267

THE UNDERWORLD OF PROBABILITY

By Sir RONALD A. FISHER
University of Cambridge

Readers of my recent book *Statistical Methods and Scientific Inference*, may have noticed that in the author's opinion, stagnation and confusion of thought over a long period have been due to the common failure to distinguish various different types of evidence, and the correspondingly various types of uncertainty appropriate to an appreciation of such evidence.

The following paper is much less than an exploration of the varieties of logical uncertainty in normal experience. It takes one short step towards a recognition of these varieties and introduces the beginnings of a classification, more detailed than that of my book. The illustrations given will, I believe, help others to engage in a deeper penetration.

1. A ranking of the levels of uncertainty

A series of meaningful mathematical statements may be made, each differing from the one before by its greater uncertainty, as follows :

Certainty. The mean distance of the Sun is x miles.

Uncertainty of Rank A. The mean distance of the Sun is known as a Random Variable. For all x we have a known function $P(x)$ giving the probability that the distance is less than x. If $P(x)$ is continuous there will thus be a known function x_P, such that the probability that the true distance is less than x_P shall be P. All percentile points of the distribution of the random variable are known with exactitude. The symbolical statement

$$Pr(x < x_P) = P.$$

is then justified by the known functional relationship between P and x_P.

Uncertainty of Rank B. The percentile points of the distribution spoken of in A above, are not known, but probability statements may be made about them. Each then is a random variable, so that $x_{PP'}$ is the value of x such that

$$Pr(x_P < x_{PP'}) = P'$$

or, in a more elaborate notation,

$$Pr\{Pr(x < x_{PP'}) > P\} = P',$$

and $x_{PP'}$ is a known function of P and P'.

Uncertainty of Rank C. If, however, $x_{PP'}$ is not known with certainty in terms of P and P', but if given P and P' it has a known distribution as a random variable, the uncertainty may be said to be of rank C.

Evidently an unending series of such ranks is defined by an extension of the same process.

Since in statistical discussion it is common to speak of a *probability* as an unknown quantity, appropriate ranks should be assigned to such statements as follows :

Certainty. The Event will occur.

Uncertainty of Rank A. The Event has a known probability, π.

Uncertainty of Rank B. For all P a function π_P is known such that the probability that π is less than π_P is exactly P.

Uncertainty of Rank C. π_P is given as a random variable, such that a known function of P and P', namely $\pi_{PP'}$ shall exceed π_P with frequency P'.

In respect to the success or failure of an Event, it will be noticed that the statements of probability *a priori* and *a posteriori* of Bayes are both statements with uncertainty of Rank B. A gambler's statement that he has a probability of one in 36 of throwing a pair of sixes leaves the Event with the simplest kind of uncertainty, of Rank A.

2. Bets and side-bets based on uncertain information

It might be thought that information at the levels of deeper uncertainty, C, D, E etc., is necessarily so tenuous and remote that there is no practical advantage to be gained by possessing it. I do not believe that such an opinion should be formed hastily. The gain immediately in view is, indeed, no more than the clarification of thought, but such clarification is not necessarily without practical advantages. The primitive method of exploiting a knowledge of probability (Rank A) for personal advantage is by laying advantageous bets; for although every particular event is recognized to be uncertain, a man (A) who believes that, in throwing a die, an Ace will be turned exactly once in six trials, will *ipso facto* be aware that if he can bet four to one against the Ace, or 4 to 21 in its favour, though he may often lose, yet on his data, and in the long run, his bet will certainly be an advantageous one.

Now it may be that the information available is not so certain as A supposes; a second gambler B may be aware that there is variability among the dice in the box from which A has chosen one at random, and although he may have no knowledge of which has been chosen, if he has, or thinks he has, information to the effect that quite 10% of the dice in this box will throw an ace more than once in five trials, he will feel justified in accepting odds at 9 : 1 or higher in betting that A will lose in the long run, if he continues with the same die in his present course of action.

