The Annals of Probability
1980, Vol. 8, No. 3, 405–418

LOG LOG LAWS FOR EMPIRICAL MEASURES[1]

BY J. KUELBS AND R. M. DUDLEY

University of Wisconsin, Madison, and Massachusetts Institute of Technology

Let $(X, \mathcal{A}, P)$ be a probability space and $\mathcal{C}$ a collection of measurable sets. Suppose $\mathcal{C}$ is a Donsker class, i.e., the central limit theorem for empirical measures holds uniformly on $\mathcal{C}$, in a suitable sense. Suppose also that suitable ($P\varepsilon$-Suslin) measurability conditions hold. Then we show that the log log law for empirical measures, in the Strassen-Finkelstein form, holds uniformly on $\mathcal{C}$.

1. Introduction and preliminaries. Let $(X, \mathcal{A}, P)$ be any probability space. Let $X_1, X_2, \cdots,$ be i.i.d. (independent and identically distributed) in X with distribution P. Let P_n be the random nth empirical measure $n^{-1}(\delta_{X_1} + \cdots + \delta_{X_n})$. Let $LLn := \ln \ln n$, $n \geqslant 3$; $LLn := 1$, $n = 1, 2$ (we use " $:=$ " to mean "equals by definition"). For any $C \in \mathcal{A}$ let

$$I(n, C) := (P_n - P)(C)(n/2LLn)^{1/2}, \sigma(C) := (P(C)(1 - P(C)))^{1/2}.$$

Khinchin (1924) proved the first log log law: for each $C \in \mathcal{A}$, almost surely

$$\limsup_{n\to\infty} I(n, C) = -\liminf_{n\to\infty} I(n, C) = \sigma(C).$$

For $\mathcal{C} \subset \mathcal{A}$ let

$$I(n, \mathcal{C}) := \sup_{C\in\mathcal{C}} I(n, C), \qquad A(n, \mathcal{C}) := \sup_{C\in\mathcal{C}} |I(n, C)|,$$

$$\sigma(\mathcal{C}) := \sup_{C\in\mathcal{C}} \sigma(C), \qquad T(\mathcal{C}) := \sup_n A(n, \mathcal{C}).$$

Given any collection $\mathcal{C} \subset \mathcal{A}$, let $LS(\mathcal{C}) := LS(\mathcal{C}, P) := \limsup_{n\to\infty} A(n, \mathcal{C})$. If the $I(n, \mathcal{C})$ are measurable, at least, then $LS(\mathcal{C})$ is a constant a.s. with $\sigma(\mathcal{C}) \leqslant LS(\mathcal{C}) \leqslant +\infty$ by Khinchin's theorem. We do not know any $\mathcal{C}$ for which $\sigma(\mathcal{C}) < LS(\mathcal{C}) < +\infty$. If $LS(\mathcal{C}) = \sigma(\mathcal{C})$, we call $\mathcal{C}$ a *log log class* for P.

Chung (1949)[2] proved that for any law P on $\mathbb{R}^1$ with a continuous distribution function, $\{]-\infty, x]: x \in \mathbb{R}\}$ is a log log class. Kiefer (1961) extended this log log law for empirical distribution functions to $\mathbb{R}^k$, $k > 1$.

For log log laws of Strassen's type, we define

$$H_0 := \{f \in L^2(X, \mathcal{A}, P): \int f\, dP = 0\}.$$

Let B be the unit ball of the Hilbert space H_0,

$$B := \{f \in H_0: \|f\|^2 := \int |f|^2\, dP \leqslant 1\}.$$

Then B defines a set $B_{\mathcal{C}}$ of functions on $\mathcal{C}$:

$$B_{\mathcal{C}} := \{C \to \int_C f\, dP, C \in \mathcal{C}: f \in B\}.$$

Received January 25, 1978; revised November 29, 1978 (before the present Editor's term began).
[1]Partially supported by National Science Foundation Grants MCS77-01098 and MCS76-07211 A01.
[2]Added in proof. See also Smirnov (1944).
AMS 1970 *subject classifications.* Primary 60F15; secondary 60F05, 28A05.
Key words and phrases. Log log law, empirical measures, Donsker class.

We call $\mathcal{C}$ a *Strassen log log class* for P iff, with probability 1, $I(n,\cdot)$ restricted to $\mathcal{C}$ form a relatively compact set whose set of limit points coincides with $B_{\mathcal{C}}$.

Finkelstein (1971) proved that $\{]-\infty, x]: x \in \mathbb{R}\}$ is a Strassen log log class, for any nonatomic P with bounded support. Richter (1974) extended this result to $\mathbb{R}^k$, $k \geqslant 1$, and arbitrary P.

For a Strassen log log class $\mathcal{C}$, we have in particular almost surely

$$\limsup_{n\to\infty} \sup_{C\in\mathcal{C}}\{I(n, C) - \sigma(C)\} = 0 = \liminf_{n\to\infty} \inf_{C\in\mathcal{C}}\{I(n, C) + \sigma(C)\},$$

and so

$$\limsup_{n\to\infty} \sup_{C\in\mathcal{C}} I(n, C) = -\liminf_{n\to\infty} \inf_{C\in\mathcal{C}} I(n, C) = \limsup_{n\to\infty} A(n, \mathcal{C}) = \sigma(\mathcal{C}).$$

Thus every Strassen log log class is a log log class. We do not know any log log class which is not a Strassen log log class.

For $\mathcal{C}$ infinite, the Banach space $l^\infty(\mathcal{C})$ of bounded real functions on $\mathcal{C}$ with supremum norm $\|\cdot\|_{\mathcal{C}}$ is nonseparable. Thus, as in [4], some measurability problems arise. We recall some definitions. If Y is a set and $\mathfrak{B}$ is a σ-algebra of subsets of Y, then $(Y, \mathfrak{B})$ is a *measurable space*. It is called *Suslin* if there is a metric on Y for which $\mathfrak{B}$ is the σ-algebra of Borel sets and such that there is a Borel measurable map of a complete separable metric space onto Y. Given a probability space $(X, \mathcal{A}, P)$, $\mathcal{C} \subset \mathcal{A}$, and a σ-algebra $\mathcal{S}$ of subsets of $\mathcal{C}$, we say $(X, \mathcal{A}; \mathcal{C}, \mathcal{S})$ is *$P\in$-Suslin* iff the following conditions hold:

(i) $(X, \mathcal{A})$ and $(\mathcal{C}, \mathcal{S})$ are both Suslin measurable spaces;

(ii) The $\in$ relation, $\{\langle x, C\rangle: x \in C \in \mathcal{C}\}$ is measurable in $X \times \mathcal{C}$ for the product σ-algebra $\mathcal{A} \times \mathcal{S}$ generated by rectangles $A \times E$, $A \in \mathcal{A}$, $E \in \mathcal{S}$;

(iii) For the pseudometric $d_P(C, D) := P(C \triangle D)$ on $\mathcal{C}$, all open sets belong to $\mathcal{S}$.

We say $\mathcal{C}$ is $P\in$-Suslin iff $(X, \mathcal{A}; \mathcal{C}, \mathcal{S})$ is for some $\mathcal{S}$.

