The Annals of Probability
1973, Vol. 1, No. 2, 263-271

A LINEAR EXTENSION OF THE MARTINGALE CONVERGENCE THEOREM

BY JAMES B. MACQUEEN

University of California, Los Angeles

Let $X_1, X_2, \cdots$ be a sequence of random variables satisfying $E(X_{n+1} \mid X_n, X_{n-1}, \cdots, X_1) = a_1 X_n + a_2 X_{n-1} + \cdots + X_{n-k-1}$, $n \geqq k$, where $a_1 + a_2 + \cdots + a_k = 1$. Under certain general conditions, mainly that $\sup_n E|X_n| < \infty$, it is shown that $X_n - Y_n \rightarrow_{\text{a.s.}} 0$, where $\{Y_n\}$ is a solution of the homogeneous equation $y_n = a_1 y_{n-1} + a_2 y_{n-2} + \cdots + a_k y_{n-k}$. Several applications of possible theoretical interest are described. Also, the results suggest some extensions of classical results in the theory of random walks which are outlined.

1. Introduction. Consider a sequence of random variables $X_1, X_2, \cdots$ satisfying for all $n \geqq k$,

$$E(X_{n+1} \mid X_n, X_{n-1}, \cdots, X_1) = a_1 X_n + a_2 X_{n-1} + \cdots + a_k X_{n-k+1}\,, \tag{1}$$

where $a_1, a_2, \cdots, a_k$ are given constants, $a_k \neq 0$, and

$$\textstyle\sum_1^k a_i = 1\,. \tag{2}$$

Such a sequence will be called—for lack of a better term—a *linear martingale*. The purpose of this paper is to provide an elementary extension of Doob's (1953) well-known martingale convergence theorem to such sequences.

Let $r_1, r_2, \cdots, r_k$ be the roots of the characteristic equation,

$$r^k - (a_1 r^{k-1} + a_2 r^{k-2} + \cdots + a_k) = 0\,, \tag{3}$$

corresponding to the homogeneous equation

$$z_{n+1} = a_1 z_n + a_2 z_{n-1} + \cdots + a_k z_{n-k+1}\,. \tag{4}$$

Our main result is the following:

THEOREM 1. *Let* $X_1, X_2, \cdots$ *be a linear martingale with* $\sup_n E|X_n| < \infty$, *and with the roots of* (3) *all having moduli* $|r_i| \leqq 1$, *those roots* r_i *with* $|r_i| = 1$ *being simple. Then there exists a sequence of random variables* $Y_1, Y_2, \cdots$, *such that with probability one, the homogeneous equation* (4) *is satisfied by the sequence and* $X_n - Y_n \rightarrow 0$.

The case of multiple roots with moduli equal to one is of a class with the case where some roots have moduli greater than one, in that both cases give rise to unbounded solutions of (4). By repeated application of (1) we observe that $E(X_n \mid X_k, X_{k-1}, \cdots, X_1)$, $n > k$, is given by the solution to (4) with $z_1 = X_1$, $z_2 = X_2, \cdots, z_k = X_k$. Thus in both cases it seems clear that a contradiction of the hypothesis $\sup_n E|X_n| < \infty$ would be reached immediately unless $X_1, X_2, \cdots, X_k$, and hence $X_{k+1}, X_{k+2}, \cdots$, all correspond precisely to one of the bounded

Received May 1, 1972.

solutions of (4), and, *ab contrario*, the conclusion of Theorem 1 would still be true. We do not treat these cases further.

Conditions (1) and (2) above may be considered as defining a stochastic difference equation and then Theorem 1 characterizes the "solution" under certain additional conditions, mainly $\sup_n E|X_n| < \infty$. Thus under these conditions X_n is asymptotically stationary and "harmonic" (cf. Wold and Jurreen [5] page 164), that is, asymptotically of the form $\sum_i \rho_i \cos(n\theta_i + \gamma_i)$ (see (9′) below). However, as will shortly be apparent, Theorem 1 is essentially a version of the martingale convergence theorem, and the process defined by (1) and (2) appears to be more closely akin to the usual martingale than the usual stationary process studied by means of such equations. This is because of the special assumption that $\sum_1^k a_i = 1$, which keeps the process centered somewhere in its recent past. Such a process has to settle down to a deterministic pattern of behavior if it is to remain bounded away from infinity. In the martingale case this is shown by the upcrossing inequality which proves that continued variation of X_n is incompatible with the condition $\sup_n E|X_n| < \infty$. Theorem 1 is based indirectly on this same relationship.

