The Annals of Probability
1979, Vol. 7, No. 1, 29–54

APPROXIMATION THEOREMS FOR INDEPENDENT AND WEAKLY DEPENDENT RANDOM VECTORS

BY ISTVÁN BERKES AND WALTER PHILIPP

Hungarian Academy of Sciences and University of Illinois

In this paper we prove approximation theorems of the following type. Let $\{X_k, k \geq 1\}$ be a sequence of random variables with values in $\mathbb{R}^{d_k}$, $d_k \geq 1$ and let $\{G_k, k \geq 1\}$ be a sequence of probability distributions on $\mathbb{R}^{d_k}$ with characteristic functions g_k respectively. If for each $k \geq 1$ the conditional characteristic function of X_k given $X_1, \cdots, X_{k-1}$ is close to g_k and if G_k has small tails, then there exists a sequence of independent random variables Y_k with distribution G_k such that $|X_k - Y_k|$ is small with large probability. As an application we prove almost sure invariance principles for sums of independent identically distributed random variables with values in $\mathbb{R}^d$ and for sums of ϕ-mixing random variables with a logarithmic mixing rate.

1. Introduction. In the last decade powerful methods have been developed to prove strong limit theorems, such as the law of the iterated logarithm, upper and lower class results, etc. for martingales and partial sums of independent and weakly dependent random variables. Central to these methods are what are now called almost sure invariance principles. These are, in our context, almost sure approximation theorems for martingales or sums of weakly dependent random variables by Brownian motion. If the approximation error is sufficiently small then many of the limit theorems known and usually more easily proved for Brownian motion will continue to hold for the partial sum process under consideration.

The first to prove such almost sure invariance principles was Strassen (1964), (1965a). In his now classical papers he used the Skorohod embedding theorem to prove almost sure invariance principles for sums of independent identically distributed random variables and a little later for martingales satisfying certain second moment conditions.

Csörgö and Révész (1975) proved an almost sure invariance principle for sums of independent identically distributed random variables using a totally different approach. Their method is based on the so-called quantile transform. It was subsequently refined by Komlós, Major and Tusnády (1975) who obtained the best possible error term in the approximation of partial sums of a large class of independent identically distributed random variables.

For dependent random variables other than martingales, Philipp and Stout (1975) developed a method to prove almost sure invariance principles for sums of weakly dependent random variables. Their method depends on the observation

Received February 14, 1977.
AMS 1970 *subject classifications.* Primary 60F05; Secondary 60B10.
Key words and phrases. Approximation of weakly dependent random variables by independent ones, almost sure invariance principles, independent random vectors, mixing random variables.

that the block sums of weakly dependent random variables constitute approximately a martingale difference sequence to which the Skorohod embedding theorem can be applied. In this way they proved almost sure invariance principles for sums of strong mixing, lacunary trigonometric, Gaussian, asymptotic martingale difference sequences and for certain Markov processes.

In the case of independent random variables the quantile transform method has a clear edge over the Skorohod embedding technique since it yields sharper estimates of the error term in the approximation theorems. On the other hand, the Skorohod embedding technique is simpler and it applies via martingale approximation to a wide variety of dependence structures.

Attempts to give these methods a multidimensional setting have proved only partially successful (see Kiefer (1972)). The extension of these methods to higher dimensions would be of interest not only for its own sake but also, as was pointed out by Kiefer, because of its importance for the investigation of empirical processes.

The purpose of this paper is to develop a new approximation method which works in any number of dimensions, in some cases even for metric space valued random variables, and which, at the same time, applies directly to weakly dependent random variables. The following two theorems constitute the basis for our method. Denote by $\langle u, v \rangle$ the inner product of the vectors u and v.

THEOREM 1. *Let $\{X_k, k \geqslant 1\}$ be a sequence of random variables with values in $\mathbb{R}^{d_k}$, $d_k \geqslant 1$ and let $\{\mathcal{F}_k, k \geqslant 1\}$ be a nondecreasing sequence of σ-fields such that X_k is $\mathcal{F}_k$-measurable. Finally, let $\{G_k, k \geqslant 1\}$ be a sequence of probability distributions on $\mathbb{R}^{d_k}$ with characteristic functions $g_k(u)$, $u \in \mathbb{R}^{d_k}$, respectively. Suppose that for some nonnegative numbers λ_k, δ_k and $T_k \geqslant 10^8 d_k$*

$$E|E\{\exp(i\langle u, X_k\rangle)|\mathcal{F}_{k-1}\} - g_k(u)| \leqslant \lambda_k \tag{1.1}$$

for all u with $|u| \leqslant T_k$ and

$$G_k\{u: |u| > \tfrac{1}{4}T_k\} \leqslant \delta_k. \tag{1.2}$$

Then without changing its distribution we can redefine the sequence $\{X_k, k \geqslant 1\}$ on a richer probability space together with a sequence $\{Y_k, k \geqslant 1\}$ of independent random variables such that Y_k has distribution G_k and

$$P\{|X_k - Y_k| \geqslant \alpha_k\} \leqslant \alpha_k \qquad k = 1, 2, \cdots \tag{1.3}$$

where $\alpha_1 = 1$ and

$$\alpha_k = 16 d_k T_k^{-1} \log T_k + 4\lambda_k^{\frac{1}{2}} T_k^{d_k} + \delta_k \qquad k \geqslant 2. \tag{1.4}$$

In particular, if $\Sigma \alpha_k < \infty$ then with probability 1,

$$\Sigma_{k=1}^{\infty} |X_k - Y_k| < \infty.$$

Under a more restrictive dependence relation we obtain a sharper approximation for random variables with values in metric space.

THEOREM 2. *Let $\{(S_k, \sigma_k), k \geqslant 1\}$ be a sequence of complete separable metric spaces. Let $\{X_k, k \geqslant 1\}$ be a sequence of random variables with values in S_k and let $\{\mathfrak{L}_k, k \geqslant 1\}$ be a sequence of σ-fields such that X_k is $\mathfrak{L}_k$-measurable. Suppose that for some $\phi_k \geqslant 0$*

$$|P(AB) - P(A)P(B)| \leqslant \phi_k P(A) \tag{1.5}$$

for all $A \in \bigvee_{j<k}\mathfrak{L}_j$ and $B \in \mathfrak{L}_k$. Then without changing its distribution we can redefine the sequence $\{X_k, k \geqslant 1\}$ on a richer probability space together with a sequence $\{Y_k, k \geqslant 1\}$ of independent random variables such that Y_k has the same distribution as X_k and

$$P\{\sigma_k(X_k, Y_k) \geqslant 6\phi_k\} \leqslant 6\phi_k \qquad k = 1, 2, \cdots.$$

Another approximation theorem is proved in Section 5.

Notice that in Theorem 1 there are no assumptions on the distribution or on the moments of the random variables X_k, except for the mild tail estimate (1.2) in conjunction with (1.1) and that in Theorem 2 not even a tail estimate is required. In comparing the two theorems we observe that if the random variables X_k assume values in Euclidean space, Theorem 1 is, in general, more versatile than Theorem 2. Suppose, for instance, that (1.5) is replaced by

$$|P(AB) - P(A)P(B)| \leqslant \rho_k.$$

Then Lemma 4.4.1 implies condition (1.1) with $\lambda_k = 4\rho_k$. Consequently (1.5) is more restrictive than (1.1).

The following two theorems are applications of Theorems 1 and 2 respectively.

THEOREM 3. *Let $\{\xi_n, n \geqslant 1\}$ be a sequence of independent, identically distributed random variables with values in $\mathbb{R}^d$, centered at expectations and with finite $(2 + \delta)$th moments where $0 < \delta \leqslant 1$. Suppose that the covariance matrix Γ of ξ_1 is nonsingular. Then without changing its distribution we can redefine the sequence $\{\xi_n, n \geqslant 1\}$ on a richer probability space together with $\mathbb{R}^d$-valued Brownian motion $X(t)$ with covariance matrix Γ such that*

$$\Sigma_{n\leqslant t}\xi_n - X(t) \ll t^{\frac{1}{2}-\delta/(80d)} \quad \text{a.s.}$$

By Brownian motion with covariance matrix Γ we mean a (Gaussian) process $X(t)$ with values in $\mathbb{R}^d$, independent increments, $X(0) = 0$ such that $X(t) - X(s)$ has normal distribution with mean 0 and covariance matrix $(t - s)\Gamma$.

A sequence $\{\xi_n, n \geqslant 1\}$ is called ϕ-mixing if there exists a sequence of real numbers $\phi(n) \to 0$ such that

$$|P(AB) - P(A)P(B)| \leqslant \phi(n)P(A) \tag{1.6}$$

for all $A \in \mathfrak{M}_1^k$, $B \in \mathfrak{M}_{k+n}^\infty$ and all $k, n \geqslant 1$. Here $\mathfrak{M}_a^b$ denotes the σ-field generated by the random variables $\xi_a, \xi_{a+1}, \cdots, \xi_b$.

THEOREM 4. *Let $\{\xi_n, n \geqslant 1\}$ be a strictly stationary sequence of random variables centered at expectations with finite $(2 + \delta)$th moments for some $0 < \delta \leqslant 1$. Suppose* *that* $\{\xi_n, n \geq 1\}$ *is* ϕ*-mixing with*

$$\phi(n) \ll (\log n)^{-160/\delta} \tag{1.7}$$

and that

$$s_N^2 = E(\Sigma_{n \leq N} \xi_n)^2 \to \infty. \tag{1.8}$$

Then without changing its distribution we can redefine the sequence $\{\xi_n, n \geq 1\}$ *on a richer probability space together with standard Brownian motion* $X(t)$ *such that*

$$\Sigma_{n \leq N} \xi_n - X(a_N) \ll a_N^{\frac{1}{2}} (\log a_N)^{-\frac{1}{4}} \quad \text{a.s.} \tag{1.9}$$

Here $\{a_N, N \geq 1\}$ *is a nondecreasing sequence of positive numbers with* $a_N \sim s_N^2$.

The novel feature in Theorem 4 is the slow rate of decay of $\phi(n)$ in (1.7). All the existing strong limit theorems require that $\phi(n) \ll n^{-\tau}$ for some $\tau > 0$. See, for instance, the laws of the iterated logarithm of Reznik (1968), Heyde and Scott (1973) and the almost sure invariance principle Theorem 4.1 of Philipp and Stout (1975). Each of these results is proved under the assumption $\Sigma \phi^{\frac{1}{2}}(n) < \infty$. When specializing Theorem 4 to the case $\Sigma \phi^{\frac{1}{2}}(n) < \infty$ then we can choose $a_N = \sigma^2 N$ (for the details see Section 4.4) and thus Theorem 4 contains Theorem 4.1 of Philipp and Stout as a special case, except for the error term in (1.9). However, this error term is sharp enough to guarantee that most of the classical results on sums of ϕ-mixing random variables remain valid (for the details see Theorems A–E in Section 1 of Philipp and Stout (1975)). In particular, we obtain under the hypotheses of Theorem 4.1

$$\limsup_{N \to \infty} (2 s_N^2 \log \log s_N)^{-\frac{1}{2}} \Sigma_{n \leq N} \xi_n = 1 \quad \text{a.s.} \tag{1.10}$$

Ibragimov (1962) proved a central limit theorem under the assumption of Theorem 4, but with (1.7) replaced by $\phi(n) \to 0$. It would be interesting to know whether the law of the iterated logarithm continues to hold under this assumption.

The fact that both Theorems 3 and 4 only deal with stationary sequences is purely a matter of convenience. They both can be extended to nonstationary sequences. Moreover, it is not difficult to weaken the mixing condition (1.6) to

$$|P(AB) - P(A)P(B)| \leq \rho(n). \tag{1.11}$$

(See Remark 4.4.4 in Section 4.4.)

