

On limit distribution theorems of linear order statistics

By BÉLA GYIRES (Debrecen)

Introduction

Let

$$x_1, \dots, x_m, y_1, \dots, y_n$$

be pairwise different real numbers. If the same real numbers rearranged in increasing order yield

$$z_1 < z_2 < \dots < z_{m+n}$$

with $x_k = z_{r_k}$, then we say that with respect to this ordering rank $x_k = r_k$.

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

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

Finally, let $\xi_1, \dots, \xi_m$ and $\eta_1, \dots, \eta_n$ be samples corresponding to ξ and η respectively, where ξ and η are independent random variables with a common, continuous distribution function.

In the theory of ordered samples, a fundamental role is played the following theorem ([6], 363. Satz 10):

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

Let

$$\begin{pmatrix} a_{11} & \dots & a_{1m+n} \\ \vdots & \dots & \vdots \\ a_{m1} & \dots & a_{mm+n} \end{pmatrix}.$$

be a matrix of real elements, then on the basis of the statistics

$$\xi_{m,n} = a_{1r_1} + \dots + a_{mr_m}, \text{ if } (\xi_1, \dots, \xi_m) \in \omega_{r_1, \dots, r_m}$$

defined with the help of this matrix, we can — by virtue of the theorem just quoted — decide whether to adopt or to reject the hypothesis

$$H_0: P(\xi < x) = P(\eta < x).$$

If $a_{kj}=j$ ($j = 1, \dots, m+n$), we get the well-known Wilcoxon-statistics.

The aim of the present paper is to investigate the asymptotic behavior of the so called linear orderstatistics ([4], 57), first for $n \rightarrow \infty$ and then for $m \rightarrow \infty$.

In Chapter 1. we obtain a theorem for the characteristic function of linear order statistics, and on the basis of this theorem, of fundamental importance for the whole paper, we obtain sufficient and, then necessary and sufficient conditions for a linear order statistics to have an asymptotic distribution. Also we give necessary and sufficient conditions for a simple linear order statistics to have an asymptotically normal distribution.

In Chapter 2., building on GÁBOR SZEGŐ's result concerning the eigenvalues of Toeplitz- and Hankel-matrices belonging to a given function, we investigate distributions representable as limits in the weak sense of discrete uniform distributions.

Making use of the results of Chapter 2. we give in Chapter 3. a method for constructing linear order statistics with given asymptotics.

In the whole paper, a fundamental role is played by the convergence in the weak sense of random variables. For different definitions of this notion see [1], 37—38, 58. As in [1], weak convergence will be denoted by $\Rightarrow$.

1. On the characteristic function of linear order statistics

1.1. Let the matrices with real elements

$$(1) \quad A_v = \begin{pmatrix} a_{11}^{(v)} & \dots & a_{1v}^{(v)} \\ a_{21}^{(v)} & \dots & a_{2v}^{(v)} \\ \vdots & \dots & \vdots \end{pmatrix} \quad (v = 1, 2, \dots)$$

be given. Let us define the random variable $\eta_j^{(v)}$ ($j=1, 2, \dots$) on the matrix A_v as follows:

If $\alpha_1, \dots, \alpha_s$ ($1 \leq s \leq v$) are pairwise different natural numbers and k_j ($j=1, \dots, s$) are arbitrary different numbers from the numbers $1, \dots, v$, then

$$P(\eta_{\alpha_1}^{(v)} = a_{\alpha_1 k_1}^{(v)}, \dots, \eta_{\alpha_s}^{(v)} = a_{\alpha_s k_s}^{(v)}) = \frac{1}{v(v-1)\dots(v-s+1)}.$$

From this definition we infer that $\eta_j^{(v)}$ is a uniformly distributed discrete random variable, namely

$$(2) \quad P(\eta_j^{(v)} = a_{jk}^{(v)}) = \frac{1}{v} \quad (k = 1, \dots, v).$$

Definition 1.1. By the linear order statistics generated by the random variables $\eta_1^{(m+n)}, \dots, \eta_m^{(m+n)}$ we mean the random variable

$$(3) \quad \xi_{m,n} = \eta_1^{(m+n)} + \dots + \eta_m^{(m+n)},$$

n being a non-negative integer.

Definition 1.2. By the linear order statistics generated by the matrices (1) we mean the ensemble of the random variables

$$\xi_{m,n} \quad (m = 1, 2, \dots; n = 0, 1, 2, \dots).$$

Definition 1.3. The linear order statistics generated by the matrices (1) are 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 4.1. The linear order statistics generated by the matrices (1) are doubly asymptotic, if there exists a random variable ξ such that

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

We are going also to speak about asymptotically ξ_m distributed ($m=1, 2, \dots$), and about doubly asymptotically ξ distributed linear order statistics respectively.

Clearly, the asymptotic order statistics generated by the matrices (1) are doubly asymptotically ξ distributed, if and only if $\xi_m \Rightarrow \xi, m \rightarrow \infty$.

Let us denote by $\Pi_{m+n}^{(m)}$ the set of variation of order m without repetition that can be formed from the elements $1, \dots, m+n$. Then, on the basis of Definition 1.1.

$$P(\xi_{m,n} = a) = \frac{1}{(n+1) \dots (m+n)} \sum_{a_{1k_1}^{(m+n)} + \dots + a_{mk_m}^{(m+n)} = a} 1, \quad (k_1, \dots, k_m) \in \Pi_{m+n}^{(m)}.$$

1.2. Let us put, for of sake brevity, $m+n = v$. We propose to determine the expectation and the variance of $\xi_{m,n}$. In order to be able to do this, we first consider the expectation and the variance of the random variables $\eta_j^{(v)}$ ($j=1, 2, \dots$). Let us prove the following

Theorem 1.1. The random variables $\eta_j^{(v)}$ ($j=1, 2, \dots$) are asymptotically uncorrelated. Moreover

$$\frac{D^2(\xi_{m,n})}{\sum_{j=1}^m D^2(\eta_j^{(m+n)})} \rightarrow 1, \quad n \rightarrow \infty.$$

