The Annals of Probability
1986, Vol. 14, No. 1, 260–270

ON STRONG INVARIANCE PRINCIPLES UNDER DEPENDENCE ASSUMPTIONS

BY ERNST EBERLEIN

Universität Freiburg

Strong invariance principles with order of approximation $O(t^{1/2-\kappa})$ are obtained for sequences of dependent random variables. The basic dependence assumptions include various generalizations of martingales such as asymptotic martingales (amarts), semiamarts, and mixingales as well as processes characterized by a condition on the Doléans measure. Provided the partial sum process is uniformly integrable, also martingales in the limit and games fairer with time are included. Sufficient conditions for linear growth of the covariance function of the partial sums are given.

1. Introduction and results. A number of strong invariance principles was obtained in recent years under various dependence assumptions on the underlying sequence. In particular, martingales, several classes of mixing sequences as well as very weak Bernoulli processes were investigated [see e.g., Berkes and Philipp (1979), Bradley (1983), Eberlein (1983), Kuelbs and Philipp (1980), Philipp and Morrow (1982), and Philipp and Stout (1975)]. As is explained in more detail in the introduction of the last mentioned reference, a strong invariance principle with error term $O(t^{1/2-\kappa})$ implies essentially all classical fluctuation results. We shall obtain this order of approximation in the following under two different sets of assumptions on the sequence.

Let $(x_k)_{k\geq 1}$ be a sequence of random variables. Denote $S_n(m) = \sum_{m+1}^{m+n} x_k$ and in particular $S_n = S_n(0)$. The basic assumption on the dependence structure of the sequence $(x_k)_{k\geq 1}$ considered in Theorem 1 is that there exists $\theta > 0$ such that uniformly in m

$$\|E[S_n(m)|\mathscr{F}_m]\|_1 \ll n^{1/2-\theta}, \tag{1.1}$$

where $\mathscr{F}_m$ is the σ-algebra generated by $x_1,\ldots,x_m$ and $\|\cdot\|_1$ denotes L^1-norm. The symbols " $\ll$ " and "$O(\cdot)$" are used with the same meaning here.

Trivially, martingale difference sequences satisfy condition (1.1). Various concepts to generalize martingales have been studied in the literature, in particular quasimartingales, asymptotic martingales or amarts, and semiamarts (Edgar and Sucheston, 1976). One has the inclusions {martingales} ⊂ {quasimartingales} ⊂ {amarts} ⊂ {semiamarts}. Let $(x_k)_{k\geq 1}$ be a real-valued semiamart difference sequence, i.e., $(S_n)_{n\geq 1}$ is a semiamart. This means if T denotes the set of bounded stopping times that $(E[S_\tau])_{\tau\in T}$ is bounded. By the Riesz decomposition theorem (Krengel and Sucheston, 1978), S_n can be written as $S_n = Y_n + Z_n$ where $(Y_n)_{n\geq 1}$ is a martingale and $(Z_n)_{n\geq 1}$ is a L^1-bounded semiamart. Using the martingale

Received March 1984; revised March 1985.

AMS 1980 *subject classifications*. Primary 60F17, 60G48; secondary 60F05, 60F15.

Key words and phrases. Strong invariance principle, martingale generalization, amart, semiamart, martingale in the limit, game fairer with time, Doléans measure, covariance function.

property and the fact that the conditional expectation is a L^1-contraction, we obtain for some constant $C > 0$

$$\|E[S_n(m)|\mathscr{F}_m]\|_1 \le \|Z_{n+m} - Z_m\|_1 \le C.$$

Thus semiamarts satisfy (1.1) even with a constant on the right side. If $(x_k)_{k\ge1}$ is a real-valued amart difference sequence again by the Riesz decomposition theorem (Edgar and Sucheston, 1976), S_n can be uniquely written as $S_n = Y_n + Z_n$ where $(Y_n)_{n\ge1}$ is a martingale and $(Z_n)_{n\ge1}$ is an amart converging to 0 almost surely as well as in L^1. This implies

$$\lim_{m\to\infty} \|E[S_n(m)|\mathscr{F}_m]\|_1 = 0 \tag{1.2}$$

uniformly in n.

Two further martingale generalizations are closely related. $(S_n)_{n\ge1}$ is called a game fairer with time if for all $\varepsilon > 0$

$$\lim_{m\to\infty} P\big[|E[S_n(m)|\mathscr{F}_m]| > \varepsilon\big] = 0$$

uniformly in n. $(S_n)_{n\ge1}$ is a martingale in the limit if in this definition convergence in probability is replaced by almost sure convergence. Edgar and Sucheston (1977) and Blake (1978) proved that every real-valued amart is a martingale in the limit. Thus one has the inclusions {amarts} ⊂ {martingales in the limit} ⊂ {games fairer with time}. Mucci (1973) [see also Subramanian (1973)] showed that a game fairer with time $(S_n)_{n\ge1}$ converges in L^1 if $(S_n)_{n\ge1}$ is uniformly integrable. (1.2) follows immediately in this case. If moreover $(|S_n|^p)_{n\ge1}$ is uniformly integrable for some $p > 1$ then $(S_n)_{n\ge1}$ converges in L^p which implies (1.2) with L^1-norm replaced by L^p-norm.