A third, more sophisticated application will deal with the approximation of empirical distribution functions of mixing random variables by Kiefer processes. This appeared in a separate paper (Berkes and Philipp (1977)).

2. Proof of the approximation theorems. One of the main tools used in the proofs of Theorems 1 and 2 is a theorem of Strassen (1965b) and Dudley (1968) on the existence of probability measures with given marginals. Before stating this result we give the definition of the Prohorov distance. Let λ and μ be two probability measures on the class $\mathfrak{B}$ of Borel sets in a metric space (S, σ). Put

$$A^\epsilon = \{x : \sigma(x, A) < \epsilon\}.$$

The Prohorov distance $\rho(\lambda, \mu)$ of λ and μ is defined as

$$\rho(\lambda, \mu) = \inf \{\epsilon > 0 : \lambda(A) < \mu(A^{\epsilon}) + \epsilon, \text{for all } A \in \mathfrak{B}\}.$$

By a remark of Strassen (1965) (see also Dudley (1968), Proposition 1) $\rho(\lambda, \mu) = \rho(\mu, \lambda)$.

The following lemma is a special case of the Strassen-Dudley theorem (see Dudley (1968) Theorem 1).

LEMMA 2.1. *Let S be a separable metric space and two measures P_1 and P_2 defined on $\mathfrak{B}$. Suppose that*

$$\rho(P_1, P_2) < \alpha.$$

Then there exists a probability measure Q on the Borel sets of $S \times S$ with marginals P_1 and P_2 such that

$$Q\{(x,y) : \sigma(x,y) > \alpha\} \leqslant \alpha.$$

The proofs of our two approximation theorems are similar and probably best understood from the proof of Theorem 2 under the additional assumption that all random variables X_k are discrete.

2.1. *Proof of Theorem* 2.

2.1.1. *Discrete case.* The proof of Theorem 2 is particularly simple in case that the random variables X_k are all discrete. We first observe that there is no loss of generality to assume that the σ-fields $\mathfrak{L}_k$ are all atomless. If they are not then we replace the probability space $(\Omega, \mathfrak{F}, P)$ by $(\Omega, \mathfrak{F}, P) \times ([0, 1], \mathfrak{B}, \lambda)$ (where the second factor is the unit interval with Lebesgue-measure) and redefine $\{X_k, k \geqslant 1\}$ and $\{\mathfrak{L}_k, k \geqslant 1\}$ on the new space in the obvious way. Then the new σ-fields $\mathfrak{L}_k$ are atomless and the hypotheses of Theorem 2 remain valid.

We construct the random variables Y_k inductively. Let $Y_1 = X_1$ and suppose that $Y_1, Y_2, \cdots, Y_{k-1}$ have already been constructed and satisfy the conclusions of the theorem. Moreover, suppose that $Y_j(1 \leqslant j < k)$ is $\mathfrak{F}_j$-measurable where $\mathfrak{F}_j = \bigvee_{i \leqslant j} \mathfrak{L}_i$. Since $X_j(1 \leqslant j < k)$ is discrete and has the same distribution as $Y_j(1 \leqslant j < k)$, the Y_j also is discrete. We consider the sets

$$D = D(b_1, \cdots, b_{k-1}) = \{Y_1 = b_1, \cdots, Y_{k-1} = b_{k-1}\} \tag{2.1.1}$$

where $b_1, \cdots, b_{k-1}$ is in the range of $Y_1, \cdots, Y_{k-1}$ respectively. We shall construct Y_k on each such set separately. Since the disjoint union of the events D equals Ω this will define Y_k on the whole space.

We fix D and consider the probability measures

$$P_1(A) = P\{X_k \in A | D\} \quad \text{and} \quad P_2(A) = P\{X_k \in A\} \tag{2.1.2}$$

defined on the Borel sets $A \subset S_k$. By induction hypothesis D is measurable with respect to $\mathfrak{F}_{k-1}$. Hence by (1.5)

$$|P_1(A) - P_2(A)| \leqslant \phi_k$$

for all Borel sets $A \subset S_k$. We conclude that the Prohorov distance of P_1 and P_2 does not exceed ϕ_k. By Lemma 2.1 there exists a probability measure Q on the Borel sets of $S_k \times S_k$ with marginals P_1 and P_2 such that

$$Q\{(x,y) : \sigma_k(x,y) \geqslant 2\phi_k\} \leqslant 2\phi_k. \tag{2.1.3}$$

Denote the range of X_k by $\{a_i, i \geqslant 1\}$. Then Q is concentrated on the pairs (a_i, a_j) $(1 \leqslant i, j < \infty)$. Write

$$p_{ij} = Q(a_i, a_j). \tag{2.1.4}$$

We observe that

$$P\{X_k = a_i|D\} = P_1\{a_i\} = \Sigma_{j=1}^{\infty} Q(a_i, a_j) = \Sigma_{j=1}^{\infty} p_{ij}.$$

Since $\mathfrak{F}_k$ is atomless and since $\{X_k = a_i\} \cap D$ is $\mathfrak{F}_k$-measurable we can partition the event $\{X_k = a_i\} \cap D$ into $\mathfrak{F}_k$-measurable sets $D_{ij}(j = 1, 2, \cdots)$ such that

$$P\{D_{ij}|D\} = p_{ij}. \tag{2.1.5}$$

(This is a standard property of atomless spaces. (See Exercise 23 on page 30 in Chung (1968).)

On D_{ij} we define $Y_k = a_j$ $(1 \leqslant i, j < \infty)$. In this way we obtain a random variable Y_k on D. Since by (2.1.4) and (2.1.5)

$$P\{(Y_k = a_j) \cap (X_k = a_i) \cap D|D\} = P\{D_{ij}|D\} = p_{ij} = Q(a_i, a_j)$$

we observe that Y_k and X_k have joint distribution Q on D. Hence by (2.1.3) (we write $|x - y|$ instead of $\sigma_k(x, y)$)

$$P\{|X_k - Y_k| \geqslant 2\phi_k|D\} \leqslant 2\phi_k$$

or equivalently

$$P\{|X_k - Y_k| \geqslant 2\phi_k, D\} \leqslant 2\phi_k P(D). \tag{2.1.6}$$

As D runs through all sets D of the form (2.1.1) we obtain a random variable Y_k defined on the whole space. We sum (2.1.6) over all possible D and obtain

$$P\{|X_k - Y_k| \geqslant 2\phi_k\} \leqslant 2\phi_k. \tag{2.1.7}$$

This is slightly stronger than the condition stated in the conclusion of Theorem 2. Since

$$\{Y_k = a_j\} = \cup_{i=1}^{\infty} D_{ij} \tag{2.1.8}$$

we conclude that Y_k is $\mathfrak{F}_k$-measurable. Using (2.1.8), (2.1.5) and (2.1.4) we obtain

$$\begin{aligned} P\{Y_k = a_j|D\} &= \Sigma_{i=1}^{\infty} P\{D_{ij}|D\} = \Sigma_{i=1}^{\infty} p_{ij} \\ &= P_2\{a_j\} = P\{X_k = a_j\}, \end{aligned}$$

since P_2 is the second marginal of Q. Hence

$$P\{Y_k = a_j, D\} = P\{X_k = a_j\}P(D). \tag{2.1.9}$$

We sum over all D and obtain $P(Y_k = a_j) = P(X_k = a_j)$, i.e., X_k and Y_k have the same distribution. This together with (2.1.9) shows that $Y_1, \cdots, Y_k$ are independent. This proves Theorem 2 in the discrete case.

2.1.2. *General case.* Using the separability of S_k we approximate each random variable X_k by a discrete random variable X_k^* such that X_k^* is $\mathcal{F}_k$-measurable and that

$$(2.1.10) \qquad |X_k - X_k^*| \leqslant \phi_k \qquad k = 1, 2, \cdots.$$

By the proof of Theorem 2 in the discrete case there is a sequence $\{Y_k^*, k \geqslant 1\}$ of discrete independent random variables such that Y_k^* has the same distribution as X_k^* respectively and satisfies

$$(2.1.11) \qquad P\{|X_k^* - Y_k^*| \geqslant 2\phi_k\} \leqslant 2\phi_k.$$

Let $[\xi_k, k \geqslant 1\}$ be a sequence of independent random variables having uniform distribution over [0, 1] and independent of $\{Y_k^*, k \geqslant 1\}$. (It might be necessary to enlarge the probability space to have such a sequence.) Let $\mathcal{G}_k^* = \sigma(Y_k^*, \xi_k)$. Then $\{\mathcal{G}_k^*, k \geqslant 1\}$ is a sequence of independent atomless σ-fields. Let c be in the range of Y_k^* and write $H = \{Y_k^* = c\}$. We consider the probability space $\{H, \mathcal{G}_k^*, P(\cdot|H)\}$. Now

$$(2.1.12) \qquad P_3(A) = P\{X_k \in A|H\}$$

defines a probability measure on the Borel sets $A \subset S_k$.

By Lemma 2.3 of Section 2.2 below P_3 can be realized by a random variable Y_k on $\{H, \mathcal{G}_k^*, P(\cdot|H)\}$. In other words, there exists a $\mathcal{G}_k^*$-measurable random variable Y_k on H such that

$$(2.1.13) \qquad P\{Y_k \in A|H\} = P\{X_k \in A|H\}$$

for all Borel $A \subset S_k$. As c runs through the range of Y_k^* we obtain a random variable Y_k defined on the whole space Ω.

Since Y_k is $\mathcal{G}_k^*$-measurable for $k = 1, 2, \cdots$ the random variables $Y_1, Y_2, \cdots$ are independent. Summation of (2.1.13) over all H shows that Y_k has the same distribution as X_k. In (2.1.13) H is any set of the form $\{Y_k^* = c\}$. Hence the joint distribution of Y_k and Y_k^* is the same as that of X_k and Y_k^*. Therefore

$$\begin{aligned} P\{|Y_k - Y_k^*| \geqslant 3\phi_k\} &= P\{|X_k - Y_k^*| \geqslant 3\phi_k\} \\ &\leqslant P\{|X_k^* - Y_k^*| \geqslant 2\phi_k\} \leqslant 2\phi_k \end{aligned}$$

by (2.1.10) and (2.1.11). Consequently

$$P\{|X_k - Y_k| \geqslant 6\phi_k\} \leqslant P\{|Y_k - Y_k^*| \geqslant 3\phi_k\} + P\{|X_k - Y_k^*| \geqslant 3\phi_k\} \leqslant 4\phi_k.$$

This concludes the proof of Theorem 2.

2.2. *Lemmas for the proof of Theorem* 1.

LEMMA 2.2. *Let F and G be two probability distributions on $\mathbb{R}^d$ with characteristic functions f and g respectively. Let $T \geqslant 10^8 d$. Then the Prohorov distance $\rho(F, G)$*

satisfies

$$\rho(F, G) < (T/\pi)^d \int_{|u| \leqslant T} |f(u) - g(u)| du + F\big(|x| \geqslant \tfrac{1}{2}T\big) + 16dT^{-1} \log T.$$

PROOF. We use a simple smoothing inequality of the form

$$\rho(F, G) \leqslant \rho(F * H, G * H) + 2 \max\{r, H(|x| \geqslant r)\} \tag{2.2.1}$$

valid for any three distributions F, G, H on a metric space and for any $r > 0$. (2.2.1) follows easily from the definition of the Prohorov distance.

