The Annals of Probability
1980, Vol. 8, No. 1, 68–82

WEAK AND L^p-INVARIANCE PRINCIPLES FOR SUMS OF B-VALUED RANDOM VARIABLES.

BY WALTER PHILIPP
University of Illinois

Suppose that the properly normalized partial sums of a sequence of independent identically distributed random variables with values in a separable Banach space converge in distribution to a stable law of index α. Then without changing its distribution, one can redefine the sequence on a new probability space such that these partial sums converge in probability and consequently even in $L^p (p < \alpha)$ to the corresponding stable process. This provides a new method to prove functional central limit theorems and related results. A similar theorem holds for stationary ϕ-mixing sequences of random variables.

1. Introduction.

1.1. Basically we know four types of invariance principles characterized by the mode of convergence to the limiting process. As an illustration we consider the partial sums S_n of a sequence $\{x_\nu, \nu \geqslant 1\}$ of independent identically distributed random variables centered at expectations and with variance 1. Then by Donsker's theorem (see Billingsley (1968), Theorem 16.1, page 137) which is a distribution invariance principle

$$n^{-\frac{1}{2}} S_{[n\cdot]} \to W \text{ in distr.} \tag{1.1}$$

where W is standard Brownian motion on [0, 1]. On the other hand Strassen's (1964) almost sure invariance principle states that after possibly passing to a richer probability space

$$S_{[t]} - X(t) = o\big((t \log \log t)^{\frac{1}{2}}\big) \quad \text{a.s.} \tag{1.2}$$

as $t \to \infty$. Here $\{X(t), t \geqslant 0\}$ is standard Brownian motion on $[0, \infty)$. But we also have a weak invariance principle saying that after possibly passing to a richer probability space

$$n^{-\frac{1}{2}} \max_{k \leqslant n} |S_k - X(k)| \to 0 \quad \text{in Pr.} \tag{1.3}$$

Since the sequences $\{\max_{k \leqslant n} S_k^2/n, n \geqslant 1\}$ and $\{\max_{k \leqslant n} X^2(k)/n, n \geqslant 1\}$ are uniformly integrable, we conclude that the convergence in (1.3) is even in L^2, a result which we will call an L^2-invariance principle.

Results of type (1.2) or (1.3) are obviously conceptually much simpler than (1.1). However (1.3) was usually considered only as an intermediate step in proofs of (1.1) via the Skorohod embedding theorem. (1.3) is implicitly contained in

Received February 14, 1978; revised February 20, 1979 and May 9, 1979.
AMS 1970 *subject classifications.* Primary 60F05; secondary 60B10.
Key words and phrases. Invariance principles, domains of attraction, stable laws, Banach space valued random variables, mixing sequences of random variables.

Breiman (1968), pages 279–281 and Freedman (1970), pages 83–84. Relation (1.3) itself appears to be due to Major (1976a), page 222.

The term weak invariance principle was coined by Simons and Stout (1978) in reference to the weak law of large numbers which in itself can be considered a weak invariance principle with convergence to the degenerate process. However, Simons and Stout (1978) use this term for the following type of result

$$(1.4) \qquad n^{-\frac{1}{2}}|S_n - X(n)| \to 0 \qquad \text{in Pr.}$$

which is obviously weaker than (1.3). In the present paper we will use this term for results of type (1.3) only.

Let $\{x_\nu, \nu \geq 1\}$ be an arbitrary sequence of random variables with partial sums S_n satisfying (1.2) with an error term $o(t^{\frac{1}{2}})$ instead of $o((t \log\log t)^{\frac{1}{2}})$. Then (1.2) implies (1.3) which in turn implies (1.1) as well as (1.4). However (1.4) does not imply (1.1), therefore standing somewhat isolated. This is the main reason why I propose to use the term weak invariance principle for results of type (1.3).

At this point one might be tempted to argue that weak as well as distribution invariance principles are redundant since both are easy consequences of a suitable almost sure invariance principle, i.e., one with an error term $o(t^{\frac{1}{2}})$. (One would have to add though the phrase "provided that one can prove such an almost sure invariance principle".) In some cases they probably are redundant, but certainly not in the case of independent identically distributed random variables with only finite second moments since Major (1976b) has shown that in this case the error term in (1.2) cannot be improved without some additional assumptions on the distribution of the random variables. Another example is provided by sequences of independent identically distributed random variables with infinite variance, but with common distribution in the domain of attraction to the normal law. No almost sure invariance principle is known in this case. However, the following corollary to Theorem 1 below gives necessary and sufficient conditions for a distribution as well as for a weak and an L^p-invariance principle under these hypotheses.

COROLLARY. *Let $\{x_\nu, \nu \geq 1\}$ be a sequence of independent identically distributed random variables. Then the following three statements are equivalent.*

(a) $$A^2P\{|x_1| > A\} = o(E\{x_1^2 I(|x_1| \leq A)\}) \qquad \text{as } A \to \infty.$$

(b) *There exist two sequences $\{a(n), n \geq 1\}$ and $\{h_\nu, \nu \geq 1\}$ such that*

$$X_n \to W \qquad \textit{in } \text{distr.}$$

where X_n is defined by

$$X_n(s) = a(n)^{-1}\Sigma_{\nu \leq ns}(x_\nu - h_\nu) \qquad 0 \leq s \leq 1.$$

(c) *Without changing its distribution we can redefine the sequence $\{x_\nu, \nu \geq 1\}$ on a new probability space on which there exists a standard Brownian motion*

$\{X(t), t \geq 0\}$ such that for the same sequences $\{a(n), n \geq 1\}$ and $\{h_\nu, \nu \geq 1\}$

$$\max_{k \leq n}|a(n)^{-1}\Sigma_{k \leq n}(x_\nu - h_\nu) - n^{-\frac{1}{2}}X(k)| \to 0 \qquad \text{as } n \to \infty$$

in probability and consequently in L^p for any $p < 2$.

The corollary follows at once from Theorem 1 below and from Theorem 1 of Gnedenko and Kolmogorov (1954), page 172.

