ON THE CONVERGENCE RATE OF THE LAW OF LARGE NUMBERS FOR LINEAR COMBINATIONS OF INDEPENDENT RANDOM VARIABLES

By D. L. Hanson[1, 2] and L. H. Koopmans[1, 3]

University of Missouri and Sandia Laboratory

1. Introduction and summary. The purpose of this paper is to establish the following theorem and several corollaries to it.

Theorem 1. *Let $\{\xi_k : k = 0, \pm 1, \cdots\}$ be an independent sequence of real valued random variables with $E\xi_k = 0$ and moment generating functions $f_k(t) = Ee^{t\xi_k}$ such that:*

(1) *for every $\beta > 0$ there exists $T_\beta > 0$ such that $f_k(t)$ exists and $|1 - f_k(t)| \leqq \beta|t|$ for $|t| \leqq T_\beta$ uniformly in k.*

Let $\{a_{n,k} : k = 0, \pm 1, \cdots ; n = 1, 2, \cdots\}$ be real numbers such that:

(2) $\sum_{k=-\infty}^{\infty} |a_{n,k}| < A < \infty$ *for* $n = 1, 2, \cdots$

(3) $f(n) = \sup_k |a_{n,k}| \to 0$ *as* $n \to \infty$.

Then

$$S_n = \lim_{a\to-\infty, b\to\infty} \sum_{k=a}^{b} a_{n,k}\xi_k$$

is defined as an almost sure limit for all n, and for every $\epsilon > 0$ there exists a positive $\rho_\epsilon < 1$ (depending on A but not on the particular $a_{n,k}$'s) such that

$$P[|S_n| \geqq \epsilon] \leqq 2\rho_\epsilon^{1/f(n)}. \tag{4}$$

This theorem is applied in Section 3 to establish exponential convergence rates for the strong law of large numbers for subsequences of linear processes of non-identically distributed random variables. In Section 4, the application of the theorem to the summability theory of sequences of independent random variables is discussed. Section 2 is devoted to proving the theorem.

2. Proof of Theorem 1. We first establish the following lemma.

Lemma. *Under Conditions* 1, 2, *and* 3 *of Theorem* 1, *there exists* $0 < T \leqq 1$ *independent of n such that*

$$g_n(t) = \lim_{a\to-\infty, b\to\infty} \prod_{k=a}^{b} f_k(a_{n,k}t)$$

exists for all n and $|t| < T$. Moreover

$$S_n = \lim_{a\to-\infty, b\to\infty} \sum_{k=a}^{b} a_{n,k}\xi_k$$

exists as an almost sure limit for all n and possesses the moment generating function $g_n(t)$.

Received 16 June 1964; revised 15 November 1964.

[1] Research partially supported by the United States Atomic Energy Commission.

[2] Research partially supported by the Air Force Office of Scientific Research.

[3] Now at the University of New Mexico.

PROOF OF LEMMA. Pick β and let T from the lemma statement equal $\min\{1, T_\beta/A, T_\beta\}$. Define $\sigma = \{z = t + is : |t| < T\}$. Since for $|t| < T_\beta$,

$$|Ee^{z\xi_k}| \leqq Ee^{t\xi_k} \leqq 1 + \beta|t| \leqq e^{\beta|t|},$$

we see that the functions

$$E \exp\{z \textstyle\sum_{k=a}^{b} a_{n,k}\xi_k\} = \prod_{k=a}^{b} f_k(a_{n,k}z)$$

are uniformly bounded in σ for all n, a, and b by

$$\exp(\beta \textstyle\sum_{k=-\infty}^{\infty} |a_{n,k}|T) \leqq e^{\beta T_\beta}.$$

It follows from the theory of the bilateral Laplace-Stieltjes transform that they are analytic in σ.

