AMERICAN JOURNAL of PHYSICS

A Journal Devoted to the Instructional and Cultural Aspects of Physical Science

VOLUME 14, NUMBER 1 JANUARY–FEBRUARY, 1946

Probability, Frequency and Reasonable Expectation

R. T. COX
The Johns Hopkins University, Baltimore 18, Maryland

Frequency or Reasonable Expectation as the Primary Concept

THE concept of probability has from the beginning of the theory involved two ideas: the idea of frequency in an ensemble and the idea of reasonable expectation. The choice of one or the other as the primary meaning of probability has distinguished the two main schools of thought in the theory.[1]

If a box contains two white balls and one black ball, indistinguishable except by color, both schools agree that the probability that a blindfolded man will draw a white ball on a single trial is $\frac{2}{3}$ and the probability that he will draw a black ball is $\frac{1}{3}$. On the frequency theory, the primary meaning of these probabilities is in terms of the ensemble. The ensemble may be an indefinitely large number of such boxes having the same contents, or it may be an indefinitely large number of drawings from the same box, the ball drawn being replaced each time. The significant point is that the initial circumstances are assumed to be capable of indefinite repetition, these repetitions constituting the ensemble. That the probability of a white ball is $\frac{2}{3}$ means simply that the number of trials giving a white ball as result is $\frac{2}{3}$ the number of trials in the whole ensemble. According to the frequency theory, this is not a prediction of the theory of probability but the definition of the probability. Probability in that theory is a characteristic of the ensemble and, without the ensemble, cannot be said to exist.

Again, according to both schools, the probability of a white ball in two successive drawings, when the first ball drawn is not replaced, is $\frac{2}{3}\times\frac{1}{2}$, or $\frac{1}{3}$. According to the frequency theory, this implies that two balls are drawn successively from each of an ensemble of boxes containing originally two white balls and one black ball. On $\frac{2}{3}$ of the trials a white ball is drawn first and one white and one black ball are left in the box. Then, in $\frac{1}{2}$ of the trials which give this result, a white ball is drawn next, so that $\frac{1}{3}$ of the whole number of trials give white balls on both drawings. These examples illustrate the general fact that, when probability is identified with frequency in an ensemble, the probabilities are calculated by arithmetic in particular examples and, in more general cases, the rules of probability are found by ordinary algebra.[2]

---

[1] If minor differences are counted, the number of schools seems to be somewhere between two and the number of authors, and probably nearer the latter number. But the clearest line of division is the one mentioned.

[2] An exposition of the frequency theory, with some comment on other theories, has been given by G. Bergmann, *Am. J. Physics* 9, 263 (1941). Readers with a wider knowledge of philosophy than mine will be better able to compare his views with those of this paper.

Probability is recognized also as providing a measure of the reasonable expectation of an event in a single trial. That the probability of drawing a white ball is $\frac{2}{3}$ and of drawing a black ball is $\frac{1}{3}$ means that a white ball is a more likely result of a trial than a black ball, and the numbers $\frac{2}{3}$ and $\frac{1}{3}$ serve to compare the likelihoods of the two results. According to the second main school of probability, this measure of reasonable expectation, rather than the frequency in an ensemble, is the primary meaning of probability.

If it could be shown that every measure of reasonable expectation is also a frequency in some ensemble and that every frequency in an ensemble measures a reasonable expectation, then the choice of one or the other as the primary meaning of probability would not be very important. I shall not attempt to discuss whether there are frequencies in an ensemble that are not measures of reasonable expectation. It is enough for my present purpose to show that the two interpretations are not always identical. For this it will suffice to point out that there are probabilities in the sense of reasonable expectations for which no ensemble exists and for which, if one is conceived, it is clearly no more than a convenient mental artifice. Thus, when the probability is calculated that more than one planetary system exists in the universe, it is barely tenable even as an artifice that this refers to the number of universes having more than one planetary system among an indefinitely large number of universes, all resembling in some way *the* universe, which by definition is all-inclusive.

Moreover, there is so gradual a transition from the cases in which there is a discoverable ensemble and those in which there is none that a theory which requires a sharp distinction between them offers serious difficulties. A few examples will illustrate this point. Let us consider the probability that the number of heads thrown in a certain number of tosses of an unbiased coin shall lie within certain limits, and let us compare with this the probability, often considered, that the true value of a physical constant lies within certain limits. The two probabilities have something in common, but there is a difference between them. The difference lies in the causes that oblige us to deal with probabilities rather than certainties in discussing the score in tossing a coin and the value of a physical constant. In discussing the score in a given number of tosses of a coin, we have to use probabilities because the score will vary from one trial to another. The true value of a physical constant, on the other hand, is unique. We have to speak of the probability that it lies within certain limits only because our knowledge is incomplete.

Sometimes, it is true, the probability that the value of a physical constant lies within certain limits is equivalent to another probability, that the error in the average of a number of measurements lies within these limits. If there are no systematic sources of error, we may imagine an ensemble of measurements and treat the measurements made as a random sample of this ensemble. The probability in question may then be found in a manner similar to that used in dealing with the coin. For example, the probability of certain limits for the true value of the Joule equivalent may perhaps be considered in this way.

The case is somewhat different with the reciprocal fine-structure constant that appears in quantum mechanics. For here, in addition to the values derived from measurements, there is evidence of another sort in the argument adduced by Eddington[3] that this constant may be expected to be an integer, having the value 137. If it should be estimated from the measurements alone that there is an equal probability that the constant lies inside or outside of certain limits which include 137, then Eddington's argument will increase the probability that it lies inside these limits and correspondingly decrease the probability that it lies outside.

As a final example, we may consider the case of a purely mathematical constant, of which the existence has been proved but the value determined only within certain limits. A problem of the theory of numbers, discussed by Hardy[4] among others, provides a good example. It concerns the equivalence of an integer to a sum

[3] A. S. Eddington, *Relativity theory of protons and electrons* (Macmillan, New York, and Univ. Press, Cambridge, 1936).

[4] G. H. Hardy, *Some famous problems of the theory of numbers and in particular Waring's problem* (Clarendon Press, Oxford, 1920).

of cubes of smaller integers. It has been proved that any integer is given by the sum of not more than 9 cubes, and that any integer beyond a certain one is given by the sum of 8 or fewer. It is expected that large enough integers may all be expressed by the sum of some still smaller number of cubes, and the problem is to find the minimum number required for all integers above a certain value. It has been proved that, if this value is taken high enough, the number in question is 4, 5, 6, 7 or 8. This is as far as rigorous proof has gone, but the evidence of computation makes some of these numbers less likely than others. In very large samples—all of the first 40,000 integers and the 2000 ending at one million—the integers requiring 8 cubes are found to drop out early in the progress to higher integers and those requiring 7 disappear somewhat farther on, while those requiring 6 occur more and more rarely until there are only two among the 2000 integers next below one million. Hardy concludes that the minimum number for large enough integers is almost certainly neither 8 nor 7 and probably not 6. There remain 5 and 4 as the likely numbers, and he seems to favor 4 as the more probable.

Let us consider now these four examples: the probability of certain limits for (i) the score in a number of throws of a coin, (ii) the value of the Joule equivalent, (iii) the value of the reciprocal fine-structure constant and (iv) the value of the least number of cubes for the expression of large integers. It will most likely be granted that other examples can be interpolated among these, so that the differences will be very slight between each example and those next before and after. We shall have then a graded series of examples of probability. At one end of the series the interpretation of probability in terms of frequency will be valid, at the other end it will be impossible. For it is certainly impossible to discuss the statistical spread of the determinations of a number which has never in fact been determined and of which the determination, when it is made, will give a single and logically inevitable value.

Nevertheless, it must be admitted that there is a kind of reasoning common to all these examples. The gambler in the first example, the physicist in the second and third, and the mathematician in the fourth are all using similar processes of inference.

In this connection it is worth while to observe how much of the theory of probability deals with relations between probabilities: between the probability that an event will not occur and the probability that it will occur, between the probability of both of two events and their separate probabilities, between these probabilities and the probability that at least one of the two events will occur. In the case of probabilities that can be identified with frequencies in an ensemble, these relations are readily obtained by ordinary algebra, as was mentioned earlier. But the same or at least similar relations are involved in inference concerned with reasonable expectation even when no ensemble is discoverable. Thus, under any definition of probability, or even without an attempt to define it precisely, there will still be agreement that the less likely an event is to occur the more likely it is not to occur. The occurrence of both of two events will not be more likely and will generally be less likely than the occurrence of the less likely of the two. But the occurrence of at least one of the events is not less likely and is generally more likely than the occurrence of either.

