A SHARPER FORM OF THE BOREL-CANTELLI LEMMA AND THE STRONG LAW[1]

BY LESTER E. DUBINS AND DAVID A. FREEDMAN

University of California, Berkeley

1. Introduction. Let $X_1, X_2, \cdots$ be a sequence of random variables taking only the values 0 and 1. Let $\mathfrak{F}_0, \mathfrak{F}_1, \cdots$ be an increasing sequence of σ-fields such that X_n is $\mathfrak{F}_n$-measurable. Let p_n be the conditional probability that X_n is 1, given $\mathfrak{F}_{n-1}$. To avoid trivial complications, suppose p_1 is a positive constant less than 1. The main object of this paper is prove:

(1) THEOREM. *As $n \to \infty$, the ratio $R_n = (X_1 + \cdots + X_n)/(p_1 + \cdots + p_n)$ converges to a finite limit L almost surely. L is 1 almost surely on the set where $\sum_1^\infty p_i$ is infinite.*

This theorem, which is related to (Chow, Corollary 7), sharpens and unifies two results of P. Lévy: his conditional form of the Borel-Cantelli Lemma [(Lévy, Corollary 68, p. 249) or (Doob, Corollary 2, p. 324)], and his martingale strong law of large numbers (Lévy, Section 69, pp. 250 ff). Both results are stated here for ease of reference:

(2) $\sum_1^\infty X_i$ *is finite (infinite) almost surely where* $\sum_1^\infty p_i$ *is finite (infinite)*;

(3) $\lim_{n\to\infty} n^{-1} \sum_1^n (X_i - p_i) = 0$ *almost surely.*

Let $\exp x = e^x$ and $e(x) = (\exp x) - 1 - x$. Let $V_1 = p_1(1 - p_1)$. Some information about L is contained in:

(4) THEOREM. *For each positive real number λ, and $n = 1, 2, \cdots$ the expectation of $\exp(\lambda R_n)$ is no more than $2 \exp[\lambda + 2V_1 e(\lambda/V_1)]$.*

In connection with the bound in (4), it is perhaps worth noting that, if a random variable Z has a Poisson distribution with mean v, then the expectation of $\exp[\lambda(Z - v)/v]$ is $\exp[ve(\lambda/v)]$.

(5) COROLLARY. *R_n converges to L in rth mean for $0 < r < \infty$.*

2. Proof of Theorem (1). Let $(\Omega, \mathfrak{F}, P)$ be a probability triple; $\mathfrak{F}_0, \mathfrak{F}_1, \cdots$ an increasing sequence of sub-σ-fields of $\mathfrak{F}$; and let Y_n' be $\mathfrak{F}_n$-measurable. It is not necessary to assume that the Y_n' have finite expectation; what is essential is to assume that m_n, the conditional expectation of Y_n' given $\mathfrak{F}_{n-1}$, is finite almost surely. Let $Y_n = Y_n' - m_n$. Let V_n be the conditional variance of Y_n' given $\mathfrak{F}_{n-1}$, that is, the conditional expectation of Y_n^2 given $\mathfrak{F}_{n-1}$. It is convenient (but not necessary) to assume V_n finite.

Here is a reformulation of (Dubins and Savage 1965a, and 1965b end of Section 1.2 and Theorem 9.4.1).

Received 15 January 1965.

[1] Partially supported by the National Science Foundation Grant GP-2593; and by a grant (to D. A. F.) from the Sloan Foundation.

(6) LEMMA. *If a and b are positive numbers, the probability that $Y_1 + \cdots + Y_n \geqq a(V_1 + \cdots + V_n) + b$ for some n is no greater than $(1 + ab)^{-1}$. This bound is sharp.*

(7) COROLLARY. *If a and b are positive numbers, the probability that $|Y_1 + \cdots + Y_n| \geqq a(V_1 + \cdots + V_n) + b$ for some n is no greater than $2(1 + ab)^{-1}$.*

PROOF. Immediate from (6). <>

A similar corollary (29), which can also be used to prove (1), will be presented later with a self-contained proof.

To avoid trivial complications, suppose V_1 is positive and constant.

