# ON LOCAL *U*-STATISTIC PROCESSES AND THE ESTIMATION OF DENSITIES OF FUNCTIONS OF SEVERAL SAMPLE VARIABLES

BY EVARIST GINÉ[1] AND DAVID M. MASON[2]

*University of Connecticut and University of Delaware*

A notion of local *U*-statistic process is introduced and central limit theorems in various norms are obtained for it. This involves the development of several inequalities for *U*-processes that may be useful in other contexts. This local *U*-statistic process is based on an estimator of the density of a function of several sample variables proposed by Frees [*J. Amer. Statist. Assoc.* **89** (1994) 517–525] and, as a consequence, uniform in bandwidth central limit theorems in the sup and in the $L_p$ norms are obtained for these estimators.

**1. Introduction.** Let $X, X_1, X_2, \ldots$ be i.i.d. random variables taking values in **R**, with common density function $f$ and consider the kernel density estimator of $f$ defined for $t \in \mathbf{R}$,

$$(1.1) \qquad f_n(t, h_n) = (nh_n)^{-1} \sum_{i=1}^n K\left(h_n^{-1}(t - X_i)\right) =: n^{-1} \sum_{i=1}^n K_{h_n}(t - X_i),$$

where $\{h_n\}_{n \geq 1}$ is a sequence of positive constants converging to zero at the rate $nh_n \to \infty$ and the kernel $K$ is an integrable (real) function of bounded variation satisfying $\int_{\mathbf{R}} K(x)\,dx = 1$ (Parzen [27]). It is easy to prove that, subject to smoothness conditions on $f$, for each $t \in \mathbf{R}$,

$$\sqrt{h_n} u_n(t) := \sqrt{nh_n}\{f_n(t, h_n) - Ef_n(t, h_n)\} \to_d N(0, \|K\|_2^2 f(t)),$$

whereas for any choice of $t_1 \neq t_2$ the random variables $\sqrt{h_n} u_n(t_1)$ and $\sqrt{h_n} u_n(t_2)$ are asymptotically independent. This means that $\sqrt{h_n} u_n$ cannot converge weakly to a continuous bounded process on any nontrivial subinterval of **R**. Since the bias $Ef_n(t, h_n) - f(t)$ can always be dealt with under the usual conditions on $K$ and $f$, this tells us that $f_n(t, h_n)$ estimates the density $f$ at a much slower rate than $n^{-1/2}$. On the other hand, Frees [15] discovered the perhaps surprising fact that the densities $f_g(t)$ of some symmetric real functions $g(X_1, \ldots, X_m)$ of $m > 1$

i.i.d. random variables can be estimated at each fixed $t$ at the rate $n^{-1/2}$, using the $U$-statistic estimator

$$\frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I_n^m} K_{h_n}(t - g(X_{i_1}, \ldots, X_{i_m})),$$

where $\mathbf{i} = (i_1, \ldots, i_m)$ and $I_n^m = \{(i_1, \ldots, i_m) : 1 \le i_j \le n, i_j \ne i_k \text{ if } j \ne k\}$. Schick and Wefelmeyer [32] consider the special case

$$g(X_1, \ldots, X_m) = \sum_{i=1}^m u_i(X_i),$$

which however is not necessarily a symmetric function. Using convolution kernels, they also obtain the (in probability) rate of $n^{-1/2}$ for the sup norm and the $L_1$ norm measure of the discrepancy between the estimator and the density (actually, they obtain limit theorems in distribution). These results require smoothness conditions on the kernel, the density and certain conditional densities.

Our aim is to extend the Schick and Wefelmeyer results to the Frees framework and with $g$ not necessarily symmetric. Moreover, we frame our results uniform in bandwidth in the sup norm and the $L_p$ norm with $p \ge 1$ so that they can be used with adaptive bandwidth estimators. Also, we do this not necessarily in one dimension, but in $\mathbf{R}^d$. These extensions substantially increase the scope of applicability of the results and, as a consequence, we can show how our results recover, extend and/or improve upon, previous work by several authors. We shall not discuss removing the bias in general but only in some examples and very briefly. The bias is not probabilistic and can always be treated by adding enough smoothness to the kernel and the density (see, e.g., the two references just cited and Ahmad and Fan [1]).

We shall begin by generalizing the setup in Giné, Mason and Zaitsev [21] (concretely, that of Examples 1.2 and 1.3 there) and Mason [25] to $U$-statistics. Throughout this paper, we let $X, X_i, i \in \mathbf{N}$, be i.i.d. random variables taking values in a measurable space $(S, \mathscr{S})$; let $g : S^m \mapsto \mathbf{R}^d$, $1 \le d < \infty$, be a measurable function; let $K : \mathbf{R}^d \mapsto \mathbf{R}$ be an integrable measurable function that integrates to 1 (a "kernel"); and let $0 < a \le b < \infty$. Then, for $t \in \mathbf{R}^d$ and $\lambda \in [a, b]$, we introduce the *local U-statistic*

$$(1.2) \qquad U_n(t, \lambda) := \frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I_n^m} K_{\lambda h_n}(t - g(X_{i_1}, \ldots, X_{i_m})),$$

where, here and elsewhere in this paper, for functions $H : \mathbf{R}^d \mapsto \mathbf{R}$ and $h > 0$, the notation

$$(1.3) \qquad\qquad H_h(t) := h^{-1} H(t/h^{1/d}), \qquad t \in \mathbf{R}^d,$$

is in force. The term "local" just reflects the fact that the $U$-process (1.2) is of a special kind, namely a convolution of an approximate identity $K_{\lambda h_n}$ with an

"empirical measure," in this case $((n - m)!/n!) \sum_{I_n^m} \delta_{g(X_{i_1}, \dots, X_{i_m})}$, and therefore, for each value of $t$, the largest contributions to the statistic come from the values $g(X_{i_1}, \dots, X_{i_m})$ closest to $t$.

Special cases of $U_n(t, \lambda)$, when $X_1, \dots, X_n$ are i.i.d. $\mathbf{R}^d$ valued, include the interpoint distance studied by Jammalamadaka and Janson [23], with $g(x, y) = |x - y|$,

$$h_n U_n(0, 1) = \frac{1}{n(n-1)} \sum_{\mathbf{i} \in I_n^2} I\{|X_{i_1} - X_{i_2}| \le h_n\}$$

and the related short distance process studied by Eastwood and Horváth [12],

$$\frac{1}{n(n-1)} \sum_{\mathbf{i} \in I_n^2} I\{\mathbf{d}(X_{i_1}, X_{i_2}) \le \lambda h_n\}, \qquad 0 \le \lambda \le 1,$$

where $\mathbf{d}$ is a distance on $\mathbf{R}^d$, as well as the $U$-statistic estimator of the density of the sum $X_1 + \cdots + X_m$,

$$\frac{(n - m)!}{n!} \sum_{\mathbf{i} \in I_n^m} K_{h_n}(t - X_{i_1} - \cdots - X_{i_m}).$$

Our goal is to obtain central limit theorems for the following *local U-statistic process* formed from $U_n(t, \lambda)$:

(1.4)
$$u_{n,\lambda}(t) := \sqrt{n}\{U_n(t, \lambda) - E K_{\lambda h_n}(t - g(X_1, \dots, X_m))\},$$
$$t \in \mathbf{R}^d, \lambda \in [a, b].$$

The case when $m = 1$ is a special case of the local empirical process studied in Mason [25]. We shall confine our attention to the case $m \ge 2$, which we shall soon see has a radically different asymptotic behavior than the case $m = 1$. Occasionally, we may restrict the process to $t \in D \subset \mathbf{R}^d$, where $D$ may even consist of a single point.

The limit theorems to be obtained for these processes will be in the sup and in the $L_p$ norms, $1 \le p < \infty$, uniformly in $\lambda \in [a, b]$ (precise definitions in the next section). The reason these results will be true will be essentially the same as in Frees [15], namely: (1) the process (1.4) is equivalent to its linear part, that is, all the terms in its Hoeffding decomposition of order higher than one tend to zero in the appropriate way; (2) that this linearization, which is a smoothed empirical process, is equivalent to the empirical process without smoothing, hence *independent* of $\lambda$ and $h_n$; (3) that this empirical process, under appropriate and quite weak conditions, satisfies the central limit theorem in the sup norm or in the $L_p$ norms.

The devil is in the details, and extending the Frees result from one point $t \in \mathbf{R}$, to uniformity in $t \in \mathbf{R}^d$ and in $\lambda \in [a, b]$, makes for very different proofs and requires a substantial amount of technique, some of it new.

The statements of the central limit theorems, along with examples, are collected in Section 2 and all the proofs are postponed to Section 3. In the process of establishing our results we shall develop some tools that should be of separate interest. Among them include tight bounds for the absolute moment of the supremum of the $U$-statistic process under a uniform covering number bound generalizing a similar bound obtained by Einmahl and Mason [13] and Giné and Koltchinskii [17] (see also Giné and Guillou [16]) for the usual empirical process. All of these results were motivated by Proposition 6.2 of Talagrand [35], which is an expectation bound for VC classes of sets, where VC stands for Vapnik and Červonenkis. We also obtain moment and exponential inequalities for Banach space valued $U$-statistics which, although not necessarily optimal, are very easy to apply and are well adapted to the problems treated in this article. In a sequel to this paper (Giné and Mason [20]), we derive the corresponding functional laws of the logarithm for $u_{n,\lambda}$.

**2. Main results and examples.** As is usual when dealing with empirical processes, we take $((S \times T)^{\mathbf{N}}, (\mathcal{S} \otimes \mathcal{T})^{\mathbf{N}}, \mathrm{Pr})$ as the underlying probability space, where $(S, \mathcal{S})$ is a measurable space, $T = \{-1, 1\}$, $\mathcal{T}$ is the family of all subsets of $T$, and $\mathrm{Pr} = P^{\mathbf{N}} \times (P')^{\mathbf{N}}$, $P$ a probability measure on $(S, \mathcal{S})$ and $P'$ the uniform distribution on $T$. Then the random variables $X_i$ are the projections $(S \times T)^{\mathbf{N}} \mapsto S$, $X_i(s_1, t_1, s_2, t_2, \ldots) = s_i$, which are i.i.d. with law $P$. We will occasionally use the random variables $\varepsilon_i(s_1, t_1, s_2, t_2, \ldots) = t_i$, which obviously satisfy $\mathrm{Pr}\{\varepsilon_i = 1\} = \mathrm{Pr}\{\varepsilon_i = -1\} = 1/2$. Note that the random variables $\{X_i, \varepsilon_j : i, j \in \mathbf{N}\}$ are independent. The variables $\varepsilon_i$ are often called Rademacher variables. Sometimes we will write $X$ for $X_1$.

All the asymptotic results on the process $u_{n,\lambda}$ in this article require the following key assumption.

(CD)    For each $i = 1, \ldots, m$, the random variable $g(X_1, \ldots, X_m)$, conditionally on $X_i = x$, $x \in S$, has a density $\overline{f}_i(t, x)$, $t \in \mathbf{R}^d$, which is jointly measurable in $t$ and $x$.

Note that, setting

$$(2.1) \qquad\qquad \overline{f} = \sum_{i=1}^{m} \overline{f}_i,$$

the function

$$(2.2) \qquad\qquad f_g(t) := E[\overline{f}(t, X)]/m, \qquad t \in \mathbf{R}^d,$$

defines a density for the random variable $g(X_1, \ldots, X_m)$. Another condition that we will require for different values of $p \in [1, \infty]$ is

$$(\mathrm{CD}p) \qquad \|\overline{f}(\cdot, x)\|_p < \infty \qquad \text{for all } x \in S \text{ and } \|f_g\|_p < \infty.$$

Here and elsewhere, $\| \cdot \|_p$ denotes the $L_p(\mathbf{R}^d)$ norm for $1 \le p < \infty$ and the sup norm on $\mathbf{R}^d$ for $p = \infty$.

The asymptotics of the processes $u_{n,\lambda}$ will turn out to be equivalent to that of the processes

$$(2.3) \qquad \overline{v}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\overline{f}(t, X_i) - E[\overline{f}(t, X_i)]), \qquad t \in \mathbf{R}^d,$$

which are empirical processes. With some abuse of notation, we say that

$$(2.4) \qquad\qquad \overline{v}_n \text{ converges in law in } L_\infty(\mathbf{R}^d)$$

if condition (CD$\infty$) holds and the class of functions $\{\overline{f}(t, \cdot) : t \in \mathbf{R}^d\}$ is $P$-Donsker. We refer to Dudley [10] or van der Vaart and Wellner [37] for the definition of Donsker classes of functions. The condition (CD$\infty$) is equivalent to the maps $t \mapsto \overline{f}(t, x)$ and $t \mapsto E\overline{f}(t, X)$ being in $\ell^\infty(\mathbf{R}^d)$, the abuse of notation consists in replacing this last space by $L_\infty(\mathbf{R}^d)$, which usually means something else, and the precise meaning of (2.4) is that $\lim_{n \to \infty} EH(v_n)^* = EH(G)$ for every $H : \ell^\infty(\mathbf{R}^d) \mapsto \mathbf{R}$ bounded and continuous, where $G = \{G(t) : t \in \mathbf{R}^d\}$ is the centered Gaussian process with the covariance of $\overline{f}(\cdot, X)$, more precisely a sample continuous version, and $H(\overline{v}_n)^*$ is the (a.s.) smallest measurable function larger than or equal to $H(\overline{v}_n)$. Likewise, for $1 \le p < \infty$, we say that

$$(2.5) \qquad\qquad \overline{v}_n \text{ converges in law in } L_p(\mathbf{R}^d)$$

if the condition (CD$p$) holds and the $L_p(\mathbf{R}^d)$-valued random variable $\overline{f}(\cdot, X)$ satisfies the central limit theorem in this space, that is, there is a centered Gaussian process $G$ with the same covariance as $\overline{f}(\cdot, X)$ and with sample paths in $L_p(\mathbf{R}^d)$, such that for every $H : L_p(\mathbf{R}^d) \mapsto \mathbf{R}$ bounded and continuous,

$$\lim_{n \to \infty} EH(\overline{v}_n) = EH(G).$$

Note that in each case, $\overline{v}_n(\cdot)$ is a random variable that takes values in $L_p(\mathbf{R}^d)$, $1 \le p \le \infty$, although in the case $p = \infty$, with abuse of notation: $\overline{v}_n(\cdot)$ is really in $\ell^\infty(\mathbf{R}^d)$, the space of bounded functions on $\mathbf{R}^d$, and $\overline{v}_n$ is not necessarily measurable.

The following definition describes the type of central limit theorem we will prove for the process $u_{n,\lambda}$.

DEFINITION 1. Let $1 \le p \le \infty$ and assume (CD) and (CD$p$). The processes $u_{n,\lambda}$ converge weakly in $L_p(\mathbf{R}^d)$, uniformly in $a \le \lambda \le b$, to the centered Gaussian process $G$ with the same covariance as $\overline{f}(\cdot, X)$ if

$$\sup_{\lambda \in [a,b]} \|u_{n,\lambda} - \overline{v}_n\|_p \to 0 \qquad \text{in pr}^*$$

and $\overline{v}_n$ converges weakly in $L_p(\mathbf{R}^d)$ in the sense of (2.4) for $p = \infty$ and (2.5) for $p < \infty$.

Convergence in pr* means convergence in probability of the measurable envelopes (the smallest dominating measurable functions).

