

Linear order statistics in the case of samples with non-independent elements

By BÉLA GYIRES (Debrecen)

Introduction

Let x_1, \dots, x_{m+n} be pairwise different real numbers. If in the rearrangement according to size $z_1 < \dots < z_{m+n}$ of these numbers $x_k = z_{r_k}$, than we say that x_k has rank r_k , $\text{rank } x_k = r_k$.

Let R_{m+n} be the vector space of dimensions $m+n$ and let one of the variations without repetition of class m of the elements $1, \dots, m+n$ be r_1, \dots, r_m . Let moreover

$$\omega_{r_1, \dots, r_m} = \{(x_1, \dots, x_{m+n}) \in R_{m+n} | x_j \neq x_k, j \neq k, \text{ and } \text{rank } x_k = r_k (k = 1, \dots, m)\}.$$

As is known, in the theory of order statistics a fundamental role is played by the following theorem ([8], 263, Satz 10):

If the common distribution function of the identically distributed random variables ξ_1, \dots, ξ_{m+n} is a symmetric function of its variables and continuous in each of the variables, then

$$(1) \quad P((\xi_1, \dots, \xi_{m+n}) \in \omega_{r_1, \dots, r_m}) = \frac{1}{(n+1) \dots (n+m)}.$$

The conditions listed will be satisfied, if ξ_1, \dots, ξ_{m+n} are identically distributed, independent random variables with continuous distribution function.

If the random variables are identically distributed and the common distribution function is continuous in each of its variables, but no symmetric function of the variables, then the above theorem fails in general to be true. In this case we have more generally

$$(2) \quad P((\xi_1, \dots, \xi_{m+n}) \in \omega_{r_1, \dots, r_m}) = p_{r_1, \dots, r_m},$$

where

$$p_{r_1, \dots, r_m} \geq 0, \quad \sum_{(r_1, \dots, r_m) \in \Pi_m^{(m+n)}} p_{r_1, \dots, r_m} = 1$$

with $\Pi_m^{(m+n)}$ denoting the set of variations without repetition of order m of the elements $1, \dots, m+n$.

In the present paper we are going to consider linear order statistics in this general case, i.e. in the case of non-independent samples too.

In the first chapter we generalize the notion of linear order statistics. In the second chapter we consider infinite stochastic matrices of special types playing a role in the construction of linear order statistics, while the third chapter is devoted to limit theorems built on the characteristic functions of linear order statistics. In the fourth chapter we give a procedure for the construction of linear order statistics with a given limit distribution. The constructions employed are built — with the exception of only two cases — on the theory of mechanical quadrature. For one of the exceptions, we use one of the criteria of Riemann-integrability. For the case when (1) is valid, constructive procedures different from those here exposed have been given by the author in his papers [4] and [5].

Throughout the whole of the exposition, a fundamental role will be played by the weak convergence of random variables. We denote it by the symbol \Rightarrow . For different definitions of this notion and their interdependence see [3], 37—38, 58.

1. Linear order statistics

1.1. Let the stochastic matrices

$$(3) \quad S_k = \begin{cases} p_{11}^{(k)} \\ p_{21}^{(k)} p_{22}^{(k)} \\ \dots\dots\dots \\ p_{v1}^{(k)} p_{v2}^{(k)} \dots p_{vv}^{(k)} \\ \dots\dots\dots \end{cases} \quad (k = 1, 2, \dots)$$

be given, i.e. let

$$p_{vj}^{(k)} \geq 0, \quad \sum_{j=1}^v p_{vj}^{(k)} = 1.$$

Let B_{mv} denote the matrix with m rows and $v = m+n$ columns:

$$B_{mv} = \begin{pmatrix} p_{v1}^{(1)} & \dots & p_{vv}^{(1)} \\ \dots\dots\dots \\ p_{v1}^{(m)} & \dots & p_{vv}^{(m)} \end{pmatrix};$$

again, let $B_{\beta_k}^{(k)}$ be the matrix having m rows and β_k columns, each column of which is equal to the k -th column of the matrix B_{mv} . Let $(B_{\beta_1}^{(1)} \dots B_{\beta_v}^{(v)})$ be the matrix with m rows and $\beta_1 + \dots + \beta_v$ columns, obtained by writing successively the matrices $B_{\beta_1}^{(1)}, \dots, B_{\beta_v}^{(v)}$. If $\beta_k = 0$, then the k -th column of the matrix B_{mv} is absent from the matrix $(B_{\beta_1}^{(1)} \dots B_{\beta_v}^{(v)})$. Let $\prod_m^{(v)}$ denote the set of variations without repetition of class m of the elements $1, \dots, v$, and we denote the permanent of a matrix by the symbol Per .

We introduce the following matrix operations:

$$G(B_{mv}) = \sum_{\substack{\beta_1 + \dots + \beta_v = m \\ \beta_1^2 + \dots + \beta_v^2 = m}} \text{Per}(B_{\beta_1}^{(1)} \dots B_{\beta_v}^{(v)}) =$$

$$= \sum_{1 \leq k_1 < \dots < k_m \leq v} \text{Per}(B_1^{(k_1)} \dots B_1^{(k_m)}) = \frac{1}{m!} \sum_{(k_1, \dots, k_m) \in \Pi_m^{(v)}} \text{Per}(B_1^{(k_1)} \dots B_1^{(k_m)}),$$

$$H(B_{mv}) = \sum_{\substack{\beta_1 + \dots + \beta_v = m \\ \beta_1^2 + \dots + \beta_v^2 > m}} \text{Per}(B_{\beta_1}^{(1)} \dots B_{\beta_v}^{(v)}).$$

Let the matrices with real elements

$$A_k = \begin{cases} a_{11}^{(k)} \\ a_{21}^{(k)} & a_{22}^{(k)} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{v1}^{(k)} & a_{v2}^{(k)} & \dots & a_{vv}^{(k)} \\ \dots & \dots & \dots & \dots \end{cases} \quad (k = 1, 2, \dots)$$

be given.

