A CONVERSE TO THE DOMINATED CONVERGENCE THEOREM

BY

David Blackwell[1] and Lester E. Dubins[2]

1. Introduction and summary

On a probability space $(\Omega, \mathcal{B}, P)$, let $\{f_n, n = 1, 2, \cdots\}$ be a sequence of nonnegative random variables in L_1 such that $f_n \to f \in L_1$ with probability 1, and define $g = \sup_n f_n$. If $g \in L_1$, the Lebesgue dominated convergence theorem asserts that $E(f_n) \to E(f)$. More generally, as noted by Doob [1, p. 23], if $g \in L_1$, then for any Borel field $\mathcal{B}_0$ contained in $\mathcal{B}$,

$$E(f_n \mid \mathcal{B}_0) \to E(f \mid \mathcal{B}_0) \quad \text{a.e.} \tag{1}$$

If one extends this result in a minor manner, Lebesgue's condition $g \in L_1$ is not only sufficient but necessary, as the following converse to the dominated convergence theorem asserts.

Theorem 1. *If $f_n \geqq 0$, $f_n \to f$ a.e., $f_n \in L_1$, $f \in L_1$, and $g = \sup_n f_n \notin L_1$, there are, on a suitable probability space, random variables $\{f_n^*, n = 1, 2, \cdots\}, f^*$, and a Borel field $\mathcal{C}$ such that $f^*, f_1^*, f_2^*, \cdots$ have the same joint distribution as $f, f_1, f_2, \cdots$, and*

$$P\{E(f_n^* \mid \mathcal{C}) \to E(f^* \mid \mathcal{C})\} = 0. \tag{2}$$

In view of this result, it is of interest to find conditions which will ensure that $g \in L_1$. As a special case of interest, let h be a nonnegative random variable in L_1, let $\mathcal{B}_n$ be a monotone sequence of Borel fields contained in $\mathcal{B}$, and let $f_n = E(h \mid \mathcal{B}_n)$. Doob [1, p. 317] has shown that if $h \log h \in L_1$, then also $g = \sup_n f_n \in L_1$. It turns out that the condition $h \log h \in L_1$ is necessary, as well as sufficient, in the following sense:

Theorem 2. *If $h \geqq 0$, $h \in L_1$, $h \log h \notin L_1$, there are, on a suitable probability space, a random variable h^* with the same distribution as h and a monotone sequence $\mathcal{B}_n^*$ of Borel fields, which can be chosen either increasing or decreasing, for which*

$$g^* = \sup_n E(h^* \mid \mathcal{B}_n^*) \notin L_1. \tag{3}$$

Theorem 2 will be an immediate consequence of the following result, which gives sharp upper bounds on the distribution of g^*, rather than only information about the expectation of g^* as in Theorem 2.

Theorem 3. *Let h^* be any nonnegative random variable in L_1, and let h be the (essentially unique) nonincreasing function on the unit interval $(0, 1]$ whose*

Received March 19, 1962.

[1] This paper was prepared with the partial support of the Office of Naval Research.

[2] Prepared with the partial support of the National Science Foundation.

distribution, with respect to Lebesgue measure m on $(0, 1]$, *is the same as that of* h^*. *Define g on* $(0, 1]$ *by*

$$g(x) = \frac{1}{x}\int_0^x h(t)\,dt. \tag{4}$$

Then

(a) *for any monotone sequence $\mathcal{B}_n$ of Borel fields contained in $\mathcal{B}$, and any* $\lambda > 0$,

$$P\{g^* > \lambda\} \leqq m\{g > \lambda\}, \tag{5}$$

where $g^* = \sup_n E(h^* \mid \mathcal{B}_n)$,

(b) *for every $\varepsilon > 0$ there is an increasing sequence $\mathcal{C}_n$ of Borel fields in the unit interval X for which*

