THE LIL FOR CANONICAL U-STATISTICS OF ORDER 2

BY EVARIST GINÉ*, STANISLAW KWAPIEŃ†, RAFAŁ LATAŁA†, AND JOEL ZINN‡

University of Connecticut, Warsaw University, Warsaw University and Texas A&M University

Abstract

Let X, X_i, $i \in \mathbf{N}$, be independent identically distributed random variables and let $h(x,y) = h(y,x)$ be a measurable function of two variables. It is shown that the bounded law of the iterated logarithm, $\limsup_n (n \log\log n)^{-1} \left| \sum_{1 \le i < j \le n} h(X_i, X_j) \right| < \infty$ a.s., holds if and only if the following three conditions are satisfied: h is canonical for the law of X (that is, $Eh(X,y)=0$ for almost all y) and there exists $C<\infty$ such that, both, $E(h^2(X_1,X_2) \wedge u) \le C \log\log u$ for all large u and $\sup\{Eh(X_1,X_2)f(X_1)g(X_2) : \|f(X)\|_2 \le 1, \|g(X)\|_2 \le 1, \|f\|_\infty < \infty, \|g\|_\infty < \infty\} \le C$.

May 1999

AMS 1991 *subject classifications*. Primary: 60F15.

Key words and phrases: U-statistics (canonical or degenerate), law of the iterated logarithm.

* Research partially supported by NSF Grant No. DMS-96–25457.

† Research partially supported by Polish Grant KBN 2 PO3A 043 15.

‡ Research partially supported by NSF Grant No. DMS-96-26778.

1. Introduction. Although U–statistics (Halmos, 1946; Hoeffding, 1948) are relatively simple probabilistic objects, namely averages over an i.i.d. sample $X_1, \ldots, X_n$ of measurable functions (kernels) $h(x_1, \ldots, x_m)$ of several variables, their asymptotic theory is only recently attaining a satisfactory degree of completeness: see e.g. Rubin and Vitale (1980), Giné and Zinn (1994), Zhang (1999) and Latała and Zinn (1999) on necessary and sufficient conditions for the central limit theorem and the law of large numbers. We are interested here in the law of the iterated logarithm for U-statistics based on canonical (or completely degenerate) kernels, that is, on kernels whose conditional expectation given any $m-1$ variables is zero, and only for $m = 2$.

U-statistics with nondegenerate kernels behave, as is well known, like sums of independent random variables, and the LIL in this case was proved by Serfling (1971). The LIL for canonical (or completely degenerate) kernels h with finite absolute moment of order $2+\delta$, $\delta > 0$, was obtained by Dehling, Denker and Philipp (1984, 1986), and with finite second moment by Dehling (1989) and Arcones and Giné (1995). Giné and Zhang (1996) showed that there exist degenerate kernels h with infinite second moment such that, nevertheless, the corresponding U-statistics satisfy the law of the iterated logarithm, and obtained a necessary integrability condition as well. This last article and Goodman's (1996) also contain LIL's under assumptions that do not imply finiteness of the second moment of h, but that fall quite short from being necessary. The LIL for finite sums of products $\sum_{i=1}^{k} \lambda_i \phi_i(x_1) \cdots \phi_i(x_m)$ is easier ($Eh^2 < \infty$ is necessary) and was considered by Teicher (1995) for $k = 1$ and by Giné and Zhang (1996) for any $k < \infty$. In the present article the bounded LIL problem is solved for kernels of order 2. Next we describe our result and comment on its (relatively involved) proof.

In what follows, X, X_i, $i \in \mathbf{N}$, are independent identically distributed random variables taking values on some measurable space $(S, \mathcal{S})$, and $h : S^2 \mapsto \mathbf{R}$ is a measurable function that we assume, without loss of generality (for our purposes), symmetric in its entries, that is, $h(x,y) = h(y,x)$ for all $x, y \in S$. When h is integrable we say that it is canonical, or degenerate, for the law of X if $Eh(X,y) = 0$ for almost all $y \in S$ (relative to the law of X). The natural LIL normalization for U-statistics corresponding to degenerate kernels of order 2 is $n \log\log n$ as is seen with the following example. A simple canonical kernel for $S = \mathbf{R}$ and X integrable with $EX = 0$ is $h(x,y) = xy$. For this example, if moreover $EX^2 < \infty$ then, by the LIL and the law of large numbers for sums of independent random variables, we have

$$\limsup_n \frac{1}{2n\log\log n}\Big|\sum_{i\neq j\leq n} X_i X_j\Big| = \limsup_n \Big[\frac{1}{\sqrt{2n\log\log n}}\sum_{i=1}^{n} X_i\Big]^2 = \mathrm{Var} X.$$

Our main result is as follows:

THEOREM 1.1. *Let X, Y, X_i, $i \in \mathbf{N}$, be i.i.d. random variables taking values in $(S, \mathcal{S})$ and let $h : S^2 \mapsto \mathbf{R}$ be a measurable function of two variables. Then,*

$$\limsup_n \frac{1}{n\log\log n}\Big|\sum_{1\leq i\neq j\leq n} h(X_i, X_j)\Big| < \infty \quad \text{a.s.} \tag{1.1}$$

if an only if the following three conditions hold:
a) h is canonical for the law of X
and there exists $C < \infty$ such that
b) for all $u \geq 10$,

$$E(h^2(X,Y) \wedge u) \leq C \log\log u, \tag{1.2}$$

and
c)

$$\sup\bigl\{Eh(X,Y)f(X)g(Y) : Ef^2(X) \leq 1, Eg^2(X) \leq 1, \\ \|f\|_\infty < \infty, \|g\|_\infty < \infty\bigr\} \leq C. \tag{1.3}$$

It is easily seen that condition b) implies

$$E\frac{h^2}{(\log\log(|h| \vee e^e)^{1+\delta}} < \infty \tag{1.4}$$

for all $\delta > 0$ (and is implied by $Eh^2/\log\log(|h| \vee e^e) < \infty$. In particular condition b) ensures the existence of the integrals in conditions a) and c). Condition c) implies that the operator defined on $L_\infty(\mathcal{L}(X))$ by $Hf(y) = Eh(X,y)f(X)$ takes values in $L_2(\mathcal{L}(X))$ and extends as a bounded operator to all of $L_2(\mathcal{L}(X))$. Moreover, if with a slight abuse of notation we set $E_X h(X,Y)f(X) := Hf(Y)$ for $f \in L_2$, then condition b) is equivalent to

$$E_Y\bigl(E_X h(X,Y)f(X)\bigr)^2 \leq C^2 Ef^2(X) \ \text{ for all } f \in L_2. \tag{1.5}$$

(Here and in what follows, E_X (resp. E_Y) indicates expectation with respect to X (resp. Y) only.)

The integrability condition b) was proved to be necessary for the LIL (1.1) by Giné and Zhang (1996), whereas the idea for condition c) comes from Dehling (1989) who showed that if $h(x,y)$ is canonical and square integrable then

$$\lim\text{set}\left\{\frac{1}{2n\log\log n}\sum_{1\leq i\neq j\leq n} h(X_i,X_j)\right\} \\ = \bigl\{Eh(X,Y)f(X)f(Y) : Ef^2(X) \leq 1\bigr\} \ \text{ a.s.}$$

We will not prove Theorem 1.1 directly, but instead we will prove first that conditions b) and c) are necessary and sufficient for a decoupled and randomized version of the LIL, namely, for

$$\limsup_n \frac{1}{n\log\log n}\left|\sum_{1\leq i,j\leq n} \varepsilon_i\tilde{\varepsilon}_j h(X_i,Y_j)\right| < \infty \ \text{ a.s.}, \tag{1.6}$$

where $\{\varepsilon_i\}$ is a Rademacher sequence independent of all the other variables. (We recall that a Rademacher sequence is a sequence of independent random variables taking on only the values 1 and -1, each with probability 1/2.) The reasons for this are multiple. One is that necessity of condition c) follows as a consequence of a recent result of Latała (1999) on estimation of tail probabilities of Rademacher chaos variables. Another reason is that, because of the Rademacher multipliers, truncation of the kernel will result in symmetric, and hence mean zero, variables; this is important since the proof of sufficiency contains several relatively complicated truncations of h. Moreover, part of the core of the proof of sufficiency consists of an iterative application of an exponential bound for sums of independent random variables and vectors, and having decoupled expressions makes this iteration possible (although we could use, alternatively, an exponential inequality for martingale differences that does not require decoupled expressions).

The exponential inequality in question is Talagrand's (1996) uniform Prohorov inequality. This inequality depends on two parameters, the L_∞ bound of the variables and the weak variance of their sum, and to apply it iteratively requires not only that h be truncated at a low level, but that the conditional second moments of these truncations of h be small as well. This explains the relatively complicated multi-step truncation procedure in the proof of sufficiency.

Finally, the limit (1.6) will imply the limit (1.1) by a two stage symmetrization argument that will also require control of the conditional expectations of the sums; this control will be achieved once more, again after multiple truncations, by means of Talagrand's exponential inequality.

Section 2 contains several known results needed in the sequel. Section 3 is devoted to the proof of the LIL for decoupled, randomized kernels, and Section 4 reduces the LIL for canonical kernels to this case. In Section 5 we complete the proof of Theorem 1.1 and make several comments about the limsup in (1.1) and the limit set of the LIL sequence.

We adhere in what follows to the following notation (some of it already set up above):

- ⋄ h is a measurable real function of two variables defined on $(S^2, \mathcal{S}\otimes\mathcal{S})$, symmetric in its entries.
- ⋄ $X, X_1, X_2, \ldots$ and $Y, Y_1, Y_2, \ldots$ denote two independent, equidistributed sequences of i.i.d. S-valued random variables.
- ⋄ We write $Ef(h)$ for $Ef(h(X,Y))$, and E_X, $\Pr_X$ (resp. E_Y, $\Pr_Y$) denote expected value and probability with respect to the random variables X, X_i (resp. Y, Y_i) only.
- ⋄ $\varepsilon_1, \varepsilon_2, \ldots$, and $\tilde\varepsilon_1, \tilde\varepsilon_2, \ldots$ are two independent Rademacher sequences, independent of all other random variables.
- ⋄ We write L_2x and L_3x instead of $L(L(x))$ and $L(L(L(x)))$, where $L(x) = \max(\log x, 1)$.
- ⋄ In all proofs $\tilde C$ denotes a universal constant which may change from line to line but does not depend on any parameters.

2. Preliminary results. For convenience, we isolate in this section several known results needed below.

(A) Hoeffding's decomposition. The U-statistics with kernel h (not necessarily symmetric in its entries) based on $\{X_i\}$ are defined as

$$U_n(h) = \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} h(X_i, X_j), \quad n \in \mathbf{N}.$$

By considering instead the kernel $\tilde{h}(x,y) = \big(h(x,y)+h(y,x)\big)/2$, we have

$$U_n(h) = U_n(\tilde{h}) = \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} \tilde{h}(X_i,X_j) = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \tilde{h}(X_i,X_j).$$

So, *we will assume h symmetric in its entries* in all that follows.

Suppose $E|h(X,Y)| < \infty$. Then,

$$\begin{aligned} h(x,y) - Eh(X,Y) &= \big[h(x,y) - E_Y h(x,Y) - E_X h(X,y) + Eh(X,Y)\big] \\ &\quad + \big[E_Y h(x,Y) - Eh(X,Y)\big] + \big[E_X h(X,y) - Eh(X,Y)\big] \\ &:= \pi_2 h(x,y) + \pi_1 h(x) + \pi_1 h(y), \end{aligned} \tag{2.1}$$

where the identities hold a.s. for $\mathcal{L}(X)\times\mathcal{L}(X)$. The kernel $\pi_2 h$ is canonical (or degenerate) for the law of X as $E_X \pi_2 h(X,Y) = E_Y \pi_2 h(X,Y) = 0$ a.s., and $\pi_1 h(X)$ is centered. This decomposition of h gives rise to *Hoeffding's decomposition* of the corresponding U-statistics,

$$\sum_{1\le i<j\le n} h(X_i,X_j) = \sum_{1\le i<j\le n} \pi_2 h(X_i,X_j) + (n-1)\sum_{i=1}^n \pi_1 h(X_i) + \binom{n}{2} Eh(X,Y), \tag{2.2}$$

and of their decoupled versions,

$$\begin{aligned} \sum_{1\le i,j\le n} h(X_i,Y_j) = \sum_{1\le i,j\le n} \pi_2 h(X_i,Y_j) + n\sum_{i=1}^n \pi_1 h(X_i) \\ + n\sum_{i=1}^n \pi_1 h(Y_i) + n^2 Eh(X,Y). \end{aligned} \tag{2.3}$$

(B) The equivalence of several LIL statements. The following lemma contains necessary randomization and integrability conditions for the LIL:

LEMMA 2.1. (Giné and Zhang, 1996). *(a) (*Integrability.*) There exists a universal constant K such that, if*

$$\sum_{n=1}^\infty \Pr\Big\{\frac{1}{2^n Ln}\Big|\sum_{1\le i,j\le 2^n} \varepsilon_i\tilde{\varepsilon}_j h(X_i,Y_j)\Big| > C\Big\} < \infty \tag{2.4}$$

for some $C<\infty$, then

$$\limsup_{u\to\infty} \frac{E\big(h^2(X,Y)\wedge u\big)}{L_2 u} \le KC^2. \tag{2.5}$$

*(b) (*Randomization and decoupling, partial.*) The LIL*

$$\limsup_n \frac{1}{nL_2 n}\Big|\sum_{1\le i<j\le n} h(X_i,X_j)\Big| \le C \quad \text{a.s.} \tag{2.6}$$

for some $C < \infty$ implies

$$\sum_{n=1}^{\infty} \Pr\left\{ \frac{1}{2^n Ln} \max_{k \le 2^n} \left| \sum_{1 \le i,j \le k} \varepsilon_i \tilde{\varepsilon}_j h(X_i, Y_j) \right| > 2^7 C \right\} < \infty.$$

In particular, the LIL implies both the integrability condition (2.5) and the randomized and decoupled LIL, that is,

$$\limsup_n \frac{1}{nL_2 n} \left| \sum_{1 \le i,j \le n} \varepsilon_i \tilde{\varepsilon}_j h(X_i, Y_j) \right| \le D \quad \text{a.s.} \tag{2.7}$$

with $D = KC$ for some universal constant K.