PROOF. Let us denote the vector formed from the elements of the j -th row of the matrix A_v by $\mathfrak{A}_j$, and let e be the vector of v -th order having all its elements equal to 1. $\mathfrak{A}_j \mathfrak{A}_k$ is the inner product of $\mathfrak{A}_j$ and $\mathfrak{A}_k$, $\mathfrak{A}_j e$ that of $\mathfrak{A}_j$ and e , and $\mathfrak{A}_j^2$ that of $\mathfrak{A}_j$ by itself. Making use of these notations, we get

$$(4) \quad E(\eta_j^{(v)}) = \frac{1}{v} (\mathfrak{A}_j e),$$

$$(5) \quad D^2(\eta_j^{(v)}) = \frac{1}{v} \mathfrak{A}_j^2 - \frac{1}{v^2} (\mathfrak{A}_j e)^2.$$

Since for $j \neq k$

$$E(\eta_j^{(v)} \eta_k^{(v)}) = \sum_{\alpha, \beta=1}^v a_{j\alpha}^{(v)} a_{k\beta}^{(v)} P(\eta_j^{(v)}) = a_{j\alpha}^{(v)}, \quad \eta_k^{(v)} = a_{k\beta}^{(v)} = \frac{1}{v(v-1)} [(\mathfrak{A}_j e)(\mathfrak{A}_k e) - \mathfrak{A}_j \mathfrak{A}_k],$$

we get

$$E[(\eta_j^{(v)} - E(\eta_j^{(v)}))(\eta_k^{(v)} - E(\eta_k^{(v)}))] = \frac{1}{v^2(v-1)} (\mathfrak{A}_j e)(\mathfrak{A}_k e) - \frac{1}{v(v-1)} \mathfrak{A}_j \mathfrak{A}_k;$$

thus the correlation coefficient of $\eta_j^{(v)}$ and $\eta_k^{(v)}$ is

$$\varrho(\eta_j^{(v)}, \eta_k^{(v)}) = \frac{1}{v-1} \frac{(\mathfrak{A}_j e)(\mathfrak{A}_k e) - v(\mathfrak{A}_j \mathfrak{A}_k)}{\sqrt{[v\mathfrak{A}_j^2 - (\mathfrak{A}_j e)^2][v\mathfrak{A}_k^2 - (\mathfrak{A}_k e)^2]}}.$$

Let the random variables η_j and η_k be defined by the formulae

$$P(\eta_j = a_{j\alpha}^{(v)}) = P(\eta_k = a_{k\alpha}^{(v)}) = \frac{1}{v},$$

$$P(\eta_j = a_{j\alpha}^{(v)}, \eta_k = a_{k\beta}^{(v)}) = \frac{1}{v} \delta_{\alpha\beta} \quad (\alpha, \beta = 1, \dots, v),$$

where $\delta_{\alpha\beta}$ is the Kronecker symbol. Then

$$E(\eta_j) = \frac{1}{v} (\mathfrak{A}_j e), \quad D^2(\eta_j) = \frac{1}{v} \mathfrak{A}_j^2 - \frac{1}{v^2} (\mathfrak{A}_j e)^2,$$

$$E(\eta_j \eta_k) = \frac{1}{v} \mathfrak{A}_j \mathfrak{A}_k,$$

and thus, if we still write $\varrho(\eta_j, \eta_k) = -\varrho_{jk}$, the equality

$$\varrho(\eta_j^{(v)}, \eta_k^{(v)}) = \frac{1}{v-1} \varrho_{jk}.$$

results. Since $|\varrho_{jk}| \leq 1$, according to the first part of our statement

$$\varrho(\eta_j^{(v)}, \eta_k^{(v)}) \rightarrow 0, \quad v \rightarrow \infty.$$

On the basis of (2) and of (3)

$$E(\xi_{m,n}) = \frac{1}{v} \sum_{j=1}^m \mathfrak{A}_j e$$

and

$$D^2(\xi_{m,n}) = \sum_{j=1}^m D^2(\eta_j^{(v)}) + \sum_{j \neq k} \varrho(\eta_j^{(v)}, \eta_k^{(v)}) D(\eta_j^{(v)}) D(\eta_k^{(v)}).$$

In view of

$$\frac{\sum_{j \neq k} \varrho_{jk} D(\eta_j^{(v)}) D(\eta_k^{(v)})}{\sum_{j=1}^m D^2(\eta_j^{(v)})} \leq \sum_{j \neq k} \frac{D(\eta_j^{(v)}) D(\eta_k^{(v)})}{\sum_{j=1}^m D^2(\eta_j^{(v)})}$$

and of

$$\frac{D(\eta_j^{(v)}) D(\eta_k^{(v)})}{\sum_{j=1}^m D^2(\eta_j^{(v)})} \leq \frac{D(\eta_j^{(v)}) D(\eta_k^{(v)})}{D^2(\eta_j^{(v)}) + D^2(\eta_k^{(v)})} \leq \frac{D(\eta_j^{(v)}) D(\eta_k^{(v)})}{2 \sqrt{D^2(\eta_j^{(v)}) D^2(\eta_k^{(v)})}} = \frac{1}{2},$$

we obtain

$$\left| \frac{D^2(\xi_{m,n})}{\sum_{j=1}^m D^2(\eta_j^{(v)})} - 1 \right| \leq \frac{1}{v-1} \frac{m(m-1)}{2}$$

and from this we infer the second part of our statement.