Let H be a distribution on $\mathbb{R}^d$ with density $\nu(x)$ and characteristic function $h(u) \in L^1$. We shall prove a general estimate for $\rho(F, G)$ in terms of H and specialize later to obtain the desired bound. Write $F_1 = F * H$ and $G_1 = G * H$. Then the characteristic functions of F_1 and G_1 respectively are given by $f_1 = fh$ and $g_1 = gh$. Both belong to L^1. Hence we obtain for the densities ϕ and γ of F_1 and G_1 respectively

$$\begin{aligned} |\phi(x) - \gamma(x)| &= (2\pi)^{-d} \Big|\int_{\mathbb{R}^d} e^{-i\langle u, x\rangle}(f_1(u) - g_1(u))\, du\Big| \\ &\leqslant (2\pi)^{-d} \int_{\mathbb{R}^d} |f(u) - g(u)||h(u)|\, du \\ &\leqslant (2\pi)^{-d}\Big(I + 2\int_{|u| \geqslant T} |h(u)|\, du\Big) \end{aligned} \tag{2.2.2}$$

for all $x \in \mathbb{R}^d$. Here

$$I = \int_{|u| \leqslant T} |f(u) - g(u)|\, du. \tag{2.2.3}$$

Hence for any Borel set B

$$\begin{aligned} F_1(B) - G_1(B) & \\ &\leqslant F_1(B \cap \{|x| \leqslant T\}) - G_1(B \cap \{|x| \leqslant T\}) + F_1(|x| \geqslant T) \\ &\leqslant \int_{|x| \leqslant T} |\phi(x) - \gamma(x)| dx + F\big(|x| \geqslant \tfrac{1}{2}T\big) + H\big(|x| \geqslant \tfrac{1}{2}T\big) \\ &\leqslant (T/\pi)^d\Big(I + 2\int_{|u| \geqslant T} |h(u)|\, du\Big) + F\big(|x| \geqslant \tfrac{1}{2}T\big) + H\big(|x| \geqslant \tfrac{1}{2}T\big) \end{aligned} \tag{2.2.4}$$

by (2.2.2). We note that the right-hand side of (2.2.4) is an upper bound for $\rho(F_1, G_1)$. Hence, if $r \leqslant \frac{1}{2}T$ we obtain from (2.2.1)

$$\begin{aligned} \rho(F, G) \leqslant (T/\pi)^d\Big(I + 2\int_{|u| \geqslant T} |h(u)| du\Big) + F\big(|x| \geqslant \tfrac{1}{2}T\big) & \\ + 3 \max\{r, H(|x| \geqslant r)\}. & \end{aligned} \tag{2.2.5}$$

It remains to choose H and r suitably. Put

$$\sigma = 3d^{\frac{1}{2}} T^{-1} \log^{\frac{1}{2}} T \quad \text{and} \quad r = 5dT^{-1} \log T$$

and let H be a normal distribution with density

$$\nu(x) = (2\pi\sigma^2)^{-\frac{1}{2}d} \exp\Big(-\tfrac{1}{2}\sigma^{-2}\Sigma_{j \leqslant d} x_j^2\Big).$$

Then the characteristic function is given by

$$h(u) = \exp\Big(-\tfrac{1}{2}\sigma^2\Sigma_{j \leqslant d} u_j^2\Big).$$

The lemma follows now from (2.2.3), (2.2.5) and the estimate

(2.2.6) $$(2\pi)^{-\frac{1}{2}d}\int_{|u|\geqslant A}\exp\left(-\tfrac{1}{2}\Sigma_{j\leqslant d}u_j^2\right)du = P\{\chi_d^2 \geqslant A^2\}$$
$$\leqslant e^{-tA^2}E\{\exp(t\chi_d^2)\}|_{t=3/8} = e^{-3A^2/8}2^d$$

via elementary calculations.

LEMMA 2.3. *Let (S, σ) be a complete separable metric space and let μ be a probability measure on the Borel sets of S. Let $(\Omega, \mathfrak{F}, P)$ be a probability space such that $\mathfrak{F}$ is atomless. Then there exists a random variable X on $(\Omega, \mathfrak{F}, P)$ with state space S and distribution μ.*

PROOF. The separability of S implies that for any $\delta > 0$ there is a partition $S = \cup_{i=1}^{\infty}A_i$ where A_i is a Borel set with diameter less than δ. Since this is also true for any Borel subset of S we can find for each $k \geqslant 1$ a partition $\{A_{ik}, i = 1, 2, \cdots\}$ of S such that the kth partition is a refinement of the $(k-1)$th one and such that for each k the diameter of the Borel sets A_{ik} is less than k^{-2} for $i = 1, 2, \cdots$. Since $\mathfrak{F}$ has no atoms we can find for any $C \in \mathfrak{F}$ and any sequence $\{p_i, i \geqslant 1\}$ of nonnegative numbers with $P(C) = \Sigma_{i=1}^{\infty}p_i$ a partition $C = \cup_{i=1}^{\infty}C_i$ with $P(C_i) = p_i$. (See the remark following (2.1.5).) Hence for the sequence $\{A_{ik}, i = 1, 2, \cdots\}_{k=1}^{\infty}$ of partitions of S there corresponds a sequence $\{B_{ik}, i = 1, 2, \cdots\}_{k=1}^{\infty}$ of partitions of Ω such that

(2.2.7) $$P(B_{ik}) = \mu(A_{ik})$$

and such that $\{B_{i,k+1}, i = 1, 2, \cdots\}$ is a refinement of $\{B_{ik}, i = 1, 2, \cdots\}$ for each $k = 1, 2, \cdots$. For each pair (i, k) we choose a point $x_{ik} \in A_{ik}$. We now define a sequence $\{X_k, k \geqslant 1\}$ of random variables on $(\Omega, \mathfrak{F}, P)$ by

(2.2.8) $$X_k(\omega) = x_{ik} \quad \text{if} \quad \omega \in B_{ik}.$$

By the construction of the sets A_{ik} and B_{ik} we conclude that

$$\sigma(X_{k+1}(\omega), X_k(\omega)) \leqslant k^{-2} \quad \text{for all} \quad \omega \in \Omega.$$

Consequently $\{X_k, k \geqslant 1\}$ converges to a random variable X, uniformly on Ω. This implies

(2.2.9) $$\sigma(X_k, X) \leqslant 2/k.$$

In what follows we shall show that X has the desired property. Let μ_k and μ^* be the distributions of X_k and X respectively. Then for each Borel set $A \subset S$,

$$\mu_k(A) = P\{X_k \in A\} \leqslant P\{X \in A^{2/k}\} = \mu^*(A^{2/k})$$

by (2.2.9). Hence we obtain for the Prohorov distance

(2.2.10) $$\rho(\mu_k, \mu^*) \leqslant 2/k.$$

On the other hand by (2.2.7)—(2.2.9), since μ_k is concentrated on x_{ik} and since A_{ik}

has diameter less than k^{-2},

$$\mu_k(A) = \Sigma_{x_{ik} \in A} \mu_k(x_{ik}) = \Sigma_{x_{ik} \in A} P(B_{ik}) = \Sigma_{x_{ik} \in A} \mu(A_{ik})$$
$$\leqslant \Sigma_i \mu(A^{1/k^2} \cap A_{ik}) = \mu(A^{1/k^2}).$$

Hence

$$\rho(\mu_k, \mu) \leqslant 1/k^2.$$

This together with (2.2.10) shows that $\mu^* = \mu$ and that thus X has distribution μ.

LEMMA 2.4. *Let $(\Omega, \mathcal{F}, P)$ be a probability space such that $\mathcal{F}$ is atomless and let S be a complete separable metric space. Let X be a discrete random variable on $(\Omega, \mathcal{F}, P)$ with state space S. Moreover, let Q be a probability measure on the Borel sets of $S \times S$ such that the distribution X is a marginal distribution of Q. Then there exists a random variable Y on $(\Omega, \mathcal{F}, P)$ with state space S such that X and Y have joint distribution Q.*

PROOF. This is an easy consequence of Lemma 2.3. Let $\{x_i, i \geqslant 1\}$ be the range of X. For fixed i we define a probability measure

$$Q_i(A) = \frac{Q(\{x_i\}, A)}{P\{X = x_i\}} \tag{2.2.11}$$

on the Borel sets $A \subset S$. We consider the probability space $\{(X = x_i), \mathcal{F} \cap (X = x_i), P(\cdot|X = x_i)\}$. By Lemma 2.3 there exists a random variable Y on this space such that

$$P\{Y \in A | X = x_i\} = Q_i(A). \tag{2.2.12}$$

As x_i runs through the range of X we obtain a random variable Y defined on the whole space $(\Omega, \mathcal{F}, P)$ and satisfying (2.2.12). By (2.2.11) Y does what is required.

LEMMA 2.5. *Let $\mathcal{L}_0$, $\mathcal{L}_1$ and $\mathcal{L}_2$ be σ-fields such that $\mathcal{L}_0 \vee \mathcal{L}_1$ is independent of $\mathcal{L}_2$. Then for each integrable random vector X with state space $\mathbb{R}^d$ and measurable with respect to $\mathcal{L}_0$,*

$$E\{X|\mathcal{L}_1 \vee \mathcal{L}_2\} = E\{X|\mathcal{L}_1\} \quad \text{a.s.}$$

PROOF. See Chung (1968), page 285.

The following lemma is well known and easy to prove.

LEMMA 2.6. *Let U be a random variable with finite expectation and let $\mathcal{F} \supset \mathcal{G}$ be two σ-fields. Then*

$$E|E(U|\mathcal{F})| \geqslant E|E(U|\mathcal{G})|.$$

2.3. *Proof of Theorem* 1. As in the proof of Theorem 2 we observe that there is no loss of generality in assuming that the σ-fields $\mathcal{F}_k$ are atomless. Indeed, we can enrich the probability space $(\Omega, \mathcal{F}, P)$ by adjoining a sequence $\{\xi_n, n \geqslant 1\}$ of independent uniformly distributed random variables such that this sequence is independent of $\bigvee_1^\infty \mathcal{F}_{n+1}$. Let $\mathcal{F}_k' = \sigma(\mathcal{F}_k, \xi_k)$. Then $\mathcal{F}_k'$ is atomless and by Lemma 2.5 all the hypotheses of Theorem 1 remain valid if we replace $\mathcal{F}_k$ by $\mathcal{F}_k'$.

2.3.1. *The totally discrete case.* We first prove Theorem 1 under the additional hypothesis that all X_k's and all G_k's are discrete. We apply Lemma 2.3 to $(\Omega, \mathcal{F}_1, P)$ and G_1 and obtain an $\mathcal{F}_1$-measurable random variable Y_1 with distribution G_1. Since $\alpha_1 = 1$ condition (1.3) is satisfied. We now assume that $Y_j (j < k)$ have already been constructed and that, in addition to the desired properties, they are $\mathcal{F}_j$-measurable. We consider the sets

(2.3.1) $$D = D(b_1, \cdots, b_{k-1}) = \{Y_1 = b_1, \cdots, Y_{k-1} = b_{k-1}\}$$

where b_i is in the range of Y_i $(1 \leqslant i < k)$. We construct Y_k on each set D separately. Since the disjoint union of all possible events D equals Ω this will then define Y_k on the whole space $(\Omega, \mathcal{F}, P)$.

