SOME INVARIANCE PRINCIPLES FOR FUNCTIONALS OF A MARKOV CHAIN[1]

By David A. Freedman

University of California, Berkeley

1. Introduction. This paper is mainly expository, and deals with the asymptotic behavior of the partial sums of functionals of a Markov chain. Some new results are proved in Sections 3 and 4. Section 4 contains (16), an arcsine law for functional processes. The invariance principles of Donsker (1951) and Strassen (1964) are discussed in Section 5 (which does not depend on the earlier sections). In view of Section 3, these theorems generalize immediately to cover functional processes. For precise statements, see (18) and (25).

One basic technique in this note, of partitioning the sample sequence according to the occurrence of a fixed state, goes back to Doeblin (1938). Finally, it is a pleasure to acknowledge my debt to Chung (1960).

2. Summary. The new results are (C), (D) and (F) below. Let $X_0, X_1, \cdots$ be a Markov chain with countable state space I and stationary transitions. Suppose I is a positive recurrent class, with stationary probability vector p. Let f be a real-valued function on I. Fix a reference state $s \varepsilon I$, and let $0 \leqq t_1 < t_2 < \cdots$ be the times n at which $X_n = s$. Let

$$Y_j = \sum \{f(X_n): t_j \leqq n < t_{j+1}\}$$

and

$$U_j = \sum \{|f(X_n)|: t_j \leqq n < t_{j+1}\}.$$

Let $V_m = \sum_{j=1}^m Y_j$ and $S_n = \sum_{j=0}^n f(X_j)$. For (C) and (D) below, assume

(A) $\sum_{i \varepsilon I} p_i f(i) = 0$;

and

(B) U_j^2 has finite expectation.

Then:

(C) Theorem. $n^{-\frac{1}{2}} \max \{|S_j - V_{jp_s}|: 1 \leqq j \leqq n\} \to 0$ *in probability;*

and

(D) Theorem. $(n \log \log n)^{-\frac{1}{2}} \max \{|S_j - V_{jp_s}|: 1 \leqq j \leqq n\} \to 0$ *almost everywhere.*

For (F), do not assume (A) and (B), but assume

(E) Y_j differs from 0 with positive probability.

Let v_m (respectively, s_n) be 1 or 0 according as V_m (respectively, S_n) is posi-

Received 19 September 1966

[1] Special Invited Paper presented at the meeting of the Institute of Mathematical Statistics, Rutgers, 31 August 1966, by invitation of the IMS Committee on Special Invited Papers. Prepared with the partial support of the National Science Foundation, Grant GP-2593, and by a grant from the Sloan Foundation.

tive or non-positive. Then

(F) THEOREM. $n^{-1} \sum \{s_j : 1 \leqq j \leqq n\} - p_s^{-1} n^{-1} \sum \{v_j : 1 \leqq j \leqq np_s\} \to 0$ *almost everywhere.*

I do not believe the convergence in (C) is a.e., but have no counter-example.

3. Proofs of (C) **and** (D). By the strong Markov property, $Y_1, Y_2, \cdots$ are independent and identically distributed. So are $U_1, U_2, \cdots$. By (A), $E(Y_j) = 0$, (Chung, 1960, pp. 81–82); by (B), $E(Y_j^2) < \infty$. On $[t_1 \leqq n]$, let: $l(n)$ be the largest j with $t_j \leqq n$; $Y'(n) = \sum \{f(X_j) : 0 \leqq j \leqq t_1 - 1\}$; $Y''(n) = \sum \{f(X_j) : t_{l(n)} \leqq j \leqq n\}$. On $[t_1 > n]$, let: $l(n) = 0$; $Y'(n) = S_n$; $Y''(n) = 0$.

Then (Chung, 1960, p. 78)

$$S_n = Y'(n) + Y''(n) + V_{l(n)-1}, \tag{1}$$

and (C) follows from (2) through (6).

(2) LEMMA. *Let $a_1, a_2, \cdots$ and $0 < b_1 \leqq b_2 \leqq \cdots$ be real numbers with $b_j \to \infty$ and $a_j/b_j \to 0$. Then $\max \{a_1, \cdots, a_n\}/b_n \to 0$.*

PROOF. Easy. □

(3) LEMMA. $n^{-\frac{1}{2}} \max \{|Y'(j)| : 1 \leqq j \leqq n\} \to 0$ a.e.

