Springer

A History of the Axiomatic Formulation of Probability from Borel to Kolmogorov: Part I

Author(s): Jack Barone and Albert Novikoff

Source: *Archive for History of Exact Sciences*, 8.III.1978, Vol. 18, No. 2 (8.III.1978), pp. 123-190

Published by: Springer

Stable URL: https://www.jstor.org/stable/41133504

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Springer is collaborating with JSTOR to digitize, preserve and extend access to *Archive for History of Exact Sciences*

JSTOR

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

A History of the Axiomatic Formulation of Probability from Borel to Kolmogorov: Part I

JACK BARONE & ALBERT NOVIKOFF

Communicated by M. KLINE

Abstract

This paper, the first of two, traces the origins of the modern axiomatic formulation of Probability Theory, which was first given in definitive form by KOLMOGOROV in 1933. Even before that time, however, a sequence of developments, initiated by a landmark paper of E. BOREL, were giving rise to problems, theorems, and reformulations that increasingly related probability to measure theory and, in particular, clarified the key role of countable additivity in Probability Theory.

This paper describes the developments from BOREL's work through F. HAUSDORFF's. The major accomplishments of the period were BOREL's Zero-One Law (also known as the BOREL-CANTELLI Lemmas), his Strong Law of Large Numbers, and his Continued Fraction Theorem. What is new is a detailed analysis of BOREL's original proofs, from which we try to account for the roots (psychological as well as mathematical) of the many flaws and inadequacies in BOREL's reasoning. We also document the increasing realization of the link between the theories of measure and of probability in the period from G. FABER to F. HAUSDORFF. We indicate the misleading emphasis given to independence as a basic concept by BOREL and his equally unfortunate association of a HEINE-BOREL lemma with countable additivity. Also original is the (possible) genesis we propose for each of the two examples chosen by BOREL to exhibit his new theory; in each case we cite a now neglected precursor of BOREL, one of them surely known to BOREL, the other, probably so. The brief sketch of instances of the "CANTELLI" lemma before CANTELLI's publication is also original.

We describe the interesting polemic between F. BERNSTEIN and BOREL concerning the Continued Fraction Theorem, which serves as a rare instance of a contemporary criticism of BOREL's reasoning. We also discuss HAUSDORFF's proof of BOREL's Strong Law (which seems to be the first valid proof of the theorem along the lines sketched by BOREL).

In retrospect, one may ask why problems of "geometric" (or "continuous") probability did not give rise to the (KOLMOGOROV) view of probability as a form of

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

measure, rather than the study of repeated independent trials, which was BOREL's approach. This paper shows that questions of "geometric" probability were always the essential guide to the early development of the theory, despite the contrary viewpoint exhibited by BOREL's preferred interpretation of his own results.

Table of Contents

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

0. Introduction

The overall purpose of this work is to sketch the highpoints in the development of the axiomatic formulation of Probability Theory in terms of measure theory. (It is not our intention to depreciate thereby other efforts to found the theory of probability on axiomatic bases that do not employ the concept of measure theory.) This axiomatic formulation was first definitively achieved by A. KOLMOGOROV in 1933. The need for an axiomatic foundation for Probability Theory had been stressed by HILBERT as part of the sixth problem in his celebrated list of 1900: "...: To treat ... by means of axioms, those physical sciences (*sic*) in which mathematics plays an important part; in the first rank are the theories of Probability and Mechanics" (HILBERT (1900: 81)). Thus this work can be considered a sketch of the history of part of one of HILBERT's problems. Here we trace the history from the work of BOREL (1909) through the work of HAUSDORFF (1914); subsequently Part II will carry the narrative forward to STEINHAUS, FRÉCHET, CANTELLI, PÓLYA, MAZURKIEWICZ and others, culminating in the work of KOLMOGOROV.

The landmark paper, initiating the modern theory of probability, is E. BOREL's "Les Probabilitiés Dénombrables et Leurs Applications Arithmétiques" of 1909. Here we are mainly concerned with the contents, background, and immediate reactions to this paper. The key figures immediately following BOREL are G. FABER, F. BERNSTEIN, and F. HAUSDORFF. Their contributions will be discussed in turn. Key predecessors are A. WIMAN (1900, 1901), E. VAN VLECK (1908) and, in fact, BOREL himself by virtue of a brief note of 1905. Both WIMAN and VAN VLECK are subjects of separate brief notes; the note on VAN VLECK has already appeared (NOVIKOFF & BARONE (1977)). Of these earlier works, only BOREL's paper of 1905 is discussed at any length here.

Any adequate discussion of the contents of BOREL's landmark paper (which we shall refer to as BOREL (1909)) is of necessity delicate and somewhat detailed. The reason for this is the ironical circumstance that BOREL, the unquestioned founder of measure theory, attempted in 1909 to found a new theory of "denumerable probability" *without* relying on measure theory. The irony is further compounded in the light of BOREL's paper of 1905 which identified "continuous probability" in the unit interval with measure theory there. Consequently, we are at great pains both to establish and to comprehend BOREL's reluctance in 1909 to accept the underlying role of countable additivity in his new theory. The first part of this paper is thus an attempt to examine BOREL's state of mind in 1909, taking into account his earlier insights and his reluctance to exploit them.

BOREL's paper, "Probabilités Dénombrables", falls into three major divisions: a "general theory" (culminating in what we have called the BOREL Zero-One Law), an application of this theory to decimal and dyadic expansions (the BOREL Strong Law, or Strong Law of Large Numbers), and a second application of this theory, this time to Continued Fractions (the BOREL Continued Fraction Theorem).

Since BOREL's paper is as interesting for its defects as for its results (as Professor M. KAC once remarked, "all of its theorems are true but almost all of the proofs are false"), we summarize its shortcomings in Section 4.6.

BOREL'S immediate successors devoted themselves to clarifying both BOREL'S Strong Law and his Continued Fraction Theorem by treating them both within

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

LEBESGUE'S theory of measure on the real line.[1] This clarification, however, was at the expense of drastically diminishing BOREL'S emphasis on the construction of a new general theory of probability concerning *repeated independent trials.*

STEINHAUS (1923) finally incorporated BOREL'S "general theory" into measure theory by exploiting an axiomatic characterization of measure theory due to SIERPIŃSKI (1918). STEINHAUS' work and more generally the increasing abstraction of measure theory itself (initiated by FRÉCHET (1915) and CARÁTHÉODORY (1914)) were events which helped pave the way to KOLMOGOROV's culminating achievement. These developments will be discussed, among others, in Part II.

While we have restrained ourselves from drawing general historiographic conclusions, we believe the contents of our paper (both here and in Part II) furnish ammunition for those who wish to illustrate (i) DIEUDONNÉ's "fusion" hypothesis of mathematical progress (DIEUDONNÉ (1975:537)), (ii) LAKATOS' thesis on the gradual rigorizing of partially understood reasoning via dispute (LAKATOS (1963–64)), and (iii) the doctrine that "special problems create general theories". The present study of BOREL'S work brings to light a stupendous instance of tunnel vision. It also shows that a foundational question (such as HILBERT's) may have unexpectedly complex responses. The solution we trace to HILBERT's problem did not come from a direct attack; it evolved from computations which successfully dealt with a series of particular, well-chosen questions, first raised by BOREL.

1. Brief Sketch of Major Results of Borel (1909)

BOREL considered an infinite sequence of trials, each having only two possible outcomes arbitrarily called "success" and "failure". The probabilities of "success" and "failure" on the n^{th} trial are p_n and q_n, respectively, where

$$0 \leqq p_n,\ q_n \leqq 1,$$

$$p_n + q_n = 1, \qquad n = 1, 2, 3, \ldots.$$

The trials are assumed independent. In contemporary terms such trials are called "Binomial" or "POISSON" trials. In what represents a decisive new step, BOREL asked for the probability A_k that exactly k successes occur in such an infinite sequence ($k = 0, 1, 2, \ldots$) and, most important, the probability A_∞ of the occurrence of infinitely many successes.

There is no notation for the sets (or "events") under consideration in BOREL (1909). In consequence, there are also no unions, complements, intersections, *etc.*, indicated or referred to as such.

[1] This statement must be qualified by emphasizing first that N. WIENER and P. LÉVY worked at a different level, not following BOREL's lead directly, and second PÓLYA, CANTELLI and MAZURKIEWICZ considered rather general "random variables" or trials" as in BOREL'S "general theory", without appealing to measure theory.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

Let us introduce the notation, therefore, that

$$E_k = \text{set of all sequences with exactly } k \text{ successes}, \quad k=0,1,2,\ldots,$$
$$E_\infty = \text{set of all sequences with infinitely many successes.}$$

BOREL then sought expressions for

$$A_k = P(E_k), \quad k=0,1,2,\ldots,$$
$$A_\infty = P(E_\infty)$$

in terms of the given numerical sequence $\{p_n\}$, $n=1,2,\ldots$. BOREL persistently treated separately the cases in which $\sum_1^\infty p_n$ converged or diverged.

The first result, concerning the case $k=0$, asserted that

$$A_0 = (1-p_1)(1-p_2)\ldots(1-p_n)\ldots.$$

Here the right-hand side represents a convergent non-vanishing infinite product if $\sum_1^\infty p_n$ converges and is to be interpreted as zero if $\sum_1^\infty p_n$ diverges $\left(\text{and hence } \prod_1^N (1-p_n) \to 0\right)$;

$$A_1 = \sum_1^\infty A_0 \frac{p_i}{1-p_i} = \sum_1^\infty A_0 u_i$$

$$u_i = \frac{p_i}{1-p_i}.$$

The series on the right is convergent if $\sum_1^\infty p_n$ converges, while it is to be interpreted as zero if $\sum_1^\infty p_n$ diverges.

More generally, for any finite k, BOREL asserted that

$$A_k = \sum A_0 u_{i_1} u_{i_2} \ldots u_{i_k}, \quad \text{the summation over } 1 < i_1 < i_2 < \cdots < i_k < \infty.$$

Here the right-hand side is to be interpreted as zero if $\sum_1^\infty p_n$ diverges, but BOREL asserted that in the convergent case

$$0 < A_k < 1, \quad k=0,1,2,3,\ldots.$$

Finally, letting

$$S = A_0 + A_1 + \cdots + A_k + \cdots,$$

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

BOREL showed that if $\sum_1^\infty p_n$ converges, then $S=1$, while if $\sum_1^\infty p_n$ diverges, $S=0$. Since $A_\infty = 1-S$, the remarkable result ("BOREL's Zero-One Law")

$$A_\infty = \begin{cases} 0 & \text{if } \sum_1^\infty p_n \text{ converges} \\ 1 & \text{if } \sum_1^\infty p_n \text{ diverges.}^1 \end{cases}$$

is obtained. The rest of the paper contains, as its important results, two applications of this curious behavior of A_∞.

The first application concerns the dyadic expansion of a real number x chosen "at random" in [0, 1]:

$$x = \cdot b_1 b_2 \ldots b_n \ldots = \sum_1^\infty \frac{b_n}{2^n}$$

where $b_n = b_n(x)$ is either 0 or 1. It is assumed that the sequence $\{b_n\}$ is generated, or the number x is "chosen", so that each binary digit $b_n(x)$ has probability $\frac{1}{2}$ of being 0 or 1, and also so that the various digits $n=1, 2, 3, \ldots$ are examples of independent trials. BOREL chooses a fixed sequence λ_n going to infinity with n but that $\lim_{n\to\infty} \frac{\lambda_n}{\sqrt{n}} = 0$.

Let $\nu_n(x)$ denote the number of ones among the first n binary digits $b_1(x)$, $b_2(x), \ldots, b_n(x)$. For each dyadic expansion BOREL considered the associated sequence of "trials" the n^{th} trial of which is defined as success if and only if

$$|\nu_{2n}(x) - n| \geqq \lambda_n \sqrt{n}, \qquad n = 1, 2, 3, \ldots.$$

Let p_n be the probability of this occurrence and $q_n = 1 - p_n$ the complementary probability. (We have reversed BOREL's notation for p_n and q_n to conform with our notation for E_∞.) By an application of the Central Limit Theorem, BOREL attempts to establish that $\sum_1^\infty p_n$ converges, and hence by an application of his main earlier result, that $A_\infty = P(E_\infty) = 0$. He concluded then that the event,

$$\lim_{n\to\infty} \frac{\nu_{2n}(x)}{2n} = \frac{1}{2},$$

has probability 1, *i.e.*,

$$P\left(\lim_{n\to\infty} \frac{\nu_n(x)}{n} = \frac{1}{2}\right) = 1.$$

[1] The formulas for $A_0, A_1, \ldots, A_k, \ldots$ are extensions of formulas for finite numbers of trials to the case of denumerably many. The case of A_∞ is utterly new, since E_∞ is empty for a finite number of trials and in that case its probability is of no interest.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

This result is called variously BOREL's Law of Large Numbers, the Strong Law of Large Numbers, or BOREL's Law of Normal Numbers. BOREL's argument will be analyzed in § 5.

The second application considers the expansion in continued fractions of a "randomly" chosen irrational number chosen from [0, 1]:

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots + \cfrac{1}{a_n + \cfrac{1}{\ddots}}}}}}$$

Here each $a_n = a_n(x)$ is a positive integer, $n = 1, 2, 3, \ldots$.

BOREL constructed a sequence of trials, and associated "success" and "failure", by choosing a fixed sequence $\phi(n)$ and defining for each irrational x the n^{th} "trial" as being a "success" or "failure" (with corresponding probabilities p_n, q_n) according as $a_n > \phi(n)$ or $a_n \leqq \phi(n)$. Thus BOREL has associated with each infinite continued fraction an instance of an infinite sequence of trials, each trial having only two outcomes; he then seeks to apply his Zero-One Law mentioned above to this collection of sequences of trials.

By adroit manipulations of continued fractions (see below), BOREL established the inequalities

$$\frac{2}{3} \frac{1}{(k+1)} < P(a_n(x) > k) < \frac{3}{k+2}.$$

Replacing k by $\phi(n)$ in this inequality shows that $\sum_1^\infty p_n$ converges if and only if $\sum_1^\infty \frac{1}{\phi(n)}$ does. Again appealing to his earlier result concerning A_∞, BOREL asserted that

if $\sum_1^\infty \frac{1}{\phi(n)}$ converges, then with probability 1, a_n will ultimately satisfy $a_n \leqq \phi(n)$

while

if $\sum_1^\infty \frac{1}{\phi(n)}$ diverges, then with probability 1, a_n will infinitely often violate that inequality.

This result will be called henceforth BOREL's *Continued Fraction Theorem.*

2. "Countable Independence" as a Key Principle

As is evident from this brief sketch, the property of *independence* of trials underlies the formulas for $A_0, A_1, \ldots, A_k, \ldots$ and A_∞, and hence also the two

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

applications. We propose to use the phrase "*countable independence*" for the principle that BOREL explicitly introduced and on which all of his results are based. This is the assertion, usually taken as a hypothesis, that a given collection of events, $B_1, B_2, \ldots, B_n, \ldots$ satisfy

$$P\left(\bigcap_1^\infty B_i\right) = \prod_1^\infty P(B_i). \tag{2.1}$$

When the collection of events is finite, the corresponding principle was known as the "loi des probabilités composées", although the notation for set intersection was not generally employed. BOREL assumed the principle if the events B_i referred to different trials for different i and, most important, assumed it to hold even if the range of the index were infinite. A particular instance of special importance, which we might call the *limited* principle of countable independence, is that (2.1) holds if each $P(B_i) = 1$.

The reader can see an analogy (which we believe must have acted powerfully on BOREL) with the behavior of length applied to disjoint intervals, B_i, on the real line [1]

$$l\left(\bigoplus_1^\infty B_i\right) = \sum_1^\infty l(B_i).$$

This extension to the infinite range and especially the interpretation of the left-hand side as a sort of generalized length if $\bigoplus_1^\infty B_i$ was not itself an interval (or even expressible as a finite union of intervals) lies at the heart of BOREL's earlier, profoundly important discovery of the theory of measure. In particular, the theory of measure, by focussing on this principle ("countable additivity") is led to extend the class $\mathscr{I}$ of intervals to a much wider and more significant class $\mathscr{B}$, with the key property that if each of the B_i is a set of this wider class $\mathscr{B}$, and the B_i are disjoint, $\bigoplus_1^\infty B_i$ is necessarily within the class $\mathscr{B}$.

BOREL in 1909 may have felt himself embarking on a similar exploration, using independence of trials as the counterpart to disjointness of intervals and with numerical products replacing numerical sums. Indeed, the very name BOREL gave his theory, *denumerable* probability, refers precisely to the range of the index i in the "loi des probabilités composées". It most certainly does *not* refer to the number of possible distinguishable outcomes (*i.e.* sequences of trials) as BOREL himself well knew. (For example, BOREL exploits the dyadic expansion of numbers in the unit interval as an example of his theory. The collection of such expansions has the cardinal number of the continuum.) Somehow BOREL felt that "denumerable probability", his new theory, was poised between classical "finite probability" and "geometric probability". Geometric probability, he well knew (*cf.* BOREL (1905)), may be generalized to make use of his own theory of measure and even to make use of LEBESGUE's new integration theory (namely for the computation of mean values). In addition to a "loi des probabilités composées", classical probability also

[1] When a union of *disjoint* sets is taken, the symbol $\bigoplus A_k$ will be used instead of $\bigcup A_k$ to emphasize that the A_k are (pair-wise) disjoint.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

has a "loi des probabilités totales":

$$P\left(\bigoplus_{1}^{N} B_i\right)=\sum_{1}^{N} P(B_i)$$

the events B_i being assumed disjoint (*i.e.* mutually incompatible). In some key instances BOREL extended this also to the infinite range,

$$P\left(\bigoplus_{1}^{\infty} B_i\right)=\sum_{1}^{\infty} P(B_i) \qquad \text{(countable additivity)},$$

but, as we shall argue, all the internal evidence is that BOREL regarded *countable independence* as the essential, new (and probabilistic) ingredient of his new theory. By contrast he used *countable additivity* seldom (often surreptitiously), and he never explored its implications; about it he had reservations so deep that he frequently offered "alternative" proofs to evade reliance on it.

This assessment, which we shall defend by suitable analysis, explains at least in part why BOREL failed to draw the conclusion, attributed to CANTELLI (1917a, 1917b), that

$$A_\infty=0 \quad \text{if } \sum_{1}^{\infty} p_n \quad \text{converges}$$

even if the trials are *not* assumed independent.

BOREL's fascination with the principle of countable independence similarly may explain his failure to use *anywhere* in BOREL (1909) an argument that

$$P\left(\bigcup_{1}^{\infty} B_i\right)\leqq\sum_{1}^{\infty} P(B_i) \qquad \text{(countable sub-additivity)}$$

which follows from countable additivity even if the B_i are *not* mutually incompatible. Whenever an event is shown to have probability 0, it is not by proof that it has "small covers", but rather by proof that its complement has probability 1. This latter reformulation becomes an assertion about the probability of an intersection and rests on the preferred principle of countable independence, often in the "limited" form.

We shall show below (§ 7.2) that BOREL knew the CANTELLI part of the BOREL-CANTELLI lemmas as early as 1903, but in the context of a geometric, not an abstract, space. More exactly, BOREL associated this type of reasoning with the HEINE-BOREL Covering Theorem as a preliminary. This further supports our thesis that BOREL did not see the full analogy of probability with measure except when the problem permitted a *geometric interpretation.* In our opinion, the analogy was imperfectly seen even then (*cf.* the discussion of his exchange with F. BERNSTEIN, § 8).

3. Denumerable Probability versus Measure Theory

Before we examine BOREL's text in some detail, it is instructive to summarize a few more recently acquired insights. In contemporary terms, each "trial" consid-

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

ered by BOREL is a 2-point measure space. The infinite sequences of outcomes are just points of the (denumerable) CARTESIAN product of such 2-point spaces. To each sequence of outcomes can be associated a sequence of 0's and 1's, where failure corresponds to 0, say, and success to 1. This provides a mapping from $\overset{\infty}{\underset{i=1}{\times}} \{0, 1\}_i$, which is the CARTESIAN product of 2-point spaces, to the unit interval. This mapping is $1-1$ except that each dyadic rational number may have two distinct pre-images, one terminating in 0's, the other in 1's. Further, the assumed independence between digits uniquely determines the probability, or as we would say equivalently now, measure, of each dyadic interval $\left[\frac{p}{2^n}, \frac{p+1}{2^n}\right]$; this measure $\mathscr{P}$ can be shown to satisfy

$$m\left(\bigoplus_1^\infty B_i\right)=\sum_1^\infty m(B_i)$$

if each B_i is a dyadic interval, and $\bigoplus_1^\infty B_i$ is again a dyadic interval. Finally it follows from all this (by insights gained after 1909) that the sequence $\{p_n\}$ determines a σ-additive measure on the σ-algebra of "BOREL sets" $\mathscr{B}$ of $[0, 1]$. If $p_n=\frac{1}{2}$, $n=1, 2, 3, \ldots$, then the measure is the very one introduced earlier by BOREL himself. If the $\{p_n\}$ are not identically $\frac{1}{2}$, the corresponding measure is a variant of the above (of the type later envisioned by RADÓN (1913)), a sort of STIELTJES measure associated with a suitable distribution function $F(x)$ but still defined at least on the BOREL sets of $[0, 1]$. Thus BOREL's new theory of "denumerable probability" is, if each $p_n=\frac{1}{2}$, essentially a disguised version of his own earlier theory of measure in the unit interval, the probabilistic impact of which he realized at least partially when the problem was one of "geometric probability" (*cf.* § 7.3). There is no evidence in favor of the supposition that BOREL understood that the case of general $\{p_n\}$ also gave rise to a "generalized" BOREL measure (*i.e.*, a different countably additive measure on the same countably additive field of sets). There is evidence that in the case $p_n=\frac{1}{2}$, $n=1, 2, 3, \ldots$, he did sense the connection between "probabilités dénombrables" and his own theory of measure. In this case he speaks of the "point de vue géométrique", referring to the interval $[0, 1]$ with its associated BOREL measure, and the "point de vue logique", referring to the infinite sequences of trials, which we denote $\overset{\infty}{\underset{i=1}{\times}} \{0, 1\}_i$. BOREL asserts that one can employ either. (*Cf.* § 5.1 for the exact citation.) He makes this comment only in discussing the specific case $p_n=\frac{1}{2}$, $n=1, 2, 3, \ldots$. Other internal evidence to support all of the above assertions concerning what BOREL realized or realized imperfectly, will be discussed below.

