THE LAW OF THE ITERATED LOGARITHM FOR THE INTEGRATED SQUARED DEVIATION OF A KERNEL DENSITY ESTIMATOR

by Evarist Giné[1] and David M. Mason[2]

[1] *University of Connecticut, Department of Mathematics, U-3009, Storrs, CT 06269, USA*
E-mail: gine@math.uconn.edu
[2] *University of Delaware, Statistics Program, 206 Townsend Hall, Newark, DE 19717, USA*
E-mail: davidm@udel.edu

Let $f_{n,K}$ denote a kernel estimator of a density f in $\mathbf{R}$ such that $\int_{\mathbf{R}} f^p(x)dx<\infty$ for some $p>2$. It is shown, under quite general conditions on the kernel K and on the window sizes, that the centered integrated squared deviation of $f_{n,K}$ from its mean, $\|f_{n,K}-Ef_{n,K}\|_2^2-E\|f_{n,K}-Ef_{n,K}\|_2^2$ satisfies a law of the iterated logarithm. This is then used to obtain an LIL for the deviation from the true density, $\|f_{n,K}-f\|_2^2-E\|f_{n,K}-f\|_2^2$. The main tools are the Komlós-Major-Tusnády approximation, a moderate deviation result for triangular arrays of weighted chi-square variables adapted from Pinsky (1966) and an exponential inequality of Giné, Latała and Zinn (2000) for degenerate U-statistics applied in combination with decoupling and maximal inequalities.

Key words and phrases: kernel density estimator, integrated squared deviation, law of the iterated logarithm.

Runninghead: The LIL for L_2 functionals of kernel density estimators

1. **Introduction.** As far as we know there are no laws of the iterated logarithm for the integrated p-th absolute deviation of a kernel density estimator from its mean, although central limit theorems do exist (Bickel and Rosenblatt (1973), Rosenblatt (1975), Nadaraya (1989), Hall (1984), Csörgő and Horváth (1988), Beirlant and Mason (1995), Mason in Eggermont and LaRiccia (2001), Giné, Mason and Zaitsev (2003)). This anomaly seems to be due to the fact that there are serious difficulties both to find the proper way in blocking and in deriving sufficiently precise moderate deviation results. In this paper we show how these difficulties can be handled in the case $p = 2$. In the process we shall spotlight a number of techniques that should be of independent interest. Unfortunately our methods do not extend to other values of p.

In order to make our aim clear let us now fix some notation and introduce our basic assumptions. Throughout this paper we shall assume that f is a probability density on the real line $\mathbf{R}$ such that

$$\int_{\mathbf{R}} f^p(x)dx < \infty \quad \text{for some } p > 2, \tag{1.1}$$

and our kernel K is a measurable function such that

$$\|K\|_v < \infty, \quad \|K\|_1 < \infty \quad \text{and} \quad \int_{\mathbf{R}} K(x)dx = 1, \tag{1.2}$$

where $\|\cdot\|_v$ denotes the total variation norm and $\|\cdot\|_r$, $1 \le r \le \infty$, the L_r norm with respect to Lebesgue measure on $\mathbf{R}$, λ. Note that condition (1.2) implies $\|K\|_r < \infty$ for all $1 \le r \le \infty$. Further, we shall assume that our window sizes $\{h_n\}$ form a sequence of positive numbers satisfying the conditions:

$$h_n \searrow 0, \quad h_n \asymp n^{-\delta} \ \text{ for some } \ \delta \in (0, 1/3), \tag{1.3}$$

and there exists an increasing sequence of positive constants $\{\lambda_k\}_{k \ge 1}$ satisfying $\lambda_{k+1}/\lambda_k \to 1$ and $\log\log \lambda_k / \log k \to 1$, as $k \to \infty$, such that

$$h_n \text{ is constant for } n \in [\lambda_k, \lambda_{k+1}), \ k \in \mathbf{N}. \tag{1.4}$$

(Note that the sequence $\lambda_k = \exp(k/\log(e+k))$ satisfies these conditions.) Without loss of generality, the numbers λ_k can be assumed to be integers, as we argue below. In (1.3) and elsewhere in this article, $A_n \asymp B_n$ means that

$$0 < \liminf_n A_n/B_n \le \limsup_n A_n/B_n < \infty.$$

Let X, X_i, $i \in \mathbf{N}$, be independent identically distributed random variables with density f. Then $f_{n,K}$, the classical density estimator of f, is defined as

$$f_{n,K}(t) := \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h_n}\right), \quad t \in \mathbf{R}. \tag{1.5}$$

In this notation we write the integrated squared deviation of a kernel density estimator from its mean as

$$\big\|f_{n,K} - Ef_{n,K}\big\|_2^2.$$

We are interested in establishing the law of the iterated logarithm [LIL] for the statistic

$$J_n := \big\|f_{n,K} - Ef_{n,K}\big\|_2^2 - E\big\|f_{n,K} - Ef_{n,K}\big\|_2^2. \tag{1.6}$$

Namely, we shall prove under the just stated conditions on f, K and $\{h_n\}$ that for $\sigma^2 > 0$ defined in Theorem 5.1 below,

$$\limsup_{n\to\infty} \pm \frac{n\sqrt{h_n}}{\sqrt{2\sigma^2 \log\log n}} J_n = 1, \quad \text{almost surely.}$$

This should be compared to a result of Mason (2003), who establishes under appropriate conditions that for some $\tau^2 > 0$

$$\limsup_{n\to\infty} \pm \frac{\sqrt{n}\left\{\big\|f_{n,K}\big\|_2^2 - \big\|Ef_{n,K}\big\|_2^2\right\}}{\sqrt{2\tau^2 \log\log n}} = 1, \quad \text{almost surely.}$$

which shows not unexpectedly that $|J_n|$ is of strictly smaller order than

$$\Big|\|f_{n,K}\|_2^2 - \|Ef_{n,K}\|_2^2\Big|.$$

We should mention that, with some abuse of notation, when we write $\log\log n$ it is understood to equal 1 if $n < e^e$ (alternatively, we could always take n to be larger than or equal to e^e).

Our proof of the LIL just described requires all the hypotheses given above. It would be particularly interesting to know whether the result also holds assuming only square integrability of the density f.

Here are the basic steps of our approach to proving our LIL. First we shall exploit the fact that the integrated squared deviation is, up to its diagonal term, a degenerate U–statistic. For such statistics there exists a recent exponential bound of the right order (up to constants) due to Giné, Latała and Zinn (2000). We shall show how to apply this inequality effectively to block the original sequence and to reduce the domain of integration of the statistic. Next we shall approximate the resulting U–statistic by a Gaussian chaos random variable via KMT and then derive a moderate deviation result for this random variable. Large deviation results for Gaussian chaos of order two can be found for instance in Ledoux and Talagrand (1991, p. 69), but they are not completely tailored for our purpose. On the other hand, since our Gaussian chaos is real and diagonalizes, we shall be able to obtain moderate deviation bounds, suitable for our needs, just by adapting an easy method of Pinsky (1969). Finally, after we have established all the necessary ingredients, we shall complete the proof in the usual way one establishes an LIL.

To clarify what we have in mind, let us introduce some additional notation. From now on we shall write

$$K_h(t-x) := K\left(\frac{t-x}{h}\right) \text{ and } \bar{K}_h(t-x) := K_h(t-x) - EK_h(t-X).$$

With this notation,

$$J_n = \frac{1}{n^2h_n^2}W_n(\mathbf{R}),$$

where, for any measurable $F \subset \mathbf{R}$, we set

$$\begin{aligned} W_n(F) &:= \int_F \Big(\sum_{i=1}^n \bar{K}_{h_n}(t-X_i)\Big)^2 dt - E\int_F \Big(\sum_{i=1}^n \bar{K}_{h_n}(t-X_i)\Big)^2 dt \\ &= \sum_{1\le i\ne j\le n} \int_F \bar{K}_{h_n}(t-X_i)\bar{K}_{h_n}(t-X_j)\, dt + \sum_{i=1}^n \int_F \Big(\bar{K}_{h_n}^2(t-X_i) - E\bar{K}_{h_n}^2(t-X)\Big)dt \\ &:= U_n(F) + L_n(F). \end{aligned} \tag{1.7}$$

We shall assume that the measurable set F satisfies the conditions

$$\int_F f(t)dt > 0 \text{ and } \lambda\Big(\big\{x+y:\ x\in F,\ |y|<\varepsilon\big\}\cap F^c\Big) \to 0 \text{ as } \varepsilon \searrow 0. \tag{1.8}$$

Basic to our proofs are asymptotic properties of $W_n(F)$ with the choices $F = \mathbf{R}$, $F = [-M, M]$ and its complement $F = [-M, M]^c$ for $M > 0$, which all satisfy (1.8).

Note that $U_n(F)$ is a canonical (degenerate) U–statistic for the law of X, and that the diagonal term $L_n(F)$ is a sum of independent random variables. As mentioned above, it is this special form of J_n that makes it treatable for the LIL, at least for us here (and also, before us, for P. Hall (1984) and Nadaraya (1989) in connection with the central limit theorem).

Finally, here is the content of the different sections. In Section 2 we present some variance computations to be used throughout. In Section 3 we show that the residual random variables that remain from the main part of the statistic when we restrict the domain of integration or when we subtract $W_m(\mathbf{R})$ from $W_n(\mathbf{R})$, $m < n$, as required for blocking, are asymptotically negligible. In Section 4 we obtain the necessary moderate deviation result for $W_n([-M, M])$. In Section 5 we state the main result, which is the LIL for the integrated squared deviation of the density estimator from its mean, and complete its proof.

From a statistical point of view, the integrated squared deviation of the kernel density estimator from the true density f , $\|f_{n,K} - f\|_2^2$, is at least as interesting as its integrated squared deviation from its mean. We shall make some remarks in Section 6 on how our results apply to the LIL for

$$\|f_{n,K} - f\|_2^2 - E\|f_{n,K} - f\|_2^2. \tag{1.9}$$

In fact J_n constitutes the degenerate U-statistic part of (1.9), which can be written as the sum of J_n and a linear term that often dominates and can be dealt with in the usual way. The same is true for $\|f_{n,K}\|_2^2 - E\|f_{n,K}\|_2^2$, as mentioned in Mason (2003).

2. Variance computations.

Hall (1984) has also similar variance computations, but under more restrictive assumptions that we are able to relax in this section mainly because of the following observation.

2.1. Lemma. *Let φ be an integrable function on $\mathbf{R}$ and set $\varphi_\varepsilon(x) = \varepsilon^{-1}\varphi(\varepsilon^{-1}x)$, $\varepsilon > 0$. Then, for all functions g in $L_p(\mathbf{R})$, $1 \le p < \infty$,*

$$\lim_{\varepsilon \searrow 0} \big\|g * \varphi_\varepsilon - g\big\|_p = \lim_{\varepsilon \searrow 0} \left(\int_{\mathbf{R}} \left| \int_{\mathbf{R}} \varphi(u)\big(g(x - \varepsilon u) - g(x)\big)du \right|^p dx \right)^{1/p} = 0 \tag{2.1}$$

Proof. The generalized Minkowski inequality gives

$$\begin{aligned}
&\left(\int_{\mathbf{R}} \left| \int_{\mathbf{R}} \varphi(u)(g(x - \varepsilon u) - g(x))du \right|^p dx \right)^{1/p} \\
&\qquad\qquad \le \int_{\mathbf{R}} \left(\int_{\mathbf{R}} \big|\varphi(u)\big(g(x - \varepsilon u) - g(x)\big)\big|^p dx \right)^{1/p} du \\
&\qquad\qquad = \int_{\mathbf{R}} |\varphi(u)| \big\|g(x - \varepsilon u) - g(x)\big\|_p du,
\end{aligned}$$

where the L_p norm is with respect to dx. Now, the last integrand is dominated by the integrable function $2\|g\|_p|\varphi(u)|$, and $\|g(x-\varepsilon u)-g(x)\|_p \to 0$ as $\varepsilon \to 0$ by the L_p-continuity of shifts. Therefore, (2.1) follows by dominated convergence. $\square$

Let us write

$$R_h(t,s) = h^{-1}\int_{\mathbf{R}} \left[\bar{K}_h\,(t-x)\,\bar{K}_h\,(s-x)\right] f(x)dx, \tag{2.2}$$

and define the operator $\mathcal{R}_h$ for $\varphi \in L_2(F)$,

$$\mathcal{R}_h\varphi(s) := \int_F R_h(s,t)\varphi(t)dt. \tag{2.3}$$

The main object in this section is to prove the following proposition.

