# ON THE PROBLEM OF CONFIDENCE INTERVALS

By J. Neyman

When discussing my paper read before the Royal Statistical Society on 19th June, 1934, Professor Fisher said that the extension of his work concerning the fiducial argument to the case of discontinuous distributions, as presented in my paper, has been reached at a great expense: that instead of exact probability statements we get only statements in the form of inequalities.

This remark raises the question whether the disadvantage of the solution which he mentioned (the inequalities instead of equalities) results from the unsatisfactory method of approach, or whether it is connected with the nature of the problem itself.

I think that the problem is of considerable general interest. For instance it may be asked whether the confidence intervals for the binomial distribution recently published by E. S. Pearson and C. J. Clopper,[1] which correspond to the probability statements in inequalities, could be bettered.

The purpose of the present note is to show, (1) that in some exceptional cases the exact probability solution of the problem exists and that then it may easily be found by the method described in Note I of my paper;[2] (2) that in the general case of discontinuous distribution exact probability statements in the problem of confidence intervals are impossible.

In particular it will be seen that exact probability statements are impossible in the case of the binomial distribution and so that the system of confidence intervals published by Clopper and Pearson could not be bettered.

In order to avoid any possible misunderstanding I shall start by restating the problem.

We shall consider a random discontinuous variate $x$, capable of having one or another of a finite, or at most denumerable set of values

$$x_1, x_2, \cdots x_n, \qquad (1)$$

We shall assume that the frequency function, say $p(x \mid \theta)$, of $x$ depends upon one parameter $\theta$, the value of which is unknown. The problem of confidence intervals consists in ascribing to every possible value of $x$ e.g. to $x_n$, $(n = 1, 2, \cdots)$ a "confidence interval," say $\theta_1(n)$ to $\theta_2(n)$ such that the probability, $P$, of our being correct in stating

$$\theta_1(n) \leqq \theta \leqq \theta_2(n) \qquad (2)$$

whenever we observe $x = x_n$ $(n = 1, 2, \cdots)$, is either:

[1] E. S. Pearson and C. J. Clopper: The Use of Confidence or Fiducial Limits in the Case of the Binomial. Biometrika Vol. XXVI, pp. 404–413.

[2] J. R. S. S. Vol. 97, p. 589.

(a) equal to a given value $\alpha < 1$ chosen in advance, or

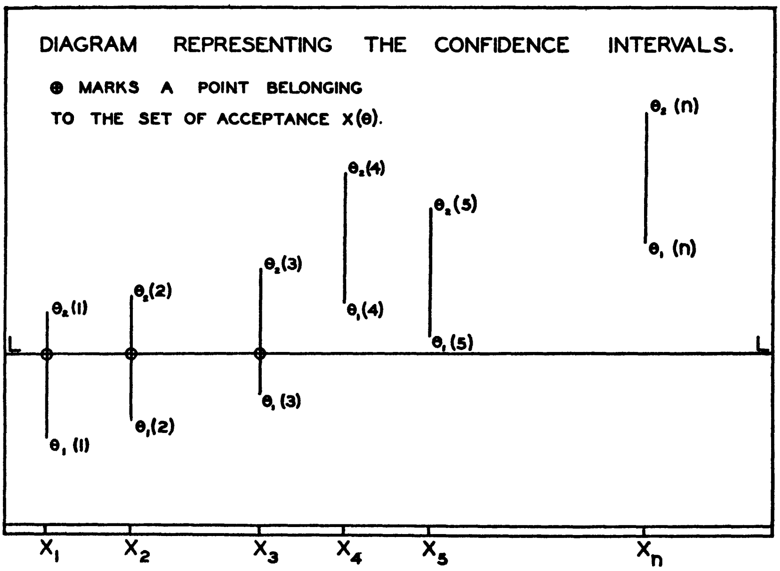
(b) *at least* equal to this value $\alpha$.

I proposed to call this chosen value $\alpha$ the confidence coefficient.

In the earlier paper I showed that the solution of the problem in its form (b) is always possible and easy to find. If the variate $x$ is continuous, then the solution of the problem (a) is equally easy. At present we shall consider whether and under what conditions the solution (a) is possible when the variate $x$ is discontinuous.

Suppose that the variate $x$ is discontinuous as described above, and that the solution of the problem in its form (a) exists and is given by the system of confidence intervals $(\theta_1(x_n), \theta_2(x_n))$ for $n = 1, 2, \cdots$.

The position is illustrated in the diagram below. On the axis of abscissae the possible values of the variate $x$ are marked. The axis of ordinates is the axis of $\theta$. The confidence intervals are marked on verticals passing through corresponding values of $x$.