It may be worth pointing out, however, that if the sequence of random variables $V_1, V_2, \cdots$ satisfies $E(V_{n+1} \mid V_n, V_{n-1}, \cdots, V_1) = b_1 V_n + b_2 V_{n-1} + \cdots + b_k V_{n-k+1}$, $n \geqq k$, then an associated linear martingale can be defined as follows: Set $X_n = V_n c^{-n}$ where c is any nonzero real root of the equation $c^k - (b_1 c^{k-1} + b_2 c^{k-2} + \cdots + b_k) = 0$, such a root being assumed to exist. Then $X_1, X_2, \cdots$ is a linear martingale with $a_1 = b_1 c^{-1}$, $a_2 = b_2 c^{-2}, \cdots, a_k = b_k c^{-k}$. Some information about the behavior of $V_1, V_2, \cdots$, may thus be provided by analysis of the latter. It may also be worth pointing out that processes satisfying (1) and (2) may be defined in the following essentially equivalent way: Take $X_1, X_2, \cdots, X_{k-1}$ and U_k as given and let $X_n = U_n + b_1 X_{n-1} + b_2 X_{n-2} + \cdots + b_{k-1} X_{n-k+1}$ for $n \geqq k$, where the random variables $U_k, U_{k+1}, \cdots$ satisfy $E(U_n \mid X_{n-1}, X_{n-2}, \cdots, X_1) = U_{n-1}$, $m \geqq k+1$, and $b_1, b_2, \cdots, b_{k-1}$ are constants. We have $X_{n+1} = U_{n+1} + b_1 X_n + \cdots + b_{k-1} X_{n-k+2} - (-X_n + U_n + b_1 X_{n-1} + \cdots + b_{k-1} X_{n-k+1}) = X_{n+1} = U_{n+1} - U_n + (1 + b_1) X_n + (b_2 - b_1) X_{n-1} + \cdots + (b_{k-1} - b_k) X_{n-k+2} - b_{k-1} X_{n-k+1}$ with the coefficients of the X's in the latter summing to one and with $E(U_{n+1} - U_n \mid X_n, X_{n-1}, \cdots, X_1) = 0$. In view of this, the process $X_1, X_2, \cdots$ so defined may be analyzed directly. We have not found it interesting or technically convenient to do so, especially in view of the possible applications and the immediate technical reduction of the problem in Lemma 1 below. The converse relation between the above definition and (1) and (2) is implicit in Lemma 1.

Examples illustrating possible application of the linear martingale model are given in Section 3.

2. Proof of Theorem 1.

LEMMA 1. *Let* $U_n = X_n + A_2 X_{n-1} + \cdots + A_k X_{n-k+1}$, $n \geqq k$, *where* $A_j = \sum_{i=j}^k a_i$. *Then the sequence of random variables* $U_k, U_{k+1}, \cdots$ *is a martingale* *sequence with* $\sup_n E|U_n| < \infty$, *and is uniformly integrable if the sequence* X_1, $X_2, \cdots$ *is uniformly integrable; moreover,* $U_n \to U$ *with probability* 1, *and in the case of uniform integrability,* $E(U \mid U_k) = U_k$.

PROOF. $E(U_{n+1} \mid U_n, U_{n-1}, \cdots, U_k) = E(E(U_{n+1} \mid X_n, X_{n-1}, \cdots, X_1) \mid U_n, U_{n-1}, \cdots, U_k) = a_1 X_n + a_2 X_{n-1} + \cdots + a_k X_{n-k+1} + A_2 X_n + A_3 X_{n-1} + \cdots + A_k X_{n-k+2} = U_n$. Using the triangle inequality it is easy to show that $\sup_n E|X_n| < \infty$ implies $\sup_n E|U_n| < \infty$ and that if the sequence $X_1, X_2, \cdots$ is uniformly integrable, so is the sequence $U_k, U_{k+1}, \cdots$. Application of the martingale convergence theorem completes the proof.

From Lemma 1 we see that

$$\delta_{n+1} = U_{n+1} - U_n = X_{n+1} - (a_1 X_n + a_2 X_{n-1} + \cdots + a_k X_{n-k+1}) \to 0 \tag{5}$$

with probability one and, in fact, $\sum_{i=k}^{n} \delta_i = U_n - U_k$ converges. However, examples will readily show that convergence of $\sum \delta_i$ alone is not enough to identify the limiting behavior of $X_1, X_2, \cdots$ as belonging to the class of solutions of (4). For this purpose we will use the following lemma, which characterizes solutions to the discrete renewal equation in a form suitable for the problem at hand.