PROOF. $|Y'(n)| \leqq \sum \{|f(X_j)| : 0 \leqq j \leqq t_1 - 1\}$. □

(4) LEMMA. $n^{-\frac{1}{2}} \max \{|Y''(j)| : 1 \leqq j \leqq n\} \to 0$ a.e.

PROOF. Plainly, $|Y''(n)| \leqq U_{l(n)}$. But $l(n) \leqq n + 1$, so $|Y''(n)| \leqq \max \{U_j : 1 \leqq j \leqq n + 1\}$. From (B) and the Borel-Cantelli lemmas, $n^{-\frac{1}{2}} U_{n+1} \to 0$ a.e., so (2) implies $n^{-\frac{1}{2}} \max \{U_j : 1 \leqq j \leqq n + 1\} \to 0$ a.e., that is, $n^{-\frac{1}{2}} |Y''(n)| \to 0$ a.e. Use (2) again. □

(5) LEMMA. $n^{-1} l(n) \to p_s$ a.e.

PROOF. Use the strong law, or see (Chung, 1960, p. 87). □

(6) LEMMA. $n^{-\frac{1}{2}} \max \{|V_{l(j)-1} - V_{jp_s}| : 1 \leqq j \leqq n\} \to 0$ *in probability.*

PROOF. Use (5) and (23). □

This completes the proof of (C). The proof of (D) is similar, using (29) instead of (23).

4. Proof of (F). Recall (E). Replacing f by $-f$ if necessary, suppose that $V_m > 0$ for infinitely many m along almost all paths.

Let $r_j = t_{j+1} - t_j$. The r_j are independent, identically distributed, and have mean $1/p_s$. Theorem (F) will be proved by establishing:

$$p_s^{-1} n^{-1} \sum \{v_j : 1 \leqq j \leqq np_s\} - n^{-1} \sum \{v_j r_{j+1} : 1 \leqq j \leqq l(n) - 2\} \to 0 \text{ a.e.} \tag{7}$$

and

$$n^{-1} \sum \{v_j r_{j+1} : 1 \leqq j \leqq l(n) - 2\} - n^{-1} \sum \{s_m : 1 \leqq m \leqq n\} \to 0 \text{ a.e.} \tag{8}$$

Relation (8) says that essentially all $m \leqq n$ are in $[t_{j+1}, t_{j+2})$ over which $s_m = v_j$.

The proofs of (7) and (8) require lemmas (9) and (10). For (9), let $r_1, r_2, \cdots$ be any independent, identically distributed random variables. Let $\mathfrak{F}_1 \subset \mathfrak{F}_2 \subset \cdots$ be σ-fields, such that r_n is $\mathfrak{F}_{n+1}$-measurable, and $\mathfrak{F}_n$ is independent of $r_n, r_{n+1}, \cdots$. Let $z_1, z_2, \cdots$ be random variables taking only the values 0 and 1, such that z_n is $\mathfrak{F}_{n+1}$-measurable, and $\sum_1^\infty z_n = \infty$ a.e.

(9) LEMMA. *If $E|r_j| < \infty$, the ratio of $\sum_{j=1}^n z_j r_{j+1}$ to $\sum_{j=1}^n z_j$ converges to $E(r_j)$* a.e.

PROOF. Let $Z_n = \sum_1^n z_j$ and $W_m = 1 +$ smallest n such that $Z_n = m$. Because $[W_m = j] \varepsilon \mathfrak{F}_j$, the joint distributon of $r_{W_1}, r_{W_2}, \cdots$ coincides with the joint distribution of $r_1, r_2, \cdots$. By the strong law, $n^{-1} \sum_1^n r_{W_j} \to E(r_1)$ a.e. Because $Z_n \to \infty$, $Z_n^{-1} \sum \{r_{W_j} : 1 \leqq j \leqq Z_n\} \to E(r_1)$ a.e. But $\sum \{z_j r_{j+1} : 1 \leqq j \leqq n\} = Z_n Z_n^{-1} \sum \{r_{W_k} : 1 \leqq k \leqq Z_n\}$. □

For (10), let $Y_1, Y_2, \cdots$ be any sequence of independent, identically distributed random variables, with $V_n = \sum_{j=1}^n Y_j$. Make no assumptions about the moments of Y_j. Let M be a positive, finite number; let $d_n = 1$ or 0 according as $|V_n| \leqq M$ or $|V_n| > M$.

