International Statistical Review, 46 (1978) 147–170 Longman Group Limited/Printed in Great Britain

Fiducial Inference

J.G. Pedersen
Department of Theoretical Statistics, University of Aarhus

Table of Contents

1 Introduction

Among R.A. Fisher's many important contributions to statistics the fiducial argument has had a very limited success and is now essentially dead. The idea, however, has its perpetuators and occasionally one finds the words 'a fiducial distribution' or 'fiducial interval' even in recent papers. Several papers explain how to derive a fiducial distribution but fail to mention all of the essential requirements for a fiducial argument to be valid. This is excusable if the paper was written before 1956; for one of the requirements was stipulated for the first time in Fisher's *Statistical Methods and Scientific Inference* from 1956 and is therefore lacking even in most of Fisher's writings on fiducial probability.

This requirement, the existence of relevant subsets for the fiducial distribution, is presented in section 2.2. It entails some restrictions on the use of fiducial probability not mentioned by Fisher, and this explains some of the inconsistencies of fiducial probability mentioned in the literature.

Statistical Methods and Scientific Inference is the best and most complete account of Fisher's views on fiducial probability, but it appears not to be easy reading. It is hoped that this paper may also serve as an introduction to Fisher's writings on this subject.

2 The Fiducial Distribution of a Real Valued Parameter

2.1 *On 'Inverse Probability'*

The concept of fiducial probability was conceived by R.A. Fisher. The first paper of his on fiducial probability was 'Inverse Probability' from 1930. Inverse probability is synonymous with Bayes posterior probability, which may be obtained when the mathematical model in addition to the parametrized family of distributions of observables contains a prior distribution on the parameter space. The distribution of the observables for given values of the parameters is considered as a conditional distribution given the parameters, and the prior distribution is considered as a marginal distribution for the parameters. This leads to a joint distribution of observables and parameters. The inversion is completed forming the conditional distribution of the parameters given the observables. The Bayes posterior distributions are obtained using simply a theorem in the calculus of probability, often referred to as Bayes' Theorem, and this is not the cause of any dispute. But the prior distributions have been questioned. First, it has been held that we cannot always express our rational belief in competing hypothesis in terms of probability. Fisher (1956–73) gives an historical selection of some points of view on this issue. Secondly, supposing the prior distributions do have a meaning, they are questioned in the cases where they are chosen as uniform distributions because nothing is known about the values of the parameters. This approach is sometimes referred to as Bayes' axiom[1] or as Bayes' postulate.

Fisher (1930) makes two points in this connection. First he tries to imagine an argument which may have led Laplace and Gauss to fall into their error of 'prime theoretical importance', namely to accept that our rational belief in competing hypotheses can always be expressed in terms of probability.

'In fact the argument runs somewhat as follows: a number of useful but uncertain facts can be expressed with exactitude in terms of probability; our judgments respecting causes or hypotheses are uncertain, therefore our rational attitude towards them is expressible in terms of probability. The assumption was almost a necessary one seeing that no other mathematical apparatus existed for dealing with uncertainties.' (Fisher, 1930, pp. 528–529).

The second point made is that Bayes' axiom is incapable of consistent use; for in the case of a continuum of parameters a reparametrisation may transform a uniform distribution for one parametrisation into a very complicated one for the other.

Fisher proceeds to give two alternatives to inverse probability: the likelihood function with maximum likelihood estimation and the fiducial argument. The requirements of the fiducial argument were not fully stipulated in the 1930 paper, but the argument given there for the fiducial distribution of a one-dimensional parameter remains an essential part of the justification of a one-dimensional fiducial distribution in a situation where the requirements are met. In this section we paraphrase this argument and defer a discussion of the precise requirements to sections 2.2 and 2.3.

Let T denote a one-dimensional statistic whose distribution depends on a one-dimensional parameter, θ, only. Suppose T has a continuous distribution, and let $F(t, \theta)$ denote the probability that T is less than t for given value of θ. For $\alpha \in (0, 1)$ let $t_\alpha(\theta)$ denote the set of α percentiles of $F(., \theta)$ and $\theta_\alpha(t)$ denote the set of θ's for which t belongs to $t_\alpha(\theta)$, or, in other words, the set of θ's for which t is an α percentile of $F(., \theta)$. Suppose now, as often happens, that:

(i) for all t and all $\alpha \in (0, 1)$ $\theta_\alpha(t)$ contains one and only one θ

(ii) for all $\alpha \in (0, 1)$ $t_\alpha(\theta)$ is an increasing function of θ.

Let t_0 be an observed value of T. $\theta_\alpha(t_0)$ which exists and is unique according to (i) is said to be the fiducial $1-\alpha$ percentile for θ corresponding to the observed t_0 with the following argument. If we have a large number of observations $t_1, t_2, \ldots$ on T corresponding to the parameters $\theta_1, \theta_2, \ldots$ the inequality

$$\theta_i < \theta_\alpha(t_i) \tag{1}$$

[1] In spite of its name it is not formulated as an axiom by Bayes (1763). Bayes proved his theorem in the case of a binomial parameter which was determined by an experiment and did have a uniform distribution. In a *scholium* he argued that the uniform prior should be used for any binomial parameter 'of which we absolutely know nothing antecedently to any trials made concerning it'; for if the parameter is uniformly distributed, the probability of any number of successes in n trials is $1/(n+1)$, and that to Bayes represented the meaning of the phrase that we know nothing of the parameter before any trials are made. Edwards (1973) has drawn attention to this argument.

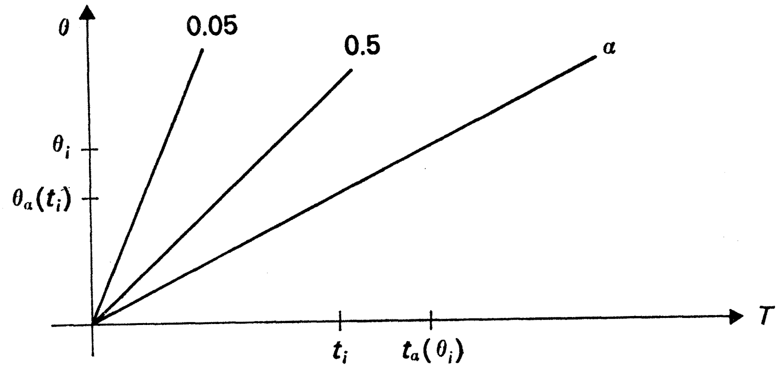

Figure 1. *Percentiles in a situation where (i) and (ii) hold. The range of T and θ is* $(0, \infty)$.

will be true with a frequency of $100\,(1-\alpha)$ per cent; for (1) is true according to (ii) (see Figure 1) if and only if

$$t_i > t_\alpha(\theta_i)$$

which is true with a frequency of $100\,(1-\alpha)$ per cent, because t_i is an observation from a random variable with distribution $F(.,\,\theta_i)$.

For an observed t the fiducial $1-\alpha$ percentiles exist for every $\alpha \in (0, 1)$. They are considered to be the percentiles of a probability distribution of θ: the fiducial distribution. The distribution function of the fiducial distribution for observed t is

$$P\{\theta' < \theta\} = 1 - F(t, \theta)$$

and the fiducial density is

$$-\frac{\partial}{\partial\theta} F(t, \theta). \tag{2}$$

If (i) holds and instead of (ii), (ii)$'$ $t_\alpha(\theta)$ is decreasing as a function of θ for every $\alpha \in (0, 1)$, the argument above may be carried out with the change that the inequality (1) is reversed such that $\theta_\alpha(t)$ is the fiducial α percentile for observed t and the fiducial distribution function is

$$F(t, \theta),$$

so that the fiducial density is

$$\frac{\partial}{\partial\theta} F(t, \theta). \tag{3}$$

We note that a distribution of θ may be obtained arguing in much the same way as above even if $\theta_\alpha(t)$ and hence $t_\alpha(\theta)$ may contain more than one point, provided $\sup_t t_\alpha(\theta)$, $\alpha \in (0, 1)$, considered as functions of θ are either increasing for all α or decreasing for all α. The density of this distribution for θ will be given by (2) or (3) according as $\sup_t t_\alpha(\theta)$, $\alpha \in (0, 1)$, are increasing or decreasing as functions of θ.

As an example Fisher considers (the fiducial 0.05 percentile of) the correlation coefficient ρ in the bivariate normal distribution. In Fisher (1915) the sampling distribution of the empirical correlation coefficient r was found, and it depends on ρ only. No expression of the fiducial distribution of ρ is given (it is given in Fisher (1956–73)). But a table for varying ρ of the 0.95 percentile of the distribution of r from a sample of 4 pairs of observations is given, and hence a table for varying r of the fiducial 0.05 percentile of the fiducial distribution of ρ based on four pairs of observations.

2.2 *Fisher's concept of probability*

The precise requirements of the fiducial argument were not stated in Fisher (1930), as already mentioned. In this section and the next the prerequisites of the fiducial argument later noted by Fisher will be discussed. We consider observations whose distribution is supposed to belong to a parametrized family of distributions, where the parameter may be a vector. Throughout θ will denote a one-dimensional component of the parameter, whose fiducial distribution is sought, and T will denote a one-dimensional statistic, whose distribution depends on θ only. Occasionally, for example in section 2.3 when discussing the role of sufficiency, we shall consider the somewhat simpler case where the one-dimensional θ parametrizes the family of distributions.

The argument used by Fisher to justify the fiducial distribution when (i) and (ii) are satisfied shows in the language of the theory of confidence intervals that $\theta_\alpha(t_i)$ is an upper $(1-\alpha)$ confidence limit. No wonder Neyman and others for some time thought that the theory of confidence intervals and fiducial probability were two names for the same thing. The difference between the two concepts is obvious today. The interpretation of a confidence interval for a parameter with confidence coefficient $1-\alpha$ is that either an event with probability less than α has occurred or the interval contains the parameter. In contrast fiducial probability uses probability in the usual sense as a measure of the uncertainty of some uncertain outcome. So even if a fiducial interval with fiducial probability $1-\alpha$ coincides with a confidence interval with confidence coefficient $1-\alpha$, their interpretations are different. Nevertheless it is important for the fiducial argument that $\theta_\alpha(t_i)$ is an upper $1-\alpha$ confidence limit for θ, for the fiducial distribution is not chosen arbitrarily.

To see the step from confidence intervals to fiducial intervals it is necessary to understand Fisher's use of the word probability. Fisher's probabilities are limiting frequencies in some aggregate of samples, which is called the reference set (Fisher, 1958a,b). This is implicit in the justification of the fiducial percentiles following (i) and (ii) in section 2.1. In his earlier writings this was the definition of probability without qualifications. For this reason the word fiducial was necessary to distinguish probabilities derived by the fiducial argument from inverse probabilities. Both were limiting frequencies but in different populations of samples. Fisher (1930) elucidates this point.

