FIDUCIAL DISTRIBUTIONS IN FIDUCIAL INFERENCE*

By S. S. Wilks

1. **Introduction.** The essential idea involved in the method of argument now known as fiducial argument, at least in a very special case, seems to have been introduced into statistical literature by E. B. Wilson [1] in connection with the problem of inferring, from an observed relative frequency in a large sample, the true proportion or probability p associated with a given attribute. Since 1930 the ideas and terminology surrounding the fiducial method have been developed by R. A. Fisher [2, 3], J. Neyman [4, 5] and others into a system for making inferences from a sample of observations about the values of parameters which characterize the distribution of the hypothetical population from which the sample is assumed to have been drawn. The functional form of the population distribution law is assumed to be known The parameters may be means, a difference between means, variances, ranges, regression coefficients, probabilities or any other descriptive indices or combinations of indices which may be considered important in specifying the distribution function of a population. In arguing fiducially about the value of a parameter, a procedure applicable to some of the simple cases begins by the calculation from the sample of an *estimate* of the parameter in question. The values of the estimate in repeated samples of the same size will theoretically cluster "near" the true value of the parameter according to a certain distribution law which can, in general, be deduced from the functional form of the population distribution law. If the distribution of the estimate involves only the one parameter, and if, as is frequently the case, one can find a function ψ of the estimate and the parameter which has a distribution not depending on the parameter, then one is able to set up, in a rather simple manner, *fiducial limits* or a *confidence interval* for the parameter corresponding to the observed value of the estimate. The limits will depend on the particular method of calculating the estimate, the value of the estimate in the sample, and on the degree of risk of being wrong which one is willing to take in stating that the limits will include between them the value of the parameter for the population under consideration. In general the smaller the degree of risk, the wider apart will be the limits. Thus for a given pair of limits there will be an associated degree of uncertainty that the true value of the parameter is actually included between those limits. This uncertainty can be expressed by a probability α calculated from the sampling distribution of the ψ function of the parameter and estimate. Under certain conditions, one can, by simply changing variables, obtain from the ψ

* An expository paper presented to the American Statistical Association on December 28, 1937, at the invitation of the Program Committee.

distribution what has been termed by Fisher a *fiducial distribution* function of the parameter. From the fiducial distribution and for a given value of the estimate one can actually determine fiducial limits of the parameter corresponding to a given risk α. It will be seen as we proceed that the fiducial distribution plays no indispensable part in fiducial inference; the ψ function and its distribution from which the fiducial distribution is derivable, are sufficient for the fiducial argument in many cases that commonly arise in statistics. We shall discuss fiducial argument and fiducial distributions from the point of view of ψ functions.

2. **Example.** To illustrate these points let us consider an example, namely, the problem of determining fiducial limits and the fiducial distribution of the range of a rectangular distribution for a given value of the range in a sample "randomly drawn" from it.

If a sample of n individuals is drawn from a population whose distribution law is $f(x, \theta) = 1/\theta$, where only values of x between 0 and θ are considered, (that is, a rectangular distribution having range θ) the probability that the range r of the sample lies between r and $r + dr$ is $\varphi(r, \theta)\, dr$, where

$$\varphi(r, \theta) = \frac{n(n-1)}{\theta^n}(\theta - r)r^{n-2}. \tag{1}$$

Here θ is the parameter under question, and r is the estimate; r is the difference between the largest and smallest variate in the sample. Thus, for a given value of θ, say θ_0, $\varphi(r, \theta_0)$ is a sampling distribution law defined for given values of r on the range $r = 0$, to $r = \theta_0$. If we let $r/\theta = \psi$, then

$$\varphi(r, \theta)\, dr = n(n-1)(1-\psi)\psi^{n-2}\, d\psi = G(\psi)\, d\psi, \tag{2}$$

