Fiducial Inference*

by

Robert J. Buehler

Technical Report No. 353

School of Statistics

University of Minnesota

July, 1979

*This is a draft of an article prepared for the Encyclopedia of Statistical Sciences, to be published by John Wiley & Sons, Inc.

It has been said that fiducial inference as put forward by R.A. Fisher is not so much a theory as a collection of examples. A gradual evolution of ideas can be seen in Fisher's published work, and he himself may have been less satisfied with his own theories than his writings would lead one to suspect. Joan Box (1978), p. 458, writes:

> "He continued to work at these problems but to the end of his days was not satisfied with the further solutions he could arrive at. .....He did not unravel the puzzle."

In view of this and in view of the lack of any clear definition of what fiducial probability is, it is not surprising that the subject has been one of confusion and controversy ever since its introduction in 1930. While interest in fiducial inference has declined since Fisher's death in 1962, there continue to be efforts to clarify and extend his ideas. See for example Fraser (1961a, b, 1966, 1968, 1976, 1979), Hacking (1965), Williams (1966), Verhagen (1966), Wilkinson (1977), Pedersen (1978).

Several key ideas can be illustrated by the case of a single observation x from a normal distribution with mean μ and unit variance, for which we use the standard notation $x \sim N(\mu,1)$. If we put $z = x - \mu$ then $z \sim N(0,1)$. A quantity like z which depends on the observation x and the parameter μ, and whose distribution is free of the parameter is called a <u>pivotal quantity</u> or <u>pivot</u>. The <u>fiducial argument</u> consists in writing $\mu = x - z$ and asserting that when we have no knowledge about μ except the value x, then our uncertainty about μ is summarized by saying μ equals x minus an unknown value of a standard normal random variable. In short $\mu \sim N(x,1)$. This is called the <u>fiducial distribution</u> of μ. The values $x \pm 1.96$ include all but five percent of the distribution and so would be called 95 percent fiducial limits for μ.

Sometimes the objection is raised that μ is not a random variable, but a fixed unknown number. Consider however that a coin which has already fallen heads or tails has achieved a fixed outcome, but until we actually view it, our uncertainty about it for any purpose of action or decision is represented probabilistically by $P(\text{heads}) = P(\text{tails}) = 1/2$.

If indeed the previously mentioned trial were repeated indefinitely with arbitrarily varying μ values to give (μ_1,x_1), $(\mu_2,x_2),\cdots$, and if the μ values were subsequently revealed and plotted on a scale relative to a fixed point x in such a way that $\mu_i - x$ equals the actual i-th difference $\mu_i - x_i$, then the plotted values would follow a normal distribution centered at x, that is, a $N(x,1)$ distribution, the fiducial distribution. Thus in the absence of a priori information, our knowledge (or uncertainty) about μ_1 say, given x_1, is summed up by stating that if its value were revealed, it would appear to be a random value from $N(x_1,1)$.

1. Relationship to Bayesian inference.

If $f(x,\theta)$ is the assumed probability law of data x and $\pi(\theta)$ is a prior density then Bayes Theorem yields the posterior density $\pi(\theta|x) = \pi(\theta)f(x,\theta) \;/\; \int\pi(\theta)f(x,\theta)d\theta$. In using the term "inverse probability" Fisher (1930) referred to the practice of taking $\pi(\theta)$ to be constant in order to represent prior ignorance (Bayes' postulate). Pointing out that this procedure was inconsistent under transformations of θ, Fisher put forward his own likelihood theory and fiducial theory to avoid the objectionable postulate.

In more recent times there has been a tendency to regard the prior density $\pi(\theta)$ as a representation of subjective belief. This view was equally distasteful to Fisher, whose constant goal was an objective theory uncontaminated by subjective elements.

