ON THE THEORY OF ORDER STATISTICS

By

ALFRÉD RÉNYI (Budapest), corresponding member of the Academy

Dedicated to A. N. KOLMOGOROV
on the occasion of his 50th birthday

Introduction

Since the beginning of the century many authors, e. g. K. PEARSON [1], L. v. BORTKIEWICZ [2], E. L. DODD [3], L. H. C. TIPPET [4], and M. FRÉCHET [5] have dealt with particular problems which may be classified as belonging to the theory of order statistics. A. N. KOLMOGOROV [6], V. I. GLIVENKO [7], N. V. SMIRNOV [8], B. V. GNEDENKO [9], and other mathematicians having recognized the great theoretical and practical importance of this set of problems, developed this subject into a systematical theory.

In the last three years a particularly great number of papers dealt with such problems; of these we mention those of B. V. GNEDENKO and V. S. KOROLUK [10], B. V. GNEDENKO and E. L. RVAČEVA [11], B. V. GNEDENKO and V. S. MIHALEVIČ [12], V. S. MIHALEVIČ [13], J. D. KVIT [14], G. M. MANIA [15], I. I. GIHMAN [16], W. FELLER [17], J. L. DOOB [18], F. J. MASSEY [19], M. D. DONSKER [20], T. W. ANDERSON and D. A. DARLING [21]. A bibliography up to 1947 is to be found in the paper of S. S. WILKS [30] enumerating 90 papers.

The purpose of the present paper is to give a new method by means of which many important results of the theory of order statistics can be obtained with surprising simplicity; the method also enables us to prove several new theorems. The essential novelty of this method is that it reduces the problems connected with order statistics to the study of sums of mutually independent random variables. § 1 contains the review of the method, § 2 is devoted to the proof of some known theorems by means of this method, and § 3 contains the formulation of some new results obtained by this method, concerning the comparison of the sample distribution function to that of the population. These results are connected with the fundamental results of A. N. KOLMOGOROV and N. V. SMIRNOV.

Let $F_n(x)$ denote the distribution function of a sample of size n drawn from a population having the continuous distribution function $F(x)$, in other words, $F_n(x)$ denotes the frequency ratio of sample values not exceeding x. KOLMOGOROV determined the limiting distribution of the supremum of $|F_n(x)-F(x)|$, SMIRNOV did the same for $F_n(x)-F(x)$; in § 4 we shall determine the limiting distribution of the supremum* of the relative deviations $\frac{F_n(x)-F(x)}{F(x)}$, and $\left|\frac{F_n(x)-F(x)}{F(x)}\right|$, respectively. To do this, besides our method, the lemmas in § 4, generalizing some results of P. ERDŐS and M. KAC [22], are needed. These lemmas are of certain interest in themselves. § 5 contains the proof of the results formulated in § 3, while § 6 contains some remarks on the numerical computation of the values of the limiting distribution functions occurring in our theorems; tables for one of these are also given. I have found the method formulated in § 1 by analysing a theorem of S. MALMQUIST [23]; § 1 contains among others a new and simple proof for this theorem of MALMQUIST. Together with G. HAJÓS, we have found another simple proof of this theorem which will be published in a joint paper of ours [24]. All these investigations have for their origin in the discussions in a seminary of the Departments of Probability Theory and of Mathematical Statistics of the Institute for Applied Mathematics of the Hungarian Academy of Sciences. I lectured on the part of the results contained in this paper in January 1953 on the congress of the Humboldt University in Berlin[1] and in September 1953 on the VIIIth Polish Mathematical Congress in Warsaw. On this last occasion A. N. KOLMOGOROV has made certain valuable remarks for which I express him my most sincere thanks. Further, I express my thanks to T. LIPTÁK who participated in the preparation of this paper by elaboration of some particular calculations, as well as to Miss I. PALÁSTI and Mrs. P. VÁRNAI for the numerical computations.

§ 1. A new method in the theory of order statistics

Let us start with the following special case: let a sample of size n be given concerning the value of a random variable ζ of exponential distribution, i. e. the results of n independent observations for its value, denoted by $\zeta_1, \zeta_2, \ldots, \zeta_n$; in other words, $\zeta_1, \zeta_2, \ldots, \zeta_n$ are mutually independent random variables with the same distribution function of exponential type. We need the following well-known property of the exponential distribution: if ζ is an exponentially distributed random variable, then

$$\mathbf{P}(\zeta < x+y \mid \zeta \geqq y) = \mathbf{P}(\zeta < x), \tag{1.1}$$

* (in an interval $0 < a \leqq F(x) \leqq b \leqq 1$).

[1] This lecture will be published in the communications of the Congress under the following title: "Eine neue Methode in der Theorie der geordneten Stichproben".

if $x>0$ and $y \geqq 0$.[2] This property characterizes uniquely the exponential distribution. Indeed, let $F(x)$ be the distribution function of ζ, then

$$\mathbf{P}(\zeta < x+y \mid \zeta \geqq y) = \frac{F(x+y)-F(y)}{1-F(y)}$$

and it follows that (1. 1) is equivalent to the following relation:

$$\Phi(x+y) = \Phi(x)\Phi(y) \tag{1.2}$$

where $\Phi(x) = 1-F(x)$. It is known however that, of all functions satisfying the condition $0 \leqq \Phi(x) \leqq 1$, except for the trivial cases $\Phi(x) \equiv 0$ and $\Phi(x) \equiv 1$, the functions $\Phi(x) = e^{-\lambda x}$ $(\lambda > 0)$ and only these satisfy the functional equation (1. 2).

The meaning of (1. 1) becomes especially clear, if the random variable ζ is interpreted as the duration of a happening having random duration. In this interpretation the proposition (1. 1) can be formulated as follows: in case of a happening of exponentially distributed random duration, being in progress at the moment y, the further duration of the happening does not depend on y, i. e. on its duration until the given moment.

Let us arrange the numbers $\zeta_1, \zeta_2, \ldots, \zeta_n$ in order of magnitude and use the notation

$$\zeta_k^* = R_k(\zeta_1, \zeta_2, \ldots, \zeta_n) \qquad (k = 1, 2, \ldots, n) \tag{1.3}$$

where the function $R_k(x_1, x_2, \ldots, x_n)$ of the n variables $x_1, x_2, \ldots, x_n$ denotes the k-th of the values $x_1, x_2, \ldots, x_n$ in order of magnitude $(k = 1, 2, \ldots, n)$; thus e. g. $\zeta_1^* = \min\limits_{1 \leqq k \leqq n} \zeta_k$ and $\zeta_n^* = \max\limits_{1 \leqq k \leqq n} \zeta_k$. Then the individual and joint distributions of the values of the *order statistics* $\zeta_1^* \leqq \zeta_2^* \leqq \cdots \leqq \zeta_n^*$ can be most easily determined. For that purpose we interpret the variables ζ_k as random durations of mutually independent happenings; then ζ_k^* denotes the duration of the happening finished as k-th of the n happenings. Let us determine, first of all, the distributions of the differences $\zeta_{k+1}^* - \zeta_k^*$. If $\zeta_k^* = y$, then

$$\mathbf{P}(\zeta_{k+1}^* - \zeta_k^* > x \mid \zeta_k^* = y) = \mathbf{P}(\zeta_{k+1}^* > x+y \mid \zeta_k^* = y) \tag{1.4}$$

where on the right side there stands the probability of the event that none of the $n-k$ happenings, being in progress at the moment y, finishes until the moment $x+y$. By virtue of (1. 1), the value of this probability is

$$(\mathbf{P}(\zeta > x))^{n-k} = e^{-(n-k)\lambda x}$$

and thus the conditional distribution function of $\zeta_{k+1}^* - \zeta_k^*$ with respect to the condition $\zeta_k^* = y$ is

$$\mathbf{P}(\zeta_{k+1}^* - \zeta_k^* < x \mid \zeta_k^* = y) = 1 - e^{-(n-k)\lambda x}. \tag{1.5}$$

[2] $\mathbf{P}(A)$ denotes the probability of the event A, and $\mathbf{P}(A \mid B)$ denotes the conditional probability of the event A with respect to the event B.

As the conditional distribution function (1.5) does not depend on y, (1.5) gives also the non-conditional distribution function of $\xi_{k+1}^*-\zeta_k^*$; indeed, by virtue of the theorem on total probability,

$$(1.6)\qquad \mathbf{P}(\zeta_{k+1}^*-\zeta_k^*<x)=\int_0^\infty \mathbf{P}(\zeta_{k+1}^*-\zeta_k^*<x\,|\,\zeta_k^*=y)\,d\mathbf{P}(\zeta_k^*<y)=1-e^{-(n-k)\lambda x}.$$

Therefore the differences $\zeta_{k+1}^*-\zeta_k^*$ are themselves exponentially distributed with the mean value $\frac{1}{(n-k)\lambda}$ and thus the variables

$$(1.7)\qquad \delta_{k+1}=(n-k)\;(\zeta_{k+1}^*-\zeta_k^*)\qquad (k=0,1,\ldots,n-1)$$

are also exponentially distributed with the mean value $\frac{1}{\lambda}$. (In the above relation by definition $\zeta_0^*\equiv 0$.)

It follows from the abovesaid that the variables $\delta_1,\delta_2,\ldots,\delta_n$ are mutually independent random variables. It is namely easy to see that the probability

$$(1.8)\qquad \mathbf{P}(\zeta_{k+1}^*-\zeta_k^*<x\,|\,\zeta_1^*=y_1,\,\zeta_2^*-\zeta_1^*=y_2,\ldots,\zeta_k^*-\zeta_{k-1}^*=y_k)$$

does not depend on the variables $y_1,y_2,\ldots,y_k$; this is evident, as the above conditions mean that $\zeta_1^*=y_1,\zeta_2^*=y_1+y_2,\ldots,\zeta_k^*=y_1+y_2+\cdots+y_k$; i. e. they give the moments of the finishing of k happenings, which finish first of the n happenings which started simultaneously at the moment $t=0$. These conditions imply that at the moment $t=y_1+y_2+\cdots+y_k$ there are still $n-k$ happenings in progress and the probability of the finishing of at least one of them before the moment $t+x$ is equal to $1-e^{-(n-k)\lambda x}$. Thus the probability of the left hand side of (1.8) equals $1-e^{-(n-k)\lambda x}$, i. e. it does not depend on the variables $y_1,y_2,\ldots,y_k$, and this is equivalent to the fact that the variables $\zeta_{k+1}^*-\zeta_k^*$ (and also the variables δ_k) are mutually independent. Thus the variables ζ_k^* can be expressed in the form

$$(1.9)\qquad \zeta_k^*=\frac{\delta_1}{n}+\frac{\delta_2}{n-1}+\cdots+\frac{\delta_k}{n-k+1}\qquad (k=1,2,\ldots,n)$$

i. e. as linear forms of mutually independent random variables having the same distribution. (1.9) can be also expressed by saying that *the variables ζ_k^* form an (additive) Markov chain*. By virtue of (1.9) the distribution of any ζ_k^*, further, the joint distribution of any number of the variables ζ_k^* can be determined in explicit form.

Consider now, how the abovesaid can be applied in general to the study of order statistics. Let ξ be any random variable having a continuous and steadily increasing[3] distribution function $F(x)$, let $(\xi_1,\xi_2,\ldots,\xi_n)$ be a

[3] By saying that $F(x)$ is steadily increasing, we mean that $F(x)$ is a strictly increasing monotone function in the least interval (a,b) where $F(a)=0$ and $F(b)=1$; it may be happen that $a=-\infty$ or $b=+\infty$.

sample of size n consisting of n independent observations of the value of ξ, that is to say, let $\xi_1, \xi_2, \ldots, \xi_n$ be mutually independent random variables with the same (continuous) distribution function $F(x)$. Let us arrange the sample values ξ_k in order of magnitude, that is to say, let us form the new variables $\xi_k^* = R_k(\xi_1, \xi_2, \ldots, \xi_n)$.

The main problem of order statistics is to study the variables ξ_k^*; this, however, can be reduced to the special case when the variables ξ_k are exponentially distributed (and therefore — by virtue of (1.9) — to the study of sums of mutually independent random variables), as follows: let us put

$$\eta_k = F(\xi_k) \quad \text{and} \quad \zeta_k = \log\frac{1}{\eta_k} \qquad (k = 1, 2, \ldots, n) \tag{1.10}$$

and let us denote by $\eta_k^* = F(\xi_k^*)$ the k-th of the variables $\eta_1, \eta_2, \ldots, \eta_n$ in order of magnitude, i. e. let us put $\eta_k^* = R_k(\eta_1, \eta_2, \ldots, \eta_n)$; further, let us put

$$\zeta_k^* = \log\frac{1}{\eta_{n+1-k}^*} \qquad (k = 1, 2, \ldots, n). \tag{1.11}$$

As $\log\dfrac{1}{x}$ is a steadily decreasing function, we obtain:

$$\zeta_k^* = R_k(\zeta_1, \zeta_2, \ldots, \zeta_n) \qquad (k = 1, 2, \ldots, n), \tag{1.12}$$

whence ζ_k^* is the k-th of the variables $\zeta_1, \zeta_2, \ldots, \zeta_n$ in order of magnitude. As we have assumed the variables ξ_k to be mutually independent, it follows that the variables ζ_k are also mutually independent.

