Laws of the Iterated Logarithm for the Local U–statistic Process

Evarist Giné* and David M. Mason†
University of Connecticut University of Delaware

Abstract

Laws of the iterated logarithm are established for the local U-statistic process. This entails the development of probability inequalities and moment bounds for U-processes that should be of separate interest. The local U-statistic process is based upon an estimator of the density of a function of several i.i.d. variables proposed by Frees (1994). As a consequence, our results are directly applicable to the derivation of exact rates of uniform in bandwidth consistency in the sup and in the L_p norms for these estimators.

Short title: Local U–statistic process LIL
AMS 2000 Subject Classifications: 60F05, 60F15, 62E20, 62G30.
Keywords: U–statistics, law of iterated logarithm, empirical process, kernel density estimation

1 Introduction and statements of results

The notion of the local empirical process has been very successful in forming the conceptual basis for recent studies of rates of consistency of kernel density estimators using modern empirical process methods. (Refer, e.g. to Deheuvels and Mason (1994, 2004), Deheuvels, Einmahl and Mason (2000), Einmahl and Mason (1997,1998, 2000, 2005), Giné and Guillou (2002) and Mason (2004).) In this paper, here is what we mean by a local empirical process. Let X, X_i, $i \in N$, be i.i.d. random variables taking values in (S, S), a measure space, with law P. Let

$$g : S \mapsto \mathbf{R}^d,\ 1 \le d < \infty,$$

be a measurable function and $K : \mathbf{R}^d \mapsto \mathbf{R}$ be an integrable function, satisfying

$$\int_{\mathbf{R}^d} K(x)dx = 1 \text{ and } 0 < \int_{\mathbf{R}^d} K^2(x)dx = ||K||_2^2 < \infty. \tag{1.1}$$

*Research partially supported by NSA Grant H98230-04-1-0075
†Research partially supported by NSA Grant MDA904–02–1–0034 and NSF Grant DMS–0503908.

For $t \in R^d$ and $\lambda \in [a,b]$, $0 < a \leq b < \infty$, we define the *local empirical process* as

$$u_n(t,\lambda) := \sqrt{n}\left\{f_n(t,\lambda h_n) - Ef_n(t,\lambda h_n)\right\},$$

where f_n is the kernel density estimator

$$f_n(t,\lambda h_n) = \frac{1}{n\lambda h_n}\sum_{i=1}^{n} K_{\lambda h_n}\left(t - g\left(X_i\right)\right) \tag{1.2}$$

and

$$K_h(\cdot) = h^{-1}K\left(\cdot\; h^{-1/d}\right),\; h > 0.$$

Recall that we say that a class of measurable P-square integrable functions $\mathcal{F}$ defined on a measure 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 ε required to cover $\mathcal{F}$, satisfies for some $A \geq 3$ and $v \geq 1$,

$$N(\mathcal{F}, L_2(Q), \varepsilon) \leq \left(\frac{A\|F\|_{L_2(Q)}}{\varepsilon}\right)^v,\quad 0 < \varepsilon \leq 2\|F\|_{L_2(Q)}, \tag{1.3}$$

for every probability measure Q on $\mathcal{S}$ for which $F \in L_2(Q)$. If (1.3) holds for $\mathcal{F}$, then we say that the VC class $\mathcal{F}$ admits the characteristics A and v. We also recall that a class of functions $\mathcal{F}$ is pointwise measurable if it contains a countable subset $\mathcal{G}$ such that for every $f \in \mathcal{F}$ there is a sequence g_m in $\mathcal{G}$ that converges pointwise to f. (See, e.g., van der Vaart and Wellner (1996).)

Introduce the class of functions formed from K,

$$\mathcal{K} = \left\{K\left(t - \cdot\gamma\right) : \gamma \geq 1, t \in \mathbf{R}^d\right\}. \tag{1.4}$$

Assume that

(F.i) $\mathcal{K}$ is a bounded point–wise measurable class.

(F.ii) $\mathcal{K}$ is VC for some for some $A \geq 3$ and $v \geq 1$.

Also assume that K has support in $[-1/2, 1/2]^d$; the density f_g of $g(X)$ is uniformly continuous on $\mathbf{R}^d$; and $\{h_n\}_{n\geq 1}$ converges to zero at the rate:

(H.i) $h_n \searrow 0$, $nh_n \nearrow \infty$; (H.ii) $nh_n/\log(1/h_n) \to \infty$;

(H.iii) $\log(1/h_n)/\log\log n \to \infty$.

Under these conditions, Theorem 1 of Mason (2004) implies the following *law of the logarithm*:

$$\lim_{n\to\infty}\sup_{a\leq\lambda\leq b}\sup_{t\in\mathbf{R}^d}\frac{\sqrt{\lambda h_n}\,|u_n(t,\lambda)|}{\sqrt{2\log(1/h_n)}} = ||K||_2 \sup_{t\in\mathbf{R}^d}\sqrt{f_g(t)},\quad \text{a.s.} \tag{1.5}$$

(Here and in what follows, $\log a$ means $\log(a \vee e)$.) This is a *uniform in bandwidth* version of a uniform consistency result of Giné and Guillou (2002) for the kernel density estimator $f_n(t, h_n)$.

Now let us generalize this setup. Given $m \geq 1$, let

$$g : S^m \mapsto \mathbf{R}^d, 1 \leq d < \infty,$$

be a measurable function. For $t \in R^d$ and $\lambda \in [a, b]$, $0 < a \leq b < \infty$, we define the *local U–statistic*

$$U_n(t, \lambda) := \frac{(n-m)!}{n!} \sum_{\mathbf{i} \in I_n^m} K_{\lambda h_n}(t - g(X_{i_1}, \ldots, X_{i_m})), \tag{1.6}$$

with

$$I_n^m = \{\mathbf{i} = (i_1, \ldots, i_m) : 1 \leq i_j \leq n, i_j \neq i_k \text{ if } j \neq k\}. \tag{1.7}$$

Notice that in the special case when $m = \lambda = 1$ and $g(X)$ has a density f_g,

$$U_n(t, 1) = f_n(t, h_n) = (nh_n)^{-1} \sum_{i=1}^n K\left(h_n^{-1/d}(t - g(X_i))\right), \tag{1.8}$$

which is the usual kernel density estimator of $f_g(t)$. When $m \geq 2$, assuming that g is symmetric in its m components and $g(X_1, \ldots, X_m)$ has a density $f_g(t)$ and conditional densities given $X_i = x$, Frees (1994) discovered the perhaps surprising fact that $U_n(t, 1)$ can be used to estimate $f_g(t)$ at each fixed t at the in probability rate of $n^{-1/2}$. (For recent closely related work, see Schick and Wefelmeyer (2004).)

Define the *local U–statistic process* for $t \in \mathbf{R}^d$, $\lambda \in [a, b]$, $0 < a \leq b < \infty$, to be

$$u_n(t, \lambda) := \sqrt{n}\left(U_n(t, \lambda) - EK_{\lambda h_n}(t - g(X_1, \ldots, X_m))\right). \tag{1.9}$$

When $m = 1$ it is a routine exercise to prove that, subject to smoothness conditions on f_g, for each $t \in \mathbf{R}$ and $\lambda \in [a, b]$,

$$\sqrt{\lambda h_n} u_n(t, \lambda) = \sqrt{n \lambda h_n}\{f_n(t, \lambda h_n) - Ef_n(t, \lambda h_n)\} \to_d N\left(0, ||K||_2^2 f_g(t)\right),$$

whereas for any choice of $t_1 \neq t_2$ and $\lambda_1, \lambda_2 \in [a, b]$ the random variables

$$\sqrt{\lambda_1 h_n} u_n(t_1, \lambda_1) \text{ and } \sqrt{\lambda_2 h_n} u_n(t_2, \lambda_2)$$

are asymptotically independent. This means that $\sqrt{\lambda h_n} u_n(t, \lambda)$ cannot converge weakly to a continuous bounded process on any compact subset of $\mathbf{R}$ with non–empty interior. The asymptotic distributional behavior of $u_n(t, \lambda)$ changes radically when $m \geq 2$. In this case, Giné and Mason (2005) have recently shown that central limit theorems [CLT] hold for $u_n(t, \lambda)$ in $L_p\left(\mathbf{R}^d\right)$, $1 \leq p \leq \infty$.

Our aim here is to derive compact laws of the iterated logarithm [LIL] in $L_p\left(\mathbf{R}^d\right)$, $1 \leq p \leq \infty$ for $u_n(t, \lambda)$ in the case when $m \geq 2$. To be more specific we shall determine conditions under which $u_n(t, \lambda)$ satisfies "the compact LIL in $L_p(\mathbf{R}^d)$, uniformly in $a \leq \lambda \leq b$", in the following sense.

Definition 1 *Let X, X_i, $i \in N$, be i.i.d. with common law P such that for some $1 \le p \le \infty$, for each $i = 1, \ldots, m$, the random variable $g(X_1, \ldots, X_m)$, conditionally on $X_i = x$, has a density $f_i(t, x)$, jointly measurable in t and x, satisfying for each $x \in \mathbf{R}^d$, $f_i(\cdot, x) \in L_p\left(\mathbf{R}^d\right)$. The processes $u_n(t, \lambda)$ satisfy the compact LIL in $L_p(\mathbf{R}^d)$, uniformly in $a \le \lambda \le b$, $0 < a \le b < \infty$, if*

$$\frac{\sup_{a \le \lambda \le b} \|u_n(t, \lambda) - \sqrt{n}(P_n - P) \sum_{i=1}^m f_i(t, \cdot)\|_p}{\sqrt{\log \log n}} \to 0, \text{ a.s.}^* \tag{1.10}$$

and the processes

$$\sqrt{n}(P_n - P) \sum_{i=1}^m f_i(t, \cdot), \; t \in \mathbf{R}^d, \tag{1.11}$$

satisfy the compact LIL in $L_p(\mathbf{R}^d)$ if $1 \le p < \infty$, and the class of functions $\left\{\sum_{i=1}^m f_i(t, \cdot) : t \in \mathbf{R}^d\right\}$ is P-separable and satisfies the compact LIL for P if $p = \infty$. (We shall call the $p = \infty$ case the compact LIL in $L_\infty(\mathbf{R}^d)$.)

Here and elsewhere $\|g\|_\infty = \sup_{t \in \mathbf{R}^d} |g(t)|$, $\|g\|_p$ denotes the L_p-norm with respect to Lebesgue measure of $\mathbf{R}^d$, $1 < p < \infty$, and a.s.* means almost sure convergence of the measurable envelopes (almost uniform convergence in Dudley (1999)). We are using, here and elsewhere, the standard empirical process notation P_n for the empirical measure based on the sample $X_1, \ldots, X_n$, that is, $P_n(f) = \sum_{i=1}^n f(X_i)/n$, and the operator notation $Pf = \int f dP = Ef(X)$. For the definitions of the compact LIL and P-separable classes of functions see Section 2.4 below.

