EXPONENTIAL CONVERGENCE RATES FOR THE LAW OF LARGE NUMBERS

BY

LEONARD E. BAUM[1], MELVIN KATZ[2], AND ROBERT R. READ[2]

Introduction. Consider a sequence of random variables $\{X_k : k = 1, 2, \cdots\}$ obeying the law of large numbers, i.e., there exists a constant c such that for every $\epsilon > 0$ the sequence of probabilities $P\{|n^{-1}\sum_{k=1}^{n} X_k - c| > \epsilon\} = P_n(\epsilon)$ converges to zero as $n \to \infty$. The object of the present paper is to study the relationships among an exponential convergence rate (i.e., $P_n(\epsilon) = O(\rho^n)$ for $\rho = \rho(\epsilon) < 1$), the existence of the individual moment generating functions and the stochastic structure of the sequence $\{X_k\}$. Papers containing related studies (e.g., [1; 3; 5]) have treated the case of independent random variables and demonstrated exponential convergence under the hypothesis that the moment generating functions exist.

The present paper studies the extent to which an exponential convergence rate implies the existence of the moment generating function and conversely. In particular, satisfactory necessary and sufficient conditions (Theorem 2) are found for exponential convergence of sequences of independent (not necessarily identically distributed) random variables.

In the first section it is proved that an exponential rate of convergence for any stationary sequence necessarily implies the existence of the moment generating function of the random variables. Conversely an example is constructed showing that restrictions on the size of the variables cannot be sufficient to insure exponential convergence for the general stationary sequence. The terms size, smallness, etc., as used throughout the paper, refer essentially to the tail probabilities. The case of independence is treated in §2 where it is shown that the existence of all the moment generating functions on the same interval is necessary and that a growth restriction on the product of the first n of these generating functions is necessary and sufficient. In §3 it is shown that the existence of the moment generating functions is necessary for exponential convergence of the averages of a function of the variables of a Markov sequence having stationary transition probabilities. If the process satisfies Doeblin's condition then this condition is proved sufficient. §4 contains an example showing that the existence of moment generating functions

generally is not necessary for broader classes of processes. The particular variables constructed are martingale differences, square summable (thus orthogonal) and form a Markov sequence. In the final section exponential convergence to $+\infty$ is treated. An example shows that under the natural definition of such convergence we cannot extend the result of the first section to assert the existence of the moment generating function of X_1^- when $\{X_k\}$ is an ergodic stationary process.

1. **Stationary random variables.** Let $\{X_k: k=1, 2, \cdots \}$ be an arbitrary sequence of random variables.

DEFINITION. The sequence $\{X_k: k=1, 2, \cdots \}$ converges exponentially rapidly to the constant c if for all $\epsilon > 0$ there exist constants A and $\rho < 1$ (depending on ϵ) such that

$$(1) \qquad P\left\{ \left| n^{-1} \sum_{k=1}^{n} X_k - c \right| \geqq \epsilon \right\} \leqq A\rho^n, \qquad n = 1, 2, \cdots .$$

THEOREM 1. *Let $\{X_k: k=1, 2, \cdots \}$ be a stationary sequence of random variables such that (1) is valid for some c, $\epsilon > 0$, A and $\rho < 1$. Then there exists $t_0 > 0$ such that*

$$Ee^{tX_1} < \infty \qquad \text{for } t \in [-t_0, t_0].$$

Proof. We assume, with no loss of generality, that $c = 0$. First observe that

$$(2) \quad P\left\{ X_1 \geqq (2n-1)\epsilon \text{ and } \sum_{k=2}^{n} X_k \geqq -(n-1)\epsilon \right\}$$

$$\geqq P\{ X_1 \geqq (2n-1)\epsilon \} + P\left\{ \sum_{k=2}^{n} X_k \geqq -(n-1)\epsilon \right\} - 1.$$

It follows from stationarity that

$$(3) \qquad P\left\{ \sum_{k=2}^{n} X_k \geqq -(n-1)\epsilon \right\} \geqq 1 - A\rho^{n-1}$$

and thus (2) yields

$$(4) \quad P\left\{ X_1 \geqq (2n-1)\epsilon \text{ and } \sum_{k=2}^{n} X_k \geqq -(n-1)\epsilon \right\}$$

$$\geqq P\{ X_1 \geqq (2n-1)\epsilon \} - A\rho^{n-1}.$$

By hypothesis $P\{ X_1 \geqq (2n-1)\epsilon$ and $\sum_{k=2}^{n} X_k \geqq -(n-1)\epsilon \}$ $\leqq P\{ \sum_{k=1}^{n} X_k \geqq n\epsilon \} \leqq A\rho^n$ and thus from (4) we obtain

$$(5) \qquad P\{ X_1 \geqq (2n-1)\epsilon \} \leqq A\rho^n + A\rho^{n-1} \leqq 2A\rho^{n-1}.$$

This bound on the measure of the sets where X_1^+ is large implies the existence of a moment generating function for X_1^+ in some nondegenerate interval

about 0. The above proof may now be repeated with the events $\{X_1 \leq -(2n-1)\epsilon\}$ and $\{\sum_{k=2}^{n} X_k \leq (n-1)\epsilon\}$ replacing $\{X_1 \geq (2n-1)\epsilon\}$ and $\{\sum_{k=2}^{n} X_k \geq -(n-1)\epsilon\}$ respectively, thus yielding a proof for the existence of the moment generating function of X_1^- in a nondegenerate interval about 0. This completes the proof since $Ee^{tX_1} \leq Ee^{tX_1^+} + Ee^{-tX_1^-}$.