'The fiducial frequency distribution will in general be different numerically from the inverse probability distribution obtained from any particular hypothesis as to *a priori* probability.... It would be perfectly possible, for example, to find an *a priori* frequency distribution for ρ such that the inverse probability that ρ is less than 0.765 when $r = 0.99$ is not 5 but 10 in 100. [0.765 is the fiducial 0.05 percentile of ρ given $r = 0.99$, based on 4 pairs of observations.] In concrete terms of frequency this would mean that if we repeatedly selected a population at random, and from each population selected a sample of four pairs of observations, and rejected all cases in which the correlation as estimated from the sample (r) was not exactly 0.99, then of the remaining cases 10 per cent would have values of ρ less than 0.765. Whereas apart from any sampling for ρ, we know that if we take a number of samples of 4, from the same or from different populations, and for each calculate the fiducial 5 per cent value for ρ, then in 5 per cent of cases the true value of ρ will be less than the value we have found. There is thus no contradiction between the two statements. The fiducial probability is more general and, I think, more useful in practice, for in practice our samples will all give different values, and therefore both different fiducial distributions and different inverse probability distributions. Whereas, however, the fiducial values are expected to be different in every case, and our probability statements are relative to such variability, the inverse probability statement is absolute in form and really means something different for each different sample, unless the observed statistic actually happens to be exactly the same.'

The definition of probability is given explicitly by Fisher in 1956 (Fisher, 1956–73, chapter I.4). It is the limiting frequency in some reference set which has the property that no subsets with different limiting frequencies can be recognized.

This is explained in terms of the probability of the event of an ace being thrown with a single die. We assign the probability 1/6 to this event if the limiting frequency of aces in all throws with the die is 1/6 and the particular throw we are about to make cannot be shown to belong to a subset of throws with a different limiting frequency of aces.

The same definition and a different example are given two years later (Fisher, 1958a): 'Probability is, I suggest, the first example of a well specified state of logical uncertainty. Let me put down a short list of three requirements, as I think them to be, for a correct statement of probability, which I shall then hope to illustrate with particular examples. I shall use quite abstract terms in listing them.

(a) There is a measurable reference set (a well-defined set, perhaps of propositions, perhaps of events).

(b) The subject (that is, the subject of a statement of probability) belongs to the set.

(c) No relevant subset can be recognized.

I expect that these words will acquire a meaning from the examples I have to give.

'Let us consider any uncertain event. A child is going to be born. I don't know enough about the present state of medical science to know whether experts exist who are really capable of saying in advance of what sex the child will be. But let us imagine ourselves in the technology of the nineteenth century, when certainly no such statement could be made with any confidence. This is my first example of a matter in which we are in the state of uncertainty; that is to say, we lack precise knowledge, but we do not lack all knowledge. On inquiry at the registrar, we may find that in his experience, or in the experience of much larger numbers recorded by registrars in different parts of the world, a fixed proportion of the births has been of boys and the remainder of girls. Let us suppose he tells us that in 51 per cent the births are those of boys (a little more than 51 per cent in most populations). To the registrar, the birth which is about to take place, though intensely important to ourselves, is just another birth. To him it belongs to this set of his experience of sex at birth, and he very properly informs us that the probability of a boy is 51 per cent, having made reference to this measurable reference set as the basis of his statement.

'Secondly, we satisfy ourselves as to the existence of relevant subsets. I need not use the word "random" because all I need say can be said under "(c)", which is the most novel in its formulation if not in its idea, the most novel of the requirements I have listed. This is a formulation which I submit to your judgment as a competent formulation of what is needed if we are to speak without equivocation of a probability of something in the real world.

'The registrar might raise such a question as this: Is it a white birth or a coloured birth? In his experience, the sex ratio might be different. Very well, then it's a white birth. We have recognized a subset of white births, and he must turn to his tables and find out what the proportion is in respect to white births, ignoring those which do not belong to the particular subset to which our event belongs. Or again, his experience might have shown that first births have a higher male sex ratio than births in general. He will then inquire whether our birth is a first birth or not. If it is a first birth, it belongs to a relevant subset. It is now recognized and takes the place of the reference set with which we started.'

The reference set of the fiducial distribution is described on pp. 57–58[1] in Fisher (1956–73). It is the set of all pairs of values (T, θ), where T has distribution with parameter θ. The pair

[1] The page numbers refer to the 1973 edition.

of values of θ and T relevant to a particular experimenter belongs to this set, and if the class of distributions satisfies (i) and (ii) the proportion of cases satisfying

$$\theta < \theta_\alpha(T)$$

is equal to $1-\alpha$. If no subsets of the reference set can be recognized to have a different proportion satisfying the inequality, the probability $1-\alpha$ will be considered as the probability that

$$\theta < \theta_\alpha(t)$$

in any particular experiment giving the outcome t, in the same way as one sixth is considered to be the probability of obtaining an ace in one throw with a die.

The question of whether relevant subsets can be recognized is most important for the choice of reference set and hence for the fiducial distribution. An observational feature of a sample is the value of any statistic calculated from the sample. Hence the recognizable subsets to which a sample belongs are defined by statistics on the sample space. If S is a statistic, one recognizable subset to which our sample, x_0, belongs is the set of pairs, (x, θ), where x has a distribution with parameter θ and $S(x) = S(x_0)$. This subset is relevant according to Fisher's definition if the conditional distribution of T given $S = S(x_0)$ is different from the distribution of T. This will be the case of almost any statistic S, but the attitude towards the reference set defined by a particular statistic will depend on whether its distribution depends on θ or not.

When the distribution of S neither depends on θ nor any other unknown parameter the reference set defined by S will take the place of the original one. But when the distribution of S depends on θ, the information in S and its distribution will not be used if attention is restricted to the reference set defined by the observed value of S. For this reason it will not be used to derive a fiducial distribution of θ; but it may show, in some cases, that fiducial probabilities cannot be based on the original reference set either; for if for some positive ε either

$$P[T \leqq t_\alpha(\theta) \mid S = S(x_0)] \geqq \alpha+\varepsilon \tag{4}$$

or

$$P[T \leqq t_\alpha(\theta) \mid S = S(x_0)] \leqq \alpha-\varepsilon \tag{5}$$

for every θ, one cannot state that

$$\theta \leqq \theta_\alpha(T) \tag{6}$$

with fiducial probability α (or $1-\alpha$), because a subset has been recognized in which (6) is true of a fraction different from α (or $1-\alpha$). If $\{S(x) = S(x_0)\}$ satisfies (4) or (5) it is a relevant subset in the sense of Buehler (1959).

When the family of distributions is parametrized by a vector parameter it may happen that although the distribution of S does not depend on θ it depends on some unknown parameter. No general comments on this situation will be given here; an example is treated in section 3.2.

Suppose T constitutes the whole sample, i.e. T is an observation from a distribution which depends on θ only. This appears to be the simplest case as far as the existence of statistics that define relevant subsets is concerned; for only functions of T have to be considered. The examples by Robinson (1975) show that even in this case relevant subsets in the sense of Buehler may exist. It is somewhat disturbing not to be sure whether one's inferences are the best possible, but there is a little comfort to be found (Buehler, 1959) if the family of distributions of T is a location parameter family. Buehler shows that if $g(t)$ is a density which is positive for every t and the density of T is $g(t-\theta)$, no subset of positive Lebesgue measure can be relevant provided g satisfies some mild regularity conditions.

With the explicit definition of probability (Fisher, 1956–73, 1958a) there is no need to distinguish between fiducial and inverse probability in the sense that both are probabilities according to Fisher's definition and when both exist they are identical. For when an *a priori* distribution for θ is known, the aggregate of all future samples has a recognizable and relevant subset, namely those samples corresponding to the observed value of T, and consequently this reference set, which is the reference set of the inverse distribution, takes the place of the reference set of all samples. Thus absence of knowledge of θ in the form of an *a priori* distribution is an essential requirement for the fiducial argument based on the reference set of all samples. It was not until 1956 it was stated that the absence of knowledge *a priori* of a distribution of θ is an essential requirement for the fiducial argument (Fisher, 1956–73, p. 54).

The following statement of Fisher's has been taken to mean that any knowledge of θ *a priori* will invalidate the fiducial argument. Fisher (1956–73, p. 59): '... it is essential to introduce the absence of knowledge *a priori* as a distinctive datum in order to demonstrate completely the applicability of the fiducial method of reasoning ...'. But Fisher has given what he terms a fiducial distribution of a normal mean θ which is known to be restricted to an interval $[M, \infty)$ (Fisher, 1956–73, pp. 138–140).

Probability is also used to qualify a function defined on a σ-algebra of events and satisfying Kolmogorov's axioms; such functions are probability measures. In this sense it is necessary to distinguish between fiducial and inverse probability. The latter is a probability measure and the calculus of probability can be applied to it, but this appears not to be the case in general of fiducial distributions. The fiducial distribution of a real valued parameter is introduced as the distribution whose α percentiles for every $\alpha \in (0, 1)$ are the fiducial α percentiles, and the fiducial percentiles, in turn, are the upper bounds of intervals $[-\infty, \theta_\alpha(t)]$, supposing (i) and (ii)′ to hold, which can be assigned a fiducial probability α. Two things are essential for the assignment of a fiducial probability α to the interval $[-\infty, \theta_\alpha(t)]$. First, it is a confidence interval with confidence coefficient α, and secondly, no relevant subsets of samples can be recognized. The coincidence that a distribution function is used both to summarize the fiducial percentiles and to assign a probability to any interval is no convincing argument for neglecting the two essential requirements of an interval with a fiducial probability and consider the probability assigned by the fiducial distribution to an arbitrary interval to be a fiducial probability. Since some intervals are not confidence intervals and consequently cannot be assigned a fiducial probability,[1] fiducial probability is not a probability in the sense of Kolmogorov.

It is useful to consider an example. Suppose T follows a $N(\mu, 1)$ distribution with unknown mean μ. The fiducial distribution of μ based on one observation t of T is $N(t, 1)$. If u_α denotes the α percentile of the standard normal distribution and $\mu_1 = u_{\alpha_1} + t$, one may state that the fiducial probability of the interval

$$\mu < \mu_1$$

is α_1. Similarly one may assign the fiducial probability α to the interval

$$\mu_2 < \mu < \mu_1 \tag{7}$$

if it is agreed that

$$\mu_1 = u_{\alpha_1} + t,\ \mu_2 = u_{\alpha_2} + t,\ \alpha = \alpha_1 - \alpha_2,$$

α_1 and α_2 being fixed. Suppose now the interval is required to be symmetric around 0, i.e. $\mu_2 = -\mu_1$. Then μ_1 is uniquely defined by the equation

$$\alpha = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\mu_1}^{\mu_1} e^{-(x-t)^2/2}\, dx = \Phi(\mu_1 - t) - \Phi(-\mu_1 - t), \tag{8}$$

[1] Note that this discussion is confined to the case of a one-dimensional fiducial distribution derived by means of (2) or (3). In some cases intervals that are not confidence intervals are indeed assigned a fiducial probability, but this is always in the more complicated cases where no statistic whose distribution depends only on the parameter of interest can be found.

where Φ denotes the distribution function of the $N(0, 1)$ distribution; and α_1 and α_2 are defined by

$$\mu_1 = u_{\alpha_1}+t, \quad -\mu_1 = u_{\alpha_2}+t. \tag{9}$$

But α_1 and α_2 now depend on t, and the interval is not a confidence interval with confidence coefficient α. In fact, as will be proved below, if P_μ denote the $N(\mu, 1)$ distribution

$$P_\mu(-\mu_1<\mu<\mu_1) = P_\mu(u_{\alpha_1(T)}<\mu-T<u_{\alpha_2(T)})>\alpha$$

for every μ.

