Probab. Th. Rel. Fields 88, 381–404 (1991)

Probability Theory and Related Fields

Nearby variables with nearby conditional laws and a strong approximation theorem for Hilbert space valued martingales

Ditlev Monrad and Walter Philipp *

Departments of Mathematics and Statistics, University of Illinois, Urbana, IL 61801, USA

Received June 11, 1990

Summary. In this paper we focus on sequences of random vectors which do not admit a strong approximation of their partial sums by sums of independent random vectors. In the first part we prove conditional versions of the Strassen-Dudley theorem. We apply these in the second part of the paper to obtain strong invariance principles for vector-valued martingales which, when properly normalized, converge in law to a mixture of Gaussian distributions.

1. Introduction

The first half of the title of this paper is borrowed from the heading "Nearby variables with nearby laws," used by Dudley [4, p. 318] in his book to summarize the Strassen-Dudley theorem: If F and G are distributions on a Polish space which are close in the Prohorov metric, then these distributions can be realized on some probability space by random variables X and Y with laws $\mathscr{L}(X)=F$ and $\mathscr{L}(Y)=G$ such that X and Y are close in probability.

Combining this theorem with Lemma 2.2.2 below, we can restate it in the following form: Let X be a random variable, defined on a rich enough probability space $(\Omega, \mathscr{S}, P)$, and with values in a Polish space B. Let G be a law on B which is close to the law $\mathscr{L}(X)$ of X in the Prohorov metric. Then there exists a random variable Y defined on $(\Omega, \mathscr{S}, P)$ with law $\mathscr{L}(Y)=G$, and such that Y is close to X in probability.

It is this form of the Strassen-Dudley theorem which is most effective in proving strong approximation theorems. It will eliminate the need to use such well-known but somewhat suspicious looking phrases as: "Without changing its distribution we can redefine the sequence of random variables on a new

---

Research partially supported by the Air Force Office of Scientific Research Contract NO. F49260 85 C 0144

* Part of this work was done while the second author was visiting the Department of Statistics and the Center for Stochastic Processes, University of North Carolina, Chapel Hill. He thanks the members of the Department of Statistics for their hospitality

probability space on which there exists a Brownian motion ...," or "Without loss of generality (in the sense of Strassen) there exists a Brownian motion," etc. In other words, in these strong approximation theorems we will be able to keep the given random variables and probability space and we will construct the approximating sequence on the same probability space.

There is a natural generalization of the Strassen-Dudley theorem to regular conditional distributions. Let X, B and $(\Omega, \mathscr{S}, P)$ be as before and let $\mathscr{G}$ be a countably generated sub-σ-field of $\mathscr{S}$. Let $G(\cdot|\mathscr{G})$ be a regular conditional distribution on B, defined on $(\Omega, \mathscr{S}, P)$ and measurable with respect to $\mathscr{G}$. Suppose that with high probability the conditional law $\mathscr{L}(X|\mathscr{G})$ of X given $\mathscr{G}$ is close in the Prohorov metric to $G(\cdot|\mathscr{G})$. Then there is a random variable Y, defined on $(\Omega, \mathscr{S}, P)$, with conditional law $\mathscr{L}(Y|\mathscr{G})=G(\cdot|\mathscr{G})$ a.s., and such that Y is close to X in probability.

However, as it happens, conditional versions of the Strassen-Dudley theorem, are much more useful if they include assertions about independence: Let $\mathscr{F}$, $\mathscr{G}$ and $\mathscr{H}$ be sub-σ-fields of $\mathscr{S}$ with $\mathscr{G}$ and $\mathscr{H}$ being countably generated and $\mathscr{G} \vee \mathscr{H} \subset \mathscr{F}$. Suppose that with high probability the conditional law $\mathscr{L}(X|\mathscr{F})$ is close to $G(\cdot|\mathscr{G})$, a regular conditional distribution on B. Then there is a random variable Y, defined on $(\Omega, \mathscr{S}, P)$ which is independent of $\mathscr{H}$ given $\mathscr{G}$, has conditional law $G(\cdot|\mathscr{G})$ given $\mathscr{G}$, and is close to X in probability.

For R^d-valued random variables all these results can be rephrased in terms of characteristic functions: If g is a characteristic function on R^d which is close to the characteristic function of X, then [2, Lemma 2.2], combined with the above version of the Strassen-Dudley theorem, yields a random variable Y, defined on $(\Omega, \mathscr{S}, P)$ which is close to X in probability, and has characteristic function g. A conditional version of this result has been known to the workers in this area for a long time. For it was recognized that the proof of [2, Theorem 1] still works if there g_k is replaced by a conditional characteristic function $g_k(\cdot|\mathscr{G}_{k-1})$ where $\{\mathscr{G}_k, k \geqq 1\}$ is a sequence of countably generated σ-fields with $\mathscr{G}_k \subset \mathscr{F}_k$. However, since there were no interesting applications apparent, this seemed a rather useless generalization. As a matter of fact, in light of Remark 2.6 below, more often than not it is.

The purpose of this paper is fourfold. First, we shall recast the conditional versions of the above mentioned theorems in a form which makes them readily applicable and, moreover, which contains most of the known approximation theorems. Second, we shall discuss in some detail to what extent these results can be generalized. For example, we will give a negative answer to the following question. If, in the above notation, with high probability, $\mathscr{L}(X|\mathscr{F})$ is close to $G(\cdot|\mathscr{G})$ in the Prohorov metric for some sub-σ-fields $\mathscr{F}$ and $\mathscr{G}$, is it always possible to construct a random variable Y with conditional law $\mathscr{L}(Y|\mathscr{G})=G(\cdot|\mathscr{G})$ which is close to X in probability? In our counterexample even $\mathscr{F} \subset \mathscr{G}$ is satisfied (Remark 2.3). Third, the utility of our results will be demonstrated in a proof of a new strong approximation theorem for Hilbert space valued martingales. When properly normalized these converge in law to a mixture of Gaussian distributions. Finally, we present counterexamples to several reasonably sounding conjectures on the strong approximation of martingales. We believe that these together with our Theorem 7 bring the subject to a certain close.

The first strong approximation theorem for martingales can be found in Strassen's fundamental paper [12]. Let $\{x_n, \mathscr{L}_n, n \geqq 1\}$ be a real-valued martingale difference sequence with finite second moments. Suppose

$V_n := \sum_{k \leq n} E(x_k^2 | \mathscr{L}_{k-1}) \to \infty$ a.s. and that $\{x_n\}$ satisfies a kind of Lindeberg condition. Using the Skorohod embedding theorem Strassen [12] proved that if the underlying probability space is rich enough then the martingale can be approximated with probability one by a standard Brownian motion scaled according to the conditional variances of the given martingale sequence, i.e.

$$\sum_{n \geq 1} x_n 1\{V_n \leq t\} - B(t) = o(t^{\frac{1}{2}}) \quad \text{a.s.} \tag{1.1}$$

The utility of strong (or almost sure) invariance principles, as they are called, is clear. If the error term in this approximation is small enough then many of the properties of standard Brownian motion are shared by the given martingale sequence. For instance, (1.1) implies the functional versions of the CLT and the LIL, but for the upper and lower class integral test for the LIL an error term $O((t/\log\log t)^{\frac{1}{2}})$ is needed.

Strassen's theorem was extended in [9] to Hilbert space valued martingales satisfying a conditional Lindeberg condition slightly stronger than Strassen's. For simplicity consider an R^d-valued martingale difference sequence $\{x_k, \mathscr{L}_k, k \geq 1\}$ with conditional covariance matrices $\sigma_k = E\{x_k x_k^T | \mathscr{L}_{k-1}\}$. Set

$$A_n = \sum_{k \leq n} \sigma_k, \qquad V_n = \operatorname{trace}(A_n) = \sum_{k \leq n} E\{|x_k|^2 | \mathscr{L}_{k-1}\}.$$

In [9] (for an improvement see [11]) it is shown that if, in addition to the Lindeberg condition,

$$\frac{A_n}{V_n} \to A, \tag{1.2}$$

where A is a non-random positive semidefinite matrix, then (1.1) continues to hold. (For the precise statement of condition (1.2), see (3.1.2) below.) But here, in contrast to (1.1), $B(t)$ is an R^d-valued Brownian motion with mean zero and covariance matrix A. Of course, if $d=1$ then (1.2) is automatically satisfied with $A=1$. In [9] an example was presented to show that for $d>1$ hypothesis (1.2) cannot be dropped if (1.1) is to hold.

Still assuming $d>1$ and (1.2) we can rewrite the d-dimensional version of (1.1) in the form

$$\Big|\sum_{n \geq 1} x_n 1\{V_n \leq t\} - A^{\frac{1}{2}} \sum_{n \leq t} y_n\Big| = o(t^{\frac{1}{2}}) \quad \text{a.s.} \tag{1.3}$$

where $\{y_n, n \geq 1\}$ is a sequence of i.i.d. standard Gaussian R^d-valued random variables. In Theorem 7, Sect. 3 below, (1.3) is established under hypothesis (1.2), but weakened to allow A to be a random covariance matrix, measurable with respect to some $\mathscr{L}_k$, $k \geq 1$. In other words, we shall construct a sequence $\{y_n, n \geq 1\}$ of i.i.d. standard Gaussian R^d-valued random variables, independent of A, such that (1.3) holds. On the other hand, as we show by example in Sect. 3.3, without the assumption that A be $\mathscr{L}_k$-measurable for some finite k (1.3) need not hold.

The more general version of (1.3) with random A is still useful, because it shows, for instance, that the martingale normalized by $t^{-\frac{1}{2}}$ converges in law to a mixture of Gaussian distributions. But it also implies, via a Fubini argument, the laws of the iterated logarithm and their upper and lower class refinements.

