The Annals of Probability
2003, Vol. 31, No. 2, 719–768

THE L_1-NORM DENSITY ESTIMATOR PROCESS

BY EVARIST GINÉ,[1] DAVID M. MASON[2]
AND ANDREI YU. ZAITSEV[3]

University of Connecticut, University of Delaware and
St. Petersburg Branch of the Steklov Mathematical Institute

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. None of the theorems imposes 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.

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 $n \to \infty$, $h_n \to 0$ and $nh_n \to \infty$. The classical kernel estimator is defined as

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

where K is a kernel with compact support satisfying

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

For notational convenience, we will usually assume

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

Since Lebesgue density functions, by definition, sit in $L_1(\mathbf{R}, \mathcal{B}, m)$, where m denotes Lebesgue measure, Devroye and Györfi have long advocated that the

Received January 2001; revised March 2002.
[1]Supported in part by NSF Grant DMS-00-70382.
[2]Supported in part by NSF Grant DMS-02-03865 and NSA Grant MDA904-02-1-0034.
[3]Supported in part 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.
AMS 2000 subject classifications. 60F05, 60F15, 60F17, 62G07.
Key words and phrases. Kernel density function estimator, L_1-norm, central limit theorem, weak convergence, Poissonization, entropy.

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:

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

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

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–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 results of Csörgő and Horváth (1988) and Horváth (1991) required numerous regularity conditions.

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, Chapter 2, in Eggermont and LaRiccia (2001)] recently applied their method to the special case of the L_1-norm of the kernel density estimator to prove that, whenever $h_n \to 0$ and $\sqrt{n}h_n \to \infty$ and K satisfies (1.2) and $\|K\|_\infty < \infty$,

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

converges in distribution to a normal random variable $\sigma(K)Z$, with

$$\sigma^2(K) := \|K\|_2^2 \int_{-1}^{1} \operatorname{Cov}\Big(\big|\sqrt{1-\rho^2(t)}\,Z_1 + \rho(t)Z_2\big|, |Z_2|\Big)\, dt,$$

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

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

The variance $\sigma^2(K)$ has an interesting alternate representation. Using the formulas for the absolute moments of a bivariate normal random variable of Nabeya (1951), we can write

$$\sigma^2(K) = \|K\|_2^2 \int_{\mathbf{R}} \varphi(\rho(t))\, dt,$$

where

$$\varphi(\rho) = \frac{2}{\pi}\Big(\rho \arcsin \rho + \sqrt{1-\rho^2} - 1\Big), \qquad \rho \in [-1, 1].$$

As a byproduct of our work, here we will extend this central limit theorem 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

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

which, by using Nabeya's (1951) formulas, becomes

$$\sigma(K_1, K_2) = \|K_1\|_2 \|K_2\|_2 \int_{-1}^{1} \varphi(\rho(K_1, K_2, t))\, dt,$$

where

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

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 $\rho(0) = 1$. Moreover, it is readily checked that $\varphi(0) = 0$ and $\varphi(\rho) > 0$ for $\rho \in [-1, 1]$, $\rho \neq 0$. 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 $p > 0$,

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

Before stating any results, we should recall a couple of definitions from empirical processes that we will use throughout. Given a metric space (T, d), for each $\varepsilon > 0$, the covering number $\mathcal{N}(T, d, \varepsilon)$ is defined as the minimal number of open d-balls of radius at most ε and centers in T required to cover T, and the packing number $D(T, d, \varepsilon)$ as the cardinality of a maximal ε-separated subset of T. Both quantities are essentially equivalent. Given processes with bounded trajectories $X_n(t)$ and $X(t)$, $t \in T$, such that the law of $X(t)$ is defined by a tight Borel measure on $\ell^\infty(T)$, we say that X_n converges weakly in $\ell^\infty(T)$ to X if $E^* H(X_n) \to E H(X)$ for all bounded continuous functions $H: \ell^\infty(T) \mapsto \mathbf{R}$, where $E^* H(X_n) = E H^*(X_n)$ denotes outer expectation of $H(X_n)$, which equals the expected value of the measurable envelope $H^*(X_n)$ of $H(X_n)$. See van der Vaart and Wellner (1996) or de la Peña and Giné (1999) for expansions and comments on these definitions. In our case, (T, d) will be $(\mathcal{K}, \partial_2)$.

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*

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

If $h_n \to 0$ and $\sqrt{n}h_n \to \infty$, 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 $\sigma(K_1, K_2)$.

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

$$\sup_{K\in\mathcal{K}} |\xi_n(K)| \xrightarrow{d} \sup_{K\in\mathcal{K}} |\xi(K)|$$

for all densities f, where the distribution of the right-hand random variable does not depend on f. Moreover, the same is true for any other continuous functional on $\ell^\infty(\mathcal{K})$.

Theorem 1.1, of course, includes the central limit theorem for one kernel or jointly for a finite number of kernels; however, in these cases, Theorem 1.2, as well as its multivariate counterpart, which we omit, is better as it does not assume boundedness of the kernel or kernels. The following example shows that condition (1.4) is satisfied by quite large, infinite classes of kernels.

EXAMPLE 1.1. Let $\mathcal{K}$ be a uniformly bounded class of measurable functions, which are 0 off the interval $[-1/2, 1/2]$. Write

$$[-\tfrac{1}{2}, \tfrac{1}{2}] = \bigcup_{i=1}^m I_i,$$

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

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

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 $D_1 > 0$,

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

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 0 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 0 on $[-1/2, 1/2]^c$, satisfies

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

for $\varepsilon \leq C^{1/2}$ and, therefore, condition (1.4). The constant C can be taken to be $C = 2MV$, where $M = \|K\|_\infty$ and V is the total variation of K on any closed interval $[-a, a]$ with $a > 1/2$, for example, $a = 1$. To see this, let $P(x)$ and $N(x)$ be respectively the positive and negative variations of K on $[-1, x]$, which are monotone nondecreasing and nonnegative functions, such that $P(x) = N(x) = 0$ for $x < 1/2$ and $P(x) = N(x) = V/2$ for $x > 1/2$; then, for $1 \leq \lambda < \mu$,

$$\begin{aligned}
&\int_{-\infty}^{\infty} |P(\lambda x) - P(\mu x)|\, dx \\
&\qquad = \left(\int_{-1}^{0} P(x)\, dx - \int_{0}^{1} P(x)\, dx + \frac{V}{2}\right)(\lambda^{-1} - \mu^{-1}),
\end{aligned}$$

with a similar equality holding for $N(x)$, so that

$$\begin{aligned}
\partial_2^2\big(K(\cdot\lambda), K(\cdot\mu)\big) &\leq M \int_{-\infty}^{\infty} |K(\lambda x) - K(\mu x)|\, dx \\
&\leq M \int_{-\infty}^{\infty} |P(\lambda x) - P(\mu x)|\, dx \\
&\qquad + M \int_{-\infty}^{\infty} |N(\lambda x) - N(\mu x)|\, dx \\
&\leq 2MV(\lambda^{-1} - \mu^{-1}).
\end{aligned}$$

Now, with this bound, it is easy to estimate the covering numbers of $\mathcal{K}$ for the ∂_2 distance: the open balls of radius ε and centers at $K(\cdot\lambda_k)$, with $\lambda_k := C/(C - k\varepsilon^2)$, $k = 0, 1, \ldots, k_0$, where k_0 is the largest integer strictly smaller than C/ε^2, cover $\mathcal{K}$.

EXAMPLE 1.3. Under certain regularity conditions on the kernel K and the density f, the optimal choice of h_n, in terms of the value of h_n that minimizes $E\|f_{n,K} - f\|_1$, is of the form $\lambda^{-1}n^{-1/5}$, where $\lambda > 0$ is a smoothing parameter, which in practical estimation problems must be estimated from the data. For details, refer to Theorem 2.21, Chapter 2, in Eggermont and LaRiccia (2001). This suggests viewing $\xi_n(K_\lambda)$, 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, that have this property, consult Berlinet and Devroye (1994) and the references therein. It is easy to see, building on Example 1.2, that if K is a 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, note that if

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

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.

