The Annals of Probability
2001, Vol. 29, No. 4, 1426–1434

THE BOREL–CANTELLI LEMMAS, PROBABILITY LAWS AND KOLMOGOROV COMPLEXITY

BY GEORGE DAVIE

University of South Africa

We formulate effective versions of the Borel–Cantelli lemmas using a coefficient from Kolmogorov complexity. We then use these effective versions to lift the effective content of the law of large numbers and the law of the iterated logarithm.

1. Introduction. Our aim is to show that ideas from Kolmogorov complexity furnish tools that may be very useful for probability theory. In particular, we will use a coefficient which is naturally associated with every Kolmogorov–Chaitin random sequence, to effectivize the Borel–Cantelli lemmas. The effectivized lemmas in their turn, lift the effective content of the law of large numbers and the law of the iterated logarithm. Similar effectivizations of many other probability laws should be possible.

2. Background. In 1966 Martin-Löf defined a set of sequences of Lebesgue measure one, the Martin-Löf random sequences, which satisfy all effective probability laws. Effective means here that the complement of the set of sequences satisfying the probability law, must be covered by each of a computably enumerable sequence of sets of intervals $O_1, O_2, \ldots$ such that $\mu(O_i) < 2^{-i}$, where μ denotes the Lebesgue measure. Since all common probability laws can be written in this form or as a countable intersection of sets of this form, all Martin–Löf random sequences satisfy laws such as the law of large numbers and the law of the iterated logarithm. In particular, the definition of Martin–Löf allows us to replace the standard formulation of probability theory:

$$\mathbf{P} \text{ holds with probability one}$$

with:

$$\mathbf{P} \text{ holds for each (Martin–Löf) random infinite sequence } \omega,$$

for all effective predicates $\mathbf{P}$. Since the Lebesgue measure of the Martin–Löf random sequences is one, this increases the effectivity of the first statement.

Chaitin proposed to identify a different set of sequences (the Kolmogorov–Chaitin random sequences, defined below) as the set of intuitively random sequences. The definitions of Chaitin and Martin–Löf were shown to define

Received September 2000; revised January 2001.
AMS 2000 *subject classifications*. 68Q30, 60A05.
Key words and phrases. Effective Borel-Cantelli lemmas, Kolmogorov complexity, compressibility coefficient, probability law.

the same set, by C. P. Schnorr, as refered to [4]. This means that every Kolmogorov–Chaitin random sequences satisfies all effective probability laws. For more details the reader is referred to the standard reference for Kolmogorov complexity, [2].

NOTATION 1. We follow the notation of [2]. We will denote the set of natural numbers by **N**. For ω an infinite binary sequence, we let $\omega_{1:n}$ denote the initial segment of ω of length n and let $S_n(\omega)$ denote the sum of the digits of $\omega_{1:n}$.

3. Definitions from Kolmogorov complexity. We will use the phrase "given n we can effectively find $f(n)$" to mean that there exists an algorithm which, on input n, will output $f(n)$, for all n.

The two definitions from Kolmogorov complexity that we will use, are the following:

DEFINITION 1. Let U be a universal prefix Turing machine (U is defined on a prefix free set) which takes binary strings as inputs and gives as output binary strings. Let x be a finite binary string. The prefix complexity $K(x)$ of x relative to U is the length of a shortest program for U which will output x on the empty input.

DEFINITION 2. An infinite binary sequence ω is Kolmogorov–Chaitin random if there exists a natural number c such that $K(\omega_{1:n}) > n - c$ for all n.

Hence, by this definition, no initial segment $\omega_{1:n}$ is the output, on the empty input, of a program of length less than $n - c$. Or roughly, no initial segment of ω is "compressible" by more than c. As mentioned above, the set thus defined consists of exactly the Martin–Löf random sequences, hence in particular, has measure one. This is somewhat surprising as no infinite binary sequence satisfies the condition in the definition above if we do *not* require our program set to be prefix free (i.e., we do not require that no program is an initial segment of any other). Intuitively, the reason it works now is that we have fewer programs of a given length and the complexity of initial segments $\omega_{1:n}$ of infinite sequences in fact grows a bit quicker than the length n.

DEFINITION 3. Let ω be random and let $c(\omega)$ be the smallest c for which the condition in Definition 2 holds. We will call $c(\omega)$ the *compressibility coefficient* of ω. We denote by K^c the set of infinite sequences ω for which $c(\omega) \leq c$.