In Theorem 2 instead of (1.1) we shall assume

$$\|E[S_n(m)|\mathscr{F}_m]\|_2 \le C \tag{1.3}$$

for all $m, n \ge 1$ and some constant $C > 0$. Adapted mixingales (McLeish, 1975) with coefficients ψ_k such that $\sum_{k\ge1}\psi_k < \infty$ satisfy (1.3). Another interesting class having this property is characterized by a simple assumption on the corresponding Doléans measure. Assume that the x_k are integrable and let $]m, n] \times F$ be a predictable rectangle, i.e., m, n are positive integers, $m < n$ and $F \in \mathscr{F}_m$. Then the associated measure or Doléans measure of $(S_n)_{n\ge1}$ [see e.g., Métivier (1982)] is defined by

$$\lambda(]m, n] \times F) = E[1_F(S_n - S_m)].$$

Now if a process $(x_k)_{k\ge1}$ has the property that there exists a constant $C > 0$ such that for all predictable rectangles

$$|\lambda(]m, n] \times F)| \le CP(F), \tag{1.4}$$

then the process is contained in the class defined by (1.3). This follows immediately from the fact that (1.4) is equivalent to the assumption

$$|E[S_n(m)|\mathscr{F}_m]| \le C \quad \text{a.s.}$$

for all $m, n \ge 1$. The following class of sequences $(x_k)_{k\ge1}$ was considered by

Peligrad (1981). Assume that for each fixed m the sequence $(E[S_n(m)|\mathscr{F}_m])_{n\geq 1}$ converges to a function U_m in L^2 and assume that the sequence $(U_m^2)_{m\geq 1}$ is uniformly integrable. It is easy to see that these assumptions imply (1.3). A subclass discussed by Peligrad are martingales in the L^2 limit. These sequences are defined in the same way as martingales in the limit just by replacing almost sure convergence by convergence in L^2. Finally let us consider again an amart difference sequence $(x_k)_{k\geq 1}$. The Riesz decomposition given above has the property that Z_n converges to 0 almost surely. Therefore if there is a function Z in L^2 such that $|Z_n| \leq Z$, by dominated convergence (1.2) holds with L^1-norm replaced by L^2-norm. In particular (1.3) is true.

The reason for introducing the more restrictive assumption (1.3) in Theorem 2 is that it has implications on the variance behavior of the partial sum process. Linear growth of the variances of S_n is an essential ingredient in the central limit theory of dependent real-valued random variables. For the different types of mixing conditions as well as for very weak Bernoulli processes, it is obtained from suitable rates of the dependence coefficients with the help of correlation inequalities [see e.g., Billingsley (1968), Bradley (1981), Eberlein (1979, 1983), and Ibragimov and Linnik (1971)]. In those cases where the dependence structure is given by conditional expectations, linear growth is usually taken as an assumption (McLeish, 1975, Peligrad, 1981, Philipp and Stout, 1975, and Serfling, 1968). It will be shown in Section 2 that for weakly stationary sequences, (1.3) implies already that $n^{-1}\mathrm{Var}(S_n)$ converges to a finite positive limit provided $\mathrm{Var}(S_n) \to \infty$ as $n \to \infty$. Moreover, the speed of convergence is of the order $n^{-1/2}$. We shall actually consider a more general situation. Throughout Section 2 let B be a real separable Banach space with norm $\|\cdot\|$. Furthermore let B^* be the dual space with norm $\|\cdot\|_*$ and let $E^* \subset B^*$ denote the subset of vectors of norm 1. For B-valued random variables x_k denote by T_n the covariance function of $n^{-1/2}S_n$, i.e.,

$$T_n(f, g) = n^{-1}E[f(S_n)g(S_n)] \qquad (f, g \in B^*).$$

We shall investigate $(T_n(f, g))_{n\geq 1}$ denoting its limit by $T(f, g)$ if it exists. Let us recall that in this context a sequence $(x_k)_{k\geq 1}$ is weakly stationary if $E[f(x_1)g(x_n)] = E[f(x_{k+1})g(x_{k+n})]$ for all $k, n \geq 1$ and $f, g \in B^*$.