The tools that we develop to prove Theorem 1.1 permit us to extend the 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* $h_n \to 0$ *and* $\sqrt{n}h_n \to \infty$, *then, as* $n \to \infty$,

$$\xi_n(K) \xrightarrow{d} \xi(K) \stackrel{d}{=} \sigma(K)Z \tag{1.6}$$

and

$$\operatorname{Var}(\xi_n(K)) \to \sigma^2(K). \tag{1.7}$$

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 $Ef_{n,K}(x)$ can be replaced by $f(x)$ in $\xi_n(K)$ in (1.6). An obvious sufficient condition is that

$$\sqrt{n}\|f - Ef_{n,K}\|_1 \to 0 \qquad \text{as } n \to \infty. \tag{1.8}$$

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

$$\sqrt{n}\|f - Ef_{n,K}\|_1 = \frac{\sqrt{n}h_n}{2},$$

which implies that (1.8) is not satisfied under the condition for asymptotic normality $\sqrt{n}h_n \to \infty$ as $n \to 0$. One set of sufficient conditions for (1.8) is the following: in addition to the assumptions that K satisfies (1.1) and (1.2) and $\|K\|_2 < \infty$, assume that $K(x) = K(-x)$, for all $x \in \mathbf{R}$,

$$\int x^2 K(x)\,dx = 0 \quad \text{and} \quad \int_{\mathbf{R}} |x|^3 |K(x)|\,dx < \infty.$$

Also assume that the density f is three times continuously differentiable on $\mathbf{R}$ and $\int_{\mathbf{R}} |f'''(x)|\,dx < \infty$. Applying Lemma 22 of Devroye and Györfi (1985), Chapter 5, we get

$$\sqrt{n}\|f - Ef_{n,K}\|_1 \le \frac{\sqrt{n}h_n^3}{6} \int_{\mathbf{R}} |x|^3 |K(x)|\,dx \int_{\mathbf{R}} |f'''(x)|\,dx.$$

Thus $\sqrt{n}h_n^3 \to 0$ and $\sqrt{n}h_n \to \infty$ as $n \to \infty$ imply both (1.8) and asymptotic normality. (Note that the choice $h_n = \lambda n^{-1/5}$, $\lambda > 0$, fulfills these conditions.) 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 difference in asymptotic behavior between the sup norm and the L_1-norm of the discrepancy between the kernel density estimator and its mean (or the density). In particular, Theorem 1.2 should be compared to the Bickel and Rosenblatt (1973) result on weak convergence of the sup norm of $f_{n,K} - Ef_{n,K}$ over compact intervals: the hypotheses are more restrictive, the centering is different and the rate is slower for the sup norm.

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 $L_1(\mathbf{R}, \mathcal{B}, m)$ satisfying* (1.1) *for all $K \in \mathcal{K}$. If $h_n \to 0$ and $nh_n \to \infty$, then*

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

The following notation, already encountered above in the definition of weak convergence in $\ell^\infty(T)$, is used in this statement and will be used thoughout: 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 nonmeasurable 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 first prove that the finite-dimensional (f.d.) distributions 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

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

[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 2–4, 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 the 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. Its idea may be traced back to Pyke and Shorack [(1968), proof of Lemma 2.2] through Einmahl (1987) and Deheuvels and Mason (1992) [see also Einmahl and Mason (1997)], and we give it here in an abstract form suitable for our purposes. See Borisov (2002) for a slight generalization of this lemma and extra historical remarks. 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* $\Pr\{X_1 \in B\} \le 1/2$. *Then*

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

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

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

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 τ_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 $Y_i \stackrel{d}{=} (X_1 | X_1 \in B)$ for all $i \ge 1$. It is easy to see that

$$I(X_1 \in B)X_1 \stackrel{d}{=} \tau_1 Y_1.$$

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

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

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

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

Now, Stirling's formula, $n! = (n/e)^n\sqrt{2\pi n}e^{\theta_n/n}$ for some $0 < \theta_n < 1/12$, gives that $\Pr\{\eta_{B^c} = [np_{B^c}]\}/\Pr\{\eta = n\} \le 2$ if $p_{B^c} > 1/2$ and $n \ge 5$; direct computation shows that this inequality is also true for $1 \le n \le 4$, 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 to 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 $f_{n,B}(x_1, \dots, x_n) = \sum_{i=1}^n I(x_i \in B)\delta_{x_i}$, $n \in \mathbf{N}$, $B \subset \mathcal{S}$. 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 $f_{n,B}^{-1}\{\mu \in D, \int h\,d\mu \le t\} = \{(x_1, \dots, x_n) : \sum I(x_i \in B)h(x_i) \le t\}$ is a measurable set of S^n]. Our functions H will have the general form

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

where A is a union of intervals and B is the $h/2$-neighborhood of A. H can be shown to be $\mathcal{D}$-measurable by approximation by Riemann sums.

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 $p \ge 2$ and $n \in \mathbf{N}$,

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

[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 nonnegative random variables ζ_i, namely: for every $p \ge 1$ and $n \in \mathbf{N}$,

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

[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, nondecreasing in each of them separately, and such that the inequality*

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

holds for all $n \in \mathbf{N}$ and all sequences of independent, identically distributed and centered random variables $\xi, \xi_1, \xi_2, \ldots$, then the inequality

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

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

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

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

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

$$\omega \stackrel{d}{=} \sum_{i=0}^{\eta_N}\zeta_i, \qquad \eta_N \stackrel{d}{=} \mathrm{Pois}\left(\frac{\gamma}{N}\right),$$

and η_N independent of $\zeta_1, \zeta_2, \ldots$. Clearly, $E\omega_i = (\gamma/N)E\zeta$. Applying inequality (2.5) to the sequence $\{\omega_i\}$, we get

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

The first argument of F in this inequality is just

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

Regarding the second argument, we have

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

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

$$
\limsup_{N\to\infty} N E\left|\omega - \frac{\gamma}{N} E\zeta\right|^r \le \gamma E|\zeta|^r.
$$

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 $\gamma > 0$ and independent of the variables $\{\zeta_i\}_{i=1}^{\infty}$, then, for every $p \ge 2$,*

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

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

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

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* $\sigma^2 < \infty$,

$$(S_n, U_n) \xrightarrow{w} \big(\sigma Z_1, \sqrt{1-\alpha} Z_2\big) \qquad \textit{as } n \to \infty.$$

Then, for all x,

$$\Pr\{S_n \leq x | N_n = n\} \to \Pr\{\sigma Z_1 \leq x\}.$$

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

PROOF OF LEMMA 2.4. Consider the characteristic function

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

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

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

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

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

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

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

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),

$$E\exp(itS_n + iuU_n) \to \phi(t,v), \tag{2.11}$$

where

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

Next, the proof of Theorem 3 on pages 490–491 of Feller (1966) shows that, as $n \to \infty$,

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

which implies that

$$\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \left| E\exp(itS_n + iuU_n)\big(E\exp(ivV_n) - \exp(-\alpha v^2/2)\big)\right| dv \to 0. \tag{2.12}$$

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

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

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 nonzero correlation. However, the version given above suffices for our needs. Lemma 2.4 is not original to Beirlant and Mason. A preliminary version was prov ed 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 $0 < \alpha < 1$,

$$N_n = \sum_{i=1}^{n} X_i, \qquad N_{2,n} = \sum_{i=1}^{[n\alpha]} X_i, \qquad N_{1,n} = N_n - N_{2,n}.$$

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 $\sup_{i\in\mathbf{N}} \int_{ih_n}^{(i+3)h_n} f(x)\, dx \le 1/2$ for all $n > N_0$. Such an N_0 exists by absolute continuity if $h_n \to 0$ as $n \to \infty$, which we assume throughout.

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

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

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

$$E\exp\left\{t\frac{|\xi_n(K_1) - \xi_n(K_2)|}{C_1\partial_2(K_1, K_2)}\right\} \le 4\exp\left\{\sum_{m=2}^{\infty}\left(\frac{720et}{\log m}\right)^m\right\} \tag{3.2}$$

for all $t \ge 0$.