Definition 1. By the linear order statistics $\{A_k, S_k\}$ we mean the stochastic process

$$\xi_{m,n} = \eta_{vm}^{(1)} + \dots + \eta_{vm}^{(m)}, \quad v = m+n \quad (m = 1, 2, \dots; n = 0, 1, \dots),$$

provided

$$P(\eta_{vm}^{(1)} = a_{vk_1}^{(1)}, \dots, \eta_{vm}^{(m)} = a_{vk_m}^{(m)}) = \frac{\text{Per}(B_1^{(k_1)} \dots B_1^{(k_m)})}{m! G(B_{mv})}, \quad (k_1, \dots, k_m) \in \Pi_m^{(v)}$$

and

$$(4) \quad G(B_{mv}) \rightarrow 1, \quad n \rightarrow \infty \quad (m = 1, 2, \dots).$$

From this can it count the probabilities

$$P(\eta_{vm}^{(k)} = a_{vj}^{(k)}) = p_{vj}^{(m,k)} \quad (j = 1, \dots, v).$$

Definition 2. The linear order statistics $\{A_k, S_k\}$ is asymptotic, if for any natural number m there exists a random variable ξ_m , such that

$$\xi_{m,n} \Rightarrow \xi_m, \quad n \rightarrow \infty.$$

Definition 3. The linear order statistics $\{A_k, S_k\}$ is doubly asymptotic, if there exists a random variable ξ such that

$$\xi_{m,n} \Rightarrow \xi \quad \text{if } n \rightarrow \infty, \quad \text{and then } m \rightarrow \infty.$$

We shall also say that the linear order statistics has asymptotically ξ_m ($m=1, 2, \dots$) distribution and doubly asymptotically ξ distribution respectively.

Clearly, a linear order statistics having asymptotically ξ_m ($m=1, 2, \dots$) distribution is doubly asymptotic if and only if $\xi_m \Rightarrow \xi, m \rightarrow \infty$.

1.2. Let ξ_1, \dots, ξ_m and η_1, \dots, η_n be samples of continuously distributed random random variables ξ and η respectively. If we suppose that in case ξ and η have the same distribution, the random vector $(\xi_1, \dots, \xi_m, \eta_1, \dots, \eta_n)$ formed from

the respective samples satisfies condition (2), where

$$(5) \quad p_{r_1 \dots r_m} = \frac{\text{Per}(B_1^{(r_1)} \dots B_1^{(r_m)})}{m! G(B_{mv})}, \quad v = m + n$$

then with respect to accepting or rejecting the hypothesis

$$P(\xi < x) = P(\eta < x)$$

we can make the following decision:

If $(\xi_1, \dots, \xi_m, \eta_1, \dots, \eta_n) \in \omega_{r_1 \dots r_m}$, then the question is settled by the falling of $a_{vr_1}^{(1)} + \dots + a_{vr_m}^{(m)}$ into the domain of acceptance or into the critical domain. The test itself will be constructed on the basis of the distribution (5), or in the case of asymptotical linear order statistics for sufficiently large n , and in the case of doubly asymptotic linear order statistics for sufficiently large n and m with the help of the limit distribution.

2. Stochastic T -matrices

2.1. The infinite matrix

$$(6) \quad \begin{pmatrix} p_{11} \\ p_{21} & p_{22} \\ \dots & \dots \\ p_{v1} & p_{v2} & \dots & p_{vv} \\ \dots & \dots & \dots & \dots \end{pmatrix}$$

is said to be a stochastic T -matrix, if ([1], 64)

$$(7) \quad p_{vj} \geq 0, \quad \sum_{j=1}^v p_{vj} = 1 \quad (j = 1, \dots, v; v = 1, 2, \dots)$$

and

$$(8) \quad \max(p_{v1}, \dots, p_{vv}) \rightarrow 0, \quad v \rightarrow \infty.$$

Examples of stochastic T -matrices:

Example 1. The so-called matrix of arithmetical means built from the elements $p_{vj} = \frac{1}{v}$ is a stochastic T -matrix.

Example 2. If $0 < p < 1$, $p + q = 1$, then the matrix (6) built from the elements

$$p_{vj} = \binom{v-1}{j-1} p^{j-1} q^{v-j}$$

is a stochastic T -matrix.

PROOF. By the theorem of LAPLACE.

$$\max(p_{v1}, \dots, p_{vv}) \sim \frac{1}{\sqrt{2\pi(v-1)pq}} = O\left(\frac{1}{\sqrt{v-1}}\right)$$

and from this our statement follows.

Example 3. Let r be a natural number. The so-called matrix of Cesaro-means built from the elements

$$p_{vj} = \binom{v+r-j-1}{r-1} \binom{v+r-1}{r}$$

is a stochastic T -matrix.

PROOF. ([1], 69). Putting in the well-known identity

$$\sum_{j=1}^v \binom{v+p-j}{v-j-1} \binom{r+j}{j} = \binom{v+p+r}{v-1}$$

first $r=0$ and then $p = r-1$, we obtain (7) for natural numbers r . Since moreover $d_{vj} \equiv p_{v1}$ and

$$\lim_{v \rightarrow \infty} p_{v1} = \lim_{v \rightarrow \infty} \frac{r}{v-1} = 0,$$

(8) also holds.

Example 4. If p_1, p_2, \dots is a sequence of positive elements, $P_v = p_1 + \dots + p_v$ and $\frac{p_v}{P_v} \rightarrow 0, v \rightarrow \infty$, then the matrix (6) built from the elements $p_{vj} = \frac{p_v - j + 1}{P_v}$, the so-called Nörlund matrix, is a stochastic T -matrix.

2.2. In this section we prove a theorem playing an important role in the construction of linear order statistics.

Theorem 1. *If the matrices S_k ($k=1, 2, \dots$) are stochastic T -matrices, then (4) holds.*

PROOF. Condition (8) being equivalent to

$$(9) \quad p_{v1}^{(k)^2} + \dots + p_{vv}^{(k)^2} \rightarrow 0, \quad v \rightarrow \infty,$$

we are going to show that (9) implies (4). The proof will proceed by induction.

For $s=2$ we have

$$G(B_{2v}) = \sum_{k \neq l} p_{vk}^{(1)} p_{vl}^{(2)} = 1 - \sum_{j=1}^v p_{vj}^{(1)} p_{vl}^{(2)}.$$

Using Cauchy's inequality, we thus obtain from (9) the statement to be proved for $s=2$.