$$P\{g^{**} \geqq k\,\varepsilon\} = m\{g \geqq k\,\varepsilon\} \quad \text{for } k = 0, 1, 2, \cdots, \text{ and } g^{**} \geqq g - \varepsilon, \tag{6}$$

where g^{**} denotes $\sup_n E(h \mid \mathcal{C}_n)$, *and*

(c) *for every $\varepsilon > 0$ and every decreasing sequence of real numbers* $\{Q_n, n = 1, 2, \cdots\}$ *with* $0 \leqq Q_n \leqq 1$ *and* $Q_n \to 0$ *as* $n \to \infty$, *there are, on a suitable probability space, a random variable f with the same distribution as h and a decreasing sequence $\mathcal{D}_n$ of Borel fields such that for every positive integer k,*

$$P\{g_1 \geqq k\varepsilon\} \geqq Q_k\, m\{g \geqq k\varepsilon\}, \tag{7}$$

where $g_1 = \sup_n E(f \mid \mathcal{D}_n)$.

The proof that Theorem 3 implies Theorem 2 will use the following result of Hardy and Littlewood [2, p. 99]: *For any nonnegative monotone decreasing function h on* $(0, 1]$, *either* $h \log h$, $h(t) \log g(t)$, *and g are all in* L_1, *or none is.*

Say that a distribution μ on the real line *dominates* a distribution ν if $\mu(x, \infty) \geqq \nu(x, \infty)$ for all x. Theorem 3(a) asserts that for any nonnegative h^* in L_1, the distribution of g, denote it by μ, dominates that of $\sup_n E[h^* \mid \mathcal{B}_n)$ for any monotone increasing or decreasing sequence of Borel fields $\mathcal{B}_n$. Part (b) asserts that μ is, in a very strong sense, best possible for increasing $\mathcal{B}_n$. Part (c) asserts that the *same* distribution μ is best possible for decreasing $\mathcal{B}_n$, though in a somewhat weaker sense.

Inequality (5) has the following consequence. Consider a fair gambling system, which terminates after N plays, and in which the bettor is not allowed credit, i.e., a sequence $X_0, X_1, X_2, \cdots, X_N$ of nonnegative random variables which form a martingale; X_k is the bettor's fortune after k plays, and for simplicity let X_0 be constant. Suppose the bettor is allowed to choose, in advance of play, either of the following options:

Option 1. He uses the system and, at the end is paid, not his final fortune X_N, but the largest fortune $Y = \max(X_0, \cdots, X_N)$ he ever had in the course of play.

Option 2. He uses the system, achieving a terminal fortune X_N. If

X_N is as high as possible, he is given X_N. If not, he is given his original fortune X_0 and tries the system repeatedly until a final fortune Z is obtained which (strictly) exceeds the final fortune X_N on his first attempt. He is then given Z.

Though the distribution of Z need not dominate that of Y, it turns out that Option 2 is always better, in the sense that $E(Z) \geqq E(Y)$.

One final easy observation. For any nonnegative martingale $X_1, \cdots, X_N$, $E(\max(X_1, \cdots, X_N)) \leqq NE(X_1)$. This bound is best possible in that for every nonnegative X_1 with finite expectation and every $\varepsilon > 0$ and $N \geqq 1$, there is a nonnegative martingale, $X_1^*, \cdots, X_N^*$, where X_1^* has the same distribution as X_1, and for which $E(\max(X_1^*, \cdots, X_N^*)) > NE(X_1) - \varepsilon$.

2. Proof of Theorem 1

The Borel field $\mathcal{C}$ will be the smallest field with respect to which some random variable Z is measurable. We first reduce the theorem to the special case in which each f_n has only two values, 0 and $v_n > 0$, and at every sample point exactly one f_n is positive. Thus, if $p_n = P\{f_n = v_n\}$, we have $0 < p_n < 1$, $\sum p_n = 1, f \equiv 0, E(g) = \sum_n p_n v_n = \infty$.