Note that in the case $p = \infty$, convergence of $u_{n,\lambda}$ to $G$ in the sense of this definition implies that the processes $(t, \lambda) \mapsto u_{n,\lambda}(t)$ converge in law in $\ell^\infty(\mathbf{R}^d \times [a, b])$ to the Gaussian process $G$ (de la Peña and Giné [7], Dudley [10], or van der Vaart and Wellner [37] for this type of convergence). However, our notion gives more. In fact if, for $l = 1, \ldots, N$, we have functions $g_l : S^{m_l} \mapsto \mathbf{R}^{d_l}$ such that the processes $u_{n,\lambda}^{(l)}$ corresponding to $g = g_l$ converge weakly in $L_\infty(\mathbf{R}^{d_l})$, uniformly in $a_l \leq \lambda \leq b_l$, to a centered Gaussian process $G_l$ with the same covariance as $F_l(\cdot, X)$, where $F_l$ is the $\overline{f}$ corresponding to $g = g_l$, then it is easy to conclude, using the obvious multivariate extension of Definition 1, that the vector-valued processes $\overrightarrow{u}_{n,\lambda_1,\ldots,\lambda_N}$ defined by

$$(2.6) \qquad \overrightarrow{u}_{n,\lambda_1,\ldots,\lambda_N}(t_1, \ldots, t_N) = \left(u_{n,\lambda_1}^{(1)}(t_1), \ldots, u_{n,\lambda_N}^{(N)}(t_N)\right),$$
$$t_1 \in \mathbf{R}^{d_1}, \ldots, t_N \in \mathbf{R}^{d_N},$$

converge weakly in $L_\infty(\mathbf{R}^{d_1}) \times \cdots \times L_\infty(\mathbf{R}^{d_N})$ uniformly in $a_l \leq \lambda_l \leq b_l$, $l = 1, \ldots, N$, to the centered vector-valued Gaussian process defined on $\mathbf{R}^{d_1} \times \cdots \times \mathbf{R}^{d_N}$,

$$(2.7) \qquad G(t_1, \ldots, t_N) = \left(G_1(t_1), \ldots, G_N(t_N)\right),$$

with the same covariance/cross covariance matrix as $F(\cdot, X)$, where

$$(2.8) \qquad F(t_1, \ldots, t_N, X) = \left(F_1(t_1, X), \ldots, F_N(t_N, X)\right),$$
$$t_1 \in \mathbf{R}^{d_1}, \ldots, t_N \in \mathbf{R}^{d_N}.$$

2.1. *Central limit theorems.* We still require another definition. We say that a class of measurable functions $\mathcal{F}$ defined on a measurable space $(S, \mathcal{S})$ is VC-type (VC for Vapnik and Červonenkis) with respect to an envelope $F$ (meaning a measurable function $F$ such that $|f| \leq F$ for all $f \in \mathcal{F}$) if the covering number $N(\mathcal{F}, L_2(Q), \varepsilon)$, defined as the smallest number of $L_2(Q)$ open balls of radius $\varepsilon$ required to cover $\mathcal{F}$, satisfies

$$(2.9) \qquad N(\mathcal{F}, L_2(Q), \varepsilon) \leq \left(\frac{A\|F\|_{L_2(Q)}}{\varepsilon}\right)^v, \qquad 0 < \varepsilon \leq 2\|F\|_{L_2(Q)},$$

for some $A \geq 3$ and $v \geq 1$, for every probability measure $Q$ on $\mathcal{S}$ for which $Q(F^2) < \infty$. If (2.9) holds for $\mathcal{F}$, then we say that the VC class $\mathcal{F}$ admits the characteristics $A$ and $v$.

THEOREM 1. *Let $p = \infty$, let* (CD) *and* (CD$\infty$) *hold, and let $K$ be bounded. Assume*:

(a) *each of the classes*

$$(2.10) \qquad \mathcal{K}_n := \big\{ K\big(h^{-1/d}(y - \cdot)\big) : y \in \mathbf{R}^d, ah_n \le h \le bh_n \big\}$$

*is VC-type for a bounded envelope $F_n$ and all the classes $\mathcal{K}_n$ admit the same characteristics $A$ and $v$;*

   (b) *the density $f_g$ of $g(X_1, \ldots, X_m)$ is bounded and the class of functions $\mathcal{F} := \{\overline{f}(t, \cdot) : t \in \mathbf{R}^d\}$ is $P$-Donsker and the identity map*

$$(\mathbf{R}^d, |\cdot|) \mapsto (\mathbf{R}^d, \rho)$$

*is uniformly continuous, where $\rho^2(u, v) = \mathrm{Var}(\overline{f}(u, X) - \overline{f}(v, X))$;*

   (c) *$h_n \to 0$ and $nh_n/(1 \vee \log(A\|F_n\|_{L_2(f_g)}/\sqrt{h_n}))^2 \to \infty$. The processes $u_{n,\lambda}$ then converge weakly in $L_\infty(\mathbf{R}^d)$, uniformly in $a \le \lambda \le b$, to the centered Gaussian process with the same covariance as $\overline{f}(\cdot, X)$.*

REMARK 1. Theorem 1 has obvious applications to the construction of confidence bands for $f_g$. It is formulated uniformly in $a \le \lambda \le b$ so as to allow the possibility for $\lambda$ to be replaced by an estimator $\widehat{\lambda}_n$. Suppose that $\widehat{\lambda}_n = \lambda_n(X_1, \ldots, X_n)$ is an adaptive bandwidth selector such that for all $0 < \varepsilon < 1$ there exist $0 < c < d < \infty$ for which for all large enough $n$

$$(2.11) \qquad \Pr\{c < \widehat{\lambda}_n < d\} \ge 1 - \varepsilon.$$

Then if assumption (a) of Theorem 1 holds for any choice of $0 < a < b < \infty$ we can immediately conclude from (2.11) that the processes $u_{n, \widehat{\lambda}_n}$ converge weakly in $L_\infty(\mathbf{R}^d)$ to the centered Gaussian process with the same covariance as $\overline{f}(\cdot, X)$. The analogous remark holds for Theorems 2 and 3 below. For a thorough discussion of bandwidth estimators that satisfy (2.11) refer to Deheuvels and Mason [5].

REMARK 2. We note that if the classes of functions $\mathcal{F}_i = \{\overline{f}_i(t, \cdot) : t \in \mathbf{R}^d\}$, $i = 1, \ldots, m$, are $P$-Donsker, then so is $\mathcal{F}$ given in (b) (e.g., Theorem 2.10.6 in van der Vaart and Wellner [37] applied to the function $\phi(t_1, \ldots, t_m) = \sum_{i=1}^m t_i/\sqrt{m}$). A sufficient condition for the class of functions $\mathcal{F}$ being $P$-Donsker is that it be of VC-type, since the function $\overline{f}$ being jointly measurable already ensures that this class of functions is measurable (e.g., Definition 2.3.3, Example 2.3.5 and Theorem 2.5.2 (Pollard's CLT) in van der Vaart and Wellner [37], or Dudley [10], Theorem 6.3.1). These two references contain many examples of classes of functions which are of VC-type. We single out one of them which is particularly convenient for condition (a) above. Let $\Psi : \mathbf{R} \mapsto \mathbf{R}$ be a function of bounded variation on $\mathbf{R}$ ($\Psi$ is the difference of two bounded nondecreasing functions). The proof of Lemma 22 in Nolan and Pollard [26] shows that if

$$(2.12) \qquad K(x) = \Psi(p(x)), \qquad x \in \mathbf{R}^d,$$

where $p$ is either a real polynomial on $\mathbf{R}^d$ or the $\alpha$th power of the absolute value of a real polynomial on $\mathbf{R}^d$, $\alpha > 0$, then the class of functions

$$\mathcal{K} = \{K(\gamma^{-1}(t - \cdot)) : t \in \mathbf{R}^d, \gamma > 0\}$$

is VC-type. Moreover, since the function $K(\gamma^{-1}(t - x))$ is jointly measurable, this class is also measurable. Most kernels of interest satisfy condition (2.12), and therefore, condition (a) in Theorem 1.

Sometimes one is interested only in weak convergence in $L_\infty(D)$ uniformly in $a \le \lambda \le b$, where $D$ is a subset of $\mathbf{R}^d$ that may even consist of a single point. For instance, this is the case for the interpoint distance process, where $K(x) = I\{|x| \le 1\}$, $S = \mathbf{R}^d$ and $g(x, y) = x - y$. The following, which is related to a result of Eastwood and Horváth [12], will be a corollary to the proof of the above theorem. We say that a subset $D$ of $\mathbf{R}^d$ is star-shaped about 0 if $x \in D$ implies $\lambda x \in D$ for all $0 \le \lambda \le 1$.

COROLLARY 1.  *Let $D$ be a measurable bounded subset of $\mathbf{R}^d$ star-shaped about 0 and with $\mathrm{Vol}(D) \ne 0$, and set $K = I_D / \mathrm{Vol}(D)$. Assume that $g(X_1, \dots, X_m)$ has a bounded density $f_g$, that $E[\overline{f}(0, X)]^2 < \infty$ and that for all $\varepsilon > 0$,*

$$(2.13) \qquad \lim_{\delta \to 0} \limsup_{n \to \infty} \mathrm{Pr}^* \left\{ \sup_{|u| \le \delta} |\overline{v}_n(u) - \overline{v}_n(0)| > \varepsilon \right\} = 0.$$

*Let $h_n \to 0$ be such that $nh_n \to \infty$. Then, the processes $\lambda u_{n,\lambda}(0)$ converge weakly in $\ell^\infty((0, 1])$ to the process $\lambda \sigma Z$, $0 \le \lambda \le 1$, where $Z$ is standard normal and $\sigma^2 = \mathrm{Var}(\overline{f}(0, X))$.*

It makes sense to define $\lambda u_{n,\lambda}(0)$ as zero for $\lambda = 0$. With this convention, in force from here on, we have weak convergence in $\ell^\infty([0, 1])$ in this corollary.

Condition (2.13) is satisfied, for instance, if the class $\mathcal{F}_\delta := \{\overline{f}(t, \cdot) : |t| \le \delta\}$ is $P$-Donsker for some $\delta > 0$, a condition weaker than the class $\mathcal{F}$ being $P$-Donsker.

Next we state the central limit theorems in the $L_p$ norm, uniform on $a \le \lambda \le b$. We need to recall the definition of Young moduli of exponential type. As in de la Peña and Giné [7], page 188, $\Psi_1(x) := e^x - 1$, but if $\alpha < 1$, since $e^{x^\alpha}$ is only convex for $x \ge x_\alpha := ((1 - \alpha)/\alpha)^{1/\alpha}$, we take as a function $\Psi_\alpha$ that is 0 at 0, convex and increasing, and of the order of $e^{x^\alpha}$ for $x$ large the following:

$$(2.14) \qquad \Psi_\alpha(x) := \tau_\alpha(x) - \alpha \exp((1 - \alpha)/\alpha),$$

where $\tau_\alpha(x)$ equals $\exp(x^\alpha)$ if $x \ge x_\alpha$, and equals the tangent line to the function $y = \exp(x^\alpha)$ at $x = x_\alpha$ for $0 \le x \le x_\alpha$. We also recall that then, for any nonnegative random variable $\xi$ for which $Ee^{(\xi/a)^\alpha} < \infty$ for some $a > 0$,

$$(2.15) \qquad \|\xi\|_{\Psi_\alpha} := \inf\{c : E\Psi_\alpha(\xi/c) \le 1\}.$$

This is a (pseudo)norm and it dominates, up to constants that depend only on $\alpha$ and $p$, all the $L_p$ (pseudo)norms. Simple standard computations show that

$$(2.16) \quad \Pr\{\xi > x\} \leq b \exp\{-(x/a)^\alpha\} \quad \text{for all } x > 0 \implies \|\xi\|_{\Psi_\alpha} \leq Ca$$

for a constant $C$ that depends only on $\alpha$ and $b$. Note also that $\Psi_\alpha^{-1}(u)$ is a constant times $u$ for $0 \leq u \leq \Psi_\alpha(x_\alpha)$ and it is the $1/\alpha$-th power of the logarithm of $u + \alpha \exp((1-\alpha)/\alpha)$ for $u \geq \Psi_\alpha(x_\alpha)$.

Given a kernel $K$ and $0 < a \leq b < \infty$, define

$$(2.17) \quad \mathcal{K}_{[a,b]} := \{K_h : h \in [a,b]\} \cup \{0\}.$$

For any two functions $f$ and $g$ in $L_p$, set

$$(2.18) \quad d_p^p(f,g) = \int_{\mathbf{R}^d} |f(t) - g(t)|^p \, dt.$$

THEOREM 2. *Let $2 \leq p < \infty$ and let* (CD) *and* (CD$p$) *hold. Assume*:

(a) *the kernel $K$ is in $L_p(\mathbf{R}^d)$ and*

$$(2.19) \quad \int_0^1 \Psi_{2/m}^{-1}\big(N\big(\mathcal{K}_{[a,b]}, d_p, \varepsilon\big)\big) \, d\varepsilon < \infty;$$

(b) *the sum of conditional densities $\overline{f}$ defined in* (2.1) *satisfies*

$$(2.20) \quad \lim_{t \to \infty} t^2 \Pr\{\|\overline{f}(\cdot, X) - E\overline{f}(\cdot, X)\|_p > t\} = 0$$

$$\text{and} \quad \int_{\mathbf{R}^d} \big[E\big(\overline{f}(u, X) - E\overline{f}(u, X)\big)^2\big]^{p/2} \, du < \infty;$$

(c) $h_n \to 0$ *and* $nh_n^{2(p-1)/p} \to \infty$.

*Then, the processes $u_{n,\lambda}$ converge weakly in $L_p(\mathbf{R}^d)$, uniformly in $a \leq \lambda \leq b$, to the centered Gaussian process with the same covariance as $\overline{f}(\cdot, X)$.*

The two conditions in (2.20) are implied by the stronger but sometimes more convenient conditions $E\|\overline{f}_i(\cdot, X) - E\overline{f}_i(\cdot, X)\|_p^2 < \infty$ for $i = 1, \ldots, m$. This follows immediately from Minkowski's inequality for integrals (e.g., Folland [14], page 194). Also, note that for $p = 2$ the first condition in (2.20) is superfluous.

We should remark here that the conditions (2.20) are precisely the necessary and sufficient conditions for the $L_p(\mathbf{R}^d)$-valued random variable $\overline{f}(\cdot, X)$ to satisfy the central limit theorem, that is, for the processes $\overline{v}_n$ to converge in law to $G$ in $L_p(\mathbf{R}^d)$, $p \geq 2$ (Pisier and Zinn [29]; see also Araujo and Giné [2], pages 206–207).

For $1 \leq p < 2$ we need $K^2$ and $f_g$ to satisfy certain moment assumptions, and for this it will be convenient to have the following notation: for $s > 0$, $p \geq 1$, we

define the Borel measure $\mu_s$ on $\mathbf{R}^d$, $L_p(\mu_s)$ and $\tilde{d}_{p,s}$ as

(2.21)
$$d\mu_s(t) = (1 + |t|)^s \, dt, \qquad L_p(\mu_s) = L_p(\mathbf{R}^d, \mathcal{B}, \mu_s),$$
$$\tilde{d}_{p,s}(f, g) = \|f - g\|_{L_p(\mu_s)},$$

the latter for functions $f, g \in L_p(\mu_s)$.

THEOREM 3. *Let* $1 \leq p < 2$ *and let* (CD) *and* (CD$p$) *hold. Assume* $K^2$ *and* $f_g$ *are in* $L_1(\mu_s)$ *for some* $s > d(2 - p)/p$. *Assume also*:

(a)  $K$ *is in* $L_p(\mathbf{R}^d)$ *and*

(2.22)
$$\int_0^1 \Psi_{1/m}^{-1}\big(N(\mathcal{K}_{[a,b]}, d_p \vee \tilde{d}_{2,s}, \varepsilon)\big) \, d\varepsilon < \infty;$$

(b)  *the sum* $\overline{f}$ *of conditional densities satisfies*

(2.23)
$$\int_{\mathbf{R}^d} \big[E\big(\overline{f}(t, X) - E\overline{f}(t, X)\big)^2\big]^{p/2} \, dt < \infty;$$

(c)  $h_n \to 0$ *and* $nh_n \to \infty$.

*Then, the processes* $u_{n,\lambda}$ *converge weakly in* $L_p(\mathbf{R}^d)$, *uniformly in* $a \leq \lambda \leq b$, *to the centered Gaussian process with the same covariance as* $\overline{f}(\cdot, X)$.

For $p = 1$, condition (2.23) is equivalent to $\int [E\overline{f}^2(t, X)]^{1/2} \, dt < \infty$.

As above, we should also mention that condition (2.23) is precisely the necessary and sufficient condition for the $L_p(\mathbf{R}^d)$-valued random variable $\overline{f}(\cdot, X)$ to satisfy the central limit theorem, that is, for the processes $\overline{v}_n$ to converge in law to $G$ in $L_p(\mathbf{R}^d)$, $1 \leq p \leq 2$ (Vakhania [36] and Jain [22]; see also Araujo and Giné [2], pages 206–207).

REMARK 3. All garden variety kernels satisfy the entropy assumptions of Theorems 2 and 3. In fact, many such kernels are of the form $K(x) = \Psi(|x|)$, where $\Psi$ is a function of bounded variation defined on $[0, \infty)$. To see this, assume that for bounded nondecreasing functions $M$ and $N$ we can write $\Psi = P - N$. Further, assume that

$$\int_0^\infty r^{d-1} |M(r)| \, dr = M_0 < \infty \quad \text{and} \quad \int_0^\infty r^{d-1} |N(r)| \, dr = N_0 < \infty.$$

Define the class of functions $\mathcal{P} = \{r \mapsto r^{d-1} \Psi(r\lambda^{1/d}) : \lambda \geq 1\}$. We claim that this class satisfies for some $A_1 > 0$,

(2.24)
$$N(\mathcal{P}, d_1, \varepsilon) \leq A_1 \varepsilon^{-1}, \qquad 0 < \varepsilon \leq 1.$$

To verify this choose $1 \leq \lambda \leq \mu$. We see that

$$\int_0^\infty |r^{d-1}\Psi(r\lambda^{1/d}) - r^{d-1}\Psi(r\mu^{1/d})|\,dr$$

$$\leq \int_0^\infty r^{d-1}[M(r\mu^{1/d}) - M(r\lambda^{1/d})]\,dr$$

$$+ \int_0^\infty r^{d-1}[N(r\mu^{1/d}) - N(r\lambda^{1/d})]\,dr$$

$$= (M_0 + N_0)(\lambda^{-1} - \mu^{-1}) =: C(\lambda^{-1} - \mu^{-1}).$$

Now as in the proof in Example 1.2 of Giné, Mason and Zaitsev [21] choose open balls with centers at $\lambda_k = C/(C - k\varepsilon)$ for $k = 0, \ldots, k_0$ where $k_0$ is the largest integer strictly less than $C/\varepsilon$. This shows (2.24).