2.2. Proposition. *Let F satisfy conditions (1.8) and assume $\|K\|_1 < \infty$, $\|K\|_2 < \infty$ and $||f||_p < \infty$ for some $p > 2$. Then, for any $0 < h \le 1$,*

$$\sup\left\{\|\mathcal{R}_h\varphi\|_2^2 : \|\varphi\|_2 = 1, \varphi \in L_2(F)\right\} \le C(K,f,r,p)h^{1+1/r}, \tag{2.4}$$

where $1/r + 2/p = 1$ and $C(K,f,r,p) = 2||f||_p^2||K||_r||K||_1^3 + 2\left(||f||_2^2\,\|K\|_1^2\right)^2$. Moreover,

$$\lim_{h\to 0} h^{-1}\int_{F^2} R_h^2(s,t)dsdt = \int_F f^2(x)dx\int_{\mathbf{R}}\left(\int_{\mathbf{R}} K(w+u)K(w)dw\right)^2 du. \tag{2.5}$$

Proof. The proof will be a consequence of the following lemmas.

2.3. Lemma. *Inequality (2.4) holds under the hypotheses of Proposition 2.2.*

Proof. Let φ be a function in $L_2(F)$. We have

$$||\mathcal{R}_h\varphi||_2^2 = h^{-2}\int_F\left\{\int_F\int_{\mathbf{R}}\left[\bar{K}_h\,(s-x)\,\bar{K}_h\,(t-x)\right] f(x)\varphi(t)dxdt\right\}^2 ds$$

$$= h^2\int_F\left\{\int_F\left[\int_{\mathbf{R}}\frac{K_h\,(s-x)}{h}\frac{K_h\,(t-x)}{h}f(x)dx - \mu_h(s)\mu_h(t)\right]\varphi(t)dt\right\}^2 ds$$

where

$$\mu_h(s) = h^{-1}\int_{\mathbf{R}} K\left(\frac{s-x}{h}\right) f(x)dx.$$

The above is less than or equal to

$$2h^2\int_F\left\{\int_F\left[\int_{\mathbf{R}}\frac{K_h\,(s-x)}{h}\frac{K_h\,(t-x)}{h}f(x)dx\right]\varphi(t)dt\right\}^2 ds$$

$$+2h^2\int_F\left\{\int_F \mu_h(s)\mu_h(t)\varphi(t)dt\right\}^2 ds.$$

Now

$$2h^2 \int_F \left\{ \int_F \mu_h(s)\mu_h(t)\varphi(t)dt \right\}^2 ds = 2h^2 \int_F \mu_h^2(s)ds \left\{ \int_F \mu_h(t)\varphi(t)dt \right\}^2$$
$$\leq 2h^2 ||\varphi||_2^2 \left(\int_{\mathbf{R}} \mu_h^2(s)ds \right)^2 .$$

By Cauchy-Schwarz,

$$\mu_h^2(s) \leq \left(h^{-1} \int_{\mathbf{R}} \left| K\left(\frac{s-x}{h}\right) \right| f(x)dx \right)^2 \leq h^{-1} \int_{\mathbf{R}} \left| K\left(\frac{s-x}{h}\right) \right| f^2(x)dx \, \|K\|_1 .$$

Thus by Fubini

$$\int_{\mathbf{R}} \mu_h^2(s)ds \leq ||f||_2^2 \, \|K\|_1^2 . \tag{2.6}$$

This gives

$$2h^2 ||\varphi||_2^2 \left(\int_R \mu_h^2(s)ds \right)^2 \leq 2h^2 ||\varphi||_2^2 \left(||f||_2^2 \, \|K\|_1^2 \right)^2 . \tag{2.7}$$

Next

$$2h^2 \int_F \left\{ \int_F \int_{\mathbf{R}} \frac{K_h\,(s-x)}{h} \frac{K_h\,(t-x)}{h} f(x)\varphi(t)dxdt \right\}^2 ds$$
$$= 2h^2 \int_F \left\{ \int_{\mathbf{R}} \left(\int_F \frac{K_h\,(t-x)}{h} \varphi(t)dt \right) f(x) \frac{K_h\,(s-x)}{h} dx \right\}^2 ds,$$

which by Cauchy–Schwarz is bounded from above by

$$2h^2 \int_F \left\{ \int_{\mathbf{R}} \left(\int_F \frac{|K_h\,(t-x)\,|}{h} |\varphi(t)|dt \right)^2 \frac{|K_h\,(s-x)\,|}{h} dx \int_{\mathbf{R}} \left[f^2(x) \frac{|K_h\,(s-x)\,|}{h} \right] dx \right\} ds. \tag{2.8}$$

Since by Hölder's inequality with $1/r + 2/p = 1$,

$$\int_{\mathbf{R}} \left[f^2(x) \frac{|K_h\,(s-x)\,|}{h} \right] dx \leq h^{1/r-1} ||f||_p^2 \left(\int_{\mathbf{R}} \frac{|K_h\,(s-x)\,|^r}{h} dx \right)^{1/r} = h^{1/r-1} ||f||_p^2 ||K||_r,$$

the bound in (2.8) is in turn bounded from above by

$$2h^{1+1/r} ||f||_p^2 ||K||_r \int_F \left\{ \int_{\mathbf{R}} \left(\int_F \frac{|K_h\,(t-x)\,|}{h} |\varphi(t)|dt \right)^2 \frac{|K_h\,(s-x)\,|}{h} dx \right\} ds. \tag{2.9}$$

Finally we note that, by Fubini,

$$\int_F \left\{ \int_{\mathbf{R}} \left(\int_F \frac{|K_h\,(t-x)\,|}{h} |\varphi(t)|dt \right)^2 \frac{|K_h\,(s-x)\,|}{h} dx \right\} ds$$

$$= \int_{\mathbf{R}} \left(\int_F \frac{|K_h\,(t-x)\,|}{h} |\varphi(t)| dt \right)^2 dx ||K||_1,$$

which by Cauchy–Schwarz and then Fubini is not larger than

$$\int_{\mathbf{R}} \left(\int_F \frac{|K_h\,(t-x)\,|}{h} \varphi^2(t) dt \right) dx ||K||_1^2 = ||K||_1^3 ||\varphi||_2^2.$$

Therefore the bound in (2.9) is dominated by

$$2h^{1/r+1} ||\varphi||_2^2 ||f||_p^2 ||K||_r ||K||_1^3.$$

Putting (2.7) together with this bound for (2.9) we get

$$\begin{aligned} ||\mathcal{R}_h \varphi||_2^2 &\leq 2h^{1/r+1} ||f||_p^2 ||K||_r ||K||_1^3 ||\varphi||_2^2 + 2h^2 ||\varphi||_2^2 \left(||f||_2^2 \, \|K\|_1^2 \right)^2 \\ &\leq h^{1/r+1} ||\varphi||_2^2 \left[2||f||_p^2 ||K||_r ||K||_1^3 + 2 \left(||f||_2^2 \, \|K\|_1^2 \right)^2 \right], \end{aligned}$$

that is, inequality (2.4). $\square$

Set

$$C_h(s,t) := h^{-1} \int_{\mathbf{R}} \left[K_h\,(s-x)\, K_h\,(t-x) \right] f(x)\; dx.$$

2.4. Lemma.

$$\lim_{h \to 0} h^{-1} \int_{F^2} C_h^2(s,t)\; ds\; dt = \int_F f^2(x)\; dx \int_{\mathbf{R}} \left(\int_{\mathbf{R}} K(w+u) K(w)\; dw \right)^2\; du.$$

Proof. We get

$$\begin{aligned} &h^{-1} \int_{F^2} C_h^2(s,t)\; ds\; dt = h^{-3} \int_{F^2} \left\{ \int_{\mathbf{R}} \left[K\left(\frac{s-x}{h}\right) K\left(\frac{t-x}{h}\right) \right] f(x)\; dx \right\}^2 ds\; dt \\ &= h^{-3} \int_{F^2} \left\{ \int_{\mathbf{R}^2} K\left(\frac{s-x}{h}\right) K\left(\frac{t-x}{h}\right) K\left(\frac{s-y}{h}\right) K\left(\frac{t-y}{h}\right) f(x) f(y) dx dy \right\} ds dt, \end{aligned}$$

which, by setting $y = x - hu$, equals

$$\frac{1}{h^2} \int_{F^2} \int_{\mathbf{R}^2} K\left(\frac{s-x}{h}\right) K\left(\frac{t-x}{h}\right) K\left(\frac{s-x}{h} + u\right) K\left(\frac{t-x}{h} + u\right) f(x) f(x-hu) dx du ds dt.$$

Next, we set $z = \frac{s-x}{h}$ and $w = \frac{t-x}{h}$ and the above becomes

$$= \int_{\mathbf{R}^2} \left\{ \int_{\mathbf{R}} \left[\int_{\mathbf{R}} K\,(z)\, K\,(w)\, 1_h(z,w,x) K\,(z+u)\, K\,(w+u)\, f(x-hu) du \right] f(x) dx \right\} dz dw,$$

where
$$1_h(z,w,x) := 1\{zh + x \in F\}\, 1\{wh + x \in F\}.$$
Let us set
$$G_h(z,w) := \int_{\mathbf{R}} \left[\int_{\mathbf{R}} K(z)\, K(w)\, 1_h(z,w,x) K(z+u)\, K(w+u)\, f(x-hu) du \right] f(x) dx,$$
$$K(z,w) = K(z)K(w) \int_{\mathbf{R}} K(z+u)\, K(w+u)\, du$$
and
$$G(z,w) = K(z,w) \int_F f^2(x) dx.$$
It follows from the definitions that
$$h^{-1} \int_{F^2} C_h^2(s,t) ds dt = \int_{\mathbf{R}^2} G_h(z,w) dz dw.$$
Claim. For all $(z,w) \in \mathbf{R}^2$,
$$G_h(z,w) \to G(z,w) \quad \text{as} \quad h \searrow 0 \tag{2.10}$$
and
$$|G_h(z,w)| \leq |K(z)K(w)| ||f||_2^2 ||K||_2^2. \tag{2.11}$$
Proof of the Claim. First we consider (2.10). We have
$$\Bigg| \int_{\mathbf{R}} \left[\int_{\mathbf{R}} K(z)K(w)K(z+u)\, K(w+u)\, f(x-hu) du \right] 1_h(z,w,x) f(x) dx \\ - K(z,w) \int_{\mathbf{R}} 1_h(z,w,x) f^2(x) dx \Bigg|$$
$$= \left| \int_{\mathbf{R}} \left[\int_{\mathbf{R}} K(z)K(w)K(z+u)\, K(w+u)\, (f(x-hu) - f(x))\, du \right] 1_h(z,w,x) f(x) dx \right|$$
$$\leq |K(z)K(w)| ||f||_2 \left(\int_{\mathbf{R}} \left[\int_{\mathbf{R}} K(z+u)\, K(w+u)\, (f(x-hu) - f(x))\, du \right]^2 dx \right)^{1/2},$$
which by Lemma 2.1 converges to zero. Now by hypothesis (1.8) on F,
$$\int_{\mathbf{R}} 1_h(z,w,x) f^2(x) dx \to \int_F f^2(x) dx,$$
which completes the proof of (2.10).

