for some l and i, we have

$$\max_{0 \leq k \leq m-1} |b_{k,m-1}(x) - b_{k,m-1}(x_l^i)| \leq \epsilon. \tag{3.2.9}$$

Therefore, for any probability measure Q, since $I_{(k/m,(k+1)/m]}(\cdot)$ are disjoint for different k, we have for any $x \in (0,1)$,

$$\min_{l,i} \int \left| \sum_{k=0}^{m-1} I_{(k/m,(k+1)/m]}(\cdot) b_{k,m-1}(x) - \sum_{k=0}^{m-1} I_{(k/m,(k+1)/m]}(\cdot) b_{k,m-1}(x_l^i) \right|^2 dQ$$

$$= \min_{l,i} \int \sum_{k=0}^{m-1} I_{(k/m,(k+1)/m]}(\cdot)(b_{k,m-1}(x) - b_{k,m-1}(x_l^i))^2 dQ \tag{3.2.10}$$
$$\leq \min_{l,i} \max_{0\leq k\leq m-1} (b_{k,m-1}(x) - b_{k,m-1}(x_l^i))^2 \leq \epsilon^2.$$

Thus the claim is true and we conclude that

$$N(\mathcal{F}_m, L_2(Q), \epsilon) \leq \frac{3m}{\epsilon} := N_m. \tag{3.2.11}$$

Furthermore, since

$$\min_{l,i} \int \left| \sum_{k=0}^{m-1} \left(F\left(\frac{k+1}{m}\right) - F\left(\frac{k}{m}\right)\right) b_{k,m-1}(x) - \sum_{k=0}^{m-1} \left(F\left(\frac{k+1}{m}\right) - F\left(\frac{k}{m}\right)\right) b_{k,m-1}(x_l^i) \right|^2 dQ$$
$$\leq \min_{l,i} \max_{0\leq k\leq m-1} (b_{k,m-1}(x) - b_{k,m-1}(x_l^i))^2 \int \left(\sum_{k=0}^{m-1} \left[F\left(\frac{k+1}{m}\right) - F\left(\frac{k}{m}\right)\right]\right)^2 dQ$$
$$\leq \min_{l,i} \max_{0\leq k\leq m-1} (b_{k,m-1}(x) - b_{k,m-1}(x_l^i))^2 \leq \epsilon^2, \tag{3.2.12}$$

then $\{\sum_{k=0}^{m-1}[I(\cdot)_{k/m,(k+1)/m} - (F((k+1)/m) - F(k/m))]b_{k,m-1}(x_l^i), l = 0, ..., m-1; i = 1, ..., N_{m,l}(\epsilon)\}$ are the centers of a covering for $\mathcal{F}_m^*$ with radius 2ϵ. For each m, we may find $x_m^1, ..., x_m^{N_m}$ so that

$$\mathcal{C}_m := \left\{ \sum_{k=0}^{m-1} \left[I\left(\frac{k}{m} < X_i \leq \frac{k+1}{m}\right) - \left(F\left(\frac{k+1}{m}\right) - F\left(\frac{k}{m}\right)\right)\right] b_{k,m-1}(x_m^j), j = 1, ..., N_m \right\} \tag{3.2.13}$$

is a $2\epsilon-$dense subset of $\mathcal{F}_m^*$. So $\bigcup_{m=m_{n_l}-1}^{m_{n_l}} \mathcal{C}_m$ are the centers of a covering for $\mathcal{H}_l$ with radius 2ϵ. It follows that, for any probability measure Q,

$$N(\mathcal{H}_l, L_2(Q), \epsilon) \leq \frac{6}{\epsilon}(m_{n_{l-1}} + 1 + m_{n_{l-1}} + 2 + ... + m_{n_l})$$
$$\leq (\sqrt{6} m_{n_l}/\epsilon)^2 \tag{3.2.14}$$

if $\epsilon \leq 1/3$ and $l > 1$. □

The following lemma helps to estimate the constant envelope U_l of the class $\mathcal{H}_l$. The derivation of the upper bound uses the method in Theorem 1.5.2, Lorentz (1986). In addition, we need to argue the bound holds uniformly for all $x \in [a, b]$, which is not implied directly by that theorem.

3.2.2 Lemma. For any $0 < a \leq x \leq b < 1$,

$$\sup_{x\in[a,b]} \max_{0\leq k\leq m} b_{k,m}(x) \leq \frac{C_{a,b}}{\sqrt{m}} \tag{3.2.15}$$

for m large enough depending on a, b.

Proof. Fix $x \in [a, b]$ and $\alpha \in (1/3, 1/2)$, we first consider $\max_{k:|k/m-x|>m^{-\alpha}} b_{k,m}(x)$. By (8), pg 15, Lorentz (1986), for any $l > 0$,

$$\max_{k:|k/m-x|>m^{-\alpha}} b_{k,m}(x) \leq Cm^{-l}, \tag{3.2.16}$$

where C depends only on α and l. Now we consider $\max_{k:|k/m-x|\leq m^{-\alpha}} b_{k,m}(x)$, where $b_{k,m}(x)$ was defined in (0.2.1). Using Stirling's formula for $m!$ (Robbins, 1955),

$$m! = \left(\frac{m}{e}\right)^m \sqrt{2\pi m}\exp(\lambda_m), \tag{3.2.17}$$

where $1/(12m + 1) < \lambda_m < 1/(12m)$. For k satisfying $|k/m - x| \leq m^{-\alpha}$, $0 \leq k \leq m$,

$$\binom{m}{k} x^k(1-x)^{m-k} = \frac{m^m\sqrt{m}}{\sqrt{2\pi k(m-k)}k^k(m-k)^{m-k}} x^k(1-x)^{m-k} H_{m,k}, \tag{3.2.18}$$

where $H_{m,k} = e^{\lambda_m}/[e^{\lambda_k}e^{\lambda_{m-k}}]$. Since $k \geq 0$, $m - k \geq 0$ for all $x \in [a, b]$,

$$\sup_{x\in[a,b]} \max_k H_{m,k} \leq e^{\lambda_m} \searrow 1 \;\; as \;\; m \to \infty. \tag{3.2.19}$$

Another factor on the right hand side of (3.2.18) is

$$\frac{m}{2\pi k(m-k)} = m^{-1} \frac{1}{2\pi \frac{k}{m}\left(1 - \frac{k}{m}\right)}. \tag{3.2.20}$$

Given the range of x and k, we have

$$\frac{k}{m} \geq a - m^{-\alpha}, 1 - \frac{k}{m} \geq 1 - b - m^{-\alpha}. \tag{3.2.21}$$

So if m is sufficiently large depending on $[a, b]$,

$$\max_{k:|k/m-x|\leq m^{-\alpha}} \frac{m}{2\pi k(m-k)} \leq \frac{1}{2\pi m(a - m^{-\alpha})(1 - b - m^{-\alpha})} \leq \frac{C_{a,b}}{m}. \tag{3.2.22}$$

The remaining factor in (3.2.18) is

$$W_{m,k}(x) := \frac{m^m}{k^k(m-k)^{m-k}} x^k(1-x)^{m-k} = \left(\frac{mx}{k}\right)^k \left(\frac{m(1-x)}{m-k}\right)^{m-k}. \tag{3.2.23}$$

And then,

$$-\log W_{m,k}(x) = k \log\left(1 + x^{-1}\left(\frac{k}{m} - x\right)\right) + (m-k)\log\left(1 - (1-x)^{-1}\left(\frac{k}{m} - x\right)\right). \tag{3.2.24}$$

Using Taylor's formula for $|u| < 1$,

$$\log(1+u) = u - \frac{1}{2}u^2\rho, \rho = 1 + \epsilon u, \tag{3.2.25}$$

and

$$\log(1-u) = -u - \frac{1}{2}u^2\rho_1, \rho_1 = 1 + \epsilon_1 u, \tag{3.2.26}$$

where ϵ and ϵ_1 remain bounded as $u \to 0$.

$$\begin{aligned}
-\log W_{m,k}(x) =& k\left[x^{-1}\left(\frac{k}{m}-x\right) - \frac{1}{2}x^{-2}\left(\frac{k}{m}-x\right)^2\rho_x\right] \\
&- (m-k)\left[(1-x)^{-1}\left(\frac{k}{m}-x\right) + \frac{1}{2}(1-x)^{-2}\left(\frac{k}{m}-x\right)^2\rho_{1x}\right] \\
=& mx\left(1 + x^{-1}\left(\frac{k}{m}-x\right)\right)\left[x^{-1}\left(\frac{k}{m}-x\right) - \frac{1}{2}x^{-2}\left(\frac{k}{m}-x\right)^2\rho_x\right] - m(1-x)\times \\
&\times\left(1-(1-x)^{-1}\left(\frac{k}{m}-x\right)\right)\left[(1-x)^{-1}\left(\frac{k}{m}-x\right) + \frac{1}{2}(1-x)^{-2}\left(\frac{k}{m}-x\right)^2\rho_{1x}\right] \\
=& m\left(\frac{k}{m}-x\right)\left\{\left(1 + x^{-1}\left(\frac{k}{m}-x\right)\right)\left[1 - \frac{1}{2}x^{-1}\left(\frac{k}{m}-x\right)\rho_x\right]\right. \\
&\left.-\left(1-(1-x)^{-1}\left(\frac{k}{m}-x\right)\right)\left[1 + \frac{1}{2}(1-x)^{-1}\left(\frac{k}{m}-x\right)\rho_{1x}\right]\right\} \\
=& m\left(\frac{k}{m}-x\right)^2\left\{\left[x^{-1} - \frac{1}{2}x^{-1}\rho_x - \frac{1}{2}x^{-2}\left(\frac{k}{m}-x\right)\rho_x\right]\right. \\
&\left.+\left[(1-x)^{-1} - \frac{1}{2}(1-x)^{-1}\rho_{1x} + \frac{1}{2}(1-x)^{-2}\left(\frac{k}{m}-x\right)\rho_{1x}\right]\right\}.
\end{aligned} \tag{3.2.27}$$

Finally, we get

$$\begin{aligned}
-\log W_{m,k}(x) =& (2x(1-x))^{-1}\, m\left(\frac{k}{m}-x\right)^2 + m\left(\frac{k}{m}-x\right)^2\times \\
&\times\left[x^{-1}\left(\frac{1}{2} - \frac{1}{2}\rho_x - \frac{1}{2}x^{-1}\rho_x\left(\frac{k}{m}-x\right)\right)\right. \\
&\left.+(1-x)^{-1}\left(\frac{1}{2} - \frac{1}{2}\rho_{1x} + \frac{1}{2}(1-x)^{-1}\rho_{1x}\left(\frac{k}{m}-x\right)\right)\right].
\end{aligned} \tag{3.2.28}$$

So we can write

$$W_{m,k}(x) = \exp(\log W_{m,k}(x)) = \exp\left\{-\frac{m\left(\frac{k}{m}-x\right)^2}{2x(1-x)}\right\}\exp\left(-Q_{m,k}(x)\right), \tag{3.2.29}$$

where $Q_{m,k}(x)$ is the second term on the right hand side in (3.2.28). By (3.2.25) and (3.2.26), ρ_x and ρ_{1x} satisfy the following:

$$\frac{1}{2} - \frac{1}{2}\rho_x = -\frac{1}{2}\epsilon_x x^{-1}\left(\frac{k}{m} - x\right) \tag{3.2.30}$$

and

$$\frac{1}{2} - \frac{1}{2}\rho_{1x} = -\frac{1}{2}\epsilon_{1,x}(1-x)^{-1}\left(\frac{k}{m} - x\right) \tag{3.2.31}$$

For all $x \in [a,b]$ and k in the specified range, as $m \to \infty$, $x^{-1}(k/m - x)$ and $(1 - x)^{-1}(k/m - x)$ tend to 0 uniformly. So ϵ_x and $\epsilon_{1,x}$ are uniformly bounded. Moreover, $x^{-1}, (1-x)^{-1}, \rho_x, \rho_{1x}$ are also bounded for $x \in [a,b]$ if m is large enough. By this argument,

$$|Q_{m,k}(x)| \leq m\left(\frac{k}{m} - x\right)^2 A_x\left|\frac{k}{m} - x\right| \leq A_x m^{1-3\alpha}, \tag{3.2.32}$$

where A_x is uniformly bounded in absolute value for $x \in [a,b]$ when m is large enough. Therefore, $|Q_{m,k}(x)| \to 0$ uniformly for all k and x in the specified range. And by (3.2.29), $W_{m,k}(x)$ is uniformly bounded by a constant when m is large enough. Combining this with (3.2.18), (3.2.19), (3.2.22) and (3.2.23), we obtain

$$\sup_{x\in[a,b]} \max_{k:|k/m-x|\leq m^{-\alpha}} b_{k,m}(x) \leq \frac{C_{a,b}}{\sqrt{m}}. \tag{3.2.33}$$

Thus, for m sufficiently large, by (3.2.16) and (3.2.33),

$$\sup_{x\in[a,b]} \max_{0\le k\le m} \binom{m}{k} x^k(1-x)^{m-k} \le \sup_{x\in[a,b]} \max_{k:|k/m-x|\le m^{-\alpha}} b_{k,m}(x) \vee \sup_{x\in[a,b]} \max_{k:|k/m-x|>m^{-\alpha}} b_{k,m}(x) \le \frac{C_{a,b}}{\sqrt{m}}. \tag{3.2.34}$$

□

We also need an estimate on σ_l^2 (defined above (3.1.3)) of the class $\mathcal{H}_l$, which is the purpose of the next lemma. We improve the result in Lemma 3.1, Babu, Chaubey, Canty (2002) so that the convergence holds uniformly for $x \in [a,b]$.

3.2.3 Lemma. Let $0 < a \le b < 1$. As $m \to \infty$,

$$2\sqrt{\pi m x(1-x)} \sum_{k=0}^{m} b_{k,m}^2(x) \to 1 \tag{3.2.35}$$

uniformly for all $x \in [a,b]$.

