

The L_1 -Norm Density Estimator Process

Evarist Giné,^{*} David M. Mason[†] and Andrei Yu. Zaitsev[‡]
University of Connecticut, University of Delaware,
St. Petersburg Branch of the Steklov Mathematical Institute

September 12, 2001

Abstract

The notion of an L_1 -norm density estimator process indexed by a class of kernels is introduced. Then a functional central limit theorem and a Glivenko–Cantelli theorem are established for this process. While assembling the necessary machinery to prove these results, a body of Poissonization techniques and restricted chaining methods is developed, which is useful for studying weak convergence of general processes indexed by a class of functions. This should be of independent interest. None of the theorems impose any condition at all on the underlying Lebesgue density f . Also, somewhat unexpectedly, the distribution of the limiting Gaussian process does not depend on f .

AMS Subject Classifications: 60F05, 60F15, 60F17, 62G07

Keywords: Kernel density function estimator, L_1 -norm, central limit theorem, weak convergence, Poissonization, entropy.

1 Introduction: The L_1 -norm density estimator process.

Let X, X_1, X_2, \dots , be a sequence of independent and identically distributed random variables in \mathbf{R} with common Lebesgue density f . Further let $\{h_n\}_{n=1}^\infty$ be a sequence of positive constants such that as $n \rightarrow \infty$, $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$. The classical kernel estimator is defined as

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

^{*}e-mail: gine@uconnvm.uconn.edu, partially supported by NSF Grant DMS-0070382.

[†]e-mail: davidm@udel.edu, research partially supported by NSF Grant DMS-9803344.

[‡]e-mail: zaitsev@pdmi.ras.ru, research partially supported by an NSF Grant while visiting the University of Delaware, Russian Foundation of Basic Research (RFBR) Grant 99-01-00112, INTAS Grant 99-01317 and RFBR-DFG Grant 99-01-04027.

where K is a kernel with compact support satisfying

$$\int_{\mathbf{R}} K(u)du = 1. \quad (1.1)$$

For notational convenience we will usually assume

$$K(u) = 0 \text{ for } |x| > 1/2. \quad (1.2)$$

Since Lebesgue density functions by definition sit in $L_1(\mathbf{R}, \mathcal{B}, \lambda)$, where λ denotes Lebesgue measure, Devroye and Györfi have long advocated that the natural distance to measure the error in estimation between a density function f and its estimator $f_{n,K}$ is the L_1 -norm of their difference,

$$J_n(K) := \|f_{n,K} - f\|_1 = \int_{\mathbf{R}} |f_{n,K}(x) - f(x)| dx. \quad (1.3)$$

In their book, Devroye and Györfi (1985), they posed the challenging problem of finding the asymptotic distribution of $\|f_{n,K} - f\|_1$.

M. Csörgő and Horváth (1988) were the first to prove a central limit theorem for $\|f_{n,K} - f\|_p$, the L_p -norm distance, $p \geq 1$. Their proof relied on the Mason and van Zwet (1987) refinement of the KMT inequality. Horváth (1991) introduced a Poissonization technique into the study of central limit theorems for $\|f_{n,K} - f\|_p$. The M. Csörgő and Horváth (1988) and Horváth (1991) results required numerous regularity conditions.

Later Beirlant and Mason (1995) developed a general method, based on a somewhat different Poissonization, coupled with Fourier inversion, for deriving the asymptotic normality of the L_p norm of empirical functionals. Mason (see Theorem 8.9 in Eggermont and LaRiccia (2000)) has recently applied their method to the special case of the L_1 -norm of the kernel density estimator to prove that whenever $h_n \rightarrow 0$ and $\sqrt{n}h_n \rightarrow \infty$, and K satisfies (1.2) and $\|K\|_\infty < \infty$, then

$$\xi_n(K) := \sqrt{n} \left\{ \|f_{n,K} - Ef_{n,K}\|_1 - E \|f_{n,K} - Ef_{n,K}\|_1 \right\}$$

converges in distribution to a normal random variable with mean zero and variance

$$\sigma^2(K) := \|K\|_2^2 \int_{\mathbf{R}} \text{Cov} \left(\left| \sqrt{1 - \rho^2(t)} Z_1 + \rho(t) Z_2 \right|, |Z_2| \right) dt,$$

where, here and elsewhere in this paper, Z_1 and Z_2 are independent standard normal random variables and

$$\rho(t) := \frac{\int_{-1}^1 K(u)K(u+t)du}{\|K\|_2^2}, \quad t \in \mathbf{R}.$$

(Throughout this paper Z will always denote a standard normal random variable.) As a by product of our work here we will extend this result by showing that $\xi_n(K)$ remains asymptotically $\sigma(K)Z$ when $\|K\|_\infty < \infty$ is replaced by $\|K\|_2^2 < \infty$.

It is natural to consider $\xi_n(K)$ as a process indexed by a class \mathcal{K} of square integrable functions on \mathbf{R} satisfying (1.2) and, in light of the asymptotic normality result, conjecture that, under

suitable assumptions on the class \mathcal{K} , the sequence of processes $\{\xi_n(K) : K \in \mathcal{K}\}_{n=1}^\infty$ converges weakly to a mean zero Gaussian process $\{\xi(K) : K \in \mathcal{K}\}$ with covariance function defined for $K_1, K_2 \in \mathcal{K}$ by

$$\sigma(K_1, K_2) := \|K_1\|_2 \|K_2\|_2 \int_{-1}^1 \text{Cov} \left(\left| \sqrt{1 - \rho^2(K_1, K_2, t)} Z_1 + \rho(K_1, K_2, t) Z_2 \right|, |Z_2| \right) dt$$

and

$$\rho(K_1, K_2, t) := \frac{\int_{\mathbf{R}} K_1(u) K_2(u+t) du}{\|K_1\|_2 \|K_2\|_2}, \quad t \in \mathbf{R}.$$

Notice that $\sigma(K, K) = \sigma^2(K)$ and $\rho(K, K, t) = \rho(t)$. Clearly, $\rho(t)$ is a continuous function of t and $\rho(0) = 1$, and it is known that $\text{Cov}(|\zeta_1|, |\zeta_2|) \geq 0$ for any two dimensional normal vector (ζ_1, ζ_2) (see Remark 1.4 in Karlin and Rinott (1981)). Therefore $\sigma^2(K) > 0$, that is, the process $\{\xi(K) : K \in \mathcal{K}\}$ is never degenerate, unless $\mathcal{K} = \{0\}$. Note also that this process does not depend on the density f .

We will use the notation for $p > 0$,

$$\partial_p(K_1, K_2) := \left(\int_{\mathbf{R}} |K_1(x) - K_2(x)|^p dx \right)^{1/p}.$$

Our main result is as follows:

Theorem 1.1. *Let \mathcal{K} be a class of measurable functions, each satisfying (1.2), uniformly bounded by a positive constant κ and such that*

$$\int_0^\infty \log N(\mathcal{K}, \partial_2, \varepsilon) d\varepsilon < \infty. \quad (1.4)$$

If $h_n \rightarrow 0$ and $\sqrt{n}h_n \rightarrow \infty$, then the sequence of processes $\{\xi_n(K) : K \in \mathcal{K}\}_{n=1}^\infty$ converges weakly in $\ell^\infty(\mathcal{K})$ to a mean zero Gaussian process $\{\xi(K) : K \in \mathcal{K}\}$ with covariance function $\sigma(K_1, K_2)$.

See van der Vaart and Wellner (1996) or de la Peña and Giné (1999) for the definition of weak convergence of processes, the covering numbers $\mathcal{N}(\mathcal{K}, \partial_2, \varepsilon)$ and the related packing numbers $D(\mathcal{K}, \partial_2, \varepsilon)$ (that we will use below).

Note that if Theorem 1.1 holds for a separable class of kernels \mathcal{K} , then, in particular, by the continuous mapping theorem,

$$\sup_{K \in \mathcal{K}} |\xi_n(K)| \rightarrow_w \sup_{K \in \mathcal{K}} |\xi(K)|,$$

for all densities f , where the distribution of the right hand random variable is distribution free, meaning that its distribution is independent of f . Moreover, the same is true for any other continuous functional on $\ell^\infty(\mathcal{K})$.

The following example shows that condition (1.4) is satisfied by quite large classes of kernels.

Example 1.1. Let \mathcal{K} be a uniformly bounded class of measurable functions, which are zero off the interval $[-1/2, 1/2]$. Write

$$[-1/2, 1/2] = \cup_{i=1}^m I_i,$$

where $I_i, i = 1, \dots, m$, are disjoint intervals. Let J_i denote the interior of I_i . Assume that each $K \in \mathcal{K}$ is differentiable on $J_i, i = 1, \dots, m$, with derivative K' satisfying

$$\sup_{x \in J_i} \sup_{K \in \mathcal{K}} |K'(x)| \leq D \text{ and } \sup_{x, y \in J_i} \sup_{K \in \mathcal{K}} \left[\frac{|K'(x) - K'(y)|}{|x - y|^\delta} \right] \leq D,$$

where $D > 0$ and $0 < \delta < 1$ are constants depending on \mathcal{K} . In this case one can apply Theorem 7.1.1 of Dudley (1984) to show that for some constant $D_1 > 0$

$$\log N(\mathcal{K}, \partial_2, \varepsilon) \leq D_1 \varepsilon^{-1/(1+\delta)},$$

which implies (1.4).

Example 1.2. Here is another class of functions that satisfies (1.4). Let K be a fixed bounded function, which is equal to zero off the interval $[-1/2, 1/2]$. Assume that K is of bounded variation. The class

$$\mathcal{K} = \{K(\cdot\lambda) : \lambda \geq 1\},$$

which consists of functions that are equal to zero on $[-1/2, 1/2]^c$, satisfies

$$N(\mathcal{K}, \partial_2, \varepsilon) \leq [C^2 \varepsilon^{-2}] + 1$$

for $\varepsilon \leq C$, and therefore, condition (1.4). The constant C can be taken to be $C^2 = M(2M - \int_{-1/2}^{1/2} (K_1 + K_2)(x)dx)$ where $M = \|K\|_\infty$ and K_1 and K_2 are the minimal increasing functions such that $K = K_1 - K_2$. (To see this just note that $\partial_2(K(\cdot\lambda), K(\cdot\mu)) \leq (C|\lambda^{-1} - \mu^{-1}|)^{1/2}$ and then make an easy estimate of the covering numbers of $[1, \infty)$ for this distance.)

Example 1.3. Often the optimal choice of h_n is of the form $\lambda^{-1}n^{-1/5}$, where $\lambda > 0$ is a smoothing parameter to be estimated from the data. Consider now $\xi_n(K_\lambda)$, with $h_n = n^{-1/5}$, as a process indexed by the class of kernels

$$\{K_\lambda(\cdot) = \lambda K(\cdot\lambda) : \lambda > 0\}.$$

Suppose one could establish that this process, restricted to any compact interval $[a, b] \subset (0, \infty)$, converges weakly to a mean zero Gaussian process $\xi(K_\lambda)$ continuous in λ . Then if λ_n is a sequence of data driven smoothing parameter estimators, which converges in probability to some fixed value λ_0 , one could conclude that $\xi_n(K_{\lambda_n})$ converges in distribution to the normal random variable $\xi(K_{\lambda_0})$. For a thorough discussion of smoothing parameter estimators, which have this property, consult Berline and Devroye (1994) and the references therein. It is easy to see, building on Example 1.2, that if K is a bounded kernel of bounded variation vanishing off a compact set, and if $0 < a < b < \infty$, then the class of kernels

$$\mathcal{K} = \{\lambda K(\cdot\lambda) : \lambda \in [a, b]\}$$

satisfies condition (1.4).

Concerning the previous example, we should remark that if

$$\mathcal{K} = \{\lambda K(\lambda) : \lambda \in [a, \infty)\} \quad (1.5)$$

with $a > 0$, then the finite dimensional distributions of the sequence of processes $\{\xi_n(K) : K \in \mathcal{K}\}$ satisfy the central limit theorem for all densities, but the processes themselves do not converge in law in $\ell^\infty(\mathcal{K})$ for any densities (the limiting Gaussian process $\xi(K)$ fails to be sample bounded). This is shown in Example 6.1 below.

The tools that we develop to prove Theorem 1.1 permit us to extend asymptotic normality of $\xi_n(K)$ to kernels satisfying $\|K\|_2 < \infty$.

Theorem 1.2. *Assume K satisfies (1.2) and $\|K\|_2 < \infty$. If $h_n \rightarrow 0$ and $\sqrt{nh_n} \rightarrow \infty$, then as $n \rightarrow \infty$,*

$$\xi_n(K) \rightarrow_w \xi(K) =_d \sigma(K)Z \quad (1.6)$$

and

$$\text{Var } \xi_n(K) \rightarrow \sigma^2(K). \quad (1.7)$$

Notice that, somewhat remarkably, neither in Theorem 1.1 nor in Theorem 1.2 did we impose any assumption on the Lebesgue density f . Also we did not require that K satisfy (1.1).

Of course one may ask when $E f_{n,K}(x)$ can be replaced by $f(x)$ in $\xi_n(K)$ in (1.6). An obvious sufficient condition is that

$$\sqrt{n}\|f - E f_{n,K}\|_1 \rightarrow 0, \text{ as } n \rightarrow \infty. \quad (1.8)$$

However (1.8) need not always hold. For instance, when $K(x) = I(x \in [-1/2, 1/2])$ and X is a Uniform $(0, 1)$ random variable, then

$$\sqrt{n}\|f - E f_{n,K}\|_1 = \frac{\sqrt{nh_n}}{2},$$

which implies that (1.8) is not satisfied under the condition for asymptotic normality $\sqrt{nh_n} \rightarrow \infty$, as $n \rightarrow 0$. One set of sufficient conditions for (1.8) is the following: In addition to the assumptions that K satisfies (1.1), (1.2) and $\|K\|_2 < \infty$, assume that $K(x) = K(-x)$ for all $x \in \mathbf{R}$. Also assume that the density f is twice continuously differentiable on \mathbf{R} with first and second derivatives f' and f'' , respectively, where $\int_{\mathbf{R}} |f''(x)| dx < \infty$. Arguing exactly as in the proofs of Lemmas 3 and 4 in Chapter 5 of Devroye and Györfi (1985) we get that

$$\sqrt{n}\|f - E f_{n,K}\|_1 \leq \frac{\sqrt{nh_n^2}}{2} \int_{\mathbf{R}} x^2 |K(x)| dx \int_{\mathbf{R}} |f''(x)| dx.$$

Thus $\sqrt{nh_n^2} \rightarrow 0$ and $\sqrt{nh_n} \rightarrow \infty$, as $n \rightarrow \infty$, imply both (1.8) and asymptotic normality. (Note that Devroye and Györfi (1985) assume that $K \geq 0$ in their lemmas. Also see Lemma

22 of Devroye and Györfi (1985) for a generalization of their Lemma 4.) A similar comment applies for the replacement of $Ef_{n,K}(x)$ by $f(x)$ in Theorem 1.1.

The results just described are further evidence of the extreme difference between the sup norm and the L_1 -norm of the discrepancy between the kernel density estimator and its mean (or the density). The central limit theorem for the L_∞ -norm does not hold, not even for a single kernel, whereas, one has central limit theorems for the L_1 -norm, even uniform over large classes of kernels.

Devroye and Györfi (1985), Chapter 3, Theorem 1, prove the law of large numbers for $J_n(K)$ (defined in (1.3)), with $K \in L_1$. Just as with the central limit theorem, we may ask for conditions under which this law of large numbers holds uniformly in $K \in \mathcal{K}$ for all densities f . This turns out to be a much easier problem than the central limit theorem, at least this is the case for the following Glivenko–Cantelli result:

Theorem 1.3. *Let \mathcal{K} be a relatively compact subset of $L_1(\mathbf{R}, \mathcal{B}, \lambda)$ satisfying (1.1) for all $K \in \mathcal{K}$. If $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$, then*

$$\lim_{n \rightarrow \infty} E^* \sup_{K \in \mathcal{K}} J_n(K) = 0. \quad (1.9)$$

The following notation is used in this statement and will be used throughout: if X is a not necessarily measurable random function, X^* denotes its measurable envelope, and we set $E^*X = EX^*$ (see van der Vaart and Wellner (1996) for calculus with non-measurable functions).