which, from a statistical point of view, shows that if we should take an aggregate of randomly drawn samples (of n items each) from rectangular populations and calculate ψ for each *sample-population combination*, then the distribution of ψ will be that given in (2). By a *sample-population combination* in this example we mean any rectangular population that may arise and a "randomly drawn" sample from it. The possible values of ψ range from 0 to 1. Thus if ψ_α is such that

$$n(n-1)\int_0^{\psi_\alpha}(1-\psi)\psi^{n-2}\, d\psi = \alpha, \quad \text{i.e.} \quad \psi_\alpha^{n-1}[n - (n-1)\psi_\alpha] = \alpha, \tag{3}$$

and if we draw a sample of n from a rectangular population, we can claim that the probability is $1 - \alpha$ that the ψ produced by this sample-population combination will satisfy the inequality

$$\psi_\alpha < \psi < 1. \tag{4}$$

It should be observed that there are many pairs of numbers, say ψ_α' and ψ_α'' such that we can claim that $\psi_\alpha' < \psi < \psi_\alpha''$, with probability $1 - \alpha$ of being correct in making the claim. ψ'_α and ψ''_α are ordinarily chosen so that the interval formed by them is as short as possible (or approximately so) in some sense. Inequality (4) is equivalent to each of the following inequalities

$$\psi_\alpha < \frac{r}{\theta} < 1, \qquad \frac{r}{\psi_\alpha} > \theta > r. \tag{5}$$

Now ψ_α can be determined from (3) when n and α are given. For example, if $\alpha = .01$ and $n = 10$, we find from (3) that $\psi_\alpha = .495$. For a given sample, the fiducial limits r/ψ_α and r can be calculated from ψ_α and the sample. It will be noticed that fiducial limits are nothing more nor less than random variables that fluctuate from sample to sample. The interval between r and r/ψ_α is called a *confidence interval* or *fiducial interval*; $1 - \alpha$ is known as the *confidence coefficient* [4] associated with the limits. Hence, in repeated samples of n from a rectangular population with range θ_0, $100(1 - \alpha)$ percent of the samples will produce fiducial limits r/ψ_α and r which include the fixed value θ_0 between them. This statement holds regardless of the value of θ_0. Hence in an aggregate of sample-population combinations, the aggregate of pairs of fiducial limits r/ψ_α and r will, in $100(1 - \alpha)$ percent of the combinations, include between them the true value of the range of the population. Furthermore, whether there is only one rectangular population for all sample-population combinations or many different rectangular populations, this statement remains true, thus showing that the method of fiducial limits for inferring the value of the parameter is independent of any *a priori* distribution of rectangular populations in an aggregate of sample-population combinations—the distribution being with respect to values of θ.

Let us look at the matter geometrically. Suppose we are drawing samples from a rectangular population with $\theta = \theta_0$. The r for each sample is represented by a dot along Or in Figure 1; corresponding to each dot there is confidence interval cutting across the V-shaped region MOR. The probability is $1 - \alpha$ that a confidence interval computed from a sample from the population having range θ_0 will cut the line θ_0K. The cutting of θ_0K by a confidence interval is equivalent to the statement that θ_0 is included between the corresponding fiducial limits.

From a practical statistical point of view what we have said has the following meaning: If on each occasion in which a randomly drawn sample of n from some rectangular population is considered, one (i) calculates the numbers r/ψ_α and r, and (ii) asserts that the range in the population producing the sample lies between these two computed limits, then in about $100(1 - \alpha)$ percent of the cases assertion (ii) will be correct (theoretically). Thus, in dealing with samples of 10 individuals from rectangular populations, one would be correct (theoretically) in about 99 percent of the cases by asserting that the population range will lie between the sample range and 2.020 $\left(= \frac{1}{.495}\right)$ times the sample range. More generally, one need not use the same value of n all the way through, provided that for the given α one evaluates ψ_α according to (3), for each n that arises. It will be seen from (3) that as n increases, the value of ψ_α tends to 1 and hence the fiducial limits r/ψ_α and r for any given sample tend to the same value, namely the sample range, thus showing that fiducial inferences about θ can be made arbitrarily certain by taking sufficiently large samples.