Let us write $S_n(m) = (S_n(m)_1, \dots, S_n(m)_d)$ in case the x_k are $\mathbb{R}^d$-valued. For the proof of the strong invariance principle itself which will be given for $\mathbb{R}^d$-valued random vectors we shall need an additional assumption. Suppose that there exists $\theta > 0$ such that uniformly in m

$$(1.5) \qquad \left\| E[S_n(m)_i S_n(m)_j | \mathscr{F}_m] - E[S_n(m)_i S_n(m)_j] \right\|_1 \ll n^{1-\theta}$$

for $1 \leq i, j \leq d$. In the following two theorems the covariances $(T_n(m))_{i,j}$ ($1 \leq i, j \leq d$) defined by $(T_n(m))_{i,j} = n^{-1}E[S_n(m)_i S_n(m)_j]$ will be considered. In particular, we shall write $(T_n)_{i,j} = (T_n(0))_{i,j}$. Denote by $\langle u, v\rangle$ the inner product of the vectors u and v.

THEOREM 1. *Let $(x_k)_{k\geq 1}$ be a sequence of mean zero $\mathbb{R}^d$-valued random vectors satisfying* (1.1) *for some* $0 < \theta < 1/2$. *We assume that there exists a* *covariance matrix T such that uniformly in m*

$$(T_n(m))_{i,j} - T_{i,j} \ll n^{-\rho} \tag{1.6}$$

for some $\rho > 0$ *and for all* $1 \le i, j \le d$. *Moreover suppose* (1.5) *and that there exists a constant* $M < \infty$ *and* $\delta > 0$ *such that* $E[\|x_k\|^{2+\delta}] \le M$, *then without loss of generality there exists a Brownian motion* $(X(t))_{t \ge 0}$ *with covariance matrix T such that*

$$\sum_{\nu \le t} x_\nu - X(t) \ll t^{1/2-\kappa} \quad a.s. \tag{1.7}$$

for some $\kappa > 0$.

The phrase "without loss of generality" here is to be understood in the sense that without changing its distribution we can redefine the sequence on a new probability space on which there exists a Brownian motion such that (1.7) holds.

THEOREM 2. *Let* $(x_k)_{k \ge 1}$ *be a weakly stationary sequence of mean zero* $\mathbb{R}^d$*-valued random vectors satisfying* (1.3). *Suppose that for each* $e \in \mathbb{R}^d$ *of length* 1, $\mathrm{Var}(\langle e, S_n \rangle) \ge r(n)$ *for some function* $r(n)$, $r(n) \to \infty$ *as* $n \to \infty$. *Then the covariances* $(T_n)_{i,j}$ *converge. Denote the limit by* $T = (T_{i,j})_{i,j}$ $(1 \le i, j \le d)$. *Suppose in addition* (1.5) *and that there exists a constant* $M < \infty$ *and* $\delta > 0$ *such that* $E[\|x_k\|^{2+\delta}] \le M$, *then without loss of generality there exists a Brownian motion* $(X(t))_{t \ge 0}$ *with covariance matrix T such that*

$$\sum_{\nu \le t} x_\nu - X(t) \ll t^{1/2-\kappa} \quad a.s.$$

for some $\kappa > 0$.

For dimension one the assumptions made in Theorem 1 are essentially those under which Serfling (1968) obtained a central limit theorem. The difference is that we assume a certain rate of convergence in (1.6). Theorem 1 will be proved in Section 3. The proof follows a method developed by Berkes and Philipp (1979) that applies directly to random vectors. It has been pointed out by the referee that in the real-valued case it is considerably shorter to prove the theorem by using a martingale approximation and then Strassen's martingale version of the Skorohod embedding theorem. This is the approach systematically exploited in Philipp and Stout (1975). Since the analysis of the covariance behavior in Section 2 provides us with (1.6) under the assumptions of Theorem 2, the latter is a consequence of Theorem 1.

2. Asymptotic covariance.

PROPOSITION 2.1. *Let* $(x_k)_{k \ge 1}$ *be a weakly stationary sequence of mean zero B-valued random variables satisfying* (1.3). *Suppose* $E[\|x_k\|^2] \le \rho$ *for all* $k \ge 1$ *and that there is a function* $r(n)$, $r(n) \to \infty$ *as* $n \to \infty$, *such that* $\mathrm{Var}(e(S_n)) \ge r(n)$ *for all* $e \in E^*$. *Then* $T(f, g)$ *exists for all* $f, g \in B^*$ *and*

$$T_m(f, g) - T(f, g) \ll \|f\|_* \|g\|_* m^{-1/2}, \tag{2.1}$$

where the constant implicitly given by $\ll$ *depends on* C, ρ, *and* $r(n)$ *only.*

REMARK. The uniform lower bound $r(n)$ for the variances of the real-valued partial sum processes $(e(S_n))_{n\geq 1}$ is needed to make the constant in (2.1) independent of f and g. The convergence result remains true if we assume only that $\mathrm{Var}(e(S_n)) \to \infty$ as $n \to \infty$ for all $e \in E^*$. But then the constant in general will depend on $f' = f/\|f\|_*$ and $g' = g/\|g\|_*$.

