Z. Wahrscheinlichkeitstheorie 1, 288—294 (1963)

A Ratio Operator Limit Theorem

By

J. L. Doob

1. Introduction

Let (X_0, μ_0) and (X_1, μ_1) be measure spaces. We omit reference to the classes of measurable sets, and all sets used below are assumed measurable even when the assumption is not made explicitly. Let T be a linear transformation from $L_1(X_0, \mu_0)$ into $L_1(X_1, \mu_1)$ which is positive (that is, takes positive functions into positive functions) and has norm $\leqq 1$. There is then an adjoint transformation T^* from $L_\infty(X_1, \mu_1)$ into $L_\infty(X_0, \mu_0)$, also positive, with L_∞ norm $\leqq 1$. If μ_i is not σ-finite, $L_\infty(X_i, \mu_i)$ has the usual supremum norm but by definition each of its functions vanishes off a set, depending on the function, which is the union of countably many sets of finite measure. The adjoint T^* is determined by

$$\int_{X_1} (T f_0) g_1 \, d\mu_1 = \int_{X_2} f_0 \, T^* g_1 \, d\mu_0 . \tag{1.1}$$

It follows that $T^* 1 \leqq 1$ a.e. on X_0. In particular, suppose that the measures are finite and that $T1 \leqq 1$ a.e. The latter condition is equivalent to the condition that T^* does not increase L_1 norms. It follows that T^* has a unique linear extension taking $L_1(X_1, \mu_1)$ into $L_1(X_0, \mu_0)$ with L_1 norm $\leqq 1$. The extended transformation will also be called T^*, and the relation between T and T^* is now symmetric. Finally, if $T1 = 1$ a.e. and if $T^*1 = 1$ a.e. T will be called 'bistochastic' following Rota [*4*] (who however assumed that the two measure spaces were the same). The transformation T is bistochastic if and only if T^* is.

Throughout this paper, (X_0, μ_0), (X_1, μ_1), ... are measure spaces, T_n is a positive linear transformation from $L_1(X_{n-1}, \mu_{n-1})$ into $L_1(X_n, \mu_n)$, of L_1 norm $\leqq 1$, and $T_{1n} = T_n \ldots T_1$, so that $T_{1n}^* = T_1^* \ldots T_n^*$. The following theorem is due to Rota, aside from certain specializations he made that were not needed in either his discussion or proof.

Theorem 1.1. *If each T_n is bistochastic and if f_0 is a function on X_0 satisfying*

$$\int_{X_0} |f_0| \log^+ |f_0| \, d\mu_0 < \infty$$

then $\lim\limits_{n\to\infty} T_{1n}^* T_{1n} f_0$ exists a.e. (μ_0) and in the $L_1(X_0, \mu_0)$ topology.

Burkholder [*2*] has given an example showing that the theorem is false if f is only supposed in $L_1(X_0, \mu_0)$.

The purpose of this paper is to give a simplified approach to Rota's method which makes its relation to standard probability reasoning clearer, and to extend the theorem to non bistochastic operators. In this extension the theorem becomes a ratio theorem.

2. The conditional expectation technique

Rota's method and that used in this paper rest on the following simple remark, and the only question is how to exploit the remark most advantageously. Let $x_0, x_1, \ldots$ be a Markov process on some probability measure space. Then if f is integrable on the x_0 space and if the usual probability notation is used,

$$E\{f(x_0)\,|\,x_n, x_{n+1}, \ldots\} = E\{f(x_0)\,|\,x_n\}, \tag{2.1}$$

because the reversed x_n sequence is also a Markov process. Moreover the conditional expectation is with respect to less and less as n increases, so the sequence of conditional expectations ordered as n decreases is a martingale. It follows that the sequence of conditional expectations converges a.e. and in L_1,

$$\lim_{n\to\infty} E\{f(x_0)\,|\,x_n\} = E\{f(x_0)\,|\,\mathscr{F}\}, \tag{2.2}$$

