Fiducial Limits of the Parameter of a Discontinuous Distribution

Author(s): W. L. Stevens

Source: *Biometrika*, Jun., 1950, Vol. 37, No. 1/2 (Jun., 1950), pp. 117-129

Published by: Oxford University Press on behalf of Biometrika Trust

Stable URL: https://www.jstor.org/stable/2332154

FIDUCIAL LIMITS OF THE PARAMETER OF A DISCONTINUOUS DISTRIBUTION

By W. L. STEVENS*

1. The problem

1·1. It has long been realized that methods of interval estimation which can be used successfully in the case of continuous distributions encounter a special kind of difficulty when we attempt to apply them to a discontinuous distribution such as the binomial or the Poisson. This difficulty has been expressed in different ways by different writers, their choice of words being mainly influenced by whether they are using Fisher's concept of fiducial probability or Neyman's of confidence intervals. While these approaches are said to be fundamentally different, their applications to the uniparametric case are so closely linked that an argument in terms of the one can usually be translated in terms of the other. We shall accordingly not hesitate to express our results in the languages of both theories.

We may begin with the simplest example of a discrete distribution—the binomial. If the probability of an event is π, the probability that it will occur f times in n trials is

$$g(f, \pi) = \frac{n!}{(n-f)!\,f!}(1-\pi)^{n-f}\pi^{f}. \tag{1·11}$$

The probability that it will happen f *or more* times is

$$G(f, \pi) = \sum_{r=f}^{n} g(r, \pi). \tag{1·12}$$

If in a given experiment it did in fact happen f times, we may define $\pi_0(f)$, corresponding to any probability level, P_0, as the root of the equation

$$G(f, \pi_0) = P_0. \tag{1·13}$$

When $f = 0$, we define $\pi_0(0) = 0.$

Then it can readily be shown that if $f \neq 0$, and if

$$\pi < \pi_0,$$

then $$G(f, \pi) < P_0.$$

Thus any hypothesis which puts $\pi < \pi_0$ is rejected at the P_0 level of significance in favour of a higher value of π. Hence π_0 is a lower limit to π and may conveniently be termed the lower P_0 limit of π. At the chosen level of significance, we accordingly write either

$$\pi > \pi_0 \quad \text{or} \quad \pi \geqslant \pi_0.$$

Similarly, we may define $\pi_1(f)$, the upper P_1 limit of π, as the root of the equation

$$G(f+1, \pi_1) = 1 - P_1, \tag{1·14}$$

implying that P_1 = probability that r is f *or fewer*.

When $f = n$, we define $\pi_1(n) = 1$. We then write either $\pi < \pi_1$ or $\pi \leqslant \pi_1$. Combining the two inequalities, we therefore have a statement, either

$$\pi_0 < \pi < \pi_1 \quad \text{or} \quad \pi_0 \leqslant \pi \leqslant \pi_1$$

* Universidade de São Paulo, Faculdade de Ciências Económicas, Rua Vila Nova, 268, São Paulo, Brazil.

corresponding to any pair of probability levels, subject to the condition that $P_0+P_1<1$. (P_0 and P_1 are usually but not necessarily equal.)

Notice that in calculating either limit the probability of the actual result, f observed in n, is included. This probability is of course finite, whereas if the observations were drawn from a continuous distribution the probability of the actual result would be an infinitesimal. It is from this difference that all the trouble springs.

1·2. *Charts and tables.* Charts of the limits of π as functions of n and f/n for $P_0 = P_1 = \frac{1}{2}\,\%$ and $2\frac{1}{2}\,\%$ were published by Clopper & Pearson (1934). The present writer noticed that if we tabulate $n\pi$ instead of π and use f and f/n instead of n and f/n, we can greatly facilitate interpolation, thus covering in a condensed table all the region where the normal approximation is unsatisfactory. Our table, which covers the range $f = 0$–15, $f/n = 0$–1 and P_0 or $P_1 = 10$, $2·5$ and $0·5\,\%$ appeared in the second edition of Fisher & Yates.

1·3. *The two fiducial distributions.* Let us now see what happens when we attempt to define the fiducial distribution of π. Differentiating $G(f,\pi)$ with respect to π, we have

$$h_0(\pi)\,d\pi = \frac{\partial G(f,\pi)}{\partial \pi}\,d\pi = \frac{n!}{(n-f)!\,(f-1)!}(1-\pi)^{n-f}\,\pi^{f-1}\,d\pi. \tag{1·31}$$

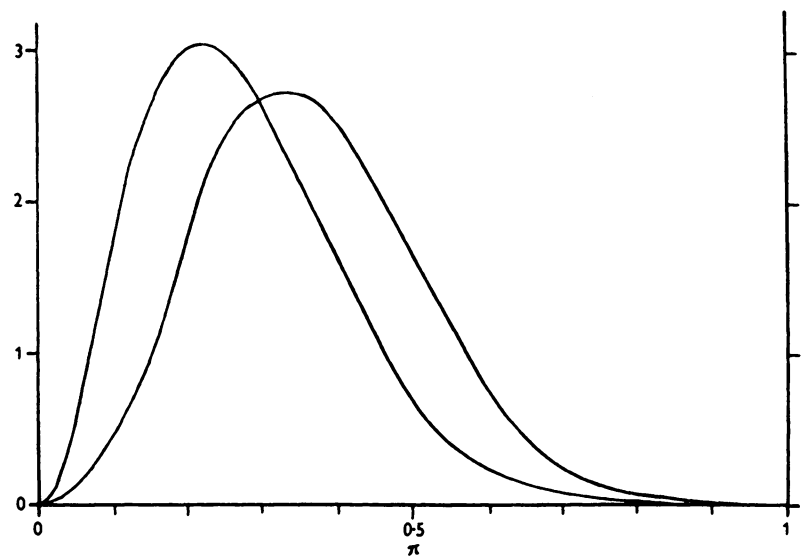

Fig. 1. The pair of fiducial distributions associated with '3 in 10'.