Note that if $u_n(t, \lambda)$ satisfies the compact LIL in $L_p(\mathbf{R}^d)$, then in particular,

$$\limsup_{n \to \infty} \sup_{\lambda \in [a,b]} \frac{\|u_n(t, \lambda)\|_p}{\sqrt{2 \log \log n}}$$

is a.s. a finite constant determined by the random process $\sum_{i=1}^m f_i(t, X)$. Specializing to the sup-norm case, we get for $m \ge 2$,

$$\limsup_{n \to \infty} \sup_{a \le \lambda \le b} \sup_{t \in \mathbf{R}^d} \frac{|u_n(t, \lambda)|}{\sqrt{2 \log \log n}} = \sup_{t \in \mathbf{R}^d} \sigma_g(t), \text{ a.s.}$$

where $\sigma_g^2(t) = Var\left(\sum_{i=1}^m f_i(t, X)\right)$. This result sharply delineates the difference between the almost sure limiting behavior as $n \to \infty$ of the local U-statistic when $m \ge 2$ and when $m = 1$, as given by the law of the logarithm in (1.5). We have *law of the logarithm* behavior when $m = 1$ and *law of the iterated logarithm* behavior when $m \ge 2$.

Surprisingly, though the CLT in L_p, $1 \le p < \infty$, for the local empirical process $u_n(t, \lambda)$, $m = 1$, is well–understood (see Giné, Mason and Zaitsev (2003) and the references therein) the establishment of the corresponding LIL remains elusive. This was first pointed out as an outstanding problem in Devroye and Györfi (1985) for the L_1 case. It remains open to this day except for the special case $p = 2$ (Giné and Mason (2004)).

1.1 Main results

All the results on the local U–statistic process $u_n(t,\lambda)$ in this article require the following assumptions:

(CD) for each $i = 1, \dots, m$, the random variable $g(X_1, \dots, X_m)$, conditionally on $X_i = x$, has a density $f_i(t,x) = f_i(t,x,g)$, $t \in \mathbf{R}^d$, $x \in S$, with respect to Lebesgue measure, which is jointly measurable in t and x;

(KI) the kernel $K : \mathbf{R}^d \mapsto \mathbf{R}$ is in $L_1(\mathbf{R}^d)$ and $\int_{\mathbf{R}^d} K(t_1, \dots, t_d) dt_1 \cdots dt_d = 1$;

(HC) the sequence $\{h_n\}$ of positive numbers is *c-regular* for some $c > 1$, that is

$$c^{-1} h_n \le h_{2n} \le c h_n, \text{ for all } n.$$

If (CD) holds, then in particular the random variable $g(X_1, \dots, X_m)$ has a density f_g with respect to Lebsegue measure, for instance, defined as

$$f_g(t) = \sum_{i=1}^{m} E f_i(t, X, g)/m. \tag{1.12}$$

Note also that condition (CD) implies $d < m$: for instance, under regularity conditions, for $d \ge m$ and x fixed, the equation $y = g(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_m)$ defines a hypersurface of $\mathbf{R}^d$ of dimension at most $m-1$, which precludes the existence of the density $f_i(t,x,d)$.

Theorem 1 *Let X, X_i, $i \in \mathbf{N}$, be i.i.d. S-valued random variables with law P, let $g : S^m \mapsto \mathbf{R}^d$ be a measurable function satisfying (CD), let $K : \mathbf{R}^d \mapsto \mathbf{R}$ be a bounded kernel satisfying (KI), and let $\{h_n\}$ be a sequence satisfying (HC). Also assume:*

a) The class of functions $\mathcal{K}$ defined by (1.4) is VC type for a bounded envelope F, with characteristics $A \ge 3$ and $v \ge 1$;

b) the density f_g is bounded and, for every $i = 1, \dots, m$, the class of functions $\mathcal{F}_i := \{f_i(t, \cdot) : t \in \mathbf{R}^d\}$ is P-separable and P-Donsker,

$$E F_i^2(X) < \infty, \text{ where } F_i(x) = \sup_{t \in \mathbf{R}^d} f_i(t,x),$$

and the identity map $(\mathbf{R}^d, |\cdot|) \mapsto (\mathbf{R}^d, \rho_i)$ is uniformly continuous, where

$$\rho_i^2(u,v) = \mathrm{Var}(f_i(u,X) - f_i(v,X));$$

c) $h_n \to 0$ and $n h_n \log\log n / (\log\log n \vee \log(1/h_n))^2 \to \infty$.

Then the processes $u_n(t,\lambda)$ satisfy the compact LIL in $L_\infty(\mathbf{R}^d)$ uniformly in $a \le \lambda \le b$.

By "compact LIL in $L_\infty(\mathbf{R}^d)$ uniformly in $a \le \lambda \le b$" we mean the case $p = \infty$ in Definition 1. For the definition of P-Donsker classes of functions (classes o functions $\mathcal{F}$ for which the empirical process satisfies the central limit theorem uniformly in $f \in \mathcal{F}$) we refer to Dudley (1999).

Remark Condition c) should be compared to the assumptions on $\{h_n\}$ for the CLT to hold for $u_n(t, \lambda)$ in Theorem 1 of Giné and Mason (2005), namely that $h_n \to 0$ and

$$nh_n / (\log(1/h_n))^2 \to \infty.$$

The LIL in the L_p norm uniform in $a \le \lambda \le b$ requires a condition on the kernel K that depends on Young moduli Ψ_α of exponential type with power exponent of order $0 < \alpha \le 1$. As in de la Peña and Giné (1999), p. 188, $\Psi_1(x) := e^x - 1$, but if $\alpha < 1$, since e^{x^α} is only convex for $x \ge x_\alpha := ((1-\alpha)/\alpha)^{1/\alpha}$, we take as Ψ_α a function that is 0 at 0, convex and increasing, and of the order of e^{x^α} for x large. It is defined to be

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

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 $0 \le x \le x_\alpha$. Note that $\Psi_\alpha^{-1}(u)$ is a constant times u for $0 \le u \le \Psi_\alpha(x_\alpha)$ and it is the $1/\alpha$-th power of the logarithm of $u + \alpha \exp((1-\alpha)/\alpha)$ for $u > \Psi_\alpha(x_\alpha)$. We also recall that then, for a non-negative random variable ξ for which

$$Ee^{(\xi/a)^\alpha} < \infty \text{ for some } a > 0,$$

the Ψ_α-Orlicz norm is defined as

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

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

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

for a constant C that depends only on α and b, and that conversely,

$$\|\xi\|_{\Psi_\alpha} \le C \quad \Rightarrow \quad \Pr\{\xi > x\} \le b \exp\{-a(x/C)^\alpha\}$$

for constants a and b that depend only on α. We will use these implications.

Given $p \in [1, \infty)$, set

$$d_p^p(f, g) := \int_{\mathbf{R}^d} |f(y) - g(y)|^p dy,$$

and, given $K : \mathbf{R}^d \mapsto \mathbf{R}$ and $0 < A < B < \infty$, set

$$\mathcal{K}_{[A,B]} := \{K_h : h \in [A, B] \cup \{0\}\},$$

with K_0 defined as $K_0 = 0$. For $\alpha > 0$ and $0 < A \le B < \infty$, we introduce the condition that will be needed to treat the $L_p(\mathbf{R}^d)$ case when $p \ge 2$.

(KI(p,α, A, B)) the kernel $K:\mathbf{R}^d\mapsto\mathbf{R}$ is in $L_p(\mathbf{R}^d)$ and

$$\int_0^1 \Psi_\alpha^{-1}(N(\mathcal{K}_{[A,B]},d_p,\tau))d\tau<\infty,$$

where $N(\mathcal{K}_{[A,B]},d_p,\tau)$ is the τ-covering number of the (pseudo) metric space $(\mathcal{K}_{[A,B]},d_p)$.

Theorem 2 *Let $2\le p<\infty$. Let X,X_i, $i\in\mathbf{N}$, be i.i.d. S-valued random variables with law P, let $g:S^m\mapsto\mathbf{R}^d$ be a measurable function satisfying (CD), let $K:\mathbf{R}^d\mapsto\mathbf{R}$ be a kernel satisfying (KI), let $\{h_n\}$ be a sequence satisfying (HC) and let $0<a\le b<\infty$. Also assume:*

a) The conditional densities $f_i(t,x)$ of the random variable $g(X_1,\dots,X_m)$ given $X_i=x$, satisfy

$$E\biggl(\int_{\mathbf{R}^d}|f_i(t,X)-Ef_i(t,X)|^p dt\biggr)^{2/p}<\infty,$$

b) the kernel K satisfies condition (KI(p,2/m,$\mathbb{A}$,$\mathbb{B}$)) for $\mathbb{A}=a/c$ and $\mathbb{B}=bc$, where c is the constant in condition (HC).

c) $h_n\to 0$ and

$$\frac{nh_n^{2(p-1)/p}}{\log\log n}\to\infty.$$

Then the processes $u_n(t,\lambda)$ satisfy the compact LIL in $L_p(\mathbf{R}^d)$ uniformly in $a\le\lambda\le b$.

For $1\le p<2$ 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 $s>0$, $p\ge 1$, we define the Borel measure μ_s on $\mathbf{R}^d$, $L_p(\mu_s)$ and $\tilde d_{p,s}$ as

$$d\mu_s(t)=(1+|t|)^s dt,\quad L_p(\mu_s)=L_p(\mathbf{R}^d,\mathcal{B},\mu_s),\quad \tilde d_{p,s}(f,g)=\|f-g\|_{L_p(\mu_s)},\quad (1.13)$$

Theorem 3 *Let $1\le p<2$. Let X,X_i, $i\in\mathbf{N}$, be i.i.d. S-valued random variables with law P, let $g:S^m\mapsto\mathbf{R}^d$ be a measurable function satisfying (CD), let $K:\mathbf{R}^d\mapsto\mathbf{R}$ be a kernel satisfying (KI), let $\{h_n\}$ be a sequence satisfying (HC) and let $0<a\le b<\infty$. Also assume:*

a) the conditional densities $f_i(t,x)$, $i=1,\dots,m$, satisfy

$$\int_{\mathbf{R}^d}\bigl[E(f_i(t,X)-Ef_i(t,X))^2\bigr]^{p/2}dt<\infty.$$

b) K^2 and f_g are in $L_1(\mu_s)$ for some $s>d(2-p)/p$.

c) The kernel K is in $L_p(\mathbf{R}^d)$ and

$$\int_0^1 \Psi_{1/m}^{-1}(N(\mathcal{K}_{[\mathbb{A},\mathbb{B}]}, d_p \vee \tilde{d}_{2,s}, \varepsilon))d\varepsilon < \infty$$

for $\mathbb{A} = a/c$ and $\mathbb{B} = bc$, where c is the constant in condition (HC).

d) $h_n \to 0$ and

$$\frac{nh_n}{\log\log n} \to \infty.$$

Then the processes $u_n(t, \lambda)$ satisfy the compact LIL in $L_p(\mathbf{R}^d)$ uniformly in $a \leq \lambda \leq b$.

The most commonly used kernels satisfy condition (KI(p,α, A, B)), in particular this condition holds (for all α, A, B) for any kernel of the form $K(x) = \Phi(|x|)$ where Φ is of bounded variation on $[0, \infty)$ and such that its positive and negative variations P and N satisfy $\int_0^\infty r^{d-1}|P(r)|dr < \infty$ and $\int_0^\infty r^{d-1}|N(r)|dr < \infty$. If K also has bounded support or decreases exponentially in a positive power of $|x|$, then condition c) of Theorem 3 holds as well. (Ref.: Giné and Mason (2005).)