Proof. Let U_i, W_j, $i, j = 1, ..., m$, be i.i.d. Bernoulli random variables satisfying

$$\Pr(U_i = 1) = x, \;\; \Pr(U_i = 0) = 1 - x. \tag{3.2.36}$$

Set $R_i := (U_i - W_i)/\sqrt{2x(1-x)}$. Then R_i has a lattice distribution with span $h = 1/\sqrt{2x(1-x)}$. We observe that,

$$\sum_{k=0}^{m} b_{k,m}^2(x) = \sum_{k=0}^{m} \Pr\left(\sum_{i=1}^{m} U_i = \sum_{i=1}^{m} W_i = k\right) = \Pr\left(\sum_{i=1}^{m} R_i = 0\right). \tag{3.2.37}$$

We may verify that $\mathbb{E}R_i = 0$ and $\mathbb{E}R_i^2 = 1$. Thus, we can apply Theorem 3, XV.5, Feller

(1971) to $\Pr(\sum_{i=1}^{m} R_i = 0)$ and get

$$\sqrt{2mx(1-x)}\sum_{k=0}^{m} b_{k,m}^2(x) - \frac{1}{\sqrt{2\pi}} \to 0. \tag{3.2.38}$$

By inspecting (5.14) in that proof, we see that for all $h > 0$, the first integral tends to zero uniformly by the proof of Theorem 2, XV.5, Feller (1971). The second integral tends to zero uniformly for all $\sqrt{n}\pi/h \to \infty$ (In this context, $\sqrt{m}\pi/h \to \infty$). This is true since for $x \in [a,b]$, h is bounded above by $1/\sqrt{2a(1-a)}$ or $1/\sqrt{2b(1-b)}$. We conclude that, the convergence in (3.2.38) is uniform for all $x \in [a,b]$. □

Now we prove the upper bound result.

3.2.4 Theorem. Let $m = m_n \asymp n^\delta$, where $0 < \delta < 2$. Assume the density f is bounded above. Then there exists a positive constant $C_{f,a,b}$, such that

$$\limsup_{n\to\infty} \sqrt{\frac{n}{m_n^{1/2}\log m_n}} \sup_{x\in[a,b]} \left|\hat{f}_{m,n}(x) - \mathbb{E}\hat{f}_{m,n}(x)\right| = C_{f,a,b} \ \ a.s. \tag{3.2.39}$$

Proof. By the definition of $\hat{f}_{m,n}(x)$ in (0.2.3), it is equivalent to proving that

$$\limsup_{n\to\infty} \sqrt{\frac{m_n^{3/2}}{n\log m_n}} \sup_{x\in[a,b]} \left|\sum_{i=1}^{n}\sum_{k=0}^{m_n-1}\left[I\left(\frac{k}{m_n} < X_i \le \frac{k+1}{m_n}\right) - \left(F\left(\frac{k+1}{m_n}\right) - F\left(\frac{k}{m_n}\right)\right)\right] b_{k,m_n-1}(x)\right|$$
$$= C_{f,a,b} \ \ a.s. \tag{3.2.40}$$

Let $n_l = 2^l$. We consider the blocking from n_{l-1} to n_l,

$$\Pr\left\{\max_{n_{l-1}<n\le n_l}\sup_{x\in[a,b]}\sqrt{\frac{m_n^{3/2}}{n\log m_n}}\left|\sum_{i=1}^{n}\sum_{k=0}^{m_n-1}\left[I\left(\frac{k}{m_n} < X_i \le \frac{k+1}{m_n}\right)\right.\right.\right.$$
$$\left.\left.\left. -\left(F\left(\frac{k+1}{m_n}\right) - F\left(\frac{k}{m_n}\right)\right)\right] b_{k,m_n-1}(x)\right| > s\right\}$$

$$\leq \Pr\left\{ \max_{n_{l-1}<n\leq n_l} \sup_{\substack{x\in[a,b] \\ m_{n_{l-1}}<m_n\leq m_{n_l}}} \left| \sum_{i=1}^{n} \sum_{k=0}^{m_n-1} \left[I\left(\frac{k}{m_n} < X_i \leq \frac{k+1}{m_n}\right) \right.\right.\right. \tag{3.2.41}$$

$$\left.\left.\left. - \left(F\left(\frac{k+1}{m_n}\right) - F\left(\frac{k}{m_n}\right)\right)\right] b_{k,m_n-1}(x) \right| > s\sqrt{\frac{n_{l-1}\log m_{n_l}}{m_{n_l}^{3/2}}} \right\}$$

since $\max_{m_{n_{l-1}}<m_n\leq m_{n_l}} \sqrt{m_n^{3/2}/n\log m_n} \leq \sqrt{m_{n_l}^{3/2}/n_{l-1}\log m_{n_l}}$. To bound the last probability, we will apply Corollary 3.1.2 to the class $\mathcal{H}_l$. Let U_l be the constant envelope of $\mathcal{H}_l$. By Lemma 3.2.2,

$$U_l \leq 2 \max_{m_{n_{l-1}}<m_n\leq m_{n_l}} \max_{0\leq k\leq m_n} b_{k,m_n-1}(x) \leq \frac{C_{a,b}}{\sqrt{m_{n_{l-1}}}} \tag{3.2.42}$$

for l sufficiently large depending on $[a,b]$. By Lemma 3.2.1, $\mathcal{H}_l$ is a VC-class of functions for each l. The VC-characteristics of $\mathcal{H}_l$ are $A_l = \sqrt{6}m_{n_l}m_{n_{l-1}}^{1/2}/C_{a,b}$, $v_l = 2$. Then we estimate σ_l^2. Using the hypothesis that f is bounded above and Lemma 3.2.3, we get

$$\begin{aligned}
&\sup_{\substack{m_{n_{l-1}}<m_n\leq m_{n_l} \\ x\in[a,b]}} \mathbb{E}\left[\sum_{k=0}^{m_n-1} I\left(\frac{k}{m_n} < X_i \leq \frac{k+1}{m_n}\right) b_{k,m_n-1}(x)\right]^2 \\
&= \sup_{\substack{m_{n_{l-1}}<m_n\leq m_{n_l} \\ x\in[a,b]}} \mathbb{E}\sum_{k=0}^{m_n-1} I\left(\frac{k}{m_n} < X_i \leq \frac{k+1}{m_n}\right) b_{k,m_n-1}^2(x) \\
&= \sup_{\substack{m_{n_{l-1}}<m_n\leq m_{n_l} \\ x\in[a,b]}} \sum_{k=0}^{m_n-1} \left[F\left(\frac{k+1}{m_n}\right) - F\left(\frac{k}{m_n}\right)\right] b_{k,m_n-1}^2(x) \qquad (3.2.43) \\
&\leq \sup_{\substack{m_{n_{l-1}}<m_n\leq m_{n_l} \\ x\in[a,b]}} \frac{\|f\|_\infty}{m_n} \sum_{k=0}^{m_n-1} b_{k,m_n-1}^2(x) \leq \frac{C_{f,a,b}}{m_{n_{l-1}}^{3/2}}.
\end{aligned}$$

So we may take $\sigma_l^2 = C_{f,a,b}m_{n_{l-1}}^{-3/2}$. To apply Corollary 3.1.2 to the class $\mathcal{H}_l$, we first check condition (3.1.9).

$$\sqrt{n_l}\sigma_l = 2^{l/2}\frac{C_{f,a,b}}{m_{n_{l-1}}^{3/4}} \asymp 2^{l/2-3\delta(l-1)/4}, \tag{3.2.44}$$

and

$$U_l\sqrt{\log\frac{A_lU_l}{\sigma_l}} \asymp m_{n_{l-1}}^{-1/2}\sqrt{\log\frac{m_{n_l}}{m_{n_{l-1}}^{-3/4}}} \asymp 2^{-\delta(l-1)/2}\sqrt{\frac{7l\delta}{4}-\frac{3\delta}{4}}\sqrt{\log 2}. \tag{3.2.45}$$

Then since $\delta < 2$,

$$\frac{\sqrt{n_l}\sigma_l}{U_l\sqrt{\log\left(A_lU_l/\sigma_l\right)}} \asymp \frac{2^{\frac{l}{2}-\frac{\delta l}{4}+\frac{\delta}{4}}}{\sqrt{\frac{7l\delta}{4}-\frac{3\delta}{4}}} \to \infty \tag{3.2.46}$$

as $l \to \infty$. So (3.1.9) holds for some λ independent of l when l is large enough. We pick s so that $s\sqrt{n_{l-1}\log m_{n_l}/m_{n_l}^{3/2}}$ in (3.2.41) is an admissible choice in (3.1.10). s should satisfy

$$C_1\sqrt{n_l}\sigma_l\sqrt{\log\left(A_lU_l/\sigma_l\right)} \le s\sqrt{\frac{n_{l-1}\log m_{n_l}}{m_{n_l}^{3/2}}} \le C_2\frac{n_l\sigma_l^2}{U_l}. \tag{3.2.47}$$

We calculate the order of the three terms in the above inequality. By (3.2.44) and (3.2.45),

$$\begin{aligned}\sqrt{n_l}\sigma_l\sqrt{\log\left(A_lU_l/\sigma_l\right)} &\asymp 2^{l/2-3\delta(l-1)/4}\sqrt{\frac{7l\delta}{4}-\frac{3\delta}{4}}\\ &\asymp 2^{l/2-3\delta l/4+3\delta/4}\sqrt{\frac{7l\delta}{4}-\frac{3\delta}{4}},\end{aligned} \tag{3.2.48}$$

$$s\sqrt{\frac{n_{l-1}\log m_{n_l}}{m_{n_l}^{3/2}}} \asymp s\sqrt{\frac{2^{l-1}l\delta}{2^{3l\delta/2}}} \asymp s2^{l/2-3l\delta/4-1/2}\sqrt{l\delta}, \tag{3.2.49}$$

and

$$\frac{n_l \sigma_l^2}{U_l} \asymp \frac{2^l m_{n_{l-1}}^{-3/2}}{m_{n_{l-1}}^{-1/2}} \asymp 2^{l-l\delta+\delta}. \tag{3.2.50}$$

So we may pick s as a positive constant (depending on f, a, b) to satisfy (3.2.47). Then according to Corollary (3.1.2),

$$\begin{aligned}
&\Pr\left\{\max_{n_{l-1}<n\le n_l} \sup_{\substack{x\in[a,b] \\ m_{n_{l-1}}<m_n\le m_{n_l}}} \left|\sum_{i=1}^{n}\sum_{k=0}^{m_n-1}\left[I\left(\frac{k}{m_n} < X_i \le \frac{k+1}{m_n}\right)\right.\right.\right. \\
&\qquad \left.\left.\left. -\left(F\left(\frac{k+1}{m_n}\right) - F\left(\frac{k}{m_n}\right)\right)\right] b_{k,m_n-1}(x)\right| > s\sqrt{\frac{n_{l-1}\log m_{n_l}}{m_{n_l}^{3/2}}}\right\} \\
&\le K\exp\left\{-C_3 \frac{s^2 n_{l-1}\left(\log m_{n_l}\right)/m_{n_l}^{3/2}}{n_l \sigma_l^2}\right\} \\
&\le K\exp\left\{-C_{f,a,b} 2^{-1-3\delta/2} l\delta \log 2\right\}.
\end{aligned} \tag{3.2.51}$$

We have shown that

$$\begin{aligned}
&\sum_l \Pr\left\{\max_{n_{l-1}<n\le n_l} \sup_{x\in[a,b]} \sqrt{\frac{m_n^{3/2}}{n\log m_n}} \left|\sum_{i=1}^{n}\sum_{k=0}^{m_n-1}\left[I\left(\frac{k}{m_n} < X_i \le \frac{k+1}{m_n}\right)\right.\right.\right. \\
&\qquad \left.\left.\left. -\left(F\left(\frac{k+1}{m_n}\right) - F\left(\frac{k}{m_n}\right)\right)\right] b_{k,m_n-1}(x)\right| > s\right\} \\
&\le \sum_l K\exp\left\{-C_{f,a,b} l\right\} < \infty.
\end{aligned} \tag{3.2.52}$$

Now by Borel-Cantelli and zero-one law, we have proved (3.2.39). □

3.2.5 Remark. An application of Proposition 2.1, Giné and Guillou (2001) to the class $\mathcal{F}_{m_n}$ gives the expectation bound directly. Under the same assumptions of Theorem 3.2.4,

$$\mathbb{E}\sup_{x\in[a,b]}\left|\hat{f}_{m,n}(x) - \mathbb{E}\hat{f}_{m,n}(x)\right| \le C_{f,a,b} m_n^{1/4}\sqrt{\frac{\log m_n}{n}}. \tag{3.2.53}$$

3.3 Lower bound

We need a modification of Proposition 2, Einmahl and Mason (2000). Let $\{a_n\}_{n\geq 1}$ be a sequence of positive constants converging to zero satisfying

$$\frac{|\log a_n|}{\log\log n} \to \infty, \quad \lim_{n\to\infty} \frac{a_n^{9/2} n}{|\log a_n|} = \infty, \quad \frac{|\log a_n|}{\sqrt{n}} \to 0 \ \ as \ \ n \to \infty. \tag{3.3.1}$$

Let $X, X_1, ..., X_n$ be a sequence of i.i.d. random variables on $\mathbb{R}$. Let $\{g_i^{(n)}\}_{i=1}^{k_n}$ be a sequence of real valued measurable functions on $\mathbb{R}$ satisfying:

$$\Pr\{g_i^{(n)}(X) \neq 0, g_j^{(n)}(X) \neq 0\} = 0, i \neq j, \tag{3.3.2}$$

$$\sum_{i=1}^{k_n} \Pr\{g_i^{(n)}(X) \neq 0\} \leq 1/2, \tag{3.3.3}$$

$$a_n k_n \to r \ \ as \ \ n \to \infty, \tag{3.3.4}$$

for some $0 < r < \infty$. Moreover, there exist $-\infty < \mu_1 < 0 < \mu_2 < \infty$ and $0 < \sigma_1 < \sigma_2 < \infty$, such that uniformly in $i = 1, ..., k_n$, when n is large enough,

$$a_n^3 \mu_1 \leq \mathbb{E} g_i^{(n)}(X) \leq a_n^3 \mu_2, \tag{3.3.5}$$

$$\sigma_1 a_n^{9/4} \leq \sqrt{\mathrm{V}ar g_i^{(n)}(X)} \leq \sigma_2 a_n^{9/4}, \tag{3.3.6}$$

and for some $0 < B < \infty$,

$$\left| g_i^{(n)} \right| \leq B \tag{3.3.7}$$

uniformly in $i = 1, ..., k_n$ for all large n. Set $G_n(g) := \frac{1}{n}\sum_{i=1}^{n} g(X_i)$.

3.3.1 Proposition. Assume that (3.3.1)-(3.3.7) hold, then with probability one, for each $0 < \epsilon < 1$,

$$\max_{1\le i\le k_n} \frac{\sqrt{n}\{G_n(g_i^{(n)}) - \mathbb{E}g_i^{(n)}(X)\}}{\sqrt{2|\log a_n| Var\left(g_i^{(n)}(X)\right)}} \ge 1-\epsilon \tag{3.3.8}$$

for all n large enough depending on ϵ.