In discussions on this subject frequently the question is raised whether for instance in relation (1.2) the Brownian motion $\{X(t), t \geq 0\}$ can be chosen such that it and the partial sum process $\{S_{[t]}, t \geq 0\}$ are independent. If (1.2) is assumed to hold then these two processes can never be independent, since if they were the law of the iterated logarithm applied to the sequence $\{X(\nu) - X(\nu - 1) - x_\nu, \nu \geq 1\}$ would contradict (1.2).

1.2. Let B be a separable Banach space. A random variable x with state space B is said to have a stable distribution G if there is a sequence $\{x_\nu, \nu \geq 1\}$ of independent random variables with common distribution G and sequences $\{a(n), n \geq 1\}$, $a(n) \in \mathbb{R}$ and $\{b_n, n \geq 1\}$, $b_n \in B$ such that for every integer $n \geq 1$

$$a(n)^{-1}(\Sigma_{\nu \leq n} x_\nu - b_n) = G \qquad \text{in distr.} \tag{1.5}$$

It is well-known that the only possible values for $a(n)$ are $n^{1/\alpha}$ with $0 < \alpha \leq 2$. We call α the index of the stable law.

We say that the distribution F on B belongs to the domain of attraction of a distribution G if there is a sequence $\{x_\nu, \nu \geq 1\}$ of independent random variables with common distribution F and constants $a(n) \in \mathbb{R}$ and $b_n \in B$ such that

$$a(n)^{-1}(\Sigma_{\nu \leq n} x_\nu - b_n) \to G \qquad \text{in distr.} \tag{1.6}$$

We call $a(n)$ and b_n the norming and centering constants respectively. It is well-known that only stable laws have nonempty domains of attraction. (See Kumar and Mandrekar (1972)). This fact will also follow from Theorem 2 below.

THEOREM 1. *Let F be a distribution on a separable Banach space B belonging to the domain of attraction of a stable law G on B with index α. Then there exist a sequence $\{h(k, n), k, n \geq 1\}$ of constants in B and two sequences $\{x_\nu, \nu \geq 1\}$ and $\{y_\nu, \nu \geq 1\}$ of independent random variables each having common distribution F and G respectively and partial sums S_n and T_n respectively such that*

$$\max_{k \leq n}\|a(n)^{-1}S_k - n^{-1/\alpha}T_k - h(k, n)\| \to 0 \tag{1.7}$$

in probability and consequently in L^p for any $p < \alpha$. (Here $a(n)$ are the norming constants for F.) Moreover, if $\alpha = 2$ and $\int_B \|x\|^2 dF(x) < \infty$ then the convergence is even in L^2 and (1.7) reduces to (1.3).

We note that the sequence $\{T_k, k \geq 1\}$ could be replaced by a stable process.

1.3. We now turn to some more applications of Theorem 1. As mentioned in the abstract, Theorem 1 provides a new method to prove distribution invariance principles. The advantage of this method lies not only in the fact that it yields theorems with stronger conclusions, but also in its simplicity. As an illustration we consider a sequence $\{x_\nu, \nu \geqslant 1\}$ of independent identically distributed B-valued random variables satisfying a central limit theorem, i.e.,

$$n^{-\frac{1}{2}}\Sigma_{\nu \leqslant n} x_\nu \to G \qquad \text{in distr.}$$

where G is a Gaussian distribution on B. By Theorem 7 of Jain (1977) we have $E\|x_1\|^p < \infty$ for each $p < 2$. Thus we can choose $h(k, n) = kEx_1$ in Theorem 1. Moreover, since $\{y_\nu, \nu \geqslant 1\}$ is a sequence of independent Gaussian random variables we can redefine, if necessary, the sequences $\{x_\nu, \nu \geqslant 1\}$ and $\{y_\nu, \nu \geqslant 1\}$ without changing their joint distribution on a richer probability space on which there exists Brownian motion $\{X(t), t \geqslant 0\}$ defined by the covariance structure of G such that

$$\sup_{t \leqslant T}\|\Sigma_{\nu \leqslant t} y_\nu - X(t)\| \ll T^{\frac{1}{4}} \qquad \text{a.s.} \tag{1.8}$$

For a proof of (1.8) see e.g., the proof of Lemma 4.2 of Kuelbs and Philipp (1980). Hence we obtain by Theorem 1

$$\sup_{0 \leqslant s \leqslant 1}\|n^{-\frac{1}{2}}\Sigma_{\nu \leqslant ns} x_\nu - n^{-\frac{1}{2}}X(ns)\| \to 0 \qquad \text{in Pr.} \tag{1.9}$$

Since $n^{-\frac{1}{2}}X(ns)$ has on [0, 1] the same distribution as $X(s)$ we conclude that

$$n^{-\frac{1}{2}}\Sigma_{\nu \leqslant n \cdot} x_\nu \to X \qquad \text{in distr.} \tag{1.10}$$

We note that although the left-hand side in (1.10) is an element of $D[0, 1]$, considerations involving weak convergence of probability measures on metric spaces never enter the picture.

Incidentally, this last argument shows that it is perhaps more convenient to approximate the sums Σx_ν by Σy_ν, rather than by the appropriate stable process since one then does not have to prove an analogue to (1.8).

We finally consider the case $B = \mathbb{R}$. There are many well-known sets of necessary and sufficient conditions on F to belong to the domain of attraction of a stable law. (See e.g., Gnedenko and Kolmogorov (1954).) Since obviously the hypothesis of Theorem 1 is also necessary these well-known sets are also necessary and sufficient conditions for the weak and the L^p-invariance principle.

Consequently, when specialized to real-valued random variables Theorem 1 contains and improves a few of the results of the above-mentioned paper by Simons and Stout (1978). On the other hand the proofs of the necessary and sufficient conditions on F to belong to the domain of attraction to a stable law are usually based on calculations involving characteristic functions in contrast to the Simons-Stout paper where the main emphasis is on so-called probabilistic methods, i.e., methods just avoiding characteristic functions.