The crucial evidence against asserting that BOREL realized that "denumerable probability" *was* measure theory in $[0, 1]$ is his repeated evasion of countable additivity and countable sub-additivity in almost all of his reasoning. This evidence seems to us decisive. A lesser evidence of the same sort is his seeing something new in the fact that sets of probability zero need not be empty. And of course there is his own assertion (in the face of his own contrary assertion for the case $p_n=\frac{1}{2}$, n

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

= 1, 2, 3, ...) that "denumerable probability" is in general a theory *intermediate* to finite or combinatoric probability, and geometric or continuous probability.[1]

We turn now to BOREL's Chapter I (pp. 247–257) to justify the above imputations.

4. The Evidence From Borel's Chapter I

4.1. Denumerable Probability Contrasted with Continuous Probability

"One generally distinguishes, in probability problems, two principal categories, according to whether the number of possible cases is finite or infinite: the first category constitutes what one calls *discontinuous probabilities*, or probabilities in a discontinuous domain, while the second category comprises *continuous probabilities* or *geometric probabilities*. Such a classification appears incomplete when one refers back to the results acquired in the theory of sets; between the cardinality of finite sets and the cardinality of the continuum stands the cardinality of denumerable sets; I propose to show briefly the interest which is attached to questions of probability in whose statement such sets intervene; I will call them, for short, *denumerable probabilities*.

Before defining more precisely denumerable probability, I wish to indicate in a few words the reasons against further failing to study it. Principal among these reasons is the importance of the notion of denumerable sets; this importance was not contested by any mathematician; but it seems to me to be greater still than one believes.

Many analysts, indeed, put in the first rank the idea of the continuum; it is this concept which intervenes more or less explicitly in their reasoning. I have indicated recently how this notion of the continuum, considered as having a cardinality greater than that of the denumerable, seems to me to be a purely negative notion. The cardinality of denumerable sets alone being what we may know in a positive manner, the latter alone intervenes *effectively* in our reasonings. It is clear, indeed, that the set of analytic elements that can be actually defined and considered can be only a denumerable set; I believe that this point of view will prevail more and more every day among mathematicians and that the continuum will prove to have been a transitory instrument, whose present-day utility is not negligible (we will supply examples at once), but it will come to be regarded only as a means of studying denumerable sets, which constitute the sole reality that we are capable of attaining." BOREL (1909: 147–248).

These opening words indicate that BOREL believes that the set of possible outcomes which he will discuss and which in modern terms is his sample space, is denumerable. Nothing could be more misleading: the sample spaces he discusses are always denumerably infinite products of finite, or at most denumerably infinite, factor spaces. Indeed, even the simplest of these, the denumerable CARTESIAN

[1] Indeed the class of denumerable sample spaces is intermediate between finite and non-denumerable ones. But the space $\bigtimes_{i=1}^{\infty} \{0, 1\}_i$ under consideration in 1909 is not among them, being rather a different way of viewing the unit interval [0, 1].

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

product of 2-point spaces, is non-denumerable, as had been shown earlier by G. CANTOR. The representation of the unit interval as a product of 2-point spaces is a brilliant idea, but it cannot alter its familiar, and to BOREL repugnant, cardinal number *c*. The evidence for the repugnance, and the misconception as to the cardinal number of the sample spaces considered re-echoes in the closing lines of the paper:

> "At such time as the theory of denumerable probabilities is developed in the manner just indicated, it will be interesting to compare the results so acquired with those obtained in the theory of continuous or geometric probability.
>
> In the geometric continuum there *exist* certainly (if it is not a misuse to employ the verb *to exist*) some elements which cannot be defined: such is the real sense of the important and celebrated proposition of Mr. Georg Cantor: the continuum is not denumerable. Should a day come when these *undefinable* elements could be put aside as no longer needed more or less implicitly, it would certainly bring great simplification in the methods of Analysis; I should be happy if the preceding pages could help arouse the interest which the study of such questions deserves." (Italics in the original.) BOREL (1909: 271).

One of the best ways to fail to see a relation (*e.g.* that $\bigtimes_{i=1}^{\infty} \{0, 1\}_i$ and $[0, 1]$ are equivalent as cardinal sets and even as measure spaces) is to yearn that no such relation hold. BOREL surely did not wish his own "denumerable probabilities" to join the continuum as a future fossil, a mere transitory device.

4.2. *The Calculation of* A_0

BOREL turned first to establishing the formula for A_0

$$A_0 = (1-p_1)(1-p_2)\dots(1-p_n)\dots. \tag{4.1}$$

BOREL excludes in advance any $p_n = 1$ so he can conclude, if $\sum_1^{\infty} p_n$ is convergent, that $0 < A_0 < 1$. "In the case of convergence, the extension of the principle of composite probabilities goes without saying, ..." (BOREL (1909: 249). Further since the limit of the partial products is positive, they approach their limit with small relative error as well as absolute error. In conclusion, "...; the passage to the limit that we have performed thus does not raise any difficulties and is entirely justified." BOREL (1909: 249).

In fact, what BOREL is skimming over is the limit relation

$$\lim_{n\to\infty} P\left(\bigcap_1^n B_i\right) = P\left(\bigcap_1^{\infty} B_i\right). \tag{4.2}$$

An assumed independence assures additionally that if each B_i has probability q_i, then

$$P\left(\bigcap_1^n B_i\right) = \prod_1^n P(B_i) = \prod_1^n (1-p_i).$$

This content downloaded from 128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

The limit relation (4.2), however, has nothing to do with independence; it is one of the many consequences, and even equivalent forms, of countable additivity. The limit relation (4.2) is, for BOREL, both too desirable to be false and too evident to require discussion or elaboration. Indeed, since he nowhere employs a notation for the algebra of sets or even for sets themselves or for set functions, it would not have been easy for him to state explicitly. Had he employed the symbolism $P\left(\bigcap_1^\infty B_i\right)$, perhaps he might have been driven to question the domain of the set function $P(\cdot)$ just as he had earlier questioned and extended the domain of "length" for point-sets in $[0, 1]$.

The divergent case here, and later, calls for special "precautions". For "probabilité discontinue" (*i.e.*, finite sample spaces) probability zero means impossibility. For "probabilité continue" this is notably false, and BOREL refers the reader to his own paper of 1905. In this paper (*cf.* § 7.3) he had *defined* the geometric probability of BOREL sets on $[0, 1]$ as being their measure. This extends its familiar definition. In this way he proves, for instance, that the probability for a number picked at random to be irrational is zero, even though this outcome is not impossible. He now says it is the same in "probabilité dénombrable": in the divergent case the formula (4.1) for A_0 gives the value zero, since the relation

$$\lim_{n\to\infty} \prod_1^n (1-p_i) = A_0$$

is evidently accepted, despite the fact that the sequence of unbroken successes happens to exist. Non-empty sets may have zero-"probabilité dénombrable" just as non-empty sets (for example, the rationals) may have measure zero; he does not claim that they *are* sets of measure zero since he introduces no measure.

4.3. The Calculation of A_1, A_2, and A_k

The derivation of the formula for A_1 involves a more overt use of countable additivity, but, like the hidden use in the limit relation (4.2), this application passes without comment. If the lone success in a sequence of trials is at the n^{th} trial, then the probability of this sequence is

$$\omega_n = \frac{p_n}{1-p_n} A_0 = p_n \prod_{\substack{i=1\\ i\neq n}}^\infty (1-p_i),$$

interpreted as zero in the divergent case. It follows, by tacit use of countable additivity, that A_1 is the sum of these ω_n. Introducing the notation

$$u_n = \frac{p_n}{1-p_n},$$

we have

$$A_1 = \sum_1^\infty (A_0 u_n) \tag{4.3}$$

which BOREL writes as

$$A_0 \sum_1^\infty u_n. \tag{4.4}$$

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

This raises the question: in the divergent case, is A_1 a sum of zeros as in (4.3), and hence zero, or is it a product of $(0)(\infty)$ as in (4.4), and hence indeterminate?[1] in fact BOREL does not find the former reasoning, a routine example of countable additivity, compelling:

> ... On the other hand, one can say that the desired probability A_1 is the sum of separate probabilities ω_n each of which is zero; however, since they are not finite in number, one cannot conclude without special care that the total probability is zero, in view of the fact that zero probability does not denote *impossibility.* BOREL (1909: 250).

This is an interesting lapse: there is *no* reason why a sum of zeros can fail to be zero, unless it be that

$$\sum_1^\infty \omega_n \neq A_1,$$

i.e. countable additivity is explicitly *false;* BOREL's subsequent argument is to counter just that possibility. This is evidence that countable additivity was *not* a presupposed trait of probability. The realization that a probability of zero need not mean impossibility should not lessen one's confidence in countable additivity so far as we can see.

BOREL, in any case, attempts an alternative argument to show that in the divergent case $A_1=0$. Let σ_n be the probability that there is exactly one success in the first n trials. Then BOREL asserts that

$$\sigma_n=(1-p_1)\dots(1-p_n)(u_1+\cdots+u_n)<e^{-(p_1+\cdots+p_n)}(u_1+\cdots+u_n)$$

and hence that

$$\lim_{n\to\infty}\sigma_n=0.$$

The conclusion, though true, does not follow from the given upper estimate.[2]

He concludes "*In the divergent case:*

[1] Since $\sum_1^\infty u_n=\infty$ if $\sum_1^\infty p_n=\infty$.

[2] To show that BOREL's original upper estimate is deficient in its intended purpose, consider, for example, the choice

$$p_n=\frac{e^{2n}}{1+e^{2n}},\quad q_n=\frac{1}{1+e^{2n}},\quad u_n=\frac{p_n}{q_n}=e^{2n}.$$

Then $\sum_1^\infty p_n$ is divergent and

$$\sum_1^n u_k>u_n=e^{2n},$$

while

$$\sum_1^n p_k<n,$$

so that

$$e^{-(p_1+\cdots+p_n)}(u_1+u_2+\cdots+u_n)>e^{-n}e^{2n}=e^n.$$

This shows that for at least some choice of p_n such that $\sum_1^\infty p_n$ is divergent $e^{-\Sigma p_i}\sum_1^\infty u_i$ need not tend to zero. $\Big($In fact this happens whenever the p_n converge sufficiently rapidly to 1, which is surely consistent with the demand for divergence of $\sum_1^\infty p_n.\Big)$ It is true, however, that $\lim_{n\to\infty}\sigma_n=0$, as is established in BARONE (1974: 69–70), where the defective upper estimate for σ_n is replaced by a valid one.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

$A_1=0$." (Italics in the original.) BOREL (1909: 250).

Why is $A_1=0$ a consequence of $\lim_{n\to\infty}\sigma_n=0$? Presumably because

$$A_1=\lim_{n\to\infty}\sigma_n.$$

Let us introduce some notation lacking in BOREL's paper: we denote by E_1 the set of sequences having exactly one success, and by E_1^n the set of sequences with exactly one success among the first n outcomes. The relation between the sets E_1 and the sets E_1^n is that, as sets,

$$E_1=\lim_{n\to\infty}E_1^n.$$

It follows, *by countable additivity*, that

$$A_1=P(E_1)=\lim_{n\to\infty}P(E_1^n)=\lim_{n\to\infty}\sigma_n.$$[1]

The fact that BOREL could imagine he had found an *alternative* argument to establish $A_1=\sum_1^{\infty}\omega_n$ in this fashion only reinforces the impression that he neither fully recorgnized when he was employing countable additivity nor appreciated its primacy in limit relations.

The cases A_2 and then A_k are treated similarly. BOREL writes

$$A_k=A_0\sum u_{i_1}\dots u_{i_k}$$

and not

$$A_k=\sum(A_0 u_{i_1}\dots u_{i_k}).$$

This gave rise to the same misgivings about the divergent case quoted above in connection with (4.3) and (4.4).

4.4. The Calculation of A_∞

The derivation of A_∞ furnishes still more evidence of BOREL's reluctance to rely fully on countable additivity. First, in the convergent case, each A_k is positive, $k=1,2,3,\dots$, and if we define S by

$$S=A_0+A_1+A_2+\dots+A_k+\cdots,$$

then

$$\begin{aligned}S&=A_0(1+\sum u_{i_1}+\sum u_{i_1}u_{i_2}+\dots+\sum u_{i_1}u_{i_2}\dots u_{i_n}+\cdots)\\&=A_0(1+u_1)(1+u_2)\dots(1+u_n)\dots.\end{aligned}$$

[1] Theorems evaluating $\lim_{n\to\infty}P(E_n)$ under various circumstances were the main results in the period before BOREL (1909). BERNOULLI's ("Weak") Law of Large Numbers, and the Central Limit Theorem are theorems of this type. They did not assert that this limit was itself the probability of any single event, however, whereas BOREL's theorem *did*. On this important point, see the discussion in § 5.3 of BOREL's ("Strong") Law of Large Numbers.

This content downloaded from 128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

It is this product for S (rather than the defining sum) which is then employed. Since

$$u_n = \frac{p_n}{1-p_n},$$

it follows that

$$1+u_n = \frac{1}{1-p_n}.$$

Thus the infinite product in the expression for S can be evaluated by observing

$$\prod_1^\infty (1+u_n) = \prod_1^\infty \left(\frac{1}{1-p_n}\right) = \frac{1}{A_0},$$

so that

$$S = 1.$$

This is a perfectly rigorous manipulation of convergent infinite products since $\sum_1^\infty u_n$ converges. The appeal to countable independence occurs in the expression for A_0. The assertion that $A_\infty = 1 - S$ "évidement" is a case of disguised countable additivity, passed over without comment.

In the divergent case "on peut induire" that each A_k is zero; hence also their sum S and hence $A_\infty = 1$. The result is "exacte", but "le raisonnement précédent manque de rigueur, pour des raisons déjà indiquées". BOREL (1909: 251).

BOREL therefore considers an alternative approach, presumably more rigorous, introducing the set of sequences with more than m successes among the first n trials. Let us call this set F_m^n, since BOREL's paper lacks a notation for it. BOREL asserts that it is easy to calculate $P(F_m^n)$ and easy to show that

$$\lim_{n\to\infty} P(F_m^n) = 1, \qquad \text{for every } m = 1, 2, 3, \ldots.$$

He leaves the argument to the reader. The following is the "obvious" proof that presents itself.

If E_m^n = the set of sequences with *exactly* m successes among the first n outcomes, then by (finite) additivity

$$P(F_m^n) + P(E_0^n) + P(E_1^n) + \cdots + P(E_m^n) = 1.$$

Since

$$\lim_{n\to\infty} E_k^n = E_k, \qquad k = 1, 2, 3, \ldots,$$

as sets, therefore

$$\lim_{n\to\infty} P(E_k^n) = P(E_k) = A_k = 0, \qquad k = 1, 2, 3, \ldots$$

(by countable additivity) and so

$$\lim_{n\to\infty} P(F_k^n) = 1, \qquad k = 1, 2, 3, \ldots.$$

If this is the argument BOREL intended to supply, it is indeed an easy one, but no improvement with respect to "lack of rigor" inasmuch as the underlying limit

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

relations are not established. Perhaps BOREL introduced the sets which we have called F_m^n because they are defined by explicit conditions on the first n outcomes (*i.e.*, they are "cylinder sets" in the product space). If so, this increase in explicitness is not accompanied by any increase in rigor: at some point a limit of probabilities, such as $\lim_{n\to\infty} P(F_k^n)$, is computed. Such a limit can be identified as being itself a probability only by employing countable additivity. That is,

$$\lim_{n\to\infty} F_m^n = F_m$$

(as sets) and

$$\lim_{n\to\infty} P(F_m^n) = P\left(\lim_{n\to\infty} \mathrm{F}_m^n\right)$$

(by countable additivity) together yield

$$P(F_m) = \lim_{n\to\infty} P(F_m^n).$$

The total absence of set notation and set algebra can only have made it more difficult for BOREL to formulate, in a clear way, the critical relation

$$\lim_{n\to\infty} F_m^n = F_m.$$

BOREL emphasized the significance of the conclusion, which in our notation is,

$$\lim_{n\to\infty} P(F_m^n) = 1:$$

> ...; this means that one can profitably bet one franc against arbitrarily large odds that the number of successes will exceed any fixed number m; this is the precise meaning of the assertion *the probability A_∞ is one.* BOREL (1909: 251).

4.5. Additional Evidence on the Primacy of Countable Independence

Before concluding his Chapter I, BOREL considers various generalizations and modifications of the above main theorem concerning $A_0, A_1, A_2, \ldots$, and A_∞ in the convergent case. As a preliminary he considers the case of a single trial in which there is a denumerable infinity of outcomes. He also considers an infinite sequence of such trials, assumed independent, the probability of the n^{th} outcome on the s^{th} trial denoted $p_{n,s}$. BOREL then assumes

$$\sum_{n=1}^{\infty} p_{n,s} = 1 \qquad \text{for } s = 1, 2, 3, \ldots$$

which means the outcomes are exhaustive in each trial. He further restricts consideration to the "fully convergent" case, meaning that $\sum_{s=1}^{\infty} p_{n,s}$ converges for $n = 1, 2, 3, \ldots$.

Though the trials are assumed independent, two events determined by the number of occurrences of two different outcomes need not be independent.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

Let us introduce the following notation:

$E_{n,k}$ = set of trial sequences for which the n^{th} outcome occurs exactly k times,

$E_{n,\infty}$ = set of trial sequences for which the n^{th} outcome occurs infinitely often,

$\hat{E}_{n,k}$ = set of trial sequences for which the n^{th} outcome occurs at most k times.

If we consider any one individual outcome, say the n^{th}, it follows from the "convergence" hypothesis that

$$P(E_{n,\infty})=0$$

or, equivalently,

$$P(E^c_{n,\infty})=1.$$

Now we encounter an example of what we have called "limited" countable independence; although the different outcomes need not in general be independent, the events $E^c_{n,\infty}$, $n=1,2,\ldots$, having probability 1, are of necessity mutually independent. In words, given the "fully convergent" hypothesis, there is probability 1 that all outcomes occur only finitely often.

In our notation, $P(E^c_{n,\infty})=1$ for $n=1,2,3,\ldots$ implies

$$P\left(\bigcap_1^\infty E^c_{n,\infty}\right)=\prod_1^\infty P(E^c_{n,\infty})=1. \tag{4.5}$$

This line of reasoning provokes BOREL to one of the rare instances in which countable independence (even in the "limited" form) gives rise to explicit doubts:

> It may help to examine the question more closely, to be sure that our reasoning is strict: the fact that there is a denumerable infinity of factors equal to 1 could indeed leave some doubt as to the value of their product. BOREL (1909: 255).

BOREL then reassures himself by providing an elaborate proof of an assertion, in two parts, which refines as well as "re-establishes" the result in question.

First, he provides an argument that for every positive integer k,

$$P\left(\bigcap_{n=1}^\infty \hat{E}_{n,k}\right)=0. \tag{4.6}$$

(In words, the probability that all outcomes occur at most k times for any fixed finite integer k is zero.) This is established by an ingenious calculation, the details of which are left to the reader (*cf.* BARONE (1974: 101–113) for a detailed discussion), together with the tacitly assumed relation

$$P\left(\bigcap_{n=1}^\infty \hat{E}_{n,k}\right)=\prod_{n=1}^\infty P(\hat{E}_{n,k}).$$

This relation is tantamount to assuming independence of the events $\hat{E}_{n,k}$, *i.e.*, of outcomes, not just of trials.

Second, given any $\varepsilon>0$, one can choose the sequence k_n so that

$$\sum_{k=1}^\infty P(\hat{E}^c_{n,k_n})<\varepsilon$$

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

since $P(E_{n,\infty})=0$ for each n. Then it follows readily, if $\varepsilon<\frac{1}{2}$, that

$$\prod_{n=1}^{\infty}(1-P(\hat{E}^c_{n,k_n}))=\prod_{n=1}^{\infty}P(\hat{E}_{n,k_n})>1-2\varepsilon. \tag{4.7}$$

BOREL concludes:

> *One can thus, for a given arbitrary ε, choose the numbers k_n in such a way that the probability that the n^{th} outcome occurs no more than k_n times differs from unity by less than* 2ε. (Italics in the original.) BOREL (1909: 257).

In other words, BOREL is interpreting the above inequality by again using independence between the events $\hat{E}_{n,k}$, this time in the form:

$$\prod_{n=1}^{\infty}P(\hat{E}_{n,k_n})=P\left(\bigcap_{n=1}^{\infty}\hat{E}_{n,k_n}\right).$$

The validity of this requires that not only events defined by different trials but also those defined by different outcomes are assumed independent. Thus BOREL has replaced reasoning based on the (faultless) principle of "limited" countable independence by reasoning which employs a tacit use of countable independence. This second use suffers from the defect that its application is valid only subject to an additional hypothesis. (Once again, the fact that verbal conclusions about events are drawn from symbolized relations between numbers, with no intervening notation for sets or events, helps disguise the transgression from a sufficiently casual reader, perhaps from BOREL himself.)

Thus BOREL employs (4.6) and (4.7) to overcome doubts about the reasoning of (4.5). Of necessity these two new assertions themselves require countable independence and hold only subject to an additional, far-reaching (and tacit) assumption, whereas (4.5) itself can be established with no additional assumptions beyond the initially stated ones.

The "obvious" proof of (4.5), making no appeal whatever to independence of outcomes, namely the complemented assertion

$$P\left(\bigcup_{1}^{\infty}E_{n,\infty}\right)\leqq\sum_{1}^{\infty}P(E_{n,\infty})=0,$$

is conspicuously absent, since it hinges on countable sub-additivity, never used by BOREL in the paper of 1909.

This concluding section of BOREL's Chapter I furnishes another instance of his predilection for avoiding arguments based on sub-additivity and countable additivity, and for offering instead arguments based on countable independence. Indeed, this insight into BOREL is the chief interest attached to this section; its results are nowhere appealed to later.

4.6. *Summary of Borel's Conceptual Shortcomings in Chapter I*

In summary, BOREL fails to employ sub-additivity and seems to doubt routine arguments involving countable additivity, offering alternatives which employ

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

sophisticated reformulations of countable additivity such as

$$P(\lim E_n) = \lim P(E_n);$$

he bases his results, sometimes needlessly, on countable independence and assumes them valid in circumstances where no independence has been assumed or established. Countable independence itself is taken as being evident by analogy with the finite case despite occasional reservations. No notation is introduced for sets, and hence none for set functions. The adjective "denumerable", appropriate for the number of trials, is occasionally and erroneously construed to refer to the size of the sample space, and it is asserted that the theory of "denumerable probabilities" is a more "effective" theory than that of the continuum, since the latter claims to treat of more than denumerably many elements as being one collection. (*Cf.* the opening remark and concluding lines of BOREL's paper, both given in § 4.1.) This last distinction is of course illusory.