Let us investigate now the distribution of the single variable ζ_k. $F(x)$ being a strictly increasing function, the inverse function of $x = F(y)$, denoted by $y = F^{-1}(x)$, is uniquely defined in the interval $0 \leqq x \leqq 1$, and thus

$$\mathbf{P}(\zeta_k < x) = \mathbf{P}\left(\log\frac{1}{F(\zeta_k)} < x\right) = \mathbf{P}(\xi_k > F^{-1}(e^{-x})) = 1 - F(F^{-1}(e^{-x})) = 1 - e^{-x},$$

if $0 \leqq x \leqq 1$. Therefore the variables $\zeta_1, \zeta_2, \ldots, \zeta_n$ are mutually independent and of exponential distribution with the mean value 1. In this way, the random variables ξ_k^* themselves can be expressed in the form

$$\xi_k^* = F^{-1}(e^{-\zeta_{n+1-k}^*}) = F^{-1}\left(e^{-\left(\frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \cdots + \frac{\delta_{n+1-k}}{k}\right)}\right) \qquad (k = 1, 2, \ldots, n); \tag{1.13}$$

where the variables $\delta_1, \delta_2, \ldots, \delta_n$ are mutually independent and of exponential distribution with the same distribution function $1 - e^{-x}$ $(x > 0)$. It also follows from this result that the quotients

$$\frac{\eta_{k+1}^*}{\eta_k^*} = e^{-\frac{\delta_{n+1-k}}{k}} \tag{1.14}$$

are mutually independent random variables (here, by definition, $\eta_{n+1}^* \equiv 1$), since the variables δ_{n+1-k} are, as we have seen, mutually independent. Another consequence of (1.13) is that *the variables* $\xi_n^*, \xi_{n-1}^*, \ldots, \xi_1^*$ *form a*

Markov chain (and thus the variables $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$ form also a Markov chain); this follows from

$$\xi_{n-k}^* = F^{-1}\left(e^{\log F(\xi_{n+1-k}^*) - \frac{\delta_{k+1}}{n-k}}\right) \tag{1.15}$$

which is obtained from (1.13) and from the fact that the variables ξ_{n+1-k}^* and δ_{k+1} are independent, since

$$\xi_{n+1-k}^* = F^{-1}\left(\exp\left[-\left(\frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \cdots + \frac{\delta_k}{n-k+1}\right)\right]\right).$$

A. N. KOLMOGOROV [6a] was the first who remarked that the variables $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$, i. e. the sequence of order statistics, form a Markov chain. The new method contained in the present paper starts from this fundamental observation, but the possibilities implied by it could be developed only after having transformed the Markov chain $\{\xi_k^*\}$ into an additive Markov chain by means of the transformation $\xi_k^* = \log \dfrac{1}{F(\xi_{n-k+1}^*)}$. In this connection it is interesting to consider the following general problem: for which Markov processes $\{\xi_t\}$ can such a family of functions $G_t(x)$ be found that the variables $\eta_t = G_t(\xi_t)$ form an additive Markov process? A necessary condition of this is that the distribution function $F(x, s, y, t) = \mathbf{P}(\xi_t < x \mid \xi_s = y)$ of the Markov process should satisfy the following differential equation:

$$\frac{\partial F}{\partial x}\frac{\partial F}{\partial y}\left(\frac{\partial F}{\partial y}\frac{\partial^3 F}{\partial x^2 \partial y} - \frac{\partial F}{\partial x}\frac{\partial^3 F}{\partial x \partial y^2}\right) = \frac{\partial^2 F}{\partial x \partial y}\left\{\frac{\partial^2 F}{\partial x^2}\left(\frac{\partial F}{\partial y}\right)^2 - \frac{\partial^2 F}{\partial y^2}\left(\frac{\partial F}{\partial x}\right)^2\right\}.$$

We hope to return to the discussion of this problem on another occasion.

The variables η_k are obviously uniformly distributed in the interval (0, 1), because, if $0 < x < 1$, then

$$\mathbf{P}(\eta_k < x) = \mathbf{P}(\xi_k < F^{-1}(x)) = F(F^{-1}(x)) = x, \tag{1.16}$$

and therefore the variables η_k^* form an ordered sample of size n drawn from a population of uniform distribution in the interval (0, 1).

It follows from (1. 14) that

$$\left(\frac{\eta_{k+1}^*}{\eta_k^*}\right)^k = e^{-\delta_{n+1-k}}$$

and as $\mathbf{P}(e^{-\delta_{n+1-k}} < x) = \mathbf{P}\left(\delta_{n+1-k} > \log \dfrac{1}{x}\right) = e^{-\log \frac{1}{x}} = x$, therefore the variables $\left(\dfrac{\eta_{k+1}^*}{\eta_k^*}\right)^k$ are mutually independent and have the same distribution, namely, they are uniformly distributed in the interval (0, 1). This is the theorem of S. MALMQUIST mentioned in the introduction.

§ 2. The theory of order statistics built up by means of the method of § 1

On the basis of what has been said above it is easy to obtain the results concerning the limiting distribution of order statistics. In order to show this, we shall prove the following theorems.[4]

THEOREM 1. *If $k \geqq 1$ is a fixed positive integer, then*

$$\lim_{n\to\infty} \mathbf{P}(n\zeta_k^* < x) = \int_0^x \frac{t^{k-1}e^{-t}}{(k-1)!}\,dt \qquad (x>0),$$

i. e. $n\zeta_k^*$ has in the limit a Γ-distribution of order k.

PROOF. As we have seen,

$$\zeta_k^* = \frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \cdots + \frac{\delta_k}{n+1-k}$$

where $\delta_1, \delta_2, \ldots, \delta_k$ are mutually independent and exponentially distributed random variables with the common distribution function $1-e^{-x}$ $(x>0)$. Therefore, the variable $\dfrac{\delta_j}{n+1-j}$ has the probability density function $(n+1-j)e^{-(n+1-j)x}$ $(x>0)$ and thus it follows by simple calculation that the probability density function of ζ_k^* is

$$g_k(t) = \binom{n}{k} k e^{-nt}(e^t-1)^{k-1};$$

hence $n\zeta_k^*$ has the frequency function

$$(2.1)\qquad \frac{1}{n}\,g_k\!\left(\frac{t}{n}\right) = \binom{n-1}{k-1} e^{-t}\left(e^{\frac{t}{n}}-1\right)^{k-1} = \frac{\left[n\left(e^{\frac{t}{n}}-1\right)\right]^{k-1} e^{-t} \prod_{j=1}^{k-1}\left(1-\frac{j}{n}\right)}{(k-1)!}.$$

As $\lim\limits_{n\to\infty} n\left(e^{\frac{t}{n}}-1\right) = t$, thus the density function of $n\zeta_k^*$ converges to $\dfrac{t^{k-1}e^{-t}}{(k-1)!}$ as $n\to\infty$, i. e. to the density function of the Γ-distribution of order k.

This result might have been expected by the following consideration. Obviously

$$(2.2)\qquad n\zeta_k^* = \delta_1+\delta_2+\cdots+\delta_k + \sum_{j=2}^{k}\frac{(j-1)\delta_j}{n+1-j};$$

the density function of $\delta_1+\delta_2+\cdots+\delta_k$ is, however, $\dfrac{t^{k-1}e^{-t}}{(k-1)!}$, on the other hand, the variable $\displaystyle\sum_{j=2}^{k}\frac{(j-1)\delta_j}{n+1-j}$ tends stochastically to 0, as $n\to\infty$.

[4] We use the notations introduced in § 1.

To develop this consideration into a precise proof, we need the following lemma, due to H. CRAMÉR ([25], p. 254).

LEMMA 1. *Let us put $\gamma_n = \alpha_n + \beta_n$ where α_n and β_n are random variables and let $F_n(x)$ denote the distribution function of α_n and further let us assume that*[5]

$$\lim_{n\to\infty} \mathbf{M}\beta_n = 0 \quad \text{and} \quad \lim_{n\to\infty} \mathbf{D}\beta_n = 0.$$

Furthermore, let us suppose that there exists the limit $F(x)$ of the distribution functions $F_n(x)$ as $n\to\infty$, i. e.

$$\lim_{n\to\infty} F_n(x) = F(x)$$

holds for all points x of continuity of the distribution function $F(x)$. Further, let us denote the distribution function of γ_n by $G_n(x)$. Then

$$\lim_{n\to\infty} G_n(x) = F(x).$$

PROOF.[6] Without loss of generality we may suppose that $\mathbf{M}\beta_n = 0$. Then, by virtue of Tchebyshev's inequality, we have

$$\mathbf{P}(|\beta_n| > \varepsilon) < \frac{\mathbf{D}^2\beta_n}{\varepsilon^2},$$

therefore, given any, arbitrarily small $\varepsilon > 0$ and $\delta > 0$, there exists a positive integer $n_0(\delta)$ such that

$$\mathbf{P}(|\beta_n| > \varepsilon) < \delta \quad \text{if} \quad n > n_0(\delta).$$

But then

$$G_n(x) = \mathbf{P}(\gamma_n < x) \leqq \mathbf{P}(\alpha_n < x+\varepsilon) + \mathbf{P}(\beta_n < -\varepsilon). \tag{2.3}$$

In fact, if $\alpha_n + \beta_n < x$, then either $\alpha_n < x + \varepsilon$ or $\alpha_n \geqq x + \varepsilon$, but in the latter case at the same time $\beta_n < -\varepsilon$ holds and we obtain (2. 3) by means of the theorem on total probability. Similarly,

$$G_n(x) = \mathbf{P}(\gamma_n < x) \geqq \mathbf{P}(\alpha_n < x-\varepsilon) - \mathbf{P}(\beta_n > \varepsilon); \tag{2.4}$$

in fact, if $\alpha_n < x - \varepsilon$, then either $\alpha_n + \beta_n < x$, or $\alpha_n + \beta_n \geqq x$, but in the latter case at the same time $\beta_n > \varepsilon$. Consequently, we have

$$F_n(x-\varepsilon) - \delta \leqq G_n(x) \leqq F_n(x+\varepsilon) + \delta.$$

Passing to the limit $n \to \infty$ and considering that ε and δ can be chosen arbitrarily small, it follows that

$$F(x-0) \leqq \varliminf_{n\to\infty} G_n(x) \leqq \varlimsup_{n\to\infty} G_n(x) \leqq F(x+0),$$

[5] Here and in what follows we shall denote then mean value of the random variable ξ by $\mathbf{M}\xi$ and its standard deviation by $\mathbf{D}\xi$.

[6] We give here the proof of this lemma of CRAMÉR because a similar method of proof is needed in the proof of Lemma 2.

i. e., that in all points of continuity of $F(x)$

$$\lim_{n\to\infty} G_n(x) = F(x).$$

This proves Lemma 1.

Theorem 1 follows immediately from this lemma, as

$$\mathbf{M}\left(\sum_{j=2}^{k}\frac{(j-1)\delta_j}{n+1-j}\right)=\sum_{j=2}^{k}\frac{j-1}{n+1-j} \quad\text{and}\quad \mathbf{D}^2\left(\sum_{j=2}^{k}\frac{(j-1)\delta_j}{n+1-j}\right)=\sum_{j=2}^{k}\frac{(j-1)^2}{(n+1-j)^2}$$

and thus the conditions of Lemma 1 are satisfied.

By means of Theorem 1 we can easily determine also the limiting distribution of η_k^*. We have $\eta_{n+1-k}^* = e^{-\zeta_k^*}$ and therefore

$$\mathbf{P}(n(1-\eta_{n+1-k}^*)<x) = \mathbf{P}\left(\zeta_k^* < \log\frac{1}{1-\frac{x}{n}}\right);$$

since

$$\lim_{n\to\infty}\frac{\log\dfrac{1}{1-\frac{x}{n}}}{\dfrac{x}{n}} = 1$$

and the Γ-distribution is continuous, it follows that $n(1-\eta_{n+1-k}^*)$ has in limit also a Γ-distribution of order k with the density function $\dfrac{t^{k-1}e^{-t}}{(k-1)!}$ $(t>0)$. Now, the random variables η_k are mutually independent and uniformly distributed in the interval $(0, 1)$. Because of the symmetry of the uniform distribution, the same holds also for the variables $1-\eta_k (k=1, 2, \ldots, n)$ and thus the variables

$$\eta_k^* = R_k(\eta_1, \eta_2, \ldots, \eta_n) \quad\text{and}\quad 1-\eta_{n+1-k}^* = R_k(1-\eta_1, 1-\eta_2, \ldots, 1-\eta_n)$$

have the same distribution. Hence it follows the following

THEOREM 2. *The distribution of the variables $n\eta_k^*$ and $n(1-\eta_{n+1-k}^*)$ in case of any fixed $k \geqq 1$, tends to the Γ-distribution of order k with the density function* $\dfrac{t^{k-1}e^{-t}}{(k-1)!}$ $(t>0)$.[7]

By means of Theorem 2 we can determine also the limiting distributions of ξ_k^* and ξ_{n+1-k}^*; these, however, — contrary to the limiting distributions of the variables η_k^* and ζ_k^* — will depend on the distribution function $F(x)$ (see [8]).