It is evident that the method of fiducial limits furnishes a satisfactory procedure for inferring the value of the population range θ from samples drawn from rectangular populations. Let us now go a step further and consider the fiducial distribution of θ and how it fits into the scene. The cumulative distribution of ψ is

$$\psi^{n-1}[n - (n - 1)\psi] \tag{6}$$

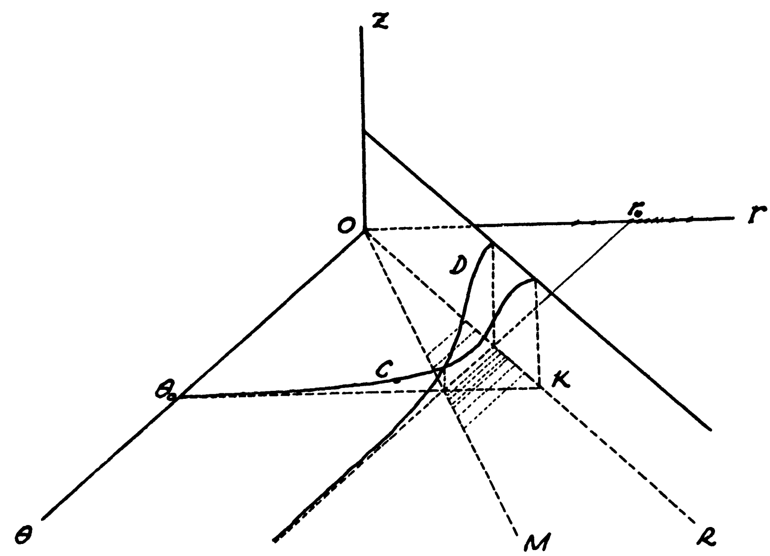

FIG. 1

and hence the cumulative distribution of r for a fixed θ, say θ_0, is

$$F(r, \theta_0) = \left(\frac{r}{\theta_0}\right)^{n-1}\left[n - (n - 1)\left(\frac{r}{\theta_0}\right)\right] \tag{7}$$

which increases from 0 to 1 as r increases from 0 to θ_0. Geometrically, $z = F(r, \theta)$ can be represented as a surface defined over the region bounded by lines $O\theta$ and OR in Figure 1, such that z is zero along O and is unity along the line OR $(r = \theta)$. $F(r, \theta)$ is continuous inside the region θOR, and for any given value $r_0 \neq 0$ of r, $F(r, \theta)$ decreases from 1 to 0 as θ increases from r_0 to ∞. The curves having the equations

$$\begin{cases} z = F(r, \theta) \\ \theta = \theta_0 \end{cases} \quad \text{and} \quad \begin{cases} z = F(r, \theta) \\ r = r_0 \end{cases}$$

(where θ_0, r_0, and α are such that $r_0/\theta_0 = \psi_\alpha$ and $F(r_0, \theta_0) = \alpha$) are the curves C and D respectively. C is the cumulative distribution of ranges of samples of n from a rectangular population with range θ_0. The curve D has the mathematical characteristics of a cumulative distribution function cumulated in the negative direction with respect to θ: its ordinates increase from 0 to 1 as θ decreases from ∞ to θ_0. Thus, if we take $-\frac{\partial}{\partial\theta} F(r_0, \theta)$ we get a function $g(\theta, r_0)$ which has the essential mathematical characteristics of a distribution function: it is non-negative, can be integrated over any interval of θ, and has total area under it equal to unity. We have

$$g(\theta, r_0) = n(n-1)\frac{r_0^{n-1}}{\theta^n}\left(1 - \frac{r_0}{\theta}\right) \tag{8}$$