As to the methodology we indicated above that Theorem 3 below is the basis for our method. However, one might ask in this context whether or not other established methods, such as the Skorohod embedding theorem, or rather a vector-valued version of it, could possibly be used, instead of Theorem 3, to prove strong approximation theorems for vector-valued martingales. In [8] we argued, no doubt very persuasively, that the canonical process to embed a general R^d-valued martingale in, must be an R^d-valued Gaussian process X indexed by $C \in \mathscr{C}$, the class of positive semidefinite $d \times d$ matrices, with the following properties:

(i) $X(C)$ is Gaussian with mean zero and covariance matrix C for each $C \in \mathscr{C}$,
(ii) X has independent increments, i.e., the vectors $X(C_1)$, $X(C_1+C_2) - X(C_1), \ldots, X(C_1+\ldots+C_{n-1}+C_n) - X(C_1+\ldots+C_{n-1})$ are independent for all $n \geqq 1$ for all $C_1, \ldots, C_n \in \mathscr{C}$.

After building a strong case in support of this process we showed that for $d>1$ it does not exist [8].

2. Nearby variables with nearby conditional laws

2.1. *Statement of results*

For convenience we introduce some notation. U will denote a random variable (defined on the underlying probability space) that is uniformly distributed over $[0, 1]$. Also $G(\cdot|\mathscr{G})$ will denote a regular conditional distribution, measurable with respect to the sigma-field $\mathscr{G}$ under consideration. If $G(\cdot|\mathscr{G})$ is such a distribution on $\mathscr{R}^d$ we define its conditional characteristic function as

(2.1.1) $$g(u|\mathscr{G}) = \int_{R^d} \exp(i\langle u, x\rangle)\, G(\mathrm{d}x|\mathscr{G})$$

Here $\langle u, x\rangle$ denotes the inner product of the vectors u and x.

Theorem 1 *Let X be an R^d-valued random variable defined on some probability space $(\Omega, \mathscr{S}, P)$ and let $\mathscr{G}$ be a countably generated sub-σ-field of $\mathscr{S}$. Assume that there exists a random variable U that is independent of the σ-field $\mathscr{G} \vee \sigma(X)$. (This makes the probability space rich enough.) Let $G(\cdot|\mathscr{G})$ be a regular conditional distribution on $\mathscr{R}^d$ with conditional characteristic function $g(\cdot|\mathscr{G})$ as defined in* (2.1.1). *Suppose that for some non-negative numbers λ, δ and $T > 10^8 d$,*

(2.1.2) $$\int_{|u| \leqq T} E|E\{\exp(i\langle u, X\rangle)|\mathscr{G}\} - g(u|\mathscr{G})|\mathrm{d}u \leqq \lambda(2T)^d$$

and that

(2.1.3) $$E\{G((x: |x| \geqq \tfrac{1}{2}T)|\mathscr{G})\} < \delta.$$

Then there exists an R^d-valued random variable Y on $(\Omega, \mathscr{S}, P)$, with the following properties:

$$G(\cdot|\mathscr{G}) \quad \text{is a conditional distribution of } Y \text{ given } \mathscr{G} \tag{2.1.4}$$

and

$$P(|X-Y|>\alpha) \leqq \alpha \tag{2.1.5}$$

where

$$\alpha = 16dT^{-1}\log T + 2\lambda^{\frac{1}{2}}T^d + 2\delta^{\frac{1}{2}}. \tag{2.1.6}$$

Theorem 1 is equivalent to the following theorem which is more convenient to apply.

Theorem 2 *Let $(\Omega, \mathscr{S}, P)$ be a probability space and let $\mathscr{F}$, $\mathscr{G}$, $\mathscr{H}$ and $\mathscr{L}$ be sub-σ-fields of $\mathscr{S}$ such that $\mathscr{G} \vee \mathscr{H} \subset \mathscr{F} \subset \mathscr{L}$. Assume that $\mathscr{G}$ and $\mathscr{H}$ are countably generated. Let X be an R^d-valued random variable defined on $(\Omega, \mathscr{S}, P)$ and measurable with respect to $\mathscr{L}$. Assume that there is a random variable U that is independent of $\mathscr{L}$. Let $G(\cdot|\mathscr{G})$ be a regular conditional distribution on $\mathscr{R}^d$ with conditional characteristic function $g(\cdot|\mathscr{G})$ as defined in* (2.1.1). *Suppose that for some non-negative numbers λ, δ and $T > 10^8 d$*

$$\int_{|u| \leqq T} E|E\{\exp(i\langle u, X\rangle)|\mathscr{F}\} - g(u|\mathscr{G})|\, du \leqq \lambda(2T)^d \tag{2.1.7}$$

and that (2.1.3) *holds.*

Then there exists an R^d-valued random variable Y, defined on $(\Omega, \mathscr{S}, P)$, measurable with respect to $\mathscr{L} \vee \sigma(U)$ such that (2.1.5) *and* (2.1.6) *hold and having the following property.*

(2.1.8) *$G(\cdot|\mathscr{G})$ is a conditional distribution of Y given $\mathscr{G} \vee \mathscr{H}$. In particular, Y is conditionally independent of $\mathscr{H}$ given $\mathscr{G}$.*

Remark 2.1 Theorem 2 is an easy consequence of Theorem 1. Since $\mathscr{G} \vee \mathscr{H} \subset \mathscr{F}$ we can replace in (2.1.7) $\mathscr{F}$ by $\mathscr{G} \vee \mathscr{H}$. This follows from [2, Lemma 2.6]. We reinterpret $G(\cdot|\mathscr{G})$ as $G(\cdot|\mathscr{G} \vee \mathscr{H})$. Thus we can apply Theorem 1 with $\mathscr{G} \vee \mathscr{H}$ in place of $\mathscr{G}$. Notice that $\mathscr{G} \vee \mathscr{H}$ is countably generated since $\mathscr{G}$ and $\mathscr{H}$ are. We then obtain an R^d-valued random variable Y satisfying (2.1.5) and (2.1.6) and such that $G(\cdot|\mathscr{G})$ is a conditional distribution of Y given $\mathscr{G} \vee \mathscr{H}$.

Remark 2.2 The following non-symmetric form of Theorem 1 may prove useful. Assume the hypotheses of Theorem 1 with $T>0$ (only) and let $r>0$. Then the conclusion of Theorem 1 remains valid with (2.1.5) and (2.1.6) replaced respectively by

$$P(|X-Y|>r) \leqq \alpha(r, T) + 2\lambda^{\frac{1}{2}}T^d + 2\delta^{\frac{1}{2}}$$

where

$$\alpha(r,T) \leqq \begin{cases} 3\left(\frac{4T^3}{r}\right)^{\frac{1}{2}d} \exp(-\frac{3}{16} rT), & \text{if } r \leqq T \\ 3(2T)^d \exp(-\frac{3}{16} T^2), & \text{if } r > T. \end{cases} \tag{2.1.9}$$

Theorem 2 can be reformulated in the same way. Proofs will be sketched in Sect. 2.2.3.

Remark 2.3 The condition $\mathscr{G} \subset \mathscr{F}$ cannot be omitted. In fact we shall give an example of a random variable X with characteristic function m, defined on $([0,1), \mathscr{B}, \lambda)$, and a family $\{g_\varepsilon(\cdot|\mathscr{G}), 0<\varepsilon \leqq \frac{1}{2}\}$ of conditional characteristic functions with respect to a σ-field $\mathscr{G}$ with the following properties: For all $\omega \in [0,1)$ and all $|u| \leqq \frac{1}{\varepsilon}$

$$|m(u) - g_\varepsilon(u|\mathscr{G})| \leqq \varepsilon,$$

yet any random variable Y_ε with conditional characteristic function $E(\exp(iuY_\varepsilon|\mathscr{G}) = g_\varepsilon(u|\mathscr{G})$ is bounded by 2 and satisfies

$$P(|X - Y_\varepsilon| \geqq \tfrac{1}{2}) = \tfrac{1}{2}.$$

Thus whereas conditions (2.1.3) and (2.1.7) hold, (2.1.6) does not. The example is as follows: We choose $X(\omega) = r_1(\omega)$, the first Rademacher function (recall $r_1(\omega) = 1$ for $0 \leqq \omega < \frac{1}{2}$; $r_1(\omega) = -1$ for $\frac{1}{2} \leqq \omega < 1$), $\mathscr{G} = \{\phi, [0,1), [0,\frac{1}{2}), [\frac{1}{2}, 1)\}$, $\mathscr{F} = \{\phi, [0,1)\}$, and

$$g_\varepsilon(u|\mathscr{G})_\omega = \begin{cases} \exp(i\varepsilon^2 u) \cos u & 0 \leqq \omega < \frac{1}{2} \\ \exp(-i\varepsilon^2 u) \cos u & \frac{1}{2} \leqq \omega < 1. \end{cases}$$

Now a random variable Y_ε with conditional characteristic function $g_\varepsilon(\cdot|\mathscr{G})$ must be of the form $Y_\varepsilon = \varepsilon^2 r_1 + r$ where $r \equiv 1$ on some sets A and B, say, with $A \subset [0, \frac{1}{2})$, $B \subset [\frac{1}{2}, 1)$ and $\lambda(A) = \lambda(B) = \frac{1}{4}$ and $r \equiv -1$ on $[0,1) \setminus (A \cup B)$. In other words, r is independent of r_1 (and of $\mathscr{G}$).

Remark 2.4 We conclude the discussion of Theorems 1 and 2 with the following observation. Let X and Y be random variables which are almost independent, say

$$|E e^{iuX + ivY} - E e^{iuX} E e^{ivY}| \tag{2.1.10}$$

is small for all $|u| \leqq T$, $|v| \leqq T$; suppose that X is bounded. Then according to [2, Lemma 2.2] ($\equiv$ Theorem 1 with $\mathscr{G} = (\phi, \Omega)$) there exist independent random variables X^* and Y^*, close to X and Y respectively and such that $\mathscr{L}(X^*) = \mathscr{L}(X)$, $\mathscr{L}(Y^*) = \mathscr{L}(Y)$.

