

PROBABILITY AS EXPRESSED BY ASYMPTOTIC
LIMITS OF PENCILS OF SEQUENCES*

BY E. L. DODD

1. *Introduction.* From time to time certain conceptions and theorems of mathematics need a restatement. Especially true is this in a subject like probability throughout which is woven so much non-mathematical material. This paper, then, is an attempt to depict probabilities, initially discrete, by the use of aggregates of sequences, in language strictly mathematical, but as simple as possible. The treatment will resemble not so much that of von Mises, † who uses single sequences, as that of Borel, ‡ Lomnicki, § and Steinhaus. || But it will differ from the latter two in avoiding the concept of measure, and indeed from all in the approach to the subject. The asymptotic limit introduced in this paper involves a notion resembling that involved in the phrase *almost everywhere*.

In pure mathematics, the word *probability* may be taken ¶ to signify simply the ratio of the number of objects in a subset to the number in the set, so long as discrete or arithmetic probability is being considered. It is, indeed, as far outside the field of mathematics to determine whether two events are equally likely as to determine whether two bodies have the same mass. Even in the applications, the role of pure mathematics is merely to count expeditiously the elements of sets and subsets, or more generally to determine certain measures of sets, which are believed by competent judges to depict adequately situations in the external world. It is generally believed, for example,

* Presented to the Society, August 27, 1929.

† *Grundlagen der Wahrscheinlichkeitsrechnung*, Mathematische Zeitschrift, vol. 5 (1919), pp. 52–99; see especially pp. 55–57.

‡ *Les probabilités dénombrables et leurs applications arithmétiques*, Rendiconti di Palermo, vol. 27 (1909), pp. 247–271.

§ *Nouveaux fondements du calcul des probabilités*, Fundamenta Mathematicae, vol. 4 (1923), pp. 34–71.

|| *Les probabilités dénombrables et leur rapport à la théorie de la mesure*, Fundamenta Mathematicae, vol. 4 (1923), pp. 286–310.

¶ See Lomnicki, loc. cit., p. 37.

that two tosses of an ordinary coin can be described adequately for most purposes by writing HH , HT , TH , TT . Granting this, the verdict of mathematics is that $2/4$ is the probability of one head and one tail. On the other hand, if a certain misshapen coin is thought to *favor heads two to one*, the following nine two-element sequences can be written: HH , HH , HT ; HH , HH , HT ; TH , TH , TT , and the probability of one head and one tail on this hypothesis is $4/9$. A supplementary exhibit would commonly appear in columnar form, the first column with two H 's and one T , each such element followed by a brace enclosing likewise two H 's and one T . The nature of the problem is unchanged if the coin is declared to favor heads 501 to 499, though the writing of the sequences becomes more laborious.

In trying to *symbolize* the *ordinary coin*, we note that with two different elements, say H and T , there can be formed exactly 2^s *different* sequences with s elements each. Let m_i be the number of H 's in the i th sequence, $i=1, 2, \dots, 2^s$. Then the percentage of the sequences in which m_i/s is close to $1/2$, when s is large, is almost 100%. This, which contains the essence of Bernoulli's theorem, is a simple *mathematical fact*, independent of any meanings or implications which philosophers or scientists may attach to the word *probability*. Furthermore, it is a correct mathematical *picture* for the proposition (whether this proposition be true or false) that if you keep on tossing an *ordinary* coin, you are *almost sure* to get heads *just about* half the time. This *pictorial correctness* and the *logical rigor* of conclusions is all that mathematicians can be held responsible for; mathematicians know nothing about the truth or falsity of propositions about coins, the implements of chance, or natural statistical phenomena.

In the so-called* Bernoulli case, the elementary probability remains constant from trial to trial; in the *Poisson case*, it changes; for example, if balls are to be drawn successively from a set of urns with different proportions of white and black balls. In the Poisson case, it becomes eventually *almost certain* that the ratio of occurrences (white balls) to trials will closely approximate the *average* of the given probabilities for the occurrence (white ball) of the event on a single trial. Make the

* Rietz, *Mathematical Statistics*, pp. 27–35, in particular, p. 28.

given probabilities all identical, and the Bernoulli case is obtained.

2. *Pencils of Sequences.* Let there be n objects, not necessarily numbers, a_1, a_2, \dots, a_n , which we may think of as arranged in a column rather widely spaced, with a brace, $\{$, following each object. Associated with a_i , and placed in a column to the right of the brace, let there be n_i objects $b_{i1}, b_{i2}, \dots, b_{in_i}$, for $i=1, 2, \dots, n$. Such a set of objects we call a *brace*. Also a_1, a_2, \dots, a_n form the *first brace*. In the same way let each b_{ij} be associated with n_{ij} objects designated by c_{ijk} and so on. From these, sequences may be formed, of which

$$a_i, b_{ij}, c_{ijk}, \dots$$

is typical. The aggregate of such sequences we call a *pencil of sequences*, and each object such as $a_i, b_{ij}, c_{ijk}, \dots$ an *element*. An exhibit of decimal fractions* or dyadic fractions would form a pencil.