2. Relationship to confidence intervals.

Like fiducial theory, Neyman's theory of confidence limits leads to probability statements about the value of θ without appealing to any prior density. Random intervals $I(x)$ of θ values are confidence intervals with confidence level γ if $\Pr[\theta \in I(x)|\theta] = \gamma$ for all θ. As we indicate later, in many examples there is a formal correspondence between the two theories in that the fiducial probability of the confidence interval equals the confidence level. Whether or not the two theories give different numerical results, there are differences in their aims and interpretations.

(i) In confidence interval theory, θ is considered a fixed constant and the interval is considered random. In fiducial theory, x is considered fixed and θ random, or perhaps more accurately, uncertain.

(ii) Confidence intervals are admittedly nonunique. In Fisher's view it was a fatal defect that different solutions could assign different confidence levels to a single interval. Uniqueness of fiducial distributions was consistently maintained by Fisher but disputed by others. Concepts like sufficiency and Fisher information presumably furnish the keys to uniqueness.

(iii) Fiducial theory yields, through the integral of the density, the fiducial probability of any interval, whereas confidence intervals only assign confidence γ to the particular interval $I(x)$. This distinction does tend to disappear however if we require a solution for every $0 < \gamma < 1$ rather than only for one fixed value such as $\gamma = 0.95$.

3. Estimating a single parameter.

Let x be either a single observation or a sufficient statistic with a CDF $F(x,\theta)$ such that $\partial F/\partial\theta$ is negative. Then the contours of F in the (θ,x) plane slope upward to the right. Consider the contours $F = 0.1, 0.2,\cdots,0.9$ which divide the (θ,x) plane into ten regions. For any fixed θ, the random value of x has equal probability of falling in each region. If x is fixed, then the set of all θ values is divided into ten intervals. From the fiducial point of view, each of these intervals has fiducial probability (given x) of 0.1. From the confidence interval point of view, the θ values in say any k contiguous intervals would constitute a confidence interval with confidence coefficient $0.1\,k$. By refining the subdivision of F values we are led to the expression

(1) $$\varphi(\theta|x) = -\partial F(x,\theta)/\partial\theta$$

for the fiducial density of θ given x, a formula given by Fisher (1930) when first introducing fiducial probability, and given again in (1956), p. 70. Fisher's 1930 explanation emphasized frequencies and hardly differs from a description of confidence intervals. Only later when the theories were extended to more complex models did differences become apparent.

Necessary and sufficient conditions for $-\partial F/\partial\theta$ to be a posterior distribution for some prior are given by Lindley (1958): There exist transformations of x to u and θ to τ such that τ is a location parameter for u. The prior on τ must then be uniform, and if regularity conditions require $-\infty < \tau < \infty$, only an improper prior distribution is capable of yielding a posterior density identical with the fiducial density.

To derive (1) by a pivotal argument, let $u = F(x,\theta)$. The transformation from x to u with θ fixed yields the uniform density $g(u) = 1$ $(0 \leq u \leq 1)$, and thus $u = F(x,\theta)$ is a pivot (Fisher (1935), p. 395). The transformation from u to θ with x fixed gives $\varphi(\theta|x) = g(u)|\partial u/\partial\theta| = -\partial F/\partial\theta$, the fiducial density.

When θ is a location parameter for x, then $F(x,\theta)$ has the form $H(x-\theta)$ and the fiducial density is $\varphi(\theta|x) = h(x-\theta)$ where $h(u) = d\,H(u)/du$. In this case the graphs of $f(x,\theta)$ and $\varphi(\theta|x)$ are mirror images.

4. Joint pivots

For problems involving two or more parameters there are several methods for deriving fiducial distributions, and it is not surprising to encounter difficulties in establishing unique solutions. For example one might use either joint pivots or a combination of conditional and marginal pivots.