Part (a) is contained in the proof of Theorem 3.1 in Giné and Zhang (1996), while part (b) is the content of Theorem 3.1 and Lemma 3.3 there.

We recall that the limsups at the left hand sides of (2.6) and (2.7) are always a.s. constant (finite or infinite) by the Hewitt-Savage zero-one law.

Decoupling gives the following equivalence between the LIL and its decoupled version.

Lemma 2.2. *(a) The LIL (2.6) is equivalent to the decoupled LIL, that is, to*

$$\limsup_n \frac{1}{nL_2 n} \left| \sum_{1 \le i \ne j \le n} h(X_i, Y_j) \right| \le D \quad \text{a.s.} \tag{2.8}$$

for some $D < \infty$, meaning that if (2.6) holds for C then (2.8) holds for $D = KC$ and that if (2.8) holds for D then (2.6) holds for $C = KD$, where K is a universal constant.
(b) The decoupled and randomized LIL (2.7) is equivalent to the randomized LIL

$$\limsup_n \frac{1}{nL_2 n} \left| \sum_{1 \le i \ne j \le n} \varepsilon_i \varepsilon_j h(X_i, X_j) \right| \le C \quad \text{a.s.} \tag{2.9}$$

for some C finite (with C and D related as in part (a)).
(c) The LIL (2.7) implies convergence of the series (2.4) for some $C = KD < \infty$, K a universal constant, hence it also implies the integrability condition (2.5) (with C replaced by D).

Proof. (a) We can equivalently write (2.6) as

$$\lim_{k \to \infty} \Pr\left\{ \sup_{n \ge k} \frac{1}{nL_2 n} \left| \sum_{1 \le i \ne j \le n} h(X_i, X_j) \right| \ge C \right\} = 0$$

for some $C < \infty$, hence as

$$\lim_{k \to \infty} \Pr\left\{ \left\| \sum_{1 \le i \ne j < \infty} h_{i \vee j, k}(X_i, X_j) \right\| \ge C \right\} = 0,$$

where

$$h_{i,k} := \left(\frac{h}{kL_2 k}, \frac{h}{(k+1)L_2(k+1)}, \ldots, \frac{h}{nL_2 n}, \ldots \right)$$

if $i \le k$ and

$$h_{i,k} := \left(0, \overset{i-k}{\ldots}, 0, \frac{h}{iL_2 i}, \frac{h}{(i+1)L_2(i+1)}, \ldots, \frac{h}{nL_2 n}, \ldots\right)$$

if $i > k$ are ℓ_∞-valued functions and $\|\cdot\|$ denotes the sup of the coordinates. Then, the decoupling inequalities of de la Peña and Montgomery-Smith (1994) apply to show that the above tail probabilities are equivalent up to constants to those of the corresponding decoupled expressions, thus giving the equivalence between (2.6) and (2.8).

(b) If (2.9) holds, then (2.7) without diagonal terms (that is, without the summands corresponding to $i = j$) holds too by the first part of the proof applied to the kernel $\alpha\beta h(x, y)$. Moreover, (2.9) implies the integrability condition (2.5) by Lemma 2.1 (note that if $\{\varepsilon_i^{(j)}\}$, $j = 1, 2, 3$, are three independent Rademacher sequences, then $\{\varepsilon_i^{(1)}\varepsilon_i^{(2)}\}$ and $\{\varepsilon_i^{(1)}\varepsilon_i^{(3)}\}$ are also independent Rademacher sequences) and, as a consequence, h is integrable. Hence, by the law of large numbers, the diagonal in (2.7) is irrelevant, showing that (2.7) holds with the diagonal included. If (2.7) holds, then we also have $E|h| < \infty$: a modification of the proof of the converse central limit theorem in Giné and Zinn (1994), consisting in replacing use of the law of large numbers by use of inequality (3.7) in Giné and Zhang (1996), shows that if the sequence $\left\{(nL_2 n)^{-1}\sum_{i,j\le n}\varepsilon_i\tilde{\varepsilon}_j h(X_i, Y_j)\right\}$ is stochastically bounded, then $Eh^2(X, Y) \wedge u \le C(L_2 u)^2$ for some $C < \infty$, in particular, that $E|h| < \infty$. So, we can delete the diagonal in (2.7), and then apply the first part of the lemma to undo the decoupling.

(c) Statement (c) follows from (b) because, by Lemma 2.1, (2.9) implies convegence of the series (2.4) for some $C < \infty$. □

The following lemma, together with the previous ones, will allow blocking and will reduce the proof of sufficiency of the LIL to showing that a series of tail probabilities converges (just as with sums of i.i.d random variables).

Lemma 2.3. *There exists a universal constant $C < \infty$ such that for any kernel h and any two sequences X_i, Y_j of i.i.d. random variables we have*

$$\Pr\left\{\max_{k\le m, l\le n}\Big|\sum_{i\le k, j\le l} h(X_i, Y_j)\Big| \ge t\right\} \le C\Pr\left\{\Big|\sum_{i\le m, j\le n} h(X_i, Y_j)\Big| \ge t/C\right\} \quad (2.10)$$

for all $m, n \in \mathbf{N}$ and for all $t > 0$.

Proof. Montgomery-Smith's (1993) maximal inequality for i.i.d. sums asserts that if Z_i are i.i.d. r.v.'s with values in some Banach space B then for some universal constant C_1 and all $t > 0$ we have

$$\Pr\left\{\max_{k\le m}\Big\|\sum_{i\le k} Z_i\Big\| \ge t\right\} \le C_1\Pr\left\{\Big\|\sum_{i\le m} Z_i\Big\| \ge t/C_1\right\}.$$

We apply this inequality to $B = \ell_\infty^n$ and $Z_i = \left(\sum_{j\le l} h(X_i, y_j) : l \le n\right)$ for fixed values of $y_1, \ldots, y_n$ to get

$$\Pr\left\{\max_{k\le m, l\le n}\left|\sum_{i\le k, j\le l} h(X_i, Y_j)\right| \ge t\right\} \le C_1 \Pr\left\{\max_{l\le n}\left|\sum_{i\le m, j\le l} h(X_i, Y_j)\right| \ge t/C_1\right\}.$$

In a similar way we may prove

$$\Pr\left\{\max_{l\le n}\left|\sum_{i\le m, j\le l} h(X_i, Y_j)\right| \ge t/C_1\right\} \le C_1 \Pr\left\{\sum_{i\le m, j\le n}\left|h(X_i, Y_j)\right| \ge t/C_1^2\right\}.$$

Thus the assertion holds with $C = C_1^2$. □

Corollary 2.4. *If*

$$\sum_{n=1}^{\infty} \Pr\left\{\frac{1}{2^n Ln}\left|\sum_{1\le i,j\le 2^n} h(X_i, Y_j)\right| > C\right\} < \infty \quad \text{a.s.} \tag{2.11}$$

for some $C < \infty$, then there is a universal constant K such that

$$\limsup_n \frac{1}{nL_2 n}\left|\sum_{1\le i,j\le n} h(X_i, Y_j)\right| \le KC \quad \text{a.s.} \tag{2.12}$$

Proof. Since, for any $0 < D < \infty$,

$$\begin{aligned}
&\Pr\left\{\sup_{n\ge N}\frac{1}{nL_2 n}\left|\sum_{1\le i,j\le n} h(X_i, Y_j)\right| > D\right\} \\
&\quad\le \Pr\left\{\sup_{k>[\log N/\log 2]}\ \max_{2^{k-1}\le n\le 2^k} \frac{3}{2^k Lk}\left|\sum_{1\le i,j\le n} h(X_i, Y_j)\right| > D\right\} \\
&\quad\le \sum_{k>[\log N/\log 2]} \Pr\left\{\max_{2^{k-1}\le n\le 2^k}\left|\sum_{1\le i,j\le n} h(X_i, Y_j)\right| > \frac{D2^k Lk}{3}\right\},
\end{aligned}$$

the result follows from Lemma 2.3. □

Applying Corollary 2.4 to the kernel $\alpha\beta h(x,y)$ we obtain the converse of Lemma 2.2(c). Hence,

Corollary 2.5. *Consider the statements*

$$\limsup_n \frac{1}{nL_2 n}\left|\sum_{1\le i,j\le n} \varepsilon_i \tilde{\varepsilon}_j h(X_i, Y_j)\right| \le C \quad \text{a.s.}$$

and

$$\sum_{n=1}^{\infty} \Pr\left\{\frac{1}{2^n Ln}\left|\sum_{1\le i,j\le 2^n} \varepsilon_i \tilde{\varepsilon}_j h(X_i, Y_j)\right| > D\right\} < \infty.$$

There is a universal constant K such that if the first statement holds for some $C < \infty$ then the second holds for $D = KC$, and conversely, if the second holds for some $D < \infty$ then so does the first, for $C = KD$.

We will also require the following partial converse to Lemma 2.1(b) regarding the regular LIL and convergence of series of tail probabilities:

COROLLARY 2.6. *Suppose $E|h| < \infty$. If*

$$\sum_{n=1}^{\infty} \Pr\left\{ \frac{1}{2^n Ln} \Big| \sum_{1 \le i,j \le 2^n} h(X_i, Y_j) \Big| > C \right\} < \infty \quad \text{a.s.}$$

for some $C < \infty$ then the LIL holds, that is, there is a universal constant K such that

$$\limsup_n \frac{1}{nL_2 n} \Big| \sum_{1 \le i < j \le n} h(X_i, X_j) \Big| \le KC \quad \text{a.s.}$$

PROOF. Convergence of the series implies (2.12), that is, the decoupled LIL with diagonal terms included. Since $E|h| < \infty$, the diagonal terms are irrelevant and therefore the decoupled LIL (2.8) holds. The result now follows from Lemma 2.2(a). □

In Section 4 we will apply the conclusion of Corollary 2.6 under the assumption that the decoupled and randomized LIL (2.7) holds: this is possible because (2.7) implies integrability of h, as indicated in the proof of Lemma 2.2(b).

(C) Inequalities. As mentioned in the Introduction, the following two inequalities will play a basic role in the proof of Theorem 1.1. The first consists of a sharp estimate of the tail probabilities of Rademacher chaos variables (it is in fact part of a sharper two sided estimate).

LEMMA 2.7. (Latała, 1999). *There exists a universal constant $c > 0$ such that, for all matrices $(a_{i,j})$ and for all $t > 0$,*

$$\Pr\left\{ \Big| \sum_{i,j} a_{i,j} \varepsilon_i \tilde{\varepsilon}_j \Big| \ge c |||(a_{i,j})|||_t \right\} \ge c \wedge e^{-t}, \tag{2.13}$$

where $|||(a_{i,j})|||_t$ is defined as

$$|||(a_{i,j})|||_t := \sup\left\{ \sum_{i,j} a_{i,j} b_i c_j : \sum_i b_i^2 \le t, \sum_j c_j^2 \le t, |b_i|, |c_j| \le 1 \text{ for all } i,j \right\}. \tag{2.14}$$

The second is a uniform Prohorov inequality due to Talagrand. It combines Theorem 1.4 in Talagrand (1996) with Corollary 3.4 in Talagrand (1994).

LEMMA 2.8. (Talagrand, 1996). *Let $\{X_i\}$, $i = 1, \ldots, n$ for any $n \in \mathbf{N}$, be independent random variables with values in a measurable space $(S, \mathcal{S})$, let $\mathcal{F}$ be a countable class of measurable functions on S and let*

$$Z := \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(X_i).$$

There exists a universal constant K such that for all $t>0$ and $n\in\mathbf{N}$, if

$$\max_{1\le i\le n}\sup_{f\in\mathcal{F}}\text{ess sup}_{\omega\in\Omega}\big|f(X_i(\omega))\big|\le U,\quad E\Big(\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}f^2(\mathbf{X}_i)\Big)\le V$$

and

$$\sup_{f\in\mathcal{F}}\sum_{i=1}^{n}Ef^2(X_i)\le\sigma^2,$$

then

$$\begin{aligned}\Pr\Big\{|Z-EZ|\ge t\Big\}&\le K\exp\Big(-\frac{t}{KU}\log\Big(1+\frac{tU}{V}\Big)\Big)\\&\le K\exp\Big(-\frac{t}{KU}\log\Big(1+\frac{tU}{\sigma^2+8UE|Z|}\Big)\Big).\end{aligned}\tag{2.15}$$

In fact, we will only use the corresponding deviation inequality, that is, the bound (2.5) for $\Pr\{Z>EZ+t\}$. Ledoux (1987) contains a simple proof of this result based on logarithmic Sobolev inequalities.