Theorem 1.3 implies that the law of large numbers for $J_n(K)$ holds uniformly over many classes of kernels. However, we will see in Example 7.1 that this is not so for the class of kernels (1.5).

The bulk of this paper is devoted to the proof of Theorem 1.1. The proof consists of two main steps. We must prove that the finite dimensional distributions [f.d.] of the processes $\{\xi_n(K) : K \in \mathcal{K}\}_{n=1}^\infty$ converge in law to the corresponding f.d. distributions of $\{\xi(K) : K \in \mathcal{K}\}$ and then we must also show that these processes are uniformly tight, that is, that they satisfy the asymptotic equicontinuity condition

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \Pr \left\{ \sup_{\partial_2(K_1, K_2) \leq \delta, K_1, K_2 \in \mathcal{K}} |\xi_n(K_1) - \xi_n(K_2)| > \varepsilon \right\} = 0 \quad (1.10)$$

(see either van der Vaart and Wellner (1996) or de la Peña and Giné (1999)). In fact we will prove a result stronger than tightness, namely, that the increments of the process satisfy a uniform exponential integrability condition. After establishing some necessary preliminary results in Sections 2, 3 and 4, tightness will be proved in Section 5 and f.d. convergence in Section 6, where the proof of Theorem 1.2 is also given. In the process of proving Theorem 1.1, we will develop a body of Poissonization techniques and restricted chaining methods useful for studying weak convergence of general processes indexed by a class of functions. These are detailed in Sections 2 and 4, and should be of independent interest. The Glivenko–Cantelli theorem is proved in Section 7.

2 Poissonization techniques.

One of the main ingredients in the proof of the central limit theorem for the sequence of processes $\{\xi_n(K) : K \in \mathcal{K}\}_{n=1}^\infty$ is Poissonization of the empirical process, the reason being that, as is well known, if η is a Poisson random variable independent of the i.i.d. sequence $X_i, i \in \mathbf{N}$, $X_0 = 0$, and if $A_k, k \in \mathbf{N}$, are disjoint measurable sets, then the processes $\sum_{i=0}^\eta I(X_i \in A_k) \delta_{X_i}, k = 1, 2, \dots$, are independent. The following lemma is basic for the tightness part of the proof. It comes from Einmahl (1987) and Deheuvels and Mason (1992) (also see Einmahl and Mason (1997)), and we give it here in an abstract form suitable for our purposes. Here is some convenient notation. We say that a set D is a (commutative) semigroup if it has a commutative and associative operation, in our case sum, with a zero element. If D is equipped with a σ -algebra \mathcal{D} for which the sum, $+: (D \times D, \mathcal{D} \otimes \mathcal{D}) \mapsto (D, \mathcal{D})$, is measurable, then we say the (D, \mathcal{D}) is a measurable semigroup.

Lemma 2.1. *Let (D, \mathcal{D}) be a measurable semigroup; let $X_0 = 0 \in D$ and let $X_i, i \in \mathbf{N}$, be independent identically distributed D -valued random variables; for any given $n \in \mathbf{N}$, let η be a Poisson random variable with mean n independent of the sequence $\{X_i\}$; and let $B \in \mathcal{D}$ be such that $\Pr\{X_1 \in B\} \leq 1/2$. Then,*

$$\Pr \left\{ \sum_{i=0}^n I(X_i \in B) X_i \in C \right\} \leq 2 \Pr \left\{ \sum_{i=0}^\eta I(X_i \in B) X_i \in C \right\} \quad (2.1)$$

for all $C \in \mathcal{D}$. In particular, if $H : D \mapsto \mathbf{R}$ is non-negative and \mathcal{D} -measurable, then

$$EH \left(\sum_{i=0}^n I(X_i \in B) X_i \right) \leq 2EH \left(\sum_{i=0}^\eta I(X_i \in B) X_i \right) \quad (2.2)$$

Proof. Set $p_B = \Pr\{X_1 \in B\}$ and $p_{B^c} = 1 - p_B$. Let $\tau_0 = 0 \in \mathbf{R}, Y_0 = 0 \in D$ and let $\tau_i, Y_i, i \in \mathbf{N}$, be independent random variables such that $\Pr\{\tau_i = 1\} = 1 - \Pr\{\tau_i = 0\} = \Pr\{X_1 \in B\}$ and $Y_i =_d (X_1 | X_1 \in B)$ for all $i \geq 1$. It is easy to see that

$$I(X_1 \in B) X_1 =_d \tau_1 Y_1.$$

Therefore, if η_B is Poisson with mean np_B independent of the variables τ_i and $Y_i, i = 1, 2, \dots$, it follows that

$$\sum_{i=0}^\eta I(X_i \in B) X_i =_d \sum_{i=0}^\eta \tau_i Y_i =_d \sum_{i=0}^{\eta_B} Y_i,$$

where the last identity is classical. We can assume $\eta = \eta_B + \eta_{B^c}$, where η_B and η_{B^c} are independent Poisson respectively with parameters np_B and np_{B^c} , independent of the other variables. We then have:

$$\Pr \left\{ \sum_{i=0}^n I(X_i \in B) X_i \in C \right\} = \Pr \left\{ \sum_{i=1}^n \tau_i Y_i \in C \right\}$$

$$\begin{aligned}
&= I(0 \in C) \Pr \left\{ \sum_{i=1}^n \tau_i = 0 \right\} + \sum_{i=1}^n \Pr \left\{ \sum_{i=1}^n \tau_i Y_i \in C, \sum_{i=1}^n \tau_i = k \right\} \\
&= \sum_{k=0}^n \binom{n}{k} p_B^k p_{B^c}^{n-k} \Pr \left\{ \sum_{i=0}^k Y_i \in C \right\} \\
&= \frac{n! e^n}{n^n} \sum_{k=0}^n \frac{(np_B)^k}{k! e^k} \frac{(np_{B^c})^{n-k}}{(n-k)! e^{n-k}} \Pr \left\{ \sum_{i=1}^k Y_i \in C \right\} \\
&= \frac{1}{\Pr\{\eta = n\}} \sum_{k=0}^n \Pr \left\{ \sum_{i=0}^k Y_i \in C, \eta_B = k \right\} \Pr\{\eta_{B^c} = n - k\} \\
&\leq \frac{\max_{0 \leq k \leq n} \Pr\{\eta_{B^c} = n - k\}}{\Pr\{\eta = n\}} \Pr \left\{ \sum_{i=0}^{\eta_B} Y_i \in C \right\} \\
&= \frac{\Pr\{\eta_{B^c} = [np_{B^c}]\}}{\Pr\{\eta = n\}} \Pr \left\{ \sum_{i=0}^{\eta} I(X_i \in B) X_i \in C \right\}.
\end{aligned}$$

Now, Stirling's formula, $n! = (n/e)^n \sqrt{2\pi n} e^{\theta n/n}$ for some $0 < \theta < 1/12$, gives that $\Pr\{\eta_{B^c} = [np_{B^c}]\} / \Pr\{\eta = n\} \leq 2$ if $p_{B^c} > 1/2$ and $n \geq 5$; direct computation shows that this inequality is also true for $1 \leq n \leq 4$, proving the first inequality in the lemma. The second inequality follows from the first by the usual integration by parts formula for expected values. \square

Remark 2.1. We will apply the preceding lemma for the semigroup D generated by the point masses, $D = \{0, \sum_{i=1}^n \delta_{x_i} : n \in \mathbf{N}, x_i \in S\}$, where (S, \mathcal{S}) is a measurable space, with the σ -algebra \mathcal{D} generated by the functions $f_{n,B}(x_1, \dots, x_n) = \sum_{i=1}^n I(x_i \in B) \delta_{x_i}$, $n \in \mathbf{N}$, $B \subset S$. It is easy to see that for any measurable function $h : S \mapsto \mathbf{R}$, the map $\mu \mapsto \int h d\mu$ is \mathcal{D} -measurable (just note that $f_{n,B}^{-1}\{\mu \in D, \int h d\mu \leq t\} = \{(x_1, \dots, x_n) : \sum I(x_i \in B) h(x_i) \leq t\}$ is a measurable set of S^n). Our functions H will have the general form

$$H \left(\sum_{i=1}^n I(X_i \in B) \delta_{X_i} \right) = \exp \left\{ \lambda \left| \int_A \left| \sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) - b(x) \right| - c(x) \right| dx \right\},$$

where A is a Borel set and B is the $h/2$ -neighborhood of A . H can be shown to be \mathcal{D} measurable by approximation of K by Riemann integrable functions.

We will need to estimate moments of Poissonized sums in both parts of the proof of the central limit theorem. The following lemma extends to Poissonized sums the sharpest bounds for moments of sums of independent random variables. Before stating it, we will recall the Johnson–Schechtman–Zinn improvement on Rosenthal's inequality: if ξ_i are independent centered random variables, then, for every $p \geq 2$ and $n \in \mathbf{N}$,

$$E \left| \sum_{i=1}^n \xi_i \right|^p \leq \left(\frac{15p}{\log p} \right)^p \max \left[\left(\sum_{i=1}^n E \xi_i^2 \right)^{p/2}, \sum_{i=1}^n E |\xi_i|^p \right] \quad (2.3)$$

(obtained by symmetrization of the inequality in Theorem 4.1 from Johnson, Schechtman and Zinn, 1985). This bound has a version for sums of independent non-negative random variables ζ_i , namely: for every $p \geq 1$ and $n \in \mathbf{N}$,

$$E \left(\sum_{i=1}^n \zeta_i \right)^p \leq 2^p \left(\frac{p}{\log p} \right)^p \max \left[\left(\sum_{i=1}^n E \zeta_i \right)^p, \sum_{i=1}^n E \zeta_i^p \right] \quad (2.4)$$

(Johnson, Schechtman and Zinn (1985), Theorem 2.5).

Lemma 2.2. *If $r \geq 2$ and F is a continuous function of two real variables, non-decreasing in each of them separately, and such that the inequality*

$$E \left| \sum_{i=1}^n \xi_i \right|^r \leq F \left(nE\xi^2, nE|\xi|^r \right), \quad (2.5)$$

holds for all $n \in \mathbf{N}$ and all sequences of independent, identically distributed and centered random variables ξ, ξ_1, ξ_2, \dots , then the inequality

$$E \left| \sum_{i=0}^{\eta} \zeta_i - \lambda E\zeta \right|^r \leq F \left(\lambda E\zeta^2, \lambda E|\zeta|^r \right), \quad (2.6)$$

holds for any $\lambda > 0$, any sequence of independent identically distributed random variables $\zeta, \zeta_1, \zeta_2, \dots$, a Poisson random variable η with mean λ independent of the variables $\{\zeta_i\}_{i=1}^{\infty}$, and $\zeta_0 = 0$.

Proof. The distribution of $\sum_{i=0}^{\eta} \zeta_i - \lambda E\zeta$ is infinitely divisible, in fact, for any $N \in \mathbf{N}$, we can write

$$\sum_{i=0}^{\eta} \zeta_i - \lambda E\zeta = \sum_{i=1}^N \left(\omega_i - \frac{\lambda}{N} E\zeta \right),$$

where $\omega, \omega_1, \omega_2, \dots$, (dependent on N) are i.i.d. with

$$\omega =_d \sum_{i=0}^{\eta_N} \zeta_i, \quad \eta_N =_d \text{Pois}(\lambda/N),$$

and η_N independent of ζ_1, ζ_2, \dots . Clearly, $E\omega_i = \frac{\lambda}{N} E\zeta$. Applying inequality (2.5) to the sequence $\{\omega_i\}$, we get

$$E \left| \sum_{i=0}^{\eta} \zeta_i - \lambda E\zeta \right|^r = E \left| \sum_{i=1}^N \left(\omega_i - \frac{\lambda}{N} E\zeta \right) \right|^r \leq F \left(NE \left(\omega - \frac{\lambda}{N} E\zeta \right)^2, NE \left| \omega - \frac{\lambda}{N} E\zeta \right|^r \right). \quad (2.7)$$

The first argument of F in this inequality is just

$$NE \left(\omega - \frac{\lambda}{N} E\zeta \right)^2 = \lambda E\zeta^2.$$

Regarding the second argument, we have

$$\begin{aligned} E \left| \omega - \frac{\lambda}{N} E\zeta \right|^r &= \sum_{j=0}^{\infty} E \left| \sum_{i=0}^j \zeta_i - \frac{\lambda}{N} E\zeta \right|^r \Pr\{\eta_N = j\} \\ &\leq \left| \frac{\lambda}{N} E\zeta \right|^r + \sum_{j=1}^{\infty} \left(E \left| \sum_{i=1}^j \zeta_i \right|^r + \frac{r\lambda}{N} |E\zeta| E \left| \sum_{i=1}^j \zeta_i \right| + \frac{\lambda}{N} |E\zeta|^{r-1} \right) \Pr\{\eta_N = j\}. \end{aligned}$$

Taking into account inequality (2.4) applied to $\{|\zeta_i|\}$ and that $\Pr\{\eta_N = j\} = e^{-\lambda/N} (\lambda/N)^j / j!$, we obtain that

$$\limsup_{N \rightarrow \infty} N E \left| \omega - \frac{\lambda}{N} E\zeta \right|^r \leq \lambda E |\zeta|^r.$$

Now (2.6) follows by combining these estimates with inequality (2.7). \square

Inequality (2.4) together with the previous lemma gives the following extension of Rosenthal's inequality to Poissonized sums:

Lemma 2.3. *If, for each $n \in \mathbf{N}$, $\zeta, \zeta_1, \zeta_2, \dots, \zeta_n, \dots$, are independent identically distributed random variables, $\zeta_0 = 0$, and η is a Poisson random variable with mean $\lambda > 0$ and independent of the variables $\{\zeta_i\}_{i=1}^{\infty}$, then, for every $p \geq 2$,*

$$E \left| \sum_{i=0}^{\eta} \zeta_i - \lambda E\zeta \right|^p \leq \left(\frac{15p}{\log p} \right)^p \max \left[(\lambda E\zeta^2)^{p/2}, \lambda E|\zeta|^p \right]. \quad (2.8)$$

Moreover, specializing to $\zeta \equiv 1$, we have for every $p \geq 2$,

$$E |\eta - \lambda|^p \leq \left(\frac{15p}{\log p} \right)^p \max \left[\lambda^{p/2}, \lambda \right]. \quad (2.9)$$

establishing f.d. convergence. Here is the basic result that we will apply in order to 'de-Poissonize' in the process of establishing f.d. convergence.

Lemma 2.4. (Beirlant and Mason (1995)) *Let $N_{1,n}$ and $N_{2,n}$ be independent Poisson random variables with $N_{1,n}$ being Poisson $(n(1 - \alpha))$ and $N_{2,n}$ being Poisson $(n\alpha)$ where $\alpha \in (0, 1)$. Denote $N_n = N_{1,n} + N_{2,n}$ and set*

$$U_n = \frac{N_{1,n} - n(1 - \alpha)}{\sqrt{n}} \quad \text{and} \quad V_n = \frac{N_{2,n} - n\alpha}{\sqrt{n}}.$$

Let $\{S_n\}_{n=1}^{\infty}$ be a sequence of random variables such that

- (i) for each $n \geq 1$, the random vector (S_n, U_n) is independent of V_n ,
- (ii) for some $\sigma^2 < \infty$,

$$(S_n, U_n) \rightarrow_w (\sigma Z_1, \sqrt{1 - \alpha} Z_2), \quad \text{as } n \rightarrow \infty.$$

Then, for all x ,

$$\Pr \{S_n \leq x \mid N_n = n\} \rightarrow \Pr \{\sigma Z_1 \leq x\}.$$

We provide a proof here for the convenience of the reader. It is somewhat different than the original one given by Beirlant and Mason (1995).