If $u_i = u_i(x_1,x_2,\theta_1,\theta_2)$ $(i = 1,2)$ where x_1,x_2 are statistics and θ_1,θ_2 are parameters and if the density $g(u_1,u_2) = f(x_1,x_2;\theta_1,\theta_2)|J_{xu}|$ (J_{xu} is the Jacobian of the transformation with fixed θ_1,θ_2) does not depend on θ_1,θ_2, then u_1,u_2 are joint pivots. Transforming from u_1,u_2 to θ_1,θ_2 with x_1,x_2 fixed gives

$$\varphi(\theta_1,\theta_2|x_1,x_2) = g(u_1,u_2)|J_{u\theta}| ,$$

which is the joint fiducial distribution of θ_1,θ_2, at least if we have chosen legitimate pivots. Fisher (1956), p. 172, cautions against an arbitrary choice of pivots but provides no comprehensive rules. An earlier discussion (1935), p. 395, ignored consistency problems.

Any fiducial distribution obtained from joint pivots is consistent with a confidence region interpretation by the following argument. Let R be any region in the u_1, u_2 plane with $\Pr[R] = \gamma$, and let $S(x_1, x_2)$ be the image of R in the θ_1, θ_2 plane depending on fixed observed values x_1, x_2. Then $S(x_1, x_2)$ is a confidence region with confidence level γ, that is $\Pr[(\theta_1, \theta_2) \in S(X_1, X_2) | \theta_1, \theta_2] = \Pr[R] = \gamma$, and $S(x_1, x_2)$ has fiducial probability γ.

5. Student's distribution

One of the best known fiducial distribution is that of the normal mean μ when the population variance σ^2 is also unknown. If $\bar{x}$ denotes the mean of a sample of size n, and $s^2 = (n-1)^{-1}\Sigma(x_i - \bar{x})^2$, then the quantity $t = \sqrt{n}\,(\bar{x}-\mu)/s$ is known to have Student's distribution. In first presenting this example in 1935 Fisher wrote:

> "It must now be noticed that t is a continuous function of the unknown parameter, the mean, together with observable values, $\bar{x}$, s and n, only. Consequently the inequality $t > t_1$ is equivalent to the inequality $\mu < \bar{x} - st_1/\sqrt{n}$, so that this last inequality must be satisfied with the same probability as the first. This probability is known for all values of t_1, and decreases continuously as t_1 is increased. Since, therefore, the right-hand side of the inequality takes, by varying t_1, all real values, we may state the probability that μ is less than any assigned value, or the probability that it lies between any assigned values, or, in short, its probability distribution, in the light of the sample observed."

Thus by the pivotal or fiducial argument, the fiducial distribution of μ is the distribution of $\bar{x} - st/\sqrt{n}$ where $\bar{x}$ and s are fixed at their observed values and t has Student's distribution.

The Student example is notable in that all routes seem to converge on a single answer. The fiducial limits are of course identical to the confidence limits found in virtually every statistics textbook. Moreover the British astronomer Jeffreys using the prior $d\mu d\sigma/\sigma$ (which he favored for reasons of invariance), noted (1948), pp. 122, 352, the correspondence of the fiducial and posterior distributions. In addition we may note some consistent variants of the fiducial method:

(i) Use joint pivots (t,u) with $u = s/\sigma$ and get the marginal density of μ from the joint density of μ and σ. (ii) Use u to get the marginal density of σ from the marginal density of s, then multiply this by the conditional density of μ given σ obtained from the pivot $\bar{x} - \mu$ conditional on σ. (Fisher (1956), page 119, calls this the "rigorous" way.) (iii) Use the Fisher-Pitman theory of location and scale parameters discussed below. (iv) Obtain the fiducial distribution of the mean and variance of a future sample of size n' and let n' tend to infinity (Fisher (1935), (1956), p. 119).

6. Behrens' distribution

The estimation of the difference of normal means $\delta = \mu_1 - \mu_2$ when the variances are not assumed equal is known as the Behrens-Fisher problem. It is of historical interest as an early example in which fiducial limits are not confidence limits. No entirely satisfactory confidence interval solution is available, and the merits of the Behrens-Fisher solution and its competitors continue to be debated.