Suppose now that the theorem is true up to the natural number m , i.e.

$$G(B_{\alpha v}) \rightarrow 1, \quad n \rightarrow \infty \quad (\alpha = 1, 2, \dots, m).$$

Let M be the matrix having v rows and m columns and having each of its elements equal to 1. By the expansion theorem of CAUCHY—BINET ([6], 579)

$$(10) \quad \frac{1}{m!} \text{Per}(B_{mv} M) = G(B_{mv}) + H(B_{mv}) = \prod_{j=1}^m (p_{v1}^{(j)} + \dots + p_{vv}^{(j)}) = 1.$$

By the induction hypothesis (10) implies

$$(11) \quad H(B_{mv}) \rightarrow 0, \quad n \rightarrow \infty.$$

Starting with the identity (10), we get

$$(12) \quad \begin{aligned} 1 &= (p_{v1}^{(m+1)} + \dots + p_{vv}^{(m+1)}) \prod_{j=1}^m (p_{v1}^{(j)} + \dots + p_{vv}^{(j)}) = \\ &= G(B_{mv})(p_{v1}^{(m+1)} + \dots + p_{vv}^{(m+1)}) + H(B_{mv}). \end{aligned}$$

On the other hand

$$(13) \quad \begin{aligned} &(p_{v1}^{(m+1)} + \dots + p_{vv}^{(m+1)})G(B_{mv}) = \\ &= \sum_{(k_1, \dots, k_m) \in \Pi_m^{(v)}} \text{Per}(B_1^{(k_1)} \dots B_1^{(k_m)})(p_{vk_1}^{(m+1)} + \dots + p_{vk_m}^{(m+1)}) + \\ &+ \sum_{(k_1, \dots, k_m) \in \Pi_m^{(v)}} \text{Per}(B_1^{(k_1)} \dots B_1^{(k_m)}) \overline{(p_{vk_1}^{(m+1)} + \dots + p_{vk_m}^{(m+1)})} \end{aligned}$$

where $\overline{p_{vk_1}^{(m+1)} + \dots + p_{vk_m}^{(m+1)}}$ is the sum of the $v-m$ members remaining from $p_{v1}^{(m+1)}, \dots, p_{vv}^{(m+1)}$ if we omit $p_{vk_1}^{(m+1)}, \dots, p_{vk_m}^{(m+1)}$. Clearly,

$$(14) \quad \sum_{(k_1, \dots, k_m) \in \Pi_m^{(v)}} \text{Per}(B_1^{(k_1)} \dots B_1^{(k_m)}) \overline{(p_{vk_1}^{(m+1)} + \dots + p_{vk_m}^{(m+1)})} = G(B_{m+1v}).$$

Moreover we see easily that

$$(15) \quad \begin{aligned} &\sum_{(k_1, \dots, k_m) \in \Pi_m^{(v)}} \text{Per}(B_1^{(k_1)} \dots B_1^{(k_m)})(p_{vk_1}^{(m+1)} + \dots + p_{vk_m}^{(m+1)}) = \\ &= \sum_{j=1}^v p_{vj}^{(m+1)} \sum_{\alpha=1}^m p_{vj}^{(\alpha)} G(B_{mv}^{(\alpha, j)}), \end{aligned}$$

where $B_{mv}^{(\alpha, j)}$ denotes the matrix obtained from B_{mv} by omitting the α -th row and the j -th column. Thus we infer from (12) by making use of the identities (13), (14) and (15) the identity

$$(16) \quad G(B_{m+1v}) + H(B_{mv}) + \sum_{j=1}^v p_{vj}^{(m+1)} \sum_{\alpha=1}^m p_{vj}^{(\alpha)} G(B_{mv}^{(\alpha, j)}) = 1.$$

Making use of the inequality $G(B_{mv}) \leq 1$ obtained from (10), we see that

$$1 \leq G(B_{m+1v}) + H(B_{mv}) + \sum_{\alpha=1}^m \sum_{j=1}^v p_{vj}^{(\alpha)} p_{vj}^{(m+1)}.$$

By virtue of (9) and of Cauchy's inequality the third member on the right hand side tends to zero if $n \rightarrow \infty$, and by (11) the same is true for the second member. Thus

$$\liminf_{n \rightarrow \infty} G(B_{m+1v}) \geq 1.$$

By one of our previous remarks

$$\limsup_{n \rightarrow \infty} G(B_{m+1v}) \leq 1$$

and so

$$G(B_{m+1v}) \rightarrow 1, \quad n \rightarrow \infty.$$

This completes the induction.

The converse of Theorem 1 fails to hold in general. Indeed, if (4) holds, then from the identity (16) we infer that for any pair of different natural numbers α, β the relation

$$\sum_{j=1}^v p_{vj}^{(\alpha)} p_{vj}^{(\beta)} \rightarrow 0, \quad n \rightarrow \infty$$

must be valid. This condition can however be satisfied even if (8) fails to hold. E.g. let

$$p_{v1}^{(\alpha)} = p_{vv}^{(\beta)} = \frac{1}{2}, \quad p_{vj}^{(\alpha)} = p_{vl}^{(\beta)} = \frac{1}{2(v-1)}$$

$$(j = 2, \dots, v; l = 1, \dots, v-1).$$

Then

$$\sum_{j=1}^v p_{vj}^{(\alpha)} p_{vj}^{(\beta)} = \frac{1}{2(v-1)} + \frac{v-2}{4(v-1)^2} \rightarrow 0, \quad v \rightarrow \infty$$

while (8) fails.

If, on the other hand, $S_k = S$ ($k=1, 2, \dots$), then the matrix B_{mv} consists of columnwise identical elements. In the detailed expression of S is given by (6), then

$$G(B_{mv}) = \sum_{(k_1, \dots, k_m) \in \Pi_m^{(v)}} p_{vk_1} \dots p_{vk_m}$$

and from (16) we obtain in case of the validity of (4) the condition $p_{v1}^2 + \dots + p_{vv}^2 \rightarrow 0, v \rightarrow \infty$. From this, however, (8) already follows. Thus we have established the following

Corollary 1. The stochastic matrix S is a stochastic T -matrix, if and only if

$$\sum_{(k_1, \dots, k_m) \in \Pi_m^{(v)}} p_{vk_1} \dots p_{vk_m} \rightarrow 1, \quad n \rightarrow \infty \quad (m = 1, 2, \dots).$$

3. Limit theorems

3.1. In this section we shall state and prove the limit theorems playing a central role in the present paper. In this first subsection we prove a theorem upon which we intend to build the proofs of the limit theorems just mentioned.