Frequently it is convenient to take the braces in a given column all alike. That is n_{ij} and b_{ij} do not change with i . In this case a brace characterizing a column will be called a *cross section of the pencil*. And, in contradistinction, the sequences will be called the *longitudinal sections*.

From a given pencil of sequences of objects, corresponding pencils of sequences of numbers can be constructed. Thus, if a sequence is A, A', A', A, \dots , with A perhaps signifying *success*, and A' *failure*,—we can write as a number sequence: $1, 1, 1, 2, \dots$, to count the A 's; or $1, 1/2, 1/3, 2/4, \dots$, to get the relative frequency of the A 's. In general, probability problems lead to such numerical sequences. The s th term of such a sequence is a function of s .

3. *Asymptotic Limits.* DEFINITION. For a pencil, or a finite number of pencils of sequences, and a given number s of terms in a sequence, let there be $N(s)$ functions, one for each sequence, $f_1(s), f_2(s), \dots$, which collectively will be designated $f(s)$. Given $\epsilon > 0$, small at pleasure, suppose that $m(s, \epsilon)$ of these functions satisfy

$$(1) \quad |f_i(s) - L| < \epsilon,$$

* See Borel, loc. cit., p. 258.

where L is a constant with respect to s and t , and where

$$(2) \quad \lim_{s \rightarrow \infty} \frac{m(s, \epsilon)}{N(s)} = 1.$$

We call L an asymptotic limit or an almost limit, and write

$$(3) \quad \operatorname{alim}_{s \rightarrow \infty} f(s) = L.$$

This means that with $\epsilon > 0$, $\eta > 0$, small at pleasure, it is possible to find s' so that if $s > s'$, we have $0 \leq 1 - m/N < \eta$.

A number of writers have recognized that problems of probability often lead to a relation resembling a limit; for example, when it becomes almost certain or asymptotically certain that an empirical ratio will approximate a theoretic ratio or *probability*. Cantelli* capitalizes the l of \lim , writing Lim for his near-limits.

PROPERTIES. Analogous to properties of ordinary limits are properties of these asymptotic limits. For example, if with a pencil of sequences we define two functions $f(s)$ and $g(s)$ having asymptotic limits L and L' , then

$$(4) \quad \operatorname{alim}_{s \rightarrow \infty} [f(s) + g(s)] = L + L'.$$

For if $m'(s, \epsilon)$ is the number of sequences for which $|f_i(s) - L| > \epsilon/2$, and $m''(s, \epsilon)$ is the number for which $|g_i(s) - L'| > \epsilon/2$, then the number for which $|f_i(s) + g_i(s) - (L + L')| > \epsilon$ cannot exceed $m'(s, \epsilon) + m''(s, \epsilon)$. But the ratio of this sum to $N(s)$ approaches zero as a limit. In a similar manner, we prove the theorem: If $F(x, \dots, y)$ is continuous at

$$x = L, \dots, y = L';$$

and $\operatorname{alim} f(s) = L, \dots, \operatorname{alim} g(s) = L'$, it follows that†

$$(5) \quad \operatorname{alim}_{s \rightarrow \infty} F[f(s), \dots, g(s)] = F[L, \dots, L'].$$

4. *The Bernoulli and Poisson Theorems.* Of the theorems of probability few are more important than the Bernoulli theorem

* *La tendenza ad un limite nel senso del calcolo della probabilità*, Rendiconti di Palermo, vol. 4 (1916), pp. 191–201.

† See Cantelli, loc. cit., p. 199.

with its extensions and its inverse. For the Bernoulli theorem all the *cross sections* of the pencil are *identical*; for the Poisson theorem, the cross sections are in general different. The Poisson theorem, in the language here used, is then a statement that the *average cross section* of a pencil gives a correct estimate of nearly all the *longitudinal sections*; that is, with the exception of a percentage of the latter, which evanesces with increasing s . To express this more fully, suppose that to form a pencil, n of the A 's, B 's, \dots , K 's are taken for the first column, n' of these elements for each identical brace in the second column, n'' for each identical brace in the third column, and so on. Let p_1 be the ratio of the number of A 's in the first column to n , let p_2 be the ratio of the number of A 's in each brace of the second column to n' , and so on. Let $f_i(s)$ be the number of A 's among the first s elements of the i th sequence. The Poisson theorem is then the statement that if