1.4. Let $\{x_\nu, \nu \geqslant 1\}$ be a sequence of random variables and let $\mathfrak{M}_a^b$ be the σ-field generated by $x_a, x_{a+1}, \cdots, x_b$. We say that $\{x_\nu, \nu \geqslant 1\}$ satisfies a strong mixing condition if there is a sequence $\alpha(n)\downarrow 0$ such that

$$|P(AB) - P(A)P(B)| \leqslant \alpha(n) \tag{1.11}$$

for all $k, n \geqslant 1$ and all $A \in \mathfrak{M}_1^k$ and $B \in \mathfrak{M}_{k+n}^\infty$. We call $\{x_\nu, \nu \geqslant 1\}$ ϕ-mixing if there is a sequence $\phi(n)\downarrow 0$ such that

$$|P(AB) - P(A)P(B)| \leqslant \phi(n)P(A) \tag{1.12}$$

for all $k, n \geqslant 1$ and all $A \in \mathfrak{M}_1^k$ and $B \in \mathfrak{M}_{k+n}^\infty$.

In Section 2 we prove the following theorem.

THEOREM 2. *Let $\{x_\nu, \nu \geqslant 1\}$ be a stationary sequence of B-valued random variables satisfying a strong mixing condition. Suppose that there is a distribution G, a sequence $\{a(n), n \geqslant 1\}$ of real numbers tending to ∞ and a sequence $\{b_n, n \geqslant 1\}$ of elements in B such that*

$$a(n)^{-1}(\Sigma_{\nu \leqslant n} x_\nu - b_n) \to G \qquad \text{in distr.} \tag{1.13}$$

Then G is a stable law of index α with $0 < \alpha \leqslant 2$ and

$$a(n) = n^{1/\alpha}L(n) \tag{1.14}$$

where L is a slowly varying function.

Recall that a function L is slowly varying if

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1 \tag{1.15}$$

for all real $t > 1$.

Except for the fact that the random variables can assume values in a Banach space, Theorem 2 looks exactly like Theorem 18.1.1 of Ibragimov and Linnik (1971), page 316. However, the slowly varying functions in Theorem 18.1.1 are slowing varying on the integers, i.e., (1.15) is required only to hold for all integers $t \geqslant 2$. Since in the proof of Theorem 1 we shall make heavy use of Karamata's theorem, the extension (1.14) and (1.15) is rather important.

Relation (1.14) is slightly simpler than

$$a(n)^{-\alpha} nL(a(n)) \to 1 \qquad n \to \infty$$

which is given in Feller (1966), pages 303–305 for independent random variables.

In Section 4 we extend part of Theorem 1 to ϕ-mixing random variables.

The proofs of most of the results of this paper are based on the following approximation theorem of Philipp (1979), which is a generalization of Theorem 2 of Berkes and Philipp (1979).

THEOREM 3. *Let $\{B_k, m_k, k \geqslant 1\}$ be a sequence of complete separable metric spaces. Let $\{X_k, k \geqslant 1\}$ be a sequence of random variables with values in B_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 sequence $\{\phi_k, k \geqslant 1\}$ of nonnegative numbers*

$$|P(AB) - P(A)P(B)| \leqslant \phi_k P(A) \tag{1.16}$$

for all $k \geqslant 1$ and all $A \in \bigvee_{j<k} \mathfrak{L}_j$ and $B \in \mathfrak{L}_k$. Denote by F_k the distribution of X_k and let $\{G_k, k \geqslant 1\}$ be a sequence of distributions on B_k such that

$$F_k(A) \leqslant G_k(A^{\rho_k}) + \sigma_k \qquad \textit{for all Borel sets } A \subset B_k. \tag{1.17}$$

Here ρ_k and σ_k are nonnegative numbers and $A^\varepsilon = \cup_{x\in A}\{y: m_k(x,y) < \varepsilon\}$. Then without changing its distribution we can redefine the sequence $\{X_k, k \geq 1\}$ on a richer probability space on which there exists a sequence $\{Y_k, k \geq 1\}$ of independent random variables Y_k with distribution G_k such that for all $k \geq 1$

$$P\{m_k(X_k, Y_k) \geq 2(\phi_k + \rho_k)\} \leq 2(\phi_k + \sigma_k). \tag{1.18}$$

It is worth mentioning that if $\rho_k = \sigma_k$ then (1.17) obviously can be replaced by

$$\rho(F_k, G_k) \leq \rho_k$$

where ρ denotes the Prohorov distance on $\{B_k, m_k\}$.

2. Proof of Theorem 2.

LEMMA 2.1. *Let f be a linear functional on B. Then $G \circ f^{-1}$ is a stable law on $\mathbb{R}$ with index α for some $0 < \alpha \leq 2$. Moreover, for all $k \in \mathbb{Z}^+$*

$$a(n)^{-1}a(kn) \to k^{1/\alpha} \qquad n \to \infty. \tag{2.1}$$

PROOF. Write

$$S_n = \Sigma_{\nu \leq n} f(x_\nu).$$

Applying f to (1.13) we obtain

$$a(n)^{-1}(S_n - f(b_n)) \to G \circ f^{-1}. \tag{2.2}$$

Since $\{f(x_\nu), \nu \geq 1\}$ is a stationary sequence of real-valued random variables satisfying a strong mixing condition, the lemma follows from Theorem 18.1.1 of Ibragimov and Linnik (1971).

LEMMA 2.2. *G is a stable law.*

PROOF. (Kuelbs). Let B^* be the topological dual of B and let Λ be a linear functional on B^* which is weak-star sequentially continuous. Then by Schaefer (1971), Corollary 3 on page 150 Λ is actually weak-star continuous. But by Rudin (1973), Theorem 3.10 and page 66 every weak-star continuous Λ on B^* is of the form

$$\Lambda f = f(b) \qquad f \in B^*$$

for some fixed $b \in B$. Hence according to Dudley and Kanter (1974) (B, B^*) is a semifull pair. The lemma follows now from Theorem 5 of Dudley and Kanter (1974) since by Lemma 2.1 $G \circ f^{-1}$ is a stable law for all $f \in B^*$. $\square$

To finish the proof of Theorem 2 it remains to prove (1.14). For $x \geq 1$ we put $a(x) = a([x])$ where $[x]$ denotes the integral part of x and

$$L(x) = x^{-1/\alpha}a(x). \tag{2.3}$$

We are to show that (1.15) holds for all $t > 1$. In what follows we denote by h_n unspecified real numbers not necessarily the same at each occurrence.