It is easy then to perceive that C, who does not know and does not think he knows the percentile points of the probability of throwing an Ace, but believes he knows how the relevant percentile point will be distributed in a consignment of boxes, one of which has been used to supply the die in question, may know with confidence what odds he can profitably accept in betting against the success of B's wager. Nor is C's information of the lowest rank that could be profitably utilized in such a sequence of side-bets.

Although the betting operations I have described cannot be exhibited, I believe, in any orthodox gaming establishment, for the reason chiefly that they take too long, yet in Commerce and Finance very similar operations are customary. A firm of contractors building a bridge will employ engineers who use tables of the strength of materials, and perhaps, empirical data on the frequency of destructive floods, hurricanes or earthquakes, of a strength capable of jeopardising their work. In view of recognised uncertainties of the strength of the structure, and its parts, they build in what are called "factors of safety" at great expense, to cover, as they hope, all reasonable risks. Nevertheless, the corporation operating such a structure will usually seek out an Insurance Office, who at an appropriate premium, or in other words at chosen odds, will underwrite the residual risk of the accidental destruction of the structure. They are in effect betting that for an assigned period the precautions of the engineers will be shown to have been sufficient.

Moreover, even the simple investor risking his savings by buying ordinary stock, perhaps in an Insurance Company, is in effect laying down a stake on the proposition that the company will collect enough in premiums to cover the large risks it has underwritten, and above this to pay a reasonable dividend on their ordinary shares.

The mathematical underworld I have introduced has a qualitative resemblance to the specification of risk in quite familiar operations in our own world. It is only the mathematical tradition that has been too shy of uncertainty, too firmly attached to the ethereal regions of mathematical certainty, to have attempted any very deep exploration.

It is important to recognize that the reference to a consignment of boxes as capable of justifying statements of uncertainty of rank C could in one respect easily be misunderstood, for an examination of a single box of dice could yield a probability integral such as the ogive below:

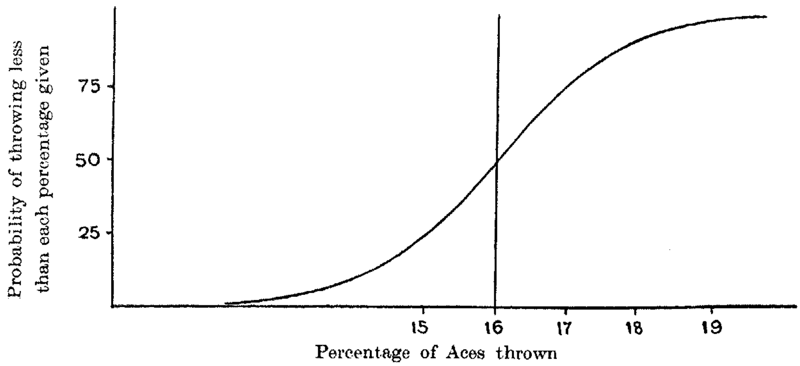

Fig. 1. Representing the kind of information on which B stakes his wager.

so that a Consignment of such boxes might be thought of as yielding of a sheaf of such ogives. This, however, is an unintended effect of an imperfect example. For uncertainty of rank C it is intended that there shall be variation, and well specified variation, in the probability integral as shown on the ordinate at 16% on the diagram, and likewise variation in the probability integral at other percentages, but it is not intended that C should possess knowledge of the covariation, or simultaneous variation, of these several ordinates, such as a sheaf of distinct ogives would imply.

Had the data supplied such a sheaf of ogives as would a consignment of boxes taken literally, the mean ordinate at each percentage would give a single ogive appropriate to the result of emptying all the boxes of a consignment into a single receptacle, and choosing a die at random from this. The uncertainty would then be of rank B. It might enable a gambler to bet successfully *against* B, and not merely *about* the success of B's wager. So long as only a single percentage of Aces is in discussion, and this percentage subject to betting at specified odds, this stipulation makes no difference. When inductive inferences have to be drawn, however, all possibilities must be comprehended.