Proof. The proof is a straightforward modification of Proposition 2, Einmahl and Mason (2000). To prove (3.3.8), let $A_i^{(n)}$ be the event

$$\frac{\sqrt{n}\{G_n(g_i^{(n)}) - \mathbb{E}g_i^{(n)}(X)\}}{\sqrt{2|\log a_n| Var\left(g_i^{(n)}(X)\right)}} < 1-\epsilon. \tag{3.3.9}$$

Then $(\bigcap_{i=1}^{k_n} A_i^{(n)})^C$ is the event

$$\max_{1\le i\le k_n} \frac{\sqrt{n}\{G_n(g_i^{(n)}) - \mathbb{E}g_i^{n}(X)\}}{\sqrt{2|\log a_n| Var\left(g_i^{(n)}(X)\right)}} \ge 1-\epsilon. \tag{3.3.10}$$

We use a Poissonization technique to bound $\Pr\{\bigcap_{i=1}^{k_n} A_i^{(n)}\}$. Let $N_n(g) :\stackrel{d}{=} \frac{1}{n}\sum_{i=1}^{\pi_n} g(X_i)$, where $\pi_n \sim Poisson(n)$ and is independent of $X_1, ..., X_n$. Let $B_i^{(n)}$ be the Poissonized version of $A_i^{(n)}$. That is, $B_i^{(n)}$ is the event

$$\frac{\sqrt{n}\{N_n(g_i^{n}) - \mathbb{E}g_i^{(n)}(X)\}}{\sqrt{2|\log a_n| Var(g_i^{(n)}(X))}} < 1-\epsilon. \tag{3.3.11}$$

By Lemma 8, Einmahl and Mason (2000), we have

$$\Pr\left\{\bigcap_{i=1}^{k_n} A_i^{(n)}\right\} \le 2\Pr\{B_1^{(n)}\}\cdots\Pr\{B_{k_n}^{(n)}\}. \tag{3.3.12}$$

To bound the right hand side of the above inequality, we consider $\Pr\{(B_i^{(n)})^C\}$. Let $\eta > 0$. Conditioning on $\pi_n = l$, we get

$$\Pr\left\{(B_i^{(n)})^C\right\} \geq \sum_{l:|n-l|\leq \eta\sqrt{n}} \Pr\left\{(B_i^{(n)})^C | \pi_n = l\right\} \Pr\{\pi_n = l\}. \tag{3.3.13}$$

It follows from the definitions of $N_n(g), G_n(g)$ and the independence of π_n and X_i's, that

$$\begin{aligned}
\Pr\left\{(B_i^{(n)})^C | \pi_n = l\right\} &= \Pr\left\{ \frac{\sqrt{n}\left[N_n(g_i^{(n)}) - \mathbb{E}g_i^{(n)}\right]}{\sqrt{2|\log a_n| Var(g_i^{(n)})}} \geq 1 - \epsilon \middle| \pi_n = l \right\} \\
&= \Pr\left\{ \frac{(\sqrt{n}/n)\left[\sum_{j=1}^{l} g_i^{(n)}(X_j) - n\mathbb{E}g_i^{(n)}\right]}{\sqrt{2|\log a_n| Var(g_i^{(n)})}} \geq 1 - \epsilon \middle| \pi_n = l \right\} \\
&= \Pr\left\{ \frac{\frac{l\sqrt{n}}{n}\left[G_l(g_i^{(n)}) - \mathbb{E}(g_i^{(n)})\right] + \frac{\sqrt{n}(l-n)}{n}\mathbb{E}g_i^{(n)}}{\sqrt{2|\log a_n| Var(g_i^{(n)})}} \geq 1 - \epsilon \right\}.
\end{aligned} \tag{3.3.14}$$

In (3.3.5), with out loss of generality, we can assume $|\mu_1| \leq \mu_2$. Combining it with (3.3.6) and $|n - l| \leq \eta\sqrt{n}$, we obtain

$$\begin{aligned}
\frac{(\sqrt{n}(n-l)/n)\mathbb{E}g_i^{(n)}}{\sqrt{2|\log a_n| Var(g_i^{(n)})}} &\leq \frac{\eta a_n^3 \mu_2}{\sigma_1 a_n^{9/4}\sqrt{2|\log a_n|}} \\
&\leq \frac{\eta\mu_2}{\sigma_1}\sqrt{\frac{a_n^{3/2}}{2|\log a_n|}} \\
&\leq \frac{\eta\mu_2}{\sigma_1}
\end{aligned} \tag{3.3.15}$$

when n is large enough. The last inequality follows since $a_n \to 0$ and then $a_n^{3/2}/(2|\log a_n|) \to 0$ as $n \to \infty$. So from the last equation in (3.3.14), we get

$$\Pr\left\{(B_i^{(n)})^C | \pi_n = l\right\} \geq \Pr\left\{ \frac{(l\sqrt{n}/n)\left[G_l(g_i^{(n)}) - \mathbb{E}(g_i^{(n)})\right]}{\sqrt{2|\log a_n| Var(g_i^{(n)})}} \geq 1 - \epsilon + \frac{\eta\mu_2}{\sigma_1} \right\}. \tag{3.3.16}$$

We choose $\eta > 0$ so that $\eta\mu_2/\sigma_1 < (\epsilon/2)^2$, then

$$\begin{aligned}\Pr\left\{(B_i^{(n)})^C|\pi_n = l\right\} &\geq \Pr\left\{\frac{(l\sqrt{n}/n)\left[G_l(g_i^{(n)}) - \mathbb{E}(g_i^{(n)})\right]}{\sqrt{2|\log a_n|\mathrm{V}ar(g_i^{(n)})}} \geq \left(1-\frac{\epsilon}{2}\right)^2\right\} \\ &= \Pr\left\{\sum_{j=1}^{l}\frac{\left[g_i^{(n)}(X_j) - \mathbb{E}g_i^{(n)}(X)\right]}{\sqrt{\mathrm{V}ar(g_i^{(n)}(X))}} \geq \left(1-\frac{\epsilon}{2}\right)^2\sqrt{2n|\log a_n|}\right\}.\end{aligned} \tag{3.3.17}$$

By (3.3.13), it suffices to find the lower bound for the last probability where l is in the range $|n-l| \leq \eta\sqrt{n}$. Let $\xi_j := (g_i^{(n)}(X_j) - \mathbb{E}g_i^{(n)}(X))/\sqrt{\mathrm{V}arg_i^{(n)}(X)}$ for $j = 1, ..., l$. ξ_j's are i.i.d. mean zero random variables. We will apply Kolmogorov's inequality to ξ_j's. Kolmogorov's inequality is stated as follows (Lemma 2, Pg 341, Chow and Teicher, 1978):

Let $\xi_1, ..., \xi_n$ be independent mean zero random variables with $s_n^2 = \sum_{i=1}^n \mathbb{E}\xi_i^2 > 0$ and $\Pr\{|\xi_i| \leq d_i\} = 1$, where $0 < d_i \uparrow$, $1 \leq i \leq n$. If $\lim_{n\to\infty} d_n x_n/s_n = 0$, where $x_n > x_0 > 0$, then for every $\gamma \in (0,1)$, there is a $C_\gamma \in (0, 1/2)$, such that for all large n,

$$\Pr\left\{\sum_{i=1}^{n}\xi_i \geq (1-\gamma)^2 s_n x_n\right\} > C_\gamma \exp(-x_n^2(1-\gamma)(1-\gamma^2)/2). \tag{3.3.18}$$

We calculate

$$s_l = \sqrt{\sum_{j=1}^{l}\mathbb{E}\xi_j^2} = \sqrt{l}, \tag{3.3.19}$$

and by (3.3.6) and (3.3.7),

$$d_j = \frac{2B}{\sqrt{\mathrm{V}arg_i^{(n)}}} \leq \frac{2B}{\sigma_1 a_n^{9/4}}. \tag{3.3.20}$$

So we may take $x_l = \sqrt{2n|\log a_n|/l}$. By (3.3.1) and the range of l, we can check that

$$\lim_{l\to\infty}\frac{d_l x_l}{s_l} = \lim_{l\to\infty}\frac{2B\sqrt{2n|\log a_n|/(a_n^{9/2}l)}}{\sigma_1\sqrt{l}} = \lim_{l\to\infty}\sqrt{\frac{n|\log a_n|}{a_n^{9/2}l^2}} = \lim_{n\to\infty}\sqrt{\frac{|\log a_n|}{a_n^{9/2}n}} = 0. \quad (3.3.21)$$

Using Kolmogorov's inequality, for every $\epsilon \in (0,1)$, there is a $C_{\epsilon/2} \in (0,1/2)$, such that

$$\begin{aligned}&\Pr\left\{\sum_{j=1}^{l}\frac{[g_i^{(n)}(X_j) - \mathbb{E}g_i^{(n)}(X)]}{\sqrt{Var(g_i^{(n)}(X))}} \geq \left(1-\frac{\epsilon}{2}\right)^2\sqrt{2n|\log a_n|}\right\}\\ >&C_{\epsilon/2}\exp\left(-\frac{n|\log a_n|}{l}\left(1-\frac{\epsilon}{2}\right)\left(1-\frac{\epsilon^2}{4}\right)\right)\\ =&C_{\epsilon/2}\exp\left(-\left(1-\frac{\epsilon}{2}\right)\left(1-\frac{\epsilon^2}{4}\right)|\log a_n|\right)\exp\left(-\left(1-\frac{\epsilon}{2}\right)\left(1-\frac{\epsilon^2}{4}\right)\left(\frac{n}{l}-1\right)|\log a_n|\right).\end{aligned} \quad (3.3.22)$$

Since $|\log a_n|/\sqrt{n} \to 0$ by hypothesis and $l \asymp n$, $\sqrt{n}|\log a_n|/l$ also converges to 0. This gives that

$$\exp\left(-\left(1-\frac{\epsilon}{2}\right)\left(1-\frac{\epsilon^2}{4}\right)\left(\frac{n}{l}-1\right)|\log a_n|\right) \geq \exp\left(-\left(1-\frac{\epsilon}{2}\right)\left(1-\frac{\epsilon^2}{4}\right)\frac{\eta\sqrt{n}}{l}|\log a_n|\right), \quad (3.3.23)$$

which is bounded below by 3/4 as $n\to\infty$. Thus we have for all $l : |n-l| \leq \eta\sqrt{n}$, when n is large enough,

$$\begin{aligned}&\Pr\left\{\sum_{j=1}^{l}\frac{[g_i^{(n)}(X_j) - \mathbb{E}g_i^{(n)}(X)]}{\sqrt{Var(g_i^{(n)}(X))}} \geq \left(1-\frac{\epsilon}{2}\right)^2\sqrt{2n|\log a_n|}\right\}\\ \geq&\frac{3}{4}C_{\epsilon/2}\exp\left(-\left(1-\frac{\epsilon}{2}\right)\left(1-\frac{\epsilon^2}{4}\right)|\log a_n|\right).\end{aligned} \quad (3.3.24)$$

On the other hand, by the central limit theorem and the fact that $\pi_n = \sum_{l=1}^{n} Y_l$, where

Y_l are i.i.d. *Poisson*(1),

$$\sum_{l:|n-l|\leq \eta\sqrt{n}} \Pr\{\pi_n = l\} = \Pr\left\{-\eta \leq \frac{\pi_n - n}{\sqrt{n}} \leq \eta\right\} \to \Pr\{|Z| \leq \eta\}, \tag{3.3.25}$$

where $Z \sim N(0,1)$. Collecting (3.3.17), (3.3.24) and (3.3.25), we get, for n large enough,

$$\sum_{l:|n-l|\leq \eta\sqrt{n}} \Pr\left\{(B_i^{(n)})^C|\pi_n = l\right\} \Pr\{\pi_n = l\} \geq \frac{3}{4}C_{\epsilon/2} \exp\left(-\left(1-\frac{\epsilon}{2}\right)\left(1-\frac{\epsilon^2}{4}\right)|\log a_n|\right)c(\eta), \tag{3.3.26}$$

where $c(\eta) := \Pr\{|Z| \leq \eta\}/2$. By (3.3.13),

$$\Pr\left\{(B_i^{(n)})^C\right\} \geq \frac{3}{4}C_{\epsilon/2} \exp(-K_\epsilon |\log a_n|)c(\eta), \tag{3.3.27}$$

where $K_\epsilon = (1-\epsilon/2)(1-\epsilon^2/4)$. Using hypothesis (3.3.4),

$$\Pr\left\{B_i^{(n)}\right\} \leq 1 - C_{\epsilon,\eta}\left(\frac{1}{a_n}\right)^{-K_\epsilon} \leq 1 - C_{\epsilon,\eta}\left(\frac{3k_n}{2r}\right)^{-K_\epsilon} \leq 1 - k_n^{-\rho}, \tag{3.3.28}$$

for any ρ strictly between K_ϵ and 1. It follows that

$$2\Pr\left\{B_1^{(n)}\right\} \cdots \Pr\left\{B_{k_n}^{(n)}\right\} \leq 2(1-k_n^{-\rho})^{k_n}. \tag{3.3.29}$$

Now we prove that this is the general term of a convergent series. The hypotheses $|\log a_n|/\log\log n \to \infty$, $a_n k_n \to r$ imply that $(1-\rho)\log k_n/\log\log n \to \infty$. Hence $k_n^{1-\rho}/\log n \to \infty$. Then since $k_n^\rho \to \infty$,

$$\frac{k_n \log(1-k_n^{-\rho})}{\log n} \asymp \frac{k_n^{1-\rho}\log(1-k_n^{-\rho})^{k_n^\rho}}{\log n} \asymp \frac{-k_n^{1-\rho}}{\log n} \to -\infty. \tag{3.3.30}$$

Therefore,

$$2\log n + k_n \log(1 - k_n^{-\rho}) \to -\infty. \tag{3.3.31}$$

It follows that $n^2(1-k_n^{-\rho})^{k_n} \to 0$. By the comparison test with $\sum_{n=1}^{\infty} 1/n^2$,

$$\sum_{n=1}^{\infty}(1-k_n^{-\rho})^{k_n} < \infty. \tag{3.3.32}$$

Then by (3.3.12) and Borel-Cantelli, we have proved the proposition. □

3.3.2 Theorem. Suppose the density f is bounded and continuous, and it is bounded away from 0 on $[0,1]$. $m_n \asymp n^\delta$ for some $\delta \in (0, 2/3)$. For any $0 < a \le b < 1$, we have, there exists a constant $C_{f,a,b} > 0$,

$$\liminf_{n\to\infty} \frac{\sqrt{n}}{m_n^{1/4}\sqrt{\log m_n}} \sup_{x\in[a,b]} |\hat{f}_{m,n}(x) - \mathbb{E}\hat{f}_{m,n}(x)| \ge C_{f,a,b}, \quad a.s. \tag{3.3.33}$$