LEMMA 2. *Let* $x_1, x_2, \cdots, x_k$ *be given numbers and for* $n \geqq k+1$, *define* x_n *recursively by*

$$x_{n+1} = \delta_{n+1} + a_1 x_n + a_2 x_{n-1} + \cdots + a_k x_{n-k+1}\,. \tag{6}$$

Then for $n \geqq k + 1$,

$$x_n = z_n + \delta_n v_k + \delta_{n-1} v_{k+1} + \cdots + \delta_{k+1} v_{n-1}\,, \tag{7}$$

where $z_1, z_2, \cdots$ *is the solution of* (4) *with initial values* $z_1 = x_1, z_2 = x_2, \cdots, z_k = x_k$, *and* $v_1, v_2 \cdots$ *is the solution of* (4) *with initial values* $v_k = 1, v_{k-1} = v_{k-2} = \cdots = v_1 = 0$.

PROOF. The sequence $x_1, x_2, \cdots$ is uniquely defined by (6), hence (7) can be verified by direct substitution.

Consider first $n = k + i$ for $1 \leqq i \leqq k$. In this range the left side of (6) becomes

$$z_{k+i+1} + \delta_{k+i+1} v_k + \delta_{k+i} v_{k+i} + \cdots + \delta_{k+1} v_{k+i}$$

and the right side becomes

$$\begin{aligned} \delta_{k+i+1} + a_1 z_{k+i} + a_2 z_{k+i-1} + \cdots + a_i z_{k+1} & \\ + a_1[\delta_{k+i} v_k + \delta_{k+i-1} v_{k+1} + \cdots + \delta_{k+1} v_{k+i-1}] & \\ + a_2[\delta_{k+i-1} v_k + \cdots + \delta_{k+1} v_{k+i-2}] & \\ \vdots \qquad & \\ + a_i[\delta_{k+1} v_k] & \\ + a_{i+1} x_k + a_{i+2} x_{k-1} + \cdots + a_k x_{i+1}\,. & \end{aligned}$$

In the above triangular array, terms with the same subscripts on δ are arranged in columns. Since $v_n = 0$ for $n < k$, $v_{k+j} = a_1 v_{k+j-1} + \cdots + a_j v_k$ for $0 \leqq j \leqq k$; hence on collecting terms in columns we see the two sides are equal. Equality of the two sides for $n > 2k$ is shown similarly by extending the above array to the right as far as necessary, and down until $i = k$.

To apply Lemma 2 we employ the well-known fact from the theory of linear recursion that every solution of (4) is of the form $\sum_i \gamma_i(n) r_i^n$ where γ_i is a polynomial in n (with complex coefficients) of degree one less than the multiplicity of r_i, there being one such polynomial taken in the sum for each distinct root. The coefficients in the polynomials, altogether k in number, are to be determined from boundary conditions. Thus let $z_n = \sum_i \alpha_i(n) r_i^n$ and let $v_n = \sum_i \beta_i(n) r_i^n$ where the coefficients in the polynomials α_i and β_i are determined from the respective initial conditions specified in Lemma 2 with $z_i = X_i$, $i = 1, 2, \cdots, k$. Then, using (7) with δ_n defined by (5), we have

$$X_n = \sum_i [\alpha_i(n) r_i^n + (\delta_n \beta_i(k) r_i^k + \delta_{n-1} \beta_i(k+1) r_i^{k+1} + \cdots + \delta_{k+1} \beta_i(n-1) r_i^{n-1})] . \tag{8}$$

For integers i such that $|r_i| < 1$ the contribution to (8) is easily seen to tend to zero with probability one. For such i, $\alpha_i(n) r_i^n \to 0$ and for the series in parentheses we have, for every fixed m, $\delta_n \beta_i(k) r_i^k + \cdots + \delta_{n-m} \beta_i(k+m) r_i^{k+m} \to 0$ since $\sum \delta_i$ converges, and for every fixed m, $|\delta_{n-m-1} \beta_i(k+m+1) r_i^{k+m+1} + \cdots + \delta_{k+1} \beta_i(n-1) r_i^{n-1}| \leqq (\max_n |\delta_n|) L_i |r_i|^{k+m}$ where L_i is the finite limit of the sum $|\beta_i(k+m+1) r_i| + |\beta_i(k+m+2) r_i^2| + \cdots$.)