We fix D and consider the probability measure F_k on $\mathbb{R}^{d_k}$ defined by

(2.3.2) $$F_k(B) = P\{X_k \in B | D\} = P_D(X_k \in B)$$

for all Borel sets B of $\mathbb{R}^{d_k}$. We now shall estimate the Prohorov distance $\rho(F_k, G_k)$ using Lemma 2.2. The characteristic function f_k of F_k is given by

(2.3.3) $$f_k(u) = \frac{1}{P(D)} \int_D \exp(i\langle u, X_k\rangle) dP.$$

Let $\mathcal{G}_{k-1}$ be the σ-field generated by $Y_1, \cdots, Y_{k-1}$. Since $\mathcal{G}_{k-1} \subset \mathcal{F}_{k-1}$ we obtain using Lemma 2.6 and (1.1)

$$E|E\{\exp(i\langle u, X_k\rangle)|\mathcal{G}_{k-1}\} - g_k(u)| \leqslant \lambda_k$$

for all $|u| \leqslant T_k$. Hence

(2.3.4) $$E\{\int_{|u| \leqslant T_k} |E\{\exp(i\langle u, X_k\rangle)|\mathcal{G}_{k-1}\} - g_k(u)| du\} \leqslant \lambda_k (2T_k)^{d_k}.$$

Let $\epsilon_k > 0$ be arbitrary to be chosen suitably later. Put

(2.3.5) $$\eta_k = \lambda_k / \epsilon_k.$$

Then by Markov's inequality

(2.3.6) $$\int_{|u| \leqslant T_k} |E\{\exp(i\langle u, X_k\rangle)|\mathcal{G}_{k-1}\} - g_k(u)| du \leqslant \epsilon_k (2T_k)^{d_k}$$

except on a set A_k with

(2.3.7) $$P(A_k) \leqslant \eta_k.$$

In the following we disregard sets of measure zero. Since D is an atom of $\mathcal{G}_{k-1}$ we conclude that on the set D the conditional characteristic function $E\{\exp(i\langle u, X_k\rangle)|\mathcal{G}_{k-1}\}$ is a nonrandom function in u and equals $f_k(u)$ by (2.3.3). Thus (2.3.6) either holds for all $\omega \in D$ or for none of them. Consequently, either $D \subset A_k$ or $D \subset A_k^c$. Let us first assume $D \subset A_k^c$. Then by (2.3.6)

$$\int_{|u| \leqslant T_k} |f_k(u) - g_k(u)| du \leqslant \epsilon_k (2T_k)^{d_k}.$$

Hence by Lemma 2.2

(2.3.8) $$\rho(F_k, G_k) < \varepsilon_k T_k^{2d_k} + \delta_k + 16 d_k T_k^{-1} \log T_k = \beta_k \quad \text{(say)}.$$

(Note that at this point we do not use the full strength of (1.2).)

If on the other hand $D \subset A_k$ then we trivially have $\rho(F_k, G_k) \leqslant 1$. Hence

(2.3.9) $$\rho(F_k, G_k) < \beta_k'$$

where

(2.3.10) $$\begin{aligned} \beta_k' &= \beta_k \quad \text{if} \quad D \subset A_k^c \\ &= 2 \quad \text{if} \quad D \subset A_k. \end{aligned}$$

By the Strassen-Dudley theorem (see Lemma 2.1) there exists a probability distribution $Q = Q_D$ on $\mathbb{R}^{d_k} \times \mathbb{R}^{d_k}$ with marginals F_k and G_k such that

(2.3.11) $$Q\{(x, y) : |x - y| \geqslant \beta_k'\} \leqslant \beta_k'.$$

We now apply Lemma 2.4 to the probability space $(D, \mathcal{F}_k^{(D)}, P_D)$. By (2.3.2) the distribution of X_k on this space is F_k, a marginal of Q. Hence there exists an $\mathcal{F}_k^{(D)}$-measurable random variable Y_k such that X_k and Y_k have joint distribution Q. In particular, by (2.3.11)

$$P_D\{|X_k - Y_k| \geqslant \beta_k'\} \leqslant \beta_k'$$

or

(2.3.12) $$P\{|X_k - Y_k| \geqslant \beta_k', D\} \leqslant P(D)\beta_k'.$$

Moreover, since the second marginal of Q is G_k, we conclude that Y_k has distribution G_k on D, i.e.,

(2.3.13) $$G_k(B) = P_D(Y_k \in B) = P(Y_k \in B | D)$$

for all Borel sets B of $\mathbb{R}^{d_k}$. As D runs through all the sets of the form (2.3.1) we obtain an $\mathcal{F}_k$-measurable random variable Y_k defined on the whole space. (2.3.13) implies that Y_k has distribution G_k and that Y_k is independent of $Y_1, \cdots, Y_{k-1}$. We sum (2.3.12) over all events $D \subset A_k^c$ and obtain, using (2.3.7) and (2.3.10),

$$P\{|X_k - Y_k| \geqslant \beta_k\} \leqslant \beta_k + \eta_k.$$

We choose $\epsilon_k = \lambda_k^{\frac{1}{2}} T_k^{-d_k}$ and obtain a result slightly stronger than (1.4), namely with (1.4) replaced by

(2.3.14) $$\alpha_k = 16 d_k T_k^{-1} \log T_k + 3\lambda_k^{\frac{1}{2}} T_k^{d_k} + \delta_k.$$

This concludes the proof of Theorem 1 in the totally discrete case.

2.3.2. *The general case.* For each X_k there is a discrete $\mathcal{F}_k$-measurable random variable $X_k^* \in \mathbb{R}^{d_k}$ such that

(2.3.15) $$|X_k - X_k^*| \leqslant \lambda_k T_k^{-1}.$$

Then for all u with $|u| \leqslant T_k$

(2.3.16) $$\begin{aligned} &|E\{\exp(i\langle u, X_k\rangle)|\mathcal{F}_{k-1}\} - E\{\exp(i\langle u, X_k^*\rangle)|\mathcal{F}_{k-1}\}| \\ &\leqslant E\{|\exp(i\langle u, X_k - X_k^*\rangle) - 1||\mathcal{F}_{k-1}\} \leqslant T_k\lambda_k T_k^{-1} = \lambda_k. \end{aligned}$$

We now approximate G_k by a discrete distribution function G_k^* with characteristic function g_k^* such that the following three conditions are satisfied.

(2.3.17) $$\rho(G_k^*, G_k) < \lambda_k T_k^{-1},$$

(2.3.18) $$|g_k^*(u) - g_k(u)| \leqslant \lambda_k \quad \text{for all} \quad |u| \leqslant T_k,$$

(2.3.19) $$G_k^*\left(|u| \geqslant \tfrac{1}{2} T_k\right) \leqslant \delta_k.$$

This can be easily achieved by choosing a random variable Z_k with distribution G_k. We then pick a discrete random variable Z_k^* such that $|Z_k - Z_k^*| \leqslant \lambda_k T_k^{-1}$. Then (2.3.17) follows from the definition of the Prohorov distance and (2.3.18) follows from the argument in (2.3.16). Finally (2.3.19) follows from (1.2) since $T_k \geqslant 10$ and thus $\lambda_k T_k^{-1} < \frac{1}{4} T_k$. Hence by (2.3.16), (2.3.18) and (1.1) we have for all u with $|u| \leqslant T_k$

$$E|E\{\exp(i\langle u, X_k^*\rangle)|\mathcal{F}_{k-1}\} - g_k^*(u)| \leqslant 3\lambda_k.$$

Then by (2.3.19) and by the totally discrete case of Theorem 1 proved in Section 2.3.1 there exists a sequence $\{Y_k^*, k \geqslant 1\}$ of independent, $\mathcal{F}_k$-measurable random variables Y_k^* with distribution G_k^* such that

(2.3.20) $$P\{|X_k^* - Y_k^*| \geqslant \alpha_k^*\} \leqslant \alpha_k^*$$

with $\alpha_1^* = 1$ and

$$\alpha_k^* = 16 d_k T_k^{-1} \log T_k + 3\lambda_k^{\frac{1}{2}} T_k^{d_k} + \delta_k \qquad k \geqslant 2.$$

(See (2.3.14). Recall that we did not use the full strength of (1.2) in the totally discrete case.)

Now let $\{\xi_k, k \geqslant 1\}$ be a sequence of independent random variables, uniformly distributed over [0, 1] and let $\mathcal{G}_k^* = \sigma(Y_k^*, \xi_k)$. Then $\{\mathcal{G}_k^*, k \geqslant 1\}$ is a sequence of independent atomless σ-fields. We enrich the probability space $(\Omega, \mathcal{F}, P)$ by adjoining the sequence $\{\xi_k, k \geqslant 1\}$ and dènote this new space also by $(\Omega, \mathcal{F}, P)$. We redefine the sequences $\{X_k, k \geqslant 1\}$, $\{X_k^*, k \geqslant 1\}$ and $\{Y_k^*, k \geqslant 1\}$ without changing their joint distribution.

By (2.3.17) and the Strassen-Dudley theorem (see Lemma 2.1) there exists a probability measure Q_k on $\mathbb{R}^{d_k} \times \mathbb{R}^{d_k}$ with marginals G_k and G_k^* such that

(2.3.21) $$Q_k\left\{(x, y) : |x - y| \geqslant \lambda_k T_k^{-1}\right\} \leqslant \lambda_k T_k^{-1}.$$

We apply Lemma 2.4 to the (enriched) space $(\Omega, \mathcal{G}_k^*, P)$. The distribution G_k^* of Y_k^* is a marginal of Q_k. Hence there exists a random variable Y_k on (Ω, G_k^*, P) such that Y_k^* and Y_k have joint distribution Q_k. In particular, by (2.3.21)

(2.3.22) $$P\left\{|Y_k - Y_k^*| \geqslant \lambda_k T_k^{-1}\right\} \leqslant \lambda_k T_k^{-1}.$$

Now Y_k has distribution G_k, the second marginal of Q_k. Since Y_k is $\mathcal{G}_k^*$-measurable the sequence $\{Y_k, k \geqslant 1\}$ is a sequence of independent random variables. Finally, (1.4) follows from (2.3.15), (2.3.20) and (2.3.22).

3. Proof of Theorem 3. The proof of Theorem 3 is based on several well-known facts. Let $f_N(u)$ be the characteristic function of the normalized sum $N^{-\frac{1}{2}}\Sigma_{n \leqslant N}\xi_n$.

LEMMA 3.1 (von Bahr (1967)). *Under the hypotheses of Theorem* 3

$$|f_N(u) - \exp(-\tfrac{1}{2}\langle u, \Gamma u\rangle)| \leqslant C_1 N^{-\frac{1}{2}\delta}|u|^{2+\delta}\exp(-\gamma u^2) \tag{3.1}$$

for all $u \in \mathbb{R}^d$ *with* $|u| \leqslant C_2 N^{\frac{1}{2}}$. *Here the constants* C_1, C_2 *and* $\gamma > 0$ *only depend on* d *and on the moments of* ξ_1.

We now define blocks H_k of k^α consecutive integers where

$$\alpha = [18d/\delta] + 1. \tag{3.2}$$

We leave no gaps between the blocks. For given $N \geqslant 1$ define $M = M_N$ by $N \in H_M$. Then

$$M^{1+\alpha} \ll \Sigma_{k \leqslant M} k^\alpha \ll N \ll M^{1+\alpha}. \tag{3.3}$$

Let h_M be the smallest number of H_M.