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 $x \in \mathbf{R}$,

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

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

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

Then

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

and we define, for convenience,

$$\overline{\xi}_n(K_1) - \overline{\xi}_n(K_2) = \int_{\mathbf{R}}\big(\Delta_n(x) - E\Delta_{\eta_n}(x)\big)\,dx. \tag{3.4}$$

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 $\lambda \in \mathbf{R}$,

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

and, by Jensen's inequality,

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

Hence

$$\begin{aligned}E\exp\{\lambda|\xi_n(K_1)-\xi_n(K_2)|\} &\le \big(E\exp\big\{\lambda|\overline{\xi}_n(K_1)-\overline{\xi}_n(K_2)|\big\}\big)^2\\ &\le E\exp\big\{2\lambda|\overline{\xi}_n(K_1)-\overline{\xi}_n(K_2)|\big\}.\end{aligned}\tag{3.5}$$

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 $|i-j|\ge 2$, and

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

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

and

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

Then condition (ii) becomes $\Pr\{X\in B_s\}\le 1/2$. 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 $\mathcal{I}_0'=\{2n:n\in\mathbf{Z}\}$; 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 $\Pr\{X\in B_s\}\le 1/2$; the same can be done with $\mathcal{I}_1'=\{2n+1:n\in\mathbf{Z}\}$.) Then we have

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

and, by Hölder's inequality,

$$\begin{aligned}&E\exp\{2\lambda|\overline{\xi}_n(K_1)-\overline{\xi}_n(K_2)|\}\\ &\qquad\le\prod_{s=1}^{6}\left(E\exp\left\{12\lambda\left|\int_{A_s}\big(\Delta_n(x)-E\Delta_{\eta_n}(x)\big)\,dx\right|\right\}\right)^{1/6}.\end{aligned}\tag{3.7}$$

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

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

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

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

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

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

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

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

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

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

Now

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

which by Lemma 2.3 is less than or equal to

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

Thus, by Jensen's inequality,

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

It follows from these estimates, from $E\delta_{j,n} = 0$ and from Stirling's formula, that $E\exp\{12\lambda\delta_{j,n}\}$

$$
\begin{aligned}
&\le 1 + \sum_{m=2}^{\infty}\Big(\frac{720e\lambda}{\log m}\Big)^m\Bigg[\Big(\int_{jh_n}^{(j+1)h_n}\frac{1}{h_n}E(K_1-K_2)^2\Big(\frac{x-X}{h_n}\Big)\,dx\Big)^{m/2} \\
&\qquad + \frac{1}{n^{m/2-1}}\int_{jh_n}^{(j+1)h_n}\frac{1}{h_n}E\Big|(K_1-K_2)\Big(\frac{x-X}{h_n}\Big)\Big|^m\,dx\Bigg],
\end{aligned}
$$

and the same bound holds as well for $E\exp\{-12\lambda\delta_{j,n}\}$. Then, plugging this estimate into (3.10) and using the elementary inequality

$$
\prod(1+x_i) \le \exp\Big(\sum x_i\Big), \qquad x_i \in \mathbf{R},
$$

we obtain

$$
\begin{aligned}
&\exp\biggl\{12\lambda\biggl|\int_{A_s}\bigl(\Delta_{\eta_n}(x)-E\Delta_{\eta_n}(x)\bigr)\,dx\biggr|\biggr\}\\
&\qquad\le 2\exp\biggl\{\sum_{j\in\mathcal{J}_s}\sum_{m=2}^{\infty}\biggl(\frac{720e\lambda}{\log m}\biggr)^m\\
&\qquad\qquad\times\biggl[\biggl(\int_{jh_n}^{(j+1)h_n}\frac{1}{h_n}E(K_1-K_2)^2\biggl(\frac{x-X}{h_n}\biggr)\,dx\biggr)^{m/2}\\
&\qquad\qquad+\frac{1}{n^{m/2-1}}\int_{jh_n}^{(j+1)h_n}\frac{1}{h_n}E\biggl|(K_1-K_2)\biggl(\frac{x-X}{h_n}\biggr)\biggr|^m\,dx\biggr]\biggr\}.
\end{aligned}
\tag{3.13}
$$

Now, by a change of variables,

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

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

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

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

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

Setting

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

in this inequality and combining it with inequalities (3.7), (3.8) 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

$$\Psi_1(x)=\frac{e^x-1}{4} \tag{3.14}$$

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 Proposition* 3.1, *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 $\sum_{m=2}^{\infty}(720et_0/\log m)^m=\log(5/4)$. 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 $n > N_0$,*

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

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

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

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

$$E\big(\xi_n(K_1) - \xi_n(K_2)\big)^2 \le 4E\big(\overline{\xi}_n(K_1) - \overline{\xi}_n(K_2)\big)^2. \tag{3.16}$$

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

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

which gives

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

Thus, from (3.6), the convexity of $y = x^2$ and Lemma 2.1 and using independence, we get

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

which by (3.17) is less than or equal to

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

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

4. Tightness of general processes. The standard theorems on the 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, e.g., the oldest one, valid only for Orlicz norms of exponential type such as, e.g., in de la Peña and Giné (1999), Theorems 5.1.4 and 5.1.5, or the ones coming from Pisier's improvement on chaining such 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 (1996)]. 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:

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

Note that Ψ_1 defined in (3.14) satisfies these conditions [as $\Psi_1^{-1}(x) = \log(4x+1)$]. In what follows, if a variable X is not necessarily measurable, we write $\|X\|_\Psi^*$ for $\||X|^*\|_\Psi$, where $|X|^*$ is the measurable envelope of $|X|$.

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 $0 < \gamma <$ $\operatorname{diam}(T)$ such that*

$$\|X_s - X_t\|_\Psi \le C\, d(s,t), \tag{4.1}$$

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

$$\begin{aligned} &\Big\| \sup_{d(s,t) \le \delta} |X_s - X_t| \Big\|_\Psi^* \\ &\qquad \le 2 \Big\| \sup_{d(s,t) \le \gamma} |X_s - X_t| \Big\|_\Psi^* + CL \int_{\gamma/2}^{\delta} \Psi^{-1}\big(D(T,d,\varepsilon)\big)\, d\varepsilon. \end{aligned} \tag{4.2}$$

PROOF. Let T_γ be a maximal subset of T satisfying $d(s,t) \ge \gamma$ for $s \ne t \in T_\gamma$. Then $\operatorname{Card}(T_\gamma) = D(T,d,\gamma)$. If $s,t \in T$ and $d(s,t) \le \delta$, 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

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

we obtain

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

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

$$\begin{aligned}\left\|\max_{\substack{d(s,t)<3\delta\\ s,t\in T_\gamma}}|X_s - X_t|/C\right\|_\Psi &\le L\int_0^{3\delta}\psi^{-1}\big(D(T_\gamma, d, \varepsilon)\big)\,d\varepsilon\\ &\le 3L\int_0^{\delta}\psi^{-1}\big(D(T_\gamma, d, \varepsilon)\big)\,d\varepsilon,\end{aligned}\tag{4.4}$$

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

$$\begin{aligned}\int_0^{\delta}\Psi^{-1}\big(D(T_\gamma, d, \varepsilon)\big)\,d\varepsilon &\le \gamma\Psi^{-1}\big(D(T, d, \gamma)\big) + \int_\gamma^{\delta}\Psi^{-1}\big(D(T, d, \varepsilon)\big)\,d\varepsilon\\ &\le 3\int_{\gamma/2}^{\delta}\Psi^{-1}\big(D(T, d, \varepsilon)\big)\,d\varepsilon,\end{aligned}$$

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 3.

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

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

then

$$\lim_{\delta\to 0}\sup_{n>N_0}\left\|\sup_{\partial_2(K_1,K_2)\le\delta,\ K_1,K_2\in\mathcal{K}}\big|\xi_n(K_1) - \xi_n(K_2)\big|\right\|_{\Psi_1}^* = 0. \tag{5.2}$$

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

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

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

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

Therefore the first term on the right-hand side of inequality (5.3) is dominated by $2\|4\rho\|_{\Psi_1} = 8\rho/\log 5$. Hence, letting first δ tend to 0 and then ρ go to 0 in (5.3), we obtain the limit (5.2). □

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 $h_n \to 0$ and $\sqrt{n}h_n \to \infty$ as $n \to \infty$ and any finite collection $\mathbf{K} := \{K_1, K_2, \dots, K_m\} \subset \mathcal{K}$, $m \ge 1$, 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 covariance matrix with entries $a_{ls} = \sigma(K_l, K_s)$,*

$l, s = 1, \ldots, m$. *Moreover, for all* $1 \le l, s \le m$,

$$
\begin{aligned}
&\lim_{n\to\infty} \operatorname{Cov}\big(\xi_n(K_l), \xi_n(K_s)\big) \\
&\quad = \lim_{n\to\infty} n \operatorname{Cov}\big(\|f_{n,K_l} - Ef_{n,K_l}\|_1, \|f_{n,K_s} - Ef_{n,K_s}\|_1\big) \qquad (6.1)\\
&\quad = \sigma(K_l, K_s).
\end{aligned}
$$