A class $\mathcal{C}$ is called a *Donsker class* iff the normalized empirical measures $\nu_n := n^{\frac{1}{2}}(P_n - P)$ converge in law (as defined in [4], Section 1) in a certain subspace $D_0(\mathcal{C}, P)$ of the space of all bounded real functions on $\mathcal{C}$ with supremum norm; [4], Section 1 gives detailed definitions.

2. Statements of results. Pisier ((1975) Theorem 4.3) proved that in any separable Banach space, if $V_1, V_2, \cdots$ are independent and identically distributed, with $EV_1 = 0$ and $E\|V_1\|^2 < \infty$, and if the central limit theorem holds, i.e., if $(V_1 + \cdots + V_n)/n^{\frac{1}{2}}$ converges in law, then the Strassen log log law holds. In our case, $V_1 := \delta_{X_1} - P$, and $\|f\| := \|f\|_{\mathcal{C}} := \sup_{C\in\mathcal{C}}|f(C)|$, so $\|V_1\| \leqslant 1$, which implies some exponential bounds, but in general $\|\cdot\|_{\mathcal{C}}$ is nonseparable.

2.1. THEOREM. *Every $P\in$-Suslin Donsker class $\mathcal{C}$ is a Strassen log log class for P, and for every $\beta > 0$, $E \exp(\beta T(\mathcal{C})^2) < \infty$.*

The proof will be given in Section 3. Here we give some corollaries using the results of [4].

2.2. COROLLARY. *For any sequence* $\mathcal{C} := \{C_m\}_{m\geq 1}$ *of measurable sets such that for some* $r < \infty$,

$$\Sigma_m P(C_m)^r (1 - P(C_m))^r < \infty,$$

$\mathcal{C}$ *is a Strassen log log class.*

PROOF. By [4], Theorem 2.1, $\mathcal{C}$ is a Donsker class. Let $\mathbb{S}$ be the σ-algebra of all subsets of $\mathcal{C}$. Then $(X, \mathcal{A}; \mathcal{C}, \mathbb{S})$ is $P\in$-Suslin if $(X, \mathcal{A})$ is Suslin. Take the measurable map f: $x \to \{1_{C_j}(x)\}_{j\geq 1}$ from X into a countable product Z_2^∞ of copies of $\{0, 1\}$, with its usual product σ-algebra, a Suslin measurable space. It is enough to prove the result for $P \circ f^{-1}$ and the sets $D_j := \{\{y_i\}: y_j = 1\}$. So we may assume $(X, \mathcal{A})$ is Suslin. □

Corollary 2.2 applies in particular to any sequence of disjoint measurable sets ($r = 1$). Thus we reobtain the theorem of Olshen and Siegmund (1971).

2.3. PROPOSITION. *If* C_m *are independent for* P, $p_m := P(C_m)$, *and* $\Sigma_m p_m^n (1 - p_m)^n = +\infty$ *for all* n, *then* $LS(\mathcal{C}) = +\infty$ a.s., *where* $\mathcal{C} := \{C_m\}_{m\geq 1}$, *so* $\mathcal{C}$ *is not a log log class.*

PROOF. Since $|I(n, C_m)| = |I(n, X \setminus C_m)|$ we can assume $p_m \leq \frac{1}{2}$ for all m. Then

$$\Pr(X_j \in C_m \text{ for } j = 1, \cdots, n) = p_m^n,$$

and $\Sigma_m p_m^n = +\infty$ implies by independence (Borel-Cantelli) that this event occurs for infinitely many m a.s. for each n. Then $|I(n, \mathcal{C})| \geq (n/2LLn)^{\frac{1}{2}}(1 - 1/2) \to \infty$ a.s. with n. □

Thus Corollary 2.2 gives a best possible result in the case of independent sets.

Given a collection $\mathcal{C}$ of subsets of a set X, let $V(\mathcal{C})$ be the smallest n such that for every set F with n elements, not every subset of F is of the form $F \cap C$, $C \in \mathcal{C}$. If $V(\mathcal{C}) < +\infty$ we call $\mathcal{C}$ a *Vapnik-Červonenkis class* (VCC). We refer to Vapnik and Červonenkis (1971, 1974) and [4], Proposition 4.5 and Section 7, for discussion of such classes. Theorem 2.1 and [4], Theorem 7.1 give:

2.4. COROLLARY. *Any* $P\in$*-Suslin* VCC $\mathcal{C}$ *is a Strassen log log class for* P.

For a real function g on X let $\text{pos}(g) := \{x \in X: g(x) > 0\}$. For a set G of such functions let $\text{pos}(G) := \{\text{pos}(g): g \in G\}$.

2.5. COROLLARY. *Let* (X, e) *be a locally compact, separable metric space and* G *a finite-dimensional real vector space of continuous real functions on* X. *Then* pos(G) *is a Strassen* log log *class for any Borel probability measure* P *on* X.

PROOF. For the usual Borel measurability structure on G, and any $x \in X$, $g \to g(x)$ is linear and hence measurable. Thus by [4], Proposition 4.5 and the remark after it, pos (G) is $P\in$-Suslin. By [4], Theorem 7.2, pos(G) is a VCC. Thus Corollary 2.4 applies, $\square$

In particular, for $X = \mathbb{R}^k$ we can take G to be the collection of all polynomials of degree $\leq d$ for any fixed d. (Philipp, 1972, showed that ellipsoids parallel to the axes form a log log class.) Steele (1978, Section 5) proved the strong law.

Given P, $\mathcal{C}$ and $\varepsilon > 0$, let $N_I(\varepsilon, \mathcal{C}, P) := \min\{m: \exists A_1, \cdots, A_m \in \mathcal{Q}:$ for all $C \in \mathcal{C}, \exists r, s: A_r \subset C \subset A_s$ and $P(A_s \setminus A_r) < \varepsilon\}$.

Then Theorem 2.1 and [4], Theorem 5.1 give:

2.6. COROLLARY. *If* $\int_0^1 (\log N_I(x^2, \mathcal{C}, P))^{\frac{1}{2}} dx < \infty$ *and* $\mathcal{C}$ *is* $P\in$*-Suslin then* $\mathcal{C}$ *is a Strassen log log class.*

In particular, [4], Theorem 5.13, gives:

2.7. COROLLARY. (Philipp, 1973). *Let* P *on* $\mathbb{R}^2$ *have a bounded density (with respect to Lebesgue measure) and let* $\mathcal{C}(U)$ *be the collection of all convex subsets of a bounded open set* U. *Then* $\mathcal{C}(U)$ *is a Strassen log log class.*

For the collection $\mathcal{C}(k)$ of all convex sets in $\mathbb{R}^k$, $k \geq 4$, and the uniform probability on the unit (hyper) cube, the log log law fails; in fact for some constant δ, any n and $X_1, \cdots, X_n$, $\sup_{C \in \mathcal{C}(k)} |\nu_n(C)| \geq \delta n^{(k-3)/(2k+2)}$ (Schmidt, 1975, Theorem 1, cf. also Stute, 1977). For $k = 3$ the question remains open as far as we know.

