

# THE INVARIANCE PRINCIPLE FOR DEPENDENT RANDOM VARIABLES

BY

PATRICK BILLINGSLEY

**1. Introduction.** The purpose of this paper is to extend to the dependent case the Erdős-Kac invariance principle [6; 7; 11], as generalized by Donsker [4]. Let  $C$  be the space of functions  $x(t)$  continuous on the closed unit interval, let  $\rho$  be the metric on  $C$  defined by  $\rho(x, y) = \sup_t |x(t) - y(t)|$ , let  $\mathcal{C}$  be the Borel field generated by the open sets and let  $W$  be Wiener measure on  $\mathcal{C}$ . Let  $\{X_1, X_2, \dots\}$  be a sequence of random variables on some probability measure space  $(\Omega, \mathfrak{B}, P)$ . Let  $S_n = X_1 + \dots + X_n$ ,  $S_0 = 0$  and let  $p_n$  be that element of  $C$  which is linear on each of the intervals  $((j-1)n^{-1}, jn^{-1})$ ,  $j = 1, \dots, n$ , and satisfies  $p_n(jn^{-1}) = S_j$  for  $j = 0, 1, \dots, n$ . That is, let  $p_n$  be the random function with the value

$$(1.1) \quad p_n(t) = S_j + (nt - j)X_{j+1}$$

if  $jn^{-1} \leq t \leq (j+1)n^{-1}$ ,  $j = 0, \dots, n-1$ . Thus  $p_n$  is a mapping of  $\Omega$  into  $C$ . It is easy to show that this mapping is measurable. Donsker's theorem, which subsumed previous results in this direction, states that if  $f$  is a function on  $C$  continuous except on a set of  $W$ -measure zero and if the sequence  $\{X_n\}$  is independent and stationary with  $E\{X_n\} = 0$  and  $E\{X_n^2\} = 1$ , then

$$\lim_{n \rightarrow \infty} P\{f(n^{-1/2}p_n) \leq \alpha\} = W\{x: f(x) \leq \alpha\}$$

at continuity points  $\alpha$  of the function  $W\{x: f(x) \leq \alpha\}$ . If  $f(x) = \sup_t x(t)$  for example, this leads to a limit theorem for  $\max_{k \leq n} S_k$ . See [4] for other functions  $f$  which lead to interesting limit theorems.

It should be pointed out that in place of the random element of  $C$  defined by (1.1), Donsker actually worked with the "random step function" with value  $S_j$  throughout the interval  $((j-1)n^{-1}, jn^{-1}]$ . There is of course no real difference between the two methods and one is led to essentially the same limit theorems.

There is another way of stating Donsker's result. Suppose there exists a sequence  $\{a_n\}$  of positive constants such that if  $P_n$  is a measure defined by setting  $P_n(A) = P\{a_n^{-1}p_n \in A\}$  for  $A \in \mathcal{C}$ , then  $P_n$  converges weakly to  $W$ . When this is true we say that the *invariance principle* holds for the sequence  $\{X_n\}$  with *norming factors*  $\{a_n\}$ . Then (cf. Theorem 2.1 below) Donsker's result is that the invariance principle holds, with norming factors  $n^{1/2}$ , provided  $\{X_n\}$  is an independent, stationary sequence with  $E\{X_n\} = 0$  and

$E\{X_n^2\}=1$ . The assumption that  $\{X_n\}$  is stationary is relatively unimportant. It is the purpose of the present paper to replace the assumption of independence by various weaker hypotheses.

In §2 we prove generalizations (Theorems 2.2 and 2.3) of some results implicit in [4]. The general line of attack is similar to that of [4], but the proofs are improved. In §3 these theorems are applied to the derivation of a general set of conditions on the sequence  $\{X_n\}$  which insure that the invariance principle holds with a suitable sequence of norming factors. These conditions are ones which can be verified for those dependent sequences of greatest interest. (The elegant method of Prohorov [13] seems difficult to apply to dependent sequences.)

In §§4 through 7 the result of §3 is specialized in various ways. In §4 the invariance principle is proved for sequences  $\{f(x_n)\}$ , where  $f$  is a function defined on the state space of a discrete Markov process  $\{x_n\}$  satisfying Doeblin's hypothesis. The conditions under which this result is obtained are identical with those under which the central limit theorem (CLT) for such sequences is proved in [5]. In §5 we prove the invariance principle for  $m$ -dependent sequences of random variables. This result is obtained under assumptions slightly stronger than those of Marsaglia's CLT for  $m$ -dependent sequences [12]. The next section treats of discrete linear processes with  $m$ -dependent residuals, processes which arise in the analysis of time series. Here we prove the invariance principle under conditions only a little stronger than those assumed by Diananda [3] in his proof of the CLT for processes of this kind. Finally, in §7 we prove the invariance principle for the number of occurrences of a recurrent event. Here we assume that the recurrence time has a finite second moment.

It is possible to prove the invariance principle in cases other than those considered here. One can, for example, prove it for martingales, as Lévy [10] has the CLT, or under the assumptions of Bernstein's *lemme fondamental* [2]. Although no applications have been essayed, the processes treated are those of greatest interest for the applications.

This paper is part of a doctoral dissertation presented to the faculty of Princeton University. The author would like to thank Professor Feller for his help and encouragement.

**2. Weak convergence of measures.** In this section we prove several useful theorems on the weak convergence of measures. Consider first an arbitrary metric space  $\mathfrak{X}$  with metric  $\rho$ . If  $P_n, P$  are probability measures on the Borel sets, we say that  $P_n$  converges weakly to  $P$  (in symbols  $P_n \Rightarrow P$ ) if  $\int f dP_n \rightarrow \int f dP$  for all bounded continuous functions  $f$ . (We omit the region of integration when it is the entire space.)

Theorem 2.1 gives several convenient sets of conditions equivalent to weak convergence. For its proof we require a lemma related to that of Urysohn.

LEMMA 2.1. *If  $A$  and  $B$  are sets with  $\rho(A, B) > 0$ , then there exists a function which is 1 on  $A$ , 0 on  $B$ , everywhere between 0 and 1 and uniformly continuous on  $\mathfrak{X}$ .*

**Proof.** We may of course assume that  $A$  and  $B$  are nonempty. With the exception of uniform continuity, it is clear that the function

$$f(x) = \frac{\rho(x, B)}{\rho(x, B) + \rho(x, A)}$$

has the required properties. It follows from standard inequalities (cf. [1, p. 57]) that  $\rho(x, A)$  and  $\rho(x, B)$  are uniformly continuous and that  $\rho(x, B) + \rho(x, A) \geq \rho(A, B)$ . The function of two non-negative real variables defined by the formula  $\xi/(\xi + \eta)$  is uniformly continuous on any domain on which the denominator is bounded away from zero. Since a uniformly continuous function of uniformly continuous functions is uniformly continuous,  $f(x)$  is uniformly continuous.

In what follows we denote the boundary of a set  $A$  by  $\tilde{A}$ . If  $P$  is a probability measure on the Borel sets of  $\mathfrak{X}$  and  $f$  is a Borel measurable function then  $P\{x: f(x) \leq \alpha\}$  is a function of  $\alpha$  which we call the  $P$ -distribution of  $f$ .

THEOREM 2.1. *The following statements are equivalent.*

- (i)  $P_n \Rightarrow P$ .
- (ii)  $\int f dP_n \rightarrow \int f dP$  for all bounded, uniformly continuous functions  $f$ .
- (iii)  $P(A) = \lim_n P_n(A)$  for any Borel set  $A$  such that  $P(\tilde{A}) = 0$ .
- (iv) *For any measurable function  $f$  which is continuous except on a set of  $P$ -measure zero, the  $P_n$ -distribution of  $f$  converges to the  $P$ -distribution of  $f$  at each continuity point of the latter.*
- (v) *For any bounded, uniformly continuous function  $f$  the  $P_n$ -distribution of  $f$  converges to the  $P$ -distribution of  $f$  at each continuity point of the latter.*

**Proof.** The implications (i)  $\rightarrow$  (ii), (iii)  $\rightarrow$  (iv), (iv)  $\rightarrow$  (v), (iv)  $\rightarrow$  (i) and (v)  $\rightarrow$  (ii) being simple to prove, we treat here only the implication (ii)  $\rightarrow$  (iii). Clearly (iii) follows if

$$(2.1) \quad P(A) \geq \limsup_n P_n(A)$$

holds for all closed sets  $A$ . Suppose (ii) holds and  $A$  is closed. Then it is possible, given  $\epsilon > 0$ , to find a  $\delta > 0$  such that if  $U = \{x: \rho(x, A) < \delta\}$  then  $P(U - A) < \epsilon$ . Clearly  $\rho(A, \mathfrak{X} - U) \geq \delta$ . Hence, by Lemma 2.1, there exists a uniformly continuous function  $f$  which is 1 on  $A$ , 0 on  $\mathfrak{X} - U$  and everywhere between 0 and 1. Now  $\int f dP_n \rightarrow \int f dP$  by assumption and  $\int f dP_n \geq P_n(A)$ , while  $\int f dP \leq P(A) + \epsilon$ . From these three relations it follows that  $\limsup_n P_n(A) \leq P(A) + \epsilon$ . Since  $\epsilon$  was arbitrary, we have (2.1).