The proof of the above theorem can be applied using rates of convergence other than exponential. The rate controls the size of the X_k. The almost everywhere convergence of $n^{-1} \sum_{k=1}^{n} X_k$ alone does not insure the smallness of the X_k (i.e., the converse of the ergodic theorem is false [6, p. 32]) essentially because of the possibility of cancellation of positive and negative values of the X_k. Other conditions under which convergence will insure the smallness of the X_k include (i) assuming the X_k to be semi-integrable [ibid.], and (ii) making assumptions on the joint distributions, e.g., independence (converse of the strong law of large numbers).

The following example demonstrates that an exponential rate of convergence for a stationary stochastic process $\{X_k\}$ does not follow from the smallness of the X_k and any properties (e.g., mixing properties) of the associated shift transformation generating the process but is essentially a property of the joint distributions themselves.

Let T be an ergodic measure preserving transformation on $(\Omega, \mathfrak{B}, P)$ and $\{A_i : i = 1, 2, \cdots\}$ be a sequence of measurable subsets of Ω such that $P(A_i) = 1/i^3$. Define $A = \bigcup_{i=2}^{\infty} \bigcup_{j=0}^{i-1} T^j A_i$. Then $P(A) \leq \sum_{i=2}^{\infty} i/i^3 = \pi^2/6 - 1$ and if χ_A denotes the characteristic function of the set A, we have $E(\chi_A) < 1$. However, for $n = 2, 3, \cdots$

$$P\left\{n^{-1} \sum_{j=0}^{n-1} \chi_A(T^j) = 1\right\} \geq P(A_n) = n^{-3}.$$

Thus since n^{-3} is eventually greater than $A\rho^n$ for any constants A and $\rho < 1$, an exponential rate of convergence for the stationary $(0, 1)$ process $\{\chi_A(T^j) \equiv X_j : j = 0, 1, \cdots\}$ to its limit $(\pi^2/6 - 1)$ is impossible for any $\epsilon < 1 - (\pi^2/6 - 1)$.

2. **Independent random variables.** Analogous to Theorem 1 we have the following lemma.

LEMMA. *Let $\{X_k : k = 1, 2, \cdots\}$ be a sequence of independent random variables such that* (1) *holds for some constants c, A, $\epsilon > 0$ and $\rho < 1$. Then each X_k has a moment generating function and these moment generating functions all exist on some common interval about zero.*

Proof. Choosing $c = 0$ as previously, we note that

$$(6) \quad A\rho^n \geq P\left\{\sum_{k=1}^{n} X_k \geq 2n\epsilon\right\} \geq P\{X_j \geq 4n\epsilon\} - P\left\{\sum_{k=1; k \neq j}^{n} X_k \leq -2n\epsilon\right\}$$

for all $n \geq j$. From the assumption of independence it follows that

$$(7) \qquad P\left\{ \sum_{k=1; k \neq j}^{n} X_k \leq -2n\epsilon \right\} P\{X_j \leq n\epsilon\} \leq P\left\{ \sum_{k=1}^{n} X_k \leq -n\epsilon \right\} \leq A\rho^n.$$

Let $1 > \delta > 0$. For each j there exists an n_j sufficiently large so that $P\{X_j \leq n\epsilon\} \geq \delta$ for all $n \geq n_j$. Thus from (6) and (7) we obtain

$$(8) \qquad P\{X_j \geq 4n\epsilon\} \leq \frac{A}{\delta} \rho^n + A\rho^n \leq \frac{2A}{\delta} \rho^n = B\rho^n$$

for $n \geq n_j$. Since for $t > 0$, $Ee^{tX_j^+} \leq 1 + \sum_k e^{(4k+1)t\epsilon} P\{X_j \geq 4(k-1)\epsilon\}$ $\leq 1 + \sum_{k=1}^{n_j+1} e^{(4k+1)t\epsilon} + Be^{t\epsilon}\rho^{-1} \sum_{k=1}^{\infty} [e^{4t\epsilon}\rho]^k < \infty$ if $e^{4t\epsilon}\rho < 1$ and a similar argument provides the same estimate for $Ee^{tX_j^-}$ it follows that $Ee^{tX_j} < \infty$ for all $t \in [-t_0, t_0]$ with t_0 determined independently of j.

Observe then that an exponential rate of convergence of an independent sequence $\{X_k\}$ (or even (1) for a single ϵ and $\rho < 1$) implies the smallness of each X_k. This result contrasts with the situation in, say, almost everywhere convergence in which any finite number of the X_k may be changed arbitrarily without affecting convergence.

