

ON THE APPLICATION OF THE BOREL-CANTELLI LEMMA

BY

K. L. CHUNG⁽¹⁾ AND P. ERDÖS

Consider a probability space (Ω, \mathcal{C}, P) and a sequence of events (\mathcal{C} -measurable sets in Ω) $\{E_k\}$, $k=1, 2, \dots$. The upper (or outer) limiting set of the sequence $\{E_k\}$ is defined by

$$\limsup E_k = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k.$$

We recall that the events E_k are said to be (mutually) independent (with respect to the probability measure P) if for any finite number of distinct subscripts k_1, \dots, k_s we have

$$P(E_{k_1} \cdots E_{k_s}) = P(E_{k_1}) \cdots P(E_{k_s}).$$

The celebrated Borel-Cantelli lemma asserts that

(A) If $\sum P(E_k) < \infty$, then $P(\limsup E_k) = 0$;

(B) If the events E_k are independent and if $\sum P(E_k) = \infty$, then $P(\limsup E_k) = 1$. In intuitive language $P(\limsup E_k)$ is the probability that the events E_k occur "infinitely often" and will be denoted by $P(E_k \text{ i.o.})$. This lemma is the basis of all theorems of the strong type in probability theory. Its application is made difficult by the assumption of independence in part (B). As Borel already noticed [1, p. 48 ff.], this assumption can be removed if we assume that⁽²⁾

$$(0) \quad \sum P(E_k | E'_1 \cdots E'_{k-1}) = \infty$$

where $P(F|E)$ denotes the conditional probability of F on the hypothesis of E and E' denotes the complement of E . Although Borel used the condition (0) successfully in his pioneering work on the metric theory of continued fractions, it is too stringent for many purposes. To overcome the difficulty one usually constructs a sequence of independent events out of the given sequence and applies (B) to the new one. This is the device used for instance in the proof of the law of the iterated logarithm and similar theorems. There is however another group of strong theorems to which this method does not

Received by the editors June 14, 1951.

⁽¹⁾ Research supported by an ONR contract.

⁽²⁾ Clearly we may suppose that $P(\bigcup_{j=1}^k E_j) < 1$ for every k so that the conditional probability is defined. *Added in proof.* Conditions like (0) were used a great deal by Paul Lévy and other authors in generalizations to dependent variables; however, that is not what we have in mind here.

seem to apply. The following theorem furnishes an alternative method which may be of fairly general applicability. On the other hand it does not seem to apply to the law of the iterated logarithm, etc. Two examples given below will serve as illustrations, of which the second concerns the arcsin law.

THEOREM 1. *Let $\{E_k\}$ be a sequence of events satisfying:*

(i) $\sum P(E_k) = \infty.$

(ii) *For every pair of positive integers h, n with $n \geq h$ there exist $c(h)$ and $H(n, h) > h$ such that for every $k \geq H(n, h)$ we have*

(1) $P(E_k | E'_h \cdots E'_n) > cP(E_k).$

(iii) *There exist two absolute constants c_1 and c_2 with the following property: to each E_j there corresponds a set of events E_{j_1}, \dots, E_{j_s} belonging to $\{E_k\}$ such that*

(2)
$$\sum_{i=1}^s P(E_j E_{j_i}) < c_1 P(E_j)$$

and if $k > j$ but E_k is not among the E_{j_i} ($1 \leq i \leq s$) then

(3) $P(E_j E_k) < c_2 P(E_j) P(E_k).$

Then $P(E_k \text{ i.o.}) = 1.$

A defense of the assumptions made seems in order. The conditions (i) and (ii) together resemble Borel's condition (0) but actually they are very much weaker. The point is that the function $H(n)$ is at our disposal and can be chosen of an infinitely greater order of magnitude than n . To put it in a picturesque way, (iii) requires only that the *arbitrarily* remote past should have no overwhelming effect on the present which is certainly a state of affairs to be hoped for in probability problems. As regards the additional conditions in (iii), they involve only joint probabilities of *pairs* of events, or what is sometimes referred to as dependence to the second order; part (2) would usually deal with the dependence at close range while (3) deals with the general situation.

Before proceeding to the proof we shall state a simple lemma.

