SUCCESSIVE CONDITIONAL EXPECTATIONS OF AN INTEGRABLE FUNCTION[1]

By D. L. Burkholder

University of Illinois and University of California, Berkeley

1. Introduction. Rota [5] has shown recently that if $\{T_n\}$ is a sequence of conditional expectation operators,

$$S_n = T_0 T_1 \cdots T_{n-1} T_n T_{n-1} \cdots T_1 T_0, \tag{1}$$

and X is a random variable such that[2]

$$E(|X| \log^+ |X|) < \infty, \tag{2}$$

then the sequence $\{S_n X\}$ converges almost everywhere to an integrable function.[3]

Here an example is given showing that this result fails to hold if (2) is replaced by $E|X| < \infty$. Moreover, it is shown that (2) is a necessary condition for the almost everywhere convergence of $\{S_n X\}$ for every $\{T_n\}$ if the underlying probability space $(\Omega, \mathcal{A}, P)$ is rich enough to support independent identically distributed random variables $X_1, X_2, \cdots$ such that $X_1 = X$.

The idea behind our approach has several other applications. One is that (2) is a necessary condition for

$$E(\sup_n |X_1 + \cdots + X_n|/n) < \infty,$$

where $X_1, X_2, \cdots$ are as above. (Sufficiency is well known.) Another is a variation on the strong law of large numbers.

The example has the following feature: The sequence of conditional expectations $T_1, T_2, \cdots$ is defined by a decreasing sequence of sub-σ-fields, hence the operator S_n of (1) satisfies $S_n = T_0 T_n T_0$.

2. Example. Let X be an integrable random variable and $\{a_n\}$ a real number sequence such that

$$\sum_{n=1}^{\infty} P(X > a_n) = \infty, \tag{3}$$

Received April 30, 1962.

[1] This research was supported by the National Science Foundation under Grant No. G19667.

[2] For real x, $\log^+ x = \log x$ if $x > 1$, $= 0$ otherwise.

[3] This significantly extends an earlier result of Chow and the present author [2]. In his paper, Rota assumes slightly more than (2), namely that $E|X|^p < \infty$ for some $p > 1$. However, as he has observed elsewhere (at the Symposium on Ergodic Theory, Tulane University, October, 1961) and as is clear from his method of proof, the condition (2) is sufficient. Rota also proves a similar proposition for more general operators by reducing it to a question about successive conditional expectations. Other closely related results have been obtained by E. M. Stein [6].

(4) $$\lim_{n\to\infty} E(X \mid X > a_n)/n = \infty.$$

Such exist as we show later. (For measurable sets A, $E(X \mid A) = \int_A X dP/P(A)$ if $P(A) > 0$ and, for definiteness, let us suppose that $E(X \mid A) = EX$ otherwise.) Let $X_1, X_2, \cdots$ be independent random variables, each with the same distribution as X. Let $Y_1 = X_1$ and if $n > 1$, let $Y_n = E(X \mid X > a_n)$ on the set

$$\{\omega \mid X_n(\omega) > a_n\},$$

$= E(X \mid X \leqq a_n)$ on the set $\{\omega \mid X_n(\omega) \leqq a_n\}$. Let $T_0 = E(\cdot \mid Y_1, Y_2, \cdots)$ and if $n > 0$, let $T_n = E(\cdot \mid X_1 + \cdots + X_k, k \geqq n)$ and S_n be defined by (1). Then, almost everywhere

$$S_n X_1 = T_0 T_n T_0 X_1 = T_0 T_n X_1 = T_0(X_1 + \cdots + X_n)/n$$
$$= (E(X_1 \mid Y_1) + \cdots + E(X_n \mid Y_n))/n = (Y_1 + \cdots + Y_n)/n.$$

