THE DIVERSE DEFINITIONS OF PROBABILITY [1]

by

MAURICE FRÉCHET (Paris)

As an introduction to the discussion of the different definitions of the probability, I shall briefly sketch a survey of these definitions, at least of those which are most common and which express different points of view.

A SKETCH OF THREE CONCEPTIONS OF PROBABILITY

There are three different fields of research which have given rise to three different conceptions of probability.

I. — Problems in games are in the background of the "a priori" school, where *apriori* conditions — that is, considerations based on the nature of the trial before its result is known, such as the symmetry of a coin or the similarity of cards — lead to the equal possibility of some events. Then when all possible results can be classified and distributed into mutually exclusive and equally possible cases, the probability of a fortuitous event E is defined as the ratio of the number of favourable (to E) cases to the number of equally possible cases.

II. — Demographic, economic and insurance problems have inspired the "statistical" school which starts from the connection between probability and frequency. If, in 100 trials, an event E appears 63 times, we say that the frequency of the event E in 100 trials is $\frac{63}{100}$. Or more generally, the frequency of E in n trials is the ratio $\frac{r}{n}$ of the number r of "favourable" (to E) trials to the total number n of trials. The connection consists in the practical fact that in statistical problems, the direct computation of proba-

[1]) Lecture read at the fourth International Congress for the Unity of Science. (Cambridge, England, 1938).

bilities consists in adopting as an approximate estimate of the probability of an event E, its frequency in n trials, the approximation being generally safer when the number n of trials is greater.

(But, as we will see further on, this school divides itself into two schools when the question is to introduce probability in an axiomatic theory).

Game problems and statistical problems lead generally to an *objective* conception of probability. When the event E and the nature of the trials have been specified, the corresponding probability is supposed to have a definite value, the same for everybody.

III. — The study of logic has sometimes led to a *subjective* conception of what probability is. What do we mean when we speak of probability of rain to-morrow? It seems that, to-morrow being a unique day in the history of the world, no frequency can be computed or even conceived in such a case to give an approximate answer; we have here an "isolated trial". Neither can the conception of equally possible events help us. In such a case, philosophers have thought that the probability of an event E is measured by the "degree of belief" in the occurrence of E. If such is the case, the probability will obviously be subjective. Some authors have maintained that as a degree of belief is not a number they may assume that probabilities cannot all be estimated numerically. Others think that — in much the same way as the degree of desirability of goods is expressed by the numerical values of prices — the degree of belief that the event E will occur may be expressed by the numerical value of odds in a bet.

If Brown bets four to three that it will rain to-morrow, we shall say that the probability of this event is $\frac{4}{4+3} = \frac{4}{7}$. But as, at the same time, Smith bets only two to three, we conclude that the degree of belief in the occurence of the event is a personal matter, being expressed numerically by $\frac{4}{7}$ for Brown and by $\frac{2}{2+3} = \frac{2}{5}$ for Smith. And the explanation of this difference is found, by those who hold this theory, in the fact that the amount of knowledge or of past experience varies from one individual to another. Some of the British scientists (by no means, all of them) have advocated the subjective theory. Their views have been, in the last few years, modified and completed by Borel (1) [1]), in an analysis of Mr. Keynes (1) book. Emile Borel (2) has expounded more fully his own doctrine in the last volume, just appeared, of his famous Treatise of Calculus of probability. De Finetti (1) has also put forward his original ideas under the title "Probabilisti di Cambridge". I ought to add that I heard a very interesting lecture by Mazurkiewicz in which he expounded a definition of the probability of propositions, thus establishing also a connection between probability and logic; but it seems that he did not consider probability as having a subjective value. As I have not yet seen a written exposition of his theory, I cannot enter safely into more details.

TWO DIFFERENT STATISTICAL DEFINITIONS OF PROBABILITY

After having reviewed the origins of the different points of view concerning the notion of probability, we have to observe that the modern tendency to axiomatisation has also prevailed recently in the Theory of probability.