To achieve this reduction, write

$$F_n = \max(f_n - f, 0), \qquad G_n = \min(f_n - f, 0).$$

Then $F_n \geqq 0$, $F_n \in L_1$,

$$\sup_n F_n \geqq g - f \notin L_1, \qquad F_n \to 0 \quad \text{a.e.,}$$

$$\sup_n |G_n| \leqq f \in L_1, \qquad G_n \to 0 \quad \text{a.e.}$$

For any Borel field $\mathcal{C}$, it follows from (1) that $E(G_n \mid \mathcal{C}) \to 0$ a.e., so that

$$P\{E(f_n \mid \mathcal{C}) \to E(f \mid \mathcal{C})\} = P\{E(f_n - f \mid \mathcal{C}) \to 0\}$$
$$= P\{E(F_n + G_n \mid \mathcal{C}) \to 0\} = P\{E(F_n \mid \mathcal{C}) \to 0\}.$$

Thus if we find, enlarging the probability space if necessary, a Borel field $\mathcal{C}$ for which $P\{E(F_n \mid \mathcal{C}) \to 0\} = 0$, it will follow that

$$P\{E(f_n \mid \mathcal{C}) \to E(f \mid \mathcal{C})\} = 0.$$

Thus we have reduced the theorem to the special case of the F_n, i.e., to the case $f = 0$.

Suppose now that $f = 0$. Denote by A_k the event

$$\{f_k \geqq g - 1, f_k < g - 1 \quad \text{for} \quad i < k\}.$$

The A_k are disjoint, and $\sum P(A_k) = 1$. Choose a simple function (i.e., one with only finitely many values) s_k such that s_k vanishes off A_k, $0 \leqq s_k \leqq f_k$ on A_k, and

$$E(s_k) \geqq \int_{A_k} f_k \, dP - \frac{1}{2^k}.$$

Then $\sup_k s_k = \sum_k s_k$, so that

$$E \sup_k s_k = \sum E(s_k) \geqq \sum_k \int_{A_k} f_k \, dP - 1$$

$$\geqq \sum_k \int_{A_k} g \, dP - 2 = E(g) - 2 = \infty.$$

Since $s_k \leqq f_k$, for any $\mathcal{C}$,

$$P\{E(s_k \mid \mathcal{C}) \to 0\} = 0 \quad \text{implies} \quad P\{E(f_k \mid \mathcal{C}) \to 0\} = 0,$$

so that we have reduced the theorem to the case of the s_k, i.e., the case in which $f = 0$, each f_n is simple, and at each sample point at most one f_n is positive. Starting from this case we represent each f_n as the sum of a finite number of nonnegative functions, each having only two values, one of which is 0, and no two of which are simultaneously positive. Rearranging these functions into a single sequence, omitting those which are 0 with probability 1 and, if the set B on which all these functions vanish has positive probability, taking the indicator I_B as an additional function, yield a sequence $f_1, f_2, \cdots$ with the properties stated at the beginning of the section, and the reduction is complete. We now prove the theorem in the special case.

Let k be the positive integer such that $1 < 2^k P_1 \leqq 2$, and let S_n, $n = 1, 2, \cdots$, denote the set of integers $i \geqq 2$ for which $2^{n+k} \leqq v_i < 2^{n+k+1}$. Define:

$$r_n = \textstyle\sum_{i \in S_n} p_i, \qquad t_n = r_n + 2^{-(n+k)};$$

$$r = \textstyle\sum_n r_n = \sum_{i \in S} p_i,$$

where $S = \bigcup S_n = \{i : i \geqq 2 \text{ and } v_i \geqq 2^{k+1}\}$;

$$t = \sum t_n = r + 2^{-k}.$$

Let W, Z_0, Z_1, $\cdots$ be independent integer-valued random variables with distributions as follows:

$$P(W = 0) = 1 - t, \qquad P(W = n) = t_n \quad \text{for} \quad n > 0.$$

$$P(Z_0 = 1) = (p_1 - 2^{-k})/(1 - t),$$

$$P(Z_0 = i) = p_i/(1 - t) \quad \text{for } i \geqq 2,\ i \notin S,$$

$$P(Z_0 = i) = 0 \qquad \text{otherwise}.$$

For $n \geqq 1$,

$$P(Z_n = 1) = 2^{-(k+n)}/t_n,$$

$$P(Z_n = i) = p_i/t_n \quad \text{for } i \in S_n,$$

$$P(Z_n = i) = 0 \qquad \text{otherwise}.$$

Define $X = Z_W$, and verify that $P(X = n) = p_n$ thus:

$$P(X = 1) = \textstyle\sum_{n=0}^{\infty} P(W = n, Z_n = 1) = p_1 - 2^{-k} + \sum_{n=1}^{\infty} 2^{-(k+n)} = p_1;$$

for $i > 1$, $i \notin S$,

$$P(X = i) = P(W = 0, Z_0 = i) = p_i;$$

for $i \in S_n$,

$$P(X = i) = P(W = n, Z_n = i) = p_i.$$

Thus, if we define $\phi_n = v_n$ on $\{X = n\}$, $\phi_n = 0$ otherwise, $\{\phi_n\}$ has the same joint distribution as $\{f_n\}$, i.e., ϕ_n has only the two values 0, v_n,

$$P\{\phi_n = v_n\} = p_n,$$

and at each sample point exactly one ϕ_n is positive.

For any Borel field $\mathcal{C}$, $E(\phi_n \mid \mathcal{C}) = v_n P\{X = n \mid \mathcal{C}\}$. It suffices to find a $\mathcal{C}$ for which the event $\{v_n P\{X = n \mid \mathcal{C}\} \geqq 1$ infinitely often$\}$ has probability one. We show that the Borel field $\mathcal{C}$ determined by $Z_0, Z_1, \cdots$ has the property. For $i \in S_n$,

$$P\{X = i \mid \mathcal{C}\} = 0 \quad \text{if} \quad Z_n \neq i,$$

$$P\{X = i \mid \mathcal{C}\} = t_n \quad \text{if} \quad Z_n = i.$$

Thus, if $Z_n = i$, $v_i P\{X = i \mid \mathcal{C}\} = v_i t_n$. Since $t_n \geqq 2^{-(n+k)}$ and, for $i \in S_n$, $v_i \geqq 2^{n+k}$, we have $v_i t_n \geqq 1$. Thus, for $n \geqq 1$ whenever $A_n = \{Z_n \neq 1\}$ occurs, so does $B_n = \{v_i P\{X = i \mid \mathcal{C}\} \geqq 1$ for some $i \in S_n\}$. The A_n are independent, with $P(A_n) = r_n/t_n$. We show that

$$\textstyle\sum_n (r_n/t_n) = \infty.$$

If $r_n < 2^{-(n+k)}$, $t_n < 2^{-(n+k-1)}$, so that

$$(r_n/t_n) \geqq 2^{n+k-1} r_n = \textstyle\sum_{i \in S_n} 2^{n+k+1} p_i/4 \geqq \sum_{i \in S_n} p_i v_i/4.$$

If $r_n \geqq 2^{-(n+k)}$ for infinitely many n, then $(r_n/t_n) \geqq \frac{1}{2}$ for infinitely many n, and the series $\sum (r_n/t_n)$ diverges. If $r_n < 2^{-(n+k)}$ for sufficiently large n, say for $n \geqq n_0$,

$$\textstyle\sum_n (r_n/t_n) \geqq \sum_{n \geqq n_0} \sum_{i \in S_n} p_i v_i/4 = \sum_{i \in T} p_i v_i/4$$

where $T = \{i \geqq 2, v_i \geqq 2^{n_0+k}\}$. Since

$$\textstyle\sum_i p_i v_i = \infty \quad \text{and} \quad \sum_{i \notin T} p_i v_i \leqq 2^{n_0+k} \sum p_i \leqq 2^{n_0+k},$$

we conclude that $\sum_{i \in T} p_i v_i$ diverges. Thus $\sum (r_n/t_n)$ diverges, so that, with probability 1, infinitely many A_n, and hence infinitely many B_n, occur. This completes the proof.