Had the paper contained no more, it would have furnished no evidence whatever to *favor* the hypothesis that BOREL understood that his "probabilités dénombrables" was a form of measure theory (in particular that countable additivity was an intrinsic part of its apparatus), and all the aforementioned to dispute it. However, the two remaining sections of the paper somewhat counterbalance this over-simple interpretation.

5. Borel's Chapter II: The Strong Law of Large Numbers

5.1. The Setting of the Problem

Consider the decimal expansion $\sum_1^\infty (b_n/10^n)$, each b_n being one of the digits $0, 1, \ldots, 9$. $\left(\text{The more general problem of "}q\text{-ary" expansions, } \sum_{n=1}^\infty (b_n/q^n)\text{, where each } b_n \text{ is among the integers } 0, 1, \ldots, q-1\text{, can of course be treated in like manner.}\right)$ BOREL proposed to study the probability that such b_n "belong to a given set" assuming (1) the digits are independent and, (2) each digit has equal probability (namely 1/10) of achieving the values $0, 1, \ldots, 9$.

> It is not necessary to emphasize the partly arbitrary nature of these two hypotheses: the first, in particular, is necessarily inaccurate when one considers, *as one if forced to in practice*, that a decimal expansion is defined by *a law*, whatever might be the nature of that law. It may nonetheless be interesting to study the consequences of this hypothesis, precisely in order to realize the extent to which things occur *as if* this hypothesis were verified. The second hypothesis, that is the equality of probabilities for the various possible values of each decimal digit, seems rather natural, granting the first.
>
> These two hypotheses are easily justified additionally by taking not the logical, but the geometric point of view: they are, indeed, equivalent to the following: *the decimal number being represented by a point of the interval* [0, 1], *the probability that it is located in a subinterval is equal to the length of that*

This content downloaded from 128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

subinterval. One could interpret and verify the results we are going to obtain from this geometric point of view; I will not do so, preferring to leave aside for the present the theory of continuous probability, which is connected, as I have shown elsewhere, to the theory of measure of sets.[1] BOREL (1909: 258).

Thus BOREL's "constructivism" leads him to assert that the first hypothesis is "nécessairement inexacte". One is led to speculate that BOREL knew where he wanted to go and arranged to get there, at the expense of his scruples if necessary.

The remark about the second hypothesis is, perhaps, even more interesting: nothing in the first chapter ("point de vue logique") demands equal probabilities $1/q$ for each b_n. The general theory of trials, each with q outcomes, accepts any choices $p_{n,s}$, $n=0, 1, \ldots, q-1$, such that

$$p_{0,s}+p_{1,s}+\cdots+p_{q-1,s}=1, \qquad s=1, 2, 3, \ldots .$$

5.2. *The Special Case* $p_n=\frac{1}{2}$ *for Dyadic Expansions*

BOREL realizes explicitly that the second hypothesis of equally likely b_n permits probability to be interpreted two equivalent ways, one of which is familiar. In the case $q=2$, this means the probability of the digit 0 and of the digit 1 in the n^{th} place are not only independent of the choice of digits in the other places but are both $\frac{1}{2}$. (It is an interesting comment on the character of mathematical evolution that before STEINHAUS (1923) no one had the temerity to consider whether any other choice of pre-assigned probabilities $\{p_n\}$ depending in general on n could also induce a measure on $[0, 1]$.[2]) BOREL now studied this equiprobable case by his new probabilistic methods in preference to measure-theoretic ones (also his own invention).

Proceeding to the case of only two outcomes for each trial, the case $q=2$, with the convention that the digit 0 is a success, or favorable case, BOREL states

One knows that, if one considers $2n$ trials, the probability that the number of favorable cases will lie between

$$n-\lambda\sqrt{n} \quad \text{and} \quad n+\lambda\sqrt{n}$$

is equal to $\theta(\lambda)$, letting

$$\theta(\lambda)=\frac{2}{\sqrt{\pi}}\int_0^{\lambda} e^{-\lambda^2}\,d\lambda.$$

BOREL (1909: 259).

---

[1] This is the sole reference to the theory of measure in the entire paper of 1909. Its role is to notify the reader that the theory of measure, an alternative approach, is *not* being employed. BOREL clearly is referring to his paper of 1905 the contents of which will be discussed in § 7.3.

[2] The study of these measures, implicitly introduced by BOREL, was first explored in any detail as recently as 1947 (WINTNER) and 1948 (HARTMAN). The notion of generalizing BOREL (or LEBESGUE) measure to other countably additive set functions had occurred within integration theory in 1913 (RADÓN) and within measure theory in 1914 (CARÁTHÉODORY).

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

(The statement is of course untrue, but the *limit* of this probability, as n tends to infinity, is given by $\theta(\lambda)$.) Assuming

$$\lim_{n \to \infty} \frac{\lambda_n}{\sqrt{n}} = 0$$

but λ_n itself grows unboundedly, BOREL considers each set of $2n$ initial trials as determining an event as follows. Let $v_n(x)$ be the number of zeros in the first n digits of the binary expansion of x. For each n, the "trial" examines $v_{2n}(x)$, and the result is a "success" if

$$|v_{2n}(x) - n| \geqq \lambda_n \sqrt{n}$$

and a "failure" if

$$|v_{2n}(x) - n| < \lambda_n \sqrt{n}.$$

The probability p_n of a favorable case is asserted to be

$$\frac{2}{\sqrt{\pi}} \int_{\lambda_n}^{\infty} e^{-\lambda^2} d\lambda$$

and the probability q_n of an unfavorable case is given by $q_n = 1 - p_n$.[1]

BOREL now focuses his attention on the set E_∞ of those "dyadic expansions" for which infinitely many "trials" have "successes". Always fascinated by sets of real numbers characterized by a denumerable set of conditions, BOREL had just found a new set, with a new "denumerable" description. Furthermore, since $\sum_1^n p_n$ converges if λ_n grows sufficiently fast (*e.g.* $n^{\frac{1}{3}}$), his Zero-One Law applied to this case (assuming the validity of its application) asserts that $A_\infty = 0$ and so $P(E_\infty^c) = 1$.

---

[1] This calculation of p_n and q_n is seriously flawed. What is true is that p_n and

$$\frac{2}{\sqrt{\pi}} \int_{\lambda_n}^{\infty} e^{-\lambda^2} d\lambda$$

are of the same order, but since λ_n is not fixed, this is not a case of the classical Central Limit Theorem. The information needed, that

$$\sum_1^\infty \left| p_n - \frac{2}{\sqrt{\pi}} \int_{\lambda_n}^{\infty} e^{-\lambda^2} d\lambda \right| < \infty,$$

is a refined and relatively recent result. Without it the convergence of

$$\sum_1^\infty \int_{\lambda_n}^{\infty} e^{-\lambda^2} d\lambda$$

(which is true if λ_n grows rapidly enough, *e.g.*, $n^{\frac{1}{3}}$) need not imply the convergence of $\sum p_n$. However, even assuming $\sum p_n$ converges, one cannot conclude from BOREL's result for A_∞ that there is a zero probability of infinitely many unfavorable cases, since that result was established on the hypothesis that the separate trials be independent. The various "trials", with probability p_n and q_n of success and failure as defined above, are by no means independent.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

To examine the set E_∞^c in more detail, note that if the n^{th} trial is unfavorable the ratio of the number of 0's to the number of 1's among the first n digits lies between

$$\frac{n-\lambda_n\sqrt{n}}{n+\lambda_n\sqrt{n}} \quad \text{and} \quad \frac{n+\lambda_n\sqrt{n}}{n-\lambda_n\sqrt{n}}$$

or, equivalently, between

$$\frac{1-\lambda_n/\sqrt{n}}{1+\lambda_n/\sqrt{n}} \quad \text{and its reciprocal} \quad \frac{1+\lambda_n/\sqrt{n}}{1-\lambda_n/\sqrt{n}}.$$

If there can be only finitely many exceptions to this, *i.e.*, if there are only finitely many "successes", the ratio of 0's to 1's must tend to the limit 1 as n tends to infinity. Thus the set E_∞^c consists of those numbers x whose dyadic expansions have asymptotically the same number of zeros as ones. Since $P(E_\infty)=A_\infty=0$, these numbers constitute a set of probability 1. This is the BOREL Law of large Numbers.

5.3. *Normal Numbers*

The corresponding reasoning applies to those numbers within whose decimal expansions the digits 0, 1, 2, ..., 9 each have a limiting frequency of 1/10: they form a set of probability 1. These numbers BOREL called "*simply normal*" to the base 10. More generally, BOREL called a number simply normal to the base q if the digits in the q-ary expansion

$$x=\sum_1^\infty \frac{b_n}{q^n}, \qquad b_n=0, 1, \ldots, q-1,$$

were such that

$$\operatorname*{limit}_{n\to\infty} \frac{(\text{number of occurrences of } i \text{ among } b_1(x), \ldots b_n(x))}{n}=\frac{1}{q}$$

for $i=0, 1, \ldots, q-1$.

For each q the set of numbers simply normal to the base q is similarly seen to be of probability 1. A number is called *entirely* normal to the base q if it is simply normal to each of the bases $q, q^2, \ldots, q^k, \ldots$. A number is called *absolutely* normal if it is simply normal to the base q for every $q=2, 3, \ldots$. If we denote the set of numbers simply normal to the base q by N_q, then the set of absolutely normal numbers is

$$N=N_2\cap N_3\cap N_4\cap\cdots\cap N_q\cap\cdots=\bigcap_{j=2}^\infty N_j$$

and the set of numbers entirely normal to the base q is

$$\bigcap_{k=1}^\infty N_{q^k}.$$

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

BOREL asserts (without argument) that both types of numbers are of probability 1. What is being tacitly affirmed is, as usual, that

$$P\left(\bigcap_{k=1}^{\infty} N_{q^k}\right)=\prod_{k=1}^{\infty} P(N_{q^k})$$

and that

$$P\left(\bigcap_{q=2}^{\infty} N_q\right)=\prod_{q=2}^{\infty} P(N_q)$$

where each factor on the right-hand side is 1. These are cases of "limited" countable independence. A laconic footnote reads

> I do not think it serves any purpose to repeat the detailed proof of the fact that one has the right to apply the theorem of composite probabilities, despite the denumerable infinity of cases. BOREL (1909: 261: Footnote (5)).

Of course, no "detailed proof of the right to apply the theorem of compound probabilities" in the denumerably infinite case has been given; recall that it was first mentioned in connection with

$$A_0=(1-p_1)(1-p_2)\dots(1-p_n)\dots$$

"Dans le cas de la convergence, l'extension du principe des probabilités composées va de soi, ..." BOREL (1909: 249). Repeated use has by this point transformed the principle from a self-evident truth to one whose proof has been demonstrated all too often.[1]

BOREL's reasoning, up to this point, is chiefly flawed in the following two ways: he employs a result based on independence of events for dependent trials (this flaw would be remedied by a "CANTELLI" modification of his Zero-One Law, *cf.* § 6.4) and he uses a form of the Central Limit Theorem more precise than was then available. HAUSDORFF's and later proofs were to avoid any appeal to the Central Limit Theorem whatever.

An additional omission is noteworthy: despite a choice of notation identical to the one he had employed for discussing the (BERNOULLI) Law of Large Numbers in his text-book on Probability, written earlier in the same year (BOREL (1909a: 63–65)), he makes no attempt to compare the BERNOULLI Law and the new result. A great opportunity is lost thereby: treating both within the theory of measure, one would have been led to the comparison of convergence "in measure" and convergence "almost everywhere", anticipating the treatments of SLUTSKY (1925), FRÉCHET (1930) and the earlier such comparisons by CANTELLI (1917) and PÓLYA (1921). It is possible that the stronger character of his result by comparison with

[1] In discussing A_0, A_k, A_∞ independence of *trials* was assumed. The independence of the events N_q, $q=2, 3, \dots$, or of the events N_{q^k}, $k=1, 2, 3, \dots$, is slightly more sophisticated. It is not to be regarded as an additional ad hoc assumption. Rather one must establish the general fact that if $P(N)=1$, then $P(N\cap A)=P(N)\,P(A)$. *i.e.*, N is independent of *any* other event A. This BOREL failed to do.

The proof involves considering the complement of $N\cap A$ and then concluding that a set contained in a set of probability zero must itself have probability zero. The former consideration has the possible drawback, for BOREL, of shifting attention away from "independence". As to the latter, BOREL is known to have rejected the corresponding reasoning when employed in the context of measure theory.

This content downloaded from 128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

BERNOULLI's was sufficiently evident to BOREL that no comment was forthcoming. Both CANTELLI and PÓLYA were to be explicit about the distinction between the Strong Law and BERNOULLI's result.

Indeed, the earlier successors of BOREL may be classified roughly as being "Probability" oriented or "Measure Theory" oriented. The former may invariably be distinguished by their interpretation of BOREL's Strong Law as an astonishing refinement of BERNOULLI's original theorem, whereas the latter fail to associate the two, since BERNOULLI's theorem seemed not to belong to measure theory.

5.4. A Possible Clue to the Genesis of Borel's Strong Law

The result of BOREL's Chapter II, namely, BOREL's Strong Law of Large Numbers, has proved to be an exceedingly fruitful application of the $A_\infty = 0$ result of Chapter I, the "general" theory of denumerable probabilities. There remains the fascinating question, what drew BOREL's attention to the instances of his general theory furnished by digits occurring in the dyadic expansion of numbers in the unit interval? One possible answer is, of course, BOREL's profound intuition. It is of interest to note, however, that in 1908 the American, E. VAN VLECK, had been led to study the set of dyadic irrationals in the unit interval defined, in the notation introduced above, by the relation

$$\overline{\lim_{n\to\infty}} \frac{v_n(x)}{n - v_n(x)} = \overline{\lim_{n\to\infty}} \frac{n - v_n(x)}{v_n(x)}$$

or, equivalently,

$$\overline{\lim_{n\to\infty}} \frac{v_n(x)/n}{1 - v_n(x)/n} = \overline{\lim_{n\to\infty}} \frac{1 - v_n(x)/n}{v_n(x)/n}.$$

Let us denote this set by V_0. BOREL's result, if stated purely in terms of measure theory, is that the set B_0 defined by

$$\lim_{n\to\infty} \frac{v_n(x)}{n} = \frac{1}{2}$$

is of measure 1. It is immediate from their definitions that $B_0 \subset V_0$. Of greatest interest in this connection is VAN VLECK's explicit and pivotal conjecture that V_0 is *not* a set of measure 1; VAN VLECK was led to the conjecture by a general criterion for nonmeasurability of sets in the unit interval. If V_0 were not measurable, or if it were to have LEBESGUE inner measure less than one, then VAN VLECK would have constructed a nonmeasurable set in the unit interval without using the Axiom of Choice; the set constructed (just the set V_0 defined above) would have been the set of dyadic irrationals defined by

$$\overline{\lim_{n\to\infty}} \frac{v_n(x)/n}{1 - v_n(x)/n} = \overline{\lim_{n\to\infty}} \frac{1 - v_n(x)/n}{v_n(x)/n}.$$

VAN VLECK's paper contains much else of interest and is discussed in detail in NOVIKOFF & BARONE (1977). There it is shown that an elementary observation (which VAN VLECK unaccountably failed to make) shows at once that V_0 is measurable and indeed, from VAN VLECK's other results, *must have measure 1.* VAN

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

VLECK had made BOREL's acquaintance in France in November 1905–January 1906, while on sabbatical leave, and the two remained on friendly terms for years thereafter. Had he communicated his results (including his conjecture about the measure of V_0) to BOREL at once, it is possible that BOREL would have seen that the question of the *probability* (not "measure") of V_0, and indeed, the more delicate question of its subset B_0, was accessible to his new theory of "denumerable probability". When one considers that VAN VLECK's main goal was to construct a nonmeasurable set (the very existence of which was discomforting to BOREL), the opportunity offered BOREL to get a better result by his "more effective" theory might well have directed his attention along the trail leading to his formulation of the Strong Law.

6. Borel's Chapter III: Continued Fractions

6.1. The Setting of the Problem

In Chapter III, BOREL returns to continued fractions, the subject of his earlier 45-page paper of 1903, "Contribution à l'Analyse Arithmétique du Continu."

The continued fraction expansion of an irrational number x in $[0, 1]$,

$$x=\cfrac{1}{a_1+\cfrac{1}{a_2+\cfrac{1}{a_3+\cfrac{1}{\ddots}}}}$$

where each element $a_n=a_n(x)$ is a positive integer, is in some ways parallel to that of the decimal (or q-ary) expansion. BOREL considered it as an instance of a sequence of infinitely many trials (one for each integer a_n) each of which may have infinitely many outcomes, *e.g.* a_n may equal $1, 2, 3, \ldots, k, \ldots$.

The probabilities $p_{i,k}$, that $a_i=k$, satisfy

$$\sum_{k=1}^{\infty} p_{i,k}=1.$$

One could, *a priori*, make arbitrary hypotheses in addition, but BOREL asserted,

> ...; we are going to study the hypotheses to which one is led when taking the geometric point of view already indicated à propos the decimal numbers. BOREL (1909: 264).

This means, in effect, that the probability for x to lie in any sub-interval of $[0, 1]$ is the length of that sub-interval.[1] Once again the motivation for the example comes from problems of "geometric" or "continuous" probability, in the most naive sense.

[1] By extension, it also means that the probability that x lies in a BOREL set of $[0,1]$ is the measure of that set, but this all important and non-naive extension of geometric probability by means of measure theory is not employed or explicitly acknowledged in the paper of 1909. See FABER's remark (§ 9.2), made after reading the paper of 1909, which explicitly questions whether such an extension can be made of "denumerable probability".

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

It is readily shown that as a consequence of this assumption

$$\begin{aligned} p_{1,k} &= \text{probability}\,\{a_1(x)=k\} \\ &= \text{probability}\left\{\frac{1}{k+1}<x<\frac{1}{k}\right\}=\frac{1}{k(k+1)}. \end{aligned}$$

The set of all x such that $\{a_2(x)=k\}$ is a union of disjoint intervals, namely

$$\bigoplus_{m=1}^{\infty} \{a_1(x)=m,\, a_2(x)=k\}.$$

The corresponding length is a cumbersome infinite series which equals $p_{2,k}$. Similarly $p_{3,k}$ is given by a double series, corresponding to the sum of lengths of the disjoint union

$$\bigoplus_{1\leqq m_1, m_2<\infty} \{a_1(x)=m_1, a_2(x)=m_2, a_3(x)=k\}.$$

In summary, the probabilities $p_{n,k}$ become increasingly unmanageable with increasing n, and attention must be directed to the *multi-indexed* probability of the set of x satisfying

$$\{a_1(x)=m_1, a_2(x)=m_2, \ldots, a_{j-1}=m_{j-1}, a_j(x)=k\}.$$

The points x for which these specified values of $a_1(x), \ldots, a_j(x)$ are achieved constitute an interval, which we shall denote

$$[a_1=m_1, a_2=m_2, \ldots, a_{j-1}=m_{j-1}, a_j=k].$$

Distinct such intervals with the same value of j are disjoint, while $[a_1=m_1, \ldots, a_j=m_j]$ is a sub-interval of $[a_1=m_1, \ldots, a_{j-1}=m_{j-1}]$ and in fact

$$[a_1=m_1, \ldots, a_{j-1}=m_{j-1}] = \bigoplus_{m_j=1}^{\infty} [a_1=m_1, \ldots, a_{j-1}=m_{j-1}, a_j=m_j].$$

In this notation

$$p_{n,k} = \sum_{m_1, \ldots, m_{n-1}} l[a_1=m_1, \ldots, a_{n-1}=m_{n-1}, a_n=k]$$

the summation being over all positive integer values of $m_1, \ldots, m_{n-1}$ and $l(J)$ = length of the interval J.

6.2. *The Derivation of the Key Inequality*

While $p_{n,k}$ cannot be precisely calculated, it can be estimated by introducing the "approximants" of a continued fraction:

$$\frac{P_n}{Q_n}=\cfrac{1}{a_1+\cfrac{1}{a_2+\cfrac{1}{\ddots+\cfrac{1}{a_n}}}}.$$

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

Here $P_n = P_n(a_1, \ldots, a_n)$ and $Q_n = Q_n(a_1, \ldots, a_n)$ satisfy the classical recursion formulae

$$P_n = P_{n-2} + a_n P_{n-1}, \qquad P_1 = 1, \qquad P_0 = 0,$$
$$Q_n = Q_{n-2} + a_n Q_{n-1}, \quad Q_1 = a_1, \quad Q_0 = 1.$$

The interval $[a_1 = m_1, \ldots, a_{n-1} = m_{n-1}, a_n = k]$ has end-points

$$\frac{P_n(m_1, \ldots, m_{n-1}, k)}{Q_n(m_1, \ldots, m_{n-1}, k)} = \frac{P_{n-2} + kP_{n-1}}{Q_{n-2} + kQ_{n-1}}$$

and

$$\frac{P_n(m_1, \ldots, m_{n-1}, k+1)}{Q_n(m_1, \ldots, m_{n-1}, k+1)} = \frac{P_{n-2} + (k+1)P_{n-1}}{Q_{n-2} + (k+1)Q_{n-1}}.$$

In virtue of the classical identity

$$\begin{vmatrix} P_{n-2} & Q_{n-2} \\ P_{n-1} & Q_{n-1} \end{vmatrix} = (-1)^{n-1}$$

the length of this interval is given by

$$l[a_1 = m_1, \ldots, a_{n-1} = m_{n-1}, a_n = k] = \frac{1}{Q_{n-1}^2} \frac{1}{\left(k + \frac{Q_{n-2}}{Q_{n-1}}\right)\left(k + 1 + \frac{Q_{n-2}}{Q_{n-1}}\right)}.$$

Here $Q_{n-1} = Q_{n-1}(m_1, m_2, \ldots, m_{n-1})$, $Q_{n-2} = Q_{n-2}(m_1, m_2, \ldots, m_{n-2})$ are convenient but dangerous abbreviations.