$$(6) \quad \lim_{s \rightarrow \infty} \frac{1}{s} [p_1 + p_2 + \dots + p_s] = p,$$

it follows that

$$(7) \quad \text{alim}_{s \rightarrow \infty} \frac{f(s)}{s} = p.$$

Not only is this true for the A 's, but the analogous relation holds for the B 's, \dots , K 's, *simultaneously*, by the addition theorem for asymptotic limits. Hence, asymptotically, that is, excepting an evanescent percentage of the sequences, the distribution of A 's, B 's, \dots , K 's in the sequences conforms to the ratios set up by the *average cross section* of the pencil. The result is unchanged if a finite number of the cross sections are altered or are *unknown*, as this would not effect the value of p . At this point there is an analogy between this *direct* form of probability, and the *indirect* or *inverse* based upon *unknown a priori* probabilities, which we consider next.

5. *The Inverse to Bernoulli's Theorem.* An urn contains b balls of which $a = pb$ are white, where a and p are unknown. In s drawings of a ball with replacement, r white balls have appeared. Then r/s may be accepted as an estimate of the unknown probability p , of getting a white ball in one drawing.

Indeed, with $\eta > 0$, small at pleasure, there is an *a posteriori* probability greater than $1 - \eta$, when s is sufficiently large, that $|r/s - p| < 1/b$, at least, if it can be assumed that there was a positive *a priori* probability, not necessarily known, for each of the $(b+1)$ possible ways of filling the urn. By means of some such schema, some precision can be given to the Bernoulli inverse. But even here are mentioned various probabilities capable of exciting philosophic perplexities. So it seems desirable to express the mathematical relations involved without using the word *probability*.

The inverse theorem, in discrete form, makes use of certain unknown *a priori* probabilities, which in the formulation which follows will be represented by proportional arbitrary* constants, N_a . The *unknown cause* is the value a/b of a probability. After r successes in s trials, the value r/s is attributed to a/b . The ϵ used is the *margin* between the discrete values considered for the probability a/b . In exact language, and thus in language unencumbered by any notions of probability, the theorem may be stated as follows.

Let ϵ and a whole number b be so chosen that

$$(8) \quad \frac{1}{b} \leq \epsilon,$$

and let $(b+1)$ whole numbers, N_0, N_1, \dots, N_b , be arbitrarily taken. For each a in turn, $a=0, 1, 2, \dots, b$, let N_a pencils of sequences be constructed, each sequence with s elements, by putting into each brace a of the A 's, and $b-a$ of the not- A 's. Let $n(r, s)$ be the number of these sequences with exactly r of the A 's among its s elements and let $m(r, s, \epsilon)$ be the number of such sequences for which

$$(9) \quad \left| \frac{r}{s} - \frac{a}{b} \right| \leq \epsilon.$$

Then

$$(10) \quad \lim_{s \rightarrow \infty} \frac{m(r, s, \epsilon)}{n(r, s)} = 1,$$

* See von Mises, *Fundamentalsätze der Wahrscheinlichkeitsrechnung*, Mathematische Zeitschrift, vol. 4 (1919), pp. 1-97, in particular, p. 83.

uniformly for r . In other words, given $\eta > 0$, small at pleasure, it is possible to find s' so that if $s > s'$, and r thereafter is given any of the values $0, 1, \dots, s$, we have

$$(11) \quad \left| \frac{m(r, s, \epsilon)}{n(r, s)} - 1 \right| < \eta.$$

This, indeed, permits r/s to have values near zero or unity where the approximation of the Laplacean integral is not so good. On account of (8), at least one integral value a satisfies (9).

If, moreover, it is understood that b is to be chosen after ϵ , the conclusion may be stated as follows: For the i th sequence which contains exactly r of the A 's, let $f_i(r, s)$ be the value of a which characterizes and initiates the sequence, and let a_{rs} be a collective symbol for $f_i(r, s)$. Then

$$(12) \quad \operatorname{alim}_{s \rightarrow \infty} \left(\frac{a_{rs}}{b} - \frac{r}{s} \right) = 0,$$

uniformly for r .

6. *Conclusion.* The foregoing is perhaps sufficient to show that, without entering into the somewhat intricate theory of measure, we may express accurately some of the most important theorems of probability in just about the usual language of mathematics, at least avoiding such notions as *cases equally likely*, a concept which seems to have significance for the external world rather than for mathematics. Of course, after the mathematical framework is understood, there is no more harm in employing the usual language of probability than in speaking of a sphere in n dimensions.

THE UNIVERSITY OF TEXAS