CONVERGENCE RATES FOR THE LAW OF LARGE NUMBERS FOR LINEAR COMBINATIONS OF MARKOV PROCESSES

By L. H. Koopmans

University of New Mexico

1. Introduction and summary. Let $\{X_k : k = 0, 1, 2, \cdots\}$ be a Markov process with state space $(\mathcal{S}, \mathcal{B})$ and denote by f a real-valued $\mathcal{B}$-measurable function on $\mathcal{S}$.

In [1] and [6] it was shown that for a Markov process with a single ergodic class, if the transition probabilities $P(x, A) = P[X_1 \varepsilon A \mid X_0 = x]$ are stationary and satisfy Doeblin's condition ([2]), if the conditional moment generating function of f, $\int_{\mathcal{S}} e^{tf(y)} P(x, dy)$, satisfies a certain boundedness condition in t uniformly in x and if $\int_{\mathcal{S}} f(y)\pi(dy) = 0$ where π is the unique stationary probability measure for the process, then an exponential bound exists for the convergence of $S_n = n^{-1}[f(X_1) + \cdots + f(X_n)]$ to zero which is independent of the initial probability measure. That is, for every $\epsilon > 0$ there exist positive constants A and $\rho < 1$ such that for all initial measures ν,

$$P_\nu[|S_n| \geqq \epsilon] \leqq A\rho^n \quad \text{for} \quad n = 1, 2, \cdots.$$

The purpose of this paper is to study the extent to which such a result holds when the Cesaro averages $n^{-1}[f(X_1) + \cdots + f(X_n)]$ are replaced by a more general class of averages $\sum_{k=1}^{\infty} a_{n,k} f(X_k)$, $n = 1, 2, \cdots$, where the infinite matrices $\{a_{n,k}\}$ satisfy the conditions

(a) $\sum_{k=1}^{\infty} |a_{n,k}| \leqq 1$ uniformly for $n = 1, 2, \cdots$;

(b) $\lambda(n) = \max_k |a_{n,k}| \to 0$ as $n \to \infty$;

(c) $\lim_{n\to\infty} \sum_{k=1}^{\infty} a_{n,k} = 1$.

Such matrices are called Toeplitz matrices (except that our Condition (b) is somewhat stronger than the one usually imposed, see e.g. [8]), and our concern will be with the rate at which the sequence of random variables $S_n = \sum_{k=1}^{\infty} a_{n,k} f(X_k)$ tends to zero in probability as $n \to \infty$.

We will obtain theorems analogous to those of [1] and [6] through a series of specializations of two basic theorems to be proved in Section 2. There, without the assumption of stationary transition functions, conditions are given under which for every $\epsilon > 0$, there exist positive constants A and $\rho < 1$ such that for all initial measures ν,

$$P_\nu[|S_n| \geqq \epsilon] \leqq A\rho^{1/\lambda(n)}, \qquad n = 1, 2, \cdots. \tag{0}$$

Here S_n is the Toeplitz average and $\lambda(n)$ is defined above in Condition (b).

In Section 3, processes with stationary transition probabilities are studied and a condition on the conditional moment generating function of f is introduced which is comparable to a condition introduced in [1] and which implies the first

Received 12 August 1965.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

theorem of Section 2. In Section 4, Doeblin's condition is introduced and additional restrictions on the process and the Toeplitz matrices sufficient to yield (0) are considered.

This study is a sequel to work begun in [3] and continued in [4] and [5]. For applications in which Toeplitz averages of stochastic processes are of interest and for further orientation on the subject of convergence rates for Toeplitz averages of random variables, the reader is referred to these papers.

2. General convergence rate theorems. The following notation will be used throughout the paper: If h is a real-valued, $\mathcal{B}$-measurable function of the random variables $X_1, X_2, \cdots$ then $E_\nu(h)$ denotes the expected value of h with respect to the distribution of the process started with the initial measure ν. E_x is used in place of E_{ν_x}, where ν_x is the initial measure assigning probability 1 to $\{x\}$. Similarly, P_ν will denote the probability distribution of the process relative to ν.

Let $\mu_x(k) = E_x f(X_k)$.