It follows that the factor $1/Q_{n-1}^2$ is common to both $l[a_1 = m_1, \ldots, a_{n-1} = m_{n-1}, a_n = k]$ and $l[a_1 = m_1, \ldots, a_{n-1} = m_{n-1}, a_n = k+1]$ so that their ratio is

$$\frac{k + \dfrac{Q_{n-2}(m_1, \ldots, m_{n-2})}{Q_{n-1}(m_1, \ldots, m_{n-1})}}{k + 2 + \dfrac{Q_{n-2}(m_1, \ldots, m_{n-2})}{Q_{n-1}(m_1, \ldots, m_{n-1})}}.$$

Since

$$0 < \frac{Q_{n-2}}{Q_{n-1}} < 1$$

for all choices of $m_1, \ldots, m_{n-1}$, this ratio lies between

$$\frac{k}{k+2} \quad \text{and} \quad \frac{k+1}{k+3}$$

This content downloaded from 128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

for all choices of $m_1, \ldots, m_{n-1}$. It follows[1] that

$$\frac{k}{k+2} < \frac{p_{n,k+1}}{p_{n,k}} < \frac{k+1}{k+3} \tag{6.1}$$

from which inequality BOREL proceeds. The set of numbers x satisfying $a_n(x) = k$ is a union of intervals and not itself an interval; the basic intervals, whose lengths are calculable in terms of P_n's and Q_n's, are of the form $[a_1 = m_1, \ldots, a_{n-1} = m_{n-1}, a_n = k]$. Since each $a_n(x)$ is thought of as a "trial" determined by x, such a basic interval represents the set of "trial" sequences with prescribed outcomes on the first n trials. In the language of Cartesian products such sets have come to be called *cylinder sets*, with "base" in the product space of the first n factor spaces.

What BOREL has calculated is, first, inequalities for the ratios of probabilities of cylinder sets

$$\frac{k}{k+2} < \frac{l[a_1 = m_1, \ldots, a_{n-1} = m_{n-1}, a_n = k+1]}{l[a_1 = m_1, \ldots, a_{n-1} = m_{n-1}, a_n = k]} < \frac{k+1}{k+3}. \tag{6.2}$$

The comparable inequality for the ratio $\frac{p_{n,k+1}}{p_{n,k}}$ of *a priori* probabilities is

$$\frac{k}{k+2} < \frac{p_{n,k+1}}{p_{n,k}} < \frac{k+1}{k+3}.$$

This last inequality yields, by recurrence,

$$\frac{(k-1)(k-2)\ldots(2)(1)}{(k+1)(k)\ldots(4)(3)} < \frac{p_{n,k}}{p_{n,1}} < \frac{(k)(k-1)\ldots(3)(2)}{(k+2)(k+1)\ldots(5)(4)}$$

i.e.

$$\frac{2}{(k)(k+1)} < \frac{p_{n,k}}{p_{n,1}} < \frac{6}{(k+1)(k+2)}.$$

---

[1] To state this more explicitly: what is established in the text (in notation that regrettably suppresses the indices $m_1, m_2, \ldots, m_{n-1}$) is that

$$\frac{k}{k+2} < \frac{\Pr[a_1 = m_1, a_2 = m_2, \ldots, a_{n-1} = m_{n-1}, a_n = k+1]}{\Pr[a_1 = m_1, a_2 = m_2, \ldots, a_{n-1} = m_{n-1}, a_n = k]} < \frac{k+1}{k+3}.$$

Here the strength of the inequalities lies in the fact that the given bounds hold for all choices of $m_1, m_2, \ldots, m_{n-1}$. It follows that if one multiplies this inequality by $\Pr[a_1 = m_1, \ldots, a_{n-1} = m_{n-1}, a_n = k]$ and then sums over all possible values of $m_1, \ldots, m_{n-1}$, one obtains

$$\frac{k}{k+2}\, p_{n,k} < p_{n,k+1} < \frac{k+1}{k+3}\, p_{n,k}$$

which is the desired statement (6.1). This follows since

$$\sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} \cdots \sum_{m_{n-1}}^{\infty} \Pr[a_1 = m_1, \ldots, a_{n-1} = m_{n-1}, a_n = k] = \Pr[a_n = k].$$

(It should be emphasized again that the set of x for which $[a_n = k]$ is a sum of *disjoint* open intervals and not itself an interval.)

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

One can free this of $p_{n,1}$ by using $\sum_{k=1}^{\infty} p_{n,k}=1$. Summing the above inequalities

$$\sum_{1}^{\infty} \frac{2}{(k)(k+1)} < \frac{\sum_{k=1}^{\infty} p_{n,k}}{p_{n,1}} < \sum_{1}^{\infty} \frac{6}{(k+1)(k+2)},$$

we obtain

$$2 < \frac{1}{p_{n,1}} < 3,$$

i.e.

$$\tfrac{1}{3} < p_{n,1} < \tfrac{1}{2}.$$

Thus inequalities are obtained for $p_{n,k}$:

$$\frac{2/3}{(k)(k+1)} < p_{n,k} < \frac{3}{(k+1)(k+2)}.$$

Let $P_{n,k}=p_{n,1}+p_{n,2}+\cdots+p_{n,k}=P\{a_n(x)\leqq k\}$. Then

$$1-P_{n,k}=\text{Probability}\,\{a_n(x)>k\}=p_{n,k+1}+p_{n,k+2}+\cdots.$$

The inequalities governing $p_{n,k}$ imply the inequality

$$\frac{2}{3(k+1)}=\sum_{j=k+1}^{\infty} \frac{2}{3(j)(j+1)} < 1-P_{n,k} < \sum_{j=k+1}^{\infty} \frac{3}{(j+1)(j+2)}=\frac{3}{k+2}.$$

At this point BOREL identifies each continued fraction with an infinite sequence of dichotomous trials so that he can apply his A_∞ result. To do this he introduces an integer valued function $\phi(n)$, and considers for each fixed irrational x in the unit interval the sequence of events defined by $a_n(x)>\phi(n)$ ("success"), $n=1,2,3,\ldots$.

If $\phi(n)$ is chosen so that

$$\sum \frac{1}{\phi(n)}$$

converges, then

$$\sum_{n=1}^{\infty} (1-P_{n,\phi(n)})$$

converges. If $\sum \frac{1}{\phi(n)}$ diverges, then $\sum_{n=1}^{\infty} (1-P_{n,\phi(n)})$ diverges. From the result on A_∞ BOREL concludes

The Borel Continued Fraction Theorem. Prob$\{a_n(x)>\phi(n)$ *infinitely often*$\}$ *is* 0 *or* 1 *according as* $\sum \frac{1}{\phi(n)}$ *converges or diverges.*

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

BOREL next observes that this sharp alternative map be put into even more striking form. Indeed, if $\sum \frac{1}{\phi(n)}$ converges, then there exists a $\psi(n)$ such that $\sum \frac{1}{\psi(n)}$ converges while $\lim_{n\to\infty} \frac{\psi(n)}{\phi(n)} = 0$, and correspondingly, if $\sum \frac{1}{\phi(n)}$ diverges, there exists a $\psi(n)$ such that $\sum \frac{1}{\psi(n)}$ diverges while $\lim_{n\to\infty} \frac{\psi(n)}{\phi(n)} = \infty$. It follows from the original formulation applied to $\psi(n)$ that the theorem can be reformulated thus:

$\operatorname{Prob}\left\{\overline{\lim} \frac{a_n}{\phi(n)} = 0\right\} = 1$ *or* $\operatorname{Prob}\left\{\overline{\lim} \frac{a_n}{\phi(n)} = \infty\right\} = 1$ *accordingly as* $\sum \frac{1}{\phi(n)}$ *converges or diverges.*

BOREL wrote of this theorem: "In this form it appears to me to be the most interesting of those we have obtained in this memoir." BOREL (1909: 269). As a historical note BOREL adds "the notation $\overline{\lim}$, due to Pringsheim, denotes 'la plus grande limite' defined by Cauchy, and made precise by du Bois Reymond and Hadamard."

6.3. Defect in Reasoning

The defect in reasoning here is like that of the case of the decimal fraction, but more grave. In the decimal case only the zero part of the Zero-One Law was used, when the conclusion $A_\infty = 0$ was asserted for a sequence of trials with $\sum_1^\infty p_n < \infty$. Though the trials were not independent, the "CANTELLI" generalization of BOREL's proof that $A_\infty = 0$, whenever $\sum_1^\infty p_n$ is convergent, suffices to validate this aspect of the reasoning.

For the case of the continued fraction, where once again the "trials" under consideration are dependent, the conclusion that $A_\infty = 0$ may again be justified in this fashion, if $\sum_1^\infty p_n$ converges. However the companion result for continued fractions, that $A_\infty = 1$ if $\sum_1^\infty p_n$ diverges, requires a different generalization of BOREL's original discussion of the divergent case to cope with the dependence of the "trials" (that is, the digits $a_1, a_2, \ldots$). That the trials are dependent is clear both geometrically and algebraically. We shall return to this point, and give the required generalization of the $A_\infty = 1$ result, in § 8.2 below which deals with the work of F. BERNSTEIN.

6.4. A Possible Clue to the Genesis of the Borel Continued Fraction Theorem

Once again, as in the case of the BOREL Strong Law, one can ask what led BOREL to the remarkable (if flawed) discussion of continued fractions as an example of his general theory of denumerable probability. As before, one can simply appeal to the intuition of a profound and fertile intellect. In this case, however, an

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

alternative of considerable weight must be considered. The relevant fact is that the Swedish mathematician A. WIMAN had obtained an analogous result dating back to 1900–1901 which asserts zero probability for a set in the unit interval. WIMAN's set (a subset of the irrationals in the unit interval) was also defined by a sophisticated condition on the asymptotic behavior of the terms $a_n(x)$ of the expansion of its members in continued fractions. The calculations showing it to have zero "probability" overlapped several of BOREL's. WIMAN's proof involved an explicit use of BOREL's new theory of measure to extend the scope of classical geometric probability and relied on a scrupulous use of countable sub-additivity of such measure in the proof (WIMAN (1900; 1901)). WIMAN himself was repairing serious defects in the previous work of his colleague T. BRODÉN (1900), in a polemic that raged between the two. The problem they debated had arisen in the context of Celestial Mechanics where it had been raised by the Swedish mathematician-astronomer H. GYLDÉN; as a result their polemic was disputed in journals read by Nordic mathematical astromers but was hardly familiar to the general European mathematical community. In contrast to the corresponding case of VAN VLECK's work on dyadic digits, there is no doubt of BOREL's acquaintance with this prior work; in 1905 BOREL explicity gave WIMAN priority for the introduction of measure into geometric probability. This acknowledgement, accompanied by a bibliographic reference to WIMAN's papers, occurs in BOREL (1905), which is discussed below (*cf.* § 7.3). However, BOREL does not, in 1905, refer to the content of WIMAN's result, nor does he even mention the occurrence of continued fractions in the formulation of the problem solved by WIMAN. The much more influential paper of 1909 lacks any reference to WIMAN's work whatsoever, consistent with its omission of any interpretation in terms of measure of the result about continued fractions.

BOREL saw his Continued Fraction Theorem as an example of a general theory of probability on abstract sets (the presumed meaning of "pointe de vue logique"). On the evidence presented above, he was at best unclear that his general theory was, in this case, equivalent to measure theory. It is entirely possible he thought denumerable probability applied to continued fractions offered a distinct, more "effective" (*i.e.*, "constructivist") alternative. In any case BOREL unquestionably knew of at least one *probabilistic* result concerning continued fractions before 1909 and this could well have been the genesis of his own example of the general theory.

The distinction between WIMAN and VAN VLECK as precursors of BOREL is marked in two ways: first, while both employed measure theory, WIMAN, unlike VAN VLECK, explicitly considered his theorem as a solution to a problem of probability; second, while VAN VLECK's work *may have* been known to BOREL, WIMAN's *unquestionably* was known.

6.5. *The Cantelli Modification of the Borel Zero-One Law*

The BOREL Zero-One Law has undergone considerable generalization since its first formulation in 1909. In particular, the "zero" case ($A_\infty = 0$ when $\sum p_n$ converges) has subsequently been established with no appeal to the independence of the events whose probabilities are $p_1, p_2, \ldots$. These events are the "trials" of the original BOREL formulation. This removal of the hypothesis of independence in the

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

"zero" case is attributed to CANTELLI (1917a, 1917b), whose work we shall discuss in Part II. However we have occasion to refer to the "CANTELLI" result and "CANTELLI-like reasoning" in what follows. Therefore we interpolate at this point the standard modern formulation, to render the immediately ensuing discussion self-contained. This formulation differs substantially from CANTELLI's original one in that it depends on an underlying σ-additive measure.

Cantelli Lemma. *Let $(\Omega, \mathscr{B}, P)$ be a probability space, so that $\mathscr{B}$ is a σ-field of sets from Ω, and P is a countably additive non-negative normalized measure defined on the sets of $\mathscr{B}$. If $E_1, E_2, \ldots, E_n, \ldots$ are sets in $\mathscr{B}$ and $\sum_1^\infty P(E_n)$ converges, then $P(\limsup_{n\to\infty} E_n)$* $=0$. The proof is given after some preliminary remarks.

Remarks. 1) $\limsup_{n\to\infty} E_n$ is defined as $\bigcap_{N=1}^\infty \bigcup_{n=N}^\infty E_n$ and so is in $\mathscr{B}$. It is the set of all points which are in infinitely many of the E_n's.

2) If $E \subset \limsup_{n\to\infty} E_n$, and E is in $\mathscr{B}$, then $P(E) \leqq P(\limsup_{n\to\infty} E_n)$, so that under the hypothesis $\sum_{n=1}^\infty P(E_n)$ converges, one concludes $P(E)=0$.

3) If $\mathscr{B}$ is complete with respect to P, then $E \subset \limsup_{n\to\infty} E_n$ implies by remark 2) that $E \in \mathscr{B}$ and further that $P(E)=0$.

4) The notation E_n here is not to be confused with the notation introduced by us in our discussion of BOREL (1909), where E_n was the occurrence of precisely n successes and $A_n = P(E_n)$. In fact, we here use E_n to play the role of the event "success at the n^{th} trial" in BOREL's terminology. The probabilities $P(E_n)$ occurring here correspond to the probabilities p_n introduced by BOREL, and $P(\limsup E_n)$ corresponds to BOREL's A_∞. It is this notation which will be used henceforth.

Proof of the Cantelli Lemma. The sets $\bigcup_{n=N}^\infty E_n$ decrease with increasing N, so that $P\left(\bigcup_{n=N}^\infty E_n\right)$ is a non-increasing sequence and

$$P(\limsup_{n\to\infty} E_n) = \lim_{N\to\infty} P\left(\bigcup_{n=N}^\infty E_n\right),$$

by the countable additivity of P. Further

$$P\left(\bigcup_{n=N}^\infty E_n\right) \leqq \sum_N^\infty P(E_n)$$

by the sub-additivity of P, a consequence of countable additivity. Since $\sum_1^\infty P(E_n)$ converges by hypothesis, $\sum_N^\infty P(E_n)$ tends to 0 with increasing N. Thus

$$P(\limsup_{n\to\infty} E_n) = \lim_{N\to\infty} P\left(\bigcup_{n=N}^\infty E_n\right) \leqq \lim_{N\to\infty} \sum_N^\infty P(E_n) = 0$$

as desired.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

If we re-introduce the notation $P(E_n)=p_n$ and $A_\infty = P(\limsup E_n)$, then we obtain the CANTELLI modification of BOREL's result: $\sum_1^\infty p_n$ converges implies $A_\infty = 0$, without any assumption as to independence.

In § 10.2 below, as well as in § 7 immediately following we give instances of this reasoning by BOREL, FRÉCHET, and HAUSDORFF, all of whom preceded CANTELLI.

7. Borel's Earlier Works

7.1. Introduction

This section deals with two earlier works of BOREL: "Contribution à l'analyse arithmétique du continu" (BOREL (1903)) and "Remarques sur certaines questions de probabilité" (BOREL (1905)).

Each of these works will be related to the considerations of BOREL's "Les probabilités dénombrables et leurs applications arithmétiques" in order to give further evidence supporting the analysis of BOREL (1909) presented above.

In the discussion of BOREL (1903), the focus will be on the infinite *sub-additivity* of geometrical volume. In particular, we shall focus on the result that, for suitable sets $E, E_1, E_2, \ldots$ in Euclidean n-dimensional space, the assumptions $\sum_1^\infty \mathrm{vol}(E_i) < \infty$, and $E \subset \limsup E_i$, imply $\mathrm{vol}(E)=0$. (This is the result of 1909, namely, $A_\infty = 0$ in the case of convergence, except in geometric and not probabilistic terms.) The absence of any assumptions corresponding to independence is noteworthy, *i.e.* this becomes a "CANTELLI" result when put in a probabilistic setting. It is proved by BOREL, using "CANTELLI-like" reasoning and some geometric hypotheses on the sets $E, E_1, E_2, \ldots$ sufficient to ensure that they have "volume" in an elementary sense.

It will be shown, by citations from BOREL (1903), that countable sub-additivity, the key to BOREL's proof of the above result, was intimately connected in BOREL's mind with the HEINE-BOREL Theorem. On this evidence it seems at least highly plausible that BOREL's reluctance to employ countable additivity and sub-additivity in denumerable probability was because of the strange new context in which the HEINE-BOREL Theorem was inapplicable. Because countable additivity and sub-additivity had not played a role in either finite or continuous probability by 1909 [1], there was no strongly suggestive evidence that either was an essential property to demand of "denumerable probability". Independence, on the other hand, was almost the characteristic feature of probability, and the extension of independence to the denumerable case exerted a correspondingly stronger appeal to BOREL as the essential ingredient for his new theory in 1909.

In summary, the contents of BOREL (1903) help considerably in accounting for BOREL's timidity about countable sub-additivity (in 1909), and in providing reasons for his needlessly restricted assertion about A_∞ in the convergent case.

[1] The exceptional discussion of continuous probability by WIMAN (1900; 1901) referred to above will be treated in a separate note.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

The paper BOREL (1905) represents BOREL's chief contribution to probability prior to 1909. This paper represented (for all its shortcomings) a significant advance in the foundations of probability theory which has been unaccountably neglected. It proposed that in the unit interval the meaning of (geometric) probability, identified with length, be substantially generalized to be identified with his new theory of measure.

BOREL (1905) underscores how significant was the gap between BOREL's view of (geometric) probability on the unit interval, on the one hand, and his view of denumerable probability on the other. With BOREL (1905) in mind one can be almost certain that BOREL's reservations (in 1909) about the role of countable additivity in "denumerable probability" stem from an incomplete recognition that the machinery of the "geometric point of view", applicable in his decimal and continued fraction examples, should be available as well in the general theory.

The aim of the detailed discussion of these earlier works is thus to shed light on the shortcomings of BOREL (1909) (noted above) by comparing them with viewpoints BOREL himself possessed before 1909.

7.2. *Borel (1903): Foreshadowings of Cantelli-Like Reasoning*

In BOREL (1903), continued fractions had been the main object of study and not merely a source, among others, of examples as they were in BOREL (1909). The general purpose of BOREL (1903) was to establish the existence of coverings of the interval [0, 1] by sub-intervals, membership in which demands a high degree of approximability by rationals. A typical such result is, *e.g.*, that *every* real number α possesses rational approximants P/Q such that

$$\left|\frac{P}{Q}-\alpha\right|<\frac{1}{\sqrt{5}}\frac{1}{Q^2}$$

where Q can be required to lie in a prescribed interval (A, B), satisfying

$$10<A<15A^2<B.$$

In consequence, one can prescribe an infinite sequence of intervals (A_n, B_n), satisfying the above, and there will exist a sequence of corresponding approximants

$$\left|\frac{P_n}{Q_n}-\alpha\right|<\frac{1}{5}\frac{1}{Q_n^2}.$$

The main tool used in BOREL (1903) is the repeated use of finite or denumarable covers of [0, 1] (or its n-dimensional analog, the n-dimensional cube). These covers are used to assert relations between the lengths of the covering intervals and the length (or, in higher dimensions, volume) of sets which are either contained in a finite subcover, or, alternatively, covered infinitely often.

These notions had already been exploited by BOREL in both his celebrated thesis (BOREL (1895)) and in his *Leçons sur la Théorie des Fonctions* (BOREL (1898)). What is now called the HEINE-BOREL Theorem had already been stated by BOREL, for dimension $n=1$, in his thesis and in BOREL (1898). There is no question that the

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

HEINE-BOREL Theorem (in one dimension) was central to BOREL's definition of measure, and to the definition of those sets (since called "BOREL-measurable") to which he applied this theory of measure. Although BOREL failed to generalize his theory of measure to higher dimensions, he did generalize the HEINE-BOREL Theorem, and from it he deduced a form of countable sub-additivity in higher dimensions. The extension of the HEINE-BOREL Theorem (using only open covers of an especially simple sort) occurred in BOREL (1903). There he gave the extension (Theorem VIII) to bounded closed sets in dimension n, restricting himself to denumerable covers. (BOREL later acknowledged LEBESGUE's generalization of the result to apply to non-denumerable covers as well.)

Of special interest is the occurrence of the theorem on volumes stated above. This is *Theorem XI bis* in the notation of BOREL (1903). As remarked in the introduction, it is a special case of what is now considered the "CANTELLI" part of the BOREL-CANTELLI Lemmas.

The specialities that surround this theorem of BOREL lie in the assumptions, deriving from the context of the paper of 1903 as a whole, that $E_1, E_2, \ldots, E_n, \ldots$ are "domains" in some fixed Euclidean spaces, R^k, of dimension k. A "domain" is defined as a closed bounded part of k-space determined by a finite number of algebraic inequalities

$$\phi_j(x_1, \ldots, x_k) \geqq 0 \qquad j = 1, 2, \ldots, M.$$

The volume of a domain is an n-dimensional (RIEMANN) integral, with integrand 1, over the *interior* of this domain. BOREL considered only domains which, in modern terminology, have non-empty interiors and which are the closures of their interiors. In particular, every domain considered has a strictly positive volume. (Thus, for example, closed spheres and closed cubes are "domains".) BOREL showed that if $\sum_1^\infty \operatorname{vol}(E_n)$ is convergent, then $\limsup_{n\to\infty} E_n = H$ has "small covering". Specifically, given $\varepsilon > 0$, there are domains $H_1, H_2, \ldots, H_n, \ldots$ such that

$$H \subset \bigcup_1^\infty (\text{interior } H_n), \quad \text{and} \quad \sum_1^\infty \operatorname{vol}(H_n) < \varepsilon.$$

This may be called BOREL's version of the CANTELLI Lemma. BOREL's original formulation will be presented below.

We note that BOREL stopped short of concluding $\operatorname{vol}(H) = 0$. Indeed he could not do this since the point set H need not be a domain, and so $\operatorname{vol}(H)$ need not exist as a RIEMANN integral. At the date of this paper the theory of measure had not been generalized to n-dimensional Euclidean space, nor had the theory of the LEBESGUE integral been established.