I. *Probability as a limit of frequencies.* — It is there that the statistical school devides itself. It is very tempting and it seems, at first sight safe and simple "to proceed to the limit" in order to go from the practice to the theory. And this is why it has been considered by many authors in the past that the only way to express axiomatically the connection between frequency and probability is the following one: to define the probability of an event E as equal to the "total frequency" of E, that is, as the limit, when it exists, of the frequency of E in n trials when n tends to infinity.

De Mises was the first to perceive that this was quite insufficient and that it overlooked many difficulties. He attempted to evolve conditions which a sequence of supposed results of trials ought to satisfy in order to represent a true random sequence. A sequence of supposed results of trials is called by him a "collective" when: 1°. the total frequency [2]) of the event E in the sequence exists (its value p being then called the probability of E in the collective); 2°. the same property holds, with the same limit p when the sequence is replaced by any sequence extracted from it. The intro-

[1]) Figures in thick print refer to the list of publications at the end of this lecture.

[2]) It is not useless to employ two expressions "total frequency" and "probability". For the total frequency has been mathematically defined, independently of the notion of random sequence, or of ,.collective".

duction of this second condition was a great improvement. But it was too strict for a collective to exist. If the trials consist for instance in throwing a coin, consider the case when in the sequence there is an infinite number of heads and an infinite number of tails. Among the selected sub-sequences will be found the subsequence giving the heads and the subsequence giving the tails. The total frequency of heads in those two subsequences will be equal to I in the first subsequence, to O in the second one. Therefore the sequence would not be a "collective". So that the only sequences which could be collectives, would be those where the event E always occurs (or never occurs), from one finite rank on; and p should always be O (or I). Obviously these consequences were not desired and de Mises had to limit the choice of the "laws of selection" (each law defining the ranks of the trials forming one of the subsequences) in order to avoid the choice of the above subsequences. This is the beginning of a series of modifications of the originally simple and intuitive definition of a collective which have made the corresponding definitions more and more intricate.

Wald (1) has succeeded in making the theory a consistent one by restricting the family S of the laws of selections of ranks (those laws by which the subsequences are defined). He imposes on S the condition to be *denumerable*, that is to be such that the laws of S can be numbered by means of integral ranks [1]) and thus represented as $L_1, L_2, \ldots\ldots L_n, \ldots\ldots$

This modification allows Wald to prove: 1° that such collectives exist, 2° that, given p and S, the probability that a random sequence of total probability p should be a collective relatively to p and S, is equal to unity.

I ought to have mentioned that, prior to Wald, some authors (Popper, Reichenbach, Copeland,...) had tried to avoid objections against De Mises' second condition. But Wald has shown that you may get a collective in the senses of these scientists by choosing accordingly the denumerable family S of the laws of selection. So that their theory is a particular form of Wald's theory.

In the new form given by Wald, the theory becomes logical but new objections arise with which we will deal later on (p. 19) or in the discussion.

[1]) It has been proved by Cantor that the set of all irrational numbers (like π, $\sqrt{2}, \ldots$) is not denumerable. It may be shown to follow that the set of all the subsequences of a given sequence is not denumerable.

II. — *Probability as a physical magnitude of which frequencies are measures.* I will now show that it was not necessary for the statistical school to proceed "to the limit" of the frequencies.

In any Science, the axiomatic theory should be preceded by an "inductive Synthesis". In this Synthesis, intuition and contact with reality are the main directions to follow and therefore rigour is not supreme.

Applied to probability, this leads us to conclude from the practical statistical processes that any frequency is to be considered as an approximate measure of one *physical constant* attached to an event E and to a category C of trials. We also know that, in practice, the greater the number n, of trials, the safer the approximation to p; this means, that for great values of n, the observed frequencies differ generally little, one from another. This constant is the probability of E over C.

Then we may deduce properties of probability from some obvious properties of frequencies.

