ON THE DISTRIBUTION OF RANKS AND OF CERTAIN RANK ORDER STATISTICS[1]

By Meyer Dwass

Northwestern University and Stanford University

1. Introduction. Suppose $X_1, \cdots, X_m$ and $X_{m+1}, \cdots, X_N$ are two independent samples from two possibly different populations, and $R_1, \cdots, R_m$ are the ranks of the first m observations in the combined sample and $R_{m+1}, \cdots, R_N$ the ranks of the remaining observations. In the first part of the paper, various moment generating functions connected with these ranks are derived. Of particular interest may be the moment generating function of the Wilcoxon statistic. The asymptotic distribution of a finite number of ranks is derived as $N \to \infty$. The remainder of the paper studies certain aspects of the distribution theory of rank order statistics of the form $\sum_{i=1}^{m} f_N(R_i/N)$. The Wilcoxon statistic and the Hoeffding c_1-statistic are special cases of such a statistic. Many previous studies have been devoted to showing its asymptotic normality. The main purpose of the last half of this paper is to show that for certain combinations of sample sizes m, n, and parent populations, the limiting distribution is non-normal as $m \to \infty$, $n \to \infty$, and $m/N \to 0$.

2. Generating functions for ranks. Throughout this paper we suppose that $X_1, \cdots, X_m, X_{m+1}, \cdots, X_{m+n}$ are $N = m + n$ independent random variables, the first m identically distributed, each with c.d.f. F_1 and the last n identically distributed, each with c.d.f. F_2. We suppose these c.d.f.'s are continuous. By the random variable R_i, the rank of X_i, we mean the number of X_j's less than or equal to X_i. The main object of this section is to write an expression for a generating function for ranks, and the following notation is intended to be useful toward that end. Let $u_0 = -\infty$, $u_{r+1} = \infty$, and

$$u_0 < u_1 < \cdots < u_r < u_{r+1}.$$

Then we denote

$$G_{i,j+1} = G_{i,j+1}(u_j, u_{j+1}) = F_i(u_{j+1}) - F_i(u_j) \qquad (i = 1, 2; j = 0, \cdots, r).$$

Let $i_1, i_2, \cdots, i_r$ be a permutation of the $r = p + q (p \leqq m, q \leqq n)$ integers $1, 2, \cdots, p, m + 1, \cdots, m + q$, and let $e_{i_1}, \cdots, e_{i_r}$ be defined by

$$e_{i_1} = \begin{cases} 1 & \text{if } i_1 \text{ is one of } 1, \cdots, p, \\ 2 & \text{if } i_1 \text{ is one of } m+1, \cdots, m+q, \end{cases}$$

with similar definitions for $e_{i_2}, \cdots, e_{i_r}$. If $d_1 < d_2 < \cdots < d_r$ is a set of positive integers, they uniquely determine a set of non-negative integers w_1,

Received June 18, 1926; revised October 22, 1956.

[1] Work performed under Office of Naval Research Contract Nonr-225(21).