To see this we examine the behaviour of u_{α_1} and u_{α_2} as functions of t. It is not difficult from (8) and (9) to extract the following facts: u_{α_1} and u_{α_2} are decreasing as functions of t; $\mathrm{d}u_{\alpha_1(0)}/\mathrm{d}t = \mathrm{d}u_{\alpha_2(0)}/\mathrm{d}t = -1$; u_{α_1} is convex and u_{α_2} is concave; $u_{\alpha_1}>u_\alpha$ and $u_{\alpha_2}<u_{1-\alpha}$ for every t; u_{α_1} has asymptote $-u_{1-\alpha}-2t$ for $t\to-\infty$ and asymptote u_α for $t\to\infty$; u_{α_2} has asymptote $u_{1-\alpha}$ for $t\to-\infty$ and u_{α_2} has asymptote $-u_\alpha-2t$ for $t\to\infty$. These facts are used to draw Figure 2, which depicts the case $\alpha<0.5$.

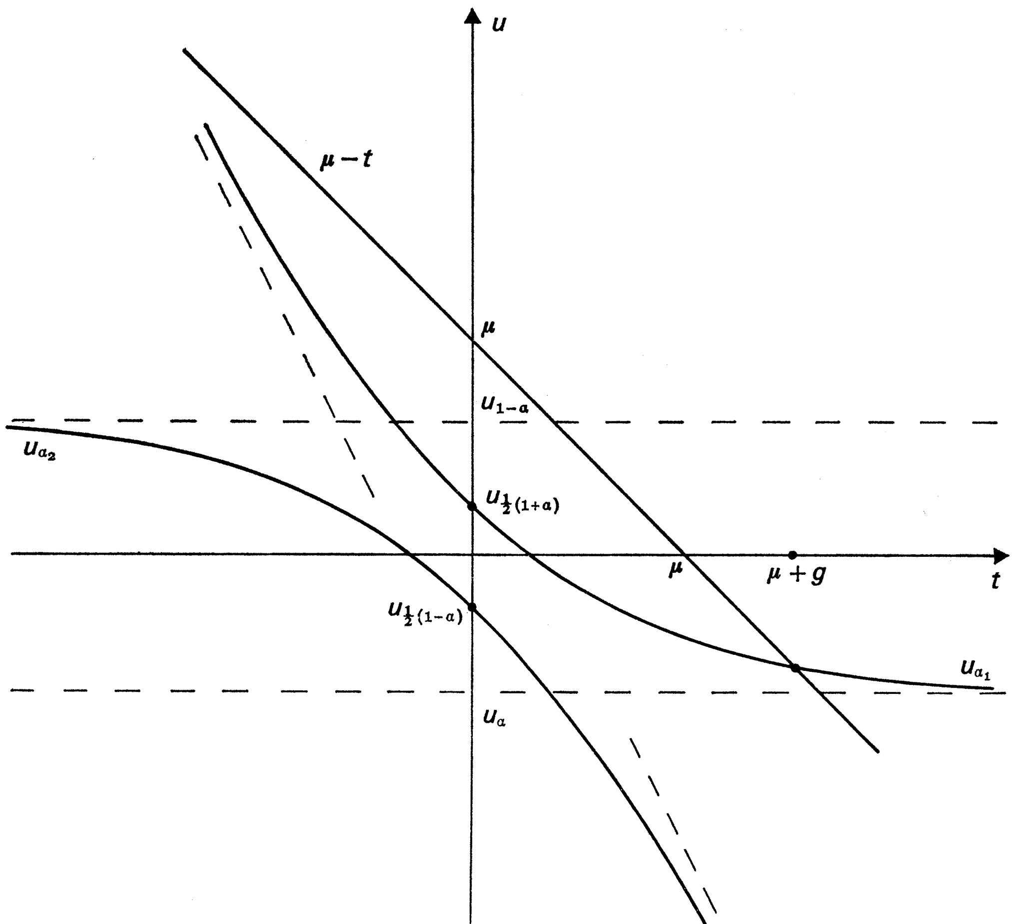

Figure 2. *u_{α_1} and u_{α_2} as functions of t in the case $\alpha<0.5$.*

If $u_{(1-\alpha)/2}<\mu<u_{(1+\alpha)/2}$, it follows that

$$P_\mu(u_{\alpha_2(T)}<\mu-T<u_{\alpha_1(T)}) = 1.$$

Consider next $\mu > u_{(1+\alpha)/2}$. If $\mu + g$ denotes the greater of the two values that solve $\mu - t = u_{\alpha_1(t)}$, then $t > \mu + g$ implies that

$$u_{\alpha_2(t)} < \mu - t < u_{\alpha_1(t)}.$$

Now $g < u_{1-\alpha}$ (see Figure 2), and it follows that

$$P_\mu\,(u_{\alpha_2(T)} < \mu - T < u_{\alpha_1(T)}) > P_\mu\,(T > \mu + g) > P_\mu\,(T - \mu > \mu_{1-\alpha}) = \alpha.$$

A similar argument applies if $\mu < u_{(1-\alpha)/2}$; and the proof is complete.

If $\mu_2 = -\mu_1$, (7) is equivalent to

$$\mu^2 < \mu_1{}^2.$$

It follows that the fiducial distribution of μ cannot be used to assign a fiducial distribution to μ^2. This has been stated earlier in unpublished papers by Wilkinson and again by Wilkinson (1977). Wilkinson's argument is based on an example by Stein (1959).

Stein considered n observations $t_1, \ldots, t_n$ from independent random variables $T_1, \ldots, T_n$, where t_i is an observation of T_i which follows a $N(\mu_i, 1)$ distribution. Assuming that the calculus of probability applies to fiducial distributions it follows that $\sum \mu_i{}^2$ has a non-central χ^2 distribution with n degrees of freedom and non-centrality parameter $\sum t_i{}^2$. Stein then showed that if $x_{\alpha,\,n}(\Sigma t^2)$ is the $(1-\alpha)$ percentile of this distribution and

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \mu_i{}^2 = 0$$

then, with obvious notation,

$$\lim_{n\to\infty} P_{\mu_1, \ldots, \mu_n}\{\sum \mu_i{}^2 \geqq x_{\alpha,\,n}(\sum T_i{}^2)\} = 0.$$

The example above shows that it is neither necessary to consider large samples nor to impose conditions on the parameters to see that the fiducial distribution of μ cannot be used to assign a fiducial distribution to μ^2.

It appears from the example of the normal mean that the most general result that can be obtained about the transformation of fiducial distributions is: A fiducial distribution of a monotone transformation of θ may be obtained transforming the fiducial distribution of θ.

It is not quite clear what Fisher's opinion on this issue was. Fisher has nowhere stated that the calculus of probability cannot be applied to fiducial probabilities, but this is, as was seen above, a consequence of the requirement (Fisher, 1956–73, p. 62): 'Probabilities obtained by a fiducial argument are objectively verifiable in exactly the same sense and exactly the same way as are the probabilities assigned in games of chance.'

There is some confusion, however, as to whether Fisher thought that the calculus of probability applied to fiducial distributions. The confusion stems partly from the fact that Fisher used fiducial distributions as prior distributions (Fisher, 1956–73, pp. 127–132; Fisher, 1962) and partly from some very general statements Fisher has made concerning the use of fiducial distributions of more than one parameter. These statements will be discussed in section 3.1.

2.3 *Sufficiency and Ancillarity and the Fiducial Argument*

In his 1930 paper Fisher had without further comments required T to be a maximum likelihood estimate of θ. In 1934 Neyman published what he considered to be an extension of the fiducial argument in that he allowed himself to use any statistic whose distribution depended on θ only. In the discussion to Neyman (1934) Fisher commented that: '. . . he would apply the fiducial argument, or rather would claim unique validity[1] for its results, only in those cases for which the problem of estimation proper had been completely solved, i.e. either when there existed a statistic of the kind called sufficient, which in itself contained the whole of the information supplied by the data, or when, though there was no sufficient statistic, yet the whole of the information could be utilized in the form of ancillary information. Both these cases were fortunately of common occurrence, but the limitation seemed to be a necessary one, if they were to avoid drawing from the same body of data statements of fiducial probability which were in apparent contradiction.'

We may note in passing that in the same contribution to the discussion of Neyman's paper, Fisher, prompted by Neyman's use of discrete statistics to derive a distribution of θ, for the first and only time gave reasons for requiring T to be continuous.

Let us consider the requirement that the statistic on which the derivation of the fiducial distribution is based must be sufficient. The mathematical delimitation usually given the word is as follows. Consider the distribution of a sample of size n. A statistic T of the sample is *sufficient* for θ if the conditional distribution of the sample given T does not depend on θ. If furthermore no non-trivial function of T is sufficient for θ, T is called *minimal sufficient*.

In cases where no one-dimensional sufficient statistic exists, it may happen that the minimal sufficient statistic is in one-to-one correspondence with (T, U), where T is one-dimensional and U is an ancillary statistic, i.e. has a distribution that does not depend on θ. T is then said to be *exhaustive* or *conditionally sufficient*.

If T is an arbitrary statistic of the sample the density function may be factorized into the marginal density of T and the conditional density of the sample given T, i.e.

$$\prod_{i=1}^{n} f(x_i, \theta) = g(T, \theta) h(x_1, \ldots, x_n, \theta \mid T).$$

If T is sufficient, h does not depend on θ, and it is obvious in an intuitive sense, that no information is lost restricting attention to the observation, T, and the marginal distribution of T. Indeed, no information is lost in the sense that the Fisher information calculated from the whole sample

$$i(\theta) = nE_\theta\left(\frac{\partial}{\partial\theta}\log f(x, \theta)\right)^2$$

is equal to the Fisher information in one observation of T

$$i_T(\theta) = E_\theta\left(\frac{\partial}{\partial\theta} g(T, \theta)\right)^2.$$

If T is not sufficient $i_T(\theta)$ will be less than $i(\theta)$, but if T is exhaustive, the whole information may be utilized in the form of the ancillary U in the sense that

$$i(\theta) = E_{\theta U}[i_T(\theta \mid U)]$$

where $i_T(\theta \mid U)$ is the information in one observation from the conditional distribution of T given U and $E_{\theta U}$ is the expectation with respect to the marginal distribution of U. These facts lie behind Fisher's statements in the discussion to Neyman (1934) quoted above.