(8) LEMMA. $\lim_{n\to\infty} (Y_1 + \cdots + Y_n)/(V_1 + \cdots + V_n) = 0$ *almost surely on the set D where $\sum_1^\infty V_i$ is infinite.*

PROOF. *Let $D(a, b)$ be the set where* $|Y_1 + \cdots + Y_n| < a(V_1 + \cdots + V_n) + b$ for all n. Plainly,

(9) $$\limsup_{n\to\infty} |(Y_1 + \cdots + Y_n)/(V_1 + \cdots + V_n)| \leqq a$$

on $D \cap D(a, b)$. Let b increase to ∞ through the integers and apply (7) to prove that (9) holds almost surely on D. Finally, let a shrink to 0 through the reciprocals of the positive integers, and conclude that (9) holds with $a = 0$, almost surely on D. <>

(10) LEMMA. $\lim_{n\to\infty} (Y_1 + \cdots + Y_n)$ *exists and is finite almost surely on the set G where $\sum_1^\infty V_i$ is finite.*

PROOF. Let $D_k(a, b)$ be the set where $|Y_{k+1} + \cdots + Y_{k+n}| < a(V_{k+1} + \cdots + V_{k+n}) + b$ for all n. Corollary (7) implies that the probability of $D_k(a, b)$ is no less than $1 - 2(1 + ab)^{-1}$. Let $G_k(c)$ be the set where $V_{k+1} + \cdots + V_{k+n} \leqq c$, for all n. For each positive number c, $\lim_{k\to\infty} P[G - G_k(c)] = 0$. On $G_k(c) \cap D_k(a, b)$,

(11) $$|Y_{k+1} + \cdots + Y_{k+n}| < ac + b \qquad \text{for all } n.$$

It is now routine to complete the proof, by exhibiting for each $\epsilon > 0$ a measurable subset $C(\epsilon)$ of G with these two properties:

(12) $$Y_1,\ Y_1 + Y_2,\ \cdots \text{ is a Cauchy sequence on } C(\epsilon);$$

(13) $$P[G - C(\epsilon)] < \epsilon.$$

For example, let $0 < b_i \to 0$. Choose $a_i > 0$ so large that $\sum_1^\infty (1 + a_i b_i)^{-1} < \epsilon/4$. Choose $c_i > 0$ so small that $a_i c_i \to 0$. Choose k_i so large that $\sum_1^\infty P[G - G_{k_i}(c_i)] < \epsilon/2$. Then put

$$C(\epsilon) = \bigcap_1^\infty [G_{k_i}(c_i) \cap D_{k_i}(a_i, b_i)].$$ <>

PROOF of (1). Specialize Y_i' to X_i of Section 1. By (8),

(14) $$\lim_{n\to\infty} \{[(X_1 - p_1) + \cdots + (X_n - p_n)]/ [p_1(1 - p_1) + \cdots + p_n(1 - p_n)]\}$$

and therefore

(15) $$\lim_{n\to\infty} \{[(X_1 - p_1) + \cdots + (X_n - p_n)]/[p_1 + \cdots + p_n]\}$$

are 0 almost surely where $\sum_1^\infty p_j(1 - p_j) = \infty$. Moreover

(16) $\lim_{n\to\infty} \sum_1^n (X_i - p_i)$ exists and is finite almost surely where

$$\sum_1^\infty p_j(1 - p_j) < \infty,$$

by (10). Thus (15) is 0 almost surely where $\sum_1^\infty p_j$ is infinite. Finally, $\sum_1^\infty p_j < \infty$ implies $\sum_1^\infty p_j(1 - p_j) < \infty$; from (16), $\sum_1^\infty X_i < \infty$ almost surely where $\sum_1^\infty p_j < \infty$. <>

3. Gambling. Proofs of some results below depend on (Dubins and Savage, 1965b, Theorem 2.12.1). A case of this theorem will be presented here (22), in language as close to that of the reference as is convenient. F is a set (the gambler's fortunes), and Σ is a σ-field of subsets of F. The real-valued function u defined on F is Σ-measurable and bounded from below ($u(f)$ is the utility of the fortune $f \varepsilon F$). For each fortune f there is a set $\Gamma(f)$ of countably additive probabilities (gambles) on (F, Σ); point mass at f is an element of $\Gamma(f)$. Informally, if the gambler's fortune is f, he may choose any gamble $\gamma \varepsilon \Gamma(f)$; his next fortune will then be chosen at random from F according to γ.