Apart from the restriction to "domains" the theorem is thus very much the CANTELLI Lemma. What is eminently remarkable is the line of reasoning employed by BOREL. The "CANTELLI" proof given in § 6.5 above, when specialized to the case at hand, provides the desired set of domains $H_1 H_2, \ldots, H_n, \ldots$ as $H_i = E_{N+i}$ where N is chosen so that $\sum_{i=1}^\infty \operatorname{vol}(E_{N+i}) < \varepsilon$.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

The interesting question is how BOREL arranged to prove this theorem. To examine this, it is necessary to read his exact formulation of this result and its immediate precursor, *Theorem XI*.

> **Theorem XI.** *Let E be a domain and $E_1, E_2, \dots, E_h, \dots$ domains such that every point of E is interior to an infinity of them; then one can assert that the volumes $v_1, v_2, \dots, v_h, \dots$ of these domains are such that the series*
>
> $$v_1 + v_2 + \cdots + v_h + \cdots$$
>
> *is divergent.*
>
> **Theorem XI bis.** *Let $E_1, E_2, \dots, E_h, \dots$, domains with volumes $v_1, v_2, \dots, v_h, \dots$ be such that the series*
>
> $$v_1 + v_2 + \cdots + v_h + \cdots$$
>
> *is convergent; then one can assert that the set H of points which belong to an infinity of these domains is such that, being given ε arbitrarily small, one can construct domains $H_1, H_2, \dots, H_\alpha, \dots$, finite or denumerably infinite in number, such that every point of H is interior to one of them and that, in addition, V_α being the volume of H_α one has*
>
> $$V_1 + V_2 + \cdots + V_\alpha + \cdots < \varepsilon.$$
>
> (Italics in the original.) BOREL (1903: 362).

The full line of development leading to these theorems provide unassailable evidence as to how BOREL, *in this context*, associated sub-additivity and the "CANTELLI Lemma" with the HEINE-BOREL result. This line of development is the following chain of theorems, occurring in Section 19 of BOREL (1903).

> **Theorem VIII.** *Let E be a given closed bounded set, and $E_1, E_2, \dots, E_p, \dots$ a denumerable infinity* [footnote omitted] *of sets such that every point of E is* INTERIOR *to at least one of them; it is possible to find among $E_1, E_2, \dots, E_p, \dots$ a* FINITE *number of sets such that every point of E is interior to at least one of them.* (Italics in the original.) BOREL (1903: 357)[1].

[1] This is as clear a rendering of the *n*-dimensional HEINE-BOREL theorem as could be wished. BOREL even supplemented it with an extension to sets which are only closed (or bounded) after a suitable projective transformation. (In modern terms, he considered *projective* 2-space, or more generally, *n*-space.) Such sets he called *projectively closed* or *projectively bounded.*

He then gave a simple example of three sets in the projective plane which are *projectively closed* and *projectively bounded* and collectively cover the plane (two dimensions being chosen for ease of exposition). Thus he could apply Theorem VIII to each of these sets. BOREL then stated a "generalized" HEINE-BOREL theorem:

> **Theorem VIII bis.** *If one has a denumerable infinity of sets $E_1, E_2, \dots, E_n, \dots$ such that every point of the plane is in the* INTERIOR *of at least one of them (the points at infinity included, of course), one can determine among the E_i a finite number of sets such that every point of the plane is interior to one of them.* (Italics in the original.) BOREL (1903: 359).

Thus BOREL was aware that the setting of the HEINE-BOREL theorem could be widened from *n*-dimensions to at least certain alternative spaces without changing the nature of the assertion.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

Next, restricting consideration to domains and their associated volumes, BOREL established his first (and key) result for sub-additivity:

> **Theorem IX.** *When a domain E is such that each point is interior to a domain* E_i $(i=1,2,3,\ldots,n\ldots)$, *one can assert that the sum of the volumes of the domains* E_i *is greater than the volume of E.* (Italics in the original.) BOREL (1903: 360).

In other words,

$$E\subset\bigcup_{1}^{\infty}(\text{interior}(E_i))$$

implies

$$\text{vol}(E)<\sum_{1}^{\infty}\text{vol}(E_i).$$

The proof, to be supplied by the reader, involves the use of the HEINE-BOREL theorem (Theorem VIII) as a preliminary to assure the existence of an N such that

$$E\subset\bigcup_{1}^{N}(\text{interior }(E_i)).$$

From this one concludes (presumably by elementary calculus)

$$\text{vol}(E)\leqq\sum_{1}^{N}\text{vol}(E_i)$$

and the result follows.

The statement of Theorem IX is to be interpreted as including the case

$$\sum_{1}^{\infty}\text{vol}(E_i)=+\infty.$$

BOREL then pointed out that one can give this same assertion an occasionally more convenient form as follows:

> **Theorem IX bis.** *Given a domain E and a denumerable infinity of domains*
>
> $$E_1, E_2, \ldots, E_n, \ldots$$
>
> *such that one has*
>
> $$\sum_{1}^{\infty}\text{vol}(E_i)<\text{vol}(E),$$
>
> *then there are points of E not in the interior of any* E_i. (Italics in the original.) BOREL (1903: 361).

This is, of course, no more than a contrapositive formulation of the original Theorem IX, requiring no further proof.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

This particular formulation can be immediately sharpened by use of an elementary argument which employs slightly larger domains E_i' containing E_i. The sharpened formulation is:

> **Theorem X.** *If E has volume v, and a denumerable infinity of domains* E_i $(i=1,2,\ldots,n,\ldots)$ *having volumes* v_i, *are such that*
>
> $$\sum_1^\infty v_i < v$$
>
> *the there are points of E belonging to none of the* E_i. (Italics in the original.) BOREL (1903: 361–362).

Theorem X has a slightly stronger conclusion than its immediate predecessor since it avoids reference to the "interiors" of the sets E_i. This non-topological version is the one which strikes the contemporary reader as foreshadowing the generalization to measure theory (or probability). As we have shown, however, it was achieved only after Theorem IX *bis* (involving "interiors"), and this in turn depended crucially on a HEINE-BOREL argument involving "open covers".

Theorem XI and XI *bis*, cited above, are now direct consequences of the preceding. BOREL leaves their proofs to the reader.

Summarizing, BOREL well knew that in certain circumstances if the sets E_k satisfy $\sum_1^\infty m(E_k) < \infty$, then the set $H = \lim\sup E_k$, of those points in infinitely many of the E_k's, has covers of arbitrarily small total volume. He did not state this in the generality of measure theory. There is no reason to suppose that BOREL so much as conceived of "abstracting" measure theory to the degree of generality needed for the CANTELLI improvement of his Zero-One Law. Indeed, he did not even consider the straightforward generalization of measure theory from 1 to n dimensions. Even if he had conceived of n-dimensional measure, the remaining step to an abstract measure remains immense. In particular, the absence of a HEINE-BOREL result in the abstract case might have proved an unbridgeable gulf, for it was this result which was basic to BOREL's methods of proof. The conclusion seems inescapable that BOREL knew the "CANTELLI" theorem in a geometric setting only, where topological considerations were relevant and available. The lack of the notation for $\lim\sup E_k$ for a collection of events further disguised the resemblance between the set H of his paper of 1903 and the nowhere designated set ("infinitely many successes") in the paper of 1909 whose probability is A_∞.

We are thus led to add an additional speculation concerning BOREL's treatment of his Zero-One Law: the absence of a topology in the space of trials prevented him from establishing a HEINE-BOREL Theorem; this absence in turn blocked the way for his use of countable sub-additivity in the setting of "probabilité dénombrable", even if he had thought of probability as analogous to volume (or measure).

In view of the *Theorem XI bis* cited above, in which sub-additivity is proved, we are simultaneously led to the surprising conclusion that BOREL was as close in 1903 to a rigorous proof of the "CANTELLI" version of his Zero-One Law as was CANTELLI in 1917. The difference is that CANTELLI's work deals explicitly with probability and so was taken up by succeeding probabilists, while BOREL's paper of

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

1903, with its intriguing Theorem XI *bis*, remained relatively neglected. Its title and even its main results give no clue to its interesting internal arguments[1].

7.3. Borel (1905): An Early Identification of Geometric Probability with Measure

The paper of 1905 contains the following four assertions of interest:

1) If one uses the convention that the probability of a set is proportional to its length (or area, or volume), it should be made explicit that this is a convention and not the intrinsic meaning of probability.

> All the preceding is well-known, but questions of probability have given rise to so many verbal controversies arising simply from a lack of agreement on the conventions of language, that it may not be superfluous to make precise the concepts which I will employ. BOREL (1905: 124).

2) For the set E of rational numbers in $[0, 1]$, and with the above convention,

$$P(E)=\int_0^1 f(x)\,dx$$

where

$$f(x)=\text{the characteristic function of } E=\begin{cases}1 & \text{if } x \text{ is rational}\\ 0 & \text{if } x \text{ is irrational.}\end{cases}$$

Similarly,

$$P(E^c)=\int_0^1 F(X)\,dx$$

where

$$F(x)=1-f(x)=\begin{cases}1 & \text{if } x \text{ is irrational}\\ 0 & \text{if } x \text{ is rational.}\end{cases}$$

These integrals produce the "réponse evidente" that $P(E)=0$, $P(E^c)=1$, but not if one restricts oneself to RIEMANN integration.

> However, if one uses the new definition of integral which is due to Lebesgue, one sees that each of the functions $f(x)$ and $F(x)$ is integrable in the sense of Lebesgue, or, more briefly, L-integrable and their L-integral provides the correct [*sic*] mean value or probability sought. Lebesgue's methods thus allow us to approach questions about probability which appear inaccessible to the classical procedures of integration. Moreover, in certain of the simplest cases, it suffices to use the theory of sets that I called *measurable* and which Lebesgue has named *B-measurable;* the application of measurable sets to probability theory was first made, to my knowledge, by Wiman[2]. BOREL (1905: 125–126).

---

[1] It should be observed that CANTELLI himself generalized the inequality

$$P\left(\bigcup E_k\right)\leqq \sum P(E_k)$$

from finite unions to denumerable ones in the same easygoing argument-by-analogy fashion as BOREL generalized finite independence to countable independence. That is, with no effort at proof, and specifically with no reference to measure theory.

[2] Authors' footnote: In a footnote BOREL cites WIMAN (1900; 1901). These papers, and more generally, the polemic between WIMAN and his colleague BRODÉN will be discussed in a separate publication.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

This shows that BOREL regarded the relation

$$P(E)=\int_0^1 f_E(x)\,dx$$

where

$$f_E(x)=\begin{cases}1 & \text{if } x\in E\\ 0 & \text{if } x\notin E\end{cases}$$

as applicable to at least all BOREL sets.

Incidentally, BOREL regards the determination $P(E)=0$ for the rationals as clearly "correct" with no sustaining argument other than appeal to the authority of POINCARÉ, who had indeed regarded it as evident, *before the existence* of a theory of measure. We have examined POINCARÉ's lecture notes, *Calcul des Probabilités (1893–1894 lectures)*, at the Institut Henri Poincaré, Paris, where this enigmatic pronouncement is to be found, already incorporated in the text at that "premature" date, some 16 years before the publication of his *Calcul des Probabilités.*

3) An example is considered, namely the set $E^{(n)}$, defined for each integer value of the parameter n as all real numbers α in $[0, 1]$ such that there are relatively prime integers p, q satisfying

$$\left|\frac{p}{q}-\alpha\right|<\frac{1}{q^n}.$$

For each integer value of n, it is evident that

$$E^{(n)}\subset\bigcup I_{p,q}^{(n)}$$

where

$$I_{p,q}^{(n)}=\left(\frac{p}{q}-\frac{1}{q^n},\ \frac{p}{q}+\frac{1}{q^n}\right),$$

p, q are relatively prime, *i.e.*, $(p, q)=1$, and the union is over all such pairs p and q. It is then asserted (without comment) that

$$P(E^{(n)})<\sum_{(p,q)=1} l(I_{p,q}^{(n)})=\sum_{q=2}^{\infty}\phi(q)\,\frac{2}{q^n}.$$

Here $\phi(q)=$ the number of integers less than and relatively prime to q, and the series clearly converges if $n>2$. This is a wonderfully clear example of BOREL's explicit use of countable sub-additivity, but only in the context of geometric probability, or what he was to call in 1909, the "point de vue géométrique".

It is unfortunate that BOREL did not relate this example in the paper of 1905 to his ideas of 1903 discussed above, by concluding further that the set of points α satisfying for any fixed $n>2$ the inequality

$$\left|\frac{p}{q}-\alpha\right|<\frac{1}{q^n}$$

for *infinitely* many distinct rationals p/q has measure zero. As a result, we lack documentary evidence of a "CANTELLI"-like assertion by BOREL that

$$P(\limsup E_n)=0 \qquad \text{if } \sum P(E_n)<\infty$$

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

in a measure-theoretic interpretation of probability by BOREL, even in the case of geometric probability when he was explicitly aware of such on interpretation.

4) A reminder that B-measurable sets form, in contemporary terms, a σ-algebra, and that the associated measure is, again in contemporary terms, σ-additive.

In sum: BOREL's identification of geometric probability with measure is explicit in 1905, and further the scope of (geometric) probability is explicitly enlarged, to apply to a σ-algebra of sets (*cf.* footnote 1, § 6.2).

Our analysis of the paper of 1909 above shows by contrast that only in the special applications (namely decimal expansions and continued fractions where the convention is explicitly made that the probability of an interval is its length) is it even marginally possible that BOREL had the same facts in view. In the paper of 1909 these examples are considered only after the fundamental Zero-One Law concerning A_∞ is obtained. This result is in turn considered independently of geometric considerations and in the seeming conviction that countable independence is the key notion, casting sub-additivity and σ-additivity aside. For BOREL, the "point de vue logique" had eclipsed the "point de vue géométrique" in 1909 even though both were applicable and in spite of the additional insights (recognized in 1905) offered by the latter.

8. The Borel-Bernstein Polemic

8.1. Introduction

BOREL's results concerning decimal (or dyadic) digits and continued fractions attracted considerable stir in the years immediately following 1909. In 1910 FABER proved again the result concerning the decimal digits, though in a substantially different way (*cf.* § 9.2). In 1911 F. BERNSTEIN attacked BOREL's proof of the Continued Fraction Theorem, supplying an alternative of his own. BOREL responded to BERNSTEIN's paper with a modification of his Zero-One Law (1912), adding that this modification, coupled with inequalities already in the paper of 1909, sufficed to validate his result on continued fractions. The exchange with BERNSTEIN offers an opportunity to examine once again BOREL's own interpretation of "probabilités dénombrables". In particular, it decisively sustains our assertion that "composite probability" lay at the heart of his theory, and that countable additivity (and its corollary, countable sub-additivity for non-disjoint unions) was in no way central to BOREL's conception of probability. BOREL (1912) also indicates how far BOREL was from applying "CANTELLI" reasoning to the case $\sum p_n$ convergent: indeed, to show his (1912) modification of his Zero-One Law validates the Continued Fraction Theorem (both cases), he chose the *convergent* case for detailed exposition, although the "CANTELLI" reasoning (*i.e.* use of sub-additivity) provides a stronger and simpler modification for this case.

Finally, he marred his own defense by insisting that a specific inequality of BOREL (1909) was equivalent to one of BERNSTEIN's (1911). In fact while the desired inequality was perhaps latent in the reasoning of BOREL's paper of 1909, only a hopelessly weakened version was explicitly given in the passage cited by BOREL.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

The paper of 1912 partially rebuts the possible claim that BOREL could perfectly well have patched up his paper of 1909 if the need had been pointed out to him. Indeed, he had the opportunity in 1912, but there failed to alter his original calculations sufficiently.

8.2. The Criticism: The Contribution of F. Bernstein (1911)

Writing in 1911, F. BERNSTEIN considered a problem in Celestial Mechanics which represented one of the several early instances of the intrusion of "point set theory" into dynamics and classical mechanics. The problem considered is: which possible configurations of a certain 3-body problem admitted a "mean motion". Put entirely in mathematical terms, and dealing first with the n-body problem, BERNSTEIN considered the real and imaginary parts of the finite sum

$$\sum_1^n r_m e^{i(g_m t + h_m)} \qquad (r_m > 0).$$

The existence of a mean motion means, for BERNSTEIN, that the sum may be put in the form

$$r(t)\, e^{i\omega(t)}$$

where

$$\omega(t) = ct + f(t)$$

and $f(t)$ is *bounded* for all t. The constant c is then called the mean motion, and $c = \lim_{t\to\infty} \frac{\omega(t)}{t}$. (The requirement that $f(t)$ be bounded which was of interest to BERNSTEIN, has since been dropped from the definition of mean motion, but we use the term as employed by BERNSTEIN.)

If $n=2$, a mean motion exists, and if $n=3$ and one of the r_1, r_2, r_3 exceeds the sum of the other two, again a mean motion exists. This was shown by LAGRANGE. If $n=3$ and LAGRANGE's condition fails, the question reduces, following BOHL (1909), to a discussion of two associated quantities, ρ, ζ, defined by

$$\rho = \frac{g_2 - g_1}{g_3 - g_1}, \qquad \zeta = \frac{1}{\pi}(\omega_3 + \rho\,\omega_2).$$

Here the notation is defined after observing that, since LAGRANGE's condition fails, there is a (unique) triangle with sides r_1, r_2, r_3; $\omega_1, \omega_2, \omega_3$ are defined as the angles opposite r_1, r_2, r_3 respectively. The result of BOHL in 1909 was that $\lim_{t\to\infty} \frac{\omega(t)}{t}$ exists for every choice of ρ, ζ, and so

$$\omega(t) = cr + f(t)$$

where $\lim_{t\to\infty} \frac{f(t)}{t} = 0$. But on the other hand, he showed there is a dense set of points (ρ, ζ) in the unit square for which the corresponding $f(t)$ is not bounded, and hence for which there is no mean motion in the sense employed by BERNSTEIN.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

BERNSTEIN's paper is devoted to investigating the *measure* of the points (ρ, ζ) in the unit square for which a mean motion exists. His result is that this set is of measure zero.

The significance of this for Celestial Mechanics is not of interest here. BERNSTEIN shows how the result reduces to determining the measure of points in the unit interval which permit approximation to high degree by means of rational approximants. Because of this, BERNSTEIN is led to discuss the measure of (irrational) numbers in the unit interval whose continued fraction expansions have infinitely many elements a_n which are large. Specifically, he establishes and employs the result that the set of x whose corresponding elements $a_n = a_n(x)$ are unbounded as n ranges over the odd integers is of measure 1.

Conveniently, all the theorems about continued fractions that he establishes are grouped in a self-contained section entitled "The geometric probability for the approximation of real numbers by rational numbers, to stronger order than continued fraction approximations, and related topics". The chief result employed for applications of mean motions is his Theorem 2:

Those irrational numbers x in (0, 1) *which satisfy*

$$a_{n_r} < k \quad \text{or } a_{n_r} \geqq k, \qquad k > 1$$

for $r = 0, 1, 2, \ldots$ *along some specified subsequence* $n_1 < n_2 < n_3 < \cdots < n_r \ldots$ *are points of measure zero.* BERNSTEIN (1911: 428).

Taking the union of these "Null-menge" for $k = 1, 2, 3, \ldots$ clearly implies the result about unbounded growth of a_n with odd indices cited above. BERNSTEIN further establishes the BOREL Continued Fraction Theorem in a new manner (BERNSTEIN (1911: 256) Theorem 4). His proof is unexceptionable, although his summing up of the argument in the form of a theorem is cloudy. He is at pains to point to the flawed character of BOREL's proof in its use of independence between "trials" $a_n \geqq \phi(n)$, and never to use such reasoning himself.

We now sketch BERNSTEIN's section 2. To the notation already introduced, we add the notation

$$P[a_n \geqq k \mid a_1 = m_1, a_2 = m_2, \ldots, a_{n-1} = m_{n-1}]$$

to denote the *conditional probability* that $a_n(x) \geqq k$ given that $a_i(x) = m_i$, $i = 1, 2, \ldots, n-1$. This is, by definition, a ratio of lengths: the numerator is

$$\sum_{m_n = k}^{\infty} l[a_1 = m_1, \ldots, a_{n-1} = m_{n-1}, a_n = m_n],$$

the "length" (*i.e.* measure) of the point-set

$$\bigoplus_{m_n = k}^{\infty} [a_1 = m_1, \ldots, a_{n-1} = m_{n-1}, a_n = m_n]$$

and the denominator is

$$l[a_1 = m_1, \ldots, a_{n-1} = m_{n-1}],$$

the length of a single interval corresponding to the set in which $a_i(x) = m_i$, $i = 1, 2, \ldots, n-1$.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

BERNSTEIN calls this ratio the "geometric probability" that $a_n \geqq k$. His results are all dependent on the preliminary inequalities:

$$\left.\begin{aligned} \frac{1}{k} < P[a_n \geqq k \mid a_1 = m_1, \ldots, a_{n-1} = m_{n-1}] < \frac{2}{k+1}, \\ 1 - \frac{2}{k+1} < P[a_n < k \mid a_1 = m_1, \ldots, a_{n-1} = m_{n-1}] < 1 - \frac{1}{k}. \end{aligned}\right\} \tag{8.1}$$

Equations (8.1) are BERNSTEIN's (1911: 426: (42) and (43)) re-written in the terminology of conditional probability. (The corresponding "global" inequalities

$$\left.\begin{aligned} \frac{1}{k} < P[a_n \geqq k] < \frac{2}{k+1}, \\ 1 - \frac{2}{k+1} < P[a_n < k] < 1 - \frac{1}{k}, \end{aligned}\right\} \tag{8.2}$$

follow immediately as weighted averages of these.)

The inequalities (8.2) on global probabilities (*cf.* BERNSTEIN (1911: 427: equation 47)) may be compared to the somewhat analogous inequalities of BOREL:

$$\frac{2}{3(k+1)} < P[a_n > k] < \frac{3}{k+2}, \tag{8.3}$$

$$\frac{k-1}{k+2} < P[a_n \leqq k] < \frac{3k+1}{3k+3} \tag{8.4}$$

(*cf.* (8.1) and its complemented form).

The distinction between these and BERNSTEIN's inequalities (8.1) is crucial: BOREL had not calculated bounds for conditional probabilities, but only for "global" probabilities. (Even where BOREL calculates bounds for ratios of global probabilities, *e.g.*

$$\frac{P[a_n > k+1]}{P[a_n > k]} \quad \text{or} \quad \frac{P[a_1 = m_1, \ldots, a_{n-1} = m_{n-1}, a_n > k+1]}{P[a_1 = m_1, \ldots, a_{n-1} = m_{n-1}, a_n > k]},$$

cf. (6.1) and (6.2), these are not probabilities on a_n conditioned on the behavior of earlier $a_1, \ldots, a_{n-1}$, *i.e.* are not BERNSTEIN's "geometric probabilities".)

