The Annals of Mathematical Statistics
1972, Vol. 43, No. 4, 1374-1379

CONVERGENCE IN DISTRIBUTION, CONVERGENCE IN PROBABILITY AND ALMOST SURE CONVERGENCE OF DISCRETE MARTINGALES[1]

BY DAVID GILAT[2]

Columbia University

Examples are provided of Markovian martingales that: (i) converge in distribution but fail to converge in probability; (ii) converge in probability but fail to converge almost surely. This stands in sharp contrast to the behavior of series with independent increments, and settles, in the negative, a question raised by Loève in 1964. Subsequently, it is proved that a discrete, real-valued Markov-chain with stationary transition probabilities, which is at the same time a martingale, converges almost surely if it converges in distribution, provided the limiting measure has a mean. This fact does not extend to non-discrete processes.

1. Statement of problem and motivation. Consider the elementary implications: (a) *Almost sure convergence* implies *convergence in probability*; (b) *Convergence in probability* implies *convergence in distribution*.

In this paper we address ourselves to an aspect of the following general question: For what classes of processes can the elementary implications be reversed? The problem is, obviously, motivated by the classical case due to P. Lévy (1937) who established the reversibility of the elementary implications (a) and (b) for partial sums of sequences of independent random variables.

In view of the multitude of results, originally proved for sums of independent random variables, and subsequently generalized to martingales, to assume that $X_1, X_2, \cdots$ form a sequence of martingale-differences, seems to be a natural departure from independence. Some of the essential similarities, as well as differences, between sums of independent random variables and martingales already emerge in the study of two random variables. For example (see Gilat (1971)), if X and Y are random variables with the same (marginal) distribution and if the joint distribution of (X, Y) makes X and $Y - X$ independent, then $X = Y$ with probability 1. Likewise, if X and Y satisfy the martingale relation $E(Y \mid X) = X$ a.s., and X and Y have a common distribution, then $X = Y$ a.s., provided X has a mean (possibly an infinite one). In view of this result, one may be tempted to state a continuity theorem, to the effect that if (X, Y) satisfies the martingale-relation and X and Y have "close" (marginal) distributions, then X and Y are "close" in probability. Unfortunately, this is far from the

Received September 14, 1971; revised December 2, 1971.

[1] Most of this paper is extracted from the author's doctoral dissertation at the University of California, Berkeley, written under the guidance of Professor Lester E. Dubins, supported in part by the Air Force Office of Scientific Research Grant AF-AFOSR-1312-67. Prepared for publication with the support of NSF-GP-29661 Grant to Columbia University.

[2] Now at Tel-Avive University.

truth. In fact, for any $a > 0$ (no matter how large) and $0 < \varepsilon < \frac{2}{3}$, there is a probability measure P on the plane, such that the coordinate mappings X and Y satisfy the martingale relation, and such that $|P[X \in B] - P[Y \in B]| < \varepsilon$ for all Borel sets B, but nevertheless $P[|X - Y| > a] = 1$. For example, take $P[X = a] = \frac{1}{2} = P[X = -a]$ and define the conditional distribution of Y given X by:

$$P[Y = \mp a \mid X = \pm a] = 1 - \varepsilon = 1 - P[Y = \pm b \mid X = \pm a],$$

where b is determined by the martingale relation:

$$E[Y \mid X = a] = (1 - \varepsilon)(-a) + \varepsilon b = a .$$

This example uses the idea of balancing a highly probable small step by a big step in the opposite direction, so that the average displacement is zero. Applying this idea repeatedly (with a proper sequence of ε's), one can easily construct martingale-processes that converge in probability but fail to converge almost surely, as well as martingales that converge in distribution but not in probability. Since the construction is straightforward, it will be omitted. However, processes so constructed will inevitably have unbounded increments. It is now natural to inquire whether the elementary implications are reversible for martingales with *uniformly bounded increments.*

2. A partial negative answer. We do not know whether there exist martingales with uniformly bounded increments that converge in distribution but do not converge in probability. The example on page 22 of Gilat (1970), designed to illustrate such a martingale, turns out to be erroneous. We now proceed to construct a martingale with uniformly bounded increments, that converges in probability but fails to converge almost surely.