Let  $C$ ,  $\rho$  and  $\mathfrak{C}$  be defined as in §1. Theorem 2.2 below gives a simple criterion for the weak convergence of measures on  $\mathfrak{C}$ . For its proof we need a

lemma which is a variation on a result due to Donsker [4]. If  $x \in C$  and  $c$  is a positive integer, define, for  $j=1, \dots, c$ ,

$$a_j(x) = \inf \{x(t) : (j-1)c^{-1} \leq t \leq jc^{-1}\},$$

$$b_j(x) = \sup \{x(t) : (j-1)c^{-1} \leq t \leq jc^{-1}\},$$

and let  $\pi_c(x) = (a_1(x), \dots, a_c(x), b_1(x), \dots, b_c(x))$ . Thus  $\pi_c$  maps  $C$  into  $2c$ -space. If  $\phi$  is bounded and continuous on  $2c$ -space, then  $\phi(\pi_c(x))$  is a bounded continuous function on  $C$ . Let  $\mathfrak{A}$  be the set of functions on  $C$  arising in this way.

LEMMA 2.2. *Let  $f$  be a bounded, uniformly continuous function on  $C$ . There exist two sequences  $\{f'_c\}$ ,  $\{f''_c\}$  of uniformly bounded functions in  $\mathfrak{A}$  such that*

$$(2.2) \quad f'_c(x) \leq f(x) \leq f''_c(x),$$

for all  $c$  and  $x$  and such that for all  $x$

$$(2.3) \quad \lim_{c \rightarrow \infty} (f''_c(x) - f'_c(x)) = 0.$$

**Proof.** For each  $c$  let  $M_{x,c}$  be the set of  $y \in C$  such that  $a_j(x) \leq y(t) \leq b_j(x)$  for  $(j-1)c^{-1} \leq t \leq jc^{-1}$ ,  $j=1, \dots, c$ . Now define

$$f'_c(x) = \inf \{f(y) : y \in M_{x,c}\},$$

$$f''_c(x) = \sup \{f(y) : y \in M_{x,c}\}.$$

It is clear that  $f'_c$  and  $f''_c$  satisfy (2.2) and that they are bounded by the bound of  $f$ . And (2.3) follows from the uniform continuity of  $f$ .

There remains only the proof that  $f'_c$  and  $f''_c$  belong to  $\mathfrak{A}$ . Consider the case of  $f'_c$ . Let  $S$  be the set of points  $\zeta = (\zeta_1, \dots, \zeta_{2c})$  of  $2c$ -space such that  $\zeta_j \leq \zeta_{c+j}$  for  $j=1, \dots, c$  and such that the closed intervals  $[\zeta_j, \zeta_{j+c}]$  and  $[\zeta_{j+1}, \zeta_{c+j+1}]$  have a nonempty intersection for  $j=1, \dots, c-1$ . Then  $S$  is closed. For  $\zeta \in S$  define  $\phi(\zeta) = \sup f(y)$ , where the supremum is extended over the set of  $y \in C$  for which  $\zeta_j \leq y(t) \leq \zeta_{c+j}$  if  $j-1 \leq tc \leq j$ , for  $j=1, \dots, c$ . Obviously  $\pi_c(x) \in S$  for all  $x \in C$ ,  $\phi(\pi_c(x)) = f'_c(x)$  and  $\phi$  is bounded. If  $\phi$  is continuous on  $S$  then it is possible by Urysohn's extension theorem [1] to extend  $\phi$  to all of  $2c$ -space in such a way that it remains bounded and continuous. Hence the proof can be completed by showing that  $\phi$  is continuous on  $S$ . This part can be established by the methods of the proof of Theorem (3.1) of [4]. We omit the details.

For any integer  $c$  and real numbers  $\alpha_1, \dots, \alpha_c, \beta_1, \dots, \beta_c$ , consider the set

$$(2.4) \quad E = \{x : \alpha_j \leq x(t) \leq \beta_j, j-1 \leq ct \leq j, j=1, \dots, c\}.$$

THEOREM 2.2. *Suppose that for probability measures  $P_n$  and  $P$  on  $\mathfrak{C}$  we have  $P_n(E) \rightarrow P(E)$  for all sets  $E$  of the form (2.4) for which  $P(\bar{E}) = 0$ . Then  $P_n \Rightarrow P$ .*

**Proof.** We show first that

$$(2.5) \quad \int f dP_n \rightarrow \int f dP$$

for any function  $f$  in  $\mathfrak{A}$ . For a fixed integer  $c$  define  $\mu_n(S) = P_n(\pi_c^{-1}(S))$  and  $\mu(S) = P(\pi_c^{-1}(S))$  for  $2c$ -dimensional Borel sets  $S$ . If  $S$  is the set of  $\zeta = (\zeta_1, \dots, \zeta_{2c})$  such that  $\zeta_j \geq \alpha_j$  and  $\zeta_{c+j} \leq \beta_j$ ,  $j=1, \dots, c$ , then  $\pi_c^{-1}(S) = E$ , so that  $\mu_n(S) \rightarrow \mu(S)$ , provided  $\mu(\tilde{S}) = 0$ . But this obviously implies  $\mu_n \Rightarrow \mu$ . Hence

$$(2.6) \quad \int \phi d\mu_n \rightarrow \int \phi d\mu$$

for any bounded continuous function  $\phi$ . But if  $f(x) = \phi(\pi_c(x))$ , (2.5) follows from (2.6) by a transformation of the integrals involved. Hence (2.5) holds for any function  $f$  of  $\mathfrak{A}$ . It now follows easily, by Lemma 2.2, that (2.5) holds for any bounded, uniformly continuous function  $f$ . Hence by Theorem 2.1 ((ii)  $\rightarrow$  (i)) we have  $P_n \Rightarrow P$ .

Let  $A_n$  be the set of functions  $x \in C$  which are linear on each of the intervals  $((i-1)n^{-1}, in^{-1})$  for  $i=1, \dots, n$ , and satisfy  $x(0)=0$ . Since the  $p_n$  defined in §1 lies in  $A_n$  it is of interest to specialize Theorem 2.2 to the case where  $P_n(A_n) = 1$ .

**THEOREM 2.3.** *Suppose that for probability measures  $P_n$  and  $P$  on  $\mathfrak{C}$  we have  $P_n(A_n) = 1$ . Suppose further that  $P_n(G_n) \rightarrow P(E)$  provided  $E$  is any set of the form (2.4) for which  $P(\tilde{E}) = 0$  and  $G_n$  is the set of  $x$  for which  $\alpha_j \leq x(in^{-1}) \leq \beta_j$  if  $j-1 \leq icn^{-1} \leq j$ ,  $j=1, \dots, c$ . Then  $P_n \Rightarrow P$ .*

**Proof.** Let  $\epsilon$  be a small positive rational and let  $E_{\epsilon, \delta}$  be the set where  $\alpha_j + \delta \leq x(t) \leq \beta_j - \delta$  if  $(j-1)c^{-1} + \epsilon \leq t \leq jc^{-1} - \epsilon$  ( $j=2, \dots, c$ ), where  $\alpha_1 + \delta \leq x(t) \leq \beta_1 - \delta$  if  $0 \leq t \leq c^{-1} - \epsilon$ , where  $\alpha_c + \delta \leq x(t) \leq \beta_c - \delta$  if  $1 - c^{-1} + \epsilon \leq t \leq 1$  and where  $\max(\alpha_j, \alpha_{j+1}) + \delta \leq x(t) \leq \min(\beta_j, \beta_{j+1}) - \delta$  if  $jc^{-1} - \epsilon \leq t \leq jc^{-1} + \epsilon$  ( $j=1, \dots, c-1$ ). For fixed  $\epsilon$  and distinct  $\delta$  the sets  $\tilde{E}_{\epsilon, \delta}$  are disjoint. It is therefore possible to find for each  $\epsilon$  a  $\delta(\epsilon)$ , with  $0 < \delta(\epsilon) < \epsilon$ , such that if  $E_\epsilon = E_{\epsilon, \delta(\epsilon)}$  then  $P(\tilde{E}_\epsilon) = 0$ . Let  $G_{n, \epsilon}$  be the set of  $x \in C$  satisfying the conditions defining  $E_\epsilon$  at points  $t$  of the form  $in^{-1}$ ,  $i=0, \dots, n$ . Since  $\epsilon$  is rational,  $E_\epsilon$  can be cast in the form (2.4) and  $G_{n, \epsilon}$  bears the same relation to  $E_\epsilon$  as  $G_n$  does to  $E$ . Then by hypothesis  $P_n(G_{n, \epsilon}) \rightarrow P(E_\epsilon)$ . Now  $E \subset G_n$ , while  $G_{n, \epsilon} A_n \subset E$  provided  $n^{-1} < \epsilon$ . Hence

$$(2.7) \quad P(E_\epsilon) \leq \liminf_{n \rightarrow \infty} P_n(E) \leq \limsup_{n \rightarrow \infty} P_n(E) \leq P(E).$$

Since  $E$  contains  $\bigcup_\epsilon E_\epsilon$ , which in turn contains the interior of  $E$ , and since  $P(\tilde{E}) = 0$ , we can conclude from (2.7), by letting  $\epsilon$  tend to 0 through rational values, that  $P_n(E) \rightarrow P(E)$ . Hence the result follows from Theorem 2.2.