When $\mathcal{F}$ consists of a single function f and the variables $f(X_i)$ are centered this inequality reduces, modulo constants, to the classical Prohorov inequality. For convenience, we will refer below to Lemma 2.8 even in cases when Prohorov's inequality suffices.

3. Symmetrized kernels. In this section we prove the following theorem, which constitutes the basic component of the proof of Theorem 1.1.

THEOREM 3.1. *The decoupled and randomized LIL holds, that is,*

$$\limsup_n\frac{1}{n\log\log n}\Big|\sum_{1\le i,j\le n}\varepsilon_i\tilde{\varepsilon}_jh(X_i,Y_j)\Big|<\infty\quad\text{a.s.}\tag{3.1}$$

if and only if the following two conditions are satisfied for some $C<\infty$:

$$E\min(h^2,u)\le CL_2u\quad\text{for all } u>0,\tag{3.2}$$

and

$$\begin{aligned}\sup\Big\{Eh(X,Y)f(X)g(Y):Ef^2(X)\le 1,Eg^2(Y)\le 1,\\\|f\|_\infty<\infty,\|g\|_\infty<\infty\Big\}\le C<\infty.\end{aligned}\tag{3.3}$$

REMARK. We recall that, by Corollary 2.5, a necessary and sufficient condition for the LIL (3.1) to hold is that

$$\sum_{n=1}^{\infty}\Pr\Big\{\frac{1}{2^nLn}\Big|\sum_{1\le i,j\le 2^n}\varepsilon_i\tilde{\varepsilon}_jh(X_i,Y_j)\Big|>C\Big\}<\infty\tag{3.4}$$

for some $C<\infty$.

PROOF OF NECESSITY. The integrability condition (3.2) is necessary for (3.1) by Lemma 2.2(c). The necessity of (3.3) will follow from Lemma 2.7. For this, we estimate first $\left|\left|\left|\big(h(X_i,Y_j) : i,j \le 2^n\big)\right|\right|\right|_{\log n}$, where $|||\cdot|||_t$ is as defined in (2.13). Suppose that $f,g \in L_\infty$ are such that $Ef^2(X) = Eg^2(X) = 1$ and set

$$K := |Eh(X,Y)f(X)g(Y)|, \tag{3.5}$$

that we can assume strictly positive. Note that the integral exists by (3.2). Then by the SLLN for i.i.d. r.v.'s and U-statistics we have a.s.

$$n^{-1}\sum_{i\le n} f^2(X_i) \to Ef^2 = 1, \quad n^{-1}\sum_{j\le n} g^2(Y_j) \to Eg^2 = 1$$

and

$$n^{-2}\Big|\sum_{i,j\le n} h(X_i,Y_j)f(X_i)g(Y_j)\Big| \to \big|Eh(X,Y)f(X)g(Y)\big|.$$

So, for large enough n,

$$\Pr\Big\{2^{-n}\sum_{i\le 2^n} f^2(X_i) \le 2\Big\} \ge \frac{3}{4}, \quad \Pr\Big\{2^{-n}\sum_{j\le 2^n} g^2(Y_j) \le 2\Big\} \ge \frac{3}{4}$$

and

$$\Pr\Big\{2^{-2n}\Big|\sum_{i,j\le 2^n} h(X_i,Y_j)f(X_i)g(Y_j)\Big| \ge K/2\Big\} \ge \frac{3}{4}$$

with K as in (3.5). Since $f,g \in L_\infty$ we have that, for large enough n,

$$\left|\sqrt{\frac{\log n}{2^{n+1}}}f(X_i)\right|, \left|\sqrt{\frac{\log n}{2^{n+1}}}g(Y_j)\right| \le 1 \quad \text{a.s.}$$

Then, it follows directly from the definition of $|||\cdot|||_t$ that, on the intersection of the above five events, we have the bound

$$\left|\left|\left|\big(h(X_i,Y_j) : i,j \le 2^n\big)\right|\right|\right|_{\log n} \ge K2^{n-2}\log n.$$

Therefore, for large n,

$$\Pr\Big\{\left|\left|\left|\big(h(X_i,Y_j) : i,j \le 2^n\big)\right|\right|\right|_{\log n} \ge K2^{n-2}\log n\Big\} \ge \frac{1}{4}.$$

Then, Lemma 2.7 implies that, for all n large enough,

$$\Pr\Big\{\Big|\sum_{i,j\le 2^n} h(X_i,Y_j)\varepsilon_i\tilde{\varepsilon}_j\Big| \ge cK2^{n-2}\log n\Big\} \ge \frac{1}{4}e^{-\log n} = \frac{1}{4n}.$$

By (3.4), this implies that if the LIL holds then K is uniformly bounded, proving necessity of condition (3.3). □

Before starting the proof of sufficiency, it is convenient to show how the integrability condition (3.2) limits the sizes of certain truncated conditional second moments. To simplify notation, we define

$$f_n(x) = E_Y \min\big(h^2(x,Y), 2^{4n}\big) \quad \text{and} \quad f_n(y) = E_X \min\big(h^2(X,y), 2^{4n}\big). \tag{3.6}$$

LEMMA 3.2. *For any kernel h satisfying condition (3.2) we have that, for all $a > 0$,*

$$\sum_n 2^n \Pr_X\Big\{E_Y \min\big(h^2(X,Y), 2^{an}\big) \geq 2^n (\log n)^2\Big\} < \infty. \tag{3.7}$$

Moreover,

$$\sum_n \frac{2^n}{(\log n)^k} \Pr\Big\{f_n(X) \geq 2^n (\log n)^{2-k}\Big\} < \infty \quad \textit{for all } k \geq 0. \tag{3.8}$$

PROOF. For a fixed, we set $\gamma_k = \exp(2^{k+1})$ and $\tilde{f}_k(X) = E_Y \min(h^2, 2^{a\gamma_k})$. Then,

$$\begin{aligned}
\sum_{2^k \leq \log n \leq 2^{k+1}} & 2^n \Pr_X\Big\{E_Y \min\big(h^2(X,Y), 2^{an}\big) \geq 2^n (\log n)^2\Big\} \\
&\leq \sum_{2^k \leq \log n \leq 2^{k+1}} 2^n \Pr_X\{\tilde{f}_k(X) \geq 2^{n+2k}\} \\
&\leq E \sum_n 2^n I\big(\tilde{f}_k(X) \geq 2^{n+2k}\big) \\
&\leq 2^{1-2k} E\tilde{f}_k(X) \leq 2^{1-2k} C L_2(2^{a\gamma_k}) \\
&\leq 2^{1-2k} C(\log a + 2^{k+1}).
\end{aligned} \tag{3.9}$$

Convergence in (3.7) follows from (3.9). Condition (3.8) is an easy consequence of (3.7) (as can be seen e.g. by making the approximate change of variables $2^n/(\log n)^k \simeq 2^m$ in (3.8) and comparing with (3.7) for $a > 4$). □

PROOF OF SUFFICIENCY. Since this is only a matter of normalization we will assume that conditions (3.2) and (3.3) are satisfied with $C = 1$. By the Remark below Theorem 3.1, proving the LIL is equivalent to showing that the series (3.4) converges for some $C < \infty$. To establish this we will show in several steps that we may suitably truncate h by proving inequalities of the form

$$\sum_n \Pr\Big\{\Big|\sum_{i,j \leq 2^n} \varepsilon_i \tilde{\varepsilon}_j h_n(X_i, Y_j)\Big| \geq C 2^n \log n\Big\} < \infty, \tag{3.10}$$

where $h_n := h I_{A_n}$ and A_n are suitably chosen subsets of the product space. Then, we will apply Lemma 2.8 conditionally to the truncated h (several times, and after some additional preparation).

STEP 1. Inequality (3.10) holds for any $C > 0$ if

$$A_n \subset \big\{(x,y) : \max\big(f_n(x), f_n(y)\big) \geq 2^n (\log n)^2\big\}.$$

In this case, by (3.8),

$$\sum_n \Pr\Big\{\Big|\sum_{i,j\le 2^n} \varepsilon_i\tilde\varepsilon_j h_n(X_i,Y_j)\Big| > C2^n\log n\Big\}$$
$$\le \sum_n \Pr\Big\{\exists\, i\le 2^n:\ f_n(X_i)\ge 2^n(\log n)^2\Big\}$$
$$+\sum_n \Pr\Big\{\exists\, j\le 2^n:\ f_n(Y_j)\ge 2^n(\log n)^2\Big\}$$
$$\le 2\sum_n 2^n \Pr\big\{f_n(X)\ge 2^n(\log n)^2\big\}<\infty.$$

STEP 2. Inequality (3.10) holds for any $C>0$ if

$$A_n\subset\big\{(x,y): h^2(x,y)\ge 2^{2n}(\log n)^2\big\}.$$

Indeed, by Chebyshev's inequality,

$$\sum_n \Pr\Big\{\Big|\sum_{i,j\le 2^n} \varepsilon_i\tilde\varepsilon_j h_n(X_i,Y_j)\Big| > C2^n\log n\Big\}$$
$$\le \sum_n \frac{1}{C2^n\log n}E\Big|\sum_{i,j\le 2^n}\varepsilon_i\tilde\varepsilon_j h_n(X_i,Y_j)\Big|$$
$$=\sum_n \frac{2^n}{C\log n}E|h|I_{\{|h|\ge 2^n\log n\}}$$
$$=C^{-1}E|h|\sum_n\frac{2^n}{\log n}I(|h|\ge 2^n\log n)$$
$$\le \tilde C E\frac{h^2}{(L_2|h|)^2}<\infty.$$

STEP 3. Inequality (3.10) holds for any $C>0$ if

$$A_n\subset\big\{(x,y): 2^{2n}n^{-4}\le h^2(x,y)<2^{2n}(\log n)^2,\, f_n(x), f_n(y)\le 2^n(\log n)^2\big\}.$$

If we use again Chebyshev's inequality, it suffices to prove that

$$\sum_n \frac{E\big|\sum_{i,j\le 2^n}\varepsilon_i\tilde\varepsilon_j h_n(X_i,Y_j)\big|^4}{2^{4n}(\log n)^4}<\infty. \tag{3.11}$$

Notice however that, by iteration of Khinchin's inequality (or by direct computation), there is $C<\infty$ (e.g. $C=18$) such that

$$\begin{aligned}C^{-1}E\Big|\sum_{i,j\le 2^n}\varepsilon_i\tilde\varepsilon_j h_n(X_i,Y_j)\Big|^4 &\le E\Big|\sum_{i,j\le 2^n}h_n^2(X_i,Y_j)\Big|^2\\ &\le \sum_{i,j}Eh_n^4(X_i,Y_j)+\sum_{i\ne i',j}Eh_n^2(X_i,Y_j)h_n^2(X_{i'},Y_j)\\ &\qquad+\sum_{i,j\ne j'}Eh_n^2(X_i,Y_j)h_n^2(X_i,Y_{j'})\\ &\qquad+\sum_{i\ne i',j\ne j'}Eh_n^2(X_i,Y_j)h_n^2(X_{i'},Y_{j'}).\end{aligned}$$

So, to prove (3.11) we have to check convergence of these four series.

First series:

$$\begin{aligned}\sum_n\frac{2^{2n}Eh_n^4}{2^{4n}(\log n)^4}&\le\sum_n\frac{1}{2^{2n}(\log n)^4}Eh^4I_{\{h^2\le 2^{2n}(\log n)^2\}}\\ &=Eh^4\sum_n\frac{1}{2^{2n}(\log n)^4}I(h^2\le 2^{2n}(\log n)^2)\\ &\le\tilde CEh^4\frac{1}{h^2(L_2|h|)^2}<\infty.\end{aligned}$$

Second series: (below we use the notation $h_n:=h_n(X,Y)$, $\tilde h_n=h_n(\tilde X,Y)$ and $\tilde X$ is an independent copy of X)

$$\begin{aligned}&\sum_n\frac{2^{3n}Eh_n^2(X,Y)h_n^2(\tilde X,Y)}{2^{4n}\log^4 n}=\sum_n\frac{Eh_n^2\tilde h_n^2}{2^n(\log n)^4}\le 2\sum_n\frac{Eh_n^2\tilde h_n^2I(|h|\le|\tilde h|)}{2^n(\log n)^4}\\ &\quad\le 2Eh^2\tilde h^2I(|h|\le|\tilde h|)\sum_n\frac{1}{2^n(\log n)^4}I(E_X\min(h^2,2^{4n})\le 2^n(\log n)^2,\tilde h^2\le 2^{4n})\\ &\quad\le 2Eh^2\tilde h^2I(|h|\le|\tilde h|)\sum_n\frac{1}{2^n(\log n)^4}I(E_X\min(h^2,\tilde h^2)\le 2^n(\log n)^2,|\tilde h|\le 2^{2n})\\ &\quad\le\tilde CEh^2\tilde h^2I(|h|\le|\tilde h|)\frac{1}{E_X\min(h^2,\tilde h^2)(L_2|\tilde h|)^2}\le\tilde CE\frac{\tilde h^2}{(L_2|\tilde h|)^2}<\infty.\end{aligned}$$

3rd series: convergence follows just as for the second.