1.3. Let the matrix with complex elements

$$Z = \begin{pmatrix} z_{11} & \dots & z_{1m+n} \\ \cdot & \dots & \cdot \\ z_{m1} & \dots & z_{mm+n} \end{pmatrix}$$

be given. Denote by $Z_{\beta_k}^{(k)}$ the matrix of m rows and β_k columns, each column of which is equal to the k -th column of the matrix Z . Denote by $(Z_{\beta_1}^{(1)} \dots Z_{\beta_{m+n}}^{(m+n)})$ the matrix of m rows and $\beta_1 + \dots + \beta_{m+n}$ columns, obtained by writing the matrices $Z_{\beta_1}^{(1)}, \dots, Z_{\beta_{m+n}}^{(m+n)}$ one after the other. If $\beta_k = 0$, then the k -th column of the matrix Z is missing from the matrix $(Z_{\beta_1}^{(1)} \dots Z_{\beta_{m+n}}^{(m+n)})$. Let

$$G_{m,n}(Z) = \frac{1}{(n+1) \dots (n+m)} \sum_{\substack{\beta_1 + \dots + \beta_{m+n} = m \\ \beta_1^2 + \dots + \beta_{m+n}^2 = m}} \text{Per}(Z_{\beta_1}^{(1)} \dots Z_{\beta_{m+n}}^{(m+n)}),$$

where Per denotes the permanent of the matrix following it. In the sequel, a fundamental role will be played by the following

Lemma. *If*

$$(6) \quad |z_{jk}| \leq 1 \quad (j = 1, \dots, m; k = 1, \dots, m+n),$$

then uniformly in z_{jn}

$$(7) \quad G_{m,n}(Z) - \frac{\prod_{j=1}^m (z_{j1} + \dots + z_{jm+n})}{(n+1) \dots (n+m)} \rightarrow 0, \quad n \rightarrow \infty \quad (m = 1, 2, \dots).$$

PROOF. Let S be the matrix with $m+n$ rows and m columns having all its elements equal to 1 then, by virtue of the Cauchy—Binet expansion theorem ([5], 579), we have

$$(8) \quad \frac{1}{(n+1) \dots (n+m)m!} \text{Per}(ZS) = G_{m,n}(Z) + H_{m,n}(Z),$$

where

$$H_{m,n}(Z) = \frac{1}{(n+1) \dots (n+m)} \sum_{\substack{\beta_1 + \dots + \beta_{m+n} = m \\ \beta_1^2 + \dots + \beta_{m+n}^2 > m}} \frac{1}{\beta_1! \dots \beta_{m+n}!} \text{Per}(Z_{\beta_1}^{(1)} \dots Z_{\beta_{m+n}}^{(m+n)}).$$

By condition (6)

$$|\text{Per}(Z_{\beta_1}^{(1)} \dots Z_{\beta_{m+n}}^{(m+n)})| \leq m!$$

and so

$$|H_{m,n}(Z)| \leq \frac{1}{(m+1)\dots(n+m)} [(n+m)^m - (n+1)\dots(n+m)] =$$

$$= \frac{\left(1 + \frac{m}{n}\right)^m}{\left(1 + \frac{1}{n}\right) \dots \left(1 + \frac{m}{n}\right)} - 1.$$

Taking now into account that

$$\text{Per}(ZS) = m! \prod_{j=1}^m (z_{j1} + \dots + z_{jm+n}),$$

we get on the basis of (8) the limit (7), as stated.

We now prove the following

Theorem 1.2. *If the random variables $\xi_{m,n}$ and $\eta_j^{(v)}$ have characteristic functions $\varphi_{m,n}(t)$ and $\varphi_j^{(v)}(t)$ respectively, then uniformly in $t \in R_1$*

$$\left[\varphi_{m,n}(t) - \frac{(n+m)^m}{(n+1)\dots(n+m)} \varphi_1^{(v)}(t) \dots \varphi_m^{(v)}(t) \right] \rightarrow 0, \quad n \rightarrow \infty \quad (m = 1, 2, \dots).$$

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

$$\frac{1}{(n+1)\dots(n+m)} \sum_{(k_1, \dots, k_m) \in \Pi_v^{(m)}} z^{a_{1k_1}^{(v)} + \dots + a_{mk_m}^{(v)}} =$$

$$= \frac{1}{(n+1)\dots(n+m)} \sum_{1 \leq k_1 < \dots < k_m \leq v} \text{Per} \begin{pmatrix} z^{a_{1k_1}^{(v)}} & \dots & z^{a_{1k_m}^{(v)}} \\ \vdots & \dots & \vdots \\ z^{a_{mk_1}^{(v)}} & \dots & z^{a_{mk_m}^{(v)}} \end{pmatrix}.$$

One easily sees that this expression is equal to the function $G_{m,n}(Z)$ occurring in our lemma, if here

$$Z = \begin{pmatrix} z^{a_{1k_1}^{(v)}} & \dots & z^{a_{1k_m}^{(v)}} \\ \vdots & \dots & \vdots \\ z^{a_{mk_1}^{(v)}} & \dots & z^{a_{mk_m}^{(v)}} \end{pmatrix}.$$

On the other hand

$$\varphi_j^{(v)}(t) = \frac{1}{v} [z^{a_{j1}^{(v)}} + \dots + z^{a_{jv}^{(v)}}], \quad z = e^{it}.$$

Since $|(e^{it})^{a_{jk}^{(v)}}| = 1$, on applying our lemma, we obtain the statement of our theorem.

From Theorem 1.2. we directly infer the following theorems:

Theorem 1.3. *The linear order statistics generated by the matrices (1) are asymptotic if and only if, the limit*

$$\varphi_1^{(m+n)}(t) \dots \varphi_m^{(m+n)}(t), \quad n \rightarrow \infty, \quad t \in R_1 \quad (m = 1, 2, \dots)$$

exists and is continuous at zero.

Theorem 1.4. *The linear order statistics generated by the matrices (1) are doubly asymptotic if and only if*

$$\varphi_1^{(m+n)}(t) \dots \varphi_m^{(m+n)}(t), \quad n \rightarrow \infty, \quad m \rightarrow \infty, \quad t \in R_1$$

exists and is continuous at zero.

Let $\mathcal{E}_1$ be the set of the uniformly distributed discrete random variables and if $\eta^{(v)} \in \mathcal{E}_1$ the index v denote that the probability belonging to $\eta^{(v)}$ is $1/v$. Let $\mathcal{E}_2$ be the set of those random variables η which can be represented in the form

$$\eta^{(v)} \Rightarrow \eta, \quad v \rightarrow \infty, \quad \eta^{(v)} \in \mathcal{E}_1$$

Theorem 1.5. *If $\eta_j \in \mathcal{E}_2$ ($j = 1, 2, \dots$), i.e. if*

$$\eta_j^{(v)} \Rightarrow \eta_j, \quad v \rightarrow \infty, \quad \eta_j^{(v)} \in \mathcal{E}_1,$$

then the linear order statistics generated by the matrices (1) determined by the points of discontinuity of the random variables $\eta_j^{(v)}$ ($v=1, 2, \dots; j=1, 2, \dots$), are asymptotically equal to the sums of the random variables $\eta_1, \dots, \eta_m$ ($m=1, 2, \dots$) independent from each other.