Proof. Consider the characteristic function

$$\begin{aligned} \phi_n(t, u) &:= E \exp(itS_n + iuN_n) \\ &= \sum_{k=0}^{\infty} e^{iuk} E(\exp(itS_n) \mid N_n = k) \Pr(N_n = k). \end{aligned}$$

From this we see by Fourier inversion that the conditional characteristic function of S_n , given $N_n = n$, is

$$\psi_n(t) := E(\exp(itS_n) \mid N_n = n) = \frac{1}{2\pi \Pr(N_n = n)} \int_{-\pi}^{\pi} e^{-iun} \phi_n(t, u) du.$$

Applying Stirling's formula, we obtain as $n \rightarrow \infty$

$$2\pi \Pr(N_n = n) = 2\pi e^{-n} n^{n-1} / (n-1)! \sim (2\pi/n)^{1/2}, \quad (2.10)$$

which, after changing variables from u to v/\sqrt{n} and using assumption (i), gives

$$\psi_n(t) = (2\pi)^{-1/2} (1 + o(1)) \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} E \exp(itS_n + ivU_n) E \exp(ivV_n) dv.$$

We shall deduce our proof from this expression for the conditional characteristic function $\psi_n(t)$, after we have collected some facts about the asymptotic behavior of the components in $\psi_n(t)$.

First, by assumption (ii)

$$E \exp(itS_n + iuU_n) \rightarrow \phi(t, v), \quad (2.11)$$

where

$$\phi(t, v) = \exp\left(-(\sigma^2 t^2 + (1 - \alpha)v^2)/2\right).$$

Next, the proof of Theorem 3 on pages 490-91 of Feller (1966), Volume II, shows that as $n \rightarrow \infty$

$$\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} |E \exp(ivV_n) - \exp(-\alpha v^2/2)| dv + \int_{|v| > \pi\sqrt{n}} \exp(-\alpha v^2/2) dv \rightarrow 0,$$

which implies that

$$\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} |E \exp(itS_n + iuU_n) (E \exp(ivV_n) - \exp(-\alpha v^2/2))| dv \rightarrow 0. \quad (2.12)$$

Now by putting (2.11) and (2.12) together with the Lebesgue dominated convergence theorem, we get

$$\psi_n(t) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} \phi(t, v) \exp(-\alpha v^2/2) dv = \exp(-\sigma^2 t^2/2).$$

□

Remark 2.2. Note that the version of Lemma 2.4 given in Beirlant and Mason (1995) is a bit more general. It allows the limiting bivariate normal random variable in (ii) to have non-zero correlation. However, the version given above suffices for our needs. Lemma 2.4 is not original to Beirlant and Mason. A preliminary version was proved by Beirlant, Györfi and Lugosi (1994). A similar partial inversion is used by Holst (1979). The idea of partial inversion, itself, goes back at least to Bartlett (1938).

Remark 2.3. Lemma 2.4 can be generalized to distributions other than the Poisson. In particular, let X_1, X_2, \dots , be i.i.d. integer valued random variables with mean 1 and variance 1 and set for any $0 < \alpha < 1$,

$$N_n = \sum_{i=1}^n X_i, \quad N_{2,n} = \sum_{i=1}^{[n\alpha]} X_i \quad \text{and} \quad N_{1,n} = N_n - N_{2,n},$$

then by using Theorem 3 on page 490 of Feller (1966) one can repeat the proof of Lemma 2.4 to show that it remains true with these definitions of N_n , $N_{1,n}$ and $N_{2,n}$.

3 Moments of the increments of the L_1 -norm kernel density estimator process.

This section contains a crucial estimate for the increments of the L_1 -norm kernel density estimator process. Then, tightness will follow by a slight modification of a standard metric entropy argument. Given a density f on \mathbf{R} , let $N_0 = N_0(f) < \infty$ be such that $\sup_{i \in \mathbf{N}} \int_{ih_n}^{(i+1)h_n} f(x) dx \leq 1/2$ for all $n > N_0$. Such an N_0 exists by absolute continuity if $h_n \rightarrow 0$ as $n \rightarrow \infty$, which we assume throughout.

Proposition 3.1. *Let K_i , $i = 1, 2$, be two bounded kernels satisfying (1.2) and such that*

$$\partial_2(K_1, K_2) \geq \frac{\rho}{n^{1/2}} \tag{3.1}$$

for some $\rho > 0$ and $n > N_0$. Set $\kappa = \max(\|K_1\|_\infty, \|K_2\|_\infty)$ and $C_1 := C_1(\kappa, \rho) = \sup_{m \geq 2} (1 + (2\kappa/\rho)^{m-2})^{1/m}$. Then,

$$E \exp \left\{ t \frac{|\xi_n(K_1) - \xi_n(K_2)|}{C_1 \partial_2(K_1, K_2)} \right\} \leq 2 \exp \left\{ \sum_{m=2}^{\infty} \left(\frac{60et}{\log m} \right)^m \right\} \tag{3.2}$$

for all $t \geq 0$.

Proof. Let $X, X_1, X'_1, X_2, X'_2, \dots$, be i.i.d. random variables. Given two bounded, integrable kernels K_1, K_2 , and $n > N_0$ we set, for $x \in \mathbf{R}$,

$$\begin{aligned} \Delta_n(x) &= \left| \frac{1}{h_n \sqrt{n}} \sum_{i=1}^n \left\{ K_1 \left(\frac{X_i - x}{h_n} \right) - EK_1 \left(\frac{X - x}{h_n} \right) \right\} \right| \\ &\quad - \left| \frac{1}{h_n \sqrt{n}} \sum_{i=1}^n \left\{ K_2 \left(\frac{X_i - x}{h_n} \right) - EK_2 \left(\frac{X - x}{h_n} \right) \right\} \right|, \end{aligned}$$

and $\Delta_0(x) = 0$. Further, for $n \geq 1$, we let η_n be a Poisson random variable with mean n , independent of X_1, X_2, \dots , and set

$$\begin{aligned} \Delta_{\eta_n}(x) &= \left| \frac{1}{h_n \sqrt{n}} \sum_{i=1}^{\eta_n} K_1 \left(\frac{X_i - x}{h_n} \right) - nEK_1 \left(\frac{X - x}{h_n} \right) \right| \\ &\quad - \left| \frac{1}{h_n \sqrt{n}} \sum_{i=1}^{\eta_n} K_2 \left(\frac{X_i - x}{h_n} \right) - nEK_2 \left(\frac{X - x}{h_n} \right) \right|. \end{aligned}$$

Then,

$$\xi_n(K_1) - \xi_n(K_2) = \int_{\mathbf{R}} (\Delta_n(x) - E\Delta_n(x)) dx, \quad (3.3)$$

and we define, for convenience,

$$\bar{\xi}_n(K_1) - \bar{\xi}_n(K_2) = \int_{\mathbf{R}} (\Delta_n(x) - E\Delta_{\eta_n}(x)) dx. \quad (3.4)$$

Let $\Delta'_n(x)$ be defined as $\Delta_n(x)$ using the X'_1, X'_2, \dots , variables, and let E' denote integration with respect to the variables X'_i only. Then, for $\lambda \in \mathbf{R}$,

$$\begin{aligned} E \exp \left\{ \lambda |\xi_n(K_1) - \xi_n(K_2)| \right\} &= E \exp \left\{ \lambda \left| \int_{\mathbf{R}} (\Delta_n(x) - E\Delta_n(x)) dx \right| \right\} \\ &\leq \exp \left\{ \lambda \left| \int_{\mathbf{R}} (E\Delta_n(x) - E\Delta_{\eta_n}(x)) dx \right| \right\} E \exp \left\{ \lambda |\bar{\xi}_n(K_1) - \bar{\xi}_n(K_2)| \right\}, \end{aligned}$$

and, by Jensen,

$$\begin{aligned} \exp \left\{ \lambda \left| \int_{\mathbf{R}} (E\Delta_n(x) - E\Delta_{\eta_n}(x)) dx \right| \right\} &= \exp \left\{ \lambda \left| \int_{\mathbf{R}} (E'\Delta'_n(x) - E\Delta_{\eta_n}(x)) dx \right| \right\} \\ &\leq E' \exp \left\{ \lambda \left| \int_{\mathbf{R}} (\Delta'_n(x) - E\Delta_{\eta_n}(x)) dx \right| \right\} \\ &= E \exp \left\{ \lambda |\bar{\xi}_n(K_1) - \bar{\xi}_n(K_2)| \right\}. \end{aligned}$$

Hence,

$$\begin{aligned} E \exp \left\{ \lambda |\xi_n(K_1) - \xi_n(K_2)| \right\} &\leq \left(E \exp \left\{ \lambda |\bar{\xi}_n(K_1) - \bar{\xi}_n(K_2)| \right\} \right)^2 \\ &\leq E \exp \left\{ 2\lambda |\bar{\xi}_n(K_1) - \bar{\xi}_n(K_2)| \right\}. \end{aligned} \quad (3.5)$$

Let \mathcal{I}_s , $s = 1, \dots, 6$, be a partition of the integers \mathbf{Z} such that:

i) if $i \neq j \in \mathcal{I}_s$ then $|i - j| \geq 2$, and

ii) for every $s = 1, \dots, 6$, $\sum_{i \in \mathcal{I}_s} \Pr\{X \in ((i - 1/2)h_n, (i + 3/2)h_n]\} \leq 1/2$,

and set

$$A_s = \cup_{i \in \mathcal{I}_s} (ih_n, (i + 1)h_n], \quad s = 1, \dots, 6,$$

and

$$B_s = \cup_{i \in \mathcal{I}_s} ((i - 1/2)h_n, (i + 3/2)h_n], \quad s = 1, \dots, 6.$$

Then, condition (ii) becomes $\Pr\{X \in B_s\} \leq 1/2$. Note that the sets A_s partition \mathbf{R} and that if $x \in A_s$ and $X \notin B_s$ then $K((X - x)/h_n) = 0$. (To see that such a partition exists for $n > N_0$, take first $\mathcal{I}'_0 = \{2n : n \in \mathbf{N}\}$; then the corresponding extended set B'_0 coincides with \mathbf{R} ; by further decomposing \mathcal{I}'_0 into three parts if necessary, we get three sets \mathcal{I}_s , $s = 1, 2, 3$, whose union is \mathcal{I}'_0 and whose corresponding B_s satisfy $\Pr\{X \in B_s\} \leq 1/2$; the same can be done with $\mathcal{I}'_1 = \{2n + 1 : n \in \mathbf{N}\}$.) Then, we have

$$|\bar{\xi}_n(K_1) - \bar{\xi}_n(K_2)| \leq \sum_{s=1}^6 \left| \int_{A_s} (\Delta_n(x) - E\Delta_{\eta_n}(x)) dx \right|, \quad (3.6)$$

and, by Hölder's inequality,

$$E \exp \left\{ 2\lambda \left| \bar{\xi}_n(K_1) - \bar{\xi}_n(K_2) \right| \right\} \leq \prod_{s=1}^6 \left(E \exp \left\{ 12\lambda \left| \int_{A_s} (\Delta_n(x) - E\Delta_{\eta_n}(x)) dx \right| \right\} \right)^{1/6}. \quad (3.7)$$

Let us fix s . Since for every measurable function g ,

$$\int_{A_s} g \left(\sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) \right) dx = \int_{A_s} g \left(\sum_{i=1}^n I(X_i \in B_s) K \left(\frac{X_i - x}{h_n} \right) \right) dx$$

(if the integrals exist), it follows from the properties of B_s that we can apply Lemma 2.1 (see also Remark 2.1 below its proof) to conclude

$$\begin{aligned} & E \exp \left\{ 12\lambda \left| \int_{A_s} (\Delta_n(x) - E\Delta_{\eta_n}(x)) dx \right| \right\} \\ & \leq 2E \exp \left\{ 12\lambda \left| \int_{A_s} (\Delta_{\eta_n}(x) - E\Delta_{\eta_n}(x)) dx \right| \right\}. \end{aligned} \quad (3.8)$$

Since the intervals building up B_s are disjoint, it follows from the properties of the Poissonized process that the random variables $\delta_{j,n}$, $j \in \mathcal{I}_s$, defined as

$$\delta_{j,n} := \int_{jh_n}^{(j+1)h_n} (\Delta_{\eta_n}(x) - E\Delta_{\eta_n}(x)) dx, \quad (3.9)$$

are independent. Therefore, since $e^{|x|} \leq e^x + e^{-x}$,

$$\begin{aligned} & E \exp \left\{ 12\lambda \left| \int_{A_s} (\Delta_{\eta_n}(x) - E\Delta_{\eta_n}(x)) dx \right| \right\} \\ & \leq \prod_{j \in \mathcal{I}_s} E \exp \{ 12\lambda \delta_{j,n} \} + \prod_{j \in \mathcal{I}_s} E \exp \{ -12\lambda \delta_{j,n} \} \end{aligned} \quad (3.10)$$

for any $\lambda > 0$. To estimate the right hand side of (3.10), we will obtain bounds on $E|\delta_{j,n}|^m$ for $m \geq 2$ (note that $E\delta_{j,n} = 0$).

Let Δ'_{η_n} be an independent copy of Δ_{η_n} . By Jensen's inequality and the generalized Minkowski inequality (e.g., Folland (1999), page 194),

$$E|\delta_{j,n}|^m \leq E \left| \int_{jh_n}^{(j+1)h_n} (\Delta_{\eta_n}(x) - \Delta'_{\eta_n}(x)) dx \right|^m \leq \left[\int_{jh_n}^{(j+1)h_n} (E|\Delta_{\eta_n}(x) - \Delta'_{\eta_n}(x)|^m)^{1/m} dx \right]^m.$$

Now,

$$\begin{aligned} & E \left| \Delta_{\eta_n}(x) - \Delta'_{\eta_n}(x) \right|^m \leq 2^m E |\Delta_{\eta_n}(x)|^m \\ &= \frac{2^m}{(\sqrt{n}h_n)^m} E \left\| \sum_{i=0}^{\eta_n} K_1 \left(\frac{X_i - x}{h_n} \right) - nEK_1 \left(\frac{X - x}{h_n} \right) - \sum_{i=0}^{\eta_n} K_2 \left(\frac{X_i - x}{h_n} \right) + nEK_2 \left(\frac{X - x}{h_n} \right) \right\|^m \\ &\leq \frac{2^m}{(\sqrt{n}h_n)^m} E \left\| \sum_{i=0}^{\eta_n} (K_1 - K_2) \left(\frac{X_i - x}{h_n} \right) - nE(K_1 - K_2) \left(\frac{X - x}{h_n} \right) \right\|^m, \end{aligned}$$

which by Lemma 2.3 is

$$\begin{aligned} &\leq \left(\frac{30m}{h_n \log m} \right)^m \left[\left(E(K_1 - K_2)^2 \left(\frac{X - x}{h_n} \right) \right)^{m/2} \right. \\ &\quad \left. + \frac{1}{n^{m/2-1}} E \left| (K_1 - K_2) \left(\frac{X - x}{h_n} \right) \right|^m \right]. \end{aligned} \quad (3.11)$$

Thus, by Jensen,

$$\begin{aligned} E|\delta_{j,n}|^m &\leq \left(\frac{60m}{\log m} \right)^m \left[\left(\int_{jh_n}^{(j+1)h_n} \frac{1}{h_n} E(K_1 - K_2)^2 \left(\frac{X - x}{h_n} \right) dx \right)^{m/2} \right. \\ &\quad \left. + \frac{1}{n^{m/2-1}} \int_{jh_n}^{(j+1)h_n} \frac{1}{h_n} E \left| (K_1 - K_2) \left(\frac{X - x}{h_n} \right) \right|^m dx \right]. \end{aligned} \quad (3.12)$$