Proof. By the definition of $\hat{f}_{m,n}(x)$ in (0.2.3), this is equivalent to proving that

$$\liminf_{n\to\infty} \sqrt{\frac{m_n^{3/2}}{n\log m_n}} \sup_{x\in[a,b]} \left| \sum_{i=1}^{n}\sum_{k=0}^{m_n-1}\left[I\left(\frac{k}{m_n} < X_i \le \frac{k+1}{m_n}\right) - \left(F\left(\frac{k+1}{m_n}\right) - F\left(\frac{k}{m_n}\right)\right)\right] b_{k,m_n-1}(x) \right|$$
$$\ge C_{f,a,b} \;\; a.s. \tag{3.3.34}$$

Fix $0 < \alpha < 1/2$, we first consider

$$\sum_{i=1}^{n} \sum_{k:|\frac{k}{m_n-1}-x|\le(m_n-1)^{-\alpha}} \left[I\left(\frac{k}{m_n} < X_i \le \frac{k+1}{m_n}\right) - \left(F\left(\frac{k+1}{m_n}\right) - F\left(\frac{k}{m_n}\right)\right)\right] b_{k,m_n-1}(x). \tag{3.3.35}$$

For z_{in} to be specified below, set

$$g_i^{(n)}(x) = g_i^{(n,\epsilon)}(x) := \sum_{k:|\frac{k}{m_n-1}-z_{in}|\le(m_n-1)^{-\alpha}} I\left(\frac{k}{m_n} < x \le \frac{k+1}{m_n}\right) b_{k,m_n-1}(z_{in}). \tag{3.3.36}$$

By this definition, $g_i^{(n)}(x)$ is nonzero, then

$$|x - z_{in}| \le \left|x - \frac{k}{m_n - 1}\right| + \left|\frac{k}{m_n - 1} - z_{in}\right| \le 2m_n^{-1} + 2m_n^{-\alpha}. \tag{3.3.37}$$

We now take $\alpha = 1/3$. Let X have the same distribution as $\{X_i\}_{i=1}^n$. Take an interval I such that $\Pr(X \in I) \le 1/2$. Since we assume that f is bounded and continuous, the interval exists. We may take $I = [c, d] \subset [a, b]$. Take $z_{in} = c+5im_n^{-1/3}$, $i = 1, 2, ..., k_n$, where $k_n := \lfloor \frac{(d-c)m_n^{1/3}}{5} \rfloor - 1$. Next we will check conditions (3.3.1)-(3.3.7) and apply Proposition 3.3.1 to $g_i^{(n)}(x)$. Set $a_n := m_n^{-1/3}$. (3.3.1) holds due to the assumption $0 < \delta < 2/3$. The construction implies that

$$\Pr\left\{g_i^{(n)}(X) \ne 0, g_k^{(n)}(X) \ne 0\right\} = 0, i \ne k, \tag{3.3.38}$$

and

$$\sum_{i=1}^{k_n} \Pr\left\{g_i^{(n)}(X) \ne 0\right\} \le \frac{1}{2}. \tag{3.3.39}$$

Set $N_i := \{k : |\frac{k}{m_n-1} - z_{in}| \le (m_n - 1)^{-1/3}\}$. Since f is bounded,

$$\begin{aligned}\left|\mathbb{E}\left(g_i^{(n)}(X)\right)\right| &= \left|\sum_{k\in N_i}\left(F\left(\frac{k+1}{m_n}\right) - F\left(\frac{k}{m_n}\right)\right) b_{k,m_n-1}(z_{in})\right| \\ &\le \sup_k \left(F\left(\frac{k+1}{m_n}\right) - F\left(\frac{k}{m_n}\right)\right)\sum_k b_{k,m_n-1}(z_{in}) \le \frac{\|f\|_\infty}{m_n}.\end{aligned} \tag{3.3.40}$$

Thus we have $|\mathbb{E}g_i^{(n)}(X)| \leq \|f\|_\infty a_n^3$. For m_n sufficiently large,

$$\begin{aligned}
\mathbb{E}\left(g_i^{(n)}(X)\right)^2 &= \mathbb{E}\left(\sum_{k\in N_i} I\left(\frac{k}{m_n} < X \leq \frac{k+1}{m_n}\right) b_{k,m_n-1}(z_{in})\right)^2 \\
&= \mathbb{E}\left(\sum_{k\in N_i} I\left(\frac{k}{m_n} < X \leq \frac{k+1}{m_n}\right) b_{k,m_n-1}^2(z_{in})\right)^2 \\
&= \sum_{k\in N_i}\left(F\left(\frac{k+1}{m_n}\right) - F\left(\frac{k}{m_n}\right)\right) b_{k,m_n-1}^2(z_{in}) \\
&\leq \frac{\|f\|_\infty}{m_n}\sum_{k\in N_i} b_{k,m_n-1}^2(z_{in}) \\
&\leq \frac{C_{f,a,b}}{m_n^{3/2}},
\end{aligned} \tag{3.3.41}$$

where we used Lemma 3.2.3 in the last inequality. And using the hypothesis that f is bounded away from 0 on $[0,1]$, we obtain

$$\begin{aligned}
\mathbb{E}\left(g_i^{(n)}(X)\right)^2 &\geq \inf_{k\in N_i}\left(F\left(\frac{k+1}{m_n}\right) - F\left(\frac{k}{m_n}\right)\right)\sum_{k\in N_i} b_{k,m_n-1}^2(z_{in}) \\
&\geq \frac{1}{m_n}\inf_{x\in[0,1]} f(x)\sum_{k\in N_i} b_{k,m_n-1}^2(z_{in}) \\
&\geq \frac{C_{f,a,b}}{m_n^{3/2}}.
\end{aligned} \tag{3.3.42}$$

This is due to Lemma 3.2.3 and inequality (8), pg 15, Lorentz (1986). So we conclude that

$$\begin{aligned}
\mathrm{V}ar\left(g_i^{(n)}(X)\right) &\geq \mathbb{E}\left(g_i^{(n)}(X)\right)^2 - \left(\mathbb{E}g_i^{(n)}(X)\right)^2 \\
&\geq \frac{C_{f,a,b}}{m_n^{3/2}} - \frac{\|f\|_\infty^2}{m_n^2} \\
&\geq \frac{C_{f,a,b}}{m_n^{3/2}}.
\end{aligned} \tag{3.3.43}$$

When n is large enough, we have

$$a_n^3\mu_1 \le \mathbb{E}g_i^{(n)} \le a_n^3\mu_2, \tag{3.3.44}$$

and

$$\sigma_1 a_n^{9/4} \le \sqrt{\mathrm{V}ar g_i^{(n)}} \le \sigma_2 a_n^{9/4}. \tag{3.3.45}$$

Therefore we have checked that a_n and the sequence of functions $g_i^{(n)}$ satisfy all the assumptions of Proposition 3.3.1. Set $N_x := \{k : |\frac{k}{m_n-1} - x| > (m_n-1)^{-1/3}\}$. An application of the proposition gives that, with probability 1,

$$\begin{aligned}
&\sup_{x\in[a,b]} \frac{m_n^{3/4}}{\sqrt{n\log m}} \sum_{i=1}^{n}\sum_{k\in N_x}\left[I\left(\frac{k}{m} < X_i \le \frac{k+1}{m}\right) - \left(F\left(\frac{k+1}{m}\right) - F\left(\frac{k}{m_n}\right)\right)\right] b_{k,m_n-1}(x) \\
&\ge \max_{1\le i\le k_n} \frac{\sqrt{n}\{G_n(g_i^{(n)}) - \mathbb{E}g_i^{(n)}\}}{\sqrt{\mathrm{V}ar(g_i^{(n)})\left|\log a_n\right|}} \ge C_{f,a,b}.
\end{aligned} \tag{3.3.46}$$

On the other hand, by (8), Pg 15, Lorentz (1986), there exists a constant C that only depends on l, such that for any $l > 0$,

$$\begin{aligned}
&\frac{m_n^{3/4}}{\sqrt{n\log m_n}} \sum_{i=1}^{n}\sum_{k\notin N_x}\left\|\left[I\left(\frac{k}{m_n} < X_i \le \frac{k+1}{m_n}\right) - \left(F\left(\frac{k+1}{m_n}\right) - F\left(\frac{k}{m_n}\right)\right)\right]\right\| b_{k,m_n-1}(x) \\
\le & 2\frac{m_n^{3/4}}{\sqrt{n\log m_n}} \sum_{i=1}^{n}\sum_{k\notin N_x} b_{k,m_n-1}(x) \\
\le & 2C\frac{m_n^{3/4}}{\sqrt{n\log m_n}} nm^{-l}.
\end{aligned} \tag{3.3.47}$$

This converges to 0 if we take $l > 3/4 + 1/(2\delta)$. (3.3.46) and (3.3.47) give (3.3.34). □

Appendix A

Variance Computations

We present here some inequalities and variance computations used throughout the paper. If the kernel $K(x, y)$ satisfies (0.1.7), we have the following estimates for all x and y, and all measurable sets F in $\mathbb{R}$.

$$\int_F \bar{K}_n^2(t,x)dt = \int_F [K_n(t,x) - \mathbb{E}K_n(t,X)]^2 dt \leq 2\Big(\int_F K_n^2(t,x)dt + \int_F (\mathbb{E}K_n(t,X))^2 dt\Big). \quad (A.0.1)$$

$$\int_F K_n^2(t,x)dt = \int_F K^2(2^{j_n}t, 2^{j_n}x)dt \leq \int_{\mathbb{R}} \Phi^2(2^{j_n}t - 2^{j_n}x)dt \leq 2^{-j_n}\|\Phi\|_2^2. \quad (A.0.2)$$

By Hölder's inequality and Fubini's theorem,

$$\begin{aligned}
\int_F (\mathbb{E}K_n(t,X))^2\, dt &= \int_F \left(\int_{\mathbb{R}} K(2^{j_n}t, 2^{j_n}x) f(x)dx\right)^2 dt \\
&\leq \int_F \int_{\mathbb{R}} K^2(2^{j_n}t, 2^{j_n}x) f(x)dx \int_{\mathbb{R}} f(y)dydt \\
&= \int_{\mathbb{R}} \left(\int_F K^2(2^{j_n}t, 2^{j_n}x)dt\right) f(x)dx \\
&\leq 2^{-j_n}\|\Phi\|_2^2.
\end{aligned} \quad (A.0.3)$$

Gathering the three inequalities above, we have for all x and y, and all measurable sets F,

$$\int_F \bar{K}_n^2(t,x)dt \leq 4 \cdot 2^{-j_n}\|\Phi\|_2^2, \quad (A.0.4)$$

$$\left| \int_F \bar{K}_n^2(t,x)dt - \mathbb{E} \int_F \bar{K}_n^2(t,X)dt \right| \leq 8 \cdot 2^{-j_n} \|\Phi\|_2^2, \tag{A.0.5}$$

and

$$\int_F \left| \bar{K}_n(t,x)\bar{K}_n(t,y) \right| dt \leq \left(\int_F \bar{K}_n^2(t,x)dt \right)^{1/2} \left(\int_F \bar{K}_n^2(t,y)dt \right)^{1/2} \leq 4 \cdot 2^{-j_n} \|\Phi\|_2^2. \tag{A.0.6}$$

A.0.1 Corollary. Assume (f), (S1) and (B1), and that F satisfies condition (1.1.10). Then there exists $n_0 = n_0(F)$ such that, for all $n \geq n_0$,

$$\mathrm{V}ar \int_F \bar{K}_n^2(t,X)dt \leq 8 \cdot 2^{-2j_n} \|\Phi\|_2^4 \int_F f(x)dx. \tag{A.0.7}$$

And for all n,

$$\mathrm{V}ar \int_F \bar{K}_n^2(t,X)dt \leq 4 \cdot 2^{-2j_n} \|\Phi\|_2^4. \tag{A.0.8}$$

Proof. Since by (A.0.3), $\int_F (\mathbb{E}K_n(t,X))^2\, dt < \infty$, we have

$$\begin{aligned}
\mathrm{V}ar \int_F \bar{K}_n^2(t,X)dt &= \mathrm{V}ar\left(\int_F (K_n(t,X) - \mathbb{E}K_n(t,X))^2 dt \right) \\
&\leq 4\mathrm{V}ar\left(\int_F \left[K_n^2(t,X) + (\mathbb{E}K_n(t,X))^2 \right] dt \right) \\
&= 4\mathrm{V}ar\left(\int_F K_n^2(t,X)dt \right) \\
&\leq 4\mathbb{E}\left(\int_F K_n^2(t,X)dt \right)^2 \\
&= 4\mathbb{E} \int_F K_n^2(t,X)dt \int_F K_n^2(s,X)ds \\
&= 4 \int_F \int_F \mathbb{E}\left[K_n^2(t,X)K_n^2(s,X) \right] dsdt.
\end{aligned} \tag{A.0.9}$$

Next we observe that by (0.1.7),

$$\begin{aligned}\int_F \int_F \mathbb{E}[K_n^2(t,X)K_n^2(s,X)]dsdt &\leq \int_F \int_F \int_{\mathbb{R}} \Phi_n^2(t-x)\Phi_n^2(s-x)f(x)dxdsdt \\ &= \int_{\mathbb{R}^3} \Phi^2(2^{j_n}(t-x))\Phi^2(2^{j_n}(s-x))f(x)1(s\in F) \\ &\quad 1(t\in F)dxdsdt \\ &= 2^{-2j_n}\int_{\mathbb{R}^3} \Phi^2(w)\Phi^2(z)f(x)1_n(z,w,x)dxdwdz,\end{aligned} \tag{A.0.10}$$

with change of variables $w = 2^{j_n}(t-x), z = 2^{j_n}(s-x)$ in the last step, where

$$1_n(z,w,x) = 1(s\in F)1(t\in F) = 1(2^{-j_n}z + x \in F)1(2^{-j_n}w + x\in F). \tag{A.0.11}$$

From condition (1.1.10) and boundedness of f,

$$\limsup_{n\to\infty} \int_{\mathbb{R}} 1_n(z,w,x)f(x)dx \leq \int_F f(x)dx \tag{A.0.12}$$

for all $z,\ w$. Moreover, for all n,

$$\Phi^2(w)\Phi^2(z)\left[\int_{\mathbb{R}} f(x)1_n(z,w,x)dx\right] \leq \Phi^2(w)\Phi^2(z). \tag{A.0.13}$$

which is integrable. So by Fatou's Lemma,

$$\limsup_{n\to\infty} \int_{\mathbb{R}^2} \Phi^2(w)\Phi^2(z)\left[\int_{\mathbb{R}} f(x)1_n(z,w,x)dx\right]dwdz \leq \|\Phi\|_2^4 \int_F f(x)dx. \tag{A.0.14}$$

This shows that for n large enough, the left hand side of (A.0.10) is less than or equal to $2\cdot 2^{-2j_n}\|\Phi\|_2^4 \int_F f(x)dx$, assuming $\int_F f(x)dx > 0$ (see (1.1.10)). So there exists $n_0 = n_0(F)$

such that, for all $n \geq n_0$,

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

(A.0.8) follows from the facts that for all n, z, w,

$$\int_{\mathbb{R}} 1_n(z, w, x) f(x)dx \leq \int_{\mathbb{R}} f(x)dx = 1 \tag{A.0.15}$$

and

$$\int_{\mathbb{R}^2} \Phi^2(w)\Phi^2(z) \left[\int_{\mathbb{R}} f(x) 1_n(z, w, x)dx \right] dwdz \leq \|\Phi\|_2^4. \tag{A.0.16}$$

□

Set $C_n(t, s) := 2^{j_n} \int_{\mathbb{R}} K_n(t, x)K_n(s, x)f(x)dx$.