LEMMA 2.3. *We have for each* $k \in \mathbb{Z}^+$

$$\lim_{n\to\infty} \frac{L(n+k)}{L(n)} = 1. \tag{2.4}$$

Let $\varepsilon_n \downarrow 0$. *Then for any* $\tau > 0$

$$\lim_{n\to\infty} \frac{L(n\varepsilon_n)}{L(n)} \varepsilon_n^\tau = \lim_{n\to\infty} \frac{L(n)}{L(n\varepsilon_n)} \varepsilon_n^\tau = 0 \tag{2.5}$$

and

$$\lim_{n\to\infty} \frac{L(n(1-\varepsilon_n))}{L(n)} = 1. \tag{2.6}$$

PROOF. Since by (2.2)

$$a(n)^{-1}(S_n + h_n) \to G \circ f^{-1} \qquad \text{in distr.}$$

since $a(n) \to \infty$ and since

$$a(n+k)^{-1}(S_n + h_n) = a(n+k)^{-1}(S_{n+k} + h_{n+k}) - a(n+k)^{-1}(S_{n+k} - S_n)$$

$$\to G \circ f^{-1} \qquad \text{in distr.}$$

the convergence of types theorem implies

$$\lim_{n\to\infty} a(n)^{-1} a(n+k) = 1.$$

(2.4) follows now from (2.3).

To prove (2.5) we note that by (2.1) and (2.2)

$$\lim_{n\to\infty} \frac{L(2n)}{L(n)} = 1.$$

Hence Ibragimov's (1962) Lemmas 1.6 and 1.7 remain valid in the present setting since (2.4) replaces his Lemma 1.5.

We finally prove (2.6). By (2.5) and (2.3)

$$a(n)^{-1} a(n\varepsilon_n) \to 0.$$

Thus

$$a(n)^{-1} \textstyle\sum_{\nu=[n(1-\varepsilon_n)]+1}^{n} f(x_\nu) - h_n \to 0 \qquad \text{in Pr.}$$

Consequently, by (2.2)

$$a(n)^{-1}(S_{[n(1-\varepsilon_n)]} - h_n) \to G \circ f^{-1} \qquad \text{in distr.}$$

Hence by the convergence of types theorem

$$a(n)^{-1} a(n(1-\varepsilon_n)) \to 1.$$

(2.6) follows now from (2.3). □

We now can finish the proof of Theorem 2. By (2.1) and (2.3) we have $L(nk)/L(n) \to 1$ for any $k \in \mathbb{Z}^+$. Hence there is an increasing sequence $\{n_k, k \geqslant 1\}$ of integers such that for all $n \geqslant n_k$

$$\left|\log \frac{L(nk)}{L(n)}\right| < k^{-1}. \tag{2.7}$$

Let $t > 1$ and define q_n by

$$q_n = k \qquad \text{if} \qquad n_k \leqslant nt < n_{k+1}.$$

Then by (2.7)

$$\log \frac{L([nt]q_n)}{L(nt)} \to 0 \qquad n \to \infty. \tag{2.8}$$

Put

$$p_n = [q_n t]. \tag{2.9}$$

Then $p_n = [kt] \geqslant k$ and hence $|\log(L(np_n)/L(n))| < k^{-1}$ for $n \geqslant n_{[kt]}$. Thus

$$\log \frac{L(np_n)}{L(n)} \to 0 \qquad n \to \infty. \tag{2.10}$$

Finally, by (2.6) and (2.9) $L([nt]q_n)/L(np_n) \to 1$. Thus by (2.8) and (2.10) $L(nt)/L(n) \to 1$. Hence by (2.4)

$$\frac{L(xt)}{L(x)} = \frac{L([x]t + o(1))}{L([x])} \to 1 \qquad x \to \infty.$$

This proves (1.14) and thus Theorem 2. □

3. Proof of Theorem 1. Let $\{x_\nu, \nu \geqslant 1\}$ be a sequence of independent identically distributed random variables with common distribution F and satisfying (1.6). For simplicity we assume that $b_n = 0$ for all $n \geqslant 1$. This does not constitute a real loss of generality since, as was demonstrated in the proof of Theorem 2, the unwanted centering constants can be collected and finally absorbed into $h(k, n)$. By Theorem 3.1 of de Acosta (1975) there is a constant C such that for all $\lambda > 0$

$$G\{x: \|x\| \geqslant \lambda\} < C\lambda^{-\alpha}. \tag{3.1}$$

Let $0 < \varepsilon < 10^{-2}$ be given. We define

$$t_k = (1 + \varepsilon^4)^k \qquad k \geqslant 1$$
$$= 0 \qquad k = 0 \tag{3.2}$$

and

$$n_k = [t_{k+1} - t_k] = [(1 + \varepsilon^4)^k \varepsilon^4]. \tag{3.3}$$

Put

$$s = 4\left[-\frac{\log \varepsilon}{\log(1 + \varepsilon^4)}\right] \tag{3.4}$$

so that

$$(1 + \varepsilon^4)^s \leqslant \varepsilon^4.$$

LEMMA 3.1. *We have for sufficiently large k*

$$\left| \frac{a(t_k)}{a(n_j)} \cdot \left(\frac{n_j}{t_k} \right)^{1/\alpha} - 1 \right| < \varepsilon^{12/\alpha} \qquad j = k - s + 1, \cdots, k \tag{3.5}$$

$$\max_{\nu \leqslant n_k} a(\nu)/a(t_k) \leqslant \varepsilon^{8/\alpha} \tag{3.6}$$

$$\max_{t_k \leqslant \nu \leqslant t_{k+1}} \left| \frac{a(\nu)}{a(t_{k+1})} \cdot \left(\frac{t_{k+1}}{\nu} \right)^{1/\alpha} - 1 \right| < \varepsilon^{8/\alpha} \tag{3.7}$$