(10) LEMMA. *Unless Y_j is* 0 a.e., $n^{-1} \sum_1^n d_j \to 0$ a.e.

PROOF. Let C_n be the concentration function of V_n. Fix r equal to one of $0, \cdots, k-1$. The conditional probability that $|V_{nk+r}| \leqq M$ given $Y_1, \cdots, Y_{(n-1)k+r}$ is at most $C_k(2M)$. Consequently, $\limsup_{n\to\infty} n^{-1} \sum_{j=1}^n d_{jk+r} \leqq C_k(2M)$ a.e., for example using (1) of (Dubins and Freedman, 1965). The result is clear unless Y_j is nondegenerate, in which case $\lim_{k\to\infty} C_k(2M) = 0$ (Rogozin, 1961). □

PROOF OF (7). As is clear from (9), with $\mathfrak{F}_j = \mathfrak{F}(X_n : n \leqq t_j)$ in the sense of (Chung, 1960, p. 72): the ratio of $\sum_1^m v_j$ to $\sum_1^m v_j r_{j+1}$ converges to p_s a.e. Set $m = l(n) - 2$:

(11) the ratio of $p_s^{-1} n^{-1} \sum \{v_j : 1 \leqq j \leqq l(n) - 2\}$

to $n^{-1} \sum \{v_j r_{j+1} : 1 \leqq j \leqq l(n) - 2\}$ converges to 1 a.e.

From (5),

(12) $p_s^{-1} n^{-1} \sum \{v_j : 1 \leqq j \leqq l(n) - 2\} - p_s^{-1} n^{-1} \sum \{v_j : 1 \leqq j \leqq np_s\} \to 0$ a.e.

Combine (11) and (12). □

PROOF OF (8). Fix $\epsilon > 0$. Choose M so large that $\int_{U_j \leqq M/2} r_j \geqq (1 - \frac{1}{2}\epsilon)/p_s$. It will now be seen that for almost all sample sequences, for all but finitely many n, for at least $(1-\epsilon)n$ of the $m \leqq n$, there is a (unique) $j = 1, 2, \cdots$ such that $t_{j+1} \leqq m < t_{j+2} \leqq n$, $|V_j| > M$, and $|U_{j+1}| \leqq M/2$. For such an m, $s_m = v_j$, provided $\sum \{|f(X_k)| : 0 \leqq k \leqq t_1 - 1\} \leqq M/2$.

Indeed, by (9) and (10) (using assumption (D)), for almost all sample sequences, for all but finitely many n, at most $\frac{1}{2}\epsilon n$ of $m \leqq n$ are in intervals $t_{j+1} \leqq m < t_{j+2} \leqq n$ with $|V_j| \leqq M$ and $j = 1, 2, \cdots$. By the strong law, $k^{-1} \sum \{r_{j+1} : 1 \leqq j \leqq k-2 \text{ and } U_j \leqq M/2\}$ has lim inf at least $(1 - \frac{1}{2}\epsilon)p_s$ a.e. Specialize $k = l(n)$ and recall (5), to see that for almost all sample sequences, for all but finitely many n, at least $(1 - \frac{1}{2}\epsilon)n$ of $m \leqq n$ are in intervals $t_{j+1} \leqq m < t_{j+2} \leqq n$ with $U_{j+1} \leqq M/2$ and $j = 1, 2, \cdots$. □

Now consider the exceptional case

(13) $$Y_j = 0 \quad \text{a.e.}$$

As is easily seen, if (13) holds for one s, it holds for all s. Suppose

(14) $$X_\theta = s \quad \text{a.e.},$$

so $t_1 = 0$ a.e. Because $V_j = S_{t_{j+1}-1} = 0$ a.e., the vectors $S_m : t_j \leqq m < t_{j+1}$ for $j = 1, 2, \cdots$ are independent and identically distributed. By the strong law,

(15) $$(1/n) \sum_1^n s_m \to p_s E[\sum \{s_m : 0 \leqq m < t_2\} \mid X_0 = s] \quad \text{a.e.}$$

The limit may depend on s.