A useful fact in the theory of distributions on the real line is that if the distributions of a sequence  $\{X_n\}$  of random variables converge weakly to some distribution, then so do the distributions of  $\{X_n + Y_n\}$ , provided  $p \lim_n Y_n = 0$ . We conclude this section with a theorem which plays an analogous role in the theory of distributions on  $\mathcal{C}$ . The theorem and its proof obviously remain unchanged if  $C$  is replaced by any Banach space.

Let  $\{X_n\}$  and  $\{Y_n\}$  be two sequences of measurable functions on some probability measure space  $(\Omega, \mathfrak{B}, P)$ , with values in  $C$ . For  $A \in \mathcal{C}$  let  $P_n(A) = P\{X_n \in A\}$ ,  $P'_n(A) = P\{Y_n \in A\}$  and  $P''_n(A) = P\{X_n + Y_n \in A\}$ . Let  $U$  be the measure on  $\mathcal{C}$  which places unit mass at the function which is identically zero. Suppose  $Q$  is a probability measure on  $\mathcal{C}$ .

**THEOREM 2.4.** *If  $P_n \Rightarrow Q$  and  $P'_n \Rightarrow U$  then  $P''_n \Rightarrow Q$ .*

**Proof.** Clearly  $P'_n \Rightarrow U$  if and only if

$$(2.8) \quad \lim_{n \rightarrow \infty} P\{\rho(0, Y_n) \geq \epsilon\} = 0.$$

for all  $\epsilon > 0$ . Let  $f$  be a bounded, uniformly continuous function on  $C$ . Given  $\epsilon$  choose  $\delta$  so that  $|f(x) - f(y)| < \epsilon$  if  $\rho(x, y) < \delta$ . Then

$$P\{|f(X_n) - f(X_n + Y_n)| \geq \epsilon\} \leq P\{\rho(X_n, X_n + Y_n) \geq \delta\} \rightarrow 0$$

by (2.8). Hence

$$(2.9) \quad p \lim_{n \rightarrow \infty} (f(X_n) - f(X_n + Y_n)) = 0.$$

Now by Theorem 2.1 ((i)  $\rightarrow$  (v)),  $\lim_n P\{f(X_n) \leq \alpha\} = Q\{x: f(x) \leq \alpha\}$  at continuity points of the latter function. Hence by (2.9) and the above-mentioned fact in the theory of distributions on the real line,  $\lim_n P\{f(X_n + Y_n) \leq \alpha\} = Q\{x: f(x) \leq \alpha\}$ . Since this holds for all bounded, uniformly continuous  $f$ , the result follows by Theorem 2.1 ((v)  $\rightarrow$  (i)).

**3. A general invariance principle.** Let  $\{X_n\}$  be a sequence of random variables on a space  $(\Omega, \mathfrak{B}, P)$  and let  $p_n$  be the random function defined in §1. In this section we give conditions under which the invariance principle holds for  $\{X_n\}$  with norming factors  $\{a_n\}$ .

For integers  $c, \nu$  and  $n$  define  $n_j = [jnc^{-1}]$ ,  $j = 0, 1, \dots, c$  and  $n_{j,u} = [n(\nu(j-1) + u)c^{-1}\nu^{-1}]$ ,  $j = 1, \dots, c$ ,  $u = 0, 1, \dots, \nu$ . For any real numbers  $\alpha_j, \beta_j$  let  $E_{n,r}$  be the  $\Omega$  set where the relations

$$(3.1) \quad \alpha_j \leq a_n^{-1} S_i \leq \beta_j \quad \text{if } n_{j-1} < i \leq n_j$$

are satisfied for  $i < r$ , but not for  $i = r$ .

**THEOREM 3.1.** *The invariance principle holds for the sequence  $\{X_n\}$  with norming factors  $\{a_n\}$  if the following two conditions are satisfied.*

*Condition (i). For each integer  $c$  the distribution of the random vector*

$a_n^{-1}(S_{n_1}, S_{n_2} - S_{n_1}, \dots, S_{n_c} - S_{n_{c-1}})$  approaches, as  $n \rightarrow \infty$ , the normal distribution having zero means and having as covariance matrix  $c^{-1}$  times the identity matrix.

Condition (ii). For each integer  $c$ , each set  $(\alpha_1, \dots, \alpha_c, \beta_1, \dots, \beta_c)$  and each  $\epsilon > 0$ ,

$$(3.2) \quad \lim_{\nu \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{r=1}^n P(E_{n,r} \cap \{ |S_{r'} - S_r| \geq \epsilon a_n \}) = 0,$$

where the  $r'$  corresponding to each  $r$  is that integer of the form  $n_{j,u+1}$  such that

$$(3.3) \quad n_{j,u} < r \leq n_{j,u+1}.$$

**Proof.** Throughout the rest of the paper we will be dealing with sums of the sort appearing in (3.2). In each instance  $r' = n_{j,u+1}$  is a function of  $r$  defined by (3.3).

We prove the theorem by a modification of an argument of [4]. Let  $E_n$  be the  $\Omega$  set where (3.1) is satisfied for all  $i = 1, \dots, n$ . Let  $E$  be the  $C$  set where  $\alpha_j \leq x(t) \leq \beta_j$  if  $(j-1)c^{-1} \leq t \leq jc^{-1}$ ,  $j = 1, \dots, c$ . Let  $D_\nu$  be the  $C$  set where  $\alpha_j \leq x(((j-1)\nu + u)c^{-1}\nu^{-1}) \leq \beta_j$  for  $j = 1, \dots, c$  and  $u = 1, \dots, \nu$ . Further, let  $F_n$  be the  $\Omega$  set where  $\alpha_j \leq a_n^{-1}S_{n_{j,u}} \leq \beta_j$  for  $j = 1, \dots, c$  and  $u = 1, \dots, \nu$ . Finally, let  $E_\epsilon$ ,  $D_{\nu,\epsilon}$  and  $F_{n,\epsilon}$  be defined in the same way as  $E$ ,  $D_\nu$  and  $F_n$ , but with  $\alpha_j$  and  $\beta_j$  replaced by  $\alpha_j + \epsilon$  and  $\beta_j - \epsilon$  respectively. For  $n_{j,u} < r \leq n_{j,u+1} = r'$  write

$$(3.4) \quad \begin{aligned} P(E_{n,r}) &= P(E_{n,r} \cap \{ |S_{r'} - S_r| \geq \epsilon a_n \}) \\ &\quad + P(E_{n,r} \cap \{ |S_{r'} - S_r| < \epsilon a_n \}). \end{aligned}$$

Obviously the set in the second term of the right member of (3.4) is contained in  $\Omega - F_{n,\epsilon}$ . Hence, since  $\Omega - E_n = \bigcup_r E_{n,r}$  and the  $E_{n,r}$  are disjoint, we have  $P(F_{n,\epsilon}) - \zeta_{\nu,n} \leq P(E_n) \leq P(F_n)$ , where  $\zeta_{\nu,n}$  is the first term of the right member of (3.4). But  $P(F_n) \rightarrow W(D_\nu)$  and  $P(F_{n,\epsilon}) \rightarrow W(D_{\nu,\epsilon})$  by Condition (i). Hence  $W(D_{\nu,\epsilon}) - \limsup_n \zeta_{\nu,n} \leq \liminf_n P(E_n) \leq \limsup_n P(E_n) \leq W(D_\nu)$ . Letting  $\nu \rightarrow \infty$  we have, by Condition (ii),  $W(E_\epsilon) \leq \liminf_n P(E_n) \leq \limsup_n P(E_n) \leq W(E)$ . Since  $\bigcup_\epsilon E_\epsilon$  is the interior of  $E$ ,  $P(E_n^*) \rightarrow W(E)$  if  $W(\bar{E}) = 0$ . Hence the result follows by Theorem 2.3.