Unfortunately, in general, we cannot choose $X^* = X$. In other words the following assertion is false: There exists a random variable Y^* independent of X, close to Y and with $\mathscr{L}(Y^*) = \mathscr{L}(Y)$. Let $0 < \varepsilon < 1$, let r assume the values $+1$ and -1, each with probability $\frac{1}{2}$ and let $X = \varepsilon r$ and $Y = r$. Then for all $|u| \leqq \varepsilon^{-\frac{1}{2}}$ and all v, (2.1.10) is bounded by $\varepsilon^{\frac{1}{2}}$. But any Y^* independent of X, with $\mathscr{L}(Y^*) = \mathscr{L}(Y)$, is also independent of Y and so $P(|Y - Y^*| \geqq 1) = \frac{1}{2}$.

Repeated applications of Theorem 2 yield the following result. Note that the existence of one random variable U independent of $\bigvee_{k \geqq 0} \mathscr{F}_k$ implies the existence of a whole sequence $\{U_k, k \geqq 1\}$ independent of $\bigvee_{k \geqq 0} \mathscr{F}_k$.

Theorem 3 *Let $\{X_k, k \geqq 1\}$ be a sequence of random variables with values in R^{d_k}, $k \geqq 1$, and defined on some probability space $(\Omega, \mathscr{S}, P)$. Let $\{\mathscr{F}_k, k \geqq 0\}$ be a non-decreasing sequence of sub-σ-fields of $\mathscr{S}$ such that X_k is $\mathscr{F}_k$-measurable for each $k \geqq 1$. Let $\{\mathscr{H}_k, k \geqq 1\}$ be a sequence of countably generated σ-fields with $\mathscr{H}_k \subset \mathscr{F}_k$, $k \geqq 1$, and let $\mathscr{G} \subset \mathscr{F}_0$ be a countably generated σ-field. Assume that there exists a random variable U that is independent of $\bigvee_{k \geqq 0} \mathscr{F}_k$. For each $k \geqq 1$, let $G_k(\cdot|\mathscr{G})$ be a regular conditional distribution on $\mathscr{R}^{d_k}$, measurable with respect to $\mathscr{G}$, and with conditional characteristic function*

$$g_k(u|\mathscr{G}) = \int_{R^{d_k}} \exp(\mathrm{i}\langle u, x\rangle)\, G_k(\mathrm{d}x|\mathscr{G}), \qquad u \in R^{d_k}.$$

Suppose that for some non-negative numbers λ_k, δ_k and $T_k \geqq 10^8 d_k$

$$\int_{|u| \leqq T_k} E|E\{\exp(\mathrm{i}\langle u, X_k\rangle)|\mathscr{F}_{k-1}\} - g_k(u|\mathscr{G})|\,\mathrm{d}u \leqq \lambda_k (2T_k)^{d_k}$$

and that

$$E\{G_k((x: |x| \geqq \tfrac{1}{2} T_k)|\mathscr{G})\} < \delta_k.$$

Then there exists a sequence $\{Y_k, k \geqq 1\}$ of R^{d_k}-valued random variables, defined on $(\Omega, \mathscr{S}, P)$ with the following properties:

(2.1.11) *Y_k is $\mathscr{F}_k \vee \sigma(U)$ measurable for each $k \geqq 1$,*

(2.1.12) *$G_k(\cdot|\mathscr{G})$ is a conditional distribution of Y_k given $\mathscr{G} \vee \mathscr{H}_{k-1}$, in particular, Y_k is conditionally independent of $\mathscr{H}_{k-1}$ given $\mathscr{G}$,*

and

$$P(|X_k - Y_k| \geqq \alpha_k) \leqq \alpha_k$$

where

$$\alpha_k = 16 d_k T_k^{-1} \log T_k + 2\lambda_k^{\frac{1}{2}} T_k^{d_k} + 2\delta_k^{\frac{1}{2}}, \qquad k \geqq 1.$$

In particular, if we choose inductively $\mathscr{H}_k = \sigma(Y_1, \ldots, Y_k)$, $k \geqq 1$ then $\{Y_k, k \geqq 1\}$ can be chosen to be a sequence of random variables conditionally independent given $\mathscr{G}$.

Remark 2.5 If $\mathscr{G}$ can be chosen to be the trivial σ-field then, except for the exponent $\frac{1}{2}$ on δ_k, Theorem 3 reduces to [2, Theorem 1]. In particular, $\{Y_k, k \geqq 1\}$ is a sequence of independent random variables with $\mathscr{L}(Y_k) = G_k$, $k \geqq 1$.

Remark 2.6 We want to spare the reader a complete report on the pitfalls that generalizations of Theorem 3 may have, except for this one: Let $\{\mathscr{G}_k, k \geqq 1\}$ be a sequence of countably generated σ-fields $\mathscr{G}_k \subset \mathscr{F}_k$, $k \geqq 1$. The proof of Theorem 3 still works if we assume that G_k and g_k are $\mathscr{G}_{k-1}$-measurable instead of $\mathscr{G}$-measurable. If, in addition, we set $\mathscr{H}_k := \sigma(Y_1, \ldots, Y_k)$ then the conclusion of (2.1.12) reads:

> $G_k(\cdot|\mathscr{G}_{k-1})$ is a conditional distribution of Y_k given $\sigma(Y_1, \ldots, Y_{k-1}) \vee \mathscr{G}_{k-1}$, in particular, Y_k is conditionally independent of $Y_1, \ldots, Y_{k-1}$ given $\mathscr{G}_{k-1}$, $k \geqq 1$.

Unfortunately, in general, this does not specify the joint distribution of the sequence $\{Y_k, k \geqq 1\}$, as the following example shows. In comparing this with Remark 2.5 this paradoxically seems to say that more information in fact yields less information. Let N_0, N_1, and N_2 be independent standard normal random variables, let $0<\rho<1$ and set $Y_1 := \rho N_0 + (1-\rho^2)^{\frac{1}{2}} N_1$, $Y_2 = N_0 + N_2$. Let $\mathscr{G} = \sigma(N_0)$. Then the conditional distribution $G_2(\cdot|\mathscr{G})$ of Y_2 given $\mathscr{G}$ is normal $\mathscr{N}(N_0, 1)$. The conditional distribution of Y_2 given $\mathscr{G} \vee \sigma(Y_1)$ is also $\mathscr{N}(N_0, 1)$. Thus Y_2 is conditionally independent of Y_1 given $\mathscr{G}$. Yet this does not determine the joint distribution of Y_1 and Y_2 since ρ is arbitrary.

Theorems 1, 2, and 3 apply to a wide variety of dependence structures including random variables which satisfy a strong mixing condition. In the following three theorems the dependence relation is more restrictive than the one implicit in (2.1.7), but the random variables are allowed to assume values in Polish space. For earlier versions see [10, Theorem 3.4] and its history given there. Given a Polish space (B, m), a set $A \subset B$ and $\rho > 0$ we write $A^{\rho]} = \{x : \inf\{m(x, y) : y \in A\} \leqq \rho\}$. As before, U is a random variable, defined on the underlying probability space, that is uniformly distributed over $[0, 1]$. Moreover, $G(\cdot|\mathscr{G})$ will denote a regular conditional distribution on $\mathscr{B}$, the Borel sigma-field on (B, m), such that $G(\cdot|\mathscr{G})$ is measurable with respect to the sigma-field $\mathscr{G}$ under consideration.

Theorem 4 *Let X be a random variable, defined on some probability space $(\Omega, \mathscr{S}, P)$ and with values on some Polish (B, m). Let $\mathscr{G}$ be a countably generated sub-sigma field of $\mathscr{S}$ and assume that there exists a random variable U that is independent of the σ-field $\mathscr{G} \vee \sigma(X)$. Let $G(\cdot|\mathscr{G})$ be a regular conditional distribution on $\mathscr{B}$ and suppose that for some non-negative numbers α and β*

$$E \sup_{A \in \mathscr{B}} \{P(X \in A|\mathscr{G}) - G(A^{\alpha]}|\mathscr{G})\} \leqq \beta. \tag{2.1.13}$$

Then there exists a random variable Y with values in B, defined on $(\Omega, \mathscr{S}, P)$ and satisfying (2.1.4) *and*

$$P\{m(X, Y) > \alpha\} \leqq \beta. \tag{2.1.14}$$

Remark 2.7 Notice that here as well as in the following two theorems the constants are sharp. Moreover, if in (2.1.13) $\mathscr{G}$ is the trivial σ-field then Theorem 4 reduces to the Strassen-Dudley theorem.

Theorem 4 is equivalent to the following theorem.

Theorem 5 *Let $(\Omega, \mathscr{S}, P)$ be a probability space and let $\mathscr{F}$, $\mathscr{G}$, $\mathscr{H}$ and $\mathscr{L}$ be sub-sigma-fields of $\mathscr{S}$ such that $\mathscr{G} \vee \mathscr{H} \subset \mathscr{F} \subset \mathscr{L}$. Assume that $\mathscr{G}$ and $\mathscr{H}$ are countably generated. Let X be a random variable, defined on $(\Omega, \mathscr{S}, P)$ and with values in some Polish space (B, m), and measurable with respect to $\mathscr{L}$. Assume that there* *exists a random variable U independent of $\mathscr{L}$. Moreover, let $G(\cdot|\mathscr{G})$ be a regular conditional distribution on $\mathscr{B}$ and suppose that for some non-negative numbers α and β*

$$E\sup_{A\in\mathscr{B}}\{P(X\in A|\mathscr{F})-G(A^{\alpha]}|\mathscr{G})\}\leqq\beta.$$

Then there exists a random variable Y with values in B, defined on $(\Omega,\mathscr{S},P)$, measurable with respect to $\mathscr{L}\vee\sigma(U)$ and such that (2.1.8) *and* (2.1.14) *hold.*

Repeated applications of Theorem 5 yield the following result.