It follows from these estimates, from $E\delta_{j,n} = 0$ and from Stirling's formula, that

$$\begin{aligned} E \exp \{ \pm 12\lambda \delta_{j,n} \} &\leq 1 + \sum_{m=2}^{\infty} \left(\frac{720e\lambda}{\log m} \right)^m \left[\left(\int_{jh_n}^{(j+1)h_n} \frac{1}{h_n} E(K_1 - K_2)^2 \left(\frac{X - x}{h_n} \right) dx \right)^{m/2} \right. \\ &\quad \left. + \frac{1}{n^{m/2-1}} \int_{jh_n}^{(j+1)h_n} \frac{1}{h_n} E \left| (K_1 - K_2) \left(\frac{X - x}{h_n} \right) \right|^m dx \right]. \end{aligned}$$

Then, plugging this estimate into (3.10) and using the elementary inequality

$$\prod (1 + x_i) \leq \exp \left(\sum x_i \right), \quad x_i \in \mathbf{R},$$

we obtain

$$\begin{aligned}
& \exp \left\{ 12\lambda \left| \int_{A_s} (\Delta_{\eta_m}(x) - E\Delta_{\eta_m}(x)) dx \right| \right\} \\
& \leq 2 \exp \left\{ \sum_{j \in \mathcal{I}_s} \sum_{m=2}^{\infty} \left(\frac{720e\lambda}{\log m} \right)^m \left[\left(\int_{jh_n}^{(j+1)h_n} \frac{1}{h_n} E(K_1 - K_2)^2 \left(\frac{X-x}{h_n} \right) dx \right)^{m/2} \right. \right. \\
& \quad \left. \left. + \frac{1}{n^{m/2-1}} \int_{jh_n}^{(j+1)h_n} \frac{1}{h_n} E \left| (K_1 - K_2) \left(\frac{X-x}{h_n} \right) \right|^m dx \right] \right\}. \tag{3.13}
\end{aligned}$$

Now, by change of variables,

$$\begin{aligned}
& \sum_{j \in \mathcal{I}_s} \left(\int_{jh_n}^{(j+1)h_n} \frac{1}{h_n} E(K_1 - K_2)^2 \left(\frac{X-x}{h_n} \right) dx \right)^{m/2} \\
& \leq \left(\sum_{j \in \mathcal{I}_s} \int_{jh_n}^{(j+1)h_n} \frac{1}{h_n} E(K_1 - K_2)^2 \left(\frac{X-x}{h_n} \right) dx \right)^{m/2} \\
& \leq \left(E \int_{\mathbf{R}} \frac{1}{h_n} (K_1 - K_2)^2 \left(\frac{X-x}{h_n} \right) dx \right)^{m/2} \\
& = \left(\int_{\mathbf{R}} (K_1 - K_2)^2(x) dx \right)^{m/2} = \partial_2^m(K_1, K_2).
\end{aligned}$$

We now restrict to n such that $\partial_2(K_1, K_2) \geq \rho/n^{1/2}$. For these n , we similarly have

$$\begin{aligned}
\frac{1}{n^{m/2-1}} \sum_{j \in \mathcal{I}_s} \int_{jh_n}^{(j+1)h_n} \frac{1}{h_n} E \left| (K_1 - K_2) \left(\frac{X-x}{h_n} \right) \right|^m dx & \leq \frac{1}{n^{m/2-1}} \partial_m^m(K_1, K_2) \\
& \leq \frac{(2\kappa)^{m-2}}{n^{m/2-1}} \partial_2^2(K_1, K_2) \\
& \leq \left(\frac{2\kappa}{\rho} \right)^{m-2} \partial_2^m(K_1, K_2).
\end{aligned}$$

Then, combining these estimates with (3.13), we obtain

$$E \exp \left\{ 12\lambda \left| \int_{A_s} (\Delta_{\eta_m}(x) - E\Delta_{\eta_m}(x)) dx \right| \right\} \leq 2 \exp \left\{ \sum_{m=2}^{\infty} \left(\frac{720e\lambda C_1(\kappa, \rho) \partial_2(K_1, K_2)}{\log m} \right)^m \right\}.$$

Setting

$$\lambda = \frac{t}{12C_1(\kappa, \rho) \partial_2(K_1, K_2)}$$

in this inequality, and combining it with inequalities (3.7) and (3.5) we obtain (3.2). \square

It is not clear that inequality (3.2) exhibits the best integrability for $|\xi_n(K_1) - \xi_n(K_2)|$, but this seems to be the best order of exponential integrability we can get using Poissonization.

Let Ψ be a Young modulus, that is, a convex increasing unbounded function $\Psi : [0, \infty) \mapsto [0, \infty)$ satisfying $\Psi(0) = 0$. For a random variable X its L_Ψ -Orlicz norm is defined to be

$$\|X\|_\Psi = \inf \left\{ c > 0 : E\Psi \left(\frac{|X|}{c} \right) \leq 1 \right\}.$$

The function

$$\Psi_1(x) = \frac{e^x - 1}{2} \tag{3.14}$$

is a Young modulus. Moreover, it is easy to see that for every $m \geq 1$ there exists $c_m < \infty$ such that the inequality $\|X\|_m \leq c_m \|X\|_{\Psi_1}$ holds for any random variable X . Proposition 3.1 implies the following bound for the Ψ_1 -norm of $\xi_n(K_1) - \xi_n(K_2)$, which will be needed to establish the tightness part of the proof of Theorem 1.1.

Corollary 3.1. *Under the hypotheses of Theorem 1.1, there exists a constant $C(\kappa, \rho) < \infty$ such that, for all $n \geq N_0$,*

$$\|\xi_n(K_1) - \xi_n(K_2)\|_{\Psi_1} \leq C(\kappa, \rho) \partial_2(K_1, K_2). \tag{3.15}$$

Proof. Let t_0 be such that $\sum_{m=2}^\infty (60et_0/\log m)^m = \log(3/2)$. Then, inequality (3.2) yields (3.15) with $C(\kappa, \rho) = C_1(\kappa, \rho)/t_0$. \square

The next moment bound will be useful for proving Theorem 1.2.

Proposition 3.2. *Let K_i , $i = 1, 2$, be two square integrable kernels satisfying (1.2). Then, for some universal constant $C > 0$ and for all $n > N_0$,*

$$E(\xi_n(K_1) - \xi_n(K_2))^2 \leq C \partial_2^2(K_1, K_2).$$

Proof. We keep the same notation as in the proof of Proposition 3.1. Observe that from (3.3) and (3.4) we get

$$\begin{aligned} \xi_n(K_1) - \xi_n(K_2) = \\ \bar{\xi}_n(K_1) - \bar{\xi}_n(K_2) + \int_{\mathbf{R}} E[\Delta_{\eta_n}(x) - \Delta_n(x)] dx, \end{aligned}$$

which by arguing as in the proof of (3.5) gives

$$E(\xi_n(K_1) - \xi_n(K_2))^2 \leq 4E(\bar{\xi}_n(K_1) - \bar{\xi}_n(K_2))^2. \tag{3.16}$$

Next, as in (3.11) but replacing Lemma 2.3 by direct computation, we obtain

$$E(\Delta_{\eta_n}(x) - \Delta'_{\eta_n}(x))^2 \leq \frac{4}{h_n^2} E(K_1 - K_2)^2 \left(\frac{X - x}{h_n} \right),$$

which gives

$$E\delta_{j,n}^2 \leq 4 \int_{jh_n}^{(j+1)h_n} \frac{1}{h_n} E(K_1 - K_2)^2 \left(\frac{X-x}{h_n} \right) dx. \quad (3.17)$$

Thus from (3.6), Lemma 2.1 and using independence we get

$$\begin{aligned} E \left(\bar{\xi}_n(K_1) - \bar{\xi}_n(K_2) \right)^2 &\leq 6 \sum_{s=1}^6 E \left(\int_{A_s} (\Delta_n(x) - E\Delta_n(x)) dx \right)^2 \\ &\leq 12 \sum_{s=1}^6 E \left(\int_{A_s} (\Delta_{\eta_n}(x) - E\Delta_{\eta_n}(x)) dx \right)^2 \\ &= 12 \sum_{s=1}^6 \sum_{j \in \mathcal{I}_s} E\delta_{j,n}^2, \end{aligned} \quad (3.18)$$

which by (3.17) is

$$\leq 48 \int_{\mathbf{R}} \frac{1}{h_n} E(K_1 - K_2)^2 \left(\frac{X-x}{h_n} \right) dx =: \frac{C}{4} \partial_2^2(K_1, K_2). \quad (3.19)$$

Inequalities (3.16) and (3.19) prove the proposition. \square

4 Tightness of general processes.

The standard theorems on tightness of processes using metric entropy (or, equivalently, packing numbers) apply to processes $\{\xi_n(K) : K \in \mathcal{K}\}$ that satisfy inequalities such as (3.15) for all $K_1, K_2 \in \mathcal{K}$, whereas in our case such an inequality holds only for K_1 and K_2 not too close in the ∂_2 pseudometric, namely, satisfying inequality (3.1). The observation we will make in order to deal with this problem applies to any of several entropy bounds for moments or for Orlicz norms in the literature (among them, for instance, the oldest one, valid only for Orlicz norms of exponential type, as e.g. in de la Peña and Giné (1999), Theorems 5.1.4, 5.1.5; or the ones coming from Pisier's improvement on chaining as e.g. Theorem 2.2.4 in van der Vaart and Wellner (1996), or its modification for any moduli based on Pisier's maximal inequality, as indicated in the notes on page 269 of van der Vaart and Wellner, loc. cit.). Here we will only consider the bound for exponential moduli as it is the one we use.

We say that a Young modulus is of *exponential type* if the following two conditions are satisfied:

$$\limsup_{x \wedge y \rightarrow \infty} \frac{\Psi^{-1}(xy)}{\Psi^{-1}(x)\Psi^{-1}(y)} < \infty, \text{ and } \limsup_{x \rightarrow \infty} \frac{\Psi^{-1}(x^2)}{\Psi^{-1}(x)} < \infty.$$

Note that Ψ_1 satisfies these conditions (as $\Psi_1^{-1}(x) = \log(2x+1)$).

Proposition 4.1. *Let Ψ be a Young modulus of exponential type, let (T, d) be a totally bounded pseudometric space and let $\{X_t : t \in T\}$ be a stochastic process indexed by T , with the property that there exist $C < \infty$ and $0 < \gamma < \text{diam}(T)$ such that*

$$\|X_s - X_t\|_\Psi \leq Cd(s, t), \quad (4.1)$$

whenever $\gamma \leq d(s, t) < \text{diam}(T)$. Then, there exists a constant L depending only on Ψ such that, for any $\gamma < \delta \leq \text{diam}(T)$

$$\left\| \sup_{d(s,t) \leq \delta} |X_s - X_t| \right\|_\Psi^* \leq 2 \left\| \sup_{d(s,t) \leq \gamma/2} |X_s - X_t| \right\|_\Psi^* + CL \int_{\gamma/2}^\delta \Psi^{-1}(D(T, d, \varepsilon)) d\varepsilon \quad (4.2)$$

Proof. Let T_γ be a maximal subset of T satisfying $d(s, t) \geq \gamma$ for $s \neq t \in T_\gamma$. Then, $\text{Card}(T_\gamma) = D(T, d, \gamma)$. If $s, t \in T$ and $d(s, t) \leq \delta$, let s_γ and t_γ be points in T_γ such that $d(s, s_\gamma) < \gamma$ and $d(t, t_\gamma) < \gamma$, which exist by the maximality property of T_γ . Then, $d(s_\gamma, t_\gamma) < \delta + 2\gamma < 3\delta$. Since

$$|X_s - X_t| \leq |X_s - X_{s_\gamma}| + |X_t - X_{t_\gamma}| + |X_{s_\gamma} - X_{t_\gamma}|,$$

we obtain

$$\left\| \sup_{d(s,t) \leq \delta} |X_s - X_t| \right\|_\Psi^* \leq 2 \left\| \sup_{d(s,t) < \gamma} |X_s - X_t| \right\|_\Psi^* + \left\| \max_{d(s,t) < 3\delta, s, t \in T_\gamma} |X_s - X_t| \right\|_\Psi. \quad (4.3)$$

Now, the process X_s restricted to the finite set T_γ satisfies inequality (4.1) for all $s, t \in T_\gamma$, and therefore we can apply to the restriction to T_γ of X_s/C the entropy bound in Theorem 5.1.4 of de la Peña and Giné (1999) (see also (5.1.10) there) to the effect that

$$\left\| \max_{d(s,t) < 3\delta, s, t \in T_\gamma} |X_s - X_t|/C \right\|_\Psi \leq L \int_0^{3\delta} \psi^{-1}(D(T_\gamma, d, \varepsilon)) d\varepsilon \leq 3L \int_0^\delta \psi^{-1}(D(T_\gamma, d, \varepsilon)) d\varepsilon, \quad (4.4)$$

where L is a constant that depends only on Ψ . Now we note that $D(T_\gamma, d, \varepsilon) \leq D(T, d, \varepsilon)$ for all $\varepsilon > 0$ and that, moreover, $D(T_\gamma, d, \varepsilon) = \text{Card}(T_\gamma) = D(T, d, \gamma)$ for all $\varepsilon \leq \gamma$. Hence,

$$\int_0^\delta \Psi^{-1}(D(T_\gamma, d, \varepsilon)) d\varepsilon \leq \gamma \Psi^{-1}(D(T, d, \gamma)) + \int_\gamma^\delta \Psi^{-1}(D(T, d, \varepsilon)) d\varepsilon \leq 3 \int_{\gamma/2}^\delta \Psi^{-1}(D(T, d, \varepsilon)) d\varepsilon,$$

and this, in combination with the previous inequalities (4.3) and (4.4), gives the proposition. \square

Proposition 4.1 constitutes an example of restricted or stopped chaining. Giné and Zinn (1984) use restricted chaining with $\gamma = n^{-1/4}$ at stage n , whereas for the processes we have in mind we will use $\gamma = \rho n^{-1/2}$, ρ arbitrary.

5 Proof of the tightness part of Theorem 1.1.

The following corollary of the above lemmas and propositions obviously implies the asymptotic equicontinuity (or tightness) of the sequence of processes $\{\xi_n(K) : K \in \mathcal{K}\}$ from Theorem 1.1. Recall the definition of N_0 given at the beginning of Section 6.

Theorem 5.1. *If the class of kernels \mathcal{K} is bounded by a constant κ , satisfies (1.2) and the entropy condition*

$$\int_0^\infty \log N(\mathcal{K}, \partial_2, \varepsilon) d\varepsilon < \infty, \quad (5.1)$$

then

$$\lim_{\delta \rightarrow 0} \sup_{n > N_0} \left\| \sup_{\partial_2(K_1, K_2) \leq \delta, K_1, K_2 \in \mathcal{K}} |\xi_n(K_1) - \xi_n(K_2)| \right\|_{\Psi_1}^* = 0, \quad (5.2)$$

Proof. Corollary 3.1 and Proposition 4.1 give that, for $n > N_0$,

$$\begin{aligned} & \left\| \sup_{\partial_2(K_1, K_2) \leq \delta, K_1, K_2 \in \mathcal{K}} |\xi_n(K_1) - \xi_n(K_2)| \right\|_{\Psi_1}^* \\ & \leq 2 \left\| \sup_{\partial_2(K_1, K_2) \leq \rho/(2n^{1/2}), K_1, K_2 \in \mathcal{K}} |\xi_n(K_1) - \xi_n(K_2)| \right\|_{\Psi_1}^* + C(\rho, \kappa)L \int_0^\delta \Psi_1^{-1}(D(\mathcal{K}, \partial_2, \varepsilon)) d\varepsilon. \end{aligned} \quad (5.3)$$

The entropy hypothesis (5.1) on \mathcal{K} readily implies that the second term at the right hand side of this inequality tends to zero as $\delta \rightarrow 0$. As for the first term, we note that, for all $n \in \mathbf{N}$ and K_1 and K_2 satisfying $\partial_2(K_1, K_2) \leq \rho/(2n^{1/2})$, we have

$$\begin{aligned} |\xi_n(K_1) - \xi_n(K_2)| & \leq \sqrt{n} \int_{\mathbf{R}} |f_{n, K_1}(x) - f_{n, K_2}(x) - E[f_{n, K_1}(x) - f_{n, K_2}(x)]| dx \\ & \quad + \sqrt{n} \int_{\mathbf{R}} E |f_{n, K_1}(x) - f_{n, K_2}(x) - E[f_{n, K_1}(x) - f_{n, K_2}(x)]| dx \\ & \leq 4\sqrt{n} \int_{-1/2}^{1/2} |K_1(x) - K_2(x)| dx \leq 4\sqrt{n} \partial_2(K_1, K_2) \\ & \leq 2\rho. \end{aligned}$$

Therefore, the first term at the right hand side of inequality (5.3) is dominated by $\Psi_1(2\rho) = 2\rho/\log 3$. Hence, letting first δ tend to zero and then ρ go to zero in (5.3) we obtain the limit (5.2). \square

So, the tightness part of the proof of Theorem 1.1 is completed, and, on the way, we have also shown that the increments of the processes $\{\xi_n(K)\}$ enjoy some uniform exponential integrability.