Let the matrix with complex elements

$$(17) \quad Z_{mv} = \begin{pmatrix} z_{11} & \dots & z_{1v} \\ \dots & \dots & \dots \\ z_{m1} & \dots & z_{mv} \end{pmatrix}$$

be given.

Theorem 2. *If the stochastic matrices having detailed from (3) satisfy condition (4), and if*

$$(18) \quad |z_{jk}| \cong 1 \quad (j = 1, \dots, m; k = 1, \dots, v),$$

then uniformly in z_{jk}

$$(19) \quad \lim_{n \rightarrow \infty} [G(Z_{mv} B_{mv}^*) - \text{Per}(Z_{mv} B_{mv}^*)] = 0 \quad (m = 1, 2, \dots).$$

PROOF. Again by the expansion theorem of Cauchy—Binet

$$(20) \quad \text{Per}(Z_{mv} B_{mv}^*) = G(Z_{mv} B_{mv}^*) + H(Z_{mv} B_{mv}^*),$$

where now we have

$$G(Z_{mv} B_{mv}^*) = \sum_{1 \leq k_1 < \dots < k_m \leq v} \text{Per}(Z_1^{(k_1)} \dots Z_1^{(k_m)}) \text{Per}(B_1^{(k_1)} \dots B_1^{(k_m)}),$$

$$H(Z_{mv} B_{mv}^*) = \sum_{\substack{\beta_1 + \dots + \beta_v = m \\ \beta_1^2 + \dots + \beta_v^2 > m}} \frac{1}{\beta_1! \dots \beta_v!} \text{Per}(Z_{\beta_1}^{(1)} \dots Z_{\beta_v}^{(v)}) \text{Per}(B_{\beta_1}^{(1)} \dots B_{\beta_v}^{(v)}).$$

By condition (18)

$$|\text{Per}(Z_{\beta_1}^{(1)} \dots Z_{\beta_v}^{(v)})| \leq m!$$

and so we get from (20)

$$|G(Z_{mv} B_{mv}^*) - \text{Per}(Z_{mv} B_{mv}^*)| \leq m! H(B_{mv}).$$

Conditions (4) being satisfied, we have $H(B_{mv}) \rightarrow 0, n \rightarrow \infty$, and this already yields the proof of our theorem.

On the basis of Theorem 1. one has the following

Corollary 2. If S_k ($k=1, 2, \dots$) is stochastic T -matrix and (18) holds, then the statement of (19) is valid.

If $S_k = S$ ($k=1, 2, \dots$) is the stochastic T -matrix having detailed from (6), then B_{mv} consists of columnwise equal elements. Thus our Theorem 2 yields the following

Corollary 3. If $S_k = S$ ($k=1, 2, \dots$) is stochastic T -matrix and (18) holds, then uniformly in z_{jk}

$$\lim_{n \rightarrow \infty} \left[\frac{1}{m!} G(Z_{mv} B_{mv}^*) - \prod_{j=1}^m (z_{j1} p_{v1} + \dots + z_{jv} p_{vv}) \right] = 0 \quad (m = 1, 2, \dots)$$

with

$$\frac{1}{m!} G(Z_{mv} B_{mv}^*) = \sum_{1 \leq k_1 < \dots < k_m \leq v} \text{Per}(Z_1^{(k_1)} \dots Z_1^{(k_m)}) p_{vk_1} \dots p_{vk_m}.$$

3.2. In the theory of linear order statistics as we have defined them, a role of fundamental importance is being played by the following

Theorem 3. Let the linear order statistics $\{A_k, S_k\}$ be given. If $\varphi_{m,n}(t)$ denotes the characteristic function of the random variable $\xi_{m,n}$, then uniformly in t

$$\lim_{n \rightarrow \infty} \left[\varphi_{m,n}(t) - \frac{1}{m! G(B_{mv})} \text{Per} \Phi_v(t) \right] = 0 \quad (m = 1, 2, \dots)$$

with

$$\Phi_v(t) = \begin{pmatrix} \varphi_{11}^{(v)}(t) & \dots & \varphi_{1m}^{(v)}(t) \\ \dots & \dots & \dots \\ \varphi_{m1}^{(v)}(t) & \dots & \varphi_{mm}^{(v)}(t) \end{pmatrix},$$

where $\varphi_{jk}^{(v)}(t)$ is the characteristic function of the random variable which takes the values $a_{v1}^{(j)}, \dots, a_{vv}^{(j)}$ with probability $p_{v1}^{(k)}, \dots, p_{vv}^{(k)}$ respectively.

PROOF. By the definition of characteristic functions, the function $\varphi_{m,n}(t)$ can be obtained by putting $z=e^{it}$ into the expression

$$\frac{1}{m! G(B_{mv})} \sum_{(k_1, \dots, k_m) \in \Pi_m^{(v)}} z^{a_{vk_1}^{(1)} + \dots + a_{vk_m}^{(m)}} \text{Per}(B_1^{(k_1)} \dots B_1^{(k_m)}).$$

Now, if

$$Z_{mv} = \begin{pmatrix} z^{a_{v1}^{(1)}} & \dots & z^{a_{vv}^{(1)}} \\ \dots & \dots & \dots \\ z^{a_{v1}^{(m)}} & \dots & z^{a_{vv}^{(m)}} \end{pmatrix},$$

then

$$\begin{aligned} \varphi_{m,n}(t) &= \frac{1}{m! G(B_{mv})} \sum_{1 \leq k_1 < \dots < k_m \leq v} \text{Per}(Z_1^{(k_1)} \dots Z_1^{(k_m)}) \text{Per}(B_1^{(k_1)} \dots B_1^{(k_m)}) = \\ &= \frac{G(Z_{mv} B_{mv}^*)}{m! G(B_{mv})}, \quad z = e^{it}. \end{aligned}$$

On the other hand,

$$[Z_{mv} B_{mv}^*]_{z=e^{it}} = \Phi_v(t).$$

Since $|e^{it a_{vj}^{(k)}}|=1$, we can establish the theorem by using formula (19) from Theorem 2.

