Z. Wahrscheinlichkeitstheorie verw. Gebiete
62, 509-552 (1983)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Invariance Principles for Sums of Banach Space Valued Random Elements and Empirical Processes*

R.M. Dudley[1] and Walter Philipp[2]

[1] Room 2-245, M.I.T., Cambridge, MA 02139, USA
[2] Department of Mathematics, University of Illinois, Urbana, IL 61801, USA

Summary. Almost sure and probability invariance principles are established for sums of independent not necessarily measurable random elements with values in a not necessarily separable Banach space. It is then shown that empirical processes readily fit into this general framework. Thus we bypass the problems of measurability and topology characteristic for the previous theory of weak convergence of empirical processes.

Contents

1. Introduction

Let F be a distribution function with mean zero and variance one. Then on some probability space there exist a sequence $\{X_j, j\geqq 1\}$ of independent random variables with common distribution function F and a sequence $\{Y_j, j\geqq 1\}$ of independent standard normal random variables such that

$$(1.1)\qquad n^{-1/2}\max_{k\leqq n}\Big|\sum_{j\leqq k} X_j - Y_j\Big| \to 0 \quad \text{in probability}$$

* Both authors were partially supported by NSF grants. This work was done while the second author was visiting the M.I.T. Mathematics Department

(Freedman, 1971, (130), p. 83; Major, 1976). This result improves upon and is conceptually much simpler than Donsker's functional central limit theorem, commonly stated in terms of laws on $C[0,1]$ (see e.g. Billingsley, 1968, p. 68).

There is an additional benefit to be gained when we formulate the corresponding result on empirical distribution functions in the spirit of (1.1). Let $\{x_j, j \geqq 1\}$ be a sequence of independent random variables each having a uniform distribution over $[0,1]$. The empirical distribution function of a sample of size n is defined as

$$F_n(s) = n^{-1} \sum_{j \leqq n} 1\{x_j \leqq s\}, \qquad 0 \leqq s \leqq 1.$$

Let

$$X_j(s) = 1\{x_j \leqq s\} - s, \qquad 0 \leqq s \leqq 1. \tag{1.2}$$

Then $\{X_j, j \geqq 1\}$ is a sequence of independent identically distributed $(D[0,1], \|\cdot\|)$-valued random variables with mean zero and uniformly bounded by 1 and $s \to n(F_n(s) - s)$ is the n-th partial sum of this sequence. (Here $\|\cdot\|$ denotes the supremum norm.) Komlós, Major and Tusnády (1975, Theorem 4) proved that: on some probability space there exist a sequence $\{x_j, j \geqq 1\}$ of independent random variables each having uniform distribution over $[0,1]$ and a sequence $\{Y_j, j \geqq 1\}$ of independent $C[0,1]$-valued random variables, each with the distribution of the standard Brownian bridge, such that for the X_j defined by (1.2)

$$\|\sum_{j \leqq n} X_j - Y_j\| = \mathcal{O}(\log n)^2 \text{ a.s.} \tag{1.3}$$

This improved on the original theorem of Kiefer (1972), the first of this type, which gave an error term $\mathcal{O}(n^{\varepsilon + 1/3})$, $\varepsilon > 0$. Such a result not only improves (it implies for instance the Smirnov-Chung law of the iterated logarithm) and is conceptually simpler than Donsker's theorem on empirical distribution functions, but also avoids the problems of measurability and topology caused by the fact that $(D[0,1], \|\cdot\|)$ is not separable (see e.g. Billingsley, 1968, p. 153).

In this paper we shall use this idea to reformulate and strengthen the results of [12, 14, 15, 43] on empirical processes while removing the measurability conditions in most of them. We do this by proving invariance principles for sums of not necessarily measurable random elements with values in a not necessarily separable Banach space and by showing that empirical processes fit easily into this setup.

Before we state our main results we need to introduce some notation. Let $(A, \mathscr{A}, P)$ be a probability space. To provide an adequate setting for our results we consider $(A^\infty, \mathscr{A}^\infty, P^\infty)$, the countable product of copies of $(A, \mathscr{A}, P)$, with coordinates $\{x_j\}_{j \geqq 1}$ and define $(\Omega, \Sigma, \Pr)$ as the product of $(A^\infty, \mathscr{A}^\infty, P^\infty)$ and a copy of the unit interval with Lebesgue measure. This last makes Ω rich enough.

Let $(S, \|\cdot\|)$ be a Banach space and let h be a mapping: $X \to S$, not necessarily measurable. We call $X_j = h(x_j)$, $j \geqq 1$, a sequence of independent *identically formed* random elements. We also define

$$\Pr^*(E)=\inf\{\Pr(B)\colon B\supset E, B\in\Sigma\}, \qquad E\subset\Omega.$$

For not necessarily measurable real-valued functions g_n on Ω we say $g_n\to 0$ *in probability* iff $\lim_{n\to\infty}\Pr^*\{|g_n|>\varepsilon\}=0$ for every $\varepsilon>0$. We say $g_n\to 0$ *in* L^p iff there is a sequence $\{f_n, n\geqq 1\}$ of measurable functions $f_n\geqq|g_n|$ with $f_n\to 0$ in $L^p(\Omega, \Sigma, \Pr)$.

Theorem 1.1. *Let $\{X_j, j\geqq 1\}$ be a sequence of independent identically formed S-valued random elements $X_j=h(x_j)$, $j\geqq 1$. Suppose that for each $m\geqq 1$ there is a mapping $\Lambda_m\colon S\to S$ with the following properties:*

(1.4) *The linear span $L_m S$ of $\Lambda_m S$ is finite-dimensional, $m\geqq 1$.*

(1.5) *For each $m\geqq 1$ there is an $n_0=n_0(m)$ such that for all $n\geqq n_0$*

$$\Pr^*\{n^{-1/2}\|\sum_{j\leqq n} X_j-\Lambda_m X_j\|\geqq 1/m\}\leqq 1/m.$$

(1.6) *For each $m\geqq 1$ the mapping $\Lambda_m\circ h$ is a measurable function from $(A, \mathscr{A})$ into $L_m S$.*

(1.7) $$E(\Lambda_m X_1)=0, \qquad E\|\Lambda_m X_1\|^2<\infty, \qquad m\geqq 1.$$

Let T be the completion of the linear span of $\bigcup_{m\geqq 1}\Lambda_m S$, so T is a separable Banach space. Then there exists a sequence $\{Y_j, j\geqq 1\}$ of independent identically distributed T-valued Gaussian random variables defined on $(\Omega, \Sigma, \Pr)$ *such that*

(1.8) $$EY_1=0$$

(1.9) $$E(s(Y_1)\,t(Y_1))=\lim_{m\to\infty} E\{s(\Lambda_m X_1)\,t(\Lambda_m X_1)\}, \qquad s, t\in T',$$

where T' is the dual space of T, and as $n\to\infty$

(1.10) $$n^{-1/2}\max_{k\leqq n}\|\sum_{j\leqq k} X_j-Y_j\|\to 0 \qquad \text{in Pr. and in } L^p \text{ for any } p<2.$$

Moreover, if there is a function $F\geqq\|X_1\|$ with $EF^2<\infty$ then the convergence in (1.10) is also in L^2.

Remarks. Condition (1.5) can be viewed as a tightness condition, close to the "flat concentration" condition of A. de Acosta (1970, p. 279) in separable Banach spaces, cf. also Pisier (1975, Théorème 3.1).

In our current applications (Theorem 1.3 and Sects. 6, 7 below) the maps Λ_m will be linear. We allow them to be non-linear in view of some difficulties of linear approximation shown to exist by Enflo (1973). Kuelbs (1976, Lemma 2.1) defines linear Λ_m in case S is separable and $E\|X_1\|^2<\infty$. In any case non-linearity of Λ_m makes little difference in our proofs.

In § 8 below we show that if S is separable in Theorem 1.1, then X_j must actually be measurable for (1.10) to hold (although for applications to empirical processes we do need non-measurable X_j in non-separable spaces S). This suggests that the finite-dimensional measurability assumption (1.6) is not too restrictive.

We also note that (1.10) and an argument given by Pisier (1975, p. III.10) imply the existence of a function $F \geqq \|X_1\|$ with $EF^p < \infty$ for any $p<2$ (proved first, as Pisier states, by Jain, 1976a, Theorem 5.7, at least for symmetric variables) but not for $p=2$ (Jain, 1976b).

We define (as have others) the function L by setting $Lx = \log(x \vee e)$ and $L_2 = LL = L \circ L$. Pisier (1975) showed that in a separable Banach space, the central limit theorem and $E\|X_1\|^2 < \infty$ imply a compact law of the iterated logarithm. Heinkel (1979) proved a refinement, replacing $E\|X_1\|^2$ by $E\|X_1\|^2/L_2\|X_1\|$, partly based on methods of Kuelbs and Zinn (1979) who also (independently, as Heinkel notes) proved the refinement (Goodman, Kuelbs and Zinn, 1981). Our next result further improves these facts in different directions, dropping separability and weakening measurability assumptions and strengthening the conclusion to an almost sure invariance principle.

Theorem 1.2. *If in addition to the hypotheses of Theorem 1.1 there exists a measurable function $F \geqq \|X_1\|$ with $E(F^2/LLF) < \infty$, then the sequence $\{Y_j, j \geqq 1\}$ can be chosen such that (instead of (1.10)) there are measurable functions U_n with*

$$\|\sum_{j \leqq n} X_j - Y_j\| \leqq U_n = o((nLLn)^{1/2}) \text{ a.s.}, \quad n \to \infty. \tag{1.11}$$

Since $\{Y_j, j \geqq 1\}$ is a sequence of independent identically distributed Gaussian random variables with values in a separable Banach space, $\{Y_j, j \geqq 1\}$ satisfies a compact as well as a functional law of the iterated logarithm. Hence it is an immediate consequence of Theorem 1.2 that $\{X_j, j \geqq 1\}$ also satisfies the compact and the functional law of the iterated logarithm. (For details and precise statements of these results in separable spaces see [44, (1.4), (1.5), (1.19), (1.20)].)

The significance of Theorems 1.1 and 1.2 primarily lies in their applications to empirical processes. The empirical measure P_n is defined as

$$P_n(B) = n^{-1} \sum_{j \leqq n} 1\{x_j \in B\}, \quad B \in \mathscr{A} \tag{1.12}$$

and the normalized empirical measure ν_n as

$$\nu_n = n^{1/2}(P_n - P). \tag{1.13}$$

Theorem 1.3. *Let $\mathscr{F} \subset \mathscr{L}^2(A, \mathscr{A}, P)$ be a class of functions such that*

(1.14) *$\mathscr{F}$ is totally bounded in $\mathscr{L}^2$.*

(1.15) *For any $\varepsilon > 0$ there is a $\delta > 0$ and n_0 such that for all $n \geqq n_0$*

$$\Pr^* [\sup\{|\int (f-g)\, d\nu_n| : f, g \in \mathscr{F}, \int (f-g)^2\, dP < \delta^2\} > \varepsilon] < \varepsilon.$$

Then there exists a sequence $\{Y_j, j \geqq 1\}$ of independent identically distributed Gaussian processes, defined on the probability space Ω, indexed by $f \in \mathscr{F}$ and with sample functions of Y_1 almost surely uniformly continuous on $\mathscr{F}$ in the $\mathscr{L}^2$ norm such that

$$EY_1(f) = 0 \quad \text{for all } f \in \mathscr{F}, \tag{1.16}$$

(1.17) $$EY_1(f)\,Y_1(g)=\int fg\,dP-\int f\,dP\cdot\int g\,dP \quad \textit{for all } f,g\in\mathscr{F}$$

and as $n\to\infty$

(1.18) $$n^{-1/2}\max_{k\leqq n}\sup_{f\in\mathscr{F}}\Big|\sum_{j\leqq k} f(x_j)-\int f\,dP-Y_j(f)\Big|\to 0$$

in Pr. *as well as in* L^p *for any* $p<2$.

If in addition, there is a function $F\geqq|f|$ *for all* $f\in\mathscr{F}$ *with* $\int F^2\,dP<\infty$ *then we also have* L^2 *convergence in* (1.18).

If only $\int F^2/LLF\,dP<\infty$ *then the* Y_j *can be chosen such that, instead of* (1.18), *we have with probability* 1 *for some measurable* U_n

(1.19) $$\sup_{f\in\mathscr{F}}\Big|\sum_{j\leqq n} f(x_j)-\int f\,dP-Y_j(f)\Big|\leqq U_n=o((nLLn)^{1/2}).$$

Remark. If $\mathscr{F}=\{1_C\colon C\in\mathscr{C}\}$ for a collection $\mathscr{C}$ of sets, then under a measurability condition on $\mathscr{C}$ Theorem 1.2 of [12] states that the central limit theorem holds for empirical measures with respect to uniform convergence on $\mathscr{C}$ if and only if both (1.14) and (1.15) hold. The corresponding result for a class of functions is [14, Theorem 1.3].

Thus Theorem 1.3 applies to all classes $\mathscr{F}$ or $\mathscr{C}$ previously proved to be "Donsker classes" [12, 14–16]. Those papers defined spaces $D_i(\mathscr{F}, P)$ of bounded functions on $\mathscr{F}$, analogous to the Skorohod space $D[0,1]$, but depending on P. On these spaces D_i special σ-algebras were defined, e.g. generated by coordinate evaluations and balls for the sup norm. The present formulation allows us to work more simply in the space of all bounded functions on $\mathscr{F}$, with no special σ-algebra. We also pass (Theorem 1.2) from a central limit theorem to a law of the iterated logarithm under a sharp moment condition without further measurability assumptions such as those in [43], cf. also Kolčinski (1981b).

Theorem 1.3 follows easily from Theorems 1.1 and 1.2 and Lemma 1.4 below which takes care of the uniform continuity: let $m\geqq 1$ and $\varepsilon=1/m$. Choose δ and n_0 according to (1.15). Let $e_P(f,g)=(\int(f-g)^2\,dP)^{1/2}$. Since by (1.14) $\mathscr{F}$ is totally bounded for e_P there exist $f_k=f_{km}\in\mathscr{F}$, $1\leqq k\leqq N(\delta)$, say, such that for each $f\in\mathscr{F}$ there is a k with

(1.20) $$e_P(f,f_k)<\delta,\ \ k=k(f)\ \text{minimal.}$$

Hence by (1.15) in view of (1.12) and (1.13), for all $n\geqq n_0$

(1.21) $$\Pr{}^*\Big\{n^{-1/2}\sup_{f\in\mathscr{F}}\Big|\sum_{j\leqq n}(f-f_k)(x_j)-\int(f-f_k)\,dP\Big|>1/m\Big\}<1/m,$$

where f_k are chosen according to (1.20).

Now let S be the space of all bounded real-valued functions on $\mathscr{F}$. For $\psi\in S$ let $\|\psi\|=\sup\{|\psi(f)|\colon f\in\mathscr{F}\}$. Then $(S,\|\cdot\|)$ is a Banach space (non-separable for $\mathscr{F}$ infinite). We define the mapping $h\colon A\to S$ by setting $h(x)(f)=f(x)-\int f\,dP$ for each $x\in A$, and the mapping $\Lambda_m\colon S\to S$ by setting $\Lambda_m\psi(f)=\psi(f_k)$ with f_k from (1.20), for each $\psi\in S$. Letting $X_j=h(x_j)$ (as usual), we have

(1.22) $$(\Lambda_m X_j)(f)=(\Lambda_m h(x_j))(f)=f_k(x_j)-\int f_k\,dP,\ \ f\in\mathscr{F}.$$

Obviously, $\dim L_m S = N(\delta) < \infty$. Then (1.5) follows from (1.20) and (1.21) since we can assume without loss of generality that $\delta(\varepsilon)$ decreases as ε decreases. Next, (1.6) and (1.7) clearly hold. Now $(T, \|\cdot\|)$ is the closed linear span in S of the ranges of Λ_m, a separable Banach space. Theorem 1.1 implies that there exist independent identically distributed Gaussian variables $Y_j \in T$ satisfying (1.8), (1.9) and (1.10).

We next state a Lemma to be proved in Sect. 3:

Lemma 1.4. *If* (1.14) *and* (1.15) *hold and* Y_j *are* i.i.d. *variables in* T *satisfying* (1.8), (1.9) *and* (1.10) *then there is a Borel set* $W \subset T$, *consisting of functions uniformly continuous on* $\mathscr{F}$ *for* e_P, *such that* $Y_j \in W$ a.s.

Using this Lemma we now obtain Y_j as desired to satisfy (1.16), (1.17) and (1.18). Theorem 1.2 (to be proved in Sect. 5) then implies (1.19), proving Theorem 1.3.

To establish an invariance principle in the form (1.18) for empirical processes (or (1.19), if e.g. $\mathscr{F}$ is uniformly bounded) it is enough to prove (1.14) and (1.15). These two conditions have been established in [12, 14–16, 37, 56b] in various cases. Here we mention only a few of these results and refer the reader to Sect. 7 for a more complete treatment of the subject.

We consider the special case $\mathscr{F} = \{1_C : C \in \mathscr{C}\}$ for a collection $\mathscr{C} \subset \mathscr{A}$ of sets. For each $\delta > 0$ we define $N_I(\delta) = N_I(\delta, \mathscr{C}, P)$ to be the smallest number d of sets $A_1, \ldots, A_d \in \mathscr{A}$ with the following property. For each $C \in \mathscr{C}$ there exist A_r and A_s, $1 \leqq r, s \leqq d$ such that $A_r \subset C \subset A_s$ and $P(A_s \setminus A_r) < \delta$. Recall that $\log N_I(\delta)$ is called a metric entropy with inclusion [12]. Inspection of the proof of Theorem 5.1 of [12] shows that, as Roy Erickson and Joel Zinn pointed out to us,

$$\int_0^1 (\log N_I(x^2))^{1/2}\, dx < \infty \tag{1.23}$$

without any further measurability assumptions implies both (1.14) and (1.15). Hence we have the following result.

Theorem 1.5. *Let* $\mathscr{C} \subset \mathscr{A}$ *be a class of sets for which the entropy condition* (1.23) *holds. Then there exists a sequence* $\{Y_j, j \geqq 1\}$ *of independent identically distributed Gaussian processes, defined on the same probability space, indexed by* $C \in \mathscr{C}$ *and with sample functions of* Y_1 *almost surely uniformly continuous on* $\mathscr{C}$ *in the* d_P*-pseudometric which is defined by*

$$d_P(C, D) = P(C \triangle D), \quad C, D \in \mathscr{A}, \tag{1.24}$$

where $\{Y_j, j \geqq 1\}$ *has the following properties:*

$$EY_1(C) = 0 \quad \text{for all } C \in \mathscr{C}, \tag{1.25}$$

$$E\{Y_1(C)\, Y_1(D)\} = P(C \cap D) - P(C)\, P(D) \tag{1.26}$$

for all $C, D \in \mathscr{C}$, *and as* $n \to \infty$

$$n^{-1/2} \max_{k \leqq n} \sup_{C \in \mathscr{C}} \Big| \sum_{j \leqq k} 1\{x_j \in C\} - P(C) - Y_j(C) \Big| \to 0 \tag{1.27}$$

in probability as well as in L^2. *Or,* Y_j *can be chosen to satisfy, instead of* (1.27), *as* $n \to \infty$

$$\sup_{C \in \mathscr{C}} \Big| \sum_{j \le n} 1\{x_j \in C\} - P(C) - Y_j(C) \Big| = o((nL_2 n)^{1/2}) \quad \text{a.s.} \tag{1.28}$$

I.S. Borisov (1981) has shown that the sufficient condition (1.23) on N_I cannot be weakened, being necessary in case $\mathscr{C}$ is the collection of all subsets of a countable set X, where it is equivalent to $\sum_{x \in X} P\{x\}^{1/2} < \infty$ (cf. also [16]).