6 Proof of the finite dimensional convergence part of Theorem 1.1 and proof of Theorem 1.2.

The next theorem gives convergence of the finite dimensional distributions in Theorem 1.1. Its proof will be divided into several lemmas.

Theorem 6.1. *For any Lebesgue density f , any sequence of positive constants $\{h_n\}_{n=1}^\infty$ satisfying $h_n \rightarrow 0$ and $\sqrt{nh_n} \rightarrow \infty$, as $n \rightarrow \infty$, and any finite collection $\mathbf{K} := \{K_1, K_2, \dots, K_m\} \subset \mathcal{K}$, $m \geq 1$, the random vector $\xi_n(\mathbf{K}) = (\xi_n(K_1), \dots, \xi_n(K_m)) \in \mathbf{R}^m$ converges in distribution to a mean zero m -variate normal distribution having the covariance matrix with entries $a_{ls} = \sigma(K_l, K_s)$, $l, s = 1, \dots, m$. Moreover, for all $1 \leq l, s \leq m$,*

$$\lim_{n \rightarrow \infty} \text{Cov}(\xi_n(K_l), \xi_n(K_s)) = \lim_{n \rightarrow \infty} n \text{Cov}(\|f_{n, K_l} - E f_{n, K_l}\|_1, \|f_{n, K_s} - E f_{n, K_s}\|_1) = \sigma(K_l, K_s). \quad (6.1)$$

In particular, for all $1 \leq s \leq m$,

$$\lim_{n \rightarrow \infty} n \text{Var}(\|f_{n, K_s} - E f_{n, K_s}\|) = \sigma^2(K_s). \quad (6.2)$$

The following lemma on convolutions will be crucial for our proof.

Lemma 6.1. *Suppose that \mathcal{H} is a finite class of uniformly bounded real valued functions H , which are equal to zero off a compact interval. Then, for any $H \in \mathcal{H}$,*

$$|f * H_h(z) - J(H) f(z)| \rightarrow 0, \text{ as } h \searrow 0, \text{ for almost all } z \in \mathbf{R}, \quad (6.3)$$

where

$$J(H) = \int_{\mathbf{R}} H(u) du$$

and

$$f * H_h(z) := h^{-1} \int_{\mathbf{R}} f(x) H\left(\frac{z-x}{h}\right) dx.$$

Moreover, for all $0 < \varepsilon < 1$, there exist $M, \nu > 0$ and a Borel set C of finite Lebesgue measure $\lambda(C)$ such that

$$C \subset [-M + \nu, M - \nu], \quad (6.4)$$

$$\int_{|x| > M} f(x) dx = \alpha > 0, \quad (6.5)$$

$$\int_C f(x) dx > 1 - \varepsilon, \quad (6.6)$$

$$f \text{ is bounded, continuous and bounded away from zero on } C, \quad (6.7)$$

and, uniformly in $H \in \mathcal{H}$,

$$\sup_{z \in C} |f * H_h(z) - J(H) f(z)| \rightarrow 0, \text{ as } h \searrow 0. \quad (6.8)$$

Proof. First (6.3) follows from Theorem 3 in Chapter 2 of Devroye and Györfi (1985). Using the continuity of our measure, we may find an interval $[-M, M]$ and a number $\nu > 0$ so that

$$\alpha = \int_{|x|>M} f(x) dx = \frac{\varepsilon}{8}$$

and

$$\int_{|x|>M-\nu} f(x) dx = \frac{\varepsilon}{4}.$$

The rest of the proof is inferred from Lusin's theorem followed by Egorov's theorem (see Dudley (1989), Theorems 7.5.1 and 7.5.2). By Lusin's theorem we can find a Borel set F such that f is continuous on F and

$$\int_F f(x) dx > 1 - \varepsilon/4.$$

Clearly we can extract a compact subset D of \mathbf{R} such that f is bounded, continuous and bounded away from zero on $D \cap F$, and

$$\int_{D \cap F} f(x) dx > 1 - \varepsilon/2.$$

Finally, using (6.3), coupled with Egorov's theorem, we can find a Borel subset C of $[-M + \nu, M - \nu] \cap D \cap F$ such that, (6.4) and (6.7) are satisfied and (6.8) holds uniformly in $H \in \mathcal{H}$. \square

In the proof of Theorem 6.1 we shall apply Lemma 6.1 with

$$\mathcal{H} = \mathcal{H}_0 := \bigcup_{s=1}^m \left\{ K_s, K_s^2, |K_s|^3 \right\}. \quad (6.9)$$

The proof consists of three basic steps. First we "truncate", then we "Poissonize" and finally we "de-Poissonize". Our next lemma provides the truncation step.

Lemma 6.2. *Whenever $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$ we have that for any Borel subset B of \mathbf{R} , any $K \in \mathcal{K}$ and any sequence of functions $a_n \in L_1(\mathbf{R}, \mathcal{B}, \lambda)$,*

$$\limsup_{n \rightarrow \infty} E \left(\sqrt{n} \int_B \{ |f_{n,K}(x) - a_n(x)| - E|f_{n,K}(x) - a_n(x)| \} dx \right)^2 \leq 4 \kappa^2 \int_B f(x) dx,$$

where $\kappa = \|K\|_\infty$.

Proof. Applying the theorem in Pinelis (1994), we get

$$\begin{aligned} & E \left(\sqrt{n} \int_B \{ |f_{n,K}(x) - a_n(x)| - E|f_{n,K}(x) - a_n(x)| \} dx \right)^2 \\ & \leq 4 E \left(\frac{1}{h_n} \int_B \left| K \left(\frac{X-x}{h_n} \right) \right| dx \right)^2 \\ & \leq 4 \|K\|_\infty E \frac{1}{h_n} \int_B \left| K \left(\frac{X-x}{h_n} \right) \right| dx \end{aligned} \quad (6.10)$$

$$\leq 4 \kappa^2 h_n^{-1} \int_B \Pr \{X \in [x - h_n/2, x + h_n/2]\} dx.$$

Now

$$\begin{aligned} & h_n^{-1} \int_B \Pr \{X \in [x - h_n/2, x + h_n/2]\} dx \\ & \leq \int_B f(x) dx + \int_{\mathbf{R}} \left| h_n^{-1} \Pr \{X \in [x - h_n/2, x + h_n/2]\} - f(x) \right| dx. \end{aligned}$$

By a special case of Theorem 1 in Chapter 2 of Devroye and Györfi (1985),

$$\int_{\mathbf{R}} \left| h_n^{-1} \Pr \{X \in [x - h_n/2, x + h_n/2]\} - f(x) \right| dx \rightarrow 0$$

as $n \rightarrow \infty$, which completes the proof of Lemma 6.2. \square

We shall apply Lemma 6.2 in the case where $a_n(x) = E f_{n,K}(x)$. Note that in this situation we could get the same bound with 16 instead of 4 applying Theorem 2.1 of de Acosta (1981). Also see Devroye (1991), who obtains the bound (6.10) with $a_n(x) = f(x)$.

Next we shall "Poissonize". Let η be a Poisson (n) random variable, i.e. a Poisson random variable with mean n , independent of X, X_1, X_2, \dots , and set

$$f_{\eta,K}(x) := \frac{1}{nh_n} \sum_{i=1}^{\eta} K\left(\frac{X_i - x}{h_n}\right), \quad K \in \mathbf{K},$$

where the empty sum is defined to be zero. Notice that

$$E f_{\eta,K}(x) = E f_{n,K}(x) = h_n^{-1} EK\left(\frac{X - x}{h_n}\right), \quad (6.11)$$

$$k_n(K, x) := n \operatorname{Var}(f_{\eta,K}(x)) = h_n^{-2} EK^2\left(\frac{X - x}{h_n}\right) \quad (6.12)$$

and

$$n \operatorname{Var}(f_{n,K}(x)) = h_n^{-2} EK^2\left(\frac{X - x}{h_n}\right) - \left\{ h_n^{-1} EK\left(\frac{X - x}{h_n}\right) \right\}^2. \quad (6.13)$$

Choose any Borel set C with $\lambda(C) < \infty$, satisfying (6.7) and (6.8) with $\mathcal{H} = \mathcal{H}_0$. Clearly for any such set C ,

$$\begin{aligned} & \sup_{x \in C, K \in \mathbf{K}} \left| \sqrt{n \operatorname{Var}(f_{\eta,K}(x))} - \sqrt{n \operatorname{Var}(f_{n,K}(x))} \right| \\ & = \sup_{x \in C, K \in \mathbf{K}} \frac{\sqrt{h_n} (f * K_{h_n})^2}{\sqrt{f * K_{h_n}^2}} = O\left(\sqrt{h_n}\right). \end{aligned} \quad (6.14)$$

We shall require the following fact that follows from Theorem 1 of Sweeting (1977).

Fact 6.1. *Let $(\omega, \zeta, \xi), (\omega_1, \zeta_1, \xi_1), (\omega_2, \zeta_2, \xi_2), \dots$, be a sequence of i.i.d. random vectors such that each component has variance 1, mean 0 and finite absolute moments of the third order. Further, let (Z_1, Z_2, Z_3) be tri-variate normal with mean vector 0, $\operatorname{Var}(Z_1) = \operatorname{Var}(Z_2) = \operatorname{Var}(Z_3) = 1$,*

and with covariance operator coinciding with that of (ω, ζ, ξ) . Then there exist universal positive constants A_1, A_2, A_3 and A_4 such that

$$\left| E \left| \frac{\sum_{i=1}^n \zeta_i}{\sqrt{n}} \right| - E|Z_1| \right| \leq \frac{A_1}{\sqrt{n}} E|\zeta|^3, \quad (6.15)$$

$$\left| E \left| \frac{\sum_{i=1}^n \omega_i}{\sqrt{n}} \cdot \frac{\sum_{i=1}^n \zeta_i}{\sqrt{n}} \right| - E|Z_1 Z_2| \right| \leq \frac{A_2}{\sqrt{n}} (E|\omega|^3 + E|\zeta|^3) \quad (6.16)$$

and

$$\left| E \left[\left| \frac{\sum_{i=1}^n \omega_i}{\sqrt{n}} \right| \cdot \left| \frac{\sum_{i=1}^n \zeta_i}{\sqrt{n}} \right| \right] \right| \leq \frac{A_3}{\sqrt{n}} (E|\omega|^3 + E|\zeta|^3). \quad (6.17)$$

The next lemma shows that the centering part of $\xi_n(K)$ is asymptotically equivalent to its Poissonized counterpart. Recall that Z denotes a standard normal random variable.

Lemma 6.3. *Whenever $h_n \rightarrow 0$, $\sqrt{n}h_n \rightarrow \infty$ and C satisfies (6.7) and (6.8) with $\mathcal{H} = \mathcal{H}_0$, we have, for all $K \in \mathbf{K}$,*

$$\lim_{n \rightarrow \infty} \int_C \left\{ \sqrt{n} E|f_{n,K}(x) - E f_{n,K}(x)| - E|Z| \sqrt{n \text{Var}(f_{n,K}(x))} \right\} dx = 0 \quad (6.18)$$

and

$$\lim_{n \rightarrow \infty} \int_C \left\{ \sqrt{n} E|f_{n,K}(x) - E f_{n,K}(x)| - E|Z| \sqrt{n \text{Var}(f_{n,K}(x))} \right\} dx = 0. \quad (6.19)$$

Proof. Let η_1 denote a Poisson random variable with mean 1, independent of X_1, X_2, \dots , and set

$$Y_n(x) = \left[\sum_{j \leq \eta_1} K \left(\frac{X_j - x}{h_n} \right) - EK \left(\frac{X - x}{h_n} \right) \right] / \sqrt{EK^2 \left(\frac{X - x}{h_n} \right)}. \quad (6.20)$$

Now $\text{Var} Y_n(x) = 1$ and for some constant $A > 0$ independent of Y_n and x

$$E|Y_n(x)|^3 \leq A \frac{h_n^{-3/2} E \left| K \left(\frac{X-x}{h_n} \right) \right|^3}{\left(h_n^{-1} EK^2 \left(\frac{X-x}{h_n} \right) \right)^{3/2}}.$$

Using (6.8) and the fact that our assumption on f implies that for some $\delta > 0$, $f(x) \geq \delta$ for all $x \in C$, we get that for all large enough n uniformly in $x \in C$ for some constant $B_0 > 0$,

$$\sup_{x \in C} E|Y_n(x)|^3 \leq h_n^{-1/2} B_0. \quad (6.21)$$

Let $Y_n^{(1)}(x), \dots, Y_n^{(n)}(x)$ be i.i.d. $Y_n(x)$. Clearly

$$T_\eta^K(x) := \frac{\sqrt{n} \{f_{\eta,K}(x) - E f_{\eta,K}(x)\}}{\sqrt{h_n^{-2} EK^2 \left(\frac{X-x}{h_n} \right)}} =_d \frac{\sum_{i=1}^n Y_n^{(i)}(x) - n EY_n(x)}{\sqrt{n}}. \quad (6.22)$$

Therefore by (6.15) of Fact 6.1,

$$\sup_{x \in C} \left| \frac{E |\sqrt{n} \{f_{\eta,K}(x) - Ef_{\eta,K}(x)\}|}{\sqrt{h_n^{-2} EK^2 \left(\frac{X-x}{h_n}\right)}} - E|Z| \right| \leq \frac{A_1}{\sqrt{n}} \sup_{x \in C} E|Y_n(x)|^3. \quad (6.23)$$

Now by (6.21), in combination with (6.23) and

$$\sup_{x \in C} \sqrt{h_n^{-2} EK^2 \left(\frac{X-x}{h_n}\right)} = \sup_{x \in C} \sqrt{n \text{Var}(f_{\eta,K}(x))} = O(h_n^{-1/2}), \quad (6.24)$$

we get then that

$$\left| \int_C \left\{ \sqrt{n} E|f_{\eta,K}(x) - Ef_{\eta,K}(x)| - E|Z| \sqrt{n \text{Var}(f_{\eta,K}(x))} \right\} dx \right| = O\left(\frac{1}{\sqrt{nh_n^2}}\right).$$

Similarly one obtains by Fact 6.1

$$\left| \int_C \left\{ \sqrt{n} E|f_{\eta,K}(x) - Ef_{\eta,K}(x)| - E|Z| \sqrt{n \text{Var}(f_{\eta,K}(x))} \right\} dx \right| = O\left(\frac{1}{\sqrt{nh_n^2}}\right),$$

which by (6.14) implies

$$\left| \int_C \left\{ \sqrt{n} E|f_{\eta,K}(x) - Ef_{\eta,K}(x)| - E|Z| \sqrt{n \text{Var}(f_{\eta,K}(x))} \right\} dx \right| = O\left(\frac{1}{\sqrt{nh_n^2}} + \sqrt{h_n}\right).$$

□

Lemma 6.4. *Whenever $h_n \rightarrow 0$, $nh_n \rightarrow \infty$ and C has finite Lebesgue measure $\lambda(C)$, we have*

$$I_C(x + h_n t) \text{ converges in measure to } I_C(x) = 1 \text{ on } C \times [-1, 1] \quad (6.25)$$

and

$$f(x + h_n t) I_C(x + h_n t) \text{ converges in measure to } f(x) \text{ on } C \times [-1, 1], \quad (6.26)$$

as functions of x and t .