The following theorems are direct consequences of Theorem 3:

Theorem 4. *The linear order statistics $\{A_k, S_k\}$ is asymptotic if and only if for $m=1, 2, \dots$ there exists the limit of the sequence*

$$\frac{1}{m!} \text{Per} \Phi_v(t) \quad (v = 1, 2, \dots), \quad t \in R_1$$

as $n \rightarrow \infty$, and this limit is continuous at the origin. Thus in this case

$$\lim_{n \rightarrow \infty} \varphi_{m,n}(t) = \lim_{n \rightarrow \infty} \frac{1}{m!} \text{Per} \Phi_v(t), \quad t \in R_1 \quad (m = 1, 2, \dots).$$

Theorem 5. *The linear order statistics $\{A_k, S_k\}$ is doubly asymptotic if and only if the sequence*

$$\frac{1}{m!} \text{Per} \Phi_v(t) \quad (v = 1, 2, \dots)$$

has a limit as $n \rightarrow \infty$ and then $m \rightarrow \infty, t \in R_1$, and this limit is continuous at the origin. Thus in this case

$$\lim_{\substack{n \rightarrow \infty \\ m \rightarrow \infty}} \varphi_{m,n}(t) = \lim_{\substack{n \rightarrow \infty \\ m \rightarrow \infty}} \frac{1}{m!} \text{Per} \Phi_v(t), \quad t \in R_1.$$

3.3. If $S_k = S$ ($k=1, 2, \dots$) then the linear order statistics corresponding to Definition 1 will be denoted by $\{A_k, S\}$.

Let $\eta_j^{(v)}$ be then random variable which takes the values $a_{v1}^{(j)}, \dots, a_{vv}^{(j)}$ with probability $p_{v1}^{(j)}, \dots, p_{vv}^{(j)}$ respectively.

From Corollary 3 we obtain the following

Theorem 6. Let the linear order statistics $\{A_k, S\}$ given. If $\varphi_{m,n}(t)$ and $\varphi_v^{(j)}(t)$ denotes the characteristic function of the random variable $\xi_{m,n}$ and $\eta_v^{(j)}$ respectively, then uniformly in t

$$\lim_{n \rightarrow \infty} \left[\varphi_{m,n}(t) - \frac{1}{G(B_{m_v})} \varphi_v^{(1)}(t) \dots \varphi_v^{(m)}(t) \right] = 0 \quad (m = 1, 2, \dots).$$

From Theorem 6 we infer the following two theorems:

Theorem 7. The linear order statistics $\{A_k, S\}$ is asymptotic if and only if for $m=1, 2, \dots$ there exists the limit of the sequence $\varphi_v^{(1)}(t) \dots \varphi_v^{(m)}(t)$ ($v=1, 2, \dots$) as $n \rightarrow \infty$, $t \in R_1$, and this limit is continuous at the origin. Then

$$(21) \quad \lim_{n \rightarrow \infty} \varphi_{m,n}(t) = \lim_{n \rightarrow \infty} [\varphi_v^{(1)}(t) \dots \varphi_v^{(m)}(t)], \quad t \in R_1 \quad (m = 1, 2, \dots).$$

Theorem 8. The linear order statistics $\{A_k, S\}$ is doubly asymptotic if and only if the sequence $\varphi_v^{(1)}(t) \dots \varphi_v^{(m)}(t)$ ($v=1, 2, \dots$) has a limit as $n \rightarrow \infty$ and then $m \rightarrow \infty$, $t \in R_1$ and this limit is continuous at the origin. Then

$$\lim_{\substack{n \rightarrow \infty \\ m \rightarrow \infty}} \varphi_{m,n}(t) = \lim_{\substack{n \rightarrow \infty \\ m \rightarrow \infty}} [\varphi_v^{(1)}(t) \dots \varphi_v^{(m)}(t)], \quad t \in R_1.$$

If for any natural number k there exists a characteristic function $\varphi^{(k)}(t)$ so that

$$\varphi_v^{(k)}(t) \rightarrow \varphi^{(k)}(t), \quad n \rightarrow \infty, \quad t \in R_1 \quad (k = 1, 2, \dots),$$

then on the basis of (21)

$$\lim_{n \rightarrow \infty} \varphi_{m,n}(t) = \varphi^{(1)}(t) \dots \varphi^{(m)}(t), \quad t \in R_1. \quad (m = 1, 2, \dots).$$

Thus we get the following

Theorem 9. If for each matrix A_k from the linear order statistics $\{A_k, S\}$ there exists a random variable $\eta^{(k)}$ such that $\eta_v^{(k)} \Rightarrow \eta^{(k)}$, $v \rightarrow \infty$, then this linear order statistics is asymptotically $\eta^{(1)} + \dots + \eta^{(m)}$ ($m=1, 2, \dots$) distributed and the random variables $\eta^{(1)}, \eta^{(2)}, \dots$ are independent.

4. The construction of linear order statistics

4.1. Theorem 9 and Corollary 1 make it possible to construct linear order statistics. For this purpose we must realize the following steps:

- a) To construct a stochastic T -matrix.
- b) To construct a sequence A_k ($k=1, 2, \dots$) of matrices, so that on each matrix A_k the sequence of discrete random variables defined with the help of the stochastic T -matrix is convergent in the weak sense.

In what follows, we shall construct linear order statistics in this manner.

Theorem 10. *If $0 < p < 1$ and $p + q = 1$ then the linear order statistics $\{A_k, S\}$ formed with the help of the quantities*

$$(22) \quad a_{vj}^{(k)} = \frac{j-1-(v-1)p}{\sqrt{(v-1)pq}}, \quad p_{vj} = \binom{v-1}{j-1} p^{j-1} q^{v-j}$$

$$(j = 1, \dots, v; k = 1, 2, \dots)$$

is asymptotically normally distributed with expectation zero and with variance m ($m=1, 2, \dots$).

PROOF. By virtue of example 2 in 2.1 the matrix (6) formed with the quantities p_{vj} of (22) is a stochastic T -matrix.

The random variable $\eta_v^{(k)}$ formed with quantities $a_{vj}^{(k)}$ of (22) has standardized binomial distribution with parameters p and $v-1$. By the theorem of Moivre—Laplace the sequence $\eta_v^{(k)}$ ($v=1, 2, \dots$) weakly converges to the normal distribution with zero expectation and unit variance. On the basis of this and with the help of Theorem 9 we obtain the proof of our theorem.