and it is called the *fiducial distribution* of θ for $r = r_0$. It must be firmly pointed out that θ is not a random variable and hence $g(\theta, r_0)$ is *not* a distribution function of a random variable, although it has the mathematical properties of such a distribution. Objections have been raised to the use of the term fiducial distribution on the grounds that the thing to which it applies is not a distribution at all. However, as long as the term is carefully defined there should be no ambiguity in using it. From an analytical point of view, the problem of obtaining the fiducial distribution of θ is only a matter of changing variables for since

$$\varphi(r, \theta)\, dr = g(\theta, r)\, d\theta = n(n-1)(1-\psi)\psi^{n-2}\, d\psi \tag{9}$$

and $\psi_\alpha = r_0/\psi_0$, we have

$$\int_{\theta_0\psi_\alpha}^{\theta_0} \varphi(r, \theta_0)\, dr = \int_{r_0}^{r_0/\psi_0} g(\theta, r_0)\, d\theta = \int_{\psi_\alpha}^{1} n(n-1)(1-\psi)\psi^{n-2} d\psi = 1 - \alpha. \tag{10}$$

We remark again that

$$\int_{r_0}^{\theta_1} g(\theta, r_0)\, d\theta \tag{11}$$

is *not* to be interpreted as probability as though θ were a random variable. Instead, the meaning is as follows: Let r_0 be the range in a sample known to be from *some* rectangular population, and let the value of r_0 be inserted in (11), and let θ_1 be determined so that the value of the integral is $1 - \alpha$. The two limits for the integral are fiducial limits associated with the sample for the confidence coefficient $1 - \alpha$, which were discussed earlier. Thus, for each sample, we can compute fiducial limits using the fiducial distribution. These limits, as we have seen by considering the ψ function, fluctuate from sample to sample in such a way that the probability is $1 - \alpha$ that they will include between them the true value of the range of the population under consideration.

3. Summary of Principles. From the point of view we have taken the essential notions involved in the method of fiducial argument and fiducial distributions for the case of a continuous variate and one parameter can be readily abstracted from the example just discussed. In general, we have the following steps:

(a) A sample is assumed to be *randomly drawn* from a population with a distribution of *known* functional form $f(x, \theta)$, θ being a parameter. Let $x_1, x_2, \cdots, x_n$ be the values of x in the sample.

(b) A function, say $\psi(x_1, \cdots, x_n, \theta)$ of the sample x's and θ is devised so that its sampling distribution $G(\psi)$ involves θ and the x's only as they enter into ψ. The value of θ in ψ is that for the population from which the sample is actually drawn.

(c) Two numerical values of ψ, say ψ'_α and ψ''_α are chosen (ordinarily as close together as possible) so that the probability computed from $G(\psi)$ is $1 - \alpha$ (e.g. 0.95) that ψ will lie between ψ'_α and ψ''_α—more briefly $P(\psi'_\alpha < \psi < \psi''_\alpha) = 1 - \alpha$.

(d) The inequality $\psi'_\alpha < \psi < \psi''_\alpha$ which contains only one unknown, namely θ, is solved for θ giving the equivalent inequality $\underline{\theta} < \theta < \bar{\theta}$ where θ and $\bar{\theta}$ are *fiducial limits* and are subject to sampling fluctuations.

(e) The expression $P(\psi'_\alpha < \psi < \psi''_\alpha) = 1 - \alpha$ is replaced by the equivalent expression $P(\underline{\theta} < \theta < \bar{\theta}) = 1 - \alpha$ which states that the probability is $1 - \alpha$ that a sample will yield values $\underline{\theta}$ and $\bar{\theta}$ which will include the true value of θ between them.

(f) The differential element for the fiducial distribution of θ is $G(\psi) \left| \dfrac{\partial\psi}{\partial\theta} \right| d\theta$ (provided $\partial\psi/\partial\theta$ is a function of θ which does not change sign for a given sample of x's) and is obtained by letting θ be the variable in $G(\psi)\, d\psi$, keeping the x's fixed.