For example, if it were a question of the credibility of a certain hypothesis for the origin of terrestrial life or of human language, one would hold it as a point against the hypothesis that it postulated the occurrence of two events, of which neither was considered very probable, but the hypothesis would gain in credibility if it could be justified by postulating merely that one or the other of these events had occurred. Generally speaking, a simple hypothesis is preferred to a complex one. If this preference is founded on a reasonable belief rather than being a mere convention, its justification would seem to be that two or more postulates are less likely to be true than a single one of about the same likelihood.

This difficulty of the frequency theory of probability may now be summarized. There is a field of probable inference which lies outside the range of that theory. The derivation of the rules of probability by ordinary algebra from the characteristics of the ensemble cannot justify the use of these rules in this outside field.

Nevertheless, the use of these rules in this field is universal and appears to be a fundamental part of our reasoning. Thus the frequency theory is inadequate in the sense that it fails to justify what is conceived to be a legitimate use of its own rules.

From a purely rational point of view, the extent of this field of inference outside the range of the frequency theory is irrelevant to the point in question. Even if the valid instances of reasoning in this field were rare and of little consequence, it would still be logically necessary to maintain the inadequacy of the frequency theory. As a practical matter, however, if these instances were few or trivial, we should probably be content to ignore them. But actually, as I have tried to suggest by the examples given, it is rather the cases in which a strictly definable ensemble exists that are exceptional. This is not to say that they are numerically few. There are many of them, and they have a particular interest, but they still do not appear to comprise in our ordinary practice the greater part of the uses of probable inference. Nor are the other uses by any means trivial. Kemble,[5] in an interesting paper which covers, among other things, some of the ground thus far traversed here, has made the point that the frequency definition of probability does not suffice to establish the connection between statistical mechanics and thermodynamics, which is certainly crucial in physical theory.

A very original and thoroughgoing development of the theory of probability, which does not depend on the concept of frequency in an ensemble, has been given by Keynes.[6] In his view, the theory of probability is an extended logic, the logic of probable inference. Probability is a relation between a hypothesis and a conclusion, corresponding to the degree of rational belief and limited by the extreme relations of certainty and impossibility. Classical deductive logic, which deals with these limiting relations only, is a special case in this more general development. Hence it follows in general that the theory of probability cannot be based entirely on concepts of classical logic. In particular, the relation of probability cannot be defined in terms of certainty, since certainty itself is a special case of probability. The frequency definition of probability is therefore invalid, since it depends on the relations of certainty involved in the knowledge of numbers of instances. Probability is taken as a primary concept, like distance or time in mechanics, not reducible to any more elementary terms.

Merely to describe Keynes' position, as I have done, without giving the reasoning by which he is led to it, does his work very poor justice. The reasoning is, to me at least, very convincing, and is the original source of a large part of the opinions given here, though the arguments I have used are not the same as his. Nevertheless, it must be conceded that his work does not bring us very far in the solution of the problem mentioned earlier, that of justifying the few basic rules of probable inference necessary for the development of the theory. These rules, in Keynes' theory, are simply taken as axiomatic. Now some primary assumptions will have to be made by anyone who accepts, as I am strongly inclined to do, his general point of view as to the nature of probability, because some rational starting point is needed to replace the frequency definition, once that has been abandoned. But Keynes' axioms seem to me, as they have doubtless seemed to others, including Kemble, somewhat too arbitrary and too sophisticated to be entirely suitable as axioms. They do not appeal very directly to common sense, and it is hard to see how they would have been formulated without considering colored balls in a box, dice, coins, or some of the other devices associated with the concept of the ensemble. It is rather as if Euclid had placed the Pythagorean theorem among the axioms of plane geometry.

Russell[7] makes a criticism somewhat different in form, but which may have the same ground as this. After conceding the strength of Keynes' argument against the frequency theory, he nevertheless prefers that theory, if it can be logically established, because of its explicit definition of probability. It is this definition that makes it possible to avoid the assumption of axioms such as characterize Keynes' theory.

[5] E. C. Kemble, *Am. J. Physics* **10**, 6 (1942).

[6] J. M. Keynes, *A treatise on probability* (Macmillan, London, 1929).

[7] B. Russell, *Philosophy* (Norton, New York, 1927).

Other authors, who, like Keynes, present an axiomatic development, choose somewhat different sets of postulates, but those I have seen still show some of the tool marks of their original derivation from the study of games of chance, with the consequent implication of an ensemble. I think this is true even of the carefully chosen postulates of Jeffreys and Wrinch, whether in their original form or as revised by Jeffreys.[8]

Relations of Reasonable Expectation Consistent with Symbolic Logic

In what follows next, I shall try to show that by employing the algebra of symbolic logic it is possible to derive the rules of probability from two quite primitive notions, which are independent of the concept of the ensemble and which, as I think, appeal rather immediately to common sense. This algebra has been applied to probability by a number of writers, including Boole,[9] who originated it. Still, its possibilities in this respect do not seem to have been fully realized. It may be well here to give a brief introduction to the Booleian algebra, at least to as much of it as the later argument will require.

Letters, **a**, **b**, **c**, $\cdots$, will denote propositions. There is an advantage in speaking of the probabilities of propositions rather than of events, partly for the sake of greater generality but mainly because speaking of events easily invokes the notion of sequence in time, and this may become a source of confusion. A proposition may, of course, assert the occurrence of an event, but it may just as well assert something else, for example, something about a physical constant. The proposition not-**a** will be denoted by $\sim\mathbf{a}$, the proposition **a**-and-**b** by $\mathbf{a}\cdot\mathbf{b}$, and the proposition **a**-or-**b** by $\mathbf{a}\vee\mathbf{b}$.

It is to be borne in mind that the proposition $\sim\mathbf{a}$ is not the particular proposition which in some sense is the opposite of **a**. Thus if **a** is the proposition, "The stranger was a short, fat old man without coat or hat," $\sim\mathbf{a}$ is not the proposition, "The stranger was a tall, thin young woman with coat and hat." To assert $\sim\mathbf{a}$ means nothing more than to answer "no" to the question, "Is **a** wholly true?" If **a** is in several parts, $\mathbf{a}_1, \mathbf{a}_2, \cdots$, to assert $\sim\mathbf{a}$ is not to affirm that $\mathbf{a}_1, \mathbf{a}_2, \cdots$ are all false but only to say that at least one of them is false.

Since the letters **a**, **b** denote propositions and not events, the order in which they appear in the symbols $\mathbf{a}\cdot\mathbf{b}$ and $\mathbf{a}\vee\mathbf{b}$ is only the order in which two propositions are stated, not the order in time in which two events occur. Also the form $\mathbf{a}\cdot\mathbf{a}$ indicates only that a proposition is twice stated, not that an event has twice occurred.

It is also to be understood that $\mathbf{a}\vee\mathbf{b}$ means **a**-or-**b** in the sense of the child who asks, "May I have a nickel or a dime?" without meaning to exclude the possibility of both a nickel and a dime, not in the sense of the orator saying "Sink or swim, survive or perish." Thus $\mathbf{a}\vee\mathbf{b}$ has the sense for which the form **a** and/or **b** is often employed.

Finally it may be noted that if the proposition, "It is raining," is true, then the proposition, "It is raining or snowing," is also true. To assert a proposition **a** is to imply every proposition $\mathbf{a}\vee\mathbf{b}$ of which **a** is one term.

