BRUNO DE FINETTI

Foresight: Its Logical Laws, Its Subjective Sources
(1937)

This article appeared in the *Annales de l'Institut Henri Poincaré,* Vol. 7 (1937).
Reprinted by permission of the author and the publisher, Gauthier-Villars.
Translated by Henry E. Kyburg, Jr.

TRANSLATOR'S NOTE

The translator wishes to express his gratitude for their time and advice on many matters to Professor L. J. Savage, Professor Bruno de Finetti, and Miss Caroline Clauser. The translation has benefited greatly from their suggestions. He particularly wishes to express his gratitude to de Finetti for suggesting the following changes.

Words: The word "equivalent" of the original has been translated throughout as "exchangeable". The original term (used also by Khinchin) and even the term "symmetric" (used by Savage and Hewitt) appear to admit ambiguity. The word "exchangeable", proposed by Fréchet, seems expressive and unambiguous and has been adopted and recommended by most authors, including de Finetti.

The word "subjectiv" was used ambiguously in the original paper, both in the sense of "subjective" or "personal", as in "subjective probability", and in the sense of "subjectivistic", as in "the subjectivistic theory of probability", where "subjectiv" does not mean subjective (personal, private) at all. The distinction between the two concepts is made throughout the translation; the word "subjectivist" is reserved to mean "one who holds a subjectivistic theory".

"Cohérence" has been translated "coherent" following the usage of Shimony, Kemeny, and others. "Consistency" is used by some English and American authors, and is perfectly acceptable to de Finetti, but it is ambiguous (from the logician's point of view) because, applied to beliefs, it has another very precise and explicit meaning in formal logic. As the words are used in this translation, to say that a body of beliefs is "consistent" is to say (as in logic) that it contains no two beliefs that are contradictory. To say that in addition the body of beliefs is "coherent" is to say that the *degrees* of belief satisfy certain further conditions.

"Nombre aléatoire" has been translated as "random quantity". Although the phrase "random variable" is far more familiar to English-speaking mathematicians and philosophers, there are excellent reasons, as de Finetti points out, for making this substitution. I shall quote two of these reasons from de Finetti's correspondence. The first reason is that emphasized repeatedly in connection with the word "event". "While frequentists speak of an event as something admitting repeated 'trials', for those who take a subjectivistic (or logical) view of probability, any trial is a different 'event'. Likewise, for frequentists, a random variable X is something assuming different values in repeated 'trials', and only with this interpretation is the word 'variable' proper. For me any single trial gives a *random quantity*; there is nothing *variable*: the value is univocally indicated; it is only *unknown*; there is only *uncertainty* (for me, for somebody) about the unique value it will exhibit." The second objection de Finetti raises to the phrase "random variable" is one that is quite independent of any particular point of view with respect to probability. "Even with the statistical conception of probability, it is unjustifiably asymmetric to speak of random points, random functions, random vectors, etc., and of random variables when the 'variable' is a number or quantity; it would be consistent to say 'random variable' *always*, specifying, if necessary, 'random variable numbers', 'random variable points', 'random variable vectors', 'random variable functions', etc., as particular kinds of random variables."

"Loi" is used in the text both in the sense of "*theorem*" (as in "the law of large numbers") and in the sense of "distribution" (as in "normal law"). This is conventional French usage, and to some extent English and American usage has followed the French in this respect. But de Finetti himself now avoids the ambiguity by reserving the word "law" for the first sense (theoremhood) only, and by introducing the term "distribution" in a general sense to serve the function of the word "law" in its second sense. "Distribution" in this general sense may refer to specific distribution functions (as in "normal distribution"), the additive function of events P(E), or distributions that are not indicated by particular functions at all. I have attempted, with de Finetti's advice and suggestions, to introduce this distinction in translation.

Notation: The original notation has been followed closely, with the single exception of that for the "conditional event", E given A, which is written in the (currently) usual way, E|A. In the original this is written $\frac{E}{A}$. I have also substituted the conventional "v" for the original "+" in forming the expression denoting the alternation of two events.

Footnotes: Professor de Finetti has very kindly provided us with new notes that give some indication of the changes that have occurred in his thinking since he wrote this paper, or which clarify points which have, since the original writing, appeared to need clarification. These new notes are indicated by italic letters; the numbered footnotes appeared in the original work.

FOREWORD

In the lectures which I had the honor to give at the Institut Henri Poincaré the second, third, eighth, ninth, and tenth of May 1935, the text of which is reproduced in the pages that follow, I attempted to give a general view of two subjects which particularly interest me, and to clarify the delicate relationship that unites them. There is the question, on the one hand, of the definition of probability (which I consider a purely subjective entity) and of the meaning of its laws, and, on the other hand, of the concepts and of the theory of "exchangeable" events and random quantities; the link between the two subjects lies in the fact that the latter theory provides the solution of the problem of inductive reasoning for the most typical case, according to the subjectivistic conception of probability (and thus clarifies, in general, the way in which the problem of induction is posed). Besides, even if this were not so, that is to say, even if the subjective point of view which we have adopted were not accepted, this theory would have no less validity and would still be an interesting chapter in the theory of probability.

The exposition is divided into six chapters, of which the first two deal with the first question, the following two with the second, and of which the last two examine the conclusions that can be drawn. The majority of the questions treated here have been dealt with, sometimes in detail, sometimes briefly, but always in a fragmentary way,[1] in my earlier works. Among these, those which treat questions studied or touched upon in these lectures are indicated in the bibliography.[2]

For more complete details concerning the material in each of these chapters, I refer the reader to the following publications.

Chapter I. The logic of the probable: [26], [34].
II. The evaluation of probability: [49], [63], [70].
III. Exchangeable events: [29], [40].
IV. Exchangeable random quantities: [46], [47], [48].
V. Reflections on the notion of exchangeability: [51], [62].
VI. Observation and prediction: [32], [36], [62].

(1) A more complete statement of my point of view, in the form of a purely critical and philosophical essay, without formulas, is to be found in [32].

(2) See page 156; the numbers in boldface type refer always to this list (roman numerals for the works of other authors; arabic numerals for my own, arranged by general chronological order).

Each of these chapters constitutes one of the five lectures,[3] with the exception of Chapters IV and V, which correspond to the fourth, in which the text has been amplified in order to clarify the notion used there of integration in function space. The text of the other lectures has not undergone any essential modifications beyond a few improvements, for example, at the beginning of Chapter III, where, for greater clarity, the text has been completely revised. For these revisions, I have profited from the valuable advice of MM. Fréchet and Darmois, who consented to help with the lectures, and of M. Castelnuovo, who read the manuscript and its successive modifications several times; the editing of the text has been reviewed by my colleague M. V. Carmona and by M. Al. Proca, who suggested to me a number of stylistic changes. For their kind help I wish to express here my sincere appreciation. Finally, I cannot end these remarks without again thanking the director and the members of the governing committee of the Institut Henri Poincaré for the great honor they have done me by inviting me to give these lectures in Paris.

Trieste, December 19, 1936

(3) Their titles are those of the six chapters, with the exception of Chapter V.

INTRODUCTION

Henri Poincaré, the immortal scientist whose name this institute honors, and who brought to life with his ingenious ideas so many branches of mathematics, is without doubt also the thinker who attributed the greatest domain of application to the theory of probability and gave it a completely essential role in scientific philosophy. "Predictions," he said, "can only be probable. However solidly founded a prediction may appear to us, we are never absolutely sure that experience will not refute it." The calculus of probability rests on "an obscure instinct, which we cannot do without; without it science would be impossible, without it we could neither discover a law nor apply it." "On this account all the sciences would be but unconscious applications of the calculus of probability; to condemn this calculus would be to condemn science entirely."[1]

Thus questions of principle relating to the significance and value of probability cease to be isolated in a particular branch of mathematics and take on the importance of fundamental epistemological problems.

Such questions evidently admit as many different answers as there are different philosophical attitudes; to give one answer does not mean to say something that can convince and satisfy everybody, but familiarity with one particular point of view can nevertheless be interesting and useful even to those who are not able to share it. The point of view I have the honor of presenting here may be considered the extreme of subjectivistic solutions; the link uniting the diverse researches that I propose to summarize is in fact the principal common goal which is pursued in all of them, beyond other, more immediate and concrete objectives; this goal is that of bringing into the framework of the subjectivistic conception and of explaining even the problems that seem to refute it and are currently invoked against it. The aim of the first lecture will be to show how the logical laws of the theory of probability can be rigorously established within the subjectivistic point of view; in the others it will be seen how, while refusing to admit the existence of an objective meaning and value for probabilities, one can get a clear idea of the reasons, themselves subjective, for which in a host of problems the subjective judgments of diverse normal individuals not only do not differ essentially from each other, but even coincide exactly. The

(1) [XXVIII], p. 183, 186.

simplest cases will be the subject of the second lecture; the following lectures will be devoted to the most delicate question of this study: that of understanding the subjectivistic explanation of the use we make of the results of observation, of past experience, in our predictions of the future.

This point of view is only one of the possible points of view, but I would not be completely honest if I did not add that it is the only one that is not in conflict with the logical demands of my mind. If I do not wish to conclude from this that it is "true", it is because I know very well that, as paradoxical as it seems, nothing is more subjective and personal than this "instinct of that which is logical" which each mathematician has, when it comes to the matter of applying it to questions of principle.

CHAPTER I

The Logic of the Probable

Let us consider the notion of probability as it is conceived by all of us in everyday life. Let us consider a well-defined event and suppose that we do not know in advance whether it will occur or not; the doubt about its occurrence to which we are subject lends itself to comparison, and, consequently, to gradation. If we acknowledge only, first, that one uncertain event can only appear to us (a) equally probable, (b) more probable, or (c) less probable than another; second, that an uncertain event always seems to us more probable than an impossible event and less probable than a necessary event; and finally, third, that when we judge an event E′ more probable than an event E, which is itself judged more probable than an event E″, the event E′ can only appear more probable than E″ (transitive property), it will suffice to add to these three evidently trivial axioms a fourth, itself of a purely qualitative nature, in order to construct rigorously the whole theory of probability. This fourth axiom tells us that inequalities are preserved in logical sums: if E is incompatible with E_1 and with E_2, then $E_1 \vee E$ will be more or less probable than $E_2 \vee E$, or they will be equally probable, according to whether E_1 is more or less probable than E_2, or they are equally probable. More generally, it may be deduced from this[2] that

(2) See [34], p. 321, note 1.

two inequalities, such as

$$E_1 \text{ is more probable than } E_2,$$

$$E_1' \text{ is more probable than } E_2',$$

can be added to give

$$E_1 \vee E_1' \text{ is more probable than } E_2 \vee E_2',$$

provided that the events added are incompatible with each other (E_1 with E_1', E_2 with E_2'). It can then be shown that when we have events for which we know a subdivision into possible cases that we judge to be equally probable, the comparison between their probabilities can be reduced to the purely arithmetic comparison of the ratio between the number of favorable cases and the number of possible cases (not because the judgment then has an objective value, but because everything substantial and thus subjective is already included in the judgment that the cases constituting the division are equally probable). This ratio can then be chosen as the appropriate index to measure a probability, and applied in general, even in cases other than those in which one can effectively employ the criterion that governs us there. In these other cases one can evaluate this index by comparison: it will be in fact a number, uniquely determined, such that to numbers greater or less than that number will correspond events respectively more probable or less probable than the event considered. Thus, while starting out from a purely qualitative system of axioms, one arrives at a quantitative measure of probability, and then at the theorem of total probability which permits the construction of the whole calculus of probabilities (for conditional probabilities, however, it is necessary to introduce a fifth axiom: see note 8, p. 109).

One can, however, also give a direct, quantitative, numerical definition of the degree of probability attributed by a given individual to a given event, in such a fashion that the whole theory of probability can be deduced immediately from a very natural condition having an obvious meaning. It is a question simply of making mathematically precise the trivial and obvious idea that the degree of probability attributed by an individual to a given event is revealed by the conditions under which he would be disposed to bet on that event.[3] The axiomatization whose general outline we have

(3) Bertrand ([1], p. 24) beginning with this observation, gave several examples of subjective probabilities, but only for the purpose of contrasting them with "objective probabilities". The subjectivistic theory has been developed according to the scheme of bets in the exposition (Chap. I and II) in my first paper of 1928 on this subject. This was not published in its original form, but was summarized or partially developed in [27], [34], [35], etc.

just indicated above has the advantage of permitting a deeper and more detailed analysis, of starting out with only qualitative notions, and of eliminating the notion of "money", foreign to the question of probability, but which is required to talk of stakes; however, once it has been shown that one can overcome the distrust that is born of the somewhat too concrete and perhaps artificial nature of the definition based on bets, the second procedure is preferable, above all for its clarity.

Let us suppose that an individual is obliged to evaluate the rate p at which he would be ready to exchange the possession of an arbitrary sum S (positive or negative) dependent on the occurrence of a given event E, for the possession of the sum pS; we will say by definition that this number p is the measure of the degree of probability attributed by the individual considered to the event E, or, more simply, that p is the probability of E (according to the individual considered; this specification can be implicit if there is no ambiguity).[a]

Let us further specify that, in the terminology that I believe is suitable to follow, an event is always a singular fact; if one has to consider several trials, we will never say "trials of the same event" but "trials of the same phenomenon" and each "trial" will be one "event". The point is obviously not the choice of terms: it is a question of making precise that, according to us, one has no right to speak of the "probability of an event" if one understands by "event" that which we have called a "phenomenon"; one can only do this if it is a question of one specific "trial".[4]

Single case

(4) This same point of view has been taken by von Kries [XIX]; see [65], [70], and, for the contrary point of view, see [XXV].

(*a*) Such a formulation could better, like Ramsey's, deal with expected *utilities*; I did not know of Ramsey's work before 1937, but I was aware of the difficulty of money bets. I preferred to get around it by considering sufficiently small stakes, rather than to build up a complex theory to deal with it. I do not remember whether I failed to mention this limitation to small amounts inadvertently or for some reason, for instance considering the difficulty overcome in the artificial situation of compulsory choice.

Another shortcoming of the definition—or of the device for making it operational—is the possibility that people accepting bets against our individual have better information than he has (or know the outcome of the event considered). This would bring us to game-theoretic situations.

Of course, a device is always imperfect, and we must be content with an idealization. A better device (in this regard) is that mentioned in B. de Finetti and L. J. Savage, "Sul modo di scegliere le probabilità iniziali," *Biblioteca del Metron*, S. C. Vol. 1, pp. 81–147 (English summary pp. 148–151), and with some more detail in B. de Finetti, "Does it make sense to speak of 'good probability appraisers'?" *The Scientist Speculates: An anthology of partly-baked ideas*, Gen. Ed. I. J. Good, Heinemann, London, 1962. This device will be fully presented by the same authors in a paper in preparation.

This being granted, once an individual has evaluated the probabilities of certain events, two cases can present themselves: either it is possible to bet with him in such a way as to be assured of gaining, or else this possibility does not exist. In the first case one clearly should say that the evaluation of the probabilities given by this individual contains an incoherence, an intrinsic contradiction; in the other case we will say that the individual is coherent.[b] It is precisely this condition of coherence which constitutes the sole principle from which one can deduce the whole calculus of probability: this calculus then appears as a set of rules to which the subjective evaluation of probability of various events by the same individual ought to conform if there is not to be a fundamental contradiction among them.

Let us see how to demonstrate, on this view, the theorem of total probability: it is an important result in itself, and also will clarify the point of view followed. Let $E_1, E_2, \ldots, E_n$ be incompatible events, of which one (and one only) must occur (we shall say: a *complete* class of incompatible events), and let $p_1, p_2, \ldots, p_n$ be their probabilities evaluated by a given individual; if one fixes the stakes (positive or negative) $S_1, S_2, \ldots, S_n$, the gains in the n possible cases will be the difference between the stake of the bet won and the sum of the n paid outlays.

$$G_h = S_h - \sum_1^n {}_i\, p_i S_i$$

By considering the S_h as unknowns, one obtains a system of linear equations with the determinant

$$\begin{vmatrix} 1 - p_1 & -p_2 & \cdots & -p_n \\ -p_1 & 1 - p_2 & \cdots & -p_n \\ \cdots & \cdots & \cdots & \cdots \\ -p_1 & -p_2 & \cdots & 1 - p_n \end{vmatrix} = 1 - (p_1 + p_2 + \cdots + p_n);$$

if this determinant is not zero, one can fix the S_h in such a way that the G_h have arbitrary values, in particular, all positive, contrary to the condition of coherence; consequently coherence obliges us to impose the condition

(*b*) To speak of coherent or incoherent (consistent or inconsistent) individuals has been interpreted as a criticism of people who do not accept a specific behavior rule. Needless to say, this is meant only as a technical distinction. At any rate, it is better to speak of coherence (consistency) of probability evaluations rather than of individuals, not only to avoid this charge, but because the notion belongs strictly to the evaluations and only indirectly to the individuals. Of course, an individual may make mistakes sometimes, often without meriting contempt.

$p_1 + p_2 \cdots + p_n = 1$.[c] This necessary condition for coherence is also sufficient, because, if it is satisfied, one has identically (whatever be the stakes S_h)

$$\sum_{1}^{n}{}_h\, p_h G_h = 0$$

and the G_h can never, in consequence, all be positive.

Thus one has the theorem of total probabilities in the following form: *in a complete class of incompatible events, the sum of the probabilities must be equal to 1*. The more general form, *the probability of the logical sum of n incompatible events is the sum of their probabilities*, is only an immediate corollary.

However, we have added that the condition is also sufficient; it is useful to make the sense of this assertion a little clearer, for in a concrete case one can throw into clear relief the distinction, fundamental from this point of view, between the logic of the probable and judgments of probability. In saying that the condition is sufficient, we mean that, a complete class of incompatible events $E_1, E_2, \ldots, E_n$ being given, all the assignments of probability that attribute to $p_1, p_2, \ldots, p_n$ any values whatever, which are non-negative and have a sum equal to unity, are admissible assignments: each of these evaluations corresponds to a coherent opinion, to an opinion legitimate in itself, and every individual is free to adopt that one of these opinions which he prefers, or, to put it more plainly, that which he *feels*. The best example is that of a championship where the spectator attributes to each team a greater or smaller probability of winning according to his own judgment; the theory cannot reject *a priori* any of these judgments unless the sum of the probabilities attributed to each team is not equal to unity. This arbitrariness, which any one would admit in the above case, exists also, according to the conception which we are maintaining, in all other domains, including those more or less vaguely defined domains in which the various objective conceptions are asserted to be valid.