PROOF. Denote by $\varphi_j^{(v)}(t)$ and by $\varphi_j(t)$ the characteristic function of $\eta_j^{(v)}$ and of η_j respectively. Since

$$\varphi_1^{(v)}(t) \dots \varphi_m^{(v)}(t) \Rightarrow \varphi_1(t) \dots \varphi_m(t), \quad n \rightarrow \infty, \quad t \in R_1$$

and $\varphi_1(t) \dots \varphi_m(t)$ is the characteristic function of the sum $\eta_1 + \dots + \eta_m$ of the independent random variables $\eta_1, \dots, \eta_m$, the validity of our statement follows from theorem 1.3.

This theorem implies

Theorem 1.6. *If $\eta_j \in \mathcal{E}_2$ ($j = 1, 2, \dots$) i.e.*

$$\eta_j^{(v)} \Rightarrow \eta_j, \quad v \rightarrow \infty, \quad \eta_j^{(v)} \in \mathcal{E}_1,$$

then the linear order statistics generated by the matrices (1) determined by the points of discontinuity of the random variables $\eta_j^{(v)}$ ($v=1, 2, \dots; j=1, 2, \dots$) are doubly asymptotically of distribution η if and only if

$$(9) \quad \eta_1 + \dots + \eta_m \Rightarrow \eta, \quad m \rightarrow \infty.$$

1.4. In what follows, as an application of theorem 1.4. we are going to consider simple linear order statistics in the special case when

$$(10) \quad A_v = \begin{pmatrix} \alpha_1 & 2\alpha_1 & \dots & v\alpha_1 \\ \alpha_2 & 2\alpha_2 & \dots & v\alpha_2 \\ \vdots & \vdots & \dots & \vdots \end{pmatrix} \quad (v = 1, 2, \dots)$$

where $\alpha_1, \alpha_2, \dots$ are different nonzero real numbers.

Accordingly, in this case the random variable $\xi_{m,n}$ satisfies

$$P(\xi_{m,n} = \alpha_1 k_1 + \dots + \alpha_m k_m) = \frac{1}{(n+1)\dots(n+m)} \quad (k_1, \dots, k_m) \in \prod_{n+m}^{(m)}$$

and on the basis of (4) and of (5)

$$(11) \quad \begin{aligned} E(\xi_{m,n}) &= \frac{v+1}{2} (\alpha_1 + \dots + \alpha_m), \\ D^2(\xi_{m,n}) &= \frac{v+1}{12} [v(\alpha_1^2 + \dots + \alpha_m^2) - (\alpha_1 + \dots + \alpha_m)^2]. \end{aligned}$$

Let us prove the following

Theorem 1.7.

$$(12) \quad P\left(\frac{\xi_{m,n} - E(\xi_{m,n})}{D(\xi_{m,n})} < x\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy$$

$$n \rightarrow \infty, \quad m \rightarrow \infty$$

if and only if the real numbers $\alpha_1, \alpha_2, \dots$ satisfy the condition

$$(13) \quad \frac{\alpha_1^4 + \dots + \alpha_m^4}{(\alpha_1^2 + \dots + \alpha_m^2)^2} \rightarrow 0, \quad m \rightarrow \infty.$$

With the help of definition 1.4., this theorem can also be formulated as follows:

Theorem 1.8. The standardized linear order statistics generated by the matrices (10) are doubly asymptotically normally distributed if and only if (13) is satisfied.

PROOF. If the characteristic function

$$\varphi_j^{(v)}(t) = \frac{1}{v} \sum_{k=1}^v e^{i\alpha_j kt} = \frac{e^{i\alpha_j t} - e^{i\alpha_j vt}}{e^{i\alpha_j t} - 1}$$

of the random variable $\eta_j^{(v)}$ we use the notations

$$\mu = E(\xi_{m,n}), \quad \sigma = D(\xi_{m,n}), \quad S(x) = \frac{\sin x}{x},$$

then a short calculation yields

$$\varphi_j^{(v)}\left(\frac{t}{\sigma}\right) = e^{i\alpha_j(v+1)\frac{t}{2\sigma}} \frac{S\left(v\alpha_j \frac{t}{2\sigma}\right)}{S\left(\alpha_j \frac{t}{2\sigma}\right)}.$$

Making use of

$$S(x) = 1 - \frac{x^2}{6} + o(x^4)$$

and replacing σ^2 by the value (11), we obtain

$$S\left(\alpha_j v \frac{t}{2\sigma}\right) \rightarrow 1 - \frac{\alpha_j^2}{\alpha_1^2 + \dots + \alpha_m^2} \frac{t^2}{2} + o\left(\frac{\alpha_j^4}{(\alpha_1^2 + \dots + \alpha_m^2)^2}\right), \quad n \rightarrow \infty$$

and

$$S\left(\frac{\alpha_j t}{2\sigma}\right) \rightarrow 1, \quad n \rightarrow \infty.$$

Denoting the characteristic function of the random variables $\xi_{m,n}$ again by $\varphi_{m,n}(t)$, we get

$$\varphi_{m,n}^*(t) = E\left(e^{i(\xi_{m,n}-\mu)\frac{t}{\sigma}}\right) = e^{-\frac{i\mu}{\sigma}t} \varphi_{m,n}\left(\frac{t}{\sigma}\right)$$

and consequently

$$\begin{aligned} \lim_{n \rightarrow \infty} \varphi_{m,n}^*(t) &= \lim_{n \rightarrow \infty} \prod_{j=1}^m \frac{S\left(v\alpha_j \frac{t}{2\sigma}\right)}{S\left(\alpha_j \frac{t}{2\sigma}\right)} = \\ &= \prod_{j=1}^m \left[1 - \frac{\alpha_j^2}{\alpha_1^2 + \dots + \alpha_m^2} \frac{t^2}{2} + o\left(\frac{\alpha_j^4}{(\alpha_1^2 + \dots + \alpha_m^2)^2}\right)\right]. \end{aligned}$$

In order to investigate the passage to the limit $m \rightarrow \infty$ we take logarithms on both sides:

$$\log\left(\lim_{n \rightarrow \infty} \varphi_{m,n}^*(t)\right) = -\frac{t^2}{2} + o\left(\frac{\alpha_1^4 + \dots + \alpha_m^4}{(\alpha_1^2 + \dots + \alpha_m^2)^2}\right),$$

whence

$$\varphi_{m,n}^*(t) \rightarrow e^{-\frac{t^2}{2}}, \quad n \rightarrow \infty, \quad m \rightarrow \infty$$

if and only if condition (13) is satisfied. Then, the simple linear order statistics generated by the matrices (10) is — by virtue of theorem 1.4. — in fact asymptotically normally distributed.