Further information about case (13) can be found in (Chung, 1960, Theorem 2 on p. 95).

To state the arcsine law (16), define F_α as follows. For $\alpha = 0$ or 1, let F_α be the distribution function of point mass at α. For $0 < \alpha < 1$, let F_α be the distribution function on [0, 1] with density proportional to $y \to y^{\alpha-1}(1 - y)^{-\alpha}$.

(16) COROLLARY. *Suppose* (E). *The distribution of* $n^{-1} \sum_1^n s_j$ *converges to* F_α *if and only if its mean converges to* α.

PROOF. This is immediate from (Spitzer, 1956, Theorem 7.1) and (F). ▯

This corollary appears in (Freedman, 1963).

5. Some invariance principles. Let g ge a real-valued function on the positive integers, whose value at n is g_n. Let $g_{(n)}$ be the continuous real-valued function on [0, 1], whose value at j/n is $n^{-\frac{1}{2}} g_j$, and which is linearly interpolated. Let $Y_1, Y_2, \cdots$ be a sequence of independent, identically distributed random variables, of mean 0 and variance σ^2. Let $V_0 = 0$ and $V_n = Y_1 + \cdots + Y_n$. Let $B_t : 0 \leqq t < \infty$ be standard Brownian motion. Let $C[0, 1]$ be the space of continuous functions on [0, 1], with the supremum norm. The distribution β_{σ^2} of $B_{t\sigma^2} : 0 \leqq t \leqq 1$ is a probability on $C[0, 1]$. Let φ be a bounded, real-valued, measurable function on $C[0, 1]$, continuous β_{σ^2}-a.e. The invariance principle of Donsker (1951) is:

(17) THEOREM. $E[\varphi(V_{(n)})] \to \int_{f \varepsilon C[0,1]} \varphi(f) \beta_{\sigma^2}(df)$.

A proof of (17) will be sketched below. With the help of a topological fact (19), Theorems (17) and (C) imply this invariance principle for functional processes:

(18) COROLLARY. *If* (A) *and* (B) *hold, then*

$$E[\varphi(S_{(n)})] \to \int_{f \varepsilon C[0,1]} \varphi(f) \beta_{\sigma^2}(df),$$

where $\sigma^2 = p_s E(Y_j^2)$.

(19) LEMMA. *Let* $(\mathfrak{X}, \rho)$ *be a metric space.*

(a) *Let* C_0 *and* C_1 *be subsets of* $\mathfrak{X}$ *with* $\rho(C_0, C_1) > 0$. *Then there is a* ρ*-uniformly continuous, real-valued function* f *on* $\mathfrak{X}$, *with* $0 \leqq f \leqq 1$, $f = 0$ *on* C_0, $f = 1$ *on* C_1.

(b) *Let* φ *be a bounded, real-valued, function on* $\mathfrak{X}$. *Then there are bounded, real-valued,* ρ*-uniformly continuous functions* f_k *and* g_k *on* $\mathfrak{X}$, *with*

(i) $f_k \leqq f_{k+1} \leqq \varphi \leqq g_{k+1} \leqq g_k$,

(ii) $\lim f_k(x) = \varphi(x)$ iff φ *is lower semicontinuous at* x,

(iii) $\lim g_k(x) = \varphi(x)$ iff *g is upper semicontinuous at x.*

Proof of (a). Let $g(x) = \rho(x, C_0)$, $\gamma = \inf \{g(x): x \varepsilon C_1\}$, $f = \min \{\gamma^{-1}g, 1\}$. □