Now for $X = \mathbb{R}^k$ let $J(k, \alpha, M)$ be the collection of compact sets with boundaries defined by functions with all partial derivatives of orders $\leq \alpha$ bounded by M, as defined in [3]. Then Theorem 2.1 above and T.-G. Sun's theorem [4], Theorem 5.12, give:

2.8. COROLLARY. (cf. Révész, 1976, for $\alpha = k$). *If* P *on* $\mathbb{R}^k$ *has a bounded density with respect to Lebesgue measure and* $\alpha > k - 1$ *then* $J(k, \alpha, M)$ *is a Strassen log log class.*

Révész (1976) considers unions of at most m sets in collections of sets with differentiable boundaries. Let $\mathcal{C}$ be any class of sets and for $m = 1, 2, \cdots,$ let

$$U(\mathcal{C}, m) := \left\{ \bigcup_{1 \leq j \leq m} U_j : U_j \in \mathcal{C} \right\}.$$

It is easily seen that for any $\varepsilon > 0$,

$$N_I(\varepsilon, U(\mathcal{C}, m), P) \leq N_I(\varepsilon/m, \mathcal{C}, P)^m.$$

So if the hypothesis on N_I in Corollary 2.6 holds for $\mathcal{C}$, it also holds for $U(\mathcal{C}, m)$ for each finite m. Likewise, unions can be replaced by intersections, etc.

We do not know in general whether the $P\in$-Suslin property for $\mathcal{C}$ implies it for $U(\mathcal{C}, m)$. The following result will suffice for some applications.

Let (X, d) be a metric space and $\mathcal{U}$ the collection of all open sets (topology) on X. Let $\mathcal{E}$ be the Effros Borel structure on $\mathcal{U}$, i.e., $\mathcal{E}$ is the σ-algebra of subsets of

$\mathfrak{U}$ generated by all sets

$$\{U \in \mathfrak{U}: U \supset V\}, V \in \mathfrak{U}.$$

2.9. PROPOSITION. *If (X,d) is a locally compact separable metric space with topology $\mathfrak{U}$, then*

(I) $$\langle U, V\rangle \to U \cup V \text{ and } \langle U, V\rangle \to U \cap V$$

are measurable: $(\mathfrak{U}, \mathfrak{S}) \times (\mathfrak{U}, \mathfrak{S}) \to (\mathfrak{U}, \mathfrak{S})$;

(II) *If $\mathcal{C} \subset \mathfrak{U}$ and $\mathcal{C}$ is a Suslin set for $\mathfrak{S}$, then for each m, $U(\mathcal{C}, m)$ is also a Suslin set.*

PROOF. For (I) we must show that for each open W,

$$S_W := \{\langle U, V\rangle: U \cup V \supset W\} \in \mathfrak{S} \times \mathfrak{S}.$$

(For intersection in place of union this is trivial, requiring no hypothesis on X.) Since W is a countable union of compact sets, it is enough to show $S_K \in \mathfrak{S} \times \mathfrak{S}$ for K compact.

Let $F^\varepsilon := \{x: d(x,y) < \varepsilon \text{ for some } y \in F\}$. For $n = 1, 2, \cdots$, let F_n be a finite subset of K with $K \subset F_n^{1/n}$. For each subset G of F_n let

$$A_G := \{\langle U, V\rangle: U \supset G^{1/n} \quad \text{and} \quad V \supset (F_n \setminus G)^{1/n}\}.$$

Then $A_G \in \mathfrak{S} \times \mathfrak{S}$ and $S_K = \cup_n \cup \{A_G: G \subset F_n\}$, proving (I).

Iterating (I), we see that $\{\langle U_1, \cdots, U_m\rangle: W \subset \bigcup_{1 \leq j \leq m} U_j\}$ is $\mathfrak{S} \times \cdots \times \mathfrak{S}$ measurable. Then, $U(\mathcal{C}, m)$ is the image of a Suslin set $\mathcal{C} \times \cdots \times \mathcal{C}$ by union, an Effros measurable map, hence $U(\mathcal{C}, m)$ is also a Suslin set. $\square$

There is also a result corresponding to (2.9) for collections of closed sets.

Hence, in Révész (1976), Theorems 2 and 2* (on collections of sets with k-times differentiable boundaries in $\mathbb{R}^k$) now follow from our Corollary 2.6 and Proposition 2.9, in the light of [3] and [4], proof of Theorem 5.12.

3. Proof that $P\in$-Suslin Donsker classes are Strassen log log classes. Given a probability space $(X, \mathcal{A}, P)$ and a $P\in$-Suslin class $\mathcal{C} \subset \mathcal{A}$, let $X(1)$, $Y(1)$, $X(2)$, $Y(2), \cdots$, be independent on a probability space $(\Omega, \Pr)$ with $\mathcal{L}(X_j) = \mathcal{L}(Y_j) = P$ for all j, where $X_j := X(j)$, $Y_j := Y(j)$, and $\mathcal{L}(X_j) := \Pr \circ X_j^{-1}$ on $(X, \mathcal{A})$.

Let $D_j := \delta_{X(j)} - \delta_{Y(j)}$ and $S_n := \Sigma_{1 \leq j \leq n} D_j$. Let $\nu_n := n^{\frac{1}{2}}(P_n - P)$ where $P_n := n^{-1}(\Sigma_{1 \leq j \leq n} \delta_{X(j)})$. Let $\|f\|_{\mathcal{C}} := \sup_{C \in \mathcal{C}} |f(C)|$ for bounded real functions f on $\mathcal{C}$. Let $d_P(A, B) := P(A \setminus B) + P(B \setminus A)$.

Here is an outline of the proof of Theorem 2.1, which is similar to proofs in [11]. Finkelstein (1971) proved the result for any finite collection $\mathcal{C}$. We first consider symmetrized variables $S_n/n^{\frac{1}{2}} := \nu_n - \nu_n'$ where ν_n' is an independent copy of ν_n. The central limit theorem (Donsker property) will imply, as in (3.4), that $\sup_n E\|S_n\|/n^{\frac{1}{2}} < \infty$. Using boundedness of the basic random variables, $\|\delta_{X_i}\|_{\mathcal{C}} \leq 1$, we will get exponential bounds (3.17) $E \exp(h\|S_n\|) \leq \exp(hE\|S_n\| + 4nh^2)$. For symmetric variables we have a Lévy inequality (3.1), allowing a proof that $\lim\sup_{n\to\infty}\|S_n\|_{\mathcal{C}}/(2nLLn)^{\frac{1}{2}} < \infty$ a.s. via a subsequence of the S_n.

Now, we approximate $\mathcal{C}$ within δ for d_P by a finite set $\mathcal{Q} = \{A_1, \cdots, A_k\}$. Let $\mathcal{C}(j) := \{C \in \mathcal{C}: d_P(C, A_j) < \delta\}$. We will replace $\mathcal{C}$ above by $\mathcal{C}(j)$ and $\|\cdot\|_{\mathcal{C}}$ by $\|\cdot\|_j$ defined (not just $\sup_{\mathcal{C}(j)}|\cdot|$) by $\|f\|_j := \sup_{C\in\mathcal{C}(j)}|f(C) - f(A_j)|$. Given $\varepsilon > 0$, for δ small enough we can then get, as in (3.14), "$< \varepsilon$" in place of "$< \infty$" above. Next, (3.16) removes the symmetry assumption, and Finkelstein's theorem will finish the proof.

Now we begin the detailed proof of (2.1). Let V be a real vector space. Recall that a *seminorm* on V is a nonnegative real function $\|\cdot\|$ satisfying $\|x + y\| \leqslant \|x\| + \|y\|$ and $\|cx\| = |c|\,\|x\|$ for all $x, y \in V$ and real c.