**4. The invariance principle for Markov processes.** In this section we prove, using Theorem 3.1, the invariance principle for discrete Markov processes satisfying Doeblin's condition. We use the definitions, notations and results of [5, Chap. V]. Let  $X$  be a space of points  $\xi$  and let  $\mathfrak{F}_X$  be a Borel field of subsets of  $X$ . Let  $\{x_n, n \geq 1\}$  be a Markov process with state space  $X$  and stationary transition probabilities

$$(4.1) \quad p(\xi, A) = P\{x_{n+1} \in A \mid x_n = \xi\}.$$

That is,  $\{x_n\}$  is a sequence of measurable functions from some probability space  $(\Omega, \mathfrak{B}, P)$  to  $X$  such that (4.1) holds, where the transition function

$p(\xi, A)$  is a measurable function of  $\xi$  for fixed  $A \in \mathfrak{F}_X$  and is a probability measure on  $\mathfrak{F}_X$  for fixed  $\xi$ . The initial distribution  $\pi$  is defined by  $\pi(A) = P\{x_1 \in A\}$  and the  $n$ -step transition probabilities by

$$p^{(n)}(\xi, A) = P\{x_{n+1} \in A \mid x_1 = \xi\}.$$

The existence problems involved here are resolved in [5].

It is known [5] that if Doeblin's condition is satisfied then the states can be classified according to their ergodic properties. It is known further that if the following hypothesis is satisfied then the  $n$ -step transition probabilities converge exponentially.

*Hypothesis (D<sub>0</sub>).*

(a) *Doeblin's condition is satisfied.*

(b) *There is only a single ergodic set and this contains no cyclically moving subsets.*

That is, it is known that if (D<sub>0</sub>) holds then there exist positive constants  $\gamma$  and  $\rho$ ,  $\rho < 1$ , and a (unique) stationary initial distribution  $p$  such that  $|p^{(n)}(\xi, E) - p(E)| \leq \gamma\rho^n$  for all  $\xi \in X$ ,  $E \in \mathfrak{F}_X$  and  $n \geq 1$ . The results of this section will be obtained under the assumption of (D<sub>0</sub>).

In what follows, the initial distribution under the assumption of which a probability or expectation is computed will be denoted by a subscript, thus:  $P_\pi(E)$ . If  $\pi = p$ , the subscript will be omitted.

It is convenient to have available the following corollary of Lemma 7.2 of [5, Chap. V]. The proof, which goes by induction, is easy.

**LEMMA 4.1.** *Suppose we have integers  $u_i, v_i$  with  $u_1 \leq v_1 < u_2 \leq v_2 < \dots < u_m \leq v_m$  and  $u_i - v_{i-1} \geq B \geq 1$ ,  $i = 2, \dots, m$ . Under Hypothesis (D<sub>0</sub>), let  $f_j$  be a random variable, with  $|f_j| \leq 1$ , on  $x_{u_j}, \dots, x_{v_j}$  sample space, for  $j = 1, \dots, m$ . Then  $|E\{f_1 \dots f_m\} - E\{f_1\} \dots E\{f_m\}| \leq 2m\gamma\rho^B$ .*

We come now to the invariance principle.

**THEOREM 4.1.** *Under Hypothesis (D<sub>0</sub>), let  $f$  be a real-valued function of  $\xi$ , measurable  $\mathfrak{F}_X$ , with  $E\{f(x_1)\} = 0$  and  $E\{|f(x_1)|^{2+\delta}\} < \infty$  for some  $\delta > 0$ . Then*

$$(4.2) \quad \lim_{n \rightarrow \infty} E \left\{ \left( n^{-1/2} \sum_{j=1}^n f(x_j) \right)^2 \right\} = \sigma_1^2$$

*exists. If  $\sigma_1^2 > 0$  then the invariance principle holds for the sequence  $\{f(x_n)\}$  with norming factors  $\sigma_1 n^{1/2}$ , no matter what the initial distribution  $\pi$ .*

**Proof.** That the limit (4.2) exists is simply a restatement of Lemma 7.3 of [5, Chap. V]. We prove the theorem first under the assumption of stationarity and remove this restriction later.

We must show that Conditions (i) and (ii) of Theorem 3.1 are satisfied. In the notation of that theorem, we must first prove that the distribution of the random vector  $\sigma_1^{-1} n^{-1/2} (S_{n_1}, S_{n_2} - S_{n_1}, \dots, S_{n_c} - S_{n_{c-1}})$  approaches the



appropriate normal distribution, where  $S_k = f(x_1) + \cdots + f(x_k)$ . Our proof of this part of the theorem follows [5]. Let  $\{\alpha_n\}$  and  $\{\beta_n\}$  be two sequences of positive integers such that if  $\{\mu_n\}$  is defined by

$$\mu_n = \left[ (\min_{j \leq c} (n_j - n_{j-1}) - \beta_n)(\alpha_n + \beta_n)^{-1} \right],$$

then  $\mu_n \beta_n \alpha_n^{-1} \rightarrow 0$  and  $\mu_n \rho^{\beta_n} \rightarrow 0$ , while  $\alpha_n, \beta_n$  and  $\mu_n$  all go to infinity. For example one can take  $\beta_n^4 \sim n c^{-1}$  and  $\alpha_n = \beta_n^3$ . Now for  $j = 1, \dots, c$  let

$$\begin{aligned} y_{j,m} &= \sum_{i=(m-1)(\alpha+\beta)+1}^{(m-1)(\alpha+\beta)+\alpha} f(x_{n_{j-1}+i}), & m &= 1, \dots, \mu, \\ y'_{j,m} &= \sum_{i=(m-1)(\alpha+\beta)+\alpha+1}^{m(\alpha+\beta)} f(x_{n_{j-1}+i}), & m &= 1, \dots, \mu, \\ y'_{j,\mu+1} &= \sum_{i=\mu(\alpha+\beta)+1}^{n_j} f(x_{n_{j-1}+i}). \end{aligned}$$

By Lemma 7.3 of [5, Chap. V] and Minkowski's inequality

$$\begin{aligned} E^{1/2} \left\{ \left( n^{-1/2} \sum_{m=1}^{\mu+1} y'_{j,m} \right)^2 \right\} &\leq n^{-1/2} \sum_{m=1}^{\mu+1} E^{1/2} \{ (y'_{j,m})^2 \} \\ &= O(n^{-1/2} (\mu \beta^{1/2} + (\alpha + 2\beta)^{1/2})) = o(1). \end{aligned}$$

Hence  $p \lim_n n^{-1/2} \sum_{m=1}^{\mu+1} y'_{j,m} = 0$ ,  $j = 1, \dots, c$ . Therefore it suffices to prove the asymptotic normality of the random vector

$$(4.3) \quad \sigma_1^{-1} n^{-1/2} \left( \sum_{m=1}^{\mu} y_{1,m}, \dots, \sum_{m=1}^{\mu} y_{c,m} \right).$$

Let  $\phi_n(u_1, \dots, u_c) = E \left\{ \exp \left( i \sum_{j=1}^c u_j \sum_{m=1}^{\mu} \sigma_1^{-1} n^{-1/2} y_{j,m} \right) \right\}$  be the characteristic function of (4.3). Now the last term  $f(x_i)$  occurring in  $y_{j,m}$  and the first occurring in  $y_{j,m+1}$  have  $\beta$  other such terms in between them. And the last term of  $y_{j,\mu}$  and the first of  $y_{j+1,1}$  have at least  $\beta$  others in between. Hence by Lemma 4.1,  $\phi_n(u_1, \dots, u_c) = \prod_{m=1}^{\mu} \prod_{j=1}^c E \left\{ \exp (i u_j \sigma_1^{-1} n^{-1/2} y_{j,m}) \right\} + \epsilon_n$ , where  $|\epsilon_n| \leq 2\gamma c \mu \rho^{\beta+1} = o(1)$ . Thus the proof of Condition (i) will be complete if we show that the distribution of the vector

$$(4.4) \quad \sigma_1^{-1} n^{-1/2} \left( \sum_{m=1}^{\mu} z_{1,m}, \dots, \sum_{m=1}^{\mu} z_{c,m} \right)$$

approaches the appropriate normal distribution, where  $z_{j,m}$  ( $j = 1, \dots, c$ ,  $m = 1, \dots, \mu$ ) are independent random variables each having the distribution of  $y_{1,1}$ . Since  $\mu/\alpha n \rightarrow c^{-1}$ , it follows that

$$(4.5) \quad \lim_{n \rightarrow \infty} E \left\{ \left( \sigma_1^{-1} n^{-1/2} \sum_{m=1}^{\mu} z_{j,m} \right)^2 \right\} = c^{-1}, \quad j = 1, \dots, c.$$

Since the components of (4.4) are independent, its covariance matrix approaches  $c^{-1}$  times the identity. By Lemma 7.4 of [5, Chap. V] there is a constant  $a$  such that  $E\{|z_{j,m}|^{2+\delta}\} \leq a^{1+(\delta/2)}$ . Hence, by (4.5), for  $n$  sufficiently large,

$$\frac{\sum_{m=1}^{\mu} E\{|\sigma_1^{-1} n^{-1/2} z_{j,m}|^{2+\delta}\}}{\left(E\left\{\left(\sigma_1^{-1} n^{-1/2} \sum_{m=1}^{\mu} z_{j,m}\right)^2\right\}\right)^{1+(\delta/2)}} \leq 2\mu^{-\delta/2} = o(1).$$

Thus Ljapunov's condition holds and Condition (i) follows.