Theorem 6 *Let $\{B_k, m_k, k\geqq 1\}$ be a sequence of Polish spaces, let $\mathscr{B}_k$ denote the Borel field of B_k, and let $\{X_k, k\geqq 1\}$ be a sequence of random variables, defined on $(\Omega,\mathscr{S},P)$ and with X_k assuming values in B_k. Let $\{\mathscr{F}_k, k\geqq 0\}$ be a non-decreasing sequence of sub-σ-fields of $\mathscr{S}$ such that X_k is $\mathscr{F}_k$-measurable for each $k\geqq 1$. Let $\{\mathscr{H}_k, k\geqq 1\}$ be a sequence of countably generated σ-fields with $\mathscr{H}_k\subset\mathscr{F}_k$, $k\geqq 1$, and let $\mathscr{G}\subset\mathscr{F}_0$ be a countably generated σ-field. Assume there exists a random variable U that is independent of $\bigvee_{k\geqq 0}\mathscr{F}_k$. For each $k\geqq 1$, let $G_k(\cdot|\mathscr{G})$ be a regular conditional distribution on $\mathscr{B}_k$, measurable with respect to $\mathscr{G}$. Suppose there exist two sequences of real numbers $\{\alpha_k\}$ and $\{\beta_k\}$ such that for all $k\geqq 1$*

$$E\sup_{A\in\mathscr{B}_k}\{P(X_k\in A|\mathscr{F}_{k-1})-G_k(A^{\alpha_k]}|\mathscr{G})\}\leqq\beta_k.$$

Then there exists a sequence $\{Y_k, k\geqq 1\}$ of random variables, defined on $(\Omega,\mathscr{S},P)$ and with Y_k assuming values in B_k such that (2.1.11) *and* (2.1.12) *hold. Moreover, for all $k\geqq 1$,*

$$P\{m_k(X_k, Y_k)>\alpha_k\}\leqq\beta_k.$$

In particular, if we choose $\mathscr{H}_k=\sigma(Y_1,\dots,Y_k)$ for $k\geqq 1$ then $\{Y_k, k\geqq 1\}$ can be chosen to be a sequence of random variables conditionally independent given $\mathscr{G}$.

Remark 2.8 In a recent paper [7] Eberlein embarks on a project similar to ours, namely to establish conditions for the approximation of a given sequence $\{X_k, k\geqq 1\}$ by another (possibly dependent) sequence $\{Y_k, k\geqq 1\}$ of prescribed distribution. In our view Eberlein's attempt has failed. For he imposes conditions on the sequence $\{X_k, k\geqq 1\}$ so strong that these guarantee that $\{X_k, k\geqq 1\}$ can be approximated by a sequence $\{Y_k, k\geqq 1\}$ of independent random variables, a case which is entirely in the domain of attraction of previous work (Sect. 2.4).

2.2. Proof of Theorem 1

The following lemma gives a random variable Z for which $G(\cdot|\mathscr{G})$ is the conditional distribution of Z given $\mathscr{G}$. Thus from the class of all such random variables Z we are to choose one, say Y, which in addition, satisfies (2.1.5) and (2.1.6).

Lemma 2.2.1 *Let $(\Omega,\mathscr{S},P)$ be a probability space with a sub-σ-field $\mathscr{G}\subset\mathscr{F}$. Let $G(\cdot,\omega)$ be a $\mathscr{G}$-measurable, regular conditional distribution on a Polish space S.*

Let U be a random variable uniformly distributed over [0, 1] *and independent of $\mathscr{G}$. Then there exists an S-valued random variable Z such that $G(\cdot, \omega)$ is a regular conditional distribution of Z given $\mathscr{G}$.*

Proof. Without loss of generality we can assume that $S=[0,1]$ with the usual metric and Borel structure. (See e.g. the proof of [5, Lemma 2.11].) For $0<u<1$ define

$$G^{-1}(u, \omega)=\inf\{t: G(t, \omega) \geqq u\}.$$

Then G^{-1} is jointly measurable since the map $u \rightarrow G^{-1}(u, \omega)$ is left-continuous and since for fixed u and t

$$\{\omega: G^{-1}(u, \omega) \leqq t\}=\{\omega: G(t, \omega) \geqq u\} \in \mathscr{G}.$$

The desired random variable is given by

$$Z(\omega):=G^{-1}(U(\omega), \omega).$$

We will also make extensive use of the following two lemmas.

Lemma 2.2.2 ([5, Lemma 2.11]) *Let S and T be Polish spaces and Q a law on $S \otimes T$, with marginal μ on S. Let $(\Omega, \mathscr{S}, P)$ 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 V and law $\mathscr{L}(U)$ on V having no atoms. Then there exists a random variable Y on Ω with values in T and $\mathscr{L}(\langle X, Y\rangle)=Q$.*

Lemma 2.2.3 ([2, Lemma A1] $\equiv$ [5, Lemma 2.13]) *Let X, Y and Z be Polish spaces. Suppose μ is a law on $X \otimes Y$ and ν a law on $Y \otimes Z$ such that μ and ν have the same marginal on Y. Then there is a law on $X \otimes Y \otimes Z$ with marginals μ on $X \otimes Y$ and ν on $Y \otimes Z$.*

Combining these two lemmas we obtain

Lemma 2.2.4 *Let R, S and T be Polish spaces and let ν be a law on $S \otimes T$. Let $(\Omega, \mathscr{S}, P)$ be a probability space and let X and Y be random variables with values in R and S respectively, such that $\mathscr{L}(Y)$ is the marginal of ν on S. Assume that there is a random variable U on $(\Omega, \mathscr{S}, P)$, uniformly distributed over* [0, 1] *and independent of Y. Then there exists a random variable Z on $(\Omega, \mathscr{S}, P)$ such that $\mathscr{L}(\langle Y, Z\rangle)=\nu$.*

Combining Lemma 2.2.2 with [2, Lemma 2.2] and the Strassen-Dudley theorem [4, Theorem 11.6.2] we obtain

Lemma 2.2.5 *Let X be a random variable with values in R^d and characteristic function f. Let G be a distribution on R^d with Fourier transform g. Assume there exists a random variable U independent of X. Then there exists a random variable Y with distribution G such that*

$$P(|X-Y|>\alpha) \leqq \alpha$$

where

$$\alpha=\left(\frac{T}{\pi}\right)^d \int_{|u|\leqq T} |f(u)-g(u)|\,\mathrm{d}u+G(x:|x|\geqq \tfrac{1}{2}T)+16dT^{-1}\log T$$

provided that $T>10^8 d$.

2.2.1. *The discrete case*

In this section we make heavy use of the ideas developed in [2, Sect. 2.3.1]. We first prove Theorem 1 under the additional hypothesis that $\mathscr{G}$ is generated by a countable partition.

Let $\varepsilon>0$ to be chosen suitably later. By (2.1.2), Fubini's theorem and Markov's inequality

$$\int_{|u|\leqq T} |E\{\exp(\mathrm{i}\langle u, X\rangle)|\mathscr{G}\}-g(u|\mathscr{G})|\,\mathrm{d}u\leqq\varepsilon(2T)^d \tag{2.2.1}$$

except on a set $A_1\in\mathscr{G}$ with $P(A_1)\leqq\frac{\lambda}{\varepsilon}$. Similarly, by (2.1.3)

$$G(x:|x|\geqq\tfrac{1}{2}T)\leqq\delta^{\frac{1}{2}} \tag{2.2.2}$$

except on a set $A_2\in\mathscr{G}$ with $P(A_2)<\delta^{\frac{1}{2}}$. Put $\eta=\frac{\lambda}{\varepsilon}+\delta^{\frac{1}{2}}$ and let $A=A_1\cup A_2$. Then the exceptional set $A\in\mathscr{G}$ and has probability $P(A)<\eta$.

Let D be any of the countably many atoms of $\mathscr{G}$ and keep it fixed. Let $\mathscr{S}^{(D)}$ denote the trace of $\mathscr{S}$ on D and define P_D by

$$P_D(E)=P(E|D), \qquad E\in\mathscr{S}^{(D)}. \tag{2.2.3}$$

Note that $X1_D$ and $U1_D$ are still P_D-independent and that the P_D-distribution of $U1_D$ is still uniform. On D the conditional characteristic function

$$E\{\exp(\mathrm{i}\langle u, X\rangle)|\mathscr{G}\}=\frac{1}{P(D)}\int_D \exp(\mathrm{i}\langle u, X\rangle)\,\mathrm{d}P=f(u), \qquad \text{say,}$$

is a non-random function in u and can be interpreted as the Fourier transform of the P_D-distribution of X. Similarly, on D, the conditional characteristic function $g(u|\mathscr{G})$ as well as $G(\cdot|\mathscr{G})$ are non-random. We denote them by $g(u)$ and $G(\cdot)$ respectively.