A.0.2 Lemma. Under the hypotheses of Corollary A.0.1,

$$\begin{aligned} \limsup_{n\to\infty} 2^{j_n} \int_{F^2} C_n^2(s, t)dsdt &\leq \int_F f^2(x)dx \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \Phi(w + u)\Phi(w)dw \right)^2 du \\ &\leq \int_F f^2(x)dx \|\Phi\|_1^2 \|\Phi\|_2^2. \end{aligned} \tag{A.0.17}$$

Proof. Again by (0.1.7),

$$\begin{aligned} 2^{j_n} \int_{F^2} C_n^2(s, t)dsdt &= 2^{3j_n} \int_{F^2} \left(\int_{\mathbb{R}} K_n(t, x)K_n(s, x)f(x)dx \right)^2 dsdt \\ &= 2^{3j_n} \int_{F^2} \Big\{ \int_{\mathbb{R}^2} K(2^{j_n}t, 2^{j_n}x)K(2^{j_n}s, 2^{j_n}x) \\ &\quad K(2^{j_n}t, 2^{j_n}y)K(2^{j_n}s, 2^{j_n}y)f(x)f(y)dxdy \Big\} dsdt \\ &= 2^{2j_n} \int_{F^2} \int_{\mathbb{R}^2} K(2^{j_n}t, 2^{j_n}x)K(2^{j_n}s, 2^{j_n}x) \end{aligned}$$

$$K(2^{j_n}t, 2^{j_n}x-u)K(2^{j_n}s, 2^{j_n}x-u)f(x)f(x-2^{-j_n}u)dxdudsdt$$

$$\leq 2^{2j_n}\int_{F^2}\int_{\mathbb{R}^2}\Phi(2^{j_n}(t-x))\Phi(2^{j_n}(s-x))\Phi(2^{j_n}(t-x)+u)$$

$$\Phi(2^{j_n}(s-x)+u)f(x)f(x-2^{-j_n}u)dxdudsdt. \tag{A.0.18}$$

By change of variables $w=2^{j_n}(t-x), z=2^{j_n}(s-x)$, the last term is bounded by

$$\int_{\mathbb{R}^2}\left\{\int_{\mathbb{R}}\left[\int_{\mathbb{R}}\Phi(z)\Phi(w)1_n(z,w,x)\Phi(z+u)\Phi(w+u)f(x-2^{-j_n}u)du\right]f(x)dx\right\}dzdw \tag{A.0.19}$$

where $1_n(z,w,x)$ is defined in (A.0.11). Now set

$$G_n(z,w):=\int_{\mathbb{R}}\left[\int_{\mathbb{R}}\Phi(z)\Phi(w)1_n(z,w,x)\Phi(z+u)\Phi(w+u)f(x-2^{-j_n}u)du\right]f(x)dx, \tag{A.0.20}$$

$$\Phi(z,w):=\Phi(z)\Phi(w)\int_{\mathbb{R}}\Phi(z+u)\Phi(w+u)du, \tag{A.0.21}$$

and

$$G(z,w):=\Phi(z,w)\int_F f^2(x)dx. \tag{A.0.22}$$

Then by the proof of (2.10) and (2.11), Giné and Mason (2004), we have

$$G_n(z,w)\to G(z,w) \text{ as } n\to\infty \tag{A.0.23}$$

and

$$G_n(z,w)\leq|\Phi(z)\Phi(w)|\|f\|_2^2\|\Phi\|_2^2. \tag{A.0.24}$$

So, by (A.0.18) and the Lebesgue dominated convergence theorem,

$$\begin{aligned}\limsup_{n\to\infty} 2^{j_n} \int_{F^2} C_n^2(s,t)dsdt &\le \limsup_{n\to\infty} \int_{\mathbb{R}^2} G_n(z,w)dzdw \\ &= \int_{\mathbb{R}^2} G(z,w)dzdw \\ &= \int_F f^2(x)dx \int_{\mathbb{R}^2} \Phi(z)\Phi(w) \int_{\mathbb{R}} \Phi(z+u)\Phi(w+u)dudzdw \\ &= \int_F f^2(x)dx \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \Phi(w+u)\Phi(w)dw\right)^2 du.\end{aligned} \tag{A.0.25}$$

Since by Hölder and Fubini's theorem,

$$\begin{aligned}&\int_{\mathbb{R}} \left(\int_{\mathbb{R}} \Phi(w+u)\Phi(w)dw\right)^2 du \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \Phi(w+u)\Phi(w)dw \int_{\mathbb{R}} \Phi(s+u)\Phi(s)dsdu \\ &\le \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \Phi^2(w+u)du\right)^{1/2} \left(\int_{\mathbb{R}} \Phi^2(s+u)du\right)^{1/2} \Phi(w)\Phi(s)dwds \\ &= \|\Phi\|_1^2 \|\Phi\|_2^2,\end{aligned} \tag{A.0.26}$$

(A.0.17) follows from this and (A.0.25). □

A.0.3 Lemma. Under the hypotheses of Corollary $A.0.1$, for $R_n(s,t)$ defined in (1.2.18),

$$\lim_{n\to\infty} 2^{j_n} \int_{F^2} (C_n(s,t) - R_n(s,t))^2 dsdt = 0. \tag{A.0.27}$$

Proof. We observe that,

$$\begin{aligned}(C_n(s,t) - R_n(s,t))^2 &= 2^{2j_n}\Big[EK_n(t,X)K_n(s,X) - E(K_n(t,X) \\ &\quad - EK_n(t,X))(K_n(s,X) - EK_n(s,X))\Big]^2 \\ &= 2^{2j_n}(EK_n(t,X))^2(EK_n(s,X))^2 \\ &= 2^{-2j_n}\mu_n^2(s)\mu_n^2(t),\end{aligned} \tag{A.0.28}$$

where

$$\mu_n(s) := 2^{j_n} EK_n(s, X) = 2^{j_n} \int K_n(s, x) f(x) dx. \tag{A.0.29}$$

Now,

$$\begin{aligned} \mu_n^2(s) &\leq 2^{2j_n} \left(\int \Phi_n(s-x) f(x) dx \right)^2 \\ &\leq 2^{2j_n} \int \Phi(2^{j_n}(s-x)) f^2(x) dx \int \Phi(2^{j_n}(s-x)) dx \\ &= 2^{j_n} \|\Phi\|_1 \int_{\mathbb{R}} \Phi(2^{j_n}(s-x)) f^2(x) dx, \end{aligned} \tag{A.0.30}$$

so that

$$\int_{\mathbb{R}} \mu_n^2(s) ds \leq 2^{j_n} \|\Phi\|_1 \int_{\mathbb{R}} \Phi(2^{j_n}(s-x)) ds \int_{\mathbb{R}} f^2(x) dx \leq \|\Phi\|_1^2 \|f\|_2^2. \tag{A.0.31}$$

Therefore,

$$\int_{\mathbb{R}^2} (C_n(s,t) - R_n(s,t))^2 ds dt = 2^{-2j_n} \int_{\mathbb{R}} \mu_n^2(s) ds \int_{\mathbb{R}} \mu_n^2(t) dt \leq 2^{-2j_n} \|\Phi\|_1^4 \|f\|_2^4. \tag{A.0.32}$$

Thus

$$2^{j_n} \int_{F^2} (C_n(s,t) - R_n(s,t))^2 ds dt \leq 2^{-j_n} \|\Phi\|_1^4 \|f\|_2^4 \to 0. \tag{A.0.33}$$

□

A.0.4 Lemma. Assume (f) and (B1) hold, and the scaling function ϕ satisfies (S1) such that the kernel K associated with ϕ is dominated by Φ whose support is contained in

$[-A, A]$, where $A > 0$ is an integer. Then for any $M > 0$,

$$\lim_{n\to\infty} 2^{j_n} \int_{[-M,M]^2} C_n^2(s,t)dsdt = \int_{-M}^{M} f^2(y)dy. \tag{A.0.34}$$

Proof. (A.0.18) with $F = [-M, M]$ gives,

$$\begin{aligned}
&2^{j_n} \int_{[-M,M]^2} C_n^2(s,t)dsdt \\
&= 2^{2j_n} \int_{[-M,M]^2} \int_{\mathbb{R}^2} K(2^{j_n}t, 2^{j_n}x)K(2^{j_n}s, 2^{j_n}x)K(2^{j_n}t, 2^{j_n}x - u)K(2^{j_n}s, 2^{j_n}x - u)f(x) \\
&f(x - 2^{-j_n}u)dxdudsdt.
\end{aligned} \tag{A.0.35}$$

By change of variables $t = 2^{-j_n}w + x, s = 2^{-j_n}z + x$, this integral is equal to

$$\begin{aligned}
&\int_{\mathbb{R}^2} \int_{\mathbb{R}^2} K(2^{j_n}x + z, 2^{j_n}x)K(2^{j_n}x + w, 2^{j_n}x)K(2^{j_n}x + z, 2^{j_n}x - u)K(2^{j_n}x + w, 2^{j_n}x - u) \\
&\quad \times f(x)f(x - 2^{-j_n}u)1\left(2^{-j_n}z + x \in [-M, M]\right)1\left(2^{-j_n}w + x \in [-M, M]\right)dxdudzdw \\
&= \int_{-A}^{A} \int_{-A}^{A} \left\{ \int_{\mathbb{R}^2} K(2^{j_n}x + z, 2^{j_n}x)K(2^{j_n}x + w, 2^{j_n}x)K(2^{j_n}x + z, 2^{j_n}x - u) \right. \\
&\quad \times K(2^{j_n}x + w, 2^{j_n}x - u)f(x)f(x - 2^{-j_n}u)1\left(2^{-j_n}z + x \in [-M, M]\right) \\
&\quad \left. \times 1\left(2^{-j_n}w + x \in [-M, M]\right)dxdu \right\} dzdw \\
&= \int_{-A}^{A} \int_{-A}^{A} \int_{\mathbb{R}} \sum_{i=-\infty}^{\infty} \int_{2^{-j_n}i}^{2^{-j_n}(i+1)} K(2^{j_n}x + z, 2^{j_n}x)K(2^{j_n}x + w, 2^{j_n}x)K(2^{j_n}x + z, 2^{j_n}x - u) \\
&\quad \times K(2^{j_n}x + w, 2^{j_n}x - u)f(x)f(x - 2^{-j_n}u)1\left(2^{-j_n}z + x \in [-M, M]\right) \\
&\quad \times 1\left(2^{-j_n}w + x \in [-M, M]\right)dxdudzdw \\
&= \int_{-A}^{A} \int_{-A}^{A} \int_{\mathbb{R}} \sum_{i=-\infty}^{\infty} \int_{0}^{2^{-j_n}} K(2^{j_n}x + z + i, 2^{j_n}x + i)K(2^{j_n}x + w + i, 2^{j_n}x + i) \\
&\quad \times K(2^{j_n}x + z + i, 2^{j_n}x - u + i)K(2^{j_n}x + w + i, 2^{j_n}x - u + i)f(x + 2^{-j_n}i)
\end{aligned}$$

$$\times f(x + 2^{-j_n} i - 2^{-j_n} u) 1\left(2^{-j_n} z + x + 2^{-j_n} i \in [-M, M]\right)$$

$$\times 1\left(2^{-j_n} w + x + 2^{-j_n} i \in [-M, M]\right) dxdudzdw. \tag{A.0.36}$$

By periodicity of the kernel K and then change of variables, it is in turn equal to

$$\int_{-A}^{A} \int_{-A}^{A} \int_{-2A}^{2A} \sum_{i=-\infty}^{\infty} \int_{0}^{2^{-j_n}} K(2^{j_n} x + z, 2^{j_n} x) K(2^{j_n} x + w, 2^{j_n} x) K(2^{j_n} x + z, 2^{j_n} x - u)$$

$$\times K(2^{j_n} x + w, 2^{j_n} x - u) f(x + 2^{-j_n} i) f(x + 2^{-j_n} i - 2^{-j_n} u) 1(2^{-j_n} z + x + 2^{-j_n} i \in [-M, M])$$

$$\times 1(2^{-j_n} w + x + 2^{-j_n} i \in [-M, M]) dxdudzdw$$

$$= \int_{-A}^{A} \int_{-A}^{A} \int_{-2A}^{2A} \sum_{i=-\infty}^{\infty} \int_{0}^{1} 2^{-j_n} K(x + z, x) K(x + w, x) K(x + z, x - u) K(x + w, x - u)$$

$$\times f(2^{-j_n}(x + i)) f(2^{-j_n}(x + i - u)) 1(2^{-j_n}(z + x + i) \in [-M, M])$$

$$\times 1(2^{-j_n}(w + x + i) \in [-M, M]) dxdudzdw. \tag{A.0.37}$$

To continue, it is convenient to write

$$\sum_{i=-\infty}^{\infty} 2^{-j_n} f(2^{-j_n}(x + i)) f(2^{-j_n}(x + i - u)) 1(2^{-j_n}(z + x + i) \in [-M, M])$$

$$\times 1(2^{-j_n}(w + x + i) \in [-M, M])$$

$$= \left\{ \sum_{i=2A}^{\infty} + \sum_{i=-2A}^{2A-1} + \sum_{-\infty}^{-2A-1} \right\} 2^{-j_n} f(2^{-j_n}(x + i)) f(2^{-j_n}(x + i - u)) \tag{A.0.38}$$

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

$$=: I_1(j_n) + I_2(j_n) + I_3(j_n) = I(j_n).$$

The next lemma proves the convergence of $I(j_n)$.