We therefore have that for a random ω there exists (at least one) n such that $K(\omega_{1:n}) \leq n - c(\omega)$ but for no $c > c(\omega)$ do we have $K(\omega_{1:n}) \leq n - c$ for any n.

4. The Borel–Cantelli lemmas. The Borel–Cantelli lemmas play the central role in the proofs of many probability laws including the law of large

numbers and the law of the iterated logarithm. We will use the above defined compressibility coefficient to lift the effective content of the Borel–Cantelli lemmas for those cases in which the associated sets are computably enumerable and have as Lebesgue measure of their union, a computable real. Recall that

DEFINITION 4. A binary real π is computable if there is an algorithm ϕ, which on input $n \in \mathbf{N}$, gives as output the first n digits in the binary expansion of π. In case there are two possible binary expansions for a computable real (e.g., 1.0 and $0.\dot{1}$) we choose the non-terminating expansion.

We state the lemmas in their conventional form:

THEOREM 1 (Borel–Cantelli lemmas). *Let $A_1, A_2, \ldots$ be an infinite sequence of events each of which depends only on a finite number of digits of a binary sequence. Denote the probability of A_k occurring by P_k.*

(i) *If $\sum P_k$ converges, then with probability one only finitely many of the events A_k occur.*
(ii) *If the events A_k are mutually independent, and if $\sum P_k$ diverges, then with probability one, infinitely many of the events A_k occur.*

We will use the notation A_i both for the event A_i and for the set of sequence in which A_i occurs. Hence $\mu(A_i)$ will denote the Lebesgue measure of the set of sequences for which A_i occurs. That is, $\mu(A_i)$ is the probability that A_i occurs. When dealing with enumerations of intervals we will often view a binary string x as the dyadic interval $[0.x, 0.x + 2^{-|x|})$.

We will now use the compressibility coefficient to prove the following effective forms:

THEOREM 2 (Borel–Cantelli lemmas–effective forms). *Let $A_1, A_2, \ldots$ be an infinite sequence of events such that each of the events depends only on a finite number of digits of a binary sequence. Suppose further that each of the sets A_i is computably enumerable and that there is an algorithm ϕ which on input (i, k) gives as output the kth interval in the enumeration of A_i, then*:

(i) *If $\sum_{i=1}^{\infty} \mu(A_i)$ converges to a computable binary real π, then we can effectively find, for each $c \in N$, an $n(c)$ such that none of the events $A_m; m > n(c)$ occur in any $\omega \in K^c$.*
(ii) *If* (a) *the events A_k are mutually independent and*
(b) *$\sum_{i=1}^{\infty} \mu(A_i)$ diverges, then we can effectively find, for each m and $c \in N$, an $n(c, m)$ such that at least one of the events $A_n; m < n < n(c, m)$ must occur for each $\omega \in K^c$.*

The idea in both parts of the proof will be that a sequence with fixed compressibility c cannot be part of a small set of low Kolmogorov complexity as we could then specify some initial segment of the sequence of length l with fewer than $l - c$ digits, giving a contradiction.

PROOF. (i) Given c and n we use the algorithm ϕ to dovetail an enumeration E of all elements of $\bigcup_{i=1}^{\infty} A_i$ until $\pi - \mu(E_s)$ is smaller than $2^{-c-2\log c}$, where E_s is the approximation to E at stage s (i.e., $E = \lim_{s\to\infty} E_s$). Consider now any sequence ω which has not appeared by this time.

Let ω appear between the following two occurrences:

The measure of the enumerated set reaches $\pi_{1:c+2\log c+s}$ and the measure of the enumerated set reaches $\pi_{1:c+2\log c+s+1}$. (Both may occur together, of course, in which case we have only one interval to consider.)

Let the maximum of the lengths of the strings enumerated between these two occurrences be l.

Now list all continuations of all listed strings up to length l and call the set thus obtained S_l.

$\omega_{1:l}$ is then contained in S_l and hence is specifiable by giving its position in S_l. Since S_l has associated Lebesgue measure less than $2^{-(c+2\log c+s)}$, there are at most $2^{l-(c+2\log c+s)}$ strings in S_l.

We can therefore specify the *position* of $\omega_{1:l}$ in S_l using at most

$$l - (c + 2\log c + s)$$

digits.