Thus, on the set D, either both (2.2.1) and (2.2.2) hold, in which case $D\subset A^C$, or else one of these two conditions fails, in which case $D\subset A$. Assume first that $D\subset A^C$. Then by (2.2.1)–(2.2.3)

$$\int_{|u|\leqq T} |f(u)-g(u)|\,\mathrm{d}u\leqq\varepsilon(2T)^d$$

and

$$G(x:|x|\geqq\tfrac{1}{2}T)\leqq\delta^{\frac{1}{2}}.$$

Hence by Lemma 2.2.5 there exists a random variable Y on $(D, \mathcal{S}^{(D)}, P_D)$ such that

(2.2.4) $$P_D(Y \in B) = G(B), \qquad B \in \mathcal{R}^d$$

and that

(2.2.5) $$P_D(|X-Y| > \beta) \leqq \beta \qquad \text{if } D \subset A^C$$

where

(2.2.6) $$\beta = 16 d T^{-1} \log T + \varepsilon T^{2d} + \delta^{\frac{1}{2}}.$$

If on the other hand $D \subset A$ we choose Y with P_D-distribution G but arbitrary otherwise. Thus

(2.2.7) $$P_D(|X-Y| > 0) \leqq 1 \qquad \text{if } D \subset A.$$

As D runs through all the atoms of $\mathcal{G}$ we obtain a random variable Y defined on the whole space $(\Omega, \mathcal{S}, P)$ such that the conditional law $\mathcal{L}(Y|\mathcal{G}) = G(\cdot|\mathcal{G})$. Moreover, summing the relations (2.2.5) and (2.2.7) over all $D \in \mathcal{G}$ we obtain by (2.2.3)

$$P(|X-Y| > \beta) \leqq \beta + \eta.$$

We choose $\varepsilon = \lambda^{\frac{1}{2}} T^{-d}$ and obtain in view of (2.2.6) a result slightly stronger than claimed in (2.1.5) and (2.1.6).

2.2.2. *The general case*

Since $\mathcal{G}$ is countably generated there exists a real-valued random variable W such that $\mathcal{G} = \sigma(W)$. For $n = 1, 2, \ldots$ let W_n denote the discrete random variable defined by

$$W_n := \sum_{-\infty < k < \infty} k 2^{-n} 1\{k 2^{-n} \leqq W < (k+1) 2^{-n}\}$$

and let $\mathcal{G}_n = \sigma(W_n)$. Let $G(\cdot|\mathcal{G}_n)$ denote the $\mathcal{G}_n$-measurable regular conditional distribution defined by

$$G(B|\mathcal{G}_n) = E\{G(B|\mathcal{G})|\mathcal{G}_n\} \qquad \text{a.s.}$$

for $B \in \mathcal{R}^d$. (Note that the verification of this as well as of several of the following claims is particularly easy if Lemma 2.2.1 is used.) Let $g(u|\mathcal{G}_n)$ denote the corresponding conditional characteristic function

$$g(u|\mathcal{G}_n) = \int_{R^d} \exp(i\langle u, x\rangle)\, G(dx|\mathcal{G}_n) = E\{g(u|\mathcal{G})|\mathcal{G}_n\}.$$

Since $\mathcal{G}_n \subset \mathcal{G}$, [2, Lemma 2.6] shows that conditions (2.2.1) and (2.2.2) are satisfied with $\mathcal{G}_n$ taking the place of $\mathcal{G}$. By the result of Sect. 2.2.1 there exists an R^d-valued random variable Y_n such that

$$G(\cdot|\mathcal{G}_n) \text{ is a conditional distribution of } Y_n \text{ given } \mathcal{G}_n$$

and

(2.2.8) $$P(|X-Y_n|>\alpha)\leqq\alpha$$

where α is given in (2.1.6).

We now show that the sequence $\{\mathscr{L}(W, Y_n), n\geqq 1\}$ of joint laws of W and Y_n converges weakly as $n\to\infty$. To see this first note that for $j=0, \pm 1, \pm 2, \ldots$, and $k=1,2,\ldots$ the events

$$\{W\in[j2^{-k},(j+1)\,2^{-k})\}\in\mathscr{G}_k.$$

Thus for each dyadic interval I_k of rank k, for all $n\geqq k$ and for all $B\in\mathscr{R}^d$

(2.2.9) $$\begin{aligned} P(W\in I_k, Y_n\in B) &= \int_{\{W\in I_k\}} G(B\,|\,\mathscr{G}_n)\,\mathrm{d}P \\ &= \int_{\{W\in I_k\}} G(B\,|\,\mathscr{G})\,\mathrm{d}P. \end{aligned}$$

This proves the claim. It follows that the sequence $\{\mathscr{L}(X, W, Y_n), n\geqq 1\}$ is a tight family of probability measures on $\mathscr{R}^{2d+1}$. Hence there exists a subsequence $\{n'\}$ such that

$$\mathscr{L}(X, W, Y_{n'})\Rightarrow Q$$

for some probability measure Q on $\mathscr{R}^{2d+1}$. Since $\mathscr{L}(X, W)$ is a marginal of Q it follows from Lemma 2.2.2 that there exists an R^d-valued random variable Y such that $\mathscr{L}(X, W, Y)=Q$. (2.2.9) implies that $G(\cdot\,|\,\mathscr{G})$ is a conditional distribution of Y given W, and (2.2.8) implies

$$P(|X-Y|>\alpha)\leqq\liminf_{n\to\infty}P(|X-Y_n|>\alpha)\leqq\alpha.$$

2.2.3. Proof of Remark 2.2 The following lemma and its proof are minor modifications of [2, Lemma 2.2] and the proof given there.

Lemma 2.2.6 *Let X be an R^d-valued random variable with distribution F and characteristic function f. Let g be a characteristic function on R^d. Moreover, suppose that there is a random variable U, uniformly distributed over* [0, 1] *and independent of X. Let r and T be positive numbers. Then there exists a (R^d-valued) random variable Y with characteristic function g such that*

$$P(|X-Y|>r)\leqq\left(\frac{T}{\pi}\right)^d\int_{|u|\leqq T}|f(u)-g(u)|\,\mathrm{d}u+F(|x|\geqq\tfrac{1}{2}T)+\alpha(r, T)$$

where $\alpha(r, T)$ is defined in (2.1.9).

Proof. We follow the proof of [2, Lemma 2.2] until [2, (2.2.4)]. If G denotes the distribution associated with g then by the argument proving [2, (2.2.1)] we obtain

$$F(B) \leqq G(B^r) + \left(\frac{T}{\pi}\right)^d \int_{|u| \leqq T} |f(u) - g(u)| \, du + F(|x| \geqq \tfrac{1}{2} T)$$
$$+ H(|x| \geqq \tfrac{1}{2} T) + H(|x| \geqq \tfrac{1}{2} r) + \left(\frac{T}{\pi}\right)^d \int_{|u| \geqq T} |h(u)| \, du$$

for all Borel sets $B \in \mathscr{R}^d$. We choose H as on [2, p. 36] with $\sigma^2 = \frac{r}{2T}$ if $r \leqq T$ and $\sigma^2 = \frac{1}{2}$ if $r > T$. We then apply the Strassen-Dudley theorem and Lemma 2.2.2 and obtain the result.

To finish the proof of Remark 2.2 we follow Sect. 2.2.1 until (2.2.4). We now apply Lemma 2.2.6 and obtain, instead of (2.2.5)

$$P(|X - Y| \geqq r, D) \leqq P(D)(T^{2d} \varepsilon + \lambda^{\frac{1}{2}} + \alpha(r, T)) \quad \text{if } D \subset A^C.$$

As in Sect. 2.2.1 we sum over all $D \in \mathscr{G}$, choose $\varepsilon = \lambda^{\frac{1}{2}} T^{-d}$ and obtain the result. The changes in Sect. 2.2.2 are minor.

2.3. Proof of Theorem 4 Again we first prove Theorem 4 under the additional hypothesis that $\mathscr{G}$ is generated by a countable partition. The proof makes use of sketches of proofs of unconditional results given in several earlier papers (see [10, Theorem 3.4] and its history given there).

Let D be any of the countably many atoms of $\mathscr{G}$ and note that on each D both

$$P(X \in A \mid \mathscr{G}) = P(X \in A \mid D) \quad \text{and} \quad G(A^{\alpha]} \mid \mathscr{G}) = G(A^{\alpha]} \mid D),$$

are non-random. Hence we can rewrite (2.1.13) in the form

$$\sum_{D \in \mathscr{G}} P(D)\, \varepsilon(D) \leqq \beta \tag{2.3.1}$$

where we set

$$\varepsilon(D) = \sup_{A \in \mathscr{B}} (P(X \in A \mid D) - G(A^{\alpha]} \mid D)). \tag{2.3.2}$$

In the context of [10, Theorem 3.4] the usefulness of this observation for obtaining sharp constants was pointed out to us by Erich Berger [1]. We thank him for this remark.

For the moment keep D fixed. We shall construct Y on each D separately. Define for all $A \in \mathscr{B}$

$$P_1(A) = P(X \in A \mid D) \quad \text{and} \quad P_2(A) = G(A \mid D)$$

Then by (2.3.2) with $\varepsilon = \varepsilon(D)$

$$P_1(A) \leqq P_2(A^{\alpha]}) + \varepsilon, \qquad \text{for all } A \in \mathscr{B}.$$

Hence by the Strassen-Dudley theorem [4, Theorem 11.6.2] there exists a probability measure $Q=Q_D$ on $B\otimes B$ with marginals P_1 and P_2 such that

$$Q_D\{(x,y): m(x,y)>\alpha\}\leqq\varepsilon.$$

Hence by Lemma 2.2.2 there exists a random variable Y on $(D,\mathscr{S}^{(D)},P_D)$ such that $\mathscr{L}(X,Y)=Q_D$, where $\mathscr{S}^{(D)}$ and P_D are defined in (2.2.3) above. It follows that

$$(2.3.3)\qquad P(m(X,Y)>\alpha, D)\leqq\varepsilon(D)\,P(D).$$

As D runs through all atoms of $\mathscr{G}$ we obtain a random variable Y defined on the whole space $(\Omega,\mathscr{S},P)$. We sum (2.3.3) over all sets D and obtain, in view of (2.3.1)

$$P(m(X,Y)>\alpha)\leqq\sum_{D\in\mathscr{G}}P(D)\,\varepsilon(D)\leqq\beta.$$

We also note that (2.1.4) holds since P_2, the second marginal of Q_D, is the P_D-distribution of Y. This proves Theorem 4 in case that $\mathscr{G}$ is generated by a countable partition.

The proof of the general case can be easily modeled after Sect. 2.2.2.