A.0.5 Lemma. Under the condition (f), for any $M > 0$,

$$I(j_n) \to \int_{-M}^{M} f^2(y)dy \tag{A.0.39}$$

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 $2^{-j_n}(-A + i) > M$. With the notation introduced above,

$$\begin{aligned} I_1(j_n) &= \sum_{i=2A}^{\lfloor 2^{j_n}M\rfloor + A} 2^{-j_n} f(2^{-j_n}(x+i)) f(2^{-j_n}(u'+i)) 1(2^{-j_n}(z'+i) \in [-M, M]) \\ &\quad 1(2^{-j_n}(w'+i) \in [-M, M]) \qquad \text{(A.0.40)} \\ &= \left(\sum_{i=2A}^{\lfloor 2^{j_n}M\rfloor - 2A - 1} + \sum_{\lfloor 2^{j_n}M\rfloor - 2A}^{\lfloor 2^{j_n}M\rfloor + A} \right) 2^{-j_n} f(2^{-j_n}(x+i)) f(2^{-j_n}(u'+i)) \\ &\quad 1\left(2^{-j_n}(z'+i) \in [0, M]\right) 1\left(2^{-j_n}(w'+i) \in [0, M]\right) \\ &=: I_4(j_n) + I_5(j_n), \end{aligned}$$

where $\lfloor 2^{j_n}M\rfloor$ 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 $x \in [0,1], u \in [-2A, 2A], z \in [-A, A], w \in [-A, A]$.

We can simplify $I_4(j_n)$ since the indicator function in the general summand of $I_4(j_n)$ must be 1. Set $\triangle y := 2^{-j_n}(4A + 1)$. Letting i jump $4A + 1$ places in each sum, we group $I_4(j_n)$ into $4A + 1$ sums.

$$
\begin{aligned}
I_4(j_n) &= \sum_{i=2A}^{\lfloor 2^{j_n}M\rfloor-2A-1} 2^{-j_n} f(2^{-j_n}(x+i)) f(2^{-j_n}(u'+i)) \\
&= \frac{1}{4A+1}\sum_{i=2A}^{6A}\sum_{j=0}^{N_i} \triangle y f(2^{-j_n}(x+i)+j\triangle y) f(2^{-j_n}(u'+i)+j\triangle y),
\end{aligned}
\tag{A.0.41}
$$

where N_i is the largest j such that for fixed i, $i+j(4A+1) \leq \lfloor 2^{j_n}M\rfloor - 2A-1$. $N_i = \lfloor M/\triangle y - 1\rfloor$ 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 N_i+3 subintervals. Except for the first and the last subintervals, whose lengths we denote respectively by $\triangle y_{i,1}$ and $\triangle y_{N_i+3}$, 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

$$
S_{i,n} =: f^2(0)\triangle y_{i,1} + \sum_{j=0}^{N_i} \triangle y f(2^{-j_n}(x+i)+j\triangle y) f(2^{-j_n}(u'+i)+j\triangle y) + f^2(M)\triangle y_{i,N_i+3}, \tag{A.0.42}
$$

we see that

$$
S_{i,n} \leq f^2(0)\triangle y_{i,1} + \sum_{j=0}^{N_i} M_{i,j}^2 \triangle y + f^2(M)\triangle y_{i,N_i+3}, \tag{A.0.43}
$$

and

$$
S_{i,n} \geq f^2(0)\triangle y_{i,1} + \sum_{j=0}^{N_i} m_{i,j}^2 \triangle y + f^2(M)\triangle y_{i,N_i+3}, \tag{A.0.44}
$$

where $M_{i,j}$ and $m_{i,j}$ denote respectively the supremum and the infimum of f on the partition $[2^{-j_n}(i-2A)+j\triangle y, 2^{-j_n}(i-2A)+(j+1)\triangle y]$.

As $n \to \infty$, the mesh of $P_{i,n}$ tends to zero. Hence, since $f^2(0)\triangle y_{i,1} + f^2(M)\triangle y_{i,N_i+3} \to 0$ and f^2 is improper Riemann integrable on $[0, M]$, it follows that $S_{i,n} \to \int_0^M f^2(y)dy$ and by (A.0.41),

$$I_4(j_n) \to \int_0^M f^2(y)dy. \tag{A.0.45}$$

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 $\lim_{n\to\infty} I_1(j_n) = \lim_{n\to\infty} (I_4(j_n) + I_5(j_n)) = \int_0^M f^2(y)dy$ 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,

$$I_2(j_n) \leq 4A \cdot 2^{-j_n} \sup_{y\in\mathbb{R}} f^2(y) \to 0 \text{ as } n \to \infty. \tag{A.0.46}$$

(A.0.39) is proved when collecting the results for $I_1(j_n)$, $I_2(j_n)$ and $I_3(j_n)$. □

A.0.6 Lemma. Assume the scaling function ϕ satisfies (S1) such that the kernel K associated with ϕ is dominated by Φ whose support is contained in $[-A, A]$, where $A > 0$ is an integer. Then

$$\int_{-A}^{A}\int_{-A}^{A}\int_{-2A}^{2A}\int_0^1 K(x+z, x)K(x+w, x)K(x+z, x-u)K(x+w, x-u)dxdudzdw = 1. \tag{A.0.47}$$

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,

$$\int_{-A}^{A}\int_{-A}^{A}\int_{-2A}^{2A}\int_{0}^{1} K(x+z,x)K(x+w,x)K(x+z,x-u)K(x+w,x-u)dxdudzdw$$
$$= \int_{0}^{1}\int_{\mathbb{R}}\int_{\mathbb{R}} K(x+z,x)K(x+z,x-u)dz \int_{\mathbb{R}} K(x+w,x)K(x+w,x-u)dwdudx. \quad \text{(A.0.48)}$$

We make the following observation:

$$\begin{aligned}\int K(x,y)K(x,z)dx &= \int \left(\sum_{k} \phi(x-k)\phi(y-k)\right)\left(\sum_{l} \phi(x-l)\phi(z-l)\right) \\ &= \int \sum_{k} \phi^2(x-k)\phi(y-k)\phi(z-k)dx \qquad \text{(A.0.49)} \\ &\quad + \int \sum_{k\neq l} \phi(x-k)\phi(y-k)\phi(x-l)\phi(z-l)dx.\end{aligned}$$

We note that by the proof of Lemma 8.6 in HKPT(1998), there exists Φ symmetric, bounded, nonnegative and compactly supported, so that $\sum_{k\in\mathbb{Z}} |\phi(x-k)\phi(y-k)| \leq \Phi(x-y)$. For some constant C,

$$\int_{\mathbb{R}} \sum_{k\in\mathbb{Z}} |\phi^2(x-k)\phi(y-k)\phi(z-k)|dx \leq C\int_{\mathbb{R}} \Phi(x-y)dx < \infty. \qquad \text{(A.0.50)}$$

So we can apply Fubini to the first integral in (A.0.49),

$$\begin{aligned}&\int_{\mathbb{R}} \sum_{k\in\mathbb{Z}} \phi^2(x-k)\phi(y-k)\phi(z-k)dx \\ &= \sum_{k\in\mathbb{Z}} \phi(y-k)\phi(z-k)\int_{\mathbb{R}} \phi^2(x-k)dx \qquad \text{(A.0.51)} \\ &= \sum_{k\in\mathbb{Z}} \phi(y-k)\phi(z-k).\end{aligned}$$

For the second integral in (A.0.49),

$$\int \sum \sum\nolimits_{k\neq l} \phi(x-k)\phi(y-k)\phi(x-l)\phi(z-l)dx \leq \int \Phi(x-y)\Phi(x-z)dx \leq \|\Phi\|_2^2. \quad \text{(A.0.52)}$$

We can also apply Fubini's theorem and orthogonality of ϕ,

$$\begin{aligned} &\int \sum \sum\nolimits_{k\neq l} \phi(x-k)\phi(y-k)\phi(x-l)\phi(z-l)dx \\ =& \sum \sum\nolimits_{k\neq l} \phi(y-k)\phi(z-l) \int_{\mathbb{R}} \phi(x-k)\phi(x-l)dx = 0. \end{aligned} \quad \text{(A.0.53)}$$

By (A.0.51) and (A.0.53),

$$\int_{\mathbb{R}} K(x,y)K(x,z)dx = \sum_{k\in\mathbb{Z}} \phi(y-k)\phi(z-k) = K(y,z). \quad \text{(A.0.54)}$$

For fixed x and u, by a change of variable $z = y - x$ and application of the above equation,

$$\int_{\mathbb{R}} K(x+z,x)K(x+z,x-u)dz = \int_{\mathbb{R}} K(y,x)K(y,x-u)dy = \sum_{k\in\mathbb{Z}} \phi(x-k)\phi(x-u-k). \quad \text{(A.0.55)}$$

Similarly,

$$\int_{\mathbb{R}} K(x+w,x)K(x+w,x-u)dw = \sum_{m\in\mathbb{Z}} \phi(x-m)\phi(x-u-m). \quad \text{(A.0.56)}$$

Now we have that, for fixed $x \in [0,1]$, by (A.0.55) and (A.0.56),

$$\begin{aligned} &\int_{\mathbb{R}} \int_{\mathbb{R}} K(x+z,x)K(x+z,x-u)dz \int_{\mathbb{R}} K(x+w,x)K(x+w,x-u)dwdu \\ =& \int_{\mathbb{R}} \left(\sum_{k\in\mathbb{Z}} \phi(x-k)\phi(x-u-k) \right) \left(\sum_{m\in\mathbb{Z}} \phi(x-m)\phi(x-u-m) \right) du \\ =& \int_{\mathbb{R}} \sum_{k\in\mathbb{Z}} \phi^2(x-k)\phi^2(x-u-k)du \end{aligned}$$

$$+\int_{\mathbb{R}}\sum_{k\neq m}\phi(x-k)\phi(x-u-k)\phi(x-m)\phi(x-u-m)du. \tag{A.0.57}$$

Again by Fubini's theorem and orthogonality, the integral is equal to

$$\sum_{k\in\mathbb{Z}}\phi^2(x-k)\int_{\mathbb{R}}\phi^2(x-u-k)du+\sum_{k\neq m}\phi(x-k)\phi(x-m)\int_{\mathbb{R}}\phi(x-u-k)\phi(x-u-m)du$$
$$=\sum_{k\in\mathbb{Z}}\phi^2(x-k). \tag{A.0.58}$$

Finally we consider

$$\begin{aligned}&\int_0^1\sum_{k\in\mathbb{Z}}\phi^2(x-k)dx\\ =&\sum_{k\in\mathbb{Z}}\int_0^1\phi^2(x-k)dx=\sum_{k\in\mathbb{Z}}\int_{-k}^{-k+1}\phi^2(x)dx=\int_{\mathbb{R}}\phi^2(x)dx=1.\end{aligned} \tag{A.0.59}$$

□

We now continue with the proof of Lemma $A.0.4$. Since the convergence is uniform for $x\in[0,1], u\in[-2A,2A], z\in[-A,A], w\in[-A,A]$,

$$|I(j_n)|\leq 2\int_{-M}^{M}f^2(t)dt \tag{A.0.60}$$

if n is sufficiently large. The quantity in (A.0.37) is bounded in absolute value by

$$\begin{aligned}&||\Phi||_\infty^4\int_{-A}^{A}\int_{-A}^{A}\int_{-2A}^{2A}\int_0^1 I(j_n)dxdudzdw\\ \leq\ &2||\Phi||_\infty^4\int_{-A}^{A}\int_{-A}^{A}\int_{-2A}^{2A}\int_0^1\int_{-M}^{M}f^2(t)dtdxdudzdw<\infty\end{aligned} \tag{A.0.61}$$

for n large. So, by Fubini, (A.0.37) is equal to

$$\int_{-A}^{A}\int_{-A}^{A}\int_{-2A}^{2A}\int_0^1 K(x+z,x)K(x+w,x)K(x+z,x-u)K(x+w,x-u)I(j_n)dxdudzdw, \tag{A.0.62}$$

which by dominated convergence theorem and Lemmas $A.0.5$, $A.0.6$ converges to $\int_{-M}^{M} f^2(y)dy$. □

A.0.7 Lemma. Under the hypotheses of Lemma $A.0.4$, and assume that, in addition, f is Hölder continuous with exponent $0 < \alpha \leq 1$ on $[-L, L]$, monotonically increasing on $(-\infty, -L]$ and monotonically decreasing on $[L, \infty)$, where $L \geq 0$. Then for all n, there exists a constant C (depending on f, ϕ and $\{j_n\}$), such that

$$\left| 2^{j_n} \int_{\mathbb{R}^2} C_n^2(s,t)dsdt - \int_{\mathbb{R}} f^2(y)dy \right| \leq Cn^{-\delta\alpha} \tag{A.0.63}$$

where $\delta \in (0, 1/3)$ is as in (1.1.2).