[7] The distribution of the variables η_k^* can be also determined exactly for finite n and after this passing to the limit Theorem 2 can be proved also in this manner by means of some simple calculations (see H. CRAMÉR [25]). We proved this theorem here by means of our method to show its application at first in a simple case.

Now we shall prove the following

THEOREM 3.[8] *The variables η_k^* and η_{n+1-j}^* are independent in the limit if $n \to \infty$ and at the same time k and j are fixed, namely*

$$\lim_{n\to\infty} \mathbf{P}\left(\eta_n^* < \frac{x}{n}, 1 - \eta_{n+1-j}^* < \frac{y}{n}\right) = \int_0^x \int_0^y \frac{u^{k-1} v^{j-1} e^{-(u+v)}}{(k-1)!\,(j-1)!}\, du\, dv \qquad (x > 0, y > 0).$$

PROOF. First of all we prove a lemma:

LEMMA 2. *Let us put $\gamma_n = \alpha_n + \beta_n$ where α_n and β_n are random variables, and let a random variable χ_n be given which is independent of α_n. Let us denote the distribution functions of α_n and χ_n by $F_n(x)$ and $H_n(x)$ respectively. Let us assume that the limit functions $F(x) = \lim_{n\to\infty} F_n(x)$ and $H(x) = \lim_{n\to\infty} H_n(x)$ exist and, further, that*

$$\lim_{n\to\infty} \mathbf{M}\beta_n = 0 \quad \textit{and} \quad \lim_{n\to\infty} \mathbf{D}\beta_n = 0.$$

In case all these conditions are satisfied, we have

$$\lim_{n\to\infty} \mathbf{P}(\gamma_n < x; \chi_n < y) = F(x)\,H(y),$$

i. e., the variables γ_n and χ_n become independent in the limit.

PROOF. Let us choose (as in the proof of Lemma 1) the value of the integer n_0 so large that

$$\mathbf{P}(|\beta_n| > \varepsilon) < \delta \text{ if } n > n_0$$

Similarly to the arguments applied in the proof of Lemma 1, it may be proved that, if $n > n_0$, where n_0 depends on the choice of the positive numbers ε and δ, then

$$\text{(2. 5)} \quad \mathbf{P}(\alpha_n < x - \varepsilon, \chi_n < y) - \delta \leqq \mathbf{P}(\gamma_n < x, \chi_n < y) \leqq \mathbf{P}(\alpha_n < x + \varepsilon, \chi_n < y) + \delta,$$

and as, by our assumption, α_n and χ_n are independent, therefore

$$\mathbf{P}(\alpha_n < x \pm \varepsilon, \chi_n < y) = F_n(x \pm \varepsilon)\,H_n(y)$$

and similarly to the proof of Lemma 1, we obtain that in all points of continuity of $F(x)$

$$\lim_{n\to\infty} \mathbf{P}(\gamma_n < x, \chi_n < y) = F(x)\,H(y)$$

holds. This completes the proof of Lemma 2.

Now

$$\zeta_{n+1-k}^* - \log n = \frac{\delta_{j+1}}{n-j} + \cdots + \frac{\delta_{n+1-k}}{k} - \log n + \zeta_j^*$$

[8] See CRAMÉR [25], p. 371. We discuss this well-known theorem here, because our method throws more light on the real ground of the fact expressed in the theorem, than its known proof.

where

$$\zeta_j^* = \frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \cdots + \frac{\delta_j}{n+1-j},$$

and therefore

$$\lim_{n\to\infty} \mathbf{M}\zeta_j^* = \lim_{n\to\infty} \mathbf{D}\zeta_j^* = 0.$$

As the sum $\frac{\delta_{j+1}}{n-j} + \cdots + \frac{\delta_{n+1-k}}{k} - \log n$ is independent of ζ_j^* and

$$\lim_{n\to\infty} n \log \frac{1}{1-\frac{y}{n}} = y,$$

further, we know that

$$\lim_{n\to\infty} \mathbf{P}(n\zeta_j^* < y) = \int_0^y \frac{t^{j-1}}{(j-1)!} e^{-t}\,dt,$$

we have

$$\text{(2. 6)}\quad \lim_{n\to\infty} \mathbf{P}\left(\eta_k^* < \frac{x}{n}, 1-\eta_{n+1-j}^* < \frac{y}{n}\right) = \lim_{n\to\infty} \mathbf{P}\left(\zeta_{n+1-k}^* - \zeta_j^* \geqq \log\frac{n}{x}\right) = \Gamma_j(y)$$

where

$$\Gamma_j(y) = \int_0^y \frac{t^{j-1}}{(j-1)!} e^{-t}\,dt,$$

On the other hand, by Lemma 1

$$\text{(2. 7)}\quad \lim_{n\to\infty} \mathbf{P}\left(\zeta_{n+1-k}^* - \zeta_j^* \geqq \log\frac{n}{x}\right) = \lim_{n\to\infty} \mathbf{P}\left(\zeta_{n+1-k}^* \geqq \log\frac{n}{x}\right)$$

and by virtue of Theorem 2

$$\text{(2. 8)}\quad \lim_{n\to\infty} \mathbf{P}\left(\zeta_{n+1-k}^* \geqq \log\frac{n}{x}\right) = \lim_{n\to\infty} \mathbf{P}\left(\eta_k^* < \frac{x}{n}\right) = \int_0^x \frac{t^{k-1}e^{-t}}{(k-1)!}\,dt.$$

From the relations (2. 6), (2. 7) and (2. 8), Theorem 3 follows.

By means of Theorem 3 we can determine the limiting distribution of the difference $\eta_n^* - \eta_1^*$. This is important, because $\eta_n^* - \eta_1^*$, the range of the sample $(\eta_1, \eta_2, \ldots, \eta_n)$, can be used to estimate the standard deviation of the population. As, for large n, the variable η_n^* is near to 1 and η_1^* is near to 0 with a probability near to 1, we obviously have to consider the variable $n[1-(\eta_n^* - \eta_1^*)]$ and as, by virtue of Theorem 3, $n\eta_1^*$ and $n(1-\eta_n^*)$ are independent in the limit, the limiting distribution of their sum equals the composition of their limiting distributions. As e^{-x} $(x>0)$ is the density function of the limiting distribution of both $n\eta_1^*$ and $n(1-\eta_n^*)$, the density function of the limiting distribution of $n[1-(\eta_n^*-\eta_1^*)]$ is

$$\int_0^x e^{-(x-y)}e^{-y}\,dy = xe^{-x} \qquad (x>0);$$

therefore $n[1-(\eta_n^*-\eta_1^*)]$ is in the limit a random variable having a Γ-distribution of order 2.

By means of the limiting distributions of the random variables η_k^*, the limiting distributions of the random variables ξ_k^* can also be determined.

Hitherto we have considered the limiting distributions of the order statistics ξ_k^* (resp. those of their transformed values η_k^* and ζ_k^*) under the condition that the index k (resp. $j=n+1-k$) is fixed and at the same time $n\to\infty$; this set of problems is called the study of the "extreme values" of the sample. We now turn to the study of the limiting distributions of the variables ξ_k^* (resp. of η_k^* and ζ_k^*) under the condition that together with n k tends also to infinity, namely so that $|k-nq|=o(\sqrt{n})$, where q is a constant $(0<q<1)$. The variable ξ_{k+1}^* (where $k=[nq]$)[9] satisfies this condition; this variable is called the *q-quantile* of the sample. In the special case $n=2m+1$ where m is an integer, the variable ξ_{m+1}^* is the median of the sample: obviously, the q-quantile of the sample is nothing else but the q-quantile of the sample distribution function and thus the median of the sample is nothing else but the median of the sample distribution function. Consequently, if n is an even integer, i. e. $n=2m$, then $\frac{1}{2}(\xi_m^*+\xi_{m+1}^*)$ is called the median of the sample.

We shall now prove the following theorem containing the proposition that the q-quantiles of the sample in the limit are normally distributed, if the distribution function $F(x)$ of the population satisfies certain simple conditions.

THEOREM 4.[10] *Let us suppose that the density function $f(x)=F'(x)$ of the common distribution function $F(x)$ of the mutually independent random variables $\xi_1,\xi_2,\ldots,\xi_n$ exists and that $f(x)$ is continuous and positive in the interval $a<x<b$; then, if $0<F(a)<q<F(b)<1$ and further, if $|k_n-nq|=o(\sqrt{n})$ (and thus, a fortiori, $\lim_{n\to\infty}\frac{k_n}{n}=q$), then $\xi_{k_n}^*$ is, in the limit, normally distributed with the mean value $Q=F^{-1}(q)$, which is the q-quantile of the distribution function $F(x)$, i. e.*

$$\lim_{n\to\infty}\mathbf{P}\left(\frac{\xi_{k_n}^*-Q}{\frac{1}{f(Q)}\sqrt{\frac{q(1-q)}{n}}}<x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}\,dt.$$

[9] $[x]$ is the largest integer for which $[x]\leqq x$.

[10] This theorem is contained in a general theorem of N. V. SMIRNOV [8g].

PROOF. First consider the limiting distribution of

$$\zeta^*_{n+1-k_n} = \sum_{j=1}^{n+1-k_n} \frac{\delta_j}{n+1-j} \tag{2.9}$$

where the variables δ_j are mutually independent and exponentially distributed with the distribution function $1-e^{-x}$ $(x>0)$, that is to say, $\mathbf{M}\delta_j = \mathbf{D}\delta_j = 1$ and

$$\mathbf{M}(|\delta_j-1|^3) = \int_0^\infty |x-1|^3 e^{-x}\,dx < \int_0^1 e^{-x}\,dx + \int_0^\infty x^3 e^{-x}\,dx \leqq 5.$$

Then, however,

$$\begin{cases} M_n = \mathbf{M}\zeta^*_{n+1-k_n} = \sum\limits_{j=1}^{n+1-k_n} \dfrac{1}{n+1-j}, \\ S_n^2 = \mathbf{D}^2\zeta^*_{n+1-k_n} = \sum\limits_{j=1}^{n+1-k_n} \dfrac{1}{(n+1-j)^2}, \\ K_n^3 = \sum\limits_{j=1}^{n+1-k_n} \mathbf{M}\left|\dfrac{\delta_j-1}{n+1-j}\right|^3 \leqq 5 \sum\limits_{j=1}^{n+1-k_n} \dfrac{1}{(n+1-j)^3}, \end{cases} \tag{2.10}$$

and thus

$$\frac{K_n}{S_n} \leqq \frac{5}{k_n}. \tag{2.11}$$

Therefore, if $n\to\infty$, then $\frac{K_n}{S_n}\to 0$; this means that the central limit theorem in LIAPUNOV's form can be applied to the sequence of sums (2.9) and thus

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{\zeta^*_{n+1-k_n} - M_n}{S_n} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt. \tag{2.12}$$

Now, it is known that

$$\sum_{k=1}^{m} \frac{1}{k} = \log m + C + \Delta_m$$

where C is the Euler constant, and an $A>0$ constant can be found such that $|\Delta_m| < \frac{A}{m}$. By means of some simple calculations it can be verified that

$$\sum_{k=h}^{H} \frac{1}{k^2} = \frac{1}{h} - \frac{1}{H} + \frac{\vartheta}{h(h-1)} \qquad (0<\vartheta<1). \tag{2.13}$$

Thus we have

$$M_n = \log\frac{n}{k_n} + \varepsilon_n', \tag{2.14}$$

$$S_n = \frac{1}{k_n} - \frac{1}{h} + \varepsilon_n'', \tag{2.15}$$

where $|\varepsilon_n'| < \frac{A}{k_n}$ and $|\varepsilon_n''| < \frac{B}{k_n^2}$, A and B being constants not depending on n. Therefore

(2. 16) $$\frac{\zeta^*_{n+1-k_n} - M_n}{S_n} = \frac{\zeta^*_{n+1-k_n} - \log \frac{n}{k_n}}{\sqrt{\frac{n-k_n}{nk_n}}} + \varepsilon_n'''$$

where $|\varepsilon_n'''| < L \sqrt{\frac{n}{k_n(n-k_n)}}$ and L is a constant not depending on n. But, if $n \to \infty$, then $\frac{k_n}{n} \to q$ and thus $\varepsilon_n''' \to 0$. Hence

(2. 17) $$\lim_{n\to\infty} \mathbf{P}\left(\frac{\zeta^*_{n+1-k_n} - \log \frac{n}{k_n}}{\sqrt{\frac{n-k_n}{nk_n}}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

For brevity, let us introduce the notation $q_n = \frac{k_n}{n}$, then, by virtue of (2. 17) and taking into account that owing to $|k_n - nq| = o(\sqrt{n})$, we have

$$\log \frac{1}{qn} - \log \frac{1}{q} = o\left(\frac{1}{\sqrt{n}}\right),$$

it follows that

(2. 18) $$\lim_{n\to\infty} \mathbf{P}\left(\frac{\zeta^*_{n+1-k_n} - \log \frac{1}{q}}{\sqrt{\frac{1-q}{nq}}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

As $\zeta^*_{n+1-k_n} = \log \frac{1}{F(\xi^*_{k_n})}$, if follows from (2. 18) that

(2. 19) $$\lim_{n\to\infty} \mathbf{P}\left(\xi^*_{k_n} > F^{-1}\left(qe^{-x\sqrt{\frac{1-q}{nq}}}\right)\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

But, in view of the mean value theorem of differential calculus, we have

(2. 20) $$F^{-1}\left(qe^{-x\sqrt{\frac{1-q}{nq}}}\right) = Q + \frac{q\left(e^{-x\sqrt{\frac{1-q}{nq}}} - 1\right)}{f(Q\vartheta_n)}$$

where $\lim_{n\to\infty} \vartheta_n = 1$.