Let $T = \{t_1, t_2, \cdots\}$ be a strictly increasing sequence of positive integers. We shall refer to members of T as *crucial times.* Let δ be point-mass at 0, and for $k \geqq 1$ let γ_k stand for the three-point distribution $\gamma_k(0) = 1 - 1/k$, $\gamma_k(-1) = 1/2k = \gamma_k(1)$. Let X_1 have γ_1 as its distribution, and for each $n \geqq 1$ define the conditional distribution $\theta_{n+1} = \theta_{n+1}(X_1, \cdots, X_n)$ of X_{n+1} given $X_1, \cdots, X_n$ by:

$$\begin{aligned} \theta_{n+1} &= \gamma_1 && \text{if } X_1 + \cdots + X_n \neq 0 \\ &= \delta && \text{if } X_1 + \cdots + X_n = 0 , && n \notin T \\ &= \gamma_k && \text{if } X_1 + \cdots + X_n = 0 , && n = t_k \in T . \end{aligned}$$

The process $X_1, X_2, \cdots$ so defined is clearly uniformly bounded ($|X_n| \leqq 1$) and satisfies $E(X_{n+1} \mid X_1, \cdots, X_n) = 0$ for all $n \geqq 1$. Its partial sums $\{S_n = X_1 + \cdots + X_n, n \geqq 1\}$ therefore form a martingale with uniformly bounded increments. $\{S_n\}$ behaves like ordinary coin-tossing except that when being at 0 between crucial times, it stays there until the next crucial time, while from zero at the kth crucial time it proceeds according to γ_k. Since $\sum \gamma_k\{-1, 1\} = \sum 1/k = \infty$, it is easy to see that $\{S_n\}$ diverges almost surely, regardless of the sequence T used in the construction. It remains to show that $P[S_n = 0]$ will converge to 1 as $n \to \infty$, for a suitably chosen sequence T of crucial times. To this end, consider the ordinary coin-tossing process $\{Z_n, n \geqq 1\}$, i.e., Z_n is the excess of heads over tails in n independent tosses of a fair coin. For each positive integer a and for each number q, $0 < q < 1$, let $j(a, q)$ be the smallest positive integer j for which

$$P[a + Z_n = 0 \text{ for some } 1 \leqq n \leqq j] = P[-a + Z_n = 0 \text{ for some } 1 \leqq n \leqq j] \geqq q\,;$$

$j(a, q)$ is well defined since $\{Z_n, n \geqq 1\}$ is recurrent; it is easily seen to be increasing in a, non-decreasing in q, and to converge to ∞ when either $a \to \infty$ or $q \to 1$. We now proceed to construct a sequence T of crucial times suitable for our purpose. Let $\{q_n, n \geqq 1\}$ be any increasing sequence of positive numbers converging to 1. Let $j_0 = 1$, $j_1 = j(1, q_1)$ and in general let

$$j_{k+1} = j(j_0 + \cdots + j_k, q_{k+1}) \qquad \text{for } k = 1, 2, \cdots .$$

Let $t_k = j_0 + \cdots + j_{k-1}$ $(k \geqq 1)$ and take $T = \{t_1, t_2, \cdots\}$.

Note that $j_k = j(t_k, q_k) = t_{k+1} - t_k$. Since the q's tend to 1, the j_k's are unbounded, so that T is a legitimate sequence of crucial times. Let $\{S_n, n \geqq 1\}$ be defined as before, with this T serving as the set of crucial times. Observe that the t_k's were constructed in such a way that the probability of getting back to zero, from any position s reachable by time t_{k-1} $(0 < |s| \leqq t_{k-1})$, in $(t_k - t_{k-1})$ or fewer steps, is at least q_k. In particular we have:

$$(1) \qquad P[S_{t_k} = 0 \mid S_{t_{k-1}} = s] \geqq q_k \quad \text{for all } s \text{ with } 0 < |s| \leqq t_{k-1} \text{ and } k > 1\,.$$

Also, by the definition of $\{S_n\}$,

$$(2) \qquad P[S_{t_k} = 0 \mid S_{t_{k-1}} = 0] \geqq 1 - 1/(k-1) \qquad \text{for } k > 1\,.$$

Let $\alpha_k = P[S_{t_k} = 0]$, $k \geqq 1$. Then: $\alpha_1 = 0$ and using (1) and (2) we get for $k > 1$:

$$(3) \qquad \begin{aligned} \alpha_k &= \alpha_{k-1} P[S_{t_k} = 0 \mid S_{t_{k-1}} = 0] + \textstyle\sum_{s \neq 0} P[S_{t_{k-1}} = s] P[S_{t_k} = 0 \mid S_{t_{k-1}} = s] \\ &\geqq \alpha_{k-1}(1 - 1/(k-1)) + (1 - \alpha_{k-1}) q_k \geqq \min(1 - 1/(k-1), q_k) \to 1 \\ &\qquad \text{as } k \to \infty\,. \end{aligned}$$