LEMMA 2.2. *If $e \in E^*$ then $T_n(e, e)$ satisfies for any $k \geq 1$*

$$|T_{kn}(e,e)/T_n(e,e) - 1| \leq 2C/r(n)^{1/2}.$$

In particular $T_n(e, e)$ is a slowly varying function of the integral variable n.

PROOF. Denote $y_j = \sum_{i=1}^{n} x_{(j-1)n+i}$ for $j = 1, \dots, k$ then by weak stationarity

$$\mathrm{Var}(e(S_{nk})) = k\,\mathrm{Var}(e(S_n)) + 2\sum_{l=1}^{k-1} E\left[e(y_1)e(y_2 + \cdots + y_{l+1})\right].$$

Because of (1.3) and Hölder's inequality

$$|E\left[e(y_1)e(y_2 + \cdots + y_{l+1})\right]| \leq C\,\mathrm{Var}(e(S_n))^{1/2},$$

which implies

$$|\mathrm{Var}(e(S_{nk}))/\mathrm{Var}(e(S_n)) - k| \leq 2(k-1)C/\mathrm{Var}(e(S_n))^{1/2}.$$

Dividing by k the result follows. □

LEMMA 2.3. *Given any $\varepsilon > 0$ there exist positive integers N and L both depending on C, ρ, and $r(n)$ only such that for all $l \geq L$ and $e \in E^*$ one has*

$$(1-\varepsilon)^2 T_N(e,e) \leq T_l(e,e) \leq (1+\varepsilon)^2 T_n(e,e).$$

PROOF. Since $r(n) \to \infty$ by the preceding lemma one can choose N large enough such that for any $m \geq 1$

$$1 - \varepsilon < T_{mN}(e,e)/T_N(e,e) < 1 + \varepsilon.$$

Now choose an integer M such that

$$\begin{aligned} 1 - \varepsilon &< (1-\varepsilon)^{1/2} - \left(N^2\rho/r(N)M\right)^{1/2} \\ &< (1+\varepsilon)^{1/2} + \left(N^2\rho/r(N)M\right)^{1/2} < 1 + \varepsilon. \end{aligned}$$

Set $L = NM$ then for any $l \geq L$ we can find an $m \geq M$ such that $mN \leq l \leq (m+1)N$. By Minkowski's inequality we have $\mathrm{Var}(e(S_j)) \leq N^2\rho$ for all $j = 1, \dots, N$. Therefore

$$\begin{aligned} (1-\varepsilon)T_N(e,e)^{1/2} &< \left((1-\varepsilon)^{1/2} - N\rho^{1/2}/(\mathrm{Var}(e(S_N))(m+1))^{1/2}\right)T_N(e,e)^{1/2} \\ &< ((m+1)N)^{-1/2} \\ &\quad \times\left(\mathrm{Var}\left(e(S_{(m+1)N})\right)^{1/2} - \mathrm{Var}\left(e(S_{(m+1)N-l})\right)^{1/2}\right) \\ &< T_l(e,e)^{1/2}. \end{aligned}$$

The other half of the inequality follows in the same way. □

PROOF OF PROPOSITION 2.1. It is an immediate consequence of Lemma 2.3 that $T(e, e)$ exists for all $e \in E^*$. Since $T_m(f, g)$ is a symmetric bilinear form this implies that $T(f, g)$ exists for all $f, g \in B^*$. Now let us notice that for $e \in E^*$ and for all $k, n \geq 1$

$$\left(1 - 2C/\operatorname{Var}(e(S_n))^{1/2}\right)T_n(e,e) \leq T_{kn}(e,e) \leq T_n(e,e)\left(1 + 2C/\operatorname{Var}(e(S_n))^{1/2}\right).$$

This can directly be seen from the proof of Lemma 2.2. The estimate

$$T_n(e,e) - T(e,e) \ll n^{-1/2}$$

follows from the last inequality by letting k tend to ∞ and by using Lemma 2.3 again. To finish the proof of (2.1) we have only to use the representation

$$T_m(f,g) = 2^{-1}(T_m(f+g, f+g) - T_m(f,f) - T_m(g,g))$$

and linearity. □

3. Proof of Theorem 1. Let $(x_k)_{k \geq 1}$ satisfy the assumptions made in Theorem 1. It is clear that we can assume $0 < \delta \leq 1$.

LEMMA 3.1. *There exists* $\varepsilon > 0$ *such that uniformly in* m

$$E\left[\|S_n(m)\|^{2+\varepsilon}\right] \ll n^{1+\varepsilon/2}.$$

ε *can be chosen as* $\varepsilon = 2\theta\delta/5$.