Now we turn to (2.11):

$$\begin{aligned}|G_h(z,w)| &\le |K(z)K(w)| \int_{\mathbf{R}}\int_{\mathbf{R}} |K(z+u)\,K(w+u)|\, f(x-hu)f(x)dxdu \\ &\le |K(z)K(w)| \int_{\mathbf{R}} |K(z+u)\,K(w+u)| \int_{\mathbf{R}} 2^{-1}\left[f^2(x-hu)+f^2(x)\right] dxdu \\ &\le ||f||_2^2 |K(z)K(w)| \int_{\mathbf{R}} |K(z+u)\,K(w+u)|\, du \\ &\le |K(z)K(w)| ||f||_2^2 ||K||_2^2,\end{aligned}$$

and the Claim is proved.

Lemma 2.4 now follows from the claim and the Lebesgue dominated convergence theorem. □

2.5. Lemma.

$$\lim_{h\to 0} h^{-1} \int_{F^2} (C_h(s,t)-R_h(s,t))^2\, dsdt \to 0.$$

Proof. Note that

$$(C_h(s,t)-R_h(s,t))^2 = h^2 \mu_h^2(s)\mu_h^2(t),$$

where $\mu_h(s) = h^{-1}\int_{\mathbf{R}} K_h(s-x)\, f(x)dx$ has been defined in the proof of Lemma 2.3. Hence, by inequality (2.6),

$$\int_{\mathbf{R}^2} (C_h(s,t)-R_h(s,t))^2\, dsdt \le h^2 ||f||_2^4\, \|K\|_1^4 .$$

□

Lemmas 2.4 and 2.5 prove the limit (2.5) thus completing the proof of Proposition 2.2.

In particular, coming back to (1.7), we have shown:

2.6. Corollary. *Under the hypotheses of Proposition 2.2,*

$$\lim_{n\to\infty} \frac{1}{n^2 h_n^3} EU_n^2(F) = \int_F f^2(x)dx \int_{\mathbf{R}} \left(\int_{\mathbf{R}} K(w+u)K(w)dw\right)^2 du. \tag{2.12}$$

and there exists $h_0 = h_0(F) > 0$ *such that*

$$E\left\{\int_F \left[\bar{K}_h(t-X_1)\,\bar{K}_h(t-X_2)\right] dt\right\}^2 \le 2h^3 ||K||_1^2 ||K||_2^2 \int_F f^2(x)dx \tag{2.13}$$

for all $0 < h \le h_0$.

Proof. The limit (2.12) follows from the limit (2.5) by noting that, by Fubini,

$$E\left\{\int_F \left[\bar{K}_h(t-X_1)\bar{K}_h(t-X_2)\right] dt\right\}^2 = h^2 \int_{F^2} R_h^2(s,t)dsdt.$$

Inequality (2.13) follows from the limit (2.12) because, by Hölder and Fubini,

$$\int_{\mathbf{R}}\left(\int_{\mathbf{R}} K(w+u)K(w)dw\right)^2 du$$
$$\leq \int_{\mathbf{R}}\int_{\mathbf{R}}\left(\int_{\mathbf{R}} K^2(w+u)du\right)^{1/2}\left(\int_{\mathbf{R}} K^2(s+u)du\right)^{1/2}|K(w)||K(s)|dwds$$
$$= \|K\|_1^2\|K\|_2^2.$$

□

We now consider the variance of the linear term L_n in (1.7). We first observe that

$$\begin{aligned}\mathrm{Var}\left[\int_F \bar{K}_h^2(t-X)dt\right] \leq\ & 4\int_F\int_F E\left[K_h^2(t-X)K_h^2(s-X)\right]dsdt\\ &+8\int_F\int_F\left[\Big(EK_h(t-X)\Big)^2 EK_h^2(s-X)\right]dsdt\\ &+4\left(\int_F\Big(EK_h(t-X)\Big)^2 dt\right)^2.\end{aligned}$$

With the change of variables $x=x$, $w=(t-x)/h$ and $z=(s-x)/h$, we see that the first integral at the right hand side of the last inequality equals

$$h^2\int_{\mathbf{R}^3}\left[K^2(w)K^2(z)1_h(z,w,x)f(x)\right]dxdwdz.$$

By condition (1.8),

$$\limsup_{h\to 0}\int_{\mathbf{R}} 1_h(z,w,x)f(x)dx \leq \int_F f(x)dx$$

so that, by Fatou,

$$\limsup_{h\to 0}\int_{\mathbf{R}^2} K^2(w)K^2(z)\left[\int_{\mathbf{R}} 1_h(z,w,x)f(x)dx\right]dwdz \leq \|K\|_2^4\int_F f(x)dx,$$

where we recall that $1_h(z,w,x)=1\{zh+x\in F\}1\{wh+x\in F\}$. Hence, for all h small enough (depending on F),

$$\int_F\int_F E\left[K_h^2(t-X)K_h^2(s-X)\right]dsdt \leq 2h^2\|K\|_2^4\int_F f(x)dx.$$

By Lemma 2.1, $h^{-1}E|K((t-X)/h)| \to \|K\|_1 f(t)$ and $h^{-1}EK^2((t-X)/h)\to\|K\|_2^2 f(t)$ in $L_r(F)$ for $1\leq r\leq p$, in particular for $r=1$ and $r=2$. This allows us to bound the other two summands in the above inequality to the effect that, for all h small enough (depending on F),

$$\int_F\int_F\left[\Big(EK_h(t-X)\Big)^2 EK_h^2(s-X)\right]dsdt \leq 2h^3\|K\|_1^2\|K\|_2^2\int_F f^2(t)dt\int_F f(x)dx$$

and

$$\left(\int_F\Big(EK_h(t-X)\Big)^2 dt\right)^2 \leq 2h^4\|K\|_1^4\left(\int_F f^2(x)dx\right)^2.$$

Thus, these estimates, Corollary 2.6, (1.7) and the fact that $h_n\to 0$ and $nh_n\to\infty$, give the following:

2.7. Corollary. *Under the hypotheses of Proposition 2.2, there exists $h_0' = h_0'(F)$ such that*

$$\mathrm{Var}\left[\int_F \bar{K}_h^2(t-X)dt\right] \le 9h^2\|K\|_2^4 \int_F f(x)dx \tag{2.14}$$

for all $0 < h \le h_0$. In particular,

$$\lim_{n\to\infty} \frac{1}{n^2h_n^3} EW_n^2(F) = \int_F f^2(x)dx \int_{\mathbf{R}} \left(\int_{\mathbf{R}} K(w+u)K(w)dw\right)^2 du. \tag{2.15}$$

We conclude this section with two easy estimates for the sup norm of the general summands in U_n and L_n that will be useful later on, namely that for all $h > 0$, x and y, we have both,

$$\left|\int_F \left[\bar{K}_h(t-x)\bar{K}_h(t-y)\right] dt\right| \le 4h\|K\|_2^2 \tag{2.16}$$

and

$$\left|\int_F \bar{K}_h^2(t-x)dt - E\int_F \bar{K}_h^2(t-X)dt\right| \le 8h\|K\|_2^2. \tag{2.17}$$

Thse estimates follow easily from the fact that, by Hölder, for all $x \in \mathbf{R}$,

$$\int_{\mathbf{R}} \bar{K}_h^2(t-x)dt \le 4h\|K\|_2^2. \tag{2.18}$$

3. Simplifying the problem: restriction of the domain of integration and blocking. In this section we obtain exponential tail estimates both for $W_n(F)$, $F = [-M, M]^c$, and for $W_n(\mathbf{R}) - W_{n,m}(\mathbf{R})$, $0 \le m < n$, where

$$W_{n,m}(\mathbf{R}) := \int_{\mathbf{R}} \left[\Big(\sum_{m<i\le n} \bar{K}_{h_n}(t-X_i)\Big)^2 - E\Big(\sum_{m<i\le n} \bar{K}_{h_n}(t-X_i)\Big)^2\right] dt. \tag{3.1}$$

The derivations are similar. To simplify notation, set

$$H_h(x,y) := \int_{\mathbf{R}} \left[\bar{K}_h(t-x)\bar{K}_h(t-y)\right] dt, \quad H_{h,F}(x,y) := \int_F \left[\bar{K}_h(t-x)\bar{K}_h(t-y)\right] dt, \tag{3.2}$$

for all $x, y \in \mathbf{R}$, $h > 0$ and any measurable set $F \subseteq \mathbf{R}$, and write H_n (resp. $H_{n,F}$) for H_{h_n} (resp. $H_{h_n,F}$). Then the variables U_n and L_n from (1.7) become

$$U_n(F) = \sum_{1\le i\ne j\le n} H_{n,F}(X_i,X_j), \quad L_n(F) = \sum_{i=1}^n \big(H_{n,F}(X_i,X_i) - EH_{n,F}(X_i,X_i)\big). \tag{3.3}$$

In analogy with the decomposition (1.7), we also have

$$W_n(\mathbf{R}) - W_{n,m}(\mathbf{R}) = 2\sum_{i=1}^m \sum_{j=m+1}^n H_n(X_i,X_j) + \sum_{1\le i\ne j\le m} H_n(X_i,X_j) + \sum_{i=1}^m \big(H_n(X_i,X_i) - EH_n(X_i,X_i)\big), \tag{3.4}$$

where the first two summands are of U-statistic type and the third is linear (a sum of centered i.i.d. random variables).

The linear terms in (1.7) and (3.4) are easy to control by Bernstein's inequality (e.g., de la Peña and Giné, 1999, p. 167), given Corollary 2.7 and the bound (2.17): under the hypotheses of Proposition 2.2, for all $\tau > 0$, n large enough and $0 \le m < n$,

$$\Pr\left\{\Big|\sum_{i=1}^{m}\big(H_n(X_i,X_i) - EH_n(X_i,X_i)\big)\Big| > \tau n h_n^{3/2}\right\}$$
$$\le 2\exp\left(-\frac{\tau^2 n^2 h_n^3}{18 m h_n^2 \|K\|_2^4 + \frac{16}{3}\tau n h_n^{5/2}\|K\|_2^2}\right). \tag{3.5}$$

To get useful bounds for the U-statistic type terms, we will use a recent exponential inequality for canonical U-statistics due to Giné, Latała and Zinn (2000), which we now describe for the particular case of i.i.d. random variables.

Let $Y, Y_i, Y_i^{(1)}, Y_j^{(2)}$, $i,j \in \mathbf{N}$, be i.i.d. random variables taking values on some measurable space $(S, \mathcal{S})$, and let $h_{i,j} : S \times S \mapsto \mathbf{R}$ be bounded canonical random variables for the law of Y, that is, $E_{Y_1} h_{i,j}(Y_1,Y_2) = E_{Y_2} h_{i,j}(Y_1,Y_2) = 0$, where E_{Y_1} (resp. E_{Y_2}) denote integration with respect to the variable Y_1 (resp. Y_2) only. Then:

3.1. Theorem. (Giné, Latała and Zinn, 2000) *There exists a universal constant $L < \infty$ such that, if A, B, C, D are as defined below, then*

$$\Pr\left\{\Big|\sum_{1\le i,j\le n} h_{i,j}(Y_i^{(1)}, Y_j^{(2)})\Big| \ge x\right\} \le L\exp\left[-\frac{1}{L}\min\left(\frac{x^2}{C^2}, \frac{x}{D}, \frac{x^{2/3}}{B^{2/3}}, \frac{x^{1/2}}{A^{1/2}}\right)\right] \tag{3.6}$$

for all $x > 0$. Moreover the same inequality holds for the un-decoupled U-statistic $\sum_{1\le i\ne j\le n} h_{i,j}(Y_i, Y_j)$. Here,

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

$$C^2 = \sum_{i,j} E h_{i,j}^2(Y_i, Y_j),$$

$$B^2 = \max_{i,j}\left[\Big\|\sum_i E h_{i,j}^2(Y_i^{(1)}, y)\Big\|_\infty, \Big\|\sum_j E h_{i,j}^2(x, Y_j^{(2)})\Big\|_\infty\right]$$

and

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

As indicated in Giné, Latała and Zinn (2000), but not precisely stated there, inequality (3.6) for $\sum_{i\ne j\le n} h_{i,j}(Y_i, Y_j)$ follows by decoupling from the inequality for the decoupled statistic by making $h_{i,i} = 0$, which does not increase the size of the parameters.