Assumption 1. There exists a positive constant $K < \infty$ such that for every $x \varepsilon \mathcal{S}$ and $k = 1, 2, \cdots$; $|\mu_x(k)| \leqq K$. Let $\gamma_x(n) = \sum_{k=1}^\infty a_{n,k}\mu_x(k)$, where $\{a_{n,k}\}$ is a Toeplitz matrix which satisfies Conditions (a), (b) and (c). It follows easily from Assumption 1 that, for each n, the series defining $\gamma_x(n)$ is absolutely summable uniformly in x and that $|\gamma_x(n)| \leqq K$ for all $x \varepsilon \mathcal{S}$ and $n = 1, 2, \cdots$.

Assumption 2. For every $\beta > 0$ there exists $T_\beta > 0$ such that if $|t| \leqq T_\beta$, then

$$E_x \exp [t \sum_{k=1}^N b_k f(X_k)] \leqq \exp [\{t \sum_{k=1}^N b_k \mu_x(k) + \beta |t| \sum_{k=1}^N |b_k|\}],$$

uniformly in $x \varepsilon \mathcal{S}$ for all integers N and all sequences $\{b_k\}$ of real numbers for which $\sum_{k=1}^\infty |b_k| \leqq 1$.

Theorem 1. *Let $\{X_k : k = 0, 1, \cdots\}$ be a Markov process (not necessarily with stationary transition probabilities) and let f be a real-valued $\mathcal{B}$-measurable function on $\mathcal{S}$ for which Assumptions* 1 *and* 2 *hold. Let*

$$S_n = \sum_{k=1}^\infty a_{n,k} f(X_k).$$

Then S_n exists as a limit-in-the-mean of order 2 *and for every $\epsilon > 0$, $\beta > 0$ and every initial measure ν,*

$$P_\nu[|S_n| \geqq \epsilon] \leqq 2 \exp [(\beta - \epsilon)|t|/\lambda(n)] \int_{\mathcal{S}} \exp [|t| \, |\gamma_x(n)|/\lambda(n)] \, d\nu,$$

for $|t| \leqq T_\beta$.

Proof. The index n will be deleted in the proof. Let $s_N = \sum_{k=1}^N a_k f(X_k)$ and let $M < N$. Then, for fixed $\beta > 0$, Assumptions 1 and 2 yield

$$E_x \exp [t(s_N - s_M)] \leqq \exp [|t|(K + \beta) \sum_{k=M+1}^N |a_k|], \qquad \text{uniformly in } x.$$

By the lemma of [5], there exists a constant γ_β (independent of x) such that $E_x(s_N - s_M)^2 \leqq \gamma_\beta \sum_{k=M+1}^N |a_k|$. The right hand side tends to zero as $M, N \to \infty$, thus for every initial measure ν, $E_\nu(s_N - s_M)^2 = \int_{\mathcal{S}} E_x(s_N - s_M)^2 \, d\nu \to 0$. Thus, $s = \sum_{k=1}^\infty a_k f(X_k)$ is determined as a limit-in-the-mean of its partial sums.

Now, for $0 < t < T_\beta$, a well known inequality ([7], p. 127) yields

$$P_\nu[\pm s_N \geqq \epsilon] = P_\nu[(\pm s_N - \epsilon)/\lambda \geqq 0] \leqq e^{-\epsilon t/\lambda} \int_{\mathbb{S}} E_x e^{\pm t s_N/\lambda}\, d\nu$$
$$\leqq e^{-\epsilon t/\lambda} \int_{\mathbb{S}} \exp\,[(\pm t \textstyle\sum_{k=1}^N a_k \mu_x(k) + |t|\, \beta \sum_{k=1}^N |a_k|)/\lambda]\, d\nu.$$
$$\leqq \exp\,[(\beta - \epsilon)|t|/\lambda] \int_{\mathbb{S}} \exp\,[\pm t \textstyle\sum_{k=1}^N a_k \mu_x(k)/\lambda]\, d\nu.$$

By the bounded convergence theorem, the last expression converges (uniformly in ν), yielding in the limit as $N \to \infty$,

$$P_\nu[\pm s \geqq \epsilon] \leqq \exp\,[(\beta - \epsilon)|t|/\lambda] \int_{\mathbb{S}} \exp\,(\pm t \gamma_x/\lambda)\, d\nu$$
$$\leqq \exp[(\beta - \epsilon)|t|/\lambda] \int_{\mathbb{S}} \exp\,[|t|\, |\gamma_x|/\lambda]\, d\nu.$$

The theorem follows from this and the inequality $P[|s| \geqq \epsilon] \leqq P[s \geqq \epsilon] + P[-s \geqq \epsilon]$.