where $\mathscr{F}$ is the tail field of the x_n sequence. Now it is known [3] that one can take limits under the conditional expectation symbol to get

$$\lim_{n\to\infty} E\{E\{f(x_0)\,|\,x_n\}\,|\,x_0\} = E\{E\{f(x_0)\,|\,\mathscr{F}\}\,|\,x_0\} \tag{2.3}$$

a.e. and in the L_1 topology if the Lebesgue dominated convergence criterion

$$E\{\sup_n |E\{f(x_0)\,|\,x_n\}|\} < \infty \tag{2.4}$$

is satisfied. (Actually the condition is now known to be necessary in general, according to a theorem of Blackwell and Dubins [1].) Moreover it is known [3] from martingale theory that (2.4) is true if $E\{|f(x_0)|\log^+|f(x_0)|\} < \infty$. Thus (2.3) is true under this restriction on f. If the conditional expectation in (2.3) can be identified with $T_{1n}^* T_{1n} f$, Theorem 1.1 follows. Rota's theorem will be proved applying the foregoing remark, and then the generalized ratio limit theorem for non bistochastic operators will be proved by reducing it to Rota's theorem.

3. A representation theorem

The following remarks are not original, but there seems to be no single place in the literature where they are readily available.

Let $\hat{X}$ be a totally disconnected Hausdorff space, that is, it is supposed that the class of clopen (simultaneously closed and open) sets is a base for the topology of the space. Let $\hat{G}$ be the Borel field of sets generated by the clopen sets. We suppose from now on that $\hat{X}$ is compact. The indicator function of a clopen set is continuous. The class of uniform limits of finite linear combinations of indicator functions of clopen sets is a closed (supremum norm) algebra of continuous functions including the constant functions and separating the points of $\hat{X}$. The class is therefore $C(\hat{X})$, the class of continuous functions on $\hat{X}$, according to the Stone-Weierstrass theorem. The sets in $\hat{G}$ will be called the Baire sets.

Let X be a space on a Borel field of whose subsets a finite-valued measure μ is defined. The Boolean algebra of measurable subsets of X, modulo sets of measure 0, is isomorphic to the algebra of clopen sets of a totally disconnected compact Hausdorff space $\hat{X}$, according to the Stone representation theorem. The measure μ is transformed into a positive finitely additive function $\hat{\mu}$ of clopen sets. If the algebra of X subsets modulo sets of measure 0 and the algebra of clopen subsets of $\hat{X}$ are metrized as usual (the distance between two sets is the measure of their symmetric difference), the first metric space is complete, so the second one is also.

Since a clopen set can be expressed as a countable union of disjunct clopen sets only if all but a finite number of the summands are empty, $\hat{\mu}$ is countably additive. Hence $\hat{\mu}$ can be extended to a measure (also denoted by $\hat{\mu}$) of Baire sets. A non-empty open Baire set has strictly positive measure. Because of the completeness remark at the end of the preceding paragraph, if $\hat{A}$ is a Baire set there is a unique clopen set at zero distance from $\hat{A}$. More generally, to any bounded function measurable with respect to the Baire sets corresponds a continuous function equal to it almost everywhere.

We have thus obtained a $1-1$ measure preserving correspondence between subsets of X and Baire subsets of $\hat{X}$, modulo sets of measure 0, preserving the Boolean operations, in which we can take a clopen set as representative of any Baire set. If f is a function on X, taking only finitely many values, $f=\sum_j a_j I_{A_j}$ where I_A is the indicator function of the set A, define $\hat{f}$ as the corresponding sum in which A_j is replaced by its image clopen set. Then $\hat{f}$ is continuous. If f is a bounded function on X it is the uniform limit of a sequence of functions of this type, and proceeding in this way we obtain a $1-1$ correspondence between $L_\infty(X, \mu)$ and $C(X)$ or, and this amounts to the same thing in the present case, between $L_\infty(X, \mu)$ and $L_\infty(\hat{X}, \hat{\mu})$. Moreover the respective norms are preserved. *If f_j is bounded and has image $\hat{f}_j$, $1 \leqq j \leqq n$, and if Φ is a bounded continuous function from Euclidean n-space to the reals, then $\Phi(f_1, \ldots, f_n)$ has image $\Phi(\hat{f}_1, \ldots, \hat{f}_n)$, and the composite function has the same integral as its image.* These facts are true for f_j taking only finitely many values and therefore as stated. We conclude that $f_1, \ldots, f_n$ and its image n-tuple have the same joint distribution.