A collection $\mathscr{C} \subset \mathscr{A}$ is called a Vapnik-Červonenkis class if for some $n < \infty$, no set F with n elements has all its subsets of the form $C \cap F$, $C \in \mathscr{C}$. The Vapnik-Červonenkis number $V(\mathscr{C})$ denotes the smallest such n.

Note. $V(\mathscr{C})$ does not depend on P.

For Vapnik-Červonenkis classes $\mathscr{C}$ satisfying some measurability conditions - which to a certain extent are also necessary [16] - it is shown [12, Theorem 7.1, Correction], [56a], that (1.14) and (1.15) hold for $\mathscr{F} = \{1_B : B \in \mathscr{C}\}$, and hence all the conclusions of Theorem 1.5 will hold. Conversely, if Y_1 exists as in Theorem 1.5 for all P on $\mathscr{A}$, then $V(\mathscr{C}) < \infty$ [16].

Bounds on the growth of $\dim L_m S$ as $m \to \infty$ in (1.4) will give, in Sect. 6 below, improvements of the error term in Theorem 1.2. Applying these results in Sect. 7 to empirical processes for classes of sets we obtain sharper error terms in (1.28), first if $\log N_I(x) \le c x^{-\tau}$ for constants $c < \infty$ and $0 \le \tau < 1$. For Vapnik-Červonenkis classes of sets we can improve the error term in (1.28) to $\mathcal{O}(n^{1/2-\lambda})$, for any $\lambda < 1/(2700\, V(\mathscr{C}))$.

Let us recall that a) for any k-dimensional real vector space $\mathscr{V}$ of functions on X, with $k < \infty$, the collection $\mathscr{C}$ of all sets of the form $\{x \in X : f(x) > 0\}$, $f \in \mathscr{V}$, is a Vapnik-Červonenkis class with $V(\mathscr{C}) = k+1$ [12, Theorem 7.2]; b) for any $\mathscr{C}$ with $V(\mathscr{C}) < \infty$ and $m < \infty$, the collection of all sets formed from elements of $\mathscr{C}$ by at most m Boolean operations (unions, intersections, and set differences) is also a Vapnik-Červonenkis class [12, Proposition 7.12]; c) if $V(\mathscr{C}_i) < \infty$, $i = 1, \ldots, m < \infty$, then $V(\bigcup_{i \le m} \mathscr{C}_i) < \infty$. Combining a), b) and c) one obtains a good supply of Vapnik-Červonenkis classes; for example, polyhedra in $\mathbb{R}^k$ with at most m faces. The sets $\{y : y_j \le x_j,\ 1 \le j \le k\}$, $x \in \mathbb{R}^k$, used in defining empirical distribution functions, form a still more special Vapnik-Červonenkis class, with $V(\mathscr{C}) = k+1$ [63, Proposition 2.3].

In Theorem 7.5 we apply Theorem 1.3 to weighted empirical distribution functions, thereby improving the compact law of the iterated logarithm, due to Goodman, Kuelbs and Zinn (1981), to an almost sure invariance principle.

Throughout, for any functions f, g, $f \ll g$ means the same as $f = \mathcal{O}(g)$, i.e. f/g is bounded, as $n \to \infty$ or under whatever is the designated condition.

2. Independent Random Elements

We need to generalize several lemmas on sums of independent random variables to random elements, not necessarily measurable, but independent in an extended sense. Let $(\Omega, \mathscr{S}, P)$ be a probability space. Let $\mathscr{L}^0(\Omega, \mathscr{S}, P)$ be the set of $\mathscr{S}$-measurable functions $f\colon \Omega \to [-\infty, \infty]$. Let $L^0(\Omega, \mathscr{S}, P)$ denote the set of equivalence classes of functions in $\mathscr{L}^0(\Omega, \mathscr{S}, P)$ for the relation of equality P-almost surely. In $L^0(\Omega, \mathscr{S}, P)$ (which, with values $\pm\infty$ allowed, is *not* a vector space) we define a metric by

$$d(f, g) = \inf\{\varepsilon > 0\colon P(|\tan^{-1} f - \tan^{-1} g| > \varepsilon) < \varepsilon\}.$$

Then (L^0, d) is a separable metric space.

For given $\mathscr{J} \subset \mathscr{L}^0$ let

$$j = \text{ess.}\inf \mathscr{J}$$

iff $j \leqq h$ almost surely for all $h \in \mathscr{J}$ and whenever $g \leqq h$ almost surely for all $h \in \mathscr{J}$ then $g \leqq j$ almost surely. For all $\mathscr{J} \subset \mathscr{L}^0$ the ess. inf $\mathscr{J}$ exists and is uniquely determined with probability 1. Indeed, we can choose a sequence $\{j_n, n \geqq 1\}$ in $\mathscr{J}$ with $\int \tan^{-1} j_n \, dP \downarrow \inf\{\int \tan^{-1} j \, dP\colon j \in \mathscr{J}\}$. Then $\min_{k \leqq n} j_k \downarrow j = \text{ess.}\inf \mathscr{J} \in \mathscr{L}^0(\Omega, \mathscr{S}, P)$ (cf. e.g. Vulikh, 1967, pp. 79–79). The next two lemmas are straightforward. The notion of "measurable cover function" f^* was defined by Eames and May (1967), by a different method (for bounded functions) which turns out to be equivalent. See also May (1973).

Lemma 2.1. *For each function* $f\colon \Omega \to [-\infty, \infty]$

$$f^* = \text{ess.}\inf\{j \in \mathscr{L}^0(\Omega, \mathscr{S}, P)\colon j \geqq f\}$$

exists and is $\mathscr{S}$*-measurable. Moreover, we can take* $f^* \geqq f$ *everywhere. Further, for all* $g\colon \Omega \to [-\infty, \infty]$,

$$(f+g)^* \leqq f^* + g^* \text{ a.s. } \textit{if both sums are defined} \text{ a.s.};$$
$$(f-g)^* \geqq f^* - g^* \text{ a.s. } \textit{if both differences are defined} \text{ a.s.}$$

Remark. If $f > -\infty$ and $g > -\infty$ everywhere then also $f^* > -\infty$ and $g^* > -\infty$ everywhere so $f+g$ and $f^* + g^*$ are everywhere defined.

Lemma 2.2. *Let* $(S, \|\cdot\|)$ *be a vector space with a seminorm* $\|\cdot\|$. *Then for all* $X, Y\colon \Omega \to S$,

$$\|X+Y\|^* \leqq (\|X\| + \|Y\|)^* \leqq \|X\|^* + \|Y\|^* \text{ a.s.},$$
$$\|cX\|^* = |c| \, \|X\|^* \text{ a.s. } \textit{for each } c \in \mathbb{R}.$$

Remarks. For many calculations Lemma 2.2 allows us to treat $\|X\|^*$ much as $\|X\|$. Can $\|\cdot\|^*$ be made into an actual seminorm on a space of S-valued functions? Let $(\Omega, \mathscr{S}, P)$ be any complete probability space and $(S, \|\cdot\|)$ any seminormed vector space. Let $\mathscr{I}^\infty(S, P)$ be the set of all bounded functions f from Ω into S. Then using a lifting of the space $\mathscr{L}^\infty(\Omega, \mathscr{S}, P)$ of real-valued functions (A. and C. Ionescu Tulcea, 1961) we can define $\|f\|^*$ so that for all $\omega, f \to \|f\|^*(\omega)$ is actually a seminorm on $\mathscr{I}^\infty(S, P)$. For unbounded functions, however, unless P is atomic one does not have a lifting of $L^p(\Omega, \mathscr{S}, P)$, $1 \leqq p < \infty$ (A. and C. Ionescu Tulcea, 1962, Theorem 7).

Lemma 2.3. *Let $(A_j, \mathscr{A}_j, P_j)$ be any probability spaces. Let $f_j\colon A_j \to [0, \infty]$ be any functions, $j=1, \ldots, n$. Then on the Cartesian product $\prod_{j=1}^{n} (A_j, \mathscr{A}_j, P_j)$ with coordinate functions x_j,*

$$\left(\prod_{j=1}^{n} f_j(x_j)\right)^* = \prod_{j=1}^{n} f_j^*(x_j) \text{ a.s.} \tag{2.0}$$

where we set $0 \cdot \infty = 0$. If $n=2$ and $f_1 \equiv 1$ then the same holds for any $f_2\colon A_2 \to [-\infty, \infty]$.

Proof. Clearly $\left(\prod_{j=1}^{n} f_j\right)^* \leqq \prod_{j=1}^{n} f_j^*$. For the converse, by induction, we can take $n=2$. Suppose g is measurable on $A_1 \times A_2$ and for $\varepsilon \geqq 0$ let

$$C(\varepsilon) := \{(x, y)\colon g(x, y) + \varepsilon < f_1^*(x) f_2^*(y)\}.$$

Suppose $(P_1 \times P_2)(C(0)) > 0$. Then for some $\varepsilon > 0$, $(P_1 \times P_2)(C(\varepsilon)) > 0$. Fix such an ε. For $m = 1, 2, \ldots$, let $B_m = \{y\colon m < f_2^*(y) < +\infty\}$. Then for some m, $(P_1 \times P_2)(C(\varepsilon) \setminus (A_1 \times B_m)) > 0$. Fix such an m and let $D := C(\varepsilon) \setminus (A_1 \times B_m)$, $D_x := \{y\colon \langle x, y \rangle \in D\}$, and $H := \{x\colon P_2(D_x) > 0\}$. Suppose $f_1(x) f_2(y) \leqq g(x, y)$ everywhere. Let $x \in H$. If $f_1(x) = +\infty$, then $f_2 \geqq 0$ and for P_2-almost all $y \in D_x$, $f_1(x) f_2(y) < f_1^*(x) f_2^*(y)$, so $f_2(y) = 0 = f_2^*(y)$, a contradiction. If $0 < f_1(x) < \infty$, then for P_2-almost all $y \in D_x$, $f_2^*(y) \leqq g(x, y)/f_1(x)$, so

$$f_2^*(y) < (f_1^*(x) f_2^*(y) - \varepsilon)/f_1(x).$$

Then $f_2^*(y) < +\infty$, so $f_2^*(y) \leqq m$. If $f_2^*(y) \leqq 0$, we get a contradiction since $f_1^*(x) \geqq f_1(x) > 0$. So for any such y, $0 < f_2^*(y) \leqq m$ and

$$f_1(x) < f_1^*(x) - \varepsilon/m.$$

If $f_1 \equiv 1$ this is a contradiction and finishes the proof in that case. In the case $f_j \geqq 0$, $j = 1, 2$, we have

$$f_1(x) \leqq \max(0, f_1^*(x) - \varepsilon/m)$$

for all $x \in H$. If $f_1^* > 0$ on some subset J of H with $P_1(J) > 0$ this allows f_1^* to be chosen smaller, a contradiction. So $f_1 = f_1^* = 0$ a.e. on H, but then $0 \leqq g < 0$ on D, again a contradiction. Q.E.D.

Lemma 2.4. *Let*

$$(\Omega, \mathscr{S}, P) = (\Omega_1 \times \Omega_2 \times \Omega_3, \mathscr{S}_1 \times \mathscr{S}_2 \times \mathscr{S}_3, P_1 \times P_2 \times P_3)$$

and denote the projections $\pi_i\colon \Omega \to \Omega_i$ $(i = 1, 2, 3)$. Then for any bounded nonnegative function f

$$E\{f^*(\omega_1, \omega_3) \mid (\pi_1, \pi_2)^{-1}(\mathscr{S}_1 \times \mathscr{S}_2)\} = E\{f^*(\omega_1, \omega_3) \mid \pi_1^{-1}(\mathscr{S}_1)\} \text{ a.s. } (P).$$

Proof. By Lemma 2.3 (with $f_2(\omega_2)=1$) $f^*(\omega_1,\omega_3)$ equals P-almost surely a measurable function not depending on ω_2 thus is independent of $\pi_2^{-1}(\mathscr{S}_2)$. □

Recall the definitions of convergence in probability or in L^p for non-measurable functions just above Theorem 1.1.

Lemma 2.5. *Let $X:\Omega\to\mathbb{R}$. Then, for all $t\in\mathbb{R}$ and $\varepsilon>0$*

$$P^*(X\geqq t)\leqq P(X^*\geqq t)\leqq P^*(X\geqq t-\varepsilon).$$

For any $X_n:\Omega\to\mathbb{R}$, $X_n\to 0$ in probability or in L^p if and only if $|X_n|^*\to 0$ in probability or in L^p, respectively.

Proof. Since $\{X\geqq t\}\subset\{X^*\geqq t\}$ which is measurable the first inequality follows. To prove the second inequality consider the sets $C_j=\{\omega: X\geqq j\varepsilon\}$, $j\in\mathbb{Z}$. Let $D_j\supset C_j$ be a measurable cover of C_j, i.e. $P^*(C_j)=P(D_j)$. Without loss of generality we can assume that the sequence $\{D_j\}$ is non-increasing. We have $\bigcup_j D_j=\bigcup_j C_j=\Omega$ since $X(\omega)>-\infty$ for all ω. Let

$$\begin{aligned} Y(\omega)&:=(j+1)\,\varepsilon \quad \text{on } D_j\setminus D_{j+1} \quad j\in\mathbb{Z}\\ &:=+\infty \quad \text{on } \bigcap_j D_j. \end{aligned}$$

We claim that for all ω

$$X^*(\omega)\leqq Y(\omega). \tag{2.1}$$

Here Y is measurable. Where $Y(\omega)=+\infty$ the result is clear. Otherwise, $\omega\in D_j\setminus D_{j+1}$ for some j and so $\omega\notin C_{j+1}$. Thus $X(\omega)<(j+1)\,\varepsilon=Y(\omega)$; this proves (2.1).

Given $t\in\mathbb{R}$ there is a unique $j\in\mathbb{Z}$ such that $j\varepsilon\leqq t<(j+1)\,\varepsilon$. Then

$$P(X^*\geqq t)\leqq P(X^*\geqq j\,\varepsilon)\leqq P(Y\geqq j\,\varepsilon).$$

But $\{Y\geqq j\,\varepsilon\}=D_{j-1}$. Thus

$$P(D_{j-1})=P^*(C_{j-1})=P^*(X\geqq (j-1)\,\varepsilon)\leqq P^*(X\geqq t-2\,\varepsilon).$$

The second sentence of the lemma follows directly. □

As in Sect. 1 let $(A^\infty,\mathscr{A}^\infty,P^\infty)$ be the product of countably many copies of $(A,\mathscr{A},P)$. Then the sequence $\{x_j, j\geqq 1\}$, where x_j is the j-th coordinate of $x\in A^\infty$, is a sequence of independent random variables with common distribution P. We extend the concept of independence by calling $\{X_j, j\geqq 1\}$ a sequence of independent elements, if we can write each $X_j=h_j(x_j)$ for some function h_j with domain X. If h_j is a measurable function from $(A,\mathscr{A})$ into some measurable space this implies the usual notion of independence.

The next lemma is an extension of Lemma 2.1 of Kuelbs (1977). Notice that neither the existence (nor the vanishing) of EX_j is needed.

Lemma 2.6. *Let $\{X_j, 1 \leqq j \leqq n\}$ be an independent sequence where the random elements $X_j = h_j(x_j)$ take values in some vector space $(S, \|\cdot\|)$. Let $S_n := \sum_{j \leqq n} X_j$ and $\tau_n \geqq \sum_{j \leqq n} E\|X_j\|^{*2}$. Suppose that*

$$\|X_j\| \leqq M, \quad 1 \leqq j \leqq n. \tag{2.2}$$

Then for $0 \leqq \gamma \leqq 1/(2M)$

$$E\exp(\gamma\|S_n\|^*) \leqq \exp(3\gamma^2\tau_n + \gamma E\|S_n\|^*). \tag{2.3}$$

It follows that for any $K > 0$

$$\Pr\{\|S_n\|^* \geqq K\} \leqq \exp(3\gamma^2\tau_n - \gamma(K - E\|S_n\|^*)), \quad 0 \leqq \gamma \leqq 1/(2M). \tag{2.4}$$

Remarks. If $K < E\|S_n\|^*$ then the infimum of the right side of (2.4) is attained at $\gamma = 0$. If $0 \leqq K - E\|S_n\|^* \leqq 3\tau_n/M$ then the infimum is attained at $\gamma = (K - E\|S_n\|^*)/(6\tau_n)$ and equals $\exp(-(K - E\|S_n\|^*)^2/(12\tau_n))$. If $K - E\|S_n\|^* > 3\tau_n/M$ then the infimum for $0 \leqq \gamma \leqq 1/(2M)$ is at $1/(2M)$ and equals

$$\exp(3\tau_n/(4M^2) - (K - E\|S_n\|^*)/(2M)) \leqq \exp(-(K - E\|S_n\|^*)/(4M)).$$

Proof. We first observe that (2.2) is equivalent to $\|X_j\|^* \leqq M$, $1 \leqq j \leqq n$. Throughout the proof of Lemma 2.1 of Kuelbs (1977) we replace $\|\cdot\|$ by $\|\cdot\|^*$ and note that $E_{k+1}\|Y_k\|^* = E_k\|Y_k\|^*$ by our Lemma 2.4 with $\omega_1 = (X_1, \ldots, X_{k-1})$, $\omega_2 = X_k$ and $\omega_3 = (X_{k+1}, \ldots, X_n)$. (Trivial correction: in the first equation of (2.9) in [41] replace "$-$" by "$+$".) After obtaining [41], (2.4) with $\|\cdot\|^*$ in place of $\|\cdot\|$ we set $\gamma = \varepsilon/(2b_n)$. Note that given $\gamma > 0$ and $M > 0$ we can choose ε, b_n and $c > 0$ satisfying the three conditions $\gamma = \varepsilon/(2b_n)$, $M \leqq b_n c$ and $\varepsilon c \leqq 1$, if and only if $\gamma \leqq 1/(2M)$. This gives our (2.3). Since for any $K \geqq 0$,

$$\Pr\{\|S_n\|^* \geqq K\} \leqq \exp(-\gamma K)\, E\exp(\gamma\|S_n\|^*)$$

relation (2.4) follows. □

We also need an extension of Ottaviani's inequality.

Lemma 2.7. *Let $\{X_j, 1 \leqq j \leqq n\}$ be an independent sequence where the random elements X_j assume values in a vector space $(S, \|\cdot\|)$. Write $S_n = \sum_{j \leqq n} X_j$ and suppose that*

$$\max_{j \leqq n} \Pr(\|S_n - S_j\|^* > \alpha) = c < 1.$$

Then

$$\Pr(\max_{j \leqq n}\|S_j\|^* > 2\alpha) \leqq (1-c)^{-1} \cdot \Pr(\|S_n\|^* > \alpha).$$

Proof. In Lemma 3.21 of Breiman (1968, p. 45) and its proof we replace $|\cdot|$ by $\|\cdot\|^*$ throughout using Lemmas 2.1 and 2.2, specifically $\|S_j\|^* \leqq \|S_n\|^* + \|S_n - S_j\|^*$. In the step where the independence is to be used we argue as follows.

Let $\omega_1=(x_{j+1},\ldots,x_n)$ and $\omega_2=(x_1,\ldots,x_j)$. Then $F(\omega_1,\omega_2)=S_n-S_j$ only depends on ω_1. By Lemma 2.3 $\|S_n-S_j\|^*$ only depends on ω_1 and thus is independent of the event $\{j^*=j\}$ in the usual sense. The remainder of the proof requires no changes. □

The following lemma is based on an argument of Kahane (1968, p. 16).