Now consider the class of functions on $\mathbf{R}^d$ given by $\mathcal{K} = \{x \mapsto K(\lambda^{1/d}x): \lambda \geq 1\}$, where $K(\lambda^{1/d}x) = \Psi(\lambda^{1/d}|x|)$. Since by changing to polar coordinates

$$\int_{R^d} |K(\lambda^{1/d}x) - K(\mu^{1/d}x)|\,dx = C_d \int_0^\infty r^{d-1}|\Psi(r\lambda^{1/d}) - \Psi(r\mu^{1/d})|\,dr,$$

where $C_d = d\pi^{d/2}/\Gamma(1 + d/2)$, we see from (2.24) that for some $B_1 > 0$

$$(2.25) \qquad N(\mathcal{K}_\lambda, d_1, \varepsilon) \leq B_1\varepsilon^{-1}, \qquad 0 < \varepsilon \leq 1.$$

From this result along with boundedness of $K$ we readily get that for all $p \geq 1$ there is a $B_p > 0$ such that

$$(2.26) \qquad N(\mathcal{K}_\lambda, d_p, \varepsilon) \leq B_p\varepsilon^{-p}, \qquad 0 < \varepsilon \leq 1.$$

It then follows easily that the class of functions $\mathcal{K}_{[a,b]}$ formed from a kernel of the form $K(x) = \Psi(|x|)$ obeying the above conditions satisfies the entropy condition of Theorems 2 and, if $K$ has bounded support or decreases exponentially in a positive power of $|x|$, that of Theorem 3 as well. Some commonly used kernels defined on $\mathbf{R}$ of this form are (1) $K(u) = 1\{u \in [-1/2, 1/2]\}$, (2) $K(u) = \frac{1}{2}\exp(-|u|)$, (3) $K(u) = \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)$, and (4) $K(u) = \frac{3}{4}(1 - x^2)_+$. See, for instance, Devroye [8].

2.2. *Examples.* In this subsection we show how the previous theorems apply to a few instances of estimation of the density of a function of several variables considered in the literature, as well as to the interpoint distance.

EXAMPLE 1 (*Linear combinations*). (Frees [15], Schick and Wefelmeyer [32]) Suppose that

$$(2.27) \qquad g(x_1, \ldots, x_m) = \sum_{i=1}^m u_i(x_i), \qquad x_i \in S,$$

for measurable functions $u_1, \ldots, u_m$ from $S$ to $\mathbf{R}^d$ such that the random variable $u_i(X)$ has a density $f_i$ for each $i = 1, \ldots, m$. Then (CD) holds with $\overline{f}_i(t, x) = \tilde{f}_i(t - u_i(x))$ and $\tilde{f}_i = f_1 * \cdots * f_{i-1} * f_{i+1} * \cdots * f_m$. Thus,

$$\overline{f}(t, x) = \sum_{i=1}^m \tilde{f}_i(t - u_i(x)), \qquad t \in \mathbf{R}^d, x \in S.$$

The process $u_{n,\lambda}$ is then given by

(2.28)
$$u_{n,\lambda}(t) = \sqrt{n} \frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I_n^m} \left[ K_{\lambda h_n}\left(t - \sum_{r-1}^m u_r(X_{i_r})\right) \right. $$
$$\left. - E K_{\lambda h_n}\left(t - \sum_{r-1}^m u_r(X_r)\right) \right].$$

We will discuss some conditions under which the central limit theorems in $L_p(\mathbf{R}^d)$, $p \in [1, \infty]$, uniform in $\lambda$, given above apply in this situation.

Let $d = 1$. If $f$ is of bounded variation on $\mathbf{R}$ then so is the convolution of $f$ with any density, as is easy to check (in fact, more is true; see, e.g., Schick and Wefelemeyer [33], Lemma 1). Hence $\tilde{f}_j$ is of bounded variation for each $j$ if at least two of the densities $f_i$ are. So, assuming the densities $f_i$ are of bounded variation, then the classes $\mathcal{F}_i = \{\tilde{f}_i(t - \cdot) : t \in \mathbf{R}^d\}$ are of VC type by Lemma 22 of Nolan and Pollard [26], as mentioned in connection with (2.12). Also, since the map $(t, x) \mapsto \tilde{f}_i(t - x)$ is jointly measurable, these classes are $Q$-Donsker for every $Q$ (e.g., Dudley [10], Theorem 6.3.1, page 208). Hence, $\mathcal{F}_i$ is $P$-Donsker for each $i = 1, \ldots, m$. As observed, for example, in the proof of Lemma 8, Schick and Wefelmeyer [34], if $f$ is a function of bounded variation on $\mathbf{R}$, with $f(-\infty+) = 0$ and $T_f$ is the total variation of its right continuous modification $f(\cdot+)$, then by Fubini's theorem applied to the product of Lebesgue measure with the measure whose cumulative distribution function is $f(\cdot+)$, $\int |f(x + s) - f(x)| \, dx \leq C|s|$ for all $s \in \mathbf{R}$. With $f = \tilde{f}_i$ this gives

$$E\left(\overline{f}_i(t, X) - \overline{f}_i(s, X)\right)^2 = \int \left(\tilde{f}_i(t - u) - \tilde{f}_i(s - u)\right)^2 f_i(u) \, du$$

$$\leq 2T_{f_i} \|\tilde{f}_i\|_\infty \|\tilde{f}_i\|_1 |t - s|.$$

Hence, since this holds for every $i = 1, \ldots, m$, the identity map $(\mathbf{R}, |\cdot|) \mapsto (\mathbf{R}, \rho)$ is uniformly continuous. So, if the densities $f_i$ are of bounded variation on $\mathbf{R}$, then condition (b) in Theorem 1 is satisfied (note that $f_g$ is of bounded variation, hence bounded).

Let $d > 1$. Finding interesting conditions on the densities $f_i$ in order for condition (b) in Theorem 1 in the CLT to be satisfied is a little more cumbersome in this case. Here are some conditions:

(1) $f_i$ is $\alpha$-Hölder continuous for some $0 < \alpha \le 1$ and of bounded support, for each $1 \le i \le m$. In this case one can easily check the VC property.

(2) $f_i$ is of bounded variation on $\mathbf{R}^d$ in the sense that it is the difference of the ($d$-dimensional) distribution functions of two positive measures (necessarily of the same mass). In this case the Donsker property follows from Dudley [10], Corollary 10.2.8, page 327.

(3) The functions $\tilde{f}_i$ satisfy condition (2.12). This condition is directly imposed on convolutions because (2.12) may not be inherited by convolution, except in particular cases. Some important examples, like full $d$-variate normal densities, satisfy it.

As a consequence of the above discussion, we have proved the following theorem that improves on Theorem 1 in Schick and Wefelmeyer [32] in that the convergence is uniform in $\lambda \in [a, b]$, the kernel is not necessarily a convolution and the window sizes $h_n$ are allowed to decrease at a smaller rate. Their result includes the bias part whereas ours does not, and we discuss this immediately below.

THEOREM 4. *Let $g$ be defined by* (2.27) *and let $K$ satisfy condition* (2.12). *If $d = 1$ assume that the densities $f_i$ of $u_i(X)$ are of bounded variation and if $d > 1$ assume that $f_i$ satisfy* (1) *or* (2) *or that $\tilde{f}_i$ satisfy* (3) *in the previous paragraph. Let $h_n \to 0$ and $nh_n/(\log h_n^{-1})^2 \to \infty$. Then the processes $u_{n,\lambda}$ defined by* (2.28) *converge weakly in $L_\infty(\mathbf{R}^d)$ uniformly in $a \le \lambda \le b$ to the centered Gaussian process with the same covariance as $\overline{f}(\cdot, x) = \sum_{i=1}^m \tilde{f}_i(\cdot - u_i(x))$.*

Next we comment on the bias part in this theorem. What we do is standard and can also be done for the rest of the results in this article, but we will refrain from doing so. Suppose that the densities $f_i$ are in $C^k(\mathbf{R}^d)$ and that their partial derivatives of order $k$ or smaller are all bounded. Then the same is true for $f_g = f_1 * \cdots * f_m$ (this follows, e.g., from Proposition 8.10 in Folland [14] and Young's inequalities) and therefore $f_g$ admits a Taylor development of the form

$$
f_g(t + \delta) - f_g(t)
$$

$$
= \sum_{r=1}^k \frac{1}{r!} \sum_{\substack{s_1 + \cdots + s_d = r \\ 0 \le s_i \le r}} \frac{\partial^r f_g}{\partial^{s_1} x_1 \cdots \partial^{s_d} x_d}(t) \delta_1^{s_1} \cdots \delta_d^{s_d} + H(t, k, \delta)|\delta|^k,
$$

where $H$ is uniformly bounded by a constant times the common bound for the $k$th partial derivatives of $f_g$ (actually, by continuity, $H \to 0$ as $\delta \to 0$). Suppose moreover that the kernel $K$ satisfies the conditions

$$
\int |t|^k |K(t)| \, dt < \infty \quad \text{and} \quad \int t_1^{s_1} \cdots t_d^{s_d} K(t) \, dt = 0 \qquad \text{for } s_i \ge 0, \sum_{i=1}^d s_i \le k.
$$

Then, integration after change of variables and use of Taylor's formula for $f_g(t + (\lambda h_n)^{1/d} u) - f_g(t)$ give

$$\sup_{\lambda \in [a,b]} \sqrt{n} \big| E K_{\lambda h_n}(t - g(X_1, \ldots, X_m)) - f_g(t) \big| = O\big(\sqrt{n} h_n^{k/d}\big).$$

Therefore, under these extra conditions on $f_i$ and $K$, if $n h_n^{2k/d} \to 0$, the conclusion of the previous theorem can be modified to convergence in law in $\ell^\infty(\mathbf{R}^d \times [a, b])$ of

$$\sqrt{n} \left\{ \frac{1}{|I_n^m|} \sum_{I_n^m} K_{\lambda h_n}\left(t - \sum_{r=1}^{m} u_r(X_{i_r})\right) - f_g(t) \right\}$$

to the same centered Gaussian process. If the partial derivatives up to order $k$ of $f_i$ are bounded and in $L_1$, then $f_g \in C^{mk}(\mathbf{R}^d)$ (e.g., by iteration in Proposition 8.10, Folland [14]), and the previous discussion applies with $k$ replaced by $mk$.

*Simultaneous estimation of convolutions.* Here is an immediate application of Theorem 4 to the simultaneous estimation of the densities of convolutions. Let $f$ be a density of bounded variation on $\mathbf{R}$. We are interested in estimating the convolutions $f^{*2} = f * f, \ldots, f^{*N}$, for $N \geq 2$. For $m = 2, \ldots, N$, $t_m \in \mathbf{R}$ and $a_m \leq \lambda_m \leq b_m$ introduce the estimators

$$\widehat{f_n^{*m}}(t_m, \lambda_m) = \frac{1}{|I_n^m| \lambda_m h_n} \sum_{I_n^m} K\left(\frac{t_m - \sum_{r=1}^{m} X_{i_r}}{\lambda_m h_n}\right)$$

and set

$$u_{n,\lambda_m}^{(m)}(t_m) = \sqrt{n}\{\widehat{f_n^{*m}}(t_m, \lambda_m) - E\widehat{f_n^{*m}}(t_m, \lambda_m)\}$$

or, under extra conditions as in the previous remark, replace $E\widehat{f_n^{*m}}(t_m, \lambda_m)$ by $f^{*m}(t_m, \lambda_m)$. [These estimators correspond to the functions $g = g_m$ defined by $g_m(x_1, \ldots, x_m) = x_1 + \cdots + x_m$.] A direct application of Theorem 4 in combination with the observations (2.6)–(2.8) gives that the vector-valued processes $(u_{n,\lambda_2}^{(2)}(t_2), \ldots, u_{n,\lambda_N}^{(N)}(t_N))$ converge weakly in $L_\infty(\mathbf{R}) \times \cdots \times L_\infty(\mathbf{R})$ uniformly in $a_m \leq \lambda_m \leq b_m$, $m = 2, \ldots, N$, to the centered vector-valued Gaussian process defined on $\mathbf{R} \times \cdots \times \mathbf{R}$, with the same covariance/cross covariance matrix as

$$\big(2f(t_2 - X), \ldots, N f^{*(N-1)}(t_N - X)\big).$$

We note here that Schick and Wefelmeyer [32] use a variation of the Frees [15] local $U$-statistic estimator of convolutions of densities. Their estimator is based on convolving kernel density estimators. Here is how their approach works in the case of estimating the density of $X_1 + X_2$, where $X_1$ and $X_2$ are i.i.d. real valued with density $f$. Consider the kernel density estimator

$$\widehat{f_n}(x) = (n h_n)^{-1} \sum_{i=1}^{n} k\big(h_n^{-1}(x - X_i)\big),$$

where $k$ is of bounded variation on $\mathbf{R}$. Their estimator is $\widehat{f}_n * \widehat{f}_n$ and can be expressed, with $K = k * k$, as

$$\widehat{f}_n * \widehat{f}_n(x) = \frac{1}{n^2 h_n} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{x - X_i - X_j}{h_n}\right)$$

$$= \frac{n-1}{n} \overline{U}_{n,h_n}(x) + \frac{1}{n^2 h_n} \sum_{i=1}^{n} K\left(\frac{x - 2X_i}{h_n}\right),$$

where $\overline{U}_{n,h_n}(x)$ is the Frees type local *U*-statistic estimator of $f * f(x)$ defined by

$$\overline{U}_{n,h_n}(x) = (h_n n(n-1))^{-1} \sum_{1 \le i \ne j \le n} K(h_n^{-1}(x - X_i - X_j)).$$

The second term in the above expression for $\widehat{f}_n * \widehat{f}_n(x)$ is asymptotically negligible and Theorem 4 applies to $\overline{U}_{nh_n}$. This remark applies as well to simultaneous estimation of convolutions of densities.

We can complement the above results for $L_\infty(\mathbf{R}^d)$ with limit theorems for the $L_p$ distance. We now do this for the cases $p = 1$ and $p = 2$.

THEOREM 5. *Assume $K$ satisfies condition* (a) *in Theorem* 2 *and let $g(x_1, \ldots, x_m) = \sum_{i=1}^{m} u_i(x_i)$, $u_i : S \mapsto \mathbf{R}^d$ be measurable, where $u_i(X)$ has density $f_i$. Assume $f_i$ is in $L_2(\mathbf{R}^d)$. Assume also $h_n \to 0$ and $nh_n \to \infty$. Then*

$$\frac{\sqrt{n}}{|I_n^m|} \sum_{I_n^m} \left\{ K_{\lambda h_n}\left(t - \sum_{r=1}^{m} u_r(X_{i_r})\right) - E K_{\lambda h_n}\left(t - \sum_{r=1}^{m} u_r(X_r)\right) \right\}$$

*converges in law in $L_2(\mathbf{R}^d)$ uniformly in $\lambda \in [a, b]$ to the centered Gaussian process with the same covariance as $\sum_{i=1}^{m} \tilde{f}_i(\cdot - u_i(X_i))$.*

This theorem follows because by Young's inequality (e.g., Folland [14], page 240), $f_i \in L_2$ for $i = 1, \ldots, m$ implies $E \int \overline{f}(t, X)^2 \, dt < \infty$. As in the previous theorem, a little more smoothness on $f_i$ and higher-order kernels allow for elimination of the bias.

We can also recover the Schick and Wefelmeyer [32], Theorem 2 on the CLT for the $L_1$ norm, with weaker assumptions (note that condition (2.22) is vacuous if $[a, b]$ reduces to a single point).

If two densities belong to $L_1(\mu_s)$ for some $s > 0$, then so does their convolution as can be seen by direct computation using the trivial observation that $(1 + |u + v|) \le (1 + |u|)(1 + |v|)$, and it is also routine to check that if any finite number of densities and their squares belong to $L_1(\mu_s)$, so does the square of their convolution. Moreover, by Lemma 1 to be proved below, if $f$ and $k$ are two nonnegative functions in $L_1(\mu_s)$ for some $s > d$, then $\int (f * k)^{1/2}(t) \, dt < \infty$. Hence Theorem 3 gives:

THEOREM 6. *Let* $g(x_1, \ldots, x_m) = \sum_{i=1}^m u_i(x_i)$, $u_i : S \mapsto \mathbf{R}^d$ *be measurable, with* $u_i(X)$ *having density* $f_i$, *for* $i = 1, \ldots, m$, *and let* $K$ *be a kernel on* $\mathbf{R}^d$. *Assume that for some* $s > d$, $K^2$, $f_i$ *and* $f_i^2$ *are in* $L_1(\mu_s)$, $i = 1, \ldots, m$, *and that* $K$ *satisfies condition* (a) *in Theorem* 3 *for this* $s$. *Assume also* $h_n \to 0$ *and* $nh_n \to \infty$. *Then*

$$\frac{\sqrt{n}}{|I_n^m|} \sum_{I_n^m} \left\{ K_{\lambda h_n}\left( t - \sum_{r=1}^m u_r(X_{i_r}) \right) - E K_{\lambda h_n}\left( t - \sum_{r=1}^m u_r(X_r) \right) \right\}$$

*converges in law in* $L_1(\mathbf{R}^d)$ *uniformly in* $\lambda \in [a, b]$) *to the centered Gaussian process with the same covariance as* $\sum_{i=1}^m \tilde{f}_i(\cdot - u_i(X_i))$.