LEMMA 3.2. *Under the hypotheses of Theorem* 3

$$\max_{h_M \leqslant N < h_{M+1}} |\Sigma_{\nu=h_M}^N \xi_\nu| \ll h_M^{\frac{1}{2}-\frac{1}{4}\delta/(18d+\delta)} \qquad \text{a.s.}$$

PROOF. Write $\xi_\nu = (\xi_{\nu 1}, \cdots, \xi_{\nu d})$. We recall that for fixed $1 \leqslant j \leqslant d$ the normalized sum $M^{-\frac{1}{2}\alpha}\Sigma_{h_M \leqslant \nu < h_{M+1}} \xi_{\nu j}$ satisfies a central limit theorem with remainder $\ll M^{-\frac{1}{2}\alpha\delta}$ since $\{\xi_{\nu j}, \nu \geqslant 1\}$ is a sequence of independent, identically distributed random variables centered at expectations and finite $(2+\delta)$th moments. Let $\lambda = (4(1+\alpha))^{-1}$. Then by Lévy's maximal inequality (see Loève (1963), page 248), by (3.2), (3.3) and by stationarity

$$\begin{aligned} P\Big\{\max_{h_M \leqslant N < h_{M+1}} |\Sigma_{\nu=h_M}^N \xi_\nu| \geqslant h_M^{\frac{1}{2}-\lambda}\Big\} & \\ &\leqslant \Sigma_{j \leqslant d} P\Big\{\max_{h_M \leqslant N < h_{M+1}} |\Sigma_{\nu=h_M}^N \xi_{\nu j}| \geqslant d^{-\frac{1}{2}} h_M^{\frac{1}{2}-\lambda}\Big\} \\ &\leqslant \Sigma_{j \leqslant d} P\Big\{|\Sigma_{h_M \leqslant N < h_{M+1}} \xi_{\nu j}| \geqslant d^{-\frac{1}{2}} h_M^{\frac{1}{2}-\lambda} - M^{\frac{1}{2}\alpha}(2E\xi_1^2)^{\frac{1}{2}}\Big\} \\ &\ll \exp(-cd^{-1}h_M^{1-2\lambda}M^{-\alpha}) + M^{-\frac{1}{2}\alpha\delta} \ll M^{-9d}, \end{aligned}$$

where $c > 0$ depends only on Γ. The lemma follows now from the Borel-Cantelli lemma. □

Put

$$X_k = k^{-\frac{1}{2}\alpha}\Sigma_{\nu \in H_k} \xi_\nu \tag{3.4}$$

and let $\mathfrak{F}_k$ be the σ-field generated by $X_1, \cdots, X_k$. Then with probability 1

$$E\{\exp(i\langle u, X_k\rangle | \mathfrak{F}_{k-1})\} = E\{\exp(i\langle u, X_k\rangle)\} \tag{3.5}$$

for all $u \in \mathbb{R}^d$. From (3.1) we obtain for $k \geqslant k_0$ and all u with $|u| \leqslant C_2 k^{\frac{1}{2}\alpha}$

$$E\{\exp(i\langle u, X_k\rangle)\} - \exp(-\tfrac{1}{2}\langle u, \Gamma u\rangle) \ll k^{-\frac{1}{2}\alpha\delta} \ll k^{-9d}.$$

This together with (3.5) shows that condition (1.1) is satisfied with $\lambda_k = \text{const.}\ k^{-9d}$ and $T_k = \text{const.}\ k^{9/4}$.

We now show that condition (1.2) is satisfied. Since Γ is positive definite all eigenvalues $\rho_1, \cdots, \rho_d$ of Γ are positive. We write $\Gamma = U\Delta U^{-1}$ where U is a unitary and Δ is a diagonal matrix. We define $\Delta^{-\frac{1}{2}}$ in the obvious way and make the substitution $t = \Delta^{-\frac{1}{2}}U^{-1}u$ in the following integral. Write $t = (t_1, \cdots, t_d)$ and $\rho = \max_{1 \leqslant j \leqslant d} \rho_j$. Then by (2.2.6)

$$\phi\{|u| \geqslant \tfrac{1}{4}T_k\} = (2\pi)^{-\frac{1}{2}d}(\det \Gamma)^{-\frac{1}{2}}\int_{|u| \geqslant \frac{1}{4}T_k} \exp\left(-\tfrac{1}{2}\langle u, \Gamma^{-1}u\rangle\right) du$$

$$= (2\pi)^{-\frac{1}{2}d}\int_{\rho_1 t_1^2 + \cdots + \rho_d t_d^2 \geqslant T_k^2/16} e^{-\frac{1}{2}t^2}\, dt$$

$$\leqslant (2\pi)^{-\frac{1}{2}d}\int_{|t| \geqslant \frac{1}{4}T_k \rho^{-\frac{1}{2}}} e^{-\frac{1}{2}t^2}\, dt < \exp\left(-3T_k^2\rho^{-1}/128\right) = \delta_k \quad \text{(say)}.$$

Hence by Theorem 1 we can redefine the sequence $\{X_k, k \geqslant 1\}$ on a new probability space together with a sequence $\{Y_k, k \geqslant 1\}$ of independent random variables Y_k with normal distribution $N(0, \Gamma)$ such that

$$P\{|X_k - Y_k| \geqslant \alpha_k\} \leqslant \alpha_k \tag{3.6}$$

where

$$\alpha_k \ll k^{-2}. \tag{3.7}$$

We now construct the Brownian motion $X(t)$. We first observe that for any Brownian motion $X(t)$ with covariance matrix Γ the vectors

$$Y_k^* = (X(h_{k+1}) - X(h_k))k^{-\frac{1}{2}\alpha} \tag{3.8}$$

are independent normal vectors with normal distribution ($N(0, \Gamma)$). Hence the sequences $\{Y_k, k \geqslant 1\}$ and $\{Y_k^*, k \geqslant 1\}$ have the same distribution. Consequently, by the Kolmogorov existence theorem we can redefine the process $\{\Sigma_{\nu \leqslant t}\xi_\nu, t \geqslant 0\}$ and the sequence $\{Y_k, k \geqslant 1\}$ on a richer probability space without changing the joint distribution of these two processes and such that on this new probability space there exists a Brownian motion $X(t)$ with covariance matrix Γ satisfying

$$Y_k = Y_k^* \qquad k = 1, 2, \cdots. \tag{3.9}$$

We now show that $X(t)$ has the desired properties. Let $t > 0$ be given. Define $N = [t]$ and M as above. Then by (3.4), (3.8) and (3.9)

$$|\Sigma_{\nu \leqslant t}\xi_\nu - X(t)| \leqslant \Sigma_{k<M}|X_k - Y_k|k^{\frac{1}{2}\alpha} + \max_{h_M \leqslant N < h_{M+1}}|\Sigma_{\nu = h_M}^{N}\xi_\nu| + \sup_{h_M \leqslant t < h_{M+1}}|X(t) - X(h_M)|. \tag{3.10}$$

By (3.6), (3.7), (3.2), (3.3) and the Borel-Cantelli lemma the first term in (3.10)

$$\ll \Sigma_{k<M}k^{\frac{1}{2}\alpha}\alpha_k \ll M^{\frac{1}{2}\alpha - 1} \ll N^{\frac{1}{2} - \delta/(20d)} \qquad \text{a.s.}$$

By Lemma 3.2 the second term in (3.10) is $\ll N^{\frac{1}{2} - \delta/(80d)}$. Finally, by a routine argument we conclude that the last term in (3.10) is also within the desired bounds. Indeed, since the components $X_j(t)$ $(1 \leqslant j \leqslant d)$ are one-dimensional Brownian motions we have with $\lambda = (4(1+\alpha))^{-1}$

$$P\left\{\sup_{h_M \leqslant t \leqslant h_{M+1}} |X(t) - X(h_M)| \geqslant h_M^{\frac{1}{2}-\lambda}\right\}$$
$$\leqslant \Sigma_{j \leqslant d} P\left\{\sup_{h_M \leqslant t \leqslant h_{M+1}} |X_j(t) - X_j(h_M)| \geqslant d^{-\frac{1}{2}} h_M^{\frac{1}{2}-\lambda}\right\}$$
$$\leqslant \Sigma_{j \leqslant d} P\left\{\sup_{0 \leqslant t \leqslant 1} |X_j(t)| \geqslant d^{-\frac{1}{2}} h_M^{\frac{1}{2}-\lambda} M^{-\frac{1}{2}\alpha}\right\}$$
$$\ll P\left\{|N(0,1)| \geqslant d^{-\frac{1}{2}} h_M^{\frac{1}{2}-\lambda} M^{-\frac{1}{2}\alpha}\right\} \ll \exp\left\{-cd^{-1} h_M^{1-2\lambda} M^{-\alpha}\right\} \ll M^{-2}$$

by the proof of Lemma 3.2. Here $c > 0$ depends only on Γ. This proves our claim and thus Theorem 3.

4. Proof of Theorem 4.

4.1. *Lemmas.* For the proof of Theorem 4 we need a few well-known facts, all valid under the hypotheses of Theorem 4.

LEMMA 4.1.1 (Ibragimov (1962)). *Let $p, q > 0$ with $p^{-1} + q^{-1} = 1$. Suppose that ξ and η are measurable with respect to $\mathfrak{M}_1^k$ and $\mathfrak{M}_{k+n}^\infty$. Moreover, suppose that $\|\xi\|_p < \infty$ and $\|\eta\|_q < \infty$. Then*

$$|E\xi\eta - E\xi \cdot E\eta| \leqslant 2\phi^{(1/p)}(n)\|\xi\|_p \|\eta\|_q.$$

LEMMA 4.1.2. *We have for some constant $a > 0$*

$$E|\Sigma_{n \leqslant N} \xi_n|^{2+\delta} \leqslant a s_N^{2+\delta}, \tag{4.1.1}$$

$$s_N^2 = N h(N) \tag{4.1.2}$$

where $h(N)$ is a slowly varying function on the integers satisfying $h(x) \ll x^\tau$ for any $\tau > 0$. Moreover, for all x

$$P\{\max_{1 \leqslant n \leqslant N} |\Sigma_{k \leqslant n} \xi_k| \geqslant x\} \leqslant 2P\{|\Sigma_{k \leqslant N} \xi_k| \geqslant x - 2s_N\} + N^{-\frac{1}{4}\delta}. \tag{4.1.3}$$

(4.1.1) and (4.1.2) are due to Ibragimov (1962) pages 360 and 357. The proof of (4.1.3) is basically the same as the proof of Lemma 3 of Reznik (1968) except that at one place (4.1.2) has to be used.