4th series: here we have by (3.2)

$$\begin{aligned}\sum_n\frac{2^{4n}(Eh_n^2)^2}{2^{4n}(\log n)^4}&\le\tilde C\sum_n\frac{Eh_n^2}{(\log n)^3}\\ &=\tilde CEh^2\sum_n\frac{1}{(\log n)^3}I(2^{2n}n^{-4}\le h^2(x,y)<2^{2n}(\log n)^2)\\ &\le\tilde CE\frac{h^2}{(L_2|h|)^2}<\infty,\end{aligned}$$

where we use the fact that

$$\mathrm{Card}\big\{n : 2^{2n}n^{-4} \le h^2(x,y) < 2^{2n}\log^2 n\big\} \sim 2L_2 h.$$

This completes the third Step.

STEP 4. Inequality (3.10) holds for any $C > 0$ if

$$A_n \subset \Big\{(x,y) : h^2(x,y) \le \frac{2^{2n}}{n^4}, \frac{2^n}{\log n} \le \max\big(f_n(x), f_n(y)\big) \le 2^n(\log n)^2\Big\}.$$

We follow the proof of the previous step. The only difference is in the proof of convergence of the fourth series. We have for $n \ge 2$

$$\begin{aligned} Eh_n^2 &\le 2\sum_{k=1}^{3} E\min(h^2, 2^{2n}) I_{\{2^n(\log n)^{2-k} \le f_n(X) \le 2^n(\log n)^{3-k}\}} \\ &\le \sum_{k=1}^{3} 2^{n+1}(\log n)^{3-k}\,\Pr\big\{f_n(X) \ge 2^n(\log n)^{2-k}\big\}. \end{aligned}$$

Thus, by (3.8),

$$\sum_n \frac{Eh_n^2}{(\log n)^3} \le \sum_{k=1}^{3}\sum_n \frac{2^{n+1}}{(\log n)^k}\Pr\big\{f_n(X) \ge 2^n(\log n)^{2-k}\big\} < \infty.$$

For the next step, we define the functions

$$g_n(x) = E_Y h I_{\{|h| \ge 2^n n^2\}}. \tag{3.12}$$

STEP 5. Inequality (3.1) holds for any $C > 0$ if

$$A_n \subset \big\{(x,y) : \max\big(g_n(x), g_n(y)\big) \ge 1\big\}.$$

Assumption (3.2) implies that $\Pr\{|h| \ge v\} \le v^{-2}L_2 v^2$. Hence, $E|h|I_{\{|h|\ge s\}} \le \tilde{C}s^{-1}L_2 s$ for $s \ge 1$. Therefore,

$$\sum_n 2^n \Pr\big\{|g_n(X)| \ge 1\big\} \le \tilde{C}\sum_n \frac{Ln}{n^2} < \infty,$$

and the same is true for $g_n(Y)$.

STEP 6. Inequality (3.10) holds for any $C > 0$ if

$$A_n \subset \Big\{(x,y) : f_n(x) \ge \frac{2^n}{n},\ f_n(y) \ge \frac{2^n}{n},\ h^2(x,y) \le \frac{2^{2n}}{n^4}\Big\}.$$

To see this we note first that

$$\begin{aligned} Eh_n^2 &\le \frac{2^{2n}}{n^4}EI_{A_n} \le \frac{2^{2n}}{n^4}\Pr\Big\{f_n(X) \ge \frac{2^n}{n}\Big\}\Pr\Big\{f_n(Y) \ge \frac{2^n}{n}\Big\} \\ &\le \frac{2^{2n}}{n^4}\left(\frac{nEf_n(X)}{2^n}\right)^2 \le \tilde{C}\frac{(\log n)^2}{n^2}, \end{aligned}$$

since $Ef_n(X) = E\min(h^2, 2^{4n}) \le \tilde{C}\log n$ by (3.2). Now we may conclude Step 6 by Chebyshev's inequality as

$$\sum_n \frac{E\big|\sum_{i,j\le 2^n} \varepsilon_i\tilde{\varepsilon}_j h_n(X_i,Y_j)\big|^2}{2^{2n}(\log n)^2} \le \sum_n \frac{Eh_n^2}{(\log n)^2} \le \tilde{C}\sum_n \frac{1}{n^2} < \infty.$$

STEP 7. Inequality (3.10) holds for some $C > 0$ if

$$A_n = \Big\{(x,y) : f_n(x) \le \frac{2^n}{\log n}, f_n(y) \le \frac{2^n}{n}, g_n(x) \le 1, g_n(y) \le 1, h^2(x,y) \le \frac{2^{2n}}{n^4}\}.$$

This is the most involved step, and the only one (except for the similar Step 8 below) where we use condition (3.3). To prove (3.10) in this case, we will use Prohorov's inequality (or Lemma 2.8) together with the following four lemmas (one of which also uses Talagrands's inequality).

LEMMA 3.3. *For all $n \in \mathbf{N}$,*

$$\Pr\Big\{\Big|\sum_{i\le 2^n} \varepsilon_i h_n(X_i, Y)\Big| \ge 2^{n+4}\Big\} \le 2^{-4n}$$

and

$$\sum_n \Pr\Big\{\max_{1\le j\le 2^n}\Big|\sum_{i\le 2^n} \varepsilon_i h_n(X_i, Y_j)\Big| \ge 2^{n+4}\Big\} < \infty$$

PROOF. We note that $A_n \subset \big\{(x,y) : |h(x,y)| \le n^{-1}2^n, f_n(y) \le n^{-1}2^n\big\}$ and then apply Bernstein's inequality or Prohorov's inequality to obtain that, for any Y,

$$\Pr_X\Big\{\Big|\sum_{i\le 2^n} \varepsilon_i h_n(X_i, Y)\Big| \ge 2^{n+4}\Big\} \le e^{-4n},$$

which clearly implies the Lemma. (Lemma 2.8 instead of Bernstein's or Prohorov's inequality would simply change multiplicative constants.) □

Before formulating the next lemma it is convenient to define a sequence c_n by the formula

$$c_n = Eh^2 I_{\{2^n n^{-2} < |h| \le 2^n n^2\}}, \quad n \in \mathbf{N}. \tag{3.13}$$

LEMMA 3.4. We have

$$\sum_n \exp\Big(-\frac{2\log n}{\sqrt{1+c_n}}\Big) < \infty.$$

PROOF. Condition (3.2) implies that, for any $k \ge 2$,

$$\sum_{k\le \log n\le k+1} c_n \le \tilde{C}kE|h|^2 I_{\{|h|\le 2^{e^{k+1}}(e^{k+1})^2\}} \le \tilde{C}k^2,$$

(where the second constant is different from the first) since the largest number of intervals $I_n = [n^{-2}2^n, n^2 2^n]$, $k \le \log n \le k+1$, that can overlap with any given one of them is not larger than $6(k+1)$. Hence,

$$\mathrm{Card}\{n : k \le \log n \le k+1, c_n \ge 1\} \le \tilde{C}k^2.$$

Condition (3.2) also implies $c_n \le 2\log n$ (note that $c_1 = 0$). So,

$$\begin{aligned}\sum_n \exp\Big(-\frac{2\log n}{\sqrt{1+c_n}}\Big) &\le \sum_n \exp(-\sqrt{2}\log n) + \sum_{c_n\ge 1}\exp\Big(-\frac{2\log n}{\sqrt{1+2\log n}}\Big)\\ &\le \sum_n \exp(-\sqrt{2}\log n) + \sum_k \tilde{C}k^2\exp(-\sqrt{k}) < \infty.\end{aligned}$$

□

The following lemma is well known but a proof is provided for the reader's convenience.

LEMMA 3.5. *If a kernel k satisfies $E_X|k(X,y)| \le 1$ and $E_Y|k(x,Y)| \le 1$ a.s., then k defines an operator on $L_2(\mathcal{L}(X))$ with norm bounded by 1, that is, condition (3.3) holds for $h = k$ and $C = 1$ (and therefore so does condition (1.5)).*

PROOF. We need to check that

$$|E_X E_Y k(X,Y)f(X)g(Y)| \le [Ef^2(X)Eg^2(Y)]^{1/2}$$

whenever $\|f\|_\infty, \|g\|_\infty < \infty$. But, assuming (without loss of generality) that k, f and g are nonnegative,

$$\begin{aligned}E_X E_Y k(X,Y)f(X)g(Y) &= E_X\Big[f(X)E_Y\big(k^{1/2}(X,Y)k^{1/2}(X,Y)g(Y)\big)\Big]\\ &\le E_X\Big[f(X)(E_Y k(X,Y))^{1/2}(E_Y k(X,Y)g^2(Y))^{1/2}\Big]\\ &\le E_X\Big[f(X)\big(E_Y k(X,Y)g^2(Y)\big)^{1/2}\Big]\\ &\le \big(E_X f^2(X)\big)^{1/2}\Big[E_X\big(E_Y k(X,Y)g^2(Y)\big)\Big]^{1/2}.\end{aligned}$$

and now the inequality follows by applying Fubini and using $E_X k(X,Y) \le 1$. □

LEMMA 3.6. *There exists $C_1 < \infty$ such that*

$$\sum_n \Pr\Big\{E_Y\Big(\sum_{i\le 2^n}\varepsilon_i h_n(X_i,Y)\Big)^2 \ge C_1\sqrt{1+c_n}\;2^n\log n\Big\} < \infty.$$

PROOF. Let H_Y be $L_2(\Omega,\sigma(Y),\Pr)$, that is, H_Y is the space of all square integable random variables $f(Y)$ where f is a Borel measurable function. Let $\mathbf{X}_i := \varepsilon_i h_n(X_i,Y)$ for $i = 1,\dots,2^n$. Then, $\mathbf{X}_i$ are symmetric i.i.d. random vectors with values in H_Y. We define

$$Z = \sup_{f\in\mathcal{F}}\sum_{i=1}^{2^n} f(\mathbf{X}_i) = \Big[E_Y\Big(\sum_{i\le 2^n}\varepsilon_i h_n(X_i,Y)\Big)^2\Big]^{1/2},$$

where $\mathcal{F}$ is a countable dense subset of the unit ball of $H'_Y = H_Y$ and we write $f(\cdot) := \langle f,\cdot\rangle$. We will apply Lemma 2.8 to Z. For this, we must estimate EZ and determine suitable U and σ^2. We have

$$EZ \le \big(EZ^2\big)^{1/2} = \big[2^n Eh_n^2\big]^{1/2} \le \sqrt{2^n\log n} \tag{3.14}$$

by (3.2). Since

$$\sup_{f\in\mathcal{F}}\left|f(\mathbf{X}_i(\omega))\right| = \|\mathbf{X}_i(\omega)\|_Y = \sqrt{E_Y h_n^2(X_i(\omega),Y)} \le \sqrt{\frac{2^n}{\log n}},$$

we can take

$$U = \sqrt{\frac{2^n}{\log n}} \tag{3.15}$$

in Lemma 2.8 for Z. Moreover, for each $f \in \mathcal{F}$,

$$Ef^2(\mathbf{X}_i) = E\big(E_Y h_n(X_i,Y)f(Y)\big)^2 \le 3\sum_{i=1}^{3} E\big(E_Y h_n^{(i)}(X_i,Y)f(Y)\big)^2,$$

where

$$h_n^{(1)} := hI_{B_n},\ h_n^{(2)} := hI_{B_n\cap\{2^n n^{-2}<|h|\le 2^n n^2\}},\ h_n^{(3)} := hI_{B_n\cap\{|h|\ge 2^n n^2\}},$$

with

$$B_n := \Big\{(x,y) : f_n(x) \le \frac{2^n}{\log n}, f_n(y) \le \frac{2^n}{n}, g_n(x) \le 1, g_n(y) \le 1\Big\},$$

since

$$h_n = h_n^{(1)} - h_n^{(2)} - h_n^{(3)}.$$

Now,

$$E\big(E_Y h_n^{(1)}(X_i,Y)f(Y)\big)^2 \le 1$$

by condition (1.5) (which is equivalent to (1.3)=(3.3)),

$$E\big(E_Y h_n^{(2)}(X_i,Y)f(Y)\big)^2 \le E(h_n^{(2)})^2 \le c_n$$

by Cauchy-Schwartz and the definition of c_n in (3.13), and

$$E\big(E_Y h_n^{(3)}(X_i,Y)f(Y)\big)^2 \le 1$$

by Lemma 3.5 (see (1.5) once more). Therefore, we can take σ^2 in Lemma 2.8 for Z to be

$$\sigma^2 = 3\cdot 2^n(2+c_n) < 6\cdot 2^n(1+c_n). \tag{3.16}$$

Then, on account of (3.14)-(3.16), Lemma 2.8 gives, with $C_2 = (\sqrt{C_1} - 1)^2$,

$$\begin{aligned}
\Pr\Big\{E_Y\Big|\sum_{i\le 2^n}\varepsilon_i h_n(X_i,Y)\Big|^2 \ge C_1\sqrt{1+c_n}\;2^n\log n\Big\}&\\
&= \Pr\Big\{Z \ge \sqrt{C_1\sqrt{1+c_n}\;2^n\log n}\Big\}\\
&\le \Pr\Big\{Z - EZ \ge \sqrt{C_2\sqrt{1+c_n}\;2^n\log n}\Big\}\\
&\le K\exp\Big(-\frac{\sqrt{C_2}}{K}\sqrt[4]{1+c_n}\;\log n\;\log\Big(1+\frac{\sqrt{C_2}\sqrt[4]{1+c_n}}{6(1+c_n)+8}\Big)\Big)\\
&\le K\exp\Big(-\frac{\sqrt{C_2}}{K}\sqrt[4]{1+c_n}\;\log n\;\log\Big(1+\frac{\sqrt{C_2}}{14(1+c_n)^{3/4}}\Big)\Big)\\
&\le K\exp\Big(-\frac{\sqrt{C_2}}{K}\log\Big(1+\frac{\sqrt{C_2}}{14}\Big)\frac{\log n}{\sqrt{1+c_n}}\Big),
\end{aligned}$$

where in the last line we have used that the function $x^{-1}\log(1+x)$ is monotone decreasing. Taking $K^{-1}\sqrt{C_2}\log(1+\sqrt{C_2}/14) \ge 2$ yields the bound

$$\Pr\Big\{E_Y\Big|\sum_{i\le 2^n}\varepsilon_i h_n(X_i,Y)\Big|^2 \ge C_1\sqrt{1+c_n}\;2^n\log n\Big\} \le K\exp\Big(-\frac{2\log n}{\sqrt{1+c_n}}\Big)$$

and Lemma 3.6 follows from Lemma 3.4. □

Now we complete the proof of Step 7. For n fixed, set

$$d(y) := \sum_{i\le 2^n}\varepsilon_i h_n(X_i,y) \quad\text{and}\quad \tilde d_j := \tilde\varepsilon_j d(Y_j) I_{\{|d|\le 2^{n+4},\, E_Y d^2(Y)\le C_1 2^n(\log n)\sqrt{1+c_n}\}}(Y_j)$$

for $1 \le j \le 2^n$. Then,

$$\begin{aligned}
\Pr\Big\{\Big|\sum_{i,j\le 2^n}\varepsilon_i\tilde\varepsilon_j h_n(X_i,Y_j)\Big| > C2^n\log n\Big\} &= \Pr\Big\{\Big|\sum_{j\le 2^n}\tilde\varepsilon_j d(Y_j)\Big| > C2^n\log n\Big\}\\
&\le \Pr\{\exists\, j\le 2^n : \tilde d_j \ne d(Y_j)\} + \Pr\Big\{\Big|\sum_{j\le 2^n}\tilde\varepsilon_j\tilde d_j\Big| > C2^n\log n\Big\}\\
&:= I_n + II_n.
\end{aligned}$$

