REFERENCES

[1] M. S. Bartlett, "Complete simultaneous fiducial distributions," *Annals of Math. Stat.*, Vol. X(1939), pp. 129–138.
[2] W.-V. Behrens, "Ein Beitrag zur Fehlerberchnung bei wenigen Beobachtungen," *Landw. Jb.*, Vol. LXVIII(1929), pp. 807–37.
[3] R. A. Fisher, "The fiducial argument in statistical inference," *Ann. Eugen.*, Vol. VI(1935), pp. 91–98.
[4] M. S. Bartlett, The information available in small samples. *Proc. Camb. Phil. Soc.*, Vol. XXXII(1936), pp. 560–66.
[5] R. A. Fisher, "On a point raised by M. S. Bartlett on fiducial probability," *Ann. Eugen.*, Vol. VIII(1937), pp. 370–75.
[6] "Student," "The probable error of a mean," *Biometrika*, Vol. VI(1908), pp. 1–25.
[7] P. V. Sukhatme, "On Fisher and Behrens' test of significance for the difference in means of two normal samples," *Sankhyā*, Vol. IV(1938), pp. 39–48.
[8] R. A. Fisher, "Samples with possibly unequal variances," *Ann. Eugen.*, Vol. IX(1939), pp. 174–180.
[9] F. Yates, "An apparent inconsistency arising from tests of significance based on fiducial distributions of unknown parameters," *Proc. Camb. Phil. Soc.* (in press).

University College,
London.

A NOTE ON NEYMAN'S THEORY OF STATISTICAL ESTIMATION[1]

By Solomon Kullback

In this note we shall examine a section of a recent paper by Neyman[1] dealing with statistical estimation. Consider the following quotation from that section[2] which deals with the statement of the problem:

"Consider the variables $[x_1, x_2, \cdots, x_n]$ and assume that the form of the probability law $[p(x_1, \cdots, x_n \mid \theta_1, \theta_2, \cdots, \theta_l)]$ is known, that it involves the parameters $\theta_1, \theta_2, \cdots, \theta_l$ which are constant (not random variables), and that the numerical values of these parameters are unknown. It is desired to estimate one of these parameters, say θ_1. By this I shall mean that it is desired to define two functions $\bar{\theta}(E)$ and $\underline{\theta}(E) \leq \bar{\theta}(E)$, determined and single valued at any point E of the sample space, such that if E' is the sample point determined by observation, we can (1) calculate the corresponding values of $\underline{\theta}(E')$ and $\bar{\theta}(E')$ and (2) state that the true value of θ_1, say θ_1^0, is contained within the limits

$$\underline{\theta}(E') \leq \theta_1^0 \leq \bar{\theta}(E') \tag{18}$$

this statement having some intelligible justification on the ground of the theory of probability.

[1] Specifically we refer to J. Neyman "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability," *Phil. Trans. Roy. Soc.*, vol. A236 (1937), pp. 333–380.

[2] J. Neyman, loc. cit., p. 347. The material in brackets are slight alterations of the original text in order that the quotation do not refer to previous matter in the original paper.

This point requires to be made more precise. Following the routine of thought established under the influence of the Bayes Theorem, we could ask that, given the sample point E', the probability of θ_1^0, *falling within the limits* (18) should be large, say $\alpha = 0.99$, etc.. If we expressed this condition by the formula

$$P\{\underline{\theta}(E') < \theta_1^0 < \bar{\theta}(E') \mid E'\} = \alpha \tag{19}$$

we see at once that it contradicts the assumption that θ_1^0 is constant. In fact, on this assumption, whatever the fixed point E' and the values $\underline{\theta}(E')$ and $\bar{\theta}(E')$, the only values the probability (19) may possess are zero and unity. For this reason we shall drop the specification of the problem as given by the condition (19)."

We believe that the following approach to the problem, emphasizes to a greater extent the fact that if the practical statistician follows the steps recommended as a result of Neyman's solution, then 'in the long run he will be correct in about 100α percent of all cases'.