Proof of (b). Without real loss, suppose $0 < \varphi \leqq 1$. The dsicussion of g_k is omitted, being similar to that for f_k. Using (a), construct a ρ-uniformly continuous, real-valued function $f(n, m, j)$ on $\mathfrak{X}$, bounded between 0 and j/n, such that: for x with $\varphi(x) \leqq j/n, f(n, m, j)(x) = 0$; for x with $\rho(x, [\varphi \leqq j/n]) \geqq 1/m$, $f(n, m, j) = j/n$. Clearly, $f(n, m, j) \leqq \varphi$. Let $f_k = \max \{f(n, m, j); 1 \leqq n, m \leqq k$ and $1 \leqq j \leqq n\}$. Clearly, $f_k \leqq \varphi$, f_k is nondecreasing with k, and f_k is bounded and ρ-uniformly continuous. Suppose φ is lower semicontinuous at x. Let $j/n < \varphi(x) \leqq (j + 1)/n$. Then x is a positive ρ-distance from $[\varphi \leqq j/n]$, and for large $m, f(n, m, j) = j/n$. □

With the help of (19), Theorem (17) follows easily from

(20) Lemma. *There is a standard Brownian motion $B_t : 0 \leqq t < \infty$, and a stochastic process $V_1^*, V_2^*, \cdots$, on the same probability triple, such that:*

(i) *the joint distribution of $V_1^*, V_2^*, \cdots$ coincides with the joint distribution of $V_1, V_2, \cdots$;*

(ii) $n^{-\frac{1}{2}} \max \{|V_j^* - B_{j\sigma^2}|: 1 \leqq j \leqq n\}$ *converges to* 0 *in probability.*

In turn, (20) follows from (21) and (22). Here (21) is trivial, since B has continuous sample functions. Proposition (22) is an elegant result of Skorokhod (1964), which I learned from Strassen.

(21) Lemma. *Let $\epsilon > 0$. The probability that* $\max \{|B_t - B_s|: 0 \leqq s \leqq n, s \leqq t \leqq rs\}$ *exceeds $\epsilon n^{\frac{1}{2}}$ does not depend on n, and converges to* 0 *as r decreases to* 1.

(22) Proposition. *Suppose there is a random variable independent of B and uniformly distributed on* [0, 1]. *Then there is a sequence $\tau_1 \leqq \tau_2 \leqq \cdots$ of stopping times for B, such that:*

(i) *$\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2, \cdots$ are independent, identically distributed, and have mean σ^2;*

(ii) *the joint distribution of $B_{\tau_1}, B_{\tau_2}, B_{\tau_3}, \cdots$ coincides with the joint distribution of $V_1, V_2, V_3, \cdots$.*

Proof. Suppose first $P(Y_j = -a) = b/(a + b)$ and $P(Y_j = b) = a/(a + b)$, where $a, b > 0$. Let $\tau_0 = 0$ and let τ_{n+1} be the least $t > \tau_n$ with $B_t = B_{\tau_n} - a$ or $B_t = B_{\tau_n} + b$. Because B is a martingale, B_{τ_1} is distributed like Y_1. Because $\{B_t^2 - t: t \geqq 0\}$ is a martingale, $E(\tau_1) = ab = E(Y_1^2)$. Use the strong Markov property to complete the proof in this case. For the general case, represent the distribution of Y_1 as an average of two-point distributions having mean 0. □

(23) Corollary. *Let $\epsilon > 0$. The probability that* $\max \{|V_j - V_k|: 1 \leqq j \leqq n, j \leqq k \leqq rj\}$ *exceeds $\epsilon n^{\frac{1}{2}}$ converges to* 0 *as r decreases to* 1, *uniformly in n.*

Proof. Use (21) and (22). □

There is a beautiful a.e. invariance principle of Strassen (1964), which gives a remarkable insight into the law of the iterated logarithm. To state it, let K_{σ^2} be the set of absolutely continuous functions f on [0, 1] satisfying: $f(0) = 0$ and $\int_0^1 |f'(t)|^2 \, dt \leqq \sigma^2$.

Strassen's result is

(24) Theorem. *For almost all sample sequences, the subset* {(2 log log

$n)^{-\frac{1}{2}}V_{(n)} : n = 3, 4, \cdots\}$, *of* $C[0, 1]$ *is relatively compact, and its set of limit points is* K_{σ^2}.

This result and (D) imply

(25) COROLLARY. *If* (A) *and* (B) *hold, then for almost all sample sequences, the subset* $\{(2 \log \log n)^{-\frac{1}{2}}S_{(n)} : n = 3, 4, \cdots\}$ *of* $C[0, 1]$ *is relatively compact, and its set of limit points is* K_{σ^2}, *where* $\sigma^2 = p_s E(Y_j^2)$.

