COMPLETE SIMULTANEOUS FIDUCIAL DISTRIBUTIONS

By M. S. Bartlett

1. Introduction. In a recent paper in these Annals, Starkey [13] has made some investigation of the distribution[1] related to the Behrens-Fisher test of the difference between two means from normal populations with unequal variances. She does not, however, give any critical discussion of the validity of this proposed test in the light of criticisms that have been made of it. It may therefore be an appropriate opportunity of reviewing the theory of fiducial distributions, as I see it, up to the present stage of development,[2] and in particular, of referring to the idea of complete simultaneous fiducial distributions. In conclusion I have made some brief comment on the particular problem at issue, in the light of this general theory; and have added a note on the use of approximate tests.

2. Fiducial Probability. If from a sample denoted symbolically by S a statistic T is obtained whose chance distribution depends on one unknown parameter θ, the distribution of T being of the form

$$p(T \mid \theta) = f(T, \theta)\, dT,$$

and if the values of T bear a regular increasing relationship with θ, (for an assigned value of the probability integral), then for any particular value $T = T_0$, we may assert that $\theta \geq \theta_0$, where

$$\int_{-\infty}^{T_0} p(T \mid \theta_0) = 1 - \epsilon,$$

and we shall know that this assertion, in the system of inferences based on the above rule, will have an exact and known probability of being wrong, given by ϵ.

The inference is thus an uncertain one, but the extent of the uncertainty is exactly known, and as stressed by Fisher [6], who first introduced this important concept of fiducial inferences and fiducial probability, is completely independent of any *a priori* notion of what value θ is likely to be.

It might be emphasized, to avoid confusion, that the inference is a *deduction* from the standpoint of logic, and still requires, if applied in practice, the necessity of inductive assumptions concerning the applicability of the mathematical theory, but its avoidance of any appeal to *a priori* probability in regard to the value of θ gives it a completely independent status district from the classical inverse probability argument, from which it should be distinguished. The

[1] This distribution has also been studied by Sukhatme [14].
[2] See also the recent expository article by Wilks [16].

interval assigned to θ is to some extent arbitrary, and we can more generally choose θ_0 and θ_1 such that the fiducial probability of

$$\theta_1 \leq \theta \leq \theta_0$$

is equal to $1 - \epsilon$. While this fiducial probability is a probability in a formal mathematical sense, I have suggested [2] that its special meaning in regard to the inference on θ might be emphasized if we distinguished it by a special symbol. Since intervals (θ_1, θ_0) can be built up for all values of ϵ, we can represent them all by the general distributional expression

$$\int_{-\infty}^{\theta} f_p(\theta \mid T) = \int_{T}^{\infty} p(T \mid \theta),$$

which defines the fiducial probability distribution $f_p(\theta \mid T)$.

From the point of view of mathematical theory T is, so far, any statistic, but Fisher restricted the term fiducial probability for those cases where T was a sufficient statistic for θ, in order that the fiducial inference should be based on a sample statistic which could justifiably claim to contain all the information on θ available from the sample.

The general theory of interval estimation, without this restriction, has been subsequently examined by Neyman (e.g. [10]) under the name of the theory of confidence intervals. In this general theory there is no particular restriction on the number of parameters involved, for it may be possible in the coordinate space represented by parameters θ_r (for which there are statistics T_s) to define a region $R(T_s)$ for which the assertion that the vector parameter θ_r lies in the region $R(T_s)$ has a known probability $1 - \epsilon$ of being correct.

A difficulty, however, in a multi-parameter theory of fiducial distributions is that it does not in general seem possible, even when T_s is a vector statistic representing a joint set of sufficient statistics for θ_r, to define a simultaneous fiducial distribution $f_p(\theta_r \mid T_s)$ which will be consistent with one-variate distributions $f_p(\theta \mid T)$ relating to one particular parameter θ. For such consistency we must have the symbolic integration

$$\int f_p(\theta_r \mid T_s)$$

over all θ_r other than θ yielding $f_p(\theta)$ as a result. A further discussion of this difficulty is given in Sections 4 and 5, after the theory of one-variate fiducial distributions has been more completely discussed.