Suppose $a \leqq -N$ and $N < b$. Moment generating functions are convex and since $f_k'(0) = E\xi_k = 0$ we see that $1 \leqq f_k(t)$ for all t. It follows that for $|t| < T$,

$$\begin{aligned} |\textstyle\prod_{k=a}^{b} f_k(a_{n,k}t) - \prod_{k=-N}^{N} f_k(a_{n,k}t)| &= [\textstyle\prod_{k=a,|k|>N}^{b} f_k(a_{n,k}t) - 1]\prod_{k=-N}^{N} f_k(a_{n,k}t) \\ &\leqq [\exp\{\beta \textstyle\sum_{k=a,|k|>N}^{b} |a_{n,k}|T\} - 1]e^{\beta T_\beta} \\ &\leqq [\exp\{\beta T \textstyle\sum_{|k|>N} |a_{n,k}|\} - 1]e^{\beta T_\beta} \end{aligned}$$

which converges to zero as $N \to \infty$. It follows that $\{\prod_{k=a}^{b} f_k(a_{n,k}t)\}$ is a Cauchy sequence for $|t| < T$ and has a limit $g_n(t)$ in this range.

By Vitali's theorem ([4], p. 168), as $a \to -\infty$ and $b \to \infty$, $\prod_{k=a}^{b} f_k(a_{n,k}z) \to g_n(z)$ uniformly in every region bounded by a contour in σ and $g_n(z)$ is analytic in σ so is continuous at 0. By Corollary 1 of [3], p. 251, this is sufficient to insure that

$$\textstyle\sum_{k=a}^{b} a_{n,k}\xi_k \to S_n$$

almost surely as $a \to -\infty$ and $b \to \infty$, and that S_n has characteristic function $g_n(is)$. Comparing coefficients in the expansions of $g_n(is)$ and $g_n(t)$ we see that $g_n(t)$ is the moment generating function of S_n.

We are now in a position to finish proving Theorem 1. Fix $\epsilon > 0$. By a fundamental inequality (see e.g. [3], p. 157), for $f(n) > 0$ and $t \geqq 0$,

$$\begin{aligned} P[\pm S_n \geqq \epsilon] = P\left[\frac{\pm S_n - \epsilon}{f(n)} \geqq 0\right] &\leqq E\exp\left(t\left\{\frac{\pm S_n - \epsilon}{f(n)}\right\}\right) \\ &= \exp\left(\frac{-t\epsilon}{f(n)}\right)\prod_{k=-\infty}^{\infty} f_k\left(\frac{\pm a_{n,k}\, t}{f(n)}\right), \end{aligned}$$

whenever the right hand side exists.

Let $\beta = \epsilon/2A$. Suppose $0 \leqq t < T$ of the lemma. Then

$$P[\pm S_n \geqq \epsilon] \leqq e^{-t\epsilon/f(n)}e^{A\beta t/f(n)} \leqq [e^{-t\epsilon/2}]^{1/f(n)}.$$

Set $t = T/2$ so that $\rho_\epsilon = e^{-\epsilon T/4}$. Inequality 4 of the theorem now follows from the fact that

$$P[|S_n| \geqq \epsilon] \leqq P[S_n \geqq \epsilon] + P[-S_n \geqq \epsilon].$$

3. Application of Theorem 1 to linear processes. Let $\{a_i : i = 0, \pm 1, \cdots\}$ be a sequence of numbers such that $\sum_{i=-\infty}^{\infty} |a_i|^2 < \infty$. Then, if $\{\xi_k : k = 0, \pm 1, \cdots\}$ is a sequence of independent identically distributed random variables with finite variances, the stochastic process

$$X_k = \sum_{i=-\infty}^{\infty} a_i \xi_{k-i}, \qquad k = 1, 2, \cdots,$$

is called a linear process. In [2] Koopmans showed, that under the additional assumptions that the moment generating function of ξ_k exists and $\sum_{i=-\infty}^{\infty} |a_i| < \infty$, the strong law of large numbers for the linear process holds with exponential convergence rate. I.e., if $E\xi_k = 0$ and $S_n = n^{-1} \sum_{k=1}^{n} X_k$, for every $\epsilon > 0$ there exist constants A and ρ, $0 < \rho < 1$, such that

$$P[|S_n| \geqq \epsilon \text{ for some } n \geqq m] \leqq A\rho^m, \qquad m = 1, 2, \cdots.$$

We will show that an application of Theorem 1 makes it possible to generalize this result to a class of linear processes of non-identically distributed random variables, ξ_k, and to arbitrary subsequences of the X_k's.