THEOREM 2. *Let* $\{X_k : k = 1, 2, \cdots\}$ *be a sequence of independent random variables and let* $S_n = \sum_1^n X_k$. *Then the sequence* $\{X_k\}$ *converges exponentially fast to some constant* c *if and only if for all* $\epsilon > 0$ *there exists a constant* M_ϵ *and* $t_\epsilon > 0$ *such that* $\prod_{k=1}^n Ee^{t(X_k - c)} = Ee^{t(S_n - nc)} \leq M_\epsilon e^{|t|\epsilon n}$ *for* $t \in [-t_\epsilon, t_\epsilon]$.

Proof. We choose $c = 0$ without loss of generality and proceed to the proof of the necessity part of the argument. Given $\epsilon > 0$ take $0 < \delta < \epsilon$ and write

$$(9) \qquad Ee^{tS_n^+} \leq 1 + \sum_{k=0}^{\infty} e^{t(k+1)\delta} P\{S_n \geq k\delta\}, \qquad\qquad t > 0.$$

Partitioning the right-hand side of (9) into blocks of length n and employing the inequality

$$\sum_{k=jn}^{(j+1)n-1} e^{t(k+1)\delta} P\{S_n \geq k\delta\} \leq ne^{t(j+1)n\delta} P\{S_n \geq jn\delta\}$$

we obtain

$$(10) \qquad Ee^{tS_n^+} \leq 1 + n\left[e^{tn\delta} + \sum_{j=0}^{\infty} e^{(j+2)tn\delta} P\{S_n \geq (j+1)n\delta\} \right].$$

To estimate $P\{S_n \geq (j+1)n\delta\}$ we first note that

$$(11) \qquad \begin{aligned} P\{S_n \geq (j+1)n\delta\} &\leq P\{S_n \geq (j+1/2)n\delta\} \\ &\leq P\{S_{jn} \geq jn\delta/2\} + P\{S_{jn} - S_n \leq -(j+1)n\delta/2\}. \end{aligned}$$

By assumption $\{X_k\}$ converges exponentially fast and hence there exist constants $A_{\delta/2}$ and $\rho_{\delta/2} < 1$ such that for all j

$$(12) \qquad\qquad P\{S_{jn} \geq jn\delta/2\} \leq A_{\delta/2}\rho_{\delta/2}^{jn}.$$

To bound the second term on the right-hand side of (11) we note that from the assumption of independence

$$(13) \quad P\{S_n < n\delta/2\}P\{S_{jn} - S_n \leq -(j+1)n\delta/2\} \leq P\{S_{jn} \leq -jn\delta/2\}$$

and thus since $P\{S_n \leq n\delta/2\} \geq 1 - A_{\delta/2}\rho_{\delta/2}^n > \eta > 0$ for n sufficiently large, it follows that

$$(14) \qquad\qquad P\{S_{jn} - S_n \leq -(j+1)n\delta/2\} \leq A_{\delta/2}\rho_{\delta/2}^{nj}\eta^{-1}$$

for n large, and $j=1, 2, \cdots$. (Note, the necessary largeness of n does not depend on j.) Combining (11), (12) and (14) we obtain

$$(15) \qquad\qquad P\{\dot{S}_n \geq (j+1)n\delta\} \leq B_\delta\rho_{\delta/2}^{nj}$$

for large n and all j and consequently it follows from (10) that

$$(16) \qquad \begin{aligned} Ee^{tS_n^+} &\leq 1 + n\left\{ e^{tn\delta} + B_\delta e^{2t\delta n} \sum_{j=0}^\infty e^{tjn\delta}\rho_{\delta/2}^{jn} \right\} \\ &\leq n\left\{ 2^{tn\delta} + B_\delta e^{2t\delta n}/[1 - (\rho_{\delta/2}e^{t\delta})^n] \right\} \end{aligned}$$

for large n and positive t such that $\rho_{\delta/2}e^{t\delta} < 1$. Consequently it is clear that there exists a constant, say d_δ, such that

$$(17) \qquad\qquad Ee^{tS_n^+} \leq nd_\delta e^{2|t|n\delta}$$

for t in some interval about zero. It is clear that an analogous argument shows

$$(18) \qquad\qquad Ee^{tS_n^-} \leq nd_\delta' e^{2|t|n\delta}$$

for t in some interval about zero. Thus it follows that there exist M_ϵ and $t_\epsilon > 0$ such that $Ee^{tS_n} \leq M_\epsilon e^{|t|\epsilon n}$ for $t \in [-t_\epsilon, t_\epsilon]$ and necessity is proved.

For the sufficiency we use the well-known inequality [9, p. 158]

$$P\{S_n \geq n\epsilon\} \leq Ee^{t(S_n - n\epsilon)} \qquad\qquad \text{for } t > 0.$$

Choosing $\delta < \epsilon$ we obtain $P\{S_n \geq n\epsilon\} \leq \exp(-t_\delta n\epsilon)M_\delta \exp(t_\delta n\delta) = M_\delta[\exp t_\delta(\delta-\epsilon)]^n$; a similar argument shows that $P\{S_n \leq -n\epsilon\}$ converges to zero exponentially fast and this completes the proof.