In the map from $L_\infty(X, \mu)$ onto $L_\infty(\hat{X}, \hat{\mu})$ L_1 norms are preserved. Hence the map can be extended to one from $L_1(X, \mu)$ onto $L_1(\hat{X}, \hat{\mu})$. As such it preserves L_1 norms and the above italicized statement remains true for $f_j \in L_1(X, \mu)$.

4. Proof of Rota's theorem

Under the hypotheses of Theorem 1.1, $\mu(X_j)$ does not change with j and is finite, so we can normalize the measures to make them all 1. In proving the theorem suppose first that T_n^* is determined by a stochastic transition function. That is, suppose that there is a function Q_n of ξ in X_{n-1} and $A \subset X_n$, with the following properties: $Q_n(., A)$ is measurable for fixed A; $Q_n(\xi, .)$ is a probability measure for each ξ;

$$(4.1) \qquad (T_n^* f)(\xi) = \int_{X_n} f(\eta)\, Q_n(\xi, d\eta) \text{ a.e. } (\mu_{n-1}).$$

Let $\Omega = X_0 \times X_1 \times \cdots$ and let x_n be the nth coordinate function on the product space. Define measure on the product space to make the x_n sequence a Markov process with initial measure μ_0 and transition measures $Q_1, Q_2, \dots$, so that

$$(4.2)\qquad E\{\Phi(x_0,\dots,x_n)\} = \int_{X_0} \mu_0(d\xi_0) \int_{X_1} Q_1(\xi_0, d\xi_1) \cdots \int_{X_{n-1}} Q_{n-1}(\xi_{n-2}, d\xi_{n-1}) \cdot \int_{X_n} \Phi(\xi_1,\dots,\xi_n)\, Q_n(\xi_{n-1}, d\xi_n).$$

(See [3].) Then applying (4.1) it is trivial that if g is a bounded function on X_n

$$(4.3)\qquad (T_{1n}^* g)(x_0) = E\{g(x_n) \mid x_0\} \text{ a.e.}$$

By definition, the distribution of x_0 is μ_0. It follows from (4.3) that the distribution of x_n is μ_n, because if $g = I_A$ is the indicator function of a set $A \subset X_n$, and if we use P for measure on the product space,

$$(4.4)\qquad P\{x_n(\omega) \in A\} = \int_{X_0} E\{I_A(x_n) \mid x_0\}\, dP = \int_{X_0} T_{1n}^* I_A\, d\mu_0 = \int_{X_n} I_A\, T_{1n} 1\, d\mu_n = \mu_n(A).$$

Then more generally, we conclude that if $g \in L_1(X_n, \mu_n)$, $E\{g(x_n)\} = \int_{X_n} g\, d\mu_n$ is finite and (4.3) is true. Finally if $f \in L_1(X_0, \mu_0)$ and if $A \subset X_n$,

$$(4.5)\qquad E\{f(x_0)\, I_A(x_n)\} = E\{E\{f(x_0)\, I_A(x_n) \mid x_0\}\} = \int_{X_0} f\, T_{1n}^* I_A\, d\mu_0 = \int_A T_{1n} f\, d\mu_n = E\{(T_{1n} f)(x_n)\, I_A(x_n)\}$$

so that, by definition of conditional expectation,

$$(4.6)\qquad (T_{1n} f)(x_n) = E\{f(x_0) \mid x_n\} \quad \text{a.e.}$$

Thus Theorem 1.1 can be deduced as in Section 2: the desired limit relation has been reduced to (2.3).