It follows that

(2. 21) $$\lim_{n\to\infty} \mathbf{P}\left(\frac{\xi^*_{k_n} - Q}{\frac{1}{f(Q)}\sqrt{\frac{q(1-q)}{n}}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt,$$

which was to be proved.

The statement of our theorem can also be characterized by that the q-quantile of a sample of size n in case of large n is approximately normally distributed around Q, i. e. around the q-quantile of the population, with the standard deviation $\frac{1}{f(Q)}\sqrt{\frac{q(1-q)}{n}}$.

In this way, by means of the sample median, an interval containing the median of the population with probability arbitrarily near to 1, can be given. In the special case of symmetrical, for example, normal distribution, the median of the distribution coincides with its mean and in this way we can estimate the mean value of the population.

The theory of order statistics has a widespread applicability in the statistical quality control of mass production,[11] which is an important field of application of probability theory. Let us assume that a certain measurement of some engine parts produced on an automatic machine displays some small random fluctuations from specimen to specimen, therefore its value can be considered as a random variable. Let us suppose that, under standard manufacturing circumstances, the distribution function of this measurement is the (continuous) function $F(x)$; to control the process of production, at regular time intervals we draw a sample of size n — e. g. of size 5.

We take the considered measurement of the values of the sample and we mark them on a perpendicular straight line drawn across the abscissa corresponding to the point of time of sampling on the „control chart" and we mark their places with dots; the values of the sample will be placed automatically in order of magnitude. In order to detect any irregularity in the process of production (e. g. the displacement of the adjustment of the automatic machine or the attrition of certain parts of the producing machine etc.), we draw 5 bands determined by parallel straight lines, giving intervals containing the least, the second, the third, the fourth, and the largest value, respectively, of the sample of size 5 at the same time with a given probability — e. g. 95 % — under standard manufacturing circumstances. The determination of these intervals is very easy by what has been said above. In fact, if ξ_k^* denotes the k-th sample value in order of magnitude ($k=1, 2, 3, 4, 5$), then, as we have seen, we can exactly determine the individual and joint distributions of the variables

$$\zeta_k^* = \log\frac{1}{F(\xi_{6-k}^*)} \qquad (k=1, 2, 3, 4, 5).$$

The practical application of this method in quality control is dealt with by the Department of Mathematical Statistics of the Institute for Applied

[11] Cf. the work of L. I. Braginsky [26]; by means of the theory of order statistics, the calculations of Braginsky which are not quite exact can be put in a precise form; in the practical application it is suitable to carry out the control charts on the basis of these precise calculations.

Mathematics of the Hungarian Academy of Sciences; tables needed for the use of the method are also prepared.

We shall not continue the enumeration of theorems obtainable by means of this method, we only emphasize, that our method consists in deducing all these theorems, by means of (1.9), from the theory of limiting distribution of sums of mutually independent random variables.

§ 3. Formulation of some new theorems concerning the comparison of the distribution function of a population and that of a sample drawn from it

Hitherto we have only shown how our method makes it possible to prove certain well-known results of order statistics. Now we shall show the new results which can be obtained by means of the same method.

A. N. KOLMOGOROV [6b] proved a fundamental theorem giving a test for the hypothesis that a sample has been drawn from a population having a given distribution. By means of this test we can infer to the unknown distribution of the population from the distribution of sample values.[12] Let us define

$$F_n(x)=\begin{cases} 0 & \text{if } x \leqq \xi_1^*, \\ \dfrac{k}{n} & \text{if } \xi_k^* < x \leqq \xi_{k+1}^*, \\ 1 & \text{if } \xi_n^* < x, \end{cases} \tag{3.1}$$

i. e. $F_n(x)$ is the distribution function of the sample, in other words, the frequency ratio of the values less than x in the sample.

KOLMOGOROV's theorem is as follows:

$$\lim_{n\to\infty} \mathbf{P}\left(\sqrt{n} \sup_{-\infty<x<+\infty} |F_n(x)-F(x)|<y\right)=\begin{cases} \sum\limits_{k=-\infty}^{\infty}(-1)^k e^{-2k^2y^2} & \text{if } y>0 \\ 0 & \text{if } y\leqq 0. \end{cases} \tag{3.2}$$

KOLMOGOROV's theorem therefore gives the limiting distribution of the supremum of absolute value of the difference between the distribution function of the sample and that of the population. This limiting distribution does not depend on the distribution function $F(x)$ of the population which is assumed, for the validity of theorem, to be continuous. KOLMOGOROV's theorem considers the difference $|F_n(x)-F(x)|$ with the same weight, regardless to the value of $F(x)$; so e. g. the difference $|F_n(x)-F(x)|=0{\cdot}01$ has the same weight in a point x with $F(x)=0{\cdot}5$ (where this difference is 2% of the value of $F(x)$) as a point x with $F(x)=0{\cdot}01$ (where this difference is 100%! of the value of $F(x)$). We can avoid this by considering the quotient $\dfrac{|F_n(x)-F(x)|}{F(x)}$ instead of $|F_n(x)-F(x)|$, that is to say, by considering the

[12] I. e. we can give confidence limits for the unknown distribution function.

relative error of $F_n(x)$. In this way, the idea arises, naturally, to consider the limiting distribution of the supremum of the quotient $\frac{|F_n(x)-F(x)|}{F(x)}$ which characterizes the relative deviation of distribution function of the population and that of the sample.

A theorem similar to that of KOLMOGOROV's was proved by N. V. SMIRNOV concerning the one-sided deviation of the sample and population distribution functions. SMIRNOV's theorem is as follows:

$$(3.3)\qquad \lim_{n\to\infty} \mathbf{P}(\sqrt{n} \sup_{-\infty<x<+\infty} (F_n(x)-F(x))<y) = \begin{cases} 1-e^{-2y^2} & \text{if } y>0, \\ 0 & \text{if } y\leqq 0. \end{cases}$$

We shall consider also the analogous problem for relative deviations.

All these problems can be successfully solved by means of the above method. In the course of solving these problems a natural limitation is to be adopted: as $F(x)$ takes on arbitrarily small values, it is not suitable to consider the supremum of $\frac{F_n(x)-F(x)}{F(x)}$ or respectively $\left|\frac{F_n(x)-F(x)}{F(x)}\right|$ taken in the whole interval $-\infty<x<+\infty$, but to restrict ourselves to an interval $x_a\leqq x<+\infty$, where the abscissa x_a is defined by the relation $F(x_a)=a>0$; the value of a, however, can be an arbitrarily small positive value. In § 5 we shall prove the following results:

THEOREM 5.

$$(3.4)\qquad \lim_{n\to\infty} \mathbf{P}\left(\sqrt{n}\sup_{a\leqq F(x)} \frac{F_n(x)-F(x)}{F(x)}<y\right)= \begin{cases} \sqrt{\frac{2}{\pi}}\int\limits_0^{y\sqrt{\frac{a}{1-a}}} e^{-\frac{t^2}{2}}dt & \text{if } y>0, \\ 0 \quad \text{if } y\leqq 0. \end{cases}$$

THEOREM 6.

$$(3.5)\qquad \lim_{n\to\infty} \mathbf{P}\left(\sqrt{n}\sup_{a\leqq F(x)} \left|\frac{F_n(x)-F(x)}{F(x)}\right|<y\right)= \begin{cases} \frac{4}{\pi}\sum\limits_{k=0}^{\infty}(-1)^k \frac{e^{-\frac{(2k+1)^2\pi^2}{8}\frac{1-a}{ay^2}}}{2k+1} & \text{if } y>0, \\ 0 \quad \text{if } y\leqq 0. \end{cases}$$

We may consider the limiting distribution of the supremum of $\frac{F_n(x)-F(x)}{F(x)}$, and of its absolute value taken in the interval $x_a\leqq x\leqq x_b$, respectively, where the abscissae x_a and x_b are defined by the relations $F(x_a)=a>0$ and $F(x_b)=b<1$ $(0<a<b<1)$. We then arrive at the following theorems.

THEOREM 7.

$$\lim_{n\to\infty} \mathbf{P}\left(\sqrt{n} \sup_{a\leqq F(x)\leqq b} \frac{F_n(x)-F(x)}{F(x)} < y\right) = \tag{3.6}$$

$$= \frac{1}{\pi} \int_{-\infty}^{y\sqrt{\frac{b}{1-b}}} e^{-\frac{u^2}{2}} \left(\int_0^{\left(\sqrt{\frac{b}{1-b}}y-u\right)\cdot\sqrt{\frac{a(1-b)}{b-a}}} e^{-\frac{t^2}{2}} dt \right) du \quad (-\infty < y < +\infty).$$

THEOREM 8.

$$\lim_{n\to\infty} \mathbf{P}\left(\sqrt{n} \sup_{a\leqq F(x)\leqq b} \left| \frac{F_n(x)-F(x)}{F(x)} \right| < y\right) =$$

$$= \begin{cases} \dfrac{4}{\pi} \displaystyle\sum_{k=0}^{\infty} (-1)^k \frac{e^{-\frac{(2k+1)^2\pi^2(1-a)}{8ay^2}}}{2k+1} E_k & \text{if } y>0, \\ 0 \quad \text{if} \quad y \leqq 0; \end{cases} \tag{3.7}$$

where

$$E_k = 1 - \frac{2}{\sqrt{2\pi}} \int_{y\sqrt{\frac{b}{1-b}}}^{\infty} e^{-\frac{u^2}{2}} du + \varrho_k$$

and

$$\varrho_k = \frac{2e^{-\frac{by^2}{2(1-b)}}}{\sqrt{2\pi}\sqrt{\frac{a}{1-a}}y} \int_0^{(2k+1)\frac{\pi}{2}} e^{\frac{(1-b)u^2}{2by^2}} \sin u\, du.$$

These theorems provide tests for verifying the hypothesis that the sample $(\xi_1, \xi_2, \ldots, \xi_n)$ has been drawn from a population of the distribution function $F(x)$. The character of these tests consists in that they give a band around $F(x)$ in which, if the hypothesis is true, the sample distribution function $F_n(x)$ have to lie with a certain probability and the width of this band in all points x being proportional to $F(x)$ This band, however, is not a symmetrical one. To overcome this difficulty, we apply the test twice, first to the sample $(\xi_1, \xi_2, \ldots, \xi_n)$ having the population distribution function $F(x)$ and then to the sample $(-\xi_1, -\xi_2, \ldots, -\xi_n)$ having the population distribution function $G(x) = 1 - F(-x)$. In order to illustrate this, let us denote the distribution function of the sample $(-\xi_1, -\xi_2, \ldots, -\xi_n)$ by $G_n(x)$ and let A be the event that

$$\sqrt{n} \sup_{a\leqq F(x)} \left| \frac{F_n(x)-F(x)}{F(x)} \right| < y,$$

and B the event that

$$\sqrt{n} \sup_{a\leqq F(x)} \left| \frac{G_n(x)-G(x)}{G(x)} \right| < y, \quad \text{i. e.,} \quad \sqrt{n} \sup_{F(x')\leqq 1-a} \left| \frac{F_n(x')-F(x')}{1-F(x')} \right| < y;$$

finally, let us denote the simultaneous occurrence of A and B by C. Taking into account that

$$\mathbf{P}(C)=\mathbf{P}(AB)=\mathbf{P}(A)+\mathbf{P}(B)-\mathbf{P}(A+B)$$

and in case of occurrence $A+B$ we have obviously at the same time

$$\sqrt{n}\sup_{0\leqq F(x)\leqq 1}|F_n(x)-F(x)|<y;$$

and this last event, by KOLMOGOROV's theorem, has in the limit the probability

$$K(y)=\sum_{k=-\infty}^{\infty}(-1)^k e^{-2k^2y^2}, \tag{3.8}$$

Therefore $\mathbf{P}(A+B)\leqq K(y)$ in the limit and in the same case

$$\mathbf{P}(C)\geqq \mathbf{P}(A)+\mathbf{P}(B)-K(y).$$

The probabilities $\mathbf{P}(A)$ and $\mathbf{P}(B)$ are equal and their common value is given by (3. 5). Thus the probability of the event that the sample distribution function $F_n(x)$ lies in the intersection of bands defined by the above two conditions corresponding to the sample $(\xi_1,\xi_2,\dots,\xi_n)$ and $(-\xi_1,-\xi_2,\dots,-\xi_n)$, respectively, is not less than $2L\left(y\sqrt{\frac{a}{1-a}}\right)-K(y)$ in the limit, where $L(z)=\sum_{k=0}^{\infty}(-1)^k\frac{e^{-\frac{(2k+1)^2\pi^2}{8z^2}}}{2(k+1)}$ $(z>0)$ and $K(y)$ is the function defined by (3. 8).