Proof. It is easy to see, for any $M > 0$,

$$\begin{aligned} 2^{j_n} \int_{\mathbb{R}^2} C_n^2(s,t)dsdt =& 2^{j_n} \int_{[-M,M]^2} C_n^2(s,t)dsdt + 2^{j_n} \int_{[-M,M]^C} \int_{[-M,M]^C} C_n^2(s,t)dsdt \\ &+ 2^{j_n} \int_{[-M,M]} \int_{[-M,M]^C} C_n^2(s,t)dsdt \\ &+ 2^{j_n} \int_{[-M,M]^C} \int_{[-M,M]} C_n^2(s,t)dsdt. \end{aligned} \tag{A.0.64}$$

(A.0.38) and lemma $A.0.6$ give that,

$$\begin{aligned} &\left| 2^{j_n} \int_{[-M,M]^2} C_n^2(s,t)dsdt - \int_{-M}^{M} f^2(y)dy \right| \\ =& \left| \int_{-A}^{A} \int_{-A}^{A} \int_{-2A}^{2A} \int_0^1 K(x+z,x)K(x+w,x)K(x+z,x-u) \right. \\ &\left. \times K(x+w,x-u) \left(I(j_n) - \int_{-M}^{M} f^2(y)dy \right) dxdudzdw \right|. \end{aligned} \tag{A.0.65}$$

Thus we estimate the convergence rate of $\left| I(j_n) - \int_{-M}^{M} f^2(y)dy \right|$. Choosing M to be an integer satisfying $M \geq L + 2^{-j_n}(4A+1)$, we may decompose $I_1(j_n)$, which is defined in

(A.0.38), into four parts as follows:

$$I_1(j_n) = \left(\sum_{i=2A}^{\lfloor 2^{j_n} L \rfloor - 2A - 1} + \sum_{i=\lfloor 2^{j_n} L \rfloor - 2A}^{\lceil 2^{j_n} L \rceil + 2A - 1} + \sum_{i=\lceil 2^{j_n} L \rceil + 2A}^{2^{j_n} M - 2A - 1} + \sum_{i=2^{j_n} M - 2A}^{2^{j_n} M + A} \right) 2^{-j_n} f(2^{-j_n}(x+i))$$
$$\times f(2^{-j_n}(u'+i)) 1(2^{-j_n}(z'+i) \in [0, M]) 1(2^{-j_n}(w'+i) \in [0, M]) \tag{A.0.66}$$
$$=: I'_4(j_n) + I'_5(j_n) + I'_6(j_n) + I'_7(j_n).$$

$I'_4(j_n)$ is essentially the same as $I_4(j_n)$ in (A.0.41). Recall the notations of $P_{i,n}$, $S_{i,n}$, M_{ij}, m_{ij} in the proof of $I_4(j_n)$. Consider the partition $P_{i,n}$ on $[0, L]$. Since now f is assumed to be continuous on $[0, L]$, there exist x_{ij}, y_{ij}, such that $f(x_{ij}) = M_{ij}$ and $f(y_{ij}) = m_{ij}$. We can rewrite (A.0.43) and (A.0.44) as

$$S_{i,n} \le f^2(0) \triangle y_{i,1} + \sum_{j=0}^{N_i} f^2(x_{ij}) \triangle y + f^2(L) \triangle y_{i,N_i+3} \tag{A.0.67}$$

and

$$S_{i,n} \ge f^2(0) \triangle y_{i,1} + \sum_{j=0}^{N_i} f^2(y_{ij}) \triangle y + f^2(L) \triangle y_{i,N_i+3}. \tag{A.0.68}$$

If $|s-t| < \triangle y$, there exists C depending on f and $\{j_n\}$, such that, for all n large depending on $\{j_n\}$,

$$|f^2(s) - f^2(t)| \le |f(s) + f(t)||f(s) - f(t)| \le C(\triangle y)^\alpha \le Cn^{-\delta\alpha}. \tag{A.0.69}$$

So we obtain

$$\left| S_{i,n} - \int_0^L f^2(y) dy \right| \le CLn^{-\delta\alpha}. \tag{A.0.70}$$

Obviously, $f^2(0)\triangle y_{i,1}$ and $f^2(L)\triangle y_{i,N_i+3}$ are both bounded by $Cn^{-\delta}$. From (A.0.41), for all $x \in [0,1]$, $u' \in [-2A, 2A+1]$, $z' \in [-A, A+1]$, $w' \in [-A, A+1]$,

$$\begin{aligned}\left|I_4'(j_n) - \int_0^L f^2(y)dy\right| &\leq \frac{1}{4A+1}\sum_{i=2A}^{6A}\left|S_{i,n} - f^2(0)\triangle y_{i,1} - f^2(L)\triangle y_{i,N_i+3} - \int_0^L f^2(y)dy\right| \\ &\leq C\left(n^{-\delta\alpha} + n^{-\delta}\right) \\ &\leq Cn^{-\delta\alpha}\end{aligned} \tag{A.0.71}$$

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

Next we'll look at $I_6'(j_n)$ and consider a partition $P_{i,n}$ on $[L, M]$.

$$\begin{aligned}I_6'(j_n) &= \sum_{i=\lceil 2^{j_n}L\rceil+2A}^{2^{j_n}M-2A-1} 2^{-j_n} f(2^{-j_n}(x+i))f(2^{-j_n}(u'+i)) \\ &= \frac{1}{4A+1}\sum_{i=\lceil 2^{j_n}L\rceil+2A}^{\lceil 2^{j_n}L\rceil+6A}\sum_{j=0}^{N_i}\triangle y f(2^{-j_n}(x+i)+j\triangle y)f(2^{-j_n}(u'+i)+j\triangle y).\end{aligned} \tag{A.0.72}$$

Let $\xi_{ij} := 2^{-j_n}(x+i)+j\triangle y$, $\xi'_{ij} = 2^{-j_n}(u'+i)+j\triangle y$. Since f is bounded and monotonically decreasing on $[L,\infty)$, it follows that

$$\sum_{j=0}^{N_i}\triangle y f(\xi_{ij})f(\xi'_{ij}) \geq \sum_{j=1}^{N_i}\triangle y f^2(M_{ij}) \geq \int_{L+\triangle y_{i,1}+\triangle y}^{M-\triangle y_{i,N_i+3}} f^2(y)dy \tag{A.0.73}$$

and

$$\begin{aligned}\sum_{j=0}^{N_i}\triangle y f(\xi_{ij})f(\xi'_{ij}) &\leq \triangle y f(\xi_{i0})f(\xi'_{i0}) + \sum_{j=0}^{N_i-1}\triangle y f^2(y_{ij}) \\ &\leq C\triangle y + \int_{L+\triangle y_{i,1}}^{M-\triangle y_{i,N_i+3}-\triangle y} f^2(y)dy,\end{aligned} \tag{A.0.74}$$

where $\triangle y_{i,1}$ and $\triangle y_{i,N_i+3}$ are the lengths of the first and last subintervals. They satisfy that $0 \leq \triangle y_{i,1} \leq \triangle y$ and $0 \leq \triangle y_{i,N_i+3} \leq \triangle y$. So we have the bound

$$\left|\sum_{j=0}^{N_i} \triangle y f(\xi_{ij}) f(\xi'_{ij}) - \int_L^M f^2(y)dy\right| \le C\triangle y \le Cn^{-\delta}. \tag{A.0.75}$$

When $M \ge L + 2^{-j_n}(4A+1)$, for all $x \in [0,1]$, $u' \in [-2A, 2A+1]$, $z' \in [-A, A+1]$, $w' \in [-A, A+1]$ and all n,

$$\left|I'_6(j_n) - \int_L^M f^2(y)dy\right| \le Cn^{-\delta}, \tag{A.0.76}$$

where C depends on f and $\{j_n\}$. We also have $|I'_5(j_n)| \le Cn^{-\delta}$ and $|I'_7(j_n)| \le Cn^{-\delta}$. Collecting these bounds,

$$\left|I_1(j_n) - \int_0^M f^2(y)dy\right| \le C(n^{-\delta\alpha} + n^{-\delta}) \le Cn^{-\delta\alpha}. \tag{A.0.77}$$

We can obtain the same bound for $|I_3(j_n) - \int_{-M}^0 f^2(y)dy|$. And $I_2(j_n) \le Cn^{-\delta}$. we have the convergence rate

$$\left|I(j_n) - \int_{-M}^M f^2(y)dy\right| \le Cn^{-\delta\alpha}. \tag{A.0.78}$$

Coming back to (A.0.65), we get

$$\left|2^{j_n} \int_{[-M,M]^2} C_n^2(s,t)dsdt - \int_{-M}^M f^2(y)dy\right| \le Cn^{-\delta\alpha}. \tag{A.0.79}$$

The derivation of a bound on $\left|2^{j_n} \int_{[-M,M]^C} \int_{[-M,M]^C} C_n^2(s,t)dsdt - \int_{[-M,M]^C} f^2(y)dy\right|$ is analogous to the above inequality. We'll analyze $I''_1(j_n)$, which is the counterpart of $I_1(j_n)$

in (A.0.38), with $[-M, M]$ replaced by $[-M, M]^C$. It can be decomposed into two terms:

$$\begin{aligned} I_1''(j_n) &= \left(\sum_{i=2^{j_n}M-A}^{2^{j_n}M+2A} + \sum_{i=2^{j_n}M+2A+1}^{\infty} \right) 2^{-j_n} f(2^{-j_n}(x+i)) f(2^{-j_n}(u'+i)) \\ &\quad 1(2^{-j_n}(z'+i) \in (M, \infty)) 1(2^{-j_n}(w'+i) \in (M, \infty)) \\ &=: I_4''(j_n) + I_5''(j_n). \end{aligned} \tag{A.0.80}$$

We now analyze $I_5''(j_n)$.

$$\begin{aligned} I_5''(j_n) &= \sum_{i=2^{j_n}M+2A+1}^{\infty} 2^{-j_n} f(2^{-j_n}(x+i)) f(2^{-j_n}(u'+i)) \\ &= \frac{1}{4A+1} \sum_{2^{j_n}M+2A+1}^{2^{j_n}M+6A+1} \sum_{j=0}^{\infty} \triangle y f(\xi_{ij}) f(\xi_{ij}'). \end{aligned} \tag{A.0.81}$$

Due to the monotonicity of f on (M, ∞), there exist x_{ij}, y_{ij} and $0 \le \triangle y_{i,1} \le \triangle y$, such that

$$\sum_{j=0}^{\infty} \triangle y f(\xi_{ij}) f(\xi_{ij}') \ge \sum_{j=1}^{\infty} \triangle y f^2(x_{ij}) \ge \int_{M+\triangle y_{i,1}+\triangle y}^{\infty} f^2(y) dy \tag{A.0.82}$$

and

$$\sum_{j=0}^{\infty} \triangle y f(\xi_{ij}) f(\xi_{ij}') \le \triangle y f(\xi_{i0}) f(\xi_{i0}') + \sum_{j=0}^{\infty} \triangle y f^2(y_{ij}) \le C\triangle y + \int_{M+\triangle y_{i,1}}^{\infty} f^2(y) dy. \tag{A.0.83}$$

Thus,

$$\left| \sum_{j=0}^{\infty} \triangle y f(\xi_{ij}) f(\xi_{ij}') - \int_M^{\infty} f^2(y) dy \right| \le C\triangle y \le Cn^{-\delta}, \tag{A.0.84}$$

where C depends on f and $\{j_n\}$. So

$$\left| I_5''(j_n) - \int_M^{\infty} f^2(y) dy \right| \le Cn^{-\delta}. \tag{A.0.85}$$

To get a bound for $|2^{j_n}\int_{[-M,M]^C}\int_{[-M,M]^C} C_n^2(s,t)dsdt - \int_{[-M,M]^C} f^2(y)dy|$ from this, we continue as in the proof of (A.0.79).

It is easier to bound the other two terms in (A.0.64). By the proof of (A.0.37),

$$\begin{aligned}
&2^{j_n}\int_{[-M,M]}\int_{[-M,M]^C} C_n^2(s,t)dsdt \qquad \text{(A.0.86)}\\
&= \int_{-A}^{A}\int_{-A}^{A}\int_{-2A}^{2A}\sum_{i=-\infty}^{\infty}\int_0^1 2^{-j_n}K(x+z,x)K(x+w,x)K(x+z,x-u)K(x+w,x-u)\\
&\times f(2^{-j_n}(x+i))f(2^{-j_n}(x+i-u))1(2^{-j_n}(z+x+i)\in[-M,M])\\
&\times 1(2^{-j_n}(x+w+i)\in[-M,M]^C)dxdudzdw.
\end{aligned}$$

As $z+x\in[-A,A+1]$, $1(2^{-j_n}(z+x+i)\in[-M,M])\neq 0$ only if $-2^{j_n}M-A-1\leq i\leq 2^{j_n}M+A$. And $1(2^{-j_n}(x+w+i)\in[-M,M]^C)\neq 0$ only if $i>2^{j_n}M-A-1$ or $i<-2^{j_n}M+A$. So there are at most finitely many summands that are not zero. Now by the boundedness of f, for n large enough,

$$\left|2^{j_n}\int_{[-M,M]}\int_{[-M,M]^C} C_n^2(s,t)dsdt\right|\leq Cn^{-\delta}. \qquad \text{(A.0.87)}$$

For the same reason,

$$\left|2^{j_n}\int_{[-M,M]^C}\int_{[-M,M]} C_n^2(s,t)dsdt\right|\leq Cn^{-\delta}. \qquad \text{(A.0.88)}$$

(A.0.63) follows by collecting the bounds on the four terms in (A.0.64) and taking C sufficiently large so that it is true for all n. □

A.0.8 Corollary. Assume the same conditions in Lemma $A.0.7$, for all n sufficiently large depending on f and $\{j_n\}$,

$$\left|2^{j_n}\int_{\mathbb{R}^2} R_n^2(s,t)dsdt - \int_{\mathbb{R}} f^2(y)dy\right| \le C(n^{-\delta/2} + n^{-\delta\alpha}), \tag{A.0.89}$$

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

Proof. Set $g_n(s,t) := 2^{j_n/2}R_n(s,t)$ and $h_n(s,t) := 2^{j_n/2}C_n(s,t)$. By (A.0.33) in the proof of Lemma $A.0.3$, for some constant C depending on f, ϕ and $\{j_n\}$,

$$\|g_n - h_n\|_2^2 \le Cn^{-\delta}, \tag{A.0.90}$$

where $\|\cdot\|_2$ denotes the L_2 norm on $\mathbb{R}^2$ in this corollary. By Minkowski inequality,

$$\left|\|g_n\|_2 - \|h_n\|_2\right| \le \|g_n - h_n\|_2 \le Cn^{-\delta/2}. \tag{A.0.91}$$

Lemma $A.0.7$ gives $\|h_n\|_2 \to \sqrt{\int_{\mathbb{R}} f^2(y)dy}$. $\|g_n\|_2 \to \sqrt{\int_{\mathbb{R}} f^2(y)dy}$ as well. So $\|g_n\|_2 + \|h_n\|_2$ is bounded by a constant if n is large depending on f. Thus,

$$\left|\|g_n\|_2^2 - \|h_n\|_2^2\right| \le (\|g_n\|_2 + \|h_n\|_2)\left|\|g_n\|_2 - \|h_n\|_2\right| \le Cn^{-\delta/2}, \tag{A.0.92}$$

$$\begin{aligned}
&\left|2^{j_n}\int_{\mathbb{R}^2} R_n^2(s,t)dsdt - \int_{\mathbb{R}} f^2(y)dy\right| \\
&\le \left|\|g_n\|_2^2 - \|h_n\|_2^2\right| + \left|2^{j_n}\int_{\mathbb{R}^2} C_n^2(s,t)dsdt - \int_{\mathbb{R}} f^2(y)dy\right| \\
&\le C(n^{-\delta/2} + n^{-\delta\alpha}).
\end{aligned} \tag{A.0.93}$$

□

A.0.9 Lemma. Assume that (f), (S1) hold and F satisfies condition (1.1.9). For the operator $\mathcal{R}_{n,F}$ defined by (1.3.37), we have

$$\sup\{\|\mathcal{R}_{n,F}\varphi\|_2^2 : \|\varphi\|_2 = 1, \varphi \in L_2(F)\} \leq 2^{-2j_n} C(\Phi, f), \tag{A.0.94}$$

where

$$C(\Phi, f) = 2\|\Phi\|_1^4 \left(\|f\|_\infty^2 + \|f\|_2^4\right). \tag{A.0.95}$$