With the meaning of the symbols thus understood, the rules for their combination may be set down as follows:

$$\sim\sim\mathbf{a}=\mathbf{a}, \tag{1}$$

$$\mathbf{a}\cdot\mathbf{b}=\mathbf{b}\cdot\mathbf{a}, \quad (2) \qquad \mathbf{a}\vee\mathbf{b}=\mathbf{b}\vee\mathbf{a}, \quad (2')$$

$$\mathbf{a}\cdot\mathbf{a}=\mathbf{a}, \quad (3) \qquad \mathbf{a}\vee\mathbf{a}=\mathbf{a}, \quad (3')$$

$$\mathbf{a}\cdot(\mathbf{b}\cdot\mathbf{c})=(\mathbf{a}\cdot\mathbf{b})\cdot\mathbf{c}=\mathbf{a}\cdot\mathbf{b}\cdot\mathbf{c}, \tag{4}$$

$$\mathbf{a}\vee(\mathbf{b}\vee\mathbf{c})=(\mathbf{a}\vee\mathbf{b})\vee\mathbf{c}=\mathbf{a}\vee\mathbf{b}\vee\mathbf{c}, \tag{4'}$$

$$\sim(\mathbf{a}\cdot\mathbf{b})=\sim\mathbf{a}\vee\sim\mathbf{b}, \tag{5}$$

$$\sim(\mathbf{a}\vee\mathbf{b})=\sim\mathbf{a}\cdot\sim\mathbf{b}, \tag{5'}$$

$$\mathbf{a}\cdot(\mathbf{a}\vee\mathbf{b})=\mathbf{a}, \quad (6) \qquad \mathbf{a}\vee(\mathbf{a}\cdot\mathbf{b})=\mathbf{a}. \quad (6')$$

These eleven rules are not all independent. From six of them it is possible to prove the remaining five, and the set of six may be chosen in various ways. It is necessary only to include the first and one from each similarly numbered pair of the others. Thus, for example, Eq. (5′) is derived as follows.

$$\begin{aligned}\sim(\mathbf{a}\vee\mathbf{b}) &= [\text{by Eq. (1)}]\sim(\sim\sim\mathbf{a}\vee\sim\sim\mathbf{b})\\ &= [\text{by Eq. (5)}]\sim\sim(\sim\mathbf{a}\cdot\sim\mathbf{b})\\ &= [\text{by Eq. (1)}]\sim\mathbf{a}\cdot\sim\mathbf{b}.\end{aligned}$$

[8] H. Jeffreys, *Theory of probability* (Clarendon Press, Oxford, 1939).
[9] G. Boole, *An investigation of the laws of thought* (Macmillan, London, 1854).

Now let the symbol $\mathbf{b}|\mathbf{a}$ denote some measure of the reasonable credibility of the proposition $\mathbf{b}$ when the proposition $\mathbf{a}$ is known to be true.[10] The term is indefinite at this point, because, if there is one such measure, there will be any number of others. If $\mathbf{b}|\mathbf{a}$ is one such measure, then an arbitrary function $f(\mathbf{b}|\mathbf{a})$ will also be a measure. Consequently, the symbol is not now to be identified with the conventional probability. To avoid that implication, I shall call $\mathbf{b}|\mathbf{a}$ the *likelihood* of the proposition $\mathbf{b}$ on the hypothesis $\mathbf{a}$, taking advantage of a suggestion made by Margenau,[11] but at the same time taking the liberty of giving the term a more inclusive meaning than the one he proposed.

It is not to be supposed that a relation of likelihood exists between any two propositions. If $\mathbf{a}$ is the proposition "Caesar invaded Britain" and $\mathbf{b}$ is "Tomorrow will be warmer than today," there is no likelihood $\mathbf{b}|\mathbf{a}$, because there is no reasonable connection between the two propositions.

It is now time to make the first of the two assumptions mentioned earlier as providing a basis for the principles of probable inference. We assume, whatever measure be chosen, that the likelihood $\mathbf{c}\cdot\mathbf{b}|\mathbf{a}$ is determined in some way by the two likelihoods $\mathbf{b}|\mathbf{a}$ and $\mathbf{c}|\mathbf{b}\cdot\mathbf{a}$, or

$$\mathbf{c}\cdot\mathbf{b}|\mathbf{a}=F(\mathbf{c}|\mathbf{b}\cdot\mathbf{a},\,\mathbf{b}|\mathbf{a}),\qquad(7)$$

where F is some function of two variables.

Written in symbolic form, this assumption may not appear very axiomatic. Actually it is a familiar enough rule of common sense, as an example will show. Let $\mathbf{b}$ denote the proposition that an athlete can run from one given place to another, and let $\mathbf{c}$ denote the proposition that he can run back without stopping. The physical condition of the runner and the topography of the course are described in the hypothesis $\mathbf{a}$. Then $\mathbf{b}|\mathbf{a}$ is the likelihood that he can run to the distant place, estimated on the information given in $\mathbf{a}$, and $\mathbf{c}|\mathbf{b}\cdot\mathbf{a}$ is the likelihood that he can run back, estimated on the initial information and the further assumption that he has just run one way. These are just the likelihoods that would have to be considered in estimating the likelihood, $\mathbf{c}\cdot\mathbf{b}|\mathbf{a}$, that he can run the complete course without stopping. In postulating only that the last-named likelihood is some function of the other two, we are making the least restrictive assumption possible.

The form of the function F is partly conventional because of the indefiniteness of the measure to be used for likelihood. But it is not wholly so, for it must be consistent with the algebra of propositions. Accordingly we make use of Eq. (4) to derive a functional equation involving F, as follows:

$$\begin{aligned}\mathbf{d}\cdot\mathbf{c}\cdot\mathbf{b}|\mathbf{a}&=[\text{by Eq. (4)}](\mathbf{d}\cdot\mathbf{c})\cdot\mathbf{b}|\mathbf{a}\\&=[\text{by Eq. (7)}]F(\mathbf{d}\cdot\mathbf{c}|\mathbf{b}\cdot\mathbf{a},\,\mathbf{b}|\mathbf{a}).\end{aligned}$$

But

$$\begin{aligned}\mathbf{d}\cdot\mathbf{c}|\mathbf{b}\cdot\mathbf{a}&=[\text{by Eq. (7)}]F[\mathbf{d}|\mathbf{c}\cdot(\mathbf{b}\cdot\mathbf{a}),\,\mathbf{c}|\mathbf{b}\cdot\mathbf{a}]\\&=[\text{by Eq. (4)}]F(\mathbf{d}|\mathbf{c}\cdot\mathbf{b}\cdot\mathbf{a},\,\mathbf{c}|\mathbf{b}\cdot\mathbf{a}).\end{aligned}$$

Hence

$$\mathbf{d}\cdot\mathbf{c}\cdot\mathbf{b}|\mathbf{a}=F[F(\mathbf{d}|\mathbf{c}\cdot\mathbf{b}\cdot\mathbf{a},\,\mathbf{c}|\mathbf{b}\cdot\mathbf{a}),\,\mathbf{b}|\mathbf{a}].$$

Also

$$\begin{aligned}\mathbf{d}\cdot\mathbf{c}\cdot\mathbf{b}|\mathbf{a}&=[\text{by Eq. (4)}]\mathbf{d}\cdot(\mathbf{c}\cdot\mathbf{b})|\mathbf{a}\\&=[\text{by Eq. (7)}]F[\mathbf{d}|(\mathbf{c}\cdot\mathbf{b})\cdot\mathbf{a},\,\mathbf{c}\cdot\mathbf{b}|\mathbf{a}]\\&=[\text{by Eqs. (4) and (7)}]\\&\qquad F[\mathbf{d}|\mathbf{c}\cdot\mathbf{b}\cdot\mathbf{a},\,F(\mathbf{c}|\mathbf{b}\cdot\mathbf{a},\,\mathbf{b}|\mathbf{a})].\end{aligned}$$

Equating these two expressions for $\mathbf{d}\cdot\mathbf{c}\cdot\mathbf{b}|\mathbf{a}$ and, for simplicity, letting $\mathbf{d}|\mathbf{c}\cdot\mathbf{b}\cdot\mathbf{a}=x$, $\mathbf{c}|\mathbf{b}\cdot\mathbf{a}=y$, and $\mathbf{b}|\mathbf{a}=z$, we have

$$F[F(x,y),z]=F[x,F(y,z)].\qquad(8)$$

The function F must be such as to satisfy Eq. (8) for arbitrary values of x, y and z. It is easily shown by substitution that this equation is satisfied if

$$Cf[F(p,q)]=f(p)f(q),$$

where f is an arbitrary function of a single variable, and C is an arbitrary constant. It is shown in the appendix that this is also the general solution, provided F has continuous second derivatives. We have then

$$Cf(\mathbf{c}\cdot\mathbf{b}|\mathbf{a})=f(\mathbf{c}|\mathbf{b}\cdot\mathbf{a})f(\mathbf{b}|\mathbf{a}).$$

The choice of the function f is purely a matter of convention. For it has already been pointed out that, if $\mathbf{b}|\mathbf{a}$ is a measure of the credibility of

---

[10] Keynes has traced the use of such a symbol to H. McColl, *Proc. Lond. Math. Soc.* **11**, 113 (1880). McColl uses the symbol x_a for the probability of the proposition x on the hypothesis a.