We turn now to Condition (ii). Let  $E_{n,r}$ ,  $r'$ ,  $c$  and  $\nu$  be as in Theorem 3.1. Define a sequence  $\{\beta_n\}$  of integers by  $\beta_n = [\log n]$ . If  $r + \beta_n < r'$  then

$$(4.6) \quad P(E_{n,r} \cap \{|S_{r'} - S_r| \geq \epsilon n^{1/2}\}) \leq P(E_{n,r} \cap \{|S_{r'} - S_{r+\beta}| \geq \epsilon n^{1/2}/2\}) \\ + P\{|S_{r+\beta} - S_r| \geq \epsilon n^{1/2}/2\},$$

and we can estimate the terms on the right separately. Now

$$P\{|S_{r'} - S_{r+\beta}| \geq \epsilon n^{1/2}/2 | x_1, \dots, x_r\} \\ \leq P\{|S_{r'} - S_{r+\beta}| \geq \epsilon n^{1/2}/2\} + 2\gamma\rho^\beta$$

by Lemma 7.2 of [5, Chap. V] and the Markov property. By Chebyshev's inequality and Lemma 7.3 of [5, Chap. V],  $P\{|S_{r'} - S_{r+\beta}| \geq \epsilon n^{1/2}/2\} \leq A\epsilon^{-2}n^{-1}(n_{j,u+1} - (r+\beta)) \leq 2A/\epsilon^2 c\nu$ , where  $A$  is a constant. By these two inequalities and the defining property of conditional probabilities,

$$(4.7) \quad P(E_{n,r} \cap \{|S_{r'} - S_{r+\beta}| \geq \epsilon n^{1/2}/2\}) \leq (2A\epsilon^{-2}c^{-1}\nu^{-1} + 2\gamma\rho^\beta)P(E_{n,r}).$$

To estimate the second term in (4.6) observe that

$$(4.8) \quad P\{|S_{r+\beta} - S_r| \geq \epsilon n^{1/2}\} \leq \sum_{i=r+1}^{r+\beta} P\{|f(x_i)| \geq \epsilon\beta^{-1}n^{1/2}/2\} \\ = \beta P\{|f(x_1)| \geq \epsilon\beta^{-1}n^{1/2}/2\}.$$

Therefore, by (4.6), (4.7) and (4.8),

$$P(E_{n,r} \cap \{|S_{r'} - S_r| \geq \epsilon n^{1/2}\}) \leq (2A\epsilon^{-2}c^{-1}\nu^{-1} + 2\gamma\rho^\beta)P(E_{n,r}) \\ + \beta P\{|f(x_1)| \geq \epsilon\beta^{-1}n^{1/2}/2\}.$$

This estimate was obtained under the assumption that  $r + \beta < r'$ , but obviously holds in the other case as well. Since the  $E_{n,r}$  are disjoint we have then

$$\sum_{r=1}^n P(E_{n,r} \cap \{|S_{r'} - S_r| \geq \epsilon n^{1/2}\}) \leq 2A\epsilon^{-2}c^{-1}\nu^{-1} + 2\gamma\rho^\beta \\ + n\beta P\{|f(x_1)| \geq \epsilon\beta^{-1}n^{1/2}/2\}.$$

Since

$$n\beta P\{|f(x_1)| \geq \epsilon\beta^{-1}n^{1/2}/2\} \leq (2/\epsilon)^{2+\delta}\beta^{3+\delta}n^{-\delta/2}E\{|f(x_1)|^{2+\delta}\},$$

(3.2) follows immediately.

We have thus proved the theorem under the assumption that the initial distribution is the stationary one, which assumption we now remove. Let  $\pi$  be any initial distribution. We show first of all that there exists a sequence  $\{\beta_n\}$  of integers going to infinity so slowly that

$$(4.9) \quad \lim_{n \rightarrow \infty} P_{\pi} \left\{ \max_{i \leq \beta_n} |S_i| \geq \epsilon n^{1/2} \right\} = 0$$

for all  $\epsilon > 0$ . In fact, for each  $k$  select an integer  $m_k$  so that

$$\sum_{i=1}^k P_{\pi} \{ |f(x_i)| \geq k^{-1}n^{1/3} \} < k^{-1} \text{ if } n \geq m_k.$$

Clearly we can choose the  $m_k$  so that  $m_k < m_{k+1}$ . If  $\beta_n = k$  for  $m_k < n \leq m_{k+1}$  then  $\{\beta_n\}$  satisfies (4.9) for all  $\epsilon > 0$ . We may at the same time choose  $\{\beta_n\}$  so that

$$(4.10) \quad \lim_{n \rightarrow \infty} P \left\{ \max_{i \leq \beta_n} |S_i| \geq \epsilon n^{1/2} \right\} = 0$$

for all  $\epsilon > 0$ .

Let  $p_n$  be the random function defined by (1.1), where  $X_i = f(x_i)$ . And let  $p'_n$  be the random function defined by

$$p'_n(t) = \begin{cases} p_n(t) & \text{if } 0 \leq t \leq \beta_n n^{-1}, \\ p_n(\beta_n n^{-1}) & \text{if } \beta_n n^{-1} \leq t \leq 1. \end{cases}$$

Finally, let  $p''_n = p_n - p'_n$ . Now (4.9) and (4.10) respectively imply

$$(4.11) \quad \lim_{n \rightarrow \infty} P_{\pi} \left\{ \max_{0 \leq t \leq 1} n^{-1/2} p'_n(t) \geq \epsilon \right\} = 0$$

and

$$(4.12) \quad \lim_{n \rightarrow \infty} P \left\{ \max_{0 \leq t \leq 1} n^{-1/2} p'_n(t) \geq \epsilon \right\} = 0.$$

Since we have shown that the invariance principle holds for the stationary initial distribution,  $P\{\sigma_1^{-1}n^{-1/2}p_n \in A\} \Rightarrow W(A)$ . By (4.12) and Theorem 2.4,  $P\{\sigma_1^{-1}n^{-1/2}p''_n \in A\} \Rightarrow W(A)$ . It is clear that the set  $\{\sigma_1^{-1}n^{-1/2}p''_n \in A\}$  is measurable on  $x_{\beta+1}, x_{\beta+2}, \dots$  sample space. Hence, by Lemma 7.2 of [5, Chap. V],

$$|P_{\pi}\{\sigma_1^{-1}n^{-1/2}p''_n \in A\} - P\{\sigma_1^{-1}n^{-1/2}p''_n \in A\}| \leq 2\gamma\rho^{\beta} = o(1).$$

Hence  $P_{\pi}\{\sigma_1^{-1}n^{-1/2}p''_n \in A\} \Rightarrow W(A)$ . But from this, (4.11) and Theorem 2.4

it follows that  $P_{\tau}\{\sigma_1^{-1}n^{-1/2}p_n \in A\} \Rightarrow W(A)$ . This completes the proof of the theorem.

**5. The invariance principle for  $m$ -dependent random variables.** A sequence  $\{X_n\}$  of random variables is said to be  $m$ -dependent if the random vectors  $(X_n, \dots, X_{n+r})$  and  $(X_{n+s}, \dots, X_{n+t})$  are independent whenever  $s-r > m$ . Sequences having this property are of interest in statistics and have been studied by various authors (cf. [2; 9; 3; 12]). In this section we concern ourselves with the invariance principle for such sequences.

Then let  $\{X_n\}$  be an  $m$ -dependent sequence of random variables with zero means and finite variances. Marsaglia [12] has proved the central limit theorem for such a sequence under the assumption that  $E\{X_n^2\}$  is bounded, that (5.2) below is satisfied and that  $ns_n^{-2} = O(1)$ , where  $s_n^2 = E\{S_n^2\}$ . The following theorem asserts the invariance principle under slightly stronger assumptions.