If $h_{i,j} = H$ independently of i and j, then the above parameters simplify a bit, and in particular, D and B become, respectively,

$$D = n \ \sup\Big\{EH(Y_1, Y_2)l(Y_1)g(Y_2) : El^2(Y_1) \le 1, \ Eg^2(Y_2) \le 1\Big\},$$

and

$$B^2 = n \ \max\Big[\big\|EH^2(Y, y)\big\|_\infty, \big\|EH^2(x, Y)\big\|_\infty\Big].$$

We will now apply Theorem 3.1 to $h_{i,j} = H_{h,F,i,j} = H_{h,F}$. We already have, from Section 2, bounds on the A and C terms of the inequality. For the D term we have:

3.2. Lemma. *Assume f satisfies (1.1) and $\|K\|_r < \infty$ for $r = 1$ and $r = q$, where where $1/p + 1/q = 1$ with p as in (1.1). Set*

$$\begin{aligned} D &= \|(H_{h,F,i,j})\|_{L^2 \to L^2} \\ &:= n \ \sup\Big\{E\left[H_{h,F}(X_1, X_2)l(X_1)g(X_2)\right] : \ El^2(X_1) \le 1, \ Eg^2(X_2) \le 1\Big\}. \end{aligned}$$

Then,

$$D \le n||f||_p||\bar{K}_h||_1\|\bar{K}_h\|_q \le 4nh^{1+1/q}||f||_p||K||_1\|K\|_q. \tag{3.7}$$

Proof. First we note that, as in (2.18), for $r \ge 1$ and all $x \in \mathbf{R}$,

$$\|\bar{K}_h\|_r^r := \int_{\mathbf{R}} |\bar{K}_h(t-x)|^r dt \le 2^{r-1} \left(\int_{\mathbf{R}} |K_h(t-x)|^r dt + E\int_{\mathbf{R}} |K_h(t-x)|^r dt\right) = 2^r h\|K\|_r^r. \tag{3.8}$$

The special form of $H_{h,F}$ implies that

$$D = n \ \sup\bigg\{\int_F \Big[E\big(\bar{K}_h(t - X)\varphi(X)\big)\Big]^2 dt : \ E\varphi^2(X) \le 1\bigg\}.$$

Then, if $E\varphi^2(X) \le 1$,

$$\begin{aligned} \int_{\mathbf{R}} \Big[E\big|\bar{K}_h(t - X)\varphi(X)\big|\Big]^2 dt &= \int_{\mathbf{R}} \left(\int_{\mathbf{R}} |\bar{K}_h(t - x)\varphi(x)\big|f(x)dx\right)^2 dt \\ &\le \int_{\mathbf{R}} \left(\int_{\mathbf{R}} |\bar{K}_h(t - x)|f(x)dx \int_{\mathbf{R}} |\bar{K}_h(t - x)|\varphi^2(x)f(x)dx\right) dt \\ &\le \|f\|_p\|\bar{K}_h\|_q \int_{\mathbf{R}}\int_{\mathbf{R}} |\bar{K}_h(t - x)|\varphi^2(x)f(x)dxdt \\ &\le \|f\|_p\|\bar{K}_h\|_q\|\bar{K}_h\|_1. \end{aligned}$$

This inequality, combined with inequality (3.8) for $r = 1$ and $r = q$, gives (3.7). □

If $p = 2$ the above inequality shows that the order of D is at most $nh_n^{3/2}$ but this is not enough for our purposes, as we will see later. On the other hand, if $\|f\|_\infty < \infty$ then the bound above gives an order of at most nh_n^2 for D. Any power of h_n larger than 3/2 is useful below.

Note that since $1 < q < 2$ in Lemma 3.2, the right hand side of inequality (3.7) is finite if (1.1) holds and $\|K\|_r < \infty$ for $r = 1$ and $r = 2$.

As for the B term, since, by (2.16),

$$EH_{h,F}^2(X,y) = E\left(\int_{\mathbf{F}} \left[\bar{K}_h\left(t - X\right)\bar{K}_h\left(t - y\right)\right] dt\right)^2 \le 16h^2\|K\|_2^4,$$

we have

$$B^2 \le 16nh^2||K||_2^4. \tag{3.9}$$

Collecting the above bounds (2.16), (3.9), (2.13) and (3.7) respectively for A, B, C and D, Theorem 3.1 and inequality (3.5) give:

3.3. Proposition. *Let X_i be i.i.d. with density f satisfying condition (1.1) for some $p > 2$. Let F be a measurable subset of $\mathbf{R}$ satisfying condition (1.8), let K be a measurable kernel such that $\|K\|_1 < \infty$ and $\|K\|_2 < \infty$, let $h_n \to 0$ and let $W_n(F)$ be defined as in (1.7). Then, there exist a constant κ_0 depending only on K and n_0 (depending on F, f, K and $\{h_n\}$) such that for all $\tau > 0$ and for all $n \ge n_0$,*

$$\begin{aligned}&\Pr\Big\{\big|W_n(F)\big| \ge \tau nh_n^{3/2}\Big\} \qquad (3.10)\\ &\quad\le \kappa_0 \exp\bigg(-\frac{1}{\kappa_0}\min\bigg[\frac{\tau^2}{\int_F f^2(x)dx}, \frac{\tau}{h_n^{1/q-1/2}}, (\tau^2 nh_n)^{1/3}, (\tau nh_n^{1/2})^{1/2}, \tau^2 nh_n, \tau nh_n^{1/2}\bigg]\bigg),\end{aligned}$$

where q is the conjugate of p. In particular, if the sequence h_n satisfies condition (1.3) for some $0 < \delta < 1$, then, for every $\eta > 0$ there exist κ_0 and n_0 as above such that

$$\Pr\left\{\big|W_n(F)\big| \ge \eta\sqrt{\log\log n}\; nh_n^{3/2}\right\} \le \kappa_0 \exp\left(-\frac{\eta^2\log\log n}{\kappa_0\int_F f^2(x)dx}\right) \tag{3.11}$$

for all $n \ge n_0$.

Similarly,

3.4. Proposition. *Let X_i be i.i.d. with density f satisfying condition (1.1) for some $p > 2$, let K be a measurable kernel such that $\|K\|_1 < \infty$ and $\|K\|_2 < \infty$, and let $h_n \to 0$. Then, there exist a constant κ_0 depending only on K and n_0 (depending on f, K and $\{h_n\}$) such that for all $\tau > 0$ and for all $n \ge n_0$, $0 \le m < n$,*

$$\begin{aligned}&\Pr\Big\{\Big|\sum_{1\le i\ne j\le m} H_n(X_i, X_j)\Big| \ge \tau nh_n^{3/2}\Big\} \qquad (3.12)\\ &\qquad\le \kappa_0 \exp\bigg(-\frac{1}{\kappa_0}\min\bigg[\frac{\tau^2 n^2}{m^2}, \frac{\tau n}{mh_n^{1/q-1/2}}, \Big(\frac{\tau^2 n^2 h_n}{m}\Big)^{1/3}, (\tau nh_n^{1/2})^{1/2}\bigg]\bigg),\end{aligned}$$

and

$$\Pr\Big\{\Big|\sum_{i=1}^{m}\sum_{j=m+1}^{n} H_n(X_i,X_j)\Big| \geq \tau n h_n^{3/2}\Big\} \tag{3.13}$$
$$\leq \kappa_0 \exp\bigg(-\frac{1}{\kappa_0}\min\bigg[\frac{\tau^2 n^2}{m(n-m)}, \frac{\tau n}{\sqrt{m(n-m)}h_n^{1/q-1/2}}, \Big(\frac{\tau^2 n^2 h_n}{m\vee(n-m)}\Big)^{1/3}, (\tau n h_n^{1/2})^{1/2}\bigg]\bigg).$$

Proposition 3.4 together with inequality (3.5) cover the three terms in the decomposition (3.4) of $W_n(\mathbf{R}) - W_{n,m}(\mathbf{R})$.

4. Moderate deviations. The object of this section is to derive a moderate deviation result for $W_n([-M,M])$. First we approximate this statistic by a diagonalizable Gaussian chaos of order two as a consequence of the KMT approximation, and then, essentially following Pinsky (1966), we derive a moderate deviation result for the approximating Gaussian chaos.

4.1. Using KMT. Let X, $X_1, X_2, \ldots,$ be a sequence of independent and identically distributed random variables in $\mathbf{R}$ with common Lebesgue density f. For each integer $n \geq 1$ let

$$F_n(t) = n^{-1}\sum_{i=1}^{n} 1\{X_i \leq t\}, \quad -\infty < t < \infty, \tag{4.1}$$

denote the empirical distribution function based on $X_1, ..., X_n$, and

$$\alpha_n(t) = \sqrt{n}[F_n(t) - F(t)], \quad -\infty < t < \infty, \tag{4.2}$$

be the corresponding empirical process. Komlós, Major, and Tusnády [KMT] (1975) proved the following Brownian bridge approximation to α_n.

4.1. Theorem. (KMT, 1975). *There exists a probability space $(\Omega, \mathcal{A}, P)$ with independent identically distributed 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 $x \in \mathbf{R}$,*

$$\Pr\Big\{D_n \geq n^{-1/2}(a\log n + x)\Big\} \leq b\exp(-cx), \tag{4.3}$$

where

$$D_n = \sup_{-\infty<t<\infty} |\alpha_n(t) - B_n(F(t))| \tag{4.4}$$

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

Here we assume K satisfies conditions (1.2), in particular, K is of finite variation, and h_n satisfies conditions (1.3). With the notation from the Introduction, we see by integrating by parts that for all $x \in \mathbf{R}$,

$$E_n(x) := \sqrt{nh_n}\,[f_{n,K}(x) - Ef_{n,K}(x)] = \sqrt{n/h_n}\int_{\mathbf{R}} K\left(\frac{x-t}{h_n}\right) d\,[F_n(t) - F(t)]$$
$$= \sqrt{n/h_n}\int_{\mathbf{R}} [F(t) - F_n(t) - (F(x) - F_n(x))]\, dK\left(\frac{x-t}{h_n}\right).$$

Thus on the probability space of the KMT theorem we have, uniformly in $x \in \mathbf{R}$,

$$\left| \sqrt{nh_n} \left[f_{n,K}(x) - Ef_{n,K}(x) \right] - h_n^{-1/2} \int_{\mathbf{R}} \left[B_n\left(F(x)\right) - B_n\left(F(t)\right) \right] dK\left(\frac{x-t}{h_n} \right) \right|$$
$$\leq \frac{2D_n}{\sqrt{h_n}} ||K||_v.$$

Define the Gaussian process

$$\begin{aligned} \Gamma_n(x) &:= -h_n^{-1/2} \int_{\mathbf{R}} \left[B_n\left(F(t)\right) - B_n\left(F(x)\right) \right] dK\left(\frac{x-t}{h_n} \right) \\ &= h_n^{-1/2} \int_{\mathbf{R}} K\left(\frac{x-t}{h_n} \right) dB_n\left(F(t)\right). \end{aligned}$$

Eventually we will be deriving a moderate deviation result for

$$\frac{1}{nh_n^{3/2}} W_n([-M, M]) = \frac{1}{\sqrt{h_n}} \int_{-M}^{M} \left[\left(E_n(t)\right)^2 - E\left(E_n(t)\right)^2 \right] dt$$

from one for

$$\frac{1}{\sqrt{h_n}} \int_{-M}^{M} \left[\left(\Gamma_n(t)\right)^2 - E\left(\Gamma_n(t)\right)^2 \right] dt.$$