We must still show that the hypothesis that T_n^* is given by a stochastic transition function can be eliminated. Map functions on X_n into functions on a Hausdorff space as described in Section 3. We define $\hat{T}_n$ and $\hat{T}_n^*$ for the new spaces, defining the latter transformation first, as follows. If the clopen subset $\hat{A}$ of $\hat{X}_n$ is the image of the subset A of X_n, and if I_A is the indicator function of A, define $\hat{Q}_n(.\,, \hat{A})$ as the unique continuous image of $T_n^* I_A$. Then $\hat{Q}_n(\xi, .)$ is an additive function of clopen sets and can be extended to be a measure of Baire sets. If $\hat{\mu}_n(\hat{A}) = 0$, $\hat{Q}_n(.\,, \hat{A}) = 0$ a.e. on $\hat{X}_{n-1}$ because for every $\varepsilon > 0$ $\hat{A}$ can be covered by a union $\bigcup_k \hat{A}_k$ of clopen sets with $\sum_k \hat{\mu}_n(\hat{A}_k) < \varepsilon$, and if $A_k \subset X_{n-1}$ is the image of $\hat{A}_k$,

$$\int_{\hat{X}_{n-1}} \hat{Q}_n(\xi, \hat{A})\, \mu_{n-1}(d\xi) \leqq \sum_k \int_{\hat{X}_{n-1}} \hat{Q}_n(\xi, \hat{A}_k)\, \mu_{n-1}(d\xi) = \sum_k \int_{X_{n-1}} T_n^* I_{A_k} \mu_{n-1}(d\xi) = \sum_k \mu_n(A_k) = \sum_k \hat{\mu}_n(\hat{A}_k) < \varepsilon.$$

Define $\hat{T}_n^* \hat{f}$ for $\hat{f}$ in $L_1(\hat{X}_n, \hat{\mu}_n)$ by

$$(4.7)\qquad \hat{T}_n^* \hat{f} = \int_{\hat{X}_n} \hat{f}(\eta)\, \hat{Q}_n(\xi, d\eta).$$

In justification of this definition first note that if the integral is well-defined for almost all ξ, for two integrands which are equal a.e. then the integrals are themselves equal a.e., according to the property of $\hat{Q}_n$ just proved. Hence the proper measure 0 ambiguities are preserved. Furthermore if $\hat{f}$ is the image of f and $(T_n^* f)^\wedge$ that of $T_n^* f$ then $\hat{T}_n^* \hat{f} = (T_n^* f)^\wedge$ when $\hat{f}$ is the indicator function of a clopen set and therefore successively when $\hat{f}$ is continuous, bounded, in the class $L_1(\hat{X}_n, \hat{\mu}_n)$. The transformation $\hat{T}_n^*$ is bistochastic. Let $\hat{T}_n$ be its adjoint and $\hat{T}_{1n} = \hat{T}_n \ldots \hat{T}_1$. Then $\hat{T}_n \hat{f} = (T_n f)^\wedge$, and the sequences $\{T_{1n}^* T_{1n} f, n \geqq 1\}$, $\{\hat{T}_{1n}^* \hat{T}_{1n} \hat{f}, n \geqq 1\}$ correspond to each other in the mapping. Hence these sequences have the same joint distributions. Thus instead of proving convergence of the first sequence it is sufficient to prove convergence of the second, and this is the case already treated, in which the x-operators are given by transition functions.

5. The case $T_n 1 = 1$, $T_n^* 1 \leqq 1$

In this section we suppose, as a first step towards our final generalization of Rota's theorem, that $\mu_n(X_n) < \infty$ for all n, that T_n is a positive linear transformation from $L_1(X_{n-1}, \mu_{n-1})$ into $L_1(X_n, \mu_n)$ and that $T_n 1 = 1$, $T_n^* 1 \leqq 1$ a.e. (See the discussion in Section 1.) Under these hypotheses, and making the extensions described in Section 1, T_n^* as a transformation of L_1 spaces is integral preserving and so has norm 1.