**THEOREM 5.1.** *If for an  $m$ -dependent sequence  $\{X_n\}$ ,  $E\{X_n\} = 0$ ,  $E\{X_n^2\}$  is bounded,*

$$(5.1) \quad |s_n^2 - n\sigma^2| = O(1)$$

*for some constant  $\sigma^2 > 0$ , and*

$$(5.2) \quad \lim_{n \rightarrow \infty} s_n^{-2-\delta} \sum_{i=1}^n E\{|X_i|^{2+\delta}\} = 0$$

*for some  $\delta > 0$ , then the invariance principle holds for the sequence  $\{X_n\}$  with norming factors  $\sigma n^{1/2}$ .*

**Proof.** We first show that Condition (i) of Theorem 3.1 is satisfied, using the technique of Marsaglia [12]. Let  $(n_1, \dots, n_c)$  be defined as in §3 and for each pair  $(n, k)$  with  $2m < k < nc^{-1}$  and each  $j = 1, \dots, c$  define

$$\begin{aligned} y_{j,i} &= \sum_{v=1}^{k-m} X_{n_{j-1}+(i-1)k+v}, & 1 \leq i \leq [k^{-1}(n_j - n_{j-1})], \\ y'_{j,i} &= \sum_{v=1}^m X_{n_{j-1}+ik-m+v}, & 1 \leq i < [k^{-1}(n_j - n_{j-1})], \\ y'_{j,i} &= \sum_{v=1}^{n_j - n_{j-1} - i + m} X_{n_{j-1}+ik-m+v}, & i = [k^{-1}(n_j - n_{j-1})]. \end{aligned}$$

Let

$$g_{n,k}^{(j)} = \sum_{i=1}^{[k^{-1}(n_j - n_{j-1})]} y_{j,i}, \quad e_{n,k}^{(j)} = \sum_{i=1}^{[k^{-1}(n_j - n_{j-1})]} y'_{j,i}.$$

Now by Hölder's inequality, if  $i < [k^{-1}(n_j - n_{j-1})]$  then

$$E\{y_{j,i}^2\} \leq m \sum_{v=1}^m E\{X_{n_{j-1}+ik-m+v}^2\} \leq m^2 B,$$

where  $B$  is the bound on  $E\{X_n^2\}$ . Using this inequality, a similar one for the case  $i = [k^{-1}(n_j - n_{j-1})]$  and the fact that the  $y'_{j,i}$  are independent, we see that

$$(5.3) \quad E\{(e_{n,k}^{(i)})^2\} \leq (n_j - n_{j-1})k^{-1}m^2B + (k+m)^2B.$$

One obtains in a similar manner the inequality

$$(5.4) \quad |E\{(S_{n_j} - S_{n_{j-1}})^2\} - E\{(g_{n,k}^{(i)})^2\}| \leq 5(n_j - n_{j-1})k^{-1}m^2B + (k+m)^2B + 2m^2B.$$

From (5.3) and (5.4) it follows that

$$(5.5) \quad \lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} n^{-1} \sigma^{-2} E\{(e_{n,k}^{(i)})^2\} = 0,$$

$$(5.6) \quad \lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} n^{-1} \sigma^{-2} |E\{(S_{n_j} - S_{n_{j-1}})^2\} - E\{(g_{n,k}^{(i)})^2\}| = 0.$$

Note that these two relations have been obtained without the use of (5.2). Now by (5.1),

$$(5.7) \quad E\{(S_{n+i} - S_n)^2\} = i\sigma^2 + \vartheta,$$

where  $\vartheta$  is bounded. From this fact and (5.6) it follows that

$$(5.8) \quad \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} cn^{-1} \sigma^{-2} E\{(g_{n,k}^{(i)})^2\} = 1,$$

where the limit is to be taken in the strong sense (cf. [12]).

We now show that if  $k$  is sufficiently large then the distribution of the vector

$$(5.9) \quad ((\tau_{n,k}^{(1)})^{-1} g_{n,k}^{(1)}, \dots, (\tau_{n,k}^{(c)})^{-1} g_{n,k}^{(c)}),$$

where  $(\tau_{n,k}^{(j)})^2 = E\{(g_{n,k}^{(j)})^2\}$ , approaches, as  $n \rightarrow \infty$ , the normal distribution with zero means having the identity as covariance matrix. But for  $k$  sufficiently large this follows, via Ljapunov's condition, from (5.7), (5.1) and the fact that

$$E\{|y_{j,i}|^{2+\delta}\} \leq k^\delta \sum_{v=1}^{k-m} E\{|X_{n_{j-1}+(i-1)k+v}|^{2+\delta}\}.$$

And now from (5.8) and the obvious multi-dimensional analogue of Theorem 2 of [12] it follows that the distribution of  $\sigma^{-1}n^{-1/2}(g_{n,k}^{(1)}, \dots, g_{n,k}^{(c)})$  approaches, if  $n \rightarrow \infty$  and then  $k \rightarrow \infty$ , the normal distribution having zero means and having as covariance matrix  $c^{-1}$  times the identity. In order to show that the distribution of  $\sigma^{-1}n^{-1/2}(S_{n_1}, S_{n_2} - S_{n_1}, \dots, S_{n_c} - S_{n_{c-1}})$  approaches, as  $n \rightarrow \infty$ , this same normal distribution it is enough, by the multi-dimensional analogue of Theorem 1 of [12], to show that

$$p \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \sigma^{-1} n^{-1/2} (e_{n,k}^{(1)}, \dots, e_{n,k}^{(e)}) = 0.$$

But this follows immediately from (5.5) and Chebyshev's inequality.

We have thus proved Condition (i) of Theorem 3.1 and pass to Condition (ii). Using all the notations of that theorem,

$$P\{|S_{r+m} - S_r| \geq \epsilon n^{1/2}/2\} \leq \sum_{v=r}^{r+m} P\{|X_v| \geq \epsilon n^{1/2}/2m\}.$$

Hence,

$$\begin{aligned} \sum_{r=1}^n P\{|S_{r+m} - S_r| \geq \epsilon n^{1/2}/2\} &\leq m \sum_{v=1}^n P\{|X_v| \geq \epsilon n^{1/2}/2m\} \\ (5.10) \qquad \qquad \qquad &\leq m(2m/\epsilon)^{2+\delta} n^{-1-(\delta/2)} \sum_{v=1}^n E\{|X_v|^{2+\delta}\} = o(1) \end{aligned}$$

as  $n \rightarrow \infty$ , by (5.2). And (making the inessential assumption that  $r+m < r'$ ) by Chebyshev's inequality and (5.7),  $P\{|S_{r'} - S_{r+m}| \geq \epsilon n^{1/2}/2\} \leq 4(\epsilon^2 c v)^{-1} + 4\vartheta(\epsilon^2 n)^{-1}$ . Hence, since  $S_{r'} - S_{r+m}$  is independent of  $E_{n,r}$ ,

$$(5.11) \quad \limsup_{n \rightarrow \infty} \sum_{r=1}^n P(E_{n,r} \cap \{|S_{r'} - S_{r+m}| \geq \epsilon n^{1/2}/2\}) \leq 4/cv\epsilon^2.$$

And now (3.2) follows from (5.10), (5.11) and

$$\begin{aligned} (5.12) \quad P(E_{n,r} \cap \{|S_{r'} - S_r| \geq \epsilon n^{1/2}\}) &\leq P\{|S_{r+m} - S_r| \geq \epsilon n^{1/2}/2\} \\ &\quad + P(E_{n,r} \cap \{|S_{r'} - S_{m+r}| \geq \epsilon n^{1/2}/2\}). \end{aligned}$$

It is possible, at the expense of complicating somewhat the proof of Condition (ii), to relax the condition (5.1). In particular, it can be replaced by  $s_n^2 \sim n\sigma^2$ .

**THEOREM 5.2.** *If  $\{X_n\}$  is a stationary  $m$ -dependent sequence of random variables with zero means and finite variances, then the invariance principle holds for  $\{X_n\}$  with norming factors  $n^{1/2}\sigma$ , where  $\sigma^2 = E\{X_1^2\} + 2 \sum_{k=1}^m E\{X_1 X_{k+1}\}$ .*

**Proof.** Define  $y_{j,i}$ ,  $y'_{j,i}$ ,  $g_{n,k}^{(j)}$  and  $e_{n,k}^{(j)}$  as in the proof of Theorem 5.1. It is a simple matter to show that (5.1) holds here. Since  $E\{X_n^2\}$  is bounded it follows that (5.5) and (5.6) hold in this case as well. In order to establish that Condition (i) holds in the present case it suffices to show that the vector (5.9) is asymptotically normal. But this follows immediately from the stationarity.