Because of this arbitrariness, the subject of the calculus of probabilities is no longer a single function $\mathbf{P}(E)$ of events E, that is to say, their probability considered as something objectively determined, but the set of all functions $\mathbf{P}(E)$ corresponding to admissible opinions. And when a *calculation* of the probability $\mathbf{P}(E)$ of an event E is wanted, the statement of the problem is to be made precise in this sense: calculate the value that one is

(c) Of course the proof might have been presented in an easier form by considering simply the case of $S_1 = S_2 = \cdots = S_n = S$ (as I did in earlier papers). On this occasion I preferred a different proof which perhaps gives deeper insight.

obliged to attribute to the event E if one wants to remain in the domain of coherence, after having assigned definite probabilities to the events constituting a certain class $\mathscr{E}$. Mathematically the function **P** is adopted over the set $\mathscr{E}$, and one asks what unique value or what set of values can be attributed to **P**(E) without this extension of **P** making an incoherence appear.

It is interesting to pose the following general question: what are the events E for which the probability is determined by the knowledge of the probabilities attributed to the events of a given class $\mathscr{E}$? We are thus led to introduce the notion (which I believe novel) of "linearly independent events" [26]. Let $E_1, E_2, \ldots, E_n$ be the events of $\mathscr{E}$. Of these n events some will occur, others will not; there being 2^n subclasses of a class of n elements (including the whole class $\mathscr{E}$ and the empty class), there will be at most 2^n possible cases $C_1, C_2, \ldots, C_s$ ($s \leq 2^n$) which we call, after Boole, "constituents". ("At most", since a certain number of combinations may be impossible.)[5] Formally, the C_h are the events obtained by starting with the logical product $E_1 \cdot E_2 \cdot \ldots \cdot E_n$ and replacing any group of E_i by the contrary events (negations) $\sim E_i$ (or, in brief notation, $\bar{E}_i$). The constituents form a complete class of incompatible events; the E_i are logical sums of constituents, and the events which are the sums of constituents are the only events logically dependent on the E_i, that is, such that one can always say whether they are true or false when one knows, for each event $E_1, \ldots, E_n$, if it is true or false.

To give the probability of an event E_i means to give the sum of the probabilities of its constituents

$$c_{i_1} + c_{i_2} + \cdots + c_{i_h} = p_i;$$

the probabilities of $E_1, \ldots, E_n$ being fixed, one obtains n equations of this type, which form, with the equation $c_1 + c_2 + \cdots + c_s = 1$, a system of $n + 1$ linear equations relating the probabilities c_h of the constituents. It may be seen that, E being an event logically dependent on $E_1, \ldots, E_n$, and thus a logical sum of constituents $E = C_{h_1} \vee C_{h_2} \vee \cdots \vee C_{h_k}$, its probability

$$p = c_{h_1} + c_{h_2} + \cdots + c_{h_k}$$

is uniquely determined when this equation is linearly dependent on the preceding system of equations. Observe that this fact does not depend on the function **P**, but only on the class $\mathscr{E}$ and the event E and can be expressed

(5) These notions are applied to the calculus of probability in Medolaghi [XXIV].

by saying that E is *linearly dependent* on $\mathscr{E}$, or—what comes to the same thing if the E_i are linearly independent—that $E_1, E_2, \ldots, E_n$ and E are linearly related among themselves.

The notion of linear independence thus defined for events is perfectly analogous to the well-known geometrical notion, and enjoys the same properties; instead of this fact being demonstrated directly, it can quickly be made obvious by introducing a geometrical representation which makes a point correspond to each event, and the notion of geometrical "linear independence" correspond to the notion of logical "linear independence". The representation is as follows: the constituents C_h are represented by the apexes A_h of a simplex in a space of $s - 1$ dimensions, the event which is the sum of k constituents by the center of gravity of the k corresponding apexes given a mass k, and finally, the certain event (the logical sum of all the s constituents) by the center 0 of the simplex, given a mass s.

This geometric representation allows us to characterize by means of a model the set of all possible assignments of probability. We have seen that a probability function $\mathbf{P}(E)$ is completely determined when one gives the relative values of the constituents, $c_1 = \mathbf{P}(C_1)$, $c_2 = \mathbf{P}(C_2), \ldots, c_s = \mathbf{P}(C_s)$, values which must be non-negative and have a sum equal to unity. Let us now consider the linear function f which takes the values c_h on the apexes A_h; at the point A, the center of gravity of $A_{h_1}, A_{h_2}, \ldots, A_{h_k}$, it obviously takes the value $f(A) = (1/k)(c_{h_1} + c_{h_2} + \cdots + c_{h_k})$, while the probability $\mathbf{P}(E)$ of the event E, the logical sum of the constituents C_{h_1}, $C_{h_2}, \ldots, C_{h_k}$ will be $c_{h_1} + c_{h_2} + \cdots + c_{h_k}$. We have, then, in general, $\mathbf{P}(E) = k \cdot f(A)$: the probability of an event E is the value of f at its representative point A, multiplied by the mass k; one could say that it is given as the value of f for the point A endowed with a mass k, writing $\mathbf{P}(E) = f(k \cdot A)$.[d] The center 0 corresponding to the certain event, one has in particular $1 = f(s \cdot 0) = s \cdot f(0)$, that is, $f(0) = (1/s)$.

It is immediately seen that the possible assignments of probability correspond to all the linear functions of the space that are non-negative on the simplex and have the value $1/s$ at the origin; such a function f being characterized by the hyperplane $f = 0$, assignments of probability correspond biunivocally to the hyperplanes which do not cut the simplex. It may be seen that the probability $\mathbf{P}(E) = f(k \cdot A)$ is the moment of the given mass point kA (distance $\times$ mass) relative to the hyperplane $f = 0$ (taking

(*d*) The notion of "weighted point", or "geometrical formation of the first kind", belongs to the geometrical approach and notations of Grassmann-Peano, to which the Italian school of vector calculus adheres.

as unity the moment of $s0$). If, in particular, the s constituents are equally probable, the hyperplane goes to infinity.

By giving the value that it takes on a certain group of points, a linear function f is defined for all those points linearly dependent on them, but it remains undetermined for linearly independent points: the comparison with the above definition of linearly dependent events thus shows, as we have said, that the linear dependence and independence of events means dependence and independence of the corresponding points in the geometric representation. The two following criteria characterizing the linear dependence of events can now be deduced in a manner more intuitive than the direct way. In the system of barycentric coordinates, where $x_i = 1$, $x_j = 0$ $(j \neq i)$ represents the point A_i, the coordinates of the center of gravity of $A_{h_1}, A_{h_2}, \ldots, A_{h_k}$ having a mass k will be

$$x_{h_1} = x_{h_2} = \cdots = x_{h_k} = 1, \qquad x_j = 0\,(j \neq h_1, h_2, \ldots, h_k);$$

the sum of the constituents can thus be represented by a symbol of s digits, 1 or 0 (for example, the sum $C_1 \vee C_3$ by $10100 \cdots 0$). Events are linearly independent when the matrix of the coordinates of the corresponding points and of the center 0 is of less than maximum rank, the rows of this matrix being the expressions described above corresponding to the events in question and—for the last line which consists only of 1's—the certain event. The other condition is that the events are linearly dependent when a coefficient can be assigned to each of them in such a way that in every possible case the sum of the coefficients of the events that occur always has the same value. If, in fact, the points corresponding to the given events and the point 0 are linearly dependent, it is possible to express 0 by a linear combination of the others, and this means that there exists a combination of bets on these events equivalent to a bet on the certain event.

An assignment of probability can be represented not only by the hyperplane $f = 0$ but also by a point not exterior to the simplex, conjugate to the hyperplane,[6] and defined as the center of gravity of s points having masses

(6) In the polarity $f\begin{pmatrix} x \\ - \\ y \end{pmatrix} = \Sigma x_i y_i = 0$ (barycentric coordinates). It is convenient here, having to employ metric notions, to consider the simplex to be equilateral. It can be specified, then, that it is a question of the polarity relative to the imaginary hypersphere $\Sigma x_i^2 = 0$, and that it makes correspond to any point A whatever the hyperplane orthogonal to the line AO passing through the point A′ corresponding to A in an inversion about the center O. In vectorial notation, the hyperplane is the locus of all points Q such that the scalar product $(A - O) \cdot (Q - O)$ gives $-R^2$, where $R = l/\sqrt{2s}$, l being the length of each edge of the simplex.

proportional to the probabilities of the events (constituents) that they represent. This representation is useful because the simplex gives an intuitive image of the space of probability laws, and above all because linear relations are conserved. The ∞^{s-1} admissible assignments of probability can in fact be combined linearly: if $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_m$ are probability functions, $\mathbf{P} = \Sigma\lambda_i\mathbf{P}_i$ $(\lambda_i \geq 0, \Sigma\lambda_i = 1)$ is also, and the point representing $\mathbf{P}$ is given by the same relation, i.e., it is the center of gravity of the representative points of $\mathbf{P}_1, \ldots, \mathbf{P}_m$ with masses $\lambda_1, \ldots, \lambda_m$; the admissible assignments of probability constitute then, as do the non-exterior points of the simplex, a closed, convex set. This simple remark allows us to complete our results quickly, by specifying the lack of determination of the probability of an event which remains when the event is linearly independent of certain others after the probability of the others has been fixed. It suffices to note that by fixing the value of the probability of certain events, one imposes linear conditions on the function $\mathbf{P}$; the functions $\mathbf{P}$ that are still admissible also constitute a closed, convex set. From this one arrives immediately at the important conclusion that when the probability of an event E is not uniquely determined by those probabilities given, the admissible numbers are all those numbers in a closed interval $p' \leq p \leq p''$. If E′ and E″ are respectively the sum of all the constituents contained in E or compatible with E, p' will be the smallest value admissible for the probability of E′ and p'' the greatest for E″.

When the events considered are infinite in number, our definition introduces no new difficulty: $\mathbf{P}$ is a probability function for the infinite class of events $\mathscr{E}$ when it is a probability function for all finite subclasses of $\mathscr{E}$. This conclusion implies that the theorem of total probability cannot be extended to the case of an infinite or even denumerable number of events[7]; a discussion of this subject would carry us too far afield.

σ- ...tivity

We have yet to consider the definition of conditional probabilities and the demonstration of the multiplication theorem for probabilities. Let there be two events E′ and E″; we can bet on E′ and condition this bet on E″: if E″ does not occur, the bet will be annulled; if E″ does occur, it will be won or lost according to whether E′ does or does not occur. One can consider, then, the "conditional events" (or "tri-events"), which are the events of a three-valued logic: this "tri-event", "E′ conditioned on E″", E′ | E″, is the logical entity capable of having three values: *true* if E″ and E′ are true; *false* if E″ is true and E′ false; *void* if E″ is false. It is clear that two tri-events $E_1' \mid E_1''$ and $E_2' \mid E_2''$ are equal if $E_1'' = E_2''$ and $E_1'E_1'' = E_2'E_2''$; we will say that E′ | E″ is written in *normal* form if E′ → E″, and it

(7) See [16], [24], [X], [28], [XI], [64].

may be seen that any tri-event can be written in a single way in normal form: $E'E'' \mid E''$. We could establish for the tri-events a three-valued logic, perfectly analogous to ordinary logic [64], but this is not necessary for the goal we are pursuing.

Let us define the probability p of E' conditioned on E'' by the same condition relative to bets, but in this case we make the convention that the bet is to be called off if E'' does not happen. The bet can then give three different results: if S is the stake, outlay paid will be pS, and the gain $(1-p)S$, $-pS$, or 0 according to whether $E' \mid E''$ will be true, false, or void, for in the first case one gains the stake and loses the outlay, in the second one loses the outlay, and in the last the outlay is returned (if $S < 0$ these considerations remain unchanged; we need only to change the terminology of debit and credit). Let us suppose that $E' \rightarrow E''$, and let p' and p'' be the probabilities of E' and E'': we will show that for coherence we must have $p' = p \cdot p''$. If we make three bets: one on E' with the stake S', one on E'' with the stake S'', and one on $E' \mid E''$ with the stake S, the total gain corresponds, in the three possible cases, to

E': $\qquad G_1 = (1-p') \cdot S' + (1-p'') \cdot S'' + (1-p)S;$

E'' and not E': $\qquad G_2 = -p'S' + (1-p'')S'' - pS;$

not E'': $\qquad G_3 = -p'S' - p''S''.$

If the determinant

$$\begin{vmatrix} 1-p' & 1-p'' & 1-p \\ -p' & 1-p'' & -p \\ -p' & -p'' & 0 \end{vmatrix} = p' - pp''$$

is not zero, one can fix S, S', and S'' in such a way that the G's have arbitrary values, in particular, all positive, and that implies a lack of coherence. Therefore $p' = pp''$, and, in general, if E' does not imply E'', this will still be true if we consider $E'E''$ rather than E': we thus have the multiplication theorem for probabilities[8]

$$\mathbf{P}(E' \cdot E'') = \mathbf{P}(E') \cdot \mathbf{P}(E'' \mid E'). \tag{1}$$

(8) This result, which, in the scheme of bets, can be deduced as we have seen from the definition of coherence, may also be expressed in a purely qualitative form, such as the following, which may be added as a fifth axiom to the preceding four (see p. 100–101): If E' and E'' are contained in E, $E' \mid E$ is more or less probable than (or is equal in probability to) $E'' \mid E$, according to whether E' is more or less probable than (or equal in probability to) E''.

The condition is not only necessary, but also sufficient, in the same sense as in the case of the theorem of total probability. According to whether an individual evaluates $\mathbf{P}(E' \mid E'')$ as greater than, smaller than, or equal to $\mathbf{P}(E')$, we will say that he judges the two events to be in a positive or negative correlation, or as independent: it follows that the notion of independence or dependence of two events has itself only a subjective meaning, relative to the particular function $\mathbf{P}$ which represents the opinion of a given individual.

We will say that $E_1, E_2, \ldots, E_n$ constitute a class of independent events if each of them is independent of any product whatever of several others of these events (pairwise independence, naturally, does not suffice); in this case the probability of a logical product is the product of the probabilities, and, the constituents themselves being logical products, the probability of any event whatever logically dependent on $E_1, \ldots, E_n$ will be given by an algebraic function of $p_1, p_2, \ldots, p_n$.

We obtain as an immediate corollary of (1), Bayes's theorem, in the form[9]

$$\mathbf{P}(E'' \mid E') = \frac{\mathbf{P}(E'') \cdot \mathbf{P}(E' \mid E'')}{\mathbf{P}(E')}, \tag{2}$$

which can be formulated in the following particularly meaningful way: The probability of E′, relative to E″, is modified in the same sense and in the same measure as the probability of E″ relative to E′.

In what precedes I have only summarized in a quick and incomplete way some ideas and some results with the object of clarifying what ought to be understood, from the subjectivistic point of view, by "logical laws of probability" and the way in which they can be proved. These laws are the conditions which characterize coherent opinions (that is, opinions admissible in their own right) and which distinguish them from others that are intrinsically contradictory. The choice of one of these admissible opinions from among all the others is not objective at all and does not enter into the logic of the probable; we shall concern ourselves with this problem in the following chapters.

(9) It is also found expressed in this form in Kolmogorov [XVII].

CHAPTER II

The Evaluation of a Probability

The notion of probability which we have described is without doubt the closest to that of "the man in the street"; better yet, it is that which he applies every day in practical judgments. Why should science repudiate it? What more adequate meaning could be discovered for the notion?

It could be maintained, from the very outset, that in its usual sense probability cannot be the object of a mathematical theory. However, we have seen that the rules of the calculus of probability, conceived as conditions necessary to ensure coherence among the assignments of probability of a given individual, can, on the contrary, be developed and demonstrated rigorously. They constitute, in fact, only the precise expression of the rules of the logic of the probable which are applied in an unconscious manner, qualitatively if not numerically, by all men in all the circumstances of life.[e]

It can still be doubted whether this conception, which leaves each individual free to evaluate probabilities as he sees fit, provided only that the condition of coherence be satisfied, suffices to account for the more or less strict agreement which is observed among the judgments of diverse individuals, as well as between predictions and observed results. Is there, then, among the infinity of evaluations that are perfectly admissible in themselves, one particular evaluation which we can qualify, in a sense as yet unknown, as *objectively correct*? Or, at least, can we ask if a given evaluation is better than another?

There are two procedures that have been thought to provide an objective meaning for probability: the scheme of equally probable cases, and the

(*e*) Such a statement is misleading if, as unfortunately has sometimes happened, it is taken too seriously. It cannot be said that people compute according to arithmetic or think according to logic, unless it is understood that mistakes in arithmetic or in logic are very natural for all of us. It is still more natural that mistakes are common in the more complex realm of probability; nevertheless it seems correct to say that, fundamentally, people behave according to the rules of coherence even though they frequently violate them (just as it may be said that they accept arithmetic and logic). But in order to avoid frequent misunderstandings it is essential to point out that probability theory is not an attempt to describe actual behavior; its subject is coherent behavior, and the fact that people are only more or less coherent is inessential.

consideration of frequencies. Indeed it is on these two procedures that the evaluation of probability generally rests in the cases where normally the opinions of most individuals coincide. However, these same procedures do not oblige us at all to admit the existence of an objective probability; on the contrary, if one wants to stretch their significance to arrive at such a conclusion, one encounters well-known difficulties, which disappear when one becomes a little less demanding, that is to say, when one seeks not to eliminate but to make more precise the subjective element in all this. In other words, it is a question of considering the coincidence of opinions as a psychological fact; the reasons for this fact can then retain their subjective nature, which cannot be left aside without raising a host of questions of which even the sense is not clear. Thus in the case of games of chance, in which the calculus of probability originated, there is no difficulty in understanding or finding very natural the fact that people are generally agreed in assigning equal probabilities to the various possible cases, through more or less precise, but without doubt very spontaneous, considerations of symmetry. Thus the classical definition of probability, based on the relation of the number of favorable cases to the number of possible cases, can be justified immediately: indeed, if there is a complete class of n incompatible events, and if they are judged equally probable, then by virtue of the theorem of total probability each of them will necessarily have the probability $p = 1/n$ and the sum of m of them the probability m/n. A powerful and convenient criterion is thus obtained: not only because it gives us a way of calculating the probability easily when a subdivision into cases that are judged equally probable is found, but also because it furnishes a general method for evaluating by comparison any probability whatever, by basing the quantitative evaluation on purely qualitative judgments (equality or inequality of two probabilities). However this criterion is only applicable on the hypothesis that the individual who evaluates the probabilities judges the cases considered equally probable; this is again due to a subjective judgment for which the habitual considerations of symmetry which we have recalled can furnish psychological reasons, but which cannot be transformed by them into anything objective. If, for example, one wants to demonstrate that the evaluation in which all the probabilities are judged equal is alone "right", and that if an individual does not begin from it he is "mistaken", one ought to begin by explaining what is meant by saying that an individual who evaluates a probability judges "right" or that he is "mistaken". Then one must show that the conditions of symmetry cited imply necessarily that one must accept the hypothesis of equal probability if one does not want to be "mistaken". But any event whatever can only happen or not happen, and neither in one case nor in the other can one decide what would be the degree of doubt with which it would be "reasonable" or "right" to expect the event before knowing whether it has occurred or not.