To give precisely the conditions under which all of these steps can be performed is a technical matter which will not be considered here. It is sufficient to remark that they can be performed in many cases of practical interest. Fiducial argument can be carried on using only the first five steps without introducing the notion of a fiducial distribution. In connection with step (a) it should be particularly noticed that the functional form $f(x, \theta)$ of the population under question is assumed to be known and that the sample under consideration is "randomly drawn" from the population. Thus, in applying the theory to practical problems it is a matter of fundamental importance that these two assumptions be valid. In cases where a sufficient amount of data exists, it can usually be satisfactorily tested by using the χ^2 test and other devices, whether or not a given functional form for $f(x, \theta)$ is a valid assumption. In cases where sufficient data do not exist for actually making such a test justification for assuming a given function form usually has to be made on the basis of theoretical considerations. From a practical point of view the notion of randomness is characterized by methods of drawing samples rather than *a posteriori* mathematical considerations of the sample after it has been drawn, and thus the question of randomly drawing samples depends largely upon the experience and sound judgment of the experimenter. However, after one or more samples have been drawn "at random," the problem of arguing from them about the populations from which they were drawn is largely mathematical.

4. **Case of large samples.** For a population with a distribution of known functional form, a fiducial distribution of the parameter clearly depends on the size of the sample and the particular estimate used. For example, in large samples, we would get a fiducial distribution of the mean of a normal population of known variance by using the sample mean which would be different from the one obtained using the median of the sample. In order to be able to make the inferences about θ as accurate as possible, a ψ function should theoretically be used which will produce fiducial limits which are closest together, on the average, or perhaps "best" in some other sense, for a given α. The fiducial distribution obtainable from such a ψ could then be referred to as the "best" fiducial distribution, and theoretically it should be used in preference to other possible fiducial distributions if fiducial distributions are to be used at all to set fiducial limits. In large samples from a population with a distribution function $f(x, \theta)$, it is known [6] that, under rather general conditions, fiducial limits which are closest together on the average can be obtained by letting

$$\psi = \frac{1}{\sqrt{n}}\left(\frac{\partial L}{\partial \theta}\right)\left[E\left\{\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2\right\}\right]^{-\frac{1}{2}} \tag{12}$$

and treating ψ as a normally distributed variate with zero mean and unit variance, where $L = \sum_{i=1}^{n} \log f(x_i, \theta)$, the logarithm of the likelihood of θ for the given sample, $x_1, x_2, \cdots, x_n$ are values of x in the sample, and E denotes mathematical expectation. For example, in the case of a binomial population where each individual belongs either to class A or class B, we have $f(x, \theta) = \theta^x(1 - \theta)^{1-x}$ where θ is the probability associated with class A, x will be 0 or 1 according to whether an individual belongs to B or A. In a sample of n individuals, $L = m \log \theta + (n - m) \log (1 - \theta)$, where m is the number of individuals in class A. $E\left\{\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2\right\} = \frac{1}{\theta(1 - \theta)}$, and we get $\psi = \frac{m - n\theta}{\sqrt{n\theta(1 - \theta)}}$. If we should want to find fiducial limits of θ for a confidence coefficient of .95, we would solve (1) the equations $\frac{m - n\theta}{\sqrt{n\theta(1 - \theta)}} = \pm 1.96$ for θ, thus getting two values of θ, say $\underline{\theta}$ and $\bar{\theta}$. We can then say that $\underline{\theta}$ and $\bar{\theta}$ will include the true value of θ between them with a probability of .95 of being correct, in the sense that if we applied this rule consistently to samples from binomial populations, we would have a procedure that would lead to a correct statement in about 95 percent of the cases (theoretically).

To illustrate the difference between the fiducial method and the commonly used method of placing limits on θ for $P = .95$, consider an example in which $m = 150$, $n = 400$. The usual procedure is to replace θ by m/n in $\theta \pm 1.96 \sqrt{\frac{\theta(1-\theta)}{n}}$, which yields .311 and .431. The fiducial procedure is to solve the equation $\frac{m - n\theta}{\sqrt{n\theta(1-\theta)}} = \pm 1.96$, for θ, thus obtaining .312 and .455. For the case of small samples, the problem of getting "best" fiducial limits becomes more complicated [5].