THEOREM 2. *If, in addition to the hypotheses of Theorem* 1, *we have* $\lim_{n\to\infty} \gamma_x(n) = 0$ *uniformly in* x, *then for every* $\epsilon > 0$ *there exist positive numbers* A *and* $\rho < 1$ *such that, uniformly for all initial measures* ν,

$$P_\nu[|S_n| \geqq \epsilon] \leqq A\rho^{1/\lambda(n)} \qquad n = 1, 2, \cdots. \tag{1}$$

PROOF. Select positive numbers β and δ such that $\beta + \delta < \epsilon/2$. Then, there exists N such that $n \geqq N$ implies $|\gamma_x(n)| < \delta$ uniformly in x. Thus, by Theorem 1, for $|t| \leqq T_\beta$,

$$P_\nu[|S_n| \geqq \epsilon] \leqq 2 \exp\,[(\beta - \epsilon + \delta)|t|/\lambda(n)],$$

for all $n \geqq N$ and all initial measures ν. Now, select A large enough to uniformly bound the first $N - 1$ terms (e.g., $A = \rho^{-1/\lambda(N)}$). Then, (1) holds with $\rho = \exp\,(-\epsilon T_\beta/2)$.

If $\gamma_x(n) \to 0$ as $n \to \infty$ but the convergence is not uniform in x, then an inequality comparable to (1) is available in many cases but the constants A and ρ will no longer be independent of ν. For example, if $\mathbb{S}$ is a topological space and ν concentrates its mass on a compact set S, then (1) will hold and the constants A and ρ will depend on ν only through S.

3. The case of stationary transition probabilities. The following results are based on the assumption that the Markov process has a stationary transition probability function $P(x, A) = P[X_k \,\varepsilon\, A \mid X_{k-1} = x]$, $k = 1, 2, \cdots$. We will show that a sufficient condition for Theorem 1 to hold in this case is a uniformity condition on the conditional moment generating function of $|f(X_1)|$ (given X_0) only slightly stronger than the one introduced in [1] for Cesaro averages.

ASSUMPTION S. There exist positive numbers T and γ, independent of x, such that for $|t| \leqq T$ and all $x \,\varepsilon\, \mathbb{S}$,

$$E_x \exp\,[t\, |f(X_1)|] \leqq \exp\,[\gamma\, |t|].$$

THEOREM 3. *If the Markov process* $\{X_k : k = 0, 1, \cdots\}$ *has stationary transition* *probabilities and Assumption* S *is satisfied for the function f, then the conclusions of Theorem* 1 *hold.*

PROOF. It suffices to establish the validity of Assumptions 1 and 2. The equality, $e^{tx} + e^{-tx} = e^{t|x|} + e^{-t|x|}$ implies that for any random variable X, $Ee^{tX} \leqq Ee^{t|X|} + Ee^{-t|X|}$ whenever the right hand side exists. Thus, Assumption S implies that

$$M_x(t) = E_x e^{tf(X_1)} \leqq 2e^{\gamma|t|} \tag{2}$$

uniformly in x for $|t| \leqq T$.

Form the functions of a complex variable $M_x(z) = E_x e^{zf(X_1)}$, $x \varepsilon \mathcal{S}$. Since $|M_x(z)| \leqq M_x(\operatorname{Re} z) < \infty$ for $|\operatorname{Re} z| < T$ by virtue of (2), it follows that $M_x(z)$ is analytic in this region. Select δ such that $0 < \delta < T$ and let C be the circle $\{z = \delta e^{i\theta}: 0 \leqq \theta < 2\pi\}$. By Cauchy's integral formula,

$$\mu_x(1) = (d/dz)M_x(z)\mid_{z=0} = (2\pi i)^{-1} \int_C (M_x(z)/z^2)\, dz.$$

Thus,

$$|\mu_x(1)| \leqq (2\pi)^{-1} \int_0^{2\pi} [|M_x(\delta e^{i\theta})|/\delta^2]\, d\theta \leqq 2e^{\gamma\delta}/\delta^2 = K < \infty. \tag{3}$$