Hence, if p_1 is a program for ϕ and p_2 is a program for π, then $K(\omega_{1:l}) < l + |p_1| + |p_2| + 2\log s + 2\log c + k - (c + 2\log c + s)$ since a program p_3 with $|p_3| = k$ outputting $\omega_{1:l}$, will need as inputs p_1, p_2, c and s.

We can therefore clearly choose, for given c, a c' such that no $\omega \in K^c$ can be enumerated after the measure of the enumerated intervals reaches $\pi_{1:c'+2\log c'+s}$.

(ii) Given n, consider the sequences of events

$$A_n, A_{n+1}, A_{n+2}, \ldots.$$

The probability that none of the events $A_n, A_{n+1}, A_{n+2}, \ldots$ take place in an arbitrary ω is of course $\lim_{k\to\infty}(\prod_{i=n}^{k}(1 - \mu(A_i)) = 0$. Given therefore c and m, we can write a program p for ϕ to dovetail an enumeration of sequences in $A_i, i \geq n$ until

$$(1 - \mu(A_n(s)))(1 - \mu(A_{n+1}(s)))(1 - \mu(A_{n+2}(s)))\cdots$$

is smaller than $2^{-c-2\log c-2\log m-|p|-1}$, where $A_i(s)$ is the approximation to A_i at stage s of the enumeration [i.e., $A_i = \lim_{s\to\infty} A_i(s)$].

Let k be the largest index for which strings in A_k are enumerated and let the maximum length of enumerated strings be l and, as in (i) pad all the strings up to length l and call this set S_l. Now make the (finite) list all strings in S_l^C.

Now, the Lebesgue measure of S_l^C is at most $2^{-c-2\log c-2\log m-|p|-1}$ and hence S_l^C contains at most $2^{l-c-2\log c-2\log m-|p|-1}$ strings.

The position of any $\omega \in S_l^C$ can therefore be specified using at most $l - c - 2\log c - 2\log m - |p| - 1$ digits. To specify S_l^C we of course also need p, m and

c hence at most a further $|p| + 2\log c + 2\log m$ digits. Our $\omega_{1:l}$ can therefore have complexity at most

$$\begin{aligned} &l - c - 2\log c - 2\log m - |p| + 2\log c + 2\log m + |p| - 1 \\ &\quad < l - c. \end{aligned}$$

Therefore any ω for which none of the events $A_n, A_{n+1}, \ldots, A_k$ hold is in K^g for $g > c$ so at least one of these events must hold for all $\omega \in K^d, d \leq c$. □

5. Application to two probability laws. In the context of binary sequences, the strong law of large numbers (first formulated by Cantelli; see [1] for references) is the following:

THEOREM 3 (Strong law of large numbers). *With probability one we have*

$$\frac{S_n}{n} \to \frac{1}{2}.$$

In the words of [1]: with probability one $\frac{S_n}{n} - \frac{1}{2}$ "becomes *and remains* small." The law of large numbers is clearly equivalent to the following:

THEOREM 4 (The law of large numbers). *For every $\varepsilon > 0$, with probability one, there occur only finitely many of the events*

$$\left| \frac{S_n}{n} - \frac{1}{2} \right| > \varepsilon.$$

The law of the iterated logarithm (due to Khintchine; see [1] for references) gives upper bounds for the fluctuations of

$$S_n^* = \frac{S_n - \frac{n}{2}}{\frac{1}{2}\sqrt{n}}.$$

THEOREM 5 (The law of the iterated logarithm–Khintchine). *With probability one we have*

$$\limsup_{n\to\infty} \frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{2}\log\log n}} = 1.$$

This means: *For $\lambda > 1$, with probability one, only finitely many of the events*

$$S_n > \frac{n}{2} + \lambda\sqrt{\frac{n}{2}\log\log n}$$

occur; *and for $\lambda < 1$, with probability one, infinitely many of the events*

$$S_n > \frac{n}{2} + \lambda\sqrt{\frac{n}{2}\log\log n}$$

occur.