BERNSTEIN then employs an ingenious argument to show that the probability of cylinder sets described by the simultaneous conditions

$$a_{n_1} \geqq k_{n_1}, a_{n_2} \geqq k_{n_2}, \ldots, a_{n_r} \geqq k_{n_r}$$

satisfy inequalities

$$\begin{aligned} \frac{1}{k_{n_1}} \frac{1}{k_{n_2}} \cdots \frac{1}{k_{n_r}} < P[a_{n_1} \geqq k_{n_1}, a_{n_2} \geqq k_{n_2}, \ldots, a_{n_r} \geqq k_{n_r}] \\ < \frac{2}{(k_{n_1}+1)} \cdots \frac{2}{(k_{n_r}+1)} \end{aligned} \tag{8.5}$$

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

and similarly

$$\left(1-\frac{2}{k_{n_1}+1}\right)\cdots\left(1-\frac{2}{k_{n_r}+1}\right)<P[a_{n_1}<k_{n_1},\ldots,a_{n_r}<k_{n_r}]$$
$$<\left(1-\frac{1}{k_{n_1}}\right)\cdots\left(1-\frac{1}{k_{n_r}}\right). \tag{8.6}$$

These bounds are for "global", not conditional probabilities. More important, they are for the probabilities of a particular kind of cylinder set, namely intersections of the individual sets (or "trials") defined by individual inequalities of the form $a_n \geqq k_n$ (or $a_n < k_n$, respectively), for various finite choices of n. Most important of all, these inequalities coincide with what one could have obtained from the individual "global" inequalities

$$\frac{1}{k_n}<P[a_n \geqq k_n]<\frac{2}{k_n+1}$$

and

$$1-\frac{2}{k_n+1}<P[a_n<k_n]<1-\frac{1}{k_n}$$

(essentially equivalent to the "global" inequalities of BOREL; *cf.* (8.3), and (8.4)) *had one been allowed to assume independence as well.*

BERNSTEIN's ingenious argument follows from the purely algebraic identity

$$\begin{aligned}P[a_1=m_1, a_2=m_2, \ldots, a_n=m_n]\\ =P[a_n=m_n \mid a_i=m_i, 1\leqq i\leqq n-1]\, P[a_{n-1}=m_{n-1} \mid a_i=m_i, 1\leqq i \leqq n-2]\\ \ldots P[a_3=m_3 \mid a_1=m_1, a_2=m_2]\, P[a_2=m_2 \mid a_1=m_1]\, P[a_1=m_1].\end{aligned} \tag{8.7}$$

This is the form of "probabilités composées" which is adequately general for the case at hand. This might well be called the (finite) *Chain Law of Probability.*

To prove BERNSTEIN's inequalities (8.5) and (8.6) from the Chain Law, denote the product on the right of the Chain Law by $\prod_n (m_1, m_2, \ldots, m_n)$. Fix $m_1, m_2, \ldots, m_{n-1}$ and sum both sides for all values of m_n ranging from a fixed lower limit to infinity. If the lower limit is 1, both sides simplify, and the result is the original assertion with all references to a_n deleted, involving $\prod_{n-1}$ on the right-hand side. If the lower limit is some integer $k_n > 1$, the left-hand side becomes

$$P[a_1=m_1, \ldots, a_{n-1}=m_{n-1}, a_n \geqq k_n]$$

while the right-hand side of (8.7) can be estimated from above and from below as the product

$$\prod_{n-1} = P[a_{n-1}=m_{n-1} \mid a_i=m_i, 1\leqq i\leqq n-2] \ldots P[a_2=m_2 \mid a_1=m_1]\, P[a_1=m_1]$$

times the upper and lower estimates, respectively, which hold for the leading factor (*cf.* (8.1))

$$P[a_n=m_n \mid a_i=m_i, 1\leqq i\leqq n-1].$$

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

In other words

$$\frac{1}{k_n}\prod_{n-1} < P[a_i = m_i,\ 1 \leqq i \leqq n-1,\ a_n \geqq k_n] < \frac{2}{k_n+1}\prod_{n-1}.$$

In a similar fashion we estimate $\prod_{n-1}$ from above and below, or reduce it to $\prod_{n-2}$, as follows: We sum over m_{n-1} from some lower limit to infinity, deleting reference to a_{n-1} if this lower limit is 1, or else obtaining

$$\frac{1}{k_{n-1}}\prod_{n-2} < \prod_{n-1} < \frac{2}{k_{n-1}+1}\prod_{n-2}$$

if the lower limit is $k_{n-1} > 1$. Proceeding in this fashion, using a precribed finite collection of indices $n_1, \ldots, n_r$ with prescribed lower limits $k_{n_1}, k_{n_2}, \ldots, k_{n_r}$, we arrive at the desired inequality (8.5). The missing intermediate indices correspond to summations with lower limit 1. The inequality (8.6) for $P[a_{n_1} < k_{n_1}, \ldots, a_{n_r} < k_{n_r}]$ is obtained similarly or, alternatively, can be viewed as an immediate corollary.

BERNSTEIN next *explicitly employs the concept of measure and the various forms of countable additivity* (referring to LEBESGUE (1906)) in order to extend previous considerations to sets defined by infinitely many inequalities. He thus calculates the probability of the set

$$[a_{n_1} \geqq k_{n_1}, a_{n_2} \geqq k_{n_2}, \ldots, a_{n_r} \geqq k_{n_r}, \ldots]$$

and of the set

$$[a_{n_1} < k_{n_1}, a_{n_2} < k_{n_2}, \ldots, a_{n_r} < k_{n_r}, \ldots].$$

These sets involve infinite sequences of inequalities and thus are no longer cylinder sets. For instance, his Theorem (2), referred to above, results by applying BERNSTEIN's inequalities (8.5) and (8.6) to a given sequence $\{n_r\}$ of indices, and choosing $k_{n_r} = k$ for $r = 1, 2, 3, \ldots$. Since

$$\lim_r \left(1 - \frac{2}{k+1}\right)^r = \lim_r \left(1 - \frac{1}{k}\right)^r = \lim_r \left(\frac{1}{k}\right)^r = \lim_r \left(1 - \frac{1}{k+1}\right)^r = 0,$$

inequalities (8.5) and (8.6) and countable additivity imply the two results:

$$\begin{aligned} P[a_{n_1} \geqq k, a_{n_2} \geqq k, \ldots, a_{n_r} \geqq k, \ldots] \\ = \lim_{r\to\infty} P[a_{n_1} \geqq k, a_{n_2} \geqq k, \ldots, a_{n_r} \geqq k] = 0 \end{aligned}$$

and

$$P[a_{n_1} < k, a_{n_2} < k, \ldots, a_{n_r} < k, \ldots] = \lim_{r\to\infty} P[a_{n_1} < k, \ldots, a_{n_r} < k] = 0.$$

The sets $[a_{n_r} \geqq k, r = 1, 2, 3, \ldots]$ and $[a_{n_r} < k, r = 1, 2, 3, \ldots]$ are both of measure zero for each value of k. Hence unions of sets such as $k = 2, 3, \ldots$ are still of measure zero; thus the probability that $\{a_{n_r}\}$ is bounded from above (or below) is zero.

To obtain BOREL's theorem on Continued Fractions, it suffices to apply (8.5) and (8.6) to the indices $n, n+1, \ldots, n+r$, and change $k_n, k_{n+1}, \ldots, k_{n+r}$ to the BOREL

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

notation $\phi(n), \phi(n+1), \ldots, \phi(n+r)$. Then from inequality (8.6) it follows that

$$\prod_{\nu=n}^{r}\left(1-\frac{2}{\phi(\nu)+1}\right)<P[a_\nu<\phi(\nu), \nu=n, n+1, \ldots, r]<\prod_{\nu=n}^{r}\left(1-\frac{1}{\phi(\nu)}\right).$$

If $\sum_1^\infty \frac{1}{\phi(\nu)}$ diverges,

$$P[a_\nu<\phi(\nu), \nu=n, n+1, \ldots]=\lim_{r\to\infty} P[a_\nu<\phi(\nu), \nu=n, n+1, \ldots, r]=0$$

(the divergent case of BOREL's theorem). If $\sum_1^\infty \frac{1}{\phi(\nu)}$ converges, (8.5) assures that

$$P[a_\nu<\phi(\nu), \nu=n, n+1, \ldots, r]$$

is positive (and less than 1) and in fact lies between

$$\prod_{\nu=n}^{\infty}\left(1-\frac{2}{\phi(\nu)+1}\right) \quad \text{and} \quad \prod_{\nu=n}^{\infty}\left(1-\frac{1}{\phi(\nu)}\right).$$

To achieve BOREL's result in this case, it is necessary to compute the probability that $a_\nu<\phi(\nu)$ hold "from some n on", that is the probability of the union of the sets $[a_\nu<\phi(\nu), \nu\geqq n]$ over all integer values of n. This probability is thus 1, again validating BOREL's result. If the phrase "from some n on" is interpreted to mean "from some *fixed value* of n on" then the answer is strictly less than one. This ambiguity of language resulted in needless, and for our purpose, irrelevant conflict between what BOREL said he proved and what BERNSTEIN actually succeeded in proving. Their results coincide, but BERNSTEIN actually supplied a valid, independently conceived proof, explicitly utilizing the language and theorems of measure theory. BOREL, by contrast, first asserted the theorem in 1909 but there supplied a non-proof, vitiated by dependence between "trials" and employing only the Zero-One Law. As we have seen, the Zero-One Law proved by BOREL has nothing to do with the language of LEBESGUE measure, and, in BOREL's exposition, has no clearly recognized link to countable additivity or sub-additivity.

After achieving his proof of BOREL's result on Continued Fractions, BERNSTEIN emphasizes that the individual events $a_n=m_n$ are not independent. He computes the four probabilities $P[a_1=1]$, $P[a_1=2]$, $P[a_1=1, a_2=1]$, $P[a_1=2, a_2=1]$ as lengths of intervals, and observes that the two ratios

$$\frac{P[a_1=1, a_2=1]}{P[a_1=2, a_2=1]} \quad \text{and} \quad \frac{P[a_1=1]}{P[a_1=2]}$$

are unequal, as true independence would require. He points out that the adequate law of "probabilités composées" (or as BERNSTEIN calls it "das Theorem der zusammengesetzten Wahrscheinlichkeiten") involves products of *composite* probabilities to compute probabilities of the simultaneous occurrence of several events. This is what we have called the Chain Law. The example he gives, in modern

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

notation, considers two events, A and B, where $A = \oplus A_i$ decomposes event A into its possible alternatives. Then

$$P(A \cap B) = P(\oplus(A_i \cap B)) = \sum P(A_i)\, P(B \mid A_i).$$

If $\underline{q} \leqq P(B \mid A_i) \leqq \bar{q}$ for all i, then

$$\underline{q} P(A) \leqq P(A \cap B) \leqq \bar{q} P(A).$$

Thus, bounds on conditional probabilities for B, valid for all alternative conditions, give rise to the same inequality concerning $P(A \cap B)$ as would have been obtained if the global probability $P(B)$ were known only to satisfy

$$\underline{q} \leqq P(B) \leqq \bar{q}$$

and the events A and B were independent.

This observation could have been applied directly to BOREL's Zero-One Law in the degree of generality of "denumerable probability". Since, however, BERNSTEIN's whole viewpoint is not to create a new theory of denumerable probability but to employ the existing theory of measure, he does not make the observation that we can make on his behalf: BERNSTEIN introduced the key element in generalizing the BOREL Zero-One Law in the case not considered by CANTELLI, *i.e.* the divergent case. CANTELLI demoted the hypothesis of independence in BOREL's Zero-One Law in the convergent case (the "zero" half) by showing it could be omitted. BERNSTEIN generalized the case of divergence (the "one" half) by showing the hypothesis of independence could be weakened to requiring adequate (upper) bounds on certain conditional probabilities, these bounds themselves forming a divergent series. In fact, had he been sufficiently thorough-going in his use of measure theory, and had he not insisted, like BOREL, on treating the convergent and divergent cases in like manner, he could have shown that in the convergent case sub-additivity estimates alone would have established the theorem in terms of the global probabilities $P[a_n \geqq \phi(n)]$. He would thus have anticipated CANTELLI. Indeed, in his section (1), he explicitly calculates the measure of the limit inferior of a collection of sets, attributing this calculation to LEBESGUE (1906). This is, in fact, a modern "CANTELLI-like" calculation. However that may be, the fact is that BERNSTEIN *gave the first valid proof of* BOREL's *result on Continued Fractions, and gave the first generalization of* BOREL's *Zero-One Law to cover the possibility of dependence.*

8.3. Borel's Response: Borel (1912)

After BERNSTEIN's paper, the status of BOREL's Continued Fraction Theorem was, for a short period, in doubt. Both BERNSTEIN and BOREL agreed that the probability that $a_n < \phi(n)$ should hold from some n on was zero if $\sum \frac{1}{\phi(n)}$ diverged. BOREL asserted this in the form that the probability that $a_n \geqq \phi(n)$ holds infinitely often, the A_∞ of his paper of 1909, is 1. In the case $\sum \frac{1}{\phi(n)}$ convergent, BOREL

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

asserted that the probability that $a_n \geqq \phi(n)$ holds infinitely often is zero; thus $a_n < \phi(n)$ holds from some n on with probability one. BERNSTEIN asserted in the convergent case the probability that $a_n < \phi(n)$ holds from some n on (meaning $a_n < \phi(n)$, $a_{n+1} < \phi(n+1)$, ... for a given n) is positive but less than one. BERNSTEIN thus criticized both BOREL's argument and his result. Stung by this, BOREL responded with a note (BOREL (1912)). There he insisted that his result was correct as stated, namely A_∞ is zero or one in the continued fraction case according as $\sum \frac{1}{\phi(n)}$ converges or diverges, but completely reworked the proof in such a fashion that dependencies were permitted. Having generalized his original Zero-One Law in this way, BOREL attempted to apply it to prove his continued fraction result. As we shall see, the application required a modification of his original calculations of 1909. This BOREL failed to undertake, perhaps unwilling to concede such an inadequacy in his calculations of 1909.

BOREL's new poof of the generalized Zero-One Law essentially coincides with the argument of BERNSTEIN. That is, the general law of "probabilités composées" (in terms of *conditional* probabilities), replaces countable independence as the essential tool. BOREL then shows that appropiate inequalities on conditional probabilities give rise to "independence-like" estimates in the form of products (*cf.* (8.5), (8.6), above). But BOREL, in 1912, places these arguments in the generality of his original Zero-One Law, *i.e.*, the space of all denumerable sequences of trials, with possible success or failure at each trial, whereas BERNSTEIN had been concerned only with the application to continued fractions.

BOREL's original notation p_n for success and q_n for failure at the n^{th} trial is inadequate for his new proof. We also need the corresponding conditional probabilities, since we explicitly allow dependence. Let us introduce the notation s_n for a "success" parameter: $s_n = 1$ means the n^{th} trial was a success, $s_n = 0$ means it was a failure. Let x be an infinite sequence of trials; then $s_n(x)$ is 1 if x has success at the n^{th} trial and 0 if x has failure at the n^{th} trial. The probability that a sequence has prescribed initial values $s_i = m_i$, $i = 1, 2, \ldots, n$ where each $m_i = 0$ or 1 is given recursively by

$$P[s_1 = m_1, \ldots, s_n = m_n] = P[s_n = m_n \mid s_1 = m_1, \ldots, s_{n-1} = m_{n-1}] \cdot P[s_1 = m_1, \ldots, s_{n-1} = m_{n-1}]$$

so that

$$P[s_1 = m_1, \ldots, s_n = m_n] = \prod_{k=1}^{n} P[s_k = m_k \mid s_1 = m_1, \ldots, s_{k-1} = m_{k-1}].$$

This is precisely the Chain Law. Similarly the probability that a sequence finish with the results $s_{n+1} = m_{n+1}, s_{n+2} = m_{n+2}, \ldots$ given the initial values $s_1 = m_1, \ldots, s_n = m_n$ is the infinite product

$$\prod_{k=1}^{\infty} P[s_{n+k} = m_{n+k} \mid s_1 = m_1, \ldots, s_{n+k-1} = m_{n+k-1}]. \tag{8.8}$$

This is the *countable* extension of the Chain Law that BOREL asserts in 1912. (BOREL offers no justification for the extension from the finite to the countable case,

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

one more instance of a tacit and extremely well disguised dependence on countable additivity.) It is employed in his response to BERNSTEIN to prove the following theorem: Let $A_0, A_1, \ldots, A_k, \ldots$ be, as in 1909, the probability that there are exactly k successes, and let $A_\infty = 1-(A_0+A_1+\cdots+A_k+\cdots)$. Let $\{p'_n\}$ and $\{p''_n\}$ be two sequences such that

$$p'_n < P[s_n = 1 \mid s_1 = m_1, \ldots, s_{n-1} = m_{n-1}] < p''_n$$

for $m_1, m_2, \ldots, m_{n-1}$ running through all choices of 0's and 1's.

$$\text{If } \sum p''_n \text{ converges, then } A_\infty = 0.$$
$$\text{If } \sum p'_n \text{ diverges, then } A_\infty = 1.$$

Equivalently,

$$\text{If } \sum p''_n \text{ converges, then } A_0+A_1+\cdots+A_k+\cdots = 1.$$
$$\text{If } \sum p'_n \text{ diverges, then } A_0+A_1+\cdots+A_k+\cdots = 0.$$

BOREL presents a proof only for the convergent case, since that is the case for which BERNSTEIN thinks (erroneously) his result is in contradiction with BOREL's. Of course "CANTELLI" reasoning renders BOREL's new proof of this case obsolete, but BOREL's proof indicates how to handle the divergent case as well, and therefore merits attention.

If $\sum p''_n$ converges, consider $A_0+A_1+\cdots+A_k$. BOREL shows that given $\varepsilon > 0$, k can be chosen so large that

$$A_0+\cdots+A_k > 1-\varepsilon.$$

The proof is a straightforward application of the generalized law of composite probabilities (8.7): $A_0+\cdots+A_k$ is the probability of having at most k successes, therefore greater than the probability of having no successes from the $(k+1)^{\text{st}}$ on. In symbolic form

$$A_0+\cdots+A_k > P[s_{k+1}=0, s_{k+2}=0, \ldots].$$

But the latter is the weighted average of the probabilities

$$P[s_{k+1}=0, s_{k+2}=0, \ldots \mid s_1 = m_1, \ldots, s_k = m_k]$$

(*cf.* (8.8)) for all 2^k choices of $(m_1, \ldots, m_k)$. By use of the countable extension of the Chain Law each of these probabilities (and hence their weighted average $P[s_{k+1} = 0, s_{k+2}=0, \ldots]$) is estimated from below by the infinite product

$$(1-p''_{k+1})(1-p''_{k+2})\cdots = \prod_{j=k+1}^{\infty} (1-p''_j).$$

If $\sum p''_j < \infty$, then k can be chosen large enough to make this product greater than $1-\varepsilon$, as desired, showing

$$A_0+A_1+\cdots+A_k+\cdots = 1.$$

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

Similarly, if $\sum p_j'$ diverges, $A_0+A_1+\cdots+A_k$ can be shown to vanish, though BOREL omits this. A proof can be based on the inequality

$$P[s_{n+1}=0, s_{n+2}=0, \ldots, s_N=0 | s_1=m_1, \ldots, s_n=m_n] < (1-p_{n+1}') \ldots (1-p_N'),$$

supplemented by continuity (*i.e.* countable additivity); *cf.* BARONE (1974).

In summary, BOREL in 1912 presents a new, more general Zero-One Law, based on a new, more general method of proof (based on the countable chain Law utilized previously and explicitly by BERNSTEIN). Countable additivity and sub-additivity are still neglected principles; countable composite probability is employed in the forms

$$P\left[\bigcap_1^\infty E_k\right] = \prod_{k=1}^\infty P[E_k | E_1, E_2, \ldots, E_{k-1}],$$

$$P\left[\bigcap_1^\infty E_k\right] = P\left[\bigcap_1^n E_k\right] \prod_{k=1}^\infty P[E_{n+k} | E_1, \ldots, E_{n+k-1}].$$

We repeat, the extension of the Chain Law from the finite to the countably infinite case is offered without explanation. By contrast, BERNSTEIN uses countable additivity explicitly to calculate the probabilities of non-cylinder sets of the form $\bigcap_1^\infty E_k$, utilizing the additional insight that these probabilities are (in every case under consideration) measures of measurable sets.

BOREL fails to remark that BERNSTEIN's proof is much the same as the one he now presents to prove the generalized Zero-One Law, but asserts that "the new proof is essentially the one that would have been given in 1909 if all the calculations had been written in full."

It only remained for BOREL to reassure the reader (and himself) that the modified Zero-One Law applies to the continued fraction case.

The needed inequalities are the upper and lower bounds for

$$P[a_n \geqq \phi(n) | a_1, \ldots, a_{n-1}]$$

independent of what constraints are placed on $a_1, \ldots, a_{n-1}$. BERNSTEIN had found exactly such bounds. BOREL now in 1912 asserts that the inequalities (which he refers to as "(23) and thereafter on page 268") of his paper of 1909 provide the same information. In fact, the essential inequalities of BOREL in 1909 are, in the order of their derivation,

$$\frac{k}{k+2} < \frac{P[a_n=k+1]}{P[a_n=k]} < \frac{k+1}{k+3},$$

$$\frac{2}{k(k+1)} < \frac{P[a_n=k]}{P[a_n=1]} < \frac{6}{(k+1)(k+2)},$$

$$\sum_{k=1}^\infty P[a_n=k]=1,$$

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

$$\frac{2}{3\,k(k+1)} < P[a_n = k] < \frac{3}{(k+1)(k+2)},$$
$$\frac{2}{3(k+1)} \leqq P[a_n \geqq k+1] < \frac{3}{k+2}.$$

None of these involve conditional probabilities of the sort required. For example, the last is not

$$\frac{2}{3(k+1)} < P[a_n \geqq k+1 \mid a_1 = m_1, \ldots, a_{n-1} = m_{n-1}] < \frac{3}{k+2}$$

as would be required for application of the generalized Zero-One Law but is only a much weakened version of it.