Therefore we will need to control the size of the following difference:

$$\begin{aligned} D_n(M) &= \frac{1}{\sqrt{h_n}} \left| \int_{-M}^{M} \left[\left(E_n(t)\right)^2 - E\left(E_n(t)\right)^2 \right] dt - \int_{-M}^{M} \left[\left(\Gamma_n(t)\right)^2 - E\left(\Gamma_n(t)\right)^2 \right] dt \right| \\ &= \frac{1}{\sqrt{h_n}} \left| \int_{-M}^{M} \left[\left(E_n(t)\right)^2 - \left(\Gamma_n(t)\right)^2 \right] dt \right| \\ &\leq \frac{4M||K||_v}{h_n} D_n \sup_x \left(|E_n(x)| + |\Gamma_n(x)| \right) \\ &\leq \frac{8M||K||_v^2}{h_n^{3/2}} D_n \left(||\alpha_n||_\infty + ||B_n||_\infty \right), \end{aligned} \tag{4.5}$$

where the last bound follows because, obviously,

$$\sup_x \left(|E_n(x)| + |\Gamma_n(x)| \right) \leq \frac{2||K||_v}{\sqrt{h_n}} \left(\|\alpha_n\|_\infty + \|B_n\|_\infty \right).$$

The Dvoretzky-Kiefer-Wolfowitz inequalities (e.g., Shorack and Wellner (1986), page 354), namely

$$\Pr\left(\|\alpha_n\|_\infty > z \right) \leq 2\exp(-2z^2) \quad \text{and} \quad \Pr\left(\|B_n\|_\infty > z \right) \leq 2\exp(-2z^2), \quad z > 0,$$

together with inequality (4.5) and the KMT inequality (4.3) readily imply:

4.2. Proposition. *Assuming K satisfies conditions (1.2) and $\{h_n\}$ satisfies conditions (1.3), for any $\gamma > 0$ there exists $c > 0$ such that*

$$\Pr\left\{D_n(M) \geq \frac{c\,(\log n)^2}{h_n^{3/2}\sqrt{n}}\right\} < n^{-\gamma}. \tag{4.6}$$

Regarding the Gaussian process $\Gamma_n(x)$, it is easily checked that $E\Gamma_n(x) = 0$ for all x and that

$$R_n(x,y) = E\left[E_n(x)E_n(y)\right] = E\left[\Gamma_n(x)\Gamma_n(y)\right],$$

with $R_n(x,y) = R_{h_n}(x,y)$ defined as in (2.2). Then, since by (2.18)

$$E\left(\int_F \Gamma_n^2(s)ds\right) = \int_F R_n(s,s)ds \leq 4\|K\|_2^2 < \infty,$$

it follows that the Gaussian process $\Gamma_n(t)$ has a version with all its sample paths in $L_2(F)$. The following well known fact about $L_2(F)$-valued Gaussian processes will be needed below.

4.3. Proposition. *A centered non-degenerate Gaussian process $\{\Gamma(t), t \in F\}$, F a Borel subset of $\mathbf{R}$, with covariance function*

$$R(s,t) = E\left(\Gamma(t)\Gamma(s)\right), \quad s,t \in F,$$

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

$$0 < \int_F R(s,s)\,ds < \infty.$$

If this is the case, then

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

and the spectrum of the operator

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

consists of a sequence of nonnegative eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ in ℓ_1, corresponding to eigenvectors $e_1, e_2, \ldots$, 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,

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

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

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

$$\sum_{k=1}^{\infty} \lambda_k = \int_F R(s,s)\, ds \quad \text{and} \quad \sum_{k=1}^{\infty} \lambda_k^2 = \int_{F^2} R^2(s,t)\, dsdt.$$

Proof. (Sketch) The condition $\int_F R(s,s)ds = \int_F E\Gamma^2(s)ds < \infty$ is clearly sufficient for Γ to have a version with almost all its sample paths in L_2, and it is necessary by the Fernique-Landau-Shepp integrability theorem (e.g., Fernique (1970)). The second condition, $\int_{F^2} (E\Gamma(s)\Gamma(t))^2\, dsdt < \infty$, follows from the previous one by Cauchy-Schwarz. Then, the operator $\mathcal{R}$ is positive semidefinite Hilbert-Schmidt (actually, trace class) and its spectrum consists of a sequence of non-negative eigenvalues λ_i with orthogonal eigenfunctions e_i such that $R(s,t) = \sum_i \lambda_i e_i(s) e_i(t)$ in the L_2 sense (e.g., Dunford and Schwartz (1963), exercises 44 and 56 on pages 1083 and 1087). The rest of the statements are now easily verified. $\square$

As a last step in the derivation of a moderate deviation result for the statistic J_n we are thus left with the estimation of the tail probabilities of random variables of the form $\sum_{k=1}^{\infty} \lambda_k \left(Z_k^2 - 1\right)$ where Z_k are i.i.d. $N(0,1)$ and $\sum \lambda_k < \infty$.

4.2. A modification of a moderate deviation result of Pinsky (1966). Let Y, Y_1, $Y_2, ...,$ be a sequence of i.i.d. random variables with mean zero, variance 1 and finite absolute $2+\eta$ moment with $0 < \eta \le 1$. Let $\lambda_{n,1}$, $\lambda_{n,2}, ...,$ be a sequence of constants indexed by $n \ge 1$ such that

$$|\lambda_{n,1}| \ge |\lambda_{n,2}|, ..., \quad \text{for all} \quad n \ge 1 \tag{4.7}$$

and for all $n \ge 1$,

$$0 < \Delta_n^2 := \sum_{k=1}^{\infty} \lambda_{n,k}^2 < \infty. \tag{4.8}$$

Set for each $n \ge 1$,

$$S_n := \frac{1}{\Delta_n} \sum_{k=1}^{\infty} \lambda_{n,k} Y_k. \tag{4.9}$$

The following lemma follows by application of the classical Lindeberg method.

4.4. Lemma. *Let g be a function with three bounded continuous derivatives. Then*

$$|Eg(S_n) - Eg(Z)| \le C||g^*|| \left(\frac{|\lambda_{n,1}|}{\Delta_n}\right)^{\eta} \left[E|Y|^{2+\eta} + E|Z|^{2+\eta}\right],$$

where Z is a standard normal random variable, $\|g^*\| := \|g''\|_\infty + \|g'''\|_\infty$ and C is a constant that only depends on η.

Proof: Let $Z_1, Z_2, ...,$ be a sequence of independent standard normal random variables. Set $\lambda_{0,1} = Y_0 = Z_0 = 0$. We see that

$$\begin{aligned} E[g(S_n) - g(Z)] &= Eg\left(\frac{1}{\Delta_n}\sum_{k=1}^{\infty} \lambda_{n,k} Y_k\right) - Eg\left(\frac{1}{\Delta_n}\sum_{k=1}^{\infty} \lambda_{n,k} Z_k\right) \\ &= \sum_{m=1}^{\infty} \left\{ Eg\left(B_{m,n} + \frac{\lambda_{n,m} Y_m}{\Delta_n}\right) - Eg\left(B_{m,n} + \frac{\lambda_{n,m} Z_m}{\Delta_n}\right)\right\} =: \sum_{m=1}^{\infty} A_m, \end{aligned}$$

where

$$B_{m,n} := \frac{1}{\Delta_n}\sum_{k=0}^{m-1}\lambda_{n,k}Z_k + \frac{1}{\Delta_n}\sum_{k=m+1}^{\infty}\lambda_{n,k}Y_k.$$

Using the Taylor estimate

$$\left|g(x+y) - g(x) - yg'(x) - \frac{y^2}{2}g''(x)\right| \leq |y|^{2+\eta}||g^*||,$$

we get

$$\begin{aligned}\sum_{m=1}^{\infty}|A_m| &\leq \|g^*\|\sum_{m=1}^{\infty}\left[E\left|\frac{\lambda_{n,m}Y_m}{\Delta_n}\right|^{2+\eta} + E\left|\frac{\lambda_{n,m}Z_m}{\Delta_n}\right|^{2+\eta}\right] \\ &\leq \|g^*\|\left[E|Y|^{2+\eta} + E|Z|^{2+\eta}\right]\sum_{m=1}^{\infty}\left|\frac{\lambda_{n,m}}{\Delta_n}\right|^{2+\eta},\end{aligned}$$

which, since $\sum_{m=1}^{\infty}\lambda_{n,m}^2/\Delta_n^2 = 1$, is

$$\leq ||g^*||\left[E|Y|^{2+\eta} + E|Z|^{2+\eta}\right]\left(\frac{|\lambda_{n,1}|}{\Delta_n}\right)^{\eta}.$$

□

Remark. In Section 6 we will make use of the following fact, whose proof differs only formally from the proof of Lemma 4.4: Assume that for each n, Y_n, $Y_{1,n}, Y_{2,n}, ..., Y_{n,n}$ are i.i.d. random variables with mean zero and variance 1 such that, for some $0 < \eta \leq 1$,

$$M_\eta := \sup_{n\geq 1} E|Y_n|^{2+\eta} < \infty.$$

Then, setting $S_n = \sum_{i=1}^{n} Y_{i,n}/\sqrt{n}$, we have, for all n and for any g as in Lemma 4.4,

$$|Eg(S_n) - Eg(Z)| \leq \frac{C\|g^*\|\left[M_\eta + E|Z|^{2+\eta}\right]}{n^{\eta/2}},$$

in the notation of that lemma.

Set, with $0 < \eta \leq 1$,

$$b_n = \left(\frac{|\lambda_{n,1}|}{\Delta_n}\right)^{\eta}. \tag{4.10}$$

4.5. Theorem. *Assume that $b_n \to 0$ as $n \to \infty$. Then, for any sequence a_n converging to infinity at the rate $a_n^2 + \log b_n \to -\infty$,*

$$\exp\left(-\frac{a_n^2}{2}(1+\varepsilon)\right) \leq \Pr\left(S_n \geq a_n\right) \leq \exp\left(-\frac{a_n^2}{2}(1-\varepsilon)\right) \tag{4.11}$$

for all $0 < \varepsilon < 1$ *and for all* n *sufficiently large depending on* ε.

Proof: Let g be any function on $\mathbf{R}$ with three bounded continuous derivatives satisfying $g(x) = 0$ for $x \le -1/2$, $0 \le g(x) \le 1$ for $x \in (-1/2, 1/2)$, and $g(x) = 1$ for $x \ge 1/2$. For example we could use

$$g(x) = \begin{cases} 1 & \text{for } x \ge 1/2, \\ \exp\left(-\left(\frac{\frac{1}{2}-x}{\frac{1}{2}+x}\right)^4\right) & \text{for } x \in (-1/2, 1/2), \\ 0 & \text{for } x \le -1/2. \end{cases}$$

We then see that

$$Eg\left(S_n - a_n - 1/2\right) \le P\left\{S_n \ge a_n\right\} \le Eg\left(S_n - a_n + 1/2\right),$$

and applying Lemma 4.4 we get that for some constant $C > 0$

$$Eg\left(Z - a_n - 1/2\right) - Cb_n \le \Pr\left\{S_n \ge a_n\right\} \le Eg\left(Z - a_n + 1/2\right) + Cb_n,$$

from which we readily obtain

$$\Pr\{Z \ge a_n + 1\} - Cb_n \le \Pr\left\{S_n \ge a_n\right\} \le \Pr\{Z \ge a_n - 1\} + Cb_n.$$

Now $-\log\Pr\{Z \ge a_n \pm 1\} = \frac{a_n^2}{2}\left(1 + o(1)\right)$ and by assumption $b_n/P\{Z \ge a_n \pm 1\} \to 0$. Thus we conclude (4.11). $\square$

We are interested in the following special case of Theorem 4.5. Set for each $n \ge 1$,

$$V_n := \frac{1}{\sqrt{2}\Delta_n}\sum_{k=1}^{\infty}\lambda_{n,k}\left(Z_k^2 - 1\right). \tag{4.12}$$

4.6. Corollary. *Assume that* $b_n \to 0$ *as* $n \to \infty$. *Then, for any sequence* a_n *converging to infinity at the rate* $a_n^2 + \log b_n \to -\infty$,

$$\exp\left(-\frac{a_n^2}{2}\left(1+\varepsilon\right)\right) \le \Pr\left\{\pm V_n \ge a_n\right\} \le \exp\left(-\frac{a_n^2}{2}\left(1-\varepsilon\right)\right) \tag{4.13}$$

for all $0 < \varepsilon < 1$ *and for all* n *sufficiently large depending on* ε.