Lemma 2.8. *Let $\{X_j, 1\leqq j\leqq n\}$ be an independent sequence where the random elements X_j assume values in a vector space $(S,\|\cdot\|)$. Write $S_n=\sum_{j\leqq n} X_j$ and $S_0=0$. Suppose that for some $K>0$*

$$\Pr(\|S_j-S_i\|^*\geqq K)\leqq \tfrac{1}{2} \quad 0\leqq i<j\leqq n. \tag{2.5}$$

Then for any $t>K$ and $s\geqq 0$

$$\Pr\{\|S_n\|^*\geqq 4t+s\}\leqq 4(\Pr\{\|S_n\|^*\geqq t\})^2+\Pr\{\max_{m\leqq n}\|X_m\|^*\geqq s\}.$$

Proof. Let $T(\omega)=\min\{j\geqq 1: \|S_j(\omega)\|^*\geqq 2t\}$. Then by Lemma 2.1

$$\begin{aligned}\Pr\{\|S_n\|^*\geqq 4t+s\} &= \sum_{1\leqq m\leqq n}\Pr\{T=m, \|S_n\|^*\geqq 4t+s\}\\ &\leqq \sum_{1\leqq m\leqq n}\Pr\{T=m, \|S_n\|^*\geqq 4t+s, \|X_m\|^*<s\}\\ &\quad +\Pr\{\max_{m\leqq n}\|X_m\|^*\geqq s\}.\end{aligned} \tag{2.6}$$

By Lemma 2.2, $\|S_n\|^*\leqq \|S_{m-1}\|^*+\|X_m\|^*+\|S_n-S_m\|^*$, so the sum on the right side of (2.6) does not exceed

$$\begin{aligned}&\sum_{1\leqq m\leqq n}\Pr\{T=m, \|S_n-S_m\|^*\geqq 2t\}\\ &\quad=\sum_{1\leqq m\leqq n}\Pr\{T=m\}\cdot\Pr\{\|S_n-S_m\|^*\geqq 2t\}\\ &\quad\leqq \Pr\{\max_{0\leqq m<n}\|S_n-S_m\|^*\geqq 2t\}\cdot\sum_{1\leqq m\leqq n}\Pr\{T=m\}\\ &\quad\leqq 2\Pr\{\|S_n\|^*\geqq t\}\cdot\Pr\{\max_{m\leqq n}\|S_m\|^*\geqq 2t\}\\ &\quad\leqq 4(\Pr\{\|S_n\|^*\geqq t\})^2\end{aligned}$$

using Lemma 2.7 twice, the first time reversing the order of summation. Also note that the independence is used in the same way as in the proof of Lemma 2.7. The lemma follows now from (2.6).

A random element X will be called *symmetric* if $X=h(x)-h(x')$ where $h: A\to S$ and x and x' are independent A-valued random variables on Ω with the same law. The proofs of the next two lemmas follow those in the given references, with $\|\cdot\|^*$ replacing $\|\cdot\|$ and other changes just as in the proofs of Lemma 2.7 and 2.8 above.

Lemma 2.9 (P. Lévy inequality). *Let $\{X_j, 1 \leqq j \leqq n\}$ be a sequence of independent symmetric random elements with values in $(S, \|\cdot\|)$ and partial sums S_k. Then*

$$\Pr(\max_{j \leqq n} \|S_j\|^* > \alpha) \leqq 2 \Pr(\|S_n\|^* > \alpha).$$

Proof. See Kahane (1968, p. 12).

Lemma 2.10. *Let $\{X_j, 1 \leqq j \leqq n\}$ be as above. Then for all $s, t > 0$*

$$\Pr\{\|S_n\|^* \geqq 2t + s\} \leqq 4(\Pr\{\|S_n\|^* \geqq t\})^2 + \Pr\{\max_{m \leqq n} \|X_m\|^* \geqq s\}.$$

Proof. See Kahane [35, p. 16], [26, p. 164], or [31, Lemma 3.4].

Lemma 2.11. *Let S and T be Polish spaces and Q a law on $S \times T$, with marginal μ on S. Let $(\Omega, \mathscr{B}, P)$ be a probability space and X be a random variable on Ω with values in S and law $\mathscr{L}(X) = \mu$. Assume that there is a random variable U on Ω, independent of X, with values in a separable metric space R and law $\mathscr{L}(U)$ on R having no atoms. Then there exists a random variable Y on Ω with values in T and $\mathscr{L}(\langle X, Y \rangle) = Q$.*

Proof. First, we may assume R is complete, hence Polish. Now, any uncountable Polish space is Borel-isomorphic with $[0, 1]$ (e.g. Parthasarathy, 1967, p. 14). Every Polish space is Borel-isomorphic with some compact subset of $[0, 1]$. Thus there is no loss of generality in assuming that $S = T = R = [0, 1]$ with the usual topology, metric and Borel structure.

Next, we take a disintegration of Q on $[0, 1] \times [0, 1]$ (e.g. N. Bourbaki, 1959, Chap. 6, pp. 58–59). Namely, there is a map $s \to Q_s$ from S into the set of all laws on T and such that $\int Q_s \, d\mu(s) = Q$, i.e. for any bounded, Borel measurable function f on $[0, 1] \times [0, 1]$ we have

$$\int f(s, t) \, dQ(s, t) = \int_0^1 \int_0^1 f(s, t) \, dQ_s(t) \, d\mu(s)$$

where all these integrals are defined (possibly for the completion of μ).

For each s let F_s be the distribution function of Q_s and for $0 \leqq t \leqq 1$ let $F_s^{-1}(t) = \inf\{z: F_s(z) \geqq t\}$. We may assume that U has uniform distribution over $[0, 1]$, since if H is the distribution function of U, which has no atoms, then $H(U)$ is uniformly distributed over $[0, 1]$. Now for each t, the map $s \to F_s^{-1}(t)$ is measurable since $F_s^{-1}(t) \leqq \alpha$ iff $F_s(\alpha) \geqq t$, and since the map $s \to F_s(\alpha) = Q_s([0, \alpha])$ is measurable by a property of the disintegration. Since $F_s^{-1}(\cdot)$ is non-decreasing and left-continuous we have

$$F_s^{-1}(t) = \lim_{n \to \infty} \sum_{j=0}^{n} F_s^{-1}(j/n) \, 1\{j/n \leqq t < (j+1) n\}$$

which is jointly measurable in (s, t). Hence

$$Y(\omega) := F_{X(\omega)}^{-1}(U(\omega))$$

is a random variable. Moreover, for any bounded Borel function g on $[0, 1] \times [0, 1]$ we have using Fubini's theorem and the fact that $\lambda \circ (F_s^{-1})^{-1} = Q_s$ (here

λ denotes Lebesgue measure)

$$\begin{aligned}\int g\,dQ &= \int_0^1 \int_0^1 g(s,t)\,dQ_s(t)\,d\mu(s)\\ &= \int_0^1 \int_0^1 g(s, F_s^{-1}(t))\,dt\,d\mu(s)\\ &= \int_0^1 \int_0^1 g(s, F_s^{-1}(t))\,d(\mu\times\lambda)(s,t)\\ &= Eg(X, F_X^{-1}(U)) = Eg(X,Y)\end{aligned}$$

since X and U are independent and thus $\mathscr{L}(\langle X, U\rangle)=\mu\times\lambda$. Consequently $\mathscr{L}(\langle X, Y\rangle)=Q$. □

Note Added in Proof

Lemma 2.11 also follows from the proof of Theorem 1 of Skorohod, Theory Prob. Appl. **21**, 628–632 (1976).

We thank E. Berger for this remark.

For two laws μ and v on a separable metric space (S,ρ) recall the Prohorov distance defined by

$$\pi(\mu, v) := \inf\{\varepsilon>0 : \mu(A) \leqq v(A^\varepsilon)+\varepsilon \text{ for all closed } A\subset S\}$$

where

$$A^\varepsilon = \{x\in S : \inf_{y\in A} \rho(x,y) < \varepsilon\}.$$

The following result is a special case of an extension by Dehling (1982, Prop. 5.1 and Lemma 5.1) of a theorem of Yurinskii (1977, Theorem 1). Where Yurinskii assumed third moments and a Euclidean (Hilbert) norm, Dehling by truncation used $(2+\delta)$th moments, $0<\delta\leqq 1$, and Banach norms (via Lindenstrauss and Tzafriri, 1977, p. 17) as follows:

Lemma 2.12. *Let $\{\xi_j, j\geqq 1\}$ be independent, identically distributed random variables in a d-dimensional Banach space $(B, \|\cdot\|)$, $d<\infty$, with $E\xi_1=0$ and $E\|\xi_1\|^{2+\delta}<\infty$ for a δ with $0<\delta\leqq 1$. Then for the Gaussian law μ with mean 0 and the covariance of ξ_1, and $T_n := \sum_{j\leqq n} \xi_j$, we have*

$$\pi(\mathscr{L}(n^{-1/2}T_n), \mu) \leqq C d^{4/3} n^{-\delta/9} (E\|\xi_1\|^{2+\delta}+1)^{1/4}$$

where C is an absolute constant.

Remarks. If in the proof of Dehling (1982, Prop. 5.1) we use the truncation $Y_i = \xi_i \cdot 1\{\|\xi_i\| \leqq n^{3/(2\delta+6)}\}$ then we can replace $n^{-\delta/9}$ by $n^{\varepsilon-\delta/(2\delta+6)}$ for any $\varepsilon>0$ or, apparently, for $\varepsilon=0$ (Senatov, 1980). Some reduction in the exponent 4/3 is also easy. But in our applications in Theorems 6.1 and 6.2 below this would lead to no major improvement.

We will use the following fact, which is essentially [4, Lemma A1], and which is also a special case of a generalized Vorob'ev theorem (Shortt, 1982, Theorem 2.6; Vorob'ev, 1962).

Lemma 2.13. *Let X, Y and Z be Polish spaces. Suppose μ is a law on $X\times Y$ and v a law on $Y\times Z$ such that μ and v have the same marginal on Y. Then there is a law on $X\times Y\times Z$ with marginals μ on $X\times Y$ and v on $Y\times Z$.*

3. Proof of Theorem 1.1 and Lemma 1.4

To prove Theorem 1.1 we first show the existence of the desired Gaussian limit distribution. Let $k, m, r \geqq 1$ and let Λ_k, Λ_m, Λ_r be the corresponding maps as given in Theorem 1.1. We consider the sequence of independent identically distributed random vectors $\{(\Lambda_k X_j, \Lambda_m X_j, \Lambda_r X_j), j \geqq 1\}$. Let $0<\varepsilon<1/2$. Applying (1.5) twice we obtain for fixed $k, m \geqq 6/\varepsilon$ and all $n \geqq n_0(k) \vee n_0(m)$

$$\Pr\{n^{-1/2} \| \sum_{j \leqq n} (\Lambda_k X_j - \Lambda_m X_j)\| > \varepsilon/2\} < \varepsilon/2. \tag{3.1}$$

We write

$$U_{nkmr} = n^{-1/2} \sum_{j \leqq n} (\Lambda_k X_j, \Lambda_m X_j, \Lambda_r X_j). \tag{3.2}$$

On the finite-dimensional space $L_k S \times L_m S \times L_r S$ we have the sum norm $\|(u, v, w)\| = \|u\| + \|v\| + \|w\|$. By the central limit theorem in this space there is a mean zero Gaussian measure μ_{kmr} such that the Prohorov distance π for the sum norm satisfies

$$\pi(\mathscr{L}(U_{nkmr}), \mu_{kmr}) < \varepsilon/2, \qquad n \geqq n_1(\varepsilon, k, m, r). \tag{3.3}$$

We denote the marginals of the Gaussian law μ_{kmr} by μ_{km}, μ_{kr}, μ_{mr}, μ_k, μ_m and μ_r correspondingly. (Letting $n \to \infty$ in (3.2) shows that the notation is consistent, e.g. μ_6 and $\mu_{4,7}$ are well-defined.) Now the measures μ_k, μ_{km} and μ_{kmr} can be regarded as Borel probability measures on the separable Banach space T, $T \times T$ or $T \times T \times T$ respectively (rather than just on finite-dimensional subspaces) for $\|\cdot\|$ on T and corresponding norms on $T \times T$, $T \times T \times T$. Then (3.1) implies

$$\mu_{mr}\{(v, w) \in T \times T \colon \|v - w\| > \varepsilon\} < \varepsilon, \qquad m, r \geqq 6/\varepsilon. \tag{3.4}$$

On $T \times T$ we take the norm $\|(u, v)\| = \|u\| + \|v\|$. We rewrite (3.4) in the form

$$\mu_{kmr}\{(u, v, w) \colon \|(u, v) - (u, w)\| > \varepsilon\} < \varepsilon, \qquad m, r \geqq 6/\varepsilon, \ k \geqq 1,$$

and obtain for the corresponding Prohorov distance π on $T \times T$

$$\pi(\mu_{km}, \mu_{kr}) \leqq \varepsilon, \qquad m, r \geqq 6/\varepsilon, \ k \geqq 1. \tag{3.5}$$

Consequently, for each $k \geqq 1$, $\{\mu_{km}, m \geqq 1\}$ is a Cauchy sequence for the Prohorov metric. Since $T \times T$ is complete we conclude by Theorem 1.11 of Prohorov (1956) that for each $k \geqq 1$ there is a law $\mu_{k\infty}$ on $T \times T$ such that

$$\mu_{km} \to \mu_{k\infty} \qquad \text{as } m \to \infty. \tag{3.6}$$

Thus by (3.4)

$$\mu_{k\infty}\{(u, v) \colon \|u - v\| > \varepsilon\} \leqq \varepsilon, \qquad k \geqq 6/\varepsilon \tag{3.7}$$

and so for some law μ_∞ on T,

$$\mu_k \to \mu_\infty \qquad \text{as } k \to \infty. \tag{3.8}$$

Since μ_{km} has marginals μ_k and μ_m we conclude from (3.6) and (3.8) that $\mu_{k\infty}$ is Gaussian with marginals μ_k and μ_∞.

We now start with the construction of the Gaussian variables Y_j. For $k \geqq 1$ fixed for the time being, let $\{Z_{kj}, Z_j), j \geqq 1\}$ be a sequence of independent identically distributed random vectors on some probability space Ω' with values in $T \times T$ and

(3.9) $$\mathscr{L}(Z_{kj}, Z_j) = \mu_{k\infty}, \quad j \geqq 1.$$

Since $\mu_{k\infty}$ is centered Gaussian we conclude from (3.7) and (3.9) that for all $n \geqq 1$

(3.10) $$\Pr\{n^{-1/2} \|\sum_{j \leqq n} (Z_{kj} - Z_j)\| > \varepsilon\} \leqq \varepsilon, \quad k \geqq 6/\varepsilon.$$

By Lévy's inequality (Lemma 2.9 above) we obtain for $n \geqq 1$

(3.11) $$\Pr\{n^{-1/2} \max_{m \leqq n} \|\sum_{j \leqq m} (Z_{kj} - Z_j)\| > \varepsilon\} \leqq 2\varepsilon.$$

We now let $k > 6/\varepsilon$. By (3.3) $\{A_k X_j, j \geqq 1\}$ satisfies the central limit theorem with limit μ_k. Next by [53, with correction] there exists on some probability space Ω'' a sequence $\{V_{kj}, j \geqq 1\}$ of independent random variables having the same distribution as $\{A_k X_j, j \geqq 1\}$ and a sequence $\{W_{kj}, j \geqq 1\}$ of independent identically distributed random variables with common distribution μ_k such that

(3.12) $$n^{-1/2} \max_{m \leqq n} \|\sum_{j \leqq m} (V_{kj} - W_{kj})\| \to 0 \quad \text{in prob.}$$

By Lemma 2.13 we may assume that $W_{kj} = Z_{kj}$, $\Omega' = \Omega''$. This together with (3.11) implies that for some $n_2(\varepsilon, k) \geqq n_0(k)$

(3.13) $$\Pr\{n^{-1/2} \max_{m \leqq n} \|\sum_{j \leqq m} (V_{kj} - Z_j)\| > 3\varepsilon\} < 3\varepsilon, \quad n \geqq n_2(\varepsilon, k).$$

In view of (3.13) and (1.5) it might appear that we are done because

(3.14) $$\mathscr{L}(\{V_{kj}, j \geqq 1\}) = \mathscr{L}(\{A_k X_j, j \geqq 1\}).$$

But there are two more hurdles to overcome. First, the sequence $\{Z_j, j \geqq 1\}$ depends on k. We use an idea of Major, also applied in [53] in the same context to construct a universal sequence $\{Z_j, j \geqq 1\}$. We choose $\varepsilon := \varepsilon_p := 2^{-p-3}$, $p = 1, 2, \ldots$ and accordingly $k = k(p) = 2^{p+6} > 6/\varepsilon_p + 1$. In view of (3.13) we obtain for each $p = 1, 2, \ldots$ two sequences $\{V_j^{(p)}, j \geqq 1\}$ and $\{Z_j^{(p)}, j \geqq 1\}$ with the following properties:

(3.15) $$V_j^{(p)} = V_{k(p)j}, \quad j \geqq 1,$$
$$\mathscr{L}(\{Z_j^{(p)}, j \geqq 1\}) = \mathscr{L}(\{Z_j, j \geqq 1\}) \quad (\text{i.i.d. } \mu_\infty)$$

and for some $n_3(p)$, which we choose to satisfy $n_3(p) \geqq n_2(2^{-p-6}, k(p))$, and all $n \geqq n_3$

(3.16) $$\Pr\{n^{-1/2} \max_{m \leqq n} \|\sum_{j \leqq m} V_j^{(p)} - Z_j^{(p)}\| > 2^{-p}\} < 2^{-p}.$$

Moreover, we can assume without loss of generality that the V-sequences are independent of one another and that the Z-sequences are independent of one another. Put $r(p)=\sum_{q\leqq p} n_3(q)$. We define

$$(3.17)\qquad V_j=V_j^{(p)} \quad\text{and}\quad Z'_j=Z_j^{(p)} \quad \text{if } r(p)<j\leqq r(p+1).$$

Then $\{V_j, j\geqq 1\}$ and $\{Z'_j, j\geqq 1\}$ are sequences of independent random variables. Moreover, it will be shown that for each $\varepsilon>0$ there is an $n_4(\varepsilon)$ such that

$$(3.18)\qquad \Pr\{n^{-1/2}\max_{m\leqq n}\|\sum_{j\leqq m}(V_j-Z'_j)\|>4\varepsilon\}<4\varepsilon, \qquad n\geqq n_4.$$

Indeed, let s be such that $2^{-s}<\varepsilon$ and let $N_0=N_0(\varepsilon)$ be so large that for all $n\geqq N_0$

$$\Pr\{n^{-1/2}\max_{m\leqq r(s)}\|\sum_{j\leqq m} V_j\|>\varepsilon\}<\varepsilon$$

and

$$\Pr\{n^{-1/2}\max_{m\leqq r(s)}\|\sum_{j\leqq m} Z'_j\|>\varepsilon\}<\varepsilon.$$

Let $n\geqq\max(N_0, n_3(s))=n_4(\varepsilon)$ and keep it fixed. We choose M such that $r(M)<n\leqq r(M+1)$. Then $n\geqq n_3(p)$ for $p\leqq M$ and by (3.16), (3.17) and stationarity

$$\begin{aligned}\max_{m\leqq n}\|\sum_{j\leqq m}(V_j-Z'_j)\| &\leqq \max_{m\leqq r(s)}\|\sum_{j\leqq m} V_j\|+\max_{m\leqq r(s)}\|\sum_{j\leqq m} Z'_j\|\\ &\quad+\sum_{p=s}^{M-1}\max_{r(p)<m\leqq r(p+1)}\left\|\sum_{j=r(p)+1}^{m}(V_j-Z'_j)\right\|\\ &\quad+\max_{r(M)<m\leqq n}\left\|\sum_{j=r(M)+1}^{m}(V_j-Z'_j)\right\|\\ &\leqq 2\varepsilon n^{1/2}+\sum_{p=s}^{M}2^{-p}\cdot n^{1/2}<4\varepsilon n^{1/2}\end{aligned}$$

except on a set of probability $<4\varepsilon$. This proves (3.18). Since the sequence $\{Z'_j, j\geqq 1\}$ does not depend on ε we have passed the first hurdle.