BOREL remained aware that his original exposition was flawed, for in 1926 he published a considerably more detailed and expanded version of his paper of 1909 under the title "Applications à L'Arithmétique et à la Théorie des Fonctions", as fascicule I of Tome II of his extensive "Traité du Calcul des Probabilités et de ses Applications". Now he presented the material of his response in 1912 to BERNSTEIN (*i.e.*, his more general Zero-One Law) *before* turning to Continued Fractions. This time he presents the above five inequalities, numbered (1), (2), (3), (4), (5), as stated (*i.e.*, *not* conditioned) and then remarks:

> The inequalities (1), (2), (3), (4), (5) remain true if one makes various hypotheses concerning the elements $a_1, a_2, \ldots, a_{n-1}$. The sum of the lengths l_k, l_{k+1} of the intervals considered, instead of ranging over all possible values of the elements $a_1, \ldots, a_{n-1}$ will only range over those values satisfying certain given conditions. The global probabilities $P[a_n = k]$, $P[a_n = k+1]$, $P[a_n \geqq k+1]$ will be replaced by the probabilities obtained by taking account of the hypotheses made on the initial $n-1$ elements, and whatever those may be, by following the same reasoning whereby they were established above the preceding inequalities will continue to be satisfied by the new probabilities. BOREL (1926: 66).

This statement, in 1926, is correct. BOREL's assertion, in 1912, that his paper of 1909 already contained the desired inequalities involving p_n' and p_n'' is false or, at least, disingenuous.

8.4. *Early observations of Lebesgue and Lévy*

One final historical comment is in order concerning BOREL's contribution to the exchange with BERNSTEIN. The second edition of BOREL's *Leçons sur la Théorie des Fonctions* (BOREL (1914)) contains many "notes" added especially for this edition. Note V consists of a reproduction of BOREL (1909) and BOREL (1912), *in toto*. A footnote was added to this reprinted paper of 1912 which does not appear in the original. It follows the sentence

> What is valid in Bernstein's objection is that the reasoning that I have given ... assumes the probabilities are independent and should be modified when they are not. BOREL (1914: 208).

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

The footnote states

> I should say that this objection was made to me in a personal letter, by Lebesgue, at the time of the publication of the *Rendiconti*. I assured myself that the results were valid and attached no importance to the objection; I had even forgotten it when, a few years later, I replied to Bernstein: only after the publication of this response, reproduced here, did I come across the old letter of Lebesgue. BOREL (1914: 208: Footnote (4)).

LEBESGUE and BERNSTEIN were not alone in observing that BOREL's paper of 1909 suffered from defects of exposition. In a letter to us P. LÉVY wrote concerning it:

> Yet, on reading it, perhaps without at once fully understanding its importance, my impression was above all one of surprise.
>
> I had no idea that such simple principles, which had been familiar to me since 1907, were new (I am speaking of the first two chapters; chapter 3 where continued fractions were discussed was new to me).
>
> I was surprised that a scholar whose work on divergent series, entire functions, Picard's theorem, and the theory of measure I admired, had given such complicated proofs of such simple theorems (this time I am speaking of the three chapters). (Letter from P. LÉVY, dated December 22, 1969.)

9. Early Re-workings of Borel's Strong Law

9.1. Introduction

This chapter is devoted to other contributions from the period immediately following BOREL's landmark paper of 1909. The mathematicians of this time, who were attracted by BOREL's results, helped to illuminate the relation between measure theory and probability by their efforts, though their interest was primarily in the direction of the former and not in the direction of probability. Three men —FABER, HAUSDORFF and RADEMACHER—reproved the BOREL Strong Law without any reference to the Central Limit Theorem; indeed, their merit, paradoxically, was that they did *not* concern themselves with the probabilistic interpretation of the theorem. By examining their original works, one can observe the shift in viewpoint from that of BOREL (1909) to a viewpoint in which probability (specifically geometric probability on [0, 1]) *meant* measure.[1]

Further, HAUSDORFF also proved a result on continued fractions closely akin to the BERNSTEIN-BOREL result, this again by interpreting the theorem in the setting of measure theory.

The first in order of occurrence, FABER, regarded the coextension of BOREL's denumerable probability and LEBESGUE measure as an open question which he was at pains to raise; HAUSDORFF was much more assertive and went so far as to offer explicitly an "arbitrary" definition of probability as (LEBESGUE) measure. RADEMACHER regarded the dual viewpoints as sufficiently well accepted as to be almost without need of comment. The work of FABER and RADEMACHER (the

[1] As was the viewpoint adopted by BOREL in 1905.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

latter unintentionally duplicating the former) will be dealt with first. To preserve chronological order we intersperse between them HAUSDORFF's brief but historically significant comments on measure as a possible *definition* of probability.

The subsequent section is an exposition of more technical aspects of HAUSDORFF's work, on both binary and continued fraction expansions.

9.2. The Contribution of G. Faber: His Query on the Relation of Probability to Measure

G. FABER (1910) and H. RADEMACHER (1918) each construct a continuous monotone function in [0, 1] whose difference quotients at a point x depend on the distribution of digits in the expansion of x. Their constructions are virtually identical. In each case a monotone function f is constructed (as the limit of an auxiliary sequence $\{f_n\}$) such that the value of the difference quotient

$$\frac{f\left(\frac{x_n}{10^n}\right)-f\left(\frac{x_n-1}{10^n}\right)}{\frac{1}{10^n}}$$

depends only on

$$\frac{\gamma_k(n)}{n}, \qquad k=0,1,2,\dots,9$$

where $\gamma_k(n)$ counts the number of k's among the first n digits in the decimal expansion of x, and where the integer x_n is such that

$$\frac{x_n-1}{10^n}<x<\frac{x_n}{10^n}.$$

The main result which follows from further specifics of the construction is that the set of points x for which

$$\lim_{n\to\infty}\frac{\gamma_k(n)}{n}\neq\frac{1}{10}$$

are points of non-differentiability. We note that a similar construction can be applied to binary expansions, or to expansions in any given base, and in every case the result is that the numbers which are not "normal" with respect to the given basis, in the sense of BOREL, are points of non-differentiability.

Both FABER and RADEMACHER then appealed to the powerful result of LEBESGUE which asserts that for *every* monotone function the set of points of non-differentiability is of measure zero.

Thus, the functions constructed by FABER and RADEMACHER, together with this theorem of LEBESGUE, imply that the set of points where for at least one choice of $k=0,1,\dots,9$

$$\lim_{n\to\infty}\frac{\gamma_k(n)}{n}\neq\frac{1}{10},$$

(that is, the set of numbers non-normal to the base 10) is of measure zero.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

A comparison between the two papers, FABER's dated 1910 and RADEMACHER's 1918, shows the extent to which the equivalence between LEBESGUE measure on [0, 1] and geometric probability on [0, 1] had shifted its status from conjectural to a virtually unanimous convention.

When FABER, whose main aim was in a somewhat different direction, achieved almost inadvertently the result that almost all numbers are normal (the "Strong Law of Large Numbers"), he paused to interpolate some comments; these serve as a primary source for examining the extent to which BOREL's paper left its readers unsure as to the relation between probability theory and measure theory[1]:

> *The set of points for which* $\overline{\lim}_{n\to\infty} \frac{\gamma_n}{\mu_n} \neq 1$ or $\underline{\lim}_{n\to\infty} \frac{\gamma_n}{\mu_n} \neq 1$, is of measure zero.
>
> This theorem appears interesting to me from many points of view.
>
> First it gives a simple example of a set *which is not only everywhere dense but also has the cardinality of the continuum in every interval, however small, and nonetheless has measure zero.*
>
> Borel recently proved, after formulating suitable definitions concerning denumerable probabilities, that the probability that a point belong to the above set is zero. The comparison of the above theorem with Borel's result suggests the question:
>
> Is the probability – according to the Borel set-up which possibly might need to be extended to answer this question – that a number belongs to a prescribed set of zero measure, always equal to zero? And conversely: is a set always of measure zero, if the probability that a point belongs to it is equal to zero? (Italics in the original.) FABER (1910: 400).

The passage just cited is evidence that, for FABER, the clear-cut identification of geometric probability on [0, 1] with LEBESGUE measure had not quite taken place by 1910. BOREL's own suggestion of 1905 that such geometric probability be identified with LEBESGUE measure seems to have been unknown to FABER. (Recall that in 1909 this suggestion, reduced to a mere aside, was dismissed by BOREL himself.) It is clear that FABER, on the heel's of BOREL's paper of 1909, was grappling with the same nascent identification, motivated by the "accidental" alternative approach he found to BOREL's Strong Law.

As quotations from BOREL have shown, all that was needed to answer FABER's question affirmatively, and to make this identification absolutely explicit, was a careful re-examination of BOREL's own paper with emphasis on the duality between the "point de vue logique" and the "point de vue géométrique".

Neither FABER nor BOREL can be realistically charged with obtuseness; BOREL achieved his result as an application of an abstract theorem on independent trials, FABER by considering functions of a real variable defined on [0, 1]. The situation may be described by saying that BOREL, perferring the "point de vue logique," was exploring the product measure available on a certain product space, and FABER was exploring the considerably more familiar territory of measurable functions defined on [0, 1].

[1] In FABER's notation $\gamma_n(\mu_n)$ is the number of 1's (0's) in the first n terms of the binary expansion of the given real number.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

The question of the mapping between the product space implicit in BOREL and the familiar measure structure of [0, 1] was no doubt of lesser interest to BOREL and FABER than their primary but distinct aims. Although it is difficult for a contemporary reader to read BOREL's paper of 1909 without employing the advantages supplied by hindsight, we take FABER's above remark as further evidence confirming our assertion that BOREL failed to see, and certainly to state, the identification in 1909. Indeed, we think WINTNER was being somewhat over-generous in his assessment of BOREL when he wrote:

> Historically, the whole development was initiated by Borel's formulation and proof of his "either 0 or 1" theorem (Rend. Palermo, vol. 29 (1909), pp. 247–271). Today it is easy, but at that time it was a true mathematical achievement, to think of the ordinary measure on the interval $0 \leqq x \leqq 1$ as a product measure in an infinite product space, the factors being the binary spaces corresponding to the dyadic expansion of x. WINTNER (1941: 182).

9.3. Rademacher and Hausdorff: The Evidence for the Evolution of a Point of View

Eight years after the publication of FABER (1910), RADEMACHER, in ignorance of FABER's result, proved it anew by a virtually identical comstruction of a monotone function whose non-differentiability was assured at the non-normal numbers. In the interim, HAUSDORFF's *Grundzüge der Mengenlehre* had appeared. This volume helped disseminate many concepts which were becoming part of the mathematical "culture" of the time, for example such concepts as fields of sets, BOREL sets, lim inf and lim sup of sets.

In particular, HAUSDORFF gave a brief introduction to the LEBESGUE theory of measure and of the LEBESGUE integral. As *examples* of measure theory HAUSDORFF presented both the BOREL result on decimals and a closely related continued fraction result, but with proofs in no way resting on BOREL's investigation of 1909 of the probability of infinitely many successes in a denumerable sequence of trials or its extension in 1912.

In contrast with the original treatment of BOREL, both the Strong Law and the result on continued fractions are treated solely as examples of the theory of LEBESGUE measure. HAUSDORFF's treatment of these theorems will be given in the next section. What is needed here, as the necessary historical setting of RADEMACHER's paper 1918, is the precise language of HAUSDORFF on the relation between measure theory and probabilistic terminology, and his restatement of BOREL's Strong Law.[1]

> We remark that many theorems concerning the measure of point-sets appear perhaps more intuitively, if one expresses them in the language of probability. If two sets P and M are measurable, and M in particular is of positive measure, then one can define, by means of the quotient $f(P)/f(M)$ if $P \subseteq M$, or more generally by $f(P \cap M)/f(M)$, the probability that a point of M belongs to P. If we consider only subsets of a fixed set M, of measure 1, then $f(P)$

[1] We have taken the liberty of altering HAUSDORFF's original notation for union and intersection, $\mathfrak{S}(P_1, P_2)$ and $\mathfrak{D}(P_1, P_2)$ respectively.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

$= p$ is the probability that a point belongs to the set P. If $P = P_1 \cup P_2$, $P' = P_1 \cap P_2$, then $p + p' = p_1 + p_2$; p is the probability that a point belongs to P_1 or P_2, p' is the probability that a point belongs to P_1 as well as to P_2. If P_1 and P_2 have no points in common, then $p = p_1 + p_2$. Furthermore $p' = p_1 \left(\frac{p'}{p_1}\right)$ (if $p_1 > 0$); the probability that a point belongs simultaneously to P_1 and P_2 is the probability that it belongs to P_1 multiplied by the probability that a point of P_1 should belong to P_2. The formula for so-called "independent" events $p' = p_1 p_2$ is of course not valid in general. It is also perfectly clear that from this (by and large arbitrary) definition, probability 0 is not the expression of impossibility, and probability 1 not that of certainty; for 0 is the probability that a point belongs to a set of zero measure (which might still be of the cardinality of the continuum). HAUSDORFF (1914: 416–417)

This citation should leave no doubt as to how early FABER's question of 1910 was answered, and how early a scrupulously clear expression of probability was published – including independence, countable additivity, conditional probability, and even the notion of sample space. HAUSDORFF's remarks, prefatory for his explicit re-working of BOREL's results, are in the same vein as BOREL's paper of 1905, but they reach considerably further.

Two pages later, HAUSDORFF states BOREL's Strong Law as follows:

II. *Sets of Dyadic Decimals.* We consider an irrational number x between 0 and 1 and expand it in a dyadic decimal [footnote omitted]:

$$x = \frac{x_1}{2} + \frac{x_2}{2^2} + \frac{x_3}{3^3} + \cdots = (x_1, x_2, x_3, \ldots) \qquad (x_n = 0, 1).$$

Among the first n digits will appear p zeros and $q = n - p$ ones. Then one has the Theorem (E. Borel):

The set of x for which $\lim \frac{p}{q} = \frac{1}{2}$ has measure 1.

Or: the complement, i.e., the set of those x, for which $\frac{p}{n}$ is either non-convergent, or does not converge to $\frac{1}{2}$, has measure 0. There is thus a probability 1 that the dyadic expansion of x has asymptotically as many zeros as ones.

This theorem is remarkable. On the one hand it seems a plausible extension of the "Law of Large Numbers" to the infinite; on the other it asserts the existence of a limit of a sequence, and indeed even a prescribed value of the limit, a very special circumstance, which one would have held *a priori* to be exceedingly unlikely. HAUSDORFF (1914: 419–420).

We may thus state that for HAUSDORFF the notion of probability is emphatically seen as a derived one, resting on the more fundamental one of measure. The subsequent proof does not refer to probabilities for notation, intuition or method, but employed solely the lengths of intervals of covering sets (see below). The theorem served HAUSDORFF only as an example, although a very interesting one, of the use of countable additivity of LEBESGUE measure.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

This is the background of RADEMACHER's paper. The Strong Law can be viewed as an assertion that a certain set has measure 0; what RADEMACHER presented is also in this setting, but instead of a direct calculation of small covering sets, in the manner of HAUSDORFF, the powerful LEBESGUE theorem (that every monotone function is almost everywhere differentiable) is employed. This complete acceptance of the two-fold character of the theorem, probabilistic in the hands of BOREL, measure-theoretic in the hands of HAUSDORFF [1] is evident in the wording of RADEMACHER's introduction, which follows.

> One owes the following remarkable theorem to Mr. E. Borel [footnote omitted]:
>
> Denoting by n_γ ($\gamma = 0, 1, 2, \ldots, 9$) the number of digits γ among the first n places in the decimal expansion of a number from the interval between 0 and 1, then
>
> $$\lim_{n \to \infty} \frac{n_\gamma}{n} = \frac{1}{10}$$
>
> except for a null-set Z, which however has the cardinality of the continuum [footnote omitted].
>
> Borel expresses the theorem thus: the probability that, in a decimal expansion, the ten digits do not asymptotically equally occur, is zero, and carries out the proof by an extension of Bernoulli's theorem on probability. A direct proof of the theorem in our formulation was given by Hausdorff, in which he gave an upper estimate of the set Z by means of coverings with intervals. The following proof is perhaps not superfluous, ... RADEMACHER (1918: 306).

To us, RADEMACHER's assertion that BOREL carried out the proof by "an extension of BERNOULLI's theorem" seems rather misleading. Indeed, the theorem is not a consequence of the Zero-One Law of 1909, and the gap is not readily filled (even by the extension of it in 1912), without recourse to countable sub-additivity in the "CANTELLI" manner as was done by HAUSDORFF.

As we have seen, even by 1912, the date of his reply to BERNSTEIN, BOREL does not seem to have been advised by LEBESGUE, BERNSTEIN, or anyone else that his Strong Law of Large Numbers does not follow from his Zero-One Law of 1909. Writing in 1912, when he evolved a more general Zero-One Law to cover the result on continued fractions, BOREL made no attempt to legitimate his Strong Law similarly. FABER's proof in 1910, incontestable as it was, must have seemed an "accident", since it bore no evident relation to probability (*cf.* FABER's own puzzlement remarked above).

This brings the narrative up to HAUSDORFF, who in 1914 simultaneously accomplished several remarkable things: he considers both the decimal and continued fraction case in turn, recasting them both as theorems in measure theory. He proves the decimal case by sub-additivity (*i.e.*, in the "CANTELLI" manner), introducing a prowerful device to evade an appeal to the Central Limit Theorem. He also proves a continued fraction result akin to the BERNSTEIN-BOREL one, as clearly as did BERNSTEIN, although without introducing the modern language of conditional probabilities.

---

[1] And also FABER, although unknown at the time to RADEMACHER and cited by HAUSDORFF.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

His proofs are all in the language of (LEBESGUE) measure, with modern notations, such as lim sup E_n for a sequence $\{E_n\}$ of sets. As to the relation of his (measure-theoretic) assertions to those of BOREL, which were couched in the language of probability, HAUSDORFF explicitly disposes of this in the prefatory remaks cited above.

10. Hausdorff's "Grundzüge der Mengenlehre": A Notable Advance in Technique

10.1. General Background

This section will cite virtually all the probabilistic references to be found in HAUSDORFF's *Grundzüge der Mengenlehre.* The first probabilistic reference, already cited, concerns the relation between the vocabulary of probability and that of measure theory. In this section we discuss HAUSDORFF's treatment of dyadic expansions and continued fractions via measure theory.

Chapter X of the *Grundzüge* (1914) is the locus of all of the above items. This chapter is titled "Content of Point-Sets." Its first paragraph begins with generalities concerning length, measure and area in EUCLIDEAN space, and a historical sketch. The older theory of content associated with CANTOR, HANKEL, PEANO, and JORDAN, concerned itself with finite additivity for disjoint sets; the newer theory, due to BOREL and LEBESGUE, specifically added countable additivity. HAUSDORFF cited LEBESGUE's problem and formulated it thus:

> ... Lebesgue formulated the problem of associating to every bounded set A in n-dimensional space E_n, as "content," a number $f(A) \geqq 0$ satisfying the following conditions:
> (α) Congruent sets have the same measure.
> (β) The unit cube has content 1.
> (γ) $f(A+B)=f(A)+f(B)$.
> (δ) $f(A+B+C+\cdots)=f(A)+f(B)+f(C)+\cdots$ for a bounded sum of countably many sets. HAUSDORFF (1914: 401).

HAUSDORFF immediately gave an example showing the impossibility of this problem in full generality. The example, still the most popular one (*cf.* ROYDEN (1965: 52–55), maps the line on to the circle; on the circle it exhibits a set which is disjoint from all its images under rational rotations. Further, this set and its (necessarily congruent) images under all rational rotations is a disjoint decomposition of the entire circle.

It is of interest to note that this example is not attributed by HAUSDORFF to any author. However, in the references near the end of the book two sources of examples of non-measurable sets are given: G. VITALI's *Sul Problema della Misura dei Gruppi di Punti di una Retta*, and A. SCHOENFLIES' *Mengenlehre*. The latter contains several examples of non-measurable sets, including the one given by HAUSDORFF. The text of SCHOENFLIES (SCHOENFLIES (1913: 377, especially footnote 2)) indicates interestingly that this simple example was in fact due to HAUSDORFF himself, and communicated directly.

HAUSDORFF further commented on the impossibility of the "easy LEBESGUE problem" (employing the terminology of NATANSON (1961: 80)) in which

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

requirement δ is dropped in spaces of three or more dimensions. The example which establishes this is given in an appendix; it also is due to HAUSDORFF, who again omits any claim of authorship or priority.

The second paragraph of HAUSDORFF (1914), Chapter X, is concerned with the theory of PEANO-JORDAN content (*cf.* HAWKINS (1970: 86–96) for a history of these developments).

Paragraph three, about nine pages in length, is a concise and complete treatment of the theory of LEBESGUE measure in the plane. Countable additivity in all of its forms is established. Countable sub-additivity is rigorously deduced (HAUSDORFF (1914: 414: formula (6)).

The next, or fourth paragraph of HAUSDORFF's Chapter X, is devoted to applications and examples. Four topics are discussed. The first concerns non-measurable sets, the fourth concerns sequences of measurable functions and their convergence properties, such as almost uniform convergence. It is topics two and three that represent HAUSDORFF's direct contribution to the development of the ideas contained in BOREL's paper of 1909. It is noteworthy that in a book of 473 pages, only nine of which are devoted to the theory of measure, two of the four topics chosen to illustrate measure theory stem from BOREL's paper of 1909. It is especially to be noted that HAUSDORFF presented these two topics, the Strong Law and continued fractions, in a completely original framework, in which he offered incontestable proofs. HAUSDORFF thereby emphasized, the clarity and scope of the methods of measure theory once the fundamental properties of measure have been clearly laid down.

10.2. Hausdorff's Proof of the Strong Law: The Use of Moments and Bienaymé-Tchebycheff Type Inequalities

HAUSDORFF's formulation of the BOREL Strong Law has been cited above. We now sketch HAUSDORFF's proof, differing only in minor notational modifications from the original.