3. Fiducial Distributions and Properties of Sufficiency. If we now consider the extension of one-variate fiducial inferences to the case where other parameters exist but are unknown, we are led to examine the various types of sufficient statistic which are related to the theory of estimation of one parameter when other parameters are unspecified (Bartlett, [1]). By a distribution of fiducial type we shall mean a distribution providing at least confidence inter-

-als in the sense of Neyman. This distribution will be defined as *the* fiducial distribution for θ if the statistic used (conditional or unconditional) satisfies the necessary sufficiency properties given in the paper just referred to (sections 6, p. 132, and 7, p. 136). This definition is understood to include the possibility mentioned in section 7, where $T_1 \mid T_2(\theta)$ is a conditional statistic of the type required for any specified value of θ ($T_1 \mid T_2$ denotes T_1, given T_2). For example, the theoretical statistic $\bar{x} \mid \Sigma(x - m)^2$ in normal theory, where $\bar{x}$ is the sample mean and $\Sigma(x - m)^2$ the sum of squares of deviations from the population mean m, is of this form, and since

$$p(\bar{x} \mid \Sigma(x - m)^2) = p(t),$$

a fiducial distribution for m is obtained from the familiar Student's t-distribution.

As other developments of fiducial theory we may note (i) its application to fiducial inferences on sufficient statistics in unknown samples (this application to *normal* samples has been discussed by more than one writer, see, for example, Fisher [8]; I have moreover indicated the general theory underlying such applications [3]) (ii) the case of discontinuous or "discrete" sampling distributions, for which the theory of exact fiducial distributions breaks down.

In the latter case, it is only possible to choose an interval for θ, such that the chance of our fiducial inference being incorrect is *not greater than* ϵ (see, for example, Clopper and Pearson [5]). This "inexact theory" I have shown [3] may also be extended to inferences on sufficient statistics in unknown samples. In particular, from the general distribution

$$p(r_1, r_2 \mid r) = \frac{n_1!}{(n_1 - r_1)!\,r_1!}\,\frac{n_2!}{(n_2 - r_2)!\,r_2!}\,\frac{(n - r)!\,r!}{n!}$$

giving the number of ways of assigning r members with some attribute A in numbers r_1 and r_2 to samples S_1 and S_2, sizes n_1 and n_2, we have for $n_2 = 1$,

$$p(r_2 \mid r) = \begin{cases} \dfrac{r}{n_1 + 1} & (r_2 = 1) \\[2ex] \dfrac{n_1 + 1 - r}{n_1 + 1} & (r_2 = 0). \end{cases}$$

Thus if S_1 contains r_1 members with a certain attribute, such that

$$\frac{r_1 + 1}{n_1 + 1} \leq \epsilon,$$

we may assert that a new member from the same population will not possess the attribute. If

$$\frac{n_1 + 1 - r_1}{n_1 + 1} \leq \epsilon,$$

we assert that the new member will possess the attribute. If r_1 does not conform to either inequality, we cannot, with the limit of error imposed, commit ourselves. The probability that our variable assertion, based on the above rule, is wrong, is then not greater than ϵ. (This type of inference may be contrasted with the Law of Succession in the theory of inverse probability. In this rather degenerate example it is not of course surprising that the nature of the inference we can make is not always very profound!)

4. Simultaneous Fiducial Distributions. It was pointed out in section 2 that an inference of fiducial type might be made regarding a joint interval containing unknown parameters θ_r, this interval or region being a variable function of the (continuous) statistics T_s. If a sufficient set of statistics T_r $(r = 1 \cdots k)$ exist for the parameters θ_r $(r = 1 \cdots k)$, that is, if we have

$$p(S \mid \theta_r) = p(T_r \mid \theta_r)p(S \mid T_r)$$

where T_r denotes the set $T_1 \cdots T_k$, and similarly for θ_r; and if we can write

$$p(T_r \mid \theta_r) = p(\varphi_r)$$

where the distribution of the set of theoretical functions φ_r of T_r and θ_r is independent to any further extent of θ_r, then we may write also