For i such that $|r_i| = 1$ the polynomials α_i and β_i are simply constants since by hypothesis such roots are simple, and we may write the corresponding contribution to (8) in the form

$$\alpha_i r_i^n + \beta_i r_i^{n+k} (\delta_{k+1} r_i^{-(k+1)} + \delta_{k+2} r_i^{-(k+2)} + \cdots + \delta_n r_i^{-n}) .$$

Since $\delta_{k+1}, \delta_{k+2}, \cdots$ are the increments of a martingale the series in parentheses is a *martingale transform*. A theorem of Burkholder applies ([2] Theorem 1, page 1496) to show that the series converges with probability one. Let φ_i be the finite limit. (Of course, in the case where r_i has a nonzero imaginary part, the integers i in question will appear in pairs corresponding to conjugate pairs of roots with corresponding conjugate coefficients in the above, and their respective limits will be conjugate.)

Finally, if we let

$$Y_n = \sum_i (\alpha_i + \beta_i r_i^k \varphi_i) r_i^n \tag{9}$$

where the sum is taken over integers i such that $|r_i| = 1$, then Y_n satisfies (4) and $Y_n - X_n \to 0$ as was to be shown.

Equation (9) is readily put in the form

$$Y_n = \sum_i \rho_i \cos (n\theta_i + \gamma_i) \tag{9'}$$

where the θ_i are amplitudes of the roots r_i with $|r_i| = 1$, while ρ_i and γ_i are (real) random variables. There will be one constant term in the sum for the root $r = 1$ (so that $\cos(n\theta_i + \gamma_i) = 1$ for that root); one term for the real root $r = -1$ if it occurs, in which case the corresponding term reduces to $\rho_i(-1)^n$; and one term for each of the conjugate pairs of roots with imaginary parts. From (9′) it is clear that the random contribution to the asymptotic behavior of X_n is a random phase and intensity for each of the basic frequencies determined by the persistent part of the solution to (4).

3. Remarks. If all roots of (3) except the single invariable root $r = 1$ have moduli less than one, then only the constant term corresponding to $r = 1$ will appear in (9) (i.e., all solutions of (4) are asymptotically constant), and we conclude $X_n \to X$ with probability one. Since $U_n = X_n + A_2X_{n-1} + \cdots + A_kX_{n-k+1} \to U$, with probability one, we have $U = X(1 + A_2 + \cdots + A_k)$ with probability one. Let $\mu = 1 + A_2 + \cdots + A_k$. Then $E(U \mid U_k) = E(X \mid U_k)\mu$. We note also that $\mu \neq 0$, since on dividing (3) by $r - 1$, the reduced equation is $r^{k-1} + A_2r^{k-2} + \cdots + A_k = 0$, and if $\mu = 0$, $r = 1$ would be a root of multiplicity of at least two. In fact $\mu \neq 0$ is already implied by the hypotheses of Theorem 1, by this same argument.

This relation between X and U appears to be particularly useful in the case uniform integrability where $E(U \mid U_k) = U_k = X_k + A_2X_{k-1} + \cdots + A_kX_1$. In this case the latter formula together with the above observations yield the following corollary of Theorem 1.

COROLLARY 1. *If $X_1, X_2, \cdots$ is a uniformly integrable linear martingale (hence $\sup_n E|X_n| < \infty$) with (3) having the one simple root $r = 1$, and with the moduli of the remaining roots being less than one, then $X_n \to X$ with probability one, and*

$$(10) \qquad E(X \mid X_k, X_{k-1}, \cdots, X_1) = (X_k + A_2X_{k-1} + \cdots + A_kX_1)/\mu\,,$$

where $\mu = 1 + A_2 + \cdots + A_k$.

As will be illustrated in the examples which follow, (10) provides a surprisingly simple method for calculating certain absorption probabilities.

In the case where the a_i are nonnegative, an easy check that the condition on the roots of (3) in Corollary 1 is satisfied is provided for by the following lemma. (In fact all the examples given below utilize this case.) If the a_i are nonnegative, and the greatest common divisor of the integers i such that $a_i > 0$ exceeds one, we will say the a_i are *arithmetic*, while if the greatest common divisor is equal to one, we will say the a_i are *aperiodic*.