Using the above LIL's one can now readily transform all the CLT examples discussed in Giné and Mason (2005) into LILs. We leave this to the interested reader.

2 Main steps for proofs

Our main tool to analyze $u_n(t, \lambda)$ will be the Hoeffding decomposition, which we recall here for ready access and reference.

2.1 Hoeffding decomposition

Let L be a function of m variables, symmetric in its entries. Then, for $1 \leq k \leq m$, the Hoeffding projections with respect to P are defined as

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

with $\pi_0 L = EL(X_1, \ldots, X_m)$. For more details refer to de la Peña and Giné (1999). The Hoeffding decomposition states the following, which is easy to check:

$$U_n^{(m)}(L) - EL = \sum_{k=1}^{m} \binom{m}{k} U_n^{(k)}(\pi_k L), \tag{2.1}$$

where for a kernel L of k variables, $1 \leq k \leq m$, we set

$$U_n^{(k)}(L) = \frac{(n-k)!}{n!} \sum_{\mathbf{i} \in I_n^k} L(X_{i_1}, \ldots, X_{i_k}). \tag{2.2}$$

Assuming L is in $L_2(P^m)$, this is an orthogonal decomposition and

$$E\left(\pi_k L \,|X_2, \ldots, X_k\right) = 0 \text{ for } k \geq 1,$$

that is, the kernels $\pi_k L$ are canonical for P (or completely degenerate, or completely centered). Also, π_k, $k \geq 1$, are nested projections, that is

$$\pi_k \circ \pi_\ell = \pi_k \text{ if } k \leq \ell.$$

2.2 Smoothed empirical process

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

$$\overline{K}_h(t, x_1, \ldots, x_m) := \frac{1}{m!} \sum_{\sigma \in \rho_m} K_h(t - g(x_{\sigma(1)}, \ldots, x_{\sigma(m)})),$$

where ρ_m are the permutations of $1, \ldots, m$. Then, clearly, for each $t \in \mathbf{R}^d$,

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

(Here the superscript designates the number of "dot" variables.) Moreover, we get

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

It turns out that the first term of this expansion,

$$\sqrt{n} m U_n^{(1)}(\pi_1 \overline{K}_{\lambda h_n}(t, \cdot) = \frac{m}{\sqrt{n}} \sum_{i=1}^{n} \pi_1 \overline{K}_{\lambda h_n}(t, X_i),$$

is a smoothed empirical process (as studied by Yukich (1992), van der Vaart (1994) and Rost (1999)). We shall soon see that it determines the LIL of $u_n(t, \lambda)$. Our first step towards showing this is through the following proposition.

2.3 A general proposition for the LIL

Proposition 1 *Assume that X, X_i, $i \in N$, are i.i.d. with common law P such that for some $1 \leq p \leq \infty$, for each $i = 1, \ldots, m$, the random variable $g(X_1, \ldots, X_m)$, conditionally on $X_i = x$, has a density $f_i(t, x)$, jointly measurable in t and x, satisfying for each $x \in R^d$, $f_i(\cdot, x) \in L_p\left(\mathbf{R}^d\right)$. Also assume*

$$\lim_{\delta \searrow 0} \limsup_{n} \sup_{|u| \leq \delta} \frac{\|\sqrt{n}(P_n - P)(f_i(t-u, \cdot) - f_i(t, \cdot))\|_p}{\sqrt{\log \log n}} = 0, \ \ a.s.; \tag{2.3}$$

$$\limsup_n \frac{\|\sqrt{n}(P_n - P)f_i(t,\cdot)\|_p}{\sqrt{\log\log n}} < \infty, \ \ a.s. \tag{2.4}$$

Then whenever $h_n \searrow 0$, we have

$$\frac{\sup_{a\le\lambda\le b}\left\|\sqrt{n}\left(mU_n^{(1)}(\pi_1\overline{K}_{\lambda h_n}(t,\cdot)\right) - \sqrt{n}(P_n-P)\sum_{i=1}^m f_i(t,\cdot)\right\|_p}{\sqrt{\log\log n}} \to 0, \ a.s. \tag{2.5}$$

Note that in the above statements (and elsewhere), the L_p norm, $1 \le p \le \infty$, is with respect to the t variable, and that, with some abuse of notion, as indicated earlier, the L_∞ norm is the sup norm. The proof of Proposition 1 follows the same lines as that of Proposition 1 in Giné and Mason (2005) and therefore will be omitted.

2.4 Compact LIL

In this section we clarify (1.11) (the compact LIL part) of Definition 1. First we consider the compact LIL for the empirical process indexed by a class of functions $\mathcal{F}$. Assume that the class $\mathcal{F}$ is separable for P (P-separable) in the following sense:

Definition 2 *A class $\mathcal{F}$ is separable for P if, for each n, the process $(P_n - P)f$, $f \in \mathcal{F}$, is separable. (This means that there exists a countable set $\mathcal{F}_0 \subseteq \mathcal{F}$ such that for each f in $\mathcal{F}$,*

$$(P_n - P)f \in \overline{\{(P_n-P)g : g \in \mathcal{F}_0, \|f-g\|_{L_2(P)} \le \varepsilon\}},$$

for every $\varepsilon > 0$, where $\overline{A}$ denotes the closure of a set A.)

We say that a P-separable class of functions $\mathcal{F}$ satisfies the compact LIL for P, whenever the sequence

$$\left\{\frac{\sqrt{n}}{\sqrt{2\log\log n}}(P_n - P)f : f \in \mathcal{F}\right\}_{n=1}^{\infty}$$

is almost surely relatively compact in $\ell^\infty(\mathcal{F})$ with set of limit points

$$\mathcal{H} = \left\{f \mapsto E\big[(f(X) - Pf)h(X)\big] : Eh^2(X) \le 1\right\}. \tag{2.6}$$

Note that, in particular, if $\mathcal{F}$ satisfies the compact LIL for P, then

$$\limsup_{n\to\infty}\sup_{f\in\mathcal{F}}\left|\frac{\sqrt{n}}{\sqrt{2\log\log n}}(P_n-P)f\right| = \sup_{f\in\mathcal{F}}\big(\mathrm{Var}(f(X))\big)^{1/2}, \ \text{a.s.} \tag{2.7}$$

Turning to the compact LIL in $L_p(\mathbf{R}^d)$, $1 \le p < \infty$, let ξ, ξ_i, $i \in \mathbf{N}$, be i.i.d. L_p-valued random variables such that $E\|\xi\|_{L_p} < \infty$. In this situation we say that ξ satisfies the compact LIL if almost surely the sequence

$$\left\{ \frac{\sum_{i=1}^{n}(\xi_i - E\xi_i)}{\sqrt{2n\log\log n}} \right\}_{n=1}^{\infty}$$

is relatively compact in L_p and has as set of limit points

$$\mathcal{H} = \Bigg\{ f \in L_p(\mathbf{R}^d) : \int_{\mathbf{R}^d} f(y)h(y)dy$$

$$\le \left(E\Big(\int_{\mathbf{R}^d} (\xi(y) - E\xi(y))h(y)dy \Big)^2 \right)^{1/2} \text{ for all } h \in L_q(\mathbf{R}^d) \Bigg\}, \tag{2.8}$$

where q is the conjugate of p, with $q = \infty$ if $p = 1$. If this happens, then $\mathcal{H}$ is a compact subset of L_p. In particular, if ξ satisfies the LIL in L_p, then

$$\limsup_{n} \frac{\left\| \sum_{i=1}^{n}(\xi_i - E\xi_i) \right\|_p}{\sqrt{2n\log\log n}} = \sup_{\|h\|_q \le 1} \left(E\Big(\int_{\mathbf{R}^d} (\xi(y) - E\xi(y))h(y)dy \Big)^2 \right)^{1/2}. \tag{2.9}$$

See e.g., Ledoux and Talagrand (1991), pp. 208-210.

2.5 In summary

It is clear by Proposition 1 and the Hoeffding decomposition that to determine conditions under which $u_n(\cdot, \lambda)$ considered as a process taking values in $L_p\left(\mathbf{R}^d\right)$ indexed by $\lambda \in [a, b]$ obeys the compact LIL uniformly in $a \le \lambda \le b$ in the sense of Definition 1, that is (1.10) and (1.11) hold, it suffices to impose conditions so that simultaneously, for $k = 2, \ldots, m$,

$$\sup_{a \le \lambda \le b} \frac{\sqrt{n} \left\| U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n}(t, \cdot, \ldots, \cdot)) \right\|_p}{\sqrt{\log\log n}} \to 0, \text{ a.s.} \tag{2.10}$$

and those needed for Proposition 1 and the compact LIL are in effect.

3 Proof of main results

3.1 A maximal inequality

For each $n \in \mathbf{N}$, set for $n \ge k$,

$$S_n(H) = \left\| \sum_{\mathbf{i} \in I_n^k} H(X_{i_1}, \ldots, X_{i_k}) \right\|, \tag{3.1}$$

where X_i are i.i.d. with probability law P, and H is P-canonical, meaning that for all $x_1, \dots, x_k \in S$ and $1 \le r \le k$,

$$EH(x_1, \dots, x_{r-1}, X_1, x_{r+1}, \dots, x_k) = 0, \tag{3.2}$$

with an obvious modification of (3.2) for $r = 1$ or $r = k$. Here $\|\cdot\|$ refers to the norm in a separable Banach space $\mathbf{B}$ if H is $\mathbf{B}$-valued, or to the sup over $f \in \mathcal{F}$ of $\left|\sum_{\mathbf{i}} f(X_{i_1}, \dots, X_{i_k})\right|$ if $H = \{f : f \in \mathcal{F}\}$, and in this last case, to avoid measurability problems, we assume $\mathcal{F}$ is P-separable (in the sense of Definition 2 with $P_n - P$ replaced by the multiple sums). Assume $\|H\|$ is integrable. For each $n \in \mathbf{N}$, let $\mathcal{A}_n$ denote the σ-field generated by $X_1, \dots, X_n$. We recall that, by convexity of $\|\cdot\|$, if a is a constant and $EZ = 0$, then

$$E\|a + Z\| \ge \|a + EZ\| = \|a\|.$$

Hence, by (3.2), we have, letting E_{n+1} denote integration with respect to the variable X_{n+1} only,

$$\begin{aligned} &E\big(S_{n+1}(H)\big|\mathcal{A}_n\big) \\ &= E_{n+1}\Bigg\| \sum_{\mathbf{i}\in I_n^k} H(X_{i_1}, \dots, X_{i_k}) + \sum_{\mathbf{i}\in I_n^{k-1}} \big(H(X_{n+1}, X_{i_1}, \dots, X_{i_{k-1}}) \\ &\quad + H(X_{i_1}, X_{n+1}, X_{i_2}, \dots, X_{i_{k-1}}) + \cdots + H(X_{i_1}, \dots, X_{i_{k-1}}, X_{n+1})\big)\Bigg\| \ge S_n(H). \end{aligned}$$

That is, the sequence $(S_n(H), \mathcal{A}_n)$, $n \ge k$, is a submartingale. (From now on to ease notation we shall write $S_n = S_n(H)$.) Then, by an inequality of Brown (1971) (see, e.g., Shorack and Wellner (1986), p. 870), we have: for any $x > 0$ and $0 < c < 1$,

$$\Pr\left\{\max_{k\le m\le n} S_m > x\right\} \le \frac{\int_{\{S_n > cx\}} S_n dP}{x(1-c)}.$$

By Hölder's inequality, this immediately gives the following inequality, which will be crucial for our proofs.