Thus the frequency $\frac{r}{n}$ is clearly such that

$$O \leqslant \frac{r}{n} \leqslant I \tag{1}$$

The frequency of an event certain to happen is

$$\frac{n}{n} = I \tag{2}$$

Now let E_I, —— E_s be s mutually exclusive events and $\frac{r_1}{n}, \ldots \frac{r_s}{n}$, their corresponding frequencies; for the frequency $\frac{r}{n}$ of the event E = (either E_I or $E_2 \ldots$ or E_s), we get obviously $r = r_I + \ldots + r_s$, whence

$$\frac{r_1}{n} + \ldots + \frac{r_s}{n} = \frac{r}{n} \tag{3}$$

Finally, let G consist in the simultaneous occurence of E and F. In n trials, let G occur ρ times and E, r times. Obviously

$$\frac{\rho}{n} = \frac{r}{n}\frac{\rho}{r}, \text{ that is to say:} \tag{4}$$

the frequency $\frac{\rho}{n}$ of the simultaneous occurence G of E and F, is equal to the product of the frequency $\frac{r}{n}$ of E in the same n trials, multiplied by the frequency $\frac{\rho}{r}$ of the occurence of F amongst those of these n trials in which E occurs.

Now if n is great, all these frequencies are generally safe measures of the corresponding probabilities. The above four properties of frequencies may then be empirically extended to the probabilities. This is *not* done by proceeding mathematically to the limit, when n increases; for there may be no limit or the limit may be different from the probability. This is done exactly as in experimental sciences where experimental measures are *generally* approximate values of physical magnitudes.

The classical axioms. — We are thus led to assume that the probability Pr (E) — to abbreviate a more complete designation Prob_C (E) — possesses the following properties

(1′) $O \leqslant \mathrm{Pr}\,(E) \leqslant I$

(2′) Pr (certitude) $= I$

(3′) Pr (either E_I or E_2, ... or E_s) $= \mathrm{Pr}\, E_I + \ldots\ldots + \mathrm{Pr}\, E_s$
where $E_I \ldots E_s$ are mutually exclusive events.

Let it be possible for two events E and F to appear in one category C of trials: The probability Pr_C (E.F) of the occurrence, E. F, of both E and F in the category C is equal to the product of the probability $\mathrm{Pr}_C E$ of E in the category C by the probability $\mathrm{Pr}_{(E.C)}F$ of F in the category $C' = E.C$ of all those trials of C where E occurs.

(4′) $\mathrm{Pr}_C(E.F) = [\mathrm{Pr}_C E]\,[\mathrm{Pr}_{(E.C)}F]$

These properties may also be obtained by starting from the classical definition of probability based on the notion of equally possible events [1]).

[1]) When applied to geometric probabilities, this definition gives rise to some mathematical difficulties which may be avoided by introducing (as I did once in one of my courses) the postulate that if E implies F the probability of F is not smaller than the probability of E. This brings the classical definition of probability into line with the definition of plane area, where it is not sufficient to assume that equal curves contain equal areas but also that if a region e contains a region f, the area of f is not smaller that that of e.

Now we have the essential of what is necessary for building up an axiomatic theory.

In a classical axiomatized theory, the probability will not be defined *constructively* as by De Mises, but *descriptively*. Thus: to each fortuitous event is attached in a given category, C, of trials a number $Pr_C(E)$, which is subject only to properties (1′), (2′), (3′), (4′). This simple definition is perhaps not explicitly stated in the classical treatises of Laplace, Bertrand, Poincaré, but they certainly perceived that the whole mathematical part of their treatises could be derived from this axiomatic definition without having to define more precisely the nature of probability and its connection with E and C.

The modernized axiomatic theory. — This theory has been modified under the simultaneous influence of the discovery of measure by Borel and of the consideration of abstract sets. Its new form (which has been a long time present in many minds) has been for the first time fully and explicitly stated by Kolmogoroff. His definition of probability may be regarded as characterizing what may be called the *modernized axiomatic definition* of probability. I know directly from him that he agrees with me that, to be complete, his axiomatic definition should be preceded by an Inductive Synthesis and followed by numerical verifications.

The modern modifications of the classical axiomatized theory are of two kinds. Thinking of the case where the probability for a point to be on a subset of a segment T (0,1) is uniform, we are led to consider in that case the measure of a subset e of T as the probability for the point to be on e. But then, as some subsets are not measurable, we have to restrict the theory to a family of events which are — as I said in 1930 — "probabilisable" ones. Thus, for a given category C of trials, the probability Pr (E) will appear as a functional (function of an event).

But an event is a somewhat vague notion. The theory of sets leads us to express Pr (E) differently. Let e be the set of the trials of C where E occurs .