To prove that Condition (ii) holds we proceed as before. In fact (5.11) and (5.12) are still valid. Finally,

$$\sum_{r=1}^m P\{|S_{r+m} - S_r| \geq \epsilon n^{1/2}/2\} \leq mnP\{|X_1| \geq \epsilon n^{1/2}/2m\},$$

and the right side of this inequality goes to zero since  $E\{X_1^2\} < \infty$ .

**6. The invariance principle for linear processes with  $m$ -dependent residuals.** Let  $\{Y_j; j=0, \pm 1, \dots\}$  be a stationary,  $m$ -dependent process such that  $Y_j$  has zero mean and a finite variance. Let  $\{A_t; t=0, 1, \dots\}$  be a sequence of constants such that

$$(6.1) \quad \sum_{t=0}^{\infty} |A_t| < \infty.$$

For a fixed  $i$  the series  $X_i = \sum_{t=0}^{\infty} A_t Y_{i-t}$  converges in probability. Consider in fact the  $m+1$  series obtained by taking every  $(m+1)$ th term starting with the first, the second and so on. These are series of independent terms so that, since  $\sum A_t^2$  converges, they converge in probability (and even with probability one). In the terminology of [3] the process  $\{X_i; i=1, 2, \dots\}$  is a *discrete linear process with  $m$ -dependent residuals*. Clearly it is stationary. Processes of this sort are of interest in the analysis of time series (cf. [3] for references to the statistical literature).

Diananda [3] has shown that the central limit theorem holds for processes which arise in this way, i.e., under the assumption that (6.1) holds. It is the purpose of this section to prove the invariance principle for such processes. We are forced, however, to make a stronger assumption on the nature of the sequence  $\{A_t\}$ , viz., we assume that

$$(6.2) \quad |A_n| = O(n^{-3}).$$

At the end of the proof we indicate some ways in which the requirement (6.2) can be relaxed.

**THEOREM 6.1.** *Let  $\{Y_j\}$  be a stationary  $m$ -dependent process with zero means and finite variances and assume that (6.2) holds. Then  $\sum_{t=0}^n A_t Y_{i-t}$  converges in probability, as  $n \rightarrow \infty$ , to some random variable  $X_i = \sum_{t=0}^{\infty} A_t Y_{i-t}$ , so that  $\{X_i; i=1, 2, \dots\}$  is a stationary discrete linear process with  $m$ -dependent residuals. The invariance principle holds for  $\{X_n\}$  with norming factors  $n^{1/2}\sigma$ , provided  $\sigma > 0$ , where*

$$\sigma^2 = \left( \sum_{i=0}^{\infty} A_i \right)^2 \left( E\{Y_0^2\} + 2 \sum_{v=1}^m E\{Y_0 Y_v\} \right).$$

**Proof.** That  $\{X_i\}$  exists and forms a process of the type asserted follows from the preceding discussion and the fact that (6.2) implies (6.1). In each of the theorems of the preceding two sections Condition (i) of Theorem 3.1 was established for a particular process by modifying appropriately the proof of the central limit theorem for the process. That Condition (i) holds for the process  $\{X_n\}$  of the present theorem can be proved by a similar modification of the proof [3] of the central limit theorem for such a process. In fact one

need only assume (6.1), rather than (6.2). To avoid repetition, we omit the details and turn to the proof of Condition (ii).

We show first of all that there exists a sequence  $\{\beta_n\}$  of integers going to infinity in such a way that

$$(6.3) \quad \lim_{n \rightarrow \infty} n\beta_n P\{|X_1| \geq \epsilon\beta_n^{-1}n^{1/2}/2\} = 0.$$

It is easy to show that  $X_1$  has a finite second moment. Hence there exists an increasing sequence  $\{m_k\}$  such that  $knP\{|X_1| \geq \epsilon k^{-1}n^{1/2}/2\} \leq k^{-1}$  if  $n \geq m_k$ . If one puts  $\beta_n = k$  for  $m_k < n \leq m_{k+1}$ , then  $\beta_n$  goes to infinity and (6.3) holds. If  $\beta_n$  goes to infinity then

$$(6.4) \quad \lim_{n \rightarrow \infty} \sum_{r=1}^n P\left\{\left|\sum_{v=0}^{\infty} Y_{-v}(A_{\beta+v} + \cdots + A_{r'-r+v})\right| \geq \epsilon n^{1/2}/4\right\} = 0.$$

In fact it follows from (6.2) that

$$(6.5) \quad R_n = O(n^{-2}),$$

where  $R_n = |A_n| + |A_{n+1}| + \cdots$ . Now

$$\begin{aligned} P\left\{\left|\sum_{v=0}^{\infty} Y_{-v}(A_{\beta+v} + \cdots + A_{r'-r+v})\right| \geq \epsilon n^{1/2}/4\right\} \\ \leq P\left\{\sum_{v=0}^{\infty} |Y_{-v}| R_{\beta+v} \geq \epsilon n^{1/2}/4\right\} \\ \leq \sum_{v=0}^{\infty} P\{|Y_{-v}| R_{\beta+v} \geq n^{1/2}(\beta+v)^{-5/4}\}, \end{aligned}$$

provided  $n$  is large enough that  $\sum_{v=0}^{\infty} (\beta+v)^{-5/4} < \epsilon/4$ . But  $P\{|Y_{-v}| R_{\beta+v} \geq n^{1/2}(\beta+v)^{-5/4}\} \leq n^{-1}(\beta+v)^{5/2} R_{\beta+v}^2$ , so that by (6.5) the sum in (6.4) is dominated by  $\sum_{v=0}^{\infty} (\beta+v)^{-3/2}$ , which goes to zero since  $\beta$  goes to infinity. Hence (6.4).

We now decompose the summand in (3.2) into

$$(6.6) \quad \begin{aligned} P(E_{n,r} \cap \{|S_{r'} - S_r| \geq \epsilon n^{1/2}\}) &\leq P\{|S_{r+\beta} - S_r| \geq \epsilon n^{1/2}/2\} \\ &\quad + P(E_{n,r} \cap \{|S_{r'} - S_{r+\beta}| \geq \epsilon n^{1/2}/2\}), \end{aligned}$$

where  $\{\beta_n\}$  satisfies (6.3) and (6.4). It follows immediately from stationarity and (6.3) that

$$(6.7) \quad \lim_{n \rightarrow \infty} \sum_{r=1}^n P\{|S_{r+\beta} - S_r| \geq \epsilon n^{1/2}/2\} = 0.$$

Let  $r' - r = w$ . Then  $S_{r'} - S_r = \xi + \eta$ , where

$$\xi = \sum_{v=0}^{\infty} (A_{\beta+v} + \cdots + A_{w+v}) Y_{r-v}$$



and

$$\eta = A_0 Y_{r+w} + (A_0 + A_1) Y_{r+w-1} + \cdots + (A_0 + \cdots + A_{w-\beta-1}) Y_{r+\beta+1} \\ + \sum_{v=0}^{\beta-1} (A_v + \cdots + A_{v+w-\beta}) Y_{r+\beta-v}.$$

Hence

$$(6.8) \quad P(E_{n,r} \cap \{ |S_{r'} - S_{r+\beta}| \geq \epsilon n^{1/2}/2 \}) \\ \leq P(E_{n,r}) P\{ |\eta| \geq \epsilon n^{1/2}/4 \} + P\{ |\xi| \geq \epsilon n^{1/2}/4 \}$$

where the factorization of the first term on the right is valid if  $n$  is large enough that  $\beta_n > m$ . Now by stationarity and (6.4) we have

$$(6.9) \quad \lim_{n \rightarrow \infty} \sum_{r=1}^n P\{ |\xi| \geq \epsilon n^{1/2}/4 \} = 0.$$

Using the  $m$ -dependence property, one verifies

$$E\{\eta^2\} \leq 2A^2 Bmw + A_0^2 + (A_0 + A_1)^2 + \cdots + (A_0 + \cdots + A_{w-\beta-1})^2 \\ + \sum_{v=0}^{\beta-1} (A_v + \cdots + A_{v+w-\beta})^2 \\ \leq w \left( 2A^2 Bm + \left( \sum_{v=0}^{\infty} |A_v| \right)^2 \right),$$

where  $A = \sup_t |A_t|$  and  $B = \max_{v \leq m} |E\{Y_0 Y_v\}|$ . Since, in the notation of Theorem 3.1,  $w \leq n/cv$ ,

$$(6.10) \quad \sum_{r=1}^n P(E_{n,r}) P\{ |\eta| \geq \epsilon n^{1/2}/4 \} \leq (\text{const.})/\epsilon^2 cv.$$

Finally, (3.2) follows from (6.6), (6.7), (6.8), (6.9) and (6.10), completing the proof of the theorem.