THEOREM 2. *Let $\{\xi_k : k = 0, \pm 1, \cdots\}$ be a sequence of independent random variables with $E\xi_k = 0$ and moment generating functions $f_k(t)$ which satisfy Condition* 1 *of Theorem* 1. *Let $\{a_i : i = 0, \pm 1, \cdots\}$ be a sequence of numbers such that*

$$\sum_{i=-\infty}^{\infty} |a_i| < \infty.$$

Then, the random variables

$$X_k = \sum_{i=-\infty}^{\infty} a_i \xi_{k-i}$$

are defined as limits of the partial sums almost surely for $k = 1, 2, \cdots$, and if $R = \{r_1, r_2, \cdots\}$ is any subsequence of distinct integers, for every $\epsilon > 0$ there exists ρ, $0 < \rho < 1$, independent of R, such that

$$P[|S_n^R| \geqq \epsilon \quad \text{for some} \quad n \geqq m] \leqq [2/(1-\rho)]\rho^m, \qquad m = 1, 2, \cdots,$$

where

$$S_n^R = n^{-1} \sum_{i=1}^{n} X_{r_i}.$$

PROOF. Note that

$$S_n^R = n^{-1} \sum_{i=1}^{n} \sum_{i=-\infty}^{\infty} a_k \xi_{r_i - k}.$$

If $a_{n,k}$ is the coefficient of ξ_k in S_n^R and $A = \sum_{i=-\infty}^{\infty} |a_i|$, then $|a_{n,k}| \leqq A/n$ and $\sum |a_{n,k}| \leqq A$. Thus, $f(n) = \sup_k |a_{n,k}| \leqq A/n$ and it follows that Conditions 2 and 3 of Theorem 1 are satisfied for this assignment of the $a_{n,k}$'s, where the constant A is independent of R.

Now, as a consequence of the theorem,

$$S_n^R = \sum_{k=-\infty}^{\infty} a_{n,k} \xi_k = n^{-1} \sum_{i=1}^{n} \sum_{k=-\infty}^{\infty} a_{r_i - k} \xi_k$$
$$= n^{-1} \sum_{i=1}^{n} X_{r_i},$$

is defined as an almost sure limit for all n and all R. Thus, taking $n = 1$ and an R with $r_1 = k$, it follows that each X_k is defined almost surely. Also, for every

$\epsilon > 0$ there exists γ depending only on the ξ_k's and on A and ϵ such that

$$P[|S_n^R| \geqq \epsilon] \leqq 2\gamma^{1/f(n)}.$$

Thus, since $f(n) \leqq A/n$,

$$P[|S_n^R| \geqq \epsilon] \leqq 2\rho^n$$

where $\rho = \gamma^{1/A}$. Now

$$P[|S_n^R| \geqq \epsilon \quad \text{for some} \quad n \geqq m] \leqq \textstyle\sum_{n=m}^{\infty} P[|S_n^R| \geqq \epsilon] \leqq 2(1 - \rho)^{-1}\rho^m,$$

$m = 1, 2, \cdots$, and the proof is complete.

4. Application of Theorem 1 to summability theory. An infinite matrix $M = (a_{n,k})$, $n, k = 1, 2, \cdots$, is called a Toeplitz matrix if the following three conditions are satisfied:

$$\textstyle\sum_{k=1}^{\infty} |a_{n,k}| \leqq A < \infty, \qquad n = 1, 2, \cdots, \tag{2'}$$

$$\lim_{n\to\infty} a_{n,k} = 0, \qquad k = 1, 2, \cdots, \tag{3'}$$

$$\lim_{k\to\infty} \textstyle\sum_{k=1}^{\infty} a_{n,k} = 1, \qquad n = 1, 2, \cdots. \tag{5}$$

An infinite column vector $u = (u_1, u_2, \cdots)'$ can be formally transformed into the column vector $Mu = (s_1, s_2, \cdots)'$,

$$s_n = \textstyle\sum_{k=1}^{\infty} a_{n,k}u_k,$$

by the usual operation of left multiplying a column vector by a matrix. If the resulting sequence $\{s_n\}$ converges to a limit s, the sequence is said to be summable M to the value s, and the expressions s_n are called Toeplitz means.

It is well known (see e.g. [5]) that Conditions $2'$, $3'$ and 5 are sufficient to guarantee that if $u_n \longrightarrow s$, then $s_n \longrightarrow s$. In fact, if $s = 0$, Condition 5 can be dropped.

The application of the matrix M to a class of sequences is called a limitation process if there exists at least one non-convergent sequence in the class which is M summable. The following theorem provides convergence rates for Toeplitz means of independent random variables.