We have thus proved that $P[S_n = 0]$ tends to 1 as n tends to ∞ through T. If $n \notin T$, then $t_k < n < t_{k+1}$ for some unique $k = k(n)$, and then:

$$\begin{aligned} P[S_n = 0] &= \alpha_k P[S_n = 0 \mid S_{t_k} = 0] + (1 - \alpha_k) P[S_n = 0 \mid S_{t_k} \neq 0] \\ &\geqq \alpha_k P[S_n = 0 \mid S_{t_k} = 0] \geqq q_k(1 - 1/k) \to 1 \qquad \text{as } n \to \infty \end{aligned}$$

because $k = k(n)$ tends to ∞ with n and by (3) $\alpha_k \to 1$. Thus $S_n \to 0$ in probability.

3. Fair Markov-chains. The example of Section 2 as well as the examples alluded to at the end of Section 1, are all Markov processes with a discrete state space. None of them however has stationary transition probabilities. In this section we show that a discrete real-valued Markovian martingale with stationary transition probabilities, cannot have a limiting distribution with a mean (possibly infinite) unless it converges almost surely. To state our theorem formally, let I be a discrete set of real numbers (i.e., a set with no real limit points), and let M be a Markov matrix on $I \times I$ (i.e., $M(x, y) \geqq 0$ and $\sum_{y \in I} M(x, y) = 1$ for all $x, y \in I$). Call M *subfair* if $\sum_{y \in I} yM(x, y) \leqq x$ for all $x \in I$; call it *superfair* if the reverse inequalities are satisfied; M is said to be *fair* if it is both subfair and superfair. Let α be a probability measure (p.m.) on I, i.e., $\alpha = \{\alpha(x) : x \in I\}$, with $\alpha(x) \geqq 0$ and $\sum_{x \in I} \alpha(x) = 1$. α together with a Markov matrix M gives rise to an I-valued Markov chain $\{S_0, S_1, \cdots\}$ in the usual way, i.e., the formula

$$P\{S_0 = x_0, S_1 = x_1, \cdots, S_n = x_n\} = \alpha(x_0) \prod_{i=1}^{n} M(x_{i-1}, x_i)$$

determines a consistent system of finite dimensional joint distributions; αM^n (usual matrix product) is then the (marginal) distribution of S_n. It seems natural to denote such a process by (α, M). Say that (α, M) is fair if M is, and similarly for sub- and superfair. (Except perhaps for the condition of integrability, which is usually incorporated into the definition of a martingale, a fair (α, M) is a martingale, and likewise sub- and superfair (α, M)'s correspond to semi-martingales.)

It will be convenient to use operator notation for integrals, i.e., μf denotes the integral of the function f w.r.t. the measure μ. Thus for example,

$$(\alpha M^n)f = \sum_{x \in I} [f(x) \sum_{y \in I} \alpha(y) M^n(y, x)] .$$

Recall the definition of weak convergence of measures: $\{\alpha_n\}$ converges weakly to α (Notation: $\alpha_n \to_w \alpha$) iff $\alpha_n f \to \alpha f$ as $n \to \infty$ for every bounded function f on I (since I is discrete, every f on I is continuous). It is well known and easy to verify that if α_n and α are p.m.'s on I then $\alpha_n \to_w \alpha$ iff $\alpha_n(x) \to \alpha(x)$ uniformly in $x \in I$.

THEOREM. *Let (α, M) be a fair Markov chain with state space I. If (α, M) converges in distribution (i.e., the sequence (αM^n) of marginals converges weakly to a* p.m. *λ), then (α, M) converges almost surely, provided the limiting distribution λ has a mean (not necessarily finite). (Note that no moment conditions are imposed on the initial distribution α.)*

REMARK 1. The proviso that λ has a mean is essential for the validity of the theorem, as can be seen from the following example:

Let $I = \{0\} \cup \{\pm 2^n, n \geqq 1\}$; let $\alpha(0) = \frac{1}{3}$ and $\alpha(2^n) = 2^{-n}/3 = \alpha(-2^n)$ for $n = 1, 2, \cdots$. Note that α does not have a mean. Let $M(0, 2) = \frac{1}{2} = M(0, -2)$ and for each nonzero x in I let $M(x, 0) = \frac{1}{2} = M(x, 2x)$. It is then easy to check that the process (α, M) is fair, that $\alpha M = \alpha$, and therefore (α, M) has the limiting distribution $\lambda = \alpha$. Nevertheless (α, M) does not converge almost surely (and even not in probability).