[11] H. Margenau, *Am. J. Physics* **10**, 224 (1942). R. A. Fisher has used the term in a quite different sense.

$\mathbf{b}$ on the hypothesis $\mathbf{a}$, then so also is $\mathrm{f}(\mathbf{b}|\mathbf{a})$. We might then continue the discussion with $\mathrm{f}(\mathbf{b}|\mathbf{a})$ as the symbol of likelihood in place of $\mathbf{b}|\mathbf{a}$ and never have to specify the function f. But this would give two symbols where one would be enough. As a matter of convenience, therefore, we write

$$C\mathbf{c}\cdot\mathbf{b}|\mathbf{a}=\mathbf{c}|\mathbf{b}\cdot\mathbf{a}\ \mathbf{b}|\mathbf{a}. \tag{9}$$

This is, of course, the same as choosing the function f to make $\mathrm{f}(\mathbf{b}|\mathbf{a})=\mathbf{b}|\mathbf{a}$. Since the choice was conventional, it follows that another choice could have been made. We might, for example, have let $\mathrm{f}(\mathbf{b}|\mathbf{a})=\exp(\mathbf{b}|\mathbf{a})$, whence it would have followed that the likelihood of $\mathbf{c}\cdot\mathbf{b}$ was, except for an arbitrary additive constant, equal to the sum of the likelihoods which determine it. This would have given us a likelihood related to the one we have as entropy is related to thermodynamic probability in statistical mechanics. It would have been an allowable choice, but a less convenient one than that which was made.

If in Eq. (9) we let $\mathbf{c}=\mathbf{b}$, and note that $\mathbf{b}\cdot\mathbf{b}=\mathbf{b}$, by Eq. (3), we obtain, after dividing by $\mathbf{b}|\mathbf{a}$,

$$C=\mathbf{b}|\mathbf{b}\cdot\mathbf{a}.$$

Thus we see that when the hypothesis includes the conclusion the likelihood has the constant value C, whatever the propositions may be. This is what we should expect, since $\mathbf{b}$ is certain on the hypothesis $\mathbf{b}\cdot\mathbf{a}$, and we do not recognize degrees of certainty.

The value to be assigned to C, the likelihood of certainty, is purely conventional. If it is desired to make the likelihoods with which we are dealing correspond as nearly as possible to ordinary probabilities, then C will be given the value 1. Other choices are often made, especially in conversation. The phrase "one chance in a hundred" may be taken to mean unit likelihood on a scale in which certainty is represented by 100. Statements that have the form of assertions about numbers in an ensemble may be merely convenient ways of stating likelihoods on a scale chosen for its aptness to the question considered. In a general discussion the most convenient value for C is unity, and we therefore write Eq. (9) in the form

$$\mathbf{c}\cdot\mathbf{b}|\mathbf{a}=\mathbf{c}|\mathbf{b}\cdot\mathbf{a}\ \mathbf{b}|\mathbf{a}. \tag{10}$$

This has the same form as the ordinary rule for the probability of two events. However, it does not make our likelihood correspond uniquely to the ordinary probability. For Eq. (10) raised to any power m is

$$(\mathbf{c}\cdot\mathbf{b}|\mathbf{a})^m=(\mathbf{c}|\mathbf{b}\cdot\mathbf{a})^m(\mathbf{b}|\mathbf{a})^m.$$

Thus any power of our likelihood satisfies an equation of the same form as Eq. (10) and corresponds equally well to the ordinary probability.

Next to be sought is a second assumption of probable inference, which is to provide a relation between the likelihoods of the propositions $\mathbf{b}$ and $\sim\mathbf{b}$ on the same hypothesis $\mathbf{a}$. Since $\sim\mathbf{b}$ is determined when $\mathbf{b}$ is specified, a reasonable assumption, and the least restrictive possible, appears to be that $\sim\mathbf{b}|\mathbf{a}$ is determined by $\mathbf{b}|\mathbf{a}$, or

$$\sim\mathbf{b}|\mathbf{a}=\mathrm{S}(\mathbf{b}|\mathbf{a}), \tag{11}$$

where S is some function of a single variable.

By Eq. (1), $\sim\sim\mathbf{b}|\mathbf{a}=\mathbf{b}|\mathbf{a}$, and therefore $\mathrm{S}[\mathrm{S}(\mathbf{b}|\mathbf{a})]=\mathbf{b}|\mathbf{a}$. Thus S must be such a function that

$$\mathrm{S}[\mathrm{S}(x)]=x, \tag{12}$$

where x may have any possible value of a likelihood between those of certainty and impossibility. This does not impose enough restriction on S to be of much use by itself. Another functional equation may be obtained by considering $\mathrm{S}(\mathbf{c}\vee\mathbf{b}|\mathbf{a})$; thus,

$$\begin{aligned}\mathrm{S}(\mathbf{c}\vee\mathbf{b}|\mathbf{a})&=\sim(\mathbf{c}\vee\mathbf{b})|\mathbf{a}\\ &=[\text{by Eq. }(5')]\sim\mathbf{c}\cdot\sim\mathbf{b}|\mathbf{a}.\end{aligned}$$

We wish to eliminate the propositions $\sim\mathbf{c}$ and $\sim\mathbf{b}$, so as to obtain an equation in the propositions $\mathbf{c}$, $\mathbf{b}$ and $\mathbf{a}$ and the function S. First we eliminate $\sim\mathbf{c}$.

$$\begin{aligned}\sim\mathbf{c}\cdot\sim\mathbf{b}|\mathbf{a}&=[\text{by Eq. (10)}]\sim\mathbf{c}|\sim\mathbf{b}\cdot\mathbf{a}\ \sim\mathbf{b}|\mathbf{a}\\ &=[\text{by Eq. (11)}]\mathrm{S}(\mathbf{c}|\sim\mathbf{b}\cdot\mathbf{a})\mathrm{S}(\mathbf{b}|\mathbf{a}).\end{aligned}$$

Thus we have

$$\mathrm{S}(\mathbf{c}\vee\mathbf{b}|\mathbf{a})=\mathrm{S}(\mathbf{c}|\sim\mathbf{b}\cdot\mathbf{a})\mathrm{S}(\mathbf{b}|\mathbf{a}),$$

or

$$\mathrm{S}(\mathbf{c}|\sim\mathbf{b}\cdot\mathbf{a})=\mathrm{S}(\mathbf{c}\vee\mathbf{b}|\mathbf{a})/\mathrm{S}(\mathbf{b}|\mathbf{a}).$$

Taking the function S of both sides of this equation and using Eq. (12), we obtain

$$\mathbf{c}|\sim\mathbf{b}\cdot\mathbf{a}=\mathrm{S}[\mathrm{S}(\mathbf{c}\vee\mathbf{b}|\mathbf{a})/\mathrm{S}(\mathbf{b}|\mathbf{a})]. \tag{13}$$

Next we eliminate $\sim\mathbf{b}$:

$$\mathbf{c}|\sim\mathbf{b}\cdot\mathbf{a}=[\text{by Eq. (10)}]\mathbf{c}\cdot\sim\mathbf{b}|\mathbf{a}/\sim\mathbf{b}|\mathbf{a}$$
$$=[\text{by Eq. (2)}]\sim\mathbf{b}\cdot\mathbf{c}|\mathbf{a}/\sim\mathbf{b}|\mathbf{a}$$
$$=[\text{by Eq. (10)}]\sim\mathbf{b}|\mathbf{c}\cdot\mathbf{a}\,\mathbf{c}|\mathbf{a}/\sim\mathbf{b}|\mathbf{a}$$
$$=[\text{by Eq. (11)}]\mathrm{S}(\mathbf{b}|\mathbf{c}\cdot\mathbf{a})\mathbf{c}|\mathbf{a}/\mathrm{S}(\mathbf{b}|\mathbf{a}).$$