Theorem 4 *Let X_i be i.i.d. S-valued with probability law P. Let $H : S^k \mapsto \mathbf{B}$, (resp. let $H = \{f : f \in \mathcal{F}\}$, where $\mathcal{F}$ is a P-separable collection of measurable functions $f : S^k \mapsto \mathbf{R}$), and let $\|\cdot\|$ denote the $\mathbf{B}$-norm (resp. let $\|\cdot\|$ denote* sup *over $f \in \mathcal{F}$ of the absolute value). Assume H is P-canonical (which, in the case of functions means that every f in $\mathcal{F}$ is P-canonical). Further assume that $E\|H(X_1, \dots, X_k)\|^r < \infty$ for some $r > 1$. Let s be the conjugate of r. Then, with S_n, $n \ge k$, as defined in (3.1), we have for all $x > 0$,*

$$\Pr\left\{\max_{k\le m\le n} S_m > x\right\} \le \frac{\left(\Pr\{S_n > cx\}\right)^{1/s} \left(ES_n^r\right)^{1/r}}{x(1-c)}. \tag{3.3}$$

To make this maximal inequality useful we need good bounds for $\Pr\{S_n > cx\}$ and ES_n^r.

3.2 Proof of Theorem 1

We begin by showing that the assumptions of Theorem 1 imply that the compact LIL part of Definition 1, namely (1.11) holds. From this we shall infer that assumptions (2.3) and (2.4) of Proposition 1 hold, which imply (2.5).

For the compact LIL in the sup-norm, first let us recall a LIL for empirical processes proved by Ledoux and Talagrand (1988) in separable Banach spaces and stated in the language of empirical processes in Theorem 9 on p. 609 of Ledoux and Talagrand (1989). Let $\mathcal{F}$ be a P-separable class of functions in the sense of Definition 2.

In this situation, a pointwise separable class $\mathcal{F} \subset L_2(P)$ such that $\sup_{f\in\mathcal{F}} |Pf| < \infty$ satisfies the compact LIL for P if and only if

a) $\mathcal{F}$ is totally bounded in $L_2(P)$,

b) $E(F^2/\log\log F) < \infty$ where $F = \sup_{f\in\mathcal{F}} |f|$, and

c) $\sup_{f\in\mathcal{F}} \left|\frac{\sqrt{n}(P_n-P)f}{\sqrt{\log\log n}}\right| \to 0$ in probability.

In particular, assuming separability, if $PF^2 < \infty$ and $\mathcal{F}$ is P-Donsker then $\mathcal{F}$ satisfies the compact LIL (since, $\mathcal{F}$ being P-Donsker implies both that $\mathcal{F}$ is totally bounded in $L_2(P)$ and that the sequence $\sup_{f\in\mathcal{F}} |\sqrt{n}(P_n - P)f|$ is stochastically bounded).

Since we assume that the classes of functions $\mathcal{F}_i = \left\{f_i(t,\cdot) : t \in \mathbf{R}^d\right\}$ are separable and P-Donsker for the law P of X, and

$$EF_i^2(X) < \infty, \text{ where } F_i(x) = \sup_{t\in\mathbf{R}^d} f_i(t,x),$$

we readily conclude that the class

$$\mathcal{F} = \left\{\sum_{i=1}^{m} f_i(y,\cdot) : y \in \mathbf{R}^d\right\}$$

is separable for P because the classes $\mathcal{F}_i$ are, and that $PF^2 < \infty$, where $F = \sum_{i=1}^m F_i$. Also, $\mathcal{F}$ is P-Donsker by Theorem 3.8.1, p. 121, and exercise 6, p. 127, in Dudley (1999). Hence, $\mathcal{F}$ satisfies the LIL for P by the Ledoux-Talagrand theorem and thus the compact LIL part of Definition 1 holds.

Next since for each $i = 1, \ldots, m$, the class $\mathcal{F}_i$ satisfies the LIL for P, we have

$$\limsup_n \sqrt{\frac{n}{\log\log n}} \|(P_n - P)f_i(y,\cdot)\|_\infty = \sqrt{\sup_y \mathrm{Var}_P(f_i(y,X))} < \infty \quad \text{a.s.} \tag{3.4}$$

Further, applying the Ledoux-Talagrand theorem again, we see that the class

$$\{f_i(u,\cdot) - f_i(v,\cdot) : |u - v| \le \delta\}$$

also satisfies the LIL for P. Therefore we have

$$\limsup_n \sqrt{\frac{n}{\log\log n}} \sup_{|u-v|\le\delta} |(P_n - P)(f_i(u,\cdot) - f_i(v,\cdot))| = \omega(\delta) \quad \text{a.s.,} \tag{3.5}$$

where

$$\omega(\delta) = \sup_{|u-v|\le\delta} \sqrt{\mathrm{Var}(f_i(u,X) - f_i(v,X))}.$$

Clearly now, since for each $i = 1, \ldots, m$, we have (3.4) and (3.5), and taking into account that by hypothesis b) of Theorem 1, $\omega(\delta) \to 0$ as $\delta \to 0$, we see that assumptions (2.3) and (2.4) of Proposition 1 hold, which implies (2.5).

Putting everything together we conclude that (1.11) of Definition 1 and (2.5) are fulfilled in the $L_\infty\left(\mathbf{R}^d\right)$ case. We next turn to verifying (2.10), which will complete the proof of Theorem 1. To accomplish this we require two main ingredients: an exponential inequality and a moment bound.

3.2.1 Ingredient 1: Major's inequality

Major (2005) recently proved the following exponential inequality:

Theorem 5 *(Major (2005)) Let X_i be i.i.d. S valued with probability P and $\mathcal{H}_k$ be a VC type class of real functions of k variables from S, symmetric in their entries, that is separable for P and consisting of P-canonical functions H bounded by 1, and let*

$$\sigma^2 \ge \sup_{H\in\mathcal{H}_k} EH^2(X_1, \ldots, X_k).$$

Then, there exist constants C_1, C_2, C_3 depending only on k and the characteristics of the class (A and v in (1.3)), such that

$$\Pr\left\{\sup_{H\in\mathcal{H}_k}\left|\frac{1}{n^{k/2}}\sum_{\mathbf{i}\in I_n^k} H((X_{i_1}, \ldots, X_{i_k})\right| > y\right\} \le C_1 \exp\left(-C_2\left(\frac{y}{\sigma}\right)^{2/k}\right) \qquad (3.6)$$

for all y such that

$$n\sigma^2 \ge \left(\frac{y}{\sigma}\right)^{2/k} \ge C_3 \log\left(\frac{2}{\sigma}\right). \qquad (3.7)$$

3.2.2 Ingredient 2: Moment bounds

For any functional Ψ defined on a class of functions $\mathcal{F}$ set

$$\|\Psi(f)\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |\Psi(f)|.$$

We begin by recording the following moment bound of Giné and Mason (2005) for $E\|U_n^{(k)}(\pi_k f)\|_{\mathcal{F}}$, when $\mathcal{F}$ is of VC type.

Theorem 6 *Let $\mathcal{F}$ be a collection of measurable 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, that without loss of generality we take so that $A \geq e^m$ and $v \geq 1$. Then there exist constants $D_1 := D_1(m, A, v, M)$ and $D_2 = D_2(m, A, v, M)$ such that,*

$$n^{k/2} E \left\| U_n^{(k)}(\pi_k f) \right\|_{\mathcal{F}} \leq D_1 \sigma \left[\log \left(\frac{A \|F\|_{L_2(P^m)}}{\sigma} \right) \right]^{k/2}, \ k = 0, 1, \ldots, m, \tag{3.8}$$

assuming

$$n\sigma^2 \geq D_2 \log \left(\frac{2\|F\|_{L_2(P^m)}}{\sigma} \right), \tag{3.9}$$

where σ^2 is any number satisfying

$$\|P^m f^2\|_{\mathcal{F}} \leq \sigma^2 \leq P^m F^2. \tag{3.10}$$

We shall need the following corollary to Theorem 6.

Corollary 1 *Let $\mathcal{F}$, P, $(S, \mathcal{S})$ (with X_i i.i.d. P), A and v be as in Theorem 6 with $F = 1$. Then there exist constants $D_1 := D_1(m, A, v)$ and $D_2 = D_2(m, A, v)$ such that*

$$n^k E \left\| U_n^{(k)}(\pi_k f) \right\|_{\mathcal{F}}^2 \leq D_1 \sigma^2 \left[\log \left(\frac{A}{\sigma} \right) \right]^k, \quad k = 0, 1, \ldots, m, \tag{3.11}$$

assuming

$$n\sigma^2 \geq D_2 \log \left(\frac{A}{\sigma} \right), \tag{3.12}$$

where σ^2 is any number satisfying

$$\|P^m f^2\|_{\mathcal{F}} \leq \sigma^2 \leq 1. \tag{3.13}$$

Proof. We shall use the notation "$x \leq' y$" to mean that $x \leq cy$ for a suitable constant $0 < c < \infty$ that depends only on $k \leq m$, hence, only on m. Also we shall write $\| \cdot \| := \| \cdot \|_{\mathcal{F}}$. It is easily checked that on $0 < x \leq 1$ the function $y = x(\log(A^2/x))^k$ is increasing and concave, which can be verified by taking derivatives. Also we have for $0 < a \leq 1$,

$$\int_0^a \left[\log \left(\frac{A}{\tau} \right) \right]^{k/2} d\tau \leq' a \left[\log \left(\frac{A^2}{a^2} \right) \right]^{k/2}.$$

These facts require A large enough, and $A \geq e^m$ suffices.