As it is known,

$$\frac{\alpha_1^4 + \dots + \alpha_m^4}{(\alpha_1^2 + \dots + \alpha_m^2)^2} \cong \frac{1}{m}$$

and here equality holds only if $\alpha_1^2 = \dots = \alpha_m^2$. Therefore and on the basis of theorem 1.7. we have the following

Corollary 1.1. If $\alpha_j^2 = \alpha^2$ ($j=1, 2, \dots$) then (12) holds, and it has the greatest rapidity of convergence among the simple linear order statistics defined by the matrices (10).

In case of the well-known Wilcoxon-test $\alpha_j = 1$ ($j=1, 2, \dots$); consequently, on the basis of corollary 1.1. — as it is well-known — the Wilcoxon distribution is doubly asymptotically normal and belongs to those simple linear order statistics defined by the matrices (10) for which the speed of convergence to the normal distribution is the greatest.

The case

$$\alpha_k = k \quad (k = 1, 2, \dots)$$

is also of some interest. Indeed, in this case we wish to find, for natural numbers a satisfying

$$1m + 2(m-1) + \dots + m1 \leq a \leq 1(n+1) + \dots + m(n+m)$$

the number of solutions of the Diophantine equation

$$1x_1 + \dots + mx_m = a, \quad (x_1, \dots, x_m) \in \prod_{m+n}^{(m)}.$$

Since in this case

$$\frac{\alpha_1^4 + \dots + \alpha_m^4}{(\alpha_1^2 + \dots + \alpha_m^2)^2} = \frac{6}{5} \frac{3m(m+1) - 1}{m(m+1)(2m+1)} \rightarrow 0, \quad m \rightarrow \infty,$$

according to the theorem 1.7. the number of solutions shows a doubly asymptotic normal distribution.

2. On distributions representable as limits in the weak sense of discrete uniform distributions

2.1. By theorem 1.5. the construction of asymptotic linear order statistics and the construction of random variables $\xi \in \mathcal{E}_2$ are equivalent problems. It is this problem that we are going to consider in the present chapter by giving a method for the representation of random variables belonging to the set $\mathcal{E}_2$.

Definition 2.1. A Hermite-symmetrical matrix-valued function of order p defined on the interval $[-\pi, \pi]$ is said to be $\mathcal{L}$ -integrable, if each of its elements is $\mathcal{L}$ -integrable. By the integral of this we understand the matrix formed from the integral of its elements.

Let the set of $\mathcal{L}$ -integrable matrix-valued functions of order p be denoted by $\mathcal{L}_p$.

With the help of the Fourier transform

$$\varphi(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{itx} f(x) dx$$

of the matrix-valued function $f(x) \in \mathcal{L}_p$ we form the Hermite-symmetrical Toeplitz-matrix

$$T_n(f) = \begin{pmatrix} \varphi(0) & \varphi(-1) & \dots & \varphi(-n) \\ \varphi(1) & \varphi(0) & \dots & \varphi(-n+1) \\ . & . & \dots & . \\ \varphi(n) & \varphi(n-1) & \dots & \varphi(0) \end{pmatrix} \quad (n = 0, 1, 2, \dots)$$

of order $(n+1)p$, and we denote the eigenvalues of this matrix by

$$(14) \quad \lambda_k^{(n)} \quad (k = 1, 2, \dots, (n+1)p).$$

Let the eigenvalues of $f(x) \in \mathcal{L}_p$ be

$$\lambda_1(x) \leq \lambda_2(x) \leq \dots \leq \lambda_p(x)$$

when these are also $\mathcal{L}$ -integrable. Let $[m, M]$ be the shortest intervall containing the range of the functions $\lambda_k(x)$ ($k=1, \dots, p$). The values $m = -\infty$ and $M = \infty$ are also permitted.

Generalizing a theorem of G. SZEGÖ, the author has show ([3], 172, Satz 1.) that

$$\lambda_k^{(n)} \in [m, M] \quad (k = 1, \dots, (n+1)p; n = 0, 1, \dots)$$

and if $F(\lambda)$ is a continuous function defined on the intervall $[m, M]$, then

$$(15) \quad \frac{1}{n+1} \sum_{k=1}^{(n+1)p} F(\lambda_k^{(n)}) \rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} [F(\lambda_1(x)) + \dots + F(\lambda_p(x))] dx, \quad n \rightarrow \infty.$$

Definition 2.2. The random variables

$$P(\eta_{np} = \lambda_k^{(n)}) = \frac{1}{(n+1)p} \quad (k = 1, \dots, (n+1)p) \quad (n = 0, 1, 2, \dots)$$

defined with the help of the eigenvalues (14) are said to be Szegő distributions of the matrix-valued function $f(x) \in \mathcal{L}_p$.

Let η denote the random variable uniformly distributed in the interval $[-\pi, \pi]$, and let $A(\eta)$ be the mixture of the random variables $\lambda_1(\eta), \dots, \lambda_p(\eta)$ with weights $\frac{1}{p}$:

$$A(\eta) = \frac{1}{p} [\lambda_1(\eta) + \dots + \lambda_p(\eta)].$$

Theorem 2.1. *If the Szegő distributions of $f(x) \in \mathcal{L}_p$ are*

$$\eta_{np} \quad (n = 0, 1, 2, \dots),$$

then

- (a) $\eta_{np} \Rightarrow A(\eta), n \rightarrow \infty$;
- (b) $E(\eta_{np}) = E(A(\eta))$;
- (c) $D^2(\eta_{np}) \uparrow D^2(A(\eta)), n \rightarrow \infty$.