If the distribution sampled depends on a vector parameter a one-dimensional component of which is of interest, the concept of sufficiency defined above does not supply any guidance

[1] 'Naturally, no rigorously demonstrable statements, such as these are, can fail to be true. They can, however, only convey the truth to those who apprehend their exact meaning; in the case of fiducial statements based on inefficient estimates this meaning must include a specification of the process of estimation employed. But this process is known to omit, or suppress, part of the information supplied by the sample. The statements based on inefficient estimates are true, therefore, so long as they are understood not to be the whole truth. Statements based on sufficient estimates are free from this drawback, and may claim a unique validity.'

as to whether or not to base a fiducial argument on a particular statistic. This is the case of the correlation coefficient from a bivariate normal distribution, which was used as example by Fisher (1930), but the example was still thought to be appropriate (Fisher, 1956–73, p. 57).

This has led Barnard (1963) to introduce a formal definition of sufficiency of a statistic for one component of a parameter in the absence of knowledge of the remaining components. r is sufficient for ρ in the absence of knowledge of the means and variances in this sense.

In the case of a univariate normal sample, the sample standard deviation is sufficient for the standard deviation in the absence of knowledge of the mean in the sense of Barnard (1963).

Sufficiency or exhaustiveness are also related to the fundamental question: Can relevant subsets be recognized? In the simple case where the order statistic of the sample is minimal sufficient and T is exhaustive, the ancillary U defines a recognizable subset. It is also relevant; for if not T would be sufficient. Consequently the conditional distribution of T given U should be used to construct the fiducial distribution. If T is sufficient it does not always happen that an ancillary U exists, such that (T, U) is in one–one correspondence with the order statistic of the sample. If V is independent of T, V is ancillary (Basu, 1955, 1958), and if V is an ancillary statistic and T is boundedly complete, T and V are independent (Basu, 1955, 1958). Thus if T is boundedly complete ancillary statistics do not define relevant subsets, and whether T is boundedly complete or not any statistic that is not ancillary does define a relevant subset. But when the statistics are not ancillary their distribution depends on the unknown parameter, and so the reference sets they define become useless as reference sets for probability statements about the parameter. If one insists on deriving a distribution of θ the only possibility is to base it on the distribution of T. If on the other hand such action is considered to be illegitimate, a fiducial distribution can only be derived in those cases where the data may be transformed by a one-to-one transformation into the sufficient or exhaustive T and an ancillary U. It is still an open question exactly when this is possible. If the parameter is k-dimensional and the sample is n-dimensional an $n-k$ dimensional ancillary statistic exists if the distribution sampled is a transformation parameter model. Those models are the ones amenable to structural inference (Fraser, 1961a, b, 1968). It seems a plausible conjecture that an $n-k$ dimensional ancillary exists only in those cases.

The problem whether or not ancillaries exist has come to be known as the Problem of the Nile since it was described by Fisher (1936) in this form: 'The agricultural land of a pre-dynastic Egyptian village is of unequal fertility. Given the height to which the Nile will rise, the fertility of every portion of it is known with exactitude, but the height of the flood affects different parts of the territory unequally. It is required to divide the area, among the several households of the village, so that the yields of the lots assigned to each shall be in pre-determined proportions, whatever may be the height to which the river rises.'

Consider as an example pairs of observations (x, y) from the distribution

$$df = e^{-(x\theta + y/\theta)}\, dx\, dy,$$

where x and y are positive. Fisher (1956–73) notes that the product xy has a distribution independent of the parameter and therefore is a 'solution of the Nile problem in the sense that the total frequency lying between any two rectangular hyperbolas

$$xy = c_1, \qquad xy = c_2,$$

shall be independent of θ, and depend only on the chosen values c_1 and c_2. Such curves therefore divide the total frequency in fixed proportions independently of the value of the unknown parameter, representing in that case the unknown height to which the Nile will rise.'

3 Pivots and Multivariate Fiducial Distributions

3.1 *Pivots*

The argument in section 2 leading to the fiducial distribution may be formulated in a slightly different way. When T has a distribution with parameter θ, $F(T, \theta)$ is distributed uniformly on $(0, 1)$. After the observation t_0 of T, $F(t_0, \theta)$ is still considered to be uniformly distributed. If in addition $F(t_0, .)$ is well-behaved, the uniform distribution may be transformed into a distribution on the parameter space. To be specific it is required that:

(i) the range of $F(t, .)$ is the same for any t, and

(ii) $F(t, .)$ is one-to-one, and hence $F^{-1}(t, .)$ exists, and

(iii) $F(t, .)$ has continuous derivative,

in which case the transformed distribution has density

$$\left|\frac{\partial}{\partial\theta} F(t_0, \theta)\right|.$$

(i) and (ii) are equivalent to (i) and (ii) or (ii)′ of section 2.1; and the only function of (iii) is to ensure the existence of the transformed density. This argument is a different formulation only of the one given in section 2.1 and the additional requirements are the same. T must contain all the relevant information of θ in the data, and no prior distribution of θ must be known. The latter requirement is related to the crucial step in the argument above that $F(T, \theta)$ is considered to be uniformly distributed with t_0 substituted for T. The reference set in probability statements concerning $F(T, \theta)$ is the set of all (T, θ) where θ is arbitrary and T follows a distribution with parameter θ. For those statements to be relevant to particular outcomes such as the ones with $T = t_0$ no recognizable subsets must be relevant, which would be the case if a prior distribution of θ was known.

In this form the argument is easily generalized to give distributions of multidimensional parameters given data. All that is needed is a function of the parameters and the data that can play the role of the distribution function in the argument. These functions are pivotal quantities. A quantity is a possibly multidimensional function of the parameters and the data defined for all admissible values of the parameters and the data. A quantity is *pivotal* or a *pivot* if it has a fixed distribution when the values of the parameters in the quantity are at the same time the values determining the distribution of the data appearing in the quantity. A quantity is *sufficient* if, when any arbitrary fixed values of the parameters are inserted as arguments, the resulting statistic is sufficient. Any function of a pivot is a pivot. As an attempt to make sure that no information is lost by a bad choice of pivot one might require that only sufficient pivots are used. This requirement, however, is very restrictive and occasionally a pivot that is not sufficient will be used. The generalization of the fiducial argument is as follows. If $p(x, \theta)$ is a pivot and satisfies (i), (ii), and (iii), or, if multidimensional and of the same dimension as θ, satisfies (i), (ii), and

(iii)′ $p(x, .)$ has continuous partial derivatives with Jacobian different from 0,

the fixed distribution of the pivot may be transformed to give a fiducial distribution for θ. The same argument may be used to find from a given sample the fiducial distribution of a future observation from the same population as the sample in hand. The future observations take the place of the parameters in the argument. This was illustrated in Fisher (1935b).

Examples 1, 2, and 3 contain pivots used by Fisher.

Example 1. The fiducial distribution of the standard deviation, σ, from n observations from a normal population with unknown mean.

The sample standard deviation

$$s = \left[\frac{1}{n-1}\sum(x_i-\bar{x})^2\right]^{\frac{1}{2}}$$

is sufficient for σ in the absence of knowledge of the mean in the sense of Barnard (1963). $(n-1)^{\frac{1}{2}}s/\sigma$ is χ distributed on $n-1$ degrees of freedom, and hence

$$\frac{s}{\sigma} \sim \frac{1}{(n-1)^{\frac{1}{2}}}\chi(n-1).$$

No recognizable subsets are relevant; for s is independent of $\bar{x}$ and the configuration $C = [(x_1-\bar{x})/s, \ldots, (x_n-\bar{x})/s]$. The result of the inversion of the distribution of the pivot may be given formally as

$$\sigma \sim (n-1)^{\frac{1}{2}}s\chi(n-1)^{-1},$$

where $\chi(n-1)$ is a random variable with a $\chi(n-1)$ distribution. This fiducial distribution was given in Fisher (1933).

Example 2. The fiducial distribution of the mean, μ, of a normal population based on n observations from a normal population with unknown mean and variance.

No statistic of a sample of size n from a normal population has a distribution that depends on μ only. The fiducial distribution of μ was obtained in Fisher (1935b) using the t-statistic

$$t = \frac{n^{\frac{1}{2}}}{s}(\bar{x}-\mu).$$

t has a t-distribution with $n-1$ degrees of freedom; it is a pivot. It will be noticed that t fulfils the requirements (i), (ii), and (iii) for pivots given in section 3.1, but that t is not a sufficient pivot.

Fisher inverts the pivot to obtain the distribution

$$\mu \sim \bar{x} + \frac{s}{n^{\frac{1}{2}}}t,$$

i.e. the fiducial distribution of μ is a t-distribution with $n-1$ degrees of freedom relocated to have median $\bar{x}$ and rescaled by the factor $s/n^{\frac{1}{2}}$.

As derived here the reference set of the fiducial distribution of μ is the set of all samples of size n. The same fiducial distribution of μ, however, has a hypothetical reference set which will be considered in section 3.2.

Example 3. The joint fiducial distribution of the mean and the standard deviation from a normal population based on n observations.

In Fisher (1941) it was noted that both the pivot

$$\chi_1 = \frac{n^{\frac{1}{2}}}{\sigma}(\bar{x}-\mu),$$

$$\chi_{n-1}{}^2 = \frac{(n-1)s^2}{\sigma^2}$$

and the pivot

$$t = \frac{n^{\frac{1}{2}}}{s}(\bar{x}-\mu)$$

$$\chi_n{}^2 = \frac{1}{\sigma^2}[(n-1)s^2+n(\bar{x}-\mu)^2]$$

may be inverted to give the same distribution for (μ, σ). The same is true of the pivots $(t, \chi_{n-1}{}^2)$ and $(\chi_1, \chi_n{}^2)$. This distribution is called the fiducial distribution of (μ, σ). It was obtained by a different argument in Fisher (1935b).

Considering the very few operations that can be meaningfully applied to one-dimensional fiducial distributions it is likely that multidimensional fiducial distributions will be of very little use. Fisher, however, has expressed great confidence in some multivariate fiducial distributions. This is the case of the joint fiducial distribution of μ and σ. In Fisher (1956–73) the marginal distributions of μ and σ^2 are found by integrating the joint fiducial distribution and on p. 124 Fisher states: 'The fiducial distribution of any chosen function of μ and σ^2 can equally be obtained.' Taken literally this is admittedly in conflict with the conclusion arrived at in section 2.2 that the only transformations of one dimensional parameters whose fiducial distribution can be found by transforming the fiducial distribution are the monotone ones; for the chosen function of μ and σ^2 might be μ^2. It is perhaps worth mentioning that the only function of μ and σ^2 Fisher in fact considers is $\mu+\alpha\sigma$ for a fixed α. Fisher does not find its fiducial distribution, but had he done so integrating the joint fiducial distribution of μ and σ^2 it would have been justified on other grounds to be given in Example 4. This relates to an important aspect of joint fiducial distributions to be considered now.