2.4. Proof of Remark 2.8 We concentrate only on one of Eberlein's results, namely on [7, Theorem 2]. We first prove the following lemma.

Lemma 2.4.1 *Let X and W be random variables defined on some probability space $(\Omega,\mathscr{S},P)$ and with values in a Polish space B. Let $\mathscr{F}$ and $\mathscr{G}$ be sub-σ-fields in $\mathscr{S}$ and assume that $\mathscr{F}$ is non-atomic. Suppose there exist two positive numbers ε and λ such that for each pair of sets $D\in\mathscr{F}$ and $E\in\mathscr{G}$ with $P(D)=P(E)$ the following relation holds:*

$$(2.4.1)\qquad P(X\in A\,|\,D)\leqq P(W\in A^{\varepsilon]}|\,E)+\lambda,\qquad \textit{for all } A\in\mathscr{B}.$$

Here $\mathscr{B}$ denotes the Borel-field of B. Then with probability one

$$(2.4.2)\qquad \sup_{A\in\mathscr{B}}\{P(W\in A\,|\,\mathscr{G})-P(W\in A^{2\varepsilon]})\}\leqq 3\lambda.$$

Proof. Let $E\in\mathscr{G}$ be any set with $\alpha:=P(E)>0$. Choose integers $n\geqq\frac{1}{\alpha\lambda}$ and $0\leqq k\leqq n$ such that $0\leqq\alpha-\frac{k}{n}\leqq\frac{1}{n}\leqq\alpha\lambda$. Partition Ω into n sets $D_1,\ldots,D_n\in\mathscr{F}$ with $P(D_j)=\frac{1}{n}$, $1\leqq j\leqq n$. For any subset M of k integers j, $1\leqq j\leqq n$ choose a set $D_M^*\in\mathscr{F}$ disjoint from $\bigcup_{j\in M}D_j$ with $P(D_M^*)=\alpha-\frac{k}{n}$. By a well-known argument (2.4.1) implies

$$P(W\in A\,|\,E)\leqq P(X\in A^{\varepsilon]}|\,D)+\lambda,\qquad \text{for all } A\in\mathscr{B}$$

and so with $D=\bigcup_{j\in M} D_j \cup D^*_M$

$$P(W\in A, E) \leqq \sum_{j\in M} P(X\in A^{\varepsilon]}, D_j)+2\alpha\lambda, \qquad A\in\mathscr{B}. \tag{2.4.3}$$

We sum (2.4.3) over all $\binom{n}{k}$ possible subsets M and obtain

$$\binom{n}{k} P(W\in A, E) \leqq \frac{k}{n}\binom{n}{k} P(X\in A^{\varepsilon]})+2\alpha\lambda\binom{n}{k}, \qquad A\in\mathscr{B}.$$

Dividing by $\alpha\binom{n}{k}$ and applying (2.4.1) with $D=E=\Omega$, and $A^{\varepsilon]}$ instead of A, we get

$$P(W\in A\,|\,E) \leqq P(W\in A^{2\varepsilon]})+3\lambda \qquad \text{for all } A\in\mathscr{B}. \tag{2.4.4}$$

Now fix $A\in\mathscr{B}$ and let $E=\{P(W\in A\,|\,\mathscr{G})-P(W\in A^{2\varepsilon]})>3\lambda\}$. Then (2.4.4) implies $P(E)=0$. Since the supremum on the LHS of (2.4.2) needs to be extended only over countably many sets $A\in\mathscr{B}$ we obtain the result.

We now recall Eberlein's [7, Theorem 2]: Let $\{B_k, m_k, k\geqq 1\}$ be a sequence of Polish spaces, let $\{X_k, k\geqq 1\}$, $\{W_k, k\geqq 1\}$ be two sequences of random variables, defined on $(\Omega, \mathscr{S}, P)$ and with X_k and W_k assuming values in B_k, $k\geqq 1$. Let $\{\mathscr{F}_k, k\geqq 1\}$ and $\{\mathscr{G}_k, k\geqq 1\}$ be two non-decreasing sequences of sub-σ-fields of $\mathscr{S}$ and assume that $\mathscr{F}_k$ is non-atomic, X_k is $\mathscr{F}_k$-measurable and W_k is $\mathscr{G}_k$-measurable for each $k\geqq 1$. Suppose there exist sequences $\{\varepsilon_k, k\geqq 1\}$, $\{\lambda_k, k\geqq 1\}$ of positive numbers such that for each pair of sets $D\in\mathscr{F}_{k-1}$, $E\in\mathscr{G}_{k-1}$ with $P(D)=P(E)$,

$$P(X_k\in A\,|\,D) \leqq P(W_k\in A^{\varepsilon_k]}\,|\,E)+\lambda_k \qquad \text{for all } A\in\mathscr{B}_k. \tag{2.4.5}$$

Here $\mathscr{B}_k$ is the Borel σ-field over B_k. Let us finally assume that there exists a random variable U, uniformly distributed over [0, 1] and independent of $\mathscr{F}_\infty \vee \mathscr{G}_\infty$. Under these assumptions Eberlein [7] proves that there exists a sequence $\{Z_k, k\geqq 1\}$ with the same law as $\{W_k, k\geqq 1\}$ such that

$$P\{m_k(X_k, Z_k)\geqq 3\varepsilon_k\} \leqq \lambda_k, \qquad k\geqq 1. \tag{2.4.6}$$

What we claim is that under these hypotheses one can do better. Namely, one can approximate $\{X_k, k\geqq 1\}$ by a sequence $\{Y_k, k\geqq 1\}$ of independent random variables with $\mathscr{L}(Y_k)=\mathscr{L}(W_k)$, for each $k\geqq 1$.

To see this note that by Lemma 2.4.1 we have with probability 1

$$P(W_k\in A\,|\,\mathscr{G}_{k-1}) \leqq P(W_k\in A^{2\varepsilon_k]})+3\lambda_k, \qquad A\in\mathscr{B}_k$$

and hence by Theorem 6 with $\mathscr{H}_{k-1}=\sigma(Y_1, \ldots, Y_{k-1})$, $k \geqq 1$ and $\mathscr{G}$, the trivial σ-field, there exists a sequence $\{Y_k^*, k \geqq 1\}$ of independent random variables with $\mathscr{L}(Y_k^*)=\mathscr{L}(W_k)$, $k \geqq 1$ such that

$$P\{m_k(Y_k^*, W_k) > 2\varepsilon_k\} \leqq 3\lambda_k, \qquad A \in \mathscr{B}_k. \tag{2.4.7}$$

As a matter of fact [10, Theorem 3.4] would in essence yield the same conclusion.

Let $B = \otimes B_k$. Consider the law $\mathscr{L}(\{W_k\}, \{Y_k^*\})$ on $B \otimes B$. Since $\mathscr{L}(\{W_k\}) = \mathscr{L}(\{Z_k\})$ we obtain from Lemma 2.2.4 a random variable $\{Y_k, k \geqq 1\}$ such that $\mathscr{L}(\{Z_k\}, \{Y_k\}) = \mathscr{L}(\{W_k\}, \{Y_k^*\})$. Hence by (2.4.6) and (2.4.7) we obtain

$$P\{m_k(X_k, Y_k) > 5\varepsilon_k\} \leqq 4\lambda_k, \qquad k \geqq 1. \tag{2.4.8}$$

Since $\mathscr{L}(\{Y_k\}) = \mathscr{L}(\{Y_k^*\})$ the sequence $\{Y_k, k \geqq 1\}$ is a sequence of independent random variables with $\mathscr{L}(Y_k) = \mathscr{L}(W_k)$, $k \geqq 1$.

We would like to add in passing that we can derive the existence of a sequence $\{Y_k, k \geqq 1\}$ of independent random variables approximating the sequence $\{X_k, k \geqq 1\}$ and satisfying (2.4.8), directly and more easily from (2.4.5) by using Theorem 6 or [10, Theorem 3.4], if we consider separately the following two cases: (1) each σ-field $\mathscr{G}_k$ is finite, (2) one σ-field $\mathscr{G}_k$ contains sets of arbitrarily small measure.

3. A strong approximation theorem for Hilbert space valued martingales

3.1. *Statement of theorem*

Let $\{x_n, \mathscr{L}_n, n \geqq 1\}$ be a square integrable martingale difference sequence defined on some probability space $(\Omega, \mathscr{S}, P)$ and with values in a real separable Hilbert space $(H, \langle \cdot, \cdot \rangle, |\cdot|)$. Suppose that $(\Omega, \mathscr{S}, P)$ supports a random variable U, uniformly distributed over [0, 1] and independent of $\{x_n, n \geqq 1\}$. We denote the conditional expectation operator $E(\cdot | \mathscr{L}_{n-1})$ by $E_n(\cdot)$. Let σ_n be the conditional covarince operator of x_n given $\mathscr{L}_{n-1}$, defined by

$$\sigma_n(u) := E_n(\langle u, x_n \rangle x_n), \qquad u \in H$$

and let

$$tr(\sigma_n) := \sum_{i \geqq 1} \langle \sigma_n(e_i), e_i \rangle = E_n |x_n|^2$$

be its trace. Here $\{e_i, i \geqq 1\}$ is a complete orthonormal basis for H. We write

$$A_n := \sum_{i \leqq n} \sigma_i$$

and put

$$V_n := tr(A_n) = \sum_{i \leqq n} E_i |x_i|^2.$$

For each $\omega\in\Omega$ let μ_ω be a mean zero measure on H such that $\int_H |x|^2\,\mu_\omega(\mathrm{d}x)<\infty$. Moreover, suppose that the map $T\colon H\times\Omega\to H$ defined by

$$T(u,\omega)=\int_H \langle u,x\rangle\, x\,\mu_\omega(\mathrm{d}x) \qquad u\in H,\ \omega\in\Omega$$

is measurable. We call T a random covariance operator. We define further a seminorm $\|\cdot\|$ on linear operators $B\colon H\to H$ by

$$\|B\|=\sup_{u\in H,\ \|u\|=1} |\langle B(u),u\rangle|$$

and observe that if T is a random covariance operator then $\|T\|$ is a random variable.

