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

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

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

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

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

By assuming that in addition to (1) the *a priori* information includes the knowledge that $\beta \geq \beta_0$ and $\gamma \leq \gamma_0$ constitute the only admissible ranges of values of these parameters, Neyman and Pearson [5, p. 122] have succeeded in showing that a uniformly most powerful test of H_0 does exist when the admissible values of β and γ are restricted in this way. At first this appears to be in contradiction to Fisher's statement referred to above, but Bartlett [1, p. 271] points out that the restrictions on the admissible values of β and γ reduce the problem *effectively* to one of testing a single parameter: In the first place, no statistical test is necessary if an observation less than γ_0 occurs, since this refutes the hypothesis H_0 immediately. Therefore, a statistical test of H_0 is needed only when none of the observations are less than γ_0, and *for such observations* the distribution law is

$$(2) \qquad p(x) = \beta e^{-\beta(x-\gamma)}/e^{-\beta(\gamma_0-\gamma)} = \beta e^{-\beta(x-\gamma_0)}, \; x \geq \gamma_0,$$

and is independent of γ. In consequence, the test reduces to testing the single parameter β in (2), for which the arithmetic mean, $\bar{x}$, is a sufficient statistic. The discovery of a uniformly most powerful test of H_0, when the above restrictions are placed on the admissible values of β and γ, is, therefore, reasonably consistent with the full meaning of Fisher's statement.

The preceding example makes quite clear how a little additional *a priori* knowledge can affect the solution of a problem in mathematical statistics. The *a priori* knowledge employed by writers in mathematical statistics usually falls into one of the following categories:

(i) The elementary probability law is taken to be continuous or discrete, as the case may be, but its mathematical form is left unspecified.

(ii) The elementary probability law is taken to be of a definite mathematical form involving one or more parameters the value(s) of which is (are) not considered as known *a priori*, and any value(s) of this (these) parameter(s) consistent with the non-negative character of a probability law is (are) admissible.

(iii) Here the information assumed known is as in (ii) except that the admissible values of the parameter(s) form (a) restricted sub-set (or sub-sets) of the values admissible in (ii), such subsets, however, being comprised of more than a single value.

(iv) The information is so complete that the admissible values of the parameter(s) have (a) known *a priori* probability distribution(s)—if a param-

---

involves only a single parameter θ, or, what comes to the same thing, when all the parameters entering into the specification of the population are definite functions of one of their number. In this case, the regions defined by the uniformly most powerful test of significance are those defined by the estimate of the maximum likelihood, T. For the test to be uniformly most powerful, moreover, these regions must be independent of θ, showing that the statistic must be of the special type distinguished as sufficient." (Words in square brackets are mine.—C. E.)

eter θ is known to have a definite value θ', then the *a priori* probability law of θ can be taken as (Prob. θ equals θ') $= 1$, (Prob. θ not equal to θ') $= 0$.[2]

As statistical theory advances it may become necessary to classify problems according to the amount of information which may be assumed known *a priori*, when proceeding to their solution. No claim is made here that the above categories are the best to choose, but it may prove fruitful to study the extent to which results obtained with a certain amount of information assumed known are useful when more, less, or perhaps different, information is taken as known *a priori*. In particular, as the preceding example shows, it may be well to investigate exactly what are the implications of restricting the ranges of the admissible values of parameters.

It is unwise to attempt to predict the outcome of such research at this time, but it is probably safe to say that an increase in *a priori* information will generally render possible better tests of significance—better in the sense that, for a given probability of rejecting the hypothesis tested when true, the probability of rejecting it when false will be greater—and narrower confidence intervals for a given confidence level. The example already given, concerned with a test of significance, supports this conjecture. As a further example, from the point of view of estimation, we may recall that it is possible with a level of confidence equal to .96875 to assert [3, p. 4] that the true *median* of the population from which a random sample of 6 was drawn lies within the observed range of the sample, and this without any assumption about the population except that it is continuous. If, however, the population is known to be of normal form with unknown mean, m, and standard deviation, σ, then Student's t will provide the narrowest confidence intervals for the median of the population, since t provides [6, p. 378] the best available confidence intervals for the mean, m, (which is also the median) of a normal population when σ is unknown—if the population is normal and σ is known, then the normal deviate $(\bar{x} - m)\sqrt{6}/\sigma$ will supply still narrower confidence limits for m.

