CONTRIBUTIONS TO THE THEORY OF RANK ORDER STATISTICS—THE TWO-SAMPLE CASE

By I. Richard Savage

National Bureau of Standards and *Stanford University*

1. Introduction. The idea of a statistical test of a hypothesis and the related concepts introduced by Neyman and Pearson have served as a model for much of modern statistics. In nonparametric work it is seldom possible to apply all of these concepts. This results from the fact that for most of the alternatives that have been considered there do not exist optimum critical regions or analytic tools for finding power functions. The sign test gives an illustration where it is possible to find the exact power function; on the other hand, this procedure is seldom optimum. The c_1 test [11] has optimum limiting properties but little is known about its power function for small samples. The Kolmogorov and Smirnov tests [6] have a certain intuitive appeal but their only justification is consistency. The Wilcoxon test [9] is justified on the basis that it is analogous to a good parametric procedure but has little direct justification.

In the course of this paper we will consider several nonparametric hypotheses that have been treated previously. In Section 5 it will be indicated that for the two-sample problem with such alternatives as slippage, there do not exist optimum nonparametric tests. In particular, we show that the class of admissible tests is too large to be of use. In Section 6 alternatives are considered involving monotone likelihood ratios and a necessary criterion for admissibility is given. In particular, two normal populations differing only in mean value are considered. It is shown that several of the previously proposed tests of this hypothesis satisfy this criterion. Section 7 deals with a special subclass of the alternatives used in Section 6. Members of this subclass are the extreme-value distribution and the exponential distribution. For these alternatives we not only have the results of the previous section on the construction of admissible tests, but also are able to carry out the construction of optimum nonparametric tests for small samples and to evaluate the operating characteristics of these tests. These small-sample tests are uniformly most powerful rank order tests and most stringent rank order tests. Also the limiting optimum test is given.

2. Notation. The main concern in the following will be the situation where there are random variables $X_1, \cdots, X_m$ independently distributed, each with continuous distribution function $F(x)$, and random variables $Y_1, \cdots, Y_n$ which are independent of the X's and are independently distributed, each with continuous distribution $G(x)$, i.e., two independent samples.

The observed values $x_1, \cdots, x_m$ of the random variables $X_1, \cdots, X_m$ will be called the first sample and the observed values $y_1, \cdots, y_n$ of the random variables $Y_1, \cdots, Y_n$ will be called the second sample. When all of the observed values are ordered from smallest to largest, they form a sequence which

Received June 1, 1954.