This is an admissible probability distribution, since its integral between 0 and 1 is unity. (See left curve in Fig. 1 for the case $n = 10$ and $f = 3$.) Moreover, if we write

$$\int_0^{\pi_0} h_0(\pi)\,d\pi = P_0, \tag{1·32}$$

the value of π_0 satisfying this equation is in fact the lower P_0 limit of π. If only we could also obtain the upper P_1 limit by solving

$$\int_0^{\pi_1} h_0(\pi)\,d\pi = 1-P_1, \tag{1·33}$$

then we could legitimately call $h_0(\pi)$ the fiducial distribution of π. But, of course, we can do no such thing. We have to define a second distribution of π,

$$h_1(\pi)\,d\pi = \frac{\partial G(f+1,\pi)}{\partial \pi}\,d\pi = \frac{n!}{(n-f-1)!\,f!}(1-\pi)^{n-f-1}\,\pi^f\,d\pi. \qquad (1\cdot34)$$

(See right curve in Fig. 1.) The upper P_1 limit of π is then the π_1 which satisfies the equation

$$\int_0^{\pi_1} h_1(\pi)\,d\pi = 1-P_1. \qquad (1\cdot35)$$

Thus we see that the unique fiducial distribution obtainable from a continuous distribution (where a sufficient statistic exists) becomes resolved in the discrete case into a *pair* of distributions.

It is a matter of taste, up to this point in the argument, whether we choose to say that there is no fiducial distribution associated with the result, f in n, or whether we say that each result has associated with it a unique pair of fiducial distributions. At any rate, these two between them perform the same service as is performed by the unique fiducial distribution in the continuous case, namely, that of supplying limits to the parameter, corresponding to any pair of significance levels. The only extra rule we need is the simple one, that when finding lower limits we use $h_0(\pi)$ and when upper limits, $h_1(\pi)$.

1·4. *Fisher's discussion of the problem.* The peculiar nature of the fiducial argument when applied to discontinuous data has been expressed differently by Fisher (1935). He was there discussing a fourfold table, showing twins of convicts classified as convicted or not-convicted and monozygotic or dizygotic. The relevant parameter is the cross-product of the cellular probabilities

$$\psi = \frac{\pi_1\pi_4}{\pi_2\pi_3},$$

but it is evident that all his remarks are applicable to the example of the binomial distribution. The passage must be quoted:

> ...We may thus infer that the observations differ significantly at the 1 % level of significance, from any hypothesis which makes ψ greater than 0·48 .... This is not a probability statement about ψ. It is a formally precise statement of the results of applying tests of significance. If, however, the data had been continuous in distribution, on the hypothesis considered, it would have been equivalent to the statement that the fiducial probability that ψ exceeds 0·48 is just one chance in a hundred. With discontinuous data, however, the fiducial argument only leads to the result that this probability does not exceed 0·01. We have a statement of inequality, and not one of equality. It is not obvious in such cases, that, of the two forms of statement possible, the one explicitly framed in terms of probability has any practical advantage. The reason why the fiducial statement loses its precision with discontinuous data is that the frequencies in our table make no distinction between a case in which two dizygotic convicts were only just convicted, perhaps on venial charges, or as first offenders, while the remaining 15 had characters above suspicion, and an equally possible case in which the two convicts were hardened offenders, and some at least of the remaining 15 had barely escaped conviction. If we knew where we stood in the range of possibilities represented by these two examples, and had similar information with respect to the monozygotic twins, the fiducial statements derivable from the data would regain their exactitude.

The explanation here advanced for the loss of precision in the fiducial statement may strike the reader as somewhat metaphysical. Although one can think of a continuous phenomenon underlying the discrete classification into convicted and not-convicted, yet in other examples such a notion would at once ring a false note. For example, if the classification is made according to whether the individual possesses or not a certain gene at a certain locus, there can surely be no underlying phenomenon representing all gradations between possessing and not possessing the gene? Be that as it may, it is curious that the method which we are going to develop can be linked up in a striking way with Fisher's concept of underlying continuity.

1·5. *Confidence intervals for π*. Let us next see how the peculiarity of the discontinuous distribution problem reveals itself when we write in terms of confidence intervals. Let us suppose that we make it a rule always to calculate the lower P_0 limit and then to state that

$$\pi > \pi_0 \quad \text{or} \quad \pi \geqslant \pi_0.$$

In following either rule, we shall sometimes make a true statement and sometimes a false one. We can therefore define

$$\text{Prob}\{\pi > \pi_0\} \quad \text{or} \quad \text{Prob}\{\pi \geqslant \pi_0\}$$

as the limit of

$$\frac{\text{number of times the statement is correct}}{\text{number of applications of the rule}}$$

as the number of applications tends to infinity.