Strassen's proof of (24) will now be outlined. Let $C_n = B_{n\sigma^2}$. The main step is this special case of (24):

(26) PROPOSITION. *For almost all sample functions, the subset* $\{(2 \log \log n)^{-\frac{1}{2}}C_{(n)} : n = 3, 4, \cdots\}$ of $C[0, 1]$ *is relatively compact, and its set of limit points is* K_{σ^2}.

The argument is omitted.

Now (24) follows from (26) with the help of (2), (22) and

(27) LEMMA. $(n \log \log n)^{-\frac{1}{2}}|B_{\tau_n} - C_n| \to 0$ a.e.

Here τ_n is the stopping time of Skorokhod's proposition (22).

In turn, (27) follows from

(28) LEMMA. *Let* $\epsilon > 0$. *There is an* $r > 1$ *such that* A_n *occurs infinitely often with probability* 0, *where* A_n *is the event that* $\max\{|B_t - B_s| : 0 \leqq s \leqq n, s \leqq t \leqq rs\}$ *exceeds* $\epsilon(n \log \log n)^{\frac{1}{2}}$.

PROOF. Let E_n be the event that $\max\{|B_t - B_s| : 1 \leqq s \leqq n, s \leqq t \leqq rs\}$ exceeds $\epsilon(n \log \log n)^{\frac{1}{2}}$. Plainly, it is enough to prove that E_n occurs infinitely often with probability 0. Let $F_k = \cup\{E_n : r^k \leqq n \leqq r^{k+1}\}$. By Borel-Cantelli, it suffices to prove that $\sum P(F_k) < \infty$. Let $F_{k,j}$ be the event that $\max\{|B_t - B_{rj}| : r^j \leqq t \leqq r^{j+2}\}$ exceeds $\frac{1}{2}\epsilon[r^k(\log k + \log \log r)]^{\frac{1}{2}}$. Clearly, $F_k \subset \bigcup_{j=1}^k F_{k,j}$, so $P(F_k) \leqq \sum_{j=1}^k P(F_{k,j})$. But $P(F_{k,j})$ is easy to estimate from the reflection principle (Doob, 1953, Theorem 2.1 on p. 392), and $P(F_k)$ is $O(k^{-a})$, $a = \epsilon^2/[8(r^2 - 1)] > 1$ *for small* $r > 1$. □

(29) COROLLARY. *Let* $\epsilon > 0$. *There is an* $r > 1$ (*depending only on* ϵ) *such that* A_n *occurs infinitely often with probability* 0, *where* A_n *is the event that*

$$\max\{|V_j - V_k| : 1 \leqq j \leqq n, j \leqq k \leqq rj\}$$

exceeds $\epsilon(n \log \log n)^{\frac{1}{2}}$.

PROOF. Use (22) and (28). □

REFERENCES

CHUNG, KAI LAI (1960). *Markov chains with stationary transition probabilities*. Springer, Berlin.

DOEBLIN, W. (1938). Sur deux problemes de M. Kolmogoroff concernant les chaînes dénombrables, *Bull. Soc. Math. France*. **66** 210–220.

DONSKER, MONROE D. (1951). An invariance principle for certain probability limit theorems. *Mem. Am. Math. Soc.* **6**.

DOOB, J. L. (1953). *Stochastic processes*. Wiley, New York.

DUBINS, L. E. and FREEDMAN, D. A. (1965). A sharper form of the Strong Law. *Ann. Math. Statist.* **36** 800–807.

FREEDMAN, D. A. (1963). An arcsine law for Markov chains. *Proc. Amer. Math. Soc.* **14** 680–684.

ROGOZIN, B. A. (1961). On an estimate of the concentration function. *Theor. Prob. Appl.* **6** 94–96.

SKOROKHOD, A. (1965). *Studies in the theory of random processes.* Addison-Wesley, Reading.

SPITZER, F. (1956). A combinatorial lemma and its applications to probability theory. *Trans. Amer. Math. Soc.* **82** 323–339.

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