In conclusion, the circumstances under which it may be desired to apply methods of statistical inference may differ considerably in the amount of knowledge available to the research worker *a priori*, and the most efficient tests of significance and methods of estimation applicable to a given case will depend upon the nature of the available information as described in the above classification. In comparing the procedures of different writers, therfore, it is most important to examine their premises and see how much information each is considering as known at the start.

REFERENCES

[1] M. S. Bartlett, "Properties of Sufficiency and Statistical Tests." *Proc. Royal Society of London*, Series A, vol. 160(1937), pp. 268-282.

[2] Category (iv) should be regarded as sharing certain 'territory' with the others since, for instance, a problem may consist in the estimation of the true mean, m, which has a known *a priori* distribution, without assuming the mathematical form of the probability law of the random variable observed, x, as known *a priori*.

[2] R. A. FISHER, "Two New Properties of Mathematical Likelihood." *Proc. Royal Society of London*, Series A, vol. 144(1934), pp. 285-307.
[3] R. A. FISHER, " 'Student' ". *Annals of Eugenics*, vol. 9(1939) pp. 1-9.
[4] J. NEYMAN AND E. S. PEARSON, "On the Problem of the Most Efficient Tests of Statistical Hypotheses." *Phil. Trans. Royal Society of London*, Series A, vol. 231(1933), pp. 289-337.
[5] J. NEYMAN AND E. S. PEARSON, "Sufficient Statistics and Uniformly Most Powerful Tests of Statistical Hypotheses." *Statistical Research Memoirs*, vol. 1(1936), pp. 113-137.
[6] J. NEYMAN, "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability." *Phil. Trans. Royal Society of London*, Series A, vol. 236(1937), pp. 333-380.

UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN

A NOTE ON COMPUTATION FOR ANALYSIS OF VARIANCE

BY MORRIS C. BISHOP

The method of computation for analysis of variance commonly favored is one which involves obtaining the total and total sum of squares in a single operation on a computing or card-punch machine,[1] in which case a check on the accuracy of the work requires complete recomputation. But the best tools available to the student, and sometimes to the experimenter, are a table of squares and perhaps a listing machine. In such a situation, a simple algorithm which embodies checks on the computations is urgently needed. The method here presented reduces the arithmetic to repeated application of a single procedure, with adequate checks; it reveals rather than obscures the sample variances, which may or may not be of primary importance; and it provides an intuitively logical portrayal of the step-by-step improvement of the estimate of population variance.

The data items and their squares may be merged into a single table by setting them down in staggered fashion, as shown in Table I. If only a single criterion of classification is to be used—classified into columns, say—the columns are summed down, and then these totals across (obtained as two sets of subtotals and totals on a listing machine). This yields the grand total (T) and total sum of squares $\left(\sum_{i=1}^{N}\sum_{j=1}^{k} X_{ij}^2\right)$. Summing across and down verifies the addition and provides material for two-way classification. The total sum of squares of deviations is obtained by the familiar formula

$$(1) \qquad \sum_{i=1}^{N}\sum_{j=1}^{k} (X_{ij} - \bar{X})^2 = \sum_{i=1}^{N}\sum_{j=1}^{k} X_{ij}^2 - \frac{T^2}{Nk}$$

where Nk is the total number of observations in N rows and k columns.

[1] See George W. Snedecor, *Analysis of Variance and Covariance*, and Paul R. Rider, *Modern Statistical Methods*.