In $\mathbf{R}$, if the densities $f_i$ are bounded (they do not need to be) then the condition imposed on $f_i$ is simply that $\int |x|^{1+\delta} f(x) \, dx < \infty$ for some $\delta > 0$.

EXAMPLE 2 (*Distribution of sample distances*). Frees [15] considers estimating the density of the interpoint functional $g(X_1, X_2) = |X_1 - X_2|$ in two dimensions, where $X_i$ are i.i.d. with a density $f$ which is bounded and of bounded support. We further assume that $f$ is $\alpha$-Hölder continuous for some $0 < \alpha \le 1$, that is, $|f(u) - f(v)| \le C|u - v|^\alpha$ for all $u, v \in \mathbf{R}^2$. Then, it is easy to see that $f_g$, the density of $g$,

$$f_g(t) = \int_0^{2\pi} \int_{\mathbf{R}^2} f(x_1 + t\cos\theta, x_2 + t\sin\theta) f(x_1, x_2) t \, dx_1 \, dx_2 \, d\theta,$$

is bounded and has bounded support. Moreover, for $x = (x_1, x_2)$ and with $R$ the radius of a ball around zero containing the support of $f$, we have, for the conditional densities $\overline{f}_i$ of $|X_1 - X_2|$,

$$\overline{f}_1(t, x) = \overline{f}_2(t, x) = \int_0^{2\pi} f(x_1 + t\cos\theta, x_2 + t\sin\theta) t \, d\theta, \qquad 0 \le t \le 2R,$$

and $\overline{f}_i(t, x) = 0$ for larger values of $t$. Then $\overline{f} = \overline{f}_1 + \overline{f}_2$ satisfies

$$|\overline{f}(t, \cdot) - \overline{f}_i(s, \cdot)| \le 8\pi RC|t - s|^\alpha + 4\pi \|f\|_\infty |t - s|,$$

and therefore, with $\mathcal{F} = \{\overline{f}(t, \cdot) : |t| \le 2R\}$ and any probability measure $Q$, $N(\mathcal{F}, L_2(Q), \varepsilon) \le C/\varepsilon^{1/\alpha}$ which, since the class $\mathcal{F}$ is image admissible Suslin by joint measurability of $\overline{f}$, implies that the class $\mathcal{F}$ is $P$-Donsker. So, if we take a kernel $K$ satisfying (2.12), and $h_n \to 0$ with $nh_n/(\log h_n^{-1})^2 \to \infty$, we get that the processes defined for $|t| \le R$, $\lambda \in [a, b]$,

$$\frac{n^{1/2}}{n(n-1)} \sum_{1 \le i \ne j \le n} \{ K_{\lambda h_n}(t - |X_i - X_j|) - E K_{\lambda h_n}(t - |X_i - X_j|) \},$$

converge in law uniformly in $t$ and $\lambda$ to the centered Gaussian process with the same covariance as $2f_1(t, X)$. Moreover, the comments following Theorem 4 regarding replacement of $(\lambda h_n)^{-1} E K((\lambda h_n)^{-1}(t - |X_i - X_j|))$ by $f_g(t)$ apply here as well.

In dimension 1,

$$f_1(t, x) = f_2(t, x) = f(t + x) + f(-t + x), \qquad t \geq 0,$$

and a sufficient condition for $\mathcal{F}$ to be $P$-Donsker is that $f$ be of bounded variation on **R**. Then, $K$ of bounded variation on **R** and $nh_n/(\log h_n^{-1})^2 \to \infty$ ensure the same CLT as above. Since

$$f_g(x) = \int f(y + x) f(y)\, dy + \int f(y - x) f(y)\, dy$$

and

$$f_g(0) = 2 \int f^2(y)\, dy,$$

as observed by Frees [15], we get a $\sqrt{n}$ consistent estimator of $\int f^2(x)\, dx$ when the extra smoothness conditions to make the bias tend to zero hold. See also Bickel and Ritov [4] or Giné and Mason [19] and references therein for other $\sqrt{n}$ consistent estimators of $\int f^2(x)\, dx$.

EXAMPLE 3 (*Local interpoint distance processes*).    (Jammalamadaka and Janson [23], Eastwood and Horváth [12].) In Corollary 1 we consider the processes

$$\frac{\sqrt{n}}{h_n \operatorname{Vol}(D)} \sum_{I_n^m} [I\{g(X_{i_1}, \ldots, X_{i_m}) \in (\lambda h_n)^{1/d} D\}$$

$$- \Pr\{g(X_1, \ldots, X_m) \in (\lambda h_n)^{1/d} D\}]$$

for $D$ star shaped about 0. Of particular interest in the literature is the case corresponding to $m = 2$, $S = \mathbf{R}^d$, $g(X_1, X_2) = X_2 - X_1$ and $D$ the (open or closed) unit ball about zero for some norm in $\mathbf{R}^d$. In this case the densities $\overline{f}_i(t, x)$, $i = 1, 2$, are, respectively, $f(x - t)$ and $f(x + t)$, where $f$ is the density of $X$. So, for the local asymptotic equicontinuity condition (2.13) to hold we only need that the class $\mathcal{F} = \{f(\cdot + t) : |t| < \delta\}$ be $P$-Donsker for some $\delta > 0$. If $f$ is Hölder continuous of order $\alpha \in (0, 1]$, and $Q$ is any probability measure on $\mathbf{R}^d$, then

$$\int \big(f(x + t) - f(x + s)\big)^2\, dQ(x) \leq C|t - s|^{2\alpha}, \qquad s, t \in \mathbf{R}^d.$$

It follows as in the previous example that $\mathcal{F}$ is VC and measurable, hence $Q$-Donsker for every probability measure $Q$. Thus, Corollary 1 implies the following slight strengthening and generalization of Theorem 1.1 in Eastwood and Horváth [12].

THEOREM 7.   *If $D$ is a bounded measurable subset of $\mathbf{R}^d$ star-shaped about zero and with $\mathrm{Vol}(D) \neq 0$, $X_i$ are i.i.d. random vectors in $\mathbf{R}^d$ with a density $f$ which is $\alpha$-Hölder continuous for some $\alpha \in (0, 1]$, and $h_n \to 0$, $nh_n \to \infty$, then the processes*

$$\frac{1}{n^{3/2}h_n \mathrm{Vol}(D)} \sum_{(i_1, i_2) \in I_n^2} \left[ I\left\{ \frac{X_{i_1} - X_{i_2}}{(\lambda h_n)^{1/d}} \in D \right\} - \Pr\left\{ \frac{X_1 - X_2}{(\lambda h_n)^{1/d}} \in D \right\} \right],$$

$$0 \leq \lambda \leq 1,$$

*converge in law in $\ell^\infty[0, 1]$ to the process $\lambda \sigma Z$, $0 \leq \lambda \leq 1$, where $Z$ is $N(0, 1)$ and $\sigma^2 = 4[\int f^3(x)\, dx - (\int f^2(x)\, dx)^2]$.*

**3. Proofs.**   In the sequel it will be helpful to introduce the following notation and facts. For a kernel $L$ of $k \geq 1$ variables we set

$$(3.1) \qquad\qquad U_n^{(k)}(L) = \frac{(n-k)!}{n!} \sum_{\mathbf{i} \in I_n^k} L(X_{i_1}, \ldots, X_{i_k})$$

[so, $U_n(t, \lambda) = U_n^{(m)}(K_{\lambda h_n}(t - g(\cdot, \ldots, \cdot)))$ even if $g$ is not symmetric in its entries]. [When $L$ is a constant function we define $U_n^{(0)}(L) = L$.] Assume now that $L$ is a function of $m \geq 1$ variables, symmetric in its entries. Then, for $1 \leq k \leq m$, the Hoeffding projections with respect to $P$ are defined as

$$(3.2) \qquad \pi_k L(x_1, \ldots, x_k) = (\delta_{x_1} - P) \times \cdots \times (\delta_{x_k} - P) \times P^{m-k}(L)$$

and $\pi_0 L = EL(X_1, \ldots, X_m)$. Then, the *Hoeffding decomposition* states the following, which is easy to check:

$$(3.3) \qquad\qquad U_n^{(m)}(L) - EL = \sum_{k=1}^{m} \binom{m}{k} U_n^{(k)}(\pi_k L).$$

For $L \in L_2(P^m)$ this is an orthogonal decomposition and $E(\pi_k L | X_2, \ldots, X_k.) = 0$ for $k \geq 1$; that is, the kernels $\pi_k L$ are canonical for $P$ (or completely degenerate, or completely centered). Also, $\pi_k$, $k \geq 1$, are nested projections, that is, $\pi_k \circ \pi_\ell = \pi_k$ if $k \leq \ell$, and $E(\pi_k L)^2 \leq E(L - EL)^2 \leq EL^2$.

The function $K_h(t - g(X_1, \ldots, X_m))$ is not necessarily symmetric in its entries, but we can symmetrize it as

$$(3.4) \qquad \overline{K}_h(t, x_1, \ldots, x_m) := \frac{1}{m!} \sum_{\sigma \in I_m^m} K_h(t - g(x_{\sigma_1}, \ldots, x_{\sigma_m})).$$

Then, clearly, for each $t \in \mathbf{R}^d$,

$$U_n(t, \lambda) - EK_{\lambda h_n}(t - g(X_1, \ldots, X_m))$$
$$= U_n^{(m)}(\overline{K}_{\lambda h_n}(t, \cdot, \ldots, \cdot)) - E\overline{K}_{\lambda h_n}(t, X_1, \ldots, X_m).$$

Moreover, by applying (3.3) to $u_{n,\lambda}(t)$ we get

$$(3.5) \qquad u_{n,\lambda}(t) = \sqrt{n} \sum_{k=1}^{m} \binom{m}{k} U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n}(t, \cdot, \ldots, \cdot)).$$

3.1. *A general proposition for the CLT.* Let us consider the first term in the expansion (3.5). Note that, by definition of $\overline{f}$ and $f_g$,

$$\begin{aligned}
m\pi_1 \overline{K}_{\lambda h_n}(t, x) &= \sum_{i=1}^{m} EK_{\lambda h_n}\big(t - g(X_1, \ldots, X_{i-1}, x, X_{i+1}, \ldots, X_m)\big) \\
&\quad - m EK_{\lambda h_n}\big(t - g(X_1, \ldots, X_m)\big) \\
&= \int K_{\lambda h_n}(t - u)\big(\overline{f}(u, x) - E\overline{f}(u, X)\big) \, du \\
&= \int \big(\overline{f}(t - u, x) - E\overline{f}(t - u, X)\big) K_{\lambda h_n}(u) \, du.
\end{aligned}$$

Hence,

$$\begin{aligned}
(3.6) \qquad & \sqrt{n} m U_n^{(1)}(\pi_1 \overline{K}_{\lambda h_n}(t, \cdot)) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \big(\overline{f}(t - u, X_i) - E\overline{f}(t - u, X)\big) K_{\lambda h_n}(u) \, du \\
&= \int \overline{v}_n(t - u) K_{\lambda h_n}(u) \, du = (\overline{v}_n * K_{\lambda h_n})(t),
\end{aligned}$$

which is a generalized version of the smoothed empirical process that has been recently investigated by several authors (e.g., Rost [31] and references therein). We shall see that it controls the asymptotic behavior of the process $u_{n,\lambda}$. In the next proposition we show it has the same asymptotic behavior as the empirical process over the class of functions $\{\overline{f}(t, \cdot), t \in \mathbf{R}^d\}$.

PROPOSITION 1. *Let* $1 \le p \le \infty$. *Assume*:

(i) *condition* (CD $p$) *holds*;
(ii) $\lim_{\delta \to 0} \limsup_{n \to \infty} \mathrm{Pr}^* \{\sup_{|u| \le \delta} \|\overline{v}_n(\cdot - u) - \overline{v}_n(\cdot)\|_p > \varepsilon\} = 0$ *for all* $\varepsilon > 0$;
(iii) *the sequence* $\|\overline{v}_n\|_p^*$, $n \in \mathbf{N}$, *is stochastically bounded*.

*Then, whenever* $h_n \to 0$ *we have*

$$(3.7) \qquad \lim_{n \to \infty} \sup_{a \le \lambda \le b} \big\| \sqrt{n} m U_n^{(1)}(\pi_1 \overline{K}_{\lambda h_n}(\cdot, \cdot)) - \overline{v}_n(\cdot) \big\|_p = 0 \qquad in \ \mathrm{pr}^*.$$

PROOF. Let $w_\delta(\overline{v}_n) = \sup_{|u| < \delta} \|\overline{v}_n(\cdot - u) - \overline{v}(\cdot)\|_p$, $\delta > 0$, denote the $L_p(\mathbf{R}^d)$ modulus of "continuity" of $\overline{v}_n$, which is defined because of (i). Then it follows

by Fubini in the case $p = \infty$ and by Minkowski's inequality for integrals (e.g., Folland [14], page 194) in the case $1 \le p < \infty$ that

$$\|\overline{\nu}_n * K_{\lambda h_n} - \overline{\nu}_n\|_p \le w_\delta(\overline{\nu}_n)\|K_{\lambda h_n}\|_1 + 2\|\overline{\nu}_n\|_p \int_{|u|>\delta} |K_{\lambda h_n}(u)|\, du$$

$$\le w_\delta(\overline{\nu}_n)\|K\|_1 + 2\|\overline{\nu}_n\|_p \int_{|u|>\delta/(\lambda h_n)^{1/d}} |K(u)|\, du.$$

Now, the result follows from this, $K \in L_1(\mathbf{R}^d)$ and (ii) and (iii), in view of (3.6).
□

COROLLARY 2. *Let $1 \le p \le \infty$. Assume* (CD$p$) *and hypothesis* (b) *in Theorem* 1 *for $p = \infty$, in Theorem* 2 *for $2 \le p < \infty$ and in Theorem* 3 *for $1 \le p < 2$. Then the processes $\overline{\nu}_n$ converge weakly in $L_p(\mathbf{R}^d)$ to the centered Gaussian process $G$ with the covariance of $\overline{f}(\cdot, X)$. If moreover $h_n \to 0$, then the limit* (3.7) *holds.*

PROOF. (a) Case $p = \infty$. In this case, the hypothesis of the class $\mathcal{F}$ being $P$-Donsker is just another way of saying that $\overline{\nu}_n$ converges in law to $G$ in $L_\infty(\mathbf{R}^d)$, and this obviously implies (iii); finally, (ii) holds by the uniform continuity of the identity map $(\mathbf{R}^d, |\cdot|) \mapsto (\mathbf{R}^d, \rho)$ together with the usual asymptotic equicontinuity condition (e.g., Dudley [10], pages 117–118).

(b) Case $1 \le p < \infty$. As mentioned in Section 2, the hypotheses (b) in Theorems 2 and 3 are precisely the necessary and sufficient conditions for the process $\overline{f}(\cdot, X)$ to satisfy the CLT in $L_p$, that is, for $\overline{\nu}_n$ to converge in law to $G$ in $L_p$. Moreover, (iii) is a direct consequence of this convergence. Finally, condition (ii) holds by the uniform tightness implied by weak convergence together with the Fréchet–Kolmogorov characterization of compact sets of $L_p(\mathbf{R}^d)$ (see, e.g., Dunford and Schwartz [11], Theorem IV.8.21, page 301). □

In view of (3.5) and Corollary 2, to complete our CLT program for $u_{n,\lambda}$, that is, the proofs of Theorems 1, 2 and 3, it only remains to show that the hypotheses in the statements of these theorems also imply

$$(3.8) \qquad \sup_{a \le \lambda \le b} \left\|\sqrt{n} U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n})\right\|_p \to 0 \qquad \text{in pr,} \quad k = 2, \dots, m.$$

Proving this constitutes the main part of our proofs, and requires new inequalities for $U$-processes that we develop in the next subsections.

3.2. *Inequalities for $U$-processes.* In the next two subsections we collect the inequalities we need to prove the limits (3.8). First we consider the case of $U$-processes indexed by VC classes of functions and obtain a moment inequality that generalizes the scope of that of Einmahl and Mason [13] and Giné and Koltchinskii [17] for empirical processes. Next, we consider $U$-statistics taking

values on separable type 2 Banach spaces such as $L_p$, $p \geq 2$, and derive exponential inequalities for them. The inequality that we get in this situation is particularly neat. Then, based on the method recently used by Giné, Latała and Zinn [18] to prove inequalities for *U*-statistics, we derive both moment and exponential inequalities for other Banach spaces, such as $L_p$, $1 \leq p < 2$. These are less clean than in the type 2 case, but are still usable. Neither of our exponential inequalities captures the Gaussian tail behavior that the statistic should have for small values of $x$; nevertheless, their application yields very strong results in the situations encountered in this article.