3. The problem of binomial data

3.1. *General formulation*

Towards the end of the seventeenth century James Bernoulli considered at great length, in his posthumously published work *Ars conjectandi*, what inferences could properly be drawn from a finite empirical record of successes and failures, on the value of the true probability of success. Either the work was never completed, or the author encountered unexpected difficulties, for the promise of the early chapters was never fulfilled. Two generations later Thomas Bayes, by introducing, in addition to the data, the notion of probabilities *a priori* (regarded as possibly axiomatic) was able to deduce a definite frequency distribution for the probability, namely, for a successes and b failures out of n trials,

$$\frac{(a+b+1)\,!}{a\,!\;b\,!}\,p^a(1-p)^b dp\;;$$

so that all percentile points were exactly definable by the incomplete Eulerian integrals.

There are, indeed, cases, such as that laid down by Bayes for demonstrating his theorem, in which information of an *a priori* type based on a previous and independent act of random sampling, are really available; and for these cases Bayes' argument, and his conclusion, with uncertainty of rank B, are exact. In most cases in the Natural Sciences, however, such knowledge *a priori* is lacking, and all efforts to express the inference proper to a rational mind, derivable from the observations only, in terms of probability statements about the unknown probability p, have been unsuccessful. I propose to show that the exact inferences properly supplied by such data can be expressed in uncertainty statements of Rank C.

In finding what probability statements of whatever rank may be inferred about the unknown probability, π, we shall be concerned with the probabilities, in random samples, of all possible conjunctions of possible values of π, with all possible observational records. The actual observations recorded are then inserted as particular cases known to be of the general class of random samples, to provide the particular set of probability statements appropriate to the particular unknown from which they were derived.

For binomial data, showing a successes out of n trials, let c_r stand for term of the binomial expansion

$$\frac{n\,!}{r\,!\,(n-r)\,!}\,\pi^r(1-\pi)^{n-r}$$

and let

$$P = c_0+c_1+\;\ldots\;+c_{a-1}+\lambda c_a.$$

P is then an explicit function determined by the observational data, by the unknown π, and by an auxiliary quantity λ which is to be distributed, independently of the observations, in a rectangular distribution from 0 to 1, so that independently of all else

$$Pr(\lambda < \lambda_1) = \lambda_1 \qquad 0 \leqslant \lambda,\, \lambda_1 \leqslant 1.$$

Simultaneous values of

(i) the observations a, b, $(a+b = n)$.

(ii) the unknown probability, π,

(iii) the random variable, λ,

constitute a contingency, and of such contingencies the probability that

$$P \equiv \sum_0^{a-1} c_r+\lambda c_a,$$

a known function of (π, λ) given the observations, is less than P' is simply equal to P'.

If P' were, for example, 1% then one per cent of all contingencies will have P less than P', and may be termed the first percentile of all contingencies. The boundary of this region, in which $P = P'$, defines the 1% value of π for each possible λ, namely $\pi_{P\lambda}$.

Since λ has a known distribution independently of all else, its variation defines the sampling distribution of $\pi_{P\lambda}$ as a random variable.

3.2. *Formal analysis*

In a more formal notation we may denote by $\pi_{P\lambda}$ that function of P and λ for which the known function of π and λ

$$\sum_{0}^{a-1} c_r + \lambda c_a$$

shall be equal to P, where $\pi_{P\lambda}$ is substituted for π. Then $\pi_{P\lambda}$ will increase monotonically with P, and decrease monotonically as λ is increased, for all values of P and λ, so that if P is less than P_1 it follows that

$$Pr(\pi < \pi_{P_1\lambda_1}) = P_1,$$

for $\lambda = \lambda_1$, and

$$Pr(\pi < \pi_{P_1\lambda_1}) > P_1$$

for $\lambda < \lambda_1$.

But, unconditionally

$$Pr(\lambda < \lambda_1) = \lambda_1.$$

Hence, dropping suffixes,

$$Pr\{Pr(\pi < \pi_{P\lambda}) > P\} = \lambda,$$

for all values of P and λ.