4.2. Let the triangular matrices with real elements

$$X = \begin{pmatrix} x_{11} & & & & \\ x_{21} & x_{22} & & & \\ \dots & \dots & \dots & \dots & \\ x_{v1} & x_{v2} & \dots & x_{vv} & \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}, \quad C = \begin{pmatrix} C_{11} & & & & \\ C_{21} & C_{22} & & & \\ \dots & \dots & \dots & \dots & \\ C_{v1} & C_{v2} & \dots & C_{vv} & \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

be given. The elements of X and of C will be called abscissas and Cotes numbers respectively. Suppose, that the elements of X fall into the interval $[a, b]$ and that the elements in each row of X are pairwise different. Then for any function $f(x)$ defined on the interval $[a, b]$, the expressions

$$Q_v(f) = \sum_{j=1}^v C_{vj} f(x_{vj}) \quad (v = 1, 2, \dots)$$

make sense. In case $f(x)$ is integrable on $[a, b]$ and

$$Q_v(f) \rightarrow \int_a^b f(x) dx, \quad v \rightarrow \infty,$$

we say that for the function $f(x)$ the quadrature process (X, C) belonging to the matrices X and C converges.

Theorem 11. *If the Cotes numbers are nonnegative and the quadrature process (X, C) is convergent for any continuous function defined in the interval $[a, b]$, then the matrix (6) formed with the quantities*

$$(23) \quad p_{vj} = \frac{C_{vj}}{Q_v(1)} \quad (j = 1, \dots, v; v = 1, 2, \dots)$$

is a stochastic T -matrix.

If, moreover, $f_k(x)$ ($k=1, 2, \dots$) are continuous functions defined on the interval $[a, b]$, and the elements of the matrix A_k are the numbers

$$(24) \quad a_{vj}^{(k)} = f_k(x_{vj}) \quad (j = 1, \dots, v; k, v = 1, 2, \dots),$$

then the linear order statistics $\{A_k, S\}$ has asymptotically

$$(25) \quad f_1(\zeta_1) + \dots + f_m(\zeta_m) \quad (m = 1, 2, \dots)$$

distribution, where ζ_1, ζ_2, \dots are random variables independent from each other, and uniform distributed in the interval $[a, b]$.

PROOF. If $f(x)=1$ then $\sum_{j=1}^v C_{vj} = Q_v(1)$, i.e. conditions (7) are valid for the numbers (23). Now the elements of the matrix C are nonnegative numbers, and the quadrature process (X, C) converges for any function defined and continuous in the interval $[a, b]$, so that (8) also holds ([7], 459, Satz 4). Thus however the matrix (6) formed with the quantities (23) must indeed be a stochastic T -matrix.

Just as before, we shall denote the discrete random variable made to correspond to the v -th row of the matrix A_k by $\eta_v^{(k)}$. Since $\cos f_k(x)$ and $\sin f_k(x)$ are continuous in the interval $[a, b]$ for continuous $f_k(x)$, we have

$$\begin{aligned} \varphi_{\eta_v^{(k)}}(t) &= E(e^{it\eta_v^{(k)}}) = E[\cos(\eta_v^{(k)}t)] + iE[\sin(\eta_v^{(k)}t)] = \\ &= \frac{1}{Q_v(1)} [Q_v(\cos[tf_k(x)]) + iQ_v(\sin[tf_k(x)])] \rightarrow \\ &\rightarrow \frac{1}{b-a} \left[\int_a^b \cos[tf_k(x)] dx + i \int_a^b \sin[tf_k(x)] dx \right] = \\ &= \frac{1}{b-a} \int_a^b e^{itf_k(x)} dx = \varphi_{f_k(\zeta_k)}(t), \quad v \rightarrow \infty, \end{aligned}$$

i.e. $\eta_v^{(k)} \Rightarrow f_k(\zeta_k)$, $v \rightarrow \infty$. The second statement of our theorem follows from this directly with the help of Theorem 9.

4.3. If between the matrices X and C there exists the relation

$$C_{vj} = \int_a^b l_{vj}(x) dx \quad (j = 1, \dots, v; v = 1, 2, \dots),$$

where

$$(26) \quad l_{vj}(x) = \frac{\omega_v(x)}{\omega'_v(x_{vj})(x - x_{vj})}, \quad \omega_v(x) = \prod_{j=1}^v (x - x_{vj})$$

is the Lagrange-polynomial belonging to the abscissas x_{v1}, \dots, x_{vv} , then we say that the quadrature process (X, C) is an interpolation quadrature.

Theorem 12. If the Cotes numbers of the interpolation quadrature (X, C) are nonnegative, then the matrix (6) formed with the quantities (23) is a stochastic T -matrix

If, moreover, $f_k(x)$ ($k=1, 2, \dots$) are bounded Riemann-integrable functions defined on the interval $[a, b]$ and the elements of A_k are the quantities (24), then the linear order statistics $\{A_k, S\}$ has asymptotically the distribution (25).

PROOF. By Fejér's theorem ([2], Satz I) an interpolation quadrature process with nonnegative coefficients converges for any bounded Riemann-integrable function. Thus we can prove our theorem by employing the method of proof of Theorem 11.

From this theorem we obtain as corollaries the following results:

a) If $a = -1$, $b = 1$ and the x_{vj} are the roots of the Čebišev polynomial $T(x)$, e.i.

$$x_{vj} = \cos \frac{2j-1}{2v} \pi \quad (j = 1, \dots, v)$$

then, as was shown by L. FEJÉR ([2], Satz II),

$$C_{vj} = \int_{-1}^1 \frac{T_v(x)}{(x-x_{vj})T_v(x_{vj})} dx > 0.$$

Thus, by putting $f_k(x)=x$ in Theorem 12 we obtain the following

Corollary 4. If

$$p_{vj} = \frac{1}{2} \int_{-1}^1 \frac{T_v(x)}{(x-x_{vj})T_v(x_{vj})} dx, \quad a_{vj}^{(k)} = x_{vj}$$

$$(j = 1, \dots, v; k, v = 1, 2, \dots),$$

where

$$x_{vj} = \cos \frac{2j-1}{2v} \pi \quad (j = 1, \dots, v)$$

is a root of the Čebišev-polynomial $T_v(x)$, then the linear order statistics $\{A_k, S\}$ has asymptotically $\zeta_1 + \dots + \zeta_m$ ($m=1, 2, \dots$) distribution, where ζ_1, ζ_2, \dots are independent random variables uniformly distributed in the interval $[-1, 1]$.