$$f_p(\theta_r \mid T_r) = p(\varphi_r)$$

as the simultaneous fiducial distribution of the θ_r (cf. Fisher, [8]). This notation allows implicitly for the formal transformation from one set of variates to another, the last equation meaning that $p(\varphi_r)$ provides the fiducial distribution of the θ_r, when it is regarded as a distribution in θ_r. For the equations to hold, however, the Jacobian of the transformations must not change sign anywhere in the sample space, this condition determining both the formal identity of the two sides of the equations and also the necessary one-to-one relationship between values of θ_r and T_r.

It has been shown by Segal [12] that if the sufficient set T_r exist, the functions φ_r also exist. For we may define φ_r by the equations

$$\begin{cases} \varphi_1 = \int_{-\infty}^{T_1} p(T_1) \\ \varphi_2 = \int_{-\infty}^{T_2} p(T_2 \mid T_1) \\ \vdots \end{cases}$$

so that

$$p(T_1)p(T_2 \mid T_1) \cdots = d\varphi_1 d\varphi_2 \cdots \qquad (\varphi_r, 0 - 1).$$

The above theory is also immediately applicable to quasi-sufficient statistics, it being merely necessary to consider the appropriate conditional distributions.

5. Complete Simultaneous Fiducial Distributions. It has been emphasized [3] that the simultaneous fiducial distribution $f_p(\theta_r \mid T_r)$ obtained from a sufficient set of statistics must not be interpreted analogously to a simultaneous distribution $p(T_r \mid \theta_r)$. For example, if the set T_r represent the sufficient statistics $\bar{x}$ and s^2 for the unknown mean and variance of a normal population we have

$$p(\bar{x}, s^2 \mid m, \sigma^2) = p\left(\frac{\bar{x} - m}{\sigma}, \frac{s^2}{\sigma^2}\right)$$
$$= f_p(m, \sigma^2 \mid \bar{x}, s^2),$$

but this does not imply that a fiducial inference could be made for one unknown parameter defined by $\theta = m + \sigma$ by integration of the above fiducial distribution after formal change of variable.

We may, however, in certain cases show that consistency relations are satisfied which justify to a much further extent our calling $f_p(\theta_r \mid T_r)$ a simultaneous fiducial distribution. Unfortunately this last expression has already been appropriated for $f_p(\theta_r \mid T_r)$ in general; we shall therefore call $f_p(\theta_r \mid T_r)$ a *complete* simultaneous fiducial distribution if (taking $k = 2$ for simplicity)

$$f_p(\theta_1, \theta_2) = f_p(\theta_1 \mid \theta_2) f_p(\theta_2)$$
$$= f_p(\theta_2 \mid \theta_1) f_p(\theta_1),$$

where the fiducial distributions on the right are known to exist, and their form determined, from the theory of one-variate fiducial distributions. For example, if we consider again the normal sample, we have

$$p\left(\frac{\bar{x} - m}{\sigma}, \frac{s^2}{\sigma^2}\right) = p\left(\frac{\bar{x} - m}{\sigma}\right) p\left(\frac{s^2}{\sigma^2}\right)$$
$$= f_p(m \mid \sigma^2) f_p(\sigma^2)$$

and also

$$= p\left(\frac{\bar{x} - m}{\sqrt{\Sigma}}\right) p\left(\frac{\Sigma}{\sigma^2}\right)$$
$$= f_p(m) f_p(\sigma^2 \mid m)$$

where $\Sigma \equiv \Sigma(x - m)^2$.