Let us point out a most surprising corollary of Theorem 7. From the theorem (3. 3) of SMIRNOV, we get

$$\lim_{n\to\infty}\mathbf{P}(\sup_{-\infty<x<+\infty}(F_n(x)-F(x))<0)=0, \tag{3.9}$$

i. e. the probability of the event that the sample distribution function does not exceed the population distribution function all along the interval $-\infty<x<+\infty$, tends to 0 as $n\to\infty$. From Theorem 5 it follows that

$$\lim_{n\to\infty}\mathbf{P}(\sup_{x_a\leqq x<+\infty}(F_n(x)-F(x))<0)=0, \tag{3.10}$$

i. e. the same is true for the interval $x_a\leqq x<+\infty$. On the other hand, by Theorem 7,

$$\lim_{n\to\infty}\mathbf{P}(\sup_{x_a\leqq x\leqq x_b}(F_n(x)-F(x))<0)=\frac{1}{\pi}\int_0^{\infty}e^{-\frac{u^2}{2}}\int_0^{u\sqrt{\frac{a(1-b)}{b-a}}}e^{-\frac{t^2}{2}}dt\,du>0, \tag{3.11a}$$

i. e. the probability of the event that the sample distribution function does not exceed the population distribution function all along the interval in which the value of $F(x)$ lies between arbitrarily fixed values a and b $(0<a<b<1)$, remains positive also in the limit. This result, obviously, is important also from the point of view of statistical practice.

The result that the limit on the left side of (3. 11a) is positive, was also proved by GIHMAN [16]; moreover, he obtained that

$$\text{(3. 11b)} \qquad \lim_{n\to\infty} \mathbf{P}(\sup_{x_a \leqq x \leqq x_b} (F_n(x) - F(x)) < 0) = \frac{1}{\pi} \text{ arc sin } \sqrt{\frac{a(1-b)}{b(1-a)}}.$$

GIHMAN mentioned further that the result (3. 11b) has been already known to GNEDENKO. The terms on the right sides of (3. 11a) and (3. 11b) are, of course, identical. This follows from the following consideration: the right side of (3. 11a) is nothing else than two times the probability of the event that a random point normally distributed in the plane (x, y) and having the probability density function $\frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$, lies in the infinite sector $0 < x < +\infty$, $0 < y < x \sqrt{\frac{a(1-b)}{b-a}}$, and this probability is equal to

$$\text{(3. 12)} \qquad \frac{2 \text{ arc tg} \sqrt{\frac{a(1-b)}{b-a}}}{2\pi} = \frac{1}{\pi} \text{ arc sin } \sqrt{\frac{a(1-b)}{b(1-a)}}.$$

Indeed, because of the circular symmetry of the normal distribution having the density function $\frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$, the probability corresponding to the infinite sector of angle φ is $\frac{\varphi}{2\pi}$.

Theorems 5—8 will be proved in § 5. First in § 4 we shall prove some auxiliary theorems which are of interest in themselves too.

§ 4. Some new limiting distribution theorems

Let a sequence be given consisting of the sets of random variables

$$\xi_{n,1}, \xi_{n,2}, \ldots, \xi_{n,N_n} \qquad (n = 1, 2, \ldots).$$

Let us assume that the random variables $\xi_{n,k}$ have the expectation 0 and a finite variance, further, that the random variables having the same first index n $(n = 1, 2, \ldots)$ are mutually independent and satisfy LINDEBERG's condition, that is to say, introducing the notations

$$F_{n,k}(x) = \mathbf{P}(\xi_{n,k} < x); \quad S_{n,k} = \sum_{\nu=1}^{k} \xi_{n,\nu}; \quad B_n^2 = \mathbf{D}^2 S_{n,N_n} = \sum_{k=1}^{N_n} \mathbf{D}^2 \xi_{n,k},$$

we suppose

$$\mathbf{M}\xi_{n,k} = \int_{-\infty}^{\infty} x\, dF_{n,k}(x) = 0,$$

and

$$\text{(4. 1)} \qquad \lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{N_n} \int_{|x| > \varepsilon B_n} x^2 dF_{n,k}(x) = 0 \quad \text{if} \quad \varepsilon > 0.$$

Concerning these sequences satisfying the above conditions we shall prove the following theorems.

THEOREM 9.

$$\lim_{n\to\infty}\mathbf{P}(\max_{1\leq k\leq N_n} S_{n,k}<xB_n)=\begin{cases}\sqrt{\dfrac{2}{\pi}}\displaystyle\int_0^x e^{-\frac{t^2}{2}}dt & \text{if } x>0,\\ 0 \text{ if } x\leqq 0.\end{cases}\tag{4.2}$$

THEOREM 10.

$$\lim_{n\to\infty}\mathbf{P}(\max_{1\leq k\leq N_n}|S_{n,k}|<xB_n)=\begin{cases}\dfrac{4}{\pi}\displaystyle\sum_{k=0}^{\infty}(-1)^k\frac{e^{-\frac{(2k+1)^2\pi^2}{8x^2}}}{2k+1} & \text{if } x>0,\\ 0 \text{ if } x\leqq 0\end{cases}\tag{4.3}$$

THEOREM 11.

$$\lim_{n\to\infty}\mathbf{P}(-yB_n\leqq\min_{1\leqq k\leqq N_n}S_{n,k}\leqq\max_{1\leqq k\leqq N_n}S_{n,k}<xB_n)=\tag{4.4}$$

$$=\begin{cases}\dfrac{4}{\pi}\displaystyle\sum_{k=0}^{\infty}\frac{e^{-\frac{(2k+1)^2\pi^2}{2(x+y)^2}}\sin(2k+1)\pi\dfrac{x}{x+y}}{2k+1} & \text{if } x>0 \text{ and } y\geqq 0,\\ 0 \text{ if either } x\leqq 0 \text{ or } y<0.\end{cases}$$

REMARK. In case $y=x$, Theorem 11 reduces to Theorem 10.

THEOREM 12. *Let* $A_n^2=\mathbf{D}^2 S_{n,M_n}$ *with* $1\leqq M_n<N_n$ *and*

$$\lim_{n\to\infty}\frac{A_n}{B_n}=\lambda \qquad (0\leqq\lambda<1).$$

Then

$$\lim_{n\to\infty}\mathbf{P}(\max_{M_n<k\leqq N_n}|S_{n,k}|<yB_n)=\tag{4.5}$$

$$=\begin{cases}\dfrac{4}{\pi}\displaystyle\sum_{k=0}^{\infty}(-1)^k\frac{e^{-\frac{(2k+1)^2\pi^2}{8y^2}}}{2k+1}\left(1-\frac{2}{\sqrt{2\pi}}\int_{\frac{y}{\lambda}}^{\infty}e^{-\frac{u^2}{2}}du+\varrho_k\right) & \text{if } y>0,\\ 0 \text{ if } y\leqq 0\end{cases}$$

where

$$\varrho_k=\frac{2\lambda e^{-\frac{y^2}{2\lambda^2}}}{\sqrt{2\pi y}}\int_0^{(2k+1)\frac{\pi}{2}}e^{\frac{\lambda^2u^2}{2y^2}}\sin u\,du.$$

REMARK. In the special case of $M_n=1$ (i. e. for $\lambda=0$), Theorem 12 is identical with Theorem 10.

For the special case in which all the considered random variables $\xi_{n,k}$ have the same distribution, Theorems 9 and 10 were proved by P. ERDŐS and M. KAC [22].[12] In the proofs of the above more general theorems, we modify their proofs inasmuch as we apply an ordinary (one-dimensional) limiting distribution theorem instead of the multi-dimensional limit theorem used by them; this enables us to generalize their results. We shall use Theorems 9—12 in § 5 to determine the limiting distribution of the random variables

$$\sup_{a \leqq F(x) \leqq b} \frac{F_n(x) - F(x)}{F(x)} \quad \text{and} \quad \sup_{a \leqq F(x) \leqq b} \left| \frac{F_n(x) - F(x)}{F(x)} \right|$$

where $F_n(x)$ denotes again the distribution function of a sample of n mutually independent observations concerning the random variable ξ having a continuous distribution function $F(x)$ and further $0 < a < b \leqq 1$.

Let us turn to the proof of Theorem 9. Let us put

$$P_n(x) = \mathbf{P}(\max_{1 \leqq k \leqq N_n} S_{n,k} < xB_n).$$

Let $\eta_1, \eta_2, \ldots, \eta_n, \ldots$ be mutually independent random variables which are normally distributed with mean 0 and variance 1, and let us introduce the random variables

$$\zeta_k = \sum_{\nu=1}^{k} \eta_\nu \qquad (k = 1, 2, \ldots).$$

First of all we shall prove that for any $\varepsilon > 0$ and for any positive integer k we have

$$\text{(4. 6)} \qquad \underline{\lim}_{n \to \infty} P_n(x) \geqq \mathbf{P}(\max(\zeta_1, \zeta_2, \ldots, \zeta_k) < (x - \varepsilon)\sqrt{k}) - \frac{1}{\varepsilon^2 k}$$

and

$$\text{(4. 7)} \qquad \overline{\lim}_{n \to \infty} P_n(x) \leqq \mathbf{P}(\max(\zeta_1, \zeta_2, \ldots, \zeta_k) < x\sqrt{k}).$$

For, let m_j be the least positive integer satisfying

$$\sum_{r=1}^{m_j} \mathbf{D}^2 \xi_{n,r} \geqq \frac{j}{k} B_n^2$$

$(j = 1, 2, \ldots, k)$. Obviously, $1 \leqq m_1 \leqq m_2 \leqq \ldots \leqq m_k = N_n$. Let us define now the following variables:

$$\text{(4. 8)} \qquad \Delta_{n,1} = S_{n,m_1}; \quad \Delta_{n,j} = S_{n,m_j} - S_{n,m_{j-1}} \qquad (j = 2, 3, \ldots, k).$$

We can see easily that for any fixed j $(j = 1, 2, \ldots, k)$ the Lindeberg condition holds for the sequence

$$\text{(4.9)} \qquad \xi_{n,m_{j-1}+1}, \xi_{n,m_{j-1}+2}, \ldots, \xi_{n,m_j} \qquad (n = 1, 2, 3, \ldots).$$

[12] Their method was generalized by M. D. DONSKER [20b]. See further the papers by A. WALD [27], [28], and K. L. CHUNG [29]. They consider the limiting distributions of the supremum of the first n partial sums under the conditions that the variables have the same distribution and that the variables have finite third moments, respectively. CHUNG gives for this latter special case an estimate for the remainder term also. ERDŐS and KAC remarked that their theorems can be proved under more general conditions.

Indeed, introducing the notation

$$B_{n,j}^2 = \mathbf{D}^2 \Delta_{n,j}$$

and using the relation

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sup_{1\leq k\leq N_n} \mathbf{D}^2 \xi_{n,k} = 0$$

which trivially follows from (4. 1), we obtain that, for any $\delta > 0$,

$$\frac{1-\delta}{k} B_n^2 \leq B_{n,j}^2 \leq \frac{1+\delta}{k} B_n^2, \quad \text{if} \quad n > n_0(\delta) \tag{4. 10}$$

holds. Thus for any $\varepsilon > 0$

$$\frac{1}{B_{n,j}^2} \sum_{r=m_{j-1}+1}^{m_j} \int\limits_{|x|>\varepsilon B_{n,j}} x^2\, dF_{n,r}(x) \leq \frac{k}{1-\delta} \frac{1}{B_n^2} \sum_{r=1}^{N_n} \int\limits_{|x|>\varepsilon' B_n} x^2\, dF_{n,r}(x)$$

if $n > n_0(\delta)$, and $\varepsilon' = \varepsilon \sqrt{\dfrac{1-\delta}{k}}$. The Lindeberg condition is therefore actually satisfied by the sequence (4. 9). Therefore, by the central limit theorem,

$$\lim_{n\to\infty} \mathbf{P}(\Delta_{n,j} < x B_{n,j}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt \qquad (j = 1, 2, \ldots, k). \tag{4. 11}$$

But the random variables $\Delta_{n,1}, \Delta_{n,2}, \ldots, \Delta_{n,k}$ are mutually independent and

$$\lim_{n\to\infty} \frac{B_{n,j}}{B_n} = \frac{1}{\sqrt{k}} \qquad (j = 1, 2, \ldots, k);$$

hence

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{\Delta_{n,j}}{B_n} < \frac{x_j}{\sqrt{k}};\ j = 1, 2, \ldots, k\right) = \tag{4. 12}$$

$$= \frac{1}{(\sqrt{2\pi})^k} \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_k} e^{-\frac{1}{2}\left(\sum\limits_{j=1}^{k} t_j^2\right)}\, dt_1 \ldots dt_k$$

i. e.