PROOF. The proof is a minor modification of the proof of Theorem 3.1 of Serfling (1968). Assumption (1.5) implies that uniformly in m

$$\left\| E\left[\|S_n(m)\|^2 | \mathscr{F}_m\right] - E\left[\|S_n(m)\|^2\right] \right\|_1 \ll dn^{1-\theta}, \tag{3.1}$$

whereas (1.6) entails that uniformly in m

$$E\left[\|S_n(m)\|^2\right] \ll dn. \tag{3.2}$$

These are the two inequalities needed in Serfling's proof. $\beta = \theta\delta/5$ is a suitable value in Lemma 2.1 of his paper, satisfying $\beta < \theta\delta/(2 + 2\theta + \delta)$. ε was chosen as 2β. □

Now define for any integer $k \geq 1$

$$N = N(k) = \left[k^{(1+1/\gamma)\lambda}\right], \qquad l = l(k) = \left[k^{(1+1/\gamma)(1-\lambda)}\right] \tag{3.3}$$

and $n = n(k) = Nl$ where $0 < \gamma \leq 1/120$ and $1/2 < \lambda < 1$ are values to be determined later and $[r]$ is the greatest integer less than or equal to the real number r. It is clear that for any $k > 1$

$$2^{-1}k^{1+1/\gamma} < n \leq k^{1+1/\gamma}. \tag{3.4}$$

We set $t_k = \sum_{i \leq k-1} n(i)$ and

$$X_k = n^{-1/2} S_n(t_k). \tag{3.5}$$

Denote furthermore $\mathscr{A}_{k-1} = \mathscr{F}_{t_k}$ then X_k is $\mathscr{A}_k$-measurable. Finally we put

$D_k = n^\alpha$ where $\alpha = \theta\delta\rho/(4 + 2d)(5\rho + \theta\delta)$. For the proof of our theorem, we will eventually apply Theorem 1 of Berkes and Philipp (1979) to the sequence $\{X_k, \mathscr{A}_k; k \geq 1\}$. The crucial part is therefore to get a good estimate λ_k such that

$$E\left[\left|E\left[\exp(i\langle u, X_k\rangle)|\mathscr{A}_{k-1}\right] - \exp(-\langle u, Tu\rangle/2)\right|\right] \leq \lambda_k \tag{3.6}$$

for all u with $\|u\| \leq D_k$. For this purpose let k be fixed. The left-hand side of (3.6) can be written as

$$\begin{aligned} &E\Big[\Big|E\left[\exp(i\langle u, X_k\rangle)(1 - \exp(\langle u, T_n(t_k)u\rangle/2 - \langle u, Tu\rangle/2))|\mathscr{A}_{k-1}\right] \\ &+E\left[\exp(\langle u, T_n(t_k)u\rangle/2 - \langle u, Tu\rangle/2 + i\langle u, X_k\rangle) - \exp(-\langle u, Tu\rangle/2)|\mathscr{A}_{k-1}\right]\Big|\Big] \\ &\quad\leq \left|1 - \exp(\langle u, (T_n(t_k) - T)u\rangle/2)\right| \\ &\quad\quad+\exp(-\langle u, Tu\rangle/2)E\left[\left|E\left[\exp(\langle u, T_n(t_k)u\rangle/2 + i\langle u, X_k\rangle) - 1|\mathscr{A}_{k-1}\right]\right|\right] \\ &\quad= H_1 + H_2. \end{aligned}$$

An estimate for H_1 is easy to obtain from the inequality

$$|e^x - 1| \leq 2|x|, \qquad x \leq 1/2. \tag{3.7}$$

Since (1.6) implies $|\langle u, (T_n(t_k) - T)u\rangle| \ll \|u\|^2 dn^{-\rho}$, we get for all u with $\|u\| \leq D_k$

$$H_1 \leq \left|\langle u, (T_n(t_k) - T)u\rangle\right| \ll n^{-(\rho - 2\alpha)} = n^{-\gamma(0)}. \tag{3.8}$$

Here we have written $\gamma(0)$ for the exponent for short. The same notation $\gamma(i)$ will be introduced for the exponents in the following estimates. In order to estimate H_2 we define for $j = 1, \dots, l$

$$X_{k,j} = n^{-1/2}S_N(t_k + (j-1)N), \qquad \mathscr{A}_{k,j} = \mathscr{F}_{t_k + jN}$$

and

$$\begin{aligned} \zeta_{n,j} = \exp&\left(i\langle u, \sum_{\nu=1}^{j-1} X_{k,\nu}\rangle + jN\langle u, T_n(t_k)u\rangle/2n\right) \\ &\cdot\left\{\exp(i\langle u, X_{k,j}\rangle) - \exp(-\langle u, T_n(t_k)u\rangle/2l)\right\} \end{aligned}$$

where $\Sigma_{\nu=1}^{0}$ is understood to be 0. Since $\Sigma_{\nu=1}^{l} X_{k,\nu} = X_k$ we deduce

$$\sum_{j=1}^{l} \zeta_{n,j} = \exp(i\langle u, X_k\rangle + \langle u, T_n(t_k)u\rangle/2) - 1. \tag{3.9}$$