Let $E^{(X_{i_1}, \cdots, X_{i_r})}$ denote conditional expectation relative to the sigma-algebra generated by $X_{i_1}, \cdots, X_{i_r}$. Then, for $k > 1$,

$$\begin{aligned} \mu_x(k) &= \int E^{(X_0)} f(X_k)\, d\nu_x \\ &= \int E^{(X_0)} E^{(X_0, \cdots, X_{k-1})} f(X_k)\, d\nu_x \\ &= \int E^{(X_0)} E^{(X_{k-1})} f(X_k)\, d\nu_x \qquad \text{by the Markov property.} \end{aligned}$$

Thus, by the stationarity of the transition probabilities and (3), $|\mu_x(k)| \leqq \int E^{(X_0)}(K)\, d\nu_x = K$. This establishes Assumption 1.

To prove that Assumption 2 is satisfied, we dominate a limited power series expansion of $E_x \exp [t \sum_{k=1}^N b_k f(X_k)]$:

$$\begin{aligned} E_x \exp [t \textstyle\sum_{k=1}^N b_k f(X_k)] &\leqq 1 + t \textstyle\sum_{k=1}^N b_k \mu_x(k) \\ &\quad + t^2 E_x(\textstyle\sum_{k=1}^N |b_k|\, |f(X_k)|)^2 \exp [|t| \textstyle\sum_{k=1}^N |b_k|\, |f(X_k)|]. \end{aligned} \tag{4}$$

The last term is dominated as follows: Let $C_k = |b_k|/\sum_{k=1}^N |b_k|$ and $R_N = \sum_{k=1}^N C_k\, |f(X_k)|$. Since for every $\delta > 0$, $y^2 e^{-\delta y} \to 0$ as $y \to \infty$, it follows that there exists $A_\delta > 0$ such that $y \geqq A_\delta$ implies $y^2 e^{|t|y} \leqq e^{(|t|+\delta)y}$ for all t. Fix $\delta = T/4$ and $|t| \leqq T/2$ (so that $0 < |t| + \delta < T$). Now, if I_B is the set characteristic function of the event $B = [R_N \leqq A_\delta]$, then the inequality $\sum_{k=1}^N |b_k| \leqq 1$ implies that

$$\begin{aligned} &E_x(\textstyle\sum_{k=1}^N |b_k|\, |f(X_k)|)^2 \exp [|t| \textstyle\sum_{k=1}^N |b_k|\, |f(X_k)|] \\ &\qquad \leqq (\textstyle\sum_{k=1}^N |b_k|)^2 E_x R_N^2 e^{|t|R_n} \\ &\qquad = (\textstyle\sum_{k=1}^N |b_k|)^2 E_x (I_B + I_{B^c}) R_N^2 e^{|t|R_N} \\ &\qquad \leqq (\textstyle\sum_{k=1}^N |b_k|)\{A_\delta^2 e^{|t|A_\delta} + E_x e^{(|t|+\delta)R_N}\}. \end{aligned}$$

But,

$$(5)\quad E_x e^{(|t|+\delta)R_N} = \int \cdots \int \prod_{k=1}^N \exp[(|t| + \delta)C_k |f(x_k)|]P(x_{k-1}, dx_k), \quad x_0 \equiv x,$$
$$\leqq \prod_{k=1}^N e^{(|t|+\delta)C_k\gamma} = e^{(|t|+\delta)\gamma}.$$

This last inequality is obtained by taking the suprema of the first arguments of the transition probabilities under the integral signs in the second expression of (5). Then, since $E_x \exp(\alpha|f(X_1)|) = \int \exp(\alpha|f(y)|)P(x, dy)$, it follows by Assumption S that $\sup_x \int \exp(\alpha|f(y)|)P(x, dy) \leqq \exp(\alpha\gamma)$ for $0 < \alpha < T$.

Substituting these expressions into (4), we have

$$E_x \exp[t \textstyle\sum_{k=1}^N b_k f(X_k)] \leqq 1 + t \sum_{k=1}^N b_k \mu_x(k) + |t| \sum_{k=1}^N |b_k|\{|t|\, G\}$$
$$\leqq \exp[t \textstyle\sum_{k=1}^N b_k \mu_x(k) + |t| \sum_{k=1}^N |b_k|\{|t|\, G\}],$$

where $G = A_\delta^2 \exp[(T/2)A_\delta] + \exp[(T/2 + \delta)\gamma]$. (Recall $|t| \leqq T/2$.)