Proof.

$$\begin{aligned}
\|\mathcal{R}_{n,F}\varphi\|_2^2 &= \left\| \int_F R_n(s,t)\varphi(t)dt \right\|_2^2 \\
&= \left\| \int_F 2^{j_n} \int_{\mathbb{R}} \bar{K}_n(t,x)\bar{K}_n(s,x) f(x) dx \varphi(t) dt \right\|_2^2 \\
&= 2^{2j_n} \int_F \left(\int_F \int_{\mathbb{R}} \bar{K}_n(t,x)\bar{K}_n(s,x) f(x)\varphi(t) dx dt \right)^2 ds \qquad \text{(A.0.96)} \\
&= \int_F \left(\int_F \int_{\mathbb{R}} \left(2^{j_n} K_n(t,x) - 2^{j_n}\mathbb{E}K_n(t,X)\right) \left(2^{j_n} K_n(s,x) \right.\right. \\
&\qquad \left.\left. -2^{j_n}\mathbb{E}K_n(s,X)\right) f(x)\varphi(t) dx dt \right)^2 ds \\
&= 2^{-2j_n} \int_F \left(\int_F \left(\int_{\mathbb{R}} 2^{2j_n} K_n(t,x) K_n(s,x) f(x) dx \right.\right. \\
&\qquad \left.\left. - 2^{2j_n}\mathbb{E}K_n(t,X)\mathbb{E}K_n(s,X) \right) \varphi(t) dt \right)^2 ds \\
&= 2^{-2j_n} \int_F \left(\int_F \left(\int_{\mathbb{R}} 2^{2j_n} K_n(t,x) K_n(s,x) f(x) dx - \mu_n(s)\mu_n(t) \right) \varphi(t) dt \right)^2 ds \\
&\leq 2 \cdot 2^{-2j_n} \int_F \left(\int_F \left(\int_{\mathbb{R}} 2^{2j_n} K_n(t,x) K_n(s,x) f(x) dx \right) \varphi(t) dt \right)^2 ds \\
&\qquad + 2 \cdot 2^{-2j_n} \int_F \left(\int_F \mu_n(s)\mu_n(t)\varphi(t) dt \right)^2 ds,
\end{aligned}$$

where $\mu_n(s)$ is defined in (A.0.29).

By Fubini and Cauchy-Schwarz,

$$\int_F \left(\int_F \mu_n(s)\mu_n(t)\varphi(t)dt \right)^2 ds \leq \|\varphi\|_2^2 \left(\int_{\mathbb{R}} \mu_n^2(s)ds \right)^2. \tag{A.0.97}$$

Using the bound (A.0.31), we get,

$$\|\varphi\|_2^2 \left(\int_{\mathbb{R}} \mu_n^2(s)ds \right)^2 \leq \|\varphi\|_2^2 \left(\|\Phi\|_1^2 \|f\|_2^2 \right)^2. \tag{A.0.98}$$

Now consider the first term in the last inequality in (A.0.96),

$$\begin{aligned}
&\int_F \left(\int_F \left(\int_{\mathbb{R}} 2^{2j_n} K_n(t,x) K_n(s,x) f(x) dx \right) \varphi(t) dt \right)^2 ds \\
&= \int_F \left(\int_{\mathbb{R}} \left(\int_F 2^{j_n} K_n(t,x) \varphi(t) dt \right) f(x) 2^{j_n} K_n(s,x) dx \right)^2 ds \\
&\leq \int_F \left\{ \left(\int_{\mathbb{R}} \left(\int_F 2^{j_n} |K_n(t,x)| |\varphi(t)| dt \right)^2 2^{j_n} |K_n(s,x)| dx \right) \left(\int_{\mathbb{R}} f^2(x) 2^{j_n} |K_n(s,x)| dx \right) \right\} ds.
\end{aligned} \tag{A.0.99}$$

Since f is bounded,

$$\begin{aligned}
\int_{\mathbb{R}} f^2(x) 2^{j_n} |K_n(s,x)| dx &\leq \|f\|_\infty^2 \int_{\mathbb{R}} 2^{j_n} |K_n(s,x)| dx \\
&\leq \|f\|_\infty^2 \int_{\mathbb{R}} 2^{j_n} |\Phi(2^{j_n} s - 2^{j_n} x)| dx \\
&= \|f\|_\infty^2 \|\Phi\|_1.
\end{aligned} \tag{A.0.100}$$

(A.0.99) is bounded from above by

$$\begin{aligned}
&\|f\|_\infty^2 \|\Phi\|_1 \int_F \left\{ \int_{\mathbb{R}} \left(\int_F 2^{j_n} |K_n(t,x)| |\varphi(t)| dt \right)^2 2^{j_n} |K_n(s,x)| dx \right\} ds \\
&\leq \|f\|_\infty^2 \|\Phi\|_1 \int_F \int_{\mathbb{R}} \left(\int_F 2^{j_n} |\Phi_n(t-x)| |\varphi(t)| dt \right)^2 2^{j_n} |\Phi_n(s-x)| dx ds
\end{aligned}$$

$$
\begin{aligned}
&= \|f\|_\infty^2 \|\Phi\|_1 \int_{\mathbb{R}} \left(\int_F 2^{j_n} |\Phi_n(t-x)||\varphi(t)| dt \right)^2 \int_F 2^{j_n} |\Phi_n(s-x)| ds dx \\
&\le \|f\|_\infty^2 \|\Phi\|_1^2 \int_{\mathbb{R}} \left(\int_F 2^{j_n} |\Phi_n(t-x)||\varphi(t)| dt \right)^2 dx \\
&\le \|f\|_\infty^2 \|\Phi\|_1^2 \int_{\mathbb{R}} \int_F 2^{j_n} |\Phi_n(t-x)||\varphi(t)|^2 dt \int_F 2^{j_n} |\Phi_n(t-x)| dt dx \\
&\le \|f\|_\infty^2 \|\Phi\|_1^3 \int_F \left(|\varphi(t)|^2 \int_{\mathbb{R}} 2^{j_n} |\Phi_n(t-x)| dx \right) dt \\
&= \|f\|_\infty^2 \|\Phi\|_1^4 \|\varphi\|_2^2. \qquad \text{(A.0.101)}
\end{aligned}
$$

Together with (A.0.96) and (A.0.98),

$$
\begin{aligned}
\|\mathcal{R}_{n,F}\varphi\|_2^2 &\le 2 \cdot 2^{-2j_n} \|\varphi\|_2^2 \left(\|\Phi\|_1^2 \|f\|_2^2 \right)^2 + 2 \cdot 2^{-2j_n} \|f\|_\infty^2 \|\Phi\|_1^4 \|\varphi\|_2^2 \\
&= 2^{-2j_n} \|\varphi\|_2^2 \|\Phi\|_1^4 \left(2\|f\|_2^4 + 2\|f\|_\infty^2 \right).
\end{aligned} \qquad \text{(A.0.102)}
$$

Then

$$
\sup\{\|\mathcal{R}_{n,F}\varphi\|_2^2 : \|\varphi\|_2 = 1, \varphi \in L_2(F)\} \le 2^{-2j_n} C(\Phi, f), \qquad \text{(A.0.103)}
$$

where $C(\Phi, f)$ is given by (A.0.95). □

A.0.10 Corollary. Under the hypotheses of Lemma *A.0.9*,

$$
\lim_{n\to\infty} \frac{2^{3j_n}}{n^2} \mathbb{E} U_n^2([-M, M]) = \int_{-M}^{M} f^2(t) dt. \qquad \text{(A.0.104)}
$$

Proof.

$$
\begin{aligned}
\lim_{n\to\infty} \frac{2^{3j_n}}{n^2} \mathbb{E} U_n^2([-M, M]) &= \lim_{n\to\infty} \frac{2^{3j_n}}{n^2} \sum_{1\le i\ne j\le n} \mathbb{E} \left\{ \int_{-M}^{M} \bar{K}_n(t, X_i) \bar{K}_n(t, X_j) dt \right\}^2 \\
&= \lim_{n\to\infty} \frac{2^{3j_n}}{n^2} \sum_{1\le i\ne j\le n} \int_{-M}^{M} \int_{-M}^{M} \mathbb{E} \bar{K}_n(t, X_i) \bar{K}_n(t, X_j) \qquad \text{(A.0.105)}
\end{aligned}
$$

$$
\begin{aligned}
&\times \bar{K}_n(s, X_i)\bar{K}_n(s, X_j)dtds \\
&= \lim_{n\to\infty} \frac{2^{3j_n}}{n^2} n(n-1) \int_{-M}^{M}\int_{-M}^{M} \mathbb{E}\bar{K}_n(t, X_i)\bar{K}_n(s, X_i) \\
&\quad \mathbb{E}\bar{K}_n(t, X_j)\bar{K}_n(s, X_j)dtds \\
&= \lim_{n\to\infty} \frac{2^{3j_n}}{n^2} n(n-1)2^{-2j_n} \int_{-M}^{M}\int_{-M}^{M} R_n^2(t,s)dsdt \\
&= \lim_{n\to\infty} 2^{j_n} \int_{-M}^{M}\int_{-M}^{M} R_n^2(t,s)dsdt.
\end{aligned}
$$

By (A.0.27), (A.0.34) and an argument similar to that in Corollary $A.0.8$, we see that

$$
\lim_{n\to\infty} 2^{j_n} \int_{-M}^{M}\int_{-M}^{M} R_n^2(t,s)dsdt = \lim_{n\to\infty} 2^{j_n} \int_{-M}^{M}\int_{-M}^{M} C_n^2(t,s)dsdt. \tag{A.0.106}
$$

Then (A.0.104) follows.

□

Bibliography

[1] Babu, G.J., Canty, A.J. and Chaubey, Y.P. (2002) *Application of Bernstein polynomials for smooth estimation of a distribution and density function*. J. Statist. Plann. Inference **105**, 377-392.

[2] Babu, G.J. and Chaubey, Y.P. (2006) *Smooth estimation of a distribution and density function on a hypercube using Bernstein polynomials for dependent random vectors*. Statist. Probab. Lett. **76**, 959-969.

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

[4] Chow, Y.S. and Teicher, H. (1978) *Probability Theory: Independence, Interchangeability, Martingales*. 1st ed. Springer Verlag, New York.

[5] Daubechies, I. (1992) *Ten lectures on wavelets*. CBMS-NSF Reg. Conf. Ser. in Appl. Math. 61. Philadelphia, Society for Industrial and Applied Mathematics.

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

[7] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995) *Wavelet shrinkage: asymptopia?* J. Roy. Statist. Soc. Ser. B **57**, 301-369.

[8] 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.

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

[10] 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.

[11] Einmahl, U. and Mason, D.M. (2000) *An empirical process approach to the uniform consistency of kernel-type function estimators*. J. Theoret. Probab. **13**, 1-37.

[12] 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.

[13] Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*. Vol II, 2nd ed. Wiley, New York.

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

[15] Ghosal, S. (2001) *Convergence rates for density estimation with Bernstein polynomials*. Ann. Statist., **29**, 1264-1280.

[16] Giné, E. and Guillou, A. (2001) *On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals*. Ann. Inst. H. Poincaré Probab. Statist. **37**, 303-522.

[17] Giné, E. and Guillou, A. (2002) *Rates of strong uniform consistency for multivariate kernel density estimators*. Ann. Inst. H. Poincaré Probab. Statist. **38**, 907-921.

[18] Giné, E., Latała, 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.

[19] Giné, E. and Mason D.M. (2000) *The law of the iterated logarithm for the integrated squared deviation of a kernel density estimator*. Bernoulli, **10**, 721-752.

[20] Giné, E. and Nickl, R. (2009) *Uniform limit theorems for wavelet density estimators*. Ann. Probab. **37**, 1605-1646.

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

[22] Hall, P. and Heyde, C.C. (1980) *Martingale limit theory and its application*. Academic Press, New York.

[23] Hall, P. and Patil, P. (1995) *Formulae for mean integrated squared error of nonlinear wavelet-based density estimators*. Ann. Statist., **23**, 905-928.

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

[25] Jones, M.C. (1993) *Simple boundary correction for kernel density estimation*. Stat. Comput. **3**, 135-146.

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

[27] 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.

[28] Kruijer, W. and van der Vaart, A. (2008) *Posterior convergence rates for Dirichlet mixtures of beta densities*. J. Statist. Plann. Inference, **138**, 1981-1992.

[29] Leblanc, A. (2010) *A bias-reduced approach to density estimation using Bernstein polynomials*. J. Nonparametr. Stat. **22**, 459-475.

[30] Leblanc, A. (2012) *On the boundary properties of Bernstein polynomial estimators of density and distribution functions*. J. Statist. Plann. Inference, **142**, 2762-2778.

[31] Lorentz, G.G. (1986) *Bernstein Polynomials*, 2nd ed. Chelsea Publishing, New York.

[32] Lu, L. (2013) *On the integrated squared error of the linear wavelet density estimator*. J. Statist. Plann. Inference, **143**, 1548-1565.

[33] Massiani, A. (2002) *Étude asymptotique locale de l'estimateur par ondelettes*. C. R. Math. Acad. Sci. Paris Ser. I. **335**, 553-556.

[34] Massiani, A. (2003) *Vitesse de convergence uniforme presque sure de l'estimateur linéaire par méthode d'ondelettes*. C. R. Math. Acad. Sci. Paris Ser. I. **337**, 67-70.

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

[36] Petrone, S. (1999) *Bayesian density estimation using Bernstein polynomials*. Canad. J. Statist. **27**, 105-126.

[37] Petrone, S. and Wasserman, L. (2002) *Consistency of Bernstein polynomial posteriors*. J. Roy. Statist. Soc. Ser. B **64**, 79-100.

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

[39] Robbins, H. (1955) *A remark on Stirling's formula*. Amer. Math. Monthly **62**, 26-29.

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

[41] Talagrand, M. (1996) *New concentration inequalities in product spaces*. Invent. Math. **126**, 505-563.

[42] Tenbusch, A. (1994) *Two-dimensional Bernstein polynomial density estimators*. Metrika **41**, 233-253.

[43] Varron, D. (2008) *Some asymptotic results on density estimators by wavelet projections*. Statist. Probab. Lett. **78**, 2517-2521.

[44] Vitale, R.A. (1975) *A Bernstein polynomial approach to density function estimation*. Statistical Inference and Related Topics **2**, 87-99.

[45] Zhang, S. and Zheng Z. (1999) *On the asymptotic normality for L_2-error of wavelet density estimator with application*. Comm. Statist. Theory Methods. **28**, 1093-1104.