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

$$\lim_{n\to\infty} n \operatorname{Var}\big(\|f_{n,K_s} - Ef_{n,K_s}\|_1\big) = \sigma^2(K_s). \tag{6.2}$$

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* 0 *off a compact interval. Then, for any* $H \in \mathcal{H}$,

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

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* $m(C)$ *such that*

$$C \subset [-M + \nu, M - \nu], \tag{6.4}$$

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

$$\int_C f(x)\, dx > 1 - \varepsilon, \tag{6.6}$$

$$f \text{ is bounded, continuous and bounded away from } 0 \text{ on } C \tag{6.7}$$

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

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

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

$$\int_F f(x)\,dx > 1 - \frac{\varepsilon}{4}.$$

Clearly, we can extract a compact subset D of $\mathbf{R}$ such that f is bounded, continuous and bounded away from 0 on $D \cap F$, and

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

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

$$\mathcal{H} = \mathcal{H}_0 := \bigcup_{s=1}^{m} \{K_s, |K_s|, K_s^2, |K_s|^3\}. \tag{6.9}$$

Let C be from this lemma. Denote

$$\varepsilon_n = \sup_{z\in C, H\in\mathcal{H}_0} \bigl|f * H_{h_n}(z) - J(H) f(z)\bigr|. \tag{6.10}$$

By Lemma 6.1,

$$\varepsilon_n \to 0 \qquad \text{as } n \to \infty. \tag{6.11}$$

Assume that $n \ge n_0$ is so large that $\varepsilon_n \le \delta \min\{J(H) : H \in \mathcal{H}_0\}/2$, where $\delta > 0$ is defined by

$$\delta = \inf_{x\in C} f(x).$$

Then, for any $z \in C$ and $H \in \mathcal{H}_0$, we have

$$f(z)J(H)/2 \le f * H_{h_n}(z) \le 2 f(z) J(H). \tag{6.12}$$

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 $h_n \to 0$ and $nh_n \to \infty$, we have that, for any Borel subset B of $\mathbf{R}$, any $K \in \mathcal{K}$ and any sequence of functions $a_n \in L_1(\mathbf{R}, \mathcal{B}, m)$,*

$$\limsup_{n\to\infty} E\biggl(\sqrt{n}\int_B \{|f_{n,K}(x) - a_n(x)| - E|f_{n,K}(x) - a_n(x)|\}\,dx\biggr)^2 \le 4\kappa^2 \int_B f(x)\,dx,$$

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

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

$$\begin{aligned} E\biggl(\sqrt{n}\int_B \{|f_{n,K}(x) - a_n(x)| - E|f_{n,K}(x) - a_n(x)|\}\,dx\biggr)^2 &\le 4E\biggl(\frac{1}{h_n}\int_B \biggl|K\biggl(\frac{x-X}{h_n}\biggr)\biggr|\,dx\biggr)^2 \\ &\le 4\|K\|_\infty E\frac{1}{h_n}\int_B \biggl|K\biggl(\frac{x-X}{h_n}\biggr)\biggr|\,dx \\ &\le 4\kappa^2 h_n^{-1}\int_B \Pr\{X \in [x - h_n/2, x + h_n/2]\}\,dx. \end{aligned} \tag{6.13}$$

Now

$$h_n^{-1}\int_B \Pr\{X \in [x - h_n/2, x + h_n/2]\}\,dx \le \int_B f(x)\,dx + \int_{\mathbf{R}} \bigl|h_n^{-1}\Pr\{X \in [x - h_n/2, x + h_n/2]\} - f(x)\bigr|\,dx.$$

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

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

as $n \to \infty$, which completes the proof of Lemma 6.2. □

We shall apply Lemma 6.2 in the case where $a_n(x) = Ef_{n,K}(x)$. Note that in this situation we could get the same bound with 16 instead of 4 by applying Theorem 2.1 of de Acosta (1981). Also see Devroye (1991), who obtained the bound (6.13) with $a_n(x) = f(x)$.

Next, we shall "Poissonize." Let η be a Poisson(n) random variable, that is, a Poisson random variable with mean n, independent of $X, X_1, X_2, \ldots$, and set

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

where the empty sum is defined to be 0. Notice that

$$E f_{\eta,K}(x) = E f_{n,K}(x) = h_n^{-1} E K\left(\frac{x-X}{h_n}\right), \tag{6.14}$$

$$k_n(K,x) := n \operatorname{Var}(f_{\eta,K}(x)) = h_n^{-2} E K^2\left(\frac{x-X}{h_n}\right) \tag{6.15}$$

and

$$n \operatorname{Var}(f_{n,K}(x)) = h_n^{-2} E K^2\left(\frac{x-X}{h_n}\right) - \left\{h_n^{-1} E K\left(\frac{x-X}{h_n}\right)\right\}^2. \tag{6.16}$$

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

$$\begin{aligned}
&\sup_{x\in C, K\in\mathbf{K}} \left|\sqrt{n \operatorname{Var}(f_{\eta,K}(x))} - \sqrt{n \operatorname{Var}(f_{n,K}(x))}\right| \\
&\qquad \le \sup_{x\in C, K\in\mathbf{K}} \frac{\sqrt{h_n}(f * K_{h_n}(x))^2}{\sqrt{f * K_{h_n}^2(x)}} = O(\sqrt{h_n})
\end{aligned} \tag{6.17}$$

[see (6.7), (6.12), (6.15) and (6.16)].

We shall require the following fact, which follows from Theorem 1 of Sweeting (1977) and is related to the classical Berry–Esseen theorem.

FACT 6.1. *Let* $(\omega,\zeta), (\omega_1,\zeta_1), (\omega_2,\zeta_2), \dots$ *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* $(\overline{Z}_1, \overline{Z}_2)$ *be bivariate normal with mean vector* 0, $\operatorname{Var}(\overline{Z}_1) = \operatorname{Var}(\overline{Z}_2) = 1$ *and* $\operatorname{Cov}(\overline{Z}_1, \overline{Z}_2) = \operatorname{Cov}(\omega,\zeta) = \rho$. *Then there exist universal positive constants* A_1, A_2 *and* A_3 *such that*

$$\left| E\left|\frac{\sum_{i=1}^n \zeta_i}{\sqrt{n}}\right| - E|\overline{Z}_1| \right| \le \frac{A_1}{\sqrt{n}} E|\zeta|^3 \tag{6.18}$$

and, *whenever* $\rho^2 < 1$,

$$\left| E\left|\frac{\sum_{i=1}^n \omega_i}{\sqrt{n}} \cdot \frac{\sum_{i=1}^n \zeta_i}{\sqrt{n}}\right| - E|\overline{Z}_1 \overline{Z}_2| \right| \le \frac{A_2}{(1-\rho^2)^{3/2}} \frac{1}{\sqrt{n}} (E|\omega|^3 + E|\zeta|^3) \tag{6.19}$$

and

$$\left| E\left[\frac{\sum_{i=1}^n \omega_i}{\sqrt{n}} \cdot \left|\frac{\sum_{i=1}^n \zeta_i}{\sqrt{n}}\right|\right] \right| \le \frac{A_3}{(1-\rho^2)^{3/2}} \frac{1}{\sqrt{n}} (E|\omega|^3 + E|\zeta|^3). \tag{6.20}$$

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

$$\lim_{n\to\infty} \int_C \left\{ \sqrt{n} E|f_{\eta,K}(x) - Ef_{n,K}(x)| - E|Z|\sqrt{n \operatorname{Var}(f_{\eta,K}(x))} \right\} dx = 0 \tag{6.21}$$

and

$$\lim_{n\to\infty} \int_C \left\{ \sqrt{n} E|f_{n,K}(x) - Ef_{n,K}(x)| - E|Z|\sqrt{n \operatorname{Var}(f_{\eta,K}(x))} \right\} dx = 0. \tag{6.22}$$

PROOF. Let η_1 denote a Poisson random variable with mean 1, independent of $X_1, X_2, \ldots$, and set

$$Y_n(x) = \left[\sum_{j \le \eta_1} K\left(\frac{x - X_j}{h_n}\right) - EK\left(\frac{x - X}{h_n}\right) \right] \Big/ \sqrt{EK^2\left(\frac{x - X}{h_n}\right)}. \tag{6.23}$$