3. Proofs of other results

For part (a) of Theorem 3 use an inequality of Doob [1, p. 314] which asserts that, for every $\lambda > 0$,

$$\lambda P\{g_n^* \geqq \lambda\} \leqq \int_{\{g_n^* \geqq \lambda\}} h^* \, dP, \tag{8}$$

where $g_n^* = \max_{1 \leqq i \leqq n} E(h^* \mid \mathcal{B}_i)$. Letting $\lambda \downarrow \lambda_0 > 0$ yields

$$\lambda_0 P\{g_n^* > \lambda_0\} \leqq \int_{\{g_n^* > \lambda_0\}} h^* \, dP.$$

Letting $n \to \infty$, and dropping the subscript in λ_0, you obtain, for every $\lambda > 0$,

$$\lambda P\{g^* > \lambda\} \leqq \int_{\{g^* > \lambda\}} h^* \, dP, \tag{9}$$

and letting $\lambda \uparrow \lambda_0$ yields an inequality like (9) with the event $\{g^* > \lambda\}$ replaced by $\{g^* \geqq \lambda_0\}$. For any λ for which $P\{h^* > \lambda\} = 0$, we have also $P\{g^* > \lambda\} = 0$, and (5) is trivial. If $P\{h^* \geqq \lambda\} > 0$, note that g is monotone, and let u be the largest number for which $g(u) \geqq \lambda$. Then for any event A for which

$$\frac{1}{P(A)} \int_A h^* \, dP \geqq \lambda,$$

we must have $P(A) \leqq u$. The event $A = \{g^* \geqq \lambda\}$ has the property, from the remark following (9), so that

$$P\{g^* \geqq \lambda\} \leqq u = m\{g \geqq \lambda\}. \tag{10}$$

Letting $\lambda \downarrow \lambda_0$ yields (5), and (a) is established.

The remark on gambling systems is a consequence of $E(g^*) \leqq E(g)$, which follows from (5). For, with $h^* = X_n$, and $\mathcal{B}_i$ the Borel field determined by $X_0, \cdots, X_i$, $g^* = Y$, and

$$E(Z) = \int_0^1 \alpha(u) \, du,$$

where $\alpha(u) = g$ (smallest v with $h(v) = h(u)$). Since $\alpha(u) \geqq g(u)$, $E(Z) \geqq E(g)$, and the proof is complete.

For part (b), let $C_n = \{(n-1)\varepsilon \leqq g < n\varepsilon\}$, and let $\mathcal{C}_n$ be the Borel field determined by $C_1, \cdots, C_{n-1}$. If C_n is nonempty, it is an interval $a < u \leqq b$. When C_n occurs, $E(h \mid \mathcal{C}_n) = E(h \mid u \leqq b) = g(b) \geqq (n-1)\varepsilon$. Thus, on C_n, $E(h \mid \mathcal{C}_n) \geqq g - \varepsilon$, and $g^{**} = \sup_n E(h \mid \mathcal{C}_n) \geqq g - \varepsilon$ everywhere.

For part (c), set $Q_0 = 1$, and define $p_n = Q_{n-1} - Q_n$, so that $p_n \geqq 0$, $\sum_1^\infty p_n = 1$. Let α be a random variable, independent of g, h (this may require extending the probability space) with $P\{\alpha = n\} = p_n$. The Borel field $\mathfrak{D}_n$ will specify the value of α and, when $\alpha = k$, will specify, for every i, $1 \leqq i \leqq k - n$, whether C_i, defined in the proof of part (b), occurs. More formally, if I_i is the indicator or characteristic function of C_i (I_i has 1 as its value on C_i and 0 off C_i), and J_k is the indicator of $\{\alpha = k\}$, $\mathfrak{D}_n$ is the Borel field determined by the functions J_k, $k = 1, 2, \cdots$, and those functions $J_k I_i$, for which $k > n$ and $1 \leqq i \leqq k - n$. Then, for any i, k, n with $i > k - n > 0$ we have, on $C_i \cap \{\alpha = k\}$,

$$E(h \mid \mathfrak{D}_n) = E(h \mid \alpha, \cup_{j > k-n} C_j) \geqq (k - n)\varepsilon.$$