For our case $V = D_0(\mathcal{C}, P)$, we will have various seminorms $\|\cdot\|_{\mathcal{C}}$ and $\|\cdot\|_j$, and even various σ-algebras (e.g., σ-algebras $\mathcal{B}_b$ generated by balls $\{x: \|x - v\| < r\}$ for the various seminorms). Random variables with values in V will be required to be measurable only by specific hypotheses in Lemmas 3.1 and 3.2.

The following fact is well known, at least for separable normed spaces with Borel measurability. In our applications, D_j will be independent and symmetric. (Measurability problems prevent this from implying directly the hypotheses of 3.1).

3.1. LEMMA. (Lévy inequality). *Let V be any real vector space and $\|\cdot\|$ a seminorm on V. Let $D_1, D_2, \cdots,$ be random variables with values in V, $S_n := D_1 + \cdots + D_n$, and $S_{mn} := S_m - (D_{m+1} + \cdots + D_n)$, $1 \leqslant m \leqslant n$. Suppose that whenever $1 \leqslant m \leqslant n$, $\|S_m\|$ and $\|S_{mn}\|$ are measurable, and that $\langle\|S_1\|, \cdots, \|S_m\|, \|S_n\|\rangle$ have the same joint distribution on $\mathbb{R}^{m+1}$ as $\langle\|S_1\|, \cdots, \|S_m\|, \|S_{mn}\|\rangle$. Then for any $K > 0$ and $n = 1,2, \cdots$, $\Pr\{\max_{j\leqslant n}\|S_j\| \geqslant K\} \leqslant 2\Pr\{\|S_n\| \geqslant K\}$.*

PROOF. As in Kahane ((1968), page 12), for $1 \leqslant m \leqslant n$ let A_m be the event $\{\|S_1\| < K, \cdots, \|S_{m-1}\| < K, \|S_m\| \geqslant K\}$. Since $S_m = (S_n + S_{mn})/2$, A_m is the union of the two measurable events $A_m \cap \{\|S_n\| \geqslant K\}$ and $A_m \cap \{\|S_{mn}\| \geqslant K\}$, which have the same probability. Thus

$$\Pr\{\|S_n\| \geqslant K\} = \Sigma_{m=1}^{n}\Pr\{A_m \cap \{\|S_n\| \geqslant K\}\}$$

$$\geqslant \tfrac{1}{2}\Sigma_{m=1}^{n}\Pr(A_m) = \tfrac{1}{2}\Pr\{\max_{m\leqslant n}\|S_m\| \geqslant K\}. \qquad \square$$

The next fact is also well known in the separable case (Kahane, 1968, page 16; Hoffmann-Jørgensen, 1974, 3.3; Jain and Marcus, 1975, Lemma 3.4).

3.2. LEMMA. *Assume the hypotheses of* 3.1 *and that for some $K < \infty$, $\|D_i\| \leqslant K$ for all i, and for $1 \leqslant m \leqslant n$, the two random variables: $\langle\|S_1\|, \cdots, \|S_m\|\rangle$ in $\mathbb{R}^m$, and $\|S_n - S_m\|$ in $\mathbb{R}^1$, are independent. Then for any $t > K$ and $n = 1,2, \cdots$, $\Pr\{\|S_n\| \geqslant 3t\} \leqslant 4\Pr\{\|S_n\| \geqslant t\}^2$.*

PROOF. As in the cited references, let $T(\omega) := \inf\{n \geqslant 1: \|S_n\, \omega\| \geqslant t\}$. Then using Lemma 3.1 twice, we get

$$\begin{aligned}\Pr\{\|S_n\| \geqslant 3t\} &= \Sigma_{m=1}^{n-2} \Pr\{T = m, \|S_n\| \geqslant 3t\}\\ &\leqslant \Sigma_{1 \leqslant m < n} \Pr\{T = m, \|S_n - S_m\| \geqslant 2t - K\}\\ &\leqslant \Sigma_{1 \leqslant m < n} \Pr\{T = m\}\Pr\{\|S_n - S_m\| \geqslant 2t - K\}\\ &\leqslant \Sigma_{1 \leqslant m < n} \Pr\{T = m\} 2 \Pr\{\|S_n\| \geqslant t\}\\ &\leqslant 2 \Pr\{\|S_n\| \geqslant t\}\Pr\{\max_{j \leqslant n}\|S_j\| \geqslant t\}\\ &\leqslant 4 \Pr\{\|S_n\| \geqslant t\}^2.\end{aligned}$$

Lemmas 3.1 and 3.2 (with $K = 1$) apply to our $D_i := \delta_{X(i)} - \delta_{Y(i)}$ and $\|\cdot\| = \|\cdot\|_{\mathcal{C}}$ by [4], Proposition 3.2.

3.3 LEMMA. *For any $P\in$-Suslin Donsker class $\mathcal{C}$ and $\varepsilon > 0$, there is some finite r and a decomposition $\mathcal{C} = \mathcal{C}_1 \cup \cdots \cup \mathcal{C}_r$, $\mathcal{C}_j$ disjoint, such that each $\mathcal{C}_j := \mathcal{C}(j)$ is $P\in$-Suslin and for some $A_j \in \mathcal{C}_j$, $N = N(\varepsilon) < \infty$, and $\|f\|_j := \sup_{C \in \mathcal{C}(j)}|f(C) - f(A_j)|$, we have for each $j = 1, \cdots, r$,*

$$\sup_{n \geqslant N} \Pr\left(\|S_n\|_j > n^{\frac{1}{2}}\varepsilon\right) < \tfrac{1}{80}.$$

PROOF. We can assume $\varepsilon < \frac{1}{80}$. By [4], Theorem 1.2b, there is a $\delta = \delta(\varepsilon) > 0$ and an $N = N(\varepsilon)$ such that for $n \geqslant N$, the outer probability

$$\Pr^*\{\sup\{|\nu_n(A) - \nu_n(B)|: A, B \in \mathcal{C}, d_P(A, B) < \delta\} > \varepsilon/2\} < \varepsilon/2.$$

Hence,

$$\Pr^*\{\sup\{|(\nu_n - \nu_n')(A) - (\nu - \nu_n')(B)|: A, B \in \mathcal{C}, d_P(A, B) < \delta\} > \varepsilon\} < \varepsilon.$$

Also, by [4], Theorem 1.2a, $\mathcal{C}$ is totally bounded for d_P. Take r and $A_1, \cdots, A_r \in \mathcal{C}$ such that for all $C \in \mathcal{C}$, $d_P(C, A_j) < \delta$ for some j. Let

$$\mathfrak{B}_j := \{C \in \mathcal{C}: d_P(C, A_j) < \delta\}$$

and $\mathcal{C}(j) := \mathcal{C}_j := \mathfrak{B}_j \setminus \bigcup_{i<j} \mathfrak{B}_i$. Then $\mathcal{C}_j$ form a decomposition of $\mathcal{C}$ and are Borel sets for d_P. Let $(X, \mathcal{Q}; \mathcal{C}, \mathbb{S})$ be $P\in$-Suslin and $\mathbb{S}_j := \{\mathfrak{S} \cap \mathcal{C}_j: \mathfrak{S} \in \mathbb{S}\}$. Then $(\mathcal{C}_j, \mathbb{S}_j)$ is Suslin (e.g., Parthasarathy, 1967, page 16, Theorems 3.1, 3.4). Hence $(X, \mathcal{Q}; \mathcal{C}_j, \mathbb{S}_j)$ is $P\in$-Suslin.