Let us now consider the other criterion, that of frequencies. Here the problem is to explain its value from the subjectivistic point of view and to show precisely how its content is preserved. Like the preceding criterion, and like all possible criteria, it is incapable of leading us outside the field of subjective judgments; it can offer us only a more extended psychological analysis. In the case of frequencies this analysis is divided into two parts: an elementary part comprised of the relations between evaluations of probabilities and predictions of future frequencies, and a second, more delicate part concerning the relation between the observation of past frequencies and the prediction of future frequencies. For the moment we will limit ourselves to the first question, while admitting as a known psychological fact, whose reasons will be analyzed later, that one generally predicts frequencies close to those that have been observed.

The relation we are looking for between the evaluation of probabilities and the prediction of frequencies is given by the following theorem. Let E_1, $E_2, \ldots, E_n$ be any events whatever.[1] Let us assign the values $p_1, p_2, \ldots, p_n$ to their probabilities and the values $\omega_0, \omega_1, \ldots, \omega_n$, to the probabilities that zero, or only one, or two, etc., or finally, all these events will occur (clearly $\omega_0 + \omega_1 + \omega_2 + \cdots + \omega_n = 1$). For coherence, we must have:

$$\underbrace{p_1 + p_2 + \cdots + p_n}_{\text{Average fr.}} = \underbrace{0 \times \omega_0 + 1 \times \omega_1 + 2 \times \omega_2 + \cdots + n \times \omega_n}_{\text{Expected RF.}} \qquad \text{Algebra.}$$

or simply

$$\bar{p} = \bar{f} \tag{3}$$

where $\bar{p}$ indicates the arithmetic mean of the p_i, and $\bar{f}$ the mathematical

(1) In order to avoid a possible misunderstanding due to the divergence of our conception from some commonly accepted ones, it will be useful to recall that, in our terminology, an "event" is always a determinate singular fact. What are sometimes called *repetitions* or *trials* of the same event are for us distinct events. They have, in general, some common characteristics or symmetries which make it natural to attribute to them equal probabilities, but we do not admit any *a priori* reason which prevents us in principle from attributing to each of these trials $E_1, \ldots, E_n$ some different and absolutely arbitrary probabilities $p_1, \ldots, p_n$. In principle there is no difference for us between this case and the case of n events which are not analogous to each other; the analogy which suggests the name "trials of the same event" (we would say "of the same phenomenon") is not at all essential, but, at the most, valuable because of the influence it can exert on our psychological judgment in the sense of making us attribute equal or very nearly equal probabilities to the different events.

expectation of the frequency (that is to say of the random quantity which takes the values $0/n$, $1/n$, $2/n$, . . . , n/n according to whether 0, 1, 2, . . . , n of the E_i occur); we note that in this respect the notion of mathematical expectation has itself a subjective meaning, since it is defined only in relation to the given judgment which assigns to the $n + 1$ possible cases the probabilities ω_h.

This relation can be further simplified in some particular cases: if the frequency is known, the second member simply represents that value of the frequency; if one judges that the n events are equally probable, the first member is nothing but the common value of the probability. Let us begin with the case in which both simplifying assumptions are correct: there are n events, m are known to have occurred or to be going to occur, but we are ignorant of which, and it is judged equally probable that any one of the events should occur. The only possible evaluation of the probability in this case leads to the value $p = m/n$. If $m = 1$, this reduces to the case of n equally probable, incompatible possibilities.

If, in the case where the frequency is known in advance, our judgment is not so simple, the relation is still very useful to us for evaluating the n probabilities, for by knowing what their arithmetic mean has to be, we have a gross indication of their general order of magnitude, and we need only arrange to augment certain terms and diminish others until the relation between the various probabilities corresponds to our subjective judgment of the inequality of their respective chances. As a typical example, consider a secret ballot: one knows that among the n voters $A_1, A_2, \ldots, A_n$, one has m favorable ballots; one can then evaluate the probabilities $p_1, p_2, \ldots, p_n$ that the different voters have given a favorable vote, according to the idea one has of their opinions; in any case this evaluation must be made in such a way that the arithmetic mean of the p_i will be m/n.

When the frequency is not known, the equation relates two terms which both depend on a judgment of probability: the evaluation of the probabilities p_i is no longer bound by their average to something given objectively, but to the evaluation of other probabilities, the probabilities ω_h of the various frequencies. Still, it is an advantage not to have to evaluate exactly all the ω_h in order to apply the given relation to the evaluation of the probabilities p_i; a very vague estimation of a qualitative nature suffices, in fact, to evaluate $\bar{f}$ with enough precision. It suffices, for example, to judge as "not very probable" that the frequency differs noticeably from a certain value a, which is tantamount to estimating as very small the sum of all the ω_h for which $|h/n - a|$ is not small, to give approximately $\bar{f} = a$.

Once $\bar{f}$ has been evaluated, nothing is changed of what we said earlier concerning the case where the frequency is known: if the n events are judged equally probable, their common probability is $p = \bar{f}$; if that is not the case, then certain probabilities will be augmented or diminished in order that their arithmetic mean will be $p = \bar{f}$.

It is thus that one readily evaluates probabilities in most practical problems, for example, the probability that a given individual, let us say Mr. A, will die in the course of the year. If it is desired to estimate directly under these conditions what stakes (or insurance, as one would prefer to say in this case) seem to be equitable, this evaluation would seem to us to be affected with great uncertainty; the application of the criterion described above facilitates the estimation greatly. For this one must consider other events, for example, the deaths, during the year, of individuals of the same age and living in the same country as Mr. A. Let us suppose that among these individuals about 13 out of 1000 will die in a year; if, in particular, all the probabilities are judged equal, their common value is $p = 0.013$, and the probability of death for Mr. A is 0.013; if in general there are reasons which make the chances we attribute to their deaths vary from one individual to another, this average value of 0.013 at least gives us a base from which we can deviate in one direction or the other in taking account of the characteristics which differentiate Mr. A from other individuals.

This procedure has three distinct and successive phases: the first consists of the choice of a class of events including that which we want to consider; the second is the prediction of the frequency; the third is the comparison between the average probability of the single events and that of the event in question. Some observations in this regard are necessary in order to clarify the significance and value that are attributed to these considerations by subjectivists' point of view, and to indicate how these views differ from current opinion. Indeed, it is only the necessity of providing some clarification about these points before continuing that makes it indispensable to spend some little time on such an elementary question.

The choice of a class of events is in itself arbitrary; if one chooses "similar" events, it is only to make the application of the procedure easier, that is to say, to make the prediction of the frequency and the comparison of the various probabilities easier: but this restriction is not at all essential, and even if one admits it, its meaning is still very vague. In the preceding example, one could consider, not individuals of the same age and the same country, but those of the same profession, of the same height, of the same profession and town, etc., and in all these cases one could observe a noticeable enough similarity. Nothing prevents *a priori* the grouping of the event which interests us with any other events whatever. One can consider, for example, the death of Mr. A during the year as a *claim* in relation to all the policies of the company by which he is insured, comprising fire insurance, transport insurance, and others; from a certain point of view, one can still maintain that these events are "similar".

This is why we avoid expressions like "trials of the same event", "events which can be repeated", etc., and, in general, all the frequency considerations which presuppose a classification of events, conceived of as rigid and essential, into classes or collections or series. All classifications of this sort have only an auxiliary function and an arbitrary value.

The prediction of the frequency is based generally on the hypothesis that its value remains nearly constant: in our example, the conviction that the proportion of deaths is 13 per 1000 can have its origin in the observation that in the course of some years past the mortality of individuals of the same kind was in the neighborhood of 13/1000. The reasons which justify this way of predicting could be analyzed further; for the moment it suffices to assume that in effect our intuition leads us to judge thus. Let us remark that such a prediction is generally the more difficult the narrower the class considered.

On the other hand, the comparison of the different probabilities is more difficult in the same proportion the events are more numerous and less homogeneous: the difficulty is clearly reduced to a minimum when the events appear to us equally probable. In practice one must attempt to reconcile as best one can these opposing demands, in order to achieve the best application of the two parts of the procedure: it is only as a function of these demands that the class of events considered can be chosen in a more or less appropriate fashion.

An illustration will render these considerations still clearer. If one must give an estimate of the thickness of a sheet of paper, he can very easily arrive at it by estimating first the thickness of a packet of n sheets in which it is inserted, and then by estimating the degree to which the various sheets have the same thickness. The thickness can be evaluated the more easily the larger the packet; the difficulty of the subsequent comparison of the sheets is on the contrary diminished if one makes the packet thinner by saving only those sheets judged to have about the same thickness as the sheet that interests us.

Thus the criterion based on the notion of frequency is reduced, like that based on equiprobable events, to a practical method for linking certain subjective evaluations of probability to other evaluations, themselves subjective, but preferable either because more accessible to direct estimation, or because a rougher estimate or even one of a purely qualitative nature suffices for the expected conclusions. *A priori*, when one accepts the subjectivistic point of view, such ought to be the effective meaning and the value of any criterion at all.

In the case of predictions of frequencies, one only relates the evaluation of the p_i to that of the ω_h and to a comparison between the p_i; the estimation of the ω_h does not need to come up to more than a rough approximation, such as suffices to determine the p_i closely enough. It must be remarked nevertheless that this prediction of the frequency is nothing else than an evaluation of the ω_h; it is not a prophecy which one can call correct if the frequency is equal or close to $\bar{f}$, and false in the contrary case. All the frequencies $0/n, 1/n, 2/n, \ldots, n/n$ are possible, and whatever the realized frequency may be, nothing can make us right or wrong if our actual judgment is to attribute to these $n + 1$ cases the probabilities ω_h, leading to a certain value

$$\bar{p} = \bar{f} = \frac{\omega_1 + 2\omega_2 + 3\omega_3 + \cdots + n\omega_n}{n} \tag{3}$$

It is often thought that these objections may be escaped by observing that the impossibility of making the relations between probabilities and frequencies precise is analogous to the practical impossibility that is encountered in all the experimental sciences of relating exactly the abstract notions of the theory and the empirical realities.[2] The analogy is, in my view, illusory: in the other sciences one has a theory which asserts and predicts with certainty and exactitude what would happen if the theory were completely exact; in the calculus of probability it is the theory itself which obliges us to admit the possibility of all frequencies. In the other sciences the uncertainty flows indeed from the imperfect connection between the theory and the facts; in our case, on the contrary, it does not have its origin in this link, but in the body of the theory itself [32], [65], [IX]. No relation between probabilities and frequencies has an empirical character, for the observed frequency, whatever it may be, is always compatible with all the opinions concerning the respective probabilities; these opinions, in consequence, can be neither confirmed nor refuted, once

(2) This point of view is maintained with more or less important variations in most modern treatises, among others those of Castelnuovo [VI], Fréchet-Halbwachs [XII], Lévy [XX], von Mises [XXV].

it is admitted that they contain no categorical assertion such as: such and such an event *must* occur or *can not* occur.

This last consideration may seem rather strange if one reflects that the prediction of a future frequency is generally based on the observation of those past; one says, "we will correct" our initial opinions if "experience refutes them." Then isn't this instinctive and natural procedure justified? Yes; but the way in which it is formulated is not exact, or more precisely, is not meaningful. It is not a question of "correcting" some opinions which have been "refuted": it is simply a question of substituting for the initial evaluation of the probability the value of the probability which is conditioned on the occurrence of facts which have already been observed; this probability is a completely different thing from the other, and their values can very well not coincide without this non-coincidence having to be interpreted as the "correction of a refuted opinion".

The explanation of the influence exercised by experience on our future predictions, developed according to the ideas that I have just expounded, constitutes the point that we have left aside in the analysis of the criterion based on frequencies. This development will be the subject of the following chapters, in which we will make a more detailed study of the most typical case in this connection: the case of exchangeable events, and, in general, of any exchangeable random quantities or elements whatever. This study is important for the development of the subjectivistic conception, but I hope that the mathematical aspect will be of some interest in itself, independently of the philosophical interpretation; in fact, exchangeable random quantities and exchangeable events are characterized by simple and significant conditions which can justify by themselves a deep study of the problems that arise in connection with them.

CHAPTER III
Exchangeable Events

Why are we obliged in the majority of problems to evaluate a probability according to the observation of a frequency? This is a question of the relations between the observation of past frequencies and the prediction of future frequencies which we have left hanging, but which presents itself anew under a somewhat modified form when we ask ourselves if a prediction of frequency can be in a certain sense confirmed or refuted by experience. The question we pose ourselves now includes in reality the problem of reasoning by induction. Can this essential problem, which has never received a satisfactory solution up to now, receive one if we employ the conception of subjective probability and the theory which we have sketched?

In order to fix our ideas better, let us imagine a concrete example, or rather a concrete interpretation of the problem, which does not restrict its generality at all. Let us suppose that the game of heads or tails is played with a coin of irregular appearance. The probabilities of obtaining "heads" on the first, the second, the hth toss, that is to say, the probabilities $\mathbf{P}(E_1), \mathbf{P}(E_2), \ldots, \mathbf{P}(E_h), \ldots$ of the events $E_1, E_2, \ldots, E_h, \ldots$ consisting of the occurrence of heads on the different tosses, can only be evaluated by calculating *a priori* the effect of the apparent irregularity of the coin.

It will be objected, no doubt, that in order to get to this point, that is to say, to obtain the "correct" probabilities of future trials, we can utilize the results obtained in the previous trials: it is indeed in this sense that—according to the current interpretation—we "correct" the evaluation of $\mathbf{P}(E_{n+1})$ after the observation of the trials which have, or have not, brought about $E_1, E_2, \ldots, E_n$. Such an interpretation seems to us unacceptable, not only because it presupposes the objective existence of unknown probabilities, but also because it cannot even be formulated correctly: indeed the probability of E_{n+1} evaluated with the knowledge of a certain result, A, of the n preceding trials is no longer $\mathbf{P}(E_{n+1})$ but $\mathbf{P}(E_{n+1} \mid A)$. To be exact, we will have

$$A = E_{i_1}E_{i_2}\cdots E_{i_r}\bar{E}_{j_1}\bar{E}_{j_2}\cdots\bar{E}_{j_s} \quad (r+s=n),$$

the result A consisting of the r throws $i_1, i_2, \ldots, i_r$ giving "heads" and the other s throws $j_1, j_2, \ldots, j_s$ giving tails: A is then one of the constituents formed with $E_1, E_2, \ldots, E_n$. But then, if it is a question of a conditional probability, we can apply the theorem of compound probability, and the interpretation of the results which flow from this will constitute our justification of inductive reasoning.

In general, we have

$$\mathbf{P}(E_{n+1} \mid A) = \frac{\mathbf{P}(A \cdot E_{n+1})}{\mathbf{P}(A)}\,; \tag{4}$$

our explanation of inductive reasoning is nothing else, at bottom, than the knowledge of what this formula expresses: the probability of E_{n+1} evaluated when the result A of $E_1, \ldots, E_n$ is known, is not something of an essentially novel nature (justifying the introduction of a new term like "statistical" or "*a posteriori*" probability). This probability is not independent of the "*a priori* probability" and does not replace it; it flows in fact from the same *a priori* judgment by subtracting, so to speak, the components of doubt associated with the trials whose results have been obtained.[f]

In order to avoid erroneous interpretations of what follows, it is best at the outset to recall once more the sense which we attribute to a certain number of terms in this work. Let us consider, to begin with, a class of events (as, for example, the various tosses of a coin). We will say sometimes that they constitute the *trials* of a given phenomenon; this will serve to remind us that we are almost always interested in applying the reasoning that follows to the case where the events considered are events *of the same type*, or which have *analogous* characteristics, without attaching an intrinsic significance or a precise value to these exterior characteristics whose definition is largely arbitrary. Our reasoning will only bring in the events, that is to say, the trials, each taken individually; the analogy of the events does not enter into the chain of reasoning in its own right but only to the degree and in the sense that it can influence in some way the judgment of an individual on the probabilities in question.

It is evident that by posing the problem as we have, it will be impossible for us to *demonstrate* the validity of the principle of induction, that is to say, the principle according to which the probability ought to be close to the observed frequency—for example, in the preceding case: $\mathbf{P}(E_{n+1} \mid A) \simeq r/n$. That this principle can only be justified in particular cases is not due to an insufficiency of the method followed, but corresponds logically and necessarily to the essential demands of our point of view. Indeed, probability being purely subjective, nothing obliges us to choose it close to the frequency; all that can be shown is that such an evaluation follows in a coherent manner from our initial judgment when the latter satisfies certain perfectly clear and natural conditions.

(*f*) This terminology derives from the time when a philosophical distinction was made between probabilities evaluated by considerations of symmetry (*a priori* probabilities), and those justified statistically (*a posteriori* probabilities); this dualistic view is now rejected not only in the subjectivistic theory maintained here, but also by most authors of other theories. With reference to current views, it is proper to speak simply of *initial* and *final* probabilities (the difference being relative to a particular problem where one has to deal with evaluations at different times, before and after some specific additional information has been obtained); the terminology has not been modernized here because the passage makes reference to the older views.

We will limit ourselves in what follows to the simplest conditions which define the events which we call exchangeable, and to fix our ideas we will exhibit these conditions in the example already mentioned; our results will nevertheless be completely general.