and

$$a(t_k) < a(t_{k+1}). \tag{3.8}$$

PROOF. Theorem 2, (2.3) and (1.14) imply

$$a(rn)/a(n) \to r^{1/\alpha} \qquad n \to \infty$$

for any $r > 1$. We put

$$r_j = t_k/n_j \qquad j = k - s + 1, \cdots, k$$

and obtain (3.5). To prove (3.7) we apply Karamata's theorem to (1.14) and obtain for sufficiently large k and for all ν with $t_k \leqslant \nu \leqslant t_{k+1}$

$$\frac{a(\nu)}{a(t_{k+1})} \left(\frac{t_{k+1}}{\nu} \right)^{1/\alpha} = \frac{L(\nu)}{L(t_{k+1})} = \exp\left(\int_{t_{k+1}}^{\nu} \varepsilon(y) y^{-1} dy \right) \tag{3.9}$$

where $\varepsilon(y) \to 0$ as $y \to \infty$. But if $k \geqslant k_0$ the integral is bounded by

$$\varepsilon^{8/\alpha} \log(t_{k+1}/t_k) = \varepsilon^{8/\alpha} \log(1 + \varepsilon^4)$$

by (3.2). Hence (3.7) follows from (3.9). The proofs of (3.6) and (3.8) are similar. □

LEMMA 3.2. *As* $k \to \infty$

$$P\left\{ \max_{j \leqslant n_k} \|\Sigma_{\nu = t_k + 1}^{t_k + j} x_\nu\| \geqslant \varepsilon a(t_k) \right\} \ll \varepsilon^6.$$

PROOF. Using stationarity and Ottaviani's inequality we obtain for the desired probability the bound

$$P\left\{ \max_{j \leqslant n_k} \|S_j\| \geqslant \varepsilon a(t_k) \right\} \leqslant (1 - c)^{-1} P\left\{ \|S_{n_k}\| \geqslant \tfrac{1}{2} \varepsilon a(t_k) \right\} \tag{3.10}$$

where

$$\begin{aligned} c &= \max_{j \leqslant n_k} P\left\{ \|S_{n_k} - S_j\| \geqslant \tfrac{1}{2} \varepsilon a(t_k) \right\} \\ &= \max_{j \leqslant n_k} P\left\{ \|S_j\| \geqslant \tfrac{1}{2} \varepsilon a(t_k) \right\}. \end{aligned} \tag{3.11}$$

The proof of Ottaviani's inequality as given in Breiman (1968), pages 45–46, shows that this inequality remains valid in the Banach space setting. Denote by K_n the distribution of $a(n)^{-1} S_n$. From (1.6) we conclude that

$$\rho(K_n, G) \to 0.$$

Hence by the definition of the Prohorov distance and by (3.1) we have for all $n \geqslant N_0$

(3.12) $$P\left\{a(n)^{-1}\|S_n\| \geqslant \tfrac{1}{2}\varepsilon^{-6/\alpha}\right\} \leqslant G\left\{x: \|x\| \geqslant \tfrac{1}{4}\varepsilon^{-6/\alpha}\right\} + \varepsilon^6 \ll \varepsilon^6.$$

Hence by (3.6)

(3.13) $$\begin{aligned}\max_{N_0 \leqslant j \leqslant n_k} P\left\{\|S_j\| \geqslant \tfrac{1}{2}\varepsilon a(t_k)\right\} &= \max_{N_0 \leqslant j \leqslant n_k} P\left\{a(j)^{-1}\|S_j\| \geqslant \tfrac{1}{2}\varepsilon a(j)^{-1}a(t_k)\right\} \\ &\leqslant \max_{N_0 \leqslant j \leqslant n_k} P\left\{a(j)^{-1}\|S_j\| \geqslant \tfrac{1}{2}\varepsilon^{-6/\alpha}\right\} \ll \varepsilon^6.\end{aligned}$$

On the other hand we trivially have for sufficiently large k

(3.14) $$\max_{j \leqslant N_0} P\left\{\|S_j\| \geqslant \tfrac{1}{2}\varepsilon a(t_k)\right\} \leqslant \varepsilon^6.$$

The lemma follows now from (3.10), (3.11), (3.13) and (3.14). □

We now define for $k \geqslant 1$

(3.15) $$H_k = [t_k, t_{k+1})$$

and

(3.16) $$X_k = a(n_k)^{-1}\Sigma_{\nu \in H_k} x_\nu.$$

Let F_k denote the distribution function of X_k. By stationarity and (1.6) we have for the Prohorov distance

$$\rho_k = \rho(F_k, G) \to 0.$$

Hence by Theorem 3 without changing its distribution we can redefine the sequence $\{X_k, k \geqslant 1\}$ on a new probability space on which there exists a sequence $\{Y_k, k \geqslant 1\}$ of independent random variables with common distribution G such that

(3.17) $$P\{\|X_k - Y_k\| \geqslant 2\rho_k\} \leqslant 2\rho_k.$$

Let k_0 be so large that for all $k \geqslant k_0$

(3.18) $$\rho_k \leqslant \varepsilon^6.$$

Moreover, let $\{y_\nu, \nu \geqslant 1\}$ be a sequence of independent random variables with distribution G. As is well-known there exist constants c_k such that

$$n_k^{-1/\alpha}\left(\Sigma_{\nu \in H_k} y_\nu - c_k\right) = G = Y_k \qquad \text{in distr.} \qquad k \geqslant 1.$$

We apply Kolmogorov's existence theorem. Without changing their joint distribution we can redefine the sequences $\{x_\nu, \nu \geqslant 1\}$ and $\{Y_k, k \geqslant 1\}$ on a new probability space on which there exists a sequence $\{y_\nu, \nu \geqslant 1\}$ of independent random variables with common distribution G such that

(3.19) $$n_k^{-1/\alpha}\left(\Sigma_{\nu \in H_k} y_\nu - c_k\right) = Y_k \qquad k \geqslant 1.$$

As before we observe that there is no real loss of generality in assuming $c_k = 0$ $(k \geqslant 0)$. Hence we can rewrite (3.19) in the form

(3.20) $$n_k^{-1/\alpha} \Sigma_{\nu \in H_k} y_\nu = Y_k \qquad k \geqslant 1.$$

Let $S(n)$ and $T(n)$ denote the nth partial sum of $\{x_\nu, \nu \geqslant 1\}$ and $\{y_\nu, \nu \geqslant 1\}$ respectively. Let M be given and put

(3.21) $$m = M - s$$

where s is defined in (3.4).