In section 2 where only the derivation of one dimensional fiducial distributions by the differentiation of the distribution function was considered it was argued that only intervals possessing the confidence property, i.e. confidence sets, can be assigned a fiducial probability.

The confidence property is conveniently formulated in terms of pivots as an invariance property. If P is the distribution of the pivot $p(x;\theta)$ and $p_x(.) = p(x;\,.)$ a set Θ_x is a confidence set with confidence coefficient α if

$$P[p_x(\Theta_x)] = \alpha.$$

To adopt as a general principle that only intervals with the confidence property can be assigned a fiducial probability certainly conflicts with Fisher's use of fiducial distributions. Section 3.3, which treats Behrens' problem, contains the most prominent example of fiducial intervals that are not confidence intervals. The fiducial distribution in section 3.3 was originally obtained as a marginal distribution from a joint fiducial distribution. The fact that a fiducial interval need not have the confidence property does not imply that fiducial probabilities need not be objectively verifiable; but it means that the reference set of a joint fiducial distribution need not be the appropriate reference set of the fiducial distributions of its components. The joint fiducial distribution does not convey the appropriate reference sets of its marginal fiducial distributions, which must be found in other ways, for example by a consideration of the pivots that generated the joint fiducial distribution. This appears to be a limitation of the utility of the joint fiducial distributions.

Example 4. The fiducial distribution of $\mu+\alpha\sigma$ for fixed α based on a sample of size n from a normal population.

The fiducial distribution of (μ, σ) may be obtained from the joint distribution of the pivots t and χ^2_{n-1} defined in Example 3. The function

$$p(\bar{x}, s; \alpha) = n^{\frac{1}{2}}\frac{\bar{x}-\mu}{s} - n^{\frac{1}{2}}\alpha\frac{\sigma}{s} = n^{\frac{1}{2}}\frac{\bar{x}-(\mu+\alpha\sigma)}{s}$$

is a function of t and $\chi_{n-1}{}^2$ and is thus a pivot. It is not difficult to see that the same fiducial distribution of $\mu+\alpha\sigma$ is obtained whether one finds the marginal distribution from the joint fiducial distribution of (μ, σ) or inverts the distribution of the pivot to a distribution of $\mu+\alpha\sigma$. By the latter method the reference set is the reference set of the distribution of p which is also the reference set of the joint fiducial distribution of (μ, σ), namely the set of all $(\bar{x}, s)$

calculated from normal samples of size n. Since the marginal fiducial distribution of $\mu+\alpha\sigma$ has the same reference set as the joint fiducial distribution it appears to be appropriate to obtain it as a marginal distribution.

The use of different pivots to generate distributions of two or more parameters in a given problem may lead to different distributions of the parameters. The first example to appear in the literature was given by Mauldon (1955) and was concerned with the positive upper triangular matrix of a multivariate normal distribution. Fisher (1956–73) alludes to this example on page 178, for the pivot in equation (231) is one of Mauldon's pivots. Another example based on two observations from independent normal distributions with unknown means and variances unity was given by Tukey (1957).

This lack of uniqueness is neither surprising nor disturbing. In the case of a real parameter it was seen that one fiducial distribution could not be used to derive fiducial distribution of functions of the parameters. With this in mind there is no reason to expect a unique fiducial distribution to exist in the multiparameter case. The important property of a pivot, p, is the class of confidence sets it generates, i.e. the class

$$\mathscr{C} = \{p_x^{-1}(B) \mid B \text{ measurable set in the range of } p\}.$$

It is this class of confidence regions together with the aspects of the parameters of current interest, which decides whether a particular pivot is at all useful.

3.2 *The Step-by-Step Argument*

Fisher appears to have been concerned about the lack of uniqueness revealed in the use of pivots; for he introduced the 'building up the simultaneous distribution rigorously by a step by step process' (Fisher, 1956–73, p. 179), and having derived the joint distribution of the parameters of the bivariate normal distribution he commented that the short-cut suggested by the use of pivots 'has no claim to validity unless it can be shown to be equivalent to a genuine fiducial argument' Fisher (1956–73, p. 179). The usual interpretation of this is as follows. Suppose $T_1, \ldots, T_k$ contain the relevant information on $\theta_1, \ldots, \theta_k$. The joint density of $T_1, \ldots, T_k$ factorizes into the density of T_1 and the conditional densities of T_i given $T_1, \ldots, T_{i-1}$ for $i = 2, \ldots, k$, and if the distribution of T_1 depends on θ_1 only and the conditional distribution of T_i given $T_1, \ldots, T_{i-1}$ depends on $\theta_1, \ldots, \theta_i$ only, the factorization may be written

$$f(t_1, \ldots, t_k; \theta_1, \ldots, \theta_k) = f(t_1; \theta_1) f(t_2; \theta_2, \theta_1 \mid t_1) \ldots f(t_k; \theta_k, \theta_1, \ldots, \theta_{k-1} \mid t_1, \ldots, t_{k-1}). \quad (10)$$

If the distribution of T_1 and the distribution of T_i given $T_1, \ldots, T_{i-1}$ for $\theta_1, \ldots, \theta_{i-1}$ fixed satisfy (i) and (ii) or (ii)$'$ of section 2.1 a fiducial distribution of θ_1 and fiducial distributions of θ_i for $\theta_1, \ldots, \theta_{i-1}$ fixed may be obtained for $i = 2, \ldots, k$. The one-dimensional fiducial distributions are considered to be marginal and conditional distributions of a joint fiducial distribution based on $T_1, \ldots, T_k$. The step-by-step argument will be considered in some detail in the simplest case where the real parameters θ_1 and θ_2 parametrize a bivariate density of T_1 and T_2 that factorizes as

$$f(t_1, t_2; \theta_1, \theta_2) = f(t_1; \theta) f(t_2; \theta_1, \theta_2 \mid t_1), \quad (11)$$

where $f(t_1; \theta_1)$ is the marginal density of T_1.

In this discussion we take for granted that no relevant subsets can be recognized, in particular we suppose that T_2 does not define a relevant subset for θ_1. We return to this question later when discussing an example by Dempster (1963a).

The investigations in this simple case will reveal two things; first, the marginal distribution of θ_2 can not always be considered to be a fiducial distribution, and secondly, the step-by-step argument is not successful in ensuring that a unique fiducial distribution is derived.

If the fiducial density of θ_1 based on the marginal distribution of T_1 and the observed value t_1 is denoted by $p(\theta_1; t_1)$ and the fiducial density of θ_2 based on the conditional distribution of T_2 given $T_1 = t_1$ and the observed value t_2 is denoted by $p(\theta_2; t_2 \mid \theta_1; t_1)$, the joint fiducial density of θ_1 and θ_2 is

$$p(\theta_1, \theta_2; t_1, t_2) = p(\theta_1; t_1)p(\theta_2; t_2 \mid \theta_1; t_1). \tag{12}$$

The corresponding cumulative distribution functions will be denoted by

$$P(\theta_1; t_1),\ P(\theta_2; t_2 \mid \theta_1; t_1),\ \text{and}\ P(\theta_1, \theta_2; t_1, t_2).$$

It is far from obvious what the joint fiducial distribution of θ_1 and θ_2 may be used for, and therefore we seek guidance in Fisher. The only bivariate fiducial distribution Fisher has considered is the one of μ and σ mentioned in Example 3.3. It may also be derived by a step-by-step argument because the joint distribution of $\bar{x}$ and s factorizes as (11) with $\sigma = \theta_1$ and $\mu = \theta_2$. Fisher shows how the fiducial distributions of μ and σ may be obtained as the marginal distribution of the joint distribution of μ and σ. This seems to be a desirable property of a joint fiducial distribution, and it is of interest to see whether it is true in general.

Obviously the marginal distribution θ_1 from (12) will be the fiducial distribution of θ_1 based on the marginal distribution of T_1; but it is not obvious that the marginal distribution of θ_2 from (12), whose density and cumulative distribution function will be denoted by $\bar{p}(\theta_2; t_2 \mid t_1)$ and $\bar{P}(\theta_2; t_2 \mid t_1)$, respectively, is a fiducial distribution; i.e. it is not obvious that there exists a possibly hypothetical reference set where the probability statements based on $\bar{P}(\theta_2; t_2 \mid t_1)$ are objectively verifiable. In the case of the fiducial distribution of μ a hypothetical reference set exists. It was given by Yates (1939) and fully endorsed by Fisher (1939b). This reference set is the set of all samples giving the observed value of s, and σ within this set of samples varies according to the fiducial distribution given the observed value of s. Inspired by this one may try to see whether the set of samples with a fixed value t_1 of T_1 and θ_1 having the fiducial distribution given t_1 within this set of samples, will be a reference set for $\bar{P}(\theta_2; t_2 \mid t_1)$, i.e. if $\bar{\theta}_{2\alpha}(t_2 \mid t_1)$ denotes the α percentile of $\bar{P}(\theta_2; t_2 \mid t_1)$, we want to see whether the inequality

$$\theta_2 \leqq \bar{\theta}_{2\alpha}(T_2 \mid t_1)$$

will be true with frequency α in this hypothetical set of samples. This set of samples will be referred to as the hypothetical reference set; it is similar to the reference set of the fiducial distribution of the difference of the means of two normal distributions with possibly unequal variances considered in Fisher (1939a, 1961a).

It turns out that we can prove this only in those cases where the percentiles $\theta_{2\alpha}(t_2 \mid \theta_1; t_1)$ of the distribution $P(\theta_2; t_2 \mid \theta_1; t_1)$ have the property that for every $\alpha \in (0, 1)$ and every θ_1 there exists a $p(\theta_1) \in (0, 1)$ such that

$$\theta_{2p(\theta_1)}(. \mid \theta_1; t_1) = \bar{\theta}_{2\alpha}(. \mid t_1). \tag{13}$$

To see this, one notes that

$$\begin{aligned}\alpha &= \bar{P}[\bar{\theta}_{2\alpha}(t_2 \mid t_1); t_2 \mid t_1] \\ &= \int_{-\infty}^{\infty} P[\bar{\theta}_{2\alpha}(t_2 \mid t_1); t_2 \mid \theta_1; t_1]p(\theta_1; t_1)\, \mathrm{d}\theta_1.\end{aligned} \tag{14}$$

Now $P[\bar{\theta}_{2\alpha}(t_2 \mid t_1); t_2 \mid \theta_1; t_1]$ is a fiducial probability which has a reference set; in fact, if

$$p(\theta_1) = P[\bar{\theta}_{2\alpha}(t_2 \mid t_1); T_2 \mid \theta_1; t_1] \tag{15}$$

the inequality

$$\theta_2 \leqq \theta_{2p(\theta_1)}(T_2 \mid \theta_1; t_1) \tag{16}$$

will be true with frequency $p(\theta_1)$ in the set of samples where T_1 has the value t_1 originally observed and θ_1 is fixed. But when (13) holds, (16) is equivalent to

$$\theta_2 \leqq \bar{\theta}_{2\alpha}(T_2 \mid t_1), \tag{17}$$

and (14) and (15) then show that if θ_1 has its fiducial distribution given t_1, (17) will indeed be true with frequency α.

