151

ON A POINT RAISED BY M. S. BARTLETT ON FIDUCIAL PROBABILITY

By R. A. FISHER, Sc.D., F.R.S.

I. Introductory

In recent years a considerable amount of attention has been attracted by a type of inductive argument to which the term "fiducial" has been applied, which was first explicitly developed in 1930 (1) in a paper entitled "Inverse Probability". Illustrations, and more or less general expositions of the principle, have been given in several recent papers. The limitation, inherent in any attempt at exact inductive reasoning, that the data discussed shall include the totality of the relevant information, has been frequently emphasized. When this is granted, it has been shown that in certain well-defined cases we may infer, for some of the unobservable parameters by which the population sampled is specified, definite frequency distributions, which are not to be conceived of, or confused with, the *a priori* and *a posteriori* distributions discussed in the theory of inverse probability. The subject, therefore, is one that requires some refinement of statement, without which paradoxes and apparent contradictions are bound to result. One fruitful source of these is the use of insufficient estimates of the parameters. A second is the introduction, into an argument of this type, of fixed values for the parameters, an introduction which is bound to conflict with the fiducial distributions derivable from the data.

In a recent paper entitled "The information available in small samples", M. S. Bartlett (2), although avowedly using the fiducial argument, is led to question the validity of a test of significance, originally put forward by W.-V. Behrens (3), and since recognized (4) as capable of justification and generalization in terms of fiducial probability. Behrens was not, I think, conscious of introducing any new logical principle, and his argument, in the terms in which it was conceived, may therefore have been fallacious, in the way that Bartlett seems to suspect. The problem he discusses of a test of significance for the difference between the means of two samples of populations, not supposed to be equally variable, affords, however, an excellent example by which the nature and implications of fiducial reasoning may be examined. Since Bartlett chooses for discussion the case in which each sample consists of two observations only, a case which presents some mathematical peculiarities, it may be well to state in advance some properties of the distribution characteristic of "Student's" t when n is unity.

 Annals of Eugenics, 7: 370-375, (1937).

II. The distribution of t for one degree of freedom

For the aggregate of all samples of two (y_1, y_1'), of the first population, the distribution of

$$\frac{y_1+y_1'-2\mu}{|y_1-y_1'|} = t_1,$$

where μ is the hypothetical mean of the population, is

$$dp = \frac{1}{\pi}\frac{dt_1}{1+t_1^2};$$

for the aggregate of all samples of two (y_2, y_2'), of the second population, the distribution of

$$\frac{y_2+y_2'-2\mu}{|y_2-y_2'|} = t_2$$

is

$$dp = \frac{1}{\pi}\frac{dt_2}{1+t_2^2},$$

independently of t_1.

It is a well-known property of the distribution of t_1 and t_2, that the distribution of any real linear function,

$$\psi = at_1+bt_2,$$

is given by

$$\frac{\psi}{|a|+|b|} = t,$$

where t has the same distribution as t_1 and t_2.

It is not true, if a and b are of opposite signs, that

$$\frac{\psi}{|a+b|}$$

is so distributed.

Consequently, when, using the fiducial argument, we put $|y_1-y_1'|$ and $|y_2-y_2'|$ for a and b, we find that the distribution of

$$\frac{y_1+y_1'-y_2-y_2'}{|y_1-y_1'|+|y_2-y_2'|}$$

is that of t. This appears in Bartlett's work as his positive solution. It corresponds with Behrens' test of significance. It is not true that