As soon as the confidence interval argument was applied to the binomial distribution, it was realized, and in fact pointed out in Clopper & Pearson, that

$$\text{Prob}\{\pi > \pi_0\} \geqslant 1 - P_0 \quad \text{and} \quad \text{Prob}\{\pi \geqslant \pi_0\} > 1 - P_0. \tag{1·51}$$

This probability is a function of the parameter π. The graph of the complementary function, $\text{Prob}\{\pi \leqslant \pi_0\}$ for the case $n = 10$ and $P_0 = 10\%$, is given in Fig. 2.

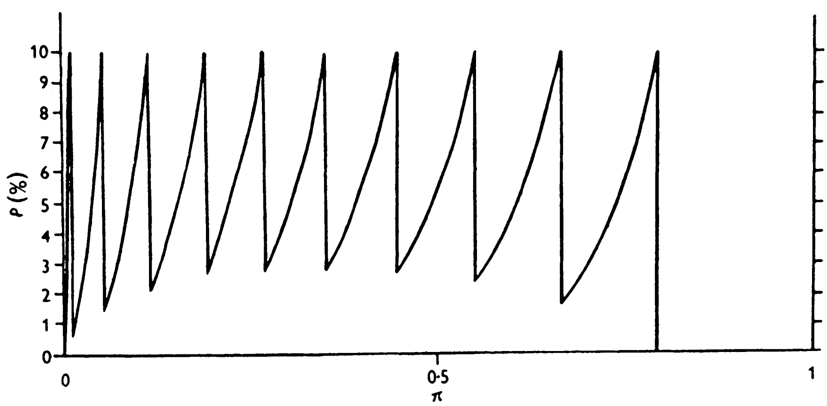

Fig. 2. Probability $\{\pi < \pi_0\}$ when $n = 10$.

The curve shows discontinuities at the n points,

$$\pi_0(1), \quad \pi_0(2), \quad \ldots, \quad \pi_0(n),$$

and it is only at these points that $\text{Prob}\{\pi \leqslant \pi_0\} = P_0$ and hence $\text{Prob}\{\pi > \pi_0\} = 1 - P_0$. Elsewhere, $\text{Prob}\{\pi \leqslant \pi_0\}$ can differ very gravely from 10%. Leaving out the extreme cases of $f = 0$ and $f = 10$ where we might have expected trouble, we are still left with a variation from $\frac{1}{2}$ to 10%.

Similarly, it can be shown that, having calculated an upper P_1 limit, π_1, then

$$\text{Prob}\{\pi < \pi_1\} \geqslant 1 - P_1 \quad \text{and} \quad \text{Prob}\{\pi \leqslant \pi_1\} > 1 - P_1. \tag{1·52}$$

The equality sign holds only at the n points of discontinuity,

$$\pi_1(0), \quad \pi_1(2), \quad \ldots, \quad \pi_1(n-1).$$

Combining (1·51) and (1·52), we conclude that

$$\left.\begin{aligned} &\text{Prob}\{\pi_0 < \pi < \pi_1\} \geqslant 1 - P_0 - P_1 \\ \text{and} \qquad &\text{Prob}\{\pi_0 \leqslant \pi \leqslant \pi_1\} > 1 - P_0 - P_1. \end{aligned}\right\} \tag{1·53}$$

We can say further that there 'always'* exists a positive quantity D greater than zero, such that

$$\text{Prob}\{\pi_0 < \pi < \pi_1\} > 1 + D - P_0 - P_1. \tag{1·54}$$

Now the use of $1 - P_0 - P_1$ (the *confidence coefficient*) as a measure of our confidence in the assertion, $\pi_0 < \pi < \pi_1$, has been justified by two lines of argument: (*a*) that in the limit, when n is large, the inequality sign may be replaced by the equality sign, and (*b*) that the calculated confidence coefficient is necessarily conservative, i.e. on the 'safe side', because the true probability of π being between the limits is 'always' greater than is stated. The first type of argument need not be taken too seriously, since it amounts to no more than saying that any small-sample test tends in the limit to the corresponding large-sample test, which, although true, would hardly be a reason for ignoring the small sample test if it were available. The second type of justification is a much more serious matter and, in the view of the present writer, is completely invalid.

It is the very basis of any theory of estimation, that the statistician shall be permitted to be wrong a certain proportion of times. Working within that permitted proportion, it is his job to find a pair of limits as narrow as he can possibly make them. If, however, when he presents us with his calculated limits, he says that his probability of being wrong is less than his permitted probability, we can only reply that his limits are unnecessarily wide and that he should narrow them until he is running the stipulated risk. Thus we reach the important, if at first sight paradoxical conclusion, that *it is a statistician's duty to be wrong* the stated proportion of times, and failure to reach this proportion is equivalent to using an inefficient in place of an efficient method of estimation.