By Borel-Cantelli, the definition of Y_n, and (3), we have that almost everywhere $Y_n = E(X \mid X > a_n)$ for infinitely many positive integers n. Therefore, lim $\sup_{n\to\infty} Y_n/n = \infty$ almost everywhere, using (4). If X is nonnegative, then almost everywhere $Y_1 + \cdots + Y_n \geqq Y_n$ and $\limsup_{n\to\infty} S_n X_1 = \infty$. Otherwise, one may proceed as follows: The condition (4) implies that $a_n \to \infty$, which implies that $E(X \mid X \leqq a_n) \to EX$, which in turn implies that there is a real number c such that $Y_n \geqq c$, $n = 2, 3, \cdots$. Therefore, almost everywhere

$$\limsup_{n\to\infty} S_n X_1 = \limsup_{n\to\infty}((Y_2 - c) + \cdots + (Y_n - c))/n + c$$
$$\geqq \limsup_{n\to\infty}(Y_n - c)/n + c = \infty.$$

Here, in short, the desired type of convergence fails to hold.

We now show that X and $\{a_n\}$ satisfying (3) and (4) exist. Let X be a random variable with the density f, relative to Lebesgue measure, where

$$f(x) = 0 \text{ if } x \leqq 4,$$
$$= c/(x^2 \log x \log^2 \log x) \quad \text{if} \quad x > 4,$$

where c satisfies $\int_{-\infty}^{\infty} f(x)\,dx = 1$. Let $\{a_n\}$ be any real number sequence such that $P(X > a_n) = c/(n \log n)$ for all large n. Then X is integrable, (3) is satisfied, and $a_n \to \infty$. The relation (4) also holds since for all large n, $\log \log a_n > 1$,

$$a_n < E(X \mid X > a_n) = \frac{n \log n}{c} \int_{a_n}^{\infty} x f(x)\,dx = \frac{n \log n}{\log \log a_n} < n^2,$$
$$\log \log a_n < \log \log n^2,$$
$$E(X \mid X > a_n)/n = \frac{\log n}{\log \log a_n} > \frac{\log n^2}{2 \log \log n^2}.$$

3. A more precise result.

THEOREM 1. *Let X be an integrable random variable and $X_1, X_2, \cdots$ independent random variables each with the same distribution as X. The following statements are equivalent:*

(i) $E(X \log^+ X) = \infty$.
(ii) $E(\sup_n X_n/n) = \infty$.
(iii) $E(\sup_n(X_1 + \cdots + X_n)/n) = \infty$.
(iv) *There is a sequence* $\{T_n\}$ *of conditional expectation operators such that* $\limsup_{n\to\infty} S_n X_1 = \infty$ *almost everywhere, where* S_n *is defined by* (1).

REMARKS. Since

$$E(|X| \log^+ |X|) = E(X \log^+ X) + E((-X) \log^+ (-X)),$$

at least one of the terms on the right is infinite if the left side is infinite. Thus, if (2) does not hold and $X_1, X_2, \cdots$ are independent random variables each with the same distribution as X, we have by Theorem 1 that

$$E(\sup_n |X_1 + \cdots + X_n|/n) = \infty,$$

there is a sequence $\{T_n\}$ of conditional expectation operators such that $\limsup_{n\to\infty} |S_n X_1| = \infty$ almost everywhere, and so forth.

Blackwell and Dubins [1] have proved recently, along with other results, that if X is a nonnegative integrable random variable such that $E(X \log^+ X) = \infty$, then there is defined on some probability space a random variable X_1 with the same distribution as X and a decreasing sequence $\{\mathcal{A}_n\}$ of sub-σ-fields such that $\sup_n E(X_1 \mid \mathcal{A}_n)$ is not integrable. The implication (i) $\Rightarrow$ (iii), in the above theorem, implies this result since almost everywhere

$$E(X_1 \mid X_1 + \cdots + X_k, k \geqq n) = (X_1 + \cdots + X_n)/n, \qquad n = 1, 2, \cdots.$$

The following result is used in the proof of Theorem 1.