THEOREM 3. A. *Let the matrix* $M = [a_{n,k}]_{n,k=1}^{\infty}$ *satisfy the conditions*

$$\textstyle\sum_{k=1}^{\infty} |a_{n,k}| \leqq A < \infty, \qquad n = 1, 2, \cdots,$$

and

$$f(n) = \sup_k |a_{n,k}| \longrightarrow 0 \quad as \quad n \longrightarrow \infty.$$

Let $\mathcal{C}$ *denote the class of sequences* $\xi = \{\xi_k : k = 1, 2, \cdots\}$ *of independent random variables with* $E\xi_k = 0$ *for all* k *and moment generating functions* $f_k(t)$ *satisfying Condition* 1 *of Theorem* 1. *Then every sequence* $\xi \in \mathcal{C}$ *is* M *summable in probability to zero with convergence rate* $\rho^{1/f(n)}$ *in the sense given by Equation* 4.

B. *Every matrix* M *satisfying the conditions of the theorem yields a limitation process on* $\mathcal{C}$ *in the sense of convergence in probability.*

C. *For every matrix M and sequence $\xi \varepsilon \mathcal{C}$ there exists a subsequence $\{r_1, r_2, \cdots\}$ of the positive integers which depends only on $f(n)$ such that if S_n is the* n*th term of the sequence $M\xi$ of Toeplitz means, then $S_{r_k} \to 0$ almost surely with rate $\rho^{1/f(r_k)}$. That is, for every $\epsilon > 0$ there exists $0 < \rho < 1$, such that*

$$P[|S_{r_l}| \geqq \epsilon \quad \textit{for some} \quad l \geqq k] \leqq [2/\rho(1-\rho)]\rho^{1/f(r_k)}.$$

PROOF. Part A follows immediately from Theorem 1 by setting $a_{n,k} = 0$ for $k \leqq 0$. Part B is a trivial consequence of the fact that $\mathcal{C}$ contains all identically distributed sequences which possess moment generating functions and have zero expectations.

To prove Part C, let n_k, $k = 1, 2, \cdots$ be those integers for which $n_k \leqq 1/f(r) < n_k + 1$ for some integer r. Let r_k be the smallest integer for which $n_k \leqq 1/f(r_k) < n_k + 1$. Then, from Part A,

$$\begin{aligned} P[|S_{r_l}| \geqq \epsilon \quad \text{for some} \quad l \geqq k] &\leqq \textstyle\sum_{l=k}^{\infty} P[|S_{r_l}| \geqq \epsilon] \\ &\leqq 2\textstyle\sum_{l=k}^{\infty} \rho^{1/f(r_l)} \leqq 2\rho^{n_k}\textstyle\sum_{l=k}^{\infty} \rho^{n_l - n_k} \\ &\leqq 2\rho^{n_k}\textstyle\sum_{r=0}^{\infty} \rho^{r} = 2(1-\rho)^{-1}\rho^{n_k} \\ &\leqq 2[\rho(1-\rho)]^{-1}\rho^{1/f(r_k)}. \end{aligned}$$

An important class of summability methods are the Cesaro methods which are ordered by a continuous index α for $\alpha \geqq 0$. The Toeplitz matrix for the Cesaro method of order α, or (C, α) method, has entries

$$\begin{aligned} a_{n,k}^{\alpha} &= A_{n-k}^{\alpha-1}/A_n^{\alpha} && \text{for } k = 0, 1, \cdots, n \\ &= 0 && \text{otherwise.} \end{aligned}$$

We extend our indexing to $k = 0$ in order to conform to the notation used in our principle reference on (C, α) summability (i.e. to the notation used in [5]). From [5], the functions A_n^{α} are defined for all real α by the expressions

(6) $\quad A_0^{\alpha} = 1 \quad \text{and} \quad A_n^{\alpha} = \binom{n+\alpha}{n} = \prod_{k=1}^{n} (1 + (\alpha/k)) \quad \text{for } n = 1, 2, \cdots.$

The A_n^{α}'s have the following properties in the α range of interest, $\alpha \geqq -1$:

(7) $$\textstyle\sum_{k=0}^{n} A_k^{\alpha-1} = A_n^{\alpha},$$

(8) For $\alpha > -1$, $A_n^{\alpha} \simeq n^{\alpha}$ in the sense that $\lim_{n\to\infty} A_n^{\alpha}/n^{\alpha}$ exists and is finite and non-zero,

(9) A_n^{α} is positive for $\alpha > -1$, is an increasing function of n for $\alpha > 0$ and is decreasing for $-1 < \alpha < 0$.