The sequences $\{X_j, j\geqq 1\}$ and $\{Z'_j, j\geqq 1\}$ are defined on probability spaces, not necessarily the same. Lemma 2.13, successfully employed earlier in the proof, cannot be applied at this point since the X_j are not necessarily measurable. We apply Lemma 2.11 instead. For $j\geqq 1$ define $p:=p(j)$ by $r(p)<j\leqq r(p+1)$ and let $\rho(j):=2^{p(j)+6}$. By (3.14), (3.15) and (3.17)

$$\mathscr{L}(\{A_{\rho(j)}X_j, j\geqq 1\})=\mathscr{L}(\{V_j, j\geqq 1\}).$$

Hence by Lemma 2.11 with

$$Q=\mathscr{L}(\{V_j, j\geqq 1\}, \{Z'_j, j\geqq 1\}), \qquad X=\{A_{\rho(j)}X_j, j\geqq 1\}$$

defined on the appropriate Banach spaces and U a random variable uniformly distributed over $[0,1]$ we obtain a random variable $\{Y_j, j\geqq 1\}$, say, defined on the original probability space Ω such that $Q=\mathscr{L}(\{A_{\rho(j)}X_j, j\geqq 1\}, \{Y_j, j\geqq 1\})$.

Thus by (3.18)

$$n^{-1/2} \max_{m \leqq n} \| \sum_{j \leqq m} (\Lambda_{\rho(j)} X_j - Y_j) \| \to 0 \quad \text{in Pr. } (n \to \infty). \tag{3.19}$$

For the proof of (1.10) it remains to show that

$$n^{-1/2} \max_{m \leqq n} \| \sum_{j \leqq m} (X_j - \Lambda_{\rho(j)} X_j) \| \to 0 \quad \text{in Pr. } (n \to \infty). \tag{3.20}$$

This follows in the same way as (3.18). Since $n_3(p) \geqq n_2(\varepsilon_p, k(p)) \geqq n_0(2^{p+6})$ for all $p \geqq 1$, we have by (1.5)

$$\Pr^* \{ n^{-1/2} \| \sum_{j \leqq n} (X_j - \Lambda_{k(p)} X_j) \| \geqq 2^{-p-6} \} \leqq 2^{-p-6}, \quad n \geqq n_3(p).$$

Hence by Lemmas 2.5 and 2.7

$$\Pr^* \{ n^{-1/2} \max_{k \leqq n} \| \sum_{j \leqq k} (X_j - \Lambda_{k(p)} X_j) \| > 2^{-p} \} < 2^{-p}, \quad n \geqq n_3(p).$$

In the proof of (3.18) we replace V_j and Z'_j by X_j and $\Lambda_{\rho(j)} X_j$, respectively and obtain for given $\varepsilon > 0$ and for some $n_5(\varepsilon)$ and all $n \geqq n_5(\varepsilon)$

$$\max_{k \leqq n} \| \sum_{j \leqq k} (X_j - \Lambda_{\rho(j)} X_j) \| \leqq 4 \varepsilon n^{1/2} \tag{3.21}$$

except on a set of Pr*-measure $< 4\varepsilon$. This proves (3.20) and thus convergence in pr. in (1.10).

Next we show (1.8) and (1.9). Recall that $\{\Lambda_k X_j, j \geqq 1\}$ satisfies the central limit theorem with limit measure μ_k. Thus μ_k is a mean zero Gaussian measure. Moreover, for each $s \in T'$

$$E\{s^2(\Lambda_k X_j)\} = E\{s^2(Z_{k1})\} \to E\{s^2(Z_1)\} = E\{s^2(Y_1)\}$$

by (3.8) and since $s(Z_{k1})$ and $s(Z_1)$ are Gaussian. But $E\{s(Y_1)\, t(Y_1)\}$ is determined by $E\{s^2(Y_1)\}$.

We now prove the statements about L^p-convergence in (1.10). As in Pisier (1975, Proposition 2.1 and Remarque 2.1), we have

$$\sup_{n \geqq 1} E \| n^{-1/2} S_n \|^{*p} < \infty \quad \text{for each } p < 2. \tag{3.22}$$

We follow the next to last paragraph of Sect. 3 on p. 80 of [53] (replacing λ by $\lambda/2$ on the right in the display) and obtain L^p-convergence for each $p < 2$.

For $p = 2$ we observe that by (1.5) with $m = 10^4$ and Chebyshev's inequality applied to $\sum_{j \leqq n} \Lambda_m X_j$

$$\Pr^* \{ \|S_n\| \geqq \alpha n^{1/2} \} < 10^{-3} \quad \text{for all } n \geqq n_0 \tag{3.23}$$

for some $\alpha > 0$. To prove L^2-convergence in (1.10) we replace (3.22) above by the following lemma.

Lemma 3.1. *Let $\{X_j, j \geqq 1\}$ be a sequence of independent identically formed random elements $X_j = h(x_j)$ with values in a seminormed vector space S and $E\|X_1\|^{*2} < \infty$. Suppose that* (3.23) *holds for some α and n_0. Then there is an $N = N(\alpha, n_0, \mathscr{L}(\|X_1\|^*))$ such that for all $n \geqq N$*

$$E\|S_n\|^{*2} \leqq 500\,\alpha^2 n.$$

Proof. By stationarity and (3.23), for $0 \leqq i \leqq j \leqq n$,

$$\Pr\{\|S_j - S_i\|^* > 2\alpha n^{1/2}\} \leqq \Pr^*\{\|S_j - S_i\| > \alpha n^{1/2}\} < 10^{-3} \tag{3.24}$$

if $j - i \geqq n_0$. But if $j - i < n_0$ then by Markov's inequality

$$\Pr\{\|S_j - S_i\|^* > 2\alpha n^{1/2}\} \leqq (j-i)\,E\|X_1\|^*/(2\alpha n^{1/2}) < 10^{-3}$$

if $n \geqq 10^6 n_0^2 E\|X_1\|^{*2}/\alpha^2$. We apply Lemma 2.8 with $K = 2\alpha n^{1/2}$ and obtain for all $t \geqq 4\alpha^2 n$

$$\begin{aligned}\Pr\{\|S_n\|^* \geqq 10 t^{1/2}\} &\leqq 4\Pr\{\|S_n\|^* \geqq 2t^{1/2}\}^2 + n\Pr\{\|X_1\|^* \geqq 2t^{1/2}\}\\ &\leqq 4\cdot 10^{-3}\Pr\{\|S_n\|^* \geqq 2t^{1/2}\} + n\Pr\{\|X_1\|^* \geqq 2t^{1/2}\}\end{aligned}$$

using (3.24). Thus with $u = 4\alpha^2 n$

$$\begin{aligned}10^{-2}E\|S_n\|^{*2} &= \int_0^\infty \Pr\{\|S_n\|^* \geqq 10t^{1/2}\}\,dt\\ &\leqq u + 4\cdot 10^{-3}\int_u^\infty \Pr\{\|S_n\|^* \geqq 2t^{1/2}\}\,dt\\ &\quad + n\int_u^\infty \Pr\{\|X_1\|^* \geqq 2t^{1/2}\}\,dt\\ &\leqq 4\alpha^2 n + 4\cdot 10^{-3}\cdot \tfrac{1}{4}\cdot E\|S_n\|^{*2} + o(n)\end{aligned}$$

since $E\|X_1\|^{*2} < \infty$. Hence for some N and $n \geqq N$

$$E\|S_n\|^{*2} \leqq n(4\alpha^2 + o(1))(10^{-2} - 10^{-3})^{-1} < 500\,\alpha^2 n. \quad \square \tag{3.25}$$

The proof of Theorem 1.1 is complete.

Proof of Lemma 1.4. Let $G_n := (Y_1 + \ldots + Y_n)/n^{1/2}$. Then $\mathscr{L}(G_n) \equiv \mathscr{L}(Y_1)$ on T and (1.10) implies that $\|G_n - v_n\| \to 0$ in probability. Given $\varepsilon > 0$, take $\delta = \delta(\varepsilon) > 0$ and n_0 from (1.15). Take $n \geqq n_0$ large enough so that

$$\Pr^*\{\|G_n - v_n\| > \varepsilon\} < \varepsilon. \tag{3.26}$$

For $\psi \in S$ let

$$p_\delta(\psi) = \sup\{|\psi(f) - \psi(g)| : f, g \in \mathscr{F}, e_P(f, g) < \delta\}.$$

Then p_δ is a seminorm on S with $p_\delta(\psi) \leqq 2\|\psi\|$ for all $\psi \in S$. Relation (1.15) gives $\Pr^*\{p_\delta(v_n) > \varepsilon\} < \varepsilon$, $n \geqq n_0$. Thus we have by (3.26)

$$\Pr^*\{p_\delta(G_n) > 3\varepsilon\} < 2\varepsilon.$$

But p_δ is continuous, hence measurable on the separable space T where the laws $\mathscr{L}(G_n)=\mathscr{L}(Y_1)$, so $\Pr\{p_\delta(Y_1)>3\varepsilon\}<2\varepsilon$. Let $\gamma(k):=\delta(2^{-k})$, and

$$W_k:=\{\psi\in S\colon p_{\gamma(k)}(\psi)\leqq 3/2^k\}.$$

Then $\Pr\{Y_1\notin W_k\}<2^{1-k}$. Let $W=\bigcup_{j\geqq 1}\bigcap_{k\geqq j} W_k$. Then W is a Borel set in T, consisting of functions uniformly continuous on $\mathscr{F}$, and $\Pr\{Y_1\in W\}=1$ by the Borel-Cantelli Lemma. □

4. A Bounded Law of the Iterated Logarithm

Let $a_n:=a(n):=(2n L_2 n)^{1/2}$ and $S_n:=\sum_{j\leqq n} X_j$.

Theorem 4.1. *Let $\{X_j, j\geqq 1\}$ be a sequence of independent identically formed random elements $X_j=h(x_j)$ with values in a seminormed vector space $(S, \|\cdot\|)$. Suppose that for some $\kappa>0$ and n_0 we have*

$$\Pr{}^*\{\|S_n\|>\kappa n^{1/2}\}<10^{-3} \quad \textit{for all } n\geqq n_0 \tag{4.1}$$

and that

$$E\{(\|X_1\|^2/L_2\|X_1\|)^*\}<\infty. \tag{4.2}$$

Then

$$\limsup_{n\to\infty} \|S_n\|^*/a_n\leqq 2^{28}\kappa \quad a.s.$$

The proof is a minor modification of the proof of Theorem 4.1 of Goodman, Kuelbs and Zinn (1981). We first observe that we can assume without loss of generality $\kappa=1$. To symmetrize the random elements X_j we choose a sequence $\{x_j', j\geqq 1\}$ of independent identically distributed A-valued random variables independent of the sequence $\{x_j, j\geqq 1\}$ with the same law. (Specifically, let x_j' also be coordinates in a product of copies of A.) Let $X_j'=h(x_j')$. As in [23] let

$$I(n):=\{2^n+1, \ldots, 2^{n+1}\}, \quad \alpha(t):=t/L_2 t,$$
$$\beta(t):=tL_2 t, \quad \alpha_n:=\beta^{-1}(2^n), \quad \beta_n:=\alpha^{-1}(2^n)$$

and for $j\in I(n)$ let

$$\begin{aligned} U_j&:=X_j 1\{\|X_j\|^{*2}<\alpha_n\},\\ V_j&:=X_j 1\{\alpha_n\leqq\|X_j\|^{*2}\leqq\beta_n\},\\ W_j&:=X_j 1\{\beta_n<\|X_j\|^{*2}\}\end{aligned} \tag{4.3}$$

and define U_j', V_j', W_j' accordingly replacing X_j by X_j'. We then set

$$u_j=U_j-U_j', \quad v_j=V_j-V_j' \quad \text{and} \quad w_j=W_j-W_j'. \tag{4.4}$$

Then u_j, v_j, w_j are symmetric and $u_j+v_j+w_j=X_j-X_j'$.

Lemma 4.1. *We have*

$$\sum_{n\geq 1} \Pr\{\|\sum_{j\in I(n)} w_j\|^* \geqq 2a(2^n)\} < \infty.$$

Proof. The series in question does not exceed $\sum_{j\geqq 1} \Pr^*\{w_j \neq 0\} < \infty$ in view of (4.2).

Lemma 4.2. *We have*

$$\sum_{n\geqq 1} \Pr\{\|\sum_{j\in I(n)} v_j\|^* \geqq a(2^n)\} < \infty.$$

Proof. As in [23] we set $Z_j = 2^n v_j/a(2^n)$ for $j\in I(n)$. We are to show

$$\sum_{n\geqq 1} \Pr\{\|\sum_{j\in I(n)} Z_j\|^* \geqq 2^n\} < \infty. \tag{4.5}$$

This can be done exactly as in the proof of Lemma 4.2 of [23] by verifying

$$j^{-1}\|Z_j\|^* \to 0 \quad \text{a.s.} \tag{4.6}$$

$$\sum_{n\geqq 1} \Lambda^2(n) < \infty, \quad \Lambda(n) := \sum_{j\in I(n)} 4^{-n} E\|Z_j\|^{*2} \tag{4.7}$$

and

$$\lim_{k\to\infty} E\|k^{-1}\sum_{j\leqq k} Z_j\|^* = 0. \tag{4.8}$$

Now (4.6) is easy [23, (4.11)] and (4.7) can be proved in the same way as [23, (4.12)], with the following changes. We replace $\|\cdot\|$ by $\|\cdot\|^*$ everywhere. Convergence of any random variables $U_n \to 0$, in probability or a.s., is replaced by $\|U_n\|^* \to 0$ in the same sense (cf. Lemma 2.5 above). After [23, (4.14)] replace the next "$\Lambda(n)=$" by "$\Lambda(n)/4 \leqq$". Then in (4.15), (4.16) and (4.19), divide $\Lambda(n)^2$ by 16.

Correct [23, (4.16)] by replacing $T := (Lm - LM)^+/\log 2$ by $T+1$. The additional term multiplied by $+1$ has finite expectation by (4.2) so all is well. Above [23, (4.18)], choose $c>1$. Correct [23, (4.20)] by changing the second "$=$" to "$\leqq$" and in its last line replace "$>$" by "$\geqq$". Note that $\gamma'(t) \leqq L_3 t/L_2 t$, $L_3 t > 1$, stated before (4.20), is actually used just after (4.21).

For the proof of [23, (4.21)] we use the proof of Proposition 2.1 and Remarque 2.1 of Pisier (1975), applying our Lemmas 2.5, 2.9 and 2.2. In the next display after (4.21), replace $\lambda(n)$ (a typo) by $\Lambda(n)/4$, obtaining our (4.7) as desired.

To prove our (4.8) we have to estimate $E\|\sum_{j\leqq k} v_j\|^*$. (For the details see [23, p. 729].) Instead of using Lemma 2.3 of [23] we use once more the proof of Pisier (1975), Proposition 2.1 and Remarque 2.1 to obtain for $N \geqq n_0$

$$\Pr\{\|S_N - S'_N\|^* \geqq 12 N^{1/2} n\} \leqq 1/n^2 \quad \text{for all } n\geqq 1.$$

By (4.4) and stationarity we obtain for all $m \geqq 0$ and $N \geqq n_0$

$$E\left\|\sum_{j=m+1}^{m+N} u_j+v_j+w_j\right\|^* < 9N^{1/2}$$

and thus by symmetry

$$E\left\|\sum_{j=m+1}^{m+N} u_j\right\|^* \leqq 9N^{1/2}, \quad E\left\|\sum_{j=m+1}^{m+N} v_j\right\|^* \leqq 9N^{1/2}. \tag{4.9}$$

We are now ready to apply the proof of Theorem 1 of Kuelbs and Zinn (1979). Fix δ, $0<\delta<1$. For $n \geqq 1$ we set

$$\begin{aligned} \xi_n &= \sum_{j\in I(n)} Z_j 1\{\|Z_j\|^* \leqq \Lambda(n)^{1/4}\, 2^{n+1}\} \\ \eta_n &= \sum_{j\in I(n)} Z_j 1\{\|Z_j\|^* \geqq 2^{n-1}\delta\} \\ \zeta_n &= \sum_{j\in I(n)} Z_j 1\{\Lambda(n)^{1/4}\cdot 2^{n+1} < \|Z_j\|^* < 2^{n-1}\delta\}. \end{aligned} \tag{4.10}$$

We bypass the question whether or not ξ_n, η_n and ζ_n are symmetric by symmetrizing them. Let $\tilde{\xi}_n = \xi_n - \xi'_n$, $\tilde{\eta}_n = \eta_n - \eta'_n$ and $\tilde{\zeta}_n = \zeta_n - \zeta'_n$ where we define ξ'_n, η'_n and ζ'_n as in (4.10) replacing Z_j by an independent copy Z'_j. Then by (4.8) and (4.10)

$$E\|\tilde{\xi}_n+\tilde{\eta}_n+\tilde{\zeta}_n\|^* \leqq 2(E\|\sum_{j\leqq 2^{n+1}} Z_j\|^* + E\|\sum_{j\leqq 2^n} Z_j\|^*) = o(2^n).$$

Thus by symmetry

$$E\|\tilde{\xi}_n\|^* = o(2^n).$$

As in the proof of Theorem 1 of [42] we obtain

$$\Pr\{\|\tilde{\xi}_n+\tilde{\eta}_n+\tilde{\zeta}_n\|^* \geqq 2^{n-1}\} \leqq A_n$$

for some A_n with $\sum A_n < \infty$. By a standard desymmetrization argument and (4.10) we obtain Lemma 4.2.

Lemma 4.3. *We have*

$$\sum_{n\geqq 1} \Pr\{\|\sum_{j\in I(n)} u_j\|^* \geqq 2^{24} a(2^n)\} < \infty.$$

Proof. It is enough to show

$$\sum_{k\geqq 1} \Pr\{\|\sum_{j\leqq 2^k} u_j\|^* \geqq 2^{22} a(2^k)\} < \infty. \tag{4.11}$$

For the proof of (4.11) we apply Proposition 4.3 of Pisier (1975) and its proof to the sequence $\{u_j, j\geqq 1\}$. By (4.9) and (4.3) the hypotheses are satisfied with $\alpha = 9$. We also note that $H = 2^{12}$ in [55, Lemma 4.1].

We now can finish the proof of Theorem 4.1. By Lemmas 4.1, 4.2, 4.3 and (4.3) we obtain

$$\sum_{n\geqq 1} \Pr\{\|\sum_{j\in I(n)} X_j - X_j'\|^* \geqq 2^{25}\, a(2^n)\} < \infty$$

and by stationarity and Lemma 2.9

$$\sum_{n\geqq 1} \Pr\{\max_{k\leqq 2^n} \|S_k - S_k'\|^* \geqq 2^{25}\, a(2^n)\} < \infty.$$

By (4.1) and Lemma 2.7 and a standard desymmetrization argument we finally obtain

$$\sum_{n\geqq 1} \Pr\{\max_{k\leqq 2^n} \|S_k\|^* \geqq 2^{26}\, a(2^n)\} < \infty.$$

Theorem 4.1 follows now from the Borel-Cantelli lemma and the triangle inequality.