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

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

Using (6.8) and that $f(x) \ge \delta > 0$ for all $x \in C$, we get that, for all large enough n uniformly in $x \in C$ for some constant $B_0 > 0$,

$$\sup_{x\in C} E|Y_n(x)|^3 \le h_n^{-1/2} B_0. \tag{6.24}$$

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

$$T_\eta^K(x) := \frac{\sqrt{n}\{f_{\eta,K}(x) - Ef_{n,K}(x)\}}{\sqrt{h_n^{-2} EK^2((x - X)/h_n)}} \stackrel{d}{=} \frac{\sum_{i=1}^n Y_n^{(i)}(x)}{\sqrt{n}}. \tag{6.25}$$

Therefore, by (6.18),

$$\sup_{x\in C} \left| \frac{E|\sqrt{n}\{f_{\eta,K}(x) - Ef_{n,K}(x)\}|}{\sqrt{h_n^{-2} EK^2((x - X)/h_n)}} - E|Z| \right| \le \frac{A_1}{\sqrt{n}} \sup_{x\in C} E|Y_n(x)|^3. \tag{6.26}$$

Now, by (6.24), in combination with (6.26) and

$$\sup_{x\in C} \sqrt{h_n^{-2} EK^2\left(\frac{x - X}{h_n}\right)} = \sup_{x\in C} \sqrt{n \operatorname{Var}(f_{\eta,K}(x))} = O(h_n^{-1/2}), \tag{6.27}$$

we get that

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

Similarly, we obtain, by Fact 6.1,

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

which by (6.17) implies

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

LEMMA 6.4. *Whenever $h_n \to 0$, $nh_n \to \infty$ and C has finite Lebesgue measure $m(C)$, we have*

(6.28) $\quad I_C(x + h_n t)$ *converges in measure to* $I_C(x) = 1 \qquad$ *on* $C \times [-1, 1]$

and

(6.29) $\quad f(x + h_n t)\, I_C(x + h_n t)$ *converges in measure to* $f(x) \qquad$ *on* $C \times [-1, 1]$,

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,

$$\tfrac{1}{2}\int_{x-h_n}^{x+h_n} h_n^{-1} I_C(y)\, dy \to I_C(x).$$

Thus, by the dominated convergence theorem,

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

which, in other words, says

$$(m \times m)\{(x, t) \in C \times [-1, 1] : 1 - I_C(x + h_n t) \neq 0\}$$
$$= 2m(C) - \int_C \int_{-1}^{1} I_C(x + h_n t)\, dt\, dx \to 0,$$

yielding (6.28).

To prove (6.29), just note that, by the continuity of f on C,

$$\big(f(x+h_nt)-f(x)\big)I_C(x+h_nt)\to 0$$

for all $(x,t)\in C\times[-1,1]$, and that, by (6.28),

$$f(x)\big(I_C(x+h_nt)-I_C(x)\big)\to 0$$

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

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

$$\begin{aligned}\sigma_n(C,K_1,K_2) := nE\prod_{s=1,2}\bigg(\int_C\{&|f_{\eta,K_s}(x)-Ef_{n,K_s}(x)| \\ &- E|f_{\eta,K_s}(x)-Ef_{n,K_s}(x)|\}\,dx\bigg),\end{aligned}\tag{6.30}$$

$$\begin{aligned}\sigma_n^2(C,K) &:= \sigma_n(C,K,K) \\ &= E\bigg(\sqrt{n}\int_C\{|f_{\eta,K}(x)-Ef_{n,K}(x)| \\ &\qquad - E|f_{\eta,K}(x)-Ef_{n,K}(x)|\}\,dx\bigg)^2\end{aligned}\tag{6.31}$$

and

$$P(C)=\int_C f(x)\,dx=\Pr\{X\in C\}.\tag{6.32}$$

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

$$\lim_{n\to\infty}\sigma_n(C,K_l,K_s)=P(C)\sigma(K_l,K_s).\tag{6.33}$$

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

$$\lim_{n\to\infty}\sigma_n^2(C,K)=P(C)\sigma^2(K).\tag{6.34}$$

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 $|f_{\eta,K_1}(x)-Ef_{n,K_1}(x)|$ and $|f_{\eta,K_2}(y)-Ef_{n,K_2}(y)|$ are independent. This follows from the fact that they are functions of independent increments of a Poisson process with intensity nf. Therefore

$$\begin{aligned}\sigma_n(C, K_1, K_2) &= n\int_C\int_C \operatorname{Cov}\bigl(|f_{\eta,K_1}(x) - Ef_{n,K_1}(x)|,\\ &\qquad\qquad |f_{\eta,K_2}(y) - Ef_{n,K_2}(y)|\bigr)\,dx\,dy\\ &= \int_C\int_C I(|x-y| \le h_n)\operatorname{Cov}\bigl(|T_\eta^{K_1}(x)|, |T_\eta^{K_2}(y)|\bigr)\\ &\qquad\qquad \times\sqrt{k_n(K_1,x)k_n(K_2,y)}\,dx\,dy.\end{aligned}$$

Keeping (6.15), (6.7), (6.8) and (6.25) in mind and noting that, with $m(C) < \infty$,

$$\int_C\int_C I(|x-y| \le h_n)\,dx\,dy \le 2h_n m(C),$$

we see that

$$\sigma_n(C, K_1, K_2) = \overline{\sigma}_n(C, K_1, K_2) + o(1),$$

where

$$\begin{aligned}\overline{\sigma}_n(C, K_1, K_2) = \int_C\int_C & I(|x-y| \le h_n)\operatorname{Cov}\bigl(|T_\eta^{K_1}(x)|, |T_\eta^{K_2}(y)|\bigr)\\ &\times h_n^{-1}\|K_1\|_2\|K_2\|_2\sqrt{f(x)\,f(y)}\,dx\,dy.\end{aligned}$$

Now let $(Z_{1,n}(x), Z_{2,n}(y))$, $x, y \in \mathbf{R}$, be a mean zero bivariate Gaussian process such for each $(x, y) \in \mathbf{R}^2$, $(Z_{1,n}(x), Z_{2,n}(y))$ and $(T_\eta^{K_1}(x), T_\eta^{K_2}(y))$ have the same covariance structure. In particular, we have

$$\bigl(Z_{1,n}(x), Z_{2,n}(y)\bigr) \stackrel{d}{=} \Bigl(\sqrt{1 - \bigl(\rho_n^*(x,y)\bigr)^2}\,Z_1 + \rho_n^*(x,y)\,Z_2, Z_2\Bigr),$$

where Z_1 and Z_2 be independent standard normal random variables, and

$$\rho_n^*(x,y) := E\bigl[T_\eta^{K_1}(x)\,T_\eta^{K_2}(y)\bigr].$$

Set

$$\begin{aligned}\overline{\tau}_n(C, K_1, K_2) = \int_C\int_C & I(|x-y| \le h_n)\operatorname{Cov}\bigl(|Z_{1,n}(x)|, |Z_{2,n}(y)|\bigr)\\ &\times h_n^{-1}\|K_1\|_2\|K_2\|_2\sqrt{f(x)\,f(y)}\,dx\,dy,\end{aligned}$$

which by the change of variables $y = x + th_n$ equals

$$\int_C\int_{-1}^1 g_n(x,t)\,dx\,dt,$$

where

$$g_n(x,t) := I_C(x) I_C(x+th_n)\,\mathrm{Cov}\big(|Z_{1,n}(x)|, |Z_{2,n}(x+th_n)|\big)$$
$$\times \|K_1\|_2 \|K_2\|_2 \sqrt{f(x)\, f(x+th_n)}.$$

Also observe that

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

We will show that, as $n \to \infty$,

$$\overline{\tau}_n(C, K_1, K_2) \to P(C)\,\sigma(K_1, K_2) \tag{6.35}$$

and then, as $n \to \infty$,

$$\overline{\tau}_n(C, K_1, K_2) - \overline{\sigma}_n(C, K_1, K_2) \to 0, \tag{6.36}$$

which will complete the proof of the lemma.