It is clear that (6.2) can be replaced by the weaker hypothesis (6.5). In fact  $R_n = O(n^{-1-\delta})$ ,  $\delta > 0$ , suffices. An examination of the proof shows that if there exists a sequence  $\{\beta_n\}$  going to infinity in such a way that (6.3) and (6.4) hold, then the result follows. This fact can be used to weaken (6.2) under the assumption that  $Y_j$  possesses some moment of order higher than two.

**7. The invariance principle for recurrent events.** Let  $\mathcal{E}$  be a recurrent event in the sense of Feller [8]. Suppose that  $\mathcal{E}$  is certain, let  $(X_1, X_2, \cdots)$  be the recurrence times of  $\mathcal{E}$  and let  $S_k = X_1 + \cdots + X_k$ . Let  $Z_n$  be 1 or 0 according as  $\mathcal{E}$  occurs or not at the  $n$ th trial and let  $N_n = Z_1 + \cdots + Z_n$  be the number of occurrences of  $\mathcal{E}$  during the first  $n$  trials. In this section we prove the invariance principle for  $\{Z_n - \mu^{-1}\}$ , where  $\mu$  is the mean recurrence time.

**THEOREM 7.1.** *If the recurrence times of a certain, aperiodic recurrent event have finite mean  $\mu$  and variance  $\sigma^2$ , then the invariance principle holds for the sequence  $\{Z_n - \mu^{-1}\}$  with norming factors  $\sigma\mu^{-3/2}n^{1/2}$ .*

**Proof.** Feller [8] has proved the central limit theorem for  $N_n$  by reducing it to the central limit theorem for  $S_k$  via the identity

$$(7.1) \quad \{N_n \geq k\} = \{S_k \leq n\}.$$

Our proof that Condition (i) of Theorem 3.1 holds proceeds in the same way. Let  $\Phi(\alpha_1, \dots, \alpha_c)$  be the normal distribution with zero means and covariance matrix  $(c^{-1} \min(i, j))$ . We must show that

$$(7.2) \quad \lim_{n \rightarrow \infty} P\{N_{n_j} - n_j\mu^{-1} \leq \alpha_j \sigma \mu^{-3/2} n^{1/2}, j = 1, \dots, c\} = \Phi(\alpha_1, \dots, \alpha_c).$$

For  $j=1, \dots, c$  let  $k_j = k_j(n)$  be one greater than the integral part of  $n_j\mu^{-1} + \alpha_j \sigma \mu^{-3/2} n^{1/2}$ . Then (7.2) reduces to

$$(7.3) \quad \lim_{n \rightarrow \infty} P\{N_{n_j} < k_j, j = 1, \dots, c\} = \Phi(\alpha_1, \dots, \alpha_c).$$

By (7.1) we see that (7.3) will follow if we can prove

$$\lim_{n \rightarrow \infty} P\left\{\frac{S_{k_j} - k_j\mu}{\sigma\mu^{-1/2}n^{1/2}} > \frac{n_j - k_j\mu}{\sigma\mu^{-1/2}n^{1/2}}, j = 1, \dots, c\right\} = \Phi(\alpha_1, \dots, \alpha_c).$$

Since  $n_j - k_j \sim -\alpha_j \sigma \mu^{-1/2} n^{1/2}$  and  $\Phi$  possesses a symmetry, it suffices to show that the distribution of  $\sigma^{-1} \mu^{1/2} n^{1/2} (S_{k_1} - k_1, \dots, S_{k_c} - k_c)$  approaches  $\Phi$ . But this follows easily from the central limit theorem for independent, identically distributed random vectors and the fact that  $k_j \sim n_j \mu^{-1}$ .

To establish Condition (ii) we define some auxiliary random variables. Let  $V_r$  be the first trial, after the  $r$ th, at which  $\varepsilon$  occurs. And define random variables  $\{\beta_{n,r}\}$  by

$$\beta_{n,r} = \begin{cases} V_r & \text{if } V_r \leq r', \\ r' & \text{if } V_r > r'. \end{cases}$$

Then

$$\begin{aligned} (7.4) \quad & \sum_{r=1}^n P(E_{n,r} \cap \{|(N_{r'} - r'\mu^{-1}) - (N_r - r\mu^{-1})| \geq \epsilon n^{1/2}\}) \\ & \leq \sum_{r=1}^n P(E_{n,r} \cap \{|(N_{\beta_{n,r}} - \beta_{n,r}\mu^{-1}) - (N_r - r\mu^{-1})| \geq \epsilon n^{1/2}/2\}) \\ & \quad + \sum_{r=1}^n P(E_{n,r} \cap \{|(N_{r'} - r'\mu^{-1}) - (N_{\beta_{n,r}} - \beta_{n,r}\mu^{-1})| \geq \epsilon n^{1/2}/2\}). \end{aligned}$$

Now  $N_{\beta_{n,r}} - N_r$  is 0 or 1, so that  $|(N_{\beta_{n,r}} - \beta_{n,r}\mu^{-1}) - (N_r - r\mu^{-1})| \leq 1 + \mu^{-1} \cdot (\beta_{n,r} - r)$ . Hence if  $n$  is large enough that  $\epsilon n^{1/2} \geq 4$ , the first sum on the right

in (7.4) does not exceed  $P(\bigcup_{r=1}^n \{\beta_{n,r} - r \geq \epsilon n^{1/2}/4\})$ . But this is less than or equal to the probability that before the  $n$ th occurrence of  $\varepsilon$  there is a sequence of at least  $\epsilon n^{1/2}/4$  consecutive trials at none of which  $\varepsilon$  occurs, which probability is clearly not greater than  $P\{\max_{i \leq n} X_i \geq \epsilon n^{1/2}/4\} \leq nP\{X_1 \geq \epsilon n^{1/2}/4\}$ . Since  $X_1$  has a finite second moment  $nP\{X_1 \geq \epsilon n^{1/2}/4\} = o(1)$  and hence the first sum on the right in (7.4) goes to 0 as  $n \rightarrow \infty$ .

By the defining properties of recurrent events, the second sum on the right in (7.4) is equal to

$$(7.5) \quad \sum_{r=1}^n P(E_{n,r}) P\{|N_{r'} - N_{\beta_{n,r}} - (r' - \beta_{n,r})\mu^{-1}| \geq \epsilon n^{1/2}/2\}.$$

It is shown in [8, p. 111] that  $E\{N_k\} = k\mu^{-1} + (\sigma^2 + \mu + \mu^2)\mu^{-2}/2 - 1 + o(1)$  and  $\text{Var}\{N_k\} \sim k\sigma^2\mu^{-3}$ , so that  $E\{(N_k - k\mu^{-1})^2\} = O(k)$ . If we now apply Chebyshev's inequality to the terms of (7.5), (3.2) follows from (7.4) and the result of the preceding paragraph.

#### REFERENCES

1. P. Alexandroff and H. Hopf, *Topologie*, Berlin, 1935.
2. S. Bernstein, *Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendants*, Math. Ann. vol. 97 (1927) pp. 1-59.
3. P. Diananda, *Some probability limit theorems with statistical applications*, Proc. Cambridge Philos. Soc. vol. 49 (1953) pp. 239-246.
4. M. Donsker, *An invariance principle for certain probability limit theorems*, Memoirs of the American Mathematical Society, no. 6, 1951.
5. J. L. Doob, *Stochastic processes*, New York, 1953.
6. P. Erdős and M. Kac, *On certain limit theorems in the theory of probability*, Bull. Amer. Math. Soc. vol. 52 (1946) pp. 292-302.
7. ———, *On the number of positive sums of independent random variables*, Bull. Amer. Math. Soc. vol. 53 (1947) pp. 1011-1020.
8. W. Feller, *Fluctuation theory of recurrent events*, Trans. Amer. Math. Soc. vol. 67 (1949) pp. 98-119.
9. W. Hoeffding and H. E. Robbins, *The central limit theorem for dependent random variables*, Duke Math. J. vol. 15 (1948) pp. 773-780.
10. P. Lévy, *Théorie de l'addition des variables aléatoires*, Paris, 1937.
11. A. M. Mark, *Some probability limit theorems*, Bull. Amer. Math. Soc. vol. 55 (1949) pp. 885-900.
12. G. Marsaglia, *Iterated limits and the central limit theorem for dependent variables*, Proc. Amer. Math. Soc. vol. 6 (1954) pp. 987-991.
13. Yu. V. Prohorov, *Probability distributions in functional spaces*, Uspehi Matematicheskikh Nauk (N.S.) vol. 8 (1953) pp. 165-167 (in Russian).

PRINCETON UNIVERSITY,  
PRINCETON, N. J