5. Proof of Theorem 1.2

We use the notation of Sect. 3. We apply Theorem 4.1 to the sequence $\{Z_{kj}-Z_j, j\geqq 1\}$ of independent identically distributed T-valued random variables. In view of (3.10), for each $k\geqq 10^4$, setting $\kappa := 6/k$ we obtain

$$\limsup_{n\to\infty} a_n^{-1} \|\sum_{j\leqq n} Z_{kj} - Z_j\| \leqq 2^{31}/k \quad \text{a.s.} \tag{5.1}$$

By (3.3) the sequence $\{\Lambda_k X_j, j\geqq 1\}$ satisfies the central limit theorem with limit measure μ_k. Next we infer that there is a sequence $\{V_{kj}, j\geqq 1\}$ of independent identically distributed random variables having the same distribution as $\{\Lambda_k X_j, j\geqq 1\}$ and a sequence $\{W_{kj}, j\geqq 1\}$ of independent identically distributed random variables with common distribution μ_k such that as $n\to\infty$

$$\|\sum_{j\leqq n} V_{kj} - W_{kj}\| = o(a_n) \quad \text{a.s.} \tag{5.2}$$

Existence of such V_{kj}, W_{kj} follows from Corollary 1 of [52]. We take the opportunity to make a few corrections and remarks on [52]. In checking the proof of Corollary 1 [52], the reader can omit the beginning of Sect. 3 through the end of Subsection 3.1 since we can set $\prod_N = \text{identity}$, $\xi_\nu = x_\nu$.

In [52, (3.22)] set $H_k=(t_k, t_{k+1}]$. On p. 179, line 4, p. 172 refers to Hartman and Wintner (1941). Both ξ_{n_k} in (3.2.16) and $\xi^*_{n_k}$ in (3.2.17) should be $\xi^*_{t_{k+1}}$, and in the next display $\mu^3_{n_k}$ should be $\mu^3_{t_{k+1}}$. In (3.2.20) 1/2 should be 1/8 and in the next display "(2" should be "2(". In the proof of Lemma 3.7 [52], last line of display, the exponent on α should be $-1/4$. On p. 285, line 9 replace "Kolmogorov's existence theorem" by "Lemma A1 of" [4]. On the right side of (3.3.6) replace α by 5α. At the beginning of the proof of Lemma 3.8 [52] insert: "We may assume without loss of generality $E\|x_1\|^2 < 1/4$. Redefine $\varepsilon(\nu)$ and $\lambda(\nu)$ in terms of $\|x_\nu\|$ rather than $|\xi_\nu|$." Below (3.4.2) add a factor 2 in the definition of c.

As a matter of fact Lemma 3.8 [52] is redundant as we shall demonstrate below when we prove (5.5). This completes our remarks on [52].

By the argument following relation (3.12) using Lemma 2.13 we can assume $W_{kj} \equiv Z_{kj}$ (on some probability space). Then (5.1) and (5.2) imply

$$\limsup_{n\to\infty} a_n^{-1} \| \sum_{j\leq n} V_{kj} - Z_j \| \leq 2^{31}/k \quad \text{a.s.} \tag{5.3}$$

Again the sequence $\{Z_j, j \geq 1\}$ may depend on k. We again use Major's idea to construct a universal sequence $\{Y_j, j \geq 1\}$. We choose $k = k(p) = 2^{p+6}$ and obtain for each $p = 1, 2, \dots$ two sequences of independent identically distributed random variables, say $\{V_j^{(p)}, j \geq 1\}$, $\{Z_j^{(p)}, j \geq 1\}$ with the following properties:

$$V_j^{(p)} = V_{k(p)j}, j \geq 1, \qquad \mathscr{L}(\{Z_j^{(p)}, j \geq 1\} = \mathscr{L}(\{Z_j, j \geq 1\})$$

and

$$\limsup_{n\to\infty} \| \sum_{j\leq n} V_j^{(p)} - Z_j^{(p)} \| / a_n \leq 2^{-p+25} \quad \text{a.s.} \tag{5.4}$$

Moreover, we can assume without loss of generality that the V-sequences, considered as T^∞-valued random variables, are independent and that the same is true for the Z-sequences. Using (5.4) we apply monotone convergence in order to obtain a sequence $\{s(p), p \geq 1\}$ of integers with $s(p) \uparrow \infty$ such that

$$\Pr\{ \sup_{n\geq s(p)} \| \sum_{j\leq n} V_j^{(p)} - Z_j^{(p)} \| / a_n \geq 2^{-p+26} \} \leq 2^{-p}$$

and thus by the Borel-Cantelli lemma

$$\sup_{n\geq s(p)} \| \sum_{j\leq n} V_j^{(p)} - Z_j^{(p)} \| / a_n \ll 2^{-p} \quad \text{a.s.} \tag{5.5}$$

Next we apply Theorem 4.1 to the sequences $\{X_j - \Lambda_{k(p)} X_j, j \geq 1\}$ and obtain for $p \geq 10$

$$\limsup_{n\to\infty} \| \sum_{j\leq n} X_j - \Lambda_{k(p)} X_j \|^* / a_n \leq 2^{-p+25} \quad \text{a.s.} \tag{5.6}$$

and thus by the above argument, for a possibly larger sequence $s(p) \uparrow \infty$

$$\sup_{n\geq s(p)} \| \sum_{j\leq n} X_j - \Lambda_{k(p)} X_j \|^* / a_n \ll 2^{-p} \text{ a.s.} \tag{5.7}$$

We take $s(p)$ large enough to satisfy both (5.5) and (5.7). We note that for each $p \geq 1$ the sequences $\{\Lambda_{k(p)} X_j, j \geq 1\}$ and $\{V_j^{(p)}, j \geq 1\}$ have the same laws. Thus by Lemma 2.11 we can assume without loss of generality that $V_j^{(p)} = \Lambda_{k(p)} X_j$, for each p, and that $\{Z_j^{(p)}, j \geq 1\}$ having the desired joint distribution with $\{V_j^{(p)}, j \geq 1\}$ are all defined on the original probability space.

Without loss of generality we can take $s(q) \geq n_3(q)$ as in (3.16). Let $r(p) := \sum_{q\leq p} s(q)$, so that $r(p) \geq s(p)$. Thus (5.5) and (5.7) remain valid if $s(p)$ is replaced by $r(p)$. Then define $\rho(j)$ and Y_j as in the proof of (3.19). By the remarks in the preceding paragraph

$$\Lambda_{\rho(j)} X_j = V_j^{(p)}, \quad Y_j = Z_j^{(p)} \quad \text{if } r(p) < j \leqq r(p+1). \tag{5.8}$$

By (3.19) and (3.20) there exists a subsequence $\{p(t), t \geqq 0\}$ such that

$$\Big\| \sum_{j \leqq r(p(t))} X_j - Y_j \Big\|^* = o(r(p(t))^{1/2}) \quad \text{a.s.} \tag{5.9}$$

We finally show that the sequence $\{Y_j, j \geqq 1\}$ has the desired properties. Let n be given and find t such that $r(p(t)) < n \leqq r(p(t+1))$. Next find h such that $r(p(t)+h) < n \leqq r(p(t)+h+1)$. Then by (5.8)

$$\begin{aligned}
\Big\| \sum_{j \leqq n} X_j - Y_j \Big\|^* \leqq & \Big\| \sum_{j \leqq r(p(t))} X_j - Y_j \Big\|^* \\
& + \sum_{p(t) \leqq p < p(t)+h} \Big\| \sum_{r(p) < j \leqq r(p+1)} X_j - \Lambda_{k(p)} X_j \Big\|^* \\
& + \sum_{p(t) \leqq p < p(t)+h} \Big\| \sum_{r(p) < j \leqq r(p+1)} \Lambda_{k(p)} X_j - Y_j \Big\| \\
& + \sup_{n \geqq m > r(p(t)+h)} \Big\| \sum_{j \leqq m} X_j - \Lambda_{k(p(t)+h)} X_j \Big\|^* \\
& + \sup_{n \geqq m > r(p(t)+h)} \Big\| \sum_{j \leqq m} \Lambda_{k(p(t)+h)} X_j - Z_j^{(p(t)+h)} \Big\| \\
& + \Big\| \sum_{j \leqq r(p(t)+h)} X_j - \Lambda_{k(p(t)+h)} X_j \Big\|^* \\
& + \Big\| \sum_{j \leqq r(p(t)+h)} \Lambda_{k(p(t)+h)} X_j - Z_j^{(p(t)+h)} \Big\| \\
\leqq & \Big\| \sum_{j \leqq r(p(t))} X_j - Y_j \Big\|^* \\
& + 2 \sum_{p(t) \leqq p \leqq p(t)+h} \Big\| \sum_{j \leqq r(p)} X_j - \Lambda_{k(p)} X_j \Big\|^* \\
& + 2 \sum_{p(t) \leqq p \leqq p(t)+h} \Big\| \sum_{j \leqq r(p)} Z_j^{(p)} - \Lambda_{k(p)} X_j \Big\| \\
& + \sup_{n \geqq m \geqq r(p(t)+h)} \Big\| \sum_{j \leqq m} X_j - \Lambda_{k(p(t)+h)} X_j \Big\|^* \\
& + \sup_{n \geqq m \geqq r(p(t)+h)} \Big\| \sum_{j \leqq m} Z_j^{(p(t)+h)} - \Lambda_{k(p(t)+h)} X_j \Big\|.
\end{aligned}$$

We divide the inequality by a_n and take the limes superior as $n \to \infty$. By (5.9) the first term on the right side of the last inequality tends to 0. The lim sup of the last two terms is $\ll \limsup 2^{-p(t)-h}$ by (5.7) and (5.5). Similarly the lim sup of the second and third term is

$$\ll \limsup \sum_{p(t) \leqq p \leqq p(t)+h} 2^{-p} \ll \limsup 2^{-p(t)} = 0.$$

Thus

$$\Big\| \sum_{j \leqq n} X_j - Y_j \Big\|^* = o(a_n) \quad \text{a.s.}$$

6. Improvement of the Error Term

In this section we improve the error term in Theorem 1.2 under two additional sets of hypotheses, thereby giving a partial extension of Theorem 1.1 of [47]. This theorem can be completely extended to conform with the underlying theme of the present paper. But we refrain from carrying out this program since we do not have reasonable applications for this more general theorem.

Theorem 6.1. *Let* $\{X_j, j \geqq 1\}$ *be a sequence of independent identically formed random elements with values in* $(S, \|\cdot\|)$. *Suppose that* $E\|X_1\|^{*2+\delta} < \infty$ *for some* $0 < \delta \leqq 1$. *Let* $\{\Lambda_m, m \geqq 1\}$ *be a sequence of mappings as described in Theorem* 1.1 *having the following additional properties.*

(6.1) *The* Λ_m *are linear maps with* $\sup_{m \geqq 1} \|\Lambda_m\| < \infty$.

If in condition (1.5)

$$n_0(m) \leqq C_1 m^D, \quad m \geqq 1 \tag{6.2}$$

and if

$$\dim \Lambda_m S \leqq C_2 \exp(C_3 m^\beta), \quad m \geqq 1 \tag{6.3}$$

for some constants $C_i \geqq 1$ $(i = 1, 2, 3)$, $D > 2$ *and* $\beta > 0$ *then the error term in* (1.11) *can be improved to* $\mathcal{O}(n^{1/2} (\log n)^{-\theta})$ *for any* $\theta < 1/(2\beta)$.

If, instead of (6.3)

$$\dim \Lambda_m S \leqq C_4 m^\gamma \tag{6.4}$$

for some C_4, $\gamma \geqq 1$ *then the error term in* (1.11) *can be improved to* $\mathcal{O}(n^{1/2-\lambda})$ *where* $\lambda = \kappa^2/(600\gamma)$, $\kappa = \min(\delta, 4/(D-2))$.

The proof of Theorem 6.1 follows a by now well-established method [3], [44, Theorem 2], [54], [47, Theorem 1.1], [2], [10, Theorem 2], etc. In all of these papers explicit bounds on the probabilities of errors easily could have been established, by collecting the relevant probability bounds before the Borel-Cantelli lemma is applied. In the case of our Theorem 6.1 the corresponding result is as follows.

Theorem 6.2. *Assume that the hypotheses of Theorem* 6.1 *hold. If* (6.3) *holds then one can choose the sequence* $\{Y_j, j \geqq 1\}$ *such that for any* $\theta < 1/(2\beta)$ *and any constant* $B < \infty$

$$\Pr\{\max_{m \leqq n} \|\sum_{j \leqq m} X_j - Y_j\|^* \geqq n^{1/2} (\log n)^{-\theta}\} \ll (\log n)^{-B}.$$

If however, condition (6.4) *holds then with* $\lambda = \kappa^2/(600\gamma)$ *the sequence* $\{Y_j, j \geqq 1\}$ *can be chosen so that for some constant* $H < \infty$

$$\Pr\{\max_{m \leqq n} \|\sum_{j \leqq m} X_j - Y_j\|^* \geqq H n^{1/2-\lambda}\} \ll n^{-\kappa/28}.$$

Theorem 6.1 is an immediate consequence of Theorem 6.2. If (6.3) holds we let $n=2^{k+1}$ and applying the Borel-Cantelli lemma we obtain

$$\max_{2^k<m\leqq 2^{k+1}} \|\sum_{j\leqq m} X_j - Y_j\| \ll 2^{k/2} k^{-\theta} \quad \text{a.s.}$$

since we can choose $B=2$. Similarly, if (6.4) holds we let $n=k^\rho$ with $\rho=[56/\kappa]$, instead.

Hence it remains to prove Theorem 6.2. Since $\kappa\leqq\delta$, $EF^{2+\kappa}<\infty$. So we may replace δ by κ and assume in the proofs that $\kappa=\delta$, i.e. that $D\leqq 2+4/\delta$ in (6.2). Then, let $\alpha:=6\delta/7$. The hypotheses made so far all hold with α in place of δ: $EF^{2+\alpha}<\infty$ and $D\leqq 2+4/\alpha$. In the conclusion under (6.4), $\lambda:=\delta^2/(600\gamma)<\alpha^2/(400\gamma)$. In Theorem 6.2 we can and will take $B\geqq 1$.

We can assume without loss of generality

$$10^6 C_1^2(1+\sup_{m\geqq 1}\|\Lambda_m\|)\|F\|_{2+\delta}\leqq\alpha. \tag{6.5}$$

Let

$$V_{jm}:=V_j(m):=X_j-\Lambda_m X_j, \qquad m,j\geqq 1$$

and

$$\begin{aligned} V'_{jm}&:=V_{jm}1\{\|X_j\|^*\leqq j^{1/(2+\alpha)}\}:=V'_j(m)\\ V''_{jm}&:=V_{jm}-V'_{jm}=V_{jm}1\{\|X_j\|^*>j^{1/(2+\alpha)}\}.\end{aligned} \tag{6.6}$$

Lemma 6.1. *Let $m\geqq 10^4$. Then for all $n\geqq C_1 m^{2+4/\alpha}$ and $s\geqq 0$*

$$\Pr\left\{\left\|\sum_{j=s+1}^{s+n} V'_{jm}\right\|^*\geqq 2n^{1/2}/m\right\}\leqq 10^{-3}.$$

Proof. We have by Lemma 2.5, (6.5) and (6.6) for all $n\geqq m^{2+4/\alpha}$ and $s\geqq 0$

$$\begin{aligned}
&\Pr\left\{\left\|\sum_{j=s+1}^{s+n} V''_{jm}\right\|^*\geqq n^{1/2}/m\right\}\\
&\quad\leqq \Pr\left\{\sum_{j=s+1}^{s+n}\|V''_{jm}\|^*\geqq n^{1/2}/m\right\}\\
&\quad\leqq n^{-1/2}m\sum_{j=s+1}^{s+n}E\{\|V_{jm}\|^*1\{\|X_j\|^*\geqq j^{1/(2+\alpha)}\}\}\\
&\quad\leqq n^{-1/2}m(1+\|\Lambda_m\|)^{2+\alpha}\cdot E\|X_1\|^{*2+\alpha}\cdot\sum_{j=s+1}^{s+n} j^{-(1+\alpha)/(2+\alpha)}\\
&\qquad<10^{-5}mn^{-\alpha/(4+2\alpha)}\leqq 10^{-5}.
\end{aligned}$$

The lemma follows now from (6.6), (1.5) and the stationarity of V_{jm}.

Lemma 6.2. *Let $m\geqq 10^4$. Then for all $n\geqq m^{14+16/\alpha}$ and $s\geqq 0$*

$$E\left\|\sum_{j=s+1}^{s+n} V'_{jm}\right\|^{*2}\leqq 500\,n/m^2.$$

Proof. Write $S'_n = \sum_{j=1}^{n} V'_{j+s,m}$. Then by Lemma 6.1 we have for all $0 \leqq i < j \leqq n$

$$\Pr\{\|S'_j - S'_i\|^* \geqq 2\, n^{1/2}/m\} \leqq 10^{-3} \tag{6.7}$$

if $j-i \geqq C_1 m^{2+4/\alpha}$. But if $j-i < C_1 m^{2+4/\alpha}$ we have by (6.5), (6.6) and Markov's inequality for $n \geqq m^{6+8/\alpha}$

$$\Pr\{\|S'_j - S'_i\|^* \geqq n^{1/2}/m\} \leqq m\, n^{-1/2}\, C_1 m^{2+4/\alpha}(1+\|A_m\|)\cdot E\,\|X_1\|^* \leqq 10^{-3}.$$

Thus (6.7) holds for all $n \geqq m^{6+8/\alpha}$. We now follow the proof of Lemma 3.1 with S_n replaced by S'_n and $\alpha = 1/m$. Then by the proof of (3.25) we have for all $n \geqq m^{14+16/\alpha}$ $(= m^2 \cdot m^{2(6+8/\alpha)} \geqq m^2 \cdot 10^6\, m^{2(6+8/\alpha)} E\,\|X_1\|^{*2}$ by (6.5) and since $\delta \leqq 1)$

$$E\,\|S'_n\|^{*2} \leqq n \left(4/m^2 + \int_u^\infty \Pr\{\|X_1\|^* \geqq 2\, t^{1/2}\}\, dt\right)(10^{-2} - 10^{-3})^{-1}. \tag{6.8}$$

The result follows now at once since the integral in (6.8) does not exceed $E\,\|X_1\|^{*2+\alpha} u^{-\alpha/2}/\alpha \leqq m^{-2}$.

From now on we assume that (6.3) holds. (The case (6.4) will be treated at the end of this section.) We put $\rho = 1/(1+\beta) < 1$ and take any ζ with $0 < \zeta < \rho$. Then $\beta\zeta < 1-\rho$. Let

$$m_0 := 1,\ m_k := m(k) := k^\zeta, \qquad k \geqq 1, \tag{6.9}$$

$$t_0 := 0,\ t_k := t(k) := [\exp(k^{1-\rho})], \qquad k \geqq 1 \tag{6.10}$$

and

$$n_k := n(k) := t_{k+1} - t_k \sim \text{const} \cdot t_k\, k^{-\rho}. \tag{6.11}$$

Let $H_k := H(k) := \{j\colon\ t_k < j \leqq t_{k+1}\} = \{t_k + 1, \ldots, t_{k+1}\}$.

Proposition 6.1. *For any $B < \infty$ there is a $C = C_{6.1}(B)$ such that as $k \to \infty$*

$$\Pr\{\|\sum_{j \in H(k)} V_j(m_k)\|^* \geqq C(n_k L_2 t_k)^{1/2}/m_k\} \ll k^{-B}.$$

Proof. Let $r_k := r(k) := m_k^{14+16/\alpha}$ and define new random elements

$$U_h := U_h(k) := \sum_{j=1}^{r(k)} V'_{t(k)+r(k)(h-1)+j}(m_k),$$

$$h = 1, 2, \ldots, l_k := l(k) := [n_k/r_k],$$

$$U_{l(k)+1} := \sum_{j=t(k)+l(k)r(k)+1}^{t(k+1)} V'_j(m_k).$$

Then we get using (6.6) and (6.9)

$$\|U_h\| < r_k\, t_{k+1}^{1/(2+\alpha)} := M_k, \qquad 1 \leqq h \leqq l_k + 1. \tag{6.12}$$

For k large, $m_k \geqq 10^4$ and by Lemma 6.2,

$$E\|U_h\|^{*2} \leqq 500\, r_k m_k^{-2}, \quad 1 \leqq h \leqq l_k,$$

and

$$E\|\sum_{h\leqq l(k)} U_h\|^* \leqq 500\, n_k^{1/2}/m_k.$$

We apply now Lemma 2.6 and the Remarks after it to the sequence $\{U_h, h \leqq l_k\}$, with $M := M_k$, $n := l_k$, $\tau := \tau_n := 500 r_k l_k m_k^{-2}$ and $K := K_k := \xi(n_k L_2 t_k)^{1/2}/m_k$ for a constant $\xi > (6000\, B/(1-\rho))^{1/2}$.