Applying Proposition 4.3 to the Gaussian process $\{\Gamma_n(x) : x \in [-M, M]\}$, where $M > 0$, we get that

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

where $\lambda_{n,1} \ge \lambda_{n,2} \ge \ldots \ge 0$ are the eigenvalues of the operator $\mathcal{R}_{h_n}$ on $L_2\left([-M, M]\right)$ defined by R_{h_n}. We recall that by Proposition 2.2,

$$\sup\left\{\|\mathcal{R}_h^2\varphi\|_2^2 : \|\varphi\|_2 = 1, \varphi \in L_2([-M, M])\right\} \le C(K, f, r, p)h^{1+1/r},$$

where p is given by condition (1.1), the constant is finite and $r = r(p) > 0$, which implies that for all large enough n,

$$\lambda_{n,1} \leq C^{1/2}(K, f, r, p) h^{1/2+1/(2r)}.$$

Moreover, by Propositions 2.2 and 4.3,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{h_n} E\Big[\int_{-M}^{M} \Big[(\Gamma_n(t))^2 - E(\Gamma_n(t))^2\Big]dt\Big]^2 &= \lim_{n\to\infty} \frac{2}{h_n} \sum_{k=1}^{\infty} \lambda_{n,k}^2 \\
&= \lim_{n\to\infty} \frac{2}{h_n} \int_{-M}^{M}\int_{-M}^{M} R_n^2(s,t)dsdt \\
&= 2\int_{-M}^{M} f^2(x)dx \int_{\mathbf{R}} \left(\int_{\mathbf{R}} K(w+u)K(w)dw\right)^2 du \\
&=: \sigma^2(M).
\end{aligned}$$

Set

$$V_n(M) := \frac{1}{\sqrt{h_n}\sigma(M)} \int_{-M}^{M} \Big[(\Gamma_n(t))^2 - E(\Gamma_n(t))^2\Big] dt. \tag{4.14}$$

Since

$$\frac{\lambda_{n,1}}{\sqrt{2\sum_{k=1}^{\infty}\lambda_{n,k}^2}} \leq \frac{C^{1/2}(K,f,r,p)h_n^{1/2+1/(2r)}}{\sqrt{2\sum_{k=1}^{\infty}\lambda_{n,k}^2}} \asymp h_n^{1/(2r)} \to 0,$$

we can apply Corollary 4.6 to conclude that whenever a_n converges to infinity at the rate $a_n^2 + \log b_n \to -\infty$,

$$\exp\left(-\frac{a_n^2}{2}(1+\varepsilon)\right) \leq \Pr\Big\{\pm V_n(M) \geq a_n\Big\} \leq \exp\left(-\frac{a_n^2}{2}(1-\varepsilon)\right), \tag{4.15}$$

for all $0 < \varepsilon < 1$ and for all n sufficiently large depending on ε.

If h_n satisfies condition (1.3) then b_n is dominated by a constant times $h_n^{1/(2r)} \asymp n^{-\delta/(2r)}$ and we can obviously take $a_n = \sqrt{2\log\log n}$. Since, for this a_n,

$$\lim_{n\to\infty} \frac{a_n}{c(\log n)^2/(nh_n^3)^{1/2}} = \infty,$$

Proposition 4.2 (that is, the KMT) together with inequality (4.15) (that is, the moderate deviation result for the Gaussian chaos) immediately give:

4.7. Proposition. *Let $a_n = C\sqrt{2\log\log n}$ with $0 < C < \infty$. Assuming K satisfies conditions (1.2), that $\{h_n\}$ satisfies conditions (1.3) and that f satisfies condition (1.1) and $\int_{-M}^{M} f^2(x)dx > 0$, we have*

$$\begin{aligned}
\exp\left(-\frac{a_n^2}{2}(1+\varepsilon)\right) - \frac{1}{n^2} &\leq \Pr\left\{\pm \frac{1}{\sigma(M)nh_n^{3/2}} W_n([-M,M]) \geq a_n\right\} \\
&\leq \exp\left(-\frac{a_n^2}{2}(1-\varepsilon)\right) + \frac{1}{n^2}
\end{aligned} \tag{4.16}$$

for all $0 < \varepsilon < 1$ *and for all* n *sufficiently large depending on* ε.

Define

$$\sigma^2 := \sigma^2(\infty) = 2\int_{\mathbf{R}} f^2(x)dx \int_{\mathbf{R}} \left(\int_{\mathbf{R}} K(w+u)K(w)dw\right)^2 du. \tag{4.17}$$

Since $\sigma(M) \to \sigma$ as $M \to \infty$, we will be able to replace $\sigma(M)$ by σ in (4.16) for M large enough.

5. The LIL for the second moment of the deviation of a kernel density estimator with respect to its mean. Here is the main result of this article:

5.1. Theorem. *Let* f, K *and* $\{h_n\}$ *satisfy hypotheses (1.1)-(1.4), and set*

$$J_n := \left\|f_{n,K} - Ef_{n,K}\right\|_2^2 - E\left\|f_{n,K} - Ef_{n,K}\right\|_2^2,$$

as in (1.6). Set $\sigma^2 := 2\|f\|_2^2 \int_{\mathbf{R}} \left(\int_{\mathbf{R}} K(w+u)K(w)dw\right)^2 du$ *as in (4.17). Then,*

$$\limsup_{n\to\infty} \pm \frac{n\sqrt{h_n}}{\sqrt{2\sigma^2 \log\log n}} J_n = 1, \quad a.s. \tag{5.1}$$

Proof. We decompose the proof into three parts.

5.1.1. The lower bound. We begin by observing that the random variable

$$\limsup_n \frac{W_n(\mathbf{R})}{\sigma n h_n^{3/2}\sqrt{2\log\log n}}$$

is measurable with respect to the tail σ-algebra of the sequence $\{X_i\}$. This follows from the fact that, by Proposition 3.4 and inequality (3.5), given $m < \infty$, there exists $\eta > 0$ and $\kappa_0 < \infty$ such that, for all $\varepsilon > 0$ and all n large enough,

$$\Pr\left\{|W_n(\mathbf{R}) - W_{n,m}(\mathbf{R})| \geq \varepsilon\sigma n h_n^{3/2}\sqrt{2\log\log n}\right\} \leq \kappa_0 \exp\left\{-\frac{\varepsilon^2 n^\eta}{\kappa_0}\right\}, \tag{5.2}$$

and therefore, for every finite m, $|W_n(\mathbf{R}) - W_{n,m}(\mathbf{R})|/\left(\sigma n h_n^{3/2}\sqrt{2\log\log n}\right) \to 0$ a.s. Note that $W_{n,m}$ does not depend on $X_1, \ldots, X_m$. This observation applies as well if we replace W_n by $|W_n|$ or by $-W_n$.

The object here is to prove the lower bound for the LIL, that is, that

$$\limsup_n \frac{W_n(\mathbf{R})}{\sigma n h_n^{3/2}\sqrt{2\log\log n}} \geq 1 \quad \text{a.s.} \tag{5.3}$$

(The same proof applies to $-W_n$, hence also to $|W_n|$.) If (5.3) is not true then, by the previous observation, there exists $c < 1$ such that

$$\limsup_n \frac{W_n(\mathbf{R})}{\sigma n h_n^{3/2}\sqrt{2\log\log n}} = c \quad \text{a.s.} \tag{5.4}$$

Let now $r_k = k^k$. Then,

$$\limsup_k \frac{W_{r_k}(\mathbf{R})}{\sigma r_k h_{r_k}^{3/2}\sqrt{2\log\log r_k}} = c' \le c \quad \text{a.s.,} \tag{5.4$'$}$$

and, by the argument used in the proof of (5.2), the same is true of $W_{r_k, r_{k-1}}$, so that, by independence and Borel-Cantelli, there exists $c'' < 1$ such that

$$\sum_k \Pr\Big\{W_{r_k, r_{k-1}}(\mathbf{R}) \ge c''\sigma r_k h_{r_k}^{3/2}\sqrt{2\log\log r_k}\Big\} < \infty. \tag{5.5}$$

If we set $m_k = r_k - r_{k-1}$ and define W'_{m_k} just as W_{m_k} but with h_{m_k} replaced by h_{r_k}, then, W'_{m_k} has the same distribution as $W_{r_k, r_{k-1}}$ and, since $m_k/r_k \to 1$, it follows from (5.5) that, with $c''' < 1$,

$$\sum_k \Pr\Big\{W'_{m_k}(\mathbf{R}) \ge c'''\sigma m_k h_{r_k}^{3/2}\sqrt{2\log\log m_k}\Big\} < \infty. \tag{5.5$'$}$$

On the other hand, for any $\delta > 0$, in particular for some $0 < C := c'''(1+\delta) < 1$, and for any $M > 0$,

$$\begin{aligned} \Pr\Big\{W'_{m_k}(\mathbf{R}) &\ge c'''\sigma m_k h_{r_k}^{3/2}\sqrt{2\log\log m_k}\Big\} \\ &\ge \Pr\Big\{W'_{m_k}([-M,M]) \ge C\sigma m_k h_{r_k}^{3/2}\sqrt{2\log\log m_k}\Big\} \\ &\quad - \Pr\Big\{\big|W'_{m_k}([-M,M]^c)\big| \ge \delta c'''\sigma m_k h_{r_k}^{3/2}\sqrt{2\log\log m_k}\Big\}. \end{aligned} \tag{5.6}$$

Since $\int_{[-M,M]^c} f^2(x)dx \to 0$ as $M \to \infty$, it follows from inequality (3.11) in Proposition 3.3 that there exists $M_0 < \infty$ such that, for all $M > M_0$,

$$\sum_k \Pr\Big\{W'_{m_k}([-M,M]^c) \ge \delta c'''\sigma m_k h_{r_k}^{3/2}\sqrt{2\log\log m_k}\Big\} < \infty. \tag{5.7}$$

Let now $\varepsilon > 0$ be such that $b := C^2(1+\varepsilon)^3 < 1$ and let $M > M_0$ be such that $\sigma/\sigma(M) < 1+\varepsilon$. Then, the left hand side of (4.16) in Proposition 4.7 gives that, for all k large enough,

$$\Pr\Big\{W'_{m_k}([-M,M]) \ge C\sigma m_k h_{r_k}^{3/2}\sqrt{2\log\log m_k}\Big\} \ge \exp\big(-b\log\log m_k\big) - \frac{1}{m_k^2},$$

and the right hand side of this inequality is the general term of a divergent series, that is,

$$\sum_k \Pr\Big\{W'_{m_k}([-M,M]) \ge C\sigma m_k h_{r_k}^{3/2}\sqrt{2\log\log m_k}\Big\} = \infty. \tag{5.8}$$

Combining (5.6)-(5.8) gives

$$\sum_k \Pr\Big\{W'_{m_k}(\mathbf{R}) \ge c'''\sigma m_k h_{r_k}^{3/2}\sqrt{2\log\log m_k}\Big\} = \infty,$$

which contradicts (5.5'), therefore proving inequality (5.3).