The above necessary and sufficient conditions are "satisfactory" (compare [9, p. 258]) in that they involve functions of quantities associated with the individual random variables themselves, namely products of integrals of

their exponentials, rather than functions of quantities associated with the sums.

If the random variables $\{X_k\}$ are Gaussian, the previous condition can be simply expressed in terms of the means and variances.

COROLLARY. *A sequence of independent Gaussian random variables converges exponentially rapidly to c if and only if the sequence of averages of expected values converge to c and the sequence of averages of the variances is bounded.*

3. **Markov processes.** In this section we shall consider a Markov process $\{X_k: k=0, 1, 2, \cdots\}$ with state space $(\mathfrak{X}, \mathfrak{B})$ and stationary transition probability function $p(x, B)$. Let f denote a real valued measurable function on the state space. In what follows we shall be interested only in exponential rates of convergence that are independent of an arbitrary initial probability measure ν on the state space of the process, i.e., the constants A and $\rho < 1$ do not depend on ν.

THEOREM 3. *Let $\{X_k: k=0, 1, 2, \cdots\}$ be a Markov process as described above. If $P\{|n^{-1}\sum_{k-1}^{n}f(X_k)-c| \geqq \epsilon | x_0\} \leqq A\rho^n$ for all x_0, for some $c, \epsilon > 0$, A and $\rho < 1$, then for each $k \geqq 1$ the moment generating functions $E_{x_0}e^{tf(X_k)}$ exist and are finite for $t \in (-t_0, t_0)$, with $t_0 > 0$ not depending on x_0 or k, and where $E_{x_0}e^{tf(X_k)}$ denotes the expected value with initial measure assigning probability one to x_0.*

Proof. The pattern of proof of Theorem 1 will be followed. The probabilities in the inequalities (2) through (5) should now be interpreted as conditional probabilities given $X_0 = x_0$. Again take $c = 0$. Now (3) follows from

$$P\left\{\sum_{k=2}^{n}f(X_k) \leqq -(n-1)\epsilon \Big| x_0\right\}$$

(19)
$$= E\left\{P\left\{\sum_{k=2}^{n}f(X_k) \leqq -(n-1)\epsilon \Big| X_1\right\} \Big| x_0\right\}$$
$$\leqq E\{A\rho^{n-1} | x_0\}$$
$$= A\rho^{n-1} \qquad \text{for all } x_0.$$

Thus as in the proof of Theorem 1 it follows that $E\{e^{tf^+(X_1)} | x_0\} < \infty$ for t in some nondegenerate interval (independent of x_0) about zero. Consequently the moment generating function of $f^+(X_1)$ for arbitrary initial measure exists in this interval. Also the above argument shows that the conditional moment generating functions of the $f^+(X_j)$ given X_{j-1} are uniformly bounded in the X_{j-1}, and it follows that for $j > 1$

(20)
$$E\{e^{tf^+(X_j)} | x_0\} = E\{E\{e^{tf^+(X_j)} | X_{j-1}\} | x_0\}$$

and hence the moment generating function of $f^+(X_j)$ exists on the same

t-interval as the moment generating function of $f^+(X_1)$. Clearly we can show the existence of the moment generating function for $f^-(X_j)$ in an analogous manner and this completes the proof.

It is known that in general the converse of Theorem 3 is not valid and that in particular, it cannot be assured for stationary Markov processes. It can be demonstrated however for processes whose transition probabilities satisfy Doeblin's condition [4]. For convenience assume that the process $\{X_k; k=0, 1, 2, \cdots\}$ has only one ergodic class. Denote the expected value of $f(X_0)$ with respect to the (now) unique stationary measure on $\mathfrak{X}$ by μ.

THEOREM 4. *Let* $\{X_k: k=0, 1, 2, \cdots\}$ *denote the Markov process determined by* $p(x, B)$, *satisfying Doeblin's condition. Then for every* $\epsilon > 0$ $P\{|n^{-1}\sum_{k=1}^{n} f(X_k) - \mu| \geqq \epsilon | x_0\} \leqq A\rho^n$ *for* $n = 1, 2, \cdots$ *and* A, $\rho < 1$ *independent of* x_0, *if and only if there exists* A_1, $\rho_1 < 1$ *such that* $P\{|f(X_k)| \geqq n\epsilon | x_0\}$ $\leqq A_1 \rho_1^n$ *for* $k, n = 1, 2, \cdots$, *independent of* x_0.