Let us return again to the condition (19) of the quotation, and write

$$\pi(E) = P\{\underline{\theta}(E) \leq \theta_1^0 \leq \bar{\theta}(E) \mid E\} \tag{1}$$

where of course $\pi(E) =$ zero or unity according as the true value of θ_1, say θ_1^0 does not or does satisfy the inequality

$$\underline{\theta}(E) \leq \theta_1^0 \leq \bar{\theta}(E) \tag{2}$$

We may however calculate the *average value* of $\pi(E)$ i.e., the percentage of cases in which in the long run the statistician will be correct.[3] In accordance with the definition of an average

$$\overline{\pi(E)} = \int_R \pi(E) p(E \mid \theta_1^0, \theta_2, \cdots, \theta_l)\, dx_1 dx_2 \cdots dx_n \tag{3}$$

where the region R is the entire sample space. If we let R_1 be that portion of the sample space for which (2) is satisfied, then since $\pi(E) = 1$ if E falls in R_1 and zero otherwise

$$\overline{\pi(E)} = \int_{R_1} p(E \mid \theta_1^0, \theta_2, \cdots, \theta_l)\, dx_1 dx_2 \cdots dx_n \tag{4}$$

Thus, if we want our rule to lead to a correct statement in 100α percent of cases in the long run, we must look for two functions $\underline{\theta}(E)$ and $\bar{\theta}(E)$ such that for the corresponding region R_1

$$\overline{\pi(E)} = \int_{R_1} p(E \mid \theta_1^0, \theta_2, \cdots, \theta_l)\, dx_1 dx_2 \cdots dx_n = \alpha \tag{5}$$

holds good whatever the value θ_1^0 of θ_1 and whatever the values of the other parameters $\theta_2, \theta_3, \cdots, \theta_l$ involved in the probability law of the X's may be.

[3] Cf. A. Werteimer, "A Note on Confidence Intervals and Inverse Probability," *Annals Math. Statistics*, Vol. X (1939), pp. 74ff.

If we apply to the preceding the calculus of probability in accordance with Neyman,[4] we find that (5) may be written as

$$\overline{\pi(E)} = P\{\underline{\theta}(E) \leq \theta_1^0 \leq \bar{\theta}(E) \mid \theta_1^0\} = \alpha \tag{6}$$

which, with the conditions stated for (5) is identical with formula (20) on page 348 of Neyman's paper.

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, D. C.

[4] J. Neyman loc. cit. pp. 333–343.

A NOTE ON A PRIORI INFORMATION

BY C. EISENHART

A survey of recent literature on mathematical statistics is sufficient to reveal the fact that in approaching certain types of problems some writers assume more information known *a priori* than do other writers. Indeed, it soon becomes evident that great care is necessary in wording (and in reading) propositions in mathematical statistics. Furthermore, propositions which are true and powerful when certain information is known *a priori* may become either useless or irrelevant according as more, or less, information is available *a priori*. Once this situation is appreciated some apparent contradictions are resolved, and certain exceptional examples can "be reasonably regarded as bearing out the principle to which formally they are anomalous."

So far as I know it was Bartlett [1, p. 271] who first clearly pointed out how a slight change in the amount of information known *a priori* can greatly alter the complexion of a problem. He was indebted to Neyman and Pearson [5, p. 122] for his problem, which was to develop a test of the statistical hypothesis, H_0, that $\beta = \beta_0$ *and* $\gamma = \gamma_0$ for a random sample from the distribution

$$p(x) = \begin{cases} \beta e^{-\beta(x-\gamma)} & \text{for } x \geq \gamma \\ 0 & \text{for } x < \gamma. \end{cases} \tag{1}$$

If (1) expresses *all* the information (about the distribution of x) that is to be considered as known *a priori*, any value of $\beta > 0$ and any finite value of γ being admissible, then it follows immediately from a result of R. A. Fisher's [2, p. 295] that no uniformly most powerful test, in the sense of Neyman and Pearson [4; 5, p. 115], can exist for H_0, since H_0 involves the simultaneous testing of two unrelated parameters.[1]

[1] Since Fisher's wording is important it will be well to quote him here: "It is evident, at once, that such a system [of maximum likelihood relations needed to insure the existence of a uniformly most powerful test] is only possible when the class of hypotheses considered