b) If $a = -1$, $b = 1$ and the abscissas are the roots of the Čebišev-polynomials of second kind

$$U_v(x) = \frac{\sin(v+1) \arccos x}{\sqrt{1-x^2}}$$

i.e.

$$x_{vj} = \cos \frac{j\pi}{v+1} \quad (j = 1, \dots, v),$$

then the Cotes numbers C_{vj} are positive ([2], Satz III.). If we again put $f_k(x)=x$ in the Theorem 12, we obtain the following

Corollary 5. If

$$p_{vj} = \frac{1}{2} C_{vj}, \quad a_{vj}^{(k)} = \cos \frac{j\pi}{v+1} \quad (j = 1, \dots, v; k, v = 1, 2, \dots),$$

where the C_{vj} are the Cotes numbers belonging to the abscissas formed from the roots of the Čebišev polynomials of second kind, then the linear order statistics $\{A_k, S\}$ has asymptotically $\zeta_1 + \dots + \zeta_m$ ($m=1, 2, \dots$) distribution, where ζ_1, ζ_2, \dots are independent random variables, uniformly distributed in the interval $[-1, 1]$.

4.4. Let us suppose that the density function $p(x)$ of the random variable ξ defined on the interval $[a, b]$ is positive outside a set of measure zero. Let $\{\omega_v(x)\}$ be the system of orthogonal polynomials belonging to the density function $p(x)$. As is known, the roots $x_{v1}, x_{v2}, \dots, x_{vv}$ of the polynomial $\omega_v(x)$ fall into the interval $[a, b]$ and have multiplicity one. If we now choose these roots as abscissas and

$$(27) \quad C_{vj} = \int_a^b p(x) l_{vj}(x) dx \quad (j = 1, \dots, v; v = 1, 2, \dots),$$

where the polynomials $l_{vj}(x)$ are defined by formula (26), then the interpolation quadrature process (X, C) will be called Gaussian quadrature.

As was shown by T. I. STIELTJES ([9]), the quantities (27) are positive, and for any function $f(x)$ defined and continuous on the interval $[a, b]$

$$(28) \quad Q_v(f) \rightarrow \int_a^b f(x)p(x) dx.$$

Theorem 13. If (X, C) is a Gaussian quadrature, then the matrix (6) formed with the quantities $p_{vj} = C_{vj}$ is a stochastic T -matrix. If, moreover, $f_k(x)$ ($k=1, 2, \dots$) are function defined and continuous in the interval $[a, b]$, and the elements of the matrix A_k are the numbers (24), then the linear order statistics $\{A_k, S\}$ is asymptotically

$$f_1(\xi_1) + \dots + f_m(\xi_m) \quad (m = 1, 2, \dots)$$

distributed, where ξ_1, ξ_2, \dots are independent random variables with the common density function $p(x)$.

PROOF. We have ([7], 438)

$$\sum_{j=1}^v C_{vj} = \int_a^b p(x) dx = 1,$$

and by the theorem quoted of Stieltjes the numbers C_{vj} are positive, so that our first statement follows by (28) from formula (8) valid in the present case too.

In order to establish the second statement of our theorem, we first show — in exactly the same manner as was done for the analogous statement in Theorem 10 — that $\eta_v^{(k)} \Rightarrow f_k(\xi)$, $v \rightarrow \infty$, where ξ is a random variable with density function $p(x)$. Hence, with the help of Theorem 9, the statement itself follows.

Now we are going to consider special cases of Theorem 13.

a) If $a = -1$, $b = 1$, $p(x) = 1$, then the orthogonal polynomials are the Legendre polynomials

$$P_v(x) = \frac{1}{(2n!)!} \frac{d^v}{dx^v} (x^2 - 1)^v, \quad v = 1, 2, \dots$$

If this has roots x_{v1}, \dots, x_{vv} , then it can be shown ([7], 443) that

$$C_{vj} = \frac{2}{1-x_{vj}^2} \frac{1}{[P'_v(x_{vj})]^2}.$$

So, by making the choice $f_k(x)=x$, we get from Theorem 13 the following

Corollary 6. If

$$p_{vj} = \frac{1}{1-x_{vj}^2} \frac{1}{[P'_v(x_{vj})]^2}, \quad a_{vj}^{(k)} = x_{vj} \quad (j = 1, \dots, v; k, v = 1, 2, \dots),$$

where x_{v1}, \dots, x_{vv} are the roots of the Legendre polynomial $P_v(x)$, then the linear order statistics $\{A_k, S\}$ is asymptotically $\zeta_1 + \dots + \zeta_m$ ($m=1, 2, \dots$) distributed, where ζ_1, ζ_2, \dots are independent random variables, uniformly distributed in the interval $[-1, 1]$.

b) If $a = -1, b = 1, p(x) = \frac{1}{\sqrt{1-x^2}}$, then the orthogonal polynomials are the Čebišev polynomials

$$T_v(x) = \cos(v \arccos x) \quad (v = 1, 2, \dots).$$

This polynomial has roots

$$x_{vj} = \cos \frac{2j-1}{2v} \pi \quad (j = 1, \dots, v)$$

and $C_{vj} = \frac{\pi}{v}$ ([7], 444—446). Thus, again by putting $f_k(x) = x$, we obtain from Theorem 13 the following

Corollary 7. If

$$p_{vj} = \frac{1}{v}, \quad a_{vj}^{(k)} = \cos \frac{2j-1}{2v} \pi \quad (j = 1, \dots, v; k, v = 1, 2, \dots),$$

then the linear order statistics $\{A_k, S\}$ is asymptotically $\xi_1 + \dots + \xi_m$ ($m=1, 2, \dots$) distributed, where ξ_1, ξ_2, \dots are independent random variables having the same distribution function $\frac{1}{\pi\sqrt{1-x^2}}, x \in (-1, 1)$.

c) If $a = -1, b = 1, p(x) = \sqrt{1-x^2}$, then the orthogonal polynomials are the Čebišev polynomials of the second kind

$$U_v(x) = \frac{\sin[(v+1) \arccos x]}{\sqrt{1-x^2}} \quad (v = 1, 2, \dots).$$

For these polynomials ([7], 447—449)

$$x_{vj} = \cos \frac{j\pi}{v+1}, \quad C_{vj} = \frac{\pi}{v+1} \sin^2 \frac{j\pi}{v+1} \quad (j = 1, \dots, v).$$

Thus, again by putting $f_k(x)=x$, we obtain from Theorem 13 the following

Corollary 8. If

$$p_{vj} = \frac{2}{v+1} \sin^2 \frac{j\pi}{v+1}, \quad a_{vj}^{(k)} = \cos \frac{j\pi}{v+1} \quad (j = 1, \dots, v; k, v = 1, 2, \dots),$$

then the linear order statistics $\{A_k, S\}$ is asymptotically $\xi_1 + \dots + \xi_m$ ($m=1, 2, \dots$) distributed, where ξ_1, ξ_2, \dots are independent random variables having the same density function $\frac{2}{\pi} \sqrt{1-x^2}$.