Proof. Clearly the proof of Theorem 3 can be modified slightly to give the proof of the necessity of the condition. To prove sufficiency we proceed as in [8]. There is no loss of generality in assuming $\mu < 0$. It is enough to show that there exist two constants C and $\gamma < 1$, each independent of x_0 such that for all n, $P\{\sum_{k=1}^{n} f(X_k) \geqq 0 | x_0\} \leqq C\gamma^n$, $(\epsilon = -\mu)$. Define

$$(21) \qquad \phi(n, t) = \sup_{x_0 \in \mathfrak{X}} E\left\{\exp\left(t \sum_{k=1}^{n} f(X_k)\right) \Big| x_0\right\}.$$

By hypothesis there exists a nondegenerate interval about zero such that for t included in this interval $\phi(n, t)$ is well defined for all n. If $n = j + l$ then

$$
\begin{aligned}
(22) \qquad & E\left\{\exp\left[t \sum_{k=1}^{n} f(X_k)\right] \Big| x_0\right\} \\
& = E\left\{E\left[\exp\left(t \sum_{k=1}^{i} f(X_k) + t \sum_{k=j+1}^{n} f(X_k)\right) \Big| X_1, \cdots, X_j\right] \Big| x_0\right\} \\
& = E\left\{\exp\left[t \sum_{k=1}^{i} f(X_k)\right] E\left(\exp\left(t \sum_{k=j+1}^{n} f(X_k)\right) \Big| X_j\right) \Big| x_0\right\} \\
& \leqq \phi(j, t)\phi(l, t).
\end{aligned}
$$

Thus $\phi(n, t) \leqq \phi(j, t)\phi(l, t)$. For any integer d and for $n \geqq d$ write $n = md + l$, $0 \leqq l \leqq d - 1$, and then

$$(23) \qquad \phi(n, t) \leqq \phi(md, t)\phi(l, t) \leqq [\phi(d, t)]^m \phi(l, t).$$

Therefore $E(\exp[t \sum_{k=1}^{n} f(X_k)] | x_0)^{1/n} \leqq [\phi(d, t)]^{m/n}[\phi(l, t)]^{1/n}$ and it follows, letting $n \to \infty$, that

(24) $$\limsup_{n} E\left(\exp\left[t\sum_{k=1}^{n}f(X_k)\right]\Big| x_0\right)^{1/n} \leqq [\phi(d, t)]^{1/d}.$$

The result will follow if there exist $t_0 > 0$ and an integer d_0 such that $\phi(d_0, t_0) < 1$, since $P\{\sum_{k=1}^{n}f(X_k) \geqq 0 | x_0\} \leqq E\{\exp[t_0\sum_{k=1}^{n}f(X_k)]| x_0\} \leqq C\gamma^n$ where $\gamma = \{\phi(d_0, t_0)^{1/d_0} + \eta\}$ and η is chosen so that $\gamma < 1$.

For sufficiently small $t > 0$ consider the limited expansion

$$
\begin{aligned}
(25) \qquad E\left\{\exp\left[t\sum_{k=1}^{d}f(X_k)\right]\Big| x_0\right\} &\leqq 1 + tE\left\{d\left(d^{-1}\sum_{k=1}^{d}f(X_k)\right)\Big| x_0\right\} \\
&+ t^2 E\left\{\left[\sum_{k=1}^{d}f(X_k)\right]^2 \exp\left[t\sum_{k=1}^{d}|f(X_k)|\right]\Big| x_0\right\}.
\end{aligned}
$$

It will be shown that the coefficient of dt in the second term of the right-hand side of (25) converges uniformly to μ and the function multiplying t^2 in the third term can be dominated independently of x_0 for sufficiently small positive t. Thus d_0 and t_0 can be chosen so that $\phi(d_0, t_0) < 1$.

First note that three functions $g_0(t)$, $g_1(t)$ and $g_2(t)$ can be found such that

$$
\begin{aligned}
(26) \qquad E\{\exp[t|f(X_1)|]| x_0\} &\leqq g_0(t) = K_0(1 - \rho_1 e^{t\epsilon})^{-1}, \\
E\{f(X_1)\exp[t|f(X_1)|]| x_0\} &\leqq g_1(t) = K_1(1 - \rho_1 e^{t\epsilon})^{-2}, \\
E\{f^2(X_1)\exp[t|f(X_1)|]| x_0\} &\leqq g_2(t) = K_2(1 - \rho_1 e^{t\epsilon})^{-3},
\end{aligned}
$$

for all sufficiently small positive t and suitably chosen constants K_0, K_1 and K_2. The existence of the g's follows from the hypothesis of the theorem upon expanding the exponentials on the right-hand side in power series about zero and summing. These functions are to be used in an iterative evaluation of

$$
\begin{aligned}
(27) \qquad &E\left\{\left[\sum_{k=1}^{d}f(X_k)\right]^2 \exp\left[t\sum_{k=1}^{d}|f(X_k)|\right]\Big| x_0\right\} \\
&= \int e^{t|f(x_1)|}p(x_0, dx_1)\int e^{t|f(x_2)|}p(x_1, dx_2)\cdots\int e^{t|f(x_d)|} \\
&\qquad\cdot\left\{\sum_{k=1}^{d}f(x_k)\right\}^2 p(x_{d-1}, dx_d).
\end{aligned}
$$