4.2. *Introduction of blocks.* Put

$$\alpha = \delta/40. \tag{4.2.1}$$

We define blocks H_k and I_k of consecutive positive integers, leaving no gaps between the blocks. The order is $H_1, I_1, H_2, I_2, \cdots$. The lengths of the blocks are defined by

$$\operatorname{card} H_k = [2\alpha e^{k^{2\alpha}} k^{2\alpha-1}], \quad \operatorname{card} I_k = [\alpha e^{k^\alpha} k^{\alpha-1}]. \tag{4.2.2}$$

We set

$$N_k = \Sigma_{j \leqslant k} \operatorname{card}(H_j \cup I_j) \sim e^{k^{2\alpha}} \tag{4.2.3}$$

and

$$u_k = \Sigma_{\nu \in H_k} \xi_\nu, \quad v_k = \Sigma_{\nu \in I_k} \xi_\nu, \quad \sigma_k^2 = E u_k^2. \tag{4.2.4}$$

LEMMA 4.2.1. *We have*

$$\sigma_k k^{\frac{1}{2}-\alpha} \ll s_{N_k} \ll \sigma_k \cdot k^{\frac{1}{2}-\alpha}.$$

PROOF. Again we define blocks H_j^* and $I_j^*(1 \leqslant j \leqslant l)$ of consecutive positive integers leaving no gaps between the blocks. Let each $H_j^*(1 \leqslant j \leqslant l)$ consist of card H_k elements and let each $I_j^*(1 \leqslant j \leqslant l)$ consist of card I_k elements. Define l to be the largest integer with

$$\bigcup_{j \leqslant l} H_j^* \cup \bigcup_{j \leqslant l} I_j^* \subset \bigcup_{j \leqslant k} H_k \cup I_k \tag{4.2.5}$$

and let I_l^* consist of these integers which belong to the set theoretic difference of the two sets in (4.2.5). Note that I_l^* can consist of as many as card $(I_k \cup H_k) - 1$ elements. Then

$$e^{k^{2\alpha}} \ll l e^{k^{2\alpha}} k^{2\alpha-1} \ll e^{k^{2\alpha}}$$

or

$$k^{1-2\alpha} \ll l \ll k^{1-2\alpha}. \tag{4.2.6}$$

By (4.1.2), (4.2.2), (4.2.4) and (4.2.6)

$$\|v_j\| \ll e^{2k^\alpha} \ll \sigma_k^{\frac{1}{2}}/l \qquad 1 \leqslant j \leqslant k. \tag{4.2.7}$$

We set for $1 \leqslant j \leqslant l$

$$u_j^* = \Sigma_{\nu \in H_j^*} \xi_\nu, \qquad v_j^* = \Sigma_{\nu \in I_j^*} \xi_\nu. \tag{4.2.8}$$

We first show that

$$\|v_l^*\| \ll \sigma_k. \tag{4.2.9}$$

If $|\text{card } H_k - \text{card } I_l^*| \leqslant \text{card } I_k$ then by Minkowski's inequality and by the argument used to prove (4.2.7)

$$\|v_l^*\| \ll \sigma_k + e^{2k^\alpha} \ll \sigma_k.$$

If, on the other hand, card $I_l^* <$ card $H_k -$ card I_k, we define a random variable z by (recall that the last summand in v_l^* is ξ_{N_k})

$$z = v_l^* + \xi_{N_k+m} + \cdots + \xi_{N_k+h}$$

where m is the smallest integer with $\phi(m) \leqslant \frac{1}{4}$ and where $h =$ card $H_k -$ card I_l^*. Since z is a sum of card $H_k - m$ terms ξ_ν we have

$$\sigma_k = \|z\| + 0(1). \tag{4.2.10}$$

By Lemma 4.1.1

$$\begin{aligned} Ez^2 = E(v_l^* + z - v_l^*)^2 &\geqslant \|v_l^*\|^2 + \|z - v_l^*\|^2 - 2\phi^{\frac{1}{2}}(m)\|v_l^*\|\|z - v_l^*\| \\ &\geqslant \tfrac{1}{2}\|v_l^*\|^2. \end{aligned} \tag{4.2.11}$$

(4.2.9) follows now from (4.2.10) and (4.2.11).

By (4.2.7), (4.2.6), (4.2.4) and (4.2.9)

$$\|\Sigma_{j \leqslant l} v_j^*\| \ll l\|v_1^*\| + \|v_l^*\| \ll l\|v_k\| + \sigma_k \ll \sigma_k. \tag{4.2.12}$$

Moreover, by Lemma 4.1.1, (4.2.6), (1.6) and (4.2.1)

$$E(\Sigma_{j \leqslant l} u_j^*)^2 - l\sigma_k^2 \ll \Sigma_{i<j \leqslant l} \sigma_k^2 \phi^{\frac{1}{2}}(e^{k^\alpha} k^{\alpha-1})$$
$$\ll l^2 \sigma_k^2 k^{-2} \ll \sigma_k^2.$$

Hence by (4.2.12)

$$s_{N_k} - l^{\frac{1}{2}} \sigma_k \ll \sigma_k. \tag{4.2.13}$$

The lemma follows now from (4.2.6).

LEMMA 4.2.2. *We have*

$$\Sigma_{j \leqslant k} v_j \ll s_{N_k}^{\frac{1}{2}} \qquad \text{a.s.}$$

PROOF. From (4.1.2) and (4.2.2) we conclude that

$$\|\Sigma_{j \leqslant k} v_j\|_2 \ll \Sigma_{j \leqslant k} e^{\frac{1}{2}(1+\tau) j^\alpha} j^{(\alpha-1)(1+\tau)} \ll e^{\frac{1}{2}(1+\tau) k^\alpha}.$$

Hence

$$P\left\{|\Sigma_{j \leqslant k} v_j| \geqslant s_{N_k}^{\frac{1}{2}}\right\} \ll s_{N_k}^{-1} e^{(1+\tau)k^\alpha} \ll e^{-(1-\tau)k^{2\alpha} + (1+\tau)k^\alpha} \ll e^{-\frac{1}{2}k^{2\alpha}}$$

by another application of (4.1.2) and (4.2.2). The lemma follows now from the Borel-Cantelli lemma.

LEMMA 4.2.3. *We have*

$$\max_{N_k \leqslant N < N_{k+1}} |\Sigma_{\nu = N_k+1}^{N} \xi_\nu| \ll s_{N_k} k^{-5\alpha} \qquad \text{a.s.}$$

PROOF. By stationarity, Lemma 4.2.1, (4.1.1), (4.1.3), (4.2.2) and (4.2.1) we obtain

$$P\left\{\max_{N_k \leqslant N < N_{k+1}} |\Sigma_{\nu = N_k+1}^{N} \xi_\nu| > s_{N_k} k^{-5\alpha}\right\}$$
$$\ll P\left\{|u_k| \geqslant s_{N_k} k^{-5\alpha} - 2\sigma_k\right\} + (N_{k+1} - N_k)^{-\frac{1}{4}\delta}$$
$$\ll P\left\{|u_k| \geqslant \tfrac{1}{2} s_{N_k} k^{-5\alpha}\right\} + e^{-k^\alpha}$$
$$\ll s_{N_k}^{-(2+\delta)} E|u_k|^{2+\delta} k^{5\alpha(2+\delta)} + e^{-k^\alpha} \ll \left(\sigma_k k^{5\alpha}/s_{N_k}\right)^{2+\delta} + e^{-k^\alpha}$$
$$\ll k^{-\left(\frac{1}{2}-6\alpha\right)(2+40\alpha)} \ll k^{-1-2\alpha}.$$

The lemma follows now from the Borel-Cantelli lemma.

LEMMA 4.2.4. *We have*

$$s_{N_{k+1}}^2 - s_{N_k}^2 = \sigma_{k+1}^2(1 + 0(k^{-1})) \tag{4.2.14}$$

and

$$s_{N_k}^2 - \Sigma_{j \leqslant k} \sigma_j^2 \ll s_{N_k}^2 k^{-1}. \tag{4.2.15}$$

PROOF. From Minkowski's inequality and (4.2.7)

$$s_{N_{k+1}} \ll s_{N_k} + \sigma_{k+1}, \qquad s_{N_k} \ll s_{N_{k+1}} + \sigma_{k+1}.$$

Thus by Lemma 4.2.1

(4.2.16) $$s_{N_k} \ll s_{N_{k+1}} \ll s_{N_k}.$$

Since

$$s^2_{N_{k+1}} = E\left(u_{k+1} + v_{k+1} + \Sigma_{\nu \leqslant N_k} \xi_\nu\right)^2$$

we have

$$s^2_{N_{k+1}} - s^2_{N_k} - \sigma^2_{k+1} \ll \sigma^{\frac{1}{2}}_{k+1} s_{N_k} + \sigma_{k+1} s_{N_k} k^{-2}$$

by Lemma 4.1.1, (1.8), (4.2.7), (4.2.4) and (4.2.1). Thus by Lemma 4.2.1, (4.1.2) and (4.2.16)

(4.2.17) $$s^2_{N_{k+1}} - s^2_{N_k} - \sigma^2_{k+1} \ll \sigma^2_{k+1} k^{-1}.$$

This proves (4.2.14). To prove (4.2.15) we sum (4.2.17) and obtain using (4.2.2) and (4.1.2)

$$s^2_{N_{n+1}} - \Sigma^n_{k=0} \sigma^2_{k+1} \ll \Sigma_{k \leqslant \frac{1}{2}n} \sigma^2_k + n^{-1} \Sigma_{\frac{1}{2}n < k < n} \sigma^2_{k+1} \ll s^2_{N_n} n^{-1}.$$

4.3. *Conclusion of proof of Theorem* 4. Let $\mathfrak{L}_k$ be the σ-field generated by u_k. We apply Theorem 2 to the sequence $\{u_k, k \geqslant 1\}$ and obtain a sequence $\{Y_k, k \geqslant 1\}$ of independent random variables having the same distribution as u_k and satisfying

(4.3.1) $$P\{|u_k - Y_k| \geqslant 12k^{-4}\} \ll k^{-4}$$

since by (1.8), (4.2.2) and (4.2.1) $\phi_k \leqslant 2k^{-160\alpha/\delta} = 2k^{-4}$ for all $k \geqslant k_0$.

We now apply Strassen's (1965a) Theorem 4.4 to the sequence $\{Y_k, k \geqslant 1\}$ and $f(t) = t(\log t)^{-5}$. Then

(4.3.2) $$V_k = \Sigma_{j \leqslant k} EY_j^2 = \Sigma_{j \leqslant k} Eu_j^2 = s^2_{N_k}(1 + 0(k^{-1}))$$

by Lemma 4.2.4. We note that by (4.1.2) and (4.2.3)

(4.3.3) $$\log s_{N_k} \ll k^{2\alpha} \ll \log s_{N_k}.$$

Thus by the last two lines of the proof of Lemma 4.2.3 and by (4.3.2)

$$\Sigma_{k \geqslant 1} V_k^{-1} (\log V_k)^5 \int_{\{Y_k^2 > V_k (\log V_k)^{-5}\}} Y_k^2 dP$$
$$\ll \Sigma_{k \geqslant 1} \left(s_{N_k}^{-2} k^{10\alpha}\right)^{1+\frac{1}{2}\delta} E|u_k|^{2+\delta} < \infty.$$

Hence by Strassen's (1965a) Theorem 4.4 we can redefine the sequence $\{Y_j, j \geqslant 1\}$, without changing its distribution, on a richer probability space together with standard Brownian motion $X(t)$ such that

(4.3.4) $$\Sigma_{j \leqslant k} Y_j - X(V_k) \ll V_k^{\frac{1}{2}} (\log V_k)^{-\frac{1}{4}} \quad \text{a.s.}$$

Theorem 3.1 follows now at once. We put

(4.3.5) $$a_N = V_k \; (N_k \leqslant N < N_{k+1}).$$

By (4.3.2) the sequence $\{V_k, k \geq 1\}$ and thus $\{a_N, N \geq 1\}$ is nondecreasing. Moreover, for $N_k \leq N < N_{k+1}$

$$\begin{aligned} a_N - s_N^2 &= V_k - E\left(\Sigma_{n \leq N_k} \xi_n + \Sigma_{n=N_k+1}^{N} \xi_n\right)^2 \\ &= V_k - s_{N_k}^2 + 0(s_{N_k} \sigma_k) \ll s_{N_k}^2 k^{-1} \ll s_N^2 k^{-1} \end{aligned}$$

by (4.3.2) and since

$$E\left(\Sigma_{n=N_k+1}^{N} \xi_n\right)^2 \ll \sigma_k^2.$$

(This last relation can be proved in the same way as (4.2.9).) Consequently, $a_N \sim s_N^2$. Finally, let $N \geq 1$ be given. Choose k so that $N_k \leq N < N_{k+1}$. Then

$$\begin{aligned} \Sigma_{n \leq N} \xi_n - X(a_N) &\ll \Sigma_{j \leq k} |u_j - Y_j| + |\Sigma_{j \leq k} v_j| \\ &\quad + \max_{N_k \leq M < N_{k+1}} |\Sigma_{n=N_k+1}^{M} \xi_n| \\ &\quad + |\Sigma_{j \leq k} Y_j - X(V_k)| \\ &\ll V_k (\log V_k)^{-\frac{1}{4}} \ll a_N (\log a_N)^{-\frac{1}{4}} \quad \text{a.s.} \end{aligned} \tag{4.3.6}$$

by (4.2.4), (4.3.1)-(4.3.4), the Borel-Cantelli lemma and Lemmas 4.2.2 and 4.2.3. This proves Theorem 4.