Therefore we may write in place of Eq. (13),

$$\mathrm{S}(\mathbf{b}|\mathbf{c}\cdot\mathbf{a})\mathbf{c}|\mathbf{a}/\mathrm{S}(\mathbf{b}|\mathbf{a})=\mathrm{S}[\mathrm{S}(\mathbf{c}\vee\mathbf{b}|\mathbf{a})/\mathrm{S}(\mathbf{b}|\mathbf{a})].$$

It is convenient now to have **a** as the common hypothesis in all the likelihoods. We note that

$$\mathbf{b}|\mathbf{c}\cdot\mathbf{a}=[\text{by Eq. (10)}]\mathbf{b}\cdot\mathbf{c}|\mathbf{a}/\mathbf{c}|\mathbf{a}$$
$$=[\text{by Eq. (2)}]\mathbf{c}\cdot\mathbf{b}|\mathbf{a}/\mathbf{c}|\mathbf{a}.$$

Substituting this expression in the preceding equation and multiplying both sides by $\mathrm{S}(\mathbf{b}|\mathbf{a})$, we obtain

$$\mathrm{S}(\mathbf{c}\cdot\mathbf{b}|\mathbf{a}/\mathbf{c}|\mathbf{a})\mathbf{c}|\mathbf{a}=\mathrm{S}[\mathrm{S}(\mathbf{c}\vee\mathbf{b}|\mathbf{a})/\mathrm{S}(\mathbf{b}|\mathbf{a})]\mathrm{S}(\mathbf{b}|\mathbf{a}). \quad (14)$$

This equation must hold for arbitrary meanings of the propositions **a**, **b** and **c**. Let $\mathbf{b}=\mathbf{c}\cdot\mathbf{d}$. Then

$$\mathbf{c}\vee\mathbf{b}=\mathbf{c}\vee(\mathbf{c}\cdot\mathbf{d})=[\text{by Eq. (6')}]\mathbf{c},$$

and

$$\mathbf{c}\cdot\mathbf{b}=\mathbf{c}\cdot(\mathbf{c}\cdot\mathbf{d})$$
$$=[\text{by Eq. (4)}](\mathbf{c}\cdot\mathbf{c})\cdot\mathbf{d}=[\text{by Eq. (3)}]\mathbf{c}\cdot\mathbf{d}.$$

Making these substitutions in Eq. (14), we obtain

$$\mathrm{S}(\mathbf{c}\cdot\mathbf{d}|\mathbf{a}/\mathbf{c}|\mathbf{a})\mathbf{c}|\mathbf{a}=\mathrm{S}[\mathrm{S}(\mathbf{c}|\mathbf{a})/\mathrm{S}(\mathbf{c}\cdot\mathbf{d}|\mathbf{a})]\mathrm{S}(\mathbf{c}\cdot\mathbf{d}|\mathbf{a}).$$

This may be written in a highly symmetric form if we let $\mathbf{c}|\mathbf{a}=x$ and $\mathrm{S}(\mathbf{c}\cdot\mathbf{d}|\mathbf{a})=y$, and make use of the fact that $\mathbf{c}\cdot\mathbf{d}|\mathbf{a}=[\text{by Eq. (12)}]\ \mathrm{S}[\mathrm{S}(\mathbf{c}\cdot\mathbf{d}|\mathbf{a})]=\mathrm{S}(y)$. In these terms we have

$$x\mathrm{S}[\mathrm{S}(y)/x]=y\mathrm{S}[\mathrm{S}(x)/y]. \quad (15)$$

This equation must be satisfied by the function S for all of the values of x and y obtainable by arbitrarily varying the propositions **c**, **d** and **a**. If the function S is twice differentiable, the solution of Eq. (15) together with Eq. (12) is, as shown in the appendix,

$$\mathrm{S}(p)=(1-p^m)^{1/m},$$

where m is an arbitrary constant. Hence by Eq. (11),

$$(\mathbf{b}|\mathbf{a})^m+(\sim\mathbf{b}|\mathbf{a})^m=1.$$

Now, whatever the value of m, if $\mathbf{b}|\mathbf{a}$ measures the credibility of **b** on the hypothesis **a**, then so also will $(\mathbf{b}|\mathbf{a})^m$. It has already been pointed out that $(\mathbf{b}|\mathbf{a})^m$ may replace $\mathbf{b}|\mathbf{a}$ in Eq. (10). Therefore we may take $(\mathbf{b}|\mathbf{a})^m$ as the symbol of likelihood without being under any necessity of assigning a value to m. This is the same as to say that the choice of a value for m is purely conventional. For simplicity of notation we let $m=1$ and write

$$\mathbf{b}|\mathbf{a}+\sim\mathbf{b}|\mathbf{a}=1. \quad (16)$$

This has the same form as the ordinary rule relating the probability of $\sim\mathbf{b}$ to that of **b**, or, as it is usually said, the rule for the probability that an event will not occur, given the probability that it will occur.

If in Eq. (16) we let $\mathbf{b}=\mathbf{a}$, then

$$\mathbf{a}|\mathbf{a}+\sim\mathbf{a}|\mathbf{a}=1.$$

The two likelihoods are now those of certainty and impossibility. Since certainty has been given the likelihood 1, it now follows that impossibility has the likelihood zero.

Two other useful theorems are easily obtained. By Eq. (10),

$$\mathbf{c}\cdot\mathbf{b}|\mathbf{a}+\sim\mathbf{c}\cdot\mathbf{b}|\mathbf{a}=(\mathbf{c}|\mathbf{b}\cdot\mathbf{a}+\sim\mathbf{c}|\mathbf{b}\cdot\mathbf{a})\mathbf{b}|\mathbf{a}.$$

By Eq. (16),

$$\mathbf{c}|\mathbf{b}\cdot\mathbf{a}+\sim\mathbf{c}|\mathbf{b}\cdot\mathbf{a}=1.$$

Therefore,

$$\mathbf{c}\cdot\mathbf{b}|\mathbf{a}+\sim\mathbf{c}\cdot\mathbf{b}|\mathbf{a}=\mathbf{b}|\mathbf{a}, \quad (17)$$

This is one of the theorems. The other is obtained as follows.

$$\mathbf{c}\vee\mathbf{b}|\mathbf{a}=[\text{by Eq. (16)}]1-\sim(\mathbf{c}\vee\mathbf{b})|\mathbf{a}$$
$$=[\text{by Eq. (5')}]1-\sim\mathbf{c}\cdot\sim\mathbf{b}|\mathbf{a}$$
$$=[\text{by Eq. (17)}]1-\sim\mathbf{b}|\mathbf{a}+\mathbf{c}\cdot\sim\mathbf{b}|\mathbf{a}.$$

Now, by Eq. (16), $1-\sim\mathbf{b}|\mathbf{a}=\mathbf{b}|\mathbf{a}$. Also,

$$\mathbf{c}\cdot\sim\mathbf{b}|\mathbf{a}=[\text{by Eq. (2)}]\sim\mathbf{b}\cdot\mathbf{c}|\mathbf{a}$$
$$=[\text{by Eq. (17)}]\mathbf{c}|\mathbf{a}-\mathbf{b}\cdot\mathbf{c}|\mathbf{a}$$
$$=[\text{by Eq. (2)}]\mathbf{c}|\mathbf{a}-\mathbf{c}\cdot\mathbf{b}|\mathbf{a}.$$

Therefore,

$$\mathbf{c}\vee\mathbf{b}|\mathbf{a}=\mathbf{c}|\mathbf{a}+\mathbf{b}|\mathbf{a}-\mathbf{c}\cdot\mathbf{b}|\mathbf{a}, \quad (18)$$

which has the same form as the ordinary rule for the probability that at least one of two events will occur. When, as is often done, the rule is stated for mutually exclusive events, the last term in the right-hand member does not appear. This conceals the rather interesting symmetry of the equation between the propositions **c ∨ b** and **c · b**.