LEMMA 1. *Let X be an integrable random variable and $\psi(a) = E(X \mid X > a)$ for real a. Consider the following statements*:

(i) $E(X \log^+ X) = \infty$.
(ii) $E\psi(X) = \infty$.
(iii) *There is a sequence* $\{a_n\}$ *satisfying* (3) *and* (4).

The implication (i) $\Rightarrow$ (ii) *holds. Also,* (ii) $\Rightarrow$ (iii) *under the further condition that the distribution function of X is continuous.*

REMARKS. There is an integrable random variable X such that $E(X \log^+ X) = \infty$ and no sequence $\{a_n\}$ of the desired type exists. That is, (i) does not imply (iii) without some further condition. On the other hand, (iii) $\Rightarrow$ (i), which follows from Section 2 and the implication (iv) $\Rightarrow$ (i) in Theorem 1. The implication sign in (i) $\Rightarrow$ (ii) cannot be reversed. Let $\varphi(a) = E(X \mid X \geqq a)$ for real a. Then the statement $E\varphi(X) = \infty$ is equivalent to (iii). Under the further condition that the distribution function of X is continuous, $\varphi = \psi$ and all four statements are equivalent. Since none of these remarks are used in the following, their proofs are omitted.

The nonequivalence of (i) and (iii) makes the proof of Theorem 1 slightly more complicated than it otherwise would be but has no effect on the content of the theorem itself.

PROOF OF LEMMA 1. (i) $\Rightarrow$ (ii): To prove this we may suppose that X is nonnegative: Letting $X^+ = \max\{0, X\}$ we have that $E(X\log^+ X) = E(X^+\log^+ X^+)$ and $E\psi(X) \geqq (EX - \psi(0))P(X \leqq 0) + E\psi(X^+)$, using the fact that ψ is bounded from below by EX. Also, we may replace X by any other random variable with the same distribution. Therefore, we may and do assume here that our probability space is the open interval $(0, 1)$ under Lebesgue measure and that our random variable X is a nonnegative nonincreasing right-continuous function on $(0, 1)$. Let $Y(\omega) = \int_0^\omega X(t)\,dt/\omega$ for $0 < \omega < 1$ and $X^{-1}(a) = \inf\{\omega \mid X(\omega) \leqq a\}$ for all real a for which the set is nonempty. Then Y is nonincreasing, $X^{-1}(X(\omega)) \leqq \omega$, and $\{\omega \mid X(\omega) > a\} = \{\omega \mid X^{-1}(a) > \omega\}$. By (i), which we assume, X is unbounded and therefore $X^{-1}(a) > 0$. Thus, $\psi(a) = Y(X^{-1}(a))$, $\psi(X(\omega)) = Y(X^{-1}(X(\omega))) \geqq Y(\omega)$, and $E\psi(X) \geqq EY$. By Hardy and Littlewood ([4], Theorem 11), (i) implies here that $EY = \infty$. Thus, (ii) holds.

We now prove that (ii) $\Rightarrow$ (iii) under the further condition that the distribution function of X is continuous. Suppose that $E\psi(X) = \infty$. Then

$$\sum_{k=1}^{\infty} P(\psi(X) \geqq k) = \infty, \tag{5}$$

since, letting $Y = \max\{0, \psi(X)\}$, we have that

$$E\psi(X) \leqq EY = \int_\Omega \int_0^Y dt\,dP = \int_0^\infty P(Y \geqq t)\,dt \leqq \sum_{k=0}^{\infty} P(Y \geqq k).$$

