

Mathematical Probability in the Natural Sciences

SIR RONALD A. FISHER

Cambridge University

Cambridge, England

Presidential address given at Symposium III of the XVIII^e Congress International des Sciences Pharmaceutiques, organized jointly by the Scientific Section of the International Pharmaceutical Federation and the Section Adolphe Quetelet—Brussels, 8-15 September, 1958. This address will also appear in *Metrika*, 1959.

1. FREQUENCY DISTRIBUTIONS AND RANDOM VARIABLES

The very existence of quantitative or biometrical work in the Pharmaceutical Sciences is in itself evidence of the extent to which during the last sixty or seventy years one of the greatest obstacles to exact thought has been effectually overcome. The obstacle to which I refer is the existence of variability in the natural world. Only one hundred years ago the inhibiting effect of this obstacle can be seen in many writers; for us to-day the obstacle does not exist. The familiar concept of a frequency distribution as used by Adolphe Quetelet comes to our minds, and we recognize that by means of this device, easily represented on paper, variability may be accurately specified, and its consequences calculated.

A close mathematical analogy, though not a logical identity, exists between the specification of a frequency distribution, and what in recent years has come to be called a Random Variable. If x stand for such a variable and if, for all values of P from 0 to 1 we can make probability statements of the form

$$Pr(x < x_p) = P,$$

where x_p is a known function of P , then x is said to be a random variable, and knowing x_p , we may say it is a random variable of known distribution, for the probability of x lying between any two values x_b and x_a may be equated conceptually to the frequency in a variable population of objects having some characteristic measurement which lies between these limits. The probability statements would then be equivalent to those appropriate to the knowledge that the subject of these statements had been chosen at random from a real population having this specification.

However, the probability statements do not imply the existence of any such population in the real world. All that they assert is that the exact nature and degree of our uncertainty is just *as if* we knew it to have been one chosen

at random from such a population. The subject of a probability statement if we know what we are talking about, is singular and unique; we have some degree of uncertainty about its value, and it so happens that we can specify the exact nature and extent of our uncertainty by means of the concept of Mathematical Probability as developed by the great mathematicians of the 17th century Fermat, Pascal, Leibnitz, Bernoulli and their immediate followers.

The emergence of the notion of mathematical probability in these great minds was undoubtedly a major step in the intellectual development of the human race. It was unknown to the Greeks, and to the Islamic mathematicians of the Middle Ages. It owes its emergence, I suppose, to the high prestige of the recreation of gambling among the nobility of France and England, and to the existence of a technology advanced enough to supply apparatus of gambling, dice, cards, etc. with a precision sufficient to justify the calculations of the mathematicians. The type of uncertainty which the new concept was capable of specifying with exactitude was that which confronts a gambler in a game played fairly, and with perfect apparatus. The probability statements refer to the particular throw or to the particular result of shuffling the cards, on which the gambler lays his stake. The state of rational uncertainty in which he is placed may be equated to that of the different situation which can be imagined in which his throw is chosen at random out of an aggregate of throws, or of shufflings, which might equally well have occurred, though such aggregates exist only in the imagination.

2. UNCERTAIN KNOWLEDGE OF THE REAL WORLD

This distinction makes clear the propriety of the application by the great Carl Gauss of the Theory of Errors to make probability statements about features of the real world, which the observations can ascertain only with some uncertainty, as in Astronomy, about the distance of the Sun. Gauss with a certain amount of approximation, which later work has removed, showed how to calculate a function x_p of any value of P from 0 to 1, such that if x stand for the unknown distance, the probability statement

$$Pr(x < x_p) = P,$$

could be asserted for all values of P . He therefore expressed the unknown distance x as what we should now call a Random Variable, about which probability statements at all levels could be asserted.

It is an indication of the state of confusion about even the most primary concepts of mathematical statistics, that more than one modern writer, starting perhaps with Neyman about 1935, should have taken exception to this form of statement on the ground that, for example, there is only one distance of the Sun, that if it is less than x_p , the probability should be unity, and if it is greater than x_p it must be zero. The error, evidently, is to suppose that additional data, such as that which provides the exact value of x , can be introduced without altering the probability statements which the data actually provide. However,

the *conclusions* of a logical argument must depend on its *premises*, and it is no criticism of the *argument* that these are different from what they would have been had the premises used in the induction been different from what they are.