3.3. *U-processes indexed by VC classes.* In this subsection we consider classes of measurable functions $\mathcal{F}$ defined on $(S^m, \mathcal{S}^m)$ taking values in $[-1, 1]$, and we assume that $0 \in \mathcal{F}$. The object is to obtain a bound for $E\|U_n^{(k)}(\pi_k f)\|_{\mathcal{F}}$ where $\mathcal{F}$ is of VC-type, and where we use the notation $\|\Psi(f)\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\Psi(f)|$ for any functional $\Psi$ defined on the class $\mathcal{F}$. This bound will require measurability on the class $\mathcal{F}$ described in de la Peña and Giné [7], page 138: the class $\mathcal{F}$ should be measurable in the sense that for every $k = 1, \ldots, m$ and every choice of $a_{i_1,\ldots,i_k} \in \{-1, 1\}$, the mapping $f \mapsto \sum_{I_n^k} a_{i_1,\ldots,i_k} P^{m-k} f(X_{i_1}, \ldots, X_{i_k})$ is measurable for the completion of $\mathcal{S}^n$. This holds, for instance, if (a) if there exists $\mathcal{F}_0$ countable such that this sup equals the sup over $\mathcal{F}_0$, or (b) if the $\sigma$-algebra $\mathcal{S}$ is countably generated and contains the singletons, and the class $\mathcal{F}$ is image admissible Suslin, for instance, if it is parametrized by a complete separable metric space $T$ in such a way that the evaluation map $(t, x_1, \ldots, x_m) \mapsto f_t(x_1, \ldots, x_m)$ is jointly measurable (Dudley [10], Section 5.3; van der Vaart and Wellner [37], Section 2.3.1; Pollard [30], page 196). These conditions allow us to randomize by independent random signs and use Fubini. If either of these two conditions is satisfied, we say that the class $\mathcal{F}$ is measurable.

The following moment inequality will be instrumental in finishing the proof of Theorem 1. The proof of this inequality has several points in common with the proofs of similar inequalities for $m = 1$ in Einmahl and Mason [13] and in Giné and Koltchinskii [17]; however the present proof does not rely on the square root trick or on the contraction principle for Rademacher processes.

THEOREM 8. *Let $\mathcal{F}$ be a measurable collection of functions $S^m \mapsto \mathbf{R}$ symmetric in their entries with an envelope function $F$ and let $P$ be any probability measure on $(S, \mathcal{S})$ (with $X_i$ i.i.d. $P$). Assume $F$ is bounded by $M > 0$ and $\mathcal{F}$ is VC with respect to $F$ with characteristics $A$ and $v$, as in (2.9). Then for every $m \in \mathbf{N}$, $A \geq e^m$, $v \geq 1$, there exist constants $C_1 := C_1(m, A, v, M)$ and $C_2 = C_2(m, A, v, M)$ such that*

$$n^{k/2} E\|U_n^{(k)}(\pi_k f)\|_{\mathcal{F}} \leq C_1 \sigma \left( \log \frac{A\|F\|_{L_2(P^m)}}{\sigma} \right)^{k/2},$$

(3.9)

$$k = 0, 1, \ldots, m,$$

*assuming*

$$(3.10) \qquad n\sigma^2 \geq C_2 \log\left(\frac{2\|F\|_{L_2(P^m)}}{\sigma}\right),$$

*where $\sigma^2$ is any number satisfying*

$$(3.11) \qquad \|P^m f^2\|_{\mathcal{F}} \leq \sigma^2 \leq P^m F^2.$$

PROOF.    Without loss of generality we assume $F \leq M = 1$. The theorem is true for all $m$ and $k = 0$ by Hölder's inequality, since $U_n^{(0)}(\pi_0 f) = Pf$, so we assume the statement to be true for all $m \in \mathbf{N}$ and $k - 1$, for some $k \geq 1$, and prove it for $k$ (and for all $m \in \mathbf{N}$ ). We shall omit symbols when no confusion is possible, so, for instance, we write $\|\cdot\|$ for $\|\cdot\|_{\mathcal{F}}$. We shall not keep track of constants; in particular, constants that depend on a subset of $m, A, \upsilon, M$, will generically be denoted by $C$ (so, the value of $C$ may change from line to line). By Theorem 3.5.3 in de la Peña and Giné [7], the definition of $\pi_k$ and Jensen's inequality,

$$E\left\|\sum_{I_n^k}(\pi_k f)(X_{i_1}, \ldots, X_{i_k})\right\| \leq CE\left\|\sum_{I_n^k}\varepsilon_{i_1}\cdots\varepsilon_{i_k}(\pi_k f)(X_{i_1}, \ldots, X_{i_k})\right\|$$

$$\leq CE\left\|\sum_{I_n^k}\varepsilon_{i_1}\cdots\varepsilon_{i_k}(P^{m-k} f)(X_{i_1}, \ldots, X_{i_k})\right\|.$$

Here $X_i, \varepsilon_j$ are all independent, the $X$'s have law $P$ and the $\varepsilon$'s are random signs (Rademacher variables). Since for any probability measure $Q$ on $S^k$,

$$Q\left(P^{m-k}(f - g)\right)^2 \leq Q \times P^{m-k}(f - g)^2,$$

it follows from the VC property of $\mathcal{F}$ that

$$N(P^{m-k}\mathcal{F}, L_2(Q), \tau) \leq \left(\frac{A\|\sqrt{P^{m-k} F^2}\|_{L_2(Q)}}{\tau}\right)^{\upsilon},$$

$$0 < \tau \leq 2\|\sqrt{P^{m-k} F^2}\|_{L_2(Q)},$$

that is, $P^{m-k}\mathcal{F}$ is VC-type with characteristics $A$ and $\upsilon$ and envelope $\sqrt{P^{m-k} F^2}$. This gives, by the entropy integral for Rademacher chaos of order $k$ (de la Peña and Giné [7], Corollary 5.1.8, upon noting that exponential Orlicz norms dominate $L_p$ norms up to constants),

$$|I_n^k|^{-1/2} E_\varepsilon \left\|\sum_{I_n^k}\varepsilon_{i_1}\cdots\varepsilon_{i_k}(P^{m-k} f)(X_{i_1}, \ldots, X_{i_k})\right\|$$

$$\leq C \int_0^{\sqrt{\|U_n^{(k)}((P^{m-k} f)^2)\|}} \left(\log \frac{A\sqrt{U_n^{(k)}(P^{m-k} F^2)}}{\tau}\right)^{k/2} d\tau,$$

where $E_\varepsilon$ denotes expectation with respect to the Rademacher variables only. [To apply Corollary 5.1.8 exactly to our process we need that, for $X_1, \ldots, X_n$ fixed, the Rademacher chaos process

$$f \mapsto \sum_{I_n^k} \varepsilon_{i_1} \cdots \varepsilon_{i_k} (P^{m-k} f)(X_{i_1}, \ldots, X_{i_k}), \qquad f \in \mathcal{F},$$

be separable, and this follows by separability of the unit cube of $\mathbf{R}^{|I_n^k|}$ for the Euclidean norm.] Then, by Fubini,

$$(3.12) \qquad |I_n^k|^{1/2} E \|U_n^{(k)}(\pi_k f)\| \le CB,$$

where

$$(3.13) \qquad B = E \int_0^{\sqrt{\|U_n^{(k)}((P^{m-k} f)^2)\|}} \left( \log \frac{A\sqrt{U_n^{(k)}(P^{m-k} F^2)}}{\tau} \right)^{k/2} d\tau.$$

Decompose the integral $B$ into two parts,

$$(I) := E\left[ \int_0^{\sqrt{\|U_n^{(k)}((P^{m-k} f)^2)\|}} \left( \log \frac{A\sqrt{U_n^{(k)}(P^{m-k} F^2)}}{\tau} \right)^{k/2} d\tau \, I_n \right],$$

where

$$I_n = I\{ U_n^{(k)}(P^{m-k} F^2) > 4 P^m F^2 \}$$

and

$$(II) := E\left[ \int_0^{\sqrt{\|U_n^{(k)}((P^{m-k} f)^2)\|}} \left( \log \frac{A\sqrt{U_n^{(k)}(P^{m-k} F^2)}}{\tau} \right)^{k/2} d\tau \, I_n^c \right],$$

so that $B = (I) + (II)$. The $(I)$ term is handled by the Arcones [3] exponential inequality (de la Peña and Giné [7], Theorem 4.1.13), which gives

$$(3.14) \quad \begin{aligned} &\Pr\{ U_n^{(k)}(P^{m-k} F^2) > 4 P^m F^2 \} \\ &\le 4 \exp\left\{ -\frac{9n(P^m F^2)^2}{2k^2 P^m F^4 + c P^m F^2} \right\} \le 4 \exp\left\{ -\frac{9n P^m F^2}{2k^2 + c} \right\} \end{aligned}$$

for a constant $c$ that depends only on $k$. In the last inequality we have used $F \le 1$. Since, by change of variables,

$$\int_0^{\sqrt{\|U_n^{(k)}((P^{m-k} f)^2)\|}} \left( \log \frac{A\sqrt{U_n^{(k)}(P^{m-k} F^2)}}{\tau} \right)^{k/2} d\tau$$

$$\le A \int_0^1 (\log u^{-1})^{k/2} \, du \, \sqrt{U_n^{(k)}(P^{m-k} F^2)},$$

and $P^k(U_n^{(k)}(P^{m-k}F^2)) \leq P^m F^2$, it follows by Hölder's inequality and (3.14) that

$$(3.15) \qquad (I) \leq C \|F\|_{L_2(P^m)} \exp\left\{-\frac{9n\|F\|_{L_2(P^m)}^2}{2(2k^2+c_1)}\right\} \leq \frac{D}{\sqrt{n}}.$$

As for $(II)$, we note

$$(3.16) \qquad (II) \leq E\left[\int_0^{\sqrt{\|U_n^{(k)}((P^{m-k}f)^2)\|}} \left(\log \frac{2A\|F\|_{L_2(P^m)}}{\tau}\right)^{k/2} d\tau \, I_n^c\right].$$

Now by integration we see that for any $0 < c < C$

$$\int_0^c (\log(C/x))^{k/2}\left[1 - \frac{k}{2}(\log(C/x))^{-1}\right]dx = c(\log(C/c))^{k/2},$$

which when $(\log(C/c))^{-1} \leq k^{-1}$ gives the inequality

$$(3.17) \qquad \int_0^c (\log(C/x))^{k/2}\,dx \leq 2c(\log(C/c))^{k/2}.$$

Thus since on $I_n^c$

$$(3.18) \qquad 2A\|F\|_{L_2(P^m)}/\sqrt{\|U_n^{(k)}((P^{m-k}f)^2)\|} \geq A \geq e^m > e^k,$$

we get from (3.17), (3.18) and (3.16) that

$$(II) \leq 2E\left[\sqrt{\|U_n^{(k)}((P^{m-k}f)^2)\|}\left(\log \frac{2A\|F\|_{L_2(P^m)}}{\sqrt{\|U_n^{(k)}((P^{m-k}f)^2)\|}}\right)^{k/2}\right].$$

Since the function $\sqrt{x}(-\log x)^{k/2}$, $0 < x < 1$, is concave on $(0, e^{-k}]$ and $A \geq e^m > e^k$, this last bound is by Jensen's inequality

$$(3.19) \qquad \leq 2\sqrt{E\|U_n^{(k)}((P^{m-k}f)^2)\|}\left(\log \frac{2A\|F\|_{L_2(P^m)}}{\sqrt{E\|U_n^{(k)}((P^{m-k}f)^2)\|}}\right)^{k/2}.$$

We are going to show that there exists a $C > 0$ such that

$$(3.20) \qquad B \leq C\left[\frac{1}{\sqrt{n}} + \sqrt{\frac{B}{n^{k/2}} + \sigma^2}\left(\log \frac{A\|F\|_{L_2(P^m)}}{\sigma}\right)^{k/2}\right].$$

We shall consider two cases. In case 1, $E\|U_n^{(k)}((P^{m-k}f)^2)\| \leq \sigma^2$. In this case, since the function $\sqrt{x}(-\log x)^{k/2}$, $0 < x < 1$, is increasing on $(0, e^{-k}]$, we get the trivial bound from (3.15) and (3.19) that

$$B \leq \frac{D}{\sqrt{n}} + 2\sigma\left(\log \frac{2A\|F\|_{L_2(P^m)}}{\sigma}\right)^{k/2},$$

which of course implies the bound (3.20) (note that $A\|F\|_{L_2(P^m)}/\sigma \geq e^m$).

Next consider case 2, $E\|U_n^{(k)}((P^{m-k}f)^2)\| > \sigma^2$. To handle this case we must bound $E\|U_n^{(k)}((P^{m-k}f)^2)\|$. It is here that we will use the induction hypothesis. By Hoeffding's decomposition (3.3) we have

$$(3.21) \qquad E\|U_n^{(k)}((P^{m-k}f)^2)\| \leq \sum_{r=0}^{k} \binom{k}{r} E\|U_n^{(r)}(\pi_r(P^{m-k}f)^2)\|.$$

The term corresponding to $r = 0$ is simply $\|P^k(P^{m-k}f)^2\| \leq \sigma^2$. Consider now the new class

$$\mathcal{G} := \{(P^{m-k}f)^2 : f \in \mathcal{F}\}$$

of functions of $k$ variables. For any probability measure $Q$ on $S^k$,

$$Q((P^{m-k}f)^2 - (P^{m-k}g)^2)^2 \leq 4Q \times P^{m-k}(f-g)^2,$$

and thus the class $\mathcal{G}$ is VC-type with constants $A$ and $\upsilon$ and envelope $2\sqrt{P^{m-k}F^2}$ as in (2.9). Also observe that since

$$\|P^k(P^{m-k}f)^4\| \leq \|P^m f^2\| \leq 4\sigma^2 \leq 4\|\sqrt{P^{m-k}F^2}\|_{L_2(P^k)}^2 = 4\|F\|_{L_2(P^m)}^2,$$

we verify that (3.11) holds. Also we trivially see that (3.10) is satisfied, that is,

$$4n\sigma^2 > n\sigma^2 \geq C_2 \log\left(\frac{2\|F\|_{L_2(P^m)}}{\sigma}\right) = C_2 \log\left(\frac{4\|F\|_{L_2(P^m)}}{2\sigma}\right).$$

Finally noting that $A \geq e^m > e^k$ and $2\sqrt{P^{m-k}F^2} \leq 2$, we are permitted to apply the induction hypothesis for $1 \leq r < k$ ($r = 0$ has already been dealt with) to get with $M = 2$,

$$(3.22) \quad E\|U_n^{(r)}(\pi_r(P^{m-k}f)^2)\| \leq C_1(k, A, \upsilon, 2)n^{-r/2}2\sigma\left(\log\frac{A\|F\|_{L_2(P^m)}}{\sigma}\right)^{r/2}.$$

Since by the hypotheses (3.10) and (3.11) on $\sigma$ [note (3.11) and $F \leq 1$ imply $\sigma^2 \leq 1$] we have

$$(3.23) \qquad \frac{1}{n^{r/2}}\sigma\left(\log\frac{A\|F\|_{L_2(P^m)}}{\sigma}\right)^{r/2} \leq C^{r/2}\sigma^{1+r} \leq C^{r/2}\sigma^2,$$

it follows from the bounds (3.22) and (3.23) that

$$(3.24) \qquad \binom{k}{r} E\|U_n^{(r)}(\pi_r(P^{m-k}f)^2)\| \leq C\sigma^2, \qquad 1 \leq r < k.$$

As for $r = k$, we randomize and use the entropy bound for Rademacher chaos, just as we did at the beginning of the proof, using the fact that $\mathcal{G}$ is VC-type. This

gives

$$E \left\| U_n^{(k)}(\pi_k (P^{m-k} f)^2) \right\|$$

$$\leq C |I_n^k|^{-1/2} E \left\| |I_n^k|^{-1/2} \sum_{I_n^k} \varepsilon_{i_1} \cdots \varepsilon_{i_k} (P^{m-k} f)^2 \right\|$$

(3.25)

$$\leq \frac{C}{|I_n^k|^{1/2}} E \int_0^{\sqrt{\|U_n^{(k)}((P^{m-k} f)^4)\|}} \left( \log \frac{A \sqrt{U_n^{(k)}(P^{m-k} F^2)}}{\tau} \right)^{k/2} d\tau$$

$$\leq C \frac{B}{n^{k/2}},$$

where $B$ is as defined in (3.13) and we use

$$\|U_n^{(k)}((P^{m-k} f)^4)\| \leq \|U_n^{(k)}((P^{m-k} f)^2)\|.$$

From (3.15), (3.19), (3.21)–(3.25) and $E\|U_n^{(k)}((P^{m-k} f)^2)\| > \sigma^2$ we get (3.20).