To begin with, the set of all irrational x whose dyadic expansions begin with the binary digits $b_1, b_2, \dots, b_n$, consists of the irrational numbers in the interval

$$\frac{b_1}{2}+\frac{b_2}{2^2}+\cdots+\frac{b_n}{2^n}<x<\frac{b_1}{2}+\frac{b_2}{2^2}+\cdots+\frac{b_n+1}{2^n}$$

and is therefore of measure $1/2^n$. (The omission of rational numbers, a set of measure zero, permits this identification of measure with length.) The set of those (irrational) x which have precisely k zeros and $n-k$ ones in the first n terms consists of $\binom{n}{k}$ pairwise disjoint "intervals," all of measure $1/2^n$, and so is of total measure

$$\binom{n}{k}\frac{1}{2^n}.$$

Further, let ε be a positive number and let $E_n(\varepsilon)$ be the set of those x for which

$$\left|\frac{k}{n}-\frac{1}{2}\right|\geqq\varepsilon.$$

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

The measure of this set, denoted $p_n(\varepsilon)$, is then given by

$$p_n(\varepsilon)=\sum^{*}\binom{n}{k}\frac{1}{2^n}$$

where the sum $\sum^{*}$ is taken over those values of k for which

$$\left|\frac{k}{n}-\frac{1}{2}\right|\geqq\varepsilon.$$

At this point HAUSDORFF introduces a technical device destined to influence successors (such as STEINHAUS) which involves estimating even-order moments of the random variable $\frac{v}{n}-\frac{1}{2}$. By means of this device (*cf.* § 10.3 below) HAUSDORFF succeeded in showing that

$$\sum_{k=0}^{n}\left(\frac{k}{n}-\frac{1}{2}\right)^4\binom{n}{k}\frac{1}{2^n}<\frac{3}{16}\frac{1}{n^2}. \qquad (10.1)$$

(In terms of the notation $v_n(x)$ for the number of zeros among $b_1, \ldots, b_n$ in the dyadic expansion of x, the sum appearing on the left-hand side of the above inequality would now be called the 4th moment of the random variable $\frac{v_n(x)}{n}-\frac{1}{2}$).

It follows that

$$\varepsilon^4\sum^{*}\binom{n}{k}\frac{1}{2^n}<\sum^{*}\left(\frac{k}{n}-\frac{1}{2}\right)^4\binom{n}{k}\frac{1}{2^n}<\frac{3}{16}\frac{1}{n^2}$$

i.e.,

$$p_n(\varepsilon)<\frac{3}{16}\frac{1}{\varepsilon^4}\frac{1}{n^2},$$

so that $\sum_1^\infty p_n(\varepsilon)$ converges for every $\varepsilon>0$.

The strong Law now follows with remarkable rapidity[1]. It must be emphasized that the inequality (10.1) plays the role for HAUSDORFF which estimates from the Central Limit Theorem had played for BOREL.

Let $E(\varepsilon)=\limsup E_n(\varepsilon)=\bigcap_{n=1}^{\infty}\bigcup_{n=N}^{\infty}E_n(\varepsilon)$. Then $E(\varepsilon)$ is exactly the set of irrational x in the unit interval for which

$$\overline{\lim_{n\to\infty}}\left|\frac{v_n(x)}{n}-\frac{1}{2}\right|\geqq\varepsilon.$$

[1] This inequality is of the sort probabilists call TCHEBYCHEFF (or BIENAYMÉ-TCHEBYCHEFF) type, although based on 4th powers rather than 2nd powers. It lies at the heart of the brilliantly brief proof given by KAC (1959: 16–17).

Fourth power estimates were also used by F. CANTELLI in his proof (CANTELLI (1917a) and (1917b)).

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

Let $A_\infty(\varepsilon)=P(E(\varepsilon))=$ measure of $E(\varepsilon)$. The sets $\bigcup_{n=N}^{\infty} E_n(\varepsilon)$ decrease with increasing N. Obviously

$$E(\varepsilon)\subset \bigcup_{n=N}^{\infty} E_n(\varepsilon) \qquad \text{for } N=1,2,3,\ldots.$$

For each fixed integer N we have, by monotonicity and countable sub-additivity, that

$$A_\infty(\varepsilon)\leqq P\left(\bigcup_{n=N}^{\infty} E_n(\varepsilon)\right)\leqq \sum_{n=N}^{\infty} P(E_n(\varepsilon))=\sum_{n=N}^{\infty} p_n(\varepsilon).$$

Having shown that $\sum_{n=N}^{\infty} p_n(\varepsilon)$ converges for every $\varepsilon>0$, the measure $E(\varepsilon)=A_\infty(\varepsilon)$ is 0 for every $\varepsilon>0$.[1]

In particular $\bigcup_{l=1}^{\infty} E\left(\frac{1}{l}\right)$ is a countable union of sets of measure 0, and by countable additivity, is itself of measure 0. But $\bigcup_{l=1}^{\infty} E\left(\frac{1}{l}\right)$ is the set of x for which $\overline{\lim_{n\to\infty}}\left|\frac{v_n(x)}{n}-\frac{1}{2}\right|>0$. Thus if $\sum p_n(\varepsilon)$ converges for all $\varepsilon>0$, there is probability (*i.e.*, measure) 1 that:

$$\overline{\lim_{n\to\infty}}\left|\frac{v_n(x)}{n}-\frac{1}{2}\right|=0$$

or equivalently

$$\lim_{n\to\infty}\left|\frac{v_n(x)}{n}-\frac{1}{2}\right|=0.$$

At the very least, HAUSDORFF gave the first (*cf.* preceding footnote) "CANTELLI"-type proof that $\sum p_n<\infty$ implies $A_\infty=0$. It seems clear that HAUSDORFF, not CANTELLI, should be given credit for this line of reasoning. Not only did HAUSDORFF anticipate CANTELLI by three years, he also gives a treatment

---

[1] This same argument shows (suppressing the ε), that if $\sum p_n<\infty$, where $p_n=P(E_n)$, then $P(\limsup E_n)=A_\infty=0$ thanks to subadditivity and with no appeal to independence. In other words, this is CANTELLI's argument, but resting on the firm base of measure theory. CANTELLI had rested the corresponding argument on the *ad hoc* assumption that sub-additivity ("BOOLE's inequality") could be extended from the finite to the countably infinite case.

As early as 1906, in his thesis, FRÉCHET proved a theorem to which the "CANTELLI" result is an immediate corollary, but couched in the language of measure (*cf.* FRÉCHET (1906: 16)). If for each integer n, E_n is a measurable subset of the unit interval whose measure $m(E_i)$ equals $1-m_i$, then FRÉCHET observed that $m\left(\bigcap_{n=1}^{\infty} E_n\right)>1-(m_1+m_2+\cdots)$. It is immediate (although it seems to have passed unremarked in the literature) that if $\sum m_n$ converges (the only case of interest) then $m\left(\bigcap_{n=N}^{\infty} E_n\right)>1-\varepsilon$ for N sufficiently large, depending on ε. This yields $m(\liminf E_n)=1$ at once without the intervention of any "independence" assumption. Since $m_n=m(E_n^c)$, this can be rephrased as follows: $\sum m(E_n^c)$ converges implies $m(\liminf E_n^c)=0$, precisely the "CANTELLI" result.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

which is in one respect superior: the sub-additivity

$$P\left(\bigcup_{n=N}^{\infty} E_n\right) \leqq \sum_{n=N}^{\infty} P(E_n)$$

is seen as a *consequence* of the general properties of measure, namely, non-negativity and countable additivity, and a domain of definition consisting of a σ-algebra of sets. By contrast, CANTELLI asserted

$$P\left(\bigcup_{n=N}^{\infty} E_n\right) \leqq \sum_{n=N}^{\infty} P(E_n)$$

with unspecified generality, reasoning by pure analogy with the corresponding inequality when n has a finite range. Thus CANTELLI assumes that the degree of generality is not limited to geometric probability (the only case then known for which measure theory was applicable). On the other hand, he ignores the fact that countable sub-additivity rests on the prior apparatus of countable additivity and σ-algebras. HAUSDORFF's *proof seems to be the first fully accurate re-working of* BOREL's "*proof*" *of the* BOREL *Law of Large Numbers.*

HAUSDORFF observed also that thanks to his method of moments, using (even) moments higher than the fourth, the result of BOREL can in fact be considerably strengthened: where BOREL showed that the set of x for which

$$\lim_{n\to\infty}\left(\frac{v_n(x)}{n}-\frac{1}{2}\right)=0$$

is 1, HAUSDORFF's method established that the set of x for which

$$\lim_{n\to\infty}\left(\frac{v_n(x)}{n}-\frac{1}{2}\right)n^{\theta}=0$$

is 1, for any θ less than $\frac{1}{2}$. He thus initiated a sequence of refinements of BOREL's Strong Law establishing more and more precise information as to the *rate* at which $\frac{v_n(x)}{n}$ approaches $\frac{1}{2}$ with probability 1.

10.3. Hausdorff's Method for Calculating Moments

Summarizing the main result: HAUSDORFF succeeded in establishing the BOREL Law of Large Numbers without appeal to the Central Limit Theorem. The corresponding tool for HAUSDORFF was the inequality:[1]

$$\sum_{k=0}^{n}\left(\frac{k}{n}-\frac{1}{2}\right)^4\binom{n}{k}\frac{1}{2^n}<\frac{3}{16}\frac{1}{n^2}. \qquad (10.2)$$

[1] The left-hand side of (10.2) would now be denoted

$$E\left[\left(\frac{X_1+\cdots+X_n}{n}-\frac{1}{2}\right)^4\right]$$

where X_j are independent random variables each with probability $\frac{1}{2}$ of having the value 0 and probability $\frac{1}{2}$ of having the value 1.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

HAUSDORFF showed generally how to express "moments" such as

$$\sum_{k=0}^{n} (2k-n)^l \binom{n}{k} \frac{1}{2^n}$$

as polynomials in n with integer coefficients. The even values of l are the most significant, as in the case $l=4$ cited above. These polynomials are easily bounded above by pure powers of n with a slightly larger leading coefficient. Thus HAUSDORFF shows how to obtain estimates such as

$$\sum_{k=0}^{n} (2k-n)^{2m} \binom{n}{k} \frac{1}{2^n} < A_{2m}^{n^m}$$

and hence

$$\sum_{k=0}^{n} \left(\frac{k}{n}-\frac{1}{2}\right)^{2m} \binom{n}{k} \frac{1}{2^n} < \frac{A_{2m}}{n^m} \frac{1}{2^{2m}}.$$

If $\overset{*}{\sum}$ means sum over those values of k satisfying $\left|\frac{k}{n}-\frac{1}{2}\right| \geqq \varepsilon$ there results finally

$$\varepsilon^{2m} \overset{*}{\sum} \binom{n}{k} \frac{1}{2^n} < \frac{A_{2m}}{n^m} \frac{1}{2^{2m}}.$$

In showing how these moments could be estimated and how advantageous they were, HAUSDORFF achieved a notable advance in method. The ability to make such estimates was destined to play a role in later developments both when $l=4$ and when $l=2m$ for large m. Indeed, CANTELLI was to make use, three years later, of similar techniques involving the 4^{th}moment, *i.e.*, $l=4$, and STEINHAUS considered $l=2m$ for large values of m in 1923.

To conclude these remarks, we sketch HAUSDORFF's method of obtaining exact polynomial expressions for

$$\sum_{k=0}^{n} (2k-n)^{2m} \binom{n}{k} \frac{1}{2^n}.$$

HAUSDORFF first considered

$$\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = (x+y)^n$$

and introduced the change of variables $u=x+y$, $v=x-y$. Then the differential operator $D = x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y}$ satisfies

$$\begin{aligned} D(f+g) &= Df + Dg, \\ D(fg) &= (Df)\,g + f(Dg), \\ D(f^n) &= nf^{n-1}(Df), \\ D(x^k y^{n-k}) &= (2k-n)\, x^k y^{n-k}. \end{aligned}$$

In particular, $D(u)=v$, $D(v)=u$.

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

The calculation of $D^{2m}(u^n)$ then presents no difficulty, and the numerical choice $x=y=\frac{1}{2}$, which corresponds to $u=1$, $v=0$, results in

$$D^{2m}(u^n)\big]_{\substack{u=1\\v=0}} = \sum_{k=0}^{n} (2k-n)^{2m} \binom{n}{k} \frac{1}{2^n},$$

an m^{th} degree polynomial in n.

10.4. Hausdorff's Continued Fraction Theorem

Let us introduce once more the notation $[a_1=m_1, a_2=m_2, \ldots, a_n=m_n]$ to stand for the set of continued fractions whose first n elements $a_1, \ldots, a_n$ are the prescribed integers $m_1, \ldots, m_n$. These form an interval, the length of which we denote by

$$P[a_1=m_1, \ldots, a_n=m_n].$$

The set defined by

$$[a_1=m_1, \ldots, a_{n-1}=m_{n-1}, k \leqq a_n \leqq m]$$

is also an interval (being a union of adjacent intervals of the above type), the length is denoted by

$$P[a_1=m_1, \ldots, a_{n-1}=m_{n-1}, k \leqq a_n \leqq m].$$

The ratio of the above to the length $P[a_1=m_1, \ldots, a_{n-1}=m_{n-1}]$ can be denoted by

$$P[k \leqq a_n \leqq m \mid a_1=m_1, \ldots, a_{n-1}=m_{n-1}]$$

using the modern notation for conditional probability although this notation is not strictly necessary (and indeed is not used by HAUSDORFF). HAUSDORFF's point of departure is an inequality

$$\rho(k, m) \leqq P[k \leqq a_n \leqq m \mid a_1=m_1, \ldots, a_{n-1}=m_{n-1}] \leqq \sigma(k, m)$$

where ρ and σ can be calculated explicitly, and do not depend on n, nor on $m_1, \ldots, m_{n-1}$. HAUSDORFF calculates "best possible" expressions for ρ and σ, including their somewhat altered form if $k=1$ or $m=\infty$ (but not both). It follows readily that

$$\rho(k_n, m_n) \leqq P[k_n \leqq a_n \leqq m_n \mid k_1 \leqq a_1 \leqq m_1, \ldots, k_{n-1} \leqq a_{n-1} \leqq m_{n-1}] \leqq \sigma(k_n, m_n)$$

by purely algebraic manipulations. We have described this manipulation in discussing BERNSTEIN's work; it is essentially identical in HAUSDORFF's.

By multiplication of these inequalities for consecutive values of n (which can be interpreted as the Chain Law of Probability) it follows that

$$\rho_1 \ldots \rho_n \leqq P[k_1 \leqq a_1 \leqq m_1, \ldots, k_n \leqq a_n \leqq m_n] \leqq \sigma_1 \ldots \sigma_n$$

where we have introduced

$$\rho_j = \rho(k_j, m_j), \qquad \sigma_j = \sigma(k_j, m_j)$$

for brevity.

By countable additivity (explicitly so stated)

$$\prod_1^\infty \rho_j \leqq P[k_j \leqq a_j \leqq m_j, j=1, 2, 3, \ldots] \leqq \prod_1^\infty \sigma_j.$$

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

HAUSDORFF is interested only in conditions on the two numerical sequences $\{k_n\}$ and $\{m_n\}$ that assure *positive* measure to the set $[k_j \leqq a_j \leqq m_j, j=1, 2, 3, \ldots]$. For this it is clearly necessary that $\prod_1^\infty \sigma_j > 0$ and sufficient that $\prod_1^\infty \rho_j > 0$. In view of the specific form of $\rho(k, m)$, $\sigma(k, m)$ it results that positive measure is obtained if and only if $k_j = 1$ except for finitely many values of j and

$$\sum_{j=1}^{\infty} \frac{1}{m_j} \text{ converges.}$$

HAUSDORFF did not pursue his calculations further, as had BERNSTEIN and BOREL, to observe that the set defined by $[k_n \leqq a_n \leqq m_n$ for all but finitely many values of $n = 1, 2, 3, \ldots]$ is necessarily of measure 1 when it is positive (although the result is immediate using his techniques). Recall, it was this result which so appealed to BOREL that he had pointed it out as the most interesting in his entire paper of 1909.

In summary, HAUSDORFF used continued fractions as an example of the power of LEBESGUE measure, especially its countable additivity and wide domain of definition. He obtained a result along the lines of the BERNSTEIN-BOREL one. Like BERNSTEIN he did not mistakenly assume or employ independence, but rather manipulated the probabilities of mutually dependent events in a fully accurate manner. Like BERNSTEIN, HAUSDORFF obtained the probabilities of non-cylinder sets by valid limit operations on the probability of approximating cylinder sets.

None of BOREL's early successors followed his lead towards the rigorous fundation of a theory of probability concerned with repeated trials (independent or not). Not until someone who possessed facility with axiomatically based measure theory and who shared his primary concern with probability (in particular with repeated trials) was BOREL to have a successor as important as himself, a successor who would fully deserve the accolade which WINTNER bestowed on BOREL. That was to wait until 1923; the successor was to be HUGO STEINHAUS.

Bibliography

BARONE, J., 1974. *An Historical Analysis of the Development of Axiomatic Probability Theory.* Doctoral Dissertation. New York University.

BERNSTEIN, F., 1911. "Über eine Anwendung der Mengenlehre auf ein aus der Theorie der säkularen Störungen herrührendes Problem," *Mathematische Annalen*, **71**, 417–439.

BOHL, P., 1909. "Über ein in der Theorie der säkularen Störungen vorkommendes Problem," *Crelle's Journal*, **135**, 189–283.

BOREL, E., 1895. "Sur quelques points de la théorie des fonctions," *Annales de l'Éçole Normale*, **3** (12), 9–55.

BOREL, E., 1898. *Leçons sur la Théorie des Fonctions*, Paris.

BOREL, E., 1903. "Contribution à l'analyse arithmétique du continu," *Journal de Mathématiques Pures et Appliquées*, **9**, 329–375.

BOREL, E., 1905. "Remarques sur certaines questions de probabilité," *Bulletin de la Société Mathématique de France*, **33**, 123–128.

BOREL, E., 1909. "Les probabilités dénombrables et leurs applications arithmétiques," *Rendiconti del Circolo Matematico di Palermo*, **27**, 247–271.

BOREL, E., 1909a. *Eléments de la Théorie des Probabilités*, Paris.

BOREL, E., 1912. "Sur un problème de probabilités relatif aux fractions continues," *Mathematische Annalen*, **72**, 578–584.

This content downloaded from 128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms

BOREL, E., 1914. *Leçons sur la Théorie des Fonctions.* (2nd ed.), Paris.
BOREL, E., 1926. *Applications à l'Arithmétique et à la Théorie des Fonctions*, Paris.
BRODEN, T., 1900. "Wahrscheinlichkeitsbestimmungen bei der gewöhnlichen Kettenbruchentwicklung reeller Zahlen," *Öfversigt af Kongl. Svenska Vetenskaps-Akademiens Förhandlingar*, **2**, 239–266.
CANTELLI, F., 1917a. "Sulla probabilità come limite della frequenza," *Accademia dei Lincei Roma. Classe di Scienze Fisiche, Mathematiche e Naturali. Rendiconti*, **26** (5), 39–45.
CANTELLI, F., 1917b. "Su due applicazioni di un teorema di G. Boole alla statistica matematica," *Accademia dei Lincei Roma. Classe di Scienze Fisiche, Matematiche e Naturali, Rendiconti*, **26** (5), 295–302.
CARÁTHÉODORY, C., 1914. "Über das lineare Maß von Punktmengen – eine Verallgemeinerung des Längenbegriffs, "*Nachrichten der Akademie der Wissenschaften zu Göttingen. II. Mathematisch-Physikalische Klasse*, **4**, 404–426.
DIEUDONNÉ, J., 1975. "Introductory remarks on algebra, topology and analysis," *Historia Mathematica*, **2**, 537–548.
FRÉCHET, M., 1906. "Sur quelques points du calcul fonctionnel." *Rendiconti del Circolo Matematico di Palermo*, **22**, 1–74.
FRÉCHET, M., 1915. "Définition de l'Intégrale sur un ensemble abstrait," *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, June 28th, 839–840.
FRÉCHET, M., 1930. "Sur la convergence en probabilité," *Metron*, **8**, 1–48.
HARTMAN, S., 1948. "Sur deux notions de fonctions indépendantes," *Colloquium Mathematicum*, **1**, 19–22.
HAUSDORFF, F., 1914. *Grundzüge der Mengenlehre*, Leipzig.
HAWKINS, T., 1970. *Lebesgue's Theory of Integration: Its Origins and Development*, Madison, Wisconsin.
HILBERT, D., 1900. "Sur les problèmes futurs des mathématiques," *Comptes Rendus du Deuxième Congrès International des Mathématiciens*, Paris, 58–114.
KAC, M., 1959. *Statistical Independence in Probability, Analysis and Number Theory*, New York.
LAKATOS, I., 1963–64. "Proofs and refutations," *British Journal for Philosophy of Science*, **14**, 1–25, 120–139, 221–243, 296–342.
LEBESGUE, H., 1906. *Leçons sur les Séries Trigonométriques*, Paris.
NATANSON, I., 1961. *Theory of Functions of a Real Variable*, Volume I (revised ed.), New York.
NOVIKOFF, A., & J. BARONE, 1977. "The Borel Law of Normal Numbers, The Borel Zero-One Law, and the Work of Van Vleck," *Historia Mathematica*, **4**, 43–65.
POINCARÉ, H., 1912. *Calcul des Probabilités* (2nd ed.), New York.
RADÓN, J., 1913. "Theorie und Anwendungen der absolut additiven Mengenfunktionen," *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse*, **122** (2), 1295–1438.
ROYDEN, H., 1965. *Real Analysis*, New York.
SCHOENFLIES, A., 1913. *Entwicklung der Mengenlehre und ihrer Anwendungen*, Leipzig.
SIERPIŃSKI, W., 1918. "Sur une définition axiomatique des ensembles mesurables (L)," *Bulletin de l'Académie des Sciences de Cracovie*, July, 173–178.
SLUTSKY, E., 1925. "Über stochastisch Asymptoten und Grenzwerte," *Metron*, **5**, 3–89.
STEINHAUS, H., 1923. "Les probabilités dénombrables at leur rapport à la théorie de la mesure," *Polska Akademia Nauk. Fundamenta Mathematicae*, **4**, 286–310.
VAN VLECK, E., 1908. "On non-measurable sets of points, with an examples," *Transactions of the American Mathematical Society*, **9**, 237–244.
VITALI, G., 1905. *Sul Problema della Misura dei Gruppi de Punti di una Retta*, Bologna.
WIMAN, A., 1900. "Ueber eine Wahrscheinlichkeitsaufgabe bei Kettenbruchentwicklungen," *Öfversigt af Kongl. Svenska Vetenskaps-Akademiens Förhandlingar*, **7**, 829–841.
WIMAN, A., 1901. "Bemerkungen über eine von Gyldén aufgeworfene Wahrscheinlichkeitsfrage, *Acta Universitatis Lundensis*, 1–19.
WINTNER, A., 1941. *The Analytical Foundations of Celestial Mechanics*, Princeton, New Jersey.
WINTNER, A., 1947. *The Fourier Transform of Probability Distributions*, Ann Arbor, Michigan.

Department of Mathematics
New York University

(Received August 30, 1976)

This content downloaded from
128.210.126.199 on Wed, 29 Dec 2021 15:17:25 UTC
All use subject to https://about.jstor.org/terms