These relations imply not only that a fiducial region for m and σ^2 can be determined from the observed values of $\bar{x}$ and s^2, but that in particular, the region can be chosen so that (i) it is some section of an area bounded by two lines parallel to the m axis (ii) alternatively it is some section of an area bounded by two lines parallel to the σ^2 axis. Integration for m and σ^2 respectively then implies extending these sections until the whole area bounded by these two parallel lines is included in the chosen region. This existence of a complete simultaneous fiducial distribution for the two population parameters corresponding to a normal sample is a special case of the complete fiducial distribution which exists for the two parameters of location and scaling for a sample from any population of the form

$$p(x \mid m, \sigma) = f\left(\frac{x - m}{\sigma}\right)\frac{dx}{\sigma},$$

as I have previously pointed out ([2], p. 564).[3]

For let T_1 and T_2 be any two algebraically independent statistics giving information on the two parameters, such that

$$p(S \mid m, \sigma) = p(T_1, T_2 \mid C, m, \sigma)p(C)$$

where C represents the configuration of the sample (the idea of specifying the configuration C was first introduced by Fisher [7]). The above equation is always possible, for if x_1 is the smallest observation, x_2 the next smallest and so on, let

$$T_1 = x_1$$

$$T_2 = x_2 - x_1$$

$$T_r = (x_r - T_1)/T_2, \qquad (r > 2).$$

Then $C \equiv (T_r)$ is independent of m and σ, and the quasi-sufficient set T_1, T_2 will determine a simultaneous fiducial distribution for m and σ, $\left(\text{the Jacobian } J\left(\frac{\varphi_1, \varphi_2}{m, \sigma}\right), \text{ where } \varphi_1 = \frac{T_1 - m}{\sigma}, \varphi_2 = \frac{T_2}{\sigma}, \text{ is } \frac{T_2}{\sigma^3}, \text{ and is always positive}\right)$.

As further necessary conditions for $f_p(m, \sigma)$ to be complete, we have the relations

$$p(T_1, T_2 \mid C, m, \sigma) = p(T_1 - m \mid \sigma, T_2, C)p(T_2 \mid C, \sigma)$$

$$= p\left(\frac{T_1 - m}{T_2}\middle| C\right)p\left(T_1 - m \,\middle|\, C, \frac{T_1 - m}{T_2}, \sigma\right).$$

The first of these relations is obvious, and since the first factor in it corresponds to the quasi-sufficient statistic T_1 for m when the configuration $C' \equiv (C, T_2)$ is given (σ known), we have

$$f_p(m, \sigma) = f_p(m \mid \sigma)f_p(\sigma).$$

For the second relation we note that the set

$$\frac{T_1 - m}{T_2} \quad \text{and} \quad T_1 - m$$

are algebraically equivalent to the set T_1 and T_2. Moreover, $\frac{T_1 - m}{T_2}$ is inde-

[3] Cf. also Pitman [11], who does not, however, consider the point with which I am concerned in this paper.

pendent of σ; and if m is known, $(T_1 - m) \mid C''$, where $C'' \equiv \left(C, \frac{T_1 - m}{T_2}\right)$, is a quasi-sufficient statistic for σ. Hence

$$f_p(m, \sigma) = f_p(m) f_p(\sigma \mid m),$$

which is the relation required.

The theory of complete simultaneous fiducial distributions may be applied to sufficient statistics in unknown samples. In particular, a complete distribution may be shown to exist for the statistics $\bar{x}_2$ and s_2^2 in an unknown normal sample S_2, or for the statistics $\bar{x}$ and s^2 for the joint sample S of which the known sample S_1 is also a part ([3]; cf. Fisher, [8]).

6. The Behrens-Fisher Test between two means. Fisher [8] showed that by integrating out the simultaneous fiducial distribution $f_p(m, \sigma^2)$ obtained from a normal sample, we obtained either $f_p(m)$ or $f_p(\sigma^2)$. He then suggested that such integration was possible for any simultaneous fiducial distribution; and hence obtained a distribution apparently appropriate for testing the difference between two means from normal populations whose variances were unequal. Since I have shown that this integration can be justified for $f_p(m, \sigma^2)$ owing to the *complete* simultaneous nature of this distribution, it is clear that integration in any other problem is so far justified merely by analogy, and no statement as to its meaning in general has been given by Fisher.