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{\Delta_{n,1} + \Delta_{n,2} + \cdots + \Delta_{n,j}}{B_n} < x;\ j = 1, 2, \ldots, k\right) = \tag{4. 13}$$

$$= \frac{1}{(\sqrt{2\pi})^k} \int\!\!\int_{(T_x)} \cdots \int e^{-\frac{1}{2}\left(\sum\limits_{j=1}^{k} t_j^2\right)}\, dt_1\, dt_2 \ldots dt_k;$$

where the integration is to be extended over the domain T_x defined by

$$T_x\colon \{-\infty < t_1 + t_2 + \cdots + t_j < x\sqrt{k};\ j = 1, 2, \ldots, k\}.$$

Hence we obtain

$$\lim_{n\to\infty} \mathbf{P}(\max_{1\leq j\leq k} S_{n,m_j} < x B_n) = \mathbf{P}(\max_{1\leq j\leq k} \zeta_j < x\sqrt{k}). \tag{4. 14}$$

Let us put

$$Q_{n,k}(x) = \mathbf{P}(\max_{1 \leqq j \leqq k} S_{n,m_j} < xB_n),$$

and let $\Pi_{n,r}(x)$ denote the probability of the event that $S_{n,r}$ is the first of the sums $S_{n,j}$ $(j = 1, 2, \ldots, r)$ which is $\geqq xB_n$, i. e. let

$$\Pi_{n,r}(x) = \mathbf{P}(S_{n,r} \geqq xB_n;\ \max_{1 \leqq j \leqq r-1} S_{n,j} < xB_n).$$

Obviously,

$$\sum_{r=1}^{N_n} \Pi_{n,r}(x) = 1 - P_n(x) \leqq 1.$$

Let us suppose $m_{j-1} < r \leqq m_j$; introducing the notations

$$\Pi_{n,r}^{(1)}(x) = \mathbf{P}(S_{n,r} \geqq xB_n;\ \max_{1 \leqq j \leqq r-1} S_{n,j} < xB_n;\ |S_{n,m_j} - S_{n,r}| \geqq \varepsilon B_n)$$

and

$$\Pi_{n,r}^{(2)}(x) = \mathbf{P}(S_{n,r} \geqq xB_n;\ \max_{1 \leqq j \leqq r-1} S_{n,j} < xB_n;\ |S_{n,m_j} - S_{n,r}| < \varepsilon B_n).$$

we get evidently

$$\Pi_{n,r}(x) = \Pi_{n,r}^{(1)}(x) + \Pi_{n,r}^{(2)}(x).$$

Let us apply TCHEBYSHEV's inequality:

$$\Pi_{n,r}^{(1)}(x) = \Pi_{n,r}(x)\mathbf{P}(|S_{n,m_j} - S_{n,r}| \geqq \varepsilon B_n) \leqq \Pi_{n,r}(x)\frac{B_{n,j}^2}{\varepsilon^2 B_n^2}$$

and consider the relation (4. 10); thus we obtain that

$$\Pi_{n,r}^{(1)}(x) \leqq \Pi_{n,r}(x)\frac{1+\delta}{\varepsilon^2 k};$$

therefore

$$1 - P_n(x) = \sum_{r=1}^{N_n} \Pi_{n,r}(x) \leqq \frac{1+\delta}{\varepsilon^2 k} + \sum_{r=1}^{N_n} \Pi_{n,r}^{(2)}(x). \tag{4. 15}$$

On the other hand,

$$\sum_{r=1}^{N_n} \Pi_{n,r}^{(2)}(x) \leqq 1 - Q_{n,k}(x-\varepsilon) \tag{4. 16}$$

as from the relations

$$S_{n,r} \geqq xB_n \quad \text{and} \quad |S_{n,m_j} - S_{n,r}| < \varepsilon B_n$$

it follows that

$$S_{n,m_j} > (x-\varepsilon)B_n.$$

Thus we have

$$1 - P_n(x) \leqq \frac{1+\delta}{\varepsilon^2 k} + 1 - Q_{n,k}(x-\varepsilon).$$

Further, on account of the trivial inequalities $P_n(x) \leqq Q_{n,k}(x)$ $(k = 1, 2, \ldots)$, we obtain

$$Q_{n,k}(x-\varepsilon) - \frac{1+\delta}{\varepsilon^2 k} \leqq P_n(x) \leqq Q_{n,k}(x). \tag{4. 17}$$

Comparing the above relation (4. 17) with (4. 14), we have just the desired inequalities (4. 6) and (4. 7).

Let us consider now the special case in which the variables $\xi_{n,k}$ assume only the values $+1$ and -1, and

$$\mathbf{P}(\xi_{n,k}=+1)=\mathbf{P}(\xi_{n,k}=-1)=\frac{1}{2} \qquad (k=1,2,\ldots,N_n=n;\ n=1,2,\ldots).$$

Then

$$\text{(4. 18)} \qquad \mathbf{P}_n(x)=\frac{1}{2^n}\sum_{\substack{-n<\nu<[x\sqrt{n}]\\ \nu\equiv n \bmod 2}}\left[\binom{n}{\frac{n+\nu}{2}}-\binom{n}{\frac{n-\nu}{2}+[x\sqrt{n}+1]}\right]$$

(except if $x\sqrt{n}$ is an integer) and thus from the Moivre—Laplace theorem we conclude that

$$\text{(4. 19)} \qquad \lim_{n\to\infty} P_n(x)=\sqrt{\frac{2}{\pi}}\int_0^x e^{-\frac{t^2}{2}}\,dt \qquad (x>0).$$

Therefore, as it follows from (4. 6) that

$$\text{(4. 20)} \qquad \overline{\lim_{k\to\infty}}\,\mathbf{P}(\max(\zeta_1,\zeta_2,\ldots,\zeta_k)<x\sqrt{k})\leqq \underline{\lim_{n\to\infty}}\,P_n(x+\varepsilon)$$

and from (4. 7) that

$$\text{(4. 21)} \qquad \underline{\lim_{k\to\infty}}\,\mathbf{P}(\max(\zeta_1,\zeta_2,\ldots,\zeta_k)<x\sqrt{k})\geqq \overline{\lim_{n\to\infty}}\,P_n(x),$$

we have

$$\text{(4. 22)} \qquad \lim_{k\to\infty}\mathbf{P}(\max(\zeta_1,\zeta_2,\ldots,\zeta_k)<x\sqrt{k})=\sqrt{\frac{2}{\pi}}\int_0^x e^{-\frac{t^2}{2}}\,dt \qquad (x>0),$$

and, applying again the relations (4. 6) and (4. 7), we obtain (4. 2).

The basic idea of this proof can be summed up as follows: we have pointed out that in case of a special choice of the variables $\xi_{n,k}$, (4. 2) holds; from this, by (4. 6) and (4. 7), we have concluded that (4. 2) is true also in case $\xi_{n,k}=\eta_n$ where the variables η_n are normally distributed; hence, again by (4. 6) and (4. 7), it followed that (4. 2) holds also for any variables $\xi_{n,k}$ satisfying the conditions of the theorem.

The proof of Theorems 10 and 11 is based on the same idea and, with suitable modifications, agrees step by step with the proof of Theorem 9.

It is sufficient to prove only Theorem 11, because this, as we have seen, includes Theorem 10 as a special case. It is unnecessary to detail the first part of the proof, therefore we shall deal only with the second.

Let us have again

$$\mathbf{P}(\xi_{n,k}=+1)=\mathbf{P}(\xi_{n,k}=-1)=\frac{1}{2} \qquad (k=1,2,\ldots,N_n=n;\ n=1,2,\ldots)$$

and let us suppose that the variables $\xi_{n,k}$ $(k=1,2,\ldots,N_n=n)$ are mutually independent. Then it follows by simple arguments, well-known in the theory of the random walk in the plane, that, if $A=[x\sqrt{n}]+1$ and $B=[y\sqrt{n}]+1$, then

(4. 23)
$$\mathbf{P}(-y\sqrt{n}<S_{n,k}<x\sqrt{n};\ k=1,2,\ldots,n)=$$
$$=\sum_{-B<k<A}\left\{v_k+\sum_{\nu=1}^{\infty}(v_{2\nu(A+B)+k}-v_{2\nu(A+B)-2B-k}+v_{-2\nu(A+B)+k}-v_{-2\nu(A+B)+2A-k})\right\}$$
where
$$v_k=\begin{cases}\binom{n}{\frac{n+k}{2}}\frac{1}{2^n} & \text{if } k\equiv n \pmod 2\\ 0 & \text{in all other cases.}\end{cases}$$
As, by the Moivre—Laplace theorem,
$$\lim_{n\to\infty}\left(\sum_{\frac{n}{2}-b\frac{\sqrt{n}}{2}<k<\frac{n}{2}+a\frac{\sqrt{n}}{2}}\binom{n}{k}\frac{1}{2^n}\right)=\frac{1}{\sqrt{2\pi}}\int_{-b}^{a}e^{-\frac{t^2}{2}}dt,$$
by simple calculation we obtain

(4. 24)
$$\lim_{n\to\infty}\mathbf{P}(-y\sqrt{n}<S_{n,k}<x\sqrt{n};\ k=1,2,\ldots,n)=$$
$$=\frac{4}{\pi}\sum_{k=0}^{\infty}\frac{e^{-\frac{(2k+1)^2\pi^2}{2(x+y)^2}}\sin(2k+1)\pi\frac{x}{x+y}}{2k+1}\qquad(x>0,\ y>0).$$
Similarly to the case of Theorem 9, it follows that the limit of the probability on the left side of (4. 4) is the same also in the general case.

Theorem 12 can be derived from Theorem 11 as follows. Owing to the independence of S_{n,M_n} and $S_{n,k}-S_{n,M_n}$ $(k>M_n)$, by the relation
$$\mathbf{P}(\max_{M_n<k\leq N_n}|S_{n,k}|<yB_n)=\mathbf{P}(-yB_n<(S_{n,M_n}+(S_{n,k}-S_{n,M_n}))<yB_n),$$
we obtain, by virtue of the theorem on total probability, that

(4. 25)
$$\mathbf{P}(\max_{M_n<k\leq N_n}|S_{n,k}|<yB_n)=$$
$$=\int_{-\infty}^{\infty}\mathbf{P}(-(x+y)B_n<S_{n,k}-S_{n,M_n}<(y-x)B_n)d\mathbf{P}\left(\frac{S_{n,M_n}}{B_n}<x\right).$$
As, in accordance with our restrictions, Lindeberg's condition is satisfied by the sums $S_{n,M_n}=\sum_{k=1}^{M_n}\xi_{n,k}$, further, $\mathbf{D}\left(\frac{S_{n,M_n}}{B_n}\right)=\frac{A_n}{B_n}\to\lambda$, as $n\to\infty$, therefore (uniformly in x in all finite intervals) we have

(4. 26)
$$\lim_{n\to\infty}\mathbf{P}\left(\frac{S_{n,M_n}}{B_n}<x\right)=\frac{1}{\sqrt{2\pi}\lambda}\int_{-\infty}^{x}e^{-\frac{u^2}{2\lambda^2}}du=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x}{\lambda}}e^{-\frac{t^2}{2}}dt.$$

Furthermore, by Theorem 11, considering the relation

$$\mathbf{D}(S_{n,N_n}-S_{n,M_n})=\sqrt{B_n^2-A_n^2},$$

it follows that

(4.27) $$\lim_{n\to\infty}\mathbf{P}(-(x+y)B_n<S_{n,k}-S_{n,M_n}<(y-x)B_n;\ k=1,2,\dots,n)=$$

$$=\begin{cases}\dfrac{4}{\pi}\displaystyle\sum_{k=0}^{\infty}\frac{e^{-\frac{(2k+1)^2\pi^2(1-\lambda^2)}{8y^2}}}{2k+1}\sin(2k+1)\pi\frac{y-x}{2y}, & \text{if } y>0 \text{ and } |x|\leqq y,\\ 0 \text{ if } y\leqq 0 \text{ or } y>0 \text{ but } |x|>y.\end{cases}$$

Therefore, finally, we obtain

(4.28) $$\lim_{n\to\infty}\mathbf{P}(\max_{M_n<k\leqq N_n}|S_{n,k}|<yB_n)=$$

$$=\frac{4}{\pi}\sum_{k=0}^{\infty}\frac{e^{-\frac{(2k+1)^2\pi^2(1-\lambda^2)}{8y^2}}}{2k+1}\int_{-y}^{+y}\frac{e^{-\frac{x^2}{2\lambda^2}}}{\sqrt{2\pi}\lambda}\sin(2k+1)\pi\frac{y-x}{2y}\,dx.$$