LEMMA 3.3. *We have for* $M \geqslant M_0 = M_0(\varepsilon)$

$$P\{\max_{m \leqslant k < M} \| a(t_M)^{-1} S(t_k) - t_M^{-1/\alpha} T(t_k) \| \geqslant \varepsilon\} \ll \varepsilon.$$

PROOF. By (3.5), (3.2), (3.3) and since $0 < \alpha \leqslant 2$ we have

(3.22) $$a(t_M)^{-1} \Sigma_{m \leqslant k < M} a(n_k) \leqslant 2 t_M^{-1/\alpha} \Sigma_{m \leqslant k < M} n_k^{1/\alpha} < \frac{2\varepsilon^{4/\alpha}}{(1 + \varepsilon^4)^{1/\alpha} - 1} \leqslant 4\varepsilon^{-2}.$$

Now the probability in question does not exceed

(3.23) $$P\{\max_{m \leqslant k < M} \| a(t_M)^{-1}(S(t_k) - S(t_m)) - t_M^{-1/\alpha}(T(t_k) - T(t_m)) \| \geqslant \tfrac{1}{2}\varepsilon\} + P\{a(t_M)^{-1} \| S(t_m) \| \geqslant \tfrac{1}{4}\varepsilon\} + P\{t_M^{-1/\alpha} \| T(t_m) \| \geqslant \tfrac{1}{4}\varepsilon\} = \mathrm{I} + \mathrm{II} + \mathrm{III} \quad \text{(say)}$$

By (3.2), (3.3) and (3.4)

(3.24) $$t_m \leqslant n_M.$$

Thus by (3.13) and (3.14)

(3.25) $$\mathrm{II} \ll \varepsilon^6.$$

Since (3.13) and (3.14) also apply to the sequence $\{y_\nu, \nu \geqslant 1\}$ we conclude that

(3.26) $$\mathrm{III} \ll \varepsilon^6.$$

Now by (3.16) and (3.20)

$$S(t_k) - S(t_m) = \Sigma_{m \leqslant j < k} a(n_j) X_j$$

and

(3.27) $$T(t_k) - T(t_m) = \Sigma_{m \leqslant j < k} n_j^{1/\alpha} Y_j.$$

Hence we obtain using (3.22), (3.18), (3.21), (3.20), (3.5), (3.1) and (3.4)

$$\begin{aligned}
\mathrm{I} &\leqslant P\left\{\Sigma_{m\leqslant k<M}\|a(t_M)^{-1}a(n_k)X_k - t_M^{-1/\alpha}n_k^{1/\alpha}Y_k\| \geqslant \tfrac{1}{2}\varepsilon\right\}\\
&\leqslant \Sigma_{m\leqslant k<M}P\left\{\|a(t_M)^{-1}a(n_k)X_k - t_M^{-1/\alpha}n_k^{1/\alpha}Y_k\| \geqslant \tfrac{1}{8}\varepsilon^3 a(n_k)a(t_M)^{-1}\right\}\\
&= \Sigma_{m\leqslant k<M}P\left\{\|X_k - a(t_M)a(n_k)^{-1}t_M^{-1/\alpha}n_k^{1/\alpha}Y_k\| \geqslant \tfrac{1}{8}\varepsilon^3\right\}\\
&\leqslant \Sigma_{m\leqslant k<M}\Big(P\left\{\|X_k - Y_k\| \geqslant \tfrac{1}{16}\varepsilon^3\right\}\\
&\qquad + P\left\{\|Y_k\|\,|a(t_M)a(n_k)^{-1}t_M^{-1/\alpha}n_k^{1/\alpha} - 1| \geqslant \tfrac{1}{16}\varepsilon^3\right\}\Big)\\
&\leqslant s\left(2\varepsilon^6 + P\left\{\|Y_k\|\varepsilon^{12/\alpha} > \tfrac{1}{16}\varepsilon^3\right\}\right) = s\left(2\varepsilon^6 + G\left(x\colon \|x\| \geqslant \tfrac{1}{16}\varepsilon^{-6/\alpha}\right)\right)\\
&\ll s\varepsilon^6 \ll \varepsilon.
\end{aligned}$$

The lemma follows now from (3.23), (3.25), (3.26) and this last estimate. □

Finally, we can finish the proof of Theorem 1. Let n be sufficiently large. Define M by $t_{M-1} \leqslant n < t_M$ and let m be defined by (3.22). Then

$$\begin{aligned}
&P\left\{\max_{j\leqslant n}\|a(n)^{-1}S(j) - n^{-1/\alpha}T(j)\| \geqslant 8\varepsilon\right\}\\
&\quad\leqslant P\left\{\max_{j\leqslant t_m}\|a(n)^{-1}S(j)\| \geqslant \varepsilon\right\} + P\left\{\max_{j\leqslant t_m}\|n^{-1/\alpha}T(j)\| \geqslant \varepsilon\right\}\\
&\qquad + P\left\{\max_{m\leqslant k<M}\|a(n)^{-1}S(t_k) - n^{-1/\alpha}T(t_k)\| \geqslant 4\varepsilon\right\}\\
&\qquad + P\left\{\max_{m\leqslant k<M}\max_{j\leqslant n_k}\|S(t_k + j) - S(t_k)\| \geqslant \varepsilon a(n)\right\}\\
&\qquad + P\left\{\max_{m\leqslant k<M}\max_{j\leqslant n_k}\|T(t_k + j) - T(t_k)\| \geqslant \varepsilon a(n)\right\}\\
&\quad= \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} + \mathrm{V}\ \text{(say)}.
\end{aligned} \tag{3.28}$$