Then for k large, $E\|S_n\|^* < K_k < 3\tau_n/M_k$, so

$$\begin{aligned}\Pr\{\|\sum_{h\leqq l(k)} U_h\|^* \geqq K_k\} &\leqq \exp(-(K - E\|S_n\|^*)^2/(12\tau_n))\\ &\leqq \exp(-\xi^2 L_2 t_k/6000) \ll \exp(-B\log k) = k^{-B}.\end{aligned}$$

Since $M_k = o(K_k)$ as $k \to \infty$, the latter bound still holds for $\sum_{h\leqq l(k)+1} U_h$, because of (6.12).

Next,

$$\begin{aligned}(6.13)\qquad &\Pr\{V_j'(m_k) \neq V_j(m_k) \quad \text{for some } j \in H(k)\}\\ &\leqq \sum_{j>t(k)} \Pr\{\|X_1\|^* > j^{1/(2+\alpha)}\}\\ &\leqq E\|X_1\|^{*2+\alpha}\, 1\{\|X_1\|^* > t_k^{1/(2+\alpha)}\}\\ &\leqq E\|X_1\|^{*2+\delta}\, t_k^{(\alpha-\delta)/(2+\alpha)} \leqq t_k^{(\alpha-\delta)/3} = t_k^{-\delta/21} \ll k^{-B}.\end{aligned}$$

This completes the proof.

Next we set

$$\begin{aligned}(6.14)\qquad X_j' &:= X_j\, 1\{\|X_j\|^* \leqq j^{1/(2+\alpha)}\},\\ X_j'' &:= X_j - X_j' = X_j\, 1\{\|X_j\|^* > j^{1/(2+\alpha)}\}.\end{aligned}$$

Lemma 6.3. *There is a constant C_5 such that for all $s \geqq 0$ and $n \geqq 1$*

$$E\left\|\sum_{j=s+1}^{s+n} X_j'\right\|^* \leqq C_5 n^{1/2}.$$

Proof. From (6.6) and Lemma 6.2 with $\mu = 10^4$ we obtain

$$\begin{aligned}E\left\|\sum_{j=s+1}^{s+n} X_j'\right\|^* &\leqq E\left\|\sum_{j=s+1}^{s+n} V_j'(\mu)\right\|^* + E\left\|\sum_{j=s+1}^{s+n} \Lambda_\mu X_j\right\|^*\\ &\quad + \sum_{j=s+1}^{s+n} E\{\|\Lambda_\mu X_j\|\, 1\{\|X_j\|^* > j^{1/(2+\alpha)}\}\\ &\ll n^{1/2} + \sum_{j=s+1}^{s+n} \|\Lambda_\mu\|\, E\|X_j\|^{*2+\alpha} \cdot j^{-(1+\alpha)/(2+\alpha)} \ll n^{1/2}.\end{aligned}$$

Note that we also have used the fact that $\{\Lambda_\mu X_j, j \geqq 1\}$ is a sequence of independent identically distributed random vectors with mean zero and finite second moment (1.7).

Proposition 6.2. *For any $B<\infty$ there is a $C=C_{6.2}(B)$ large enough so that as $k\to\infty$*

$$\Pr\{\max_{n\in H(k)}\|\sum_{t(k)<j\leqq n} X_j\|^*\geqq C(n_k L_2 t_k)^{1/2}\}\ll k^{-B}.$$

Proof. Using (6.13) it suffices to prove this for X_j' in place of X_j. Let $S_n':=\sum_{j\leqq n} X_j'$. By (6.11) and Markov's inequality we have for all k

$$\max_{n\in H(k)} \Pr\{\|S_{t(k+1)}'-S_n'\|^*>2C_5 n_k^{1/2}\}\leqq 1/2.$$

Hence by Lemma 2.7 we have for all k, and $\xi>2\max(C_5, 12B/(1-\rho))^{1/2}$

$$\begin{aligned}(6.16)\qquad &\Pr\{\max_{n\in H(k)}\|S_n'-S_{t(k)}'\|^*\geqq 2\xi(n_k L_2 t_k)^{1/2}\}\\ &\leqq 2\Pr\{\|S_{t(k+1)}'-S_{t(k)}'\|^*\geqq \xi(n_k L_2 t_k)^{1/2}\}.\end{aligned}$$

To estimate this last probability we apply Lemma 2.6 to the sequence $\{X_j', j\in H_k\}$ with $K=\xi(n_k L_2 t_k)^{1/2}$, $n=\tau_n=n_k$ and $M=t_{k+1}^{1/(2+\alpha)}$ and for k large obtain the bound

$\exp(-\xi^2 L_2 t_k/48)\ll k^{-B}$. Q.E.D.

Let $0<\varepsilon<1/4$. For $m\geqq 6/\varepsilon$ let $\{(Z_{mj}, Z_j), j\geqq 1\}$ be any sequence of independent random variables with $\mathscr{L}(Z_{mj}, Z_j)=\mu_{m\infty}, j\geqq 1$ (as defined in (3.6)).

Proposition 6.3. *For any $B<\infty$ there is a $C=C_{6.3}(B)$ large enough so that as $k\to\infty$*

$$\Pr\{\|\sum_{j\in H(k)} Z_{m(k)j}-Z_j\|\geqq C(n_k L_2 t_k)^{1/2}/m_k\}\ll k^{-B}.$$

Proof. By (3.10) we have for each $m\geqq 6/\varepsilon$

$$\Pr\{n^{-1/2}\|\sum_{j\leqq n} Z_{mn}-Z_j\|\geqq\varepsilon\}\leqq\varepsilon<1/4.$$

Hence by the Fernique-Landau-Shepp inequality [19, p. 1699]

$$\Pr\{\sum_{j\leqq n}\|Z_{mj}-Z_j\|\geqq\eta\,\varepsilon\, n^{1/2}\}\leqq\exp(-\eta^2/24)$$

for any $\eta\geqq 1$. Putting $n=n_k$, $\varepsilon=6/m_k$ and $\eta=(24B(1-\rho)^{-1}L_2 t_k)^{1/2}$ we obtain the result.

Next, let $F_k:=\mathscr{L}(n_k^{-1/2}\sum_{j\in H(k)}\Lambda_{m(k)}X_j)$ and

$$G_k:=\mathscr{L}(n_k^{-1/2}\sum_{j\in H(k)} Z_{m(k)j})=\mathscr{L}(Z_{m(k)1})=\mu_{m(k)}.$$

Then by Lemma 2.12, (6.3), (6.9), (6.10) and (6.11) we obtain for the Prohorov distance

$$\pi(F_k, G_k) \ll n_k^{-\alpha/9} \exp(\tfrac{4}{3} C_2 m_k^\beta) \ll n_k^{-\alpha/10}, \tag{6.17}$$

i.e. for some constant C_6,

$$\pi(F_k, G_k) \leqq C_6 n_k^{-\alpha/10}, \qquad k \geqq 0.$$

Let $\mathcal{S}_k := L_{m(k)} S$. Then by a classic theorem of Strassen (1965) there is a law J_k on $\mathcal{S}_k \times \mathcal{S}_k$ with marginals F_k and G_k on $\mathcal{S}_k$ such that

$$J_k \{\langle x, y\rangle : \|x-y\| > C_6 n_k^{-\alpha/10}\} \leqq C_6 n_k^{-\alpha/10}, \qquad k \geqq 0.$$

Thus by Lemma 2.13 applied to $X = \mathcal{S}_k^{n(k)}$, $Y = Z = \mathcal{S}_k$, there are independent $\mathcal{S}_k$-valued random variables X_{kj}, $j = 1, \ldots, n_k$, all with law $\mathcal{L}(X_{kj}) = \mathcal{L}(\Lambda_{m(k)} X_1)$, and a random variable Γ_k with law G_k such that

$$\Pr\left\{\left\|\Gamma_k - n_k^{-1/2} \sum_{j=1}^{n(k)} X_{kj}\right\| > C_6 n_k^{-\alpha/10}\right\} \leqq C_6 n_k^{-\alpha/10}.$$

Now we apply Lemma 2.13 to $X = \mathcal{S}_k^{n(k)}$, $Y = \mathcal{S}_k$, and $Z = (\mathcal{S}_k \times T)^{n(k)}$ to obtain, for each $k \geqq 0$, independent random variables $(\Gamma_{kj}, W_{kj}) \in \mathcal{S}_k \times T$, $j = 1, \ldots, n_k$, each with law $\mu_{m(k)\infty}$ (as in (3.6)), so that each Γ_{kj} has the Gaussian law $G_k = \mu_{m(k))}$ and such that

$$\Pr\left\{n_k^{-1/2} \left\|\sum_{j=1}^{n(k)} \Gamma_{kj} - X_{kj}\right\| > C_6 n_k^{-\alpha/10}\right\} \leqq C_6 n_k^{-\alpha/10}. \tag{6.18}$$

(The above two applications of Lemma 2.13 could be replaced by one application of the generalized Vorob'ev theorem: Shortt, 1982, Theorem 2.6.)

Now we take the countable product of the laws

$$\mathcal{L}\{\langle X_{kj}, \Gamma_{kj}, W_{kj}\rangle : j = 1, \ldots, n_k\}_{k \geqq 0}$$

on the Polish space $\prod_{k \geqq 0} (\mathcal{S}_k \times \mathcal{S}_k \times T)^{n(k)}$ to obtain a joint law for all these random variables; they are independent for different values of k.

The law of $\{X_{kj}\}_{k \geqq 0,\, j=1, \ldots, n(k)}$ equals that of

$$\{\Lambda_{m(k)} X_{j+t(k)}\}_{k \geqq 0,\, j=1, \ldots, n(k)}.$$

So by Lemma 2.11 we may in fact take

$$X_{kj} \equiv \Lambda_{m(k)} X_{j+t(k)} \tag{6.19}$$

and define (Γ_{kj}, W_{kj}) on the original probability space Ω. Then we define

$$Y_j := W_{k, j-t(k)} \quad \text{for } t_k < j \leqq t_{k+1}, \qquad k \geqq 0. \tag{6.20}$$

Thus Y_j are independent and identically distributed with law μ_∞ on T as desired. Then as $M\to\infty$ we let $\mu := [M^{1/2}]$ and estimate

$$\begin{aligned}
&\max_{t(M)<n\leqq t(M+1)} \|\sum_{j\leqq n} X_j - Y_j\|^* \\
&\leqq \max_{n\in H(m)} \|\sum_{t(M)<j\leqq n} X_j\|^* + \max_{n\in H(m)} \|\sum_{t(M)<j\leqq n} Y_j\| \\
&\quad + \sum_{\mu\leqq k<M} \|\sum_{r\leqq n(k)} \Gamma_{kr} - W_{kr}\| + \sum_{\mu\leqq k<M} \|\sum_{j\in H(k)} X_j - \Lambda_{m(k)} X_j\|^* \\
&\quad + \sum_{\mu\leqq k<M} \|\sum_{r\leqq n(k)} X_{kr} - \Gamma_{kr}\| + \sum_{j\leqq t(\mu)} \|X_j\|^* + \|Y_j\| \\
&:= A_1 + A_2 + A_3 + A_4 + A_5 + A_6,
\end{aligned}$$

say, using (6.19) and (6.20), with $1\leqq r=j-t_k$. We will prove that for $1\leqq B<\infty$, $\theta<1/(2\beta)$, and $i=1,2,\ldots,6$, as $M\to\infty$

$$\Pr\{A_i \geqq t_M^{1/2}(Lt_M)^{-\theta}\} \ll M^{-B}. \tag{6.21}$$

For $i=1$, Proposition 6.2 gives for $C\leqq C_{6.2}(B)$

$$\Pr\{A_1 \geqq C(n_M L_2 t_M)^{1/2}\} \ll M^{-B}.$$

Now for $\theta<\varphi<\rho/2(1-\rho)=1/(2\beta)$,

$$(n_M L_2 t_M)^{1/2} \ll (t_M M^{-\rho} LM)^{1/2} \ll t_M^{1/2}(Lt_M)^{-\varphi}. \tag{6.22}$$

This implies (6.21) for $i=1$.

Since Y_j are Gaussian, the Landau-Shepp-Fernique inequality ([19], [20], or [48]) implies $E\|Y_1\|^3<\infty$. Thus, replacing X_j by Y_j in Prop. 6.2 and its proof, we obtain (6.21) for $i=2$.

For $i=3$, Proposition 6.3 gives, for $C\geqq C_{6.3}(2B+1)$,

$$\Pr\{\|\sum_{r\leqq n(k)} \Gamma_{kr} - W_{kr}\| \geqq C(n_k L_2 t_k)^{1/2}/m_k\} \ll k^{-2B-1}. \tag{6.23}$$

We have

$$\begin{aligned}
\sum_{k<M} (n_k L_2 t_k)^{1/2}/m_k &\ll \sum_{M/2<k\leqq M} \exp(k^{1-\rho}/2)\, k^{-\zeta-\rho/2} Lk \\
&\ll \exp(M^{1-\rho}/2)\, M^{-\zeta+\rho/2} LM \\
&\ll t_M^{1/2}(Lt_M)^{(-\zeta+\rho/2)/(1-\rho)} L_2 t_M.
\end{aligned}$$

For $\theta<\varphi<1/(2\beta)=\rho/2(1-\rho)$, and ζ close enough to ρ, we have $(-\zeta+\rho/2)/(1-\rho)<-\varphi<-\theta$. Then

$$\sum_{k<M} (n_k L_2 t_k)^{1/2}/m_k \ll t_M^{1/2}(Lt_M)^{-\varphi}, \tag{6.24}$$

and

$$\sum_{k\geqq\mu} k^{-2B-1} \leqq (\mu-1)^{-2B} \ll M^{-B}. \tag{6.25}$$

Thus (6.21) holds for $i=3$.

For $i=4$, Proposition 6.1 gives, for $C \geqq C_{6.1}(2B+1)$,

$$\Pr\{\|\sum_{j\in H(k)} X_j - A_{m(k)} X_j\|^* \geqq C(n_k L_2 t_k)^{1/2}/m_k\} \ll k^{-2B-1}.$$

Then by (6.24) and (6.25), we have (6.21) also for $i=4$.

For A_5 we use (6.18) and note that

$$\sum_{k<M} n_k^{(5-\alpha)/10} \ll M n_M^{(5-\alpha)/10} \ll n_M^{(10-\alpha)/20}$$
$$\ll t_M^{1/2}(\log t_M)^{-\varphi}, \quad \varphi < 1/(2\beta),$$

while

$$\sum_{k\geqq\mu} n_k^{-\alpha/10} \ll \sum_{k\geqq\mu} k^{-2B-1} \ll \mu^{-2B} \ll M^{-B}.$$

So (6.21) holds for $i=5$. For $i=6$,

$$\Pr\{\sum_{j\leqq t(\mu)} \|X_j\|^* + \|Y_j\| \geqq n_M^{1/3}\} \leqq \sum_{j\leqq t(\mu)} E(\|X_j\|^* + \|Y_j\|)/n_M^{1/3}$$
$$\ll t(\mu)\, n_M^{-1/3} \ll M^{-B},$$

while $n_M^{1/3} \ll t_M^{1/2}(Lt_M)^{-\varphi}$, $\theta < \varphi < 1/(2\beta)$, so (6.21) holds for $i=6$.

As $M\to\infty$, $t_M < n \leqq t_{M+1}$ implies $n \sim t_M$ and $\log n \sim M^{1-\rho}$, so $M^{-B} \ll (\log n)^{-B}$. Again letting $\theta \uparrow 1/(2\beta)$ gives Theorem 6.2 in the (6.3) case.

If instead of (6.3), condition (6.4) holds, then *instead of* (6.9), (6.10) and (6.11) we define, for $\rho := 189\gamma/\alpha^2$;

(6.26) $$m_k := k, \quad n_k := [k^\rho] \quad \text{and} \quad t_k := \sum_{j<k} n_j.$$

Then $t_k \sim k^{\rho+1}/(\rho+1)$ as $k\to\infty$. Other definitions in terms of these remain the same. In place of Props. 6.1 and 6.2 we now will have

Proposition 6.4. *There is a $C = C_{6.4}$ large enough so that as $k\to\infty$*

$$\Pr\{\|\sum_{j\in H(k)} V_j(m_k)\|^* \geqq C(n_k L t_k)^{1/2}\} \ll k^{-10\gamma/\alpha}$$

and

$$\Pr\{\max_{n\in H(k)} \|\sum_{t(k)<j\leqq n} X_j\|^* \geqq C(n_k L t_k)^{1/2}\} \ll k^{-10\gamma/\alpha}.$$

Proof. Wherever $L_2 t_k$ appeared in the (6.3) case we put Lt_k in the (6.4) case. Our choice of ρ in (6.26) implies

$$14 + 16/\alpha + (\rho+1)/(2+\alpha) < \rho/2 - 1,$$

so following the proof of Proposition 6.1 we have $M_k = o(K_k)$. If we take $\xi > \xi_1(B)$ large enough, $B = 10\gamma/\alpha$, we still get a bound k^{-B} from Lemma 2.6. In the last terms of (6.13) we now get $t_k^{-\delta/21} \ll k^{-10\gamma/\alpha}$. The proof of Proposition 6.2 adapts likewise. Q.E.D.

Next, in Proposition 6.3 we (as always) replace $L_2 t_k$ by Lt_k. In its proof we need only replace η by $5(BLt_k)^{1/2}$.

Then Lemma 2.12, (6.4) and (6.26) give, in place of (6.17),

$$\pi(F_k, G_k) \ll n_k^{-\alpha/9} k^{4\gamma/3} \ll k^{-R} \tag{6.27}$$

where $R := (\alpha\rho - 2\gamma)/9 > 1$, so $\pi(F_k, G_k) \leqq C_7 k^{-R}$ for some constant C_7. The latter bound replaces $C_6 n_k^{-\alpha/10}$ everywhere; specifically, in place of (6.18) we have

$$\Pr\left\{n_k^{-1/2}\left\|\sum_{j=1}^{n(k)} \Gamma_{kj} - X_{kj}\right\| > C_7 k^{-R}\right\} \ll k^{-R}. \tag{6.28}$$

The proof then runs unchanged until (6.21) except that we redefine $\mu := [M^{4/9}]$. In place of (6.21) it will be shown that for some constants D_i,

$$\Pr\{A_i \geqq D_i t_M^{-\lambda+1/2}\} \ll M^{-8\gamma/\alpha}. \tag{6.29}$$

We first note that

$$\rho/2 < (\rho+1)(\tfrac{1}{2} - \lambda). \tag{6.30}$$

Then letting $D_1 = C_{6.4}$ and applying Proposition 6.4 (latter half) gives (6.29) for $i=1$. As in the (6.3) case, the proof of Propositions 6.2 and 6.4 adapts to Y_j in place of X_j to give D_2 for which (6.29) holds, $i=2$. For $i=3$,

$$\sum_{k<M} (n_k L t_k)^{1/2}/m_k \ll \sum_{k<M} k^{-1+\rho/2} Lk \ll M^{\rho/2} LM \ll t_M^{-\lambda+1/2} \tag{6.31}$$

by (6.30) again. Since $\lambda < 1/6$, using Proposition 6.3 as adapted (L in place of L_2) gives a D_3 large enough such that, noting

$$\sum_{k \geqq \mu} k^{-3B-1} \ll \mu^{-3B} \ll M^{-B}, \tag{6.32}$$

we obtain (6.29) for $i=3$. The case $i=4$ follows likewise, using Prop. 6.4 (first half).