It appears from the proof that there is no reason to believe $\bar{P}(\theta_2; t_2 \mid t_1)$ to have the hypothetical reference set unless (13) holds. It may of course happen that another hypothetical reference set exists in which θ_1 has a distribution different from the fiducial distribution given t_1, but such a reference set is arbitrary and in conflict with the fiducial distribution of θ_1 given t_1.

Thus when (13) does not hold it is very unlikely that $\bar{P}$ can be considered to be a fiducial distribution, and in those cases the interpretation, if any can be given, of the joint fiducial distribution of θ_1 and θ_2 must be different from the interpretation when (13) holds.

It is of interest to be able to see from the joint density of T_1 and T_2 whether (13) will hold. It is not difficult to prove that a sufficient condition for (13) to hold is that there exist functions g and h such that

$$f(t_1, t_2; \theta_1, \theta_2) = g\,[t_1, h(t_1, t_2; \theta_2); \theta_1]\frac{\partial}{\partial t_2} h(t_1, t_2; \theta_2). \tag{18}$$

This is for example the case of the joint distribution of $\bar{x}$ and s with $h(s, \bar{x}; \mu) = \bar{x} - \mu$, which proves that the marginal distribution of μ of the joint distribution of μ and σ has the hypothetical reference set.

As mentioned previously the step-by-step argument was proposed by Fisher after several examples had shown that the use of pivots to generate simultaneous fiducial distributions might lead to inconsistencies. After explaining the step-by-step argument in the case of μ and σ Fisher commented (Fisher, 1956–73, pp. 123–124): 'Several writers have adduced instances in which, when the formal requirements of the fiducial argument are ignored, the results of the projection of frequency elements using artificially constructed pivotal quantities may be inconsistent. When the fiducial argument itself is applicable, there can be no such inconsistency.'

Remarks such as this provoked Dempster (1963a) to give an example which showed that the step-by-step argument could be applied to two parametrizations of the normal distribution and that the resulting distributions were inconsistent. Dempster considered the distribution of $\bar{x}/s$ and s, which is parametrized by μ/σ and σ, and using the factorization

$$f\left(\frac{\bar{x}}{s}, s; \frac{\mu}{\sigma}, \sigma\right) = f\left(\frac{\bar{x}}{s}; \frac{\mu}{\sigma}\right) f\left(s; \frac{\mu}{\sigma}, \sigma \,\middle|\, \frac{\bar{x}}{s}\right), \tag{19}$$

where $f(\bar{x}/s; \mu/\sigma)$ is the marginal density of $\bar{x}/s$, he derived a joint distribution of μ/σ and σ. Comparing it to the distribution of μ/σ and σ obtained from the joint distribution of μ and σ based on the step-by-step argument using the joint distribution of $\bar{x}$ and s, Dempster found that the same distribution was assigned to μ/σ in the two cases whereas the conditional distributions of σ given μ/σ were different.

Bearing in mind that only monotone transformations of one dimensional fiducial distribution are fiducial distributions, this is not very surprising and there is no *a priori* reasons to believe the transformed distribution of μ and σ to be a fiducial distribution. It is much more interesting that the joint distribution of $\bar{x}/s$ and s may be factorized,

$$f\left(\frac{\bar{x}}{s}, s; \frac{\mu}{\sigma}, \sigma\right) = g(s; \sigma) g\left(\frac{\bar{x}}{s}; \frac{\mu}{\sigma}, \sigma \mid s\right), \tag{20}$$

where $g(s;\sigma)$ is the marginal density of s, and that the step-by-step argument based on this factorization gives a joint distribution of μ/σ and σ which is different from the one obtained by the step-by-step argument based on (19). In fact the distribution of μ/σ and σ based on (20) is the transformed distribution of μ and σ. This is seen as follows. The conditional distribution of $\bar{x}/s$ given s is $N(\mu/s, \sigma^2/ns^2)$, and hence the distribution of $(s/\sigma).(\bar{x}/s)$ given s and for fixed σ is $N(\mu/\sigma, 1/n)$; it follows that the fiducial distribution of μ/σ given σ, s, and the observed $\bar{x}/s$ is $N[(s/\sigma)(\bar{x}/s), 1/n]$. Denoting the fiducial density of σ based on the marginal distribution of s by $p(\sigma;s)$, the fiducial density of μ/σ and σ based on (20) is

$$\frac{n^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}}e^{-(n/2)[(\mu/\sigma)-(s/\sigma)(\bar{x}/s)]^2}p(\sigma;s). \tag{21}$$

Similarly the fiducial density of μ and σ is seen to be

$$n^{\frac{1}{2}}e^{-n(\mu-\bar{x})^2/(2\sigma^2)}p(\sigma;s); \tag{22}$$

and transforming (22) to give a density of μ/σ and σ obviously gives (21). It follows from Dempster (1963a) that the same distribution of μ/σ is obtained whether (19) or (20) is used but the conditional distributions of σ given μ/σ are different.

It is not possible to proclaim one of the two distributions as the genuine fiducial distribution as will be seen now.

First the reference sets of marginal distributions will be considered. The joint density of $\bar{x}/s$ and s is

$$f\left(\frac{\bar{x}}{s}, s; \frac{\mu}{\sigma}, \sigma\right) = K\left(\frac{s}{\sigma}\right)^{n-1}\frac{1}{\sigma}e^{-[s^2(n-1)/2\sigma^2]-(n/2)[(\bar{x}/s)(s/\sigma)-(\mu/\sigma)]^2}, \tag{23}$$

i.e. (18) holds with $t_1 = \bar{x}/s$, $t_2 = s$, $\theta_2 = \sigma$, and $h(t_1, t_2; \theta_2) = t_2/\theta_2$ which means that the marginal distribution of σ from the distribution based on the factorization (19) will have the hypothetical reference set. It is seen from (23) that (18) does not hold with $t_1 = s$, $t_2 = \bar{x}/s$, and $\theta_2 = \mu/\sigma$, and indeed, (13) does not hold; for the fiducial distribution of μ/σ based on the conditional distribution of s for σ fixed is $N[(s/\sigma)(\bar{x}/s), (1/n)]$, which means that the fiducial percentiles are linear in $\bar{x}/s$; but the percentiles of the marginal distribution of μ/σ are those of the fiducial distribution based on the marginal distribution of $\bar{x}/s$, because the fiducial distribution of μ/σ is the same in the two cases. $n^{\frac{1}{2}}(\bar{x}/s)$ has a non-central t distribution with $n-1$ degrees of freedom and non-centrality parameter $n^{\frac{1}{2}}(\mu/\sigma)$, and the percentiles of the non-central t distribution are not linear in the non-centrality parameter; it follows that the fiducial percentiles of μ/σ based on the distribution of $\bar{x}/s$ are not linear in $\bar{x}/s$. When (13) does not hold there is no reason to believe that the marginal distribution of μ/σ based on the factorization (20) has the hypothetical reference set, but on the other hand it is known to have the reference set based on the marginal distribution of $\bar{x}/s$, so the distribution of μ/σ based on (20) cannot be discarded because its marginal distribution cannot be considered to be fiducial distributions.

Another thing which might exclude a step-by-step argument is the existence of a relevant subset which invalidates the first step. Obviously, the examination of whether relevant subsets exist is to be considered as part of the step-by-step argument. Unfortunately no general methods exist for finding relevant subsets. In this case the sample, $x_1, \ldots, x_n$, is in one-to-one correspondence with the ancillary $C = [(x_1-\bar{x})/s, \ldots, (x_n-\bar{x})/s]$, $\bar{x}$, and s. $(\bar{x}, s)$ is sufficient for (μ, σ) and boundedly complete, so C and $(\bar{x}, s)$ are independent (Basu, 1955, 1958), and therefore only functions of $\bar{x}$ and s can define relevant subsets. The least one can do is to see whether s defines a relevant subset which invalidates the fiducial distribution of μ/σ based on $\bar{x}/s$, or $\bar{x}/s$ defines a relevant subset for the fiducial distribution of σ based on s.

To see whether s defines a relevant subset one considers the conditional distribution of $\bar{x}/s$ given s, which is normal with mean μ/s and variance σ^2/ns^2; and so the α percentiles of the conditional distribution are

$$t_\alpha\left(\frac{\mu}{\sigma}\middle|\sigma, s\right) = \frac{\mu}{s} + \frac{1}{n^{\frac{1}{2}}}\frac{\sigma}{s}u_\alpha$$

$$= \frac{\sigma}{s}\left(\frac{\mu}{\sigma} + \frac{1}{n^{\frac{1}{2}}}u_\alpha\right),$$

and the fiducial α percentiles are

$$\left(\frac{\mu}{\sigma_\alpha}\right)\left(\frac{\bar{x}}{s}\middle|\sigma; s\right) = \frac{s}{\sigma}\frac{\bar{x}}{s} + \frac{1}{n^{\frac{1}{2}}}u_\alpha.$$

The fiducial percentiles are straight lines which, depending on the unknown σ, may have any positive slope. Thus no exact fiducial distribution can be based on the conditional distribution of $\bar{x}/s$ given s when σ is unknown; but if for some values of $\bar{x}/s$, $\mu/\sigma_\alpha\,[(\bar{x}/s)\mid .\,; s]$ was bounded away from the fiducial α percentile of μ/σ, $\mu/\sigma_\alpha(\bar{x}/s)$, based on the marginal distribution of $\bar{x}/s$, the fiducial distribution based on $\bar{x}/s$ would be invalid. This is not the case, however, for $\mu/\sigma_\alpha(.)$ is increasing for each α and

$$\left(\frac{\mu}{\sigma_\alpha}\right)(\sigma \mid \sigma; s) = \frac{1}{n^{\frac{1}{2}}}\mu_\alpha = \frac{\mu}{\sigma_\alpha}(0).$$

To see that $\mu/\sigma_\alpha(0) = u_\alpha/n^{\frac{1}{2}}$ one notes that the fiducial α percentile is the inverse of the $1-\alpha$ percentile of the distribution of $\bar{x}/s$ considered as a function of $\bar{x}/s$, i.e. one has to find the value of μ/σ such that $1-\alpha = P[(\bar{x}/s)<0]$; now

$$1-\alpha = P\left(\frac{\bar{x}}{s}<0\right) = P(\bar{x}<0)$$

$$= P\left(\frac{\bar{x}}{\sigma}<0\right),$$

and $\bar{x}/\sigma$ has a $N[(\mu/\sigma), (1/n)]$ distribution with $1-\alpha$ percentiles $(\mu/\sigma)+(u_{1-\alpha}/n^{\frac{1}{2}})$. It follows that $\mu/\sigma = -u_{1-\alpha}/n^{\frac{1}{2}} = u_1/n^{\frac{1}{2}}$.