Proof. Notice that

$$\int_C \int_{-1}^1 I_C(x + h_n t) dt dx = \int_C \int_{x-h_n}^{x+h_n} h_n^{-1} I_C(y) dy dx.$$

Now by Theorem 3 in Chapter 2 of Devroye and Györfi (1985), applied to $K(x) = I_{[-1,1]}(x)$ and $f(x) = I_C(x)$, for almost every x

$$\frac{1}{2} \int_{x-h_n}^{x+h_n} h_n^{-1} I_C(y) dy \rightarrow I_C(x).$$

Thus by the dominated convergence theorem

$$\int_C \int_{x-h_n}^{x+h_n} h_n^{-1} I_C(y) dy dx \rightarrow 2\lambda(C),$$

which in others words says

$$\begin{aligned} & \lambda \{(x, t) \in C \times [-1, 1] : 1 - I_C(x + h_n t) \neq 0\} \\ &= 2\lambda(C) - \int_C \int_{-1}^1 I_C(x + h_n t) dt dx \rightarrow 0. \end{aligned}$$

yielding (6.25).

To prove (6.26) just note that, by continuity of f on C ,

$$(f(x + h_n t) - f(x)) I_C(x + h_n t) \rightarrow 0$$

for all $(x, t) \in C \times [-1, 1]$, and that, by (6.25),

$$f(x) (I_C(x + h_n t) - I_C(x)) \rightarrow 0$$

in measure on $C \times [-1, 1]$. \square

Set, for $K, K_1, K_2 \in \mathcal{K}$,

$$\sigma_n(C, K_1, K_2) := n E \prod_{s=1,2} \left(\int_C \{|f_{\eta, K_s}(x) - E f_{\eta, K_s}(x)| - E |f_{\eta, K_s}(x) - E f_{\eta, K_s}(x)|\} dx \right), \quad (6.27)$$

$$\begin{aligned} & \sigma_n^2(C, K) := \sigma_n(C, K, K) \\ &= E \left(\sqrt{n} \int_C \{|f_{\eta, K}(x) - E f_{\eta, K}(x)| - E |f_{\eta, K}(x) - E f_{\eta, K}(x)|\} dx \right)^2 \end{aligned} \quad (6.28)$$

and

$$P(C) = \int_C f(x) dx = \Pr \{X \in C\}. \quad (6.29)$$

Lemma 6.5. *Whenever $h_n \rightarrow 0$, $nh_n \rightarrow \infty$ and C satisfies (6.7) and (6.8) with $\mathcal{H} = \mathcal{H}_0$, we have, for $K_l, K_s \in \mathbf{K}$, $l, s = 1, \dots, m$,*

$$\lim_{n \rightarrow \infty} \sigma_n(C, K_l, K_s) = P(C) \sigma(K_l, K_s). \quad (6.30)$$

In particular, for $K \in \mathbf{K}$,

$$\lim_{n \rightarrow \infty} \sigma_n^2(C, K) = P(C) \sigma^2(K). \quad (6.31)$$

Proof. Without loss of generality we shall only consider the case $l = 1$, $s = 2$. Notice that whenever $|x - y| > h_n$, the random variables $|f_{\eta, K_1}(x) - E f_{\eta, K_1}(x)|$ and $|f_{\eta, K_2}(y) - E f_{\eta, K_2}(y)|$

are independent. This follows from the fact that they are functions of independent increments of a Poisson process with intensity nf . Therefore

$$\begin{aligned}\sigma_n(C, K_1, K_2) &= n \int_C \int_C E\{|f_{\eta, K_1}(x) - Ef_{\eta, K_1}(x)| \cdot |f_{\eta, K_2}(y) - Ef_{\eta, K_2}(y)|\} dx dy \\ &\quad - n \int_C \int_C \{E|f_{\eta, K_1}(x) - Ef_{\eta, K_1}(x)| E|f_{\eta, K_2}(y) - Ef_{\eta, K_2}(y)|\} dx dy \\ &= \int_C \int_C I(|x - y| \leq h_n) \text{Cov}(|T_\eta^{K_1}(x)|, |T_\eta^{K_2}(y)|) \sqrt{k_n(K_1, x) k_n(K_2, y)} dx dy.\end{aligned}$$

Now let Z_1 and Z_2 be independent standard normal random variables and set

$$(Z_n(x), Z_n(y)) = \left(\sqrt{1 - (\rho_n^*(x, y))^2} Z_1 + \rho_n^*(x, y) Z_2, Z_2 \right),$$

with

$$\rho_n^*(x, y) := ET_\eta^{K_1}(x) T_\eta^{K_2}(y).$$

By an application of Fact 6.1, keeping (6.12), (6.7), (6.8), (6.18), (6.21) and (6.22) in mind, we see that

$$\begin{aligned}\sigma_n(C, K_1, K_2) &= \int_C \int_C I(|x - y| \leq h_n) \text{Cov}(|Z_n(x)|, |Z_n(y)|) h_n^{-1} \|K_1\|_2 \|K_2\|_2 \sqrt{f(x) f(y)} dx dy \\ &\quad + O\left(\int_C \int_C I(|x - y| \leq h_n) dx dy \frac{1}{h_n \sqrt{nh_n}} + o(1)\right),\end{aligned}$$

which by the change of variables $y = x + th_n$ and using the fact that $\lambda(C) < \infty$, equals

$$\tau_n^2(C, K) + o(1),$$

where

$$\tau_n^2(C, K) := \int_C \int_{-1}^1 g_n(x, t) dx dt,$$

with

$$g_n(x, t) := I_C(x) I_C(x + th_n) \text{Cov}(|Z_n(x)|, |Z_n(x + th_n)|) \|K_1\|_2 \|K_2\|_2 \sqrt{f(x) f(x + th_n)}.$$

Also observe that

$$\rho_n^*(x, x + th_n) = \frac{h_n^{-1} E \left[K_1 \left(\frac{X-x}{h_n} \right) K_2 \left(\frac{X-x}{h_n} + t \right) \right]}{\sqrt{h_n^{-1} E K_1^2 \left(\frac{X-x}{h_n} \right) h_n^{-1} E K_2^2 \left(\frac{X-x}{h_n} + t \right)}}.$$

Applying (6.3) of Lemma 6.1, with $H(u) = K_1(u) K_2(u + t)$, we get for each t , as $n \rightarrow \infty$, for almost every $x \in C$,

$$h_n^{-1} E \left[K_1 \left(\frac{X-x}{h_n} \right) K_2 \left(\frac{X-x}{h_n} + t \right) \right] \rightarrow f(x) \int_{\mathbf{R}} K_1(u) K_2(u + t) du.$$

Moreover, we get with $H(u) = K_1^2(u)$ and $H(u) = K_2^2(u + t)$, respectively, for almost every $x \in C$, both

$$h_n^{-1}EK_1^2\left(\frac{X-x}{h_n}\right) \rightarrow f(x) \|K_1\|_2^2, \quad \text{and} \quad h_n^{-1}EK_2^2\left(\frac{X-x}{h_n} + t\right) \rightarrow f(x) \|K_2\|_2^2.$$

Thus for each t and almost every $x \in C$, as $n \rightarrow \infty$,

$$\rho_n^*(x, x + th_n) \rightarrow \rho(K_1, K_2, t),$$

and

$$\text{Cov}(|Z_n(x)|, |Z_n(x + th_n)|) \rightarrow \text{Cov}\left(\left|\sqrt{1 - \rho^2(K_1, K_2, t)}Z_1 + \rho(K_1, K_2, t)Z_2\right|, |Z_2|\right).$$

Combining these observations with Lemma 6.4, we readily conclude that $g_n(x, t)$ converges in measure on $C \times [-1, 1]$ to

$$I_C(x) \|K_1\|_2 \|K_2\|_2 \text{Cov}\left(\left|\sqrt{1 - \rho^2(K_1, K_2, t)}Z_1 + \rho(K_1, K_2, t)Z_2\right|, |Z_2|\right) f(x).$$

Since f is bounded on C , the function $g_n(x, t)$ is for all $n \geq 1$ uniformly bounded on $C \times [-1, 1]$. Thus we get by the Lebesgue bounded convergence theorem, as $n \rightarrow \infty$,

$$\tau_n^2(C, K) \rightarrow P(C) \|K_1\|_2 \|K_2\|_2 \int_{-1}^1 \text{Cov}\left(\left|\sqrt{1 - \rho^2(K_1, K_2, t)}Z_1 + \rho(K_1, K_2, t)Z_2\right|, |Z_2|\right) dt,$$

which, since $\rho(K_1, K_2, t) = 0$ whenever $|t| > 1$, equals $P(C) \sigma(K_1, K_2)$. \square

Suppose that our set C satisfies Lemma 6.1 with $\mathcal{H} = \mathcal{H}_0$. Let M, ν, α be the numbers from (6.4) and (6.6). Let, for $K \in \mathcal{K}$,

$$\Delta_n^K(x) := \sqrt{n} \{|f_{\eta, K}(x) - Ef_{\eta, K}(x)| - E|f_{\eta, K}(x) - Ef_{\eta, K}(x)|\}.$$

We shall prove the asymptotic normality of the random vector

$$\Delta_n(\mathbf{K}) := \left(\int_C \Delta_n^{K_1}(x) dx, \dots, \int_C \Delta_n^{K_m}(x) dx\right).$$

It suffices to prove it for any linear combination of the form

$$\mu_1 \int_C \Delta_n^{K_1}(x) dx + \dots + \mu_m \int_C \Delta_n^{K_m}(x) dx,$$

where $\mu_1, \dots, \mu_m \in \mathbf{R}$, without loss of generality, satisfy

$$|\mu_1| + \dots + |\mu_m| = 1.$$

Assume that n is so large that $h_n \leq \nu$ and $h_n \leq M/2$. Define $m_n = [M/h_n]$, $h_n^* = M/m_n$, where $[x]$ denotes the integer part of x . Clearly, we have $M/(2h_n) \leq m_n \leq M/h_n$. Hence,

$$h_n \leq h_n^* \leq 2h_n. \quad (6.32)$$

Set for any integer i

$$\alpha_{i,n} := \frac{\int_{ih_n^*}^{(i+1)h_n^*} I_C(x) \Delta_n(x) dx}{\sigma_n(C, \mu, \mathbf{K})},$$

where

$$\begin{aligned} \Delta_n(x) &:= \mu_1 \Delta_n^{K_1}(x) + \cdots + \mu_m \Delta_n^{K_m}(x), \\ \sigma_n^2(C, \mu, \mathbf{K}) &:= \text{Var} \left(\mu_1 \int_C \Delta_n^{K_1}(x) dx + \cdots + \mu_m \int_C \Delta_n^{K_m}(x) dx \right) = \sum_{l,s=1}^m \mu_l \mu_s \sigma_n(C, K_l, K_s). \end{aligned}$$

By Lemma 6.5, we have

$$\lim_{n \rightarrow \infty} \sigma_n^2(C, \mu, \mathbf{K}) = P(C) \sum_{l,s=1}^m \mu_l \mu_s \sigma(K_l, K_s). \quad (6.33)$$

Therefore we can from now on assume without loss of generality that

$$P(C) \sum_{l,s=1}^m \mu_l \mu_s \sigma(K_l, K_s) > 0. \quad (6.34)$$

Lemma 6.6. *Whenever $h_n \rightarrow 0$, $nh_n \rightarrow \infty$ and C satisfies Lemma 6.1 with $\mathcal{H} = \mathcal{H}_0$, there exists a constant $B_1 > 0$ such that, uniformly in i and for all n sufficiently large,*

$$E|\alpha_{i,n}|^3 \leq B_1 h_n^{3/2}. \quad (6.35)$$

Proof. Notice that

$$\begin{aligned} &\sigma_n^3(C, \mu, \mathbf{K}) E|\alpha_{i,n}|^3 \\ &\leq \int_{I_{i,n}} I_C(x) I_C(y) I_C(z) E|\Delta_n(x) \Delta_n(y) \Delta_n(z)| dx dy dz, \end{aligned}$$

where $I_{i,n} = [ih_n^*, (i+1)h_n^*]^3$.

Clearly,

$$E|\Delta_n(x) \Delta_n(y) \Delta_n(z)| \leq E\{|\Delta_n(x)| + |\Delta_n(y)| + |\Delta_n(z)|\}^3,$$

which by repeated applications of Jensen's inequality is for some constant $D_m > 0$,

$$\leq D_m n^{3/2} \sum_{j=1}^m \left(E|f_{\eta, K_j}(x) - E f_{\eta, K_j}(x)|^3 + E|f_{\eta, K_j}(y) - E f_{\eta, K_j}(y)|^3 + E|f_{\eta, K_j}(z) - E f_{\eta, K_j}(z)|^3 \right).$$

Notice that by Lemma 2.3 we get, for any $1 \leq j \leq m$ and $w \in C$,

$$\begin{aligned} & n^{3/2} E |f_{\eta, K_j}(w) - E f_{\eta, K_j}(w)|^3 \\ & \leq \left(\frac{45}{\log 3} \right)^3 \max \left\{ \left[\frac{1}{h_n^2} E \left(K_j^2 \left(\frac{X-w}{h_n} \right) \right) \right]^{3/2}, E \left(\frac{1}{\sqrt{n} h_n^3} |K_j|^3 \left(\frac{X-w}{h_n} \right) \right) \right\}, \end{aligned}$$

which by Lemma 6.1 with the choice of $\mathcal{H} = \mathcal{H}_0$ is for some constant $B > 0$, uniformly in $1 \leq j \leq m$, $w \in C$ and all $n \geq n_0$, with n_0 large enough,

$$\leq B \left[\frac{1}{h_n^{3/2}} + \frac{1}{\sqrt{n} h_n^2} \right].$$

Thus uniformly in $x, y, z \in C$ and all $n \geq n_0$

$$E |\Delta_n(x) \Delta_n(y) \Delta_n(z)| \leq 3mBD_m \left[\frac{1}{h_n^{3/2}} + \frac{1}{\sqrt{n} h_n^2} \right],$$

which implies that uniformly in i

$$\begin{aligned} & \int_{I_{i,n}} I_C(x) I_C(y) I_C(z) E |\Delta_n(x) \Delta_n(y) \Delta_n(z)| dx dy dz \\ & \leq 6mBD_m \left[h_n^{3/2} + \frac{h_n}{\sqrt{n}} \right]. \end{aligned}$$

This last bound is for some $B_0 > 0$, uniformly in n , less than or equal to $B_0 h_n^{3/2}$. Now, (6.35) follows from (6.33) and (6.34). \square

Our goal now is to set things up to apply Lemma 2.4. Define

$$S_n = \sum_{i=-m_n}^{m_n-1} \alpha_{i,n}, \tag{6.36}$$

$$U_n = \frac{1}{\sqrt{n}} \left\{ \sum_{j=1}^{\eta} I(X_j \in [-M, M]) - n \Pr\{X \in [-M, M]\} \right\} \tag{6.37}$$

and

$$V_n = \frac{1}{\sqrt{n}} \left\{ \sum_{j=1}^{\eta} I(X_j \notin [-M, M]) - n \Pr\{X \notin [-M, M]\} \right\}. \tag{6.38}$$

Clearly (S_n, U_n) is independent of V_n . Recall that we have $C \subset [-M + \nu, M - \nu]$ and $h_n \leq \nu$. Observe that

$$\text{Var}(S_n) = 1 \text{ and } \text{Var}(U_n) = 1 - \alpha, \tag{6.39}$$

where $1 > \alpha = \Pr\{X \notin [-M, M]\} > 0$.