This is an uncertain, though mathematically precise, statement about π of Rank B, and therefore of Rank C, about the Event having probability π.

3.3. *Verification*

A concrete notion of the probability statements inferred from simple binomial data may be gained from the experimental process of verification. Such processes are always laborious and their specification is no more than a means of making clear the exact meaning of the inferences to one who wishes to understand their relation to a possible experimental test.

First, a random sample of n is taken and the number of successes counted. Secondly, a value of λ is chosen at random so that it shall fall with equal probability within all equal ranges from 0 to 1.

Next $\pi_{P\lambda}$ is calculated with P equal, for example, to .99, and B bets 99 to 1 that π, the true probability of success, is less than $\pi_{P\lambda}$. Comparing B's calculated value with the true value π, it is ascertained whether or not he has won his bet.

If for every possible sample he can encounter B gives to λ all possible values from 0 to 1 with even frequency, then it is demonstrable, and will be verified experimentally that B loses his bet with a frequency of just 1%. $\pi_{P\lambda}$ may properly be called the upper 1% value of π deducible from the observations. It is, however, not known save as a random variable.

Gambler C may now observe that, in the particular circumstances in which B is betting $\pi_{P\lambda}$ has a definite distribution determined by the known frequency distribution of λ. He has the material for betting with confidence on the 1% value found by B for any particular sample.

If it were thought that the values $\pi_{P\lambda}$, for the same value of λ for different possible samples, could be associated together as a recognizable sub-set of possibilities, this would associate the percentile points for different values of P so as to specify their covariation in a way which would imply uncertainty of no more than rank B. No probability statements, however, true for all π, can be deduced for such sub-sets.

3.4. *A numerical example*

As an example we may take a case such as that in which an event has happened three times only among 14 random tests. The possible upper one per cent points ($P = .99$) must lie in the range bounded by the value of p at which the first three terms

$$q^{14}+14pq^{13}+91\ p^2q^{12}$$

constitute exactly 1% of the distribution, and that for which the inclusion of the fourth term

$$364p^3q^{11}$$

brings the total to exactly this percentage. The value of p required in the first case is

$$p_1 \quad .478244,$$

and in the second case

$$p_2 \quad .556655.$$

These constitute the limits of the distribution; they provide the absolute inequalities satisfied by the upper 1% point of the possible values of the probability in the population from which the sample was drawn.

The simplest method of arriving at values of p (or q) which make the partial sum of a binomial series equal to an assigned percentage of the whole consists in the use of Table V of *Statistical Tables* as explained in the *Introduction* Ex. 2.

In the first case three out of the 15 terms of the binomial expansion are to be contrasted with the remaining twelve terms. Doubling these numbers we use 6 for n_1 and 24 for n_2. The table for the Variance Ratio at the 1% point then gives 3.67. Multiplied by 6/24 this corresponds to the ratio of "sums of squares" 0.9175, and so to $p = .9175/1.9175$ or

$$p_1 \quad .47849.$$

A more accurate value is obtained from the corresponding value of z given on the page facing, namely .6496; doubling this and taking the Naperian antilog, the variance ratio comes to 3.66637, and the lower limit of the distribution of the upper 1% point, to

$$p_1 \quad .478240$$

correct to two more places. The table of z, for the range of percentile values for which it is available, will be suitable for determining the end points of such distributions.

For chosen values of p between these end points, the corresponding values of λ will be required. These are calculated directly from

$$\lambda = (0.01 - c_0 - c_1 - c_2)/c_3$$

in which c_0, c_1, etc., stand for the successive binomial terms.