Now, by the usual decoupling and randomization (Theorem 3.5.3 on p. 140 of de la Peña and Giné (1999)) and an entropy bound for Rademacher chaos (Corollary 5.1.8 on p. 220 of de la Peña and Giné (1999)), we have, using the above integral bound and concavity:

$$|I_n^k| E \left\| U_n^{(k)}(\pi_k f) \right\|^2 \leq CE \left\{ \int_0^{\sqrt{\|U_n^{(k)}((P^{m-k}f)^2)\|}} \left[\log\left(\frac{A}{\tau}\right) \right]^{k/2} d\tau \right\}^2$$

$$\leq' E\|U_n^{(k)}((P^{m-k}f)^2)\| \left[\log\left(\frac{A^2}{E\|U_n^{(k)}((P^{m-k}f)^2)\|} \right) \right]^k. \tag{3.14}$$

By the Hoeffding decomposition

$$E\|U_n^{(k)}((P^{m-k}f)^2)\| \leq \sum_{r=0}^{k} \binom{k}{r} E\|U_n^{(r)}(\pi_r(P^{m-k}f)^2)\|. \tag{3.15}$$

The term corresponding to $r = 0$ is simply $\|P^k(P^{m-k}f)^2\| \leq \sigma^2$. And we can apply Theorem 6 to the other terms because we have $\mathrm{VC}(A, v)$ classes bounded by 1 that admit the same σ, as argued in the proof of Theorem 6 given in Giné and Mason (2005), to get:

$$E\|U_n^{(r)}(\pi_r(P^{m-k}f)^2)\| \leq' n^{-r/2}\sigma \left[\log\left(\frac{A\|F\|_{L_2(P^m)}}{\sigma} \right) \right]^{r/2}, \quad r = 1, \ldots, k. \tag{3.16}$$

But by $F = 1$, hypothesis (3.12) on $n\sigma^2$ and (3.13) we obtain

$$\frac{1}{n^{r/2}}\sigma \left[\log\left(\frac{A\|F\|_{L_2(P^m)}}{\sigma} \right) \right]^{r/2} = \frac{1}{n^{r/2}}\sigma \left[\log\left(\frac{A}{\sigma} \right) \right]^{r/2} \leq' \sigma^{1+r} \leq' \sigma^2, \tag{3.17}$$

and we get from (3.15), (3.16) and (3.17) that

$$E\|U_n^{(k)}((P^{m-k}f)^2)\| \leq' \sigma^2, \tag{3.18}$$

in fact, $E\|U_n^{(k)}((P^{m-k}f)^2)\| \leq 1 \wedge D\sigma^2$ for some $D = D(m) < \infty$. Now, by monotonicity of the function $y = x(\log(A^2/x))^k$ for $0 < x \leq 1$, (3.18) and (3.14) give for some $D_1 > 0$,

$$|I_n^k| E \left\| U_n^{(k)}(\pi_k f) \right\|^2 \leq D_1 \sigma^2 \left[\log\left(\frac{A}{\sigma} \right) \right]^k,$$

assuming (3.12). q.e.d.

3.2.3 Completion of the proof of Theorem 1

We shall now use Ingredients 1 and 2 to show that for $k = 2, \ldots, m$, (2.10) holds. For each $k = 2, \ldots, m$ and $n \geq 1$ let

$$\mathcal{H}_{k,n} = \left\{ \frac{\lambda h_n \pi_k \overline{K}_{\lambda h_n}(t, \cdot, \ldots, \cdot)}{\|\pi_k \overline{K}\|_\infty} : t \in \mathbf{R}^d, a \leq \lambda \leq b \right\}.$$

Set $n_l = 2^l$ for $l \geq 0$. Using assumption (HC), we can choose $0 < \mathbb{A} \leq a \leq b \leq \mathbb{B} < \infty$ so that for $l \geq 1$ and $n_{l-1} \leq n < n_l$,

$$\mathcal{H}_{k,n} \subset \mathcal{G}_{k,l} := \left\{ \frac{\lambda h_{n_l} \pi_k \overline{K}_{\lambda h_{n_l}}(t, \cdot, \ldots, \cdot)}{\|\pi_k \overline{K}\|_\infty} : t \in \mathbf{R}^d, \mathbb{A} \leq \lambda \leq \mathbb{B} \right\} \tag{3.19}$$

Define for $n_{l-1} < n \leq n_l$,

$$S_n = \sup_{H \in \mathcal{G}_{k,l}} \left| \frac{1}{n_l^{k/2}} \sum_{\mathbf{i} \in I_n^k} H(X_{i_1}, \ldots, X_{i_k}) \right|.$$

Arguing as in the proof of Theorem 1 in Giné and Mason (2005), we get that each $\mathcal{G}_{k,l}$ is of VC type with characteristic constants A and v that can be taken independently of $\mathbb{A}$, $\mathbb{B}$ and $l \geq 1$, and furthermore, we get for each $l \geq 0$,

$$\sigma_l^2 := C h_{n_l} \geq \sup_{H \in \mathcal{G}_{k,l}} EH^2$$

with

$$C = \mathbb{B} \|f_g\|_\infty \|K\|_2^2 / \|\pi_k \overline{K}\|_\infty^2.$$

This is where the assumption that $g(X_1, \ldots, X_m)$ has a bounded density f_g is used in the proof of Theorem 1.

Applying inequality (3.3) with $r = 2$ and $c = 1/2$ we get for $l \geq 1$ so that $n_{l-1} \geq k$,

$$\Pr\left\{ \max_{n_{l-1} < n \leq n_l} S_n > x \right\} \leq \frac{2 \left(\Pr\left\{ S_{n_l} > x/2 \right\} \right)^{1/2} \left(ES_{n_l}^2 \right)^{1/2}}{x}$$

Since the functions in $\mathcal{G}_{k,l}$ are bounded by 1, Major's inequality gives for all y satisfying

$$n_l \sigma_l^2 = n_l C h_{n_l} \geq \left(\frac{y}{C h_{n_l}} \right)^{2/k} \geq C_3 \log \left(\frac{2}{C h_{n_l}} \right) \tag{3.20}$$

the inequality

$$\Pr\left\{ S_{n_l} > y \right\} \leq C_1 \exp\left(-C_2 \left(\frac{y}{\sigma_l} \right)^{2/k} \right).$$

Thus if $y = x/2$ satisfies (3.20) we obtain

$$\Pr\left\{\max_{n_{l-1}<n\le n_l} S_n > x\right\} \le \frac{2\sqrt{C_1}\exp\left(-\frac{C_2}{2}\left(\frac{x}{2\sigma_l}\right)^{2/k}\right)\left(E\left(S_{n_l}\right)^2\right)^{1/2}}{x}.$$

Recall that we are assuming that

$$nh_n \log\log n / \left(\log\log n \vee \log\left(1/h_n\right)\right)^2 \to \infty. \tag{3.21}$$

This implies that for all large l assumption (3.12) of Corollary 1 holds and therefore for some constant C_4,

$$\left(E\left(S_{n_l}\right)^2\right)^{1/2} \le C_4 h_{n_l}^{1/2}\left(\log\left(1/h_{n_l}\right)\right)^{k/2}. \tag{3.22}$$

Thus assuming that $y = x/2$ satisfies (3.20), this yields the further bound, noting that $\sigma_l = \sqrt{Ch_{n_l}}$,

$$\Pr\left\{\max_{n_{l-1}<n\le n_l} S_n > x\right\}$$

$$\le \frac{2\sqrt{C_1}C_4}{x}\exp\left(-\frac{C_2}{2}\left(\frac{x}{2\sqrt{Ch_{n_l}}}\right)^{2/k}\right) h_{n_l}^{1/2}\left(\log\left(1/h_{n_l}\right)\right)^{k/2}. \tag{3.23}$$

Setting

$$x = 2\sqrt{Ch_{n_l}}\left[\beta\left(\log\log n_l \vee \log\left(1/h_{n_l}\right)\right)\right]^{k/2} \tag{3.24}$$

for some $\beta > 0$, we get that for a large enough β for $l \ge 1$ so that $n_{l-1} \ge k$,

$$\left(\frac{x}{2\sigma_l}\right)^{2/k} = \beta\left(\log\log n_l \vee \log\left(1/h_{n_l}\right)\right) > C_3 \log\left(2/h_{n_l}\right)$$

and keeping in mind (3.21) we also have for all large l,

$$n_l\sigma_l^2 = Cn_l h_{n_l} \ge \frac{\log\log n_l \vee \log\left(1/h_{n_l}\right)}{\log\log n_l}\beta\left(\log\log n_l \vee \log\left(1/h_{n_l}\right)\right)$$

$$\ge \beta\left(\log\log n_l \vee \log\left(1/h_{n_l}\right)\right) = \left(\frac{x}{2\sigma_l}\right)^{2/k}.$$

Thus by our choice of x, $y = x/2$ satisfies (3.20) and we are permitted to apply inequality (3.23) to get for all large l,

$$\Pr\left\{\max_{n_{l-1}<n\le n_l} S_n > 2\sqrt{Ch_{n_l}}\left[\beta\left(\log\log n_l \vee \log\left(1/h_{n_l}\right)\right)\right]^{k/2}\right\}$$

$$\le \frac{\sqrt{C_1}C_4\exp\left(-\frac{C_2\beta}{2}\left(\log\log n_l \vee \log\left(1/h_{n_l}\right)\right)\right)\left(\log\left(1/h_{n_l}\right)\right)^{k/2}}{\sqrt{C}\left(\beta\left(\log\log n_l \vee \log\left(1/h_{n_l}\right)\right)\right)^{k/2}},$$

which with $C_\beta = C_4\sqrt{C_1/C}/\beta^{k/2}$ is

$$\leq C_\beta \exp\left(-\frac{C_2\beta}{2}\left(\log\log n_l \vee \log\left(1/h_{n_l}\right)\right)\right).$$

From this last bound, we readily check that for all large enough β,

$$\sum_{l=1}^{\infty} \Pr\left\{\max_{n_{l-1}<n\leq n_l} S_n > 2\sqrt{Ch_{n_l}}\left[\beta\left(\log\log n_l \vee \log\left(1/h_{n_l}\right)\right)\right]^{k/2}\right\} < \infty,$$

which by the Borel–Cantelli lemma and some routine bounds using (3.19) implies that for $k = 2, \ldots, m$, a.s.

$$\frac{\sqrt{n}}{\sqrt{\log\log n}} \sup_{a\leq\lambda\leq b} \sup_{t\in\mathbf{R}^d} \left|U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n}(t,\cdot,\ldots,\cdot))\right| = O\left(\frac{\left(\log\log n \vee \log\left(1/h_n\right)\right)^{k/2}}{n^{\frac{k-1}{2}} h_n^{1/2}\sqrt{\log\log n}}\right). \tag{3.25}$$

Clearly from (3.25) and (3.21) we get (2.10). This completes the proof of Theorem 1. q.e.d.

3.3 Proof of Theorem 2

We begin by showing that assumption a) of Theorem 2 implies that the compact LIL part of Definition 1, namely (1.11) holds. From this we shall infer that assumptions (2.3) and (2.4) of Proposition 1 hold, which imply (2.5).

Let us recall that the Ledoux-Talagrand (1988) LIL (see e.g. Theorem 8.6 on p. 213 in Ledoux and Talagrand (1991)) states that the compact LIL holds for a random vector ξ taking values in a separable Banach space $\mathbf{B}$ if and only if

i) $E\|\xi\|^2/\log\log\|\xi\| < \infty$,

ii) $E\xi = 0$ and the set of random variables $\{f^2(\xi) : f \in \mathbf{B}', \|f\| \leq 1\}$ is uniformly integrable,

iii) $\sum_{i=1}^n \xi_i/\sqrt{n\log\log n} \to 0$ in probability.