4.5. The simplest and most general method for constructing linear order statistics with a given limit distribution is probably the one based on one of the criterion of Riemann-integrability.

The decomposition of the interval $[0, 1]$ into disjoint subintervals realized by the point of division

$$(29) \quad 0 = x_{v0} < x_{v1} < \dots < x_{vv} = 1$$

will be called a distinguished decomposition sequence, if the matrix (6) formed with the numbers

$$(30) \quad p_{vj} = x_{vj} - x_{vj-1} \quad (j = 1, \dots, v; v = 1, 2, \dots)$$

is a stochastic T -matrix.

Let y_{vj} ($j=1, \dots, v$) be an arbitrary point in the interval determined by the points of decomposition x_{vj-1} and x_{vj} .

The function $f(x)$ defined and bounded in the interval $[0, 1]$ is Riemann-integrable if and only if the sequence

$$\sum_{j=1}^v p_{vj} f(y_{vj}) \quad (v = 1, 2, \dots)$$

converges for any distinguished decomposition sequence. The limit is then the Riemann-integral of the function $f(x)$ on the interval $[0, 1]$.

Let the random variable η_v be defined by

$$P[\eta_v = f(y_{vj})] = p_{vj} \quad (j = 1, \dots, v).$$

Since together with $f(x)$ the functions $\cos f(x)$ and $\sin f(x)$ are also Riemann-integrable in the interval $[0, 1]$, we infer by following word for word the second statement of Theorem 11, that $\eta_v \Rightarrow f(\eta)$, $v \rightarrow \infty$, where η is a random variable uniformly distributed in the interval $[0, 1]$.

On the basis of Theorem 9 we have the following

Theorem 14. Let (29) be a distinguished decomposition sequence of the interval $[0, 1]$ and let S denote the stochastic T -matrix formed with the elements (30). Let $y_{vj}^{(k)}$ be an arbitrary point in the interval determined by the points x_{vj-1} and x_{vj} .

If $f_k(x)$ ($k=1, 2, \dots$) is a Riemann-integrable function in the interval $[0, 1]$ and the elements of the matrix A_k are given by

$$a_{vj}^{(k)} = f_k(y_{vj}^{(k)}) \quad (j = 1, \dots, v; v = 1, 2, \dots),$$

then the linear order statistics $\{A_k, S\}$ is asymptotically $f_1(\eta_1) + \dots + f_m(\eta_m)$ ($m=1, 2, \dots$) distributed, where η_1, η_2, \dots are independent random variables uniformly distributed in the interval $[0, 1]$.

If, in particular, the decomposition sequence is equistant, i.e. $x_{vj} = \frac{j}{v}$, then by example 1 from 2.1 the matrix (6) built from the elements $p_{vj} = \frac{1}{v}$ is a stochastic T -matrix, and so the decomposition sequence is distinguished. If, moreover, $y_{vj} = \frac{j}{v}$, then we obtain the following

Corollary 9. If $f(x)$ is a function Riemann-integrable in the interval $[0, 1]$ and

$$p_{vj} = \frac{1}{v}, \quad a_{vj}^{(k)} = f\left(\frac{j}{v}\right) \quad (j = 1, \dots, v; k, v = 1, 2, \dots)$$

then the linear statistics $\{A_k, S\}$ is asymptotically $f(\eta_1) + \dots + f(\eta_m)$ ($m = 1, 2, \dots$) distributed, where η_1, η_2, \dots are independent random variables, uniformly distributed in the interval $[0, 1]$.

By putting $f(x) = x$ in Corollary 9, we obtain the following

Corollary 10. If

$$p_{vj} = \frac{1}{v}, \quad a_{vj}^{(k)} = \frac{j}{v} \quad (j = 1, \dots, v; k, v = 1, 2, \dots),$$

then the linear order statistics $\{A_k, S\}$ is asymptotically $\eta_1 + \dots + \eta_m$ ($m = 1, 2, \dots$) distributed, where η_1, η_2, \dots are independent random variables, uniformly distributed in the interval $[0, 1]$.

The corollary just formulated is a case of Wilcoxon statistics. Since η_1 has expectation $\frac{1}{2}$ and variance $\frac{1}{12}$, $\eta_1 + \dots + \eta_m$ is asymptotically normally distributed with expectation $\frac{m}{2}$ and variance $\frac{m}{12}$.

References

- [1] R. G. COOKE, Infinite matrices and sequence spaces. *London*, 1950.
- [2] L. FEJÉR, Mechanische Quadraturen mit positiven Cotesschen Zahlen. *Math. Zeitschr.* **37** (1933), 287—309.
- [3] B. V. GNEDENKO—A. N. KOLMOGOROV, Limit distribution of sums of independent random variables. (Russian), *Moscow*, 1949. English transl., *Cambridge, Mass.*, 1954.
- [4] B. GYIRES, Limit distribution of linear rankstatistics. *Publ. Math. (Debrecen)* **21** (1974), 95—112.
- [5] B. GYIRES, The construction of linear rankstatistics with the help of pseudo random numbers. *Publ. Math. (Debrecen)* **21** (1974), 225—232.
- [6] M. MARCUS—M. MINC, Permanents. *Amer. Math. Monthly* **72** (1965), 577—591.
- [7] J. P. NATANSON, Konstruktive Funktionentheorie. *Berlin*, 1955.
- [8] L. SCHMETTERER, Einführung in die mathematische Statistik. *Wien*, 1956.
- [9] T. I. STIELTJES, Quelques recherches sur la théorie des quadratures dites mécaniques. *Ann. de l'École Norm. Sup.* **1** (1884), 409—426.

(Received Oktober 5, 1973.)