Now for any $c > 0$, $\|S_n\|_j > c$ iff either

$$\sup_{C \in \mathcal{C}(j)} S_n(C) > t > S_n(A_j) + c$$

or

$$\inf_{C \in \mathcal{C}(j)} S_n(C) < t < S_n(A_j) - c$$

for some rational t. The events $S_n(A_j) < t - c$ and $S_n(A_j) > t + c$ are clearly measurable. The set of $\langle \omega, C \rangle$ such that $S_n(C)(\omega) > t$ is jointly measurable in

$\Omega \times \mathcal{C}$ by the $P\in$-Suslin assumption, so $\sup_{C\in\mathcal{C}(j)} S_n(C)$ is measurable as in [4], proof of Proposition 3.2. Likewise, $\inf_{C\in\mathcal{C}(j)} S_n(C)$ is measurable, hence so is $\|S_n\|_j$. Then by choice of δ, for each j

$$\Pr\left\{\|S_n\|_j > n^{\frac{1}{2}}\varepsilon\right\} < \varepsilon \leqslant \tfrac{1}{80}. \qquad \square$$

3.4 LEMMA. *For each* $\varepsilon > 0$, *each* $j = 1, \cdots, r$ *and* $n \geqslant \max(2\varepsilon^{-2}, N(\varepsilon))$ *with* $N(\varepsilon)$ *and* $\mathcal{C}_j$ *as in* 3.3 *we have* $E\|S_n\|_j^2 \leqslant 17\varepsilon^2 n$.

PROOF. Let $u := n\varepsilon^2 \geqslant 2$. From the measurability of $\|S_n\|_j$, in the last proof, and the form of the D_i, Lemmas 3.1 and 3.2 with $K = 2$ hold for $\| \quad \|_j$ in our case. Thus

$$\begin{aligned} E\|S_n\|_j^2/9 &= \textstyle\int_0^\infty \Pr\left(\|S_n\|_j > 3t^{\frac{1}{2}}\right)dt \\ &\leqslant u + \textstyle\int_u^\infty \Pr\left(\|S_n\|_j > 3t^{\frac{1}{2}}\right)dt \\ &\leqslant u + \textstyle\int_u^\infty 4\Pr\left(\|S_n\|_j > t^{\frac{1}{2}}\right)^2 dt && \text{(by Lemma 3.2)} \\ &\leqslant u + \textstyle\int_u^\infty \Pr\left(\|S_n\|_j > t^{\frac{1}{2}}\right)dt/20 && \text{(by Lemma 3.3)} \\ &\leqslant u + \textstyle\int_0^\infty \Pr\left(\|S_n\|_j > t^{\frac{1}{2}}\right)dt/20, \end{aligned}$$

so

$$E\|S_n\|_j^2 \leqslant 180u/11 < 17n\varepsilon^2. \qquad \square$$

Of the estimates in the next lemma, (3.9) and (3.10) are the most important.

3.5. LEMMA. *For* $\varepsilon > 0$ *and* $N(\varepsilon)$ *as in Lemma* 3.3 *take* $N \geqslant \max(N(\varepsilon), 2\varepsilon^{-2})$. *Let* $T_k := S_{kN}$ *and* $Z_k := T_k - T_{k-1}$ *for* $k = 1, 2 \cdots$ (*where* $T_0 := 0$). *Then for all* $j = 1, \cdots, r$ *and* $k = 1, 2, \cdots$,

(3.6) $$\|Z_k\|_j \leqslant 2N,$$

(3.7) $$E\|Z_k\|_j^2 \leqslant 17N\varepsilon^2,$$

and

(3.8) $$E\|T_k\|_j \leqslant (17\varepsilon^2 kN)^{\frac{1}{2}}.$$

(3.9) *For all* $m \geqslant 1$ *and real* h,

$$E\exp(h\|T_m\|_j) \leqslant \exp\left(hE\|T_m\|_j + 2h^2 e^{4Nh}\textstyle\sum_{k=1}^m E\|Z_k\|_j^2\right).$$

(3.10) *For any* $b > 0$ *and* $\lambda > 0$, *we have*

$$\begin{aligned} \Pr(\|T_m\|_j \geqslant 2\lambda b) \leqslant{}& \\ &\exp\left\{-\lambda^2 + \lambda^2 2^{-1}\left[E\|T_m\|_j/(\lambda b) + \exp(2\lambda N/b)b^{-2}\textstyle\sum_{k=1}^m E\|Z_k\|_j^2\right]\right\} \\ &\leqslant \exp\left\{-\lambda^2 + \lambda^2 2^{-1}\left[(17\varepsilon^2 mN)^{\frac{1}{2}}/(\lambda b) + \exp(2\lambda N/b)b^{-2}17\varepsilon^2 mN\right]\right\}. \end{aligned}$$

PROOF. $\|D_i\|_j \leqslant 2$ for all i and j implies (3.6). Lemma 3.4 implies (3.7) and (3.8).

Let $\mathfrak{B}_k$ be the smallest σ-algebra for which X_i and Y_i are measurable for $1 \leqslant i \leqslant kN$, so that $\mathfrak{B}_0$ is the trivial σ-algebra $\{\phi, \Omega\}$. Let E_k denote conditional expectation with respect to $\mathfrak{B}_k$, so that $E_0X = EX$ whenever $E|X| < \infty$. Let $F_k := (E_k - E_{k-1})\|T_m\|_j$. Then $EF_k = 0$ and

$$\|T_m\|_j = E\|T_m\|_j + \Sigma_{1 \leqslant k \leqslant m} F_k.$$

Thus for any real h,

$$E\exp(h\|T_m\|_j) = E\exp(hE\|T_m\|_j + h\Sigma_{1 \leqslant k \leqslant m}F_k) \tag{3.11}$$
$$= E\big(\exp(hE\|T_m\|_j + h\Sigma_{1 \leqslant k < m}F_k)E_{m-1}\exp(hF_m)\big)$$

since for $k < m$, F_k is $\mathfrak{B}_{m-1}$ measurable. Let $U_k := T_m - Z_k$, $1 \leqslant k \leqslant m$. Then

$$\|U_k\|_j - \|Z_k\|_j \leqslant \|T_m\|_j \leqslant \|U_k\|_j + \|Z_k\|_j,$$

so

$$E_k\|T_m\|_j \leqslant E_k\|U_k\|_j + E_k\|Z_k\|_j,$$
$$E_{k-1}\|T_m\|_j \geqslant E_{k-1}\|U_k\|_j - E_{k-1}\|Z_k\|_j,$$

and

$$F_k \leqslant E_k\|U_k\|_j + E_k\|Z_k\|_j - E_{k-1}\|U_k\|_j + E_{k-1}\|Z_k\|_j.$$

Now $E_k\|Z_k\|_j = \|Z_k\|_j$, $E_{k-1}\|Z_k\|_j = E\|Z_k\|_j$, and $E_k\|U_k\|_j = E_{k-1}\|U_k\|_j$. Thus

$$F_k \leqslant \|Z_k\|_j + E\|Z_k\|_j.$$

Likewise, $F_k \geqslant -\|Z_k\|_j - E\|Z_k\|_j$, and

$$|F_k| \leqslant \|Z_k\|_j + E\|Z_k\|_j. \tag{3.12}$$

Using (3.11) and iterating, (3.9) will follow from

$$E_{k-1}\exp(hF_k) \leqslant \exp\big(2h^2e^{4Nh}E\|Z_k\|_j^2\big), \qquad k = 1, 2, \cdots, m. \tag{3.13}$$