LEMMA. *Let $\{F_k\}, k = 1, \dots, N,$ be an arbitrary sequence of events in $(\Omega, \mathcal{C}, P).$ We have, if $P(\bigcup_{k=1}^N F_k) > 0,$*

(4)
$$2 \sum_{1 \leq j < k \leq N} P(F_j F_k) \geq \left[P\left(\bigcup_{k=1}^N F_k\right) \right]^{-1} \left(\sum_{k=1}^N P(F_k) \right)^2 - \sum_{k=1}^N P(F_k).$$

Proof. Define random variables $X_k(\omega), \omega \in \Omega,$ as follows:

$$X_k(\omega) = \begin{cases} 0 & \text{if } \omega \notin F_k, \\ 1 & \text{if } \omega \in F_k. \end{cases}$$

The following identity is evident:

$$(5) \quad 2 \sum_{1 \leq j < k \leq N} P(F_j F_k) = E\{(X_1 + \dots + X_N)^2\} - E(X_1^2 + \dots + X_N^2).$$

Now by the Schwarz inequality we have

$$(6) \quad [E(X_1 + \dots + X_N)]^2 \leq P(X_1 + \dots + X_N > 0) E\{(X_1 + \dots + X_N)^2\}.$$

Since $E(X_k) = E(X_k^2) = P(F_k)$, $P(X_1 + \dots + X_N > 0) = P(\bigcup_{k=1}^N F_k)$ by definition, (4) follows from (5) and (6).

Proof of Theorem 1. Let

$$B_h = \bigcup_{k=h}^{\infty} E_k.$$

Since $(E_k \text{ i.o.}) = \lim_{h \rightarrow \infty} P(B_h)$, it is sufficient to prove that $P(B_h) = 1$ for every h . Suppose that this is not true for a certain h ; let $P(B_h) = 1 - \delta$, $\delta > 0$. Thus

$$(7) \quad P\left(\bigcap_{k=h}^{\infty} E'_k\right) = \delta > 0.$$

Given any ϵ , $0 < \epsilon < 1 - \delta$, we can find n such that $P(\bigcup_{k=h}^n E_k) > 1 - \delta - \epsilon$ so that if we write $D_{h,n} = \bigcup_{k=h}^{\infty} E_k - \bigcup_{k=h}^n E_k$, we have

$$(8) \quad P(D_{h,n}) < \epsilon.$$

We have by (1) and (7), if $k > H(n)$,

$$(9) \quad P(E_k E'_h \dots E'_n) > c\delta P(E_k).$$

Hence by (i), $\sum_{k=H(n)}^{\infty} P(E_k E'_h \dots E'_n) = \infty$. Therefore there exists an integer $H'(n) > H(n)$ such that $(H = H(n), H' = H'(n))$

$$(10) \quad 1 < \sum_{k=H}^{H'} P(E_k E'_h \dots E'_n) \leq 2.$$

From (9) and (10) we obtain

$$(11) \quad \sum_{k=H}^{H'} P(E_k) < \frac{2}{c\delta}.$$

From (2), (3), and (11) we have

$$(12) \quad \begin{aligned} \sum_{H \leq j < k \leq H'} P(E_j E_k) &\leq c_1 \sum_{j=H}^{H'} P(E_k) + c_2 \sum_{H \leq j < k \leq H'} P(E_j) P(E_k) \\ &< \frac{2c_1}{c\delta} + \frac{c_2}{2} \left(\frac{2}{c\delta}\right)^2 = c_3(\delta), \end{aligned}$$

where $c_3(\delta)$ is a constant defined by the last equality.

Now let $F_k = E_k E'_k \cdots E'_n$, $H \leq k \leq H'$. It is obvious that $\bigcup_{k=H}^{H'} F_k$ is a subset of $D_{h,n}$, hence by (8),

$$(13) \quad P\left(\bigcup_{k=H}^{H'} F_k\right) < \epsilon.$$

From (10) and (11) we have

$$(14) \quad 1 < \sum_{k=H}^{H'} P(F_k) \leq \sum_{k=H}^{H'} P(E_k) < \frac{2}{c\delta}.$$

Applying the lemma to $\{F_k\}$, $H \leq k \leq H'$, we obtain using (13) and (14)

$$(15) \quad 2 \sum_{H \leq j < k \leq H'} P(E_j E_k) \geq 2 \sum_{H \leq j < k \leq H'} P(F_j F_k) \geq \frac{1}{\epsilon} - \frac{2}{c\delta}.$$

But (12) and (15) are incompatible for sufficiently small ϵ . This contradiction proves that $\delta = 0$. Hence $P(B_h) = 1$. q.e.d.