With this notation we have the following extension of [9, Theorem 1] and of [11, Theorem 2].

Theorem 7 *Let $\{x_n,\mathscr{L}_n,n\geqq 1\}$ be a square-integrable martingale difference sequence with values in a real separable Hilbert space H of dimension $d\leqq\infty$, and defined on $(\Omega,\mathscr{S},P)$. Let f be a non-decreasing function with $f(x)\to\infty$ as $x\to\infty$, and such that $\dfrac{f(x)(\log x)^\alpha}{x}$ is non-increasing for some $\alpha>50d$. (If $d=\infty$ we interpret this last condition to mean that it holds for all large α.) Suppose that $V_n\to\infty$ a.s. and that*

$$D:=\sum_{n\geqq 1} E\{|x_n|^2\, 1\{|x_n|^2>f(V_n)\}/f(V_n)\}<\infty. \tag{3.1.1}$$

Moreover, suppose that there exists some covariance operator A, measurable with respect to $\mathscr{L}_k$ for some $k\geqq 0$, and some $0<p\leqq 1$ such that

$$E\sup_{n\geqq 1}\{\|A_n-AV_n\|/f(V_n)\}^p<\infty. \tag{3.1.2}$$

Finally, let I be an arbitrary non-singular, non-random covariance operator.

Then there exists a sequence $\{y_n,n\geqq 1\}$ of i.i.d. *Gaussian H-valued random variables, defined on $(\Omega,\mathscr{S},P)$, with mean zero and covariance operator I, and independent of A such that with probability* 1

$$\Big|\sum_{n\geqq 1} x_n\, 1\{V_n\leqq t\}-(AI^{-1})^{\frac12}\sum_{m\leqq t} y_m\Big|=\begin{cases} O\big(t^{\frac12}(f(t)/t)^{p/50d}\big) & \text{if } d<\infty\\ o\big((t\log\log t)^{\frac12}\big) & \text{if } d=\infty.\end{cases}$$

Remark 3.1 For $d=1$ our result is somewhat weaker than Strassen's [12] because our class of functions is somewhat smaller. Moreover, instead of (3.1.1), Strassen only assumed the almost sure convergence of the series in (3.1.1) with $E(\cdot)$ replaced by $E(\cdot\,|\,\mathscr{L}_{n-1})$.

Remark 3.2 Collecting the probability bounds before the Borel Cantelli lemma is applied we can obtain for $d<\infty$

$$P\{\max_{s\leqq t}|\sum_{n\geqq 1} x_n 1\{V_n\leqq s\}-A^{\frac{1}{2}}\sum_{m\leqq s} y_m|\geqq t^{\frac{1}{2}}(f(t)/t)^{\frac{p}{100d}}\}\ll(f(t)/t)^{\frac{p}{100d}}.$$

Here $\{y_m, m\geqq 1\}$ is a sequence of i.i.d. $\mathcal{N}(0,I)$ random vectors independent of A, and I denotes the identity matrix.

Remark 3.3 Influenced by Lévy's proof of the CLT for martingale differences (see e.g. [3, 498–501]) one of our initial goals was to establish strong approximations of the type

(3.1.3) $$|\sum_{k\leqq n} x_k-\sum_{k\leqq n}\sigma_k^{\frac{1}{2}} y_k|=o((V_n\log\log V_n)^{\frac{1}{2}})\quad\text{a.s.}$$

or

(3.1.4) $$|\sum x_k 1\{V_k\leqq t\}-\sum\sigma_k^{\frac{1}{2}} y_k 1\{V_k\leqq t\}|=o((t\log\log t)^{\frac{1}{2}})\quad\text{a.s.}$$

where $\{y_k, k\geqq 1\}$ is a sequence of i.i.d. standard normal random variables, independent of the sequence $\{\sigma_k, k\geqq 1\}$. But even if $d=1$ neither (3.1.3) nor (3.1.4) can hold in general. To see this let $\{r_k, k\geqq 1\}$ be the sequence of Rademacher function, i.e., $P(r_k=\pm 1)=\frac{1}{2}$, $\mathcal{L}_k=\sigma(r_1,\dots,r_k)$ and $x_k=(1+r_{k-1})r_k$. Then $\sigma_k=E(x_k^2|\mathcal{L}_{k-1})=(1+r_{k-1})^2$. Write as above $V_n=\sum_{k\leqq n}\sigma_k$. If $\{y_k\}$ is independent of $\{\sigma_k\}$, then $\{y_k\}$ is also independent of $\{r_k\}$ and so

(3.1.5) $$x_k-\sigma_k^{\frac{1}{2}} y_k=\sigma_k^{\frac{1}{2}}(r_k-y_k),\qquad k\geqq 1$$

is a martingale difference sequence with respect to the natural filtration. Hence (3.1.5) satisfies the LIL with quadratic variation $2V_n$. This contradicts both (3.1.3) and (3.1.4).

Of course if $d=1$ we obtain for some i.i.d. $\mathcal{N}(0,1)$ sequence $\{y_j, j\geqq 1\}$

$$\sum x_k 1\{V_k\leqq t\}-\sum_{j\leqq t} y_j=o(t^{\frac{1}{2}-\lambda})\quad\text{a.s.}$$

by Strassen's theorem [12].

[11, Theorem 1] and, a fortiori, [6, Theorem 1] easily extend to the case of random covariance operators T.

Theorem 8 *Let $\{\xi_j, j\geqq 1\}$ be a sequence of random variables, defined on $(\Omega,\mathcal{S},P)$, with values in a real separable Hilbert space H of dimension $d\leqq\infty$ and with*

$$\sup_{j\geqq 1} E|\xi_j|^{2+\delta}<\infty$$

for some $\delta>0$. Let U be a random variable independent of $\{\xi_j, j\geqq 1\}$ and let $\{\mathcal{M}_j, j\geqq 1\}$ be a non-decreasing sequence of σ-fields such that ξ_j is $\mathcal{M}_j$-measurable for each $j\geqq 1$. Denote

$$S_n(m):=\sum_{j=m+1}^{m+n}\xi_j$$

and for $m \geqq 0$ *and* $n \geqq 1$ *define the conditional covariance operators* $C(n,m)$ *by*

$$C(n,m;u) := E\{\langle u, S_n(m)\rangle S_n(m) | \mathcal{M}_m\}, \qquad u \in H.$$

Suppose that there exist $\theta > 0$ *and* $p > 0$ *such that uniformly in* $m \geqq 0$

$$E|E(S_n(m)|\mathcal{M}_m)|^p \ll n^{(\frac{1}{2}-\theta)p}$$

and suppose that there exists a (possibly random) covariance operator T, measurable with respect to some $\mathcal{M}_j$, $j \geqq 1$ *such that uniformly in* $m \geqq 0$

$$E\|C(n,m) - nT\| \ll n^{1-\theta}.$$

Finally, let I be an arbitrary non-singular, non-random covariance operator.

Then there exists a sequence $\{y_n, n \geqq 1\}$ *of* i.i.d. *Gaussian H-valued random variables, defined on* $(\Omega, \mathcal{S}, P)$ *with mean zero, covariance operator I, and independent of T such that with probability* 1

$$\Big|\sum_{j \leqq n} \xi_j - (TI^{-1})^{\frac{1}{2}} \sum_{j \leqq n} y_j\Big| = \begin{cases} O(n^{\frac{1}{2}-\lambda}) & \text{if } d < \infty \\ o((n \log\log n)^{\frac{1}{2}}) & \text{if } d = \infty. \end{cases}$$

Here $\lambda > 0$ *is a constant depending only on d,* δ, *p and* θ.

3.2. Proof of Theorem 7 The proof of Theorem 7 follows in essence the proof of [9, Theorem 1] except that for the construction of the random variables $\{y_j, j \geqq 1\}$ Theorem 3 instead of [2, Theorem 1] will be applied. Throughout the proof we shall use the same notation as in [9], wherever possible.

We first observe that there is no loss of generality in assuming that A is $\mathcal{L}_0$-measurable.

3.2.1. *The case* $d < \infty$

Except for one minor change we follow [9, Sect. 2.1–2.3]. Starting with [9, (2.1)] we replace d by

$$d' := \frac{d}{p}.$$

This will compensate for the fact that our hypothesis (3.1.8) is weaker than the corresponding [9, (1.5)]. The changes in [9, Sect. 2.3] precipitated by this weakening of the hypothesis [9, (1.5)] have been dealt with in [11, p. 230 to p. 231, line 4]. In this context it is perhaps helpful to observe that the argument in [9, (2.11)] requires no change, because A is assumed to be $\mathcal{L}_0$-measurable. Hence [9, Proposition 1] remains valid with the appropriate interpretation of A: As $k \to \infty$

$$\text{(3.2.1)} \qquad \sup_{|u| \leqq k^2} E|E\{\exp(i\langle u, Z_k\rangle) | \mathcal{F}_{k-1}\} - \exp(-\tfrac{1}{2}\langle u, Au\rangle)| \ll k^{-5d}.$$

Remark 3.4 For the proof of (3.2.1) the hypothesis that A be $\mathscr{L}_0$-measurable is not needed. As a matter of fact the same argument shows that if $\mathscr{G}_k$ denotes the σ-field generated by $\{\sigma_j, j \leqq \tau(k)\}$ then

$$\sup_{|u| \leqq k^2} E|E\{\exp(i\langle u, Z_k\rangle)|\mathscr{F}_{k-1}\} - E\{\exp(-\tfrac{1}{2}\langle u, Au\rangle)|\mathscr{G}_{k-1}\}| \ll k^{-5d}.$$

This together with some routine calculations imply the CLT with a mixture of Gaussian distributions as limit.