First, consider (6.35). Applying (6.3) of Lemma 6.1, with $H(u) = K_1(u) \times K_2(u+t)$, we get, for each t, as $n \to \infty$, for almost every $x \in \mathbf{R}$, hence for almost every $x \in C$,

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

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

$$h_n^{-1} E K_1^2\left(\frac{x-X}{h_n}\right) \to f(x) \|K_1\|_2^2$$

and

$$h_n^{-1} E K_2^2\left(\frac{x-X}{h_n} + t\right) \to f(x) \|K_2\|_2^2.$$

Notice that we do not need the just-mentioned functions $H(\cdot)$ to belong to $\mathcal{H}_0$. The limit result (6.3) is applied to each of these functions separately. Thus, for each t and almost every $x \in C$, as $n \to \infty$,

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

and thus

$$\mathrm{Cov}\big(|Z_{1,n}(x)|, |Z_{2,n}(x+th_n)|\big)$$
$$\to \mathrm{Cov}\Big(\big|\sqrt{1-\rho^2(K_1,K_2,t)}Z_1 + \rho(K_1,K_2,t) Z_2\big|, |Z_2|\Big).$$

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

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

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 $n \to \infty$, $\overline{\tau}_n(C,K_1,K_2)$ converges to

$$P(C)\|K_1\|_2\|K_2\|_2 \int_{-1}^{1} \operatorname{Cov}\Big(\big|\sqrt{1-\rho^2(K_1,K_2,t)}Z_1 \\ + \rho(K_1,K_2,t)Z_2\big|, |Z_2|\Big)\,dt,$$

which, since $\rho(K_1,K_2,t)=0$ whenever $|t|>1$, equals $P(C)\sigma(K_1,K_2)$. This completes the proof of (6.35).

Now we turn to (6.36). Set

$$G_n(x,t) = \|K_1\|_2\|K_2\|_2 I_C(x) I_C(x+th_n)\sqrt{f(x)f(x+th_n)}.$$

Notice that

$$\int_C \int_{-1}^{1} G_n(x,t)\,dx\,dt \leq 2m(C)\|K_1\|_2\|K_2\|_2 B =: \beta, \tag{6.37}$$

where B is the bound of f on C. We see that

$$\begin{aligned}
&|\overline{\tau}_n(C,K_1,K_2) - \overline{\sigma}_n(C,K_1,K_2)| \\
&\quad \leq \int_C \int_{-1}^{1} \big|E|Z_{1,n}(x)|E|Z_{2,n}(x+th_n)| \\
&\qquad\qquad - E|T_\eta^{K_1}(x)|E|T_\eta^{K_2}(x+th_n)|\big| G_n(x,t)\,dx\,dt \\
&\qquad + \int_C \int_{-1}^{1} \big||EZ_{1,n}(x)Z_{2,n}(x+th_n)| \\
&\qquad\qquad - E|T_\eta^{K_1}(x)T_\eta^{K_2}(x+th_n)|\big| G_n(x,t)\,dx\,dt \\
&\quad =: \Delta_n(1) + \Delta_n(2).
\end{aligned}$$

First, using (6.37) and (6.18) of Fact 6.1 with (6.24), we get

$$\Delta_n(1) = O\bigg(\frac{1}{\sqrt{nh_n}}\bigg).$$

Choose any $0 < \varepsilon < 1$ and set

$$A_n(\varepsilon) = \big\{(x,t) : 1 - \big(\rho_n^*(x, x+th_n)\big)^2 \geq \varepsilon\big\}.$$

Now

$$\begin{aligned}\Delta_n(2) \le & \int_C \int_{-1}^{1} \big|1 - |EZ_{1,n}(x)Z_{2,n}(x+th_n)|\big| \mathbb{1}_{A_n^c(\varepsilon)}(x,t) G_n(x,t)\,dx\,dt \\ & + \int_C \int_{-1}^{1} \big|1 - E|T_\eta^{K_1}(x)T_\eta^{K_2}(x+th_n)|\big| \mathbb{1}_{A_n^c(\varepsilon)}(x,t) G_n(x,t)\,dx\,dt \\ & + \int_C \int_{-1}^{1} \big||EZ_{1,n}(x)Z_{2,n}(x+th_n)| - E|T_\eta^{K_1}(x)T_\eta^{K_2}(x+th_n)|\big| \\ & \qquad \times \mathbb{1}_{A_n(\varepsilon)}(x,t) G_n(x,t)\,dx\,dt \\ =:& \,\overline{\Delta}_{n,1}(2,\varepsilon) + \overline{\Delta}_{n,2}(2,\varepsilon) + \Delta_n(2,\varepsilon) \\ =:& \,\overline{\Delta}_n(2,\varepsilon) + \Delta_n(2,\varepsilon).\end{aligned}$$

To bound $\overline{\Delta}_n(2,\varepsilon)$, we use the elementary fact that if X and Y are mean-zero and variance-one random variables with $\rho = E(XY)$, then $1 - E|XY| \le 1 - |\rho| \le 1 - \rho^2$, in combination with (6.37), to get that

$$\overline{\Delta}_n(2,\varepsilon) \le 2\varepsilon\beta.$$

Next, we use (6.37) and (6.19) of Fact 6.1 with (6.24) to get

$$\Delta_n(2,\varepsilon) = O\bigg(\frac{1}{\sqrt{nh_n}}\bigg).$$

Thus, for all $0 < \varepsilon < 1$,

$$\limsup_{n\to\infty} \big|\overline{\tau}_n(C,K_1,K_2) - \overline{\sigma}_n(C,K_1,K_2)\big| \le 2\varepsilon\beta,$$

which, since $\varepsilon > 0$ can be chosen arbitrarily small, yields (6.36). This finishes the proof of Lemma 6.5. $\square$

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 $K \in \mathcal{K}$,

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

We shall prove the asymptotic normality of the random vector

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

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

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

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

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

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

$$h_n \le h_n^* \le 2h_n. \tag{6.38}$$

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

$$\Delta_n(x) := \mu_1\,\Delta_n^{K_1}(x) + \cdots + \mu_m \Delta_n^{K_m}(x),$$

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

By Lemma 6.5, we have

$$\lim_{n\to\infty} \sigma_n^2(C,\mu,\mathbf{K}) = P(C)\sum_{l,s=1}^m \mu_l\mu_s\sigma(K_l,K_s). \tag{6.39}$$

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

$$P(C)\sum_{l,s=1}^m \mu_l\mu_s\sigma(K_l,K_s) > 0. \tag{6.40}$$

LEMMA 6.6. *Whenever $h_n \to 0$, $nh_n \to \infty$ 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,*

$$E|\alpha_{i,n}|^3 \le B_1 h_n^{3/2}. \tag{6.41}$$

PROOF. Notice that

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

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

Clearly,

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

which by repeated applications of Jensen's inequality is, for some constant $D_m > 0$, less than or equal to

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

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

$$n^{3/2} E\big|f_{\eta,K_j}(w) - Ef_{n,K_j}(w)\big|^3 \le \left(\frac{45}{\log 3}\right)^3 \max\left\{\left[\frac{1}{h_n^2} E\left(K_j^2\left(\frac{w-X}{h_n}\right)\right)\right]^{3/2},\ E\left(\frac{1}{\sqrt{n}h_n^3}|K_j|^3\left(\frac{w-X}{h_n}\right)\right)\right\},$$

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

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

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

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

which implies that, uniformly in i,

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

This last bound is, for some $B_0 > 0$, uniformly in n, less than or equal to $B_0 h_n^{3/2}$. Now (6.41) follows from (6.39) and (6.40). $\square$

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

$$S_n = \sum_{i=-m_n}^{m_n-1} \alpha_{i,n}, \tag{6.42}$$

$$U_n = \frac{1}{\sqrt{n}}\left\{\sum_{j=1}^{\eta} I(X_j \in [-M, M]) - n\Pr\{X \in [-M, M]\}\right\} \tag{6.43}$$

and

$$V_n = \frac{1}{\sqrt{n}}\left\{\sum_{j=1}^{\eta} I(X_j \notin [-M, M]) - n\Pr\{X \notin [-M, M]\}\right\}. \tag{6.44}$$

Clearly, (S_n, U_n) is independent of V_n. To check this assertion, it suffices to use the independence described at the beginning of Section 2 for the case $A_1 = [-M, M]$ and $A_2 = \mathbf{R}\backslash[-M, M]$ (recall that we have $C \subset [-M + \nu, M - \nu]$ and $h_n \le \nu$). Observe that

$$\mathrm{Var}(S_n) = 1 \quad \text{and} \quad \mathrm{Var}(U_n) = 1 - \alpha, \tag{6.45}$$