The problem is to evaluate the probabilities of all the possible results of the n first trials (for any n). These possible results are 2^n in number, of which $\binom{n}{n} = 1$ consist of the repetition of "heads" n times, $\binom{n}{n-1} = n$ of $n - 1$ occurrences of "heads" and one occurrence of "tails", . . . , and in general $\binom{n}{r}$ of r occurrences of "heads" and $n - r$ occurrences of "tails". If we designate by $\omega_r^{(n)}$ the probability that one obtains in n tosses, in any order whatever, r occurrences of "heads" and $n - r$ occurrences of "tails", $\omega_r^{(n)}$ will be the sum of the probabilities of the $\binom{n}{r}$ distinct ways in which one can obtain this result; the average of these probabilities will then be $\omega_r^{(n)} \Big/ \binom{n}{r}$. Having grouped the 2^n results in this way, we can distinguish usefully, though arbitrarily, two kinds of variation in the probabilities: to begin with we have an average probability which is greater or smaller for each frequency, and then we have a more or less uniform subdivision of the probabilities $\omega_r^{(n)}$ among the various results of equal frequency that only differ from one another in the order of succession of favorable and unfavorable trials. In general, different probabilities will be assigned, depending on the order, whether it is supposed that one toss has an influence on the one which follows it immediately, or whether the exterior circumstances are supposed to vary, etc.; nevertheless it is particularly interesting to study the case where the probability does not depend on the order of the trials. In this case every result having the same frequency r/n on n trials has the same probability, which is $\omega_r^{(n)} \Big/ \binom{n}{r}$; if this condition is satisfied, we will say that the events of the class being considered, e.g., the different tosses in the example of tossing coins, are *exchangeable* (in relation to our judgment of probability). We will see better how simple this condition is and the extent to which its significance is natural, when we have expressed it in other forms, some of which will at first seem more general, and others more restrictive.

It is almost obvious that the definition of exchangeability leads to the following result: the probability that n determinate trials will all have a favorable result is always the same, whatever the n-tuple chosen; this probability will be equal to $\omega_n = \omega_n^{(n)}$, since the first n cases constitute a particular n-tuple. Conversely, if the probabilities of the events have this property, the events are exchangeable, for, as will be shown a little later, it follows from this property that all the results having r favorable and s unfavorable results out of n trials have the same probability, that is:

$$\frac{\omega_r^{(n)}}{\binom{n}{r}} = \sum(-1)^s \Delta^s \omega_r \tag{5}$$

Another conclusion has already been obtained: the probability that r trials will be favorable and s unfavorable will always be $\omega_r^{(n)} \Big/ \binom{n}{r}$ (with $n = r + s$), not only when it is a question of the first n trials in the original order, but also in the case of any trials whatever.

Another condition, equivalent to the original definition, can be stated: the probability of any trial E whatever, conditional on the hypothesis A that there have been r favorable and s unfavorable results on other specific trials, does not depend on the events chosen, but simply on r and s (or on r and $n = r + s$).[g] If

$$\mathbf{P}(\mathrm{A}) = \frac{\omega_r^{(n)}}{\binom{n}{r}} \quad \text{and} \quad \mathbf{P}(\mathrm{A} \cdot \mathrm{E}) = \frac{\omega_{r+1}^{(n+1)}}{\binom{n+1}{r+1}}$$

then we will have

$$\mathbf{P}(\mathrm{E} \mid \mathrm{A}) = \frac{r+1}{n+1}\left(\frac{\omega_{r+1}^{(n+1)}}{\omega_r^{(n)}}\right) = p_r^{(n)} \tag{6}$$

a function of n and r only; if, on the other hand, one supposes that $\mathbf{P}(\mathrm{E} \mid \mathrm{A}) = p_r^{(n)}$, a function of n and r only, it follows clearly that for every n-tuple the probability that all the trials will be favorable is

$$\omega_n = p_0^{(0)} \cdot p_1^{(1)} \cdots p_{n-1}^{(n-1)}. \tag{7}$$

In general it may easily be seen that in the case of exchangeable events, the whole problem of probabilities concerning $\mathrm{E}_{i_1}, \mathrm{E}_{i_2}, \ldots, \mathrm{E}_{i_n}$ does not depend on the choice of the (distinct) indices $i_1, \ldots, i_n$, but only on the probabilities $\omega_0, \omega_1, \ldots, \omega_n$. This fact justifies the name of "exchangeable events" that we have introduced: when the indicated condition is satisfied, any problem is perfectly well determined if it is stated for *generic* events.

(g) This may also be expressed by saying that the observed frequency r/n and n give a *sufficient statistic*, or that the likelihood is only a function of r/n and n.

Single card de re

de dicto

de

This same fact makes it very natural to extend the notion of exchangeability to the larger domain of random quantities: We shall say that $X_1, X_2, \ldots, X_n, \ldots$ are exchangeable random quantities if they play a symmetrical role in relation to all problems of probability, or, in other words, if the probability that $X_{i_1}, X_{i_2}, \ldots, X_{i_n}$ satisfy a given condition is always the same however the distinct indices $i_1 \cdots i_n$ are chosen. As is the case for exchangeable *events*, any problem of probability is perfectly determined when it has been stated for *generic* random quantities; in particular if $X_1, X_2, \ldots, X_n, \ldots$ are exchangeable random quantities, the events $E_i = (X_i \leq x)$ (where x is any fixed number) or more generally $E_i = (X_i \in I)$ (I being any set of numbers) are exchangeable. This property will be very useful to us, as in the following case: the mathematical expectation of any function of n exchangeable random quantities does not change when we change the n-tuple chosen; in particular there will be values $m_1, m_2, \ldots, m_k, \ldots$ such that $\mathscr{M}(X_i) = m_1$, whatever i may be; $\mathscr{M}(X_iX_j) = m_2$, whatever be i and j $(i \neq j)$, and in general $\mathscr{M}(X_{i_1}X_{i_2} \cdots X_{i_k}) = m_k$ whatever be the distinct $i_1, i_2, \ldots, i_k$. This observation has been made by Khinchin[1] who has used it to simplify the proofs of some of the results that I have established for exchangeable events. I have used this idea in the study of exchangeable random quantities, and I will avail myself of it equally in this account.

One can, indeed, treat the study of exchangeable events as a special case of the study of exchangeable random quantities, by observing that the events E_i are exchangeable only if that is also true of their "indicators", that is to say, the random quantities X_i such that $X_i = 1$ or $X_i = 0$ according to whether E_i occurs or not. We mention in connection with these "indicators" some of the simple properties which explain their usefulness.

Characteristic Functions

The indicator of $\bar{E}_i$ is $1 - X_i$; that of E_iE_j is X_iX_j; that of $E_i \vee E_j$ is $1 - (1 - X_i)(1 - X_j) = X_i + X_j - X_iX_j$—it is not, as it is in the case of incompatible events where $X_iX_j = 0$, simply $X_i + X_j$. The indicator of $E_{i_1}E_{i_2} \cdots E_{i_r}\bar{E}_{j_1}\bar{E}_{j_2} \cdots \bar{E}_{j_s}$ is then

$$
\begin{aligned}
X_{i_1}X_{i_2} \cdots X_{i_r}&(1 - X_{j_1})(1 - X_{j_2}) \cdots (1 - X_{j_s}) \\
&= X_{i_1}X_{i_2} \cdots X_{i_r} - \sum_{h=1}^{s} X_{i_1}X_{i_2} \cdots X_{i_r}X_{jh} \\
&\quad + \sum_{k,h=1}^{s} X_{i_1}X_{i_2} \cdots X_{i_r}X_{j_h}X_{j_k} - \cdots \pm X_1X_2 \cdots X_n
\end{aligned}
$$

(1) [XV]; also see [XVI].

The mathematical expectation of the indicator is only the probability of the corresponding event; thus the possibility of transforming the logical operations on the events into arithmetical operations on the indicators greatly facilitates the solution of a certain number of problems. One infers immediately, in particular, the formula (5) stated for $\omega_r^{(n)}$ in the case of exchangeable events: if the product of h trials always has the probability ω_h, then the probability $\omega_r^{(n)} \Big/ \binom{n}{r}$ of $E_{i_1}E_{i_2}\cdots E_{i_r}\overline{E}_{j_1}\overline{E}_{j_2}\cdots\overline{E}_{j_s}$ is deduced from the above development of the indicator of this event and one obtains

$$\frac{\omega_r^{(n)}}{\binom{n}{r}} = \omega_r - \binom{s}{1}\omega_{r+1} + \binom{s}{2}\omega_{r+2} - \cdots (-1)^s\omega_{r+s} = \sum(-1)^s\,\Delta^s\omega_r. \quad (5)$$

Putting $\omega_0 = 1$, the formula remains true for $r = 0$.

Leaving aside for the moment the philosophical question of the principles which have guided us here, we will now develop the study of exchangeable events and exchangeable random quantities, showing first that the law of large numbers and even the strong law of large numbers are valid for exchangeable random quantities X_i, and that the probability distribution of the average Y_n of n of the random quantities X_i tends toward a limiting distribution when n increases indefinitely. It suffices even, in the demonstration, to suppose

$$\mathscr{M}(X_i) = m_1, \qquad \mathscr{M}(X_i^2) = \mu_2, \qquad \mathscr{M}(X_iX_j) = m_2$$

for all i and j $(i \neq j)$, a condition which is much less restrictive than that of exchangeability. We remark again that it suffices to consider explicitly random quantities, the case of events being included by the consideration of "indicators"; an average Y_n is identical, in this case, with the frequency on n trials.

The "law of large numbers" consists of the following property: *if Y_h and Y_k are respectively the averages of h and of k random quantities X_i (the two averages may or may not contain some terms in common), the probability that $|Y_h - Y_k| > \epsilon$ $(\epsilon > 0)$ may be made as small as we wish by taking h and k sufficiently large*; this follows immediately from the calculation of the mathematical expectation of $(Y_h - Y_k)^2$:

$$\mathscr{M}(Y_h - Y_k)^2 = \frac{h+k-2r}{hk}(\mu_2 - m_2)$$
$$= \left(\frac{1}{h} + \frac{1}{k} - \frac{2r}{hk}\right)(\mu_2 - m_2) \leq \left(\frac{1}{h} + \frac{1}{k}\right)(\mu_2 - m_2), \quad (8)$$

where r is the number of common terms, i.e., the X_i that occur in Y_h as well as in Y_k. In particular, if it is a question of "successive" averages, that is to say, if all the terms in the first expression appear also in the other, as for example if

$$Y_h = (1/h)(X_1 + X_2 - \cdots + X_h), = Y_k = (1/k)(X_1 + X_2 + \cdots + X_k)$$
$(h < k)$

we will have $r = h$, and

$$\mathscr{M}(Y_h - Y_k)^2 = \left(\frac{1}{h} - \frac{1}{k}\right)(\mu_2 - m_2) \qquad (9)$$

When successive averages are considered, we have in addition the following result, which constitutes the strong law of large numbers: *ϵ and θ being given, it suffices to choose h sufficiently great in order that the probability of finding the successive averages $Y_{h+1}, Y_{h+2}, \ldots, Y_{h+q}$ all between $Y_h - \epsilon$ and $Y_h + \epsilon$ differs from unity by a quantity smaller than θ, q being as great as one wants.* If one admits that the probability that all the inequalities

$$|Y_h - Y_{h+i}| < \epsilon \qquad (i = 1, 2, 3, \ldots)$$

are true is equal to the limit of the analogous probability for $i = 1, 2, \ldots, q$, when $q \to \infty$, then one can say that *all* the averages Y_{h+i} $(i = 1, 2, \ldots)$ fall between $Y_h - \epsilon$ and $Y_h + \epsilon$, excepting in a case whose probability is less than θ; I prefer however to avoid this sort of statement, for it presupposes essentially the extension of the theorem of total probabilities to the case of a denumerably infinite number of events, and this extension is not admissible, at least according to my point of view (see p. 108).

The proof of the strong law of large numbers can be obtained easily, by considering the variation among the terms Y_{h+i} with the index $(h + i)$ square, and then the variation in the segments between two successive square indices. If the Y's with square indices do not differ among themselves by more than $\epsilon/3$, and the Y's with indices falling between two successive square indices do not differ from each other by more than $\epsilon/3$, the deviations among the Y_{h+i} obviously cannot exceed ϵ. But it suffices to apply the Bienaymé-Tchebycheff inequality to succeed in overestimating the probability of an exception to one of these partial limitations,[2] and

(2) The formula $\mathscr{M}(Y_h - Y_k)^2 = \dfrac{h-k}{hk}(\mu_2 - m_2)$ gives, by the Bienaymé-Tchebycheff inequality, $\mathbf{P}(|Y_h - Y_k| < \epsilon) < \dfrac{1}{\epsilon^2}(\mu_2 - m_2)\dfrac{h-k}{hk}$; applying the theorem of total probabilities in the manner indicated in my note [47], it is possible to draw from this inequality the conclusions that follow in the text.

the corresponding probabilities come out less than $36(\mu_2 - m_2)\epsilon^{-2}\sum_{i=s}^{\infty} i^{-2}$ (s = the integral part of $\sqrt{h}$); the probability of an exception to one or the other of the partial limitations cannot therefore exceed

$$72(\mu_2 - m_2)\epsilon^{-2}\sum_{i=s}^{\infty} i^{-2}.$$

This value does not depend on q, and tends toward zero when $s \to \infty$ (that is to say, when $h \to \infty$); the strong law of large numbers is thus demonstrated.

From the fact that the law of large numbers holds, the other result stated follows easily: *the distribution* $\mathbf{\Phi}_n(\xi) = \mathbf{P}(Y_n \leq \xi)$ *tends to a limit as* $n \to \infty$. If the probability that $|Y_h - Y_k| > \epsilon$ is smaller than θ, the probability that $Y_h < \xi$ and $Y_k > \xi + \epsilon$ will *a fortiori* be smaller than θ, and one will have $\mathbf{\Phi}_h(\xi) > \mathbf{\Phi}_k(\xi + \epsilon) + \theta$, and similarly $\mathbf{\Phi}_h(\xi) < \mathbf{\Phi}_k(\xi - \epsilon) - \theta$. As ϵ and θ can be chosen as small as we wish, it follows that there exists a limiting distribution $\mathbf{\Phi}(\xi)$ such that $\lim_{n\to\infty} \mathbf{\Phi}_n(\xi) = \mathbf{\Phi}(\xi)$ except perhaps for points of discontinuity.[3]

If, in particular, the random quantities $X_1, X_2, \ldots, X_n, \ldots$ are the indicators of exchangeable trials of a given phenomenon, that is to say, if they correspond to the exchangeable events $E_1, E_2, \ldots, E_n, \ldots$, the hypothesis will be satisfied; it would suffice even that the events be equally probable $[\mathbf{P}(E_i) = \mathscr{M}(X_i) = m_1 = \omega_1]$ and have the same two-by-two correlation $[\mathbf{P}(E_iE_j) = \mathscr{M}(X_iX_j) = m_2 = \omega_2]$. We remark that for the indicators one has $X^2 = X$ (since $0^2 = 0$ and $1^2 = 1$) so that $\mu_2 = m_1 = \omega_1$. For Y_h, the frequency on h trials, we then have

$$\mathscr{M}(Y_h) = \omega_1, \quad \mathscr{M}(Y_h - Y_k)^2 = (1/h + 1/k - 2r/hk)(\omega_1 - \omega_2) \qquad (10)$$

and the demonstrated results show simply that the frequencies of two sufficiently numerous groups of trials are, almost surely, very close [even if it is a question of disjoint groups ($r = 0$); if there are some common events ($r > 0$), so much the better]. The same results further signify that the

(3) We remark that if the X_i are exchangeable, the distribution $\mathbf{\Phi}_n(\xi)$ is the same for all the averages Y_n of n terms; one then has a sequence of functions $\mathbf{\Phi}_n$ depending solely on n and tending toward $\mathbf{\Phi}$; with a less restrictive hypothesis than the demonstration assumes, two averages Y_n' and Y_n'' formed from distinct terms can have two different distributions $\mathbf{\Phi}_n'$ and $\mathbf{\Phi}_n''$, but the result will still hold in the sense that all the $\mathbf{\Phi}_n(\xi)$ concerning the average of any n terms whatever will differ very little from $\mathbf{\Phi}(\xi)$ (and thus from one another) if n is sufficiently large.

successive frequencies in the same sequence of experiments oscillate almost surely within a quantity less than a given ϵ, beginning from a rank h sufficiently large, whatever be the number of subsequent events; and finally that there exists a probability distribution $\Phi(\xi)$ differing only slightly from that of a frequency Y_h for very large h.

In order to determine completely the limiting distribution $\Phi(\xi)$, the knowledge of m_1, m_2, μ_2, is evidently no longer sufficient, except in the limiting case where there is no two-by-two correlation and $m_2 = m_1^2$; here $\Phi(\xi)$ is degenerate and reduces to the distribution where the probability is concentrated in one point, $\xi = m_1$. In this case the law of large numbers and the strong law of large numbers reduce to the laws of Bernoulli and Cantelli [III], [V], according to which the deviation between Y_h and the value m_1, fixed in advance, tends stochastically toward zero in a "strong" way. In the general case of a class of exchangeable random quantities, Φ is determined by the knowledge of the complete sequence $m_1, m_2, \ldots, m_n, \ldots$, for these values are the *moments* relative to the distribution Φ:

indep.

$$m_n = \int_0^1 \xi^n d\Phi(\xi) \tag{11}$$

and then

$$\psi(t) = \sum_{0=n}^{\infty} \frac{i^n t^n}{n!} m_n \tag{12}$$

is the characteristic function of Φ.