PROOF OF THE THEOREM. We present the proof in the form of three lemmas.

LEMMA 1. *Under the conditions of the theorem, λ is invariant for M, i.e., $\lambda M = \lambda$.*

PROOF. It suffices to prove

$$(\lambda M)f = \lambda f \qquad \text{for all bounded } f \text{ on } I . \tag{4}$$

It is a matter of routine to check

$$(\alpha M^{n+1})f = \textstyle\sum_z (\alpha M^n)(z)(M(z, \cdot)f) \qquad \text{for every bounded function } f \text{ on } I . \tag{5}$$

(Recall that for each z, $M(z, \cdot)$ is a p.m. on I.) Let g be the function $z \to M(z, \cdot)f$, then g is bounded, so from (5) we get

$$(\alpha M^{n+1})f = (\alpha M^n)g . \tag{6}$$

By weak convergence, the left-hand side of (6) converges to λf whereas its right-hand side tends to λg, which is easily seen to equal $(\lambda M)f$. This establishes (4) and hence the lemma.

REMARK 2. Had I not been discrete, the boundedness of g would not have implied $(\alpha M^n)g \to \lambda g$, because g is not continuous unless the Markov operator $M(z, \cdot)$ varies continuously with z, a condition which is automatically satisfied in the discrete case. Indeed Lemma 1 is false in the non-discrete case, as is clear from the following example.

EXAMPLE. Take $S = \{0, 1, \frac{1}{2}, \frac{1}{3}, \cdots\}$ as the state space and define M by $M(0, 1) = 1$, $M(1/n, 1/(n + 1)) = 1$ for $n \geqq 1$. Let α be point-mass at 0. Then: $\alpha M^n \to_w \alpha$ (in a trivial deterministic way), but α is not M-invariant. The trouble here is that $M(s, \cdot)$ is not continuous at $s = 0$ with respect to the usual topology on S.

LEMMA 2. *Under the conditions of the theorem, every point x in the support of λ (i.e., such that $\lambda(x) > 0$) is an absorbing state of the process, i.e., $M(x, x) = 1$.*

PROOF. By fairness of M and by the preceding lemma, the process (λ, M) is a stationary generalized semi-martingale (if the mean of λ is finite, it is in fact a stationary martingale). The lemma now follows from Corollary 2 of Gilat (1971).

LEMMA 3. *For every initial distribution α, the process (α, M) eventually gets to the support of λ, with probability* 1.

PROOF. Let S be the support of λ. Suppose that

$$\text{Prob}\{(\alpha, M) \text{ never gets to } S\} = q > 0 . \qquad \text{Then}$$

$$1 - q \geqq \text{Prob}\{(\alpha, M) \text{ is in } S \text{ at time } n\} = (\alpha M^n)(S) \to \lambda(S) = 1 .$$

This is a contradiction which establishes the lemma and thereby concludes the proof of the theorem.

REMARK 3. Lemma 3 is false when the state space is not discrete, as can be seen from the example succeeding Remark 2. We use discreteness in our proof of the lemma to conclude $(\alpha M^n)(S) \to \lambda(S)$ from $\alpha M^n \to_w \lambda$, while in general we could conclude only $\limsup(\alpha M^n)(S) \leqq \lambda(S)$ (because the support S of a measure is, by definition, a closed set). This would not lead to any contradiction in the argument. Since in the discrete case S is also open (as is every set) we are lucky to also have the dual relation $\liminf(\alpha M^n)(S) \geqq \lambda(S)$.

REMARK 4. The theorem, with essentially the same proof, remains valid if in its statement "fair" is replaced by "superfair" ("subfair"), provided we specify that the positive (negative) part of λ has a finite mean.

REMARK 5. Although, as indicated in Remarks 2 and 3, the proof presented above hinges heavily on the discreteness of the process, I have originally conjectured that the theorem remains valid for general real-valued fair Markov chains. It was recently pointed out to me by Burgess Davis that this is not the case, even when the Markov operator $M(x, \cdot)$ varies continuously in x.

REFERENCES

[1] GILAT, D. (1970). On the convergence of random series, Part I of Ph. D. dissertation, Univ. of California, Berkeley.

[2] GILAT, D. (1971). Some conditions under which two random variables are equal almost surely and a simple proof of a theorem of Chung and Fuchs. *Ann. Math. Statist.* **42** 1647–1655.

[3] LÉVY, P. (1937). *Theorie de l'addition des Variables Aleatoires*. Gauthier-Villars, Paris.

[4] LOÈVE, M. (1964). On stochastic processes. Mimeographed notes of a talk given at George Washington University.