Introducing the notation (for fixed d)

$$I_d(x_{d-1}) = \int e^{t|f(x_d)|}\left\{f^2(x_d) + 2f(x_d)\sum_{k=1}^{d-1}f(x_k) + \left[\sum_{k=1}^{d-1}f(x_k)\right]^2\right\}p(x_{d-1}, dx_d)$$

and $y_j = \sum_{k=1}^{j}f(x_k)$ for $j = 1, \cdots, d-1$ we obtain

(28) $$I_d(x_{d-1}) \leqq g_2(t) + 2g_1(t)y_{d-1} + g_0(t)y_{d-1}^2.$$

Define $I_{d-k}(x_{d-k-1}) = \int e^{t|f(x_{d-k})|} I_{d-k+1}(x_{d-k}) p(x_{d-k-1}, dx_{d-k})$ for $k=1, \cdots, d-1$, then from (28)

$$
\begin{aligned}
(29) \quad I_{d-1}(x_{d-2}) &\leq g_2(t)g_0(t) + 2g_1(t) \int e^{t|f(x_{d-1})|} [f(x_{d-1}) + y_{d-2}] p(x_{d-2}, dx_{d-1}) \\
&\quad + g_0(t) \int e^{t|f(x_{d-1})|} [f^2(x_{d-1}) + 2f(x_{d-1}) y_{d-2} + y_{d-2}^2] p(x_{d-2}, dx_{d-1}) \\
&\leq g_0(t)g_2(t) + 2g_1^2(t) + 2g_1(t)g_0(t)y_{d-2} + g_0(t)g_2(t) \\
&\quad + 2g_0(t)g_1(t)y_{d-2} + g_0^2(t)y_{d-2}^2 \\
&= 2g_0(t)g_2(t) + 2g_1^2(t) + 4g_0(t)g_1(t)y_{d-2} + g_0^2(t)y_{d-2}^2.
\end{aligned}
$$

Proceeding inductively we obtain

$$
\begin{aligned}
(30) \quad I_2(x_1) &\leq (d-1)g_0^{d-2}(t)g_2(t) + (d-2)(d-1)g_0^{d-3}(t)g_1^2(t) + g_0^{d-1}(t)y_1^2 \\
&\quad + 2(d-1)g_0^{d-2}(t)g_1(t)y_1
\end{aligned}
$$

and finally

$$
\begin{aligned}
(31) \quad E\left\{ \left[\sum_{k=1}^d f(X_k) \right]^2 \exp\left[t \sum_{k=1}^d |f(X_k)| \right] \Big| x_0 \right\} &\leq \int e^{t|f(x_1)|} I_2(x_1) p(x_0, dx_1) \\
&\leq dg_0^{d-1}(t)g_2(t) + d(d-1)g_0^{d-2}(t)g_1^2(t).
\end{aligned}
$$

It remains to prove that $E\{n^{-1}\sum_{k=1}^n f(X_k)|x_0\} \to \mu$ uniformly in x_0. Observe that

$$
\begin{aligned}
(32) \quad E\left\{ n^{-1} \sum_{k=1}^n f(X_k) \Big| x_0 \right\} &= n^{-1} \sum_{k=1}^n E\{f(X_k) \mid x_0\} \\
&= n^{-1} \sum_{k=1}^n \int f(y) \int p(x, dy) p^{(k-1)}(x_0, dx) \\
&= n^{-1} \sum_{k=1}^n \int \left[\int f(y) p(x, dy) \right] p^{(k-1)}(x_0, dx) \\
&= \int \left[\int f(y) p(x, dy) \right] n^{-1} \sum_{k=1}^n p^{(k-1)}(x_0, dx)
\end{aligned}
$$

where all integrals are taken over $\mathfrak{X}$ and $p^{(0)}(x_0, dx)$ assigns probability one to the point x_0. The interchange of the order of integration follows from an extension of the Fubini theorem which is given as problem 13 [9, p. 140] and whose proof is like that of the Fubini theorem. From Doeblin's condition it follows that $n^{-1}\sum_{k=1}^n p^{(k-1)}(x, B) \to \pi(B)$ uniformly in x and B, where $\pi(\cdot)$ denotes the unique stationary initial measure. From (26) we have that $|\int f(y) p(x, dy)| \leq g_0(0) + g_2(0)$ and thus it follows that

$$(33) \qquad E\left\{n^{-1}\sum_{k=1}^{n}f(X_k)\mid x_0\right\} \to \int\left[\int f(y)p(x,\,dy)\right]\pi(dx)$$

uniformly in x_0. However

$$\int\left[\int f(y)p(x,\,dy)\right]\pi(dx) = \int f(y)\int p(x,\,dy)\pi(dx)$$

$$(34)$$

$$= \int f(y)\pi(dy) = \mu < 0$$

and consequently we may choose an integer d_0 such that

$$E\left\{d_0^{-1}\sum_{k=1}^{d_0}f(X_k)\mid x_0\right\} \leqq \delta < 0 \qquad\qquad \text{for all } x_0.$$

4. Existence of moment generating functions not necessary. For nonstationary processes the necessity theorems of §§2 and 3 for independent and Markov processes do not extend to other familiar classes of stochastic sequences. Specifically we construct a martingale difference sequence of square summable (thus orthogonal) functions which is also a Markov process (with nonstationary transition probabilities) which converges exponentially rapidly although none of the moment generating functions exist on any interval.