Indeed,

$$Y_h^n = \frac{1}{h^n}(X_1 + X_2 + \cdots + X_n)^n = \frac{1}{h^n} \sum X_{i_1} X_{i_2} \cdots X_{i_n};$$

among the h^n products there are $h(h-1)(h-2)\cdots(h-n+1)$ that are formed from distinct factors; the products containing the same term more than one time constitute a more and more negligible fraction as h is increased, so that

$$\mathscr{M}(Y_h^n) = \frac{h(h-1)\cdots(h-n+1)}{h^n} m_n + \mathscr{O}\left(\frac{1}{h}\right) \to m_n \qquad (h \to \infty). \tag{13}$$

If, in particular, the X_i are the indicators of exchangeable trials of a phenomenon Y_h, the frequency on h trials, then m_n is the probability ω_n that n trials will all have a favorable result: it is the mean of the nth power of the relative frequency on a large number of trials. The characteristic function of $\mathbf{\Phi}(\xi)$ is

$$\psi(t) = \sum_{n=0}^{\infty} \frac{i^n t^n}{n!} \omega_n \tag{14}$$

and we have

$$\mathbf{\Phi}(\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{it} - e^{it\xi}}{it} \psi(t)\, dt \tag{15}$$

for—the Y_h signifying frequencies—the probability distribution can only fall between 0 and 1, and thus $\mathbf{\Phi}(-1) = 0$. The characteristic function of $\mathbf{\Phi}_h(\xi)$ is

$$\psi_h(t) = \Omega_h\left(e^{i\frac{t}{h}}\right), \tag{16}$$

where Ω_h is the polynomial

$$\Omega_h(z) = \sum_{k=0}^{h} \binom{h}{k} \omega_k (z-1)^k, \tag{17}$$

and $\Omega_h(t)$ converges uniformly to $\psi(t)$. This fact can be proved directly; it is from this standpoint that I developed systematically the study of exchangeable events in my first works [29], [40], and demonstrated the existence of the limiting distribution $\mathbf{\Phi}$, and of ψ, which I call the "characteristic function of the phenomenon".[4]

To give the limiting distribution $\mathbf{\Phi}$, or the characteristic function ψ, is, as we have seen, equivalent to giving the sequence ω_n; it follows that this suffices to determine the probability for any problem definable in terms of exchangeable events. All such problems lead, indeed, in the case of exchangeable events, to the probabilities $\omega_r^{(n)}$ that on n trials, a number r will be favorable; we have (putting $s = n - r$)

$$\omega_r^{(n)} = (-1)^s \Delta^s \omega_r = \binom{n}{r} \int_0^1 \xi^r (1-\xi)^s \, d\mathbf{\Phi}(\xi), \tag{18}$$

and an analogous formula having the same significance is valid for the general case. Indeed, let $\mathbf{P}_\xi(E)$ be the probability attributed to the generic event E when the events $E_1, E_2, \ldots, E_n, \ldots$ are considered independent and equally probable with probability ξ; the probability $\mathbf{P}(E)$ of the same generic event, the E_i being exchangeable events with the limiting

(4) I had then reserved the name "phenomenon" for the case of exchangeable trials; I now believe it preferable to use this word in the sense which is commonly given to it, and to specify, if it should be the case, that it is a question of a phenomenon whose trials are judged exchangeable.

distribution $\mathbf{\Phi}(\xi)$, is

$$\mathbf{P}(\mathrm{E}) = \int_0^1 \mathbf{P}_\xi(\mathrm{E})\, d\mathbf{\Phi}(\xi).^5 \tag{19}$$

This fact can be expressed by saying that the probability distributions **P** corresponding to the case of exchangeable events are linear combinations of the distributions $\mathbf{P}_\xi$ corresponding to the case of independent equiprobable events, the weights in the linear combination being expressed by $\mathbf{\Phi}(\xi)$.

This conclusion exhibits an interesting fact which brings our case into agreement with a well known scheme, with which it even coincides from a formal point of view. If one has a phenomenon of exchangeable trials, and if $\mathbf{\Phi}$ is the limiting distribution of the frequencies, a scheme can easily be imagined which gives for every problem concerning this phenomenon the same probabilities; it suffices to consider a random quantity X whose probability distribution is $\mathbf{\Phi}$ and events which, conforming to the hypothesis $\mathrm{X} = \xi$ (ξ any value between 0 and 1), are independent and have a probability $p = \xi$; the trials of a phenomenon constructed thus are always exchangeable events. Further on, we will analyze the meaning of this result, after having examined its extension to exchangeable random quantities. For the moment, we will limit ourselves to deducing the following result: in order that $\mathbf{\Phi}$ may represent the limiting distribution corresponding to a class of exchangeable events, it is necessary and sufficient that the distribution be limited to values between 0 and 1 [so that $\mathbf{\Phi}(-\epsilon) = 0$, $\mathbf{\Phi}(1 + \epsilon) = 1$ when $\epsilon > 0$]; in other words it is necessary that the ω_h be the moments of a distribution taking values between 0 and 1, or again that $(-1)^s \Delta^s \omega_r \geq 0$ ($r, s = 1, 2, \ldots$), as results from the expression for $\omega_n^{(r)}$.

If only the probabilities of the various frequencies on n trials, $\omega_0^{(n)}$, $\omega_1^{(n)}, \omega_2^{(n)}, \ldots, \omega_n^{(n)}$, are known, the condition under which there can exist a phenomenon consisting of exchangeable trials for which the $\omega_r^{(n)}$ have the given values, will clearly be that the corresponding $\omega_1, \omega_2, \ldots, \omega_n$ be the first n moments of a distribution on (0, 1); these ω_h can be calculated as a function of the $\omega_r^{(n)}$ by the formula

$$\omega_h = \sum_{r=h}^{n} \omega_r^{(n)} \frac{r!\,(n-h)!}{n!\,(r-h)!}\,; \tag{20}$$

(5) It is clear that the particular case just mentioned—formula (18)—is obtained by putting E = "on n (given) trials, r results are favorable"; then, indeed,

$$\mathbf{P}_\xi(\mathrm{E}) = \binom{n}{r} \xi^r (1-\xi)^{n-r}, \quad \mathbf{P}(\mathrm{E}) = \omega_r^{(n)}$$

finally, the condition that $\omega_1, \ldots, \omega_n$ be the first n moments of a distribution on (0, 1) is that all the roots of the polynomial

$$f(\xi) = \begin{vmatrix} 1 & \xi & \xi^2 & \cdots & \xi^k \\ \omega_0 & \omega_1 & \omega_2 & \cdots & \omega_k \\ \omega_1 & \omega_2 & \omega_3 & \cdots & \omega_{k+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \omega_{k-1} & \omega_k & \omega_{k+1} & \cdots & \omega_{2k-1} \end{vmatrix} \quad \text{if } n = 2k - 1. \tag{21}$$

$$f(\xi) = \begin{vmatrix} 1 & \xi & \xi^2 & \cdots & \xi^k \\ \omega_1 & \omega_2 & \omega_3 & \cdots & \omega_{k+1} \\ \omega_2 & \omega_3 & \omega_4 & \cdots & \omega_{k+2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \omega_k & \omega_{k+1} & \omega_{k+2} & \cdots & \omega_{2k} \end{vmatrix} \quad \text{if } n = 2k. \tag{22}$$

fall in the interval (0, 1), including the endpoints.[6]

CHAPTER IV

Exchangeable Random Quantities

Thus, as we have seen, in any problem at all concerning the exchangeable events $E_1, E_2, \ldots, E_n$, the probability will be completely determined either by the sequence of probabilities ω_n or by the limiting distribution of the frequency $\mathbf{\Phi}(\xi)$ [or, what amounts to the same thing, by the corresponding characteristic function $\psi(t)$]. We have thus completely characterized the families of exchangeable events, and we have, in particular, elucidated the essential significance of $\mathbf{\Phi}(\xi)$ connected with the fundamental result we have demonstrated: the probability distributions **P**, corresponding to the case of exchangeability, are linear combinations of the probability distributions $\mathbf{P}_\xi$ corresponding to the case of independence and equiprobability (probability $= \xi$). We have, indeed,

$$\mathbf{P}(\mathrm{E}) = \int \mathbf{P}_\xi(\mathrm{E})\, d\mathbf{\Phi}(\xi) \tag{19}$$

(6) This result follows from Castelnuovo [VII] (see also [VIII]), as we have noted in [29].

where $d\Phi(\xi)$ represents the distribution of weights in the linear combination.

We are going to extend this fundamental result to the case of exchangeable random quantities for which, up to now, we have only demonstrated the preliminary theorems, which we have used to establish certain results concerning the events themselves, rather than to solve the analogous problem, i.e. to characterize completely families of exchangeable random quantities.

Let us now consider the case of exchangeable random quantities and let us take an example to fix our ideas. In the study of exchangeable events, we have taken as an example the case of a game of heads or tails; let us now suppose that $X_1, X_2, \ldots, X_n$ represent measurements of the same magnitude; it suffices that the conditions under which the measurements are made do not present any apparent asymmetry which could justify an asymmetry in our evaluation of the probabilities, in order that we be able to consider them as exchangeable random quantities.

The extension of our earlier conclusions to this case will clearly be less easy than in the case of events, a *random quantity* being no longer characterized, from the probabilistic point of view, by a number (probability) as are the events, but by a function (for example, a distribution function or a characteristic function, etc.). Here the case of independence and equiprobability corresponds to the hypothesis of the independence of the random quantities X_i and the existence of a general distribution function $V(x)$; by calling $\mathbf{P}_V(E)$ the probability attributable to a generic event E, when the X_i are considered to be independent and to have the same distribution function V, the linear combinations will be distributions of the type

$$\mathbf{P}(E) = \sum c_i \mathbf{P}_i(E)$$

(with the weights $c_i > 0$, $\Sigma\, c_i = 1$); in the limit

$$\mathbf{P}(E) = \int \mathbf{P}_V(E)\, d\mathscr{F}(V), \tag{23}$$

the integral being extended over the function space of distribution functions, and the distribution of weights being characterized by the functional $\mathscr{F}(V)$, in a manner which will be made precise in what follows. Even before knowing the exact meaning of this integration, one is led to notice immediately that if $\mathbf{P}(E)$ is a linear combination of the $\mathbf{P}_V(E)$ one has the case of exchangeability: it suffices to observe that each $\mathbf{P}_V(E)$ giving the same probability to the events defined in a symmetrical[1] fashion in relation to $X_1, \ldots, X_n, \ldots$, the same condition will necessarily be satisfied by every linear conbination $\mathbf{P}(E)$; it is a question then only of proving the inverse, i.e. of showing that, in the case of exchangeability, $\mathbf{P}(E)$ is necessarily of the form $\int \mathbf{P}_V(E)\, d\mathscr{F}(V)$.[2]

The definition of the integral

$$\int f(V)\, d\mathscr{F}(V)$$

that we must introduce over the function space is only a completely natural generalization of the Stieltjes-Riemann integral:[3] by subdividing the space of distribution functions into a finite number of partial domains in any way whatever, we consider the expressions $\Sigma \overline{f_i} c_i$ and $\Sigma \underline{f_i} c_i$ where c_i is the weight of a generic element of these parts, and $\overline{f_i}$ and $\underline{f_i}$ are respectively the upper and lower bounds of the values taken by the function f in these domains. The lower bound of $\Sigma \overline{f_i} c_i$ and the upper bound of $\Sigma \underline{f_i} c_i$, when the subdivision is changed in all possible ways, are respectively the superior and inferior integral of f, extended to the function space of distribution functions in relation to the distribution of weights $\mathscr{F}$; when they coincide, their common value is precisely the integral $\int f(V)\, d\mathscr{F}(V)$ that we are going to examine more closely.

(1) Symmetric in the sense that, for example, the event E = "the point determined by the coordinates $X_1, X_2, \ldots, X_n$ will fall in the domain D" (in Euclidean space of n dimensions) is symmetrical to the events consisting in the same eventuality for one of the $n!$ points $(X_{i_1}, X_{i_2}, \ldots, X_{i_n})$ corresponding to the $n!$ permutations of the coordinates. In particular:

(for D rectangular):

$$E = \text{“}a_h < X_h < b_h \quad (h = 1, 2, \ldots, n)\text{”}$$

and

$$\text{“}a_h < X_{i_h} < b_h \quad (h = 1, 2, \ldots, n)\text{”};$$

(for D spherical):

$$E = \text{“}\textstyle\sum (X_h - a_h)^2 < \rho^2\text{”}$$

and

$$\text{“}\textstyle\sum (X_{i_h} - a_h)^2 < \rho^2\text{”};$$

(for D a half-space):

$$E = \text{“}\textstyle\sum a_h X_h > a\text{” and “}\sum a_h X_{i_h} > a\text{”}, \cdots$$

(2) One can accept this result and omit the following developments which are devoted to proving it and making it precise (toward the end of Chap. IV), without prejudice to an overall view of the thesis maintained in these lectures.

(3) For the reasons which make us regard the Stieltjes-Riemann integral as more appropriate to the calculus of probability, see [58] and [64].

We are going to show that, in the circumstances that interest us, this integral exists, and that in order to determine its value, it suffices to know the weight for some very simple functional domains of distribution functions. Suppose to begin with the $f(V)$ depends only on the values

$$y_1 = V(x_1), y_2 = V(x_2), \ldots, y = V(x_s)$$

which the function V takes on a finite given set of abscissas $x_1, x_2, \ldots, x_s$; $f(V)$ is thus the probability that n random variables following the distribution V will all fall in a rectangular domain D, the first falling between x_1 and x_1', the second between x_2 and $x_2' \cdots$, the last between x_n and x_n'. This probability is[4]

$$f(V) = [V(x_1') - V(x_1)][V(x_2') - V(x_2)] \cdots [V(x_n') - V(x_n)]$$
$$= (y_1' - y_1)(y_2' - y_2) \cdots (y_n' - y_n) \qquad (s = 2n) \tag{24}$$

It is clear that in order to evaluate the integral of such a function, it is sufficient to know the weights of the functional domains defined only by the ordinates $y_1, \ldots, y_s$ corresponding to the abscissas $x_1, \ldots, x_s$, i.e. the weights of the domains of the space of s dimensions defined by $y_1, \ldots, y_s$; if f is a continuous function of the y_i it will suffice to know the weights of the domains defined by the inequalities $y_i < a_i$ $(i = 1, 2, \ldots, s)$. The significance of these domains is the following: they comprise the distribution functions V whose representative curve $y = V(x)$ remains below each of the s points (x_i, a_i). Let $\mathbf{\Phi}(x)$ be the stepwise curve of which the points (x_i, a_i) are the lower corners; the above condition can now be expressed by $V(x) < \mathbf{\Phi}(x)$ [for all x],[5] and the weights of the set of distribution functions V such that $V(x) < \mathbf{\Phi}(x)$ will be designated by $\mathscr{F}(\mathbf{\Phi})$; thus we give a concrete meaning to $\mathscr{F}$ which until now has represented a distribution of weights in a purely symbolic way. In this case the integral $\int f(V)\, d\mathscr{F}(V)$ is only the ordinary Stieltjes-Riemann integral in the space of s dimensions. If $f(V)$ does not depend solely on the ordinates of $V(x)$ for a finite set of abscissas $x_1, \ldots, x_s$ we will consider the case where it is possible to

(4) It is not necessary to be particularly concerned with the discontinuity points: indeed, a determinate function V is continuous almost everywhere (better: everywhere except, at most, on a denumerable set of points), and likewise in the integration, the weight of the set of distribution functions having x as a point of discontinuity is always zero, except, at most, for a denumerable set of points x; it suffices to observe that these points are the points of discontinuity of $\mathbf{\Phi}(x) = \int V(x)\, d\mathscr{F}(V)$, and that $\mathbf{\Phi}(x)$ is a distribution function.

(5) It is always understood that an inequality like $f(x) < g(x)$ between two functions means that it holds for all x (unless one has explicitly a particular case in mind when it is a question of a determinate value x).

approach $f(V)$, from above and below, by means of functions of the preceding type, in such a way that the value of the integral will be uniquely determined by the values approached from above and below. In other words, it will be necessary that, for an arbitrary ϵ, one be able to find two functions $f'(V)$ and $f''(V)$ depending on a finite number of values $V(x_i)$, such that

$$f'(V) \leq f(V) \leq f''(V) \quad \text{and} \quad \int f'(V)\, d\mathscr{F}(V) > \int f''(V)\, d\mathscr{F}(V) - \epsilon.$$

We return to the case of n independent random quantities having the distribution $V(x)$: if $f(V)$ is the probability that the point $(X_1, X_2, \ldots, X_n)$ falls in a domain D which is not reducible to a sum of rectangular domains, f' and f'' can represent the analogous probabilities for the domains D′ contained in D, and D″ containing D, each formed from a sum of rectangular domains.

We have no need to pursue to the end the analysis of the conditions of integrability; we will content ourselves with having shown that they are satisfied in some sufficiently general conditions which contain all the interesting cases. We now return to the problem concerning the exchangeable random quantities $X_1, X_2, \ldots, X_n, \ldots$ in order to show the existence of the functional $\mathscr{F}$ having a meaning analogous to $\Phi(\xi)$ for exchangeable events. Let V be a stepwise function of which the lower corners are the s points

$$(x_i, y_i)\ (i = 1, 2, \ldots, s;\ x_{i+1} > x_i;\ y_{i+1} > y_i);$$

we will designate by $\mathscr{F}_h(V)$ the probability that, of h numbers $X_1, X_2, \ldots, X_h$, hy_1 at the most will exceed x_1, hy_2 at the most will exceed $x_2, \ldots, hy_s$ at the most will exceed x_s, or, in other words, the probability

$$\mathbf{P}\{G_h(x) \leq V(x)\}$$

that the distribution function $G_h(x)$ of the values of $X_1, X_2, \ldots, X_h$ never exceeds $V(x)$. More precisely, the function $G_h(x)$ is the "observed distribution function" resulting from the observation of $X_1, \ldots, X_h$; it represents the stepwise curve of which the ordinate is zero to the left of the smallest of the h numbers $X_1, \ldots, X_h$, equal to $1/h$ between the smallest and the second, equal afterwards to $2/h, 3/h, \ldots, (h-1)/h$, and finally equal to unity to the right of the largest of the h numbers considered. The steps of $G_h(x)$ are placed on the points of the axes of abscissas which correspond to the values X_i; before knowing these values, $G_h(x)$ is a random function, since these abscissas are random quantities.

It is easy to show, by extending a theorem given by Glivenko[6] for the case of independent random quantities to the case of exchangeable random quantities, that it is very probable that for h and k sufficiently large, $G_h(x)$ and $G_k(x)$ differ very little, and, in the case of a set of *successive* averages $G_h(x), G_{h+1}(x), \ldots$, we have a strong stochastic convergence. By dividing the x axis into a sufficiently large finite number of points $x_1, \ldots, x_s$ the proof can be based on that given for the analogous properties in the case of exchangeable events. For a given x, $G_h(x)$ and $G_k(x)$ give respectively the frequencies Y_h and Y_k for the h and k trials of the set of exchangeable events $E_i = (X_i < x)$; the difference between $G_h(x)$ and $G_k(x)$ then has standard deviation less than $\sqrt{\frac{1}{h} + \frac{1}{k}}$ [see formula (10)], and the probability that it exceeds ϵ can be made as small as one wishes by choosing h and k larger than a sufficiently large number N. By taking N so that the probability of a difference greater than ϵ is less than θ/s for each of the abscissas

$$x = x_1, x_2, \ldots, x_s,$$

we see that, except in a case whose total probability is less than θ, the two functions $G_h(x)$ and $G_k(x)$ will not differ by more than ϵ for any of the abscissas $x_1, \ldots, x_s$.

Under these conditions, the probability $\mathscr{F}_k(V - \epsilon)$ that $G_k(x)$ will not exceed the stepwise curve $V(x) - \epsilon$ for any x, which is to say the probability of having

$$G_k(x_i) < y_i - \epsilon \qquad (i = 1, 2, \ldots, s)$$

can not be more than $\mathscr{F}_h(V) + \theta$, for, in order to satisfy the imposed conditions, it is necessary either that $G_h(x)$ not exceed $V(x)$ for any x, or that we have $G_h(x) - G_k(x) > \epsilon$ for at least one of the abscissas $x_1 \cdots x_s$. We thus have

$$\mathscr{F}_k(V - \epsilon) - \theta \leq \mathscr{F}_h(V) \leq \mathscr{F}_k(V + \epsilon) + \theta \tag{25}$$

(the second inequality can be proved in the same way); by defining convergence in an appropriate way (as for distribution functions[7]), one concludes that $\mathscr{F}_h \to \mathscr{F}$; it is the functional $\mathscr{F}$ which allows us to characterize the family of exchangeable random quantities we have in mind.