In an obvious extension of the notation of the previous section we can write

$$\delta = \mu_1-\mu_2 = \bar{x}_1-\bar{x}_2 - s_1t_1/\sqrt{n_1} + s_2t_2/\sqrt{n_2}.$$

From this Fisher (1935) argued that δ is fiducially distributed like a constant, $\bar{x}_1-\bar{x}_2$, plus a variable equal to a weighted sum of two independent Student variables (a Behrens distribution). While the exact coverage probability of the resulting fiducial intervals cannot exactly equal the corresponding fiducial probability (see for example Kendall and Stuart (1961), page 149), numerical evidence indicates the procedure is conservative (see Robinson (1976) and also footnote 28 (by John Pratt) to Savage (1976)).

The fiducial distribution of δ is known to equal a posterior distribution corresponding to the improper prior $d\mu_1 d\mu_2 d\sigma_1 d\sigma_2/\sigma_1\sigma_2$.

7. Ancillary statistics

An ancillary statistic is one whose distribution does not depend on the parameter. For example with a sample of size n from a location family $f(x-\theta)$ the vector y of $n-1$ spacings of the ordered observations has a distribution which is free of θ. The statistic y has the curious property that by itself it contains no information about θ, but when used in conjunction with the maximum likelihood estimator $\hat{\theta}$, it carries information about the precision of $\hat{\theta}$. Fisher called the spacings the "complexion" of the sample and argued that inference should be made conditionally on the observed value of y. For any fixed y, $\hat{\theta}-\theta$ is a pivot, and a conditional pivotal argument gives a fiducial density proportional to the likelihood function: $\varphi(\theta|x_1,\cdots,x_n) \propto \prod_{i=1}^{n} f(x_i-\theta)$.

A forceful if artificial example in which the conditional treatment makes sense is furnished by case of two observations from a uniform distribution with mean θ and unit range. If $x_{(1)}$ and $x_{(2)}$ denote the smaller and larger observations respectively, then the possible values of θ are $x_{(2)} - 1/2 \leq \theta \leq x_{(1)} + 1/2$, and the fiducial distribution is uniform over this range. The unconditional pivot $\bar{x} - \theta$ would not give a legitimate fiducial distribution because $\bar{x}$ is not sufficient; nevertheless it would yield confidence intervals. Symmetric limits, $\bar{x} \pm c$, have the seemingly objectionable property of covering θ with certainty for large enough values of the ancillary statistic $x_{(2)} - x_{(1)}$.

Logical problems associated with ancillary statistics are not fully settled. Even if it is granted that inference should be conditional on an ancillary statistic, the problem of showing existence or nonexistence of ancillaries remains unsolved. Fisher (1936), (1956), p. 118, termed this the "problem of the Nile," likening the partition of the sample space to the partitioning of land of a Nile village in such a way that the yields of the lots would be in predetermined proportion whatever the height of the flood. Here height of flood corresponds to θ and lot boundaries to contours of the ancillary statistic.

8. Location and scale models

For a sample of size n from a location-scale model $\sigma^{-1}f((x-\theta)/\sigma)$, the $n-2$ quotients of the $n-1$ spacings of the ordered observations are distributed independently of (θ,σ) and so are jointly ancillary. Conditional joint pivots can be found which yield the Fisher-Pitman fiducial distribution

$$\varphi(\theta,\sigma|x_1,\cdots,x_n) \propto \sigma^{-n-1} \prod_{i=1}^{n} f((x_i-\theta)/\sigma) .$$

Fisher (1934) gave likelihood theory relevant to this model, but the fiducial distribution was first given explicitly by Pitman (1939). The example is discussed again in Fisher (1956), pp. 159-163. The fiducial distribution is evidently equivalent to a posterior distribution corresponding to the improper prior $d\theta d\sigma/\sigma$, and it is known that marginal distributions of θ and σ can be used to obtain confidence intervals. By transformation the results can be applied to distributions not initially in location-scale form, such as the Weibull (Lawless, 1978).

9. Difficulties

The following examples, paradoxical in varying degrees, show why circumspection is needed in interpretations and manipulations of fiducial probability.