According to our hypothesis the intervals $(\theta_1(x_n), \theta_2(x_n))$ are so chosen that

$$P = \alpha. \qquad (3)$$

$P$ is the probability of an event, say $E$, which we shall describe in some detail. Let us denote generally the probability of any event $a$ by $P\{a\}$. $P\{a \mid b\}$ will denote the probability of an event, $a$, calculated under the assumption that another event, $b$, has already occurred.

Now

$P = P\{E\} =$ the probability that {either $(x = x_1)$ and then $\theta_1(1) \leqq \theta \leqq \theta_2(1)$

$$\text{or } (x = x_2) \text{ and then } \theta_1(2) \leqq \theta \leqq \theta_2(2)$$
$$\dots\dots\dots\dots\dots\dots\dots\dots$$
$$\dots\dots\dots\dots\dots\dots\dots\dots$$
$$\text{or } (x = x_n) \text{ `` `` } \theta_1(n) \leqq \theta \leqq \theta_2(n)$$
$$\dots\dots\dots\dots\dots\dots\dots\dots$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\}$$

$$= P\{x = x_1\}P\{\theta_1(1) \leqq \theta \leqq \theta_2(1) \mid (x = x_1)\}$$
$$+ P\{x = x_2\}P\{\theta_1(2) \leqq \theta \leqq \theta_2(2) \mid (x = x_2)\}$$
$$+ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$= \sum_{n=1}^{\infty} P\{x = x_n\}P\{\theta_1(n) \leqq \theta \leqq \theta_2(n) \mid (x = x_n)\} = \alpha \dots\dots\dots\dots\dots (4)$$

The calculation of the probability $P$ in the above form is not convenient, as both multipliers in each term of the sum in (4) depend upon the unknown probability function *a priori* of $\theta$. Therefore we shall present $P$ in another form, giving to the event $E$ a geometrical interpretation. Let us denote by $CB$ the set of all confidence intervals $(\theta_1(n), \theta_2(n))$, as marked on the plane of $x$ and $\theta$. Thus $CB$ will be composed of points with co-ordinates $x$ and $\theta$, where

$$\left.\begin{aligned} x &= x_n \\ \theta_1(n) \leqq \theta &\leqq \theta_2(n) \end{aligned}\right\} n = 1, 2, \dots\dots\dots\dots\dots\dots\dots (5)$$

The set $CB$ will be called the confidence belt.

Denote by $A$ any point of the plane of $x$ and $\theta$, having any values for its co-ordinates.

It is easily seen that the event, which we denote by $E$, and the probability of which is $P = \alpha$, consists in the point $A$ belonging to the confidence belt $CB$. In fact the event $E$ occurs if and only if the co-ordinates of $A$ fulfil the conditions (5). But just these conditions define the points belonging to $CB$.

The above circumstance allows us to calculate $P$ by means of a formula which discloses its connection with $p(x \mid \theta)$.

Fix any possible value of $\theta = \theta'$ and draw the straight line $LL$ the points of which have just this fixed value $\theta'$ for their ordinates. The line $LL$ will cut some of the confidence intervals. Denote by $X(\theta')$ the set of points of intersection, and by $\phi(\theta)$ the unknown frequency function of $\theta$. The set $X(\theta)$ will be called the set of acceptance corresponding to the specified value of $\theta$.

The function $\phi(\theta)$ may be continuous or not. So may be $p(x \mid \theta)$ considered as a function of $\theta$. These cases may be treated together if we agree that $\sum_{\theta} F(\theta)$ will denote either the sum or the integral of $F(\theta)$ extending over all values of $\theta$, whenever $F(\theta)$ is integrable.

Using this notation we may write

$$P = P\{E\} = \sum_{\theta} \left\{ \phi(\theta) \sum_{X(\theta)} (p(x \mid \theta)) \right\} \ldots\ldots\ldots\ldots\ldots (6)$$

where $\sum_{X(\theta)}$ denotes the summation over all values of $x$ belonging to $X(\theta)$.

From the formula (6) may be deduced the following important proposition. *The probability $P$ may possess a constant value $\alpha$, independent of the properties of the unknown function $\phi(\theta)$, if and only if for each $\theta$*

$$\sum_{X(\theta)} (p(x \mid \theta)) = \alpha. \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

The condition (7) is obviously sufficient to have $P = \alpha$. In fact, if it is satisfied, then we should get from (6)

$$P = \alpha \sum_{\theta} (\phi(\theta)) = \alpha \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

since $\sum_{\theta} (\phi(\theta)) = 1$ whatever the frequency distribution of $\theta$. It is equally easy to see that the condition (7) is necessary for having $P = \alpha$ whatever the function $\phi(\theta)$. For suppose that for $\theta = \theta_1$ we have

$$\sum_{X(\theta_1)} (p(x \mid \theta_1)) = \beta \neq \alpha \ldots\ldots\ldots\ldots\ldots\ldots\ldots (9)$$

Then if it happens, that

$$\phi(\theta_1) = 1 \qquad \text{for } \theta = \theta_1 \tag{10}$$

and

$$\phi(\theta) = 0 \qquad \text{for } \theta \neq \theta_1 \tag{11}$$

the only term in the sum $\sum_{\theta}$ which is different from zero will be that corresponding to $\theta = \theta_1$ and the formula (6) will reduce to

$$P = \sum_{X(\theta_1)} (p(x \mid \theta_1)) = \beta \neq \alpha\,. \tag{12}$$

The original question, whether the solution of the form (a) is possible when the variate $x$ is discontinuous is thus put in the following form: is it possible to define for every possible value of $\theta$ a set of acceptance $X(\theta)$ such that the equation (7) holds good?