Now $E_{k-1}F_k = 0$. For $r \geqslant 2$, we have by (3.12)

$$E_{k-1}F_k^r \leqslant E_{k-1}\big((\|Z_k\|_j + E\|Z_k\|_j)^r\big) = E\big((\|Z_k\|_j + E\|Z_k\|_j)^r\big)$$
$$\leqslant 2^rE\|Z_k\|_j^r$$

(since for $f := \|Z_k\|_j \geqslant 0$ and $s = 0, 1, \cdots, r$, $Ef \leqslant (Ef^r)^{1/r}$, $Ef^s \leqslant (Ef^r)^{s/r}$, and so $Ef^s(Ef)^{r-s} \leqslant Ef^r$). Thus using (3.6), and $1 + x \leqslant e^x$,

$$E_{k-1}\exp(hF_k) \leqslant 1 + 2h^2E\|Z_k\|_j^2Y$$

where

$$Y := 1 + 2h3^{-1}2N + 2^2h^2(2N)^2/(3 \cdot 4) + \cdots \leqslant e^{4Nh},$$

so

$$E_{k-1}\exp(hF_k) \leqslant 1 + 2h^2E\|Z_k\|_j^2e^{4Nh}$$
$$\leqslant \exp\big(2h^2E\|Z_k\|_j^2e^{4Nh}\big),$$

proving (3.13) and hence (3.9). Then in (3.9) set $h := \lambda/(2b)$ to get the first inequality in (3.10), then apply (3.7) and (3.8) for the second. □

3.14. LEMMA. *For $\mathcal{C}_j$ as in Lemma 3.3 we have*

$$\Pr\left\{\limsup_n \max_{1 \leqslant j \leqslant r} \|S_n\|_j (2nLLn)^{-\frac{1}{2}} \leqslant 26\varepsilon\right\} = 1.$$

PROOF. It is enough to prove the result for each j, and to consider the subsequence $\{T_m\} := \{S_{mN}\}$ since $2N/(2mNLL(mN))^{\frac{1}{2}} \to 0$ as $m \to \infty$. Let $m_k := m(k) := 2^k$. For any $\Lambda > 0$ we have by Lemma 3.1

$$\begin{aligned} P_k &:= \Pr\left\{\max_{m(k-1) < m < m(k)} \|T_m\|_j > \Lambda\varepsilon(2m_{k-1}NLL(m_{k-1}N))^{\frac{1}{2}}\right\} \\ &\leqslant 2\Pr\left\{\|T_{m(k)}\|_j > \Lambda\varepsilon(2^kNLL(2^{k-1}N))^{\frac{1}{2}}\right\} \\ &\leqslant 2\Pr\left\{\|T_{m(k)}\|_j > \Lambda\varepsilon(2^{k+1}NLL(2^kN))^{\frac{1}{2}}/3\right\} \end{aligned}$$

for k large enough so that

$$2LL(2^kN) < 3LL(2^{k-1}N).$$

Now we apply (3.10) with $b := (17mN\varepsilon^2)^{\frac{1}{2}}$ and $\lambda := \Lambda(LL(mN))^{\frac{1}{2}}/18$ to get for $m := m(k)$,

$$P_k \leqslant 2\exp\left\{-\Lambda^2(LL(2^kN))/324 + \Lambda^2\left((LL(2^kN))(648)^{-1}\left[\lambda^{-1} + \exp(2\lambda N/b)\right]\right)\right\}.$$

As $k \to \infty$, $\lambda^{-1} \to 0$ and $2\lambda N/b \to 0$, so eventually

$$P_k < 2\exp\{-\Lambda^2(LL(2^kN)/649)\}. \tag{3.15}$$

Setting $\Lambda := 26$ gives $\Sigma_k P_k < \infty$, so by the Borel-Cantelli lemma, 3.14 follows. □

The next step in the proof will be a "desymmetrization." Let $\nu_n := \nu_n(\omega)$ be the normalized empirical measure $n^{\frac{1}{2}}(P_n - P)$ based on $X_1(\omega), \cdots, X_n(\omega)$. Likewise let $\nu_n' := \nu_n'(\omega')$ be the independent normalized empirical measure $n^{\frac{1}{2}}(P_n' - P)$ where

$$P_n' := n^{-1}\Sigma_{j=1}^n \delta_{Y(j)(\omega')}.$$

Here we may assume Ω is a product space, $\langle\omega, \omega'\rangle \in \Omega$ where ω and ω' are independent for $\Pr = \Pr_1 \times \Pr_2$.

3.16. LEMMA. *If for some K and $\varepsilon > 0$,*

$$\Pr\left\{\limsup_{n\to\infty} \|\nu_n - \nu_n'\|_j (2LLn)^{-\frac{1}{2}} < K\varepsilon\right\} = 1,$$

then

$$\Pr\left\{\limsup_{n\to\infty} \|\nu_n\|_j (2LLn)^{-\frac{1}{2}} < (K+2)\varepsilon\right\} = 1.$$

PROOF. If not, then by Fubini's theorem there exists an ω such that

$$\limsup_{n\to\infty} \|\nu_n\|_j / (2LLn)^{\frac{1}{2}} \geqslant (K+2)\varepsilon$$

and such that

$$\Pr_2\left\{\omega': \limsup_{n\to\infty} \|\nu_n(\omega) - \nu_n'(\omega')\|_j (2LLn)^{-\frac{1}{2}} < K\varepsilon\right\} = 1.$$

Take a subsequence $n(k) \uparrow \infty$ as such that for the given ω, we have

$$\|\nu_{n(k)}\|_j > (K + 1)\varepsilon(2LLn(k))^{\frac{1}{2}}.$$

Then almost surely

$$\liminf_k \|\nu'_{n(k)}\|_j (2LLn(k))^{-\frac{1}{2}} > \varepsilon,$$

so for $k \geqslant M$ large enough,

$$\Pr_2\left\{\|\nu'_{n(k)}\|_j > \varepsilon(2LL(n(k)))^{\frac{1}{2}}\right\} > \tfrac{1}{2}$$

and

$$E\|\nu'_{n(k)}\|_j > \varepsilon(2LL(n(k)))^{\frac{1}{2}}/3.$$

But by Lemma 3.4,

$$\lim_{n\to\infty} E\|S_n\|_j / (2nLLn)^{\frac{1}{2}} = 0.$$

Let E_i denote expectation $\int d\Pr_i$, $i = 1, 2$. Then

$$\begin{aligned} E\|S_n\|_j &= E_2 E_1 \sup_{C\in\mathcal{C}(j)} |S_n(C) - S_n(A_j)| \\ &\geqslant E_2 \sup_{C\in\mathcal{C}(j)} E_1 |S_n(C) - S_n(A_j)| \\ &\geqslant E_2 \sup_{C\in\mathcal{C}(j)} |E_1 S_n(C) - E_1 S_n(A_j)| \\ &= E_2 \sup_{C\in\mathcal{C}(j)} |n(P - P_n')(C) - n(P - P_n')(A_j)| \\ &= E_2 \|n(P - P_n')\|_j = n^{\frac{1}{2}} E\|\nu_n'\|_j. \end{aligned}$$