Now, if $\beta > 0$, Assumption 2 is satisfied by selecting $T_\beta = \min\{T/2, \beta/G\}$.

Corollary 1. *If the Markov process has stationary transition probabilities and the function f is bounded, then the conclusions of Theorem 1 are valid.*

Proof. Assumption S is clearly true for bounded functions.

The theorem for Cesaro averages stated in the introduction was first proved in [6] under the assumption that f is bounded.

In order to obtain the conclusions of Theorem 2 in the present context it is only necessary that $\gamma_x(n) \to 0$ uniformly in x as $n \to \infty$. This is not, however, a consequence of Assumption S. To see this, consider the normal Markov process with $\mathcal{S}$ = Reals and k-step (stationary) transition probability function

$$P^{(k)}(x, A) = [2\pi(1 - \rho^{2k})]^{-\frac{1}{2}} \int_A \exp -[(y - \rho^k x)^2/2(1 - \rho^{2k})]\, dy, \quad 0 < \rho < 1,$$

and let $f(y) = \arctan y$. The conclusions of Theorem 1 are valid by virtue of Corollary 1. Also, by a change of variables,

$$(6)\qquad \mu_x(k) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \arctan[(1 - \rho^{2k})^{\frac{1}{2}} y + \rho^k x] e^{-y^2/2}\, dy$$

from which it is easily seen that $\mu_x(k) \to 0$. However, it is also seen from (6) that for every $\epsilon > 0$ and integer K there exists an $x > 0$ such that $\pi/2 - \epsilon < \mu_x(k) < \pi/2$ for $1 \leqq k \leqq K$. Now, consider any Toeplitz matrix $\{a_{n,k}\}$ of non-negative terms and assume, for convenience, that $\sum_{k=1}^\infty a_{n,k} = 1$ for all n. Then, $\gamma_x(n) = \sum_{k=1}^\infty a_{n,k}\mu_x(k) \to 0$ as $n \to \infty$. But, if K_n is an integer such that $\sum_{k=1}^{K_n} a_{n,k} > 1 - 2\epsilon/\pi$, by selecting this to be the above defined integer K, there is an $x > 0$ such that $|\gamma_x(n) - \sum_{k=1}^{K_n} a_{n,k}\mu_x(k)| < \epsilon$ and $\pi/2 - 2\epsilon \leqq \sum_{k=1}^{K_n} a_{n,k}\mu_x(k) \leqq \pi/2 - \epsilon$. It follows that for every n, there exists an $x > 0$ such that $\gamma_x(n) > \pi/2 - 3\epsilon$.

In the next section we will impose a further restriction on the process that will ensure the uniformity of convergence of $\gamma_x(n)$.

4. The case of stationary transition probabilities which satisfy Doeblin's condition. As in the last section, the transition probabilities of the process are assumed to be stationary. We will impose Doeblin's condition (Hypothesis (D) pg. 192 [2]) on the transition probabilities. We list here only the definitions and results explicitly needed in our development. For a complete account of the pertinent theory, the reader is referred to [2]. In the notation of [2], ${}_aC_\alpha$ will denote the αth cyclic subclass of the ath ergodic class E_a, $\alpha = 1, 2, \cdots, d_a$; $a = 1, 2, \cdots, A$.

$\rho(x, {}_aC_\alpha)$ is the conditional probability that the process, initially at x, will finally be in ${}_aC_\alpha$. It is true that for all $x \varepsilon \mathcal{S}$,

$$\sum_{a,\alpha} \rho(x, {}_aC_\alpha) = 1. \tag{7}$$

${}_a\Pi_\alpha(E)$ is the stationary probability measure associated with ${}_aC_\alpha$ and is determined by

$$ {}_a\Pi_{\alpha+m}(A) = \lim_{n\to\infty} P^{(nd_a+m)}(x, A) \quad \text{for} \quad x \varepsilon {}_aC_\alpha, \tag{8}$$

where, in this and future expressions, the index $\alpha + m$ is to be reduced (mod d_a). This limit is uniform in x and A.

Finally, if $\{a_{n,k}\}$ is a Toeplitz matrix, we let $\gamma_{m,a} = \lim_{n\to\infty} \sum_{k=0}^{\infty} a_{n,kd_a+m}$.