But,

$$\begin{aligned}
I_n \le \Pr\Big\{\max_{j\le 2^n}\Big|\sum_{i\le 2^n}\varepsilon_i h_n(X_i,Y_j)\Big| > 2^{n+4}\Big\}&\\
+\Pr\Big\{E_Y\Big|\sum_{i\le 2^n}\varepsilon_i h_n(X_i,Y)\Big|^2 > C_1 2^n(\log n)\sqrt{1+c_n}\Big\}&
\end{aligned}$$

and Lemma 3.3 and Lemma 3.6 show that

$$\sum_n I_n < \infty. \tag{3.17}$$

To estimate II_n we can apply Bernstein's or Prokhorov's inequality conditionally on the sequence $\{X_i\}$. For convenience we will use Lemma 2.8. We can take $U = 2^{n+4}$ and $V = C_1 2^{2n}(\log n)\sqrt{1+c_n}$ to get

$$\begin{aligned}\Pr_Y\Big\{\Big|\sum_{j\le 2^n}\tilde\varepsilon_j\tilde d_j\Big| > C2^n\log n\Big\}& \\ &\le K\exp\Big(-\frac{1}{K}\frac{C2^n\log n}{2^{n+4}}\log\Big(1+\frac{C2^{2n+4}\log n}{C_1 2^{2n}(\log n)\sqrt{1+c_n}}\Big)\Big)\\ &\le K\exp\Big(-\frac{C}{2^4K}\log\Big(1+\frac{2^4C}{C_1}\Big)\frac{\log n}{\sqrt{1+c_n}}\Big)\Big).\end{aligned}$$

Taking C so that

$$\frac{C}{2^4K}\log\Big(1+\frac{2^4C}{C_1}\Big)\ge 2$$

shows, by Lemma 3.4, that

$$\sum_n II_n < \infty. \tag{3.18}$$

(3.17) and (3.18) complete the proof of Step 7.

STEP 8. Inequality (3.10) holds for some $C<\infty$ if

$$A_n = \Big\{(x,y) : f_n(x)\le\frac{2^n}{n}, f_n(y)\le\frac{2^n}{\log n}, g_n(x)\le 1, g_n(y)\le 1, h^2(x,y)\le\frac{2^{2n}}{n^4}\}.$$

This can be done in the same way as Step 7.

It is clear that we can write $S\times S = \cup_{i=1}^8 A_n^i$ with $A_n^1,\dots,A_n^8$ disjoint, and A_n^i satisfying the conditions in Step i for each n. Then, $h=\sum_{i=1}^8 hI_{A_n^i} = \sum_{i=1}^8 h_n^i$. Since for each i the kernels h_n^i satisfy condition (3.10) for some $C<\infty$, it follows by the triangle inequality that the series (3.4) for h converges for some $C<\infty$, proving the sufficiency part of Theorem 3.1. □

4. Canonical kernels. In this section we show that, for canonical kernels, the LIL (1.1) is equivalent to the decoupled and randomized LIL. The preliminary results in Section 2(B) yield that the regular LIL implies the decoupled and randomized one. The converse implication, however, seems to require Theorem 3.1. The first step consists of the following simple inequality, rooted in known symmetrization techniques.

LEMMA 4.1. *For any kernel h, and for any $n\in\mathbf{N}$ and $t>0$, we have*

$$\begin{aligned}\Pr\Big\{\Big|\sum_{i,j\le n}h(X_i,Y_j)\Big|\ge 10t\Big\}\le 16\Pr\Big\{\Big|\sum_{i,j\le n}\varepsilon_i\tilde\varepsilon_j h(X_i,Y_j)\Big|\ge t\Big\}&\\ +4\Pr\Big\{E_Y\Big|\sum_{i,j\le n}\varepsilon_i h(X_i,Y_j)\Big|\ge t\Big\}&\\ +\Pr\Big\{E_X\Big|\sum_{i,j\le n}h(X_i,Y_j)\Big|\ge t\Big\}&.\end{aligned}$$

PROOF. Let $\{Z_i\}$ be a sequence of independent random variables such that $E|\sum_i Z_i| \leq s$ and let $\{Z_i^{'}\}$ be an independent copy of $\{Z_i\}$. Then, by Chebyshev's inequality, $\Pr\{|\sum_i Z_i^{'}| \leq 2s\} \geq 1/2$. So, for any $t > 0$,

$$\begin{aligned}
\Pr\Big\{\Big|\sum_i Z_i\Big| \geq 2t + 2s\Big\} &\leq 2\Pr\Big\{\Big|\sum_i Z_i'\Big| \leq 2s, \Big|\sum_i Z_i\Big| \geq 2t + 2s\Big\} \\
&\leq 2\Pr\Big\{\Big|\sum_i (Z_i - Z_i^{'})\Big| \geq 2t\Big\} \\
&= 2\Pr\Big\{\Big|\sum_i \varepsilon_i (Z_i - Z_i^{'})\Big| \geq 2t\Big\} \\
&\leq 2\Pr\Big\{\Big|\sum_i \varepsilon_i Z_i\Big| \geq t\Big\} + 2\Pr\Big\{\Big|\sum_i \varepsilon_i Z_i^{'}\Big| \geq t\Big\} \\
&= 4\Pr\Big\{\Big|\sum_i \varepsilon_i Z_i\Big| \geq t\Big\}.
\end{aligned}$$

Using the above inequality conditionally we get

$$\Pr\Big\{\Big|\sum_{i,j} h(X_i, Y_j)\Big| \geq 10t\Big\} \leq 4\Pr\Big\{\Big|\sum_{i,j} \varepsilon_i h(X_i, Y_j)\Big| \geq 4t\Big\} + \Pr\Big\{E_X\Big|\sum_{i,j} h(X_i, Y_j)\Big| \geq t\Big\}$$

and

$$\begin{aligned}
\Pr\Big\{\Big|\sum_{i,j} \varepsilon_i h(X_i, Y_j)\Big| \geq 4t\Big\} \leq 4\Pr\Big\{\Big|\sum_{i,j} \varepsilon_i \tilde{\varepsilon}_j h(X_i, Y_j)\Big| \geq t\Big\} \\
+ \Pr\Big\{E_Y\Big|\sum_{i,j} \varepsilon_i h(X_i, Y_j)\Big| \geq t\Big\}.
\end{aligned}$$

□

The next lemma shows that if the second moment and the conditional second moment of a canonical kernel h are suitably truncated, then Talagrand's inequality (Lemma 2.8) allows control of the last two terms on the right hand side of the inequality in Lemma 4.1.

LEMMA 4.2. *Let h be a canonical kernel such that*

$$Eh^2(X, Y) \leq c^2 \log n \quad \textit{and} \quad E_Y h^2(X, Y) \leq c^2 2^n \quad X - a.s.$$

for some $c < \infty$. Then we have that, for some universal constant C,

$$\Pr\Big\{E_Y\Big|\sum_{i,j \leq 2^n} h(X_i, Y_j)\Big| \geq cC2^n \log n\Big\} \leq n^{-2}.$$

PROOF. We can assume $c = 1$. If we define

$$Z := E_Y\Big|\sum_{i,j \leq 2^n} h(X_i, Y_j)\Big|$$

then

$$Z = \sup\Big\{\sum_{i\le 2^n} E_Y\Big(\sum_{j\le 2^n} h(X_i,Y_j)g(\mathbf{Y})\Big)\Big\},$$

where the supremum is taken over all $g(\mathbf{Y}) = g(Y_1,\dots,Y_{2^n})$ with $\|g\|_\infty \le 1$, actually over a countable L_1-norm determining subset of such functions. Thus Z has the same form as in Lemma 2.8. Then, since

$$\Big\|E_Y\Big|\sum_{j=1}^{2^n} h(x,Y_j)\Big|\Big\|_\infty \le \Big\|\Big(\sum_{j=1}^{2^n} E_Y h^2(x,Y_j)\Big)^{1/2}\Big\|_\infty \le 2^n$$

and

$$\sum_{i=1}^{2^n} E\Big(E_Y\Big|\sum_{j=1}^{2^n} h(X_i,Y_j)\Big|\Big)^2 \le \sum_{i=1}^{2^n} E\Big(\sum_{j=1}^{2^n} E_Y h^2(X_i,Y_j)\Big) = 2^{2n}Eh^2 \le 2^{2n}\log n,$$

we can take

$$U = 2^n \quad\text{and}\quad V = 2^{2n}\log n \tag{4.1}$$

in Talagrand's exponential bound for Z. Moreover

$$EZ \le \Big(E\Big|\sum_{i,j\le 2^n} h(X_i,Y_j)\Big|^2\Big)^{1/2} = 2^n\big(Eh^2\big)^{1/2} \le 2^n\log n. \tag{4.2}$$

Now the statement follows by (4.1), (4.2) and the exponential bound in Lemma 2.8. □

The following lemma will allow us to carry out truncations for canonical kernels exactly in the same way as we did for randomized kernels in the first four steps of the sufficiency proof of Theorem 3.1.

LEMMA 4.3. *For any integrable kernel h, $n \in \mathbf{N}$ and $p \ge 1$ we have*

$$\Big\|\sum_{i,j\le n} \pi_2 h(X_i,Y_j)\Big\|_p \le 4\Big\|\sum_{i,j\le n} \varepsilon_i\tilde\varepsilon_j h(X_i,Y_j)\Big\|_p.$$

PROOF. Since $\pi_2 h$ is canonical, by Jensen's inequality we have that, for all $\{Y_j\}$,

$$\begin{aligned}
E_X\Big|\sum_{i,j\le 2^n} \pi_2 h(X_i,Y_j)\Big|^p &\le E_X\Big|\sum_{i,j\le 2^n}\big(\pi_2 h(X_i,Y_j) - \pi_2 h(X_i',Y_j)\big)\Big|^p \\
&= E_X\Big|\sum_{i,j\le 2^n} \varepsilon_i\big(\pi_2 h(X_i,Y_j) - \pi_2 h(X_i',Y_j)\big)\Big|^p \\
&= E_X\Big|\sum_{i,j\le 2^n} \varepsilon_i\big(h(X_i,Y_j) - E_Y h(X_i,Y_j) \\
&\qquad - h(X_i',Y_j) + E_Y h(X_i',Y_j)\big)\Big|^p.
\end{aligned}$$

Thus, by the triangle inequality,

$$\Big\|\sum_{i,j\le 2^n}\pi_2 h(X_i,Y_j)\Big\|_p \le \Big\|\sum_{i,j\le 2^n}\varepsilon_i\big(h(X_i,Y_j)-E_Y h(X_i,Y_j)\big)\Big\|_p + \Big\|\sum_{i,j\le 2^n}\varepsilon_i\big(h(X_i^{'},Y_j)-E_Y h(X_i^{'},Y_j)\big)\Big\|_p = 2\Big\|\sum_{i,j\le 2^n}\varepsilon_i\big(h(X_i,Y_j)-E_Y h(X_i,Y_j)\big)\Big\|_p.$$

In a similar way we may prove that

$$\Big\|\sum_{i,j\le 2^n}\varepsilon_i\big(h(X_i,Y_j)-E_Y h(X_i,Y_j)\big)\Big\|_p \le 2\Big\|\sum_{i,j\le 2^n}\varepsilon_i\tilde{\varepsilon}_j h(X_i,Y_j)\Big\|_p.$$

□

Now we can prove the main result of this section.