In the two applications given below we shall treat only the simplest Bernoullian case, since we are more interested in the principle involved than the technical difficulties. It is not hard to generalize Theorems 2 and 3 to fairly general lattice cases or even continuous cases. It will be seen from their proofs that only certain asymptotic formulas and a kind of boundedness of S_n , with probability one or even in probability, are required. These are available in more general cases, thanks to various modern limit theorems.

THEOREM 2. *Let $\{X_k\}$, $k = 1, 2, \dots$, be independent random variables and each X_k assume the values $+1$ and -1 with probabilities $1/2$ and $1/2$. Let $S_n = \sum_{k=1}^n X_k$. Let $\{n_i\}$, $i = 1, 2, \dots$, be an increasing sequence of even integers such that there exists an absolute constant A with the property that*

$$(16) \quad n_{i+1} - n_i > A n_i^{1/2}.$$

Then

$$P(S_{n_i} = 0 \text{ i.o.}) = \begin{cases} 0 \\ 1 \end{cases}$$

according as

$$\sum_i n_i^{-1/2} \begin{cases} < \\ = \end{cases} \infty.$$

REMARK. The theorem in the divergence case is not true without some such condition as (16). Example: Take $\{n_i\}$ to be the sequence of even integers in the intervals $[k^8, k^8 + k^5]$, $k = 1, 2, \dots$. For an alternative condi-

tion and proof see [3, p. 1009].

Proof. The convergence case follows directly from part (A) of the Borel-Cantelli lemma without the condition (16).

Next, let E_i denote the event $S_{n_i} = 0$. We know that

$$P(E_i) = P(S_{n_i} = 0) \sim (2/\pi n_i)^{1/2}.$$

Hence condition (i) in Theorem 1 is satisfied.

To verify the condition (ii) in Theorem 1 we notice that $|S_{n_i}| \leq n_i$, hence

$$\begin{aligned} P(S_{n_k} = 0 \mid S_{n_h} \neq 0, \dots, S_{n_i} \neq 0) \\ &\geq \text{Min}_{|x| \leq n_i} P(S_{n_k} = 0 \mid S_{n_i} = x) = \text{Min}_{|x| \leq n_i} P(S_{n_k} - S_{n_i} = -x) \\ &= \text{Min}_{|x| \leq n_i} P(S_{n_k - n_i} = -x). \end{aligned}$$

Now we have, if $x^2 = o(n)$,

$$P(S_n = x) = C \binom{n}{(n-x)/2} \frac{1}{2^n} \sim \left(\frac{2}{\pi n}\right)^{1/2} e^{-x^2/2n} \sim \left(\frac{2}{\pi n}\right)^{1/2}.$$

We choose $H(i)$ sufficiently large so that if $k > H(i)$, then $n_i^2 = o(n_k)$. Then we have for all $|x| \leq n_i$,

$$P(S_{n_k - n_i} = -x) \sim \left(\frac{2}{\pi n_k}\right)^{1/2}.$$

Therefore we have for all $h, i \geq h$ and $k \geq H(i)$ and any fixed $c < 1$, if $H(1)$ is sufficiently large,

$$P(S_{n_k} = 0 \mid S_{n_h} \neq 0, \dots, S_{n_i} \neq 0) > cP(S_{n_k} = 0).$$

Thus condition (ii) in Theorem 1 is satisfied.

To verify the condition (iii) in Theorem 1 we have

$$\begin{aligned} P(E_j E_k) &= P(S_{n_j} = 0)P(S_{n_k - n_j} = 0) \\ &\sim P(S_{n_j} = 0) \left(\frac{2}{\pi(n_k - n_j)}\right)^{1/2} \\ (17) \quad &\sim P(S_{n_j} = 0)P(S_{n_k} = 0) \left(\frac{n_k}{n_k - n_j}\right)^{1/2} \\ &= P(E_j)P(E_k) \left(\frac{n_k}{n_k - n_j}\right)^{1/2}. \end{aligned}$$