By definition $\mathscr{A}_{k-1} \subset \mathscr{A}_{k,j}$ for all j, therefore,

$$\exp(\langle u, Tu\rangle/2)H_2 \leq \sum_{j=1}^{l} E\left[\left|E\left[\zeta_{n,j}|\mathscr{A}_{k,j-1}\right]\right|\right]. \tag{3.10}$$

The first factor in the definition of $\zeta_{n,j}$ is $\mathscr{A}_{k,j-1}$-measurable. This together with the fact that $jN/n \le 1$ yields

$$\left|E\left[\zeta_{n,j}|\mathscr{A}_{k,j-1}\right]\right| \le \exp(\langle u, T_n(t_k)u\rangle/2) \times \left|E\left[\exp(i\langle u, X_{k,j}\rangle) - \exp(-\langle u, T_n(t_k)u\rangle/2l)|\mathscr{A}_{k,j-1}\right]\right|.$$

Introducing this in (3.10) we see that

$$H_2 \le \exp(\langle u, (T_n(t_k) - T)u\rangle/2) \cdot \left(\sum_{j=1}^{l} E\left[\left|E\left[\exp(i\langle u, X_{k,j}\rangle) - \exp(-\langle u, T_n(t_k)u\rangle/2l)|\mathscr{A}_{k,j-1}\right]\right|\right]\right). \tag{3.11}$$

Again with the help of (3.7) we see that the first factor is less than $1 + n^{-\gamma(0)} \le 2$. For the second factor we use the expansions

$$e^{ix} = 1 + ix - x^2/2 + x^2Q(x)/2 \quad \text{and} \quad e^{-x} = 1 - x + N(x),$$

where $|N(x)| \le x^2/2$ for all $x \ge 0$ and $|Q(x)| \le \min(|x|/3, 2)$ for all x [see e.g., Gänssler and Stute (1977), page 364] and get the following upper bound for the jth term in this sum

$$\begin{aligned} &\|u\|E\left[\left|E\left[\langle e, X_{k,j}\rangle|\mathscr{A}_{k,j-1}\right]\right|\right] \\ &\quad + 2^{-1}\|u\|^2E\left[\left|E\left[\langle e, X_{k,j}\rangle^2|\mathscr{A}_{k,j-1}\right] - \langle e, T_n(t_k)e\rangle/l\right|\right] \\ &\quad + 2^{-1}\|u\|^2E\left[\left|E\left[\langle e, X_{k,j}\rangle^2 Q(\langle u, X_{k,j}\rangle)|\mathscr{A}_{k,j-1}\right]\right|\right] \\ &\quad + \left|N(\langle u, T_n(t_k)u\rangle/2l)\right| = I_{j,1} + I_{j,2} + I_{j,3} + I_{j,4}. \end{aligned}$$

We have introduced $e = u/\|u\|$ here. Remember that we consider u with $\|u\| \le n^{\alpha}$. Assumption (1.1) implies

$$\begin{aligned} \sum_{j=1}^{l} I_{j,1} &\ll \|u\| l n^{-1/2} N^{1/2-\theta} \ll n^{\alpha + (1-\lambda) - 1/2 + \lambda(1/2-\theta)} \\ &= n^{-(\lambda/2 - 1/2 - \alpha + \lambda\theta)} = n^{-\gamma(1)}. \end{aligned} \tag{3.12}$$

Here we made use of the relations

$$n^{\lambda} \ll N \ll n^{\lambda} \quad \text{and} \quad n^{1-\lambda} \ll l \ll n^{1-\lambda},$$

which follow from (3.3) and (3.4). (1.6) implies that $|\langle e, T_n(t_k)e\rangle| \ll d$. Using the estimate for $N(x)$ we derive

$$\begin{aligned} \sum_{j=1}^{l} I_{j,4} &\le \sum_{j=1}^{l} \|u\|^4 |\langle e, T_n(t_k)e\rangle|^2/8l^2 \\ &\ll \|u\|^4 l^{-1} \ll n^{-(1-\lambda-4\alpha)} = n^{-\gamma(4)}. \end{aligned} \tag{3.13}$$

Now we turn to $I_{j,2}$.

$$\begin{aligned} 2\|u\|^{-2} I_{j,2} \le n^{-1}E\Big[\Big|&E\left[\langle e, S_N(t_k + (j-1)N)\rangle^2|\mathscr{F}_{t_k+(j-1)N}\right] \\ &- E\left[\langle e, S_N(t_k + (j-1)N)\rangle^2\right]\Big|\Big] \\ &+ l^{-1}|\langle e, T_N(t_k + (j-1)N)e\rangle - \langle e, T_n(t_k)e\rangle|. \end{aligned}$$

By (1.5) the first term is bounded by $C_1 n^{-1} dN^{1-\theta}$ for some constant $C_1 > 0$.