Of course, this criticism begs the fundamental question of whether we can, while still not exceeding the stipulated probabilities of being wrong at either limit, find an interval lying within the interval found by the usual method. We shall now show that we can do this and more. In fact, it is possible to find a pair of limits π_0 and π_1 such that

$$\begin{aligned} \text{Prob}\{\pi \geqslant \pi_0\} &= 1 - P_0, \\ \text{Prob}\{\pi \leqslant \pi_1\} &= 1 - P_1, \\ \text{and hence} \qquad \text{Prob}\{\pi_0 \leqslant \pi \leqslant \pi_1\} &= 1 - P_0 - P_1, \end{aligned} \tag{1·55}$$

and these limits will 'always' be wholly inside the limits customarily calculated. This means that we can recover for the discontinuous problem the exact type of fiducial or confidence interval statement which is possible in continuous problems.

* We shall use inverted commas and write 'always' when we mean that the probability of the event is unity but that it need not happen. Similarly, we shall say an event 'never' happens when it can happen but its probability is zero. Thus if x is drawn at random from the rectangular distribution $0 \leqslant x < 1$, we say that x is 'always' greater than zero, 'never' equal to zero, but never equal to unity.

2. The solution

2·1. Suppose that in n trials the event has occurred f times. Then define a variable y by the equation

$$y = f + x, \tag{2·11}$$

where x is any number chosen at random from the rectangular distribution, $0 \leqslant x < 1$. (This means, in practice, that x is a decimal point followed by a sequence of digits taken from a table of random numbers.) The quantity y is accordingly any real number in the range $0 \leqslant y < n+1$. We may also note that y uniquely determines f and x, since f is the whole and y the fractional part of y.

Since y is the sum of two quantities, each drawn from a known distribution, it follows that its own distribution is determinate, and can in fact be found without difficulty. Now since y uniquely determines f, it is evidently a sufficient statistic for π. We therefore need not hesitate to apply the fiducial probability argument. Thus, if

$$y = f + x$$

and

$$y_0 = f_0 + x_0,$$

$$\begin{aligned} \text{Prob}\{y \geqslant y_0\} &= \text{Prob}\{f > f_0\} + \text{Prob}\{f = f_0\}\,\text{Prob}\{x \geqslant x_0\} \\ &= G(f_0+1, \pi) + g(f_0, \pi)(1 - x_0) \\ &= x_0 G(f_0+1, \pi) + (1-x_0)\,G(f_0, \pi). \end{aligned} \tag{2·12}$$

Differentiating this probability with respect to π, and dropping the suffices on f and x, we thus define a distribution function

$$h_x(\pi)\,d\pi = x h_0(\pi)\,d\pi + (1-x)\,h_1(\pi)\,d\pi, \tag{2·13}$$

where $h_0(\pi)$ and $h_1(\pi)$ are the functions already defined in (1·31) and (1·34).

We propose to call $h_x(\pi)\,d\pi$ the fiducial distribution of π. Notice, however, that it is defined only after we have drawn our value of x. Given merely the number f, then there exists an infinite continuum of such distributions ranging between the pair of distributions $h_0(\pi)\,d\pi$ and $h_1(\pi)\,d\pi$ which appeared in the earlier discussion. Now it is essential that the fiducial distribution of a parameter should be unique. This, however, can be assured by the following simple rule:

When any experiment has been performed (or series of observations taken), the investigator is allowed once and once only to select, at random, his value of x; the distribution thus determined will be called *the* fiducial distribution of π, and neither he nor anyone else is permitted another drawing of the number x.

As a result of this rule, we find that a fiducial distribution is uniquely determined for each experiment, though not for each value of f. Now it will readily be seen that the fiducial limits found by integrating $h_x(\pi)\,d\pi$ 'always' lie wholly inside the fiducial limits found by the usual method. Moreover, since y is a continuous variable, the statements in fiducial probability become statements of equality instead of statements of inequality.

2·2. *Demonstration.* To clarify these remarks, let us first examine the functional relation between y and π. Corresponding to any probability P (for a lower limit) or $1-P$ (for an upper limit), we have the following equation relating y and π:

$$xG(f+1, \pi) + (1-x)\,G(f, \pi) = P. \tag{2·21}$$

Differentiating totally with respect to x, we get

$$\frac{dP}{dx} = \frac{\partial P}{\partial x} + \frac{\partial P}{\partial \pi}\frac{d\pi}{dx} = 0.$$

Hence $$-g(f)+h(\pi)\frac{d\pi}{dx}=0. \tag{2·22}$$

Hence $d\pi/dx$ is everywhere finite and positive.

However, if $y<1-P$, i.e. if $f=0$ and $x<1-P$, the equation becomes

$$(1-\pi)^n=(1-P)/x,$$

thus yielding a negative value for π. We must therefore define π as zero, when $y<1-P$. Similarly, we define π as unity when $y>n+1-P$.

Next we note that as $x\to 1$ the equation tends to

$$G(f+1,\pi)=P,$$

which is the equation appropriate to $y=f+1$. Hence π is a continuous function of y, although the derivative $d\pi/dy$ is discontinuous at the integral values of y. The graph of π against y is in fact a series of ascending loops, like a telegraph wire going up a mountain side. We conclude therefore that y uniquely determines π and that π, if not equal to zero or unity, also uniquely determines y. Under this same restriction, we also see that $\pi(y)$ and $y(\pi)$ are ever-increasing functions of each other. Moreover, since the lower limit, as usually calculated, can be found from the method here proposed by putting $x=0$, it follows that the new method 'always' yields a higher value for the lower limit. Similarly, it can be shown that it yields a lower value for the upper limit.