Recall that $\mathfrak{F}_0, \mathfrak{F}_1, \cdots$ is an increasing sequence of sub-σ-fields of $\mathfrak{F}$. A *Γ-process starting from f* is a stochastic process $\mathbf{f}_0, \mathbf{f}_1, \cdots$ with these three properties:

(17) $\mathbf{f}_n$ is measurable from $(\Omega, \mathfrak{F}_n)$ to (F, Σ);

(18) $$\mathbf{f}_0 \equiv f;$$

(19) given $\mathfrak{F}_n$, the conditional distribution of $\mathbf{f}_{n+1}$ is a gamble $\varepsilon\ \Gamma(\mathbf{f}_n)$.

Informally, a Γ-process starting from f is the sequence of fortunes of some gambler whose initial fortune is f, who does not foresee the future, and who gambles measurably using only gambles available in the gambling house Γ.

A random variable s is a *stopping time* if:

(20) s is a nonnegative integer with probability 1,

and

(21) the set where s takes the value n is in $\mathfrak{F}_n$, for $n = 0, 1, 2, \cdots$.

Informally, if the gambler stops gambling at time s, he receives payoff $u(\mathbf{f}_s)$. Introduce E for expectation.

(22) THEOREM. *Let Q be a real-valued, Σ-measurable function on F. If*

(23) $$u \leqq Q$$

and

(24) *for each fortune $f \varepsilon F$ and gamble $\gamma \varepsilon \Gamma(f)$, the γ-expectation of*

$$Q \leqq Q(f);$$

then for each fortune $f \varepsilon F$, for each Γ-process $\mathbf{f}_0, \mathbf{f}_1 \cdots$ *starting from f, and each stopping time s:*

$$E(u(\mathbf{f}_s)) \leqq Q(f). \tag{25}$$

Proof. (24) implies that $Q(\mathbf{f}_0)$, $Q(\mathbf{f}_1)$, $\cdots$ is an expectation-decreasing martingale. Taking (23) into account,

$$E(u(\mathbf{f}_s)) \leqq E(Q(\mathbf{f}_s)) \leqq Q(f). \qquad <>$$

4. Bounds on L^2 norm. Recall the meaning of Y_n and V_n from Section 2.

(26) Theorem. *Let y and v be real numbers with $v > 0$, and let s be a stopping time. Then the expectation of $(y + Y_1 + \cdots + Y_s)^2/(v + V_1 + \cdots + V_s)^2$ is no more than*

$$(y^2/v^2) + (1/v). \tag{27}$$

This bound is sharp for $y = 0$ and all v.

Proof. The proof that (27) is a bound is an application of (22). Let F be the set of pairs (y, v) of real numbers with $v > 0$. Let Σ be the σ-field of Borel subsets of F. Define the utility function u by: $u(y, v) = y^2/v^2$. As a preliminary to defining the gambling house Γ, let $\mathcal{P}$ be the set of pairs (Y, V), where V is a nonnegative real number, and Y is a random variable with expectation 0 and variance V. If (y, v) is a fortune in F, then a gamble γ is in $\Gamma(y, v)$ if and only if γ is the distribution of $(y + Y, v + V)$, for some pair (Y, V) in $\mathcal{P}$. Plainly, (y, v), $(y + Y_1, v + V_1)$, $\cdots$ is a Γ-process starting from (y, v). Define $Q(y, v)$ as (27).

Clearly, (23) is satisfied. To verify (24), let $(Y, V) \varepsilon \mathcal{P}$. The expectation of $Q(y + Y, v + V)$ is

$$[(y^2 + V)/(v + V)^2] + [1/(v + V)]. \tag{28}$$

Obviously, (28) decreases as V increases through the positive reals, and is $Q(y, v)$ for $V = 0$. Therefore (24) holds, and by (22), so does Inequality (25).