(6) [XIII], see also Kolmogorov [XVIII], and [45].
(7) See Lévy [XX], p. 194.

To prove the fundamental formula

$$\mathbf{P}(E) = \int \mathbf{P}_V(E)\, d\mathscr{F}(V) \tag{23}$$

we remark that we have, for all h,

$$\mathbf{P}(E) = \int \mathbf{P}_{h,V}(E)\, d\mathscr{F}(V) \tag{26}$$

where $\mathbf{P}_{h,V}(E)$ is the probability of E, given the hypothesis

$$G_h(x) = V(x).$$

If the event E depends on the n first random quantities $X_1, \ldots, X_n$ (to fix our ideas by a simple example, let us imagine that the event E consists of X_1 falling between a_1 and b_1, X_2 between a_2 and $b_2, \ldots, X_n$ between a_n and b_n), it will naturally be necessary to suppose $h \geq n$; if h is very large in relation to n, it is clear that $\mathbf{P}_{h,V}(E) \cong \mathbf{P}_V(E)$, for the probability $\mathbf{P}_{h,V}(E)$ is obtained by supposing $X_1, \ldots, X_n$ chosen by chance, simultaneously (that is, without repetition) from among the h values where $G_h = V$ is discontinuous, whereas $\mathbf{P}_V(E)$ is the analogous probability obtained by considering all the combinations possible on the supposition of independent choices. The fact of including or excluding repetitions has a more and more negligible influence as $h \to \infty$; thus $\mathbf{P}_{h,V}(E) \to \mathbf{P}_V(E)$. This relation and the relation $\mathscr{F}_h(V) \to \mathscr{F}(V)$ provide the proof that

$$\mathbf{P}(E) = \int \mathbf{P}_{h,V}(E)\, d\mathscr{F}_h(V) = \int \mathbf{P}_V(E)\, d\mathscr{F}(V).$$

We shall consider a particular type of event E, which will permit us to analyze the relation between the functional distribution given by $\mathscr{F}$, relative to the exchangeable random quantities X_i, and the linear distributions $\mathbf{\Phi}_x(\xi)$, that is to say, the limiting distributions $\mathbf{\Phi}(\xi)$, related to the events $E_i = (X_i < x)$. An event E will belong to the particular type envisaged if it expresses a condition depending solely on the fact that certain random quantities $X_1, \ldots, X_n$ are less than or greater than[8] a unique given number x. For example, E = "X_1, X_3, X_8 are $> x$, X_2 and X_7 are $< x$"; E = "among the numbers X_2, X_3, X_9, X_{12} there are three which are $> x$ and one $< x$"; E = "in the sequence $X_1 X_2 \cdots X_{100}$ there are no more than three consecutive numbers $> x$"; etc. In other words, the event E is a logical combination of the $E_i = (X_i < x)$ for a unique given x.

(8) This case of equality has a zero probability, except for particular values of x finite or at most denumerable in number; we neglect this case for simplicity, observing, besides, that it does not entail any essential difficulty, but only the consideration of two distinct values of the distribution function to the left and right of x.

The theory of exchangeable events tells us that the probability of any event E of this type is completely determined by the knowledge of $\mathbf{\Phi}_x(\xi)$, and we can express this probability with the aid of $\mathscr{F}(V)$; we can then express $\mathbf{\Phi}_x(\xi)$ by means of $\mathscr{F}(V)$, and we then have precisely

$$\mathbf{\Phi}_x(\xi) = \int_{V(x)<\xi} d\mathscr{F}(V); \quad d\mathbf{\Phi}_x(\xi) = \int_{\xi<V(x)<\xi+d\xi} d\mathscr{F}(V). \tag{27}$$

Indeed, let $E^{(n)}$ be the event consisting in this: that the frequency of values $< x$ on the first n trials $X_1, \ldots, X_n$ not exceed ξ; by definition $\mathbf{\Phi}_x(\xi) = \lim \mathbf{P}(E^{(n)})$, and moreover $\mathbf{P}_V(E^{(n)})$, for n very large, is very close[9] to 0, if $V(x) > \xi$, or to 1, if $V(x) < \xi$; we will have, then,

$$\begin{aligned}\mathbf{\Phi}_x(\xi) &= \lim \int \mathbf{P}_V(E^{(n)})\, d\mathscr{F}(V) \\ &= \int_{V(x)<\xi} 1 \cdot d\mathscr{F}(V) + \int_{V(x)>\xi} 0 \cdot d\mathscr{F}(V) = \int_{V(x)<\xi} d\mathscr{F}(V). \end{aligned} \tag{27}$$

One can deduce from this (or, better, obtain directly) the following result: $\omega_r^{(n)}(x)$, the probability that r out of n random quantities $X_{i_1}, \ldots, X_{i_n}$, chosen in advance, should be $< x$, is equal to

$$\omega_r^{(n)}(x) = \binom{n}{r} \int [V(x)]^r [1 - V(x)]^{n-r}\, d\mathscr{F}(V), \tag{28}$$

and in particular for $r = n$:

$$\omega_n(x) = \int [V(x)]^n\, d\mathscr{F}(V), \qquad \omega_1(x) = \int V(x)\, d\mathscr{F}(V); \tag{29}$$

this last formula giving, in particular, the probability that a general fixed number X_i is less than x; and this is the distribution function attributed to each of the X_i before beginning the trials.

Up to now, in $\mathbf{\Phi}_x(\xi)$, $\omega_r^{(n)}(x)$, $\omega_n(x)$, we have considered x only as a parameter which determines the events $E_i = (X_i < x)$, but which does not vary; if, on the contrary, these expressions are considered as functions of x, certain remarks can be made which throw a new light on them. Let us consider n of the given random quantities: $X_1, X_2, \ldots, X_n$; $\omega_n^{(n)}$ is the probability that none of these numbers exceed x, and thus constitutes the

(9) We recall that $\mathbf{P}_V(E^{(n)})$ is simply the probability of a frequency $< \xi$ on n independent trials, with constant probability $p = V(x)$, and therefore has the value

$$\sum_{\nu<\xi n} \binom{n}{\nu} p^\nu q^{n-\nu} [q = 1 - p = 1 - V(x)].$$

distribution function of the maximum value among $X_1, \ldots, X_n$; $\omega_n^{(n)}(x) + \omega_{n-1}^{(n)}(x)$ is in an analogous way the distribution function of the next-to-largest of the numbers $X_1, \ldots, X_n$ arranged in increasing order; $\omega_n^{(n)}(x) + \omega_{n-1}^{(n)}(x) + \cdots + \omega_r^{(n)}$ that of the rth; and finally $\omega_n^{(n)}(x) + \cdots + \omega_1^{(n)}(x) = 1 - \omega_0^{(n)}(x)$ that of the smallest of the X_i. As the identity

$$\omega_1^{(n)}(x) + 2\omega_2^{(n)}(x) + 3\omega_3^{(n)}(x) + \cdots + n\omega_n^{(n)}(x) = n\omega_1(x) \qquad (30)$$

shows, the average of these n distribution functions is $\omega_1(x)$, that is to say, the distribution function of any one whatever of the X_i: this fact is very natural, for, according to the definition of exchangeability, each number X_i has the same probability $1/n$ of being the smallest, or the second, . . . , or the greatest, and, in general, all the permutations of the X_i have the same probability $1/n!$ of being disposed in increasing order (if there exists a probability different from zero that the n values are not distinct, the modifications to make in these statements are obvious).

There exists a close relation between the distribution functions of a random quantity of determinate rank and the function $\mathbf{\Phi}_x(\xi)$: by definition, $\mathbf{\Phi}_x(\xi)$ is the limiting value, for $n \to \infty$, of the probability that of n random quantities $X_1, \ldots, X_n$ there will be at most ξn which are $< x$; this probability is equal to $\sum_r \omega_r^{(n)}(x)$, the sum being extended over the indices $r < \xi n$. But this sum is the distribution function of those of the numbers $X_1, \ldots, X_n$ which occupy the rank "whole part of ξn" where the random quantities are arranged in order of increasing magnitude: by considering ξ fixed, $\mathbf{\Phi}_x(\xi)$ is (as a function of x) the distribution function of the number of rank $\cong \xi n$ on a very large number n of given random quantities.

It is easily seen that $\mathbf{\Phi}_x(\xi)$ is a never-decreasing function of ξ and x, that $\mathbf{\Phi} = 0$ if $\xi < 0$, and $\mathbf{\Phi} = 1$ if $\xi > 1$ ($\mathbf{\Phi}$ is thus defined substantially only on the interval $0 \leq \xi \leq 1$), and finally that $\mathbf{\Phi} \to 0$ and $\mathbf{\Phi} \to 1$ respectively for $x \to -\infty$ and $x \to +\infty$. Conversely, each function $\mathbf{\Phi}_x(\xi)$ having these properties can be associated in an infinite number of ways with a probability distribution for exchangeable random quantities; one such function $\mathbf{\Phi}_x(\xi)$ being given, one can always construct a distribution of weights $\mathscr{F}(V)$ in function space, such that formula (27) holds. The simplest way of doing this is the following: let $V_\lambda(x) = \xi$ be the explicit equation of the contour line $\mathbf{\Phi}_x(\xi) = \lambda$, which represents, due to the properties of $\mathbf{\Phi}_x(\xi)$, a distribution function, and define the distribution $\mathscr{F}(V)$ by attributing the weights $\lambda' - \lambda$ ($\lambda' > \lambda$) to the set of $V(x)$ such that

$$V_\lambda(x) < V(x) < V_{\lambda'}(x)$$

for all x; in this way the integration in function space is reduced to a simple integral:

$$\int f(\mathrm{V})\, d\mathscr{F}(\mathrm{V}) = \int_0^1 f(\mathrm{V}_\lambda)\, d\lambda. \tag{31}$$

We have, for example,

$$\omega_n(x) = \int_0^1 [\mathrm{V}_\lambda(x)]^n\, d\lambda = \int_0^1 \xi^n\, d\mathbf{\Phi}_x(\xi); \tag{32}$$

this suffices to show that the distribution we have obtained satisfies the desired condition; it results directly from the calculation of

$$\begin{aligned}\mathbf{\Phi}_x(\xi) &= \int_{\mathrm{V}(x)<\xi} d\mathscr{F}(\mathrm{V}) \\ &= \int_{\mathrm{V}_\lambda(x)<\xi} d\lambda = \{\text{the value of } \lambda \text{ for which } \mathrm{V}_\lambda(x) = \xi\}\end{aligned} \tag{33}$$

However, there always exists an infinity of other distributions $\mathscr{F}(\mathrm{V})$ corresponding to the same function $\mathbf{\Phi}_x(\xi)$: it suffices to observe, for example, that if one puts in any way at all

$$\mathbf{\Phi}_x(\xi) = c_1\mathbf{\Phi}_x^{(1)}(\xi) + c_2\mathbf{\Phi}_x^{(2)}(\xi) + \cdots + c_k\mathbf{\Phi}_x^{(k)}(\xi)$$

with $c_i > 0$, $\Sigma\, c_i = 1$, the $\mathbf{\Phi}_x^{(i)}(\xi)$ satisfying the same conditions as $\mathbf{\Phi}$, and if one introduces the corresponding $\mathrm{V}_\lambda^{(i)}(x)$, one will always have

$$\omega_n(x) = \sum c_i \int_0^1 [\mathrm{V}^{(i)}(x)]^n\, d\lambda. \tag{34}$$

The function $\mathbf{\Phi}_x(\xi)$ thus characterizes neatly all the families of exchangeable events $\mathrm{E}_i = (\mathrm{X}_i < x)$ for any x whatever, but this does not suffice in problems where the interdependence of these various families comes into play: complete knowledge of the distribution $\mathscr{F}(\mathrm{V})$ in function space is then indispensable.

It should be noted once more that if one were to consider exchangeable random elements of any space whatever, one would arrive at perfectly analogous results: implicitly, we have already indeed considered some exchangeable random functions, since, for example, the $\mathrm{G}_h(x)$, the distribution functions of $\mathrm{X}_{i_1}, \mathrm{X}_{i_2}, \ldots, \mathrm{X}_{i_h}$ constitute a family of exchangeable random functions, when all possible groups $i_1, i_2, \ldots, i_h$ are considered. The general result which has been established for events and for random quantities, and which could be demonstrated for random elements of any space whatever, may be expressed by saying that *the probability distribution of classes of exchangeable random elements are "averages" of the probability distributions of classes of independent random elements.*[h]

CHAPTER V

Reflections on the Notion of Exchangeability

We have thus established the general notion of stochastic exchangeability, and obtained the fundamental result which permits the characterization of probability distributions corresponding to the case of exchangeability as some linear combination of the distributions corresponding to the case of independence and equiprobability. We will now try to show how this would be interpreted according to the current conceptions, to exhibit the reasons which keep us from holding such an opinion, and to explain the significance of the notion of exchangeability as well as the role that it is called to play, according to our point of view, in the calculus of probability.

Let us consider, for example, an urn of unknown composition, and let us draw out balls, replacing each ball after it is drawn. The ratio of the number of white balls to the total number can have various possible values ρ_i, to which we attribute certain probabilities c_i. These drawings are exchangeable events, and the probability of a given outcome is $\mathbf{P}(\mathrm{E}) = \sum_i c_i \mathrm{P}_i(\mathrm{E})$, where there is a probability $\mathbf{P}_i(\mathrm{E})$ corresponding to the

(*h*) The rigorous proof of this result under weak conditions on the space concerned has been the subject of several developments; for example, Hewitt, E., and Savage, L. J. (1955), "Symmetric measures on Cartesian products," *Trans. Amer. Math. Soc.*, **80,** 470–501 (and papers cited in the bibliography of this article), and, among later contributions, Bühlmann, H. (1960), "Austauschbare stochastische Variablen und ihre Grenzwertsätze," *Univ. Calif. Publ. Stat.*, **3,** 1–35, and Freedman, D. A. (1962), "Invariants under mixing which generalize de Finetti's theorem," *Ann. Math. Stat.*, **33,** 916–923. An extension in a different sense (so as to include, for instance, cases of the type of Markov chains) has been dealt with by the present author in "Sur la condition d'"équivalence partielle' " (Colloque Genève, 1937), *Act. Sci. Ind. No. 739,* Hermann, Paris, 1938, 5–18, and in "La probabilità e la statistica nei rapporti con l'induzione, secondo i diversi punti di vista" (Summer course at Varenna, 1959), *Induzione e Statistica,* Istituto Matematico dell'Università, Cremonese, Rome, 1959 (particularly, §8, pp. 92–100). An English translation of this last paper is almost ready and will probably be published as soon as some additional material can be prepared.

composition ρ_i; vice versa, as we have already remarked, if the E_i are exchangeable events, the distribution $\mathbf{P}(E)$ is always of the form $\mathbf{P}(E) = \sum_i c_i P_i(E)$ $(c_i > 0, \sum_i c_i = 1)$ or of the limiting form $\mathbf{P}(E) = \int \mathbf{P}_\xi(E)\, d\Phi(\xi)$ (Stieltjes' integral), where $\mathbf{P}_\xi$ is the probability distribution corresponding to the hypothesis of independence and constant probability ξ. The "exchangeable events" correspond then to those which we would ordinarily distinguish as "independent events with constant but unknown probability p", and $\Phi(\xi)$ would be "the probability that this 'unknown probability' p will be smaller than ξ". In a similar fashion, "exchangeable random quantities" correspond to those that we would call "independent random quantities with the constant but unknown probability distribution $G(x)$", and the functional $\mathscr{F}$ would give "the probability distribution of this unknown distribution", $\mathscr{F}(V)$ being the probability that $G(x) \leq V(x)$[1] (for V stepwise with a finite number of steps). But then, you say, why proceed from a novel definition, and make such efforts to conclude that it characterizes nothing else than this well-known case?

It is not without reason that we have considered ourselves obliged to proceed in this way. The old definition cannot, in fact, be stripped of its, so to speak, "metaphysical" character: one would be obliged to suppose that beyond the probability distribution corresponding to our judgment, there must be another, unknown, corresponding to something real, and that the different hypotheses about the unknown distribution—according to which the various trials would no longer be dependent, but independent—would constitute *events* whose probability one could consider. From our point of view these statements are completely devoid of sense, and no one has given them a justification which seems satisfactory, even in relation to a different point of view. If we consider the case of an urn whose composition is unknown, we can doubtless speak of the probability of different compositions and of probabilities relative to one such composition; indeed the assertion that there are as many white balls as black balls in the urn expresses an objective fact which can be directly verified, and the conditional probability, relative to a given objective event, has been well defined. If, on the contrary, one plays heads or tails with a coin of irregular appearance, as in the example of Chap. III, one does not have the right to consider as distinct hypotheses the suppositions that this imperfection has a more or less noticeable influence on the "unknown probability", for this "unknown probability" cannot be defined, and the hypotheses that

(1) See note (1), page 123.

one would like to introduce in this way have no objective meaning. The difference between these two cases is essential, and it cannot be neglected; one cannot "by analogy" recover in the second case the reasoning which was valid in the first case, for this reasoning no longer applies in the second case. If, after numerous drawings, the observed frequency of the white balls is f, why do we attribute a value close to f to the probability that the ball will be white in one of the drawings which is going to follow? It can be answered that after the observation of such a frequency we attribute a very large value to the probability that the number of white balls will come very close to the fraction f of the total, and further, by supposing this fraction to be ρ, we judge that the drawings are independent and have all the same probability $p = \rho$. This explanation is perfectly satisfactory even from the subjectivistic point of view, and does not differ formally from that which is ordinarily given and which reduces, finally, to the theorem of compound probabilities. But in the preceding case of heads or tails, it is otherwise: the corresponding terms which would allow analogous reasoning do not exist. If, nevertheless, we want to reason in an identical and rigorous way in the two cases, it is necessary to begin by looking for the common elements which characterize them, and for those elements which differentiate them.