This theorem implies $A(\eta) \in \mathcal{E}_2$.

PROOF. If we write the expression (15) into the form

$$(16) \quad E[F(\eta_{np})] \rightarrow E[F(A(\eta))], \quad n \rightarrow \infty$$

then applying first the substitution $F(\lambda) = \cos \lambda t$ and then the substitution $F(x) = \sin \lambda t$ and adding the two obtained formulae, we get

$$(17) \quad \varphi_{\eta_{np}}(t) \rightarrow \frac{1}{p} \sum_{j=1}^p \varphi_{\lambda_j(\eta)}(t),$$

where $\varphi_{\cdot}(t)$ is the characteristic function of the random variable occurring in the index. Formula (17) however is a tantamount to our statement (a).

On the basis of formula (16)

$$E(\eta_{np}^k) = \frac{1}{(n+1)p} \operatorname{tr} T_n^k(f) \rightarrow E(A^k(\eta)), \quad n \rightarrow \infty.$$

Hence

$$E(\eta_{np}) = \frac{1}{(n+1)p} \operatorname{tr} T_n(f) = E(A(\eta))$$

and this is our statement (b).

Starting again with formula (16), we get

$$\begin{aligned} E(\eta_{np}^2) &= \frac{1}{(n+1)p} \operatorname{tr} T_n^2(f) = \frac{1}{p} \operatorname{tr} \varphi^2(0) + \\ &+ \frac{2}{(n+1)p} \sum_{k=1}^n (n-k+1) \operatorname{tr} \varphi(k) \varphi^*(k); \end{aligned}$$

thus

$$\begin{aligned} E(\eta_{np}^2) - E(\eta_{(n-1)p}^2) &= \frac{2}{n(n+1)p} \sum_{k=1}^{n-1} k \operatorname{tr} \varphi(k) \varphi^*(k) + \\ &+ \frac{2}{(n+1)p} \sum_{k=1}^n \operatorname{tr} \varphi(k) \varphi^*(k) \geq 0, \end{aligned}$$

i.e.

$$E(\eta_{np}^2) \uparrow E(A^2(\eta)), \quad n \rightarrow \infty,$$

thus statement (c) is also valid.

In what follows we shall need theorem 2.1. in the special case $p=1$. Accordingly we now formulate this special statement as the separate

Theorem 2.2. *If the Szegő distributions of the function $f(x) \in \mathcal{L}_1$ are*

$$\eta_n \quad (n = 0, 1, 2, \dots),$$

then

- (a) $\eta_n \Rightarrow f(\eta), \quad n \rightarrow \infty;$
- (b) $E(\eta_n) = E(f(\eta));$
- (c) $D^2(\eta_n) \uparrow D^2(f(\eta)), \quad n \rightarrow \infty.$

The following theorem also follows from formula (16).

Theorem 2.3. *If $f(x) \in \mathcal{L}_p$ is positive semidefinite and, outside a set of measure zero, is positive definite then*

$$E(\log \eta_{np}) = \log [\text{Det } T_n(f)]^{\frac{1}{(n+1)p}} \rightarrow E\left(\log [\text{Det } f(\eta)]^{\frac{1}{p}}\right), \quad n \rightarrow \infty,$$

and $\log (\text{Det } f(\eta))^{\frac{1}{p}} \in \mathcal{E}_2$. In particular

$$\varphi_{\log \eta_{np}}(t) \rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} [\text{Det } f(x)]^{\frac{it}{p}} dx, \quad n \rightarrow \infty, \quad t \in R_1.$$

Let $\mathcal{C}$ denote the set of those continuous distribution functions $F(x)$ which are strictly monotone increasing in some interval $[a, b]$, and satisfy $F(a)=0$, $F(b)=1$, where the values $a=-\infty$ and $b=\infty$ are also permitted. The inverse of $y=F(x)$ will be denoted by $F^{-1}(y)$.

Theorem 2.4. *If the expectation of the random variable ξ exists, and if its distribution function belongs to the set $\mathcal{C}$, then the Szegő distributions of the function $F^{-1}\left(\frac{x+\pi}{2\pi}\right) \in \mathcal{L}_1$ weakly converge to the random variable ξ , i.e. $\xi \in \mathcal{E}_2$.*

PROOF. Since

$$\int_0^1 F^{-1}(y) dy = \int_{-\infty}^{\infty} x dF(x),$$

$F^{-1}(y)$ is $\mathcal{L}$ -integrable on the interval $[0, 1]$ and $F^{-1}\left(\frac{x+\pi}{2\pi}\right)$ on $[-\pi, \pi]$.

On the other hand, $F(\xi)$ being uniformly distributed in the interval $[0, 1]$, $\eta = 2\pi F(\xi) - \pi$ is uniformly distributed on the interval $[-\pi, \pi]$. Now, by theorem 2.2., the Szegő distributions of the function $F^{-1}\left(\frac{x+\pi}{2\pi}\right)$ weakly converge to the random variable $F^{-1}\left(\frac{\eta+\pi}{2\pi}\right) = \xi$.

In this case the Fourier transform generating the Toeplitz matrices is

$$\varphi(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F^{-1}\left(\frac{x+\pi}{2\pi}\right) e^{itx} dx = e^{-int} \int_a^b x e^{2\pi it F(x)} dF(x), \quad t \in R_1.$$

Theorem 2.4. implies the following

Corollary 2.1. *If the distribution function of the bounded random variable ξ belongs to the set $\mathcal{C}$, then $\xi \in \mathcal{E}_2$.*

Corollary 2.2. *If for the random variable ξ with an absolutely continuous distribution function there exist expectation and the density function is positive on an interval $[a, b]$ (with, possibly, $a=-\infty$, $b=\infty$) and vanish outside it, then $\xi \in \mathcal{E}_2$.*

This yields the following

Corollary 2.3. Random variables of the normal, the χ^2 , the Student ($n > 1$), and the exponential distribution resp. belong to the set $\mathcal{E}_2$.