4.4. *Miscellaneous ramarks.*

4.4.1. We first prove the claim that Theorem 4 implies the law of the iterated logarithm (1.10). Since $a_N \sim s_N^2$ this will follow from

$$\limsup_{N \to \infty} (2 a_N \log \log a_N)^{-\frac{1}{2}} X(a_N) = 1 \quad \text{a.s.} \tag{4.4.1}$$

The law of the iterated logarithm for standard Brownian motion implies at once that $\lim\sup(\cdots) \leq 1$. To prove the reverse inequality we recall that, by the standard proof, for given $\delta > 0$ there is a sufficiently large q such that

$$\limsup_{n \to \infty} (2 q^n \log \log q^n)^{-\frac{1}{2}} X(q^n) \geq 1 - \delta \quad \text{a.s.} \tag{4.4.2}$$

But for given n there is a k such that

$$V_k \leq q^n < V_{k+1}.$$

This implies by (4.3.2), (4.2.4) and Lemma 4.2.1

$$q^n - V_k \leq V_{k+1} - V_k = \sigma_{k+1}^2 \ll V_k k^{-\frac{1}{2}+\alpha}. \tag{4.4.3}$$

Hence

$$\limsup_{n \to \infty} (2 V_k \log \log V_k)^{-\frac{1}{2}} X(q^n) \geq 1 - \delta \quad \text{a.s.} \tag{4.4.4}$$

Now $X(q^n) - X(V_k)$ is a normal random variable with variance $\ll V_k k^{-\frac{1}{2}+\alpha}$ by (4.4.3). Thus

$$P\left\{|X(q^n) - X(V_k)| \geq V_k^{\frac{1}{2}}\right\} \ll \exp(-k^{\frac{1}{4}}).$$

Consequently by (4.4.4) and the Borel-Cantelli lemma

$$\limsup_{k\to\infty}(2V_k \log\log V_k)^{-\frac{1}{2}}X(V_k) \geqslant 1-\delta \qquad \text{a.s.}$$

This together with (4.3.5) proves (4.4.1) and thus our claim.

4.4.2. We now show that in the standard case $\Sigma\phi^{\frac{1}{2}}(n) < \infty$, the quantity a_N can be replaced by $\sigma^2 N$. Hence in this case Theorem 4 reduces to Theorem 4.1 of Philipp and Stout (1975) (except for the error term). Indeed, if $\Sigma\phi^{\frac{1}{2}}(n) < \infty$ then by Lemma 4.2.2 of the just mentioned paper we obtain after a proper normalization of the random variables

$$s_N^2 - N \ll N^{2/(2+\delta)}.$$

Hence if $N_k \leqslant N < N_{k+1}$, then

$$\begin{aligned} a_N - N &= s_{N_k}^2(1 + 0(k^{-1})) - N \\ &= N_k(1 + 0(k^{-1})) - N \\ &\ll N_k(k^{2\alpha-1} + k^{-1}) \ll N_k(\log N_k)^{-2} \ll N(\log N)^{-2} \end{aligned}$$

by (4.3.5), (4.3.2), (4.2.2), (4.2.3) and (4.2.1). Consequently,

$$\Sigma_{k\leqslant N}\xi_k - X(N) \ll N(\log N)^{-\frac{1}{4}} \qquad \text{a.s.}$$

by Theorem 4, the Borel-Cantelli lemma and since $X(a_N) - X(N)$ is a normal random variable with variance $\ll N(\log N)^{-2}$.

4.4.3. It is an easy matter to extend Theorem 4 to functions $\eta_n = f(\xi_n, \xi_{n+1}, \cdots)$ of sequences of ϕ-mixing random variables. A proof can be modeled after Section 7 in Philipp and Stout (1975).

4.4.4. There is no difficulty in extending Theorem 4 by our method to sequences of random variables satisfying a strong mixing condition (1.11). However, in this case we have to assume that the mixing rate $\rho(n)$ satisfies $\rho(n) \ll n^{-c/\delta}$ for some $c > 0$. Since such a result was proved by Philipp and Stout (1975) using martingale approximation techniques it is not worthwhile to present here a proof in full detail. We only give a short sketch instead.

We define the blocks u_k and v_k as well as σ_k by (4.2.4). Instead of Theorem 2 we apply Theorem 1 to the sequence $\{u_k\sigma_k^{-1}, k \geqslant 1\}$. We note that the left-hand side of (1.1) does not exceed

$$\text{(4.4.5)} \qquad E\left|E\left\{\exp\left(i\langle u, u_k\sigma_k^{-1}\rangle\right)\middle|\mathfrak{F}_{k-1}\right\} - E\left\{\exp\left(i\langle u, u_k\sigma_k^{-1}\rangle\right)\right\}\right| + \left|E\left\{\exp\left(i\langle u, u_k\sigma_k^{-1}\rangle\right)\right\} - e^{-\frac{1}{2}u^2}\right|.$$

The second term in (4.4.5) is small for $|u| \leqslant T_k$ since, as is well known, the central limit theorem with remainder holds for properly normalized sums of strong mixing random variables. The first term in (4.4.5) is small by (4.2.4), (4.2.2) and the following lemma due to Dvoretzky (1970).

LEMMA 4.4.1. *Let X be a random variable with $|X| \leqslant 1$ and let $\mathcal{F}$ be the σ-field generated by X. Then for any σ-field $\mathcal{G}$*

$$E|E(X|\mathcal{G}) - EX| \leqslant 2\pi\rho(\mathcal{F}, \mathcal{G})$$

where

$$\rho(\mathcal{F}, \mathcal{G}) = \sup|P(AB) - P(A)P(B)|$$

the supremum being extended over all $A \in \mathcal{F}$ and $B \in \mathcal{G}$. If X is a real-valued random variable then 2π can be replaced by 4.

Hence by (4.4.5) we see that (1.1) is satisfied with $g_k(u) = e^{-\frac{1}{2}u^2}$ for all $|u| \leqslant T_k$. Since G_k is normal (1.2) also holds if T_k was suitably chosen. Thus by Theorem 2.2 there is a sequence of independent standard normal random variables Y_k such that

$$P\left\{|Y_k - u_k\sigma_k^{-1}| \geqslant \alpha_k\right\} \leqslant \alpha_k$$

for some α_k with $\Sigma\alpha_k < \infty$. Thus

$$P\{|u_k - Y_k\sigma_k| \geqslant \sigma_k\alpha_k\} \leqslant \alpha_k \tag{4.4.6}$$

where $\{Y_k, k \geqslant 1\}$ is a sequence of independent normal $N(0, \sigma_k^2)$ random variables. The partial sums of this sequence can be embedded into standard Brownian motion by Kolmogorov's existence theorem. By (4.4.6)

$$\Sigma_{j \leqslant k} u_j - \Sigma_{j \leqslant k} Y_k\sigma_k \ll \Sigma_{j \leqslant k}\sigma_k\alpha_k \ll t_k^{\frac{1}{2}-\lambda} \qquad \text{a.s.}$$

As in Lemma 4.2.2 it is easy to see that the small blocks v_k can be discarded. Finally, we recall that by a maximal inequality, similar to (4.1.3), Lemma 4.2.3 also remains valid in the strong mixing case. However, in its proof one has to apply the central limit theorem with remainder instead of Markov's inequality for the $(2 + \delta)$th moments.

5. One more approximation theorem. The basic hypotheses in both Theorems 1 and 2 involve the conditional characteristic function or the conditional distribution of X_k. However, in some instances the joint characteristic function of $X_1, \cdots X_k$ is easier to deal with.

THEOREM 5. *Let $\{X_k, k \geqslant 1\}$ be a sequence of random variables and let $\{G_k, k \geqslant 1\}$ be a sequence of probability distributions with characteristic functions $g_k(u)$, $u \in \mathbb{R}$. For fixed $k \geqslant 1$ denote $f(u_1, \cdots, u_k)$ the joint characteristic function of $X_1, \cdots, X_k$ and $f^*(u_1, \cdots, u_{k-1})$ the joint characteristic function of $X_1, \cdots, X_{k-1}$. Suppose for each $k \geqslant 2$*

$$|f(u_1, \cdots, u_k) - f^*(u_1, \cdots, u_{k-1})g_k(u_k)| \leqslant \rho_k \tag{5.3}$$

for all $u = (u_1, \cdots, u_k)$ with $|u| \leqslant U_k$ where

$$U_k > 10^4k^2. \tag{5.4}$$

Moreover, suppose that for some $\eta_k \geqslant 0$

$$\max_{1 \leqslant j \leqslant k} P\left\{|X_j| \geqslant U_k^{\frac{1}{4}}\right\} \leqslant \eta_k. \tag{5.5}$$

Then the conclusion of Theorem 1 *remains valid with*

$$\alpha_k \ll U_k^{-\frac{1}{4}} \log U_k + \eta_k^{\frac{1}{2}} k^{\frac{1}{2}} U_k^{\frac{1}{4}} + \rho_k^{\frac{1}{2}} U_k^{k+\frac{1}{4}} \qquad k \geq 2.$$

PROOF. Let F_k be the distribution and f_k be the characteristic function of X_k. We put $u_1 = \cdots = u_{k-1} = 0$ in (5.3) and obtain

$$|f_k(u_k) - g_k(u_k)| \leq \rho_k \tag{5.6}$$

for all u_k with $|u_k| \leq U_k$. Hence by Lemma 2.2 with $T = 2U_k^{\frac{1}{4}}$

$$\rho(F_k, G_k) < \rho_k U_k^{5/4} + F_k\big(|u| \geq U_k^{\frac{1}{4}}\big) + 8U_k^{-\frac{1}{4}} \log U_k.$$

Thus by the definition of the Prohorov distance and since by (5.4) $U_k^{\frac{1}{4}} < 2U_k^{\frac{1}{4}} - 1 \leq U_k^{\frac{1}{4}} - \rho(F_k, G_k)$, we obtain

$$\begin{aligned} G_k\big(|u| > 2U_k^{\frac{1}{4}}\big) &\leq F_k\big(|u| \geq U_k^{\frac{1}{4}}\big) + \rho(F_k, G_k) \\ &\leq \rho_k U_k^{\frac{5}{4}} + 2\eta_k + 8U_k^{-\frac{1}{4}} \log U_k = \delta_k(\text{say}). \end{aligned} \tag{5.7}$$

This shows that condition (1.2) of Theorem 1 is satisfied with $T_k = 8U_k^{\frac{1}{4}}$.