Then Pr (E) will be a function p (e) of the set e. Besides in (3′), E_1, $E_2, \dots E_n$, $E =$ (either E_1 or $E_2 \dots$ or E_n), will correspond to $e_1, e_2, \dots e_n$. Thus p(e) is an *additive function* of sets. Furthermore the family f of the sets e which correspond to "probabilisable" events clearly must be such that if $e_1, e_2, \dots, e_n$ belong to f, $e_1 + \dots + e_n$ belongs also to f. So that f is an additive family of sets.

There remains the notion of trials, which are the elements of e and C. As an axiomatic theory must not be limited in its interpretations to one single real well defined concrete field, we will introduce here the basic idea of the theory of abstract sets and extend the theory to the case when C is a fundamental set of elements of any nature whatsoever: trials, points, propositions...

Finally, observe that the essential novelty of the measure notion was the admission of the *complete* additivity of measure, that is the extension of the additivity of the lengths of a finite number of segments to the case of an infinite number of segments. Similarly, in the modernized axiomatic theory of probability, we shall admit complete additivity, that is, if in a denumerable sequence of sets, $e_1, e_2 \ldots e_n \ldots$, each belongs to f, then $e_1 + e_2 + \ldots$ will also belong to f and

$$(3'') \qquad p\,(e_1 + e_2 + \ldots + e_n + \ldots) = p\,(e_1) + p\,(e_2) + \ldots\, + p\,(e_n) + \ldots$$

where the sets $e_1, e_2 \ldots$ are in infinite number and do not overlap.

The postulate of *complete* additivity of p (e) had been objected to. It is true that this postulate is of a different nature from the previous ones. Properties (1'), (2'), (3'), (4') have been arrived at from the experimental notion of frequency. On the contrary the extension of (3') to (3'') is based on mathematical convenience, exactly in the same way as was the introduction of irrational numbers. Empirical measurement of a physical constant (length, weight, ...) cannot prove the magnitude of this constant to be rational (ratio of two integers) or irrational. But most mathematical statements and proofs would become extremely complicated if it were thought better to ignore — and this would be perfectly possible — the irrational numbers and dispense with them. Similarly, some authors may prefer to limit the property (3') to a finite number of events, but great simplifications occur when it is extended to an infinite number of events.

Induction of axioms and verification. — As has been said before, an axiomatic theory concerning undefined elements may have many interpretations in the real world. But each of these interpretations may be more or less adequate. For instance, Euclidean geometry is sufficient to describe the space in engineering and in practical astronomy, whereas relativity theory is supposed to be more correct when we come to more precise astronomical measurements.

Therefore no science can be restricted to an axiomatic theory.

An inductive synthesis has led to this axiomatic theory and must not be lost sight of; in this inductive synthesis a bold step has been taken by schematizing empirical notions (such as a ball) by means of idealized abstract notions (such as a sphere). Pure logic has then permitted the deduction of exact numerical relations between these abstract notions. Then, *if* the step mentioned above has been well chosen, the inverse step from these abstract notions to the corresponding empirical notions will make the above mentioned numerical relations appear — no more as exactly, but still — as approximately verified by these last empirical notions. For any given axiomatic theory we have thus to verify whether this expected result really occurs and whether the approximation is sufficient.

Thus, in practical geometry the ratio of the length of a thread surrounding a wheel to the diameter of this wheel is verified to be, not of course exactly equal to $\pi = 3{,}14159\ldots$, but approximately equal to π. Such approximate values of π are obtained exactly in the same way through the theory of probability in the famous needles-problem of Buffon by computing certain frequencies: let $l < a$ and let parallel lines be drawn on the floor at mutual distances equal to a. If in throwing n times a needle of length l, this needle is found to cross r times one or the other of the lines, then — provided that n is not too small —, $\frac{2l}{a}\frac{n}{r}$ must be approximately equal to π. You may easily repeat this experiment. Making it for $a = 6{,}14$ centimetres, $l = 5{,}55$ cm, $n = 100$, we found $r = 51$, whence π is found approximately equal to 3,5.

