ON BAYES PROCEDURES FOR A PROBLEM WITH CHOICE OF OBSERVATIONS[1]

By T. W. Anderson

Columbia University

1. Introduction. In establishing statistical means to decide between two hypotheses, H_0 and H_1, an experimenter may have the choice of observing a variable X alone or of observing two variables X and Y. While observation of the two variables is more informative than observation of the one variable, it is also more expensive. The question is whether it is worthwhile for the experimenter to pay the greater cost necessary for the two variables. He must make the decision whether to observe Y before the observation on X is made.

As an example, in a medical study X may be a sufficient statistic for a sample of m individuals who have been treated with a drug and are undergoing observation. Because the treatment and observation take a long period of time, before he has completed study of the first m individuals treated the investigator may consider treating n additional individuals, Y being the sufficient statistic for the second sample. This example is covered by the general study if X and Y are the sums (or means) of the observations in the two samples, respectively, and the measurements are normally distributed with known and common variance.

The situation considered here is different from the usual two-sample or sequential situation in that the decision whether to observe Y is made independently of the observation X. As a matter of fact, this problem arose as a simple analogue of a problem of finding Bayes and admissible procedures for deciding between two hypotheses H_0 and H_1 when observations are taken sequentially and after the decision to stop observation has been taken, m more observations (corresponding to X) are obtained, as for instance in clinical trials [Anderson (1964)].

This study also applies to a problem of classification in multivariate statistical analysis. Suppose that an investigator wants to classify an individual as coming from one of two populations. The measurements he may make have joint normal distributions in the two populations; the populations are the same in variances and correlations, but differ in means. The investigator may be able to observe either the set of measurements $z_1, \cdots, z_m$ or the set of measurements $z_1, \cdots, z_{m+n}$; for example, the first m measurements may be made by one device and are required, and the last n measurements may be made by another device and are optional. Does it pay the investigator to observe the second set of measurements in addition to the first set? This problem will be treated explicitly at the end of Section 2 as a special case of the general problem.

We formulate our problem more precisely by assuming that there is a loss

Received 17 December 1963; revised 4 March 1964.

[1] Research sponsored by Contract AF 41(609)-1534 between the USAF School of Aerospace Medicine and Teachers College, Columbia University.