5. **Extensions of Fiducial Argument**. It will be observed that it is not necessary for ψ to be a function of only one statistic and θ in order to be able to argue fiducially about θ. For example, if a sample of n is drawn from a normal population with mean θ, it is well known that if $\bar{x}$ is the sample mean then

$$\psi = \frac{(\bar{x} - \theta)\sqrt{n(n-1)}}{\left[\sum_1^n (x_i - \bar{x})^2\right]^{\frac{1}{2}}} \tag{13}$$

(which is Fisher's t function), has the distribution

$$\frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}(n-1))\sqrt{\pi(n-1)}} \frac{d\psi}{[1 + \psi^2/(n-1)]^{\frac{1}{2}n}}. \tag{14}$$

Here ψ is a function of two statistics, namely $\bar{x}$ and $\sum_{i=1}^n (x_i - \bar{x})^2$, and the fiducial distribution of θ for this ψ function is obtained at once by applying rule (f).

The ideas of fiducial argument may be extended in other directions, but these cannot be considered in any detail here. For example ψ may be a function of $x_1, \cdots, x_n$ and two or more population parameters, in which case one could set up fiducial *regions* for the several parameters. From a practical point of view, the fiducial argument for two or more parameters simultaneously, had hardly been touched. Again ψ may be a function of statistics from two samples, one observed and the other not yet observed, and not involving population parameters, at all, in which case one can argue fiducially about the statistic in question for the unobserved sample [3]. The notion of a fiducial distribution has been extended to several parameters taken simultaneously [3, 7], but the problem of working out relations between fiducial distributions of several parameters and fiducial regions is yet to be investigated. The principles may be readily applied in situations in which the x's involved in ψ take on discrete values. In this case the equality signs in the probability expressions in steps (c) and (d) would be replaced by greater than or equal signs ($\geq$). Two excellent examples of the application of principles of fiducial argument to the discrete case are furnished: (i) by a paper by Pearson and Clopper [8] on fiducial limits of the probability P from samples from a binomial population, and (ii) by a paper by Ricker [9] on fiducial limits of m in the Poisson distribution $f(x, m) = m^x e^{-m}/x!$.

REFERENCES

[1] E. B. Wilson, "Probable Inference, the Law of Succession, and Statistical Inference," *Jour. American Statistical Association*, Vol. 22 (1927), pp. 209–212.

[2] R. A. Fisher, "The Concepts of Inverse Probability and Fiducial Probability Referring to Unknown Parameters," *Proc. Royal Society of London*, Series A, Vol. 139 (1933), pp. 343–348.

[3] R. A. Fisher, "The Fiducial Argument in Statistical Inferences," *Annals of Eugenics*, vol. 6 (1935), pp. 391–398.

[4] J. Neyman, "On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection," *Royal Statistical Society*, vol. 97, 1934, pp. 558–625.

[5] J. Neyman, "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability," *Phil. Trans. Roy. Soc. London*, Series A, Vol. 236 (1937), pp. 333–380.

[6] S. S. Wilks, "Shortest Average Confidence Intervals from Large Samples," *Annals of Mathematical Statistics*, Vol. 9 (1938), pp. 166–175.

[7] I. E. Segal, "Fiducial Distribution of Several Parameters with Application to a Normal System," *Proc. Cambridge Phil. Soc.*, vol. 34 (1938), pp. 41–47.

[8] C. J. Clopper and E. S. Pearson, "The Use of Confidence or Fiducial Limits in the Case of the Binomial," *Biometrika*, vol. 24, 1934, pp. 404–413.

[9] William E. Ricker, "Fiducial Limits of the Poisson Frequency Distribution," *Jour. American Statistical Association*, vol. 32, 1937, pp. 349–356.

Princeton University.