Acta Math. Acad. Sci. Hungar.
40 (1—2), (1982), 55—63.

DOUBLY ORDERED LINEAR RANK STATISTICS

By

B. GYIRES (Debrecen)

1. Introduction

Let $N=m+n$, where m and n are positive integers. Let the real numbers $x_1, \dots, x_N$ be pairwise different from each other. Let Rank x_k and rank x_k denote the rank of x_k within the sequences $x_1, \dots, x_N$, and $x_1, \dots, x_n$, respectively. In other words, if the rearrangement according to size $z_1 < \dots < z_N$ of those numbers $x_k = z_{r_k}$, and if the rearrangement according to size $y_1 < \dots < y_n$ of the numbers $x_1, \dots, x_n$, $x_k = y_{i_k}$, then we say that x_k has rank r_k and i_k with respect to these orders, and we write Rank $x_k = r_k$, and rank $x_k = i_k$ $(k=1, \dots, m)$, respectively.

Denote by $\Pi_m^{(N)}$ the set of all $(r_1, \dots, r_m)$ chosen without repetition from the elements $1, \dots, N$, and by P_m the set of all permutations of the elements $1, \dots, m$ without repetition.

Let the distribution functions of the real random variables $X_1, \dots, X_N$ be continuous. Then $P(X_j = X_k) = 0, j \neq k$.

Denote by $\{r_1, \dots, r_m\}$ and $[i_1, \dots, i_m]$ the vectors with the components $r_1, \dots, r_m$ and $i_1, \dots, i_m$, where $(r_1, \dots, r_m) \in \Pi_m^{(N)}$, $(i_1, \dots, i_m) \in P_m$, respectively.

DEFINITION 1.2. The vector $\{r_1, \dots, r_m\}$ is said to be the outer-rank of the random variables $X_1, \dots, X_m$ with respect to the random variables $X_1, \dots, X_N$, if

$$\{\text{Rank}\, X_1, \dots, \text{Rank}\, X_m\} = \{r_1, \dots, r_m\}.$$

DEFINITION 1.1. The vector $[i_1, \dots, i_m]$ is said to be the inner-rank of the random variables $Y_1, \dots, Y_m$, if

$$[\text{rank}\, Y_1, \dots, \text{rank}\, Y_m] = [i_1, \dots, i_m].$$

Obviously, the random events $\{r_1, \dots, r_m\}$ and $[i_1, \dots, i_m]$ are independent if the random variables $X_1, \dots, X_N$ and $Y_1, \dots, Y_m$ are independent, i.e. in this case

$$P(\{r_1, \dots, r_m\}, [i_1, \dots, i_m]) = P(\{r_1, \dots, r_m\}) P([i_1, \dots, i_m]).$$

By the help of this formula we get the following theorem on the basis of [3] (p. 369, Satz 10).

THEOREM 1.1. *Let the random variables $Z_1 = (X_1, Y_1), \dots, Z_m = (X_m, Y_m)$ and the random variables $X_{m+1}, \dots, X_N$ be given. If the random vector variables $(X_1, \dots, X_N)$ and $(Y_1, \dots, Y_m)$ are independent and if the joint distribution functions of the random variables $X_1, \dots, X_N$ and the random variables $Y_1, \dots, Y_m$ are symmetric functions of* *their variables, and they are continuous in each of the variables, then*

$$P(\{r_1, \ldots, r_m\}, [i_1, \ldots, i_m]) = \frac{1}{m!(n+1)\ldots(n+m)},$$

$$(r_1, \ldots, r_m)\in \Pi_m^{(N)}, \quad (i_1, \ldots, i_m)\in P_m.$$

The conditions of Theorem 1.1 will be satisfied if $X_1, \ldots, X_N$ and if $Y_1, \ldots, Y_m$ are samples with continuous distribution functions, and these random variables are independent.

Let the matrices

$$A_j = \begin{pmatrix} a_{11}^{(j)} \ldots a_{1N}^{(j)} \\ \cdot \;\ldots\; \cdot \\ a_{m1}^{(j)} \ldots a_{mN}^{(j)} \end{pmatrix} \quad (j = 1, \ldots, m) \tag{1.1}$$

with real elements be given, and let $A=A_1 \ldots A_m$ be the $m \times mN$ matrix with blocks (1.1).

On the basis of Theorem 1.1 we give the following definition.

DEFINITION 1.3. The random variable $X_{m,n}^{(N)}$ is said to be a doubly ordered linear rank statistics generated by the matrix A if

$$P(X_{m,n}^{(N)} = a_{i_1 r_1}^{(1)} + \ldots + a_{i_m r_m}^{(m)}) = \frac{1}{m!(n+1)\ldots(n+m)}, \tag{1.2}$$

where $(r_1, \ldots, r_m)$ and $(i_1, \ldots, i_m)$ run over the sets $\Pi_m^{(N)}$ and P_m, respectively.

Let the random vector variables $Z_1=(X_1, Y_1), \ldots, Z_m=(X_m, Y_m)$ and the random variables $X_{m+1}, \ldots, X_N$ be given. Suppose that the random variables $X_1, \ldots, X_N$, $Y_1, \ldots, Y_m$ are independent with continuous distribution functions. Suppose that $X_1, \ldots, X_m$ and $X_{m+1}, \ldots, X_N$ are samples with distribution functions $F(x)$ and $G(x)$, respectively. Then the doubly ordered linear rank statistics $X_{m,n}^{(N)}$ defined by (1.2) give us the possibility to decide on the acceptance or the rejection of the joint hypothesis

a) the second components of the random vector variables $Z_1, \ldots, Z_m$ have a common distribution function;

b) $F(x) = G(x)$, $x\in R_1$.

If all rows of the matrix A are equal, then $X_{m,n}^{(N)}$ give us the possibility to take a decision on the acceptance or rejection of the hypothesis b).

Denote by $\Psi_{m,n}^{(N)}(t)$ the characteristic function of the random variable $X_{m,n}^{(N)}$ defined by (1.2).

The aim of this paper is to investigate the characteristic function $\Psi_{m,n}^{(N)}(t)$. Beside the Introduction the paper contains two sections. In Section 2 the characteristic function $\Psi_{m,n}^{(N)}(t)$ will be approximated by the permanents of simpler characteristic functions. On the basis of this approximation theorem we give an asymptotic formula for $\Psi_{m,n}^{(N)}(t)$. The theorems of Section 3 are dealing with the construction of doubly ordered linear rank statistics with given limit distribution. To do this it is necessary to extend te well-known Koksma's inequality for arbitrary distributions.