Let $X_1 = 0,\ 2^n,\ -2^n$ on sets $E_0,\ E_n^+,\ E_n^-$ having measures $13/15,\ 2^{-4n},\ 2^{-4n}$ respectively. Thus $EX_1^2 = 2\sum_{n=1}^{\infty}2^{2n}2^{-4n} = 2/3$. Next split each X_1 atom into two sets, e.g. $E_n^{++},\ E_n^{+-}(E_n^+ = E_n^{++}\cup E_n^{+-})$ of equal measure. Define $X_2 = 0,\ 2^n,\ -2^n,\ 2^n,\ -2^n$ on the sets $E_0,\ E_n^{++},\ E_n^{+-},\ E_n^{-+},\ E_n^{--}$ with measures $13/15$ for E_0 and 2^{-4n-1} for each of the others. Thus $EX_2^2 = EX_1^2$ and $E(X_2\mid X_1) = 0$. Let each X_2 atom where $X_1 + X_2 = 2^{n+1}$ or -2^{n+1} be halved and let $X_3 = 2^{n+1}$ and -2^{n+1} on each such pair (each member having measure 2^{-4n-2}); and on the set where $X_1 + X_2 = 0$ (of measure $14/15$) let $X_3 = 0$. Thus $EX_3^2 = 4\sum_1^{\infty}2^{2n+2}2^{-4n-2} = 4/3$ and $E[X_3\mid X_2, X_1] = E(X_3\mid X_2) = 0$. In general define, for $j \geqq 2$,

$$X_j = \begin{cases} \pm\ 2^{j-2}2^n \text{ on the halves of the set where} \\ \qquad \sum_{i=1}^{j-1} X_i = 2^{j-2}2^n \text{ each half having measure } 2^{-4n-j+1}, \\ \pm\ 2^{j-2}2^n \text{ on the halves of the set where} \\ \qquad \sum_{i=1}^{j-1} X_i = -\ 2^{j-2}2^n \text{ each half having measure } 2^{-4n-j+1}, \\ 0 \text{ on the set where } \sum_{i=1}^{j-1} X_i = 0 \text{ having measure } 1 - \dfrac{2}{15}2^{-j+2}. \end{cases}$$

Thus $EX_j^2 = 4 \sum_{n=1}^{\infty} 2^{2(j-2)+2n} 2^{-4n-j+1} = 2^{j-1}/3$ $(j = 2, 3, \cdots)$.

On the subset where $X_{j-1} \neq 0$ the σ-field induced by $X_{j-1}, X_{j-2}, \cdots, X_1$ is the σ-field induced by X_{j-1} itself. Since $X_{j-1} = 0$ implies $X_j = 0$, $\{X_k\}$ is Markov and a martingale difference sequence: $E(X_j | X_{j-1}, \cdots, X_1) = E(X_j | X_{j-1}) = 0$. Since $P\{\sum_{k=1}^{j} X_k \neq 0\} = (4/15) \, 2^{-j}$ the sequence has an exponential convergence rate although for any j and $t \neq 0$

$$Ee^{tX_j} \geq 4 \sum_{n=1}^{\infty} \exp(|t| \, 2^{j-2+n}) 2^{-4n-j+1} = \infty$$

since the nth term becomes infinite. In the reverse direction the example of §1 shows that size limitations alone on the random variables do not insure an exponential rate of convergence. It is known however that uniformly bounded sequences will converge exponentially rapidly if they are martingale differences [2]. The sufficiency results of §2 suggest the question of exactly what weakening of the uniform boundedness condition suffices to insure an exponential rate of convergence.

5. **Convergence to $+\infty$.**

DEFINITION. The sequence $\{X_k : k = 1, 2, \cdots\}$ converges exponentially rapidly to $+\infty$ if for all $M > 0$ there exist constants A and $\rho < 1$ (depending on M) such that

$$P\left\{n^{-1} \sum_{k=1}^{n} X_k < M\right\} \leq A\rho^n \qquad n = 1, 2, \cdots.$$

PROPOSITION. *Let $\{X_k : k = 1, 2, \cdots\}$ be a sequence of independent and identically distributed random variables such that $EX_1^+ = +\infty$ and the moment generating function of X_1^- exists on some interval about 0. Then $\{X_k\}$ converges to $+\infty$ exponentially rapidly.*

Proof. It follows from the hypothesis that

$$P\left\{n^{-1} \sum_{k=1}^{n} X_k^- > E(X_1^-) + 1\right\} \leq \prod_{1}^{n} Ee^{t(X_k^- - E(X_1^-) + 1)}$$

$$\leq A_- \rho_-^n$$

if $t > 0$ is chosen sufficiently close to 0. Given any constant Q, it follows from $EX_1^+ = +\infty$ that

$$P\left\{n^{-1} \sum_{k=1}^{n} X_k^+ < Q\right\} \leq \prod_{1}^{n} Ee^{t(Q - X_k^+)} \leq A_+ \rho_+^n$$

for $t > 0$ sufficiently close to 0. Then given M,

$$P\left\{\sum_{k=1}^{n} X_k < nM\right\} \leqq P\left\{\sum_{1}^{n} X_k^+ < n(M + E(X_1^-) + 1)\right\}$$

$$+ P\left\{\sum_{1}^{n} X_k^- > n(E(X_1^-) + 1)\right\}$$

$$\leqq A_+\rho_+^n + A_-\rho_-^n \leqq B\rho^n.$$

However, exponential convergence to $+\infty$ is not as well behaved as exponential convergence to a finite constant in that a theorem in the converse direction for stationary processes (the analog of Theorem 1) is not available as is demonstrated by the following counterexample.

