The Annals of Mathematical Statistics
1969, Vol. 40, No. 3, 1011–1020

THE ASYMPTOTIC DISTRIBUTION OF SOME NON LINEAR FUNCTIONS OF THE TWO-SAMPLE RANK VECTOR[1]

BY SIEGFRIED SCHACH[2]

University of Minnesota and Stanford University

1. Introduction. Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be independent samples from the same continuous distribution. Let R_{Ni} $(i = 1, 2, \cdots, m)$ be the rank of X_i in the combined sample $(N = m + n)$, and define

$Z_i = Z_{Ni} = 1$ if the ith element of the combined ordered sample is an X observation
$= 0$ otherwise.

Let $A^{(N)} = (a_{i,j}^{(N)})$ be a sequence of symmetric matrices. We find conditions under which a statistic of the form

$$S_N = \sum_{i=1}^m \sum_{j=1}^m a_{R_{Ni},R_{Nj}}^{(N)} = \sum_{i=1}^N \sum_{j=1}^N a_{i,j}^{(N)} Z_i Z_j \tag{1.1}$$

converges in distribution as $N \to \infty$ and $mn^{-1} \to \lambda$, $0 < \lambda < 1$.

Several examples of non-parametric test statistics of the form (1.1) can be found in the literature. In fact, any two-sided symmetric test based on a linear rank statistic is trivially equivalent to a test of the form $S_N \geqq C$. More interesting examples arise in the area of non-parametric statistics for circular distributions. Wheeler and Watson [6] proposed a two-sample non-parametric test for circular distributions of the form (1.1), which is related to the usual parametric test (likelihood ratio test for the class of v. Mises distributions) in much the same way as the Wilcoxon test is related to the two-sample t-test. The author [4] found that in detecting rotation alternatives for circular distributions a locally most powerful invariant test under a suitable group of transformations is based on a statistic of the form (1.1). Matthes and Truax [2] obtained a test statistic related to (1.1) when deriving locally most powerful invariant tests for two-sided shift alternatives.

2. Step function representation. Degenerate and non-degenerate limit functions. Let $h_N(\cdot, \cdot)$ be a step function on the unit square, which is constant on subsquares of the form $(i-1)/N < x \leqq i/N$, $(j-1)/N < y \leqq j/N$ and which is symmetric, i.e., $h_N(x, y) = h_N(y, x)$. If we set $h_N(i/N, j/N) = a_{i,j}^{(N)}$ then obviously every statistic of the form (1.1) can be written as

$$S_N = \sum_{i=1}^m \sum_{j=1}^m h_N(R_{Ni}/N, R_{Nj}/N) = \sum_{i=1}^N \sum_{j=1}^N h_N(i/N, j/N) Z_i Z_j \tag{2.1}$$

and vice versa. Since conditions on covergence can be stated more easily in terms of sequences $\{h_N\}$ we will use representation (2.1) in the sequel.

Assume that there exists a function $h(\cdot, \cdot)$ such that $h_N \to h$ in $L_2(I^2)$, i.e.,

Received 15 May 1968.

[1] This report was supported by the National Science Foundation (GP-6859) and by U. S. Public Health Service Grant (USPHS-GM-14554-02).

[2] Now at The Johns Hopkins University.