5.1.2. Blocking for the upper bound. Let $\{\lambda_k\}$ be the sequence specified by condition (1.4). Replacing λ_k by $n_k := \min\{n \in \mathbf{N} : n \geq \lambda_k\}$ produces a sequence of natural numbers with the same properties as the original sequence $\{\lambda_k\}$ except for strict monotonicity, however, $\{n_k\}$ is non-decreasing and eventually strictly monotone, that is, there is $k_0 \in \mathbf{N}$ such that $\{n_k\}$ is strictly monotone on $[k_0, \infty)$. So, without loss of generality we assume that there exists a non-increasing sequence $\{n_k\}$ of natural numbers, strictly increasing on $[k_0, \infty)$, $k_0 < \infty$, such that $n_{k+1}/n_k \to 1$ and $\log\log n_k / \log k \to 1$, as $k \to \infty$, and that the sequence $\{h_n\}$ satisfies

$$h_n \text{ is constant for } n \in [n_k, n_{k+1}),\ \ k \in \mathbf{N}. \tag{1.4'}$$

For each $k \in \mathbf{N}$, let I_k be the blocks

$$I_k := [n_k, n_{k+1}) \cap \mathbf{N},$$

and notice that, by (1.4'), h_n, as a function of n, is constant on I_k for all k. Also, $I_k \neq \emptyset$ for $k \geq k_0$. In order to prove the upper bound for the LIL, that is,

$$\limsup_n \frac{|W_n(\mathbf{R})|}{\sigma n h_n^{3/2}\sqrt{2\log\log n}} \leq 1 \quad \text{a.s.}, \tag{5.9}$$

it clearly suffices to prove that for every $\delta > 0$

$$\sum_{k\geq k_0} \Pr\Big\{\max_{n\in I_k} |W_n(\mathbf{R})| > (1+\delta)\sigma n_k h_{n_k}^{3/2}\sqrt{2\log\log n_k}\Big\} < \infty. \tag{5.10}$$

In this subsection we prove:

5.2. Lemma. *Under the hypotheses of Theorem 5.1,*

$$\sum_{k\geq k_0} \Pr\Big\{\max_{n\in I_k} |W_n(\mathbf{R}) - W_{n_k}(\mathbf{R})| > \tau\sigma n_k h_{n_k}^{3/2}\sqrt{2\log\log n_k}\Big\} < \infty \tag{5.11}$$

for every $\tau > 0$.

This lemma clearly reduces proving (5.10) to showing that

$$\sum_{k\geq k_0} \Pr\Big\{|W_{n_k}(\mathbf{R})| > (1+\delta)\sigma n_k h_{n_k}^{3/2}\sqrt{2\log\log n_k}\Big\} < \infty \tag{5.12}$$

for every $\delta > 0$.

Proof of Lemma 5.2. For $n \in I_k$, $k \geq k_0$, we have

$$W_n(\mathbf{R}) - W_{n_k}(\mathbf{R}) = 2\sum_{j=n_k+1}^{n}\sum_{i=1}^{n_k} H_{n_k}(X_i, X_j) + W_{n,n_k}(\mathbf{R}), \tag{5.13}$$

and here we are making use of (1.4'). Conditionally on $X_1, \dots, X_{n_k}$, the random variables $\sum_{i=1}^{n_k} H_{n_k}(X_i, X_j)$, $j = n_k + 1, \dots, n_{k+1} - 1$, are i.i.d. and therefore, by Montgomery-Smith's maximal inequality (Montgomery-Smith (1993); see e.g. de la Peña and Giné (1999), page 6), we have

$$\Pr\Big\{\max_{n\in I_k}\Big|\sum_{j=n_k+1}^{n}\sum_{i=1}^{n_k} H_{n_k}(X_i,X_j)\Big| > t\Big\} \le 9\Pr\Big\{\Big|\sum_{j=n_k+1}^{n_{k+1}-1}\sum_{i=1}^{n_k} H_{n_k}(X_i,X_j)\Big| > \frac{t}{30}\Big\} \tag{5.14}$$

for all $t > 0$. Let us consider now the second summand in (5.13),

$$\begin{aligned} W_{n,n_k}(\mathbf{R}) &= \sum_{n_k<i\ne j\le n} H_{n_k}(X_i,X_j) + \sum_{i=n_k+1}^{n}\big(H_{n_k}(X_i,X_i) - EH_{n_k}(X_i,X_i)\big) \\ &:= U_{n,n_k} + L_{n,n_k}. \end{aligned}$$

On the way to proving (5.11) we must eliminate the maximum from probabilities of the form $\Pr\{\max_{n\in I_k}|U_{n,n_k}| > t\}$. This can be achieved by decoupling, adding the diagonal, and then applying Montgomery-Smith's maximal inequality twice (iteratively). Typically one decouples norms of Banach space valued U-statistics (perhaps with varying kernels), so we must show that $\max_{n\in I_k}|U_{n,n_k}|$ is such a norm. Proceeding as in page 108 from de la Peña and Giné (1999), we set

$$\tilde{H}_{n_k,r} := \big(0, \overset{r-1}{\dots}, 0, H_{n_k}, \dots, H_{n_k}\big) \in \mathbf{R}^{n_{k+1}-1},$$

meaning that the first $r-1$ coordinates are zero and the remaining ones up to $n_{k+1}-1$ are H_{n_k}. We consider $\tilde{H}_{n_k,r}$ as a function with values in $\ell^\infty_{n_{k+1}-1}$, that is, in $\mathbf{R}^{n_{k+1}-1}$ with norm the maximum of the absolute values of the coordinates, $\|(a_1,\dots,a_{n_{k+1}-1})\| = \max|a_i|$. With this notation, it is easy to see that

$$\max_{n\in I_k}|U_{n,n_k}| = \Big\|\sum_{n_k<i\ne j\le n_{k+1}-1}\tilde{H}_{n_k,i\vee j}\Big\|.$$

Then, by direct application of the decoupling result of de la Peña and Montgomery-Smith (1994) (e.g., de la Peña and Giné (1999), pages 125-126) to this norm of a generalized vector-valued U-statistic, we obtain that there exists a universal constant C such that, for all $t > 0$,

$$\Pr\Big\{\max_{n\in I_k}|U_{n,n_k}| > t\Big\} \le C\Pr\Big\{\max_{n\in I_k}|U^{\mathrm{dec}}_{n,n_k}| > \frac{t}{C}\Big\},$$

where

$$U^{\mathrm{dec}}_{n,n_k} = \sum_{n_k<i\ne j\le n} H_{n_k}(X_i^{(1)}, X_j^{(2)}),$$

with the random variables $X_i^{(1)}$ and $X_j^{(2)}$, $i,j\in\mathbf{N}$, being i.i.d. copies of X_1. This is not directly treatable by Montgomery-Smith's maximal inequality, which requires i.i.d.

random variables, but adding the diagonal (that we can subtract later), we have:

$$
\begin{aligned}
&\Pr\Big\{\max_{n\in I_k}\Big|\sum_{n_k<i,j\le n} H_{n_k}(X_i^{(1)},X_j^{(2)})\Big|>t\Big\}\\
&\quad\le \Pr{}^{(2)}\Pr{}^{(1)}\Big\{\max_{n\in I_k}\max_{m\in I_k}\Big|\sum_{i=n_k+1}^{n}\Big(\sum_{j=n_k+1}^{m} H_{n_k}(X_i^{(1)},X_j^{(2)})\Big)\Big|>t\Big\}\\
&\quad\le 9\Pr{}^{(2)}\Pr{}^{(1)}\Big\{\max_{m\in I_k}\Big|\sum_{i=n_k+1}^{n_{k+1}-1}\Big(\sum_{j=n_k+1}^{m} H_{n_k}(X_i^{(1)},X_j^{(2)})\Big)\Big|>\frac{t}{30}\Big\}\\
&\quad= 9\Pr{}^{(1)}\Pr{}^{(2)}\Big\{\max_{m\in I_k}\Big|\sum_{j=n_k+1}^{m}\Big(\sum_{i=n_k+1}^{n_{k+1}-1} H_{n_k}(X_i^{(1)},X_j^{(2)})\Big)\Big|>\frac{t}{30}\Big\}\\
&\quad\le 81\Pr\Big\{\Big|\sum_{j=n_k+1}^{n_{k+1}-1}\sum_{i=n_k+1}^{n_{k+1}-1} H_{n_k}(X_i^{(1)},X_j^{(2)})\Big|>\frac{t}{900}\Big\},
\end{aligned}
$$

where $\Pr^{(1)}$ and $\Pr^{(2)}$ refer to conditional probabilities given respectively the $X^{(2)}$ and the $X^{(1)}$ variables. In the second inequality we have applied the Montgomery-Smith maximal inequality to the ℓ^∞ norms of the successive sums of vectors $(\sum_{j=n_k+1}^{m} H_{n_k}(X_i^{(1)},X_j^{(2)}) : m\in I_{n_k})$, which are i.i.d. conditionally on the $X^{(2)}$ variables, and in the last inequality, to the absolute values of the successive sums of random variables $\sum_{i=n_k+1}^{n_{k+1}-1} H_{n_k}(X_i^{(1)},X_j^{(2)})$, which are i.i.d. conditionally on the $X^{(1)}$ variables.

Using the previous bound, adding and subtracting the diagonal to the U-statistics U_{n,n_k}, and aplying Montgomery-Smith to the resulting sums of i.i.d. random variables and to L_{n,n_k}, we finally obtain that, for all $t>0$,

$$
\begin{aligned}
\Pr\Big\{\max_{n\in I_k}|W_{n,n_k}(\mathbf{R})|>t\Big\} &\le C\Pr\Big\{\max_{n\in I_k}|U^{\mathrm{dec}}_{n,n_k}|>\frac{t}{2C}\Big\}+\Pr\Big\{\max_{n\in I_k}|L_{n,n_k}|>\frac{t}{2}\Big\}\\
&\le C\Pr\Big\{\max_{n\in I_k}\Big|\sum_{n_k<i,j\le n} H_{n_k}(X_i^{(1)},X_j^{(2)})\Big|>\frac{t}{4C}\Big\}\\
&\quad + C\Pr\Big\{\max_{n\in I_k}\Big|\sum_{n_k<i\le n} H_{n_k}(X_i^{(1)},X_i^{(2)})\Big|>\frac{t}{4C}\Big\}\\
&\quad + \Pr\Big\{\max_{n\in I_k}|L_{n,n_k}|>\frac{t}{2}\Big\}\\
&\le 81C\Pr\Big\{\Big|\sum_{n_k<i,j<n_{k+1}} H_{n_k}(X_i^{(1)},X_j^{(2)})\Big|>\frac{t}{3600C}\Big\}\\
&\quad + 9C\Pr\Big\{\Big|\sum_{n_k<i<n_{k+1}} H_{n_k}(X_i^{(1)},X_i^{(2)})\Big|>\frac{t}{120C}\Big\} \qquad (5.15)\\
&\quad + 9C\Pr\Big\{\Big|\sum_{n_k<i<n_{k+1}} \big(H_{n_k}(X_i,X_i)-EH_{n_k}(X_i,X_i)\big)\Big|>\frac{t}{60C}\Big\}.
\end{aligned}
$$