3. THE MEANING OF MATHEMATICAL PROBABILITY

The error has, however, had some troublesome consequences. To obviate these and to clarify the matter further, I will set out what I understand to be the meaning of the term Mathematical Probability, as used both by the Old Masters and by modern scientific practitioners, for it seems to be only in Mathematical Departments insulated from practical research in the Natural Sciences, that confusion and misapprehensions abound. The requirements of a correct statement of Mathematical Probability are, I believe, only three:

(a) A conceptual Reference Set, which may be *conceived* as a population of possibilities of which an exactly known fraction possess some chosen characteristic. To the extent that this is a meaningful fraction of the whole, the set must be measurable, but it need not be measurable in all respects. The set of possible throws with a die may be conceived to have exactly $1/6$ Aces, and if this proportion is the same for throws made on any day of the week, it is unnecessary that the specification of the set shall tell us what proportion come on Tuesdays. The distribution among days of the week is irrelevant and may remain unknown.

(b) It must be possible to assert that the subject of the probability statement belongs to this Set. To the mathematician this may seem trivial, though obviously a logical necessity. Tasks of identification, however, belong to the scientist, and may require his full attention. We must rely on a Chemist to tell us, if an atomic weight has been determined, whether it is that of Potassium or of Rubidium. This second requirement puts our probability statement into the real world.

(c) No subset can be *recognized* having a different probability. Such subsets must always exist; it is required that no one of them shall be recognizable. This is a postulate of ignorance, and therefore unfamiliar to deductive reasoning, though characteristic of inductive logic. Let me illustrate it briefly using the probability of throwing an Ace.

(i) The subset of throws made on a Tuesday is easily recognizable, it has, however, the same probability as the whole set, and is therefore irrelevant.

(ii) The subset of throws of odd numbers (1, 3, 5) is obviously relevant, since all Aces are odd numbers; this subset is, however, unrecognizable before the die is cast.

(iii) A parapsychologist claiming the gift of precognition might recognize a subset in which the Ace is *foreseen*. If the gift is veridical the proportion of Aces in this subset must exceed one in six. In such a case the subset will supersede the original set as the basis of the probability statement. If, however, experience shows that the proportion in the subset is just one in six as it would

be without precognition, the subset is recognized to be irrelevant, and the original set is reinstated.

Now I submit that these conditions are Necessary and Sufficient; that there is nothing lacking in the three conditions I have set out for a valid probability statement in the real world; and that there is nothing superfluous. In other words, all genuine probability statements fulfil these conditions and any statement which fulfils these conditions is a genuine statement of mathematical probability.

On this basis we may clear the air somewhat. The concept of Mathematical Probability is shown to be unitary. There are not two or more logically distinct kinds of probability statements, for if there were they would be susceptible of distinct logical specification. Some writers have, however, misunderstood the phrase "fiducial probability", forgetting that it was introduced only to mean probability derived by the particular kind of argument to which the term fiducial has been applied, and imagining that it signified a special kind of probability, for which therefore a special definition would be required. No such definition, however, has been proposed. In the minds of these writers we are left with a distinction without a difference.

4. THE FIDUCIAL ARGUMENT

I propose to illustrate the fiducial argument in a simple case, which utilizes a small but important improvement added in this century to the method of Gauss. In 1908 more than 50 years after Gauss's death, a research chemist, W. S. Gosset, better known under his pseudonym of "Student", decided that the observations of his laboratory at Guinness's brewery required for their interpretation exact knowledge about the true mean of a normal distribution from which only a sample, and perhaps a small sample of observations was available.

His mathematical approach to this problem was to calculate the distribution in random samples of the error of the mean divided by the best possible estimate of that error. Symbolically, if μ stand for the true mean, and \bar{x} for the mean of the sample calculated from

$$\bar{x} = S(x)/N$$

where S stands for summation over the sample observed, and if we define s by

$$s^2 = S(x - \bar{x})^2/N(N - 1)$$

so that s is the estimated standard error of the mean, then Student determined the exact distribution of a quantity t defined so that

$$(\mu - \bar{x})/s = t$$

The distribution is independent both of the unknown parameter μ , and of the unknown variance of the population sampled, so that value t , can be tabulated such that for any random sample from any normal distribution

$$Pr(t < t_p) = P,$$

for all values of P from 0 to 1.

If, therefore, a chemist has N observations, being independent determinations of equal precision of some unknown of importance to his work, he can calculate \bar{x} and s from his sample, and substitute their numerical values in the expression for t so obtaining the probability statement

$$Pr\{\mu < (\bar{x} + st_p)\} = P.$$

The subject of these probability statements is the unknown μ , a property of the real world to be determined by observation and experiment like an atomic weight, and the series of probability statements about this unique value is strictly of the form aimed at by Gauss.