The indefiniteness of the concept of likelihood, defined only as a measure of reasonable credibility, has been removed by the conventions which have been adopted. The symbol **b | a** stands now for a particular measure of credibility. Since this measure has been shown to be subject to the ordinary rules of probability, it is appropriate to call it the probability of the proposition **b** on the hypothesis **a**, discarding the term likelihood, which was less definitely defined.

The rules obtained, being only relations between probabilities, do not of themselves assign numerical values to all the probabilities arising in specific problems. The only numerical values thus far obtained are those corresponding to certainty and impossibility, and these were assigned by convention rather than required by the rules of symbolic logic. It is hardly to be supposed that every reasonable expectation should have a precise numerical value. In a number of cases, however, the familiar rule of insufficient reason may be employed. If there are n propositions of which, with respect to a given hypothesis, one and no more than one can be true, and if the hypothesis gives no reason for considering any one of them more likely than another, then, by the rules obtained, each of them has the probability $1/n$.

Probability and Frequency

The whole discussion thus far has consisted of two parts. The first part was intended to show that the rules of probable inference are credited by common sense with a wider validity than can be established by deducing them from the frequency definition of probability. In the second part they were derived without reference to this definition, from rather elementary postulates. It remains now to see what is the connection between probability, as here understood, and the frequency of an event.

Let us suppose that two capsules contain equal masses of radon, but that the contents are of different ages, one having been produced by the very recent decay of radium, and the other having been drawn from a vessel in which radon has been accumulating for a long time over radium in solution. Suppose there are two identical ion counters, each receiving radiation from one capsule, and each placed with respect to its capsule in the same relative position as the other. One of the two capsules will be the first to cause 1000 discharges in its ion counter. More than one hypothesis will ascribe to each capsule the same probability of being first. A physicist will estimate equal probabilities on the ground of the many observations which have been made on rates of radioactive transformation, with some additional evidence from quantum mechanics that the stability of such an aggregate of elementary particles as an atomic nucleus is independent of its age. Another person, quite unfamiliar with all this, will estimate equal probabilities merely on the ground that he does not know which capsule contains the older radon and has therefore no possible reason to suppose that one sample rather than the other will be the first to cause 1000 discharges.

These are the extreme cases, the first estimate being highly significant and the second quite trivial, but each is right on the hypothesis given. Kemble calls the estimate of the physicist one of *objective* probability and that of the other person one of *subjective*, or *primary*, probability. The latter term seems preferable to me, as it does to him. It is true that the estimate of the nonphysicist is subjective in the sense that it is relative to his limited information, but it is objective in the sense that another person with the same information would reasonably make the same estimate. It seems questionable whether there is a real difference in the *kind* of judgment made by the nonphysicist and the physicist. The nonphysicist bases his estimate on the fact that the capsules are indistinguishable. The physicist bases *his* estimate on the accumulated evidence that the atoms of radon themselves are indistinguishable. The difference seems to be not so much a difference in the nature of the evidence as in its amount and relevance or, to use Keynes' suggestive term, its weight.

Now let the experiment with the radon capsules be tried a number of times, with the old and new samples identified in advance of each trial. Even a long run of instances in which the older sample is first will not change the probabilities as estimated by the physicist. The evidence on which he made his first estimate had so much weight that no additional number of instances, not enormously large, could require a new estimate. The probabilities have for practical purposes become stable. Strictly speaking, since probability is relative to an experience that is never complete, it is always subject to change by new experience. A stable probability is a limit that is not strictly attainable, but that can in certain cases be approximated as nearly as necessary for practical use. It is to be expected that a stable probability will give a better basis for prediction than will an unstable one.

Let **a** be a hypothesis of which a number of instances may be examined. Let $\mathbf{b}_r$ mean that a certain proposition **b** is valid in the rth instance of **a**. Unless the hypothesis **a** itself assigns a stable probability to **b**, then $\mathbf{b}_s|\mathbf{a}$, $\mathbf{b}_s|\mathbf{a}\cdot\mathbf{b}_r$ and $\mathbf{b}_s|\mathbf{a}\cdot\sim\mathbf{b}_r$ will generally all be different; the knowledge that **b** is valid or that it is not valid in one instance will affect the reasonable expectation of its validity in another instance. But now let there be included in the hypothesis a proposition **p**, which asserts that the probability is stable and equal to p, some number between 0 and 1. This means that the probability of $\mathbf{b}_s$ is the same whether $\mathbf{b}_r$ or $\sim\mathbf{b}_r$ or neither is included in the hypothesis. Thus

$$\mathbf{b}_s|\mathbf{a}\cdot\mathbf{p}\cdot\mathbf{b}_r=\mathbf{b}_s|\mathbf{a}\cdot\mathbf{p}\cdot\sim\mathbf{b}_r=\mathbf{b}_s|\mathbf{a}\cdot\mathbf{p}=p.$$

Then by Eqs. (10) and (16) we obtain

$$\begin{aligned}\mathbf{b}_s\cdot\mathbf{b}_r|\mathbf{a}\cdot\mathbf{p}&=p^2,\\ \mathbf{b}_s\cdot\sim\mathbf{b}_r|\mathbf{a}\cdot\mathbf{p}&=p(1-p),\\ \sim\mathbf{b}_s\cdot\sim\mathbf{b}_r|\mathbf{a}\cdot\mathbf{p}&=(1-p)^2.\end{aligned}$$

Let $\mathbf{n}_N$ mean that the number of instances of **b** in N instances of **a** is exactly n. Then by Eqs. (10), (16) and (18) it is possible to derive the well-known result of Bernoulli, that

$$\mathbf{n}_N|\mathbf{a}\cdot\mathbf{p}=p^n(1-p)^{N-n}N!/n!(N-n)!.$$

This is a maximum when $p=n/N$, and the maximum becomes sharper as N is increased. Thus, when there is a stable probability, the frequency may confidently be expected to approach it as a limit.

There will sometimes be questions in which the existence of a stable probability is known but its value is undetermined. As a rather artificial but simple example, let it be supposed that there are two dice, both dynamically symmetric, but one of them defectively marked, having two faces instead of one stamped with four dots. Then for either of these dice there is a stable probability of throwing a four, equal to $\frac{1}{6}$ if it is the true die and to $\frac{1}{3}$ if it is the defective one. Suppose one die of the pair is picked up at random and, without being examined, is tossed N times. If a four turns up on n of these throws, what is the probability of a four on the next throw?

The problem may be generalized as follows. Let it be supposed that in the ensemble of instances of a proposition **a**, another proposition **b** is known to have a stable probability, but the value of this stable probability is unknown. As before, **p** will denote the proposition that the probability is stable and equal to a number p, but in the present case **p** is not a part of the hypothesis. Instead, the hypothesis contains a weaker proposition which only assigns a probability to the proposition **p** corresponding to every value of p.

We may let the single symbol **a** represent the entire initial hypothesis, including this proposition. Thus, in the example of the dice, **a** will describe the two dice and will also assert that one is chosen at random and tossed without being identified. There are then only two possible stable probabilities, $\frac{1}{6}$ and $\frac{1}{3}$, and they are equally probable at the beginning. Hence in this example $\mathbf{p}|\mathbf{a}$ has the value $\frac{1}{2}$ if p is either $\frac{1}{6}$ or $\frac{1}{3}$ and the value zero if p is any other number.

Returning to the general problem, we suppose that N instances of **a** are observed, and **b** is found valid in n of them and invalid in the rest. What is now the probability of **b** in the $N+1$th instance of **a**? As before, let $\mathbf{n}_N$ denote the proposition that n is the number of instances of **b** in N instances of **a**. The problem is to find $\mathbf{b}_{N+1}|\mathbf{a}\cdot\mathbf{n}_N$, given n and N, and also $\mathbf{p}|\mathbf{a}$ for every value of p.

The theorems available are enough to give the result

$$\mathbf{b}_{N+1}|\mathbf{a}\cdot\mathbf{n}_N=\frac{\sum p^{n+1}(1-p)^{N-n}\mathbf{p}|\mathbf{a}}{\sum p^n(1-p)^{N-n}\mathbf{p}|\mathbf{a}},$$

where the summations are over all values of p.