Now let π $(0<\pi<1)$ represent the true parametric value, and let $y=y(\pi)$ be the corresponding y at a chosen level P_0. Suppose we draw at random a sample of n observations and encounter the event f_0 times in the sample. To this f_0 we add our randomly chosen x_0 in order to produce y_0, which then uniquely determines π_0, the lower P_0 limit. In repeated applications of this procedure, we may accordingly define

$$\text{Prob}\{\pi_0<\pi\}.$$

However, $\pi_0<\pi$ implies $y_0<y$. Hence

$$\text{Prob}\{\pi_0<\pi\}=\text{Prob}\{y_0<y\}=1-\text{Prob}\{y_0\geqslant y\}=1-P_0. \tag{2·23}$$

Hence $\text{Prob}\{\pi_0<\pi\}$, in the sense of the Neyman-Pearson theory, is *exactly* equal to $1-P_0$. Since y is continuous it is immaterial whether we write

$$\text{Prob}\{\pi_0<\pi\}=1-P_0 \quad \text{or} \quad \text{Prob}\{\pi_0\leqslant\pi\}=1-P_0.$$

Similarly, if π_1 is the upper P_1 limit, then

$$\text{Prob}\{\pi<\pi_1\}=\text{Prob}\{\pi\leqslant\pi_1\}=1-P_1. \tag{2·24}$$

Combining the two statements, we therefore have

$$\text{Prob}\{\pi_0<\pi<\pi_1\}=\text{Prob}\{\pi_0\leqslant\pi\leqslant\pi_1\}=1-P_0-P_1. \tag{2·25}$$

These equations hold for all values of π in the open interval $0<\pi<1$. Is it possible to close the interval? Yes, if we adopt the following rules:

$$\begin{array}{llll}
\text{When} & \pi_0=0, & \text{we must write} & 0\leqslant\pi. \\
\text{When} & \pi_0=1, & \text{we must write} & 1<\pi. \\
\text{When} & \pi_1=0, & \text{we must write} & \pi<0. \\
\text{When} & \pi_1=1, & \text{we must write} & \pi\leqslant 1.
\end{array}$$

Under these rules, it can be shown that the equations are true, even for $\pi=0$ and $\pi=1$. However, the solution is a little artificial, since in two instances we have to make statements which we know to be false. Perhaps the proper answer is that in practice we can rule out the possibility of either $\pi = 0$ or $\pi = 1$, for if, in fact, we thought that π was small and possibly zero (or large and possibly unity) we would have employed a different experimental technique, namely that of inverse sampling.

2·3. *Alternative solutions.* Once the idea of introducing a random variate has been admitted, one may inquire whether it need have been drawn from a rectangular distribution. Could it not have been drawn from some other distribution, and if so, could we not find some distribution which would result in yet narrower limits?

Suppose z to be drawn from *any* continuous distribution, $j(z)\,dz$, over *any* range. Then we may consider the joint distribution of the pair of sampling variables, f and z. Let us say, conventionally, that

$$(f, z) > (f_0, z_0)$$

means that either

$$f > f_0$$

or

$$f = f_0 \quad \text{and} \quad z > z_0.$$

From this we have

$$\text{Prob}\,\{(f,z) > (f_0, z_0)\} = \int_{-\infty}^{z_0} j(z)\,dz\, G(f_0+1, \pi) + \int_{z_0}^{\infty} j(z)\,dz\, G(f_0, \pi). \tag{2·31}$$

Differentiating with respect to π and removing the suffices, we would again have a fiducial distribution of π, which might be offered as an alternative to $h_x(\pi)\,d\pi$.

However, a moment's thought will show that this method yields the same solution in a different guise. In order to draw z_0 at random from $j(z)\,dz$ we would have to select x_0 at random from the rectangular distribution $0 \leqslant x_0 < 1$, and then solve the equation

$$x_0 = \int_{-\infty}^{z_0} j(z)\,dz. \tag{2·32}$$

Replacing the integrals in equation (2·31), we obtain

$$\text{Prob}\,\{(f,z) > (f_0, z_0)\} = x_0 G(f_0+1, \pi) + (1-x_0)\, G(f_0, \pi). \tag{2·33}$$

This is the same equation as we used before (2·12). Hence we would obtain the same fiducial distributions with the same frequency, the only difference being that they would be differently labelled.

On the other hand, let us consider what happens when the distribution from which we draw z contains points of concentration, e.g. a point $z = Z$, where $\text{Prob}\,\{z = Z\} = \alpha$, a finite quantity. If then we write

$$x_0 = \int_{-\infty}^{z_0} j(z)\,dz,$$

there will be a one-one relation between x_0 and z_0 everywhere except when $z_0 = Z$, in which point there is a series of values of x_0, $X \leqslant x_0 \leqslant X + \alpha$, corresponding to the one value of z_0. As a result, we find that in this range of x_0, the lower limit remains constant at the value $\pi(f+X)$, i.e. instead of continuing to rise smoothly, the graph of π_0 against x_0 runs horizontally for a space. Within this interval, it is evident that the calculated lower limits will 'always' be below those found from the rectangular distribution. It can also be shown that the inequality sign would have to be returned to the fiducial probability statement if the distribution had points of concentration.