THEOREM 4.4. *For any canonical kernel h the following two conditions are equivalent:*

$$\limsup_{n\to\infty}\frac{1}{n\log\log n}\Big|\sum_{1\le i<j\le n} h(X_i,X_j)\Big| < \infty \quad a.s. \tag{4.3}$$

and

$$\limsup_{n\to\infty}\frac{1}{n\log\log n}\Big|\sum_{1\le i,j\le n}\varepsilon_i\tilde{\varepsilon}_j h(X_i,Y_j)\Big| < \infty \quad a.s. \tag{4.4}$$

Here, again, each of the two limsups is a.s. bounded by a universal constant times the other.

PROOF. (4.3) implies (4.4) (even without degeneracy of the kernel) by Lemma 2.1(b).

To prove the opposite implication, by Corollary 2.6 it is enough to show that if (4.4) holds (which is equivalent to the two conditions (3.2) and (3.3) by Theorem 3.1), then

$$\sum_n \Pr\Big\{\Big|\sum_{i,j\le 2^n} h(X_i,Y_j)\Big| \ge C2^n\log n\Big\} < \infty.$$

Since h is canonical, we may replace h by $\pi_2 h$ in this series ($h = \pi_2 h$). As in the case of decoupled and randomized kernels, convergence of the series will follow in a few steps by showing that

$$\sum_n \Pr\Big\{\Big|\sum_{i,j\le 2^n}\pi_2 h_n(X_i,Y_j)\Big| \ge C2^n\log n\Big) < \infty, \tag{4.5}$$

where $h_n = hI_{A_n}$ for suitably chosen sequences of sets A_n. We can assume, as in Theorem 3.1, that $C = 1$ in conditions (3.2) and (3.3).

STEP 1. The series in (4.5) converges for

$$A_n = \big\{(x,y):\ f_n(x) > 2^n(\log n)^2 \ \text{ or } \ f_n(y) > 2^n(\log n)^2\big\}.$$

By the degeneracy of h we have

$$\begin{aligned}|Eh_n| &= \big|EhI_{\{f_n(x)>2^n(\log n)^2\}} + EhI_{\{f_n(y)>2^n(\log n)^2\}} \\ &\qquad\qquad - EhI_{\{f_n(x)>2^n(\log n)^2, f_n(y)>2^n(\log n)^2\}}\big| \\ &= \big|EhI_{\{f_n(x)>2^n(\log n)^2, f_n(y)>2^n(\log n)^2\}}\big| \\ &\le \Pr\{f_n(X)>2^n(\log n)^2\}^{1/2}\Pr\{f_n(Y)>2^n(\log n)^2\}^{1/2} \\ &\le \tilde{C}2^{-n}, \end{aligned} \tag{4.6}$$

where the last two inequalities follow by (3.3) and (3.8) respectively. We also have

$$\pi_1 h_n(x) = \pi_1 hI_{\{f_n(y)>2^n(\log n)^2, f_n(x)\le 2^n(\log n)^2\}}(x),$$

as can be seen using the decomposition of h_n given in the first line of (4.6) together with the fact that $E_Y hI_{\{f_n(x)>2^n(\log n)^2\}} = 0$. Thus, by Chebyshev's inequalty,

$$\begin{aligned}\sum_n \Pr\Big\{\Big|\sum_{i\le 2^n}\pi_1 h_n(X_i)\Big| \ge c\log n\Big\} \\ &\le \sum_n \frac{2^n}{c^2(\log n)^2} E\Big|\pi_1 hI_{\{f_n(y)>2^n(\log n)^2, f_n(x)\le 2^n(\log n)^2\}}(X)\Big|^2 \\ &\le \sum_n \frac{2^n}{c^2(\log n)^2} E_X\Big[\big(E_Y hI_{\{f_n(y)>2^n(\log n)^2\}}\big)^2 I_{\{f_n(X)\le 2^n(\log n)^2\}}\Big] \\ &\le \sum_n \frac{2^n}{c^2(\log n)^2} E_X\big(E_Y hI_{\{f_n(y)>2^n(\log n)^2\}}\big)^2 \\ &\le \sum_n \frac{2^n}{c^2(\log n)^2} \Pr\{f_n(Y)>2^n(\log n)^2\} < \infty, \end{aligned} \tag{4.7}$$

where in the last line we used (1.5) with $C=1$ (that is, condition (3.3)) and (3.8). Finally, as in step 1 of the proof of sufficiency of the symmetrized LIL,

$$\sum_n \Pr\Big\{\Big|\sum_{i,j\le 2^n} h_n(X_i,Y_j)\Big| \ge C2^n\log n\Big\} < \infty. \tag{4.8}$$

Inequalities (4.6)-(4.8) imply (4.5) by Hoeffding's decomposition ((2.1)).

STEP 2. The series in (4.5) converges for

$$\begin{aligned}A_n \subset \big\{(x,y) : |h(x,y)| > 2^n\log n \ \text{ or } \ f_n(x) > 2^n \ \text{ or } \ f_n(y) > 2^n\big\} \\ \cap \big\{(x,y) : \max(f_n(x), f_n(y)) \le 2^n(\log n)^2\big\}.\end{aligned}$$

To prove this we may proceed just as in steps 2-4 of the proof of the symmetrized LIL, with only formal changes: note that in steps 2-4 there we used only Chebyshev's inequality to bound probabilities; thus Lemma 4.3 reduces proving inequality (4.5) here to steps 2-4 in that proof, where the lower bounds for h and f_n are even smaller.

STEP 3. The series in (4.5) converges for

$$A_n = \{(x,y) : |h(x,y)| \le 2^n\log n, f_n(x) \le 2^n, f_n(y) \le 2^n\}.$$

The LIL (4.4) implies that

$$\sum_n \Pr\Big\{\Big|\sum_{i,j\le 2^n} \varepsilon_i\tilde\varepsilon_j h(X_i,Y_j)\Big| \ge C2^n\log n\Big\} < \infty$$

for some $C < \infty$ by Lemma 2.2(c). Steps 1-4 from the proof of sufficiency in Theorem 3.1 show that

$$\sum_n \Pr\Big\{\Big|\sum_{i,j\le 2^n} \varepsilon_i\tilde\varepsilon_j hI_{D_n}(X_i,Y_j)\Big| \ge C2^n\log n\Big\} < \infty,$$

for any $D_n \subset \big\{(x,y) : |h(x,y)| > 2^n/n \;\text{ or }\; \max(f_n(x), f_n(y)) > 2^n/\log n\big\}$, in particular for $D_n = A_n^c$. Therefore we have

$$\sum_n \Pr\Big\{\Big|\sum_{i,j\le 2^n} \varepsilon_i\tilde\varepsilon_j h_n(X_i,Y_j)\Big| \ge C2^n\log n\Big\} < \infty \tag{4.9}$$

for some $C < \infty$. In order to deduce (4.5) from (4.9) we show first that we can replace h_n by $\pi_2 h_n$ in (4.9), and then apply Lemmas 4.1 and 4.2 to $\pi_2 h_n$. So, we begin by proving (4.9) for $h_n - \pi_2 h_n$ or, what is the same by Hoeffding's decomposition, we prove (4.9) with h_n replaced by $\pi_1 h_n$ and by Eh_n. We can write h_n as

$$\begin{aligned} h_n = h - hI_{\{f_n(x)>2^n\}} - hI_{\{f_n(y)>2^n\}} \\ + hI_{\{f_n(x)>2^n, f_n(y)>2^n\}} - hI_{\{|h|>2^n\log n, f_n(x)\le 2^n, f_n(y)\le 2^n\}}. \end{aligned}$$

Then, by the degeneracy of h and (3.3) we have

$$\begin{aligned} \Big|\sum_{i,j\le 2^n} \varepsilon_i\tilde\varepsilon_j Eh_n\Big| &\le 2^{2n}\Big(\big|EhI_{\{f_n(x)>2^n, f_n(y)>2^n\}}\big| + E|h|I_{\{|h|>2^n\log n\}}\Big) \\ &\le 2^{2n}\Big(\Pr\big\{f_n(X) > 2^n\big\}^{1/2}\Pr\big\{f_n(Y) > 2^n\big\}^{1/2} + E|h|I_{\{|h|>2^n\log n\}}\Big). \end{aligned}$$

Now, we note that (3.2) implies $E|h|I_{\{|h|>2^n\log n\}} \le \tilde C 2^{-n}$ (as $\Pr\{|h| > u\} \le u^{-2}L_2 u$) and

$$\Pr\big\{f_n(X) > 2^n\big\} \le \frac{E(h^2\wedge 2^{4n})}{2^n} \le \tilde C\frac{Ln}{2^n}.$$

Hence,

$$\Big|\sum_{i,j\le 2^n} \varepsilon_i\tilde\varepsilon_j Eh_n\Big| \le \tilde C 2^n\log n. \tag{4.10}$$

The above decomposition of h_n together with the degeneracy of h also give

$$\pi_1 h_n(x) = -\pi_1 hI_{\{f_n(y)>2^n, f_n(x)\le 2^n\}}(x) - \pi_1 hI_{\{|h|>2^n\log n, f_n(x), f_n(y)\le 2^n\}}(x).$$

So, by Chebyshev's inequality and (3.2), we have

$$\sum_n \Pr\Big\{\Big|\sum_{i,j\le 2^n} \varepsilon_i\tilde\varepsilon_j \pi_1 hI_{\{|h(x,y)|>2^n\log n, f_n(x), f_n(y)\le 2^n\}}(X_i)\Big| \ge c2^n\log n\Big\}$$

$$\leq \sum_n \frac{2^n}{c\log n} E\left|\pi_1 h I_{\{|h(x,y)|>2^n\log n, f_n(x), f_n(y)\leq 2^n\}}(X)\right|$$

$$\leq \sum_n \frac{1}{c\log n} 2^{n+1} E|h| I_{\{|h|>2^n\log n\}}$$

$$\leq c^{-1} E|h| \sum_n \frac{2^{n+1}}{\log n} I(|h| > 2^n \log n)$$

$$\leq \tilde{C} E \frac{h^2}{(L_2|h|)^2} < \infty. \tag{4.11}$$

Also, by Chebyshev's inequality, (1.5) with $C = 1$ and (3.8),

$$\sum_n \Pr\left\{\left|\sum_{i,j\leq 2^n} \varepsilon_i \tilde{\varepsilon}_j \pi_1 h I_{\{f_n(y)>2^n, f_n(x)\leq 2^n\}}(X_i)\right| \geq c 2^n \log n\right\}$$

$$\leq \sum_n \frac{1}{c^2\log^2 n} E\left|\pi_1 h I_{\{f_n(y)>2^n, f_n(x)\leq 2^n\}}(X)\right|^2$$

$$\leq \sum_n \frac{1}{c^2\log^2 n} E_X\Big(E_Y h I(f_n(y) > 2^n)\Big)^2$$

$$\leq \sum_n \frac{1}{c^2\log^2 n} \Pr\{f_n(Y) > 2^n\} < \infty. \tag{4.12}$$

Inequalities (4.9)-(4.12) imply, by the Hoeffding's decomposition,

$$\sum_n \Pr\left\{\left|\sum_{i,j\leq 2^n} \varepsilon_i \tilde{\varepsilon}_j \pi_2 h_n(X_i, Y_j)\right| \geq C 2^n \log n\right\} < \infty \tag{4.13}$$

for some $C < \infty$. By (3.2), $E(\pi_2 h_n)^2 \leq E h_n^2 \leq \tilde{C}\log n$, and, by the definition of A_n and (3.2), $E_Y(\pi_2 h_n)^2(x) \leq 2E_Y h_n^2 + 2E h_n^2 \leq 2^{n+1} + \tilde{C}\log n$, and likewise for $E_X(\pi_2 h_n)^2$. Then, it follows from Lemma 4.2 that

$$\sum_n \Pr\left\{E_Y\left|\sum_{i,j\leq 2^n} \varepsilon_i \pi_2 h_n(X_i, Y_j)\right| \geq C 2^n \log n\right\} < \infty \tag{4.14}$$

for some $C < \infty$, and that, likewise,

$$\sum_n \Pr\left\{E_X\left|\sum_{i,j\leq 2^n} \pi_2 h_n(X_i, Y_j)\right| \geq C 2^n \log n\right\} < \infty. \tag{4.15}$$

Then, (4.13)-(4.15) give (4.5) by Lemma 4.1, concluding the proof of Step 3.

Steps 1-3 together show that

$$\sum_n \Pr\left\{\left|\sum_{i,j\leq 2^n} \pi_2 h(X_i, Y_j)\right| \geq C 2^n \log n\right) < \infty,$$

concluding the proof of the theorem. □

5. Arbitrary kernels. Final comments. We conclude with the proof of Theorem 1.1, a conjecture on the LIL for kernels of more than two variables, and several remarks on the limsup in (1.1) and the limit set of the LIL sequence.

PROOF OF THEOREM 1.1. Conditions (1.2) and (1.3) are sufficent for the LIL for degenerate kernels by Theorems 3.1 and 4.4.

If the kernel h satistifies the LIL (1.1), then it satisfies the decoupled and randomized LIL by Lemma 2.1(b). Then, by Theorem 3.1, it also satisfies conditions (1.2) and (1.3). So, it suffices to prove that if the LIL (1.1) holds then the kernel h is canonical.