For $i=5$ we use (6.28). Then

$$\sum_{k<M} n_k^{1/2} k^{-R} \ll \sum_{k<M} k^{-R+\rho/2} \ll M^{1-R+\rho/2} \ll t_M^{-\lambda+1/2}$$

since $1 + \rho\lambda < R$. Also,

$$\sum_{k \geqq \mu} k^{-R} \ll \mu^{1-R} \ll M^{-8\gamma/\alpha}$$

giving (6.29) for $i=5$. For $i=6$, Markov's inequality gives for $D_6 := 1$

$$\begin{aligned}\Pr\{\sum_{j\leqq t(\mu)} \|X_j\|^* + \|Y_j\| \geqq t_M^{-\lambda+1/2}\} \\ \leqq \sum_{j\leqq t(\mu)} E(\|X_j\|^* + \|Y_j\|)\, t_M^{\lambda-1/2} \\ \ll t_\mu t_M^{\lambda-1/2} \ll M^{(\rho+1)(\lambda-1/18)} \ll M^{-10\gamma/\alpha}\end{aligned}$$

and (6.29) follows for $i=6$. Now $t_M < n \leqq t_{M+1}$ implies

$$M^{-8\gamma/\alpha} \ll t_M^{-8\gamma/(\alpha(\rho+1))} \ll n^{-\delta/28}.$$

Setting $H := \sum_{i\leqq 6} D_i$ we obtain Theorem 6.2 in the (6.4) case. Q.E.D.

Remarks. In the above proof, the large size of ρ in the (6.4) case was primarily used in the proof of Proposition 6.4 (first display). Once ρ is large, then to satisfy (6.30), λ must be small. Since we suspect that the result is far from best possible we did not seek the largest possible λ using our methods.

7. Application to Empirical Processes

In this section Theorems 6.1 and 6.2 will be applied to empirical processes, giving speeds of convergence in Theorem 1.3 uniformly over suitable families $\mathscr{F}$ of functions. Our rates of convergence, proved in some generality, are relatively slow, but are sufficient to imply, for example, some "upper and lower class" results (cf. Corollary 4 of [44]). On the other hand for special classes of functions on, or sets in, Euclidean spaces, defined by differentiability conditions, and under some more or less severe restrictions on the underlying probability laws, much faster rates have been obtained (Révész, 1976; Ibero, 1979a, b).

For a collection $\mathscr{C}$ of sets we take $\mathscr{F} = \{1_B : B \in \mathscr{C}\}$. Let $(A, \mathscr{A}, P)$ be a probability space. First, under hypotheses on $\log N_I$, the metric entropy with inclusion, stronger than (1.23), here are rates of convergence in Theorem 1.5.

Theorem 7.1. *Let* $\mathscr{C}$ *be a collection of measurable sets with* $\log N_I(x, \mathscr{C}, P) \leqq c x^{-\tau}$, $x > 0$, *for some constants* $c > 0$ *and* $0 \leqq \tau < 1$. *Then for any* $H < \infty$ *and* $\theta < (1-\tau)/(4\tau)$, *we can choose* Y_j *in Theorem* 1.3 *to improve* (1.28) *to*

$$(7.1)\quad \Pr^*\{n^{-1/2} \max_{k\leqq n} \sup_{B\in\mathscr{C}} |\sum_{j\leqq k} 1_B(x_j) - P(B) - Y_j(1_B)| > (\log n)^{-\theta}\} \ll (\log n)^{-H},$$

and for some measurable U_n, almost surely

$$(7.2)\qquad \sup_{B\in\mathscr{C}} |\sum_{j\leqq n} 1_B(x_j) - P(B) - Y_j(1_B)| \leqq U_n = o(n^{1/2}(\log n)^{-\theta}).$$

Proof. Given $m \geqq 1$, let $\varepsilon = 1/(3m)$. In the proof of Theorem 5.1 of [12], to satisfy (5.2) and (5.3) there we can set $\alpha = \gamma = c_1 \varepsilon^{2/(1-\tau)}$ for a small enough $c_1 = c_1(\tau, c) > 0$. We have (5.4) as corrected, i.e.

$$\sum_{i\geqq u} (2^{-i} \log N_I(2^{-i}))^{1/2} < \varepsilon/96,$$

if $2^{-u} < c_2 \varepsilon^{2/(1-\tau)}$ for some $c_2 = c_2(\tau, c)$. There is a constant $K = K(\tau)$ such that

$$\exp(-K\varepsilon^{-2\tau/(1-\tau)}) < \varepsilon/200, \quad 0 < \varepsilon \leqq 1.$$

Then there is a $c_3(\tau) \leqq \min(c_1(\tau), c_2(\tau))$ small enough so that for all $j \geqq 0$,

$$2^j > c_3(\tau)\, K(\tau)\, 9000(j+1)^5.$$

Then both (5.4) and (5.5) of [12] hold if $2^{-u} < c_3(\tau)\, \varepsilon^{2/(1-\tau)}$. Thus we can take $\delta_0 := 2^{-u}$ for the least such u. Then

$$n_0(\varepsilon) > \varepsilon^{-(2+2\tau)/(1-\tau)}\, c_3(\tau)^{-2}/256 \geqq n_0(\varepsilon)/4.$$

We then obtain (1.15) above with its ε replaced by $1/m$ and $\delta^2 = \delta_0 \geqq c_3(\tau)(3m)^{-2/(1-\tau)}/2$, according to [12, p. 917, line 2] with its ε corrected to $3\varepsilon = 1/m$. Then in the proof of Theorem 1.3 above we can let

$$\dim \Lambda_m S = N_I(\delta_0) \leqq \exp(c\delta_0^{-\tau}) \leqq \exp(c_4(\tau)\, m^{2\tau/(1-\tau)})$$

for some $c_4(\tau) > 0$, and $n_0(m) \leqq c_5(\tau)\, m^{(2+2\tau)/(1-\tau)}$ for some $c_5(\tau) > 0$.

So condition (6.3) holds with $\beta = 2\tau/(1-\tau)$ and (6.2) holds with $D = 2(1+\tau)/(1-\tau)$. Thus Theorem 6.2 applies with θ as stated. Q.E.D.

Corollary 7.2. *Let P on $\mathbb{R}^2$ have a bounded density with respect to the Lebesgue measure and let $\mathscr{C} = \mathscr{C}(U)$ be the collection of all convex subsets of a bounded open set $U \subset \mathbb{R}^2$. Then for any $\theta < 1/4$ and $H < \infty$ we can choose Y_j in Theorem* 1.5 *such that* (7.1) *and* (7.2) *hold.*

Proof. We can apply Theorem 7.1 with $\tau = 1/2$ according to Bronštein (1976). □

In $\mathbb{R}^3$, Theorem 1.5 does not hold for the convex sets [13].

Let $J(k, \alpha, M)$ be the collection of compact subsets of $\mathbb{R}^k$ with boundaries defined by functions with all partial derivatives of orders $\leqq \alpha$ bounded by M, as defined in [11, with Correction] or [12, p. 917].

Corollary 7.3. *For P with bounded support and bounded density with respect to Lebesgue measure on $\mathbb{R}^k$, $k \geqq 1$, if $\alpha > k-1$ then for any $r > (k-1)/\alpha$, $\theta := (1-r)/(4r)$ and $H < \infty$,* (7.1) *and* (7.2) *hold for $\mathscr{C} = J(k, \alpha, M)$, $M < \infty$.*

Proof. We may assume $r < 1$. Then Theorem 7.1 applies to any τ such that $(k-1)/\alpha < \tau < r < 1$ [11] and the rest follows. □

Remark. Révész (1976) obtained in Corollary 7.3, if $k = \alpha = 2$, for regions with 1-1 boundary curves, if P is the uniform measure on the unit square or has a sufficiently regular density, (7.2) with $\mathcal{O}(n^{1/2}(\log n)^{-\theta})$ replaced by $\mathcal{O}(n^{12/25})$, a much better result in that case. We suppose that for such P, Révész' method would also yield $\mathcal{O}(n^{12/25})$ in (7.2) for the convex sets as in Corollary 7.2. But we do not see how to extend Révész' construction to the generality of Theorem 7.1.

Next, for Vapnik-Červonenkis classes of sets (mentioned late in Sect. 1 above) we do need [16] some measurability condition, such as the following [15].

Given a probability space $(A, \mathscr{A}, P)$ and $\mathscr{C} \subset \mathscr{A}$, we say $\mathscr{C}$ is *image P∈-Suslin* iff there is a measurable space $(W, \mathscr{S})$ and a map G of W onto $\mathscr{C}$ such that:

1) $\{\langle x, w\rangle : x \in G(w)\}$ is $\mathscr{A} \times \mathscr{S}$-measurable;

2) for the pseudo-metric $d_P(B, C) = P(B \triangle C)$ and any d_P-open set $U \subset \mathscr{C}$, $G^{-1}(U) \in \mathscr{S}$;

3) $(W, \mathscr{S})$ is Suslin, i.e. for some Polish space V there is a map H of V onto W with $H^{-1}(B)$ Borel measurable for all $B \in \mathscr{S}$. Also, $(A, \mathscr{A})$ is Suslin.

Theorem 7.4. *Suppose $(A, \mathscr{A}, P)$ is a probability space, $\mathscr{C} \subset \mathscr{A}$, $V(\mathscr{C}) < \infty$ and $\mathscr{C}$ is image P∈-Suslin. Then for any $\lambda < 1/(2700\, V(\mathscr{C}))$*

$$\Pr{}^*\{n^{-1/2} \max_{k \leq n} \sup_{B \in \mathscr{C}} |\sum_{j \leq k} 1_B(x_j) - P(B) - Y_j(1_B)| > n^{-\lambda}\} \ll n^{-1/50} \tag{7.3}$$

and for some measurable U_n, almost surely

$$\sup_{B \in \mathscr{C}} |\sum_{j \leq n} 1_B(x_j) - P(B) - Y_j(1_B)| \leq U_n = \mathcal{O}(n^{1/2 - \lambda}). \tag{7.4}$$

Proof. Examination of the proof of [12, Theorem 7.1, Correction] (clarified in [15]) shows that given $\varepsilon > 0$ we can take the δ there to satisfy, for a small enough constant $c > 0$,

$$\delta = c\,\varepsilon^2/|\ln \varepsilon| \tag{7.5}$$

(specifically see [12, Correction, p. 909, last display]). By the last two lines of the proof of [12, Theorem 7.1], we have for $n \geq n_0(\delta)$ large enough

$$\Pr\{\sup\{|v_n(B \setminus C)| : B, C \in \mathscr{C}, P(B \setminus C) < \delta\} \geq 6\varepsilon\} < \varepsilon.$$

Then in the proof of Theorem 1.3 above, replacing δ^2 in (1.15) by δ and taking $m = 6/\varepsilon$, we apply [12, Lemma 7.13] to obtain a linear mapping Λ_m onto a subspace of dimension $\ll \delta^{-w}$ for any $w > V(\mathscr{C})$. Thus by (7.5) we have $\dim \Lambda_m S \ll \varepsilon^{-2w}$ for each $w > V(\mathscr{C})$. So in (6.4) we can take any $\gamma > 2V(\mathscr{C})$. In the hypothesis of Theorem 6.1 we have $\delta = 1$.

In the proof of [12, Theorem 7.1] we must take $n \geq \max(n_0(\delta), n_1(\varepsilon))$ where for ε small it suffices to take $n_0(\delta) \geq \delta^{-r}$ for any $r > 8$ and $n_1(\varepsilon) \geq \varepsilon^{-r}$ for $r > 2$, so (6.2) holds for any $D > 8$. Thus the conclusions of Theorems 6.1 and 6.2 in the (6.4) case hold for any $\kappa < 2/3$, so for $\lambda < 1/(2700\, V(\mathscr{C}))$ and $\kappa/28 < 1/42$. Q.E.D.

Remarks. Let $(A, \mathscr{A}, P)$ be the unit interval with Lebesgue measure and let $\mathscr{C}_k$ be the collection of all unions of at most k intervals. Then $V(\mathscr{C}_k) = 2k+1$. Yet from Komlós, Major and Tusnády (1975, Theorem 4) one can get (7.4) with $\mathcal{O}(n^{.5-\lambda})$ replaced by $\mathcal{O}(\log n)^2$, for any P. So this rate holds for nontrivial classes with arbitrarily large $V(\mathscr{C})$. It is not known, in fact, whether it or even the rate $\mathcal{O}(\log n)$ may hold for arbitrary Vapnik-Červonenkis classes. Révész (1976, Lemma 12) considers the class $P(\kappa, 1)$ of polygons with at most κ sides where P is uniform on the unit square in $\mathbb{R}^2$ or has a regular enough density, and obtains (7.4) with $(n^{.5-\lambda})$ replaced by $\mathcal{O}(n^{3/8} \log^2 n)$, which although much better than our rate, depends on the specific assumptions made on P and is much slower than the Komlós-Major-Tusnády rate.

To obtain empirical distribution functions one just takes the collection $\mathscr{C}$ of all sets $B_x \subset \mathbb{R}^k$ where

$$B_x = \{y: y_j \leqq x_j, j=1, \ldots, k\}, \qquad x \in \mathbb{R}^k.$$

Here $V(\mathscr{C})=k+1$ [63, Proposition 2.3], so that Theorem 7.4 improves somewhat on Theorem 1 of Philipp and Pinzur (1980), again for arbitrary P (for P e.g. uniform on the unit cube, Révész [58] obtains a much smaller error term even over a much larger collection $\mathscr{C}$ of sets).

Given a set $\mathscr{F} \subset \mathscr{L}^2(A, \mathscr{A}, P)$ of functions, and $\varepsilon > 0$, let $N_I(\varepsilon, \mathscr{F}, P)$ be the smallest m such that for some $f_1, \ldots, f_m \in \mathscr{L}^2(A, \mathscr{A}, P)$, for all $f \in \mathscr{F}$ there are $i, j \leqq m$ with $f_i \leqq f \leqq f_j$ and $\int f_j - f_i \, dP < \varepsilon$. Then $\log N_I(\varepsilon, \mathscr{F}, P)$ has been called a *metric entropy with bracketing* [14].

Suppose that for some $F \in \mathscr{L}^2$, $|f| \leqq F$ for all $f \in \mathscr{F}$. In [14] it is shown that if $F \in \mathscr{L}^p$ for some $p > 2$ and $\log N_I(\varepsilon, \mathscr{F}, P) = \mathcal{O}(\varepsilon^{-r})$ for some $r < 1 - 2/p$ as $\varepsilon \downarrow 0$, then (1.14) and (1.15) hold. For this, or any such future result under weaker conditions on p and r, one can again obtain invariance principles (1.18) or (1.19) without needing further measurability conditions on the class $\mathscr{F}$. For such classes, the proof in [14] gives dimensions too large for our Theorems 6.1 and 6.2 to apply.

Pollard [56b] shows that if $\mathscr{C} \subset \mathscr{A}$ is a countable Vapnik-Červonenkis class, $F \in \mathscr{L}^2(A, \mathscr{A}, P)$, and $\mathscr{F} = \{F 1_B : B \in \mathscr{C}\}$, then (1.14) and (1.15) hold for $\mathscr{F}$. Thus Theorem 1.3 applies to give either (1.18) or (1.19).

We end this section with an application to weighted empirical distribution functions, which improves somewhat on one direction of results of O'Reilly (1974, Theorem 2), James (1975), and Goodman, Kuelbs and Zinn (1981, Theorem 6.1).

Theorem 7.5. *Let* $\{W_j, j \geqq 1\}$ *be a sequence of independent random variables with uniform distribution on* $[0, 1]$. *Define*

$$\begin{aligned} X_j &= \omega(s)(1\{W_j \leqq s\} - s), && 0 < s < 1 \\ &= 0 && \textit{else}, \end{aligned}$$

where ω *is a real function with the following properties:*

(i) ω *is continuous and positive on* $(0, 1)$

(ii) *for some* $\gamma > 0$, *we have* ω *is nonincreasing (nondecreasing) on* $(0, \gamma]$ ($[1-\gamma, 1)$ *respectively)*

(iii) *For all* $\varepsilon > 0$ *and* $i = 1, 2$

$$\int_0^1 s^{-1} \exp(-\varepsilon / k_i(s)) \, ds < \infty$$

where $k_1(s) = s\omega^2(s)$ *and* $k_2(s) = s\omega^2(1-s)$.

(iv) *We have*

$$\int_0^1 \omega^2(s) / L_2\left(\frac{1}{s(1-s)}\right) ds < \infty.$$

Then there exists a sequence $\{Y_j, j \geqq 1\}$ *of independent identically distributed* $C[0,1]$*-valued Gaussian random variables, indexed by* $s \in [0,1]$, *and with sample functions almost surely continuous (in s) such that*

(i)′ $EY_1(s)=0,\ 0 \leqq s \leqq 1,$

(ii)′ $E\{Y_1(s)\,Y_1(t)\}=\omega(s)\,\omega(t)\,s(1-t),\ 0 \leqq s \leqq t \leqq 1,$

(iii)′ $n^{-1/2}\, E\{\max_{k \leqq n} \sup_{0 \leqq s \leqq 1} |\sum_{j \leqq k} X_j(s)-Y_j(s)|\} \to 0$

or, instead of (iii)′, *for some measurable* V_n

(iv)′ $\sup_{0 \leqq s \leqq 1} |\sum_{j \leqq n} X_j(s)-Y_j(s)| \leqq V_n = o((nLLn)^{1/2})$ *a.s.*

Proof. Under conditions (i), (ii) and (iii), O'Reilly (1974, Theorem 2) proves that the law of $Z_n := (\sum_{1 \leqq j \leqq n} X_j)/n^{1/2}$ converges to that of Y_1 in the Polish space $D[0,1]$ with Skorohod topology. By Skorohod's theorem [61] there exist U_n with $\mathscr{L}(U_n)=\mathscr{L}(Z_n)$, $n \geqq 1$, $\mathscr{L}(U_0)=\mathscr{L}(Y_1)$ and $U_n \to U_0$ a.s. for the Skorohod metric. Since the limit process has continuous sample functions, also $U_n \to U_0$ for the sup norm. Now $\langle f, g\rangle \to \sup|f-g|$ is jointly measurable on $D[0,1] \times D[0,1]$ since we can restrict the supremum to rational arguments. So $\sup|U_n-U_0| \to 0$ a.s. and in probability. Since $\mathscr{L}(U_0)$ is tight on a Polish space $C[0,1]$, for any $m \geqq 1$ there is a map Λ_m of $D[0,1]$ into $C[0,1]$ with finite-dimensional range (consisting of piecewise linear functions with given vertices) such that

$$\Pr\{\sup|U_0-\Lambda_m U_0| > 1/(3m)\} < 1/(3m).$$

For some n_0 we have for $n \geqq n_0$

$$\Pr\{\sup|U_n-U_0| > 1/(3m)\} < 1/(3m).$$

Since the Λ_m can be defined by interpolation and are linear we have $\sup|\Lambda_m f| \leqq \sup|f|$ for all $f \in D[0,1]$, so

$$\Pr\{\sup|\Lambda_m U_n-\Lambda_m U_0| > 1/(3m)\} < 1/(3m).$$

Combining we have

$$\Pr\{\sup|\Lambda_m U_n-U_n| > 1/m\} < 1/m,$$

and likewise for Z_n, i.e. we have the tightness condition (1.5). The other hypotheses in Theorem 1.1 are easily checked, so we obtain (iii)′. Now $E\|X_1\|^2/L_2\|X_1\| < \infty$ is equivalent to (iv), assuming (i), (ii) and (iii): Goodman, Kuelbs and Zinn (1981, Lemma 6.2) give a proof, on which we have the following comments.