Now by (3.7), (3.24), stationarity and Lemma 3.2

$$\mathrm{I} \leqslant P\left\{\max_{j\leqslant n_M}\|a(n)^{-1}S(j)\| \geqslant \tfrac{1}{2}\varepsilon a(t_M)\right\} \ll \varepsilon \tag{3.29}$$

as $M \to \infty$. Similarly,

$$\mathrm{II} \ll \varepsilon. \tag{3.30}$$

Also by (3.7), (3.2), (3.8), (3.4) and Lemma 3.2

$$\begin{aligned}
\mathrm{IV} &\leqslant \Sigma_{m\leqslant k<M}P\left\{\max_{j\leqslant n_k}\|S(t_k + j) - S(t_k)\| \geqslant \tfrac{1}{2}a(t_M)\right\}\\
&< \Sigma_{m\leqslant k<M}P\left\{\max_{j\leqslant n_k}\|S(t_k + j) - S(t_k)\| \geqslant \tfrac{1}{2}\varepsilon a(t_k)\right\} \ll s\varepsilon^6 \ll \varepsilon.
\end{aligned} \tag{3.31}$$

Similarly

$$\mathrm{V} \ll \varepsilon. \tag{3.32}$$

Finally by (3.7), Lemma 3.3, (3.8), and (3.1)

$$
\begin{aligned}
\mathrm{III} = P\{&\max_{m\leqslant k<M}\|a(t_M)^{-1}S(t_k) - a(n)a(t_M)^{-1}n^{-1/\alpha}T(t_k)\| \geqslant 4\varepsilon a(n)a(t_M)^{-1}\}\\
&\leqslant P\{\max_{m\leqslant k<M}\|a(t_M)^{-1}S(t_k) - t_m^{-1/\alpha}T(t_k)\| \geqslant \varepsilon\}\\
&\quad + P\{\max_{m\leqslant k<M}\|T(t_k)\|\,|t_M^{-1/\alpha} - a(n)a(t_M)^{-1}n^{-1/\alpha}| \geqslant \varepsilon\}\\
&\leqslant \varepsilon + P\{\max_{m\leqslant k<M}\|T(t_k)\| \geqslant \varepsilon^{-6/\alpha}t_M^{1/\alpha}\}\\
&\ll \varepsilon + \textstyle\sum_{m\leqslant k<M} P\{\|T_k\| \geqslant \varepsilon^{-6/\alpha}t_k^{1/\alpha}\}\\
&\ll \varepsilon + s\cdot G\{x: \|x\| \geqslant \varepsilon^{-6/\alpha}\} \ll \varepsilon + s\cdot\varepsilon^6 \ll \varepsilon
\end{aligned}
\tag{3.33}
$$

since $t_k^{-1/\alpha}T_k$ has distribution G. This proves convergence in probability.

We now prove the assertions about L^p-convergence. By Theorem 6.1 of de Acosta and Giné (1979) and by Theorem 5.4 of Billingsley (1968), the sequence $\{\|a(n)^{-1}S_n\|^p, n \geqslant 1\}$ is uniformly integrable for $p < \alpha$. (Recall that we still assume $b_n = 0$.) By Ottaviani's inequality and by the proof of Lemma 3.2

$$P\{\max_{k\leqslant n}\|a(n)^{-1}S_k\| \geqslant \lambda\} \ll P\{\|a(n)^{-1}S_n\| \geqslant \lambda\}$$

for any $\lambda > 0$. Consequently, by relation (3) on page 223 of Billingsley (1968), the sequence $\{\max_{k\leqslant n}\|a(n)^{-1}S_k\|^p, n \geqslant 1\}$ is also uniformly integrable for $p < \alpha$. This implies that $\{\max_{k\leqslant n}\|n^{-1/\alpha}T_k\|^p, n \geqslant 1\}$ and hence that $\{\max_{k\leqslant n}\|a(n)^{-1}S_k - n^{-1/\alpha}T_k\|^p, n \geqslant 1\}$ are uniformly integrable. In view of the already established convergence of (1.7) in probability, this proves L^p-convergence for any $p < \alpha$.

If $\alpha = 2$ and $\int_B\|x\|^2dF(x) < \infty$ the same argument but with Theorem 6.1 of de Acosta and Giné (1979) replaced by their Theorem 3.3 yields convergence in L^2.

4. Extension of Theorem 1 to ϕ-mixing sequences of random variables. We say that the sequence $\{x_\nu, \nu \geqslant 1\}$ belongs to the domain of attraction of a law G if there exist real numbers $a(n)$ and $b_n \in B$ such that (1.6) holds. Of course, if $\{x_\nu, \nu \geqslant 1\}$ is stationary and ϕ-mixing then, by Theorem 2, G is necessarily stable.

THEOREM 4. *Let $\{x_\nu, \nu \geqslant 1\}$ be a stationary, ϕ-mixing sequence of random variables with values in a separable Banach space B. Suppose that $\{x_\nu, \nu \geqslant 1\}$ belongs to the domain of attraction of a stable law with index $0 < \alpha \leqslant 2$. Moreover, suppose that either one of the following conditions holds.*

$$\phi(1) < 1 \tag{4.1}$$

or

(4.2) *For some $r \geqslant 1$ with $\phi(r) < 1$ the subsequence $\{x_{\nu r}, \nu \geqslant 1\}$ belongs to the domain of attraction to some stable law G_r of the same index α.*

Then without changing its distribution we can redefine the sequence $\{x_\nu, \nu \geqslant 1\}$ on a new probability space on which there exists a sequence $\{y_\nu, \nu \geqslant 1\}$ of independent *random variables with common distribution G and having the following property. Let S_n and T_n be the* nth *partial sum of $\{x_\nu, \nu \geq 1\}$ or $\{y_\nu, \nu \geq 1\}$ respectively. Then for some $h(k, n) \in B$*

$$\max_{k \leq n} \|a(n)^{-1} S_k - n^{-1/\alpha} T_k - h(k, n)\| \to 0 \qquad \textit{in } \text{Pr}.$$

where a(n) are the norming constants for $\{x_\nu, \nu \geq 1\}$.