2.2. If, instead of Toeplitz matrices, we start with the Hankel matrices of a function, we are able to give a new method for the construction of elements of the set $\mathcal{E}_2$.

Let $f(x)$ be a function, $\mathcal{L}$ -integrable on the interval $[-1, 1]$. Let $P_n(x)$ be the n -th Legendre polynomial, and with its help us form the polynomials

$$p_n(x) = \left(n + \frac{1}{2}\right)^{\frac{1}{2}} P_n(x) \quad (n = 0, 1, 2, \dots).$$

The symmetrical matrix of order $n+1$

$$H_n(f) = \left(\int_{-1}^1 p_\alpha(x) p_\beta(x) f(x) dx \right) \quad (\alpha, \beta = 0, 1, \dots, n; n = 0, 1, 2, \dots)$$

is said to be the n -th Hankel matrix of the function $f(x)$. Let the eigenvalues of $H_n(f)$ be

$$\lambda_k^{(n)} \quad (k = 1, \dots, n+1).$$

Definition 2.3. The random variables

$$P(\eta_n = \lambda_k^{(n)}) = \frac{1}{n+1} \quad (k = 1, \dots, n+1)$$

defined with the help of the eigenvalues of $H_n(f)$ are said to be the Szegő distributions of the function $f(x)$.

The following result is due to G. Szegő ([2], 88—89):

The eigenvalues of $H_n(f)$ fall into the narrowest interval determined by the range of $f(x)$. If, moreover, $F(\lambda)$ is a continuous function defined on this interval then, in a probabilistic terminology,

$$E(F(\eta_n)) \rightarrow \frac{1}{\pi} \int_0^\pi F(f(\cos \lambda)) d\lambda, \quad n \rightarrow \infty.$$

Starting with this theorem we can prove by a method similar to that used in establishing the first statement of Theorem 2.1. that the following holds:

Theorem 2.5. If η is a uniformly distributed random variable defined on the interval $[0, \pi]$, then

$$\eta_n \Rightarrow f(\cos \eta), \quad n \rightarrow \infty,$$

i.e. $f(\cos \eta) \in \mathcal{E}_2$.

From the result of G. Szegő just quoted one also infers that

$$E(\eta_n^k) = \frac{1}{n+1} \operatorname{tr} H_n^k(f) \rightarrow E(f^k(\cos \eta)), \quad n \rightarrow \infty \quad (k = 1, 2, \dots),$$

in particular

$$E(\eta_n^2) \uparrow E(f^2(\cos \eta)) \quad n \rightarrow \infty.$$

3. On linear order statistics with a given asymptotics

3.1. In what follows, we shall make use of the results of 2.1.

Definition 3.1. We speak about linear order statistics generated by the sequence of functions

$$(18) \quad f_k(x) \in \mathcal{L}_1 \quad (k = 1, 2, \dots)$$

if the k -th rows of the generating matrices (1) contain the eigenvalues of the $(v-1)$ -th Toeplitz-matrix of the function $f_k(x)$.

On the basis of Theorems 1.5. and 2.2. we have the following

Theorem 3.1. The linear order statistics generated by the sequence (18) have asymptotically distribution $f_1(\eta_1) + \dots + f_m(\eta_m)$ ($m=1, 2, \dots$), where $\eta_1, \dots, \eta_m$ are random variables independent from each other, and uniformly distributed in the interval $[-\pi, +\pi]$.

It is clear that with respect to Hankel matrices we can formulate a definition and a theorem similar to Definition 3.1. and to Theorem 3.1. respectively.

On the basis of Corollary 2.3. Theorem 3.1. enables us to construct linear order statistics, having e.g. asymptotically normal, χ^2 , Student ($n>1$) or exponential distribution.

On the basis of Theorem 1.6. we moreover have

Theorem 3.2. If

$$f_1(\eta_1) + \dots + f_m(\eta_m) \Rightarrow \xi, \quad m \rightarrow \infty,$$

then the linear order statistics generated by (18) are doubly asymptotically ξ distributed.

We now prove the following

Theorem 3.3. Let $f(x)$ be a function square-integrable on the interval $[-\pi, \pi]$, and let α_k ($k=1, 2, \dots$) be nonzero real numbers. The standardized statistics belonging to the linear order statistics generated by the sequence

$$\alpha_k f(x) \quad (k = 1, 2, \dots)$$

are doubly asymptotically normally distributed if and only if

$$\frac{|\alpha_1|^3 + \dots + |\alpha_m|^3}{(\alpha_1^2 + \dots + \alpha_m^2)^{3/2}} \rightarrow 0, \quad m \rightarrow \infty.$$

PROOF. If we use the notations of chapter 1. and if the random variable $\alpha_k f(\eta)$ has characteristic function $\varphi_k(t)$, then

$$\varphi_k(t) = \varphi(\alpha_k t),$$

where

$$\varphi(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{itf(x)} dx.$$

By Theorem 3.1.

$$\varphi_{m,n}(t) \rightarrow \varphi_{m^*}(t) = \varphi_1(t) \dots \varphi_m(t), \quad n \rightarrow \infty,$$

where $\varphi_{m^*}(t)$ is the characteristic function of the random variable

$$\xi_{m^*} = \alpha_1 f(\eta_1) + \dots + \alpha_m f(\eta_m).$$

Here $\eta_1, \dots, \eta_m$ are independent random variables identically distributed with η . We now put

$$E(\xi_{m^*}) = (\alpha_1 + \dots + \alpha_m) E(f(\eta)) = a,$$

$$D^2(\xi_{m^*}) = (\alpha_1^2 + \dots + \alpha_m^2) D^2(\eta) = \sigma^2.$$

By what has been said previously

$$\varphi_{m^*}^*(t) = e^{-\frac{at}{\sigma}} \varphi_{m^*}\left(\frac{t}{\sigma}\right) = e^{-\frac{at}{\sigma}} \varphi\left(\alpha_1 \frac{t}{\sigma}\right) \dots \varphi\left(\alpha_m \frac{t}{\sigma}\right).$$

$f(x)$ being square integrable, one has

$$\varphi(t) = 1 + iE(f(\eta))t - E(f^2(\eta))\frac{t^2}{2} + o(|t|^3)$$

and, consequently,

$$\log \varphi\left(\frac{\alpha}{\sigma} t\right) = i\frac{\alpha}{\sigma} E(f(\eta))t - \frac{\alpha^2}{\sigma^2} D^2(f(\eta))\frac{t^2}{2} + o\left(\left|\frac{\alpha}{\sigma} t\right|^3\right).$$

Using this we get

$$\log \varphi_{m^*}^*(t) = -\frac{iat}{\sigma} + \sum_{k=1}^m \log \varphi\left(\alpha_k \frac{t}{\sigma}\right) = -\frac{t^2}{2} + o\left(\frac{|\alpha_1|^3 + \dots + |\alpha_m|^3}{\sigma^3} |t|^3\right);$$

hence our statement follows.