Using (1.6) twice we see that the second term is bounded by

$$C_2 l^{-1} d(N^{-\rho} + n^{-\rho}) \ll l^{-1} N^{-\rho} \quad \text{where } C_2 > 0 \text{ is another constant.}$$

Consequently,

$$\begin{aligned} \sum_{j=1}^{l} I_{j,2} &\ll 2^{-1}\|u\|^2 (N^{1-\theta} l n^{-1} + N^{-\rho}) \\ &\ll n^{-(\theta\lambda - 2\alpha)} + n^{-(\lambda\rho - 2\alpha)} \\ &= n^{-\gamma(2)} + n^{-\gamma(5)}. \end{aligned} \tag{3.14}$$

For the remaining term $I_{j,3}$ we shall need the inequality given for $Q(x)$. Now

$$2\|u\|^{-2} I_{j,3} \le E\left[|\langle e, X_{k,j}\rangle|^2 |Q(\langle u, X_{k,j}\rangle)|\right].$$

We split this expectation in the two parts $A = \{\|X_{k,j}\| > 1\}$ and $A^c = \{\|X_{k,j}\| \le 1\}$ and apply the inequality $|Q(x)| \le 2$ over A and $|Q(x)| \le |x|/3$ over A^c. By Lemma 3.1 we get in the first case

$$\begin{aligned} \int_A &\le 2n^{-1} E\left[\|S_n(t_k + (j-1)N)\|^2 1_A\right] \\ &\ll n^{-(1+\varepsilon/2)} N^{1+\varepsilon/2} = l^{-(1+\varepsilon/2)} \end{aligned}$$

and in the second case

$$\begin{aligned} \int_{A^c} &\le 3^{-1}\|u\| E\left[\|X_{k,j}\|^{2+\varepsilon}\right] \\ &\ll \|u\| n^{-(1+\varepsilon/2)} N^{1+\varepsilon/2} = \|u\| l^{-(1+\varepsilon/2)}. \end{aligned}$$

Putting the two results together yields

$$\begin{aligned} \sum_{j=1}^{l} I_{j,3} &\ll \|u\|^3 l^{-\varepsilon/2} \le n^{3\alpha} l^{-\theta\delta/5} \\ &\ll n^{-((1-\lambda)\theta\delta/5 - 3\alpha)} = n^{-\gamma(3)}. \end{aligned} \tag{3.15}$$

From (3.8) and (3.12)–(3.15) we conclude that there exists a constant M_0 such that

$$\lambda_k = M_0 n^{-\gamma'}, \tag{3.16}$$

where $\gamma' = \min_{0 \le i \le 5} \gamma(i)$. By Theorem 1 of Berkes and Philipp (1979) we can redefine the sequence $(X_k)_{k \ge 1}$ on a richer probability space together with a sequence $(Y_k)_{k \ge 1}$ of independent $N(0, T)$-distributed random vectors such that

$$P\left[\|X_k - Y_k\| \ge \alpha_k\right] \le \alpha_k,$$

where

$$\alpha_k \ll d n^{-\alpha} \log n^{\alpha} + n^{-\gamma'/2 + \alpha d} + P\left[\|N(0,T)\| > n^{\alpha}/4\right].$$

Since (1.6) remains true if we make ρ smaller we can assume $\rho \le 2\theta^2\delta/(5 + 2\theta\delta)$. Now if we choose $\lambda = \theta\delta/(5\rho + \theta\delta)$ and $\gamma = \alpha/2$, we see by elementary computations from (3.8) and (3.12)–(3.15) that

$$\gamma \le \gamma(i)/2 - \alpha d \quad \text{for each } i = 0, \dots, 5. \tag{3.17}$$

This means $\gamma \le \gamma'/2 - \alpha d$ therefore

$$\alpha_k \ll n^{-\gamma} \ll k^{-(1+\gamma)},$$

which implies by the Borel–Cantelli lemma

$$\|X_k - Y_k\| \ll k^{-(1+\gamma)} \quad \text{a.s.} \tag{3.18}$$

Since $(Y_k)_{k \ge 1}$ are independent Gaussian random vectors, we can assume without loss of generality that there exists a $\mathbb{R}^d$-valued Brownian motion $(X(t))_{t \ge 0}$ with mean zero and covariance matrix T satisfying

$$(t_{k+1} - t_k)^{-1/2}(X(t_{k+1}) - X(t_k)) = Y_k \qquad (k \ge 1).$$

Therefore, by (3.4) and (3.18)

$$\left\| \sum_{\nu = t_k + 1}^{t_{k+1}} x_\nu - (X(t_{k+1}) - X(t_k)) \right\| = n(k)^{1/2} \|X_k - Y_k\| \ll k^{(1+\gamma)(1/2\gamma - 1)} \quad \text{a.s.}$$

If we sum over k and use $k^{2+\gamma^{-1}} \ll t_{k+1}$ we get

$$\begin{aligned} \left\| \sum_{\nu=1}^{t_{k+1}} x_\nu - X(t_{k+1}) \right\| &\ll \sum_{j=1}^{k} j^{(1+\gamma)(1/2\gamma - 1)} \le k^{(1+\gamma)(1/2\gamma - 1) + 1} \\ &\ll t_{k+1}^{1/2 - \gamma/2} \quad \text{a.s.} \end{aligned}$$