Making also the geometrical experiment mentioned above with a round tin and a ribbon marked in centimetres, we found for π the value $\frac{18{,}4}{5{,}8} = 3{,}17$.

Category of trials, (population, universe). — Let us now return to the notion of a category of trials.

Very often, as we did in the first part of this lecture, the probability of an event is referred to without mentioning the category of trials. Very often, indeed it is not necessary to mention it, because there is no doubt about which category it is. Such is the case in most games: throwing coins, dice, drawing a card and so on. However even in those games it may be useful to be more precise. In some card games 32 cards are used, in others 52: the probability of drawing a king has accordingly two values: $\frac{4}{32}$, $\frac{4}{52}$, which correspond to two different categories of trials. In most statistical problems, it is even essential to state in which population, in which universe the occurrence of the event is considered. The probability of dying at age 45 is very indefinite; the trials consist in the observation of the ages of deaths amongst one "population". The value of this probability is not the same if the population consists for instance of new-born children or if it consists of men aged 44.

A famous paradox due to Joseph Bertrand shows also that the so-called "geometrical" probabilities are not well defined when the category of trials is not specified precisely. The question was: what is the probability that a chord l of a circle S, be smaller than the side L of an equilateral triangle inscribed in S. Bertrand showed that three classical methods gave three different values for this probability. It is easy to verify that his three methods corresponded to three categories of trials. In one of them, the two extremities of the chord were chosen independently on the circumference and the material choice was such that the chances on the circumference were distributed uniformly. In another method, the middle of the chord and its direction were chosen independently and here again uniformly; and so on.

The uniformity of distribution of probability is very often too easily admitted.

A point A is chosen at random in a square S; what is the probability p that it will be in a subarea s of S? It depends on how the experiment is actually made: it depends on the category of trials. Suppose S is divided into n equal subareas, s. If a number is assigned to each and if the number corresponding to s is chosen at random from identical cards in a bag, p will be $=\frac{1}{n}$. If S is a very great hall and the point A is the point where a coin falls when thrown at random from the centre, p will be smaller near the walls.

These remarks may appear too particular. They have however an important bearing, not only on the numerical computation of probability, but also on the philosophical meaning of probability.

We have now explained how a modernized axiomatic theory can be built on an inductive synthesis based on the usual statistical practices and without proceeding to the limit.

Return to the "proceed-to-the-limit" definition. — As many statisticians, actuaries and philosophers still believed recently that the only alternative to Laplace's definition of probability was the "proceed-to-the-limit" definition, it is perhaps not sufficient to have proved above that it is not so: it might be useful also to state that *most of the mathematicians* who have made important contributions to the progress of the theory of probability really base explicitly or implicitly their mathematical proofs on the modernized set of axioms introduced above and explicitly stated by Kolmogoroff (1).

However, as amongst those who contribute effectively to the development of the Calculus of Probability, a small minority still exist who adhere to the "proceed-to the limit" definition, we shall return to it and present a few other objections to it.

In the "proceed-to-the-limit" definition of probability, the category of trials is always supposed to be an ordered infinite denumerable sequence of trials. Of course, such sequences are of great importance. However, there is no reason to consider them exclusively and there are obvious reasons for not doing so. For instance, if somebody shows me a coin and asks me what is the probability of getting heads in throwing it, I may give the answer without thinking of a particular sequence of throws, (such as the one which would be obtained for instance by myself in throwing the coin every 10 seconds). This would be much too restricted. Perhaps at a certain time in the future, paper money only will be in use, all copper or metal money having disappeared and then never again will a coin be thrown. However, the probability of getting heads will not cease to have a meaning and be equal to 1/2: the category of trials consists implicitly, in our minds, of all real or imaginable throws of similar coins in similar conditions, simultaneously or successively, in the past, present or future.

Those who advocate the "proceed to the limit" definition think that they have in this way made the theory nearer practice. In fact, they have made it more remote.

De Finetti has shown this by considering the case when a sequence S of H (eads) and T (ails) consists of 10.000 Ts followed by a sequence S' of H's and T's which would be a "collective" (in the sense of either De Mises or Reichenbach, Popper, Copeland, Wald,. .) and would correspond to $p = 1/2$. Clearly S should be a "collective" in the same sense as S' and would still correspond to $p = 1/2$.