For $n = n(k) \to \infty$ we get a contradiction. □

Now to finish the proof of Theorem 2.1, recall the set $B_{\mathcal{C}}$ of functions on $\mathcal{C}$ defined by the unit ball B of the Hilbert space H_0. For any finite collection $\mathcal{Q} := \{A_j\}_{1\leqslant j\leqslant r} \subset \mathcal{C}$, the set of restrictions of functions in $B_{\mathcal{C}}$ to $\mathcal{Q}$ is exactly $B_{\mathcal{Q}}$. Recall that $\mathcal{C}$ is totally bounded for d_P by [4], Theorem 1.2a. For $f \in B$ and $C, D \in \mathcal{C}$ we have

$$\begin{aligned} \left|\int f(1_C - 1_D)\,dP\right| &\leqslant \left(\int f^2\,dP\right)^{\frac{1}{2}} d_P(C, D)^{\frac{1}{2}} \\ &\leqslant d_P(C, D)^{\frac{1}{2}}, \end{aligned}$$

and $|\int f 1_C\, dP| \leqslant 1$. Thus $B_{\mathcal{C}}$ is a uniformly bounded, uniformly equicontinuous collection of functions on $(\mathcal{C}, d_P)$. Hence by the Arzelà-Ascoli theorem, $B_{\mathcal{C}}$ is totally bounded for $\|\cdot\|_{\mathcal{C}}$. If $f_n \to f$ weakly in $B \subset H_0$, then for the corresponding g_n, $g \in B_{\mathcal{C}}$, we have $g_n \to g$ pointwise on $\mathcal{C}$ and hence for $\|\cdot\|_{\mathcal{C}}$. Since B is weakly sequentially compact, $B_{\mathcal{C}}$ is a compact set for $\|\cdot\|_{\mathcal{C}}$.

Given $\varepsilon > 0$, choose r large enough and $\mathcal{C}_j$ small enough, $1 \leqslant j \leqslant r$, in Lemma 3.3 such that $d_P(C, A_j) < \varepsilon^2$ for all $C \in \mathcal{C}_j$. Then $|f(C) - f(A_j)| < \varepsilon$ for all $f \in B_{\mathcal{C}}$. Let $\mathcal{Q} := \{A_j\}_{1 \leqslant j \leqslant r}$. Any finite collection $\mathcal{Q}$ is a Strassen log log class (Finkelstein, 1971, Lemma 2). So, almost surely for n large enough there is an $f \in B_{\mathcal{C}}$ (depending on n and ω) such that $\|I(n, \cdot) - f\|_{\mathcal{Q}} < \varepsilon$. Then by Lemmas 3.14 and 3.16 for $K = 26$ for n large enough there is an $f \in B_{\mathcal{C}}$ such that

$$\|I(n, \cdot) - f\|_{\mathcal{C}} < \max_j(\|I(n, \cdot)\|_j + \|I(n, \cdot) - f\|_{\mathcal{Q}} + \|f\|_j)$$
$$< 28\varepsilon + \varepsilon + \varepsilon = 30\varepsilon.$$

Letting $\varepsilon \downarrow 0$, we see that almost surely $\{I(n,\cdot)\}_{n \geqslant 1}$ is relatively compact for $\|\cdot\|_{\mathcal{C}}$ and its set $K(\omega)$ of cluster points is included in $B_{\mathcal{C}}$. For a fixed countable d_P-dense subset Q of $\mathcal{C}$, almost surely the restrictions of functions in $K(\omega)$ to $\mathcal{Q}$ are exactly $B_{\mathcal{Q}}$ for every finite $\mathcal{Q} \subset Q$ (Finkelstein, 1971, Lemma 2). Let $\mathcal{Q}_m$ be finite and $\mathcal{Q}_m \uparrow Q$. For any $f \in B_{\mathcal{C}}$ there exist $g_m \in K(\omega)$ almost surely such that $g_m = f$ on $\mathcal{Q}_m$ for all m. Then $\{g_m\}$ has a convergent subsequence $g_{m(j)} \to g \in K(\omega)$, and then $g = f$ on $\mathcal{Q}_{m(j)}$ for all j, so $g = f$ on $\mathcal{C}$ by continuity. Taking a countable dense set of $f \in B_{\mathcal{C}}$ gives $K(\omega) = B_{\mathcal{C}}$ a.s., so $\mathcal{C}$ is a Strassen log log class.

Now for $\beta > 0$ and $T := T(\mathcal{C})$ we will show that $E\exp(\beta T^2) < \infty$.

3.17. LEMMA. *For all $n \geqslant 1$, all real h, and $\|\cdot\| := \|\cdot\|_{\mathcal{C}}$*

$$E\exp(h\|S_n\|) \leqslant \exp\big(hE\|S_n\| + 4nh^2\big).$$

PROOF. We follow the proof of (3.9) above, replacing T_m by S_n, N by 1, and $\|\cdot\|_j$ by $\|\cdot\|$, so that

$$F_k := (E_k - E_{k-1})\|S_n\|, \qquad k = 1, \cdots, n.$$

Then $E_{k-1}F_k = 0$ and for $d \geqslant 2$, $E_{k-1}F_k^d \leqslant 2^d$, so

$$E_{k-1}\exp(hF_k) \leqslant 1 + \Sigma_{d \geqslant 2}(2h)^d/d!$$
$$= e^{2h} - 2h \leqslant \exp(4h^2),$$

which takes the place of (3.13). □

Now by Lemma 3.4, and since $\sup_n \max_j E|S_n(A_j)|/n^{\frac{1}{2}} < \infty$, there is a $C < \infty$, where we take $C \geqslant 5$, such that $\sup_n E\|S_n\|/n^{\frac{1}{2}} \leqslant C$. Thus by Lemma 3.17,

$$E\exp(h\|S_n\|) \leqslant \exp\big(Chn^{\frac{1}{2}} + 4nh^2\big).$$

Thus for any u and $x := hn^{\frac{1}{2}} > 0$,

$$\Pr\big(\|S_n\| \geqslant un^{\frac{1}{2}}\big) \leqslant e^{-xu}E\exp\big(x\|S_n\|/n^{\frac{1}{2}}\big)$$
$$\leqslant \exp(-xu + Cx + 4x^2).$$

For $u := 8x + C$ this gives

$$\Pr\big(\|S_n\| \geqslant un^{\frac{1}{2}}\big) \leqslant \exp(-(u - C)^2/16) \tag{3.18}$$
$$\leqslant \exp(-u^2/64) \quad \text{for} \quad u \geqslant 2C.$$

Let $a_n := (2nLLn)^{\frac{1}{2}}$, $M_v := \sup_{n \geqslant v} \|S_n/a_n\|$, $n(k) := 2^k$, and $k(v) := [\log_2 v]$. For any $t > 0$ we have

$$\Pr(M_v \geqslant t) \leqslant \Sigma_{k > k(v)} \Pr(\max_{n(k-1) < n < n(k)} \|S_n/a_n\| \geqslant t). \tag{3.19}$$

By the Lévy inequality (3.1) and since $a_n \uparrow$,

$$\Pr(\max_{n(k-1) < n < n(k)} \|S_n/a_n\| \geqslant t) \leqslant 2 \Pr(\|S_{n(k)}\| \geqslant t a_{n(k-1)}).$$