LEMMA 3. *If the a_i are nonnegative and aperiodic then there is only one root of (3) with a modulus of one and the remaining roots have moduli less then one.*

PROOF. If $|r| > 1$, we have $1 < |r^k| = |a_1r^{k-1} + \cdots + a_k| \leqq a_1|r|^{k-1} + a_2|r|^{k-2} + \cdots + a_k \leqq |r|^{k-j}$ where j is the first integer such that $a_j > 0$, which is impossible. If $|r| = 1$, then evidently $|r|^k = |a_1r^{k-1} + \cdots + a_k| = a_1|r|^{k-1} + a_2|r|^{k+2} + \cdots + a_k$ which is possible only if the arguments of the terms r^{k-j} for which $a_j > 0$ are all equal to the argument of r^k. Since $r^{i-j} = r^{k-j}/r^{k-i}$ has modulus one and argument zero, $r^{i-j} = 1$. Thus r is an nth root of unity where n is no larger than the largest common divisor of the differences $i - j$ among pairs i, j for which a_i and a_j are both positive. By hypothesis this largest divisor is unity. Hence $r = 1$. But there is only one root $r = 1$, since on dividing out $r - 1$ from (3) we obtain the reduced equation $r^{k-1} + A_2 r^{k-2} + \cdots + A_k$ which is, of course, positive when $r = 1$ since the a_i are all nonnegative. Thus all the roots but one have moduli less than one as was to be shown.

We remark that if the a_i are nonnegative and aperiodic, a proof of the situation covered by Corollary 1 and Lemma 3 can be obtained directly from familiar results in the study of random walks. Consider (5) ($X_{n+1} = \delta_{n+1} + a_1 X_n + a_2 X_{n-1} + \cdots + a_k X_{n-k+1}$, $n \geqq k$) as the equation for the expected total future income, X_{n+1}, of a man who starting from state $n + 1$, $n \geqq k$, receives δ_{n+1} immediately and moves to state $n + 1 - i$ with probability a_i, and who receives (the initially given) X_j immediately if he is in state $j \leqq k$, but his income stops the first time he reaches such a state. Thus his expected income will be

$$X_{n+1} = \delta_{n+1} + p_{n+1}(n)\delta_n + p_{n+1}(n-1)\delta_{n-1} + \cdots + p_{n+1}(k+1)\delta_{k+1} + p^0_{n+1}(k)X_k + p^0_{n+1}(k-1)X_{k-1} + \cdots + p^0_{n+1}(1)X_1,$$

where $p_{n+1}(j)$ is the probability the man visits state j, $j \geqq k + 1$, starting from $n + 1$, and $p^0_{n+1}(i)$, $i = 1, 2, \cdots, k$ is the probability the man hits state i on entering the set $\{1, 2, \cdots, k\}$ for the first time. It is clear from the interpretation that this formula will satisfy (5). From the renewal theorem it follows that as $n \to \infty$, $p_n(j) \to 1/\mu$, and a basic result on hitting probabilities (see, e.g., Spitzer [4] P7, page 285) yields $p_n{}^0(i) \to A_i/\mu$ where we set $A_1 = 1$. Thus the series $\delta_{n+1} + p_{n+1}(n)\delta_n + \cdots + p_{n+1}(k+1)\delta_{k+1} \to (\sum_{i=1}^{\infty} \delta_{k+i})/\mu = (U - U_k)/\mu$ (using Lemma 1) and $X_n \to X = (U - U_k)/\mu + U_k/\mu = U/\mu$. From this we get the formula $E(X \mid U_k) = U_k/\mu$ in the uniformly integrable case as above.

In the arithmetic case with the a_i nonnegative, the zeros which appear between positive a_i make it possible to decompose the process into several separate linear martingales. The number of such martingales required will be equal to the greatest common divisor of the i for which a_i is positive. Thus a complete proof of Theorem 1 in the case of nonnegative a_i is possible along the above lines.

The following simple example illustrates the above decomposition and the more general phenomenon of a random harmonic limit as well.

EXAMPLE 1. Let each X_n be either 0 or 1, and let $a_1 = a_3 = 0$, $a_2 = a_4 = \frac{1}{2}$. Let $X_2 = X_1 = 0$, $X_4 = X_3 = 1$. Then given this initial sequence, the sequences $X_1, X_3, \cdots$ and $X_2, X_4, \cdots$ both converge immediately independently of one another to either 1 or 0, and in each case the limit is 1 with probability $\frac{2}{3}$. There are thus four limiting sequences: two all 0's, two all 1's and two oscillating between 0 and 1 but differing in phase. Their respective probabilities are $\frac{1}{9}$, $\frac{4}{9}$, $\frac{2}{9}$, $\frac{2}{9}$.