Hence, by simple calculations, we obtain Theorem 12. In fact,

(4.29) $$\int_{-y}^{+y}\frac{e^{-\frac{x^2}{2\lambda^2}}}{\sqrt{2\pi}\lambda}\sin(2k+1)\pi\frac{y-x}{2y}\,dx=$$

$$=(-1)^k e^{-\frac{(2k+1)^2\pi^2\lambda^2}{8y^2}}\left(\int_{-\frac{y}{\lambda}}^{+\frac{y}{\lambda}}\frac{e^{-\frac{1}{2}\left(t-\frac{(2k+1)i\lambda}{2y}\pi\right)^2}}{\sqrt{2\pi}}\,dt\right).$$

Now, since the integral of $e^{-\frac{t^2}{2}}$ on any closed curve vanishes, therefore, if $a>0$ and b is real, then

$$\int_{-a}^{+a}\frac{e^{-\frac{(t-ib)^2}{2}}}{\sqrt{2\pi}}\,dt=\int_{-a-ib}^{a-ib}\frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\,dt=\int_{-a}^{+a}\frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\,dt+\int_{-a-ib}^{-a}\frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\,dt+\int_{a}^{a-ib}\frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\,dt=$$

$$=\int_{-a}^{a}\frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\,dt+\frac{2e^{-\frac{a^2}{2}}}{\sqrt{2\pi}}\int_0^b e^{\frac{v^2}{2}}\sin av\,dv.$$

Consequently,

$$\int_{-y}^{+y}\frac{e^{-\frac{x^2}{2\lambda^2}}}{\sqrt{2\pi}\lambda}\sin(2k+1)\pi\frac{y-x}{2y}\,dy=$$

$$=(-1)^k e^{-\frac{(2k+1)^2\pi^2\lambda^2}{8y^2}}\left(1-\frac{2}{\sqrt{2\pi}}\int_{\frac{y}{\lambda}}^{\infty}e^{-\frac{t^2}{2}}\,dt+\frac{2e^{-\frac{y^2}{2\lambda^2}}\lambda}{\sqrt{2\pi}y}\int_0^{(2k+1)\frac{\pi}{2}}e^{-\frac{\lambda^2v^2}{2y^2}}\sin v\,dv\right).$$

This completes the proof of Theorem 12.

§ 5. Proof of the tests analogous to those of Kolmogorov and Smirnov

Let ξ be a random variable having the continuous distribution function $F(x)$ and let $\xi_1, \xi_2, \ldots, \xi_n$ denote the results of n independent observations for the value of ξ, i. e. let $\xi_1, \xi_2, \ldots, \xi_n$ be mutually independent random variables having the same continuous distribution function $F(x)$. Let us denote the distribution function of this sample by $F_n(x)$.

We shall prove the theorems formulated in § 3 by means of the method exposed in § 1, using the theorems of § 4.

Let us put $\eta_k = F(\xi_k)$ and $\zeta_k = \log \frac{1}{\eta_k}$, further, $\eta_k^* = F(\xi_k^*)$ and $\zeta_k^* = R_k(\zeta_1, \zeta_2, \ldots, \zeta_n)$. In this case the variables η_k are uniformly distributed in the interval (0, 1) and their sample distribution function is

$$G_n(x) = F_n(F^{-1}(x))$$

where $y = F^{-1}(x)$ is the inverse function of $x = F(y)$. But it is easily seen that

$$\sup_{a \leqq F(x)} \left(\frac{F_n(x) - F(x)}{F(x)} \right) = \sup_{a \leqq x \leqq 1} \frac{G_n(x) - x}{x}, \tag{5. 1}$$

therefore, instead of the variable on the left side of (5. 1), we may consider the variable $\sup\limits_{a \leqq x \leqq 1} \frac{G_n(x) - x}{x}$ identical with it. The variables η_k^* — as we have seen — form a Markov chain. Further, we have seen that the variables $\delta_{k+1} = (n-k)(\zeta_{k+1}^* - \zeta_k^*)$ are mutually independent and exponentially distributed with mean 1, i. e.

$$\mathbf{P}(\delta_k < x) = 1 - e^{-x} \qquad (x > 0).$$

We have also seen how the variables $\zeta_k^* = \log \frac{1}{\eta_{k+1-k}^*}$ may be decomposed into sums of mutually independent random variables by means of the δ_j's:

$$\zeta_k^* = \sum_{j=1}^{k} \frac{\delta_j}{n+1-j}.$$

Let us turn to the proof of Theorem 8. First of all, it is easy to see that instead of the relation (3. 4) it is enough to prove

$$\lim_{n \to \infty} \mathbf{P}\left(\sqrt{n} \sup_{a \leqq G_n(x)} \frac{G_n(x) - x}{x} < y \right) = \sqrt{\frac{2}{\pi}} \int_0^{y\sqrt{\frac{a}{1-a}}} e^{-\frac{t^2}{2}} dt \qquad (y > 0). \tag{5. 2}$$

For, if $|G_n(x) - x| \leqq \varepsilon$, then from $G_n(x) \geqq a + \varepsilon$ it follows that $x \geqq G_n(x) - \varepsilon \geqq a$ and thus

$$\sup_{a \leqq x} \frac{G_n(x) - x}{x} \geqq \sup_{G_n(x) \geqq a+\varepsilon} \frac{G_n(x) - x}{x}$$

i. e. from $\sup_{a \leqq x} \frac{G_n(x)-x}{x} < \frac{y}{\sqrt{n}}$ it follows that $\sup_{G_n(x) \geqq a+\varepsilon} \frac{G_n(x)-x}{x} < \frac{y}{\sqrt{n}}$. But if A, A' and B are any events and $AB \subset A'B$, then

$$\mathbf{P}(A) = \mathbf{P}(A\bar{B}) + \mathbf{P}(AB) \leqq \mathbf{P}(\bar{B}) + \mathbf{P}(A'B) \leqq \mathbf{P}(\bar{B}) + \mathbf{P}(A').$$

Applying this inequality to the case when A is the event $\sup_{a \leqq x} \frac{G_n(x)-x}{x} < \frac{y}{\sqrt{n}}$, B is the event $|G_n(x)-x| \leqq \varepsilon$, and, finally, A' denotes the event $\sup_{G_n(x) \geqq a+\varepsilon} \frac{G_n(x)-x}{x} < \frac{y}{\sqrt{n}}$, we obtain

$$\mathbf{P}\left(\sqrt{n} \sup_{a \leqq x} \frac{G_n(x)-x}{x} < y\right) \leqq \mathbf{P}(|G_n(x)-x| > \varepsilon) + \mathbf{P}\left(\sqrt{n} \sup_{a+\varepsilon \leqq G_n(x)} \frac{G_n(x)-x}{x} < y\right).$$

It can be similarly shown that

$$\mathbf{P}\left(\sqrt{n} \sup_{a-\varepsilon \leqq G_n(x)} \frac{G_n(x)-x}{x} < y\right) \leqq \mathbf{P}(|G_n(x)-x| > \varepsilon) + \mathbf{P}\left(\sqrt{n} \sup_{a \leqq x} \frac{G_n(x)-x}{x} < y\right).$$

As

$$\lim_{n\to\infty} \mathbf{P}(|G_n(x)-x| > \varepsilon) = 0 \qquad (\varepsilon > 0),$$

it follows that if (5. 2) is satisfied then

$$\overline{\lim_{n\to\infty}} \mathbf{P}\left(\sqrt{n} \sup_{a \leqq x} \frac{G_n(x)-x}{x} < y\right) \leqq \sqrt{\frac{2}{\pi}} \int_0^{y\sqrt{\frac{a+\varepsilon}{1-a-\varepsilon}}} e^{-\frac{t^2}{2}} dt$$

and

$$\underline{\lim_{n\to\infty}} \mathbf{P}\left(\sqrt{n} \sup_{a \leqq x} \frac{G_n(x)-x}{x} < y\right) \geqq \sqrt{\frac{2}{\pi}} \int_0^{y\sqrt{\frac{a-\varepsilon}{1-a+\varepsilon}}} e^{-\frac{t^2}{2}} dt.$$

Since ε can be chosen arbitrarily small and the integral is a continuous function of its upper limit, it follows that

$$\lim_{n\to\infty} \mathbf{P}\left(\sqrt{n} \sup_{a \leqq x} \frac{G_n(x)-x}{x} < y\right) = \sqrt{\frac{2}{\pi}} \int_0^{y\sqrt{\frac{a}{1-a}}} e^{-\frac{t^2}{2}} dt.$$

Therefore, (3. 4) actually follows from (5. 2) and thus, to prove Theorem 8, it is enough to show that (5. 2) holds.

Further, we shall need the following relation:

$$(5.3) \qquad \sqrt{n} \sup_{a \leqq G_n(x)} \frac{G_n(x)-x}{x} = \sqrt{n} \max_{an \leqq k \leqq n} \left(\frac{\frac{k}{n}}{\eta_k^*} - 1\right).$$

This follows from the fact that $G_n(x)$ is a constant lying between η_k^* and

η_{k+1}^*, and so in any interval $\eta_k^* < x < \eta_{k+1}^*$ the supremum of $\dfrac{G_n(x)-x}{x} = \dfrac{G_n(x)}{x} - 1$ is equal to

$$\frac{G_n(\eta_k^*+0)}{\eta_k^*} - 1 = \frac{\frac{k}{n}}{\eta_k^*} - 1.$$

Now, let us apply Theorem 9 to the sequence consisting of the following sequence of random variables:

$$\frac{\delta_j - 1}{n+1-j} \qquad (j = 1, 2, \ldots, [n(1-a)]+1);$$

this sequence satisfies LINDEBERG's condition (and, moreover, even LIAPUNOV's condition) if $0 < a < 1$. Then, by (1.11), for any $z > 0$, we obtain

$$\text{(5.4)} \quad \lim_{n\to\infty} \mathbf{P}\left(\max_{an \le k \le n}\left(\log\frac{1}{\eta_k^*} - \sum_{\nu=k}^{n}\frac{1}{\nu}\right) < z\sqrt{\sum_{an \le k = n}\frac{1}{k^2}}\right) = \sqrt{\frac{2}{\pi}}\int_0^z e^{-\frac{t^2}{2}}\,dt.$$

As in case of $k \ge an$ and $0 < a < 1$,

$$\sum_{\nu=k}^{n}\frac{1}{\nu} = \log\frac{n}{k} + O\left(\frac{1}{n}\right) \quad \text{and} \quad \sqrt{\sum_{an \le k \le n}\frac{1}{k^2}} = \sqrt{\frac{1-a}{an}} + O\left(\frac{1}{n}\right),$$

from (5.4) one concludes

$$\text{(5.5)} \qquad \lim_{n\to\infty} \mathbf{P}\left(\max_{an \le k \le n}\log\frac{\frac{k}{n}}{\eta_k^*} < z\sqrt{\frac{1-a}{an}}\right) = \sqrt{\frac{2}{\pi}}\int_0^z e^{-\frac{t^2}{2}}\,dt;$$

therefore, finally, introducing the notation $y = z\sqrt{\dfrac{1-a}{a}}$, we have

$$\text{(5.6)} \qquad \lim_{n\to\infty} \mathbf{P}\left(\sqrt{n}\max_{an \le k \le n}\log\left(\frac{\frac{k}{n}}{\eta_k^*} - 1\right) < y\right) = \sqrt{\frac{2}{\pi}}\int_0^{y\sqrt{\frac{a}{1-a}}} e^{-\frac{t^2}{2}}\,dt.$$

In view of (5.2) and (5.3), Theorem 5 follows from (5.6).

Let us now turn to the proof of Theorem 7. We may obtain the random variable

$$\text{(5.7)} \qquad \tau = \sqrt{n}\max_{an \le k \le bn}\left(\log\frac{1}{\eta_k^*} - \sum_{\nu=k}^{n}\frac{1}{\nu}\right) = \sqrt{n}\max_{an \le k \le bn}\sum_{j=1}^{n+1-k}\frac{\delta_j - 1}{n+1-j}$$

as the sum of to independent random variables τ_1 and τ_2 where

$$\text{(5.8)} \qquad \tau_1 = \sqrt{n}\sum_{1 \le j \le n+1-bn}\frac{\delta_j - 1}{n+1-j}$$

and

$$\text{(5.9)} \qquad \tau_2 = \sqrt{n}\max_{an \le n+1-k \le bn}\sum_{j=1}^{k}\frac{\delta_j - 1}{n+1-j}.$$

It is evident that in the limit τ_1 is a normally distributed random variable with the standard deviation $\sqrt{\frac{1-b}{b}}$; further, from the proof of Theorem 5 we can see that

$$\lim_{n\to\infty} \mathbf{P}(\tau_2\sqrt{b}<z) = \sqrt{\frac{2}{\pi}} \int_0^{z\sqrt{\frac{a}{b-a}}} e^{-\frac{t^2}{2}}\,dt \qquad (z>0). \tag{5.10}$$

Considering further that τ_1 and τ_2 are independent, it follows from (5.8) and (5.10) that

$$\lim_{n\to\infty} \mathbf{P}(\tau<y) = \frac{1}{\pi}\sqrt{\frac{b}{1-b}} \int_{-\infty}^{y} e^{-\frac{bu^2}{2(1-b)}} \int_0^{(y-u)\sqrt{\frac{ab}{b-a}}} e^{-\frac{t^2}{2}}\,dt\,du. \tag{5.11}$$

This completes the proof of Theorem 7.