We shall use the inequalities

$$T_n^* 1 = T_{1n-1}^* T_n^* 1 \leqq T_{1n-1}^* 1 \text{ a.e.}, \quad \mu_n(X_n) = \int_{X_{n-1}} T_{n-1}^* 1 \, d\mu_{n-1} \leqq \mu_{n-1}(X_n). \tag{5.1}$$

Adjoin a new element ϱ_n to X_n for $n \geqq 0$. The measurable subsets of the enlarged space X_n' are to be the measurable subsets of X_n with or without ϱ_n and we define $\mu_n(\varrho_n) = \mu_0(X_0) - \mu_n(X_n)$. Define S_n, S_n^* by

$$\begin{aligned} S_n f &= T_n f \quad \text{on} \quad X_n \\ &= \frac{\int_{X_{n-1}} (1 - T_n^* 1) f \, d\mu_{n-1} + f(\varrho_{n-1}) [\mu_0(X_0) - \mu_{n-1}(X_{n-1})]}{\mu_0(X_0) - \mu_n(X_n)} \quad \text{at } \varrho_n \end{aligned} \tag{5.2}$$

where $f \in L_1(X_{n-1}', \mu_{n-1})$,

$$\begin{aligned} S_n^* g &= g(\varrho_n)(1 - T_n^* 1) + T_n^* g \quad \text{on} \quad X_{n-1} \\ &= g(\varrho_n) \qquad\qquad\qquad\qquad \text{at} \quad \varrho_{n-1} \end{aligned} \qquad f \in L_\infty(X_n', \mu_n). \tag{5.3}$$

Here $T_n f$ and $T_n^* g$ mean the application of the indicated transformations to the restrictions of f and g to the unenlarged spaces. If the denominator in the second line of (5.2) vanishes, the line may be omitted, because ϱ_n then has measure 0. The transformations S_n and S_n^* are bistochastic and adjoint to each other. We write S_{1n} for $S_n \ldots S_1$ and find that

$$S_{1n}^* S_{1n} f = \frac{\int_{X_0} (1 - T_{1n}^* 1) f \, d\mu_0}{\mu_0(X_0) - \mu_n(X_n)} (1 - T_{1n}^* 1) + T_{1n}^* T_{1n} f \quad \text{on} \quad X_0, \tag{5.4}$$

where the first term on the right is omitted if the denominator vanishes. By Rota's theorem the left side of (5.4) converges almost everywhere and in the L_1

topology when $n \to \infty$. In view of (5.1), the first term on the right in (5.4) converges in the same way. Hence we have obtained the following useful but rather trivial extension of ROTA's theorem, which we shall use in the next section.

Theorem 5.1. *Theorem 1.1 remains true if T_n is not necessarily bistochastic, but if T_n is positive-linear from $L_1(X_{n-1}, \mu_{n-1})$ into $L_1(X_n, \mu_n)$, of norm $\leqq 1$ and with $T_n 1 = 1$ a.e.*

6. The general theorem: the case $T_n^* 1 \leqq 1$

In this section we no longer suppose that μ_n is a finite measure. We assume that T_n is a positive linear operator from $L_1(X_{n-1}, \mu_{n-1})$ into $L_1(X_n, \mu_n)$ of norm $\leqq 1$, that is with $T_n^* 1 \leqq 1$ a.e. Let $h \geqq 0$ be a function in $L_1(X_0, \mu_0)$ and define $h_n = T_{1n} h$, $d\mu_n' = h_n \, d\mu_n$. The relation