The improvement on the method of Gauss achieved in the course of 100 years consists of two elements: (a) Gauss did not know the exact distribution of t but took instead as an approximation appropriate for large samples that it was normally distributed; (b) In the confusion of the early nineteenth century Gauss had indeed introduced into his demonstration the assumption of probability *a priori* supposedly axiomatic as taught by Laplace. He later regretted this and in a letter to Bessel recognized its arbitrariness. He was not, however, in a position to remedy this logical defect.

Now, on what conditions do a system of statements such as I have inferred by the fiducial argument represent genuine statements of probability in the real world? As regards the Reference Set we may recognize that the triad of values

$$(\mu, \bar{x}, s)$$

must exist for every sample from every normal population, that for some of these samples, but not for all, the inequality

$$\mu < \bar{x} + st_p$$

is satisfied, and that whatever may be the true values of the mean and variance of the population sampled, the proportion of random samples which satisfy the inequality is exactly P .

The second requirement, that our data really belong to this reference set, depends as I emphasized above on the competence of the Scientist and particularly on the adequacy of his experimental design.

The third requirement, which has been the most often ignored, is, however, more interesting. Can any subset be recognized within the general set? If a distribution *a priori* of μ were available, this could be used by Bayes' method to recognize a particular subset characterized by

$$\bar{x} + st_p$$

and to make probability statements appropriate to this subset, different in general from those of the fiducial argument. We must therefore specify that if

a Bayesian probability *a priori* is available we shall use the method of Bayes, and that the first condition for the applicability of the fiducial argument is that no probability *a priori* of the form needed for Bayes' theorem shall be available.

Next, it might be thought that some feature calculable from the observations could define a relevant subset. However, it has been shown that for Normal samples the two statistics \bar{x} and s are *exhaustive*, meaning that the sampling distribution of any such function, conditional upon given values of \bar{x} and s shall be completely independent of the true mean and variance of the population sampled. There is therefore no relevant and recognizable subset, and all the conditions for a genuine statement of probability are satisfied.

It is sometimes asserted that the fiducial method generally leads to the same results as the method of Confidence Intervals. It is difficult to understand how this can be so, since it has been firmly laid down that the method of confidence intervals does not lead to probability statements about the parameters of the real world, whereas the fiducial argument exists for this purpose. Moreover, the arguments of Neyman and Pearson are deliberately not restricted to cases where exhaustive estimation is possible, and this is a definite condition for an accurate fiducial argument. It is perhaps only the frequent occurrence of exhaustive estimation in elementary examples which is responsible for the frequent coincidence of Confidence with Fiducial limits. The meaning however is always different.

5. BAYES' THEOREM

Since the time of Laplace the use by Bayes of probabilities *a priori* has been a stumbling block. For Laplace believed, as Bayes did not, that such probabilities were axiomatic *a priori*. Bayes' resistance to this notion is shown by two phrases in Price's introductory letter, which explains that in Bayes' opinion the *postulate* "might not perhaps be looked on by all as reasonable", and that therefore he had chosen to demonstrate his proposition in another way "rather than to bring into his mathematical reasoning anything that might admit dispute."

We have these demurrers only at second hand through Price; but if there were any doubt we could resolve it by reference to Theorem 8 of Bayes' own text, and see there that the probability statements *a priori* are not assumed axiomatically, but are the result of an auxiliary *experiment*.

Bayes' problem was that presented by Bernoullian or binomial data in which a successes and b failures have been recorded out of n trials. His solution took the form that if prior to and independently of these observations the unknown probability p of success were known as a random variable, so that the probability of it falling in the range dp were known to be $f(p) dp$ then the probability of this event combined with the outcome of the observations could be expressed as

$$\frac{n!}{a! b!} p^a q^b f(p) dp,$$

where q stands for $1 - p$, and this expression divided by its integral from 0 to 1, as normalizing factor, would give the probability *a posteriori*, from which probability statements to the effect that p is less than any given value, or lies between given limits, would be derived by direct integration. No one in that century objected that p could have only one true value, that this must either lie between the given limits or not, and that the probability statements were therefore meaningless. Such an argument requires a sophistry peculiar to the twentieth century!

How did Bayes obtain his probability *a priori*? His apparatus was an idealised billiard table. He says "I suppose the square table $ABCD$ be so made and levelled that if either of the balls O or W be thrown upon it, there shall be the same probability that it rests upon any one equal part of the plane as another, and that it must necessarily rest somewhere upon it.

I suppose that the ball W be first thrown and through the point where it rests a line shall be drawn" parallel to the ends of the table "and that afterwards the ball O shall be thrown n times", and that its resting beyond the line should constitute a success.