It has not been possible to see whether $\bar{x}/s$ defines a relevant subset for the fiducial distribution of σ based on s, but for the fiducial argument it has the same consequences as had $\bar{x}/s$ been shown not to be relevant; no relevant subset has been recognized and hence the distribution of σ based on the marginal distribution of s will be considered to be a fiducial distribution.

3.3 *Behrens' Problem*

The fiducial distribution of the difference between the means of two normal populations with possibly different variances is perhaps one of the most famous fiducial distributions due to its associated significance test which Fisher has called the Behrens' test.

The mean and the variance of the first population will be denoted by μ_1 and $\sigma_1{}^2$. Suppose the sample of size n_1 from the first population has given the statistics $\bar{x}_1$ and $s_1{}^2$. It follows from section 3.2 that in the hypothetical reference set of the fiducial distribution of μ_1 the distribution of $\bar{x}_1-\mu_1$ will be $t_1 s_1/n_1{}^{\frac{1}{2}}$, where t_1 is a random variable with a t-distribution with n_1-1 degrees of freedom. We note that the hypothetical reference set is the set of samples of size n_1 from normal populations all samples giving the same value of $s_1{}^2$ and $\sigma_1{}^2$ within this set of samples has the fiducial distribution given $s_1{}^2$.

Similarly for the sample of size n_2 from the second population one has, with obvious notation, that the distribution of $\bar{x}_2-\mu_2$ in the hypothetical reference set of μ_2 is $t_2 s_2/n_2^{\frac{1}{2}}$, where t_2 is a random variable having a t-distribution independent of t_1 with n_2-1 degrees of freedom. Hence $\mu_1-\mu_2-(\bar{x}_1-\bar{x}_2)$ will have the distribution of the random variable

$$\frac{s_1}{\sqrt{n_1}}t_1-\frac{s_2}{\sqrt{n_2}}t_2=\sqrt{\frac{s_1}{n_1}+\frac{s_2}{n_2}}(t_1 \sin R-t_2 \cos R), \tag{24}$$

where

$$\tan R=\frac{s_1\sqrt{n_2}}{s_2\sqrt{n_1}},$$

and thus

$$d=\frac{\mu_1-\mu_2-(\bar{x}_1-\bar{x}_2)}{\left(\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}\right)^{\frac{1}{2}}}$$

will have the distribution of the random variable

$$t_1 \sin R-t_2 \cos R. \tag{25}$$

The distributions of (24) or (25) may be inverted to give a fiducial distribution of $\mu_1-\mu_2$. It is obvious from the derivation given here what the reference set is.

The fiducial distribution of $\mu_1-\mu_2$ was given by Fisher (1935b), who used the calculus of probability on the independent fiducial distribution of μ_1 and μ_2 without any mentioning of the reference set involved. Fisher suggested that to test the hypothesis $\mu_1=\mu_2$ one should calculate $(\bar{x}_2-\bar{x}_1)/[(s_1^2/n_1)+(s_2^2/n_2)]^{\frac{1}{2}}$ and consult the distribution of $t_1 \sin R-t_2 \cos R$ to see whether the deviation from zero was significant. This is Behrens' test.

To facilitate the use of Behrens' test the distribution of $t_1 \sin R-t_2 \cos R$ has been tabulated by Behrens (1929), Sukkatme (1938), Fisher (1941), and Fisher and Healy (1956).

Bartlett (1936) noted that Behrens' test is not similar, and this prompted another derivation of the fiducial distribution of $\mu_1-\mu_2$ by Fisher (1939a). The reference set of the distribution of $\mu_1-\mu_2$ in this derivation is different from the one above. It has received special attention by Fisher (1961a).

The second derivation of the fiducial distribution of $\mu_1-\mu_2$ is as follows. Considering the sampling distribution of d one notes that the ratio of the observed variances defines a relevant subset; for the conditional distribution of

$$d\left(\frac{(n_1+n_2-2)\left(\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}\right)}{\left(1+\dfrac{1}{w}\right)\left(\dfrac{n_1-1}{n_1}s_1^2+w\dfrac{n_2-1}{n_2}s_2^2\right)}\right)^{\frac{1}{2}}$$

given s_1^2/s_2^2 is $t\,(n_1+n_2-2)$ where $w=\sigma_1^2 n_1/\sigma_2^2 n_2$. This was shown by Fisher (1939a, 1956). If w is known, the conditional distribution of d given s_1^2/s_2^2 does not depend on the parameters and one can obtain a fiducial distribution of $\mu_1-\mu_2$ given w. If w is unknown one may eliminate w integrating the conditional distribution of d given s_1^2/s_2^2 with respect to the fiducial distribution of w given s_1^2/s_2^2. This is equivalent to considering the sampling distribution of d in the hypothetical population of samples where the ratio s_1^2/s_2^2 of observed variances is constant and furthermore the ratio of σ_1^2/σ_2^2 within this set of samples has the fiducial distribution given s_1^2/s_2^2. This is the reference set of Behrens' test considered by Fisher (1936, 1961a).

The distribution of d in the reference set of Behrens' test is the one given by (25). This is not obvious, at least not to the present author, but it follows from the fact that in both cases the distribution of d is a mixture of normal distributions with mean 0 and variance

$$\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)\bigg/\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)$$

which may be expressed as

$$\left[\frac{s_1^2}{\sigma_1^2}\left(1+\frac{s_2^2}{s_1^2}\frac{n_1}{n_2}\right)\frac{w}{w+1}\right]^{-1}. \tag{26}$$

In the original reference set of d the distribution of (26) is determined by:

(A) s_1^2 and s_2^2 being constants and σ_1^2 and σ_2^2being independent random variables

$$\sigma_i^2 \sim \frac{s_i^2\,(n_i-1)}{\chi^2\,(n_i-1)} \qquad i = 1, 2,$$

and in the reference set of Behrens' test (26) has the conditional distribution given s_1^2/s_2^2, where

(B) s_1^2, s_2^2, σ_1^2, and σ_2^2 are random variables; the distribution of s_1^2 and s_2^2 given σ_1^2 and σ_2^2 being independent and

$$s_i^2 \sim \frac{\sigma_i^2}{n_i-1}\,\chi^2\,(n_i-1), \qquad i = 1, 2,$$

and the distribution of σ_1^2/σ_2^2 given s_1^2/s_2^2 being

$$\frac{s_1^2}{s_2^2}\,v^2\,(n_2-1,\,n_1-1).$$

It is not difficult to see that the marginal distribution of $(s_1^2/\sigma_1^2\ w)$ is the same in the two reference sets, whence (26) and d has the same distribution in the two reference sets.

Behrens' test is not generally accepted as witnessed by the abundance of attempts to find similar or approximately similar tests. For a recent proposal and references to earlier ones see Lee and Gurland (1975). Linnik (1963, 1968) has shown that no similar tests with some desirable properties exist. None of the tests mentioned by Lee and Gurland are similar. Robinson (1976) advocates Behrens' test.

4 Conclusions

The account of fiducial probability presented in this paper is based on the concept of relevant subsets, which seems to explain some, at least, of the known examples of inconsistencies in fiducial inference. But the paper has not presented the theory of fiducial inference; for fiducial inference is not a theory but a collection of examples based on an insufficiently explored concept of a relevant subset and a vague formulation of the concept of probability involved. To make fiducial inference into a theory a lot of questions concerning relevant subset need to be answered; such as, when do relevant subset (not) exist, and if they do, how can they be found. These questions are also of great importance to a satisfactory theory of confidence intervals. The concept of probability also needs some clarification. The example with the probability of the birth of a boy quoted in section 2.2 is very suggestive but there do seem to be a huge logical step from the reference sets involved in that example and the reference set of a fiducial distribution of a parameter. Any development or destruction of Fisher's fiducial argument should be within those areas.

Acknowledgements

I am grateful to G.A. Barnard, O. Barndorff-Nielsen, K. Pedersen, and G. Schou for valuable comments on earlier versions of this paper.

Bibliography

The bibliography lists several papers not referred to in the text. It supplements the bibliographies given in Tukey (1957) and Brillinger (1962).

Discussions

Discussion on symposium of interval estimation. *Journal of the Royal Statistical Society, Series B* **16**, 204–222.
Discussion on fiducial probability. *Bulletin of the International Statistical Institute*, **40**, 919–939.