If $x \sim N(\theta,1)$, the fiducial density of θ^2 derived from $\theta \sim N(x,1)$ is different from that derived from the density of x^2.

Tukey (1957) has shown that different pairs of joint pivots can lead to different fiducial distributions.

Lindley (1958) considered two observations from $f(x,\theta) = \theta^2(x+1)e^{-\theta x}/(\theta+1)$ $(x > 0)$ and showed that the fiducial distribution $\varphi(\theta|x_1,x_2)$ is not equal to the posterior distribution of θ given x_2 when the prior is taken to equal the fiducial distribution $\varphi(\theta|x_1)$. If the fiducial distribution is to be interpreted like a prior distribution, as one might infer for example from Fisher (1956), p. 125, then one would have expected equality. A related fact is that $\varphi(\theta|x_1)$ is not a posterior density for any prior.

Stein (1959) obtained the marginal fiducial distribution of $\Sigma_1^n \theta_i^2$ from the joint fiducial distribution of $\theta_1, \cdots, \theta_n$ given $x_1, \cdots, x_n$

where $x_i \sim N(\theta_i,1)$, and showed that there could be arbitrarily large discrepancies between the resulting fiducial probabilities and confidence levels arrived at by using the statistic $\Sigma_1^n x_i^2$.

In the estimation of μ_2/μ_1 given a sample from a bivariate normal population, Creasy (1954) and Fieller (1954) obtained different solutions by using different pivots (the "Creasy-Fieller paradox").

Given x_1, x_2 from $N(\mu,\sigma)$ the interval $\min(x_1,x_2) < \mu < \max(x_1,x_2)$ has fiducial probability 0.5, but in the subset of cases where $3|x_1-x_2| > 2|x_1+x_2|$ the conditional probability exceeds 0.518 for all μ,σ (Buehler and Feddersen (1963)). Brown (1967) gives generalizations; Yates (1964) defends Fisher's theory.

10. Invariance

Many standard parametric models have the following invariance property: If x has a distribution in the given family, so does gx, where g is an element of a transformation group G, and there is a one-to-one correspondence between the elements g of G and the parameter values θ. Fraser (1961 a,b) attempted to set up a rigorous mathematical framework for fiducial theory for such models, and his later theory of structural inference (1968) is a continuation of this work. In these models the orbit of any point x is the set of all gx with g ranging over G. The orbit label turns out to be an ancillary statistic, and by using a pivotal argument conditional on the ancillary one obtains a fiducial distribution which equals the posterior distribution when the prior measure equals the right Haar measure on G. The Haar measure is improper in the most familiar examples, but not for distributions on the circle or sphere (Fraser, 1979).

If $\psi(\theta)$ is a real-valued function of a vector parameter θ , then a sufficient condition for fiducial limits for ψ , obtained from its marginal distribution, to be confidence limits is the following: $\psi(\theta_1) = \psi(\theta_2)$ implies $\psi(g\theta_1) = \psi(g\theta_2)$ for all g (Hora and Buehler, 1966). Further discussion of these invariant models is found in Zacks (1971), Sections 7.2., 7.3 .

11. Further reading.

In view of the unresolved difficulties, it is understandable that textbook authors tend to shy away from fiducial theory. The fiducial advocate Quenouille (1958) is one exception. The less partisan writers Kendall and Stuart (1961) describe their approach like this (p. 152):

> "There has been so much controversy about the various methods of estimation we have described that, at this point, we shall have to leave our customary objective standpoint and descend into the arena ourselves."

Savage (1976), p. 467, gives a list of all examples of fiducial distributions in Fisher's published work. Fairly extensive bibliographies are given by Brillinger (1962), Savage (1976) and Pedersen (1978). Fisher (1956) (or better, the third edition dated 1973) is the most authoritative source, but not the easiest to read.

References

Box, Joan Fisher (1978). R. A. Fisher, the Life of a Scientist. Wiley, New York.