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

LEMMA 6.7. *Whenever $h_n \to 0$, $nh_n \to \infty$ 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,*

$$|\operatorname{Cov}(S_n, U_n)| \le \frac{B_2}{\sqrt{n} h_n}. \tag{6.46}$$

PROOF. Notice that

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

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,

$$\left|\operatorname{Cov}\left(\sqrt{n} \int_C |f_{\eta,K}(x) - Ef_{n,K}(x)|\, dx, U_n\right)\right| \le \frac{B_3}{\sqrt{n} h_n}. \tag{6.47}$$

Now, for any $x \in C$,

$$\left(\frac{\sqrt{n}(f_{\eta,K}(x) - Ef_{n,K}(x))}{\sqrt{k_n(K, x)}}, \frac{U_n}{\sqrt{P[-M, M]}}\right) \stackrel{d}{=} \left(\sum_{i=1}^n (Y_n^{(i)}(x), U^{(i)})\right),$$

where $(Y_n^{(i)}(x), U^{(i)})$, $i = 1, \dots, n$, are i.i.d. $(Y_n(x), U)$, with

$$Y_n(x) = \left[\sum_{j \le \eta_1} K\left(\frac{x - X_j}{h_n}\right) - EK\left(\frac{x - X}{h_n}\right)\right] \Big/ \sqrt{EK^2\left(\frac{x - X}{h_n}\right)} \tag{6.48}$$

and

$$U = \left[\sum_{j \le \eta_1} I(X_j \in [-M, M]) - \Pr\{X \in [-M, M]\}\right] \Big/ \sqrt{\Pr\{X \in [-M, M]\}},$$

η_1 denoting a Poisson random variable with mean 1, independent of $X_1, X_2, \dots$.

Notice that $EY_n(x) = EU = 0$, $\operatorname{Var} Y_n(x) = \operatorname{Var} U = 1$ and

$$\begin{aligned}\big|\operatorname{Cov}(Y_n(x), U)\big| &= \left|\frac{E[K((x-X)/h_n)I(X \in [-M, M])]}{\sqrt{EK^2((x-X)/h_n)}\sqrt{\Pr\{X \in [-M, M]\}}}\right| \\ &\le \frac{E|K((x-X)/h_n)|}{\sqrt{EK^2((x-X)/h_n)}\sqrt{\Pr\{X \in [-M, M]\}}}.\end{aligned}$$

This last bound is, for some $D > 0$ and all $n \ge 1$, uniformly for $x \in C$, less than or equal to $D\sqrt{h_n}$, which, in turn, is less than or equal to ε for all large enough n and any $0 < \varepsilon < 1$. This, in combination with (6.9), (6.12) and (6.24), gives, by using (6.18) and (6.20) in Fact 6.1 and Lemma 2.3, that, for some constant A, uniformly on $x \in C$,

$$\begin{aligned}&\big|\operatorname{Cov}(\sqrt{n}|f_{\eta,K}(x) - Ef_{n,K}(x)|, U_n)\big| \\ &\quad = \left|\operatorname{Cov}\left(\frac{\sqrt{n}|f_{\eta,K}(x) - Ef_{n,K}(x)|}{\sqrt{k_n(K,x)}}, \frac{U_n}{\sqrt{P[-M,M]}}\right)\right| \\ &\qquad \times \sqrt{k_n(K,x)P[-M,M]} \\ &\quad \le \frac{A}{\sqrt{nh_n}}\sqrt{k_n(K,x)}.\end{aligned} \tag{6.49}$$

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

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

which, when combined with (6.49) and $m(C) < \infty$, completes the proof of (6.47). This in turn, by (6.39) and (6.40), gives (6.46). $\square$

Hence, by (6.45) and (6.46), as $n \to \infty$,

$$\operatorname{Var}(\lambda_1 S_n + \lambda_2 U_n) \to \lambda_1^2 + \lambda_2^2(1-\alpha). \tag{6.50}$$

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

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

(i) $\operatorname{Var}(\sum_{i=1}^{k_n} X_{i,k_n}) \to 1$ *as* $n \to \infty$, *and*
(ii) *for some* $2 < s \le 3$, $\sum_{i=1}^{k_n} E|X_{i,k_n}|^s \to 0$ *as* $n \to \infty$.

Then

$$\sum_{i=1}^{k_n} X_{i,k_n} \xrightarrow{d} Z,$$

where Z *is a standard normal random variable.*

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

$$(S_n, U_n) \xrightarrow{d} (Z_1, \sqrt{1-\alpha}Z_2) \tag{6.51}$$

as $n \to \infty$, *where* Z_1 *and* Z_2 *are independent standard normal random variables.*

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

$$\lambda_1 S_n + \lambda_2 U_n \xrightarrow{d} \lambda_1 Z_1 + \lambda_2 \sqrt{1-\alpha} Z_2. \tag{6.52}$$

Set

$$u_{i,n} := \frac{1}{\sqrt{n}} \left\{ \sum_{j=1}^{\eta} I\bigl(X_j \in [ih_n^*, (i+1)h_n^*]\bigr) - n \Pr\bigl\{X \in [ih_n^*, (i+1)h_n^*]\bigr\} \right\} \tag{6.53}$$

and

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

Now, by Jensen's inequality,

$$E|y_{i,n}|^3 \le 4\bigl[|\lambda_1|^3 E|\alpha_{i,n}|^3 + |\lambda_2|^3 E|u_{i,n}|^3\bigr]. \tag{6.54}$$

By Lemma 6.6,

$$\sum_{i=-m_n}^{m_n-1} E|\alpha_{i,n}|^3 \le 2m_n B_1 h_n^{3/2} \to 0. \tag{6.55}$$

Set

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

By Lemma 2.3, there is a universal constant A such that

$$\begin{aligned} \sum_{i=-m_n}^{m_n-1} E|u_{i,n}|^3 &\le A n^{-3/2} \sum_{i=-m_n}^{m_n-1} \bigl((np_{i,n})^{3/2} + np_{i,n}\bigr) \\ &\le A \max_{-m_n \le i \le m_n-1} \bigl(\sqrt{p_{i,n}} + n^{-1/2}\bigr) \to 0. \end{aligned} \tag{6.56}$$

Combining (6.54)–(6.56), we obtain

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

Moreover, note that the sequence $y_{i,n}$, $-m_n \le i \le m_n - 1$, is one-dependent and, by (6.50),

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

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

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

Since

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

we can prove (6.51) by the Cramér–Wold device [e.g., Billingsley (1968)]. $\square$

Set

$$L_n(C) = \frac{\sqrt{n}}{\sigma_n(C,\mu,\mathbf{K})} \sum_{s=1}^{m} \mu_s \int_C \{|f_{n,K_s}(x) - Ef_{n,K_s}(x)| - E|f_{n,K_s}(x) - Ef_{n,K_s}(x)|\}\,dx. \tag{6.57}$$

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

$$L_n(C) \xrightarrow{d} Z \tag{6.58}$$

as $n \to \infty$, where Z is a standard normal random variable.

PROOF. Recall the notation in (6.42). Note that

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

and, conditioned on $\eta = n$,

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

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

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

Assertion (6.58) now follows from Lemma 6.3. □

COMPLETION OF THE PROOFS OF THEOREMS 1.1 AND 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

$$\lim_{n\to\infty} \int_{C_k^c} f(x)\, dx = 0. \tag{6.59}$$

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

$$L_n(C_k) \stackrel{d}{\to} Z,$$

and, by (6.39),

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

Further, by Lemma 6.2,

$$\limsup_{n\to\infty} E\Bigg(\sqrt{n} \sum_{s=1}^{m} \mu_s \int_{C_k^c} \{ |f_{n,K_s}(x) - Ef_{n,K_s}(x)| \\ - E|f_{\eta,K_s}(x) - Ef_{n,K_s}(x)| \} \, dx \Bigg)^2 \\ \leq 4m \max_{1\leq s\leq m} \|K_s\|_\infty^2 \int_{C_k^c} f(x)\, dx.$$

Now, by (6.59), combined with a standard argument [see Theorem 4.2 of Billingsley (1968)], we conclude, as $n \to \infty$,

$$\sqrt{n} \sum_{s=1}^{m} \mu_s \{ \|f_{n,K_s} - Ef_{n,K_s}\|_1 - E\|f_{n,K_s} - Ef_{n,K_s}\|_1 \} \\ \stackrel{d}{\to} \sqrt{\sum_{l,s=1}^{m} \mu_l \mu_s \sigma(K_l, K_s)}\, Z. \tag{6.60}$$