To see that the bound is sharp for $y = 0$, let a be large and positive. Since the bound in (6) is sharp, there is a process $Y_1, Y_2, \cdots$ and a natural number n with $P[(Y_1 + \cdots + Y_n) \geqq a(v + V_1 + \cdots + V_n)]$ nearly $(1 + a^2 v)^{-1}$. Then $E[(Y_1 + \cdots + Y_n)^2/(v + V_1 + \cdots + V_n)^2]$ is nearly as large as $a^2/(1 + a^2 v)$.

It is easy to deduce from (26) that $E|(y + Y_1 + \cdots + Y_s)/(v + V_1 + \cdots + V_s)| < |y/v| + v^{-\frac{1}{2}}$. This provides an alternative proof of (Lemma 1, Chow and Robbins), a result about fair coin tossing.

(29) Corollary. *Let a be a positive number. The probability that*

$$|y + Y_1 + \cdots + Y_n| \geqq a(v + V_1 + \cdots + V_n)$$

for some n is no more than $(y^2 + v)/(av)^2$.

This inequality is similar to (7)—set $y = 0$ and $v = b/a$—and may replace (7) in the proof of (1).

PROOF. (29) is immediate from (26), Chebychev's inequality, and (30) below. <>

(30) LEMMA. *Let* Z_n *be* $\mathfrak{F}_n$*-measurable,* $n = 1, 2, \cdots$, *and let* b, z *be nonnegative real numbers. Suppose* $P[Z_s \geqq z] \leqq b$ *for all stopping times* s. *Then* $P[Z_n \geqq z$ *for some* $n = 1, 2, \cdots] \leqq b$.

PROOF. Let $t = \infty$ if $Z_j < z$ for all j. Otherwise, let t be the least j with $Z_j \geqq z$. Let $t(k) = \min [t, k]$. Then $t(k)$ is a stopping time, and $[Z_{t(k)} \geqq z] \uparrow \bigcup_1^\infty [Z_n \geqq z]$ as $k \uparrow \infty$. <>

5. Proofs of Theorem (4) and Corollary (5). The next proposition may be of independent interest; here it serves only to clarify (34), which is used in the proof of (4). Recall the definition of Y_n and V_n from Section 2.

(31) PROPOSITION. *Let* λ *and* b *be positive real numbers, and* n *a positive integer or even a stopping time. Suppose* $\sum_1^n V_i \leqq b$, *and* $|Y_i| \leqq 1$ *for* $i \leqq n$, *almost surely. Then the expectation of* $\exp [\lambda(Y_1 + \cdots + Y_n)]$ *is no greater than* $\exp [be(\lambda)]$. *This bound is sharp.*

PROOF. The proof is an application of (22). Let F be the set of pairs (y, v) of real numbers, with $0 \leqq v \leqq b$. Let Σ be the σ-field of Borel subsets of F. Define the utility function u by: $u(y, v) = e^{\lambda y}$. As a preliminary to defining the gambling house Γ, let $\mathcal{P}$ be the set of all pairs (Y, V), where V is a nonnegative real number, and Y is a random variable with expectation 0, variance V, and $|Y| \leqq 1$ almost surely. If (y, v) is a fortune in F, then a gamble γ is in $\Gamma(y, v)$ if and only if γ is the distribution of $(y + Y, v + V)$, for some pair (Y, V) in $\mathcal{P}$ with $v + V \leqq b$. Define Q by

$$(32) \qquad Q(y, v) = \exp [\lambda y + (b - v)e(\lambda)].$$

Clearly, (23) is satisfied. (24) is immediate from:

$$(33) \quad \text{If } (Y, V) \in \mathcal{P}, \text{ then } E[\exp (\lambda Y)] \leqq 1 + Ve(\lambda) \leqq \exp [Ve(\lambda)].$$

To verify (33), expand $\exp (\lambda Y)$ as a power series in (λY). The inequality in (31) is obtained by setting $y = v = 0$ in (25).

To see that the bound is sharp, take $Y_i : 1 \leqq i \leqq n$ to be independent, with this common distribution: Y_i is $-b/n$ with probability $1 - (b/n)$, and is $1 - (b/n)$ with probability b/n. Let n increase to ∞. <>

Recall that $2 \cosh x = \exp x + \exp (-x)$.