The result at which we have arrived gives us the looked-for answer, which is very simple and very satisfactory: the nebulous and unsatisfactory definition of "independent events with fixed but unknown probability" should be replaced by that of "exchangeable events". This answer furnishes a condition which applies directly to the evaluations of the probabilities of individual events and does not run up against any of the difficulties that the subjectivistic considerations propose to eliminate. It constitutes a very natural and very clear condition, of a purely qualitative character, reducing to the demand that certain events be judged equally probable, or, more precisely, that all the combinations of n events $E_{i_1}, \ldots, E_{i_n}$ have always the same probability, whatever be the choice or the order of the E_i. The same simple condition of "symmetry" in relation to our judgments of probability defines exchangeable random quantities, and can define, in general, exchangeable random elements in any space whatever. It leads in all cases to the same practical conclusion: a rich enough experience leads us always to consider as probable future frequencies or distributions close to those which have been observed.

Following the demonstration of the existence of a limiting distribution, this fact can be explained by reasoning almost parallel to that which one ordinarily employs when one takes account of the "unknown probability", but which does not give rise to the same criticisms. For the "unknown probability" we substitute the frequency on the first N trials, with N large enough so that the corresponding probability distribution coincides practically with the limiting distribution: $\mathbf{\Phi}_N(\xi) \cong \mathbf{\Phi}(\xi)$, and so that the number of trials with which one is concerned is negligible in relation to N; thus $\binom{R}{r}\binom{S}{s}:\binom{N}{n}$ (the probability that on n trials chosen at random among the N of which R are favorable and S unfavorable there are r favorable and s unfavorable) is practically equal to $\binom{n}{r}\xi^r(1-\xi)^s$ with $\xi = \frac{R}{N}$, $1 - \xi = \frac{S}{N}$. One can then reason as follows: consider as possible hypotheses the N + 1 possible frequencies on N trials, their probabilities being the $\omega_h^{(N)}$ ($h = 0, 1, \ldots, N$), and observe that the hypotheses for which R/N is close to r/n are precisely those according to which the probability of a frequency r/n on n trials is the strongest, and in consequence,[2] those for which the probability conditioned on the observation of the frequency r/n on n trials is the most strongly augmented in relation to the unconditional probability. We conclude finally that the hypotheses closest to the observed result take on a more and more preponderate importance when the number of observations increases, and that this leads us necessarily to make our prediction approach the observation. More precisely, we demonstrate that the limiting distribution $\overline{\mathbf{\Phi}}$, given the observation of the frequency r/n on n trials, is such that

$$d\overline{\mathbf{\Phi}}(\xi) = \alpha\xi^r(1-\xi)^s\, d\mathbf{\Phi}(\xi) \quad \left[\alpha \text{ such that } \int_0^1 \alpha\xi^r(1-\xi)^s\, d\mathbf{\Phi} = 1\right] \qquad (35)$$

and that the corresponding characteristic function is

$$\bar{\psi}(t) = \alpha \mathrm{D}^r(i - \mathrm{D})^s \psi(t) \quad [\alpha \text{ such that } \bar{\psi}(0) = 1;\ \mathrm{D} = d/dt;\ i = \sqrt{-1}]. \qquad (36)$$

In particular, the probability on an additional trial, relative to this hypothesis, will be

$$p_r^{(n)} = \int \xi\, d\overline{\mathbf{\Phi}} = \int \xi \cdot \alpha\xi^r(1-\xi)^s\, d\mathbf{\Phi}; \qquad (37)$$

that is to say, the mean of the ξ from (0,1) with the weights $\alpha\xi^r(1-\xi)^s\, d\mathbf{\Phi}$ in place of the weights $d\mathbf{\Phi}$; the ξ around the maximum $\xi = r/n$ of $\xi^r(1-\xi)^s$ are evidently strengthened more and more.

(2) See formula (2), p. 110, and the relevant explanation.

As a function of the $\omega_h^{(n)}$ we have easily

$$p_r^{(n)} = \frac{r+1}{n+1} \cdot \frac{\omega_{r+1}^{(n+1)}}{\omega_r^{(n)}} = \frac{r+1}{n+2+(s+1)\left(\dfrac{\omega_r^{(n+1)}}{\omega_{r+1}^{(n+1)}} - 1\right)}; \tag{38}$$

it is interesting to see that this formula—which already explains, though incompletely, the influence of the observed frequency on the evaluation of the probability—can be derived in a direct and very elementary way from the definition of "exchangeable events". From this definition it can indeed be inferred that

$$\frac{\omega_r^{(n)}}{\binom{n}{r}} = \frac{\omega_r^{(n+1)}}{\binom{n+1}{r}} + \frac{\omega_{r+1}^{(n+1)}}{\binom{n+1}{r+1}}, \tag{39}$$

for the first member expresses the probability that r given events on the first n trials are favorable, and the second gives the sum of the probabilities that the said combinations will occur with, respectively, a favorable or unfavorable result on the $n + 1$'st trial. Simplifying, we have

$$\omega_r^{(n)} = \frac{s+1}{n+1}\,\omega_r^{(n+1)} + \frac{r+1}{n+1}\,\omega_{r+1}^{(n+1)} \tag{40}$$

and with the help of this identity we obtain

$$p_r^{(n)} = \frac{\omega_{r+1}^{(n+1)}}{\binom{n+1}{r+1}} : \frac{\omega_r^{(n)}}{\binom{n}{r}} = \frac{r+1}{n+1}\,\frac{\omega_{r+1}^{(n+1)}}{\omega_r^{(n)}} = \frac{(r+1)\omega_{r+1}^{(n+1)}}{(s+1)\omega_r^{(n+1)} + (r+1)\omega_{r+1}^{(n+1)}}$$

$$= \frac{r+1}{n+2+(s+1)\left(\dfrac{\omega_r^{(n+1)}}{\omega_{r+1}^{(n+1)}} - 1\right)} \quad \text{Q.E.D.} \tag{38}$$

This formula acquires a particular significance in Laplace's case where the $\omega_r^{(n)}$ do not depend on r, and where one has $\omega_r^{(n)} = \dfrac{1}{n+1}$ (it can be verified immediately that this hypothesis is admissible, and that it corresponds to a homogeneous limiting law: $\mathbf{\Phi}(\xi) = \xi$, $d\mathbf{\Phi}(\xi) = d\xi$, for $0 \leq \xi \leq 1$). In the case of Laplace one has simply, as is well known, $p_r^{(n)} = \dfrac{r+1}{n+2}$, for the other term in the denominator vanishes. The result is still the same if $\omega_{r+1}^{(n+1)} = \omega_r^{(n+1)}$; if on the contrary $\omega_{r+1}^{(n+1)}$ is greater or smaller than $\omega_r^{(n+1)}$, the probability $p_r^{(n)}$ will be respectively greater or smaller than in the Laplacean case. In any case, $p_r^{(n)}$ is close to $\frac{r+1}{n+2}$, and hence close to the frequency r/n, if the ratio differs little from unity; it thus suffices to admit this condition in order to justify easily the influence of observation on prediction in the case of exchangeable events.

With regard to exchangeable random quantities, or any exchangeable elements at all, the results and the demonstrations would be perfectly analogous; for these, as for events, the subjectivistic theory solves the problem of induction completely in the case of exchangeability, corresponding to the case which is most usually considered, and leads to the same conclusions generally admitted or demonstrated by means of vague and imprecise reasoning.

Every hypothesis, however, can be studied in the same style and by the same procedure. One cannot exclude completely *a priori* the influence of the order of events, and in consequence the attribution of probabilities differing more or less among themselves to the $\binom{n}{r}$ various combinations of r favorable results on the $n = r + s$ first trials. There would then be a number of degrees of freedom and much more complication, but nothing would be changed in the setting up and the conception of the problem, which would remain that presented at the beginning of Chap. III, before we restricted our demonstration to the case of exchangeable events, and which is essentially condensed in formula (4).

The influence of the order on the evaluation of the probabilities of A and $A \cdot E_{n+1}$ (*see* p. 119) does not modify, indeed, the way in which the problem is posed and answered according to the subjectivistic conception; one will only be led, in the general case, to take account of the circumstances which in the case of exchangeability are (by definition) neglected. One can indeed take account not only of the observed frequency, but also of regularities or tendencies toward certain regularities which the observations can reveal. Suppose, for example, that the n first trials give alternately a favorable result and an unfavorable result. In the case of exchangeability, our prediction for the following trial will be the same after these n trials as after any other experience of the same frequency of $\frac{1}{2}$,[3] but with a completely irregular sequence of different results; it is indeed the absence of any influence of the order on the judgments of a certain

(3) In order that this should be exact, we suppose that n is even.

individual which characterizes, by definition, the events that he will consider "exchangeable". In the case where the events are not conceived of as exchangeable, we will, on the other hand, be led to modify our predictions in a very different way after n trials of alternating results than after n irregularly disposed trials having the same frequency of $\frac{1}{2}$; the most natural attitude will consist in predicting that the next trial will have a great probability of presenting a result opposite to that of the preceding trial.

It would doubtless be possible and interesting to study this influence of order in some simple hypotheses, by some more or less extensive generalization of the case of exchangeability, and some developments tied up with that generalization,[4] but this study is still to be done. What is essential in connection with our conception—and it is this that we want to insist on somewhat more—we have already learned through the theory of exchangeable events. Whatever be the influence of observation on predictions of the future, it never implies and never signifies that we *correct* the primitive evaluation of the probability $\mathbf{P}(E_{n+1})$ after it has been *disproved* by experience and substitute for it another $\mathbf{P}^*(E_{n+1})$ which *conforms* to that experience and is therefore probably *closer to the real probability*; on the contrary, it manifests itself solely in the sense that when experience teaches us the result A on the first n trials, our judgment will be expressed by the probability $\mathbf{P}(E_{n+1})$ no longer, but by the probability $\mathbf{P}(E_{n+1} \mid A)$, i.e. that which our initial opinion would already attribute to the event E_{n+1} considered as conditioned on the outcome A. Nothing of this initial opinion is repudiated or corrected; it is not the function $\mathbf{P}$ which has been modified (replaced by another $\mathbf{P}^*$), but rather the argument E_{n+1} which has been replaced by $E_{n+1} \mid A$, and this is just to remain faithful to our original opinion (as manifested in the choice of the function $\mathbf{P}$) and coherent in our judgment that our predictions vary when a change takes place in the known circumstances.

In the same way, someone who has the number 2374 in a lottery with 10,000 tickets will attribute at first a probability of 1/10,000 to winning the first prize, but will evaluate the probability successively as 1/1000, 1/100, 1/10, 0, when he witnesses the extraction of the successive chips which give, for example, the number 2379. At each instant his judgment

(4) One could in the first place consider the case of classes of events which can be grouped into Markov "chains" of order 1, 2, . . . , m, . . . , in the same way in which classes of exchangeable events can be related to classes of equiprobable and independent events.

is perfectly coherent, and he has no reason to say at each drawing that the preceding evaluation of probability was not right (at the time when it was made). In the last analysis, each evaluation of probability different from 0 and 1 will surely be abandoned, for a well-determined event can only happen or not happen; an evaluation of probability only makes sense when and as long as an individual does not know the result of the envisaged event; given that he does not know this result (and therefore that he is not led to the definitive value 0 or 1), he can take account of successively more circumstances which would modify his judgment, in one sense or another, without its being a question of correction or rejection. It is in just the same way that we envisage the influence of observation on prediction in the general case of judgments founded on experience.

It is thus that when the subjectivistic point of view is adopted, the problem of induction receives an answer which is naturally subjective but in itself perfectly logical, while on the other hand, when one pretends to *eliminate* the subjective factors one succeeds only in *hiding* them (that is, at least, in my opinion), more or less skillfully, but never in avoiding a gap in logic. It is true that in many cases—as for example on the hypothesis of exchangeability—these subjective factors never have too pronounced an influence, provided that the experience be rich enough; this circumstance is very important, for it explains how in certain conditions more or less close agreement between the predictions of different individuals is produced,[5] but it also shows that discordant opinions are always legitimate. This does not make any change in the purely subjective character of the whole theory of probability.

We will return in the next and last chapter to these very general questions of principle, after a review of everything we have said so far, and some additional matters which throw more light on their *raison d'être* and scope.

(5) It is the same point of view on which Poincaré several times insisted (and which inspired his well-known examples from roulette, shuffling cards, the distribution of small planets, etc.) (see, for example, [XVIII]); the only difference is in the fact that we do not admit the conception that our initial opinion concerns "unknown distributions".

CHAPTER VI
Observation and Foresight

The need for clarity in scientific and philosophical thought has never appeared to be so essential as today: the most extensive critical analysis of the clearest intuitive concepts can no longer be considered a game for sophists, but is one of the questions which touch most directly on the progress of science. With each of our assertions, a question invariably surges into our mind: has this assertion really any meaning? To give only one example, we know that the notion of simultaneity seemed, not very long ago, perfectly clear and sure, to the point that it had been thought possible to consider time as a notion given *a priori*. Why do we no longer believe this today? Because we have been taught the necessity of conceiving of every notion from a point of view which can be called "operational".[1] Every notion is only a word without meaning so long as it is not known how to verify practically any statement at all where this notion comes up; in the example given above, this practical verification is furnished us by Einstein's procedure employing light signals. An analogous evolution took place some time ago in the mathematical sciences: once, for example, the problem of knowing if $1 - 1 + 1 - 1 + \cdots = \frac{1}{2}$ or not was considered in a nebulous, mysterious, metaphysical way; it sufficed to define what was to be understood by "limit" (for example ordinary limit, limit in the sense of Cesaro) and all the obscurities vanished.

It is perfectly natural that this need for clarity is felt deeply in the domain of probability, whether because this notion is very interesting from the mathematical point of view as well as from the experimental point of view, or whether because it seems recalcitrant to all attempts to make it precise. In that which concerns the mathematical side of the question, opinions do not seem to differ too much: formally,[2] the theory of probability is the theory of additive and non-negative functions of events; opinions diverge only on one point, the question of whether these functions need to be simply or completely additive (that is to say, additive only on finite sets, or

(1) See, for example, Bridgman [II], and particularly the paragraphs *The Operational Character of Concepts* and *General Comments on the Operational Point of View* (p. 3–33).

(2) See, for example, Cantelli [IV], [V], Kolmogorov [XVII], Lomnicki [XXI], etc.

also on denumerable sets). However, the truly essential aspect of the problem is naturally the question of the meaning and the value of the notion of probability, and on this terrain opinions differ strongly. Two completely opposed points of view are possible: the first, the most commonly accepted, considers the subjective element of the naïve notion of probability which is found in our everyday life as a dangerous element which ought to be eliminated in order that the notion of probability be able to attain a truly scientific status; the opposite point of view considers, on the contrary, that the subjective elements are essential, and cannot be eliminated without depriving the notion and theory of probability of all reason for existing. The difference between the two points of view is also very sharp from the philosophical point of view: according to the one, probability is an element which partakes of the physical world and exists outside of us; according to the other, it only expresses the opinion of an individual and cannot have meaning except in relation to him.

Both of these two points of view seek to give a well-defined meaning to probability statements, but the domains in which these concepts should receive a meaning are completely different. To give a verifiable meaning to probability statements in the external world would be to consider them not as something genuinely new, but as particular statements concerning the physical world, for example, as statements about the limits of certain frequencies. However, if one wanted to interpret the requirements of the operational point of view only within the framework of the external world, in a way which could be called positivistic, I think that the goal of making all our ideas clear could never be completely attained. We are sometimes led to make a judgment which has a purely subjective meaning, and this is perfectly legitimate; but if one seeks to replace it afterward by something objective, one does not make progress, but only an error. Rather than by seeking to bring everything back to the objective, one can attain clarity by reducing any such concept systematically to the subjective; the value of a concept would then result from the analysis of the deep and essential reasons which have made us, perhaps unconsciously, introduce it, and which furnish us with the explanation of its usefulness.

This point accepted, it should not be difficult to see that the definition based on the frequency limit[3] is far from clarifying the notion of probability; indeed, even if such a definition is accepted, one will not employ

(3) See, for example, von Mises [XXV], [XXVI], Reichenbach [XXIX], [XXX], [XXXI]; see also Dörge [IX], where this conception has been modified according to criticisms which are also justified from our point of view.

the calculus of probability with the object of knowing certain values of limits of frequencies; the object of employing the calculus will always be that of judging as more or less likely the occurrence of certain facts, more or less complex, but verifiable in a finite time. These are the only events that interest us, and in regard to them the stated definition teaches us nothing. We only apply the notion of probability in order to make likely predictions: if I want to justify by the practical observation of a frequency the conviction that a neighboring frequency will appear in a certain group of *subsequent* trials, and if, for that, I proceed by estimating to begin with that the limiting frequency will be close to the observed frequency, and then that it is reasonable to expect a frequency close to the limiting frequency, I only introduce a mysterious intermediate notion through which the premises and conclusion are related indirectly by two subjective judgments in place of being related directly by a single subjective judgment [62], [63]. It does not help me at all to give the name of probability to the limiting frequency, or to any other objective entity, if the connection between these considerations and the subjective judgments which depend on them remains subjective. It is worth more, then, to seek to analyze directly the subjective element to which the notion of probability is directly anchored: it is this road that I have followed.

I know very well the doubts which are raised currently concerning this point of view, and it is for this reason that I propose to express, as clearly as it is possible for me, the way in which the problems for which the ordinary objections assume the most striking form are to be conceived and set up according to the subjectivistic conception.

There are three essential objections: it is doubted that the subjectivistic conception permits the definition of probability, the demonstration of the logical laws which govern it, and finally the explanation and justification of the applications that are made of it to the most unlike problems. Let us review rapidly our answers to these three objections.

The definition of probability that we have given is entirely irreproachable from the operational point of view, provided one admits that the latter is equally applicable in the psychological domain. The scheme of bets gives in principle a method of direct experimental measure of the degree of doubt relative to a given event. If the practical application sometimes runs up against an indeterminateness of this subjective degree of doubt which is to be measured, that is only a consequence of that limited degree of idealization without which it would always be impossible to attain precision and to develop any theory at all. The indeterminateness is doubtless stronger here than in the physical sciences because of the fact that the magnitude measured is subjective, but the difference is not essential; another definition, equally subjective and very similar, that it may perhaps be useful to compare with the subjective definition of probability, is that of the "utility" of Vilfredo Pareto [XXVII], who, in making it follow from "indifference curves", applied the operational point of view to psychological facts with perspicacity.