Given $t > 0$ we choose k such that $t_k < t \le t_{k+1}$ then

$$\left\| \sum_{\nu \le t} x_\nu - X(t) \right\| \le \left\| \sum_{\nu=1}^{t_k} x_\nu - X(t_k) \right\| + \left\| \sum_{\nu = t_k + 1}^{t} x_\nu \right\| + \|X(t) - X(t_k)\|.$$

Our proof is finished if we show that there exists $\kappa > 0$ such that with probability 1

$$\max_{t_k < t \le t_{k+1}} \left\| \sum_{\nu = t_k + 1}^{t} x_\nu \right\| \ll t_k^{1/2 - \kappa}$$

and

$$\sup_{t_k < t \le t_{k+1}} \|X(t) - X(t_k)\| \ll t_k^{1/2 - \kappa}.$$

The first of these two statements is shown analogous to the proof of Proposition 2.2 in Kuelbs and Philipp (1980). Here we need Lemma 3.1. The second statement is standard. □

REMARK. The verification of (3.17) shows that the relevant terms in the estimation are (3.15) and the second term in (3.14), i.e., the exponents $\gamma(3)$ and $\gamma(5)$. This is because we considered without loss of generality small values of ρ. For large ρ, $\gamma(5)$ is no longer relevant. In order to get the best (i.e., largest) value of γ in this case, one has to choose λ and α appropriately such that $\gamma(1)$ and $\gamma(3)$ become large. The best choice for λ in this situation is $(2\theta\delta + 5)/(2\theta\delta + 5 + 10\theta)$.

Acknowledgment. I wish to thank the referee for his clarifying comments which helped to improve the paper.

REFERENCES

BERKES, I. and PHILIPP, W. (1979). Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.* **7** 29–54.

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

BLAKE, L. (1978). Every amart is a martingale in the limit. *J. London Math. Soc.* **18** 381–384.

BRADLEY, R. (1981). A sufficient condition for linear growth of variances in a stationary random sequence. *Proc. Amer. Math. Soc.* **83** 586–589.

BRADLEY, R. (1983). Approximation theorems for strongly mixing random variables. *Michigan Math. J.* **30** 69–81.

EBERLEIN, E. (1979). An invariance principle for lattices of dependent random variables. *Z. Wahrsch. verw. Gebiete* **50** 119–133.

EBERLEIN, E. (1983). Strong approximation of very weak Bernoulli process. *Z. Wahrsch. verw. Gebiete* **62** 17–37.

EDGAR, G. and SUCHESTON, L. (1976). Amarts: a class of asymptotic martingales. A. Discrete parameter. *J. Multivariate Anal.* **6** 193–221.

EDGAR, G. and SUCHESTON, L. (1977). Martingales in the limit and amarts. *Proc. Amer. Math. Soc.* **67** 315–320.

GÄNSSLER, P. and STUTE, W. (1977). *Wahrscheinlichkeitstheorie.* Springer, Berlin.

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

KRENGEL, U. and SUCHESTON, L. (1978). On semiamarts, amarts, and processes with finite value. In *Probability on Banach Spaces.* (J. Kuelbs, ed.) *Advances in Probability and Related Topics* **4** 197–266. Marcel Dekker, New York.

KUELBS, J. and PHILIPP, W. (1980). Almost sure invariance principles for partial sums of mixing B-valued random variables. *Ann. Probab.* **8** 1003–1036.

MCLEISH, D. L. (1975). Invariance principles for dependent variables. *Z. Wahrsch. verw. Gebiete* **32** 165–178.

MÉTIVIER, M. (1982). *Semimartingales.* de Gruyter, Berlin, New York.

MUCCI, A. G. (1973). Limits for martingale-like sequences. *Pacific J. Math.* **48** 197–202.

PELIGRAD, M. (1981). An invariance principle for dependent random variables. *Z. Wahrsch. verw. Gebiete* **57** 495–507.

PHILIPP, W. and MORROW, G. (1982). An almost sure invariance principle for Hilbert space valued martingales. *Trans. Amer. Math. Soc.* **273** 231–251.

PHILIPP, W. and STOUT, W. (1975). Almost sure invariance principles for partial sums of weakly dependent random variables. *Mem. Amer. Math. Soc.* **161** 1–140.

SERFLING, R. J. (1968). Contributions to central limit theory for dependent variables. *Ann. Math. Statist.* **39** 1158–1175.

SUBRAMANIAN, R. (1973). On a generalization of martingales due to Blake. *Pacific J. Math.* **48** 275–278.

INSTITUT F. MATH. STOCHASTIK
UNIVERSITÄT FREIBURG
HEBELSTRASSE 27
D-7800 FREIBURG I. BR., WEST GERMANY