Here ξ_i are i.i.d. copies of ξ. If $\mathbf{B}$ is of type 2, then the condition $E\|\xi\|^2 < \infty$ implies the central limit theorem. Hence, if ξ is also centered, then iii) holds; since i) and ii) are automatically satisfied if $E\|\xi\|^2 < \infty$, it follows that $E\xi = 0$ and $E\|\xi\|^2 < \infty$ suffice for the LIL when $\mathbf{B}$ is of type 2, which is the case of $\mathbf{B} = L_p(\mathbf{R}^d)$, $p \geq 2$ (cf. Ledoux and Talagrand (1991)). Therefore, assumption a) of Theorem 2, namely $E\|f_i(\cdot,X) - Ef_i(\cdot,X)\|_p^2 < \infty$, implies that the L_p-valued random variable $\sum_{i=1}^m (f_i(\cdot,X) - Ef_i(\cdot,X))$ satisfies the compact LIL, or, in the form we say it also in this article, the random variables $\sqrt{n}(P_n - P)\sum_{i=1}^n f_i(t,\cdot)$ satisfy the compact LIL, that is, the LIL part (1.11) of Definition 1 holds, and in particular so does condition (2.4) in Proposition 1. Moreover, the compact LIL for $f_i(\cdot,X)$ together with the Fréchet-Kolmogorov characterization of compact sets in $L_p(\mathbf{R}^d)$ (Dunford and Schwartz (1966), Theorem IV.8.21, p. 301) imply condition (2.3) in the same proposition: see the remark following the completion of this proof for details. Hence, Proposition 1 applies, which establishes (2.5). So, as explained in Subsection 2.5, the theorem will be proved if we show that the limit (2.10) holds.

To prove (2.10) we rely on the maximal inequality in Theorem 4 and a result from Giné and Mason (2005) based on an exponential bound for U-processes taking values in Banach spaces of type 2 together with the usual entropy bound. We describe first this second result, which is equation (3.50) in the mentioned reference:

Lemma 1 *For $p \geq 2$, $2 \leq k \leq m$ and K satisfying condition (KI(p,2/m, $\mathbb{A}$, $\mathbb{B}$)), there is a constant D that depends only on k and a constant C depending only on $\mathbb{A}$ such that*

$$\Big\| \sup_{\mathbb{A}\leq\lambda\leq\mathbb{B}} \Big\| \frac{h_n^{(p-1)/p}}{n^{k/2}} \sum_{\mathbf{i}\in I_n^k} \pi_k \overline{K}_{\lambda h_n}(\cdot, X_{i_1}, \dots, X_{i_m}) \Big\|_p \Big\|_{\Psi_{2/k}}$$

$$\leq D \int_0^{C\|K\|_p} \Psi_{2/k}^{-1}(N(\mathcal{K}_{[\mathbb{A},\mathbb{B}]}, d_p, \tau)) d\tau.$$

To apply Theorem 4, set for $2^{l-1} = n_{l-1} < n \leq n_l = 2^l$,

$$S_n := \sqrt{\frac{n_l}{\log\log n_l}} \sup_{\mathbb{A}\leq\lambda\leq\mathbb{B}} \Big\| \frac{1}{n_l^k} \sum_{\mathbf{i}\in I_n^k} \pi_k \overline{K}_{\lambda h_{n_l}}(\cdot, X_{i_1}, \dots, X_{i_m}) \Big\|_p .$$

and note that, by the regularity condition (HC), for $n > 2k$,

$$\sqrt{\frac{n}{\log\log n}} \sup_{a\leq\lambda\leq b} \big\| U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n}(t, \cdot, \dots, \cdot) \big\|_p \leq' S_n. \tag{3.26}$$

By Lemma 1, condition (KI(p,2/m, $\mathbb{A}$, $\mathbb{B}$)) (which implies condition (KI(p,2/k, $\mathbb{A}$, $\mathbb{B}$)) for any $2 \leq k \leq m$) and the properties of $\|\cdot\|_{\Psi_{2/k}}$, we have both,

$$(ES_{n_l}^2)^{1/2} \leq \frac{C_1}{n_l^{(k-1)/2} h_{n_l}^{(p-1)/p} \sqrt{\log\log n_l}}$$

and

$$\Pr\{S_{n_l} > x\} \leq b \exp\Big\{ -C_2 \big(x n_l^{(k-1)/2} h_{n_l}^{(p-1)/p} \sqrt{\log\log n_l} \big)^{2/k} \Big\}, \quad x > 0,$$

where C_1 and C_2 are constants (independent of n). Using the regularity property (HC) of the sequence h_n, these two inequalities and Theorem 4 give

$$\Pr\Big\{ \max_{n_{l-1}<n\leq n_l} S_n > 2x \Big\} \leq \frac{\exp\Big\{ -C_3 \big(x n_l^{(k-1)/2} h_{n_l}^{(p-1)/p} \sqrt{\log\log n_l} \big)^{2/k} \Big\}}{C_4 x n_l^{(k-1)/2} h_{n_l}^{(p-1)/p} \sqrt{\log\log n_l}}$$

for suitable constants C_3 and C_4. Since $nh_n^{2(p-1)/p}/\log\log n \to \infty$, for $k = 2$ and any $x > 0$, the exponent is eventually smaller than $-C \log\log n_l$ for a constant C larger than 1 (for any large constant, in fact) and the denominator tends to infinity; for $k > 2$, for each $x > 0$, the exponent is eventually smaller than $-Cn_l^{(k-2)/k}$ and the denominator tends to infinity. So, in either case we have

$$\sum_l \Pr\Big\{\max_{n_{l-1}<n\le n_l} S_n > 2x\Big\} < \infty$$

for all $x > 0$. Because of (3.26), this implies that, for $2 \le k \le m$,

$$\sup_{a\le\lambda\le b} \frac{\sqrt{n}\left\|U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n}(t,\cdot,\ldots,\cdot))\right\|_p}{\sqrt{\log\log n}} \to 0, \text{ a.s,}$$

that is, (2.10) holds. The proof of Theorem 2 is now complete. q.e.d.

Remark For the sake of completeness, we show here in detail how to derive condition (2.3) from the compact LIL for $f_i(t,X)$, and we note that the same argument works in the case $1 \le p < 2$. For i fixed (the argument is the same for each $1 \le i \le m$), set

$$V_n(t) = \sqrt{\frac{n}{2\log\log n}}(P_n - P)f_i(t,\cdot), \quad n \in \mathbf{N}.$$

The compact LIL described in Subsection 2.4 implies, by a theorem of Kuelbs (1976) (see e.g. Theorem 8.5 in Ledoux and Talagrand (1991), p. 210), that

$$\lim_{n\to\infty} d_p(V_n, \mathcal{H}) = 0 \ \text{ a.s.,} \tag{3.27}$$

where $\mathcal{H}$ is a compact subset of $L_p(\mathbf{R}^d)$ and $d_p(f,A) := \inf_{g\in A}\|f-g\|_{L_p}$ is the distance between functions and sets in L_p induced by the L_p norm. The compactness criterion in $L_p(\mathbf{R}^d)$ in Dunford and Schwartz (1966), p. 301 (Theorem V.8.21 there) implies

$$\lim_{\delta\to 0} \sup_{|u|\le\delta} \sup_{f\in\mathcal{H}} \int |f(t-u) - f(t)|^p dt = 0. \tag{3.28}$$

Let ω be in the set of probability 1 where (3.27) holds. Then, there are $f_n \in \mathcal{H}$ such that

$$\int_{\mathbf{R}^d} |V_n(t,\omega) - f_n(t)|^p dt \to 0,$$

and therefore, by (3.28),

$$\lim_{\delta\to 0} \limsup_n \sup_{|u|\le\delta} \left(\int_{\mathbf{R}^d} |V_n(t-u,\omega) - V_n(t,\omega)|^p dt\right)^{1/p}$$
$$\le 2\lim_n \|V_n(\omega) - f_n\|_p + \lim_{\delta\to 0} \sup_{|u|\le\delta} \sup_n \|f_n(\cdot - u) - f_n(\cdot)\|_p \to 0,$$

proving (2.3).

3.4 Proof of Theorem 3

Condition a) of Theorem 3, concretely $\int_{\mathbf{R}^d}\left[E(f_i(t,X)-Ef_i(t,X))^2\right]^{p/2}dt<\infty$, implies that for each $i=1,\dots,m$, the L_p-valued random variable $f_i(\cdot,X)-Ef_i(\cdot,X)$ (with $1\le p<2$) both, satisfies the CLT and has square integrable norm, and therefore, the Ledoux-Talagrand LIL implies, as in the proof of the previous theorem, that the random variable $\sum_{i=1}^m(f_i(\cdot,X)-Ef_i(\cdot,X))$ satisfies the compact LIL in L_p. This, in turn, by the argument given in the last subsection, implies that the two hypotheses in Proposition 1 hold. Hence, as in the previous proof, it only remains to show that the limit (2.10) holds.

The fact that L_p is not of type 2 for $1\le p<2$ complicates matters somewhat because the exponential inequality for L_p-valued U-statistics in this case, which is one of the two key ingredients for Lemma 1, is not as good. Here we state the exponential bound in this situation and complete the proof of the theorem, but postpone the proof of the bound to the Appendix.

Lemma 2 *Let $1\le p<2$. Under the assumptions of Theorem 3, there exists a constant C independent of n such that, for $2\le k\le m$ and for all $u\ge 0$,*

$$\Pr\left\{\sup_{\mathbb{A}\le\lambda\le\mathbb{B}}\sqrt{n}\left\|U_n^{(k)}(\pi_k\overline{K}_{\lambda h_n}(t,\cdot,\dots,\cdot)\right\|_p\ge u\right\}$$

$$\le C\exp\left\{-\frac{1}{C}\left[(un^{k/2-1/2}h_n^{1/2})^{2/k}\wedge(un^{1/2}h_n^{(p-1)/p})^{1/k}\right]\right\} \tag{3.29}$$

and

$$E\sup_{\mathbb{A}\le\lambda\le\mathbb{B}}\left(\sqrt{n}\left\|U_n^{(k)}(\pi_k\overline{K}_{\lambda h_n}(t,\cdot,\dots,\cdot)\right\|_p\right)^2\le\frac{C}{nh_n}. \tag{3.30}$$

With the same notation as in the previous proof, these two inequalities and the maximal inequality in Theorem 4 give, with the help of (HC), that

$$\Pr\left\{\max_{n_{l-1}<n\le n_l}S_n>2x\right\}\le\frac{\exp\left\{-(a_{n,1}\wedge a_{n,2})\right\}}{C_4x\sqrt{n_lh_{n_l}\log\log n_l}}$$

with

$$a_{n,1}=C_3\big(xn_l^{k/2-1/2}h_{n_l}^{1/2}\sqrt{\log\log n_l}\big)^{2/k}\ \text{and}\ a_{n,2}=C_3\big(xn_l^{1/2}h_{n_l}^{(p-1)/p}\sqrt{\log\log n}\big)^{1/k},$$

for suitable constants C_3 and C_4. Since $nh_n/\log\log n\to\infty$, it readily follows from this inequality that, for any $2\le k\le m$,

$$\sum_l\Pr\left\{\max_{n_{l-1}<n\le n_l}S_n>2x\right\}<\infty$$

for all $x>0$, which implies (2.10) (see (3.26)). Theorem 3 is thus proved modulo the proof of Lemma 2. q.e.d.

3.5 Appendix: Proof of Lemma 2

Our point of departure is a moment inequality for **B**-valued U-statistics and some estimations from Giné and Mason (2005).