The answer is: in some cases it may be possible, but this depends upon the nature of the function $p(x \mid \theta)$. It is very easy to *invent* functions $p(x \mid \theta)$ for which the equation (7) for a definite value of $\alpha$ holds good, and we may even fix in advance the sets of acceptance $X(\theta)$. However the important question is not whether there may exist elaborately invented cases of discontinuous distributions where the solution (a) exists, but rather whether this solution exists always, or at least whether it exists frequently and in cases which are practically important.

This question must be answered in the negative on the basis of the following example concerning the most important of the discontinuous distributions, the Binomial.

In fact it will be seen below that if $x$ is a variate following the binomial frequency law, then whatever the arrangement of the sets of acceptance $X(\theta)$, corresponding to different values of $\theta$, the left hand side of the equation ( 7 ) cannot be constantly equal to the confidence coefficient $\alpha < 1$. It will follow that in the case of the binomial distribution, the solution of the problem ( a ) is impossible.

To prove this we shall consider the variate, $x$, following the binomial frequency law. That is to say we shall assume that $x$ may have values 0, 1, 2, $\cdots n$, and that

$$p(x \mid \theta) = \frac{n!}{x!(n-x)!}\,\theta^x(1-\theta)^{(n-x)} \tag{13}$$

while $0 < \theta < 1$. Since the set of possible values which $x$ may have is finite, therefore the set of all confidence intervals must be finite also. It follows that there is possible only a finite number of sets of acceptance $X(\theta)$. Therefore there must be at least one set of acceptance, say $X^0$, which will be common to an infinite number of values of $\theta$, say $\theta_1, \theta_2, \cdots \theta_n, \cdots$ so that for each it will be $X(\theta_n) = X^0$.

Now

$$\sum_{X(\theta_n)} (p(x \mid \theta_n)) \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \tag{14}$$

for all these values of $\theta = \theta_n$ will be *the same* polynomial in $\theta$ of the order $n$. If it has the same value $\alpha$ for a number of values of $\theta$ exceeding $n$, it means that this polynomial is an absolute constant. Therefore if it were possible to give a solution of the type (a) in the case of the binomial distribution, it would be possible to construct a sum (14), the terms of which are all different and have the form (13), and such that after all possible reductions and simplifications all terms involving $\theta$ would cancel and we should be left only with one constant term $\alpha < 1$. This, however, is impossible, since the only term of the form (13) which involves a constant, is the term corresponding to $x = 0$

$$p(0 \mid \theta) = (1-\theta)^n = 1 - n\theta + \frac{n(n-1)}{2}\theta^2 \;\ldots\ldots\ldots\ldots \tag{15}$$

and then this constant is 1. Other terms of the form (13) involve $\theta^x$ as a multiplier. Therefore there exists only one sum of the form (14) which is an absolute constant, but this includes all the terms (13)

$$\sum_{x=0}^{n} (p(x \mid \theta)) = 1 \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \tag{16}$$

and thus is of no value. It follows that whatever the sets of acceptance $X(\theta)$

the corresponding sum (14) will have values varying with the value of $\theta$ and hence the solution of the type (a) in the case of the binomial does not exist.

This, I think, gives the solution of the question raised by Professor Fisher. It is clear also that whenever the solution of the type (a) exists, it may be found by a suitable choice of sets of acceptance, and thus by the method explained in my earlier paper.

I should like now to raise another question. Past experience shows that the general problem of estimation may be formulated in different ways. The form of this problem as it appears in Bayes theorem, required for its solution the knowledge of the probabilities *a priori*.

The form of the same problem treated by R. A. Fisher in his theory of estimation was solved in terms of a new conception, that of likelihood.

The problem of estimation in its form of confidence intervals stands entirely within the bounds of the theory of probability, without involving any conception not already inherent in this theory. In the case of continuous distribution the problem also allows the solution (a) entirely independent of the probabilities *a priori*. Now it is shown that the necessity of the solution (b) is bound up with the nature of the problem if the distributions are discontinuous.

My question is: is it possible to formulate the problem of estimation in a fourth form, leading to a solution which (1) stands entirely on the grounds of the classical theory of probability, and (2) is not depending upon the probabilities *a priori*—whatever the conditions of the problem?