The basic result of this section is the following theorem:

THEOREM 4. *Let the Markov process* $\{X_k : k = 0, 1, \cdots\}$ *have a stationary transition probability function which satisfies Doeblin's condition, and let the function* f *satisfy Assumption* $\mathcal{S}$. *Then, if* $\{a_{n,k}\}$ *is a Toeplitz matrix of non-negative elements,*

$$\lim_{n\to\infty} \gamma_x(n) = \sum_{a,\alpha} \rho(x, {}_aC_\alpha) \sum_{m=1}^{d_a} \gamma_{m,a} \int f(y)\, {}_a\Pi_{\alpha+m}(dy),$$

uniformly in x.

PROOF. If we let $h(x) = \mu_x(1)$, by Theorem 3, $h(x)$ is bounded on $\mathcal{S}$. Then, by an extension of the Fubini theorem, [7], p. 140, $\mu_x(k)$ can be written in the form

$$\begin{aligned}\mu_x(k) &= \int f(y)P^{(k)}(x, dy)\\ &= \int h(z)P^{(k-1)}(x, dz),\end{aligned}$$

where $P^{(k)}(x, A)$ is the k-step transition function for the Markov process.

The following easily proved lemma will be used in the proof.

LEMMA. *If* $\{P_\gamma^{(n)}(E) \mid n = 1, 2, \cdots; \gamma \varepsilon \Gamma\}$ *is a parameterized sequence of probability measures such that* $P_\gamma^{(n)}(E) \to P_\gamma(E)$ *as* $n \to \infty$ *uniformly for all measurable sets* E *and all* $\gamma \varepsilon \Gamma$ *and if* h *is a bounded, measurable function, then*

$$\int h(x)P_\gamma^{(n)}(dx) \to \int h(x)P_\gamma(dx),$$

uniformly in γ.

Since the $a_{n,k}$'s are taken to be non-negative, $\gamma_{m,a} \geqq 0$. We assume that $\gamma_{m,a} > 0$. Only a trivial modification in the proof is needed if $\gamma_{m,a} = 0$. Then, for $m = 1, 2, \cdots, d_a$, the matrices $\{(\gamma_{m,a})^{-1}a_{n,kd_a+m}\}$ satisfy the conditions

(a)$'$ $\sum_{k=0}^{\infty} (\gamma_{m,a})^{-1}a_{n,kd_a+m} \leqq (\gamma_{m,a})^{-1} < \infty$ for all n;

(b)$'$ $\max_k (\gamma_{m,a})^{-1}a_{n,kd_z+m} \leqq \lambda(n)/\gamma_{m,a} \to 0$ as $n \to \infty$;

(c)$'$ $\lim_{n\to\infty} \sum_{k=0}^{\infty} (\gamma_{m,a})^{-1}a_{n,kd_a+m} = 1$.

I.e., they are again Toeplitz matrices ([8]). Now, by an application of the lemma to the sequence $\{\sum_{k=0}^{N} (\gamma_{m,a})^{-1} a_{n,kd_a+m} P^{(kd_a+m-1)}(x, A) : N = 1, 2, \cdots\}$, we have

$$\sum_{k=0}^{\infty} (\gamma_{m,a})^{-1} a_{n,kd_a+m} \int h(y) P^{(kd_a+m-1)}(x, dy) = \int h(y) Q_{m-1}^{(n)}(x, dy),$$

where $Q_{m-1}^{(n)}(x, A) = \sum_{k=0}^{\infty} (\gamma_{m,a})^{-1} a_{n,kd_a+m} P^{(kd_a+m-1)}(x, A)$.

But, by (8) and the easily established fact that uniform convergence is preserved under Toeplitz summation, it follows that for $x \in {}_aC_\alpha$,

$$Q_{m-1}^{(n)}(x, A) \to {}_a\Pi_{\alpha+m-1}(A) \quad \text{as} \quad n \to \infty,$$

uniformly in x and A.