If, on the hypothesis **a**, the stable probability has a continuous range of possible values from 0 to 1, and if $\mathrm{f}(p)\mathrm{d}p$ denotes the probability of a value between p and $p+\mathrm{d}p$, the summations are replaced by integrals, and we have

$$\mathbf{b}_{N+1}|\mathbf{a}\cdot\mathbf{n}_N=\frac{\int_0^1 p^{n+1}(1-p)^{N-n}\mathrm{f}(p)\mathrm{d}p}{\int_0^1 p^n(1-p)^{N-n}\mathrm{f}(p)\mathrm{d}p}.$$

It was assumed by Laplace that an unknown probability is equally likely to have any value from 0 to 1. On this assumption, $\mathrm{f}(p)$ is constant in the last equation. The integrals in this case are known; and the result, sometimes called the rule of succession, is simply

$$\mathbf{b}_{N+1}|\mathbf{a}\cdot\mathbf{n}_N=(n+1)/(N+2),$$

or approximately, for large values of n and N,

$$\mathbf{b}_{N+1}|\mathbf{a}\cdot n_N=n/N.$$

Several authors have pointed out that an unknown stable probability is not necessarily one for which all values from 0 to 1 are equally likely, and the rule of succession has been shown to lead to some absurd results. Nevertheless, we should expect that for large numbers it will generally be right. For we know, by the theorem of Bernoulli given earlier, that, when there is a stable probability, the ratio n/N is very likely to be nearly equal to it when N is large. Also we have understood a stable probability to be the limit that the probability approaches as the weight of the evidence is increased, and usually the surest way to increase the weight of evidence is to increase the number of observed instances. If the ratio n/N and the probability approach a common limit, then certainly they must approach each other.

If the absurdities to which Laplace's rule has led are examined, they are found to fall into three classes: those in which N is not a very large number, those in which $n/N=1$, and those in which $n/N=0$. (The last two are really one class, since to say that **b** is valid in N out of N instances of **a** is the same as to say that $\sim$**b** is valid in none of N instances.) If these conditions are excluded, Laplace's rule may be derived from a much less drastic assumption than the assumption that all values of the stable probability are equally likely.

If from the general equation for $\mathbf{b}_{N+1}|\mathbf{a}\cdot\mathbf{n}_N$ we eliminate n by letting $n/N=\nu$, we obtain

$$\mathbf{b}_{N+1}|\mathbf{a}\cdot\mathbf{n}_N=\frac{\int_0^1 p[p^\nu(1-p)^{1-\nu}]^N\mathrm{f}(p)\mathrm{d}p}{\int_0^1 [p^\nu(1-p)^{1-\nu}]^N\mathrm{f}(p)\mathrm{d}p}.$$

If $0<\nu<1$, then $p^\nu(1-p)^{1-\nu}$ has a maximum value when $p=\nu$. The Nth power of this expression, when N is large enough, will have so pronounced a maximum that its values when p is more than slightly different from ν will be relatively negligible. Hence, unless $\mathrm{f}(p)$ is extremely small when $p=\nu$, its only values of importance in the integrals will be those for which p and ν are very nearly equal. Therefore, unless $\mathrm{f}(p)$ is rapidly varying around this point, it may be replaced in the integrals by the constant $\mathrm{f}(\nu)$.

Thus we arrive again at Laplace's rule. Its generality is much less than Laplace supposed. But it serves to show how a probability approaches stability as the number of instances is increased, and this is all we should expect of it.[12]

* * *

Professor K. O. Friedrichs, of New York University, read a preliminary draft of this paper. I wish to thank him for this kindness and for his help in correcting some mathematical inaccuracies. He is not responsible, of course, for any errors that remain or for the opinion expressed as to the nature of probability.

Appendix: The Solution of the Functional Equations

The first equation to be solved is

$$\mathrm{F}[\mathrm{F}(x,y),z]=\mathrm{F}[x,\mathrm{F}(y,z)]. \tag{8}$$

Let $\mathrm{F}(x,y)=u$, and let $\mathrm{F}(y,z)=v$. Then Eq. (8) becomes $\mathrm{F}(u,z)=\mathrm{F}(x,v)$. Differentiating this with respect to x, y and z in turn, and writing $\mathrm{F}_1(p,q)$ for $\partial\mathrm{F}(p,q)/\partial p$ and $\mathrm{F}_2(p,q)$ for $\partial\mathrm{F}(p,q)/\partial q$, we obtain

$$\mathrm{F}_1(u,z)\partial u/\partial x=\mathrm{F}_1(x,v), \tag{19}$$
$$\mathrm{F}_1(u,z)\partial u/\partial y=\mathrm{F}_2(x,v)\partial v/\partial y, \tag{20}$$
$$\mathrm{F}_2(u,z)=\mathrm{F}_2(x,v)\partial v/\partial z. \tag{21}$$

Differentiating Eq. (20) with respect to x, y and z in turn,

[12] The problem of inverse probability when $n/N=1$ (or 0), which is important in the application of probability to inductive reasoning, is discussed at length by Jeffreys in reference 8.

writing $F_{11}(p, q)$ for $\partial F_1(p, q)/\partial p$, and similarly representing the other second derivatives, we obtain

$$F_{11}(u, z)(\partial u/\partial x)(\partial u/\partial y)+F_1(u, z)\partial^2 u/\partial x\partial y = F_{12}(x, v)\partial v/\partial y, \quad (22)$$

$$F_{11}(u, z)(\partial u/\partial y)^2+F_1(u, z)\partial^2 u/\partial y^2 = F_{22}(x, v)(\partial v/\partial y)^2+F_2(x, v)\partial^2 v/\partial y^2, \quad (23)$$

$$F_{12}(u, z)\partial u/\partial y = F_{22}(x, v)(\partial v/\partial y)(\partial v/\partial z) + F_2(x, v)\partial^2 v/\partial y\partial z. \quad (24)$$

Differentiating Eq. (19) with respect to z, or Eq. (21) with respect to x, we obtain

$$F_{12}(u, z)\partial u/\partial x = F_{12}(x, v)\partial v/\partial z. \quad (25)$$

Among Eqs. (20), (22), $\cdots$, (25) we can now eliminate the functions of u and v other than their derivatives. Thus, eliminating $F_{12}(u, z)$ and $F_{12}(x, v)$ among Eqs. (22), (24) and (25), we find

$$[F_{11}(u, z)(\partial u/\partial y)^2-F_{22}(x, v)(\partial v/\partial y)^2](\partial u/\partial x)(\partial v/\partial z) = F_2(x, v)(\partial^2 v/\partial y\partial z)(\partial v/\partial y)(\partial u/\partial x) - F_1(u, z)(\partial^2 u/\partial x\partial y)(\partial u/\partial y)(\partial v/\partial z)$$

Combining this with Eq. (23), we can eliminate $F_{11}(u, z)$ and $F_{22}(x, v)$ together, obtaining

$$F_1(u, z)(\partial v/\partial z)[(\partial^2 u/\partial y^2)(\partial u/\partial x)-(\partial^2 u/\partial x\partial y)(\partial u/\partial y)] = F_2(x, v)(\partial u/\partial x)[(\partial^2 v/\partial y^2)(\partial v/\partial z)-(\partial^2 v/\partial y\partial z)(\partial v/\partial y)].$$

Combining this with Eq. (20), we eliminate $F_1(u, z)$ and $F_2(x, v)$ together and obtain

$$\frac{\partial^2 u/\partial x\partial y}{\partial u/\partial x}-\frac{\partial^2 u/\partial y^2}{\partial u/\partial y}=\frac{\partial^2 v/\partial y\partial z}{\partial v/\partial z}-\frac{\partial^2 v/\partial y^2}{\partial v/\partial y}.$$

This may be written in the form,

$$\frac{\partial}{\partial y}\ln\left(\frac{\partial u/\partial x}{\partial u/\partial y}\right)=-\frac{\partial}{\partial y}\ln\left(\frac{\partial v/\partial y}{\partial v/\partial z}\right).$$

Now $u=F(x, y)$ and $v=F(y, z)$, so that

$$\frac{\partial u/\partial x}{\partial u/\partial y}=\frac{F_1(x, y)}{F_2(x, y)}, \quad \text{and} \quad \frac{\partial v/\partial y}{\partial v/\partial z}=\frac{F_1(y, z)}{F_2(y, z)}.$$