So, by (5.13)-(5.15), the proof of Lemma 5.2 reduces to showing that for every $\tau > 0$ we have

$$\sum_{k\geq k_0} \Pr\left\{\Big|\sum_{j=n_k+1}^{n_{k+1}-1}\sum_{i=1}^{n_k} H_{n_k}(X_i,X_j)\Big| > \tau\sigma n_k h_{n_k}^{3/2}\sqrt{2\log\log n_k}\right\} < \infty, \tag{5.16}$$

$$\sum_{k\geq k_0} \Pr\left\{\Big|\sum_{n_k<i,j<n_{k+1}} H_{n_k}(X_i^{(1)},X_j^{(2)})\Big| > \tau\sigma n_k h_{n_k}^{3/2}\sqrt{2\log\log n_k}\right\} < \infty, \tag{5.17}$$

$$\sum_{k\geq k_0} \Pr\left\{\Big|\sum_{n_k<i<n_{k+1}} H_{n_k}(X_i^{(1)},X_i^{(2)})\Big| > \tau\sigma n_k h_{n_k}^{3/2}\sqrt{2\log\log n_k}\right\} < \infty, \tag{5.18}$$

and

$$\sum_{k\geq k_0} \Pr\left\{\Big|\sum_{n_k<i<n_{k+1}} \big(H_{n_k}(X_i,X_i) - EH_{n_k}(X_i,X_i)\big)\Big| > \tau\sigma n_k h_{n_k}^{3/2}\sqrt{2\log\log n_k}\right\} < \infty, \tag{5.19}$$

By (3.5), the general term of the series in (5.19) is dominated from some k on by $\exp(-cn_kh_{n_k})$ for some $c > 0$, which is the general term of a convergent series. Likewise, by Bernstein's inequality and the variance estimate (2.13), the general term of the series (5.18) is eventually dominated by $\exp(-cn_kh_{n_k}^{1/2})$ for some $c > 0$, which is also the general term of a convergent series. Proposition 3.4 will take care of (5.16) and (5.17). For instance, if we look at the four quantities in the exponent at the right hand side of inequality (3.13) in the present case of (5.16), we see that the first term is of the order of a constant times

$$\frac{n_k}{n_{k+1}-n_k}\log\log n_k \sim M_k \log k$$

for some sequence $M_k \to \infty$ as $k \to \infty$, and the other three terms are of the order of positive powers of n_k. So, we can take k large enough to overwhelm the constant (that may be large, depending on τ) and get that, given $\tau > 0$, from some k on,

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

proving (5.18). The same argument, this time based on (3.12), proves (5.17). □

5.1.3. The upper bound. By the parts 5.1.1 and 5.1.2 of this proof, Theorem 5.1 will be proved if we show that the series in (5.12) converges for all $\delta > 0$, as mentioned above. For $M > 0$,

$$\begin{aligned}\Pr\Big\{|W_{n_k}(\mathbf{R})| > (1+\delta)\sigma n_k h_{n_k}^{3/2}\sqrt{2\log\log n_k}\Big\}\\ \leq \Pr\Big\{|W_{n_k}([-M,M])| > (1+\delta/2)\sigma n_k h_{n_k}^{3/2}\sqrt{2\log\log n_k}\Big\}\\ + \Pr\Big\{|W_{n_k}([-M,M]^c)| > \frac{\delta}{2}\sigma n_k h_{n_k}^{3/2}\sqrt{2\log\log n_k}\Big\}\end{aligned} \tag{5.20}$$

Given $\delta > 0$, since $\|f\|_2 < \infty$, there is $M_1 < \infty$ such that

$$\int_{[-M,M]^c} f^2(x)dx < \frac{\delta^2\sigma^2}{4\kappa_0}$$

for all $M \geq M_1$, where κ_0 is the constant in inequality (3.11), which gives, by Proposition 3.3, that

$$\Pr\Big\{|W_{n_k}([-M,M]^c)| > \frac{\delta}{2}\sigma n_k h_{n_k}^{3/2}\sqrt{2\log\log n_k}\Big\} \leq \kappa_0 \exp\big(-2\log\log n_k\big),$$

from some k on, and this is the general term of a convergent series. As for the first series in (5.20), if we choose $\varepsilon > 0$ such that $(1+\delta/2)^2(1-\varepsilon) = \gamma > 1$, then, the right hand side of inequality (4.16) in Proposition 4.7 for

$$a_{n_k} = (1+\delta/2)\sqrt{2\log\log n_k} \leq (1+\delta/2)\sigma\sqrt{2\log\log n_k}/\sigma(M),$$

gives that, from some k on,

$$\Pr\Big\{|W_{n_k}([-M,M])| > (1+\delta/2)\sigma n_k h_{n_k}^{3/2}\sqrt{2\log\log n_k}\Big\} \leq \frac{1}{n_k^2} + \exp\big(-\gamma\log\log n_k\big),$$

which is the general term of a convergent series. Combining the last two estimates with (5.20) gives that for every $\delta > 0$,

$$\sum_k \Pr\Big\{|W_{n_k}(\mathbf{R})| > (1+\delta)\sigma n_k h_{n_k}^{3/2}\sqrt{2\log\log n_k}\Big\} < \infty,$$

that is, (5.12). Together with Lemma 5.2, this proves the upper part of the LIL, (5.9), and therefore, together with Subsection 5.1.1, Theorem 5.1. $\square$

6. Remarks on the LIL for the integrated squared deviation of a kernel density estimator from the true density. The integrated squared deviation of $f_{n,K}$ from f, defined as

$$I_n = \int_{\mathbf{R}} \big(f_{n,K}(t) - f(t)\big)^2 dt, \tag{6.1}$$

constitutes a measure of global performance for the estimator $f_{n,K}$ of f. It is not our aim here to study this statistic, however Theorem 5.1 may be seen as the main step in the derivation of an LIL for I_n, and it turns out that it is the only step with interesting difficulties (the rest is more or less routine, except the case $h_n \asymp n^{-1/5}$). In this section we describe how to apply Theorem 5.1 and Theorem 4.5 to derive such an LIL. We will not make any efforts to get it under best conditions, but only under the 'stated conditions' of Hall (1984) for the CLT, and then, we will only get an approximate result for $h_n \asymp n^{-1/5}$. Most details will be left to the reader.

Consider the following decomposition obtained from I_n by adding and subtracting $Ef_{n,K}(t)$ inside the square in (6.2):

$$I_n - EI_n = J_n + \frac{2}{nh_n}\int_{\mathbf{R}} \big(Ef_{n,K}(t) - f(t)\big)\sum_{i=1}^{n} \bar{K}_{h_n}(t - X_i)dt. \tag{6.2}$$

Theorem 5.1 applies to J_n. For the second term at the right of (6.2), assuming (1.2) and

$$K \geq 0, \ \int_{\mathbf{R}} xK(x)dx = 0, \ \int_{\mathbf{R}} x^2K(x)dx := 2k < \infty \tag{6.3}$$

for K and that

$$f, \ f', \ \textit{and } f'' \textit{ are bounded and uniformly continuous on } \mathbf{R} \tag{6.4}$$

for f (these are Hall's 'stated conditions'), Lemma 1 from Hall (1984) shows that

$$E\Big[\int_{\mathbf{R}} \big(Ef_{n,K}(t) - f(t)\big)\bar{K}_{h_n}(t - X_i)dt\Big]^2 \simeq h_n^6k^2v^2,$$

where

$$v^2 := \int_{\mathbf{R}} \big(f''(x)\big)^2 f(x)dx - \Big(\int_{\mathbf{R}} f''(x)f(x)dx\Big)^2 \tag{6.5}$$

and k is defined in (6.3), and that the fourth moments of these random variables are $O\big(h_n^{12}\big)$. Hence, by the remark following Lemma 4.4, we can apply Theorem 4.5 to

$$S_n := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\int_{\mathbf{R}} \big(Ef_{n,K}(t) - f(t)\big)\bar{K}_{h_n}(t - X_i)dt}{h_n^3kv}$$

with $b_n \asymp n^{-1/2}$, $a_n = \sqrt{2\log\log n}$ and all n large enough. Now, given the moderate deviation inequality (4.11) for S_n, we can proceed in a standard way (as, given (1.4), blocking for these sums of independent random variables offers no problems) to obtain

$$\limsup_n \frac{\sqrt{n}}{2kvh_n^2\sqrt{2\log\log n}}\left|\frac{2}{nh_n}\int_{\mathbf{R}} \big(Ef_{n,K}(t) - f(t)\big)\sum_{i=1}^{n} \bar{K}_{h_n}(t - X_i)dt\right| = 1 \ \text{ a.s.} \tag{6.6}$$

Theorem 5.1 and the limit (6.6) then give the following:

6.1. Proposition. *Assume (1.3) and (1.4) for $\{h_n\}$, (1.2) and (6.3) for K and (6.4) for f. Then,*

$$\limsup_n \frac{n\sqrt{h_n}}{\sigma\sqrt{2\log\log n}}|I_n - EI_n| = 1 \text{ a.s. } \ \textit{if} \ \ h_n \asymp \frac{1}{n^\delta} \ \ \textit{and} \ \ \frac{1}{5} < \delta < \frac{1}{3},$$

$$\limsup_n \frac{\sqrt{n}}{2kvh_n^2\sqrt{2\log\log n}}|I_n - EI_n| = 1 \text{ a.s. } \textit{if} \;\; h_n \asymp \frac{1}{n^\delta} \;\; \textit{and} \;\; 0 < \delta < \frac{1}{5},$$

and there exists $C < \infty$ *such that*

$$\limsup_n \frac{n^{9/10}}{\sqrt{\log\log n}}\big|I_n - EI_n\big| = C \text{ a.s. } \textit{if} \;\; h_n \asymp \frac{1}{n^{1/5}}.$$

Acknowledgement. This research was partially supported by NSF Grant No. DMS-0070382 (E. Giné) and by NSF Grant No. DMS–0203865 and NSA Grant MSPF–01G-037 (D. Mason). We thank Anton Schick for a useful conversation regarding Lemma 2.1.

References

Beirlant, J. and Mason, D.M. (1995). On the asymptotic normality of L_p–norms of empirical functionals. *Math. Methods Statist.* **4** 1–19.

Bickel, P.J. and Rosenblatt, M. (1973). On some global measures of the deviations of density estimation. *Ann. Statist.* **1** 1071-95.

Csörgő, M. and Horváth, L. (1988). Central limit theorems for L_p–norms of density estimators. *Z. Wahrsch. Verw. Gebiete.* **80** 269–291.

de la Peña, V. and Giné, E. (1999). *Decoupling, from dependence to independence.* Springer-Verlag, New York.

de la Peña, V. and Montgomery-Smith, S. (1994). Bounds for the tail probabilities of U-statistics and quadratic forms. *Bull. Amer. Math. Soc.* **31** 223-227.

Dunford, N. and Schwartz, J.T. (1963). *Linear Operators, Part II.* Second Printing, 1964. Wiley, New York.

Eggermont, P.P.B. and LaRiccia, V.N. (2001). *Maximum Penalized Likelihood Estimation, Volume 1; Density Estimation.* Springer, New York.

Fernique, X. (1970). Integrabilité des vecteurs gaussiens. *C. R. Acad. Sci. Paris, Sér. A* **270** 1698-1699.

Giné, E.; Latała, R.; and Zinn, J. (2000). Exponential and moment inequalities for U-statistics. In *Progress in Probability* **47**, *High Dimenisonal Probability II*, pp. 13-38.

Giné, E.; Mason, D.M.; and Zaitsev, A. Yu. (2003). The L_1-norm density estimation process. *Ann. Probab.* **31** 719–768.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate non-parametric density estimators. *J. Multivariate Analysis* **14** 1-16.

Komlós, J.; Major, P.; and Tusnády, G. (1975). An approximation of partial sums of independent rv's and the sample distribution function, I. *Zeits. Wahrsch. verw. Geb.* **32** 111-131.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces.* Springer, Berlin.

Mason, D.M. (2003). Representations for estimators of integral functionals of the density function. *Austrian Journal of Statistics*, **32** 131-142.

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

Nadaraya, N.A. (1989). *Nonparametric Estimation of Probability Densities and Regression Curves.* Kluwer, Amsterdam.

PINSKY, M. (1966). An elementary derivation of Khintchine's estimate for large deviations. *Proc. Amer. Math. Soc.* **22** 288-290.

ROSENBLATT, M. (1975). A quadratic measure of the deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3** 1-14.

SHORACK, G. AND WELLNER, J. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.