In the present case I find :

p%	λ
48	.01101
49	.07966
50	.15890
51	.25156
52	.36121
53	.49258
54	.65160
55	.84584

The values of λ supply the probabilities of the random variable falling below the values of p indicated. The ordinates of the curve are to be found from the value of

$$y = d\lambda/dp,$$

for constant values of P, and these in the case of the binomial distribution take the simple form

$$y = \frac{3}{p}(1-\lambda) + \frac{11}{q}\lambda,$$

in which the coefficients are the observed numbers of occurrence and non-occurrence of the primary event. The ordinates take the finite values,

$$\frac{3}{p_1} \text{ and } \frac{11}{1-p_2}$$

at the termini where p takes the values p_1 and p_2. The range must, generally speaking, be commensurate with the reciprocal of the total number observed. This is an order of magnitude which, as the sample size increases, tends to zero compared with the standard deviation of the estimate

$$\sqrt{\pi(1-\pi)/n}\ ;$$

the recognition of uncertainty of Rank C may therefore be regarded as a refinement appropriate to the exact treatment of such samples.

Fig. 2 shows the distribution of the upper 1% point in the case considered.

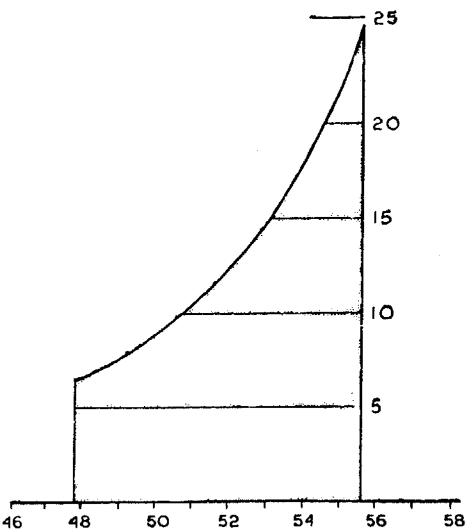

Fig. 2. Distribution of the upper 1% point indicated by the data for an unknown probability. (lower limit 47.82, upper limit 55.67)

π is the true value of the probability of an event of which all that is known is that it has been observed 3 times in 14 trials.

(a) No value for π can be asserted.

(b) No value can be stated such that the probability of π exceeding that value is 1% (or any other percentage).

(c) Such a 1% value can, however, be specified as a random variable, as shown in this Figure.

It would not be legitimate to sample the distribution (*c*) provided by the observations with the intention of asserting that the particular value found was the 1% point provided by them for the distribution of π; it would be equally illegitimate, and for exactly the same reason, in cases such as that of Bayes, in which the data provide a distribution (*b*) of this unknown, to choose a value at random from this distribution, and to assert that the particular value found in this way was the value of π provided by the data; or, finally, if the true value of π were known, but the happening of the event were still uncertain, it would do nothing to diminish this uncertainty to use the apparatus of gambling to see whether an event with exactly this probability did or did not occur in a random trial. Such so-called randomisation tests, proposed in the name of Decision Functions, serve only to obscure the elucidation of the particular nature of the uncertainty in which observations of a given kind leave the rational observer.

In binomial sampling the case in which all the experimental tests give like results, as, for example, if no success had been recorded in 14 trials, deserves special notice. For in such cases it would be possible for π to be zero, so that the range of any percentile point is from zero to

$$p = 1 - P^{1/14}$$

as its upper terminus.

Since now,

$$\lambda q^{14} = P,$$

it is clear that λ cannot b e less than P, and that in a fraction P of all cases the percentile of the distribution of π represented by π_P is included in a condensation at zero. The remainder of the distribution may be represented as a frequency curve with area $1-P$, between zero and its finite upper bound.

The ordinate,

$$y = \frac{14}{q}\lambda$$

varies from $y = 14P$ to $14/P^{1/14}$.

References

Bayes, T. (1763): An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc.*, **53**, 370.

Bernoulli, J. (1713): *Ars conjectandi*, opus posthumum, Basileae.

Fisher, R. A. (1956): *Statistical Method and Scientific Inference*, Oliver and Boyd, Ltd., Edinburgh.

Fisher R. A. and Yates, F. (1953): *Statistical Tables for Biological, Agricultural and Medical Research*, 4th edition. Oliver and Boyd, Ltd., Edinburgh.

Paper received : September, 1957.