For $k \leqq i$, we choose $n = 1$; for $k > i$, we choose $n = k - i + 1$. We then see that on $C_i \cap \{\alpha = k\}$ either $g_1 \geqq (k - 1)\varepsilon$ or $g_1 \geqq (i - 1)\varepsilon$ according as $k \leqq i$ or $k > i$. We conclude that, on $(\bigcup_{i>j} C_i) \cap \{\alpha > j\}$, $g_1 \geqq j\varepsilon$. Since $\bigcup_{i>j} C_i = \{g \geqq j\varepsilon\}$, we obtain

$$P\{g_1 \geqq j\varepsilon\} \geqq P\{g \geqq j\varepsilon\}\, P\{\alpha > j\} = Q_j\, P\{g \geqq j\varepsilon\},$$

which is assertion (7).

The proof of Theorem 2 is now easy. We may suppose that h is a nonincreasing function on the unit interval, and that probability is Lebesgue measure. Since $h \log h$ is not in L_1, the result of Hardy and Littlewood referred to following (7) implies that g, defined as in Theorem 3, is not in L_1. To choose $\mathcal{B}_n^*$ increasing, with $g^* \notin L_1$, choose as $\mathcal{B}_n^*$ the $\mathcal{C}_n$ of part (b) of Theorem 3. Then $g^* \geqq g - \varepsilon$, so that $g^* \notin L_1$. To choose $\mathcal{B}_n^*$ decreasing, note first that, since $g \notin L_1$, $\sum_k m(g \geqq k\varepsilon) = \infty$. We may then choose a monotone sequence Q_k converging to 0 with $1 \geqq Q_k \geqq 0$, $k = 1, 2 \cdots$, for which

$$\sum_k Q_k\, m(g \geqq k\varepsilon) = \infty. \tag{11}$$

For this choice of Q_k, and $\mathcal{B}_n^*$ chosen as the $\mathfrak{D}_n$ of (c) of Theorem 3, the g^* of Theorem 2 is the g_1 of Theorem 3. From (7) and (11), clearly,

$$\sum P\{g^* \geqq k\varepsilon\} = \infty,$$

so that $g^* \notin L_1$. This completes the proof.

As for the final remarks about nonnegative martingales, let

$$Y = \max(X_1, \cdots, X_N),$$

and note that $Y \leqq \sum X_i$, so $E(Y) \leqq NE(X_1)$. To find a process where $E(Y) > NE(X_1) - \varepsilon$, let $X_1^*, \cdots, X_N^*$ be the successive fortunes of a gambler who at time j gambles as follows. He stakes his entire fortune X_j^* on a long shot, so that with small probability, namely t^{-1}, his fortune increases to tX_j^*, and with high probability, namely $1 - t^{-1}$, his fortune decreases to 0. It is easy to verify that $E(Y^*) = nE(X_1) - (n - 1)t^{-1}E(X_1)$. This completes the proofs.

Added in proof. Theorem 2, with a particularly interesting choice of $\mathcal{B}_n^*$, has also been obtained by D. L. Burkholder, in *Successive conditional expectations of an integrable function*, Ann. Math. Statistics, vol. 33 (1962), pp. 887–893.

Bibliography

1. J. L. Doob, *Stochastic processes*, New York, Wiley, 1953.
2. G. H. Hardy and J. E. Littlewood, *A maximal theorem with function-theoretic applications*, Acta Math., vol. 54 (1930), pp. 81–116.

University of California
Berkeley, California