The fact that a direct estimation is not always possible constitutes the reason for the utility of the logical rules of probability: their practical end is to relate an evaluation, itself not very directly accessible, to others by means of which the determination of the first evaluation is made easier and more precise. By adopting the subjectivistic definition, these logical rules follow with rigor and ease from a single and very natural condition, that of *coherence*, which obliges us to take care in evaluating probabilities not to allow an adversary who bets against us the possibility of winning with certainty, whatever be the event that occurs, by a judicious combination of his stakes on the various events. The fundamental theorems (total probability, compound probability) are only the immediate corollaries of this fundamental condition. It can be seen that one could even eliminate everything quantitative, whether in the condition of coherence or in the definition of probability, in order to keep only the purely qualitative aspect of the definition (inequality between two probabilities) and of the condition of coherence (a small number of very simple axioms). The application of these logical rules in every case reduces to distinguishing whether, the probabilities being evaluated arbitrarily (but satisfying the condition of coherence) for the events of a certain class $\mathscr{E}$, the probabilities of other events are univocally determined by the condition of coherence, or whether there exists a limitation, or, finally, whether any values at all remain admissible.

From the logical point of view, the theory of probability would be only a polyvalent logic with a continuous scale of modalities,[4] superimposed on a logic of two values. This is to say that, for each event, one admits only two possible *results* (three for "conditional events", but that has only a formal significance); the infinity of intermediate modalities does not stem from an insufficiency of the logic of two values in this respect, but only serves to

(4) [62]; according to the opinions of Lukasiewicz [XXII], Mazurkiewicz [XXIII], and Reichenbach [XXIX], it would also be a question of a logic of a continuous scale of modalities, which would not, however, be conceived as superimposed on a logic of two values. Some criticisms, to which this point of view lends itself, are developed in Hosiasson [XIV].

6

measure our doubt when we do not yet know which of the two objective modalities is correct.

The subjectivistic explanation of the most important applications of the calculus of probabilities constitutes a very delicate problem. It would not be difficult to admit that the subjectivistic explication is the only one applicable in the case of practical predictions (sporting results, meteorological facts, political events, etc.) which are not ordinarily placed in the framework of the theory of probability, even in its broadest interpretation. On the other hand it will be more difficult to agree that this same explanation actually supplies rationale for the more scientific and profound value that is attributed to the notion of probability in certain classical domains, and doubts will be expressed about the possibility that it offers of unifying the various conceptions of probability, appropriate to various domains, that until now it has been thought necessary to introduce. Our point of view remains in all cases the same: *to show that there are rather profound psychological reasons which make the exact or approximate agreement that is observed between the opinions of different individuals very natural, but that there are no reasons, rational, positive, or metaphysical, that can give this fact any meaning beyond that of a simple agreement of subjective opinions.*

The case of games of chance leads only to the observation of how the character of symmetry presented by the various "possible cases" can force us to judge them equally possible, but not to *impose* such an evaluation of probability logically. The frequency case, on the other hand, requires an elaborate analysis, which leads us to some fairly extended mathematical developments. Similar reasoning, hardly more complicated than that concerning games of chance, suffices to explain the dependence between the evaluation of probabilities of certain events and the prediction of the number among them that will occur—that is to say, the bond between the evaluation of probabilities and the prediction of frequencies. The essential question, and the only one which is a little less elementary, is the justification and the explanation of the reasons for which in the prediction of a frequency one is generally guided, or at least influenced, by the observation of past frequencies. It is a question of showing that there is no need to admit, as it is currently held, that *the probability of a phenomenon has a determinate value* and that it suffices to get to know it. On the contrary, the question can be posed in a way which has a perfectly clear sense from the subjectivistic point of view, by distinguishing on the one hand the probability of a trial considered as isolated, and on the other the probability of the same trial preceded by some others of which the result is supposed (by hypothesis) to be known.

We have studied the case where, in the evaluation of conditional probabilities, one is influenced only by the observed frequency. This case can be characterized in an equivalent, but simpler and more intuitive way, as the case where the several trials of the phenomenon considered—or, in general, the event considered—play a symmetric role in relation to any problem of probability; or better yet, as the case where the probability that a given r trials have a favorable result and s others an unfavorable result depends only on r and s; or finally, as the case where the probability that n trials all have a favorable result is the same however the n-tuple is chosen. These conditions, which define "exchangeable events", have an immediate and very clear meaning from the point of view of the subjectivistic theory of probability, and there are numerous practical cases where they present themselves spontaneously to our minds. This suffices to explain our belief in the stability of the frequency, for, on this hypothesis, the probability of a subsequent trial, relative to the observation of a certain frequency, tends to coincide with the value of the latter. There is, however, a particular case, that of independence, for which the influence of past observations is rigorously zero. This case constitutes an exception; in all the other cases the influence of the acquired results tends to predominate when the number of observed cases is increased, though naturally in a way which is not uniform (one can have, for example, evaluations close to those which correspond to the case of independence for which this influence is zero). This *subjective* evaluation thus plays an essential role; the condition of "exchangeability" itself has, from the beginning, only a *subjective* value.

This reasoning is not applicable only to frequencies: in the case of exchangeable random quantities (and in general in the case of random elements of any space at all) a certain stability in the distribution of values can be justified in the same way and with similar reservations. Even the problem of smoothing a curve should be studied from this point of view:[5] the adjusted distribution curve would then be the probability distribution, conditioned on the observation of the values actually observed. This curve would depend on a subjective opinion, but to a smaller degree the richer the experience; on the other hand, one can see in this conception the true reason for the procedure of smoothing, and the conditions of "regularity" and "closeness to observed behavior" which suggest it. From this point of view, the conditions are no longer arbitrary or formal conditions, but the

(5) [56]; the point of view of Poincaré [XXVIII, pp. 204–206] is very similar.

consequences of theorems on exchangeable random quantities and of the natural tendencies of our minds in the evaluation of probabilities of different kinds.

What I have said and shown for the case of exchangeability can clearly be repeated, with the necessary modifications, for less simple and less typical conditions which have been merely alluded to (end of Chap. V). The meaning of these conclusions is always the same: observation cannot confirm or refute an opinion, which is and cannot be other than an opinion and thus neither true nor false; observation can only give us information which is capable of *influencing* our opinion. The meaning of this statement is very precise: it means that to the probability of a fact conditioned on this information—a probability very distinct from that of the same fact not conditioned on anything else—we can indeed attribute a different value.

Thus, I think I have succeeded, if not in persuading those who are far from accepting the subjectivistic point of view, at least in proving that this point of view gives an irreproachable answer to all the usual questions, and that it permits their combination into a single coherent conception. Certain minds—convinced in other respects that the subjectivistic theory of probability constitutes a coherent conception, complete and perfectly acceptable in itself—will refuse to rally to it for reasons of a philosophical sort. One might indeed think that scientific concepts ought always to have a real meaning, that science ought to occupy itself exclusively with realities, and that the subjectivistic point of view would lead further away each day from this principle through the more and more extended application of probability to the physical sciences; not only that particular branch of mathematics which constitutes the calculus of probability, not only its applications to games and statistics, but also a greater part each day of the concepts of physics would cease to correspond to an objective reality. And one might say that the deterministic laws of the classical type and the statistical laws which have been substituted[6] for them would no longer have even that common, essential characteristic which has bound them together until now, namely, the connection with reality. Would there not then be an uncrossable abyss separating the two types of laws which coexist today in physics?

In order to bridge this abyss, the point of view adopted here leads us in a completely natural way to a solution which is exactly the opposite of that which we habitually envisage: in place of extending the character of reality of the classical laws to the probability laws, we can try on the contrary to make even the classical laws participate in the subjective character of the

(6) For certain aspects of our opinions on this question, see also [36].

statistical laws. I have already cited this sentence of Poincaré: "However solidly based a prediction may seem to us, we are never absolutely sure that experience will not refute it." Laws only have value for us, in that—and only in the sense that—we estimate it *very improbable*, after experience and the scientific analysis of its results, that a "law" should be disproved by the occurrence of an event contradicting a result that it had predicted. Rigid laws are only *proven* by experience in the sense of the verification of an agreement between them and a certain number of facts. To ask if these facts occur *because the law is true*, or to ask if the true law is not *different from that which we have*, with which it coincides only in these particular cases, or finally to ask if the law *does not exist*, are questions which from the operational point of view have no sort of meaning. There is always an infinity of explanations possible for the same group of observations; if we choose one of them, and if we state a law, it can only be for subjective reasons that make us consider it worthy of confidence. Rigid laws are formulated and accepted by our minds for the same reasons that lead us to formulate and accept any judgment of probability whatever; the only difference consists in the very high probability that we attribute, in the case of rigid laws, to their exact agreement with experimental facts. The probability is so high that we can call it "practically absolute certainty", or, simply, "certainty", understanding all the while the qualification that is essential from the philosophical and logical point of view.

The notion of "cause" thus depends on the notion of probability, and it follows also from the same subjective source as do all judgments of probability [32]: this explanation seems to constitute the true logical translation of the conception of "cause" advanced by David Hume, which I consider the highest peak that has been reached by philosophy. The subjectivistic theory of probability will thus be able to open the field of science to this conception, whose significance and value seem not to have been sufficiently understood nor appreciated until now.

It is for these reasons that the theory of probability ought not to be considered an auxiliary theory for the branches of science which have not yet discovered the deterministic mechanism that "must" exist; instead it ought to be regarded as constituting the logical premises of all reasoning by induction. Just as the ordinary logic of two values is the necessary instrument of all reasoning where only the fact that an event happens or does not happen enters in, so the logic of the probable, the logic of a continuous scale of values, is the necessary instrument of all reasoning into which enters, visible or concealed, a degree of doubt, a judgment of practical certainty or practical impossibility, or finally, an estimation of the likelihood of any event whatever. Everything that does not reduce to a simple statement, to an isolated historical truth, everything that councils us for the future, even the belief that in leaving our room we will see as on other days the same streets and the same houses in their same places, all that constitutes a judgment of probability which is based, perhaps unconsciously and indistinctly, on the principles of the calculus of probability. This calculus thus constitutes the foundation of the greatest part of our thought, and we can well repeat with Poincaré, "Without it, Science would be impossible."

BIBLIOGRAPHY

[I]	Bertrand (J.)	*Calcul des probabilités*, Paris, 1889.
[II]	Bridgman (P.)	*The Logic of Modern Physics*, New York, 1927.
[III]	Cantelli (F. P.)	"Sulla probabilità come limite della frequenza," *Rendiconti Reale Accademia Nazionale dei Lincei*, Series 5, Vol. XXVI, 1st Semester (1917).
[IV]	———	"Una teoria astratta del calcolo delle probabilità," *Giornale dell'Istituto Italiano degli Attuari*, Vol. III-2 (1932).
[V]	———	"Considérations sur la convergence dans le calcul des probabilités," *Annales de l'Institut Henri Poincaré*, Vol. V, fasc. 1 (1935).
[VI]	Castelnuovo (G.)	*Calcolo delle probabilità*, Zanichelli, Bologna, 1925.
[VII]	———	"Sul problema dei momenti," *Gior. Ist. Ital. Attuari*, Vol. I-2 (1930).
[VIII]	———	"Sur quelques problèmes se rattachant au calcul des probabilités," *Ann. Inst. H. Poincaré*, Vol. III, fasc. 4 (1933).
[IX]	Dörge (K.)	"Ueber das Anwendung der Wahrscheinlichkeitsrechnung und das Induktionsproblem," *Deutsche Mathematik*, **1** (1936).
[X], [XI]	Fréchet (R. M.)	"Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'événements," *Rendiconti Reale Istituto Lombardo di Scienze Lettere ed Arte*, Series 2, Vol. LXIII, first note, fasc. 11–15, second note, fasc. 16–18 (1930).
[XII]	Fréchet (R. M.) and Halbwachs (M.)	*Le calcul des probabilités à la portée de tous*, Dunod, Paris, 1925.
[XIII]	Glivenko (V.)	"Sulla determinazione empirica delle leggi di probabilità," *Gior. Ist. Ital. Attuari*, Vol. IV-1 (1933).
[XIV]	Hosiasson (J.)	"La théorie des probabilités est-elle une logique généralisée?" Vol. IV of *Actes du Congrès International de Philosophie Scientifique* (1935), Act. Scient. Indust. #391, Hermann, Paris, 1936.
[XV]	Khinchin (A.)	"Sur les classes d'événements équivalentes," *Matematičeskii Sbornik*, *Recueil. Math. Moscou*, **39**–3 (1932).

[XVI] ——— "Remarques sur les suites d'événements obéissant à la loi des grandes nombres," *Ibid.*

[XVII] Kolmogorov (A.) *Foundations of the Theory of Probability*, 2nd English edition (translation edited by N. Morrison), New York, Chelsea, 1956.

[XVIII] ——— "Sulla determinazione empirica di una legge di distribuzione," *Gior. Ist. Ital. Attuari*, Vol. IV-1 (1933).

[XIX] von Kries (J.) *Die Prinzipien der Wahrscheinlichkeitsrechnung*, Freiburg i.b., 1886.

[XX] Lévy (P.) *Calcul des probabilités*, Paris, Gauthier-Villars, 1925.

[XXI] Lomnicki (A.) "Nouveaux fondements de la théorie des probabilités," *Fundamenta Mathematica*, Vol. 4 (1923).

[XXII] Lukasiewicz (J.) "Philosophische Bermerkungen zu mehrwertigen Systemen des Aussagenkalküls," *Comptes Rendus Soc. Sci. Lett. Varsovie*, Classe III, Vol. XXIII (1930).

[XXIII] Mazurkiewicz (S.) "Zur Axiomatik der Wahrscheinlichkeitsrechnung," *C. R. Soc. Sci. Lett. Varsovie*, Classe III, Vol. XXV (1932).

[XXIV] Medolaghi (P.) "La logica mathematica e il calcolo delle probabilità," *Bolletino Assoc. Ital. Attuari*, **18** (1907).

[XXV] von Mises (R.) *Probability, Statistics and Truth*, 2nd revised English edition (prepared by H. Geringer), New York, Macmillan, 1957.

[XXVI] ——— "Théorie des probabilités: fondements et applications," *Ann. Inst. H. Poincaré*, Vol. III, fasc. 2 (1932).

[XXVII] Pareto (V.) *Manuel d'économie politique*, Paris, Giard et Brière, 1909.

[XXVIII] Poincaré (H.) *Science and Hypothesis* (translated by W. J. Greenstreet), Chapter: The Calculus of Probability, New York, Dover, 1952.

[XXIX] Reichenbach (H.) *The Theory of Probability* (English edition translated by Ernest Hutten and Maria Reichenbach), Berkeley, 1949.

[XXX] ——— "Die Induktion als Methode der wissenschaftliche Erkenntnis," Vol. IV of *Actes du Congrès Int. de Phil. Scient.* (1935), Act. Scient. Indust. #391, Hermann, Paris, 1936.

[XXXI] ——— "Wahrscheinlichkeitslogik als form wissenschaftlichen Denkens," *Ibid.*

---

[16] "Sui passagi al limite nel calcolo delle probabilità," *Rend. R. Ist. Lombardo*, Series 2, Vol. LXIII, fasc. 2–5 (1930).

[24] "A proposito dell'estensione del teorema delle probabilità totali alle classi numerabili," *ibid.*, fasc. 11–15 (1930).

[26] "Problemi determinati e indeterminati nel calcolo delle probabilità," *Rend. R. Acc. Naz. Lincei*, Series 6, Vol. XII, fasc. 9 (Nov., 1930).

[27] "Fondamenti logici del ragionamento probabilistico," *Boll. Un. Mat. Ital.*, Vol. IX-5 (Dec., 1930).

[28] "Ancora sull'estensione alle classi numerabili del teorema delle probabilità totali," *Rend. R. Ist. Lombardo*, Series 2, Vol. LXIII, fasc. 16–18 (1930).

[29] "Funzione caratteristica di un fenomeno aleatorio," *Memorie R. Acc. Naz. Lincei*, Series 6, Vol. IV, fasc. 5 (1930).

[32] "Probabilismo: Saggio critico sulla teoria delle probabilità e sul valore della scienze," *Biblioteca di Filosofia diretta da Antonio Aliotta*, Naples, Perrella (1931).

[34] "Sul significato soggettivo della probabilità," *Fund. Math.*, Vol. 17, Warsaw (1931).

[35] "Sui fondamenti logici del ragionamento probabilistico," *Atti. Soc. Ital. Progr. Scienze*, Riunione Bolzano-Trento del 1930, Vol. II, Rome (1931).

[36] "Le leggi differenziali e la rinunzia al determinismo," *Rend. Semin. Mat. R. Univ. Roma*, Series 2, Vol. VII (1931).

[38] "Probabilità fuori dagli schemi di urne," *Period. di Mat.*, Series 4, Vol. XII-1 (1932).

[40] "Funzione caratteristica di un fenomeno aleatorio," *Att. Congr. Int. Mathem., Bologna*, 1928, Vol. 6, Bologna, Zanichelli ed. (1932).

[45] "Sull'approssimazione empirica di una legge di probabilità," *Gior. Ist. Ital. Attuari*, Vol. 4–3 (1933).

[46 to 48] "Classi di numeri aleatori equivalenti. La legge dei grandi numeri nel caso dei numeri aleatori equivalenti. Sulla legge di distrubuzione dei valori in una successione di numeri aleatori equivalenti," *Rend. R. Acc. Naz. Lincei*, Series 6, Vol. XVIII, 2e sem., fasc. 3–4 (1933).

[49] "Sul concetto di probabilità," *Riv. Ital. Statist.*, Vol. 1 (1933).

[51] "Indipendenza stocastica ed equivalenza stocastica," *Atti. Soc. Ital. Prog. Scienze*, Riunione Bari del 1933, Vol. II, Rome (1934).

[R. 2.] Review of Reichenbach, H. [XXIX] *Zentralblatt für Mathematik und ihre Grenzegebiete*, **10**-8 (1935).

[56] "Il problema della perequazione," *Atti. Soc. Ital. Progr. Scienze*, Riunione Napoli del 1934, Vol. II, Rome (1935).

[58] (In collaboration with M. M. Jacob). "Sull'integrale di Stieltjes-Riemann," *Gior. Ist. Ital. Attuari*, Vol. VI-4 (1935).

[62] "La logique de la probabilité" (Communication to the Congrès Int. de Philosophie Scient., Paris, 1935), Act. Scient. Indust. #391, Hermann, 1936.

[64] "Les probabilités nulles," *Bull. Sci. Math.*, Second Series, **60** (1936).

[65] "Statistica e probabilità nella concezione di R. von Mises," *Suppl. Statis. Nuovi Probl.*, Vol. II-3 (1936).

[70] "Riflessioni teoriche sulle assicurazioni elementari," Vol. 2 of *Comptes Rendus du Onzième Congrès International d'Actuaires* (Paris, June 17–24, 1936), Gauthier-Villars, Paris, 1937.