$$\frac{y_1+y_1'-y_2-y_2'}{|y_1-y_1'|-|y_2-y_2'|}$$

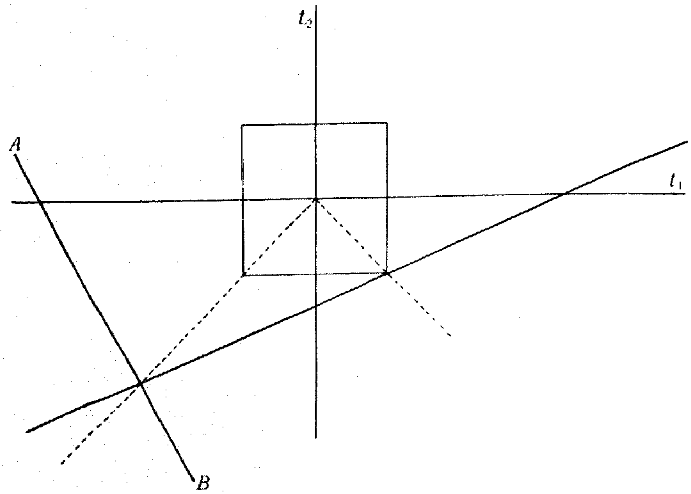
is distributed in the t distribution, as is implied by Bartlett's negative solution. It is clearly not accurate in this context to say as Bartlett does, doubtless with quite other ideas in mind, that "the sign + or − is distributed at random".

From the fiducial point of view the distinction between the positive and the negative solutions of Bartlett is clearly shown by a diagram on the t_1, t_2 plane (p. 372).

If t_1 and t_2 are the rectangular co-ordinates of a point, the frequency density over the plane is determinate. Although both t_1 and t_2 depend on the unobservable common mean of the populations, μ, yet, if we eliminate μ, we find

$$t_1 \,|\, y_1 - y_1' \,| - t_2 \,|\, y_2 - y_2' \,| = y_1 + y_1' - y_2 - y_2',$$

defining a line on the plane, given by direct observation. The slope of this line will always be positive. It will proceed from the bottom left to the top right quadrant. If the means indicated by the two samples are closely alike, the line will pass near to the origin; if they are widely discrepant, it will lie wholly in the outer regions where the frequency density is sparse. For any given line it is possible to say how much of the total frequency of the distribution lies on the same side as the origin, and how much on the other side. This probability is a direct fact of observation.



Now, in the case considered by Bartlett, where each sample is of two observations, each coordinate is based on one degree of freedom, and it follows from the property stated above that the envelope of lines cutting off a fixed percentage of the total frequency is a square. For higher values of n we have, of course, convex curves. The envelope corresponding to any given line is found from its intersection with

$$t_1 + t_2 = 0;$$

so that

$$t_1 = \frac{y_1 + y_1' - y_2 - y_2'}{|\, y_1 - y_1' \,| + |\, y_2 - y_2' \,|} = -t_2.$$

Naturally, also, the observed line will usually meet the other diagonal,

$$t_1 = t_2,$$

at the point

$$t_1 = \frac{y_1 + y_1' - y_2 - y_2'}{|\, y_1 - y_1' \,| - |\, y_2 - y_2' \,|} = t_2,$$

but the line having t_2 constantly equal to this value will not cut off a total frequency equal to that of the observed line, but a smaller frequency, as shown in Bartlett's table. Bartlett's negative solution, in fact, treats the observed line as though it were one of those, such as AB in the diagram, touching the *outer* square.

III. The introduction of hypothetical variances

What is, I believe, the main cause of confusion, not only to Bartlett, but to others engaged in exploring the possibilities of the fiducial type of argument, is its independence of the background of hypothetical parameters, in terms of which the populations sampled may be specified.