The following application of Corollary 1 is suggested by the classical problem of achieving a given goal by means of fair gambles, starting with a given initial fortune.

EXAMPLE 2. Let X_n satisfy $0 \leqq X_n \leqq 1$, with the a_i nonnegative and aperiodic, and consider the problem of determining for given $X_1, X_2, \cdots, X_k$ a sequence of random variables $X_{k+1}, X_{k+2}, \cdots$ subject to (1) which maximizes $p = \lim_{n\to\infty} P[1 = X_n = X_{n+1} = \cdots]$ (that is, the probability that X_n is equal to 1 for all n sufficiently large). Clearly, the distribution of the limiting random variable X subject to (10) which assigns maximum probability to $X = 1$ will concentrate the remainder on 0, whence the maximal probability is $(X_k + A_2X_{k-1} + \cdots + A_k)/\mu = E(X \mid X_k, X_{k-1}, \cdots, X_1)$. This is easily achievable, e.g., by choosing each succeeding distribution of X_n to concentrate on 0 or 1 subject to (1).

In the above case where $0 \leqq X_n \leqq 1$ and with probability one $X_n \to 1$ or else $X_n \to 0$, the concept of "total variance" for a martingale (see Dubins (1971)) extends as follows. We want to find $S = E(\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots \mid X_k, X_{k-1}, \cdots, X_1)$ where $\sigma_{n+1}^2 = E([X_{n+1} - E(X_{n+1} \mid X_n, X_{n-1}, \cdots, X_1)]^2 \mid X_n, X_{n-1}, \cdots, X_1) = E(\delta_{n+1}^2 \mid X_n, X_{n-1}, \cdots, X_1) = E([U_{n+1} - U_n]^2 \mid X_n, X_{n-1}, \cdots, X_1)$. We have $E([U_n - U_k]^2 \mid X_k, X_{k-1}, \cdots, X_1) = E([U_n - U_{n-1}) + \cdots + (U_{k+1} - U_k)]^2 \mid X_k, X_{k-1}, \cdots, X_1) = E(\sigma_n{}^2 + \sigma_{n-1}^2 + \cdots + \sigma_{k+1}^2 \mid X_k, X_{k-1}, \cdots, X_1)$ since $U_k, U_{k+1}, \cdots$ is a martingale sequence. But $E([U_n - U_k]^2 \mid X_k, X_{k-1}, \cdots, X_1) \to E((U - U_k)^2 \mid X_k, X_{k-1}, \cdots, X_1) = (\mu - U_k)^2 U_k/\mu + U_k{}^2(1 - U_k/\mu)$ since $U = \mu$ or 0, and $P[U = \mu \mid X_k, X_{k-1}, \cdots, X_1] = U_k/\mu$. This reduces to

$$S = (\mu - U_k)U_k \,. \tag{11}$$

The above argument can be applied in certain first passage time problems. To illustrate, consider the following process: Let $X_{n+1} = (X_n + X_{n-1} + \cdots, X_{n-k+1})/k + \varepsilon_{n+1}$, $n \geqq k$, where the rv's ε_n are independent with $P[\varepsilon_n = +1] = P[\varepsilon_n = -1] = \frac{1}{2}$. Given $X_k, X_{k-1}, \cdots, X_1$ we want to find[1] $T = E(t \mid X_k, X_{k-1}, \cdots, X_1) - k$ where t is the least integer i such that $|X_i| \geqq c$. It is easy to obtain a reasonably sharp approximation for T using the fact that $E((U_{n+1} - U_n)^2 \mid X_n, X_{n-1}, \cdots, X_1) = 1$. The process $X_1, X_2, \cdots$ is a linear martingale with $a_i = 1/k$, $i = 1, 2, \cdots, k$. If we "stop" the process at t and replace $X_{t+1}, X_{t+2}, \cdots$ with $\hat{X}_{t+1}, \hat{X}_{t+2}, \cdots$ where the latter sequence is the deterministic projection of $X_1, X_2, \cdots, X_t$ into the future, that is, $\hat{X}_{t+1} = (X_t + X_{t-1} + \cdots + X_{t-k+1})/k$, $\hat{X}_{t+2} = (\hat{X}_{t+1} + X_t + \cdots + X_{t-k+2})/k$, $\hat{X}_{t+3} = (\hat{X}_{t+2} + \hat{X}_{t+1} + \cdots + X_{t-k+3})/k, \cdots$, then the modified process, call it $X_1', X_2', \cdots$, where $X_n' = X_n$, $n \leqq t$, $X_n' = \hat{X}_n$, $n > t$, is also a linear martingale with $a_i = 1/k$, $i = 1, 2, \cdots, k$, and is uniformly integrable. Let $U_n' = X_n' + A_2X_{n-1}' + \cdots + A_kX_{n-1}' = X_n' + [(k-1)/k]X_{n-1}' + \cdots + (1/k)X_n'$. Then $U_n' \to U'$ with probability one. Also $E((U' - U_k')^2 \mid X_k, X_{k-1}, \cdots, X_1) = T$. If the first time X_n leaves the interval $(-c, c)$