Lemma 6.7. *Whenever $h_n \rightarrow 0$, $nh_n \rightarrow \infty$ and C satisfies Lemma 6.1 with $\mathcal{H} = \mathcal{H}_0$, there exists a constant $B_2 > 0$ such that, for all n sufficiently large,*

$$|\text{Cov}(S_n, U_n)| \leq \frac{B_2}{\sqrt{nh_n}}. \quad (6.40)$$

Proof. Notice that

$$\sigma_n(C, \mu, \mathbf{K}) |\text{Cov}(S_n, U_n)| = \left| \text{Cov} \left(\sum_{i=1}^m \mu_i \sqrt{n} \int_C |f_{\eta, K_i}(x) - Ef_{\eta, K_i}(x)| dx, U_n \right) \right|.$$

Therefore it suffices to show that there exists a constant B_3 such that for any $K \in \mathbf{K}$, for all n sufficiently large

$$\left| \text{Cov} \left(\sqrt{n} \int_C |f_{\eta, K}(x) - Ef_{\eta, K}(x)| dx, U_n \right) \right| \leq \frac{B_3}{\sqrt{nh_n}}. \quad (6.41)$$

Now for any $x \in C$

$$\begin{aligned} & \left| \text{Cov} \left(\sqrt{n} |f_{\eta, K}(x) - Ef_{\eta, K}(x)|, U_n \right) \right| \\ &= \left| \text{Cov} \left(\frac{\sqrt{n} |f_{\eta, K}(x) - Ef_{\eta, K}(x)|}{\sqrt{k_n(K, x)}}, \frac{U_n}{\sqrt{P[-M, M]}} \right) \right| \sqrt{k_n(K, x) P[-M, M]}, \end{aligned}$$

which by (6.17) in Fact 6.1 and Lemma 2.3 is

$$\leq \frac{A_3}{\sqrt{n}} \left[E|Y_n(x)|^3 + \frac{1}{\sqrt{P[-M, M]}} \right] \sqrt{k_n(K, x)}.$$

Notice that by Lemma 6.1 we get for some $B_4 > 0$ for all large enough n

$$\sup_{x \in C} \sqrt{k_n(K, x)} \leq h_n^{-1/2} B_4,$$

which, when combined with (6.21), completes the proof of (6.41). This, in turn, by (6.33) and (6.34) gives (6.40). \square

Hence by (6.39) and (6.40), as $n \rightarrow \infty$,

$$\text{Var}(\lambda_1 S_n + \lambda_2 U_n) \rightarrow \lambda_1^2 + \lambda_2^2(1 - \alpha). \quad (6.42)$$

The proof of the next lemma uses a version of the central limit theorem for 1-dependent random variables, which we state here for the reader's convenience.

Fact 6.2. (Shergin (1979), Corollary 2) *Let $\{X_{i, k_n} : i = 1, \dots, k_n, n \geq 1\}$ denote a triangular array of mean zero 1-dependent random variables such that for all $n \geq 1$*

(i) $\text{Var} \left(\sum_{i=1}^{k_n} X_{i, k_n} \right) \rightarrow 1$ as $n \rightarrow \infty$, and

(ii) for some $2 < s \leq 3$, $\sum_{i=1}^{k_n} E|X_{i,k_n}|^s \rightarrow 0$ as $n \rightarrow \infty$.

Then

$$\sum_{i=1}^{k_n} X_{i,k_n} \rightarrow_w Z.$$

Lemma 6.8. Whenever $h_n \rightarrow 0$, $\sqrt{n}h_n \rightarrow \infty$ and C satisfies Lemma 6.1 with $\mathcal{H} = \mathcal{H}_0$, we have

$$(S_n, U_n) \rightarrow_w (Z_1, \sqrt{1-\alpha}Z_2) \quad (6.43)$$

as $n \rightarrow \infty$.

Proof. We will show that for any λ_1 and λ_2 , as $n \rightarrow \infty$,

$$\lambda_1 S_n + \lambda_2 U_n \rightarrow_w \lambda_1 Z_1 + \lambda_2 \sqrt{1-\alpha} Z_2. \quad (6.44)$$

Set

$$u_{i,n} := \frac{1}{\sqrt{n}} \left\{ \sum_{j=1}^{\eta} I(X_j \in [ih_n^*, (i+1)h_n^*]) - n \Pr\{X \in [ih_n^*, (i+1)h_n^*]\} \right\} \quad (6.45)$$

and

$$y_{i,n} := \lambda_1 \alpha_{i,n} + \lambda_2 u_{i,n}.$$

Now by Jensen's inequality

$$E|y_{i,n}|^3 \leq 4 \left[|\lambda_1|^3 E|\alpha_{i,n}|^3 + |\lambda_2|^3 E|u_{i,n}|^3 \right]. \quad (6.46)$$

By Lemma 6.6,

$$\sum_{i=-m_n}^{m_n-1} E|\alpha_{i,n}|^3 \leq 2m_n B_1 h_n^{3/2} \rightarrow 0. \quad (6.47)$$

Set

$$p_{i,n} = \Pr\{X \in [ih_n^*, (i+1)h_n^*]\}.$$

By Lemma 2.3 there is a universal constant A such that

$$\begin{aligned} \sum_{i=-m_n}^{m_n-1} E|u_{i,n}|^3 &\leq A n^{-3/2} \sum_{i=-m_n}^{m_n-1} \left((np_{i,n})^{3/2} + np_{i,n} \right) \\ &\leq A \max_{-m_n \leq i \leq m_n-1} \left(\sqrt{p_{i,n}} + n^{-1/2} \right) \rightarrow 0. \end{aligned} \quad (6.48)$$

Combining (6.46)–(6.48), we obtain

$$\sum_{i=-m_n}^{m_n-1} E|y_{i,n}|^3 \rightarrow 0.$$

Moreover, note that the sequence $y_{i,n}$, $-m_n \leq i \leq m_n - 1$, is 1-dependent and, by (6.42),

$$\text{Var} \left(\sum_{i=-m_n}^{m_n-1} y_{i,n} \right) \rightarrow \lambda_1^2 + \lambda_2^2(1 - \alpha).$$

Thus we can apply Fact 6.2 to infer that, as $n \rightarrow \infty$,

$$\sum_{i=-m_n}^{m_n-1} y_{i,n} \rightarrow_w \sqrt{\lambda_1^2 + \lambda_2^2(1 - \alpha)} Z.$$

Since

$$\sum_{i=-m_n}^{m_n-1} y_{i,n} = \lambda_1 S_n + \lambda_2 U_n,$$

we can conclude (6.43) by the Cramér–Wold device (e.g. Billingsley (1968)). \square

Set

$$L_n(C) = \frac{\sqrt{n}}{\sigma_n(C, \mu, \mathbf{K})} \sum_{s=1}^m \mu_s \int_C \{|f_{n,K_s}(x) - E f_{n,K_s}(x)| - E|f_{n,K_s}(x) - E f_{n,K_s}(x)|\} dx. \quad (6.49)$$

Lemma 6.9. *Whenever $h_n \rightarrow 0$, $\sqrt{n}h_n \rightarrow \infty$ and C satisfies Lemma 6.1 with $\mathcal{H} = \mathcal{H}_0$, we have*

$$L_n(C) \rightarrow_w Z. \quad (6.50)$$

as $n \rightarrow \infty$.

Proof. Recall the notation in (6.36). Note that

$$S_n = \frac{\sqrt{n}}{\sigma_n(C, \mu, \mathbf{K})} \sum_{s=1}^m \mu_s \int_C \{|f_{\eta,K_s}(x) - E f_{\eta,K_s}(x)| - E|f_{\eta,K_s}(x) - E f_{\eta,K_s}(x)|\} dx$$

and, conditioned on $\eta = n$,

$$S_n =_d \frac{\sqrt{n}}{\sigma_n(C, \mu, \mathbf{K})} \sum_{s=1}^m \mu_s \int_C \{|f_{n,K_s}(x) - E f_{n,K_s}(x)| - E|f_{n,K_s}(x) - E f_{n,K_s}(x)|\} dx.$$

Next by Lemma 6.8, we can apply Lemma 2.4 to S_n and conclude that

$$\frac{\sqrt{n}}{\sigma_n(C, \mu, \mathbf{K})} \sum_{s=1}^m \mu_s \int_C \{|f_{n,K_s}(x) - E f_{n,K_s}(x)| - E|f_{n,K_s}(x) - E f_{n,K_s}(x)|\} dx \rightarrow_w Z.$$

Assertion (6.50) now follows from Lemma 6.3. \square

Completion of the proofs of Theorem 1.1 and Theorem 6.1. To finish the proof, we obtain by a straightforward application of Lemma 6.1 with $\mathcal{H} = \mathcal{H}_0$ a sequence of Borel sets $\{C_k\}_{k \geq 1}$, each with finite Lebesgue measure such that for each $k \geq 1$, both (6.7) and (6.8) hold and

$$\lim_{n \rightarrow \infty} \int_{C_k^c} f(x) dx = 0. \quad (6.51)$$

Notice that for each $k \geq 1$, by Lemma 6.9, as $n \rightarrow \infty$,

$$L_n(C_k) \rightarrow_w Z,$$

and by (6.33),

$$\lim_{n \rightarrow \infty} \sigma_n^2(C_k, \mu, \mathbf{K}) = P(C_k) \sum_{l,s=1}^m \mu_l \mu_s \sigma(K_l, K_s).$$

Further, by Lemma 6.2,

$$\begin{aligned} \limsup_{n \rightarrow \infty} E \left(\sqrt{n} \sum_{s=1}^m \mu_s \int_{C_k^c} \{|f_{n,K_s}(x) - E f_{n,K_s}(x)| - E |f_{n,K_s}(x) - E f_{n,K_s}(x)|\} dx \right)^2 \\ \leq 4m \max_{1 \leq s \leq m} \|K_s\|_\infty^2 \int_{C_k^c} f(x) dx. \end{aligned}$$

Now by (6.51), combined with a standard argument (see Theorem 4.2 of Billingsley (1968)), we conclude, as $n \rightarrow \infty$,

$$\sqrt{n} \sum_{s=1}^m \mu_s \{\|f_{n,K_s} - E f_{n,K_s}\|_1 - E \|f_{n,K_s} - E f_{n,K_s}\|_1\} \rightarrow_w \sqrt{\sum_{l,s=1}^m \mu_l \mu_s \sigma(K_l, K_s)} Z. \quad (6.52)$$

This provides the needed weak convergence by the Cramér–Wold device. Finally, by Theorem 2.3 of Pinelis (1990), we have, for all $r > 2$, $K \in \mathcal{K}$,

$$E \left| \sqrt{n} \{\|f_{n,K} - E f_{n,K}\|_1 - E \|f_{n,K} - E f_{n,K}\|_1\} \right|^r \leq 2^{1+r/2} \Gamma(1+r/2) \|K\|_2^r < \infty,$$

which permits us to infer the relation

$$\lim_{n \rightarrow \infty} \text{Var} \left(\sum_{s=1}^m \mu_s \xi_n(K_s) \right) = \sum_{l,s=1}^m \mu_l \mu_s \sigma(K_l, K_s) \quad (6.53)$$

from (6.52) (see Theorem 6.4 of Billingsley (1968)). Taking into account that

$$\text{Cov}(\xi_n(K_l), \xi_n(K_s)) = \frac{1}{4} \text{Var}(\xi_n(K_l) + \xi_n(K_s)) - \frac{1}{4} \text{Var}(\xi_n(K_l) - \xi_n(K_s)),$$

we derive (6.1) from (6.53). This completes the proof of Theorem 6.1. Now Theorem 1.1 follows from Theorem 5.1 and Theorem 6.1 by well know facts on the weak convergence of processes (e.g., Theorem 5.1.2 in de la Peña and Giné (1999).) \square

The following example shows that not all classes of bounded kernels satisfy Theorem 1.1.

Example 6.1. Let $\mathcal{K} = \{K_\ell := \ell I_{(-1/(2\ell), 1/(2\ell))} : \ell \in \mathbf{N}\}$. We will prove that the sequence of processes $\{\xi_n(K_\ell) : \ell \in \mathbf{N}\}_{n=1}^\infty$ does not converge weakly in $\ell^\infty(\mathbf{N})$. Note that the finite dimensional distributions do converge by Theorem 6.1, hence, if these processes converge then the limiting process is the Gaussian process $\xi(K_\ell), \ell \in \mathbf{N}$, prescribed by Theorem 1.1, and then, in particular, this process must be sample continuous with respect to its intrinsic L_2 -distance. Therefore, it is enough to prove that ξ is not sample continuous. In fact, we show that ξ is not even sample bounded. For this, by Sudakov's minorization (e.g., Ledoux and Talagrand (1991), pp. 79-81), it suffices to see that $N([0, \infty), d_2, \varepsilon) = \infty$ for some $\varepsilon > 0$, where $d_2^2(\ell, s) = E(\xi(K_\ell) - \xi(K_s))^2$. It readily follows from the definition of $\sigma^2(K)$ that

$$\sigma^2(K_\ell) = 2 \int_0^1 \text{Cov} \left(\left| \sqrt{1 - (1-u)^2} Z_1 + (1-u) Z_2 \right|, |Z_2| \right) du < \infty$$

independently of $\ell \in \mathbf{N}$. As indicated in the Introduction, $\sigma^2(K_\ell) > 0$. Moreover, for $s > \ell$,

$$\begin{aligned} \sigma(K_\ell, K_s) &= \left(\sqrt{\frac{s}{\ell}} - \sqrt{\frac{\ell}{s}} \right) \text{cov} \left(\left| \sqrt{1 - \frac{\ell}{s}} Z_1 + \sqrt{\frac{\ell}{s}} Z_2 \right|, |Z_2| \right) \\ &+ 2 \int \frac{\sqrt{\frac{s/\ell + \sqrt{\ell/s}}{2}}}{\sqrt{\frac{s/\ell - \sqrt{\ell/s}}{2}}} \text{Cov} \left(\left| \sqrt{1 - \left[\frac{1}{2} \left(\sqrt{\frac{s}{\ell}} + \sqrt{\frac{\ell}{s}} \right) - u \right]^2} Z_1 + \left[\frac{1}{2} \left(\sqrt{\frac{s}{\ell}} + \sqrt{\frac{\ell}{s}} \right) - u \right] Z_2 \right|, |Z_2| \right) du. \end{aligned}$$

So, $d_2^2(\ell, s) = 2\sigma^2(K_1) - 2\sigma(K_\ell, K_s) = f(s/\ell)$. If we show that $f(u) \rightarrow 2\sigma^2(K_1)$ as $u \rightarrow \infty$, then there will exist a large enough integer A such that $d_2(A^k, A^m) \geq \sigma(K_1) > 0$ for all $1 \leq k < m < \infty$, and therefore, $N([0, \infty), d_2, \sigma(K_1)/2) = \infty$. And this is indeed the case: it is easy to show that $\lim_{s \rightarrow \infty} \sigma(K_1, K_s) = 0$ by a straightforward computation based on the fact that

$$\begin{aligned} \lim_{x \rightarrow \infty} \text{Cov}(|xZ_1 + Z_2|, |Z_2|) &= \lim_{x \rightarrow \infty} E[|Z_2|(|xZ_1 + Z_2| - x|Z_1|)] + \lim_{x \rightarrow \infty} \frac{2}{\pi} (x - \sqrt{x^2 + 1}) \\ &= \lim_{x \rightarrow \infty} E \left[\frac{2xZ_1Z_2|Z_2| + |Z_2|^3}{|xZ_1 + Z_2| + x|Z_1|} \right] \\ &= E \frac{Z_1Z_2|Z_2|}{|Z_1|} = 0, \end{aligned}$$

which is justified by dominated convergence.