One easily sees that

$$\frac{|\alpha_1|^3 + \dots + |\alpha_m|^3}{(\alpha_1^2 + \dots + \alpha_m^2)^{3/2}} \cong \sqrt{\frac{1}{m}},$$

with equality only for $|\alpha_1| = \dots = |\alpha_m|$. Thus it is in this case that the speed of convergence to the normal distribution is the greatest. Comparing this with the speed of convergence to the normal distribution obtained for the Wilcoxon distribution in connection with Theorem 1.7., we see that the convergence of the latter is stronger by a factor $\sqrt{m}$ than the convergence just obtained.

Both Theorem 1.7. and Theorem 3.3. give necessary and sufficient conditions in order that linear order statistics have doubly asymptotically normal distribution. None of these two theorems can be reduced to any other. Indeed, there does not exist a function $f(x) \in \mathcal{L}_1$, the n -th Toeplitz-matrix of which would have eigenvalues $\alpha, 2\alpha, \dots, (n+1)\alpha$ where α is a nonzero real number because this would yield

$$\frac{1}{n+1} \operatorname{tr} T_n(f) = \varphi(0) = \frac{\alpha}{2} (n+2)$$

for any nonnegative integer n , what is impossible.

3.2. In what follows we are going to use the eigenvalues of the Toeplitz matrices generated by the functions

$$f(x) = a - 2b \cos x, \quad x \in [-\omega, \pi]$$

(a real, b positive) for constructing linear order statistics.

The eigenvalues of the matrix $T_n(f)$ ([2], 67) are

$$\lambda_k^{(n)} = a + b\alpha_k^{(n)} \quad (k = 1, \dots, n+1)$$

with

$$\alpha_k^{(n)} = -2 \cos \frac{k\pi}{n+2}.$$

If ζ_n denotes the n -th Szegő distribution of $f(x)$, then by Theorem 2.2.

$$\zeta_n \Rightarrow a - 2b \cos \eta, \quad n \rightarrow \infty,$$

moreover

$$E(\zeta_n) = E(f(\eta)) = a$$

and

$$D^2(\zeta_n) = \frac{2n}{n+1} b^2 \uparrow 2b^2 = D^2(f(\eta)), \quad n \rightarrow \infty.$$

Now, if a_k and b_k ($k=1, 2, \dots$) are fixed real and positive numbers, respectively, then the linear order statistics generated by the matrices

$$(19) \quad \begin{pmatrix} a_1 + b_1 \alpha_1^{(n)} & a_1 + b_1 \alpha_2^{(n)} & \dots & a_1 + b_1 \alpha_{n+1}^{(n)} \\ a_2 + b_2 \alpha_1^{(n)} & a_2 + b_2 \alpha_2^{(n)} & \dots & a_2 + b_2 \alpha_{n+1}^{(n)} \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix} = A_n \quad (n = 0, 1, 2, \dots)$$

are on the basis of Theorem 3.1. asymptotically

$$(20) \quad (a_1 + \dots + a_m) + b_1 \cos \eta_1 + \dots + b_m \cos \eta_m$$

distributed, where $\eta_1, \dots, \eta_m$ are independent random variables having the same distribution as η . Moreover, the linear order statistics are doubly asymptotic if the random variables (20) are weakly convergent for $m \rightarrow \infty$. If in Theorem 3.3. we put $\alpha_k = \beta_k$, we get a condition, necessary and sufficient in order that the standardized statistics generated by the matrices (19) be asymptotically normally distributed.

If we put $a=2b=\frac{1}{4}$, we obtain a special case of some interest. Indeed, in this case

$$f(x) = \cos^2 \frac{x}{2}, \quad x \in [-\pi, \pi];$$

$$\lambda_k^{(n)} = \cos^2 \frac{k\pi}{2(n+2)} \quad (k = 1, \dots, n+1).$$

By what has been said before, the linear order statistics generated by the matrices

$$A_n = \begin{pmatrix} \cos^2 \frac{\pi}{2(n+2)} & \cos^2 \frac{2\pi}{2(n+2)} & \dots & \cos^2 \frac{(n+1)\pi}{2(n+2)} \\ \cos^2 \frac{\pi}{2(n+2)} & \cos^2 \frac{2\pi}{2(n+2)} & \dots & \cos^2 \frac{(n+1)\pi}{2(n+2)} \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix} \quad (n = 0, 1, 2, \dots)$$

are asymptotically

$$\cos^2 \frac{\eta_1}{2} + \dots + \cos^2 \frac{\eta_m}{2} \quad (m = 1, 2, \dots)$$

distributed and the corresponding standardized statistics are doubly asymptotically normally distributed.

References

- [1] B. V. GNEDENKO—A. N. KOLMOGOROV, Limit distributions of sums of independent random variables. (*Russian.*) Moscow, 1949.
- [2] U. GRENANDER—G. SZEGÖ, Toeplitz forms and its applications. *Berkeley and Los Angeles*, 1958.
- [3] B. GYIRES, Eigenwerte verallgemeinerter Toeplitzscher Matrizen. *Publ. Math. Debrecen* 4 (1956), 171—179.
- [4] J. HAJEK—Z. SIDAK, Theory of rank tests. *Prague*, 1967.
- [5] M. MARCUS—M. MINC, Permanents. *Amer. Math. Monthly* 72 (1965), 577—591.
- [6] L. SCHMETTERER, Einführung in die mathematische Statistik. *Wien*, 1956.

(Received January 27, 1973.)