If $n_k > 2n_j$, then

$$(18) \quad \left(\frac{n_k}{n_k - n_j}\right)^{1/2} < 2^{1/2}.$$

We call the events E_k with $n_j < n_k \leq 2n_j$ the events E_{j_i} ($1 \leq i \leq s$) associated with each E_j . We have as before

$$(19) \quad \sum' P(E_j E_k) \sim P(E_j) \sum' (n_k - n_j)^{-1/2}$$

where the summation extends to those k for which $n_j < n_k \leq 2n_j$. From (16) we deduce that if $k > j$ (A_1 denoting an absolute constant),

$$n_k - n_j > A_1(k^2 - j^2).$$

Let N denote the number of k 's satisfying $n_j < n_k \leq 2n_j$. From the last inequality we deduce that $n_j + A_1(N^2 + 2jN) \leq n_{j+N} \leq 2n_j$. Hence we have

$$(20) \quad N \leq \left(\frac{n_j}{A_1} \right)^{1/2}.$$

Now using the Schwarz inequality, (16), and (20) we obtain

$$(21) \quad (\sum' (n_k - n_j)^{-1/2})^2 \leq N \sum' (n_k - n_j)^{-1} \leq \left(\frac{n_j}{A_1} \right)^{1/2} \frac{1}{A_1 n_j^{1/2}} = \left(\frac{1}{A_1 A^2} \right)^{1/2}.$$

(19) and (21) give (2) while (17) and (18) give (3). q.e.d.

THEOREM 3. *Let $\{X_n\}$ be as in Theorem 2 and let N_n denote the number of positive terms among S_1, \dots, S_n . Let $\phi(n)$ be an increasing function of n . Then*

$$(22) \quad P\left(N_n \leq \frac{n}{\phi(n)} \text{ i.o.} \right) = \begin{cases} 0 \\ 1 \end{cases}$$

according as

$$(23) \quad \sum \frac{1}{n(\phi(n))^{1/2}} \begin{cases} < \\ = \end{cases} \infty.$$

REMARK. This is the strong theorem corresponding to the now celebrated arcsin law. On grounds of symmetry we may replace the left side in (22) by $P(N_n \geq n(1 - 1/\phi(n)) \text{ i.o.})$.

Proof. Standard arguments⁽³⁾ show that we may suppose that $\phi(n) \leq n^\epsilon$ for some $0 < \epsilon < 1/2$. The convergence case follows easily from the arcsin law for Bernoullian variables (see [2, p. 252]; the convention made there regarding the "positiveness" of S_n makes no difference in the asymptotic formula below), which asserts that

$$P\left(N_n \leq \frac{n}{\phi(n)}\right) \sim \frac{2}{\pi(\phi(n))^{1/2}}.$$

To prove the theorem in the divergence case we note first that the di-

(3) Cf e.g. [3. p. 1010].

vergence of the series in (23) implies that of

$$\sum_n \frac{1}{n(\phi(n^r))^{1/2}}$$

for any $r > 0$ (proof by the integral test). Let $\phi(n^2) = \psi(n)$. Define E_k to be the event

$$S_{2k} = 0, \quad S_i < 0 \quad \text{for } 2k < i \leq 2k\psi(k).$$

Obviously E_k implies that $N_{2k\psi(k)} \leq 2k$. Writing $2k\psi(k) = n$, we have, since $\psi(k) \leq k^{2\epsilon}$, $k \geq n'$ where $n' = (2^{-1}n)^{1/(1+2\epsilon)}$. Hence $N_n \leq n/\psi(n')$. For all sufficiently large n , $\psi(n') \geq \phi(n)$. Hence in order to prove the second part in (22) it is sufficient to prove that $P(E_k \text{ i.o.}) = 1$.

It is known that (see e.g. [2, p. 252]) $P(S_i < 0 \text{ for } 0 < i \leq n) \sim bn^{-1/2}$ for some absolute constant $b > 0$. Hence we have

$$(24) \quad \begin{aligned} P(E_k) &= P(S_{2k} = 0)P(S_i < 0 \text{ for } 0 < i \leq 2k\psi(k) - 2k) \\ &\sim bk^{-1/2}(k\psi(k))^{-1/2} = bk^{-1}(\psi(k))^{-1/2}. \end{aligned}$$

Hence condition (i) in Theorem 1 is satisfied.