There is a positive integer sequence $\{k_n\}$ such that

$$\sum_{n=1}^{\infty} P(\psi(X) \geqq k_n) = \infty, \tag{6}$$

$$\lim_{n\to\infty} k_n/n = \infty, \tag{7}$$

as we show below. The function ψ is unbounded since $E\psi(X) = \infty$. Also, ψ is nondecreasing, right-continuous, and $\psi(a) \to EX$ as $a \to -\infty$. Therefore, for $k > EX$, $\psi^{-1}(k) = \inf\{x \mid \psi(x) \geqq k\}$ satisfies $\psi(\psi^{-1}(k)) \geqq k$ and $\{x \mid \psi(x) \geqq k\} = \{x \mid x \geqq \psi^{-1}(k)\}$. Letting $a_n = \psi^{-1}(k_n)$, where we may suppose that $k_n > EX$, we have that $\psi(a_n)/n = \psi(\psi^{-1}(k_n))/n \geqq k_n/n$ and $P(\psi(X) \geqq k_n) = P(X \geqq a_n)$. Now suppose that the distribution function of X is continuous. Then $P(X \geqq a_n) = P(X > a_n)$ and $\{a_n\}$ is a sequence of the desired type by (6) and (7).

To prove the existence of $\{k_n\}$ satisfying (6) and (7), let $c(k) = P(\psi(X) \geqq k)$. By (5) and the monotonicity of c,

$$\sum_{s=1}^{\infty} c(rs) = \infty, \qquad r = 1, 2, \cdots.$$

Therefore, positive integer sequences $\{r_n\}$ and $\{s_n\}$ exist satisfying $s_1 = 1$,

$$r_n > n(t_n + 1), \qquad \sum_{s=1}^{s_{n+1}} c(r_n s) > n, \tag{8}$$

where $t_n = s_1 + \cdots + s_n$, $n = 1, 2, \cdots$. Let $k_1 = 1$ and

$$k_{t_n+s} = r_n s,\ s = 1, \cdots, s_{n+1}, \qquad n = 1, 2, \cdots.$$

From (8) it follows that the sequence $\{k_n\}$ satisfies (6); (7) follows from

$$k_{t_n+s}/(t_n + s) = r_n s/(t_n + s) \geqq r_n/(t_n + 1) > n.$$

Proof of Theorem 1. (i) $\Rightarrow$ (ii): Suppose that (i) holds. Let U_{nk} denote the random variable in the nth row and kth column of the following array:

$$\begin{array}{ccccc} X_1 & X_2 & X_6 & X_7 & \cdots \\ X_3 & X_5 & X_8 & \cdot & \cdots \\ X_4 & X_9 & \cdot & \cdot & \cdots \\ X_{10} & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \end{array}$$

Let $\hat{X}_n = U_n - V_n$ where $U_n = U_{n1}$, $V_n = \sum_{k=2}^{\infty} g(U_{nk})/2^k$, and g is the characteristic function of (EX, ∞), $n = 1, 2, \cdots$. Then $\hat{X}_1, \hat{X}_2, \cdots$ are integrable, independent, have a common continuous distribution function, and $E(\hat{X}_1 \log^+ \hat{X}_1) = \infty$. By Lemma 1, there is a real number sequence $\{a_n\}$ satisfying (3) and (4) with $\hat{X}_1$ substituted for X therein. Let Y_n be defined as in Section 2 using $\hat{X}_n$ in place of X_n. Let $T_0 = E(\cdot \mid Y_1, Y_2, \cdots)$. Then $T_0\hat{X}_n = Y_n$ almost everywhere and

$$\begin{aligned} E(\sup_n X_n/n) &= E(\sup_n U_n/n) \geqq E(\sup_{1 \leqq j \leqq k} \hat{X}_j/j) = ET_0(\sup_{1 \leqq j \leqq k} \hat{X}_j/j) \\ &\geqq E(\sup_{1 \leqq j \leqq k} Y_j/j) = E(\sup_{1 \leqq j \leqq k} Y_j/j - Y_1) + EY_1 \qquad (9) \\ &\rightarrow E(\sup_n Y_n/n) = \infty, \end{aligned}$$

using the monotone convergence theorem and the fact, following from Section 2, that $\sup_n Y_n/n = \infty$ almost everywhere.

(i) $\Rightarrow$ (iii): In (9), replace X_n by $X_1 + \cdots + X_n$ and proceed similarly.