$$\int_{X_{n-1}} (T_n^* f) h_{n-1} \, d\mu_{n-1} = \int_{X_n} f h_n \, d\mu_n, \quad f \in L_\infty(X_n, \mu_n) \tag{6.1}$$

shows that, in terms of the primed measures, T_n^* preserves integrals. It follows that $T_n^* f = 0$ a.e. (μ_{n-1}') if f vanishes on the set where h_n is strictly positive. That is, T_n^* defines a positive linear transformation $T_n^{*\prime}$ from $L_\infty(X_n, \mu_n')$ into $L_\infty(X_{n-1}, \mu_{n-1}')$ by the rule $T_n^{*\prime} f = T_n^* f$; the measure zero ambiguities match as they should. Define

$$T_n' f = \frac{T_n(f h_{n-1})}{h_n}, \quad f \in L_1(X_{n-1}, \mu_{n-1}'), \quad h_0 = h, \tag{6.2}$$

where the quotient is defined arbitrarily on the set where the denominator vanishes. Note that $T_n(f h_{n-1})$ vanishes a.e. (μ_n) where $h_n = 0$. This fact is obviously true if $f = 1$, and is therefore true if f is bounded, and hence if $f h_{n-1} \in L_1(X_{n-1}, \mu_{n-1})$, that is if $f \in L_1(X_{n-1}, \mu_{n-1}')$. The transformation T_n' is linear, positive, of L_1 norm $\leqq 1$, and $T_n' 1 = 1$ a.e. Moreover the adjoint transformation is $T_n'^* = T_n^{*\prime}$. The transformation $T_{1n}' = T_n' \ldots T_1'$ is well-defined, and $T_{1n}' f$ is determined by the restriction of f to the set of strict positivity of h. Moreover $T_{1n}' f = [T_{1n}(fh)]/h_n$ so that

$$T_{1n}'^* T_{1n}' f = T_{1n}'^* \left(\frac{T_{1n}(fh)}{h_n} \right). \tag{6.3}$$

An application of Theorem 5.1 now yields the main theorem of this paper.

Theorem 6.1. *Let $(X_0, \mu_0), (X_1, \mu_1), \ldots$ be measure spaces and let $h \in L_1(X_0, \mu_0)$, $h \geqq 0$. Let T_n be a positive linear transformation from $L_1(X_{n-1}, \mu_{n-1})$ into $L_1(X_n, \mu_n)$ of norm $\leqq 1$, that is with $T_n^* 1 \leqq 1$ a.e. Then if $d\mu_n' = T_{1n} h \, d\mu_n$, and if f is a function on X_0 satisfying*

$$\int_{X_0} |f| \log^+ |f| \, h \, d\mu_0 < \infty,$$

it follows that

$$\lim_{n \to \infty} T_{1n}'^* \left(\frac{T_{1n}(fh)}{T_{1n} h} \right) \tag{6.4}$$

exists a.e. (μ_0) where $h > 0$ and also in the $L_1(X_0, h \, d\mu_0)$ topology. Here $T_{1n}'^*$ is the linear extension of L_1 norm $\leqq 1$ of T_{1n}^* to $L_1(X_n, \mu_n')$.

We note that the numerator in (6.4) vanishes a.e. where the denominator does, and the values on this set do not affect $T_{1n}^{\prime *}$. In particular, if $h = 1$ and $T_n 1 = 1$ a.e. the theorem reduces to Theorem 5.1.

References

[1] Blackwell, D., and L. E. Dubins: A converse to the dominated convergence theorem. Ill. J. Math. (to appear).

[2] Burkholder, D. L.: Successive conditional expectations of an integrable function. Ann. math. Statistics **33**, 887—893 (1962).

[3] Doob, J. L.: Stochastic processes. New York: Wiley 1952.

[4] Rota, G. C.: An „Alternierende Verfahren" for general positive operators. Bull. Amer. math Soc. **68**, 95—102 (1962).

Department of Mathematics
University of Illinois
Urbana, Illinois (USA)

(Received October 8, 1962)