[1] This problem was suggested by a result of Blackwell's (1963) characterizing the expected time until an n-dimensional random walk leaves a spherical set.

occurs at $+c$, then $U' = U_t' \geqq ((k + 1)/2)c - (k - 1)$. The latter inequality follows from the fact that $X_t' = 1 + (X'_{t-1} + \cdots + X'_{t-k})/k \geqq c$ while the variables $X'_{t-1}, X'_{t-2}, \cdots, X'_{t-k}$ are all less then c. Subject to these conditions the minimal value of $X_t' + ((k - 1)/k)X'_{t-1} + \cdots + (1/k)X'_{t-k+1}$ occurs when $X'_{t-k+i} = c$, $i = 1, \cdots, k - 2$, $X'_{t-1} = c - (k - 1)$. (A somewhat sharper bound is clearly possible.) Considering that $X_t \leqq c + 1$, we have also $U' \leqq 1 + ((k + 1)/2)c$. Analogous inequalities hold at $-c$ so we find, then, considering $E(U' - U_k)^2 = E(U')^2 - (U_k)^2$, that

$$\left[\frac{k+1}{2}c - (k-1)\right]^2 - U_k^2 \leqq T \leqq \left[\frac{k+1}{2}c + 1\right]^2 - U_k^2 . \tag{12}$$

The left side is exact if $k = 1$ and X_1 and c are both integers.

While the above bound on T was first noted in connection with the total variance concept, another somewhat more direct derivation is possible along the following lines: Let $T_1 = -(U_k)^2$, $T_2 = 1 - (U_{k+1})^2, \cdots, T_n = n - 1 - (U_{k+n})^2$. Then T_n is a martingale. If T_n is stopped at t where t is the least n for which $|X_{k+n}| \geqq c$ and $T_{t+1}, T_{t+2}, \cdots$ are all replaced by T_t, then the modified martingale, say $T_1', T_2', \cdots$ has $T_n' \to T'$ with probability one, and $E(T' \mid X_k, X_{k-1}, \cdots, X_1) = -U_k^2 = E(t \mid X_k, X_{k-1}, \cdots, X_1) - E(U^2_{k+t} \mid X_k, X_{k-1}, \cdots, X_1)$. By bounding U_{k+1} as above, using the facts that $|X_{k+1}| \geqq c$, while $|X_{k+t-i}| < c$, and $|\varepsilon_n| \leqq 1$, the above bound is again obtained.

If the above process is modified so that $X_{n+1} = \varepsilon_{n+1} - d + (X_n + X_{n-1} + \cdots + X_{n-k+1})/k$ where d is a positive constant, then the same method provides an equally elementary bound for the expected number of steps for X_n to fall below zero for the first time. The relevant martingale here is $T_n = n - 1 + (U_{k+n})/d$ and the inequality for $T^* = E[t \mid X_k, X_{k-1}, \cdots, X_1]$ where t is the least n such that $X_{k+n} \leqq 0$, is

$$\frac{U_k}{d} - \frac{(d+1)}{d} \leqq T^* \leqq \frac{U_k}{d} + \left(\frac{d+1}{d}\right)(k-1) .$$