The verification of condition (1.1) has some similarities with the proof of Lemma 2.2. Let $\{\xi_k, k \geq 1\}$ be a sequence of independent random variables with distribution H_k and characteristic function $h_k \in L^1$ to be chosen suitably later. Put $Z_j = X_j + \xi_j$, $1 \leq j \leq k$. Since $h_k \in L^1$, the joint density $p(z_1, \cdots, z_k)$ of $Z_1, \cdots, Z_k$ is given by

$$\begin{aligned} p(z_1, \cdots, z_k) = (2\pi)^{-k} \textstyle\int_{\mathbf{R}^k} e^{-i\langle u, z\rangle} f(u_1, \cdots, u_k) \\ \times h_1(u_1) \cdots h_k(u_k) du_1 \cdots du_k \end{aligned} \tag{5.8}$$

where $u = (u_1, \cdots, u_k)$ and $z = (z_1, \cdots, z_k)$. Similarly the joint density $p^*(z_1, \cdots, z_{k-1})$ of $Z_1, \cdots, Z_{k-1}$ is given by

$$\begin{aligned} p^*(z_1, \cdots, z_{k-1}) = (2\pi)^{-k+1} \textstyle\int_{\mathbf{R}^{k-1}} e^{i\langle u^*, z^*\rangle} f^*(u_1, \cdots, u_{k-1}) \\ \times h_1(u_1) \cdots h_{k-1}(u_{k-1}) du_1 \cdots du_{k-1} \end{aligned} \tag{5.9}$$

where $u^* = (u_1, \cdots, u_{k-1})$ and similarly $z^* = (z_1, \cdots, z_{k-1})$. Finally, we note that Z_k has density

$$p_k(z_k) = (2\pi)^{-1} \textstyle\int_{\mathbf{R}} e^{-iu_k z_k} f_k(u_k) h_k(u_k) du_k. \tag{5.10}$$

Thus for all real u_k

$$\begin{aligned} &E|E\{e^{iu_k Z_k}|Z_1, \cdots, Z_{k-1}\} - E\{e^{-iu_k Z_k}\}| \\ &= \textstyle\int_{\mathbf{R}^{k-1}} \left|\int_{\mathbf{R}} e^{iu_k z_k} \left(\dfrac{p(z_1, \cdots, z_k)}{p^*(z_1, \cdots, z_{k-1})} - p_k(z_k)\right) dz_k\right| \\ &\quad \times p^*(z_1, \cdots, z_{k-1}) dz_1 \cdots dz_{k-1} \\ &\leq \textstyle\int_{\mathbf{R}^k} |p(z_1, \cdots, z_k) - p^*(z_1, \cdots, z_{k-1}) p_k(z_k)| dz_1 \cdots dz_k \\ &\leq \textstyle\int_{|z| \leq U_k} |p(z_1, \cdots, z_k) - p^*(z_1, \cdots, z_{k-1}) p_k(z_k)| dz_1 \cdots dz_k \\ &\quad + P\{Z_1^2 + \cdots + Z_k^2 \geq U_k^2\} + P\{Z_1^2 + \cdots + Z_{k-1}^2 \geq \tfrac{1}{2} U_k^2\} \\ &\quad + P\big(Z_k^2 > \tfrac{1}{2} U_k^2\big). \end{aligned} \tag{5.11}$$

By (5.8)–(5.10) the last integral is bounded by

$$(5.12)\quad \begin{aligned} &(2\pi)^{-k}(2U_k)^k \int_{\mathbf{R}^k} |f(u_1, \cdots, u_k) - f^*(u_1, \cdots, u_{k-1}) f_k(u_k)| \\ &\qquad \times |h_1(u_1) \cdots h_k(u_k)| du_1 \cdots du_k \\ &\qquad \leqslant (U_k/\pi)^k \int_{|u| \leqslant U_k} \{ |f(u_1, \cdots, u_k) - f^*(u_1, \cdots, u_{k-1}) g_k(u_k)| \\ &\qquad + |f_k(u_k) - g_k(u_k)| \} du_1, \cdots du_k \\ &\qquad + U_k^k \int_{|u| \geqslant U_k} |h_1(u_1) \cdots h_k(u_k)| du_1 \cdots du_k \\ &\qquad \leqslant U_k^{2k} \rho_k + U_k^k \int_{|u| \geqslant U_k} |h_1(u_1) \cdots h_k(u_k)| du_1 \cdots du_k. \end{aligned}$$

We put

$$(5.13)\qquad \sigma_j = U_j^{-\frac{3}{4}}$$

and choose H_j a normal distribution with mean 0 and variance σ_j^2. Then the last integral in (5.12) is by (2.2.6)

$$\ll \int_{|u|>U_k} \exp\left(-\tfrac{1}{2}\Sigma_{j\leqslant k}\sigma_j^2 u_j^2\right) du_1 \cdots du_k \ll U_k^k \int_{|u|>U_k\sigma_k} \exp\left(-\tfrac{1}{2}\Sigma_{j\leqslant k} u_j^2\right) du_1 \cdots du_k$$

$$\ll U_k^k \exp(-3U_k^2\sigma_k^2/8).2^k \ll \exp\left(-\tfrac{1}{4} U_k^{\frac{1}{2}}\right).$$

Thus by (5.12) the last integral in (5.11) is bounded by

$$(5.14)\qquad \ll U_k^{2k}\rho_k + \exp\left(-\frac{1}{8} U_k^{\frac{1}{2}}\right).$$

It remains to estimate the three tail probabilities in (5.11).

By (2.2.6), (5.4), (5.5) and (2.4.7) we obtain

$$(5.15)\quad \begin{aligned} P\{Z_1^2 + \cdots + Z_k^2 \geqslant U_k^2\} &\leqslant P\{2\Sigma_{j\leqslant k}(X_j^2 + \xi_j^2) \geqslant U_k^2\} \\ &\leqslant \Sigma_{j\leqslant k} P\{X_j^2 \geqslant U_k^2/4k\} + P\{\Sigma_{j\leqslant k}\xi_j^2 \geqslant \tfrac{1}{4}U_k^2\} \\ &\leqslant \Sigma_{j\leqslant k} P\{|X_j| \geqslant U_k^{\frac{1}{4}}\} + P\{\Sigma_{j\leqslant k}\xi_j^2\sigma_j^{-2} > \tfrac{1}{4}U_k^2\} \\ &\ll k\eta_k + e^{-U_k}. \end{aligned}$$

For the other two tail probabilities in (5.11) we obtain the same bounds in the same way. Hence by (5.11), (5.14) and (5.15)

$$(5.16)\qquad E|E\{e^{iuZ_k}|Z_1, \cdots, Z_{k-1}\} - E\{e^{iuZ_k}\}| \ll U_k^{2k}\rho_k + k\eta_k + \exp\left(-\frac{1}{8} U_k^{\frac{1}{2}}\right)$$

for all real u.

Finally, by (5.6) and (5.13)

$$(5.17)\qquad \begin{aligned}|E\{e^{iuZ_k}\} - g_k(u)| &= |f_k(u)h_k(u) - g_k(u)| \\ &\ll |f_k(u) - g_k(u)| + |h_k(u) - 1| \\ &\ll \rho_k + u^2\sigma_k^2 \ll \rho_k + U_k^{-1}\end{aligned}$$

for all u with $|u| \leqslant 8U_k^{\frac{1}{4}}$.

Hence (5.16), (5.17) and (5.7) show that conditions (1.1) and (1.2) of Theorem 1 are satisfied with

$$\lambda_k \ll U_k^{2k}\rho_k + k\eta_k + U_k^{-1},$$

$$T_k = U_k^{-\frac{1}{4}},$$

and

$$\delta_k = \rho_k U_k^{\frac{5}{4}} + 2\eta_k + 8U_k^{-\frac{1}{4}} \log U_k.$$

Hence by (1.4) we can choose α_k a constant multiple of

$$U_k^{-\frac{1}{4}} \log U_k + \eta_k^{\frac{1}{2}} k^{\frac{1}{2}} U_k^{\frac{1}{4}} + \rho_k^{\frac{1}{2}} U_k^{k+\frac{1}{4}}.$$

Thus by Theorem 1 there exists a sequence of independent random variables Y_k with distribution G_k such that

$$P\{|Z_k - Y_k| \geqslant \alpha_k\} \leqslant \alpha_k.$$

Consequently, by (5.13) and since $Z_k = X_k + \xi_k$

$$\begin{aligned}P\{|X_k - Y_k| \geqslant 2\alpha_k\} &\leqslant P\{|Z_k - Y_k| \geqslant \alpha_k\} + P\{|\xi_k| \geqslant \alpha_k\} \\ &\leqslant \alpha_k + \alpha_k^{-2}\sigma_k^2 \ll \alpha_k.\end{aligned}$$ □

Note added in proof. In the course of the proofs of Theorems 3 and 4 we several times redefined sequences of random variables thereby applying without explicityly mentioning the following simple lemma.

LEMMA A1. *Let S_i, $i = 1, 2, 3$ be separable Banach spaces. Let F be a distribution on $S_1 \times S_2$ and let G be a distribution on $S_2 \times S_3$ such that the second marginal of F equals the first marginal of G. Then there exist a probability space and three random variables Z_i, $i = 1, 2, 3$ defined on it such that the joint distribution of Z_1 and Z_2 is F and the joint distribution of Z_2 and Z_3 is G.*

This lemma is used implicitly in most papers on the subject, past, present (and future!). The fact that there is perhaps a difficulty when redefining sequences of random variables was observed and settled by Philipp and Stout (1975), page 23. To verify the lemma we modify their basic idea: Simply choose Z_1 and Z_3 conditionally independent given Z_2. Of course when applying Kolmogorov's existence theorem (which remains valid in the Banach space setting) we have to check the consistency condition. Let μ be the common marginal of F and G. The conditional probabilities $P(Z_i \in A_i | Z_2 = .)$ $i = 1, 3$ are defined for all Borel sets $A_i \subset S_i$, $i = 1, 3$. (Although we are to construct the random variables Z_i, $i = 1, 2, 3$ we can use this notation for convenience since these conditional probabilities can be defined in terms of F and G respectively.) Then F and G are marginals of the following probability measure υ defined on $S_1 \times S_2 \times S_3$ by

$$\upsilon(A_1 \times A_2 \times A_3) = \int_{A_2} P(Z_1 \in A_1 | Z_2 = z) P(Z_3 \in A_3 | Z_2 = z) d\mu(z).$$

REFERENCES

BERKES, ISTVÁN and PHILIPP, WALTER (1977). An almost sure invariance principle for the empirical distribution function of mixing random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **41** 115–137.

CHUNG, KAI LAI (1974). *A Course in Probability Theory*, 2nd ed. Academic Press, New York.

CSÖRGÖ, M. and RÉVÉSZ, P. (1975). A new method to prove Strassen type laws of invariance principle I + II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **31** 255–269.

DUDLEY, R. M. (1968). Distances of probability measures and random variables. *Ann. Math. Statist.* **39** 1563–1572.

DVORETZKY, ARYEH (1970). Asymptotic normality for sums of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2** 513–535.

HEYDE, C. C. and SCOTT, D. J. (1973). Invariance principles for the law of the iterated logarithm for martingales and processes with stationary increments. *Ann. Probability* **1** 428–437.

IBRAGIMOV, I. A. (1962). Some limit theorems for stationary processes. *Theor. Probability Appl.* **7** 349–382.

KIEFER, J. (1972). Skorohod embedding of multivariate R.V.'s and the sample D.F. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **24** 1–35.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent R.V.'s and the sample DF, I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** 111–131.

LOÈVE, MICHEL (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton.

PHILIPP, WALTER and STOUT, WILLIAM, (1975). Almost sure invariance principles for partial sums of weakly dependent random variables. *Mem. Amer. Math. Soc.* **161**, Providence, R.I.

REZNIK, M. KH. (1968). The law of the iterated logarithm for some classes of stationary processes. *Theor. Probability Appl.* **8** 606–621.

STRASSEN, V. (1964). An almost sure invariance principle for the law of the iterated logarithm. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **3** 211–226.

STRASSEN, V. (1965a). Almost sure behaviour of sums of independent random variables and martingales. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **2** 315–343.

STRASSEN, V. (1965b). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

VON BAHR, BENGT (1967). Multi-dimensional integral limit theorems. *Ark. Mat.* **7** 71–88.

MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES
H 1053 BUDAPEST
REÁLTANODA U. 13–15
HUNGARY

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS 61801