Thus

$$\begin{aligned} \Pr(M_v \geqslant t) &\leqslant 2\Sigma_{k > k(v)} \Pr(\|S_{n(k)}\| \geqslant t a_{n(k-1)}) \\ &= 2\Sigma_{k > k(v)} \Pr\left\{\|S_{n(k)} 2^{-k/2}\| \geqslant t(LL2^{k-1})^{\frac{1}{2}}\right\}. \end{aligned} \tag{3.20}$$

By (3.18), for $t \geqslant 4C$,

$$\Pr(M_v \geqslant t) \leqslant 2\Sigma_{k > k(v)} \exp(-t^2(LL2^{k-1})/64). \tag{3.21}$$

For $v \geqslant 4$, $k(v) \geqslant 2$, so for $k > k(v)$, $LL2^{k-1} = \log((k-1)\log 2)$. Let $\zeta := (\log \log 2)/64$. Then

$$\Pr(M_v \geqslant t) \leqslant 2 \exp(-t^2\zeta)\Sigma_{k > k(v)}(k-1)^{-t^2/64}.$$

Thus if $t > 4C$, $\Pr(M_v \geqslant t) \to 0$ as $v \to +\infty$. From its definition, then, $M_v < +\infty$ a.s. for all v. For any $\beta < +\infty$,

$$E \exp(\beta M_v^2) = \int_0^\infty \Pr(\exp(\beta M_v^2) > t)dt.$$

Let $\alpha := \exp(8\beta C^2)$. Then

$$\begin{aligned} E \exp(\beta M_v^2) &\leqslant \alpha + \int_\alpha^\infty \Pr(\exp(\beta M_v^2) > t)dt \\ &\leqslant \alpha + 2\int_{4C}^\infty \beta \Pr\{M_v > s\} s \exp(\beta s^2)ds \\ &\leqslant \alpha + 4\beta \int_{4C}^\infty s \exp(\beta s^2 - s^2\zeta)\Sigma_{m \geqslant k(v)} m^{-s^2/64}\, ds. \end{aligned}$$

Let $w := s^2/64$. Then since $C \geqslant 5$, $w > 1$ and

$$\Sigma_{m \geqslant k(v)} m^{-w} \leqslant \int_{k(v)-1}^\infty x^{-w}dx = (k(v)-1)^{1-w}/(w-1).$$

Thus there is a $v = v(\beta)$ large enough so that the above estimates give $E \exp(\beta M_v^2) < \infty$. Since the $\|S_n\|$ for finitely many n are uniformly bounded, and $\exp(\max(f, g)) \leqslant e^f + e^g$, we have $E \exp(\beta M_1^2) < \infty$.

Let E_1 (resp. E_2) denote expectations with respect to $\{X_j\}$ (resp. $\{Y_j\}$). Then for any $\beta < 0$, by Jensen's inequality,

$$\begin{aligned} \infty &> E_1 E_2 \exp(\beta \sup_n \|S_n/a_n\|^2) \\ &\geqslant E_1 \exp(\beta E_2 \sup_n \|S_n/a_n\|^2) \\ &\geqslant E_1 \exp\left(\beta \sup_n (\sup_{C \in \mathcal{C}} |\Sigma_{1 \leqslant j \leqslant n} \delta_{X(j)}(C) - P(C)|/a_n)^2\right) \\ &= E \exp(\beta T^2). \end{aligned}$$

□

Acknowledgment. We thank W. Philipp for pointing out papers [14] and [18] to us.

REFERENCES

[1] CHUNG, K. L. (1949). An estimate concerning the Kolmogoroff limit distribution. *Trans. Amer. Math. Soc.* **67** 36–50.

[2] DONSKER, M. D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **23** 277–281.

[3] DUDLEY, R. M. (1974). Metric entropy of some classes of sets with differentiable boundaries. *J. Approximation Theory* **10** 227–236.

[4] DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probability* **6** 899–929 (correction in *Ann. Probability* **7** 909–911).

[5] FINKELSTEIN, H. (1971). The law of the iterated logarithm for empirical distributions. *Ann. Math. Statist.* **42** 607–615.

[6] HOFFMAN-JØRGENSEN, J. (1974). Sums of independent Banach space valued random variables. *Studia Math.* **52** 159–186.

[7] JAIN, N. C. and MARCUS, M. B. (1975). Integrability of infinite sums of independent vector-valued random variables. *Trans. Amer. Math. Soc.* **212** 1–36.

[8] KAHANE, J.-P. (1968). *Some Random Series of Functions.* D. C. Heath, Lexington, Mass.

[9] KHINCHIN, A. (1924). Über einen Satz der Wahrscheinlichkeitsrechnung. *Fund. Math.* **6** 9–20.

[10] KIEFER, J. (1961). On large deviations of the empiric d.f. of vector chance variables and a law of the iterated logarithm. *Pacific J. Math.* **11** 649–660.

[11] KUELBS, J. (1978). Some exponential moments of sums of independent random variables. *Trans. Amer. Math. Soc.* **240** 145–162.

[12] OLSHEN, R. and SIEGMUND, D. (1971). On the maximum likelihood estimate of cell probabilities. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* **19** 52–56.

[13] PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces.* New York, Academic Press.

[14] PHILIPP, W. (1973). Empirical distribution functions and uniform distribution mod 1. In *Diophantine Approximation and its Applications.* 211–234. Academic Press, New York.

[15] PISIER, G. (1975). Le théorème de la limite centrale et la loi du logarithme itéré dans les espaces de Banach. In *Séminaire Maurey-Schwartz,* 1975–1976, exposés 3–4. École Polytechnique, Centre de Mathématiques, Paris.

[16] RÉVÉSZ, P. (1976). Three theorems of multivariate empirical process. *Lecture Notes in Math.* **566** 106–126.

[17] RICHTER, H. (1974). Das Gesetz vom iterierten Logar-ithmus für empirische Verteilungsfunktionen im $\mathbf{R}^k$. *Manuscripta Math.* **11** 291–303.

[18] SCHMIDT, WOLFGANG M. (1975). Irregularities of distribution IX. *Acta Arith.* **27** 384–396.

[19] SMIRNOV, N. V. (1944). Approximation of the laws of distribution of random variables by empirical data (in Russian). *Uspehi Mat. Nauk.* (Old series). **10** 179–206.

[20] STEELE, J. MICHAEL (1978). Empirical discrepancies and subadditive processes. *Ann Probability* **6** 118–127.

[21] STRASSEN, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrscheinlich-keitstheorie und Verw. Gebiete* **3** 211–226.

[22] STUTE, W. (1977). Convergence rates for the isotrope discrepancy. *Ann. Probability* **5** 707–723.

[23] VAPNIK, V. N. and ČERVONENKIS, A. YA. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Teor. Verojatnost. i Primenen* **16** 264–279, (Transl. in *Theor. Probability Appl.* **16** 264–280.

[24] VAPNIK, V. N. and ČERVONENKIS, A. YA. (1974). *Teoriya Raspoznavaniya Obrazov* (Theory of Pattern Recognition). In Russian. Nauka, Moscow.

MATHEMATICS DEPARTMENT
UNIVERSITY OF WISCONSIN
MADISON, WI 53706

DEPARTMENT OF MATHEMATICS
ROOM 2-245
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139