The auxiliary experiment consists of the first throw, that made with the ball W . From this it is inferred that if p be the proportion of the table beyond the line drawn, then

$$Pr(p < p_1) = p_1$$

for all values of p_1 , and therefore that the probability that p should lie within the range dp must simply be equal to dp , so that $f(p)$ in the analysis above may be equated to unity.

6. A RESTATEMENT OF BAYES' THEOREM

The ingenuity of Bayes' demonstration consists in showing that the single probability statement derived from the ball W can be combined with the evidence of the ball O , which by itself provides no probability statement about p , to give *a posteriori* an exact expression for p as a random variable.

There is, however, an element of artificiality about this approach, which has perhaps been an obstacle to its understanding. For, it should be noted, after the line is drawn we might infer not only the probability *a priori* of the value p , but its actual value, by measuring the distance of the line from the ends of the table.

This artificiality may be easily removed by introducing improved apparatus. The billiard table was no doubt sufficiently familiar in the eighteenth century, but in the twentieth we can randomize more accurately with a radioactive source.

Let us imagine then a radioactive source emitting α particles at random with unknown frequency. This replaces the billiard table. The two balls are replaced by detectors of α -particles of two types. The first, which I shall call W , is designed to measure the exact time interval between two successive emissions. The first

of these starts the clock, the second stops it. The auxiliary experiment with W consists in taking a single reading of x seconds. The probability that it should exceed x seconds is represented by P , and it is easily seen that

$$e^{-\theta x} = P,$$

where θ is the unknown rate of emission (in the right direction) supplied by the source. Moreover as with Bayes' ball it is obvious that

$$Pr(P < P_1) = P_1,$$

for all values of P_1 from 0 to 1.

After the auxiliary experiment therefore, we do not know the value of P , but we do know its distribution *a priori*; what has been actually observed is the time interval, x .

The second detector O , which is to be used n times, is set for a known fixed time ξ , and will detect whether in this time interval either no α -particle arrives, which will be counted a success, or on the other hand one or more particles are detected, which will be a failure. The apparatus is not supposed to be a counter; it only distinguishes these two possibilities. The unknown probability of success is evidently

$$p = e^{-\theta \xi}.$$

Now the ratio of the times x and ξ is known, and if

$$x/\xi = \lambda,$$

it follows that

$$P = p^\lambda,$$

and

$$dP = \lambda p^{\lambda-1} dp,$$

which is now the probability *a priori* of p falling in the range dp . The case treated by Bayes is that in which λ is unity, as could easily be realized in practice by setting the value ξ equal to the value of x first observed.

I suggest, however, that the mild generalization supplied by λ is of value in reminding mathematicians, who often seem to forget it, that Bayes' probability *a priori* was obtained, not axiomatically, but by an auxiliary experiment, which with the improved apparatus I suggest, could lead to more than one result.

7. SOURCES OF PROBABILITY *a Priori*

Finally, it may perhaps be useful to set out a brief summary of the cases which arise in connection with probabilities *a priori*.

(a) The commonest case is that in which no probabilities *a priori* exist as, for example, with the chemical determination of an atomic weight. In these

cases, when the experiment can be so designed that exhaustive estimation is possible, we can infer probability statements *a posteriori* by the fiducial argument.

(b) Probabilities *a priori* may be inherent in the data. This situation is typical in Genetics, where before a mouse is born we may be able from its genealogy to state in advance the probability of each of its possible genotypes, and to combine these probabilities, in the manner of Bayes, with the evidence of the frequencies of different types of progeny obtained from test matings.

(c) It may be possible to follow Bayes' procedure and to carry out an auxiliary experiment which will supply probabilities *a priori*. The types of observation which will do this are indeed the same as will supply by themselves probabilities *a posteriori* by the fiducial argument.

(d) It has often been imagined that probabilities *a priori* can be set up axiomatically; this has led to many inconsistencies and paradoxes, and I am forced to conclude that the process is completely bogus.

This mid twentieth century is not the first period of grave confusion in the teaching of the Theory of Probability. For fifty years from the publication of Laplace's *Théorie analytique* books in France and England were full of rubbish concerning the veracity of witnesses, and the probability of correctness of the findings of tribunals. The present confusion seems largely to be a hangover from that period from which nineteenth century discussion had largely extricated mathematical thought in France and England, but perhaps less completely in some more distant countries.

There is the difference, however, that whereas in the 19th Century error could be rife in mathematical departments, without doing greater harm than to confuse their students, in our time really important matters such as the standardization of drugs, the control of epidemics, and the precision of ballistic missiles are liable in the future to be influenced by young men now leaving these departments armed with erroneous numerical tables, as well as with confused and obsolete ideas. This, in some sort, concerns us all.