Since we have now shown that alternative methods will either yield the same result, or a result which is in two respects less satisfactory, we conclude that we were justified in calling $h_x(\pi)\,d\pi$ *the* fiducial distribution of π.

2·4. '*Underlying continuity.*' We have seen that the method here proposed replaces integral values of f in the range $0 \leqslant f \leqslant n$, by the continuous variable y in the range $0 \leqslant y < n+1$. We notice at once the link with Fisher's notion (quoted earlier) of the discrete variable being merely an incomplete manifestation of a continuous phenomenon. At the same time, we must emphasize that the method does not imply, nor does it rest on any idea of, underlying continuity. This is proved by the discussion of the preceding section, which showed that the x could be drawn from any continuous distribution. Nevertheless, it is interesting that the method, in its simplest form, is expressible in terms of a continuous variable over the range of f.

One should also notice here a similarity with a method proposed by Fisher (1935) for recovering exact fiducial statements when estimating bacterial density by the dilution method. If the dilution ratio is r, Fisher showed that the fiducial statements may be made exact if the first dilution ratio is taken as r^x instead of r, where x is drawn at random between 0 and 1. There is, however, this essential difference: in Fisher's example a random element is introduced before the experiment is performed, whereas here we propose the introduction of a random element after the experiment is completed.

2·5. *Choice of the x for different limits.* A question which may be asked is whether, when calculating a number of limits, we should select the x once and use it for all limits, or whether we can select a different x for each limit. For the purpose of defining a fiducial distribution, it is obvious that x must be chosen once only. On the other hand, if we wish to determine one or more *confidence intervals*, we may choose, at random, a fresh x for every limit. The intervals so determined will still be genuine confidence intervals in the sense used by Neyman. Nevertheless, it should not be overlooked that such a procedure can lead to paradoxical results. Thus we might find a 90 % confidence interval lying wholly inside an 89 % confidence interval, i.e. we would have to attach more confidence to the more precise statement. Paradoxical though it may sound, the result would be genuine; one would, however, be obliged to make both statements in order to maintain the right proportions of true statements in the long run. Still one would not like to be the statistician who is called upon to explain such a conundrum to a factory manager who has merely been inquiring into the percentage of defective articles in production! Perhaps it would be wise to adopt the rule that the same x is to be used for all limits (in any given experiment), thus bringing the confidence intervals into line with the fiducial intervals.

3. APPLICATION

3·1. There is no particular obstacle, except time, to tabulating the proposed solution, as applied to the problem of the binomial distribution, or indeed to any other discontinuous distribution. Table 1 gives a sample tabulation for the binomial with index = 10. Since the fiducial distribution is now unambiguous, the percentile points are more logically labelled continuously from 0 to 100 %. The upper 10 % limit is of course tabled under 90 %.

Thus suppose 3 events were encountered in 10 trials. A table of random numbers begins: 337 .... We round this to 34 and enter the table with $f = 3$ and $x = 0{\cdot}34$, finding the upper and lower 10 % limits

$$13{\cdot}1 < \pi < 49{\cdot}4 \quad (\%).$$

For comparison, the usual method gives the wider limits,

$$11{\cdot}6 < \pi < 55{\cdot}2 \quad (\%).$$

Table 1. *Fiducial limits of the parameter, π (%), of a binomial distribution*

Ten, fifty and ninety percentile points for $n = 10$ and $f = 0$–10.

x	$f=0$			$f=1$			$f=2$			x
	10	50	90	10	50	90	10	50	90	
0·0	0	0	0	1·05	6·7	20·6	5·45	16·2	33·7	0·0
0·1	0	0	0	1·16	7·5	22·5	5·73	17·1	35·3	0·1
0·2	0	0	6·7	1·31	8·4	24·3	6·07	18·0	36·8	0·2
0·3	0	0	10·4	1·48	9·3	26·0	6·46	19·0	38·1	0·3
0·4	0	0	13·0	1·72	10·2	27·5	6·91	19·9	39·4	0·4
0·5	0	0	14·9	2·01	11·1	28·9	7·4	20·9	40·6	0·5
0·6	0	1·79	16·4	2·39	12·1	30·1	8·0	21·9	41·7	0·6
0·7	0	3·31	17·7	2·87	13·1	31·2	8·7	22·9	42·7	0·7
0·8	0	4·59	18·8	3·50	14·1	32·2	9·5	23·9	43·5	0·8
0·9	0	5·06	19·7	4·34	15·2	33·0	10·5	24·9	44·3	0·9
1·0	1·05	6·70	20·6	5·45	16·2	33·7	11·6	26·0	45·0	1·0

x	$f=3$			$f=4$			$f=5$			x
	10	50	90	10	50	90	10	50	90	
0·0	11·6	26·0	45·0	18·8	35·5	55·2	26·7	45·2	64·6	0·0
0·1	12·0	26·9	46·4	19·2	36·5	56·5	27·3	46·2	65·8	0·1
0·2	12·4	27·8	47·7	19·8	37·4	57·7	27·9	47·1	67·0	0·2
0·3	12·9	28·7	49·0	20·4	38·4	58·9	28·5	48·1	68·1	0·3
0·4	13·5	29·7	50·1	21·1	39·3	59·9	29·3	49·0	69·0	0·4
0·5	14·1	30·6	51·2	21·8	40·3	60·9	30·1	50·0	69·9	0·5
0·6	14·8	31·6	52·2	22·6	41·3	61·8	31·0	51·0	70·7	0·6
0·7	15·7	32·5	53·1	23·5	42·2	62·6	31·9	51·9	71·5	0·7
0·8	16·6	33·5	53·9	24·4	43·2	63·4	33·0	52·9	72·1	0·8
0·9	17·6	34·5	54·6	25·5	44·2	64·0	34·2	53·8	72·6	0·9
1·0	18·8	35·5	55·2	26·7	45·2	64·6	35·4	54·8	73·3	1·0