6.1 Proof of Theorem 1.2.

Choose any constant $M > 0$ and set $K_M(x) = K(x)I(|K(x)| \leq M)$. Since K_M is bounded, $\xi_n(K_M)$ is asymptotically normal with variance $\sigma^2(K_M)$. Applying Proposition 3.2, we get

$$E(\xi_n(K) - \xi_n(K_M))^2 \leq C \partial_2^2(K_M, K),$$

and clearly

$$\partial_2^2(K_M, K) \rightarrow 0, \text{ as } M \rightarrow \infty.$$

Moreover it can be easily checked that $\|K\|_2 < \infty$ implies $\sigma^2(K) < \infty$. Thus we readily conclude that

$$\sigma^2(K_M) \rightarrow \sigma^2(K) \text{ as } M \rightarrow \infty, \text{ Var } \xi_n(K) \rightarrow \sigma^2(K) \text{ as } n \rightarrow \infty$$

and

$$\xi_n(K) \rightarrow_w \sigma(K)Z \text{ as } n \rightarrow \infty.$$

□

7 Proof of Theorem 1.3.

Replacing $Ef_{n,K}(x)$ by $f(x)$ in Theorem 1.3 is much easier than in Theorems 1.1 and 1.2. For any integrable kernel K and $n \in \mathbf{N}$, set

$$D_n(K) := \int_{\mathbf{R}} |f_{n,K}(x) - Ef_{n,K}(x)| dx.$$

So, we will prove Theorem 1.3 and the following proposition together:

Proposition 7.1. *Let \mathcal{K} be a relatively compact subset of $L_1(\mathbf{R}, \mathcal{B}, \lambda)$. If $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$, then*

$$\lim_{n \rightarrow \infty} E^* \sup_{K \in \mathcal{K}} D_n(K) = 0. \quad (7.1)$$

Proof of Theorem 1.3 and Proposition 7.1. Theorem 1, Chapter 3 in Devroye and Györfi (1985) shows that

$$\lim_{n \rightarrow \infty} J_n(K) = 0 \text{ in pr. and } \lim_{n \rightarrow \infty} D_n(K) = 0 \text{ in pr.} \quad (7.2)$$

for all $K \in L_1$ ($\int K = 1$ is not needed for the second limit). Now,

$$\begin{aligned} D_n(K) &= \frac{1}{nh_n} \int_{\mathbf{R}} \left| \sum_{i=1}^n K \left(\frac{X_i - x}{h_n} \right) - nEK \left(\frac{X - x}{h_n} \right) \right| dx \\ &\leq \frac{1}{nh_n} \sum_{i=1}^n \int_{\mathbf{R}} \left| K \left(\frac{X_i - x}{h_n} \right) \right| dx + \frac{1}{h_n} E \int_{\mathbf{R}} \left| K \left(\frac{X - x}{h_n} \right) \right| dx \\ &= 2 \int_{\mathbf{R}} |K(u)| du, \end{aligned}$$

and

$$\begin{aligned}
J_n(K) &= \int_{\mathbf{R}} \left| \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) - f(x) \right| dx \\
&\leq \int_{\mathbf{R}} \left| \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) \right| dx + 1 \\
&\leq \int_{\mathbf{R}} |K(u)| du + 1,
\end{aligned}$$

by change of variables. Since $\sup_{K \in \mathcal{K}} \int_{\mathbf{R}} |K(u)| du < \infty$ by relative compactness, the random variables in (7.2) are dominated by a constant and, by (7.2), the bounded convergence theorem then gives

$$\lim_{n \rightarrow \infty} EJ_n(K) = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} ED_n(K) = 0. \quad (7.3)$$

A computation similar to the ones above gives that, for all $K, K' \in L_1$,

$$|J_n(K) - J_n(K')| \leq \int_{\mathbf{R}} |f_{n,K}(x) - f_{n,K'}(x)| dx \leq \partial_1(K, K') \quad (7.4)$$

and

$$|D_n(K) - D_n(K')| \leq 2 \int_{\mathbf{R}} |K(x) - K'(x)| dx = 2\partial_1(K, K'). \quad (7.5)$$

Since \mathcal{K} is relatively compact in L_1 , it is totally bounded in L_1 . Given $\varepsilon > 0$, let T_ε be a maximal subset of \mathcal{K} satisfying that if $K, K' \in T_\varepsilon$ and $K \neq K'$, then $\partial_1(K, K') \geq \varepsilon$. Then, $\text{Card } T_\varepsilon = D(\mathcal{K}, \partial_1, \varepsilon) < \infty$ by total boundedness. We then have, by (7.4) and (7.5),

$$\begin{aligned}
E^* \sup_{K \in \mathcal{K}} J_n(K) &\leq E \max_{K \in T_\varepsilon} J_n(K) + E^* \sup_{K_1, K_2 \in \mathcal{K}, \partial_1(K_1, K_2) \leq \varepsilon} |J_n(K_1) - J_n(K_2)| \\
&\leq D(\mathcal{K}, \partial_1, \varepsilon) \max_{K \in T_\varepsilon} EJ_n(K) + \varepsilon
\end{aligned}$$

and

$$E^* \sup_{K \in \mathcal{K}} D_n(K) \leq D(\mathcal{K}, \partial_1, \varepsilon) \max_{K \in T_\varepsilon} ED_n(K) + 2\varepsilon$$

Now both, Theorem 1.3 and Proposition 7.1 follow from (7.3) and the finiteness of $D(\mathcal{K}, \partial_1, \varepsilon)$ for all $\varepsilon > 0$, by first letting n tend to infinity and then ε tend to zero in these inequalities. \square

Note that, in contrast with Theorems 1.1 and 1.2, the kernels $K \in \mathcal{K}$ in Theorem 1.3 and in Proposition 7.1 need not be compactly supported.

Remark 7.1. A subset \mathcal{K} of $L_1(\mathbf{R}, \mathcal{B}, \lambda)$ is relatively compact if and only if

- (i) $\sup_{K \in \mathcal{K}} \int_{\mathbf{R}} |K(x)| dx < \infty$,
- (ii) $\lim_{M \rightarrow \infty} \sup_{K \in \mathcal{K}} \int_{[-M, M]^c} |K(x)| dx = 0$, and
- (iii) $\lim_{y \rightarrow 0} \sup_{K \in \mathcal{K}} \int_{\mathbf{R}} |K(x+y) - K(x)| dx = 0$.

(e.g., Dunford and Schwartz (1966), page 298.) In particular, if \mathcal{K} satisfies conditions (i) and (ii), and moreover $\sup_{K \in \mathcal{K}} \sup_{x \neq y} |K(x) - K(y)|/|x - y| \leq C|x - y|^\beta$ for some $C < \infty$ and $\beta \in (0, 1)$, then \mathcal{K} satisfies Theorem 1.3 and Proposition 7.1.

Remark 7.2. Suppose

$$\left(\sup_{K \in \mathcal{K}} J_n(K) \right)^* \rightarrow 0 \text{ in pr.} \quad (7.6)$$

where in addition we are assuming $h_n \rightarrow 0$, $nh_n \rightarrow \infty$ and $\int_{\mathbf{R}} K(x)dx = 1$ for all $K \in \mathcal{K}$. Let $\{\varepsilon_i\}$ be a Rademacher sequence independent of the sequence $\{X_i\}$. Then,

$$\sup_{K \in \mathcal{K}} \int_{\mathbf{R}} |K(x)|dx < \infty, \quad (7.7)$$

$$E^* \sup_{K \in \mathcal{K}} \int_{\mathbf{R}} \left| \frac{1}{nh_n} \sum_{i=1}^n \varepsilon_i K \left(\frac{X_i - x}{h_n} \right) \right| dx \rightarrow 0, \quad (7.8)$$

and

$$E^* \sup_{K \in \mathcal{K}} J_n(K) \rightarrow 0. \quad (7.9)$$

To see this, we first note that by a comparison theorem of Montgomery-Smith (1993) (see e.g. de la Peña and Giné (1999), Corollary 1.1.6 and Remarks 1.1.7 and 1.1.8, page 7), if (7.6) holds, then

$$\left(\sup_{K \in \mathcal{K}} \int_{\mathbf{R}} \left| \frac{1}{nh_n} \sum_{i=1}^n \varepsilon_i \left(K \left(\frac{X_i - x}{h_n} \right) - h_n f(x) \right) \right| dx \right)^* \rightarrow 0 \text{ in pr.}$$

But the term in $f(x)$ tends to zero in probability because $\sum_{i=1}^n \varepsilon_i/n \rightarrow 0$, so that we have

$$\left(\sup_{K \in \mathcal{K}} \int_{\mathbf{R}} \left| \frac{1}{nh_n} \sum_{i=1}^n \varepsilon_i K \left(\frac{X_i - x}{h_n} \right) \right| dx \right)^* \rightarrow 0 \text{ in pr.} \quad (7.10)$$

Then, by Lévy's inequality,

$$\max_{1 \leq i \leq n} \left(\sup_{K \in \mathcal{K}} \int_{\mathbf{R}} \left| \frac{1}{nh_n} \varepsilon_i K \left(\frac{X_i - x}{h_n} \right) \right| dx \right)^* \rightarrow 0 \text{ in pr.,} \quad (7.11)$$

but, by change of variables, this implies (7.7). Now, (7.8) and (7.9) follow from this last observation and (7.10) and (7.6) respectively, both by Hoffmann-Jørgensen's inequality (e.g. de la Peña and Giné (1999), Theorem 1.2.3 and Remarks 1.2.4 and 1.2.9). A similar remark applies to the processes D_n .

Example 7.1. Let

$$\mathcal{K} = \left\{ K_\ell := \ell I_{(-1/(2\ell), 1/(2\ell))} : \ell \in \mathbf{N} \right\},$$

be the class of kernels from Example 6.1 (and from (1.5)). We will show that none of the sequences

$$\left\{ \sup_{K \in \mathcal{K}} J_n(K) \right\}_{n=1}^{\infty} \quad \text{and} \quad \left\{ \sup_{K \in \mathcal{K}} D_n(K) \right\}_{n=1}^{\infty}$$

converge in probability. By the previous remark, it suffices to show that the sequence

$$E \sup_{K \in \mathcal{K}} \frac{1}{nh_n} \int_{\mathbf{R}} \left| \sum_{i=1}^n \varepsilon_i K \left(\frac{X_i - x}{h_n} \right) \right| dx, \quad n \in \mathbf{N},$$

does not tend to zero. Given $X_1(\omega), \dots, X_n(\omega)$, we choose ℓ so that the intervals $I_i := (X_i(\omega) - h_n/(2\ell), X_i(\omega) + h_n/(2\ell))$, $i = 1, \dots, n$, are disjoint which we can do, for any given n , for almost every ω . Then,

$$\begin{aligned} \int_{\mathbf{R}} \left| \sum_{i=1}^n \varepsilon_i K_\ell \left(\frac{X_i(\omega) - x}{h_n} \right) \right| dx &\geq \int_{\cup_{j=1}^n I_j} \left| \sum_{i=1}^n \varepsilon_i K_\ell \left(\frac{X_i(\omega) - x}{h_n} \right) \right| dx \\ &= \sum_{j=1}^n \int_{I_j} \left| \sum_{i=1}^n \varepsilon_i K_\ell \left(\frac{X_i(\omega) - x}{h_n} \right) \right| dx \\ &= \sum_{j=1}^n \int_{I_j} K_\ell \left(\frac{X_j(\omega) - x}{h_n} \right) dx = nh_n. \end{aligned}$$

Therefore, for all $n \geq 1$

$$\sup_{K \in \mathcal{K}} \frac{1}{nh_n} \int_{\mathbf{R}} \left| \sum_{i=1}^n \varepsilon_i K \left(\frac{X_i - x}{h_n} \right) \right| dx \geq 1 \quad \text{a.s.},$$

which contradicts (7.8).

References

- Bartlett, M.S. (1938). The characteristic function of a conditional statistic. *J. London Math. Soc.* **13** 62–67.
- Beirlant, J., Györfi, L. and Lugosi, G. (1994). On the asymptotic normality of L_1 - and L_2 -errors in histogram density estimation. *Canad. J. Statist.* **22** 309–318.
- Beirlant, J. and Mason, D.M. (1995). On the asymptotic normality of L_p -norms of empirical functionals. *Math. Methods Statist.* **4** 1–19.
- Berlinet, A. and Devroye, L. (1994). A comparison of kernel density estimates. *Publ. Inst. Statist. Univ. Paris* **38** 3–59.
- Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- Csörgő, M. and Horváth, L. (1988). Central limit theorems for L_p -norms of density estimators. *Z. Wahrsch. Verw. Gebiete.* **80** 269–291.
- de Acosta, A. (1981). Inequalities for B -valued random vectors with applications to the strong law of large numbers. *Ann. Probab.* **9** 157–161.

- Deheuvels, P. and Mason, D.M. (1992). Functional laws of the iterated logarithm for the increments of empirical and quantile processes. *Ann. Probab.* **20** 1248–1287.
- de la Peña, V. and Giné, E. (1999). *Decoupling, from Dependence to Independence*. Springer, New York.
- Devroye, L. (1991). Exponential inequalities in nonparametric estimation. In: *Nonparametric Functional Estimation and Related Topics* (G. Roussas ed.) 31–44, Kluwer Academic Publishers, Dordrecht.
- Devroye, L., and Györfi, L. (1985). *Nonparametric Density Estimation: The L_1 View*. Wiley, New York.
- Dudley, R.M. (1984). A Course on empirical processes. Ecole d’été de probabilités de St.-Flour, 1982. Lecture Notes in Math. (Springer) **1097** 1–142.
- Dudley, R.M. (1989). *Real Analysis and Probability*. Wadsworth & Brooks/Cole, Pacific Grove, California.
- Dunford, N. and Schwartz, J.T. (1958). *Linear Operators, Part I*. Third Printing, 1966. Wiley, New York.
- Eggermont, P.P.B. and LaRiccia, V.N. (2000). *Maximum Penalized Likelihood Estimation, Volume 1; Density Estimation*. Springer, New York.
- Einmahl, J.H.J. (1987). *Multivariate Empirical Processes*. CWI Tract **33** Cent. Wisk. Inform., Amsterdam.
- Einmahl, U. and Mason, D.M. (1997). Gaussian approximation of local empirical processes indexed by functions. *Probab. Theory Rel. Fields* **107** 283–301.
- Feller, W. (1966). *An Introduction to Probability Theory and Its Applications, Volume II*. Wiley, New York.
- Folland, G.B. (1999). *Real Analysis*. Second Ed. Wiley, New York.
- Holst, L. (1979). Asymptotic normality of sum spacings. *Ann. Probab.* **7** 1066–1072.
- Horváth, L. (1991). On L_p -norms of multivariate density estimators. *Ann. Statist.* **19** 1933–1949.
- Johnson, W.B., Schechtman, G. and Zinn, J. (1985). Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Probab.* **13** 234–253.
- Karlin, S., and Rinott, Y. (1981). Total positivity properties of absolute value multinormal variables with applications to confidence interval estimates and related probabilistic inequalities. *Ann. Statist.* **9** 1035–1049.
- Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces*. Springer, Berlin.
- Mason, D.M. and van Zwet, W. (1987). A refinement of the KMT inequality for the uniform empirical process. *Ann. Probab.* **15** 871–884.
- Montgomery-Smith, S. (1993). Comparison of sums of independent identically distributed random variables. *Probab. Math. Statist.* **14** 281–285.

- Pinelis, I.F. (1990). Inequalities for sums of independent random vectors and their application to estimating a density. *Theory Probab. Appl.* **35** 605–607 (Translated from Russian).
- Pinelis, I.F. (1994). On a majorization inequality for sums of independent random variables. *Probab. Statist. Letters* **19** 97–99.
- Shergin, V.V. (1979). On the convergence rate in the central limit theorem for m -dependent random variables. *Theory Probab. Appl.* **24** 782–796 (Translated from Russian).
- Sweeting, T.J. (1977). Speeds of convergence in the multidimensional central limit theorem. *Ann. Probab.* **5** 28–41.
- van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Department of Mathematics and Statistics
Math. Science Building 302
Storrs, CT 06269-3009
USA
e-mail: gine@uconnvm.uconn.edu

Department of Food and Resource Economics
206 Townsend Hall
University of Delaware
Newark, DE 19717
USA
e-mail: davidm@udel.edu

Laboratory of Statistical Methods
St. Petersburg Branch of the Steklov Mathematical Institute
27 Fontanka
St. Petersburg 191011
RUSSIA
e-mail: zaitsev@pdmi.ras.ru