Proposition 2 *(Theorem 10 in Giné and Mason (2005)) Let* $\mathbf{B}$ *be a separable Banach space and let* $H : S^k \mapsto \mathbf{B}$ *be a bounded* P*–canonical random vector symmetric in its entries. Define* κ *and* χ_n *to be any pair of numbers such that*

$$\kappa \geq \sup_{x_1,\dots,x_k \in S} \|H(x_1,\dots,x_k)\|, \quad \chi_n \geq \left(E \left\| \sum_{\mathbf{i}\in I_n^k} H(X_{i_1}^{(1)},\dots,X_{i_k}^{(k)}) \right\|^2 \right)^{1/2}. \quad (3.31)$$

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

$$E \left\| \sum_{\mathbf{i}\in I_n^k} H(X_{i_1},\dots,X_{i_k}) \right\|^r \leq C^r \left[(r^{kr/2}\chi_n^r) \ \vee \ (r^{kr} n^{(k-1)r} \kappa^r) \right]. \quad (3.32)$$

(Actually, their result is slightly better.) Next we translate Proposition 2 into exponential inequalities. For ease of notation we write $H_{\mathbf{i}} := H(X_{i_1},\dots,X_{i_k})$ when $\mathbf{i} = (i_1,\dots,i_k)$.

Corollary 2 *Under the assumptions of Proposition 2, there exists* $C = C(k) < \infty$ *such that with,*

$$A_n = \frac{eC\chi_n^2}{n^{k-1}\kappa},$$

for all $n \in \mathbf{N}$,

$$\Pr\left\{ \Big\| \sum_{\mathbf{i}\in I_n^k} H_{\mathbf{i}} \Big\| I\Big(\Big\| \sum_{\mathbf{i}\in I_n^k} H_{\mathbf{i}} \Big\| \leq A_n \Big) \geq x \right\} \leq e^2 \exp\left\{ -\Big(\frac{x}{eC\chi_n} \Big)^{2/k} \right\}, \quad (3.33)$$

for all $x \leq A_n$ *(and this probability is zero if* $x > A_n$*), and*

$$\Pr\left\{ \Big\| \sum_{\mathbf{i}\in I_n^k} H_{\mathbf{i}} \Big\| I\Big(\Big\| \sum_{\mathbf{i}\in I_n^k} H_{\mathbf{i}} \Big\| \geq A_n \Big) \geq x \right\} \leq e^2 \exp\left\{ -\Big(\frac{x \vee A_n}{eC\kappa n^{k-1}} \Big)^{1/k} \right\}, \quad (3.34)$$

for all $x > 0$.

Proof. (Sketch) In this proof we follow Giné, Latała and Zinn (2000), p. 29. Take first $x \leq A_n$. Set for any $x > 0$,

$$r = r(x) = \left(\frac{x}{eC\chi_n} \right)^{2/k}.$$

Note whenever $x \le A_n$ we have

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

Assume that $r = r(x) \ge 2$. The previous proposition gives

$$\Pr\Big\{\Big\|\sum H_{\mathbf{i}}\Big\| I\Big(\Big\|\sum H_{\mathbf{i}}\Big\| \le A_n\Big) \ge x\Big\} \le \Pr\Big\{\Big\|\sum H_{\mathbf{i}}\Big\| \ge x\Big\} \le \frac{E\Big\|\sum H_{\mathbf{i}}\Big\|^r}{x^r}$$

$$\le \frac{C^r r^{kr/2}\chi_n^r}{x^r} = e^{-r} = \exp\Big\{-\Big(\frac{x}{eC\chi_n}\Big)^{2/k}\Big\}.$$

Obviously, the above probability is $\le e^2 e^{-r}$ whenever $r < 2$. This establishes (3.33)

For $x \ge A_n$ we take

$$r = r(x) = \Big(\frac{x}{eCn^{k-1}\kappa}\Big)^{1/k}$$

and proceed by analogy with the previous case. q.e.d.

These inequalities will be applied in combination with the following estimates:

Lemma 3 *(Equation (3.38) and Lemma 2 in Giné and Mason (2005)) There exists a constant $C < \infty$ such that for any kernel K in L_p, $1 \le p < \infty$, $1 \le k \le m$ and $h > 0$,*

$$\sup_{x_1,\dots,x_k \in S} \big\|\pi_k \overline{K}_h(\cdot, x_1, \dots, x_k)\big\|_{L_p} \le \frac{C\|K\|_p}{h^{(p-1)/p}}. \tag{3.35}$$

If K^2 and f_g are in $L_1(\mu_s)$ for some $s > d(2-p)/p$. for $1 \le p < 2$, there exists $C < \infty$, such that, for $1 \le k \le m$ and all $0 < h \le b$,

$$\Big(E\Big\|\sum_{I_n^k} \pi_k \overline{K}_h(\cdot, X_{i_1}, \dots, X_{i_k})\Big\|_p^2\Big)^{1/2} \le \frac{C\|K\|_{L_2(\mu_s)} n^{k/2}}{h^{1/2}} \tag{3.36}$$

Because we have two exponential inequalities with different exponents for $\big\|\sum_{\mathbf{i}\in I_n^k} H_{\mathbf{i}}\big\|$, if we apply directly the usual entropy bound we will lose some ground, and, to get a sharper result, we must instead combine these two bounds with the actual proof of the entropy bound. For this we choose to adapt the simple chaining argument in de la Peña and Giné (1999), proof of Theorem 5.1.4, pp. 216-217. This adaptation, although straightforward, may have some independent interest and can be generalized to the case of more than two tail regimes in the exponential inequality.

Proof of Lemma 2: Chaining argument. For ease of notation, set

$$Y_{K_\lambda,n} := \sqrt{n}\big\|U_n^{(k)}(\pi_k \overline{K}_{\lambda h_n}(t, \cdot, \dots, \cdot)\big\|_{L_p}, \quad \lambda > 0,$$

and, consistent with the definition of K_0, $Y_{K_0,n} = 0$. This process is proven to be separable in Giné and Mason (2005), Remark 5. Hence, there exists a countable set $T_0 \subset [\mathbb{A}, \mathbb{B}] \cup \{0\}$ such that

$$\Pr\Big\{ \sup_{\mathbb{A} \le \lambda \le \mathbb{B}} Y_{K_\lambda,n} > x \Big\} = \Pr\Big\{ \sup_{\lambda \in T_0} Y_{K_\lambda,n} > x \Big\}$$

and therefore,

$$\Pr\Big\{ \sup_{\mathbb{A} \le \lambda \le \mathbb{B}} Y_{K_\lambda,n} > x \Big\} = \sup_{\substack{T \text{ finite} \\ T \in [\mathbb{A},\mathbb{B}] \cup \{0\}}} \Pr\Big\{ \sup_{\lambda \in T} Y_{K_\lambda,n} > x \Big\}. \tag{3.37}$$

Set

$$d := d_p \vee \tilde{d}_{2,s} \text{ and } \|K\|_{p,2,s} = \|K\|_p \vee \|K\|_{L_2(\mu_s)},$$

where μ_s and $\tilde{d}_{2,s}$ are as in (1.13) Then using the trivial inequality

$$N(\mathcal{K}_{[\mathbb{A},\mathbb{B}]}, d, \tau) \le N(\mathcal{K}_{[\mathbb{A},\mathbb{B}]}, d_p, \tau) + N(\mathcal{K}_{[\mathbb{A},\mathbb{B}]}, \tilde{d}_{2,s}, \tau),$$

the logarithmic character of Ψ_α^{-1} and its monotonicity in α, it is easy to see that the two entropy conditions in b) of Theorem 3 imply

$$\int_0^{2\mathbb{A}\|K\|_{p,2,s}} \Psi_{i/m}^{-1}(N(\mathcal{K}_{[\mathbb{A},\mathbb{B}]}, d, \tau)) d\tau < \infty, \quad i = 1, 2. \tag{3.38}$$

For $s_1, s_2 \in [\mathbb{A}, \mathbb{B}] \cup \{0\}$ define, with some abuse of notation,

$$d(s_1, s_2) := d(Y_{K_{s_1},n}, Y_{K_{s_2},n}).$$

By multiplying K by a convenient factor if necessary (this part of the proof does not require K to integrate to 1) we can assume that $\mathcal{K}_{[\mathbb{A},\mathbb{B}]}$ is contained in a d-ball of radius 1 about zero. Choose any $T \subset [\mathbb{A}, \mathbb{B}] \cup \{0\}$ finite, which as in the proof of Theorem 5.1.4 of de la Peña and Giné (1999), we can assume without loss of generality contains zero. For each $r \in \mathbf{N}$, we let $T_r = \{s_1^r, \ldots, s_{N_r}^r\} \subset T$ be the set of centers of $N_r := N(\mathcal{K}_{[\mathbb{A},\mathbb{B}]}, d, 2^{-r})$ open balls of radius at most 2^{-r} that cover T, and we take $T_0 := \{0\}$. For each r let $p_r : T \mapsto T_r$ be a function satisfying $d(s, p_r(s)) < 2^{-r}$ for all $s \in T$ (this function exists by construction of T_r). Also since T is finite, there exists $r_T < \infty$ such that for all $r \ge r_T$ and $s \in T$, $d(p_r(s), s) = 0$ (in particular then, $Y_{K_{p_r(s)},n} = Y_{K_{s,n}}$ a.s.). Hence, setting

$$Y_{K_s,n} := Y_s,$$

we have

$$Y_s = \sum_{r=1}^{r_T} \big(Y_{p_r(s)} - Y_{p_{r-1}(s)} \big), \text{ a.s.}$$

Note also

$$d(p_r(s), p_{r-1}(s)) \le d(p_r(s), s) + d(p_{r-1}(s), s) < 3 \cdot 2^{-r}.$$

Therefore

$$
\begin{aligned}
\Pr\Big\{\sup_{\lambda\in T} Y_{K_\lambda,n} > x\Big\} &\le \Pr\Big\{\sum_{r=1}^{r_T} \max_{\substack{s\in T_r, t\in T_{r-1}\\ d(s,t)\le 3\cdot 2^{-r}}} |Y_s - Y_t| > x\Big\} \\
&\le \Pr\Big\{\sum_{r=1}^{r_T} \max_{\substack{s\in T_r, t\in T_{r-1}\\ d(s,t)\le 3\cdot 2^{-r}}} |Y_{K_s-K_t,n}| > x\Big\}. \qquad (3.39)
\end{aligned}
$$