Thus, if we treat v_1, the variance of the first population as *given*, we immediately violate the conditions of the fiducial argument, for this is information not supplied by the sample. The distribution of t_1 is now not independent of $|y_1 - y_1'|$. On the contrary, that of

$$X_1 = y_1 + y_1' - 2\mu = t_1 |y_1 - y_1'|$$

is independent of $|y_1 - y_1'|$, and is given by

$$dp = \frac{1}{2\sqrt{(\pi v_1)}} e^{X_1^2/4v_1} dX_1,$$

for every value of $|y_1 - y_1'|$. Hence $X_1 - X_2$ also has a normal distribution about zero, independently of $|y_1 - y_1'|$ and $|y_2 - y_2'|$, with variance $2(v_1 + v_2)$. If we do not know $v_1 + v_2$, this affords no test of significance. Bartlett, however, points out that either of the quantities

$$(y_1 - y_1') \pm (y_2 - y_2')$$

will, on the same theory, be distributed in the same manner. From this it would be right to infer that both the ratios

$$\frac{X_1 - X_2}{|(y_1 - y_1') + (y_2 - y_2')|} \quad \text{and} \quad \frac{X_1 - X_2}{|(y_1 - y_1') - (y_2 - y_2')|}$$

will be distributed as is t for one degree of freedom. In fact "Student's" test was first developed for *paired* observations. If we have two pairs only, y_1, y_2 and y_1', y_2', the second expression will be "Student's" t. This argument does not, however, lead to the statements that

$$\frac{X_1 - X_2}{|y_1 - y_1'| + |y_2 - y_2'|} \quad \text{and} \quad \frac{X_1 - X_2}{||y_1 - y_1'| - |y_2 - y_2'||}$$

are so distributed, for the denominators of these have much more complicated distributions, involving the ratio of the theoretical variances.

In the fiducial approach, on the contrary, we regard v_1 and v_2 as integrated over their fiducial distributions. This requires that, for given *estimated* variances, t_1 and t_2 have the distribution given above. It is, in consequence, not necessary to introduce v_1 and v_2 at any stage.

IV. The actual distributions of Bartlett's ratios

To clear up a confusing point, it may be worth while to pursue Bartlett's approach a little further.

The distribution of $$s = |y_1 - y_1'| + |y_1 - y_2'|$$

is $$dp = \frac{4}{\sqrt{\{2\pi(v_1+v_2)\}}} e^{-\frac{s^2}{2(v_1+v_2)}} \left\{ p\left(\frac{s}{\sqrt{(v_1+v_2)}} \cdot \sqrt{\frac{v_2}{v_1}}\right) + p\left(\frac{s}{\sqrt{(v_1+v_2)}} \cdot \sqrt{\frac{v_1}{v_2}}\right) - 1 \right\} ds,$$

where $$p(x) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt.$$

From this it follows that the distribution of

$$T = \frac{X_1 - X_2}{s}$$

is $$dp = \frac{1}{\pi(1+T^2)} \left\{ \frac{\sqrt{\phi}}{\sqrt{(1+\phi+T^2)}} + \frac{1}{\sqrt{(1+\phi+\phi T^2)}} \right\} dT,$$

where ϕ stands for v_2/v_1.

The probability integral is

$$\frac{1}{2} + \frac{1}{\pi} \left\{ \sin^{-1} T \sqrt{\left(\frac{\phi}{(1+\phi)(1+T^2)}\right)} + \sin^{-1} T \sqrt{\left(\frac{1}{(1+\phi)(1+T^2)}\right)} \right\}.$$

It is therefore easy to find that the 5 per cent value of T rises from 3·4272 at $\phi = 1$, to 12·7062 at $\phi = 0$ or ∞. The fiducial test requires that T shall exceed 12·7062 (tan 85° 30′).

On the other hand, if $$s' = ||y_1 - y_1'| - |y_2 - y_2'||,$$

the distribution of s' is

$$dp = \frac{4}{\sqrt{\{2\pi(v_1+v_2)\}}} e^{-\frac{s'^2}{2(v_1+v_2)}} \left\{ 2 - p\left(\frac{s'}{\sqrt{(v_1+v_2)}} \cdot \sqrt{\frac{v_2}{v_1}}\right) - p\left(\frac{s'}{\sqrt{(v_1+v_2)}} \cdot \sqrt{\frac{v_1}{v_2}}\right) \right\} ds'.$$