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

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

which permits us to infer the relationship

$$\lim_{n\to\infty} \mathrm{Var}\left(\sum_{s=1}^{m} \mu_s \xi_n(K_s)\right) = \sum_{l,s=1}^{m} \mu_l \mu_s \sigma(K_l, K_s) \tag{6.61}$$

from (6.60) [see Theorem 6.4 of Billingsley (1968)]. Taking into account that

$$\mathrm{Cov}\big(\xi_n(K_l), \xi_n(K_s)\big)$$
$$= \tfrac{1}{4}\mathrm{Var}\big(\xi_n(K_l) + \xi_n(K_s)\big) - \tfrac{1}{4}\mathrm{Var}\big(\xi_n(K_l) - \xi_n(K_s)\big),$$

we derive (6.1) from (6.61). This completes the proof of Theorem 6.1. Now Theorem 1.1 follows from Theorems 5.1 and 6.1 by well-known 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), pages 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

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

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

$$\sigma(K_\ell, K_s) = \left(\sqrt{\frac{s}{\ell}} - \sqrt{\frac{\ell}{s}}\right) \mathrm{Cov}\left(\left|\sqrt{1-\frac{\ell}{s}}Z_1 + \sqrt{\frac{\ell}{s}}Z_2\right|, |Z_2|\right)$$
$$+ 2\int_{(\sqrt{s/\ell}-\sqrt{\ell/s})/2}^{(\sqrt{s/\ell}+\sqrt{\ell/s})/2} \mathrm{Cov}\left(\left|\sqrt{1-\left[\frac{1}{2}\left(\sqrt{\frac{s}{\ell}}+\sqrt{\frac{\ell}{s}}\right)-u\right]^2}Z_1 + \left[\frac{1}{2}\left(\sqrt{\frac{s}{\ell}}+\sqrt{\frac{\ell}{s}}\right)-u\right]Z_2\right|, |Z_2|\right) du.$$

So $d_2^2(\ell, s) = 2\sigma^2(K_1) - 2\sigma(K_\ell, K_s) := f(s/\ell)$. If we show that $f(u) \to 2\sigma^2(K_1)$ as $u \to \infty$, 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 $\lim_{s\to\infty} \sigma(K_1, K_s) = 0$ by a straightforward computation based on the fact that

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

which is justified by dominated convergence.

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

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

and clearly

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

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

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

and

$$\xi_n(K) \xrightarrow{d} \sigma(K)Z \qquad \text{as } n \to \infty. \qquad \square$$

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 $L_1(\mathbf{R}, \mathcal{B}, m)$. If $h_n \to 0$ and $nh_n \to \infty$, then*

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

PROOF OF THEOREM 1.3 AND PROPOSITION 7.1. Theorem 1, Chapter 3, in Devroye and Györfi (1985) shows that

$$\begin{aligned} &\lim_{n\to\infty} J_n(K) = 0 \quad \text{in probability,} \\ &\lim_{n\to\infty} D_n(K) = 0 \quad \text{in probability,} \end{aligned} \tag{7.2}$$

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

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

and

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

by a 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 theoremthen gives

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

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

$$|J_n(K) - J_n(\bar{K})| \le \int_{\mathbf{R}} |f_{n,K}(x) - f_{n,\bar{K}}(x)|\,dx \le \partial_1(K, \bar{K}) \tag{7.4}$$

and

$$|D_n(K) - D_n(\bar{K})| \le 2\int_{\mathbf{R}} |K(x) - \bar{K}(x)|\,dx = 2\partial_1(K, \bar{K}). \tag{7.5}$$

Since $\mathcal{K}$ is relatively compact in L_1, it is totally bounded in L_1 [e.g., Folland (1999), page 15]. Given $\varepsilon > 0$, let T_ε be a maximal subset of $\mathcal{K}$ satisfying that if $K, \bar{K} \in T_\varepsilon$ and $K \neq \bar{K}$, then $\partial_1(K, \bar{K}) \ge \varepsilon$. Then Card $T_\varepsilon = D(\mathcal{K}, \partial_1, \varepsilon) < \infty$ by total boundedness. We then have, by (7.4) and (7.5),

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

and

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

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 ∞ and then ε tend to 0 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 $L_1(\mathbf{R}, \mathcal{B}, m)$ is relatively compact if and only if:

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

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

REMARK 7.2. Suppose

$$\Bigl(\sup_{K \in \mathcal{K}} J_n(K)\Bigr)^* \to 0 \qquad \text{in probability,} \tag{7.6}$$

where, in addition, we are assuming $h_n \to 0$, $nh_n \to \infty$ and $\int_{\mathbf{R}} K(x)\,dx = 1$ for all $K \in \mathcal{K}$. Let $\{\varepsilon_i\}$ be an i.i.d. sequence of Rademacher variables ($\Pr\{\varepsilon_i = 1\} = \Pr\{\varepsilon_i = -1\} = 1/2$), independent of the sequence $\{X_i\}$. Then

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

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

and

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

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

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

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

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

Then, by Lévy's inequality,

$$\max_{1 \le i \le n} \left(\sup_{K \in \mathcal{K}} \int_{\mathbf{R}} \left| \frac{1}{nh_n} \varepsilon_i K\left(\frac{x - X_i}{h_n}\right)\right| dx\right)^* \to 0 \qquad \text{in probability,} \tag{7.11}$$

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

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

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

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

does not tend to 0. 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

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

Therefore, for all $n \geq 1$,

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

which contradicts (7.8).

Acknowledgment. We thank Jon Wellner for pointing out the Nabeya (1951) reference.

REFERENCES

BARTLETT, M. S. (1938). The characteristic function of a conditional statistic. *J. London Math. Soc.* **13** 62–67.

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

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.

BERLINET, A. and DEVROYE, L. (1994). A comparison of kernel density estimates. *Publ. Inst. Statist. Univ. Paris* **38** 3–59.

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BORISOV, I. (2002). Moment inequalities connected with accompanying Poisson laws in Abelian groups. Preprint.

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, 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. *École d'été de Probabilités de Saint-Flour XII. Lecture Notes in Math.* **1097** 1–142. Springer, Berlin.

DUDLEY, R. M. (1989). *Real Analysis and Probability.* Wadsworth and Brooks/Cole, Pacific Grove, CA.

DUNFORD, N. and SCHWARTZ, J. T. (1958). *Linear Operators*, *Part I*. Wiley, New York.

EGGERMONT, P. P. B. and LARICCIA, V. N. (2001). *Maximum Penalized Likelihood Estimation* **1**. *Density Estimation*. Springer, New York.

EINMAHL, J. H. J. (1987). *Multivariate Empirical Processes.* CWI, Amsterdam.

EINMAHL, U. and MASON, D. M. (1997). Gaussian approximation of local empirical processes indexed by functions. *Probab. Theory Related Fields* **107** 283–301.

FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

FOLLAND, G. B. (1999). *Real Analysis*, 2nd ed. Wiley, New York.

GINÉ, E. and ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **12** 929–989.

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.

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.

NABEYA, S. (1951). Absolute moments in 2-dimensional normal distributions. *Ann. Inst. Statist. Math.* **3** 2–6.

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. Lett.* **19** 97–99.

PYKE, R. and SHORACK, G. R. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff–Savage theorems. *Ann. Math. Statist.* **39** 755–771.

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.

E. GINÉ
DEPARTMENT OF MATHEMATICS
AND STATISTICS
MATH. SCIENCE BUILDING 402
STORRS, CONNECTICUT 06269-3009
E-MAIL: gine@uconnvm.uconn.edu

D. M. MASON
DEPARTMENT OF FOOD
AND RESOURCE ECONOMICS
206 TOWNSEND HALL
UNIVERSITY OF DELAWARE
NEWARK, DELAWARE 19717
E-MAIL: davidm@udel.edu

A. YU. ZAITSEV
LABORATORY OF STATISTICAL METHODS
ST. PETERSBURG BRANCH
OF THE STEKLOV MATHEMATICAL INSTITUTE
27 FONTANKA
ST. PETERSBURG 191011
RUSSIA
E-MAIL: zaitsev@pdmi.ras.ru