From (7) it follows that Property $2'$ is satisfied by the (C, α) coefficients for all $\alpha > 0$ with $A = 1$. Moreover, for $0 < \alpha < 1$

(10) $$f(n) = A_0^{\alpha-1}/A_n^{\alpha} = 1/A_n^{\alpha} \simeq 1/n^{\alpha},$$

and for $\alpha \geqq 1$,

(11) $$f(n) = A_n^{\alpha-1}/A_n^{\alpha} \simeq 1/n,$$

because of (6), (8), and (9). Hence (3′) is satisfied for all $\alpha > 0$.

Note from (6) that as $\alpha \to 0$, $a_{n,k}^{\alpha} \to 0$ for $0 \leqq k \leqq n - 1$ and $a_{n,n}^{\alpha} \to 1$. Thus, (C, α) summability blends "continuously" into ordinary convergence as $\alpha \to 0$. In this sense, we will now show that every sequence $\xi \varepsilon \mathcal{C}$ (defined in Theorem 3) is "almost" convergent in that it is (C, α) summable in probability to zero for all $\alpha > 0$.

THEOREM 4. *Let $\xi \varepsilon \mathcal{C}$, $\alpha > 0$ and let S_n^{α} be the nth Toeplitz partial (C, α) sum of ξ. Then for every $\epsilon > 0$ there exist constants $B > 0$ and $0 < \gamma < 1$ such that*

$$P[|S_n^{\alpha}| \geqq \epsilon] \leqq B\gamma^{n^{\delta(\alpha)}}, \qquad n = 1, 2, \cdots,$$

where $\delta(\alpha) = \alpha$ or 1 according as $0 < \alpha < 1$ or $\alpha \geqq 1$.

PROOF. The proof is an immediate consequence of Theorem 3 and the above discussed Properties (10) and (11) of the (C, α) Toeplitz coefficients.

As a corollary to Part C of Theorem 3 we state the following result.

COROLLARY. *$S_n^{\alpha} \to 0$ almost surely with exponential convergence rate for all $\alpha \geqq 1$. For $0 < \alpha < 1$ there exist subsequences $\{r_1, r_2, \cdots\}$ (dependent on α) such that $S_{r_n}^{\alpha} \to 0$ almost surely with rate $\gamma^{(r_n)^{\alpha}}$.*

Baum, Katz and Read [1] established almost sure $(C, 1)$ convergence with exponential convergence rate for sequences of random variables satisfying a somewhat weaker uniformity condition on the moment generating functions than Condition 1. They were also able to prove their condition to be necessary.

It is to be noted that the restriction $E\xi_k = 0$ is, in part, only a convenience. If $E\xi_k = \mu_k$, and if

$$\sigma_n = \lim_{a,b\to\infty} \textstyle\sum_{k=-a}^{b} a_{n,k}\mu_k$$

exists for all n and $\lim \sigma_n = \sigma$ exists and is finite, the convergence of the partial sums $\sum_k a_{n,k}\xi_k$ can be interpreted as essential convergence with the centering constants σ_n. Then, wherever S_n appears, it is to be replaced by $S_n - \sigma_n$ and it follows that $S_n \to \sigma$ in the given mode. In Theorem 3, to insure that $\sigma_n \to \mu$ if $\mu_n \to \mu$ we reinstate Condition 5.

REFERENCES

[1] BAUM, LEONARD E., KATZ, MELVIN, and READ, ROBERT R. (1962). Exponential convergence rates for the law of large numbers. *Trans. Amer. Math. Soc.* **102** 187–199.

[2] KOOPMANS, L. H. (1961). An exponential bound on the strong law of large numbers for linear stochastic processes with absolutely convergent coefficients. *Ann. Math. Stat.* **32** 583–586.

[3] LOÈVE, M. (1960). *Probability Theory*, 2nd Ed. Van Nostrand, New York.

[4] TITCHMARSH, E. C. (1939). *The Theory of Functions*, 2nd Ed. Oxford Univ. Press, Cambridge.

[5] ZYGMUND, ANTONI (1955). *Trigonometrical Series*. Dover, New York.