Now, for $s \in T_r, t \in T_{r-1}$ with $d(s,t) \le 3 \cdot 2^{-r}$, we will apply the exponential inequalities in Corollary 2 for the kernel

$$
H = \frac{1}{n^{k-1/2}}\pi_k(\overline{K}_s - \overline{K}_t)
$$

for appropriate $\kappa(r)$ and $\chi_n(r)$ independent of s and t in this set. By Lemma 3, we can take

$$
\kappa = \kappa(r) := \frac{c2^{-r}}{n^{k-1/2}h_n^{(p-1)/p}} \quad \text{and} \quad \chi_n = \chi_n(r) := \frac{c2^{-r}}{n^{(k-1)/2}h_n^{1/2}}
$$

for a conveniently chosen constant c (independent of r and n). Define $A_n(r)$ to be the constant A_n from Corollary 2 corresponding to these values of κ and χ_n. Set

$$
Y^{(1,r)}_{K_s-K_t,n} = Y_{K_s-K_t,n} I(Y_{K_s-K_t,n} \le A_n(r))
$$

and

$$
Y^{(2,r)}_{K_s-K_t,n} = Y_{K_s-K_t,n} I(Y_{K_s-K_t,n} > A_n(r)).
$$

Then, the exponential bounds in Corollary 2 imply

$$
\|Y^{(1,r)}_{K_s-K_t,n}\|_{\Psi_{2/k}} \le C\chi_n(r) =: \frac{C_1 2^{-r}}{n^{(k-1)/2}h_n^{1/2}} \qquad (3.40)
$$

and

$$
\|Y^{(2,r)}_{K_s-K_t,n}\|_{\Psi_{1/k}} \le C\kappa(r)n^{k-1} =: \frac{C_2 2^{-r}}{n^{1/2}h_n^{(p-1)/p}}. \qquad (3.41)
$$

For $i = 1, 2$, set

$$
M_i = \sum_{r=1}^{r_T} \max_{\substack{s\in T_r, t\in T_{r-1}\\ d(s,t)\le 3\cdot 2^{-r}}} |Y^{(i,r)}_{K_s-K_t,n}|.
$$

Since the Ψ_α functions are increasing, we can use Chebyshev after applying

$$
\Psi_\alpha(\cdot/\|M_i\|_{\Psi_\alpha}),\ i = 1, 2,
$$

to get, from (3.39),

$$
\Pr\Big\{\sup_{\lambda\in T} Y_{K_\lambda,n} > x\Big\} \le \Pr\Big\{\sum_{i=1}^{2} M_i > x\Big\}
$$

$$\leq \frac{1}{\Psi_{2/k}\left(\frac{x}{2\|M_1\|_{\Psi_{2/k}}}\right)} + \frac{1}{\Psi_{1/k}\left(\frac{x}{2\|M_2\|_{\Psi_{1/k}}}\right)}. \tag{3.42}$$

Now, using the maximal inequality (5.1.9) in de la Peña and Giné (1999), page 216, namely,

$$\Big\| \max_{1\leq k\leq N} |\xi_k| \Big\|_{\Psi_\alpha} \leq C_\alpha \Psi_\alpha^{-1}(N) \max_{1\leq k\leq N} \|\xi_k\|_{\Psi_\alpha},$$

the estimates in (3.40) and (3.41), and the fact that Ψ^{-1} is of logarithmic type, we get with C denoting an appropriate constant that may change from one line to the next,

$$\begin{aligned}
\Big\| \sum_{r=1}^{r_T} \max_{\substack{s\in T_r, t\in T_{r-1}\\ d(s,t)\leq 3\cdot 2^{-k}}} Y^{(1,r)}_{K_s-K_t,n} \Big\|_{\Psi_{2/k}} &\leq \sum_{r=1}^{r_T} \Big\| \max_{\substack{s\in T_r, t\in T_{r-1}\\ d(s,t)\leq 3\cdot 2^{-r}}} Y^{(1,r)}_{K_s-K_t,n} \Big\|_{\Psi_{2/k}} \\
&\leq C \sum_{r=1}^{r_T} \Psi_{2/k}^{-1}(N_r N_{r-1}) \max_{\substack{s\in T_r, t\in T_{r-1}\\ d(s,t)\leq 3\cdot 2^{-r}}} \|Y^{(1,r)}_{K_s-K_t,n}\|_{\Psi_{2/k}} \\
&\leq \frac{C}{n^{(k-1)/2} h_n^{1/2}} \sum_{r=1}^{r_S} 2^{-r} \Psi_{2/k}^{-1}(N_r) \\
&\leq \frac{C}{n^{(k-1)/2} h_n^{1/2}} \int_0^1 \Psi_{2/k}^{-1}(N(\mathcal{K}_{[a,b]}, d, \varepsilon) d\varepsilon \\
&\leq \frac{C}{n^{(k-1)/2} h_n^{1/2}},
\end{aligned}$$

where in the last step we are using (3.38). Similarly,

$$\Big\| \sum_{r=1}^{r_T} \max_{\substack{s\in T_r, t\in T_{r-1}\\ d(s,t)\leq 3\cdot 2^{-r}}} Y^{(2,r)}_{K_s-K_t,n} \Big\|_{\Psi_{1/k}} \leq \frac{C}{n^{1/2} h_n^{(p-1)/p}}.$$

Substitution of these estimates in (3.42), gives

$$\Pr\Big\{\sup_{\lambda\in T} Y_{K_\lambda,n} > x\Big\} \leq \frac{1}{\Psi_{2/k}(Cxn^{(k-1)/2} h_n^{1/2})} + \frac{1}{\Psi_{1/k}(Cxn^{1/2} h_n^{(p-1)/p})}. \tag{3.43}$$

Since $\Psi_\alpha(u) \geq 2^{-1} e^{u^\alpha}$ for all u larger than a number u_α depending only on α, and the arguments of the Ψ functions tend to infinity, it follows from the previous inequality that there exists a constant C independent of n such that, for all $x \geq 0$,

$$\Pr\Big\{\sup_{\lambda\in T} Y_{K_\lambda,n} > x\Big\} \leq C \exp\Bigg\{ -\frac{1}{C}\Big[(xn^{(k-1)/2} h_n^{1/2})^{2/k} \wedge (xn^{1/2} h_n^{(p-1)/p})^{1/k}\Big]\Bigg\}.$$

By (3.37), this gives the exponential inequality (3.29) in Lemma 2. The moment inequality (3.30) follows from it upon observing that

$$E\Big(\sup_{\lambda\in T} Y_{K_\lambda,n}\Big)^2 = 2\int_0^\infty x \Pr\Big\{\sup_{\lambda\in T} Y_{K_\lambda,n} > x\Big\} dx$$

and

$$\Pr\Big\{\sup_{\lambda\in T} Y_{K_\lambda,n} > x\Big\}$$

$$\leq C\exp\Big\{-\frac{1}{C}(xn^{(k-1)/2}h_n^{1/2})^{2/k}\Big\} + C\exp\Big\{-\frac{1}{C}(xn^{1/2}h_n^{(p-1)/p})^{1/k}\Big\}$$

and noting that for $n \geq n_0$ (for some finite n_0) and $k \geq 2$,

$$(n^{(k-1)/2}h_n^{1/2})^{2/k} \wedge (n^{1/2}h_n^{(p-1)/p})^{1/k} \geq (n^{1/2}h_n^{1/2})^{1/k}.$$

q.e.d.

References

de la Peña, V. H. and Giné, E. (1999). *Decoupling. From dependence to independence. Randomly stopped processes. U−statistics and processes. Martingales and beyond.* Probability and its Applications. Springer–Verlag, New York.

Brown, B (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42**, 59-66.

Deheuvels, P., Einmahl, U. and Mason, D. M. (2000). Asymptotic independence of the local empirical process indexed by functions. High Dimensional Probability, II (Seattle, WA, 1999), 183–206, *Progr. Probab.*, **47**, Birkhäuser, Boston, MA.

Deheuvels, P. and Mason, D. M. (1992) Functional laws of the iterated logarithm for increments of empirical and quantile processes. *Ann. Probab.* **20**, 1248–1287.

Deheuvels, P. and Mason, D. M. (1994). Functional laws of the iterated logarithm for local empirical processes indexed by sets. *Ann. Probab.* **22** 1619-1661.

Deheuvels, P. and Mason, D. M. (2004). General confidence bounds for nonparametric functional estimators. *Statistical Inference for Stochastic Processes,* **7**, 225-277.

Devroye, L. and Györfi, L. (1985) *Nonparametric Density Estimation: The L_1 View.* Wiley, New York.

Dudley, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge Studies in Advanced Mathematics, 63. Cambridge University Press, New York.

Dunford, N. and Schwartz, J. T. (1966). *Linear Operators, Part I.* Third Printing. Wiley-Interscience, New York.

Einmahl, U. and Mason, D. M. (1997). Gaussian approximation of local empirical processes indexed by functions. *Probab. Th. Rel. Fields,* **107,** 283–311.

Einmahl, U. and Mason, D. M. (1998). Strong approximations for local empirical processes. High Dimensional Probability, II (Oberwolfach, 1996), 75–02, *Progr. Probab.*, **43**, Birkhäuser, Basel

Einmahl, U. and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoretical Prob.* **13**, 1–37.

Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33**, 1380–1403.

Frees, W. (1994). Estimating densities of functions of observations. *J. Amer. Statist. Assoc.* **89**, 517–525.

Giné, E. and Guillou, A. (2002). Rates of strong consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré* **38,** 907-921.

Giné, E., Latała, R.and Zinn, J. (2000). Exponential and moment inequalities for $U-$statistics. High Dimensional Probability, II (Seattle, WA, 1999), 13–38, *Progr. Probab.*, **47**, Birkhäuser, Boston, MA.

Giné, E. and Mason, D. M. (2004). The law of the iterated logarithm for the integrated squared deviation of a kernel density estimator. *Bernoulli* **10,** 721-752.

Giné, E. and Mason, D. M. (2005). On local $U-$statistic processes and the estimation of densities of functions of several sample variables. Submitted for publication.

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

Kuelbs, J. (1976). A strong convergence theorem for Banach space valued random variables. *Ann. Probab.* **4**, 744-771.

Ledoux, M. and Talagrand, M. (1988) Characterization of the law of the iterated logarithm in Banach spaces. *Ann. Probab.* **16**, 1242–1264.

Ledoux, M. and Talagrand, M. (1989) Comparison theorems, random geometry and some limit theorems for empirical processes. *Ann. Probab.* **17**, 596–631.

Ledoux, M. and Talagrand, M. (1991). *Probability on Banach Spaces,* Springer–Verlag, Berlin.

Major, P. (2005). An estimate on the supremum of a nice class of stochastic integrals and U-statistics. *Probab. Th. Rel. Fields.*, the online version.

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

Rost, D. (2000). Limit theorems for smoothed empirical processes. High Dimensional Probability, II (Seattle, WA, 1999), 107–113, *Progr. Probab.,* **47**, Birkhäuser, Boston, MA.

Schick, A. and Wefelmeyer, W. (2004). Root n consistent density estimators for sums of independent random variables. *J. Nonparametr. Stat.* **16**, 925–935.

Shorack, G. R. and Wellner, J. A. (1986). *Emprical Processes with Applications to Statistics.* Wiley Series in Probability and Mathematical Statistics. John Wiley, New York.

van der Vaart, A. (1994). Weak convergence of smoothed empirical processes. *Scand. J. Statist.* **21**, 501–504.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics,* Springer–Verlag, New York.

Yukich, J. E. (1992). Weak convergence of smoothed empirical processes. *Scand. J. Statist.* **19**, 271–279.

Departement of Mathematics
University of Connecticut
196 Auditorium Road
Storrs, CT 06269-3009
USA
gine@math.uconn.edu

Food and Resource Economics
University of Delaware
206 Townsend Hall
Newark, DE 19717
USA
davidm@Udel.Edu