Since by (1.2) $E|\pi_2 h|^p < \infty$ for any $p < 2$, we have by the Marcinkiewicz type strong law of large numbers for U-statistics (Giné and Zinn, 1992, theorem 2),

$$\lim_{n\to\infty} \frac{1}{n^{2/p}} \sum_{i\neq j\leq n} \pi_2 h(X_i, Y_j) = 0 \ \text{ a.s. for all } 0<p<2. \tag{5.1}$$

The LIL for h implies the decoupled LIL (2.8) by Lemma 2.2(a), and therefore also that

$$\lim_{n\to\infty} \frac{1}{n^{2/p}} \sum_{i\neq j\leq n} h(X_i, Y_j) = 0 \ \text{ a.s. for all } 0<p<2. \tag{5.2}$$

Subtracting (5.1) from (5.2) and using the Hoeffding decomposition we obtain

$$\lim_{n\to\infty} n^{1-2/p} \Big| \sum_{i\leq n} \big(\pi_1 h(X_i) - \frac{1}{2} Eh\big) + \sum_{j\leq n} \big(\pi_1 h(Y_j) - \frac{1}{2} Eh\big) \Big| = 0 \ \text{ a.s.}$$

However if $p \geq 4/3$ this yields, by the CLT or the LIL in $\mathbf{R}$, that

$$\pi_1 h(X) - \frac{1}{2} Eh = 0 \ \text{ a.s.}$$

Since $\pi_1 h$ is centered, it follows that $Eh = 0$ and $\pi_1 h(X) = 0$ a.s. Hence $h = \pi_2 h$ is canonical for the law of X. □

The following conjecture for kernels of more than two variables seems only natural.

CONJECTURE 5.1. *Let h be a kernel of d variables symmetric in its entries. Then h satisfies the law of the iterated logarithm*

$$\limsup_{n\to\infty} \frac{1}{(n \log\log n)^{d/2}} \Big| \sum_{1\leq i_1<i_2<\ldots<i_d\leq n} h(X_{i_1}, \ldots, X_{i_d}) \Big| < \infty \ \ a.s. \tag{5.3}$$

if and only if the following conditions hold:
a) h is canonical for the law of X, that is $E_{X_i} h(X_1, \ldots, X_d) = 0$ a.s.
and there exists $C < \infty$ such that
b)

$$E \min(h^2, u) \leq C (L_2 u)^{d-1} \tag{5.4}$$

for all $u > 0$, and
c)

$$\sup\{E[h(X_1,\dots,X_d)\prod_{i=1}^{d} f_i(X_i)] : Ef_i^2(X) \le 1, \|f_i\|_\infty < \infty, i = 1,\dots,d\} < \infty. \tag{5.5}$$

We know at present that the necessity part of this conjecture is true.

The problem of determining the lim sup in (1.1) when $Eh^2 = \infty$ is open and, a fortiori, so is the problem of determining the limit set of the LIL sequence. We now briefly comment on these questions. The previous results do give the order of the limsup in (1.1) up to constants as we show next. In the theorem that follows we denote the quantity in (1.3) as $\|h\|_{L_2 \mapsto L_2}$.

THEOREM 5.2. *Suppose that $h(x, y)$ is canonical for the law of X. Then there exists a universal constant C such that, almost surely,*

$$\begin{aligned} C^{-1}\left[\|h\|_{L_2\mapsto L_2} + \limsup_{u\to\infty}\sqrt{\frac{E(h^2\wedge u)}{L_2 u}}\right] &\le \limsup_{n\to\infty}\frac{1}{nL_2 n}\Big|\sum_{1\le i\le j\le n} h(X_i,X_j)\Big| \\ &\le C\left[\|h\|_{L_2\mapsto L_2} + \limsup_{u\to\infty}\sqrt{\frac{E(h^2\wedge u)}{L_2 u}}\right]. \end{aligned} \tag{5.6}$$

The same inequality holds true if h is arbitrary and $h(X_i, X_j)$ is replaced in (5.6) by the randomized $\varepsilon_i\varepsilon_j h(X_i, X_j)$, or by the decoupled versions.

PROOF. Lemma 2.1 and the proof of necessity of Theorem 3.1 (see also Corollary 2.4) give the left hand side bound for decoupled and randomized kernels. The right hand side bound, also for decoupled and randomized kernels, follows from the proof of sufficiency of Theorem 3.1: let

$$K := \max\left[\|h\|_{L_2\mapsto L_2}, \limsup_{u\to\infty}\sqrt{\frac{E(h^2\wedge u)}{L_2 u}}\right];$$

if $K = 1$, the proof of Theorem 3.1 produces (3.4) for a fixed constant C that could be computed if necessary, as can be seen from steps 7 and 8 (the only ones that contribute to the limsup), and if $K \neq 1$, (3.4) with C replaced by CK is obtained by considering the kernel h/K. Then, Corollary 2.5 yields the right hand side of (5.6). De-randomization as in Section 4 gives the bounds (5.6) for canonical kernels. □

We know that when $Eh^2 < \infty$ and h is a canonical kernel of d variables, the limsup in (5.3) is just the quantity in (5.5), and even more, that the limit set of the sequence

$$\left\{\frac{d!}{(2n\log\log n)^{d/2}}\sum_{1\le i_1<i_2<\dots<i_d\le n} h(X_{i_1},\dots,X_{i_d})\right\}$$

is a.s.

$$\{E[h(X_1,\ldots,X_d)\prod_{i=1}^{d} f(X_i)] : Ef^2(X) \le 1\}$$

(Dehling, 1989, for $d = 2$ and Arcones and Giné, 1995, in general). Then, restricting to kernels of two variables, several concrete questions arise: 1) is any of the two summands in the bounds (5.6) superfluous?; 2) at least in the case when the kernel h defines a compact operator of L_2, can we determine the limit set of the LIL sequence from the limit set for finite rank h by operator approximation?, and of course, 3) what is the limit set in general? We will answer 1) by means of examples showing that, in general, both summands in the bound (5.6) are essential, and, regarding question 2) we will also determine the limit set for a class of kernels that induce compact operators in L_2. We wil show, moreover, that there are kernels h that give non-compact operators for which the LIL holds (the examples in Giné and Zhang (1996) define compact operators and suitable modifications will give non-compact ones). Finally, question 3) will remain open but we will show that the limit set is always an interval.

EXAMPLE 5.3. We consider the kernel

$$h(x,y) = \sum_{n=1}^{\infty} \frac{a_n}{b_n} I_n(x) I_n(y), \tag{5.7}$$

where $\{I_n\}$ is a sequence of functions on $\mathbf{R}$ with disjoint supports contained in $[0,1]$ such that $\int_{\mathbf{R}} I_n(u)du = 0$, $I_n(x) \in \{-1,0,1\}$ for each $x \in \mathbf{R}$, the sequence $\{b_n\}$ is defined by $b_n = \int_{\mathbf{R}} I_n^2(u)du$ and $\{a_n\}$ is an arbitrary bounded sequence of real numbers. Then, if, as will be the case, for X, Y i.i.d. uniform on $[0,1]$, $E|h(X,Y)| < \infty$, h is a canonical kernel for the uniform distribution on $[0,1]$. Since $\{b_n^{-1/2} I_n\}$ is an orthonormal sequence in $L_2 := L_2(\mathcal{L}(X))$, we have

$$\|h\|_{L_2 \mapsto L_2} = \sup_{n\in\mathbf{N}} |a_n|. \tag{5.8}$$

If we further assume that $\{a_n/b_n\}$ is an incresing sequence, then

$$\limsup_{u\to\infty} \frac{E(h^2 \wedge u)}{L_2 u} = \limsup_{n} \frac{\sum_{k=1}^{n} a_k^2 + \frac{a_n^2}{b_n^2}\left(\sum_{k=n+1}^{\infty} b_k^2\right)}{L_2(b_n^{-1})}.$$

So, if we choose $a_n = a$ for all n and I_n such that $b_n = \exp\left[-\exp(a^2 n/b)\right]$ for large n, then

$$\limsup_{u\to\infty} \frac{E(h^2 \wedge u)}{L_2 u} = b. \tag{5.9}$$

Thus, in this case, the kernel h satisfies the LIL by Theorem 3.1. Moreover, (5.8) and (5.9) show that the two quantities appearing in the bounds (5.6) are not comparable (and, in particular, neither of them is superfluous). In this type of examples, the operator in L_2 with kernel h is compact if and only if $\lim_n a_n = 0$, thus showing that there are canonical kernels h which satisfy the LIL but that do not define a compact operator on L_2.

If $Eh^2 < \infty$, then the operator norm dominates the bound in (5.6), as the limsup of the normalized truncated second moments of h is zero. Even for kernels h defining compact operators we may have that it is this second term that dominates the bound: for $a_n = 1/\sqrt{n}$ and $b_n = 2^{-n}$, consider the kernels $h_m(x,y) = \sum_{n=m}^{\infty} a_n b_n^{-1} I_n(x) I_n(y)$; then we have $\|h_m\|_{L_2 \mapsto L_2} = 1/\sqrt{m} \to 0$ whereas $\limsup_{u\to\infty} \frac{E(h_m^2 \wedge u)}{L_2 u} = 1$ for all m.

There is, however, a class of canonical kernels h satisfying the LIL and defining compact operators for which the limit set of the LIL sequence is the numerical range of the operator defined by h, as is the case when h has finite second moment. In the next proposition H will denote the operator on L_2 defined by extension of the equation $Hf(y) = Eh(X,y)f(X)$, $f \in L_\infty(\mathcal{L}(X))$ (this operator exists under condition (1.3)).

PROPOSITION 5.4. *Let h be a canonical kernel for the law of X such that*
a)

$$\limsup_{u\to\infty} \frac{E(h^2 \wedge u)}{L_2 u} = 0 \tag{5.10}$$

and
b) the operator H is a compact operator on $L_2(\mathcal{L}(X))$.
Then, the limit set of the sequence

$$\left\{ \frac{1}{2nL_2 n} \sum_{1\le i\ne j\le n} h(X_i, X_j) \right\} \tag{5.11}$$

is almost surely the closure of the set

$$\Big\{ Eh(X,Y)f(X)f(Y) : Ef^2(X) \le 1, \|f\|_\infty < \infty \Big\}, \tag{5.12}$$

that is, the numerical range of the operator H, $\{E(f(X)Hf(X)) : Ef^2(X) \le 1\}$.

PROOF. We set, from now on, $L_2 := L_2(\mathcal{L}(X))$. The proof consists in approximating the operator H with kernel h by suitable operators H_m with simple kernels, in particular, square integable kernels. We begin by showing that there exists an increasing sequence $\mathcal{G}_m$ of finite sub-σ-algebras of $\mathcal{S}$ such that, if P_m denotes the orthonormal projection onto the subspace of $\mathcal{G}_m$-measurable functions,

$$\|P_m Hf - Hf\|_{L_2} \to 0, \quad f \in L_2.$$

Indeed, H being a compact operator, its range is a separable set in L_2. Therefore we can find a sequence $\{g_i\} \subset L_2$ of simple functions such that the range of H is contained in the closure of the sequence $\{g_i\}$. Now, it is enough to set

$$\mathcal{G}_m := \sigma(g_1, \ldots, g_m)$$

to get the desired property. This is so because, obviously, $P_m g_i \to g_i$ for each $i \in \mathbf{N}$, and the set $\{f \in L_2 : P_m f \to f \text{ in } L_2 \text{ norm}\}$ is closed in view of $\|P_m\|_{L_2 \mapsto L_2} \le 1$.

For each $m \in \mathbf{N}$ we define

$$h_m(x,y) = \sum_{\substack{A,B \ atoms \ of \ \mathcal{G}_m \\ \Pr\{X\in A, Y\in B\}\neq 0}} \frac{Eh(X,Y)I_A(X)I_B(Y)}{\Pr\{X\in A\}\Pr\{Y\in B\}} I_A(x)I_B(y),$$

where, as usual, Y is an independent copy of X. In other words, h is defined by the condition

$$h_m(X,Y) = E\big(h(X,Y)\big|\sigma(X^{-1}(\mathcal{G}_m), Y^{-1}(\mathcal{G}_m))\big).$$

The operator H_m of L_2 with kernel h_m satisfies $H_m = P_m H P_m$, as is seen from its definition. Then, since $\|P_m Hf - Hf\|_{L_2} \to 0$ for any $f \in L_2$, and since H is a compact operator in L_2, we obtain that

$$\lim_{n\to\infty} \|H_m - H\|_{L_2\mapsto L_2} = 0. \tag{5.13}$$

To see this, we note that, since $(P_m - I)H$ is the adjoint of $H(P_m - I)$ and P_m has norm 1,

$$\|H_m - H\|_{L_2\mapsto L_2} = \|P_m H(P_m - I) + (P_m - I)H\|_{L_2\mapsto L_2} \le 2\|(P_m - I)H\|_{L_2\mapsto L_2};$$

now (5.13) follows by a simple compactness argument.