To verify condition (ii) in Theorem 1 we note that (without loss of generality we may suppose $n\psi(n)$ to be an integer for all n), if $k > H(n)$,

$$\begin{aligned} P(E_k \mid S_1 = x_1, \dots, S_{2n\psi(n)} = y) &= \sum_x P(E_k \mid S_{H(n)} = x)P(S_{H(n)} = x \mid S_{2n\psi(n)} = y) \\ &= \sum_x P(E_k \mid S_{H(n)} = x)P(S_{H(n)-2n\psi(n)} = x - y) \end{aligned}$$

where x_1, \dots, y, x are integers. Now $|y| \leq 2n\psi(n)$, hence if we choose $H(n)$ sufficiently large, $P(S_{H(n)-2n\psi(n)} = x - y) \sim P(S_{H(n)} = x)$ as $n \rightarrow \infty$, at least if x is within a certain range, say $|x| \leq H(n)^{1/2+\eta}$, $\eta > 0$. (This is because of the limitations of the Gaussian approximation.) But the other range of x is negligible in the sense that

$$\begin{aligned} \sum_{|x| > H(n)^{1/2+\eta}} P(S_{H(n)-2n\psi(n)} = x - y) &= O\left(\sum_{|x| \leq H(n)^{1/2+\eta}} P(S_{H(n)-2n\psi(n)} = x - y)\right). \end{aligned}$$

Hence we have

$$\begin{aligned} \text{Min}_{|v| \leq 2n\psi(n)} P(E_k \mid S_1 = x_1, \dots, S_{2n\psi(n)} = y) &\sim \sum_x P(E_k \mid S_{H(n)} = x)P(S_{H(n)} = x) = P(E_k). \end{aligned}$$

This implies condition (ii) in Theorem 1.

To verify condition (iii) in Theorem 1, let $j < k$. If $k \leq j\psi(j)$, then $P(E_j E_k) = 0$. If $k > j\psi(j)$ we have

$$(25) \quad \begin{aligned} P(E_k | E_j) &= P(S_{2k} = 0 | S_{2j} = 0, S_i < 0 \text{ for } 2j < i \leq 2j\psi(j)). \\ P(S_i < 0 \text{ for } 2k < i \leq 2k\psi(k) | S_{2k} = 0) &= P_1 \cdot P_2. \end{aligned}$$

Now for every x we have

$$P(S_{2k} = 0 | S_{2j\psi(j)} = x) = P(S_{2k-2j\psi(j)} = -x) \leq b(k - j\psi(j))^{-1/2}.$$

P_1 being a probability mean of such probabilities we have

$$P_1 \leq b(k - j\psi(j))^{-1/2}.$$

As for P_2 we have as in (24),

$$P_2 \sim b(2k\psi(k) - 2k)^{-1/2} \sim b(2k\psi(k))^{-1/2}.$$

Therefore we obtain from (25),

$$(26) \quad P(E_j E_k) \leq b_1 P(E_j) (k - j\psi(j))^{-1/2} (k\psi(k))^{-1/2},$$

where b_1 (as b_2, b_3 later) is an absolute constant. Now for every E_j we define $E_{j,i}$, $1 \leq i \leq s$, to be those E_k with $j\psi(j) < k \leq 2j\psi(j)$. We have then by (26)

$$\begin{aligned} \sum_{i=1}^s P(E_j E_{j,i}) &\leq b_1 P(E_j) (k\psi(k))^{-1/2} \sum_{i=1}^{j\psi(j)} i^{-1/2} \\ &\leq b_2 P(E_j). \end{aligned}$$

On the other hand if $k > 2j\psi(j)$, then $k - j\psi(j) > k/2$, hence by (26) and (24)

$$P(E_j E_k) \leq b_3 P(E_j) P(E_k).$$

Therefore condition (iii) in Theorem 1 is satisfied. q.e.d.

REFERENCES

1. E. Borel, *Traité du calcul des probabilités et de ses applications*, vol. 2, no. 1, *Applications à l'arithmétique et à la théorie des fonctions*, Paris, Gauthier-Villars, 1926.
2. W. Feller, *An introduction to probability theory and its applications*, vol. 1, New York, Wiley, 1950.
3. K. L. Chung and P. Erdős, *On the lower limit of sums of independent random variables*, Ann. of Math. vol. 48 (1947) p. 1003.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.
UNIVERSITY OF ABERDEEN,
ABERDEEN, SCOTLAND.