DISCUSSION OF PAPERS ON PROBABILITY THEORY

By R. von Mises and J. L. Doob

1. **Comments by R. von Mises.** Professor Doob outlines a new theory of probability starting with the following three basic conceptions. First, he uses the notion of an infinite sequence of trials or better: of an infinite sequence of numbers $x_1, x_2, x_3, \cdots$ which can be considered as the outcomes of infinitely repeated uniform experiments. Second, he introduces (in his Theorem A) the limit of the relative frequency of a particular outcome α. Third, (in his Theorem B) the notion of place selection defined by a sequence of functions $f_n(x_1, x_2, \cdots x_{n-1})$ is employed. All these three concepts are completely strange to the so called classical theory as developed by Bernoulli, Laplace, Poisson, etc. They have been introduced and made the corner stone of probability theory in my papers published since 1919. I daresay that in no probability investigation before 1919 any of those notions even were mentioned.

This concerns what Professor Doob calls the Problem I or the purely mathematical aspect of the question. As to his Problem II or the relationship between the formal calculus and real facts Professor Doob stresses that the actual values for probabilities that enter as data into a particular argument have to be drawn from long, finite sequences of experiments. This is in complete accordance with the standpoint of my theory and in strict contradiction to the classical conception which knows only "a priori" probabilities determined by "equally likely cases."

In both theories, Professor Doob's and mine (not in the classical) a mathethematical model or picture is associated with a long sequence of uniform experiments. These models are different in both theories. My model (the "kollektiv") consists of one infinite sequence ω: $x_1, x_2, x_3, \cdots$ in which the limit of the relative frequency of each possible outcome α exists and is indifferent to a place selection; the value of this limit is called the probability of α.

On the other hand Professor Doob's model implies all logically possible sequences which form a space Ω and he shows that in this space a measure function can be introduced which fulfills the following conditions: (1) If m is a positive integer, the set of all sequences the mth element of which is α has a measure p_α independent of m; (2) the set of all sequences in which the relative frequency of α-results has either no limit or a limit different from p_α is zero; (3) if S is any place selection, the set of all sequences ω for which the relative frequency of α in $S(\omega)$ has either no limit or a limit different from p_α is likewise zero; this value p_α is called the probability of the outcome α. It then can be shown that a probability in this sense can be ascribed to certain events, i.e. to certain types of experiments which in some way are connected with the sequence of basic experiments. E.g. if the original sequence consists of the single successive tossings of a die, the derived sequence may consist of pairs of tossings with the sum of the outcoming points as new value of α. The new probabilities p'_α are found as measures of certain sets in the original measure system established in Ω.

There is no doubt that the model used by Professor Doob for representing empirical sequences of uniform experiments is logically consistent. Its practical usefulness depends on how the usual problems of combining different kollektivs and so on can be settled within this scheme. This has to be shown in detail. It seems to me that my conception is simpler in its application and closer to reality, while his model may be considered more satisfactory from a logical standpoint since it avoids the difficulties connected with the concept of "all place selections." At any rate, however, there is no contradiction or irreconcilable contrast: both theories are essentially statistical or frequency theories, equally far from the classical conception based on "equally likely cases." In both theories probabilities are, of course, measures of sets.

2. **Comments by J. L. Doob.** It is perhaps unfortunate that Professor von Mises' treatment of probability problems, based on typical sequences ("collectives," "admissible numbers"), is commonly called the "frequency theory."[1] It is clear to any reader of our papers (identified as M and D below) that the idea of frequency, at least in the discussion of the relation of mathematics to practice, is no more fundamental to one approach than to the other. In one mathematical treatment frequency notions first appear in the theorems, whereas in the other they first appear in the axioms; but they appear in both. The principal objection the measure advocates have to the frequency approach is that it is awkward mathematically. Anyone who doubts this awkwardness need only examine various books published recently, using this approach, to see what a lot of fussy detail is involved merely in proving such elementary results as the Tchebycheff inequality or the Bernoulli theorem. One author considers it necessary to have his chance variables so restricted that if x is a chance variable, the event $x < k$ has a probability assigned to it only if k is not in some exceptional set, which may be infinite. To take another example, consider the coin tossing game discussed in both M and D, in which two out of three wins at tosses win a game. Apparently the probability analysis of this game is somewhat difficult in terms of the frequency theory. As the quite elementary treatment outlined in D shows, there is no difficulty involved, using the measure approach. The question is simple: a set of chance variables is given (corresponding to the original tosses); a new set is determined from them (corresponding to the grouping into games). Only elementary algebraic manipulation is required to verify that the new chance variables are mutually independent in the mathematical sense, (Cf. D), and have the same distribution, so the law of large numbers is applicable. Professor von Mises considers that the measure theory cannot handle this problem. I on the other hand consider that this problem exhibits the mathematical disadvantages of the frequency theory.

[1] This identifying name will be used below also.

The frequency theory reduces everything to the study of sequences of mutually independent chance variables, having a common distribution. "Probability theory is the study of the transformations of admissible numbers" writes Professor von Mises. This point of view is extremely narrow. Many problems of probability, say those involved in time series, can only be reduced in a most artificial way to the study of a sequence of mutually independent chance variables, and the actual study is not helped by this reduction, which is merely a tour de force.

It is claimed in M that the axioms of measure theory only describe the distribution within one collective (M, p. 00). This statement seems to mean that only the measure relations (using the notation of D) of the first coordinate function $x_1(\omega)$ can be discussed in the measure theory, that is only probabilities of the type: the probability that $x_1 < k$ (in the language of practice, "the probability that the result of the first experiment is less than k") are discussed. Actually, however, (Cf. D) the measure theory can discuss any number of experiments simultaneously, using the appropriate space Ω.

Many of the debates between the advocates of the various probability theories have been wasted, because some of the debaters talk mathematics, others physics. With this in mind, I should like to stress again[2] that (except for a few philosophically inclined Englishmen) everyone calculates probability numbers in the same way—a combination of reasoning based on experience and helped by theory, with examination of the experimental conditions and the results of trials. Frequency considerations necessarily play a large part. The fact that almost everyone calculates probability numbers in the same way does not alter the fact that one mathematical theory may be more useful or convenient than another in dealing with these probability numbers.

In closing, it seems proper to call attention to what the measure advocates consider the real services and contributions of the approach of Professor von Mises. Professor von Mises was the first to stress the importance of the second of two fundamental generalizations of experience in dealing with repeated mutually independent experiments of the same character: (1) the clustering of success ratios and (2) the fact that this clustering is unaffected by a system of rejection as described in M and D. These two generalizations of experience are certainly fundamental. The only point under discussion here is how such generalizations are to be put into a mathematical setting. The original such setting of Professor von Mises was criticized as not really mathematical. The setting now proposed by Copeland and others is criticized by the measure advocates as mathematically inflexible and clumsy. But it is significant that even in a treatment of the measure approach, as in D, it was felt essential to stress the mathematical interpretation of the two empirical generalizations of Professor von Mises. In the terminology of D, the measure advocates consider the contribution of Professor von Mises' approach to be a contribution to a solution of Problem II, not to Problem I, the mathematical problem.

---

[2] We are not talking mathematics now, but the application of mathematics.