(34) THEOREM. *Let* y, v, λ *be real numbers with* $v > 0$ *and* $\lambda > 0$, *and let* n *be a stopping time. Suppose* $|Y_i| \leqq 1$ *for* $i \leqq n$, *almost surely. Then the expectation of*

$$\cosh [\lambda(y + Y_1 + \cdots + Y_n)/(v + V_1 + \cdots + V_n)]$$

is no greater than

$$(35) \qquad \cosh (\lambda y/v) \exp [ve(\lambda/v)].$$

PROOF. The proof is an application of (22). Let F be the set of pairs (y, v) of real numbers, with $v > 0$. Let Σ be the σ-field of Borel subsets of F. Define the utility function u by:

(36) $$u(y, v) = \cosh(\lambda y/v).$$

Recall the definition of $\mathcal{P}$ from the proof of (31). Define the gambling house Γ as follows. If (y, v) is a fortune in F, then a gamble γ is in $\Gamma(y, v)$ if and only if γ is the distribution of $(y + Y, v + V)$ for some pair (Y, V) in $\mathcal{P}$. Define $Q(y, v)$ as (35).

Plainly, (23) is satisfied. (24) will be easy to verify with the help of (33) once the next two facts are checked:

(37) $\exp\{(v + 2V)e[\lambda/(v + V)]\}$ decreases as V increases through the positive reals, and in particular has its maximum at $V = 0$;

(38) $\cosh x$ increases as x increases through the positive reals.

Relation (37) can be verified by expanding $\exp[\lambda/(v + V)]$ in powers of $[\lambda/(v + V)]$, and noticing that

(39) for $n \geqq 2$, $(v + 2V)/(v + V)^n$ decreases as V increases through the positive reals. <>

(40) Corollary. *If* $|Y_i| \leqq 1$ *almost surely for* $1 \leqq i \leqq n$, *and* λ *is a positive real number, the expectation of*

(41) $$\cosh[\lambda(Y_1 + \cdots + Y_n)/(V_1 + \cdots + V_n)]$$

is no more than $2\exp[2V_1 e(\lambda/V_1)]$.

Proof. Given $\mathfrak{F}_1$, the conditional expectation of (41) is no more than $\cosh(\lambda Y_1/V_1)\exp[V_1 e(\lambda/V_1)]$, by (34). Integrate out Y_1, and apply (33).

Proof of (4). Immediate from (38) and (40).

Proof of (5). For nonnegative x, positive λ, and $k = 0, 1, 2, \cdots$

(42) $$x^k \leqq k!\lambda^{-k}e^{\lambda x}.$$

If Z is a nonnegative random variable, (42) implies

(43) $$E(Z^k) \leqq k!\lambda^{-k}E[\exp(\lambda Z)].$$

Relations (4) and (43) imply $E(R_n^k)$ is bounded in n for each k. This, together with (1), gives the Corollary (for a note on uniform integrability, see Doob, p. 629).

6. Heuristics. In this paper, the main difficulty in applying (22) was guessing suitable Q's. This section tries to explain how we were led to the Q's of Sections 4 and 5; it can be omitted without logical loss.

The Q of (26). After some fruitless attempts to guess a Q for bounding

(44) $$E[(y + Y_1 + \cdots + Y_n)/(v + V_1 + \cdots + V_n)]^2,$$

we decided to put $y = 0$ in (44). According to (6),

(45) $$P[(Y_1 + \cdots + Y_n)/(v + V_1 + \cdots + V_n) \geqq a] \leqq 1/(1 + a^2 v),$$

and this bound is sharp. Use $\doteq$ for "nearly equal". It seemed plausible that a Y-process nearly maximizing the left side of (45) would have

$$P[(Y_1 + \cdots + Y_n)/(v + V_1 + \cdots + V_n) \doteq a] \doteq 1/(1 + a^2 v)$$

and

$$P[(Y_1 + \cdots + Y_n)/(v + V_1 + \cdots + V_n) \doteq 0] \doteq 1 - [1/(1 + a^2 v)],$$

so

$$E[(Y_1 + \cdots + Y_n)/(v + V_1 + \cdots + V_n)]^2 \doteq a^2/(1 + a^2 v) \leqq 1/v.$$

Since $(c + d)^2 \leqq 2(c^2 + d^2)$, this suggested $2(y^2 + v)/v^2$ as a bound for (44), that is, as a candidate Q. Therefore, even (27) was a candidate Q.