6. Ineffectivity of the probability laws. Note that these two laws are non-effective on (at least) two counts. Given a randomly chosen infinite binary sequence ω, the laws firstly only hold with *probability one* and, secondly, we are told nothing about the *waiting times* involved. That is, for a given ω, nothing is stated about any of the following:

1. In the strong law of large numbers: For a given ε, the largest m for which

$$\left|\frac{S_m}{m} - \frac{1}{2}\right| > \varepsilon$$

or

2. In the law of the iterated logarithm:
   (i) For a given $\lambda > 1$, the largest n for which

$$S_n > \frac{n}{2} + \lambda\sqrt{\frac{n}{2}\log\log n}$$

or
   (ii) For a given $\lambda < 1$ and n, the smallest $n' > n$ such that

$$S_{n'} > \frac{n'}{2} + \lambda\sqrt{\frac{n'}{2}\log\log n'}.$$

We now consider the above objections. As mentioned in the first paragraph, the definitions of Martin–Löf and Chaitin means that we can change the statement

P holds with probability one

to

P holds for each random ω.

We are now also able to address the waiting time objection using the effectivized Borel–Cantelli lemmas.

7. Effective forms of the probability laws. We use the standard proofs by Feller [1], as reference when discussing the effective forms of the two laws.

THEOREM 6 (Strong law–effective form). *For any given c and ε we can find effectively an $n(c, \varepsilon)$ such that, if $\omega \in K^c$ then, for all $n > n(c, \varepsilon)$,*

$$\left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| < \varepsilon.$$

PROOF SKETCH. Following [1], let $a > 1$ and let A_k be the event

$$|S_k^*| = \left| \frac{S_k - \frac{k}{2}}{\sqrt{\frac{k}{4}}} \right| \geq \sqrt{2a \log k}.$$

Now, since

$$\frac{(2a \log k)^{\frac{3}{2}}}{\sqrt{k}} \to 0$$

it follows (see, e.g., page 192 of [1]) that we can choose n large enough such that

$$P\left\{|S_n^*| > \sqrt{2a \log n}\right\} < e^{-a \log n} = \frac{1}{n^a}.$$

Since $\sum_{k=1}^{\infty} \frac{1}{k^a}$ converges for $a > 1$, we can find for any given σ, an n such that $\sum_{k=n}^{\infty} P\{|S_k^*| > \sqrt{2a \log k}\} < \sum_{k=n}^{\infty} \frac{1}{k^a} < 2^{-\sigma}$. So for given ε and a, the measure of the union of the events A_k converges effectively, hence is a computable real. The law of large numbers then follows by Theorem 2(i). □

THEOREM 7 (Law of the iterated logarithm–effective form). (i) *For a given c and $\lambda > 1$, we can find effectively an $n(c, \lambda)$ such that, if $\omega \in K^c$ then, for all $n > n(c, \lambda)$,*

$$S_n(\omega) \leq \frac{n}{2} + \lambda \sqrt{\frac{n}{2} \log \log n}.$$

(ii) *For a given c, $\lambda < 1$ and $m \in \mathbf{N}$, we can find effectively an $n(c, \lambda, m)$ such that, if $\omega \in K^c$ then, for some n such that $m \leq n \leq n(c, \lambda, m)$,*

$$S_n(\omega) > \frac{n}{2} + \lambda \sqrt{\frac{n}{2} \log \log n}.$$

PROOF SKETCH OF (i). Following [1], let $\lambda > 1$, let γ be a number between 1 and λ and let n_r be the integer nearest to γ^r. Let B_r be the event that the inequality

$$S_n - \frac{n}{2} > \lambda \sqrt{\frac{n_r}{2} \log \log n_r}$$

holds for at least one n with $n_r \leq n \leq n_{r+1}$.

We will show that $\sum P\{B_r\}$ converges effectively and hence is a computable real.

We now have (see, e.g., page 205 of [1]) that

$$P\{B_r\} \leq \sigma^{-1} P\left\{ S_{n_{r+1}} - \frac{n_{r+1}}{2} > \lambda \sqrt{\frac{n_r}{2} \log \log n_r} \right\}$$

$$= \sigma^{-1} P\left\{ S_{n_{r=1}}^* > \lambda \sqrt{2 \frac{n_r}{n_{r+1}} \log \log n_r} \right\}.$$

Now $n_{r+1}/n_r \sim \gamma < \lambda$, and we can therefore choose r large enough such that

$$P\{B_r\} \leq \sigma^{-1} P\left\{S^*_{n_{r+1}} > \sqrt{2\lambda \log\log n_r}\right\}.$$

And we can thus choose r large enough such that

$$P\{B_r\} \leq \sigma^{-1} e^{-\lambda \log\log n_r} = \frac{1}{\sigma(\log n_r)^\lambda}$$

and $\frac{1}{\sigma(\log n_r)^\lambda}$ is as close as we like to $\frac{1}{\sigma(r\log\gamma)^\lambda}$. Thus we can choose for any given k, an s such that $\sum_{r=s}^\infty P\{B_r\} < 2^{-k}$ and hence $\sum_{r=s}^\infty P\{B_r\}$—the sum of the Lebesgue measures of the events B_r—is a computable number. The first half of the law then follows by Theorem 2(i).