To show more explicitly the extent to which the proposed solution is open to criticism, I examined in particular [2] the case where each estimated variance had only one degree of freedom. The Behrens-Fisher solution implies a fiducial distribution

$$f_p(\delta) = \frac{(s_1 + s_2)\, d\psi}{\pi\{(s_1 + s_2)^2 + \psi^2\}}$$

where δ is the difference in population means $m_1 - m_2$, $\psi = (m_1 - x_1) - (m_2 - x_2)$, where x_1 and x_2 are sample means with estimated variances s_1^2 and s_2^2 each based on only one degree of freedom. By direct argument, I derived a distribution of fiducial type

$$f_p(\delta) = \frac{|s_1 \pm s_2|\, d\psi}{\pi\{(s_1 \pm s_2)^2 + \psi^2\}}$$

where the sign + or − is to be decided at random. It is irrelevant to my argument whether we are justified in calling this distribution *the* fiducial distribution of δ; it is also irrelevant what distribution would ensue if the + and − signs were considered separately. It is sufficient to note that the distribution certainly provides us with an exact inference of fiducial type, as Fisher himself confirmed ([9], p. 375); and this inference clashes with the apparent inference to be drawn from the Behrens-Fisher solution. In general it is of course true that different distributions might validly lead to different inferences of fiducial type, but here the distributions are sufficiently similar mathematically for it to be possible to assert that they cannot both be correct. The direct distribution of $\psi/(s_1 + s_2)$ is in fact known to be dependent on the unknown ratio ϕ of the population variances (Fisher, [9], p. 374). While Fisher suggests that this in no way invalidates his fiducial argument, in my view if an inference is to be independent of an unknown parameter, it should in particular be independent of it if we imagine that we are being supplied with pairs of samples, for all of which the ratio ϕ has the same value.

7. **Approximate Tests.** I have shown ([2], p. 565) that while $f_p(\delta)$ in general does not appear to exist, we have

$$f_p(\delta, \phi) = f_p(\delta \mid \phi) f_p(\phi)$$

where

$$f_p(\delta \mid \phi) = C\left\{1 + \frac{\psi^2\phi}{(1+\phi)(n_1 s_1^2 + n_2 s_2^2 \phi)}\right\}^{-\frac{1}{2}(n_1+n_2+1)} \cdot d\left\{\frac{\psi}{\sqrt{(n_1 s_1^2 + n_2 s_2^2\phi)}}\sqrt{\frac{\phi}{1+\phi}}\right\}$$

where n_1 and n_2 are the degrees of freedom of s_1 and s_2, and C is a constant. For $n_1 = n_2$, the fiducial limits for δ (if ϕ were known) were shown to be insensitive to changes in ϕ, as has also been shown by Welch [15] in more detail. For $n_1 \neq n_2$ this is no longer the case. If we tried to get an approximate solution we might consider inserting $\theta = s_1^2/s_2^2$ for ϕ in the above distribution; this would be equivalent to considering the (direct) distribution of

$$T = \frac{x_1 - x_2}{\sqrt{(s_1^2 + s_2^2)}}$$

as a t-distribution with $n_1 + n_2$ degrees of freedom. This is therefore a first approximation to the true distribution of T, which has been obtained by Welch [15] to a further approximation involving ϕ.

Sometimes it is sufficient in practice if we can assign limits to the true significance level of T in any problem, as was illustrated in my own paper ([2], p. 566). A formal proof of the inequality used there is as follows.

The actual distribution of T for $n_1 = n_2 = n$, say, depends on the integral

$$I(\phi) = \int_0^\infty \frac{\lambda}{(1 + \lambda^2 T^2/n)^{n+\frac{1}{2}}} \frac{(\theta/\varphi)^{\frac{1}{2}n-1} d(\theta/\varphi)}{(1+\theta/\varphi)^n}$$

where

$$\lambda^2 = \frac{2\varphi(1+\theta)}{(1+\varphi)(\theta+\varphi)}$$

and hence the significance level of T on the integral

$$J(\phi) = \int_0^{|T|} I(\varphi)\, dT.$$

If we write

$$u = \begin{cases} \theta/\varphi, & (0 \leq \theta/\varphi \leq 1) \\ \varphi/\theta, & (1 \leq \theta/\varphi < \infty) \end{cases}$$