The Q of (31). We began by trying to bound $E\{\exp [\lambda(Y_1 + \cdots + Y_n)]\}$ when

$$\sum_1^n V_i \leqq 1. \tag{46}$$

Using (33),

$$\begin{aligned} &E\{\exp [\lambda(Y_1 + \cdots + Y_n)] \mid \mathfrak{F}_{n-1}\} \\ &\qquad = \exp [\lambda(Y_1 + \cdots + Y_{n-1})]E\{\exp (\lambda Y_n) \mid \mathfrak{F}_{n-1}\} \\ &\qquad \leqq \exp [\lambda(Y_1 + \cdots + Y_{n-1})] \exp [V_n e(\lambda)]. \end{aligned} \tag{47}$$

Take conditional expectations given $\mathfrak{F}_{n-2}$, $\mathfrak{F}_{n-3}$, $\cdots$ to obtain, at least formally,

$$E\{\exp [\lambda(Y_1 + \cdots + Y_n)]\} \leqq \exp [(V_1 + \cdots + V_n)e(\lambda)]; \tag{48}$$

so, in view of (46),

$$E\{\exp [\lambda(Y_1 + \cdots + Y_n)]\} \leqq \exp [e(\lambda)]: \tag{49}$$

which is rigorous when each V_i is constant almost surely. But (49) does not involve V_i explicitly; so it seemed plausible that (49) held for non-constant V_i satisfying (46). (At this point we had no proof of (49) for non-constant V_i, because Y_i need not have conditional mean 0 given $\mathfrak{F}_{i-1}$ together with $V_1 \cdots V_n$.) To obtain a proof, it seemed desirable to use (22). To express Condition (46), it seemed necessary to consider pairs (y, v) as fortunes. Change of scale in (49) gave (32) as a candidate Q.

The Q of (34). We tried to bound

$$E[\exp (\lambda(Y_1 + \cdots + Y_n)/(v + V_1 + \cdots + V_n))].$$

If $V_1 + \cdots + V_n$ is a constant, say b, Proposition (31) gives

$$\begin{aligned} &E\{\exp \lambda[(Y_1 + \cdots + Y_n)/(v + V_1 + \cdots + V_n)]\} \\ &\qquad \leqq \exp \{be[\lambda/(v + b)]\}. \end{aligned} \tag{50}$$

Clearly, $x \to xe(1/x)$ decreases as x increases through positive values. The right side of (50), being no more than $\exp \{(v + b)e[\lambda/(v + b)]\}$, is no more than exp $[ve(\lambda/v)]$. This function depends neither on b nor on V_i; so it was a plausible upper bound to the left side of (50) for arbitrary V_i. A candidate Q for bounding

$$E\{\exp [\lambda(y + Y_1 + \cdots + Y_n)/(v + V_1 + \cdots + V_n)]\}$$

was then exp $[(\lambda y/v) + ve(\lambda/v)]$. This was not completely plausible, and we could not verify it, because exp $(\lambda y/v)$ is a decreasing function of v for $y > 0$, and increasing for $y < 0$. When we remembered (38), it seemed natural to replace exp $(\lambda y/v)$ by cosh $(\lambda y/v)$, leading to (35) as a candidate Q.

REFERENCES

CHOW, Y. S. (1965). Local convergence of martingales and law of large numbers. *Ann. Math. Statist.* **36** 552–558.

CHOW, Y. S. and ROBBINS, H. E. (1965). On optimal stopping rules for s_n/n. To appear in *Illinois J. Math.*

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

DUBINS, L. E. and SAVAGE, L. J. (1965a). A Tchebycheff-like inequality for stochastic processes. *Proc. Nat. Acad. Sci. U.S.A.* **53** 274–275.

DUBINS, L. E. and SAVAGE, L. J. (1965b). *How to Gamble if You Must*. McGraw-Hill, New York.

LÉVY, PAUL (1937). *Theorie de l'addition des variables aléatoires*. Gauthier-Villars, Paris.

LOÈVE, MICHEL (1963). *Probability Theory, 3rd edition*. Van Nostrand, New York.