However, in practice, any gambler, any statistician, when he had seen T repeated so consistently would disregard S' and would say that the beginning of S is such that in practice it cannot correspond to $p = 1/2$.

Even the mere restriction to sequences called "collectives" (where the frequency r/n of the event E considered in n trials necessarily tends to a limit p), is contrary to the assumption that E is fortuitous. For, by that assumption, it is admitted that each trial *may* give E or its contrary C. Therefore, each conceivable sequence of results, E E C E C C . . . ., is supposed to be possible. It is true that in the modernized axiomatic theory, if we consider as element of a universe, each infinite sequence of, for instance, throws of a coin, then in this universe, there is a probability equal to zero that in such a sequence the "total" frequency does not exist or is not equal to 1/2. The interpretation of this probability as a physical constant shows that such particular sequences occur very rarely. But this interpretation does not exclude their consideration. As I have insisted elsewhere on that objection, this has happily led De Mises to state that he admits the consideration of such exceptional sequences but does not deal with them. We are thus now very near each other; for I am also ready to admit that the consideration of these exceptional sequences is more a question of theory than of practice. Just as in a similar way, everybody admits that most of the continuous functions met with in practice are differentiable. But it was none the less a step forward to recognise that continuity does not imply differentiability.

There remains the difficulty pointed out above, that a contradiction arises unless in De Mises' second condition, some laws of selection are not admitted, for instance the law S_1 which would extract from a given collective C_0, precisely that subsequence formed by the trials where the event E occurred. Now De Mises (3) in a recent rejoinder states that he restricted the set of the laws of selection on which his second condition operates, in such a way as to exclude L_0. But it is difficult to see how this exclusion is implied in the wording of his second condition. In his Treatise of Probability (De Mises, (1), p. 12) we read "Aus einer unendlichen Folge von Beobachtungen wird durch "Stellenauswahl" eine Teilfolge gebildet indem man eine Vorschrift angibt durch die ueber die Zugehoerigkeit oder Nichtzugehorigkeit der n^{ten} Beobachtung und hoechstens unter Benützung der Kenntnisse der vorhergegangenen Beobachtungsergebnisse entschieden wird". We fail to see why this should exclude S_1.

In the rejoinder mentioned above, de Mises says:

„Dans l'ensemble de toutes les sélections imaginables se trouvent la sélection S_1 qui ne retient que les épreuves où l'évènement s'est produit et la sélection S_2 ne retenant que les autres cas; évidemment S_1 et S_2 changent les valeurs de la fréquence relative. Mais, ni S_1, ni S_2 ne sont des „choix de position" d'après ma définition. Donc cette objection ne frappe pas ma théorie".

Perhaps the misunderstanding came from the idea that S_1 should extract from *every* collective C a sequence E E E. . . . ; whereas we consider a single collective C_0 where E occurs an infinity of times and we say that amongst the laws of selection, (as they are defined in de Mises' Treatise in the above quotation), there is one, S_1, which selects precisely the ranks where E occurs *in* C_0.

Wald has avoided the above difficulty by assuming that the laws of selection which operate in De Mises' second condition are so restricted that they form a denumerable (infinite) family of laws. Thanks to that restriction, he can now prove that collectives in the new sense exist. This is a new and very important result. But the Wald modification, while disposing of a mathematical objecttion raises new ones.

According to the Wald–Mises definition, the probability of an event E in a collective S is a "relative notion", since in order to know what it means, we have still to know relative to which denumerable family F of laws of selection it has been defined. De Mises has observed that whatever F may be, the value p of the probability is the same when E and the sequence S are given. We agree that the *value* of the probability is not relative to F, but it is its nature, its meaning which are relative to F.

For example, suppose that a straight line of a given length on a map represents a road. The meaning of this length to a walker will be very different according as this road is horizontal or has various slopes.

Wald's answer to the previous objection is very different. He observes that the meaning of the probability will be unique, when it has been decided to take for F the family, F_1 of all laws of selection which can be defined. And he considers that this is only possible when the definition contains a finite number of words. Then since, as is known, the set of all discourses containing a finite (not fixed) number of words is denumerable, it follows that the particular family F_1, considered by Wald is denumerable (as it should be according to his own general definition of collectives).