Notes. (*a*) The last digit is not guaranteed as correct.
(*b*) If $f > 5$, enter with $10 - f$ and find limits of $1 - \pi$.

3·2. *Single estimate of* π. When quoting a pair of fiducial limits, it is usual at the same time to give also a single estimate of the parameter. We suggest that the most appropriate single estimate in this case is the 50 % fiducial limit, i.e. the median of the fiducial distribution of the parameter. It may be remarked that the fiducial median is unbiased in a more absolute sense than are those estimates which it is common to call unbiased. For example, any estimate of the variance of a Normal distribution is implicitly an estimate of the standard deviation, but if we say that an estimate is unbiased when its expected value is equal to the parameter, we make it impossible to find any estimate which, on undergoing the appropriate transformation, is at one and the same time unbiased both for the variance and the standard deviation. In short, the definition of unbiased is purely relative to the form in which we chance to express the parameter. On the other hand, if by 'unbiased' we mean that the estimate is equally likely to be above or below the parameter, then the estimate so defined will remain unbiased under any transformation of the parameter.

In the example considered above, we therefore take as our single estimate of π the value found for $f = 3$, $x = 0\cdot34$ in the 50 % column

$$\pi = 29\cdot1\ \%.$$

Our conclusions about the value of the parameter may therefore be summarized diagramatically as

10		40		40		10	$= P(\%)$
	13·1		29·1		49·4		$= \pi(\%)$

It will be observed that it is impossible to find a median on the basis of the simpler tabulation, since any choice between the two relevant fiducial distributions would be entirely arbitrary.

3·3. *Approximate solutions*. As the great labour of tabulation of the exact solution hardly appears to be justified by the practical importance of the problem, one may inquire whether any rapid approximate solution is possible. One suggestion is that we might treat the relation between π and x as linear. This means that for a lower limit we find from the present tables the limits corresponding to f and $f+1$, and take a random linear interpolate between them, i.e. use

$$x\pi_0(f)+(1-x)\,\pi_0(f+1),$$

where x is a number chosen at random between 0 and 1. Similarly, for an upper limit, we find the limits corresponding to $f-1$ and f, and use

$$x\pi_1(f-1)+(1-x)\,\pi_1(f),$$

where x is the same x as was used for the lower limit.

If we do this, we shall find that the true probability that π will be below the lower limit (or above the upper limit) is 'always' greater than nominal. This is unfortunate because it implies that we tend to overvalue our confidence that the parameter lies between the calculated limits. On the other hand, the variations in the true probability are much less than when the usual method is followed. The maximum amount of error in the probability is more a function of f than of n. If we make it a rule only to apply this random interpolate method when $f > 3$ and $n \geqslant 20$ (otherwise using the usual method) then the maximum true probabilities that π will lie outside the calculated limits are:

	Lower limits		Upper limits		
Nominal	2·5	10	10	2·5	%
Maximum true	3·1	11·2	10·8	2·8	%

Usually, of course, the true probability will lie appreciably below its maximum, and the maxima themselves could be further reduced by raising f. The method seems then to be a reasonable one. It is also possible with the aid of a supplementary table to improve this approximation and make the maxima equal to the nominal probabilities, thus putting ourselves once more 'on the safe side', but we are not convinced that even this supplementary table is worth the trouble of computing.

3·4. *Example*. Given $f = 8$ and $n = 32$, to find the 2·5 % fiducial limits.

Entering our table in Fisher & Yates with $f\ (=a) = 8$ and $p = 0\cdot25$, we find $n\pi_0 = 3\cdot67$. Entering with $f = 9$ and $p = 0\cdot28$, we find $n\pi_0 = 4\cdot40$. Taking a random number 759 ..., we therefore have

$$\begin{aligned} n\pi_0 &= (0\cdot76)(3\cdot67)+(0\cdot24)(4\cdot40) \\ &= 3\cdot85. \end{aligned}$$

Hence $$\pi_0 = 0{\cdot}120.$$

Similarly, the upper 2·5 % fiducial limit is found to be

$$\begin{aligned} n\pi_1 &= (0{\cdot}76)(12{\cdot}78)+(0{\cdot}24)(14{\cdot}64) \\ &= 13{\cdot}23. \end{aligned}$$

Hence $$\pi_1 = 0{\cdot}413.$$

4. Discussion

4·1. *The correction for continuity.* Since tests of significance and methods of estimation are two poles of the same problem, it is to be anticipated that the idea of introducing a random increment will also be relevant to tests of significance, when the data are discontinuous. That the fiducial or confidence intervals, as usually calculated, are unnecessarily wide would imply that the usual tests of significance are more stringent than they appear. An examination of a typical problem, such as the test of independence of a 2×2 table, confirms this. If f, the entry in the first cell, is above expectation, then we calculate $G(f)$, the probability (on the hypothesis of independence) of obtaining the actual result or any larger value of f.