REMARK. Condition (4.2) might seem overly restrictive. But in practice the situation will prove much less serious. Since $\{x_{\nu r}, \nu \geq 1\}$ is also stationary and ϕ-mixing (with an even smaller mixing coefficient) any proof of (1.5) under these assumptions will most likely also yield (1.5) for the sequence $\{x_{\nu r}, \nu \geq 1\}$. If, as an illustration, we combine Corollary 1 or Corollary 2 of Kuelbs and Philipp (1980) with Theorem 4 we obtain a functional central limit theorem for sums of ϕ-mixing random variables with values in B or in a separable Hilbert space respectively.

The proof of Theorem 4 is, apart from a few minor modifications, the same as the proof of Theorem 1. We first show that Lemma 3.2 remains valid under the hypotheses of Theorem 4.

Suppose first that (4.1) holds. Then Ottaviani's inequality remains valid in the Banach space setting (see e.g., Lemma 1.1.6 of Iosifescu and Theoderescu (1969) and its proof). Since this was the only place in the proof of Lemma 3.2 where independence was used, the lemma is proved under the assumption (4.1).

Suppose now that (4.2) holds. Fix r such that $\phi(r) < 1$. Then by the above argument the conclusion of the lemma is true for the sequence $\{x_{\nu r}, \nu \geq 1\}$, i.e.,

$$P\{\max_{j \leq n_k} \|\Sigma_{t_k \leq \nu r \leq t_k + j} x_{\nu r}\| \geq \varepsilon a(t_k)\} \to 0 \tag{4.3}$$

where $a(t_k) = a_r(t_k)$. By stationarity and (4.3) we have for each integer $0 \leq d < r$

$$P\{\max_{j \leq n_k} \|\Sigma_{t_k \leq \nu r \leq t_k + j} x_{\nu r + d}\| \geq \varepsilon a(t_k)\} \to 0.$$

We add these inequalities over all $0 \leq d < r$ and obtain the result. □

We now define

$$H_k = [t_k + k, t_{k+1}), \; I_k = [t_k, t_k + k)$$

and

$$U_k = a(n_k)^{-1} \Sigma_{\nu \in H_k} x_\nu, \; V_k = a(n_k)^{-1} \Sigma_{\nu \in I_k} x_\nu.$$

Since by Lemma 3.1 $a(n_k)^{-1} a(n_k - k) \to 1$ we obtain

$$U_k \to G \qquad \text{in distr.}$$

Since the U_k are separated by a gap of length k we can put $\phi_k = \phi(k)$ in Theorem 3. Hence without loss of generality there exists a sequence $\{Y_k, k \geq 1\}$ of independent random variables with common distribution G such that

$$P\{\|U_k - Y_k\| \geq 2(\rho_k + \phi_k)\} \leq 2(\rho_k + \phi_k). \tag{4.4}$$

Similarly $a(n_k)^{-1} a(k) \to 0$ and thus $V_k \to 0$ in Pr. We put

$$X_k = U_k + V_k = a(n_k)^{-1} \Sigma_{\nu \in H_k \cup I_k} x_\nu$$

and using (4.4) we obtain for all $k \geqslant k_0$

$$P\{\|X_k - Y_k\| \geqslant \varepsilon^6\} \leqslant \varepsilon^6.$$

We pick up the proof of Theorem 1 at (3.18) and obtain Theorem 4.

Acknowledgment. I am grateful to Jim Kuelbs for providing me with the proof of Lemma 2.2 and to Evarist Giné for a helpful discussion.

REFERENCES

BERKES, ISTVÁN and PHILIPP, WALTER (1979). Approximation theorems for independent and weakly dependent random vectors. *Ann. Probability* **7** 29–54.

BILLINGSLEY, PATRICK (1968). *Convergence of Probability Measures*. Wiley, New York.

BREIMAN, LEO (1968). *Probability*. Addison-Wesley, Reading, Mass.

DE ACOSTA, ALEJANDRO (1975). Stable measures and seminorms. *Ann. Probability* **3** 865–875.

DE ACOSTA, ALEJANDRO and GINÉ, EVARIST (1979). Convergence of moments and related functionals in the general central limit theorem in Banach spaces. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **48** 213–231.

DUDLEY, R. M. and KANTER, MAREK (1974). Zero-one laws for stable measures. *Proc. Amer. Math. Soc.* **45** 245–252.

FELLER, WILLIAM (1966). *An Introduction to Probability Theory and Its Applications*, vol II. Wiley, New York.

FREEDMAN, DAVID (1971). *Brownian Motion and Diffusion*. Holden-Day, San Francisco.

GNEDENKO B. V. and KOLMOGOROV, A. N. (1954). *Limit Distribution for Sums of Independent Random Variables*. Addison-Wesley, Reading, Mass.

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

IBRAGIMOV, I. A. and LINNIK, YU. V (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.

IOSIFESCU, M. and THEODORESCU, R. (1969). *Random Processes and Learning*. Springer, Berlin.

JAIN, NARESH C. (1977). Central limit theorem and related questions in Banach space. *Proc. Symp. Pure Math.* **31** 55–65, AMS, Providence, R.I.

KUELBS, J. and PHILIPP, WALTER (1980). Almost sure invariance principles for partial sums of mixing B-valued random variables. *Ann. Probability* **8**.

KUMAR, A. and MANDREKAR, V. (1974). Stable probability measures on Banach spaces. *Studia Math.* **42** 133–144.

MAJOR, PÉTER (1976a). Approximation of partial sums of i.i.d. rv's when the summands have only two moments. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **35** 221–229.

MAJOR, PÉTER (1976b). The approximation of partial sums of independent rv's. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **35** 213–220.

PHILIPP, WALTER (1979). An almost sure invariance principle for sums of B-valued random variables. *Lecture notes in Math.* **709** 171–193. Springer, Berlin, Heidelberg, New York.

RUDIN, WALTER (1973). *Functional Analysis*. McGraw Hill, New York.

SCHAEFER, HELMUT (1971). *Topological Vector Spaces*. Springer, Berlin.

SIMONS, GORDON and STOUT, WILLIAM (1978). A weak invariance principle with applications to domains of attraction. *Ann. Probability* **6** 294–315.

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

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS 61801