Specifically we define an ergodic transformation T and a measurable function f so that the stationary sequence $\{f \cdot T^{i-1}: i = 1, 2, \cdots\}$ converges exponentially fast to $+\infty$ although $Ef^- = +\infty$.

Let T be the transformation on $\{[0, 1), \mathfrak{B}, P \equiv \text{Leb. measure}\}$ that maps $[(2^n-2)/2^n, (2^n-1)/2^n)$ linearly by translation onto $[1/2^n, 2/2^n)$ for $n = 1, 2, \cdots$. Then T is a one to one, bi-measurable, measure preserving transformation on $[0, 1)$ and is also ergodic. (In fact if E is an invariant set of positive measure, then for any $\epsilon > 0$ there exists an interval (a, b) such that $P\{E\cap(a, b)\} > (1-\epsilon)P\{(a, b)\}$, and hence there is a diadic interval $(i/2^n, j/2^n)$ such that $P\{E\cap(i/2^n, j/2^n)\} > (1-2\epsilon)P\{(i/2^n, j/2^n)\}$. Then necessarily for some k, $i \leqq k \leqq j$, $P\{E\cap(k/2^n, (k+1)/2^n)\} \geqq (1-2\epsilon)/2^n$. But the transformation T is such that the sets $T^i(k/2^n, (k+1)/2^n)$, $i = 0, 1, \cdots$, 2^n-1, are disjoint and of equal measure so that the invariance of E implies $P\{E\} > 2^n(1-2\epsilon)/2^n = 1-2\epsilon$. Thus $P\{E\} = 1$ and T is ergodic).

Next we define a function f on $\{[0, 1), \mathfrak{B}, P\}$

$$f(x) = \begin{cases} 2^{2n} \cdot 2^{2^n} & x \in [(2^n - 2)/2^n, (2^n - 1)/2^n), & n = 2, 3, \cdots \\ -2^{n-1} & x \in [1/2^n, 2/2^n), & n = 2, 3, \cdots \end{cases}$$

and observe that the transformation T is such that for all $x \in [0, 1)$ one of the images $x, Tx, \cdots, T^{(2^{n-1}-1)}x$ is included in $[(2^{n-1}-1)/2^{n-1}, 1)$. Hence at least one of the numbers $f(T^jx)$, $j = 0, 1, \cdots, 2^{n-1}-1$, is greater than or equal to $2^{2n} \cdot 2^{2^n}$. Therefore for $2^{n-1} \leqq m < 2^n$, we have that $\sum_{i=0}^{m-1} f(T^ix) \geqq 2^{2n} \cdot 2^{2^n} - (2^n-1)(2^n 2^{2^n}) = 2^n 2^{2^n}$, which is eventually greater than $2^n M$ for any given M, unless one of the summands $f(T^ix) < -2^n \cdot 2^{2^n}$, that is unless $T^ix \in [0, 1/2^n \cdot 2^{2^n})$ for some i, $i = 0, \cdots, m-1$. But $P\{T^ix \in [0, 1/2^n \cdot 2^{2^n})$ for some $i = 0, 1, \cdots, 2^n-1\} \leqq 2^n/2^n \cdot 2^{2^n}$ and consequently $P\{\sum_{i=0}^{m-1} f(T^ix) \leqq Mm\} \leqq (1/2)^m$ for all sufficiently large m. Thus we have an ergodic stationary sequence $\{f \cdot T^i: i = 0, 1, 2, \cdots\}$ converging exponentially fast to $+\infty$ although $Ef^- = \sum_{n=0}^{\infty} 2^{n-1} 2^{-n} = \infty$.

References

1. David Blackwell and J. L. Hodges, *The probability in the extreme tail of a convolution*, Ann. Math. Statist. **30** (1959), 1113–1120.

2. David Blackwell, *On optimal systems*, Ann. Math. Statist. **25** (1954), 394–397.

3. H. Cramér, *Sur un nouveau théorème-limite de la théorie des probabilités*, Actualités Sci. Ind., No. 736, Hermann, Paris, 1938.

4. J. L. Doob, *Stochastic processes*, Wiley, New York, 1953.

5. W. Feller, *Generalization of a probability limit theorem of Cramér*, Trans. Amer. Math. Soc. **54** (1943), 361–372.

6. P. R. Halmos, *Lectures on ergodic theory*, The Math. Soc. of Japan, Tokyo, 1956.

7. ———, *Measure theory*, Van Nostrand, New York, 1950.

8. Melvin Katz and A. J. Thomasian, *A bound for the law of large numbers for discrete Markov processes*, Ann. Math. Statist. **32** (1961), 336–337.

9. M. Loève, *Probability theory*, Van Nostrand, New York, 1960.

UNIVERSITY OF CHICAGO,
    CHICAGO, ILLINOIS