It is dangerous to be dogmatic about what the average research worker does next. One might say that he draws his conclusion solely on an inspection and appraisal of the value of $G(f)$. We suggest, however, that what he usually does is to compare $G(f)$ with some suitable level of probability, such as $P = 2\frac{1}{2}$ %, and to say that the table shows significant departure from independence if

$$G(f) < P.$$

It is, of course, obvious that

$$\operatorname{Prob}\{G(f) < P\} < P, \tag{4·11}$$

and is in fact often much less than P. Hence if the nul-hypothesis is true, the probability of declaring the result significant at the $2\frac{1}{2}$ % level is less, perhaps considerably less, than $2\frac{1}{2}$ %, or in other words the test is more stringent than it appears. One may ask whether this matters very much. But the whole idea of 'level of significance' is based on the principle of running an accurately known risk of declaring the hypothesis false when it is in fact true. With continuous data, the declared level of significance is an exact measure of this risk. It seems to us to be of some, though perhaps not of great, importance that the same should be made true when the data are discontinuous, in order that the measure of the research worker's confidence that he has overthrown the nul-hypothesis should have the same meaning in all problems.

This can be assured if we measure P by the formula

$$P = G(f+1) + xg(f), \tag{4·12}$$

where x is chosen at random between 0 and 1, and P is compared with the desired level of significance. This means that instead of always counting in the probability of the actual result, we include only a random portion of this probability.

When numbers are large, we note that this is approximately equivalent to adding to f a number x chosen at random between $-\frac{1}{2}$ and $+\frac{1}{2}$, afterwards calculating χ^2 in the usual way. This may be contrasted with the usually recommended correction for continuity, which consists in always subtracting $\frac{1}{2}$ from f.

The continuity correction here proposed may lead to a slight overvaluation of the significance of the table, instead of a considerable undervaluation (assuming of course that the

P obtained is compared with a standard significance level). The approximation might be improved by using the exact sampling variance of a cell entry after correction, thus writing χ^2 as

$$\chi^2 = \frac{(ad-bc+nx)^2\,(n-1)}{\{(a+b)\,(c+d)\,(a+c)\,(b+d)\}+\{n^2(n-1)\}/12},$$

where a, b, c, d are the four cell entries and x, as before, is drawn at random between $-\frac{1}{2}$ and $+\frac{1}{2}$.

It would, however, need a rather extensive investigation to discover how much the significance level thus calculated can diverge from the true probability of declaring a table significant.

4·2. *Conclusion*. We suppose that most people will find repugnant the idea of adding yet another random element to a result which is already subject to the errors of random sampling. But what one is really doing is to eliminate one uncertainty by introducing a new one. The uncertainty which is eliminated is that of the true probability that the parameter lies within the calculated interval. It is because this uncertainty is eliminated that we no longer have to keep 'on the safe side', and can therefore reduce the width of the interval, i.e. increase the precision of our estimation.

We confess, however, that we were less interested in this practical gain than in overcoming an embarrassing situation which one meets when teaching statistics. The interval estimates can be given a vivid meaning if one can say (with a demonstrated justification): 'I am willing to bet £$(1-P_0-P_1)$ against £(P_0+P_1) that the true answer is between such and such limits.' When the class reaches the binomial distribution, however, this had to be recast in the form: 'I am willing to bet *at least* £$(1-P_0-P_1)$ against £(P_0+P_1) ...', to which any intelligent pupil will reply: 'Since you say *at least*, you imply that you are able to raise the odds. How much higher can you make them?' It is really quite intolerable for the teacher then to have to reply, 'I don't know.' It therefore seemed worth while to show how this dilemma can be avoided.

REFERENCES

Clopper, C. J. & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, **26**, 404–13.

Fisher, R. A. (1935). The logic of inductive inference. *J.R. Statist. Soc.* **98**, 39–82.

Fisher, R. A. & Yates, F. (1943). *Statistical Tables for Biological, Agricultural and Medical Research*. London: Oliver and Boyd.

Addendum

That the author has to take a plane to Rio, every time he wishes to consult a back number of the *Journal of the Royal Statistical Society*, may be advanced as an excuse for his having overlooked a mention of the same method as is described here, in a discussion at the Royal Statistical Society (see F. J. Anscombe (1948), *J. Roy. Statist. Soc.*, Series A, **109**, 181–211 and the discussion which followed his paper). It was there dismissed rather briefly as being unsatisfactory. This may be granted but since, for a reason indicated in the last paragraph above, some solution is necessary, it seems that this one deserves to be studied and to be used by teachers of Statistics until a better one can be found.

[In addition to the present paper by W. L. Stevens and the following paper by K. D. Tocher, a paper on a related subject by Mark W. Eudey, entitled 'On the treatment of discontinuous random variables' has recently been printed as Technical Report No. 13, issued from the Statistical Laboratory, University of California, under a contract with the U.S. Office of Naval Research. Ed.]