Papers and Monographs

Barnard, G.A. (1963a). Some logical aspects of the fiducial argument. *Journal of the Royal Statistical Society, Series B*, **25**, 111–114.
Barnard, G.A. (1963b). Fisher's contributions to mathematical statistics. *Journal of the Royal Statistical Society, Series A*, **126**, 162–166.
Barnard, G.A. (1964). Logical aspects of the fiducial argument. *Bulletin of the International Statistical Institute*, **40**, 870–883.
Barnard, G.A. (1974a). The foundations of statistics 1964–1974. *Bulletin of the Institute of Mathematics and Its Applications*, **10**, 344–347.
Barnard, G.A. (1974b). Can we all agree on what we mean by estimation? *Utilitas Mathematica*, **6**, 1–22.
Bartlett, M.S. (1936). The information available in small samples. *Proceedings of the Cambridge Philosophical Society*, **32**, 560–566.
Basu, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā*, **15**, 377–380.
Basu, D. (1958). On statistics independent of a sufficient statistic. *Sankhyā*, **20**, 223–226.
Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *The Philosophical Transactions*, **53**, 370–418. Reproduced in *Biometrika* (1958), **45**, 296–315.
Behrens, W.V. (1929). Ein Beitrag zur Fehlerberechnung bei wenig Beobachtungen. *Landw. Jb.*, **68**, 807–837.
Bennet, G.W. and Cornish, E.A. (1964). A comparison of the simultaneous fiducial distributions derived from the multivariate normal distribution. *Bulletin of the International Statistical Institute*, **40**, 902–919.
Brillinger, D.R. (1962). Examples on the definition of fiducial probability with a bibliography. *Annals of Mathematical Statistics*, **33**, 1349–1355.
Brown, L. (1967). The conditional level of Student's t test. *Annals of Mathematical Statistics*, **38**, 1068–1071.
Buehler, R.J. (1959). Some validity criteria for statistical inferences. *Annals of Mathematical Statistics*, **30**, 845–863.
Buehler, R.J. and Feddersen, A.P. (1963). Note on a conditional property of Student's t. *Annals of Mathematical Statistics*, **34**, 1098–1100.
Cornfield, J. (1969). The Bayesian outlook and its applications (with discussion). *Biometrics*, **25**, 617–657.
Dempster, A.P. (1963a). Further examples of inconsistencies in the fiducial argument. *Annals of Mathematical Statistics*, **34**, 884–891.
Dempster, A.P. (1963b). On a paradox concerning inference about a covariance matrix. *Annals of Mathematical Statistics*, **34**, 1414–1418.
Dempster, A.P. (1964). On the difficulties inherent in Fisher's fiducial argument. *Journal of the American Statistical Association*, **59**, 56–66.
Edwards, A.W.F. (1973). A problem in the doctrine of chances. Proceedings of conference on foundational questions in statistical inference, 43–57. Memoirs no. 1. Department of theoretical statistics. University of Aarhus.
Edwards, A.W.F. (1976). Fiducial probability. *The Statistician*, **25**, 15–35.
Fisher, R.A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society*, **26**, 528–535.
Fisher, R.A. (1933). The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proceedings of the Royal Society of London, Series A*, **139**, 343–348.
Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proceedings of the Royal Society of London, Series A*, **144**, 285–307.
Fisher, R.A. (1935a). The logic of inductive inference, (with discussion). *Journal of the Royal Statistical Society*, **98**, 39–54.
Fisher, R.A. (1935b). The fiducial argument in statistical inference. *Annals of Eugenics*, **6**, 391–398.
Fisher, R.A. (1936). Uncertain inference. *Proceedings of the American Academy of Arts and Sciences*, **71**, 245–258.
Fisher, R.A. (1937). On a point raised by M. S. Bartlett on fiducial probability. *Annals of Eugenics*, **7**, 370–375.

Fisher, R.A. (1939a). The comparison of samples with possibly unequal variances. *Annals of Eugenics*, **9**, 174–180.

Fisher, R.A. (1939b). A note on fiducial inference. *Annals of Mathematical Statistics*, **10**, 383–388.

Fisher, R.A. (1941). The asymptotic approach to Behrens' integral with further tables for the d test of significance. *Annals of Eugenics*, **11**, 141–172.

Fisher, R.A. (1945). The logical inversion of the notion of the random variable. *Sankhyā*, **7**, 129–132.

Fisher, R.A. (1948). Conclusions fiduciaires. *Annales de l'Institut Henri Poincaré*, **10**, 191–213.

Fisher, R.A. (1953). Dispersion on a sphere. *Proceedings of the Royal Society of London, Series A*, **217**, 295–305.

Fisher, R.A. (1955). Statistical methods and scientific induction. *Journal of the Royal Statistical Society, Series B*, **17**, 69–78.

Fisher, R.A. (1956). On a test of significance in Pearson's biometrika tables (No. 11). *Journal of the Royal Statistical Society, Series B*, **18**, 55–60.

Fisher, R.A. (1956 – 3rd Ed. 1973). *Statistical Methods and Scientific Inference*. New York: Hafner.

Fisher, R.A. (1957a). The underworld of probability. *Sankhyā*, **18**, 201–210.

Fisher, R.A. (1957b). Comment on the notes by Neyman, Bartlett, and Welch in this journal. (Vol. **18**, No. 2, 1956). *Journal of the Royal Statistical Society, Series B*, **19**, 179.

Fisher, R.A. (1958a). The nature of probability. *The Centennial Review*, **2**, 261–274.

Fisher, R.A. (1958b). Mathematical probability in the natural sciences. 18th Int. Congr. of Pharm. Sciences, Brussels. Printed in *Technometrics*, **1**, 21–29.

Fisher, R.A. (1960a). On some extensions of Bayesian inference proposed by Mr Lindley. *Journal of the Royal Statistical Society, Series B*, **22**, 299–301.

Fisher, R.A. (1960b). Scientific thought and the refinement of human reasoning. *Journal of the Operations Research Society of Japan*, **3**, 1–10.

Fisher, R.A. (1961a). Sampling the reference set. *Sankhyā A*, **23**, 3–8.

Fisher, R.A. (1961b). The weighted mean of two normal samples with unknown variance ratio. *Sankhyā A*, **23**, 103–114.

Fisher, R.A. (1962). Some examples of Bayes' method of the experimental determination of probabilities *a priori*. *Journal of the Royal Statistical Society, Series B*, **24**, 118–124.

Fisher, R.A. and Healy, M.J.R. (1956). New tables of Behrens' test of significance. *Journal of the Royal Statistical Society, Series B*, **18**, 212–216.

Fraser, D.A.S. (1961a). On fiducial inference. *Annals of Mathematical Statistics*, **32**, 661–676.

Fraser, D.A.S. (1961b). The fiducial method and invariance. *Biometrika*, **48**, 261–280.

Fraser, D.A.S. (1962). On the consistency of the fiducial method. *Journal of the Royal Statistical Society, Series B*, **24**, 425–434.

Fraser, D.A.S. (1964a). On inference for location and scale parameters. *Biometrika*, **51**, 17–24.

Fraser, D.A.S. (1964b). On the definition of fiducial probability. *Bulletin of the International Statistical Institute* **40**, 842–856.

Fraser, D.A.S. (1965). Fiducial consistency and group structure. *Biometrika*, **52**, 55–65.

Fraser, D.A.S. (1968). *The Structure of Inference*. Wiley, New York.

Healy, M.J.R. (1963). Fiducial limits for a variance component. *Journal of the Royal Statistical Society, Series B*, **25**, 128–130.

James, A.T., Wilkinson, G.N. and Venables, W.N. (1974). Interval estimates for a ratio of means. *Sankhyā A*, **36**, 177–183.

Kyburg, Jr., H.E. (1964). Logical and fiducial probability. *Bulletin of the International Statistical Institute*, **40**, 884–901.

Lee, A.F.S. and Gurland, J. (1975). Size and power of tests for equality of means of two normal populations with unequal variances. *Journal of the American Statistical Association*, **70**, 933–941.

Linnik, Yu.V. (1963). Remarks on the Behrens–Fisher problem. *Sankhyā A*, **25**, 377–380.

Linnik, Yu.V. (1964). On the Behrens–Fisher problem. *Bulletin of the International Statistical Institute*, **40**, 833–841.

Linnik, Yu.V. (1966). Recent developments on the Behrens problem. *Sankhyā A*, **28**, 15–24.

Linnik, Yu.V. (1968). *Statistical problems with nuisance parameters*. Volume 20. American Mathematical Society translations of mathematical monographs. Translated from the Russian.

Mauldon, J.G. (1955). Pivotal quantities for Wishart's and related distributions, and a paradox in fiducial theory. *Journal of the Royal Statistical Society, Series B*, **17**, 79–85.

Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, **97**, 558–606.

Pinkham, R.S. (1966). On a fiducial example of C. Stein. *Journal of the Royal Statistical Society, Series B*, **28**, 53–54.

Robinson, G.K. (1975). Some counterexamples to the theory of confidence intervals. *Biometrika*, **62**, 155–161.

Robinson, G.K. (1976). Properties of Student's *t* and of the Behrens–Fisher solution to the two means problem. *Annals of Statistics*, **4**, 963–971.

Sprott, D.A. (1963). Fiducial distributions associated with independent normal variates. *Sankhyā A*, **25**, 403–406.

Sprott, D.A. (1964). A transformation model for the investigation of fiducial distributions. *Bulletin of the International Statistical Institute*, **40**, 856–869.

Sprott, D.A. (1965). Transformations and sufficiency. *Journal of the Royal Statistical Society, Series B*, **27**, 479–485.

Stein, C. (1959). An example of wide discrepancy between fiducial and confidence intervals *Annals. of Mathematical Statistics*, **30**, 877–880.
Sukhatme, P.V. (1938). On Fisher and Behrens' test of significance for the difference in means of two normal samples. *Sankhyā*, **4**, 39–48.
Tan, P. (1970). On Sprott's example of an ancillary statistic. *Bulletin*, **13**, 205–207.
Tan, P. (1973). On Fisher's problem of the Nile in uncertain inference. *Communications in Statistics*, **2**, 45–58.
Tukey, J.W. (1957). Some examples with fiducial relevance. *Annals of Mathematical Statistics*, **28**, 687–695.
Wilkinson, G.N. (1977). On resolving the controversy in statistical inference (with discussion). *Journal of the Royal Statistical Society, Series B*, **38**, 119-171.
Wallace, D.L. (1959). Conditional confidence level properties. *Annals of Mathematical Statistics*, **30**, 864–876.
Williams, E.J. (1962). Exact fiducial limits in non-linear estimation. *Journal of the Royal Statistical Society, Series B*, **24**, 125–139.
Williams, E.J. (1963). A comparison of the direct and the fiducial arguments in the estimation of a parameter. *Journal of the Royal Statistical Society, Series B*, **25**, 95–99.
Williams, J.S. (1966). The rôle of probability in fiducial inference. *Sankhyā A*, **28**, 271–296.
Yates, F. (1939). An apparent inconsistency arising from tests of significance based on fiducial distributions of unknown parameters. *Proceedings of the Cambridge Philosophical Society*, **35**, 579–591.
Yates, F. (1964). Fiducial probability, recognizable subsets and Behrens' test. *Biometrics*, **20**, 343–360.

Résumé

Parmi les nombreuses contributions importantes de R.A. Fisher à la statistique, l'argument fiduciel n'a eu qu'un succès très limité et n'a pas survécu. Cependant l'idée s'est perpétuée et l'on trouve à l'occasion des mots tels que 'distribution fiducielle' ou 'intervalle fiduciel' encore dans des articles récents. Plusieurs articles expliquent comment obtenir une distribution fiducielle, mais omettent de mentionner toutes les exigences essentielles concernant la validité d'un argument fiduciel. Omission excusable si l'article a été écrit avant 1956, car l'une de ces exigences fût formulée pour la première fois dans le livre de Fisher, *Statistical Methods and Scientific Inference*, de 1956 et (par suite) est absente même dans la plus part des écrits de Fisher sur l'argument fiduciel.

Cette exigence, à savoir l'existence de sous-ensembles convenables à une distribution fiducielle, est formulée ici à la section 2.2. Elle entraine quelques restrictions à l'emploi des probabilités fiducielles que Fisher n'avait pas mentionnées, et ceci explique quelques unes des incohérences des probabilités fiducielles qu'on rencontre dans la littérature.

Statistical Methods and Scientific Inference est l'exposé le meilleur et le plus complet des idées de Fisher sur les probabilités fiducielles, mais il est apparu que sa lecture n'était pas facile. On souhaite que le présent article serve aussi d'introduction aux oeuvres de Fisher sur ce sujet.