we obtain

$$J(\phi) = \int_0^{|T|} \int_0^1 \left\{ \frac{\lambda_1}{(1 + \lambda_1^2 T^2/n)^{n+\frac{1}{2}}} + \frac{\lambda_2}{(1 + \lambda_2^2 T^2/n)^{n+\frac{1}{2}}} \right\} \frac{u^{\frac{1}{2}n-1}}{(1+u)^n} dT,$$

that is,

$$\lambda_1^2 = \frac{2(1 + u\varphi)}{(1 + u)(1 + \varphi)}, \qquad \lambda_2^2 = \frac{2(u + \varphi)}{(1 + u)(1 + \varphi)};$$

where

$$J(\varphi) = \int_0^1 \{F(t_1) + F(t_2)\} u^{\frac{1}{2}n-1}(1 + u)^{-n} du$$

where $t_1 = \lambda_1 | T |$, $t_2 = \lambda_2 | T |$, and $F(t)$ is proportional to the probability integral of a t with n degrees of freedom. Since

$$\frac{\partial(F(t_1) + F(t_2))}{\partial\varphi} \propto \left\{ \frac{1}{\lambda_2(1 + \lambda_2^2 T^2/n)^{n+\frac{1}{2}}} - \frac{1}{\lambda_1(1 + \lambda_1^2 T^2/n)^{n+\frac{1}{2}}} \right\} \frac{(1 - u) | T |}{(1 + u)(1 + \varphi)^2}$$

this differential coefficient, from the relations

$$(1 - u)(1 - \varphi) \gtreqless 0,$$

$$1 + u\varphi \gtreqless u + \varphi,$$

$$\lambda_1 \gtreqless \lambda_2,$$

is never negative for all u and $\varphi(\varphi \leq 1)$. Hence $J(\varphi)$ is a steadily increasing function in the range (0, 1) for all values of T; or the significance level of T lies between its values for $\varphi = 0$ and $\varphi = 1$, as previously stated.

More generally, for $n_1 \neq n_2$, the effective number of degrees of freedom for T would be expected to lie between n_1 $(n_1 \leq n_2)$ and $n_1 + n_2$ (cf. Welch, [15], p. 360), though I have not succeeded in establishing this rigorously by a modification of the above proof.

REFERENCES

[1] M. S. Bartlett, *Proc. Roy. Soc.* Vol. A154 (1936), pp. 124–137.
[2] M. S. Bartlett, *Proc. Camb. Phil. Soc.* Vol. 32 (1936), pp. 560–566.
[3] M. S. Bartlett, *Proc. Roy. Soc.* Vol. A160 (1937), pp. 268–282.
[4] W. V. Behrens, *Landw. Jb.* Vol. 68 (1929), pp. 807–837.
[5] C. J. Clopper and E. S. Pearson, *Biometrika*, Vol. 26 (1934), pp. 404–413.
[6] R. A. Fisher, *Proc. Camb. Phil. Soc.* Vol. 26 (1930), pp. 528–535.
[7] R. A. Fisher, *Proc. Roy. Soc.* Vol. A144 (1934), pp. 285–307.
[8] R. A. Fisher, *Ann. Eugen.* Vol. 6 (1935), pp. 391–398.
[9] R. A. Fisher, *Ann. Eugen.* Vol. 7 (1937), pp. 370–375.

[10] J. Neyman, *Phil. Trans.* Vol. A236 (1937), pp. 333–380.
[11] E. J. Pitman, *Biometrika* Vol. 30 (1939), pp. 391–421.
[12] I. E. Segal, *Proc. Camb. Phil. Soc.* Vol. 34 (1938), pp. 41–47.
[13] D. M. Starkey, *Ann. Math. Stat.* Vol. 9 (1938), pp. 201–213.
[14] P. V. Sukhatme, *Sankhya* Vol. 4 (1938), pp. 39–48.
[15] B. L. Welch, *Biometrika* Vol. 29 (1938), pp. 350–361.
[16] S. S. Wilks, *Ann. Math. Stat.* Vol. 9 (1938), pp. 272–280.

Cambridge University,
England