The result follows from the previous observation together with Theorem 5.2 applied to h_m and to $h - h_m$, by a standard approximation argument that we now sketch. Before we do this, we should note that the closure in L_2 of the set (5.12) is the numerical range of H because bounded functions are dense in L_2, the unit ball of L_2 is weakly compact and if $f_n \to f$ weakly, with $\|f_n\|_{L_2} \le 1$, then, by compactness of H, $Hf_n \to Hf$ weakly. Let us write $\langle\cdot,\cdot\rangle$ for the inner product in L_2, set

$$L := \{\langle Hf, f\rangle : \|f\|_{L_2} \le 1\}$$

and, for any kernel $g(x,y)$ of two variables,

$$\alpha_n(g) := \frac{1}{2nL_2 n} \sum_{1\le i\neq j\le n} g(X_i, X_j).$$

If $x \in L$ let $f \in L_2$ with $\|f\|_{L_2} \le 1$ be such that $x = \langle Hf, f\rangle$. Then, by the LIL for kernels with finite second moment, given $m \in \mathbf{N}$, for almost every ω there is a subsequence $n_{k(\omega)}$ such that

$$\alpha_{n_{k(\omega)}}(h_m(\omega)) \to \langle H_m f, f\rangle. \tag{5.14}$$

Also, since h satisfies (5.10) and h_m has finite second moment, Theorem 5.2 gives

$$\limsup_n |\alpha_n(h_m - h)| \le K\|H_m - H\|_{L_2\mapsto L_2} \quad \text{a.s.}. \tag{5.15}$$

Moreover, by (5.13),

$$\langle H_m g, g\rangle \to \langle Hg, g\rangle, \quad g \in L_2. \tag{5.16}$$

Combining these three limits we obtain that x is a.s. a limit point of the sequence $\{\alpha_n(h)\}$. Conversely, suppose now that x is a limit point of this sequence. Then, by (5.15), given $\varepsilon > 0$, for all m large enough and for almost every ω there exists a subsequence $n_{k(\omega)}$ such that

$$|x - \alpha_{n_{k(\omega)}}(h_m(\omega))| < \frac{\varepsilon}{2}.$$

Therefore, by the LIL for square integrable kernels and (5.16), there is $f \in L_2$ with $\|f\|_{L_2} \leq 1$ such that

$$|x - \langle Hf, f\rangle| < \varepsilon.$$

So, taking $\varepsilon = 1/n$, there is a sequence f_n in the unit ball of L_2 such that

$$x = \lim_n \langle Hf_n, f_n\rangle.$$

Since the unit ball of L_2 is weakly compact, the sequence $\{f_n\}$ has a subsequence $\{f_{n_k}\}$ that converges weakly to a function f in the unit ball of L_2. It then follows by compactness of H that $x = \langle Hf, f\rangle$, that is, $x \in L$. □

For example the previous proposition applies to the kernels h of Example 5.2 for $a_n = n^{-1/2}\ell(n)$ and $b_n = 2^{-n}$, where $\ell(n)$ is any slowly varying function tending to zero as $n \to \infty$. However, if $\ell(n) = 1$ then h still satisfies the LIL (1.1) by Theorem 1.1 and defines a compact operator in L_2, but Proposition 5.4 does not apply to it; actually, we do not know what the limit set is in this case.

As mentioned, the problem of determining the a.s. limit set of the sequence (5.11) in the general case remains open but we can show that it is an interval.

PROPOSITION 5.5. *Let h be a canonical kernel satisfying conditions (1.2) and (1.3). Then, the limit set of the LIL sequence (5.11) is an interval.*

PROOF. To prove that the limit set of the sequence (5.11) is an interval, it suffices to show that the difference of two consecutive terms of the sequence tends to zero a.s. By (1.2) and the law of large numbers for U-statistics (or by the LIL), this reduces to showing that

$$\frac{1}{n\log\log n}\sum_{1\leq i<n} h(X_i, X_n) \to 0 \quad \text{a.s.} \tag{5.17}$$

We will first prove

$$\frac{1}{n\log\log n}\sum_{1\leq i<n} \varepsilon_i h(X_i, Y_n) \to 0 \quad \text{a.s.} \tag{5.18}$$

and then will show that ε_i can be removed and that Y_n can be replaced by X_n.

To prove (5.18), it is enough to prove that for all $\delta > 0$

$$\sum_n \Pr\Big\{\max_{2^{n-1}<k\leq 2^n} \frac{1}{2^n \log n}\Big|\sum_{1\leq i<k} \varepsilon_i h(X_i, Y_k)\Big| > \delta\Big\} < \infty \tag{5.19}$$

(see e.g. the proof of Corollary 2.4). Let $h_n = hI_{A_n}$ and $\tilde{h}_n = h - h_n$, where

$$A_n = \big\{(x, y) : |h(x, y)| \leq 2^n \log n,\, f_n(y) \leq 2^n(\log n)^2\big\}.$$

Then as in Steps 1 and 2 of the proof of Theorem 3.1 we get

$$\sum_n \Pr\Big\{\max_{2^{n-1}<k\le 2^n} \frac{1}{2^n \log n}\Big|\sum_{1\le i<k} \varepsilon_i \tilde{h}_n(X_i, Y_k)\Big| > \delta\Big\} < \infty.$$

In order to prove

$$\sum_n \Pr\Big\{\max_{2^{n-1}<k\le 2^n} \frac{1}{2^n \log n}\Big|\sum_{1\le i<k} \varepsilon_i h_n(X_i, Y_k)\Big| > \delta\Big\}$$
$$\le \sum_n 2^n \Pr\Big\{\Big|\sum_{1\le i<2^n} \varepsilon_i h_n(X_i, Y)\Big| > \delta 2^n \log n\Big\} < \infty.$$

we apply Chebyshev's inequality as in Step 3, reducing the above inequality to convergence of the two series

$$\sum_n \frac{1}{2^{2n}(\log n)^4} E h_n^4(X, Y) < \infty,$$

$$\sum_n \frac{1}{2^n(\log n)^4} E h_n^2(X_1, Y) h_n^2(X_2, Y) < \infty.$$

But these two series converge, just like the first and second series in Step 3. (5.19) is thus proved.

Next we show that we can remove the Rademacher variables from (5.18), that is, that (5.18) implies

$$\frac{1}{n \log\log n} \sum_{1\le i<n} h(X_i, Y_n) \to 0 \quad \text{a.s.} \tag{5.20}$$

Let $\{\tilde{X}_i\}$ be a copy of $\{X_i\}$, independent of $\{X_i\}$ and $\{Y_i\}$, and set

$$\xi_n := \frac{1}{n \log\log n} \sum_{1\le i<n} h(X_i, Y_n), \quad \tilde{\xi}_n := \frac{1}{n \log\log n} \sum_{1\le i<n} h(\tilde{X}_i, Y_n).$$

If (5.18) holds, then $\xi_n - \tilde{\xi}_n \to 0$ a.s. by Fubini's theorem and the equidistribution of the variables X_i. Hence, (5.20) will follow by a standard argument if $\xi_n \to 0$ in probability conditionally on the sequence $\{Y_i\}$. So, assuming (wlog) that the variables X and Y are defined on different factors of a product probability space $\Omega' \times \Omega$, we must show that

$$\frac{1}{a_n} \sum_{1\le i<n} h(X_i, Y_n(\omega)) \to 0 \ \text{ in pr.}, \ \omega - \text{a.s.}, \tag{5.21}$$

where, for ease of notation, we set $a_n := (nL_2 n)^{-1}$. Now, since

$$\frac{1}{a_n} \sum_{1\le i<n} \varepsilon_i h(X_i, Y_n) \to 0 \ \text{ in pr.}, \ \omega - \text{a.s.}$$

by (5.18), Lévy's inequality applied conditionally on $\{Y_i\}$ gives

$$n\Pr_X\{|h(X,Y_n)| > a_n\} \to 0 \text{ a.s.} \tag{5.22}$$

and then, Hoffmann-Jørgensen's inequality applied conditionally after truncation, yields

$$\frac{n}{a_n^2} E_X h^2(X,Y_n) I_{\{|h(X,Y_n)|\le a_n\}} \to 0 \text{ a.s.} \tag{5.23}$$

Moreover,

$$\frac{n}{a_n} E_X h(X,Y_n) I_{\{|h(X,Y_n)|\le a_n\}} \to 0 \text{ a.s.} \tag{5.24}$$

To prove that this last limit holds, note first that, since $E_X h = 0$,

$$E_X h(X,Y_n) I_{|h(X,Y_n)|\le a_n} = E_X h(X,Y_n) I_{\{|h(X,Y_n)|>a_n\}},$$

and then that

$$\sum_n \frac{n}{a_n} E|h(X,Y)| I_{|h(X,Y)|>a_n} < \infty$$

because, after exchanging expectation and sum and then summing on n, we see that this series is bounded by a constant times $E\frac{h^2}{L_2^2|h|}$, which is finite. Now, (5.22)-(5.24) give that, for all $\varepsilon > 0$,

$$\Pr_X\left\{\frac{1}{a_n}\Big|\sum_{1\le i\le n} h(X_i,Y_n)\Big| > \varepsilon\right\} \le n\Pr_X\{|h| > a_n\} + I_{\{na_n^{-1}|E_X h I_{\{|h|\le a_n\}}|>\varepsilon/2\}}$$
$$+ \frac{4}{\varepsilon^2}\frac{n}{a_n^2} E_X h^2 I_{\{|h|\le a_n\}} \to 0 \quad \text{a.s.},$$

proving (5.21), hence, (5.20).

Finally, to undecouple, assume (5.20) holds. By Theorem 1.1 and the $0-1$ law we know that

$$\limsup_n \frac{1}{n\log\log n}\Big|\sum_{1\le i<n} h(X_i,X_n)\Big| = C \quad \text{a.s.} \tag{5.25}$$

for some $C < \infty$, and must show that $C = 0$. Then, we can assume that this limsup is attained by the sequence of even terms, that is,

$$\limsup_n \frac{\left|\sum_{1\le i<2n} h(X_i,X_{2n})\right|}{2n\log\log(2n)} = C \quad \text{a.s.} \tag{5.26}$$

(otherwise we can take the subsequence of odd terms from (5.25) and continue in the same way as we will now proceed). But

$$\begin{aligned}\limsup_n \frac{1}{2n\log\log(2n)}\Big|\sum_{1\le i<2n} h(X_i,X_{2n})\Big| &\le \limsup_n \frac{1}{2n\log\log(2n)}\Big|\sum_{\substack{1<i<2n\\ i\ even}} h(X_i,X_{2n})\Big| \\ &\quad + \limsup_n \frac{1}{2n\log\log(2n)}\Big|\sum_{\substack{1\le i<2n\\ i\ odd}} h(X_i,X_{2n})\Big| \\ &= \limsup_n \frac{1}{2n\log\log(2n)}\Big|\sum_{1\le i<n} h(X_i,X_n)\Big| \\ &\quad + \limsup_n \frac{1}{2n\log\log(2n)}\Big|\sum_{1\le i<n+1} h(X_i,Y_{n+1})\Big| \\ &= \frac{C}{2}\end{aligned}$$

by (5.25) and (5.20). This contradicts (5.26) unless $C = 0$, proving (5.17). □

Acknowledgements. The second and third named authors carried out the research for the present article respectively at the departments of Mathematics of Texas A&M University and Georgia Tech, and wish to acknowledge their hospitality.

References

Arcones, M. and Giné, E. (1995). On the law of the iterated logarithm for canonical U–statistics and processes. *Stoch. Proc. Appl.* **58** 217-245.

Dehling, H. (1989). Complete convergence of triangular arrays and the law of the iterated logarithm for degenerate U–statistics. *Stat. Probab. Letters* **7** 319–321.

Dehling, H.; Denker, M. and Philipp, W. (1984). Invariance principles for von Mises and U-statistics. *Zeits. Wahrsch. verw. Geb.* **67** 139-167.

Dehling, H.; Denker, M. and Philipp, W. (1986) A bounded law of the iterated logarithm for Hilbert space valued martingales and its application to U–statistics. *Prob. Th. Rel. Fields* **72** 111-131.

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

Giné, E. and Zhang, C.-H. (1996). On integrability in the LIL for degenerate U–statistics. *J. Theoret. Probab.* **9** 385–412.

Giné, E. and Zinn, J. (1992). Marcinkiewicz type laws of large numbers and convergence of moments for U–statistics. *Probability in Banach Spaces 8* 273–291. Birkhäuser, Boston.

Giné, E. and Zinn, J. (1994). A remark on convergence in distribution of U–statistics. *Ann. Probab.* **22** 117–125.

Goodman, V. (1996). A bounded LIL for second order U–statistics. Preprint.

Halmos, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Statist.* **17** 34–43.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293-325.

Latała, R. (1999). Tails and moment estimates for some type of chaos. *Studia Math.*, to appear.

Latała, R. and Zinn, J. (1999). Necessary and sufficient conditions for the strong law of large numbers for U-statistics. Preprint.

Ledoux, M. (1996). On Talagrand's deviation inequalities for product measures. ESAIM, P&S,**1** 63-87. (http://www.emath.fr/Maths/Ps)

Montgomery–Smith, S. (1993). Comparison of sums of independent identically dsitributed random variables. *Prob. Math. Statist.* **14** 281–285.

Rubin, M. and Vitale, R. A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8** 165-170.

Serfling, R. J. (1971). The law of the iterated logarithm for U–statistics and related von Mises functionals. *Ann. Math. Statist.* **42** 1794.

Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22** 28-76.

Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505-563.

Teicher, H. (1995). Moments of randomly stopped sums revisited. *J. Theoret. Probab.* **8** 779-794.

Zhang, C.-H. (1999). Sub-Bernoulli functions, moment inequalities and strong laws for nonnegative and symmetrized U-statistics. *Ann. Probab.* **27** 432-453.

Department of Mathematics
and Department of Statistics
University of Connecticut
Storrs, CT 06269
USA
gine@uconnvm.uconn.edu

Institute of Mathematics
Warsaw University
Banacha 2
02-097 Warszawa
Poland
kwapstan@mimuw.edu.pl
rlatala@mimuw.edu.pl

Department of Mathematics
Texas A&M University
College Station, TX 77843
jzinn@math.tamu.edu