Since $\log[A\|F\|_{L_2(P^m)}/\sigma] > 1$, hypothesis (3.10) on $\sigma$ gives

$$\frac{1}{\sqrt{n}} \leq C \frac{\sigma}{(\log A\|F\|_{L_2(P^m)}/\sigma)^{1/2}} \leq C\sigma \left( \log \frac{A\|F\|_{L_2(P^m)}}{\sigma} \right)^{k/2},$$

and therefore it follows from inequality (3.20) that with perhaps a different value of $C$,

$$B \leq C \sqrt{\frac{B}{n^{k/2}} + \sigma^2} \left( \log \frac{A\|F\|_{L_2(P^m)}}{\sigma} \right)^{k/2}.$$

Taking squares and solving this inequality, we get that there exists a constant $C$ (easy to evaluate) such that

$$B \leq C \left[ \frac{(\log A\|F\|_{L_2(P^m)}/\sigma)^k}{n^{k/2}} + \sigma \left( \log \frac{A\|F\|_{L_2(P^m)}}{\sigma} \right)^{k/2} \right].$$

But, again by condition (3.10) on $\sigma$ as in (3.23),

$$\frac{(\log A\|F\|_{L_2(P^m)}/\sigma)^{k/2}}{n^{k/2}} \leq C\sigma^k \leq \sigma,$$

and therefore

$$B \leq C\sigma \left( \log \frac{A\|F\|_{L_2(P^m)}}{\sigma} \right)^{k/2},$$

which, by inequality (3.12), proves the theorem. $\square$

REMARK 4. An analogue of the previous theorem holds if we replace the function $(\log(A\|F\|_{L_2(Q)}/\tau)$ by $H(\|F\|_{L_2(Q)}/\tau)$ with $H$ an increasing regularly varying function of exponent $0 \leq \alpha < 2/m$, very much as in Theorem 3.1 in Giné and Koltchinskii [17].

### 3.4. *U-statistics taking values in separable Banach spaces.*

3.4.1. *The $L_p$ case, $2 \leq p < \infty$.* We shall begin by establishing an exponential bound for the tail of the norm of a $U$-statistic taking values in a separable type 2 Banach space **B**. Let us recall that a separable Banach space **B** is of type 2 if for any finite number of points $x_i \in \mathbf{B}$ and independent Rademacher variables $\varepsilon_i$ (independent random signs), we have $E \|\sum_{i=1}^n \varepsilon_i x_i\|^2 \leq C \sum_{i=1}^n \|x_i\|^2$, and the smallest such constant $C$ is the type 2 constant of **B**. It is well known that the $L_p$ spaces are of type 2 if (and only if) $p \geq 2$. If $Z_i$ are independent, centered **B**-valued random vectors with a square integrable norm and **B** is of type 2, then

$$(3.26) \qquad E \left\| \sum_{i=1}^n Z_i \right\|^2 \leq C \sum_{i=1}^n E \|Z_i\|^2$$

(Araujo and Giné [2] or Ledoux and Talagrand [24]).

THEOREM 9. *Let $H$ be a function of $k$ variables, $P$-canonical, symmetric, with values in a type 2 Banach space. We have, for all $x \geq 0$,*

$$(3.27) \qquad \Pr\left\{ \left\| \sum_{\mathbf{i} \in I_n^k} H(X_{i_1}, \ldots, X_{i_k}) \right\| \geq x \right\} \leq D \exp\left\{ -\left( \frac{x}{\lambda_0 n^{k/2} \kappa} \right)^{2/k} \right\},$$

*where $\kappa$ is a number satisfying*

$$(3.28) \qquad \kappa \geq \sup_{x_1, \ldots, x_k \in S} \|H(x_1, \ldots, x_k)\|$$

*and $\lambda_0$ and $D$ are constants that depend only on $k$ and the type 2 constant $C$ of **B**.*

PROOF. (This inequality was mentioned on page 252 of de la Peña and Giné [7] and its precise statement and proof were left to the reader.) Here is the proof.

Let

$$S_n = \frac{1}{\kappa \binom{n}{k}^{1/2}} \sum_{1 \leq i_1 < \cdots < i_k \leq n} H(X_{i_1}, \ldots, X_{i_k}),$$

and let

$$S_n' = \frac{1}{\kappa \binom{n}{k}^{1/2}} \sum_{1 \leq i_1 < \cdots < i_k \leq n} \varepsilon_{i_1}^1 \cdots \varepsilon_{i_k}^k H(X_{i_1}^{(1)}, \ldots, X_{i_k}^{(k)}),$$

where $\kappa$ is as in (3.28), $X_i^{(j)}$ are i.i.d. with law $P$, and $\varepsilon_i^j$ are i.i.d. Rademacher variables independent of the collection of variables $\{X_i^{(j)}\}$.

By decoupling (Theorem 3.1.1 of de la Peña and Giné [7]) and convexity (e.g., Theorem 3.5.3 on page 140 of de la Peña and Giné [7], as it applies to a nonnegative, nondecreasing convex function $\Psi$), there is a constant $C_k$ such that

(3.29)                        $\| \|S_n\| \|_{\Psi_{2/k}} \le C_k \| \|S_n'\| \|_{\Psi_{2/k}}.$

By hypercontractivity of Rademacher chaos (Khinchin's inequality for Rademacher chaos), for example, Theorem 3.2.2 on page 113 of de la Peña and Giné [7], for all $r \ge 2$,

$$E_\varepsilon \|S_n'\|^r \le r^{kr/2}(E_\varepsilon \|S_n'\|^2)^{r/2},$$

and by the type 2 inequality applied one sequence $\{\varepsilon_i^j\}$, $j = 1, \ldots, k$, at a time, we get

$$(E_\varepsilon \|S_n'\|^2)^{r/2} \le C^{kr/2},$$

where $C$ is the type 2 constant of the Banach space, and therefore,

$$E_\varepsilon \|S_n'\|^r \le C^{kr/2} r^{kr/2}.$$

This inequality yields, by Taylor expansion of the exponential, that for some $c > 0$,

$$E_\varepsilon \Psi_{2/k}(c\|S_n'\|) < \infty$$

and therefore that there exists a constant $\lambda_1 = \lambda_1(C, k)$ depending only on $C$ and $k$, such that

$$E_\varepsilon \Psi_{2/k}(\|S_n'\|/\lambda_1) \le 1.$$

Integrating with respect to the $X$'s,

$$E\Psi_{2/k}(\|S_n'\|/\lambda_1) \le 1, \qquad \text{that is, } \| \|S_n'\| \|_{\Psi_{2/k}} \le \lambda_1,$$

which, combined with (3.29), gives

$$\| \|S_n\| \|_{\Psi_{2/k}} \le C_k \lambda_1 := \lambda_0,$$

a constant. Of course this implies, by the definition of the Orlicz norm $\Psi_{2/k}$, that there is a constant $D(k, C)$ such that

$$\Pr\{\|S_n\| > x\} \le De^{-(x/\lambda_0)^{2/k}}. \qquad \qquad \square$$

3.4.2. *The $L_p$ case*, $1 \le p < 2$.  Our aim now is to obtain a useful exponential inequality for **B**-valued $U$-statistics, when **B** is not necessarily a type 2 Banach space. We will generalize to **B**-valued $U$-statistics (via decoupling followed by iteration) the following sharp inequality for sums of independent random vectors: there is a constant $L < \infty$ such that if **B** is a separable Banach space, $Z_i$, $i \in \mathbf{N}$, are

independent mean zero random vectors taking values in **B** and $r \geq 2$, then, setting $S_n = \sum_{i=1}^{n} Z_i$,

$$(3.30) \qquad E\|S_n\|^r \leq L^r \left[ r^{r/2}(E\|S_n\|^2)^{r/2} + r^r E \max_{1 \leq i \leq n} \|Z_i\|^r \right].$$

This inequality was obtained by Pinelis [28], and it also follows easily from the sharper inequality in Giné, Latała and Zinn [18], Proposition 3.1. Next we extend inequality (3.30) to **B**-valued *U*-statistics.

THEOREM 10. *Let* **B** *be a separable Banach space, let* $H : S^k \mapsto \mathbf{B}$ *be a bounded P-canonical random vector symmetric in its entries and let* $X_i, X_i^{(j)}$ *be i.i.d. S-valued random variables,* $1 \leq i \leq n$ *and* $j = 1, \ldots, k$. *Define* $\kappa$ *and* $\chi_n$ *to be any pair of numbers such that*

$$\kappa \geq \sup_{x_1, \ldots, x_k \in S} \|H(x_1, \ldots, x_k)\|$$

$$(3.31)$$

$$\text{and} \quad \chi_n \geq \left( E \left\| \sum_{\mathbf{i} \in I_n^k} H(X_{i_1}^{(1)}, \ldots, X_{i_k}^{(k)}) \right\|^2 \right)^{1/2}.$$

*Then there exists a constant C depending only on k such that, for all* $n \in \mathbf{N}$ *and* $r \geq 2$,

$$E \left\| \sum_{\mathbf{i} \in I_n^k} H(X_{i_1}^{(1)}, \ldots, X_{i_k}^{(k)}) \right\|^r$$

$$(3.32)$$

$$\leq C^r \big[ r^{kr/2} \chi_n^r + r^{(k+1)r/2} n^{(k-1)r} \kappa^r + r^{kr} \kappa^r \big].$$

*Moreover, there exists a constant* $D_0$ *depending only on k such that, for all* $n \in \mathbf{N}$ *and* $r \geq 2$,

$$(3.33) \quad E \left\| \sum_{\mathbf{i} \in I_n^k} H(X_{i_1}, \ldots, X_{i_k}) \right\|^r \leq D_0^r \big[ r^{kr/2} \chi_n^r + r^{(k+1)r/2} n^{(k-1)r} \kappa^r + r^{kr} \kappa^r \big],$$

*where* $\kappa$ *and* $\chi_n$ *are defined as in* (3.31).

PROOF. Inequality (3.30) gives the result for $k = 1$. Assume the result is true for $k$ and for every Banach space, and let $H$ be a function of $k + 1$ variables satisfying the conditions in the statement of the theorem. Before starting the induction, we describe some simplifying notation. We will denote by $\leq'$ inequality up to a multiplicative constant $C^r$ with $C$ depending only on $k$. Also, we will write $H_{\mathbf{i}} = H$ if the coordinates of the multi-index $\mathbf{i} \in \{1, \ldots, k+1\}^n$ are all different, and $H_{\mathbf{i}} = 0$ otherwise, and we will drop the arguments of $H$ or $H_{\mathbf{i}}$, so $\sum_{\mathbf{i}} H_{\mathbf{i}}$ will mean $\sum_{\mathbf{i} \in \{1, \ldots, k+1\}^n} H_{\mathbf{i}}(X_{i_1}^{(1)}, \ldots, X_{i_{k+1}}^{(k+1)})$. Finally, for $A \subset \{1, \ldots, k+1\}$, $E_A$ will mean integration with respect to the variables $X_i^{(r)}$ only, for $r \in A$ and $i \leq n$.

Applying (3.30) conditionally on the variables $X_i^{(j)}$, $j = 1, \ldots, k$, we obtain

$$E\left\|\sum_{\mathbf{i}} H_{\mathbf{i}}\right\| \le' E_{1,\ldots,k}\left[r^{r/2}\left(E_{k+1}\left\|\sum_{i_{k+1}}\left(\sum_{i_1,\ldots,i_k} H_{\mathbf{i}}\right)\right\|^2\right)^{r/2}\right]$$

(3.34)

$$+ E_{1,\ldots,k}\left[r^r E_{k+1}\max_{i_{k+1}}\left\|\sum_{i_1,\ldots,i_k} H_{\mathbf{i}}\right\|^r\right].$$

In order to deal with the first term in (3.34), for each $x_1, \ldots, x_k \in S$ fixed we consider the random variable $\sum_{i_{k+1}} H_{\mathbf{i}}(x_1, \ldots, x_k, X_{i_{k+1}}^{(k+1)})$ as a function from $S^k$ into the Banach space $L_2(\Omega, \Sigma, P; \mathbf{B})$ of $\mathbf{B}$-valued random variables whose $\mathbf{B}$ norms are square integrable, with norm

$$\left|F(x_1, \ldots, x_k, X_{i_{k+1}}^{(k+1)})\right|_* := \left(E_{k+1}\|F(x_1, \ldots, x_k, X_{i_{k+1}}^{(k+1)})\|^2\right)^{1/2}.$$

To apply the induction hypothesis to the statistic $\sum_{i_1,\ldots,i_k}(\sum_{i_{k+1}} H_{\mathbf{i}})$ in this Banach space with the norm $|\cdot|_*$, we first note that, if we denote by $\tilde{\kappa}$ and $\tilde{\chi}$ the corresponding quantities associated with this statistic, then

$$\tilde{\kappa} = \max_{x_1,\ldots,x_k\in S}\left|\sum_{i_{k+1}} H(x_1, \ldots, x_k, X_{i_{k+1}}^{(k+1)})\right|_* \le n\kappa$$

and

$$\tilde{\chi}_n = \left(E_{1,\ldots,k}\left|\sum_{i_1,\ldots,i_k}\left(\sum_{i_{k+1}} H_{\mathbf{i}}\right)\right|_*^2\right)^{1/2} = \left(E\left\|\sum_{\mathbf{i}} H_{\mathbf{i}}\right\|^2\right)^{1/2} = \chi_n.$$

Hence, the induction hypothesis gives

$$r^{r/2}E_{1,\ldots,k}\left(E_{k+1}\left\|\sum_{i_{k+1}}\left(\sum_{i_1,\ldots,i_k} H_{\mathbf{i}}\right)\right\|^2\right)^{r/2}$$

$$= r^{r/2}E_{1,\ldots,k}\left|\sum_{i_1,\ldots,i_k}\left(\sum_{i_{k+1}} H_{\mathbf{i}}\right)\right|^r$$

$$\le' r^{r/2}\left[r^{kr/2}\tilde{\chi}_n^r + r^{(k+1)r/2}n^{(k-1)r}\tilde{\kappa}^r + r^{kr}\tilde{\kappa}^r\right]$$

$$\le r^{(k+1)r/2}\chi_n^r + r^{(k+2)r/2}n^{kr}\kappa^r + r^{(k+1/2)r}n^r\kappa^r.$$

Now, if $n \le r^{1/2}$, then $r^{(k+1/2)r}n^r \le r^{(k+1)r}$ and if $n > r^{1/2}$, then $r^{(k+1/2)r}n^r \le r^{(k+2)r/2}n^{kr}$, so that

$$r^{r/2}E_{1,\ldots,k}\left(E_{k+1}\left\|\sum_{i_{k+1}}\left(\sum_{i_1,\ldots,i_k} H_{\mathbf{i}}\right)\right\|^2\right)^{r/2}$$

(3.35)

$$\le' r^{(k+1)r/2}\chi_n^r + r^{(k+2)r/2}n^{kr}\kappa^r + r^{(k+1)r}\kappa^r.$$

Finally, we apply the induction hypothesis conditionally on the variables $X_i^{(k+1)}$ to the second term of (3.34) by considering that term to be a *U*-statistic with values in the Banach space $\ell_n^\infty(\mathbf{B}) := \{(v_1, \ldots, v_n) : v_i \in \mathbf{B}\}$, with norm $|(v_1, \ldots, v_n)| := \max_{1 \le i \le n} \|v_i\|$. In fact, with this definition,

$$\max_{i_{k+1}} \left\| \sum_{i_1, \ldots, i_k} H_{\mathbf{i}} \right\|$$

$$= \left| \sum_{i_1, \ldots, i_k} \left( H_{\mathbf{i}}\big(X_{i_1}^{(1)}, \ldots, X_{i_k}^{(k)}, X_1^{(k+1)}\big), \ldots, H_{\mathbf{i}}\big(X_{i_1}^{(1)}, \ldots, X_{i_k}^{(k)}, X_n^{(k+1)}\big) \right) \right|,$$

and if we denote by $\overline{\kappa}$ and $\overline{\chi}_n$ the corresponding parameters, we have, for each value of $X_i^{(k+1)}$, $j = 1, \ldots, n$,

$$\overline{\kappa} = \max_{x_1, \ldots, x_k \in S} \left\| \big( H\big(x_1, \ldots, x_k, X_1^{(k+1)}\big), \ldots, H\big(x_1, \ldots, x_k, X_n^{(k+1)}\big) \big) \right\| \le \kappa$$

and

$$\overline{\chi}_n = \left( E_{1, \ldots, k} \max_{i_{k+1}} \left\| \sum_{i_1, \ldots, i_k} H_{\mathbf{i}} \right\|^2 \right)^{1/2} \le n^k \kappa.$$

So, induction gives

$$r^r E_{1, \ldots, k} \max_{i_{k+1}} \left\| \sum_{i_1, \ldots, i_k} H_{\mathbf{i}} \right\|^r \le' r^r \big[ r^{kr/2} \overline{\chi}_n^r + r^{(k+1)r/2} n^{(k-1)r} \overline{\kappa}^r + r^{kr} \overline{\kappa}^r \big]$$

$$\le \kappa^r \big[ r^{(k+2)r/2} n^{kr} + r^{(k+3)r/2} n^{(k-1)r} + r^{(k+1)r} \big].$$

By considering the cases $n \le r^{1/2}$ and $n > r^{1/2}$ we see that

$$r^{(k+3)r/2} n^{(k-1)r} \le \max\big( r^{(k+2)r/2} n^{kr}, r^{(k+1)r} \big),$$

from which it follows that

$$(3.36) \qquad r^r E_{1, \ldots, k} \max_{i_{k+1}} \left\| \sum_{i_1, \ldots, i_k} H_{\mathbf{i}} \right\|^r \le' \kappa^r \big[ r^{(k+2)r/2} n^{kr} + r^{(k+1)r} \big].$$

Now, the first part of the theorem follows by substituting the estimates (3.35) and (3.36) into inequality (3.34). The proof of the theorem is completed by noting that (3.33) follows from (3.32) via de la Peña's [6] decoupling inequality (see Theorem 3.1.1 in de la Peña and Giné [7]). □

We note that this theorem could have been proved, with only formal changes, for $H_{\mathbf{i}}$ depending on the subindices $\mathbf{i}$. Giné, Latała and Zinn, unpublished, have a moment inequality for **B**-valued canonical *U*-statistics of order 2 without boundedness assumptions that contains (3.32) for $k = 2$. Our proof is inspired by theirs.