In their statement of Theorem 2 of O'Reilly (1974), specifically in (6.3), $k_2(t)$ should be $t\omega^2(1-t)$ (or else in the k_2 case, t^{-1} should be replaced by $(1-t)^{-1}$). By (i), $\sup_{\gamma \leqq s \leqq 1-\gamma} \omega(s) < \infty$. Next, $\sup_{0<s<\gamma} s\omega(s) < \infty$ by (iii) (cf. [23, Lemma 6.1]), and likewise $\sup_{1-\gamma \leqq s < 1} (1-s)\,\omega(s) < \infty$. Thus integrability conditions for $\|X_1\|$ depend only on what happens for $W_1<\gamma$ or $W_1>1-\gamma$; by symmetry, we need only consider $W_1<\gamma$. Then $\sup\{X_1(s)\colon W_1\leqq s<\gamma\}$ is attained at $s=W_1$ and equals $\omega(W_1)(1-W_1)$. Since the suprema over other intervals are uniformly bounded as above and we can assume $W_1<\gamma\leqq 1/3$ here, $E\|X_1\|^2/L_2\|X_1\|<\infty$ is equivalent to

$$(7.6)\qquad E(\omega^2(W_1)/L_2\,\omega(W_1))<\infty.$$

Since $\omega(t)\ll 1/t$ as $t\downarrow 0$, (7.6) implies (iv). Conversely if (iv) holds, let $x_n=x(n)=\sup\{x\leqq\gamma\colon \omega(x)\geqq n\}$. If $\omega(\cdot)$ is bounded for $x\leqq\gamma$, then (7.6) holds, so we can assume that the x_n are all defined. Then $x_n\downarrow 0$. Now we have

$$(7.7)\qquad \sum_n n^2 \int_{x(n+1)}^{x(n)} dx/L_2(1/x)<\infty$$

and (7.6) is equivalent to

$$(7.8)\qquad \sum_n n^2(x_n-x_{n+1})/L_2\,n<\infty.$$

To prove (7.8), first note that the sum of those terms such that $x_n\leqq 2n^{-4}$ clearly converges. For the terms with $x_{n+1}\geqq x_n/2>n^{-4}$, note that for $x_{n+1}\leqq x$, $1/L_2(1/x)\geqq 1/L_2(n^4)\sim 1/L_2\,n$ as $n\to\infty$, so the sum of such terms in (7.8) also converges. Terms in (7.8) with $x_{n+1}<x_n/2>n^{-4}$ are at most doubled if we replace x_{n+1} by $x_n/2$, and then their sum converges as in the last case. So (7.8) and hence (7.6) are proved. Thus we can apply Theorem 1.2, completing the proof of Theorem 7.5.

8. Necessity of Measurability in Separable Normed Spaces

It is well known that, for example, the law of an empirical distribution function in $D[0,1]$ need not be defined on the Borel σ-algebra for the (non-separable) supremum norm. Thus results such as our Theorem 1.1 really need to allow non-measurability of the variables X_j. On the other hand we assumed in (1.6) that suitable finite-dimensional variables are measurable. We now clarify the roles of such assumptions by showing that in any separable Banach space, a weak central limit theorem (or *a fortiori*, invariance principles such as ours) can only hold for measurable variables. The passage from one-dimensional to separable Banach spaces will be easy. First we have a one-dimensional result.

We have the usual inner measure $\Pr_*(B):=\sup\{\Pr(C)\colon C\subset B\}$ and set $f_*:=-((-f)^*)=\operatorname{ess.sup}\{g\colon g\leqq f,\ g \text{ measurable}\}$.

Theorem 8.1. *Let $X_n=h(x_n)$, $n=1,2,\ldots$, where x_n are i.i.d. random variables with values in some measurable space A and h is any real-valued function (not assumed measurable) on A. Let $S_n:=X_1+\ldots+X_n$. Suppose $\mathscr{L}(S_n/n^{1/2})\to N(0,1)$ in the sense that for all t,*

$$\lim_{n\to\infty}\Pr{}^*(S_n/n^{1/2}\leqq t)=\lim_{n\to\infty}\Pr{}_*(S_n/n^{1/2}\leqq t)=N(0,1)(]-\infty,t]).$$

Then h is measurable for the completion of $\mathscr{L}(x_1)$, so that X_i are measurable, $EX_i=0$ and $EX_i^2=1$.

Proof. It is classical that if h is measurable then $EX_i=0$ and $EX_i^2=1$ (Gnedenko and Kolmogorov, 1968, §35, Theorem 4, p. 181). Suppose h is non-measurable for $\mathscr{L}(x_1)$, so that X_n are non-measurable, and consider $X_{n*}\leqq X_n\leqq X_n^*$.

Let B be the measurable set on which $X_1^*=+\infty$. If $P(B)>0$ then P restricted to subsets of B must be non-atomic. We always have $X_1^*>-\infty$ everywhere. For some numbers $M_n\uparrow+\infty$ we have $P(X_1^*\leqq -M_n)\leqq n^{-3}$. Then $P(\min_{j\leqq n} X_j^*\leqq -M_n)\leqq n^{-2}$. Then by the Borel-Cantelli lemma, almost surely for n large enough $X_j^*\geqq -M_n$ for all $j\leqq n$, so $\sum_{1\leqq j\leqq n}\{X_j^*: X_j^*<0\}\geqq -nM_n$.

On B define a measurable, finite valued $Y_1\geqq 0$ such that for n large enough $P(Y_1\geqq nM_n+2n)\geqq n^{-1/2}$. Let $Y_1=X_1^*-1$ outside B. Define Y_n from X_n for all n just as Y_1 from X_1. Then

$$P(\max_{j\leqq n} Y_j\geqq nM_n+2n)\geqq 1-(1-n^{-1/2})^n\to 1$$

rapidly as $n\to\infty$, so almost surely for n large enough, there is a $j\leqq n$ with $Y_j\geqq nM_n+2n$ and thus $\sum_{1\leqq j\leqq n} Y_j\geqq n$. Since $Y_j<X_j^*$ for all j, $P^*(X_j\geqq Y_j,\ j=1,\ldots,n)=1$ by independence and Lemma 2.3. Then $P^*(S_n/n^{1/2}\geqq n^{1/2})\to 1$ as $n\to\infty$, a contradition. Thus $P(B)=0$ and X_1^* is finite valued a.s.

Let $B_j:=B(j):=\{X_j\geqq X_j^*-2^{-j}\}$. Then $P^*(B_j)=1$, $j=1,2,\ldots$. Let $C_n=\bigcap_{j=1}^{n} B_j$. We apply Lemma 2.3 to $P_j:=\mathscr{L}(x_j)$ and $f_j:=1_{B(j)}$, giving $P^*(C_n)=1$. On C_n, $S_n\leqq X_1^*+\ldots+X_n^*\leqq S_n+1$. Thus $\mathscr{L}((X_1^*+\ldots+X_n^*)/n^{1/2})\to N(0,1)$. Hence $EX_1^*=0$. Likewise $EX_{1*}=0$. Thus $X_{1*}=X_1=X_1^*$ a.s., i.e. X_1 is completion measurable. Q.E.D.

Corollary 8.2. *Let S be any real vector space and $X_n=h(x_n)$ where x_n are independent, identically distributed random variables with values in some measurable space $(A,\mathscr{A})$ and h is any function from A into S (not assumed measurable). Let F be a collection of linear functionals: $S\to\mathbb{R}$. Suppose that for each $f\in F$, the central limit theorem as in Theorem* 8.1 *holds for the $f(h(x_j))$, with some limit law $N(0,\sigma_f^2)$, $\sigma_f^2>0$. Then each $f\circ h$, $f\in F$, is measurable for the completion of $\mathscr{L}(x_1)$ on $\mathscr{A}$.*

Corollary 8.3. *If S is a separable normed space, the hypotheses of Theorem* 1.1 *imply that the X_j are completion measurable for the Borel σ-algebra on S.*

Proof. Let $F=S'$, the dual Banach space. The conclusion of Theorem 1.1 implies the hypotheses of Corollary 8.2. Let $\mathscr{I}$ be the σ-algebra of all subsets B of S such that $h^{-1}(B)$ is measurable for the completion of $\mathscr{L}(x_1)$ on A. Then all elements of S' are $\mathscr{I}$ measurable. Since S is separable, $\mathscr{I}$ must contain all Borel sets. □

Corollary 8.4. *If the hypotheses of Theorem* 1.1 *hold and H is a bounded linear operator from S into a separable Banach space, then the $H(X_i)$ must be measurable for the completion of $\mathscr{L}(x_1)$.*

References

1. Acosta, Alejandro de: Existence and convergence of probability measures in Banach spaces. Trans. Amer. Math. Soc. **152**, 273–298 (1970)
2. Berkes, István, Morrow, G.: Strong invariance principles for mixing random fields. Z. Wahrscheinlichkeitstheorie verw. Geb. **57**, 15–37 (1981)
3. Berkes, I., Philipp, Walter: An almost sure invariance principle for the empirical distribution function of mixing random variables. Z. Wahrscheinlichkeitstheorie verw. Geb. **41**, 115–137 (1977)
4. Berkes, I., Philipp, Walter: Approximation theorems for independent and weakly dependent random vectors. Ann. Probability **7**, 29–54 (1979)
5. Billingsley, Patrick: Convergence of Probability Measures. New York: Wiley 1968
6. Borisov, I.S.: Some limit theorems for empirical distributions. Abstracts of Reports. Third Vilnius Conf. Probability. Theory Math. Statist. **I**, 71–72 (1981)
7. Bourbaki, N.: Eléments de math. Première partie, livre VI, Intégration. Paris: Hermann 1959
8. Breiman, Leo: Probability. Reading Mass.: Addison-Wesley 1968
9. Bronštein, E.M.: ε-entropy of convex sets and functions. Siberian Math. J. **17**, 393–398 (transl. from Sibirsk. Mat. Ž. **17**, 508–514) (1976)
10. Dehling, Herold: Limit theorems for sums of weakly dependent Banach space valued random variables. Z. Wahrscheinlichkeitstheorie verw. Geb., to appear (1983)
11. Dudley, R.M.: Metric entropy of some classes of sets with differentiable boundaries. J. Approximation Th. **10**, 227–236 (1974); Correction, ibid. **26**, 192–193 (1979)
12. Dudley, R.M.: Central limit theorems for empirical measures. Ann. Probability **6**, 899–929 (1978); Correction, ibid. **7**, 909–911 (1979)
13. Dudley, R.M.: Lower layers in $\mathbb{R}^2$ and convex sets in $\mathbb{R}^3$ are not GB classes. Probability in Banach Spaces II (Proc. Conf. Oberwolfach, 1978). Lecture Notes in Math. **709**, 97–102. Berlin-Heidelberg-New York: Springer 1979
14. Dudley, R.M.: Donsker classes of functions, Statistics and Related Topics (Proc. Symp. Ottawa, 1980), pages 341–352. New York Amsterdam: North-Holland 1981
15. Dudley, R.M.: Vapnik-Červonenkis Donsker classes of functions. Aspects statistiques et aspects physiques des processus gaussiens (Proc. Colloque C.N.R.S. St.-Flour, 1980), C.N.R.S., Paris, 251–269 (1981b)
16. Durst, Mark, Dudley, R.: Empirical processes, Vapnik-Chervonenkis classes and Poisson processes. Probability and Mathematical Statistics (Wrocław) **1**, 109–115 (1981)
17. Eames, W., May, L.E.: Measurable cover functions. Canad. Math. Bull. **10**, 519–523 (1967)
18. Enflo, Per: A counterexample to the approximation problem in Banach spaces. Acta Math. **130**, 309–317 (1973)
19. Fernique, Xavier: Intégrabilité des vecteurs gaussiens. C.R. Acad. Sci. Paris Sér. A **270**, A1698–1699 (1970)
20. Fernique, X.: Régularité de processus gaussiens. Invent. Math. **12**, 304–320 (1971)
21. Freedman, D.: Brownian motion and diffusion. San Francisco; Holden-Day 1971
22. Gnedenko, B.V., Kolmogorov, A.N.: Limit distributions for sums of independent random variables (transl. K.L. Chung). Reading, Mass.: Addison-Wesley 1949, 1954, 1968
23. Goodman, Victor, Kuelbs, James, Zinn, Joel: Some results on the LIL in Banach space with applications to weighted empirical processes. Ann. Probability **9**, 713–752 (1981)
24. Hartman, Philip, Wintner, Aurel: On the law of the iterated logarithm. Amer. J. Math. **63**, 169–176 (1941)
25. Heinkel, B.: Relation entre théorème central-limite et loi du logarithme itéré dans les espaces de Banach. Z. Wahrscheinlichkeitstheorie verw. Geb. **49**, 211–220 (1979)
26. Hoffmann-Jørgensen, J.: Sums of independent Banach space valued random variables. Studia Math. **52**, 159–186 (1974)
27. Ibero, M.: Approximation forte du processus empirique fonctionnel multidimensionnel. Bull. Sci. Math. **103**, 409–422 (1979a)
28. Ibero, M.: Intégrale stochastique par rapport au processus empirique multidimensionnel. C.R. Acad. Sci. Paris Sér. A **288**, A165–168 (1979b)
29. Ionescu Tulcea, A., and C.: On the lifting property I. J. Math. Anal. Appl. **3**, 537–546 (1961)
30. Ionescu Tulcea, A. and C.: On the lifting property II: Representation of linear operators on spaces L_{E^r}, $1 \leqq r < \infty$. J. Math. Mech. (Indiana) **11**, 773–795 (1962)

31. Jain, N.C., Marcus, M.B.: Integrability of infinite sums of independent vector-valued random variables. Trans. Amer. Math. Soc. **212**, 1–36 (1975)
32. Jain, N.C.: Central limit theorem in a Banach space, Probability in Banach Spaces (Proc. Conf. Oberwolfach, 1975). Lecture Notes in Math. **526**, 113–130. Berlin-Heidelberg-New York: Springer 1976
33. Jain, N.C.: An example concerning CLT and LIL in Banach space. Ann. Probability **4**, 690–694 (1976b)
34. James, B.R.: A functional law of the iterated logarithm for weighted empirical distributions. Ann. Probability **3**, 762–772 (1975)
35. Kahane, J.-P.: Some random series of functions. Lexington, Mass.: D.C. Heath 1968
36. Kiefer, J.: Skorohod embedding of multivariate rv's, and the sample df. Z. Wahrscheinlichkeitstheorie verw. Geb. **24**, 1–35 (1972)
37. Kolčinski, V.I.: On the central limit theorem for empirical measures. Teor. Verojatnost. i. Mat. Statist. (Kiev) **1981**, no. 24, 63–75 (in Russian). Theory Probability and Math. Statist. no. 24, 71–82 (1981a)
38. Kolčinski, V.I.: On the law of the iterated logarithm in Strassen's form for empirical measures. Teor. Verojatnost. i. Mat. Statist. (Kiev) **1981**, no. 25, 40–47. Theory Probability and Math. Statist. no. 25, 43–50 (1981b)
39. Komlós, J., Major, P., Tusnády, G.: An approximation of partial sums of independent $RV'-s$, and the sample DF. I. Z. Wahrscheinlichkeitstheorie verw. Geb. **32**, 111–131 (1975)
40. Kuelbs, J.: A strong convergence theorem for Banach space valued random variables. Ann. Probability **4**, 744–771 (1976)
41. Kuelbs, J.: Kolmogorov's law of the iterated logarithm for Banach space valued random variables. Illinois J. Math. **21**, 784–800 (1977)
42. Kuelbs, J., Zinn, J.: Some stability results for vector valued random variables. Ann. Probability **7**, 75–84 (1979)
43. Kuelbs, J., Dudley, R.M.: Log log laws for empirical measures. Ann. Probability **8**, 405–418 (1980)
44. Kuelbs, J., Philipp, Walter: Almost sure invariance principles for partial sums of mixing B-valued random variables. Ann. Probability **8**, 1003–1036 (1980)
45. Lindenstrauss, J., Tzafriri, L.: Classical Banach Spaces I. Berlin-Heidelberg-New York: Springer 1977
46. Major, Péter: Approximation of partial sums of i.i.d. rv's when the summands have only two moments. Z. Wahrscheinlichkeitstheorie verw. Geb. **35**, 221–229 (1976)
47. Marcus, Michael, B., Philipp, Walter: Almost sure invariance principles for sums of B-valued random variables with applications to random Fourier series and the empirical characteristic process. Trans. Amer. Math. Soc. **269**, 67–90 (1982)
48. Marcus, M.B., Shepp, L.A.: Sample behavior of Gaussian processes. Proc. 6th Berkeley Sympos. Math. Statist. Probab. **2**, 423–442. Univ. Calif. Press (1971)
49. May, L.E.: Separation of functions. Canad. Math. Bull. **16**, 245–250 (1973)
50. O'Reilly, Neville E.: On the convergence of empirical processes in sup-norm metrics. Ann. Probability **2**, 642–651 (1974)
51. Parthasarathy, K.R.: Probability measures on metric spaces. New York: Academic Press 1967
52. Philipp, Walter: Almost sure invariance principles for sums of B-valued random variables, Probability in Banach Spaces II (Proc. Conf. Oberwolfach, 1978). Lecture Notes in Math. **709**, 171–193. Berlin-Heidelberg-New York: Springer 1979
53. Philipp, Walter: Weak and L^p-invariance principles for sums of B-valued random variables. Ann. Probability **8**, 68–82 (1980); Correction, ibid. (to appear)
54. Philipp, Walter, Pinzur, Laurence: Almost sure approximation theorems for the multivariate empirical process. Z. Wahrscheinlichkeitstheorie verw. Geb. **54**, 1–13 (1980)
55. Pisier, G.: Le théorème de la limite centrale et la loi du logarithme itéré dans les espaces de Banach. Séminaire Maurey-Schwartz 1975–76, Exposé IV, École Polytechnique, Paris (1975)
56a. Pollard, D.: Limit theorems for empirical processes. Z. Wahrscheinlichkeitstheorie verw. Geb. **57**, 181–195 (1981)
56b. Pollard, D.: A central limit theorem for empirical processes. [To appear in J. Australian Math. Soc. Ser. A]
57. Prohorov, Yu.V.: Convergence of random processes and limit theorems in probability theory. Theor. Probability Appl. **1**, 157–214 (1956)

58. Révész, Pál: On strong approximation of the multidimensional empirical process. Ann. Probability **4**, 729–743 (1976)
59. Senatov, V.V.: Some estimates of the rate of convergence in the multidimensional central limit theorem. Soviet Math. Doklady **22**, 462–466 = Dokl. Akad. Nauk SSSR **254** no. 4, 809–812 (1980)
60. Shortt, R.M.: Existence of laws with given marginals and specified support. Ph.D. thesis, M.I.T. (1982)
61. Skorohod, A.V.: Limit theorems for stochastic processes. Theor. Probability Appl. **1**, 261–290 (1956)
62. Strassen, V.: The existence of probability measures with given marginals. Ann. Math. Statist. **36**, 423–439 (1965)
63. Vorob'ev, N.N.: Consistent families of measures and their extensions. Theor. Probability Appl. **7**, 147–163 (1962)
64. Vulikh, B.Z.: Introduction to the theory of partially ordered spaces (transl. by L.F. Boron). Gröningen: Walters-Noordhoff 1967
65. Wenocur, R.S., Dudley, R.M.: Some special Vapnik-Chervonenkis classes. Discrete Math. **33**, 313–318 (1981)
66. Yurinskii, V.V.: On the error of the Gaussian approximation for convolutions. Theor. Probability Appl. **22**, 236–247 (1977)

Received March 24, 1982

Added in Proof

For a better proof of Theorem 8.1 and other improvements see R.M. Dudley, Ecole d'été de probabilités de St.-Flour, 1982, Theorem 3.3.1, etc., Lecture Notes in Math. (to appear).