(i) $\Rightarrow$ (iv): Keeping the notation of the previous paragraphs, let $T_n = E(\cdot \mid \hat{X}_1 + \cdots + \hat{X}_k, k \geqq n)$ and S_n be defined by (1), $n = 1, 2, \cdots$. Since $X_1 \geqq \hat{X}_1$, we have that

$$\limsup_{n\to\infty} S_n X_1 \geqq \limsup_{n\to\infty} S_n \hat{X}_1 = \infty$$

almost everywhere, using Section 2.

(ii) $\Rightarrow$ (i), (iii) $\Rightarrow$ (i): Suppose that X is nonnegative. The general case can be reduced to this one by replacing X by $\max\{0, X\}$. In the nonnegative case, (ii) $\Rightarrow$ (iii). Therefore, we need to show only that (iii) $\Rightarrow$ (i). Let $Z_n = (X_1 + \cdots + X_n)/n$. Then $\{Z_{n+1-j}\}_{j=1}^{n}$ is a nonnegative martingale and by a theorem of Doob ([3], Theorem 3.4, p. 317) we have that

$$E(\sup_{1 \leqq j \leqq n} Z_j) \leqq [e/(e-1)] + [e/(e-1)]E(Z_1 \log^+ Z_1).$$

The desired result follows by the monotone convergence theorem.

(iv)$\Rightarrow$ (i): Again it suffices to consider nonnegative X. In this case, $E(X \log^+$

$X) < \infty$ implies (2), which by the already mentioned result of Rota, implies the almost everywhere convergence of $\{S_n X_1\}$ for every $\{T_n\}$, contradicting (iv).

This completes the proof of the theorem.

4. A variation on the strong law of large numbers.

THEOREM 2. *Let $(\Omega, \mathfrak{A}, P)$ be a probability space, $\{\mathfrak{A}_n\}$ a sequence of independent sub-σ-fields of $\mathfrak{A}$, and $\{X_n\}$ a sequence of identically distributed integrable random variables such that X_n is $\mathfrak{A}_n$-measurable, $n = 1, 2, \cdots$. Consider the following statements*:

(i) $E(|X_1| \log^+ |X_1|) < \infty$.

(ii) *For every sequence $\{\mathfrak{B}_n\}$ of σ-fields such that $\mathfrak{B}_n \subset \mathfrak{A}_n$, $n = 1, 2, \cdots$,*

$$\lim_{n\to\infty}(E(X_1 \mid \mathfrak{B}_1) + \cdots + E(X_n \mid \mathfrak{B}_n))/n = EX_1$$

almost everywhere.

Then (i) $\Rightarrow$ (ii). *Also,* (ii) $\Rightarrow$ (i) *under the further condition that the restriction of P to $\mathfrak{A}_n$ is nonatomic, $n = 1, 2, \cdots$.*

Note that $E(X_1 \mid \mathfrak{B}_1), E(X_2 \mid \mathfrak{B}_2), \cdots$ are independent but are not necessarily identically distributed.

PROOF. (i) $\Rightarrow$ (ii): Let $\mathfrak{B}_n$ be a sub-σ-field of $\mathfrak{A}_n$, $n = 1, 2, \cdots$, and $T = E(\cdot \mid \mathfrak{B}_1 \vee \mathfrak{B}_2 \vee \cdots)$. Then, letting $Z_n = (X_1 + \cdots + X_n)/n$, we have that

$$TZ_n = (E(X_1 \mid \mathfrak{B}_1) + \cdots + E(X_n \mid \mathfrak{B}_n))/n$$

almost everywhere. By (i), which we assume, and by Theorem 1, the right side of

$$|Z_k| \leqq \sup_n(|X_1| + \cdots + |X_n|)/n$$

is integrable. Thus, almost everywhere ([3], p. 23)

$$\lim_{n\to\infty} TZ_n = T(\lim_{n\to\infty} Z_n) = TEX_1 = EX_1 .$$

The following lemma will be used in the proof of the second part of this theorem.