We shall apply (3.33) to the special case when $\mathbf{B} = L_p(\mathbf{R}^d)$, $1 \leq p < 2$, and $H(x_1, \ldots, x_k) = \pi_k \overline{K}_h(\cdot, x_1, \ldots, x_k)$, to obtain the following. Recall the definition (2.21) of $\mu_s$.

COROLLARY 3. *Let $K$, $f_g$ and $s > 0$ satisfy the conditions of Theorem 3 for $1 \leq p < 2$. Then there exist $\gamma > 0$ and $D > 0$ such for all $n \geq 1$ and $x > 0$,*

$$(3.37) \quad \Pr\left\{ \frac{\|\sum_{I_n^k} \pi_k \overline{K}_h(\cdot, X_{i_1}, \ldots, X_{i_k})\|_p}{Dn^{k-1}/\sqrt{h}} > x \right\} \leq \gamma \exp\left(-\left(\frac{x}{\|K\|_{p,2,s}}\right)^{1/k}\right),$$

*where $\|K\|_{p,2,s} = \|K\|_p \vee \|K\|_{L_2(\mu_s)}$.*

For the proof, first, note the following easy bound for $\|\|\pi_k \overline{K}_h\|_p\|_\infty$, namely, we can choose some $A_1 > 0$, so that

$$(3.38) \qquad \kappa := A_1 \|K\|_p / h^{1-1/p} \geq \sup_{x_1, \ldots, x_k \in S} \|\pi_k \overline{K}_h(\cdot, x_1, \ldots, x_k)\|_p.$$

To see this, just observe that $\pi_k \overline{K}_h(t, x_1, \ldots, x_k)$ is a linear combination of terms of the form $h_\ell(t, x_{i_1}, \ldots, x_{i_\ell})$ with $(i_1, \ldots, i_\ell) \in I_n^\ell$ and $0 \leq \ell \leq k$, where $h_0(t) = P^m K_h$ and, for $\ell \geq 1$,

$$h_\ell(t, x_{i_1}, \ldots, x_{i_\ell}) = P^{m-\ell} \overline{K}_h(t, x_1, \ldots, x_\ell, X_{\ell+1}, \ldots, X_m)$$
$$= P^{m-\ell} \frac{1}{m!} \sum_{\sigma \in I_m^m} K_h\big(t - g_\sigma(x_1, \ldots, x_\ell, X_{\ell+1}, \ldots, X_m)\big)$$

and $g_\sigma(y_1, \ldots, y_m) = g(y_{\sigma_1}, \ldots, y_{\sigma_m})$. Then Jensen's inequality for expectations and averages and a substitution yield

$$\int |h_\ell(t, x_1, \ldots, x_l)|^p \, dt \leq \int |K_h(t)|^p \, dt = h^{1-p} \int |K(t)|^p \, dt.$$

We shall also need a good bound for $E\|\sum_{I_n^k} \pi_k \overline{K}_h(\cdot, X_{i_1}, \ldots, X_{i_k})\|_p$ in order to estimate $\chi_n$ in (3.31). Such a bound will be based on the following lemma, which is an extension of ideas and results in Chapter 7 in Devroye [8] and Section 3 of Devroye [9].

LEMMA 1. *Let $1 \leq p < 2$. If $f$ and $k$ are two nonnegative functions on $\mathbf{R}^d$ such that $f, k \in L_1(\mu_s)$ for some $s > d(2-p)/p$, then, for any $0 < b < \infty$,*

$$\sup_{h \in (0,b]} \int (k_h * f)^{p/2}(y) \, dy \leq C \big(\|f\|_{L_1(\mu_s)} \|k\|_{L_1(\mu_s)}\big)^{p/2}$$

*for some constant $C$ that depends only on $d$, $p$ and $s$.*

PROOF.    Let $q = sp/(2 - p)$ and note that $v(y) := (1 + |y|)^{-q}$ is integrable. Then, by Jensen with respect to the probability measure $v(y)\,dy/\|v\|_1$, we have

$$\int (k_h * f)^{p/2}(y)\,dy = \int v(y)^{-1}(k_h * f)^{p/2}(y)v(y)\,dy$$

$$\leq \|v\|_1^{1-p/2}\left(\int (1 + |y|)^s(k_h * f)(y)\,dy\right)^{p/2}.$$

Since $(1 + |u + v|) \leq (1 + |u|)(1 + |v|)$ and $1 + |hu| \leq (1 \vee h)(1 + |u|)$ for $h > 0$, we also have

$$\int (1 + |y|)^s(k_h * f)(y)\,dy$$

$$\leq \int\int (1 + |y - x|)^s(1 + |x|)^s k_h(y - x)f(x)\,dx\,dy$$

$$\leq (1 \vee b^{s/d})\|f\|_{L_1(\mu_s)}\|k\|_{L_1(\mu_s)}.$$

The lemma follows from these inequalities.    $\square$

LEMMA 2.    *If $K \in L_2(\mathbf{R}^d)$ and $h > 0$, then, for $1 \leq k \leq m$ and $t \in \mathbf{R}^d$,*

$$(3.39) \qquad E(\pi_k\overline{K}_h(t, X_1, \ldots, X_k))^2 \leq \frac{((K^2)_h * f_g)(t)}{h}.$$

*Let now $K$, $f_g$ and $s > 0$ satisfy the conditions of Theorem 3 for $1 \leq p < 2$, and let $0 < b < \infty$. Then there exists $C < \infty$, such that, for $1 \leq k \leq m$ and all $0 < h \leq b$,*

$$(3.40) \qquad \left(E\left\|\sum_{I_n^k}\pi_k\overline{K}_h(\cdot, X_{i_1}, \ldots, X_{i_k})\right\|_p^2\right)^{1/2} \leq \frac{C\|K\|_{L_2(\mu_s)}n^{k/2}}{h^{1/2}}.$$

PROOF.    Using the fact that $\pi_k$ is a projection in $L_2(\text{Pr})$, and applying convexity of the square function to the symmetrization of $K_h$, we have

$$E(\pi_k\overline{K}_h(t, X_1, \ldots, X_k))^2 \leq \frac{1}{h^2}EK^2\left(\frac{t - g(X_1, \ldots, X_m)}{h^{1/d}}\right)$$

$$= \frac{1}{h}((K^2)_h * f_g)(t),$$

that is, (3.39). Next, by orthogonality,

$$E\left(\sum_{I_n^k}\pi_k\overline{K}_h(\cdot, X_{i_1}, \ldots, X_{i_k})\right)^2 = \frac{k!n!}{(n - k)!}E(\pi_k\overline{K}_h(t, X_1, \ldots, X_k))^2.$$

Then the Minkowski inequality for integrals (e.g., Folland [14], page 194) and the above results yield

$$\left( E \left\| \sum_{I_n^k} \pi_k \overline{K}_h(\cdot, X_{i_1}, \ldots, X_{i_k}) \right\|_p^2 \right)^{1/2}$$

$$\leq \left\| \left( E \left( \sum_{I_n^k} \pi_k \overline{K}_h(\cdot, X_{i_1}, \ldots, X_{i_k}) \right)^2 \right)^{1/2} \right\|_p$$

$$\leq \left( \frac{m! n^k}{h} \right)^{1/2} \left( \int ((K^2)_h * f_g)^{p/2}(t) \, dt \right)^{1/p}.$$

The result (3.40) follows now from Lemma 1.  □

PROOF OF COROLLARY 3.   It follows from Lemma 2 and from (3.38) that we can take, for some $A_2 > 0$,

$$\chi_n := A_2 n^{k/2} \|K\|_{p,2,s} / \sqrt{h} \quad \text{and} \quad \kappa := A_1 \|K\|_{p,2,s} / h^{1-1/p}.$$

Assume that $0 < h \leq 1$ and $r \geq 2$. Then we have the bounds

$$r^{k/2} \chi_n = A_2 \|K\|_{p,2,s} r^{k/2} n^{k/2} / \sqrt{h} \leq A_2 \|K\|_{p,2,s} r^k n^{k-1} / \sqrt{h},$$

$$r^{(k+1)/2} n^{k-1} \kappa = A_1 \|K\|_{p,2,s} r^{(k+1)/2} n^{k-1} / h^{1-1/p} \leq A_1 \|K\|_{p,2,s} r^k n^{k-1} / \sqrt{h}$$

and

$$r^k \kappa = r^k A_1 \|K\|_{p,2,s} / h^{1-1/p} \leq A_1 \|K\|_{p,2,s} r^k n^{k-1} / \sqrt{h}.$$

By (3.33), this says that for some $D > 0$, for all $r \geq 2$,

$$E \left\| \sum_{I_n^k} \pi_k \overline{K}_h(\cdot, X_{i_1}, \ldots, X_{i_k}) \right\|_p^r \leq \left( D \|K\|_{p,2,s} r^k n^{k-1} / \sqrt{h} \right)^r$$

and thus

$$E \left\| \sum_{I_n^k} \pi_k \overline{K}_h(\cdot, X_{i_1}, \ldots, X_{i_k}) \right\|_p^{r/k} \leq \left( E \left\| \sum_{I_n^k} \pi_k \overline{K}_h(\cdot, X_{i_1}, \ldots, X_{i_k}) \right\|_p^r \right)^{1/k}$$

$$\leq \left( D \|K\|_{p,2,s} r^k n^{k-1} / \sqrt{h} \right)^{r/k}$$

$$= \left( 3^k D \|K\|_{p,2,s} n^{k-1} / \sqrt{h} \right)^{r/k} r^r / 3^r.$$

The same bound holds for $r = 1$. From this moment bound we easily get that for all $k \geq 2$,

$$E \exp \left( \left[ \frac{\| \sum_{I_n^k} \pi_k \overline{K}_h(\cdot, X_{i_1}, \ldots, X_{i_k}) \|_p}{3^k D \|K\|_{p,2,s} n^{k-1} / \sqrt{h}} \right]^{1/k} \right) < \sum_{r=0}^{\infty} \frac{r^r}{3^r r!} =: \gamma < \infty.$$

The desired result (3.37) follows now from Markov's inequality and a renaming of $D$. $\square$

Notice for future use that by (3.37) we get via (2.16) that for some $C > 0$,

$$(3.41) \qquad \left\| \frac{\sqrt{n} \| \sum_{I_n^k} \pi_k \overline{K}_h(\cdot, X_{i_1}, \ldots, X_{ik}) \|_p}{n^k} \right\|_{\Psi_{1/k}} \le \frac{CD \|K\|_{p,2,s}}{\sqrt{nh}}.$$

3.5. *Completion of the proofs of the CLT.* We are now ready to prove (3.8) for $1 \le p \le \infty$, which will complete the proofs of Theorems 1, 2 and 3 (Section 3.1). We begin with the case $p = \infty$.

PROOF OF THEOREM 1. Let

$$\mathcal{K}_{g,n} := \big\{ K\big((y - g(\cdot))/h^{1/d}\big) : y \in \mathbf{R}^d, ah_n \le h \le bh_n \big\}.$$

Since $N(\mathcal{K}_{g,n}, L_2(\tilde{Q}), \varepsilon) = N(\mathcal{K}_n, L_2(\tilde{Q} \circ g^{-1}), \varepsilon)$ for every probability measure $\tilde{Q}$ on $\mathbf{R}^m$, it follows from condition (a) in Theorem 1 that the classes $\mathcal{K}_{g,n}$ are VC-type with the same $A$ and $v$ and with envelopes $F_n(g)$. Moreover, since the map

$$(y, h, x_1, \ldots, x_m) \mapsto \overline{K}\big((y - g(x_1, \ldots, x_m))/h^{1/d}\big)$$

is jointly measurable, these classes are image admissible Suslin, hence measurable. So, we can apply Theorem 8 to them. Since by Minkowski's inequality

$$(\lambda h_n)^2 E \overline{K}^2_{\lambda h_n}(y, X_1, \ldots, X_m) \le E K^2 \bigg( \frac{y - g(X_1, \ldots, X_m)}{(\lambda h_n)^{1/d}} \bigg)$$

$$= \int K^2 \bigg( \frac{y - w}{(\lambda h_n)^{1/d}} \bigg) f_g(w) \, dw \le \|f_g\|_\infty \|K\|_2^2 bh_n,$$

we can take $\sigma^2 = Ch_n$ with $C = b \|f_g\|_\infty \|K\|_2^2 / \|\pi_k \overline{K}\|_\infty^2$. Hence, Theorem 8 gives that there is a constant $C_k$ such that, for all $n$ and $k$,

$$E \bigg\{ \sup \bigg| \frac{1}{n^{k/2}} \lambda h_n \sum_{I_n^k} \pi_k \overline{K}_{\lambda h_n}(t, X_{i_1}, \ldots, X_{ik}) \bigg| : t \in \mathbf{R}^d, a \le \lambda \le b \bigg\}$$

$$\le C_k \sqrt{h_n} \big[ 1 \vee \log\big(A \|F_n\|_{L_2(f_g)} / \sqrt{h_n}\big) \big].$$

Hence, if $nh_n/[1 \vee \log(A\|F_n\|_{L_2(f_g)}/\sqrt{h_n})]^2 \to \infty$, then for all $k = 2, \ldots, m$,

$$(3.42) \quad E \sup_{a \le \lambda \le b} \big\| \sqrt{n} U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n}) \big\|_\infty \le \frac{C[1 \vee \log(A\|F_n\|_{L_2(f_g)}/\sqrt{h_n})]}{n^{(k-1)/2} h_n^{1/2}} \to 0,$$

proving (3.8) in the $L_\infty$ case. $\square$

PROOF OF COROLLARY 1.    By the Hoeffding decomposition (3.5), it suffices to show [in analogy with Proposition 1 and (3.8)] that

$$(3.43) \qquad \sup_{0<\lambda\leq 1} \big|\lambda\big(\sqrt{n}m U_n^{(1)}(\pi_1 \overline{K}_{\lambda h_n}(0, \cdot, \ldots, \cdot)) - \overline{v}_n(0)\big)\big| \to 0 \qquad \text{in pr}$$

and

$$(3.44) \qquad \sup_{0<\lambda\leq 1} \lambda\sqrt{n}\big|U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n}(0, \cdot, \ldots, \cdot))\big| \to 0, \qquad k = 2, \ldots, m.$$

As in the proof of Proposition 1, (3.43) reduces to proving

$$\sup_{0<\lambda\leq 1} \lambda \int |\overline{v}_n(-u) - \overline{v}_n(0)| |K_{\lambda h_n}(u)|\, du \to 0$$

for $i = 1, \ldots, m$. Let $M$ be such that $D$ is contained in the ball of radius $M$ about the origin in $\mathbf{R}^d$. Since $K_{\lambda h_n}(u) = I\{u \in (\lambda h_n)^{1/d} D/(\lambda h_n \operatorname{Vol}(D))\}$ is zero for $|u| > (\lambda h_n)^{1/d} M$, we have

$$\sup_{0<\lambda\leq 1} \lambda \int |\overline{v}_n(-u) - \overline{v}_n(0)| |K_{\lambda h_n}(u)|\, du \leq \sup_{|u|\leq h_n^{1/d}M} |\overline{v}_n(-u) - \overline{v}_n(0)|,$$

which tends to zero in probability by condition (2.13), proving (3.43). For each $n$, the class of functions $\{I_{\lambda h_n D} : 0 \leq \lambda \leq 1\}$ is VC because it is linearly ordered (these are the simplest VC classes), and its envelope is contained in the class. The same is true for the class $\{\lambda h_n \overline{K}_{\lambda h_n}(0, \cdot, \ldots, \cdot)/\|\pi_k \overline{K}(0, \cdot, \ldots, \cdot)\|_\infty : 0 < \lambda \leq 1\}$. Then the logarithm in the bound (3.9) in Theorem 8 is simply a constant, and that theorem gives

$$n^{k/2} \frac{h_n}{\|\pi_k \overline{K}(0, \cdot, \ldots, \cdot)\|_\infty} E \sup_{0<\lambda\leq 1} \big|\lambda U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n}(0, \cdot, \ldots, \cdot))\big| \leq C\sigma$$

as long as $n\sigma^2 > C'$, for fixed constants $C$ and $C'$ and for $k = 2, \ldots, m$, where we can take

$$\sigma^2 = \frac{h_n \|f_g\|_\infty}{\operatorname{Vol}(D)\|\pi_k \overline{K}(0, \cdot, \ldots, \cdot)\|_\infty^2}$$

because, by change of variables,

$$\sup_{0<\lambda\leq 1} E\big[\lambda^2 h_n^2 K_{\lambda h_n}^2(-g(X_1, \ldots, X_m)/(\lambda h_n)^{1/d})\big] \leq h_n \|f_g\|_\infty / \operatorname{Vol}(D).$$

Since $nh_n \to \infty$ by hypothesis, the condition $n\sigma^2 > C'$ is satisfied, and therefore, we can apply Theorem 8 and obtain

$$\sup_{0<\lambda\leq 1} \lambda\sqrt{n}\big|U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n}(0, \cdot, \ldots, \cdot))\big| \leq \frac{C_k}{n^{(k-1)/2}h_n^{1/2}} \to 0$$

for all $2 \leq k \leq m$, proving (3.44).    $\square$

To finish the proof of Theorems 2 and 3 is more complicated because we must deal with a mixed norm, the sup over $\lambda$ of an $L_p$ norm. The proof will consist in showing that the exponential inequalities from the above subsections lead to an entropy bound of the random variables in (3.8).