EXAMPLE 3. *Imitation processes.* Suppose there are N persons in a group who on the nth occasion, $n = 1, 2, \cdots$, each chose one of a finite set of actions. Let $I_n^{\alpha,v} = 1$ if the vth person chooses action α on the nth occasion and 0 otherwise so that $\sum_\alpha I_n^{\alpha,v} = 1$. The probability of each person making the various choices is influenced by the past choices of the others (possibly himself) according to the following relation:

$$P[I_{n+1}^{\alpha,v} = 1 \mid H_n] = \textstyle\sum_{z=1}^N w_{v,z}(a_1 I_n^{\alpha,z} + a_2 I_{n-1}^{\alpha,z} + \cdots + a_k I_{n-k+1}^{\alpha,z}) , \tag{13}$$

where H_n stands for the complete history of all past choices of all persons as of time n, and w_{ij} is an $N \times N$ stochastic array, i.e., $\Sigma_z w_{v,z} = 1$ for all v and $w_{v,z} \geqq 0$. Thus $w_{v,z}$ is the "weight" the vth person gives to the zth person's behavior, and the a_i represent weighing of past actions according to their recency. We assume also that, given H_n, the $I_{n+1}^{\alpha,v}$ are (conditionally) independent with respect to α. Thus our process is actually Markov in the space of all the histories, of length k, of the past choices of the N persons. With $a_i > 0$ and each $w_{v,z} > 0$ it is clear that eventually everyone will be choosing the same action on each occasion. Making this assumption, we show how to calculate p_α, the probability that action α is eventually the common choice, as a function of the initial k choices of each of the persons. The method is very simple.[2] If $w_{v,z} > 0$, then there is a set of positive numbers w_v satisfying $\Sigma_v w_v = 1$, $w_v = \Sigma_z w_z w_{z,v}$. (If we consider $w_{v,z}$ as the transition distribution of a Markov chain, w_v is the stationary distribution.) We multiply both sides of (13) by w_v and sum over v. Then $X_n{}^\alpha = \Sigma_v w_v I_n{}^{\alpha,v}$ is seen to satisfy (1), and $X_n{}^\alpha \to X^\alpha$ with probability one. The only possible values for X^α are 1 or 0; hence

$$(14) \qquad p_\alpha = E(X^\alpha \mid X_k{}^\alpha, X_{k-1}^\alpha, \cdots, X_1{}^\alpha) = (X_k{}^\alpha + A_2 X_{k-1}^\alpha + \cdots + A_k X_1{}^\alpha)/\mu .$$

Acknowledgments. The author first encountered the linear martingale problem in about 1965 in connection with some processes related to Example 3 above, and at that time discussed the problem with Leo Breiman. Breiman, using the martingale $X_n + (X_{n-1} - E(X_n \mid X_{n-1}, \cdots, X_1)) + (X_{n-2} - E(X_{n-1} \mid X_{n-2}, \cdots, X_1)) + \cdots + (X_k - E(X_{k+1} \mid X_k, \cdots, X_1))$ obtained a partial characterization of the asymptotic behavior somewhat like Lemma 1, but valid for certain infinite sequences of weights $a_1, a_2, \cdots$. However, at that time the author did not realize what now appears to be obvious, that the correct asymptotic statement was that X_n tends to one of the solutions of (4). Upon realizing this quite recently, the author happened to discuss the problem with Lloyd Shapley, who noticed the somewhat simpler relationships expressed in Lemma 1.

The author also benefited from several discussions with Dave Cantor who, among other things, pointed out the simple proof of Lemma 3. Finally the author is particularly indebted to Tom Ferguson for several valuable discussions of the entire problem.

REFERENCES

[1] Blackwell, David (1963). Probability bounds via dynamic programming. Paper given at joint *AMS-SIAM* Symposium, May 2, 1963.

[2] Burkholder, D. L. (1966). Martingale transforms. *Ann. Math. Statist.* **37** 1494–1504.

[3] Dubins, Lester E. (1972). Sharp bounds for the total variance of bounded semimartingales. *Ann. Math. Statist.* **43** 1559–1565.

[4] Spitzer, Frank (1964). *Principles of Random Walk.* Van Nostrand, Princeton.

[5] Wold, Herman and Lars Jureen (1953). *Demand Analysis.* Wiley, New York.

Graduate School of Management
University of California
Los Angeles, California 90024

[2] If $w_{z,v}$ is doubly stochastic ($\Sigma_z w_{z,v} = 1$, $\Sigma_v w_{z,v} = 1$) a linear martingale is obtained directly on summing both sides of (13). Tom Ferguson made the useful observation that by first multiplying both sides by w_v, the condition $\Sigma_z w_{z,v} = 1$ could be dropped.