Here is an answer difficult to deal with because it touches a point on which all mathematicians are not in agreement.

The choice of F_1 is very ingenious. However, consider a sequence S (such as E E C E C C ....) and the family f_1, of sequences σ of rising ranks of S defined by the laws of F_1.

Each sequence σ: 2, 5, 9, 13....., for instance, corresponds to a sequence s of positive integers 2, $3 = 5 - 2$, $4 = 9 - 5$, $4 = 13 - 9$ ....; and σ corresponds to a continued fraction:

$$X_\sigma = \cfrac{1}{2 + \cfrac{1}{3 + \cfrac{1}{4 + \cfrac{1}{4 + \cfrac{1}{\ddots}}}}}$$

then, f_1 consists of $\sigma_1, \sigma_2, \ldots$ to which correspond $x_1 = X_{\sigma_1}$, $x_2 = X_{\sigma_2}, \ldots$

The restriction of Wald consists in replacing the set I of all the X_σ (I is the set of all real numbers between O and I) by a particular denumerable set i extracted from I. The fact that each number from the set i may be defined by a finite number of words is considered as a justification of this choice: for it is the greatest set we can think of.

This is a much debated question. Let α be the set of all real numbers of which each may be "defined in a finite number of words".

Cantor has shown us how to define a number not pertaining to α. If α is defined for instance by the sequences:

s_1	3,	12,	5,	1,	132,	....
s_2	4,	╲ 2,	9,	23,	4,	....
s_3	81,	9,	╲34,	4,	7,	....
s_4	7,	51,	6,	╲92,	1234,	....
....						

and if in the diagonal we replace each number by a number obtained by adding I, we obtain a sequence s': 4, 3, 35, 93,.... corresponding to a rising sequence: σ': 4, 7, 42, 135,... which defines a number $X_{\sigma'}$ certainly not belonging to α. We have thus very precisely defined this number in such a way that it cannot belong to the set α of all numbers (defined each one in a finite number of words), for s' cannot be identical with one of $s_1, s_2, \ldots$..

And yet, on the other hand, our rule for defining this number $X_{\sigma'}$ has been given in a finite number of words. Hence a contradiction arises.

It is true that this definition of σ' is based on the definition of α. But can we not say that α, being defined as the set of all numbers defined in a finite number of words, is thus itself defined in a finite number of words?

All this may explain why some mathematicians doubt whether the definition of α has any meaning at all.

Anyhow, there is no doubt that the second modification (the restriction to definitions given in a finite number of words) is no more intuitive than the first (the restriction to a denumerable family of laws of selection) proposed by Wald. So that in the modification of the De Mises' definition, logic is obtained at the cost of its original simplicity.

Is a mathematical model of randomness possible? — Besides, the consistency of Wald's axioms is only one of the qualities to be required of any axiomatic theory. Above all, it must be a fairly good — though, of course, never fully adequate — model of some of the set of facts which it purports to explain; the aim of Wald's as well as of De Mises' definition of collectives has obviously been to give mathematical rules allowing us to imitate true random sequences. Since sequences of results of trials usually show complete disorganisation, lack of any regularity, these authors have tried to formulate definitions of sequences possessing this character of disorganisation. It remained to be seen whether they have succeeded in doing it or even whether, indeed, it could be done at all. The first part of this question has been answered in the negative by Ville (1). However, as his work on the subject is of an abstract and mathematical nature, we shall only here refer to his book or to the very short summary of this book which we have given at Geneva (Fréchet, (3), p. 38—41). But there is one consequence of Ville's results which is easily grasped without the introduction of elaborate mathematical reasoning, and this we shall state here. According to a corollary of Ville: given a number p ($0 \leqslant p \leqslant 1$) and a denumerable family F of laws of selection, there is at least one collective (in the sense of Wald) relative to p and F, such that the frequency in n trials converges to p, when n increases, *by values always greater than p*.

Now, this is clearly one of those regularities which, though theoretically possible, in practise never occur in truly random sequences.

