IV THE THEORY OF CONFIDENCE INTERVALS

The procedure of estimation, as I formulated it here, is also called estimation by a point. For practical applications the estimation by intervals seems to be much more important. That is to say, we have to construct two functions of the observations $\underline{\theta}$ (E) and $\overline{\theta}$ (E), where E denotes a point of the sample space, and we estimate the parameter to be within the interval $\delta(E) = \left[\underline{\theta}(E), \overline{\theta}(E)\right]$. In connection with the theory of interval estimation, R. A. Fisher introduced the notion of fiducial probability and fiducial limits, while Neyman[8] developed the theory of interval estimation based on the classical theory of probability. I shall give here a brief outline of Neyman's theory.

Before the sample has been drawn the point E is a random variable and, therefore, the values of $\underline{\theta}$ (E) and $\overline{\theta}$ (E) are also random variables. Hence, before the sample has been drawn we can speak of the probability that

$$(3) \qquad \underline{\theta}(E) \leq \theta \leq \overline{\theta}(E)$$

even if θ is considered merely as an unknown constant. After the sample has been drawn and we have obtained a particular sample point, say E_0, it does not make sense to speak of the probability that

$$(4) \qquad \underline{\theta}(E_0) \leq \theta \leq \overline{\theta}(E_0),$$

if θ is merely an unknown constant. Each term in the inequality (4) is a fixed constant, and the inequality (4) is either

---

8) See reference 15

right or wrong for those particular constants. It would be proper to talk about the probability of (4) if θ itself could be considered as a random variable having a certain probability distribution, called an a priori probability distribution. In this case we understand by the probability that (4) holds the conditional probability, called also a posteriori probability, under the assumption that $E = E_0$ occurred. If an a priori distribution of θ exists and if it is known then, using Bayes' formula, we can easily calculate the a posteriori probability distribution of θ. However, in practical applications we seldom meet cases where the assumption of the existence of an a priori probability distribution seems to be justified; and even in those rare cases in which the latter assumption can be made, we usually do not know the shape of the a priori probability distribution and this makes the application of Bayes' theorem impossible. For these reasons the theory of interval estimation has to be developed in such a way that its validity should not depend on the existence of an a priori probability distribution. Hence, in this theory we shall speak only of the probability of (3) but never of the probability of (4).

For any relationship R we will denote by $P\left[R \mid \theta\right]$ the probability of R calculated under the assumption that θ is the true value of the parameter.

A pair of functions $\underline{\theta}$ (E) and $\overline{\theta}$ (E) is called <u>a confidence interval of $\underline{\theta}$</u> if

1) $\underline{\theta}(E) \leq \overline{\theta}(E)$ for all points of E

2) $P\left[\underline{\theta}(E) \leq \theta \leq \overline{\theta}(E) \mid \theta\right] = \alpha$ for all values of θ,

where α is a fixed constant called the <u>confidence coefficient</u>.

The practical meaning and importance of the notion of the confidence interval is this: If a large number of samples are drawn and if in each case we make the statement that θ is included in the interval $\left[\underline{\theta}(E), \overline{\theta}(E)\right]$, then the relative frequency of correct statements will approximately be equal to α.

In general, there exist infinitely many confidence intervals corresponding to a fixed confidence coefficient α, and we have to set up some principle for choosing from among them. It is obvious that we want the confidence interval corresponding to a fixed confidence coefficient to be as "short" as possible. We have to give a precise definition of the notion <u>"shortest" confidence interval</u>.

A confidence interval $\delta(E) = \left[\underline{\theta}(E), \overline{\theta}(E)\right]$ is called a shortest confidence interval corresponding to the confidence coefficient α if

(a) $P\left[\underline{\theta}(E) \leq \theta \leq \overline{\theta}(E) \mid \theta\right] = \alpha$ and

(b) for any confidence interval $\delta'(E)$ which satisfies (a)

$P\left[\underline{\theta}(E) \leq \theta' \leq \overline{\theta}(E) \mid \theta''\right] \leq P\left[\underline{\theta}'(E) \leq \theta' \leq \overline{\theta}'(E) \mid \theta''\right]$

for all values θ' and θ'' of θ.

If a shortest confidence interval exists, it seems to be the most advantageous. Unfortunately, shortest confidence intervals exist only in quite exceptional cases. Therefore, we have to introduce some further principles on which the choice should be based. Such a principle is the <u>principle of unbiasedness.</u>

A confidence interval $\delta(E)$ is called an unbiased confidence interval corresponding to the confidence coefficient α if

$$P\left[\underline{\theta}(E) \leq \theta \leq \overline{\theta}(E) \mid \theta\right] = \alpha$$

and $\quad P\left[\underline{\theta}(E) \leq \theta' \leq \overline{\theta}(E) \mid \theta''\right] \leq \alpha$ for all values θ' and θ''.

A confidence interval $\delta(E)$ is called a shortest unbiased confidence interval corresponding to the confidence coefficient α if $\delta(E)$ is an unbiased confidence interval with the confidence coefficient α and if for any unbiased confidence interval $\delta'(E)$ with the same confidence coefficient, we have

$$P\left[\underline{\theta}(E) \leq \theta' \leq \overline{\theta}(E) \mid \theta''\right] \leq P\left[\underline{\theta}'(E) \leq \theta' \leq \overline{\theta}'(E) \mid \theta''\right]$$

for all values θ' and θ''.

If we accept the principle of unbiasedness, the shortest unbiased confidence interval seems to be the most favorable one. Even shortest unbiased confidence intervals exist only in a restricted, but important, class of cases. If a shortest unbiased confidence interval does not exist, Neyman proposes the use of a third type of confidence interval, which he calls <u>"short unbiased" confidence interval</u>. An unbiased confidence interval $\delta(E)$ with the confidence coefficient α is called a short unbiased confidence interval if

$$\frac{\partial^2}{\partial\theta''^2} P\left[\underline{\theta}(E) \leq \theta' \leq \overline{\theta}(E) \mid \theta''\right]\Bigg|_{\theta''=\theta'} \leq \frac{\partial^2}{\partial\theta''^2} P\left[\underline{\theta}'(E) \leq \theta' \leq \overline{\theta}'(E) \mid \theta''\right]\Bigg|_{\theta''=\theta'}$$

for all θ' and for all unbiased confidence intervals $\delta'(E)$ with the confidence coefficient α.

I have discussed only the case of a single unknown parameter. In the case of several unknown parameters some new problems arise, which do not occur in the case of a single parameter. However, I shall not discuss them, since the case of a single parameter already provides a good illustration of the basic ideas of the theories of Fisher, Neyman and Pearson.