Given a kernel $L$ in $L_p(\mathbf{R}^d)$, $1 \leq p < \infty$, and $h > 0$, let $\overline{L}_h(t, x_1, \ldots, x_m)$ be the symmetrization of $L_h(t - g(x_1, \ldots, x_m))$, as in (3.4). We observe that, just as in (3.38), there is a constant $c < \infty$ that depends only on $k$ such that

$$(3.45) \qquad \|\|\pi_k \overline{L}_h\|_p\|_\infty \leq \frac{c\|L\|_p}{h^{(p-1)/p}}, \qquad k = 1, \ldots, m.$$

Because of (3.45), it makes sense to define

$$(3.46) \qquad \mathbf{X}_{L,n} := \left\| \frac{h_n^{(p-1)/p}}{n^{k/2}} \sum_{\mathbf{i} \in I_n^k} \pi_k \overline{L}_{h_n}(\cdot, X_{i_1}, \ldots, X_{i_m}) \right\|_p.$$

PROOF OF (3.8) FOR $2 \leq p < \infty$. For $p \geq 2$, $L_p$ is of type 2 and Theorem 9 then gives that, for a constant $C$ that depends only on $k$ and $p$, for all $x \geq 0$,

$$(3.47) \qquad \Pr\{\mathbf{X}_{L,n} \geq x\} \leq D \exp\left\{ -\left( \frac{x}{C\|L\|_p} \right)^{2/k} \right\},$$

and then, by (2.16), we have for another constant $C$ that depends on $k$ and $p$,

$$(3.48) \qquad \|\mathbf{X}_{L,n}\|_{\Psi_{2/k}} \leq C\|L\|_p,$$

where $\Psi_{2/k}$ is the Young modulus of exponential type defined in (2.14) [but with $\Psi_1(x) = e^x - 1$] and $\|\cdot\|_{\Psi_{2/k}}$ is the associated (pseudo)norm (2.15). Applying (3.48) to $L = K_\lambda - K_{\lambda'}$, we obtain

$$(3.49) \qquad \|\mathbf{X}_{K_\lambda, n} - \mathbf{X}_{K_{\lambda'}, n}\|_{\Psi_{2/k}} \leq \|\mathbf{X}_{K_\lambda - K_{\lambda'}, n}\|_{\Psi_{2/k}} \leq C\|K_\lambda - K_{\lambda'}\|_p.$$

Then this bound and (3.45) allow us to apply the usual entropy integral bound, for example, in the version given in de la Peña and Giné [7], Corollary 5.1.5, and conclude that for some constant $D$, keeping in mind that $0 \in \mathcal{K}_{[a,b]}$,

$$(3.50) \qquad \left\| \sup_{a \leq \lambda \leq b} \mathbf{X}_{K_\lambda, n} \right\|_{\Psi_{2/k}} \leq D \int_0^{C\|K\|_p} \Psi_{2/k}^{-1}\big( N\big( \mathcal{K}_{[a,b]}, d_p, \varepsilon \big) \big) \, d\varepsilon,$$

where $\mathcal{K}_{[a,b]}$ is defined by (2.17) and $d_p$ is the $L_p(\mathbf{R}^d)$ distance defined in (2.18) (technically, this only holds for any separable version of the process $\mathbf{X}_{K_\lambda, n}$, but we see in a remark below that this process itself is separable). Now, since up to constants $\Psi_\alpha$ is increasing in $\alpha$, hypothesis (a) in Theorem 2 implies that this integral is finite for all $2 \leq k \leq m$. Taking into account that the Orlicz distances of exponential type dominate (up to constants) the $L_r(\Pr)$ distances, inequality (3.50) implies that for all $2 \leq k \leq m$,

$$E \sup_{a \leq \lambda \leq b} \left\| \sqrt{n} U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n}) \right\|_p \leq \frac{C}{n^{(k-1)/2} h_n^{(p-1)/p}},$$

for some constant $C$ that depends on $k$, $p$, $a$ and $b$. The condition $nh_n^{2(p-1)/p} \to \infty$ implies that this expectation tends to zero for $2 \le k \le m$, proving (3.8) for $p \ge 2$. $\qquad\square$

PROOF OF (3.8) FOR $1 \le p < 2$. Since $f_g$ and $K^2$ are in $L_1(\mu_s)$ for some $s > d(2 - p)/p$, Lemma 1 applies and therefore Corollary 3 and inequality (3.41) hold. Now, the proof of (3.8) and subsequently Theorem 3 follow exactly as the proof of Theorem 2, but using Corollary 3 and its consequence (3.41) instead of Theorem 9 and its consequence for $\mathbf{X}_{L,n}$. These give

$$\|\mathbf{X}_{L,n}\|_{\Psi_{1/k}} \le Cn^{k/2-1}h_n^{1/2-1/p}\|L\|_{p,2,s}$$

instead of (3.48), which yields, as in the previous proof, for some constant $D$,

$$\left\| \sup_{a \le \lambda \le b} \mathbf{X}_{K_\lambda,n} \right\|_{\Psi_{1/k}}$$
$$\le Dn^{k/2-1}h_n^{1/2-1/p} \int_0^{C\|K\|_{p,2,s}} \Psi_{1/k}^{-1}\big(N\big(\mathcal{K}_{[a,b]}, d_p \vee \tilde{d}_{2,s}, \varepsilon\big)\big)\, d\varepsilon$$

instead of (3.50). Hence, for some constant $C$,

$$E \sup_{a \le \lambda \le b} \big\| \sqrt{n} U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n}) \big\|_{L_p} \le' \frac{1}{n^{k/2-1/2}h_n^{1-1/p}} E \sup_{a \le \lambda \le b} \mathbf{X}_{K_\lambda,n}$$
$$\le \frac{C\|K\|_{p,2,s}}{(nh_n)^{1/2}} \to 0$$

if $nh_n \to \infty$. $\quad\square$

REMARK 5 (Separability of the process $\mathbf{X}_{K_\lambda,n}$). In the previous subsection technically we must make sure that

$$\left\| \sup_{\lambda \in [a,b]} \mathbf{X}_{K_\lambda,n} \right\|_{\Psi_\alpha} = \sup\left\{ \left\| \max_{\lambda \in C} \mathbf{X}_{K_\lambda,n} \right\|_{\Psi_\alpha} : C \text{ finite}, C \subset [a, b] \right\}$$

in order to ensure that the entropy bound applies exactly to the process $\mathbf{X}_{K_\lambda,n}$ and not to a modification thereof. For this, by standard arguments (basically the monotone convergence theorem), it suffices that there exist $D \subset [a, b]$ countable such that

$$\text{(3.51)} \qquad \sup_{\lambda \in [a,b]} \mathbf{X}_{K_\lambda,n} = \sup_{\lambda \in D} \mathbf{X}_{K_\lambda,n} \qquad \text{a.s.}$$

Let $D$ denote the rationals in $[a, b]$. To show (3.51) it suffices to prove that, with probability 1, for each $\lambda \in [a, b]$ and any sequence $\lambda_m$ in $D$ such that $\lambda_m \to \lambda$ we have $\lim_{m\to\infty} \mathbf{X}_{K_{\lambda_m} - K_\lambda, n} = 0$. In turn to verify this it is enough to check that

$$\text{(3.52)} \qquad \lim_{m\to\infty} \left( \int_{\mathbf{R}^d} \left| \sum_{I_n^k} \pi_k \overline{K_{\lambda_m h_n} - K_{\lambda h_n}}(t, X_{i_1}, \ldots, X_{i_k}) \right|^p dt \right)^{1/p} = 0.$$

Now, each of these $L_p$ norms is bounded by a finite number of terms of the form

$$\left( \int_{\mathbf{R}^d} |E(K_{\lambda_m h_n} - K_{\lambda h_n})(t - V)|^p \, dt \right)^{1/p},$$

where $V$ is a random variable. Observe that

$$\left( \int_{\mathbf{R}^d} |E(K_{\lambda_m h_n} - K_{\lambda h_n})(t - V)|^p \, dt \right)^{1/p}$$

$$\leq \left( E \int_{\mathbf{R}^d} |(K_{\lambda_m h_n} - K_{\lambda h_n})(t - V)|^p \, dt \right)^{1/p}$$

$$= \left( E \int_{\mathbf{R}^d} \left| \frac{1}{\lambda_m h_n} K\left( \frac{t - V}{(\lambda_m h_n)^{1/d}} \right) - \frac{1}{\lambda h_n} K\left( \frac{t - V}{(\lambda h_n)^{1/d}} \right) \right|^p \, dt \right)^{1/p},$$

which by the change of variables inside the integral $u = (t - V)/(\lambda_m h_n)^{1/d}$ is equal to

$$\left( \lambda_m h_n \int_{\mathbf{R}^d} \left| \frac{1}{\lambda_m h_n} K(u) - \frac{1}{\lambda h_n} K\left( \left( \frac{\lambda_m}{\lambda} \right)^{1/d} u \right) \right|^p \, du \right)^{1/p}$$

$$= \left( \int_{\mathbf{R}^d} \left| K(u) - \frac{\lambda_m}{\lambda} K\left( \left( \frac{\lambda_m}{\lambda} \right)^{1/d} u \right) \right|^p \, du \right)^{1/p} (\lambda_m h_n)^{1/p - 1}.$$

Now since $K \in L_p(\mathbf{R}^d)$ we have $\lim_{\gamma \to 1} \int_{\mathbf{R}^d} |K(u) - K(\gamma u)|^p \, du = 0$, and this, in turn, implies that

$$\lim_{m \to \infty} \int_{\mathbf{R}^d} \left| K(u) - \frac{\lambda_m}{\lambda} K\left( \left( \frac{\lambda_m}{\lambda} \right)^{1/d} u \right) \right|^p \, du = 0,$$

which gives (3.51).

## REFERENCES

[1] AHMAD, I. A. and FAN, Y. (2001). Optimal bandwidths for kernel density estimators of functions of observations. *Statist. Probab. Lett.* **51** 245–251. MR1822731

[2] ARAUJO, A. and GINÉ, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables.* Wiley, New York. MR0576407

[3] ARCONES, M. A. (1995). A Bernstein-type inequality for $U$-statistics and $U$-processes. *Statist. Probab. Lett.* **22** 239–247. MR1323145

[4] BICKEL, P. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393. MR1065550

[5] DEHEUVELS, P. and MASON, D. M. (2004). General asymptotic confidence bands based on kernel-type function estimators. *Stat. Inference Stoch. Process.* **7** 225–277. MR2111291

[6] DE LA PEÑA, V. H. (1992). Decoupling and Khintchine's inequalities for $U$-statistics. *Ann. Probab.* **20** 1877–1892. MR1188046

[7] DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling. From Dependence to Independence.* Springer, New York. MR1666908

[8] DEVROYE, L. (1987). *A Course in Density Estimation.* Birkhäuser, Boston. MR0891874

[9] DEVROYE, L. (1992). A note on the usefulness of superkernels in density estimation. *Ann. Statist.* **20** 2037–2056. MR1193324

[10] DUDLEY, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge Univ. Press. MR1720712

[11] DUNFORD, N. and SCHWARTZ, J. T. (1988). *Linear Operators. Part I. General Theory.* Wiley, New York. MR1009162

[12] EASTWOOD, V. R. and HORVÁTH, L. (1999). Limit theorems for short distances in $R^m$. *Statist. Probab. Lett.* **45** 261–268. MR1718038

[13] EINMAHL, U. and MASON, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.* **13** 1–37. MR1744994

[14] FOLLAND, G. B. (1999). *Real Analysis*: *Modern Techniques and Their Applications*, 2nd ed. Wiley, New York. MR1681462

[15] FREES, E. W. (1994). Estimating densities of functions of observations. *J. Amer. Statist. Assoc.* **89** 517–525. MR1294078

[16] GINÉ, E. and GUILLOU, A. (2001). On consistency of kernel density estimators for randomly censored data: Rates holding uniformly over adaptive intervals. *Ann. Inst. H. Poincaré Probab. Statist.* **37** 503–522. MR1876841

[17] GINÉ, E. and KOLTCHINSKII, V. (2006). Concentration inequalities and asymptotic results for ratio type empirical processes. *Ann. Probab.* **34** 1143–1216. MR2243881

[18] GINÉ, E., LATAŁA, R. and ZINN, J. (2000). Exponential and moment inequalities for $U$-statistics. In *High Dimensional Probability*, *II* (E. Giné, D. M. Mason and J. A. Wellner, eds.) 13–38. Birkhäuser, Boston. MR1857312

[19] GINÉ, E. and MASON, D. M. (2006). Uniform in bandwidth estimation of integral functionals of the density function. Preprint.

[20] GINÉ, E. and MASON, D. M. (2007). Laws of the iterated logarithm for the local $U$-statistic process. *J. Theoret. Probab.* To appear.

[21] GINÉ, E., MASON, D. M. and ZAITSEV, A. YU. (2003). The $L_1$-norm density estimator process. *Ann. Probab.* **31** 719–768. MR1964947

[22] JAIN, N. (1977). Central limit theorem and related questions in Banach space. In *Probability* (J. L. Doob, ed.) 55–65. Amer. Math. Soc., Providence, RI. MR0451328

[23] JAMMALAMADAKA, S. R. and JANSON, S. (1986). Limit theorems for a triangular scheme of $U$-statistics with applications to inter-point distances. *Ann. Probab.* **14** 1347–1358. MR0866355

[24] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces.* Springer, Berlin. MR1102015

[25] MASON, D. M. (2004). A uniform functional law of the logarithm for the local empirical process. *Ann. Probab.* **32** 1391–1418. MR2060302

[26] NOLAN, D. and POLLARD, D. (1987). $U$-processes: Rates of convergence. *Ann. Statist.* **15** 780–799. MR0888439

[27] PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076. MR0143282

[28] PINELIS, I. F. (1995). Optimum bounds on moments of sums of independent random vectors. *Siberian Adv. Math.* **5** 141–150. MR1387858

[29] PISIER, G. and ZINN, J. (1978). On the limit theorems for random variables with values in the spaces $L_p$, $2 \le p < \infty$. *Z. Wahrsch. Verw. Gebiete* **41** 289–304. MR0471010

[30] POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York. MR0762984

[31] ROST, D. (2000). Limit theorems for smoothed empirical processes. In *High Dimensional Probability*, *II* (E. Giné, D. M. Mason and J. A. Wellner, eds.) 107–113. Birkhäuser, Boston. MR1857318

[32] SCHICK, A. and WEFELMEYER, W. (2004). Root *n* consistent density estimators for sums of independent random variables. *J. Nonparametr. Stat.* **16** 925–935. MR2094747

[33] SCHICK, A. and WEFELMEYER, W. (2006). Pointwise convergence rates and central limit theorems for kernel density estimators in linear processes. *Statist. Probab. Lett.* **76** 1756–1760. MR2274137

[34] SCHICK, A. and WEFELMEYER, W. (2007). Root-*n* consistent density estimators of convolutions in weighted $L_1$-norms. *J. Statist. Plann. Inference* **137** 1765–1774.

[35] TALAGRAND, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22** 28–76. MR1258865

[36] VAKHANIA, N. (1981). *Probability Distributions on Linear Spaces*. North-Holland, Amsterdam. MR0626346

[37] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics.* Springer, New York. MR1385671

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CONNECTICUT
196 AUDITORIUM ROAD
STORRS, CONNECTICUT 06269-3009
USA
E-MAIL: gine@math.uconn.edu

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