PROOF SKETCH OF (ii). Following [1], let $\lambda < 1$ and choose a number η so close to 1 that

$$1 - \eta < \left(\frac{\eta - \lambda}{2}\right)^2$$

and choose for γ an integer so large that $\frac{\gamma-1}{\gamma} > \eta > \lambda$. Put $n_r = \gamma^r$. Let $D_r = S_{n_r} - S_{n_{r-1}}$ and let A_r be the event

$$D_r - \frac{n_r - n_{r-1}}{2} > \eta\sqrt{\frac{n_r}{2}\log\log n_r}.$$

Note that the events A_r are independent.

Now

$$P\{A_r\} = P\left\{\frac{D_r - (n_r - n_{r-1})^{\frac{1}{2}}}{(n_r - n_{r-1})^{\frac{1}{4}}} > \eta\sqrt{2\frac{n_r}{n_r - n_{r-1}}\log\log n_r}\right\}.$$

Here $n_r/(n_r - n_{r-1}) = \gamma/(\gamma - 1) < \eta^{-1}$. Hence

$$P\{A_r\} \geq P\left\{\frac{D_r - (n_r - n_{r-1})^{\frac{1}{2}}}{\sqrt{(n_r - n_{r-1})^{\frac{1}{4}}}} > \sqrt{2\eta\log\log n_r}\right\}.$$

We can therefore choose r large enough such that

$$P\{A_r\} > \frac{1}{\log\log n_r} e^{-\eta\log\log n_r} = \frac{1}{(\log\log n_r)(\log n_r)^\eta}$$

(using [1], page 192 again.) Since $n_r = \gamma^r$ and $\eta < 1$, we can choose r large enough such that $P\{A_r\} > \frac{1}{r}$. (Showing that $S_{n_{r-1}}$ can be dropped will follow a similar argument to that of the proof of the first half of the law.) □

REMARK. Of course, we may be "given" non-computable numbers for ε and/or λ. We need then merely take $\varepsilon' < \varepsilon$ with ε' computable for the law of large numbers or $1 < \lambda' < \lambda$ (for the first part of the law of the iterated

logarithm), or $1 > \lambda' > \lambda$ (for the second part of the law of the iterated logarithm) with λ' computable and the method will prove the theorem for the non-computable ε and λ.

It is well known that one can examine and prove version of the laws of large numbers and of the iterated logarithm directly from their being random. For example, in the paper [3], the law of the iterated logarithm is examined for sequences of which the Kolmogorov complexity of each initial segment is high. The author's main goal is to examine the range of compressibility which will force the probability laws to hold.

In [2] a general theorem is proved to the effect that if the deficiency function (roughly the deviation from maximally complex) is of a certain form, then the difference between the number of 0's and the number of 1s in each $\omega_{1:n}$ is bounded by another given function in n, implying that the law of large numbers and the first part of the law of the iterated logarithm hold for random sequences.

It is the author's opinion that these results and those of this paper contribute to the idea that Kolmogorov complexity gives us sharper insight into probabilistic phenomena.

REFERENCES

[1] FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications* **1**, 3rd ed. Wiley, New York.

[2] LI, M. and VITÁNYI, P. (1993). *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, New York.

[3] VOVK, V. G. (1987). The law of the iterated logarithm for random Kolmogorov, or chaotic, sequences. *Theory Probab. Appl.* **32** 413–425.

[4] CHAITIN, G. J. (1975). A theory of program size formally identical to information theory. *J. Assoc. Comput. Mach.* **22** 329–340.

[5] CHAITIN, G. J. (1987). *Algorithmic Information Theory*. Cambridge Univ. Press.

DEPARTMENT OF MATHEMATICS,
APPLIED MATHEMATICS AND ASTRONOMY
UNIVERSITY OF SOUTH AFRICA
0003 PRETORIA
SOUTH AFRICA
E-MAIL: davieg@unisa.ac.za