This corollary is then quite sufficient to show that with their definitions, De Mises and Wald not only *have failed* to eliminate all regularities but even have not succeeded in eliminating one of the most easily recognizable of them.

It is only fair to mention that in the rejoinder referred to above, de Mises says he cannot admit that the corollary of Ville creates a difficulty in his theory. We shall leave the reader to decide between them.

Besides, if De Mises and Wald have failed to solve the problem, it is surely not due to lack of ability, but perhaps, as some writers assert, because this problem is insoluble.

In several papers E. Borel and P. Lévy have expressed the opinion that it is impossible to realize mathematically random sequences. (See particularly, Borel's (3) note in the C.R.).

There are other difficulties in the Wald–De Mises' definition, for which in order to avoid being too long or too technical, we shall refer to our Geneva lecture and to Ville's book.

Conclusion: We shall summarise our objections thus: in De Mises theory, the definition of probability does not satisfy the requirements of logic; Wald's restatement of the definition is logical, but lacks the original simplicity and does not satisfiy certain other requirements of a good axiomatization.

In the rejoinder mentioned previously, De Mises states that this increase in complexity at the same time as in rigour is what has also happened with other axiomatized sciences, as mechanics, physics and so on. This is quite true and we would have to accept the „proceed-to-the-limit" definition if there were no alternative. We hope however to have shown that not only another alternative, but a more satisfactory one, consists in the modernized axiomatic theory as summarised, for instance in Kolmogoroff's set of axioms, followed by numerical verifications and preceded by a Inductive synthesis founded on the notion of probability as a physical magnitude of which frequencies are measures.

We repeat that in a full study of the diverse definitions of probability much more space ought to be devoted than has been done above to the notion of probability as a degree of belief or as a subjective notion, and to the notion of isolated trial. Along with Borel's (2) exposition of these ideas, a note by de Finetti, to be found in these Proceedings, will help to give an idea of the modern tendencies in these directions.

LISTE DES AUTEURS CITES

E. Borel: *A propos d'un livre récent de M. Keynes* (Revue de philosophie, tome XCVIII, 1924).

E. Borel: Traité de Calcul des Probabilités: dernier fascicule: *Valeur pratique et Philosophi du Calcul des Probabilités* (Gauthier-Villars; 1939).

E. Borel: *Sur l'imitation du hasard,* C.R. Acad. Sc.p,. 203—205, vol. 204, 1937.

de Finetti: *Probabilisti di Cambridge* (Supplemento di Statistica, anno IV, p. 1—19, 1938).

de Finetti: Voir Genève.

Feller: Voir Genève.

Fréchet: Fr. et Halbwachs: *Le Calcul des Probabilités à la portée de tous* (Dunod, 297 pages, 1925).

Fréchet: *Recherches théoriques modernes sur le Calcul des Probabilités* (Premier Livre: Généralités sur les probabilités. Variables aléatoires, XVI + 308 pages). (Gauthier-Villars, 1937).

Fréchet: Voir Genève.

Jeffrey: *Scientific inference,* second edition.

Genève: Voir: Fréchet (3), Steffensen (1), De Mises (3), Wald (1), dans la *cinquième* partie, et de Finetti (1) dans la *huitième* partie du volume: *Colloque consacré au Calcul des Probabilités*" dans la collection des "*Conférences internationales de mathématiques de Genève*". (Hermann, 1938).

Kolmogoroff: "*Grundbegriffe der Wahrscheinlichkeitsrechnung* (62 pages, Springer, 1933).

de Mises: *Wahrscheinlichkeitsrechnung* (574 pages, chez Deutike, 1931).

de Mises : *Wahrscheinlichkeit, Statistik und Wahrheit* (2e édition, 282 pages, Springer, 1936).

de Mises: Voir Genève.

Reichenbach: *Les fondements logiques du Calcul des Probabilités* (Annales de l'Institut Henri Poincaré, t. VII, p. 267—348, 1937).

Steffensen: Voir Genève.

Ville: *Essai sur la théorie des collectifs,* 144 pages, 1939 (Collection de monographies sur le Calcul des Probabilités sous la direction de M. Borel, chez Gauthier-Villars).

Wald: Voir Genève.