Now, another application of the lemma yields the uniform (for $x \in {}_aC_\alpha$) limit

$$\lim_{n\to\infty} \int h(y) Q_{m-1}^{(n)}(x, dy) = \int h(y)\, {}_a\Pi_{\alpha+m-1}(dy).$$

Thus, uniformly for $x \in {}_aC_\alpha$,

$$\gamma_x(n) = \sum_{k=1}^{\infty} a_{n,k}\mu_x(k) = \sum_{m=1}^{d_a} \sum_{k=0}^{\infty} a_{n,kd_a+m} \int h(y) P^{(kd_a+m-1)}(x, dy)$$
$$\to \sum_{m=1}^{d_a} \gamma_{m,a} \int h(y)\, {}_a\Pi_{\alpha+m-1}(dy).$$

Multiplying each term of this expression by $\rho(x, {}_aC_\alpha)$ and summing over all a and α, it follows from (7) that uniformly for all $x \in \mathcal{S}$,

$$\gamma_x(n) \to \sum_{a,\alpha} \rho(x, {}_aC_\alpha) \sum_{m=1}^{d_a} \gamma_{m,a} \int h(y)\, {}_a\Pi_{\alpha+m-1}(dy).$$

Now, it is easily shown by another application of the lemma and (8) that

$${}_a\Pi_{\alpha+m}(A) = \int P(y, A)\, {}_a\Pi_{\alpha+m-1}(dy).$$

Thus,

$$\int h(y)\, {}_a\Pi_{\alpha+m-1}(dy) = \int f(z) \int P(y, dz)\, {}_a\Pi_{\alpha+m-1}(dz) = \int f(z)\, {}_a\Pi_{\alpha+m}(dz),$$

and the theorem is proved.

Because of this theorem, to obtain the conclusions of Theorem 2 in the present context it is sufficient to have

$$(9) \qquad \sum_{a,\alpha} \rho(x, {}_aC_\alpha) \sum_{m=1}^{d_a} \gamma_{m,a} \int f(y)\, {}_a\Pi_{\alpha+m}(dy) = 0 \quad \text{for all} \quad x \in \mathcal{S}.$$

An immediate (but not too interesting) sufficient condition for (9) to hold is that $\int f(y)\, {}_a\Pi_\alpha(dy) = 0$ for all a and α.

To obtain conditions comparable to those made in [1] and [6] which guarantee (9), we restrict attention to the class of Markov processes with only one ergodic class. We hereafter drop the index a. Now there is a unique stationary measure for each process defined by

$$(10) \qquad \Pi(E) = d^{-1} \sum_{\alpha=1}^{d} \Pi_\alpha(E).$$

If there are no cyclically moving subclasses, which corresponds to the case $d = 1$ in (10), then the condition $\int f(y)\Pi(dy) = 0$ is clearly sufficient for (9) to hold:

COROLLARY 2. *If, in addition to the conditions of Theorem 4, the Markov process* *has only one ergodic class with no cyclically moving subclasses and if*

$$\int f(y)\Pi(dy) = 0, \tag{11}$$

then the conclusions of Theorem 2 hold.

For Cesaro averages, Condition (11) is sufficient to guarantee the conclusions of Theorem 2 even when cyclically moving subclasses are present. The reason for this is that in the Cesaro case, $\gamma_m = d^{-1}$ for $m = 1, 2, \cdots, d$. Thus, for all $x \varepsilon \mathcal{S}$,

$$\begin{aligned} \textstyle\sum_\alpha \rho(x, C_\alpha) \sum_{m=1}^{d} \gamma_m \int f(y)\Pi_{\alpha+m}(dy) \\ &= \textstyle\sum_\alpha \rho(x, C_\alpha) \int f(y) d^{-1} \sum_{m=1}^{d} \Pi_{\alpha+m}(dy) \\ &= \int f(y)\Pi(dy) = 0, \end{aligned} \tag{12}$$

and (9) is satisfied. For general Toeplitz averages this is no longer the case. It is easily seen that for any processes with cyclically moving subclasses for which (11) holds but $\int f(y)\Pi_\alpha(dy) \neq 0$ for some α, many Toeplitz matrices can be constructed such that $\lim_n \gamma_x(n) \neq 0$ for some $x \varepsilon \mathcal{S}$. In fact, a study of the null spaces of the cyclic matrices with ijth elements $\int f(y)\Pi_{i+j}(dy)$ ($i + j$ reduced (mod d)) for which (11) is satisfied, shows that, in the most general situation, the equal assignment is the only assignment of the γ_m's for which (9) holds for all $x \varepsilon \mathcal{S}$. This motivates the introduction of the following class of Toeplitz matrices:

DEFINITION. A Toeplitz matrix $\{a_{n,k}\}$ with nonnegative elements is said to be *Cesaro Regular* if, for each $d = 1, 2, \cdots$,

$$\lim_{n\to\infty} \textstyle\sum_{k=0}^{\infty} a_{n,kd+m} = d^{-1}, \qquad m = 1, 2, \cdots, d.$$

COROLLARY 3. *If, in addition to the conditions of Theorem 4, the Markov process has only one ergodic class,* (11) *is satisfied, and the Toeplitz matrix is Cesaro Regular, then the conclusions of Theorem 2 hold.*

PROOF. If the Toeplitz matrix is Cesaro Regular, then (8) follows from (11) as in the case of Cesaro averages.

Cesaro regularity is difficult to check in practice and the extent of the class of Cesaro regular Toeplitz matrices is not known. However, this class contains other than the usual Cesaro $(C, 1)$ matrix as is demonstrated by the following construction which yields a subclass based on the Poisson distribution.

Let $H(n)$ be an arbitrary, strictly positive function of n such that $H(n) \to \infty$ as $n \to \infty$. Define

$$a_{n,k} = e^{-H(n)}[H(n)]^k/k!, \qquad k = 0, 1, 2, \cdots.$$

It is easily established that $\{a_{n,k}\}$ satisfies Conditions (a), (b) and (c) of the Introduction. In fact, it is a Toeplitz matrix of positive terms for which $\sum_{k=0}^{\infty} a_{n,k} = 1$ for all n.

Fix an integer $d \geqq 1$. Let $\delta_j = \exp(2\pi i j/d)$, the jth of the d-roots of unity, $j = 0, 1, \cdots, d - 1$, and define the functions

$$f_m(n) = d^{-1} \textstyle\sum_{j=0}^{d-1} \delta_j^{\,m} e^{\delta_j H(n)}, \qquad m = 0, 1, \cdots, d - 1.$$

It is easily seen by expanding $e^{\delta_j z}$ in its MacLaurin series and forming the linear combination defining $f_m(n)$, that

$$f_m(n) = \sum_{k=0}^{\infty} [[H(n)]^{kd+m}/(kd + m)!], \qquad m = 0, 1, \cdots, d - 1.$$

Thus,

$$\sum_{k=0}^{\infty} a_{n,kd+m} = e^{-H(n)} f_m(n) = d^{-1}[1 + \sum_{j=1}^{d-1} \delta_j^{\;m} e^{(\delta_j - 1)H(n)}].$$

But, for all $0 < j \leqq d - 1$, Re $(\delta_j - 1) < 0$. Hence, $|\exp [(\delta_j - 1)H(n)]| \to 0$ as $n \to \infty$, and

$$(13) \qquad \sum_{k=0}^{\infty} a_{n,kd+m} \to d^{-1} \quad \text{as} \quad n \to \infty \quad \text{for} \quad m = 0, 1, \cdots, d - 1.$$

Note that in the special case $d = 2$, $f_0(n) = \cosh H(n)$ and $f_1(n) = \sinh H(n)$ and (13) is easily obtained from the familiar exponential representations and series expansions of these functions.

Acknowledgment. I am indebted to my collaborator D. L. Hanson for introducing me to this area of research and for suggesting the study pursued in this paper.

REFERENCES

[1] Baum, L. E., Katz, Melvin and Read, R. R. (1962). Exponential convergence rates for the law of large numbers. *Trans. Amer. Math. Soc.* **102** 187–199.

[2] Doob, J. L. (1953). *Stochastic Processes*. Wiley, New York.

[3] Hanson, D. L. and Koopmans, L. H. (1965). On the convergence rate of the law of large numbers for linear combinations of independent random variables. *Ann. Math. Statist.* **36** 559–564.

[4] Hanson, D. L. and Koopmans, L. H. (1965). Convergence rates for the law of large numbers for linear combinations of exchangeable and *-mixing stochastic processes. *Ann. Math. Statist.* **36** 1840–1852.

[5] Hanson, D. L. and Koopmans, L. H. (1965). A probability bound for integrals with respect to stochastic processes with independent increments. *Proc. Amer. Math. Soc.* **16** 1173–1177.

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

[7] Loève, M. (1960). *Probability Theory* (2nd ed.). Van Nostrand, New York.

[8] Zygmund, Antoni (1955). *Trigonometric Series*. Dover, New York.