The distribution of $$T' = \frac{X_1 - X_2}{s'}$$

is therefore $$dp = \frac{1}{\pi(1+T'^2)} \left\{ 2 - \frac{\sqrt{\phi}}{\sqrt{(1+\phi+T'^2)}} - \frac{1}{\sqrt{(1+\phi+\phi T'^2)}} \right\} dT'.$$

The probability integral is

$$\frac{1}{2} + \frac{1}{\pi} \left\{ 2\tan^{-1} T' - \sin^{-1} T' \sqrt{\left(\frac{\phi}{(1+\phi)(1+T'^2)}\right)} - \sin^{-1} T' \sqrt{\left(\frac{1}{(1+\phi)(1+T'^2)}\right)} \right\},$$

from which it follows that the 5 per cent point falls from 25·0599 to 12·7062 as ϕ changes from 1 to 0.

From any pair of samples of 2, both T and T' are available. T' is never the smaller. If ϕ were known, although in this case T would not be used in practice, yet some pairs of samples could be judged to be significantly different by the value of T, which, without this knowledge, could not be so judged. If any pair were judged significant by the fiducial test, which is available in the absence of knowledge of ϕ, it would also have been so judged had ϕ been known to have any definite value. An experimenter who, after drawing the samples, tossed up to decide whether to use T or T', and then judged the difference between the pairs to be significant if his chosen value exceeded 12·7062, would certainly find, in the absence of any real difference, that the judgment of significance was given in 5 per cent of trials. If the choice were determined by physical pairing, this would be "Student's" test (5). Bartlett seems to suggest this as a correct test of significance, in the absence of physical pairing, and to judge Behrens' test to be incorrect by contrast. Bartlett's test of significance appears to require, either that the knowledge of whether the larger value, T', or the smaller value, T, had been chosen by lot has been obliterated from the experimenter's mind, or that he is supposed to accept the judgment of significance with the same confidence, when he knows that chance has chosen T', and that T is much smaller, as he might feel if T also had been large. Knowing that T' increases with *greater* disparity between the members of one sample, the other being unchanged, so that it reverses the order of significance ascribed to different sets of possible observations, he would, I think, rightly reject the tests based on T'; and, if ϕ were unknown, require T to satisfy the fiducial test.

That this test corresponds with the most stringent of those based on T for any known value of ϕ is, I suppose, a consequence of there being only the minimum number of degrees of freedom, 1, available for the estimation of v_1 and v_2. With samples of more than 2, I should expect some differences fiducially significant to be found insignificant, if tested for some particular values of the variance ratio, these being ratios which the data themselves had shown to be unlikely.

V. Summary

The criticism of Behrens' test of significance, recently put forward by Bartlett, on the ground that it differs from a possible alternative test, overlooks the inconsistency of assuming for the unknown variances both (*a*) fiducial distributions in accordance with the samples observed, and (*b*) values fixed from sample to sample.

The alternative test of significance proposed involves, when the variance ratio of the two populations sampled is unknown, the choice by lot between the value T, used in Behrens' test, and a second value T', which reverses the order of significance of different possible sets of observations. High values of T' are not, therefore, by themselves evidence of inequality of the means.

For known values of the variance ratio, the distributions and probability integrals of T and T', separately, are given in this note.

REFERENCES

(1) Fisher, R. A. (1930). "Inverse probability." *Proc. Camb. Phil. Soc.* **26**, 528–35.
(2) Bartlett, M. S. (1936). "The information available in small samples." *Proc. Camb. Phil. Soc.* **32**, 560–6.
(3) Behrens, W.-V. (1929). "Ein Beitrag zur Fehlerberechnung bei weinige Beobachtungen." *Landw. Jb.* **68**, 807–37.
(4) Fisher, R. A. (1935). "The fiducial argument in statistical inference." *Ann. Eugen.* **6**, 91–398.
(5) "Student" (1908). "The probable error of a mean." *Biometrika*, **6**, 1–25.