We now apply Theorem 3 to the sequence $\{X_k, k \geqq 1\} = \{Z_k, k \geqq 1\}$, $T_k = k^{\frac{5}{4}}$, $\mathscr{G} = \sigma(A)$ and $g_k(u|\mathscr{G}) = \exp(-\frac{1}{2}\langle u, Au\rangle)$. We obtain sequence $\{Y_k, k \geqq 1\}$ of R^d-valued random variables, defined on $(\Omega, \mathscr{S}, P)$ with the following properties:

Conditional on A the sequence $\{Y_k, k \geqq 1\}$ is a sequence of independent random variables with (conditional) characteristic function $\exp(-\frac{1}{2}\langle u, Au\rangle)$ such that

$$P(|Z_k - Y_k| \geqq \alpha_k) \leqq \alpha_k$$

with

$$\alpha_k \ll k^{-\frac{9}{8}}.$$

Next we apply Lemma 2.2.4 with the random variables

$$X = \{Z_k, k \geqq 1\}, \qquad Y = (A, \{Y_k, k \geqq 1\})$$

and the law

$$\nu = \mathscr{L}\left(\{N_j, j \geqq 1\}; \left(A, \left\{A^{\frac{1}{2}} h_k^{-\frac{1}{2}} \sum_{j=t_{k-1}+1}^{t_k} N_j, k \geqq 1\right\}\right)\right)$$

defined on the appropriate Polish spaces $R^{d\infty}$, $R^{d^2} \otimes R^{d\infty}$ and $R^{d\infty}$. Here t_k and h_k are defined in [9, (2.1)] and $\{N_k, k \geqq 1\}$ is a sequence of i.i.d. standard Gaussian R^d-valued random variables, independent of A. Since the marginal on $R^{d^2} \otimes R^{d\infty}$ of ν,

$$\mathscr{L}(A, \{A^{\frac{1}{2}} h_k^{-\frac{1}{2}} \textstyle\sum N_j, k \geqq 1\})$$

equals $\mathscr{L}(Y)$ there exists a random variable $Z = \{y_j, j \geqq 1\}$, defined on $(\Omega, \mathscr{S}, P)$ with R^d-valued, i.i.d. standard Gaussian components y_j, independent of A such that for all $k \geqq 1$

$$P\left(\left|Z_k - A^{\frac{1}{2}} h_k^{-\frac{1}{2}} \sum_{j=t_{k-1}+1}^{t_{k+1}} y_j\right| \geqq \alpha_k\right) \leqq \alpha_k. \tag{3.2.2}$$

Summing these relations over $k = 1, \ldots, M$ we obtain the analogue of [9, (2.20)]. The remaining changes in [9, Sect. 2.4] are routine.

We note that if A is invertible the proof of (3.2.2) can be simplified, because it is easily checked that $\{A^{-\frac{1}{2}} Y_k, k \geqq 1\}$ is a sequence of i.i.d. standard Gaussian R^d-valued random variables.

3.2.2. The case $d=\infty$

The proof of Theorem 7 in the infinite-dimensional case is almost identical to the proof given in [9, Sect. 3], as ammended in [11, pp. 231–232] to take care of the weakened hypothesis (3.1.8). The idea behind the proof is this: One approximates the H-valued martingale difference sequence x_n be a finite dimensional martingale difference sequence $\pi_k x_n$ of ever increasing dimension d_k, which, by the way, will be random. One then applies the results of Sect. 3.2.1 to $\pi_k x_n$ to construct finite dimensional approximations of $\pi_k x_n$ by mixtures of Gaussian random variables. Finally, a bounded law of the iterated logarithm for $x_n - \pi_k x_n$ is proven to show that the approximation errors are negligible.

As noted above, this program has been carried out in [9, Sect. 3] and [11, pp. 231–232]. There are three items which need attention. First, from the middle of [9, p. 247] on a factor h_k^{-1} is missing. Second, in three lines in the lower half of [9, p. 247] the symbol Q_k^M erroneously got omitted. (Recall that $\|\cdot\|$ denotes the seminorm defined in [9, p. 232, line 8] and not the operator norm.) Third, for random A the estimate of III in [9, p. 248, line 4] needs proof. In other words, we need to show that for all $\omega\in\Omega$

$$\|Q_k A Q_k\| \to 0 \tag{3.2.3}$$

as $k\to\infty$. There are at least two ways to see this. By [13, p. 326, Remark] we can approximate A (for each fixed ω) in the operator norm by a finite dimensional operator on H with finite-dimensional domain. Hence (3.2.3) follows. We thank Loren Pitt for this remark. But (3.2.3) also can be proved by observing that for each fixed ω, A is the covariance operator of some square integrable vector ξ, say, and that

$$\|Q_k A Q_k\| \leqq E|Q_k \xi|^2 \to 0.$$

3.3. A counterexample

We shall show now that in Theorem 7 the hypothesis that A is $\mathscr{L}_k$-measurable for some $k\geqq 1$ cannot be omitted.

Let $([0,1), \mathscr{B}, \lambda)$ be the unit interval with Lebesgue measure and let $\{r_n, n\geqq 1\}$ be the sequence of Rademacher functions, defined on $[0,1)$. Let $\mathscr{L}_0$ be the trivial σ-field and let $\mathscr{L}_n$ be the σ-field generated by the dyadic intervals of rank $2n$. For $n\geqq 0$ define 2×2 random matrices

$$\sigma_{n+1}(\omega)=\begin{bmatrix} k2^{-n} & 0 \\ 0 & 1-k2^{-n}\end{bmatrix} \quad \text{if } k2^{-n}\leqq\omega<(k+1)\,2^{-n}, \qquad 0\leqq k<2^n.$$

Set

$$x_n=\sigma_n^{\frac{1}{2}}\begin{pmatrix} r_{2n-1} \\ r_{2n}\end{pmatrix} \qquad n\geqq 1$$

Then $\{x_n, \mathscr{L}_n, n \geqq 1\}$ is a 2-dimensional martingale difference sequence. Now

$$A_n := \sum_{k \leqq n} \sigma_k$$

has trace $V_n = \operatorname{tr} A_n = n$. Thus for all $0 \leqq \omega < 1$ the limit

$$A := \lim_{n \to \infty} A_n / V_n = \begin{bmatrix} \omega & 0 \\ 0 & 1-\omega \end{bmatrix}$$

exists. As a matter of fact both conditions (3.1.7) and (3.1.8) are satisfied with $f(x) = x^{\frac{1}{2}}$.

Suppose there exists a sequence $\{y_k, k \geqq 1\}$ of i.i.d. 2-dimensional standard Gaussian random variables, independent of A such that with probability 1

$$|\sum_{k \leqq n} x_k - A^{\frac{1}{2}} \sum_{k \leqq n} y_k| = o((n \log\log n)^{\frac{1}{2}}), \qquad n \to \infty$$

or what amounts to the same

$$|A^{-\frac{1}{2}} \sum_{k \leqq n} x_k - \sum_{k \leqq n} y_k| = o((n \log\log n)^{\frac{1}{2}}), \qquad n \to \infty.$$

Since A generates $\mathscr{B}$, the sequence $\{y_k, k \geqq 1\}$ is independent of $\{x_k, k \geqq 1\}$ and thus of $\{A^{-\frac{1}{2}} x_k, k \geqq 1\}$. Hence we have by a Fubini type argument that for some sequence $\{\alpha_n, n \geqq 1\}$ of constant 2-dimensional vectors

$$|\sum_{k \leqq n} y_k - \alpha_n| = o((n \log\log n)^{\frac{1}{2}}) \qquad \text{a.s.}$$

But this must also hold for any independent copy $\{y_k^*, k \geqq 1\}$ of $\{y_k, k \geqq 1\}$ and so

$$|\sum_{k \leqq n} (y_k - y_k^*)| = o((n \log\log n)^{\frac{1}{2}}) \qquad \text{a.s.}$$

This contradicts the classical LIL.

References

1. Berger, E.: personal communication (1982)
2. Berkes, I., Philipp, W.: Approximation for independent and weakly dependent random vectors. Ann. Probab. **7**, 29–54 (1979)
3. Billingsley, P.: Probability and measure, 2nd edn., New York: Wiley 1986
4. Dudley, R.M.: Real Analysis and Probability, Wadsworth, Belmont, California (1989)
5. Dudley, R.M., Philipp, W.: Invariance principles for sums of Banach space valued random elements and empirical processes. Z. Wahrscheinlichkeitstheor. Verw. Geb. **62**, 509–552 (1983)
6. Eberlein, E.: On strong invariance principles under dependence assumptions. Ann. Probab. **14**, 260–270 (1986)
7. Eberlein, E.: Strong approximation of continuous time stochastic processes. J. Multivariate Anal. **31**, 220–235 (1989)

8. Monrad, D., Philipp, W.: The problem of embedding vector-valued martingales in a Gaussian process. Teor. Veroyatn. Primen. **35**, 384–387 (1990)
9. Morrow, G.J., Philipp, W.: An almost sure invariance principle for Hilbert space valued martingales. Trans. Am. Math. Soc. **273**, 231–251 (1982)
10. Philipp, W.: Invariance principles for independent and weakly dependent random variables. In: Eberlein, E., Taqqu, M.S. (eds.) Dependence in probability and statistics, pp. 225–268. Boston: Birkhäuser 1986
11. Philipp, W.: A note on the almost sure approximation of weakly dependent random variables. Monatsh. Math. **102**, 227–236 (1986)
12. Strassen, V.: Almost sure behavior of sums of independent random variables and martingales. Proc. Fifth Berkeley Symp. Math. Stat. Prob. Vol. II, Part 1, pp. 315–343 (1967)
13. Yosida, K.: Functional analysis. (Grundlehren der mathematischen Wissenschaften, Bd. 123) 3rd edn., Berlin Heidelberg New York: Springer 1971