We have then,

$$\frac{\partial}{\partial y}\ln\left[\frac{F_1(x, y)}{F_2(x, y)}\right]=-\frac{\partial}{\partial y}\ln\left[\frac{F_1(y, z)}{F_2(y, z)}\right].$$

Since x appears only in the left-hand member and z only in the right-hand member of this equation, it follows that each member is a function only of the remaining variable y. It will be convenient to denote the integral of this function by $\ln \Phi(y)$, so that we have

$$\frac{\partial}{\partial y}\ln\left[\frac{F_1(x, y)}{F_2(x, y)}\right]=\frac{d}{dy}\ln \Phi(y), \quad (26)$$

and

$$\frac{\partial}{\partial y}\ln\left[\frac{F_1(y, z)}{F_2(y, z)}\right]=-\frac{d}{dy}\ln \Phi(y). \quad (27)$$

Permuting x, y and z in Eq. (27), we obtain

$$\frac{\partial}{\partial x}\ln\left[\frac{F_1(x, y)}{F_2(x, y)}\right]=-\frac{d}{dx}\ln \Phi(x). \quad (28)$$

Multiplying Eq. (28) by dx and Eq. (26) by dy and adding, we obtain

$$\frac{\partial}{\partial x}\ln\left[\frac{F_1(x, y)}{F_2(x, y)}\right]dx+\frac{\partial}{\partial y}\ln\left[\frac{F_1(x, y)}{F_2(x, y)}\right]dy = -d\ln\Phi(x)+d\ln\Phi(y).$$

The left-hand member being now a complete differential, we may integrate and so find

$$F_1(x, y)/F_2(x, y)=h\Phi(y)/\Phi(x), \quad (29)$$

where h is a constant of integration.

To make use of this result, we divide Eq. (20) by Eq. (21), obtaining

$$\frac{F_1(u, z)}{F_2(u, z)}\frac{\partial u}{\partial y}=\frac{\partial v/\partial y}{\partial v/\partial z}.$$

The right-hand member is simply $F_1(y, z)/F_2(y, z)$, and, with the aid of Eq. (29), the equation may be written as

$$\frac{\Phi(z)}{\Phi(u)}\frac{\partial u}{\partial y}=\frac{\Phi(z)}{\Phi(y)}.$$

Replacing in this equation u by its value $F(x, y)$ we have

$$\partial F(x, y)/\partial y=\Phi[F(x, y)]/\Phi(y). \quad (30)$$

Similarly, from Eqs. (19) and (20) we obtain

$$\partial F(y, z)/\partial y=\Phi[F(y, z)]/\Phi(y),$$

which becomes, when x and y are written for y and z,

$$\partial F(x, y)/\partial x=\Phi[F(x, y)]/\Phi(x). \quad (31)$$

Combining Eqs. (30) and (31) to obtain the differential dF (the variables being understood as x and y) we find

$$dF/\Phi(F)=dx/\Phi(x)+dy/\Phi(y).$$

If we denote $\int[dp/\Phi(p)]$ by $\ln f(p)$, we obtain, by integrating and taking the exponentials of both members of this equation,

$$Cf(F)=f(x)f(y),$$

where C is a constant of integration. This then is the solution of Eq. (8).

The solution of the equation

$$xS[S(y)/x]=yS[S(x)/y] \quad (15)$$

is obtained in a similar manner, but more quickly. This equation and the three derived from it by differentiation with respect to x, to y, and to x and y may be written as follows, when $S(y)/x$ is denoted by u and $S(x)/y$ by v:

$$xS(u)=yS(v), \quad (32)$$

$$uS'(u)-S(u)=-S'(v)S'(x), \quad (33)$$

$$S'(u)S'(y)=-vS'(v)+S(v), \quad (34)$$

$$uS''(u)S'(y)/x=vS''(v)S'(x)/y. \quad (35)$$

Multiplying Eq. (32) by Eq. (35), we eliminate x and y simultaneously, obtaining

$$uS''(u)S(u)S'(y)=vS''(v)S(v)S'(x).$$

With this equation, together with Eqs. (33) and (34), it is possible to eliminate $S'(x)$ and $S'(y)$. The result is the equation

$$\frac{uS''(u)S(u)}{[uS'(u)-S(u)]S'(u)}=\frac{vS''(v)S(v)}{[vS'(v)-S(v)]S'(v)}.$$

Since each member of the foregoing equation is the same function of a different variable, this function must be equal to a constant. Calling this constant k, we have

$$uS''(u)S(u)=k[uS'(u)-S(u)]S'(u).$$

This may be put in the form

$$dS'/S'=k(dS/S-du/u),$$

whence, by integration,

$$S'=A(S/u)^k,$$

where A is a constant.

The variables being separable, another integration gives

$$S^m=Au^m+B,$$

where m has been written for $1-k$, and B is a constant of integration. It is now found by substitution that Eq. (15) can be satisfied for arbitrary values of x and y only if $B=A^2$. Finally, if the solution of Eq. (15) is also to satisfy the equation $S[S(x)]=x$, it is necessary that $A=-1$. Thus we obtain $[S(u)]^m+u^m=1$.

Application of Group Theory to the Calculation of Vibrational Frequencies of Polyatomic Molecules[1]

Arnold G. Meister and Forrest F. Cleveland
Illinois Institute of Technology, Chicago, Illinois

Wilson[2] has devised a method for obtaining the vibrational frequencies of polyatomic molecules in which group theory is used to simplify the calculations. The method is especially good for molecules having considerable symmetry and several equivalent atoms, that is, atoms with identical nuclei that transform into one another for all operations of the point group of the molecule. A further advantage of the method is that it requires no coordinate system, but only bond distances, interbond angles and unit vectors directed along the bonds.

Since one who is beginning calculations of vibrational frequencies may find the symbolism of Wilson's papers difficult, and since other papers involving the method omit many of the details, it seems worth while to give an elementary treatment of a few typical molecules for those desiring to start work in this field. The H_2O molecule is considered first because it has only a small number of atoms, has no degenerate frequencies, and permits the reader to concentrate on the method without being confused by the complexity of the molecule.[3] Then the CH_3Cl, CH_4 and CD_4 molecules are treated to show how the method is applied when doubly or triply degenerate frequencies are present.[4]

THE H_2O MOLECULE

Symmetry Coordinates

The methods given in a previous paper[5] are used to determine the point group of the molecule as well as the number of fundamental vibrations of each type. It is found that the H_2O molecule belongs to the point group $\mathbf{C}_{2v}$ and that there are two vibrations of type A_1 and one of type B_2. Since a nonlinear molecule containing N atoms has $3N-6$ vibrational degrees of freedom, $3N-6$ coordinates are necessary to describe the vibrations of the molecule. To attain the simplification made possible by the use of group theory, it is

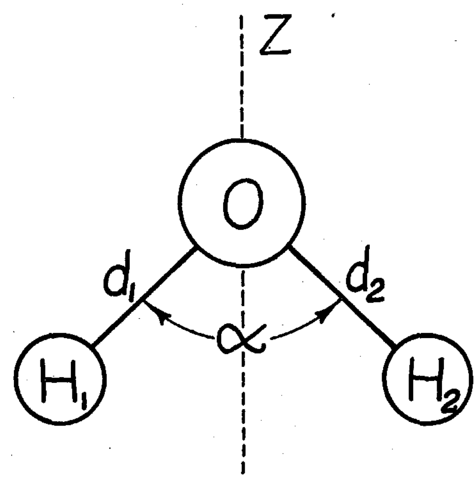

Fig. 1. Bond distances d_1 and d_2, interbond angle α, and principal symmetry axis Z for the H_2O molecule.

[1] Communication No. 43 from the *Spectroscopy Laboratory*.

[2] E. B. Wilson, Jr., *J. Chem. Physics* **7**, 1047 (1939); 9, 76 (1941).

[3] A more complicated molecule, CH_2Cl_2, involving only nondegenerate frequencies has been discussed by G. Glockler, *Rev. Mod. Physics* **15**, 125 (1943).

[4] A treatment in outline form of the CH_3Cl molecule is given at the end of Wilson's second paper, reference 2.

[5] A. G. Meister, F. F. Cleveland and M. J. Murray, *Am. J. Physics* **11**, 239 (1943).