LEMMA 2. *Suppose that X is a random variable on a nonatomic probability space $(\Omega, \mathfrak{A}, P)$. Then there is a random variable V such that $0 \leqq V \leqq 1$, $X + V$ has a continuous distribution function, and this distribution function depends on $(\Omega, \mathfrak{A}, P)$ only through the distribution function of X.*

PROOF OF LEMMA 2. If the distribution function F of X is continuous, let $V = 0$. Otherwise, let D be the set of discontinuity points of F. By the theory of nonatomic measure spaces, for $d \in D$, there is a random variable U_d from Ω into $[0, 1]$ such that

$$P(U_d \leqq t, X = d) = tP(X = d), \qquad 0 \leqq t \leqq 1.$$

Let $V(\omega) = U_d(\omega)$ if $X(\omega) = d$ and $d \in D$, $= 0$ otherwise. Let $H(t) = 0$ if $t < 0$, $= t$ if $0 \leqq t \leqq 1$, $= 1$ if $t > 1$. For each real number a we have that

$$\begin{aligned} P(X + V \leqq a) &= \sum_{d \in D} P(X + V \leqq a, X = d) + P(X + V \leqq a, X \notin D) \\ &= \sum_{d \in D} P(U_d \leqq a - d, X = d) + P(X \leqq a, X \notin D) \\ &= \sum_{d \in D} H(a - d) P(X = d) + P(X \leqq a, X \notin D) \end{aligned}$$

which is expressible in terms of H and F. Both $H(a - d)$ and $P(X \leqq a, X \notin D)$ are continuous functions of a. The desired result is implied.

Returning to Theorem 2, suppose that (ii) holds, the restriction of P to $\mathcal{A}_n$ is nonatomic, $n = 1, 2, \cdots$, and that (i) does not hold. Then either $E(X_1 \log^+ X_1) = \infty$ or $E((-X_1) \log^+ (-X_1)) = \infty$. Here assume the former is true, the other case being similar. Let V_n be an $\mathcal{A}_n$-measurable random variable chosen relative to X_n and $(\Omega, \mathcal{A}_n, P)$ as V in Lemma 2 is chosen relative to X and $(\Omega, \mathcal{A}, P)$. Let $\hat{X}_n = X_n + V_n$. Then $X_n \leqq \hat{X}_n \leqq X_n + 1$ and $\hat{X}_1, \hat{X}_2, \cdots$ are independent integrable random variables each having the same continuous distribution function. Clearly, $E(\hat{X}_1 \log^+ \hat{X}_1) = \infty$. By Lemma 1 and Section 2 (replacing X_n by $\hat{X}_n$ therein), it follows that there exists a sub-σ-field $\mathcal{B}_n$ of $\mathcal{A}_n$ such that

$$\infty = \limsup_{n \to \infty} \sum_{k=1}^{n} E(\hat{X}_k \mid \mathcal{B}_k)/n \leqq 1 + \limsup_{n \to \infty} \sum_{k=1}^{n} E(X_k \mid \mathcal{B}_k)/n$$

almost everywhere, contradicting (ii). Consequently, under the stated additional condition (ii) $\Rightarrow$ (i).

REFERENCES

[1] BLACKWELL, DAVID and DUBINS, LESTER E. A converse to the dominated convergence theorem. *Illinois J. Math.* To appear

[2] BURKHOLDER, D. L. and CHOW, Y. S. (1961). Iterates of conditional expectation operators. *Proc. Amer. Math. Soc.* **12** 490–495.

[3] DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.

[4] HARDY, G. H. and LITTLEWOOD, J. E. (1930). A maximal theorem with function-theoretic applications. *Acta Math.* **54** 81–116.

[5] ROTA, GIAN-CARLO (1962). An "alternierende Verfahren" for general positive operators. *Bull. Amer. Math. Soc.* **68** 95–102.

[6] STEIN, E. M. (1961). On the maximal ergodic theorem. *Proc. Nat. Acad. Sci. USA.* **47** 1894–1897.