Brillinger, D. R. (1962). "Examples on the definition of fiducial probability with a bibliography." Ann. Math. Statist. 33, 1349-1355.

Brown, L. (1967). "The conditional level of Student's t test." Ann. Math. Statist. 38, 838-848.

Buehler, R.J. and Feddersen (1963). "Note on a conditional property of Student's t ." Ann. Math. Statist. 34, 1098-1100.

Creasy, Monica A. (1954). "Limits for the ratio of means." J. Roy. Statist. Soc. Ser. B 16, 186-194.

Fieller, E. C. (1954). "Some problems in interval estimation." J. Roy. Statist. Soc. Ser. B 16, 175-185.

Fisher, R. A. (1930). "Inverse probability." Proc. Cambridge Philos. Soc. 26, 528-535.

Fisher, R. A. (1934). "Two new properties of mathematical likelihood." Proc. Roy. Soc. London, Ser. A., 144, 285-307.

Fisher, R. A. (1935). "The fiducial argument in statistical inference." Annals of Eugenics 6, 391-398.

Fisher, R. A. (1936). "Uncertain inference." Proceedings of the American of Arts and Sciences 71, 245-258.

Fisher, R. A. (1956). Statistical Methods and Scientific Inference. London, Oliver and Boyd.

Fraser, D. A. S. (1961a). "On fiducial inference." Ann. Math. Statist. 32, 661-676.

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

Fraser, D. A. S. (1966). "Structural probability and a generalization." Biometrika 53, 1-9.

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

Fraser, D. A. S. (1976). "Necessary analysis and adaptive inference." J. Amer. Statist. Assoc. 71, 99-113.

Fraser, D. A. S. (1979). Inference and Linear Models. New York, McGraw-Hill.

Hacking, Ian (1965). Logic of Statistical Inference. Cambridge University Press.

Hora, R. B. and Buehler, R. J. (1966). "Fiducial theory and invariant estimation." Ann. Math. Statist. 37, 643-656.

Jeffreys, Harold (1948). Theory of Probability (second ed.). Oxford University Press.

Kendall, M. G. and Stuart, Alan (1961). The Advanced Theory of Statistics, Vol. 2. Charles Griffin & Co. Ltd., London.

Lawless, J. F. (1978). "Confidence interval estimation for the Weibull and extreme value distributions." Technometrics 20, 353-368.

Lindley, D. V. (1958). "Fiducial distributions and Bayes' theorem." J. Roy. Statist. Soc. Ser. B 20, 102-107.

Pedersen, J. G. (1978). "Fiducial inference." International Statistical Review 46, 147-170.

Pitman, E. J. G. (1939). "The estimation of location and scale parameters of a continuous population of any given form." Biometrika 30, 391-421.

Quenouille, M. H. (1958). Fundamentals of Statistical Reasoning. Charles Griffin & Co. Ltd., London.

Robinson, G. K. (1976). "Properties of Student's t and of the Behrens-Fisher solution to the two means problem." Ann. Statist. 4, 963-971.

Savage, L. J. (1976). "On rereading R. A. Fisher." Ann. Statist. 4, 441-500.

Stein, Charles (1959). "An example of a wide discrepancy between fiducial and confidence intervals." Ann. Math. Statist. 30, 970-979.

Tukey, J. W. (1957). "Some examples with fiducial relevance." Ann. Math. Statist. 28, 687-695.

Verhagen, A. M. W. (1966). "The notion of induced probability in statistical inference." Commonwealth Scientific and Industrial Organization, Australia, Division of Mathematical Statistics Technical Paper No. 21.

Wilkinson, G. N. (1977). "On resolving the controversy in statistical inference." J. Roy. Statist. Soc. Ser. B 39, 119-171.

Williams, J. S. (1966). "The role of probability in fiducial inference." Sankhya A 28, 271-296.

Yates, F. (1964). "Fiducial probability, recognizable subsets and Behrens' test." Biometrics 20, 343-360.

Zacks, Shelemyahu (1971). The Theory of Statistical Inference. Wiley, New York.