In the same way we can also prove Theorem 6. Here the relation (5.3) is replaced by

$$\sqrt{n} \sup_{a\leqq G_n(x)} \left|\frac{G_n(x)-x}{x}\right| = \sqrt{n} \max_{an\leqq k\leqq n} \left(\left|\frac{\frac{k}{n}}{\eta_k^*}-1\right|, \left|\frac{\frac{k}{n}}{\eta_{k-1}^*}-1\right| \right) \tag{5.12}$$

from which it follows

$$\begin{aligned} \sqrt{n} \max_{an\leqq k\leqq n} \left|\frac{\frac{k}{n}}{\eta_k^*}-1\right| &\leqq \sqrt{n} \sup_{a\leqq G_n(x)} \left|\frac{G_n(x)-x}{x}\right| \leqq \\ &\leqq \sqrt{n} \max_{an\leqq k\leqq n} \left|\frac{\frac{k}{n}}{\eta_k^*}-1\right| + \frac{1}{a\sqrt{n}}. \end{aligned} \tag{5.13}$$

For, if $\eta_{k+1}^* < \frac{k}{n}$, then $\left|\frac{\frac{k}{n}}{\eta_{k+1}^*}-1\right| = \frac{\frac{k}{n}}{\eta_{k+1}^*}-1 < \frac{\frac{k+1}{n}}{\eta_{k+1}^*}-1 = \left|\frac{\frac{k+1}{n}}{\eta_{k+1}^*}-1\right|$; if, however, $\eta_{k+1}^* \geqq \frac{k}{n}$, then in case $k\geqq an$, $\eta_{k+1}^*\geqq a$ and thus

$$\left|\frac{\frac{k}{n}}{\eta_{k+1}^*}-1\right| = 1-\frac{\frac{k}{n}}{\eta_{k+1}^*} \leqq 1-\frac{\frac{k+1}{n}}{\eta_{k+1}^*}+\frac{1}{n\eta_{k+1}^*} \leqq \left|\frac{\frac{k+1}{n}}{\eta_{k+1}^*}-1\right| + \frac{1}{an}.$$

Consequently, in either case

$$\left|\frac{\frac{k}{n}}{\eta_{k+1}^*}-1\right| \leqq \left|\frac{\frac{k+1}{n}}{\eta_{k+1}^*}-1\right| + \frac{1}{an} \qquad (k\geqq an)$$

and since $\max\limits_{an \leq k \leq n} \left| \dfrac{\frac{k+1}{n}}{\eta^*_{k+1}} - 1 \right| \leqq \max\limits_{an \leq k \leq n} \left| \dfrac{\frac{k}{n}}{\eta^*_k} - 1 \right|$, we get

$$\max_{an \leq k \leq n} \left(\left| \frac{\frac{k}{n}}{\eta^*_k} - 1 \right|, \left| \frac{\frac{k}{n}}{\eta^*_{k+1}} - 1 \right| \right) \leqq \max_{an \leq k \leq n} \left| \frac{\frac{k}{n}}{\eta^*_k} - 1 \right| + \frac{1}{an}.$$

The limiting distribution of the variable

$$\sqrt{n} \max_{an \leq k \leq n} \left| \frac{\frac{k}{n}}{\eta^*_k} - 1 \right|$$

occurring here is identical with that of the variable

$$\sqrt{n} \max_{an \leq k \leq n} \left| \log \frac{1}{\eta^*_k} \sum_{\nu=k}^{n} \frac{1}{\nu} \right| = \sqrt{n} \max_{an \leq k \leq n} \left| \sum_{j=1}^{n+1-k} \frac{\delta_j - 1}{n+1-j} \right|,$$

which can be determined by means of Theorem 10.

To prove Theorem 8, those steps have to be applied simultaneously which have been used in the proof of Theorems 6 and 7; in this proof we shall use Theorem 12 instead of Theorem 10.

The basic idea of the proof is as follows. The limiting distribution of the variable

$$\sqrt{n} \sup_{a \leq F(x) \leq b} \left| \frac{F_n(x) - F(x)}{F(x)} \right|$$

is identical with that of the variable

$$\varkappa_1 = \sqrt{n} \max_{an \leq k \leq bn} \left| \frac{\frac{k}{n}}{\eta^*_k} - 1 \right|,$$

and therefore it is identical also with that of the variable

$$\varkappa_2 = \sqrt{n} \max_{an \leq k \leq bn} \left| \log \frac{\frac{k}{n}}{\eta^*_k} \right| = \sqrt{n} \max_{an \leq k \leq bn} \left| \sum_{j=1}^{n+1-k} \frac{\delta_j - 1}{n+1-j} \right|.$$

Thus Theorem 12 is applicable, namely, since the values of the constants A_n and B_n occurring in it, are as follows:

$$A_n = \sqrt{\frac{1-b}{bn}} + o\left(\frac{1}{n}\right) \quad \text{and} \quad B_n = \sqrt{\frac{1-a}{an}} + o\left(\frac{1}{n}\right),$$

therefore

$$\lambda = \lim_{n \to \infty} \frac{A_n}{B_n} = \sqrt{\frac{a(1-b)}{b(1-a)}},$$

and thus introducing the notation $S_{n,r}=\sum_{j=1}^{r}\frac{\delta_j-1}{n+1-j}$, we have

$$\lim_{n\to\infty}\mathbf{P}\left(\sqrt{n}\sup_{a\leq F(x)\leq b}\left|\frac{F_n(x)-F(x)}{F(x)}\right|<y\right)=$$

$$=\lim_{n\to\infty}\mathbf{P}\left(\max_{n+1-bn\leq r\leq n+1-an}|S_{n,r}|<y\sqrt{\frac{a}{1-a}B_n}\right)=$$

$$=\frac{4}{\pi}\sum_{k=0}^{\infty}(-1)^k\frac{e^{-\frac{(2k+1)^2\pi^2}{8}\frac{1-a}{ay^2}}}{2k+1}\left(1-\frac{2}{\sqrt{2\pi}}\int_{y\sqrt{\frac{b}{1-b}}}^{\infty}e^{-\frac{u^2}{2}}du+\varrho_k\right)$$

where

$$\varrho_k=\frac{2\sqrt{\frac{1-b}{b}}e^{-\frac{by^2}{2(1-b)}}}{\sqrt{2\pi}y}\int_0^{(2k+1)\frac{\pi}{2}}e^{\frac{u^2(1-b)}{2by^2}}\sin u\,du.$$

This completes the proof of Theorem 8.

§. 6. Remarks on the limiting distribution functions occurring in Theorems 5—8

The values of the limiting distribution function occurring in Theorem 5 may be read from the tables of the normal distribution function. The values of the limiting distribution function occurring in Theorem 6 can be computed by substituting the values $z=y\sqrt{\frac{a}{1-a}}$ into the function

$$L(z)=\frac{4}{\pi}\sum_{k=0}^{\infty}(-1)^k\frac{\exp\left(-\frac{(2k+1)^2\pi^2}{8z^2}\right)}{2k+1}\qquad(z>0). \tag{6.1}$$

At the end of this paper, we give the table of the function $L\left(y\sqrt{\frac{a}{1-a}}\right)$ for certain values of a. The curve of the distribution function $L\left(y\sqrt{\frac{a}{1-a}}\right)$ can for certain values of a be seen on Fig. 1.

The values of the limiting distribution function occurring in Theorem 7 can be approximatively computed in the following manner:

$$F(y,a,b)=\frac{1}{\pi}\int_{-\infty}^{y\sqrt{\frac{b}{1-b}}}e^{-\frac{u^2}{2}}\int_0^{\left(y-u\sqrt{\frac{1-b}{b}}\right)\sqrt{\frac{ab}{b-a}}}e^{-\frac{v^2}{2}}dv\,du=\frac{1}{\pi}\iint_T e^{-\frac{u^2+v^2}{2}}du\,dv \tag{6.2}$$

where T is an infinite triangular domain in the plane (u, v) defined by the following inequalites:

(6. 3) $$T: \left\{-\infty < u < y\sqrt{\frac{b}{1-b}};\ 0 \leqq v \leqq \sqrt{\frac{a(1-b)}{b-a}}\left(y\sqrt{\frac{b}{1-b}}-u\right)\right\}.$$

Introducing the polar coordinates $r=\sqrt{u^2+v^2}$, $\varphi=\operatorname{arc\,tg}\frac{v}{u}$ we obtain

(6. 4) $$F(y, a, b)=\frac{1}{\pi}\int_0^{\pi-\alpha}\left(1-e^{-\frac{ay^2}{2(1-a)\sin^2(\varphi+\alpha)}}\right)d\varphi$$

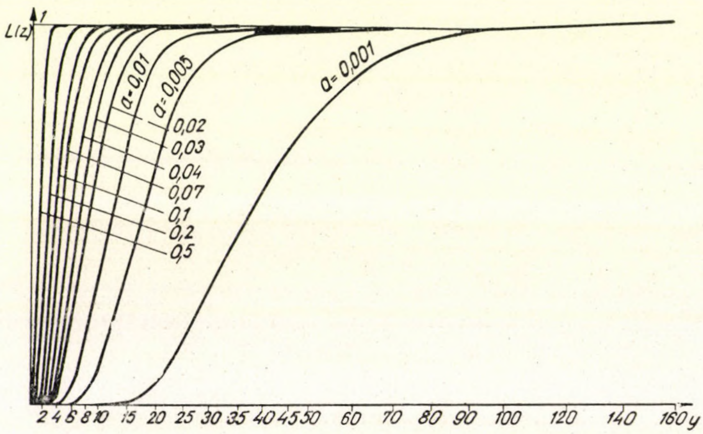

Fig. 1.

where $\operatorname{tg}\alpha=\sqrt{\frac{a(1-b)}{b-a}}$ and $0<\alpha<\frac{\pi}{2}$; therefore

(6. 5) $$F(y, a, b)=\frac{1}{\pi}\int_\alpha^{\pi}\left(1-e^{-\frac{ay^2}{2(1-a)\sin^2\beta}}\right)d\beta.$$

As

(6. 6) $$\frac{1}{\pi}\int_0^{\pi}\left[1-\exp\left(-\frac{ay^2}{2(1-a)\sin^2\beta}\right)\right]d\beta=\sqrt{\frac{2}{\pi}}\int_0^{y\sqrt{\frac{a}{1-a}}}e^{-\frac{u^2}{2}}du,$$

we have finally the following approximative expression:

(6. 7) $$F(y, a, b)=\sqrt{\frac{2}{\pi}}\int_0^{y\sqrt{\frac{a}{1-a}}}e^{-\frac{u^2}{2}}du-\frac{\operatorname{arc\,tg}\sqrt{\frac{a(1-b)}{b-a}}}{\pi}(1-R)$$

where

(6. 8) $$R=\frac{1}{\alpha}\int_0^{\alpha}\exp\left(\frac{ay^2}{2(1-a)\sin^2\beta}\right)d\beta \quad \text{and} \quad \alpha=\operatorname{arc\,tg}\sqrt{\frac{a(1-b)}{b-a}}.$$

If $1-b=\varepsilon$ is small, then R is, in most cases, negligible, except for extremely small values of y, since

(6. 9) $$R\leqq \exp\left(-\frac{ay^2}{2(1-a)\sin^2\alpha}\right)=\exp\left(-\frac{by^2}{2(1-b)}\right).$$

The values of the limiting distribution function occurring in Theorem 8 can be approximatively computed in the following manner: it follows from the second mean value theorem of the integral calculus that

(6. 10) $$|\varrho_k|=\frac{2e^{-\frac{by^2}{2(1-b)}}\sqrt{\frac{1-b}{b}}}{\sqrt{2\pi y}}\left|\int_0^{(2k+1)\frac{\pi}{2}} e^{-\frac{(1-b)u^2}{2by^2}}\sin u\, du\right|\leqq$$

$$\leqq\frac{2e^{-\frac{by^2}{2(1-b)}}\sqrt{\frac{1-b}{b}}}{\sqrt{2\pi y}}\exp\left(\frac{(2k+1)^2\pi^2(1-b)}{8by^2}\right).$$

In this way, using the notation (6. 1) the limiting distribution function occurring in Theorem 8 can be expressed in the following form

(6. 11) $$\frac{4}{\pi}L\left(y\sqrt{\frac{a}{1-a}}\right)\left(1-\frac{2}{\sqrt{2\pi}}\int_{y\sqrt{\frac{b}{1-b}}} e^{-\frac{u^2}{2}}du\right)+\Delta$$

where, as is seen by way of a simple calculation.

(6. 12) $$\Delta<\frac{2\lambda e^{-\frac{by^2}{2(1-b)}}}{\pi\sqrt{2\pi y}}\log\frac{1+\exp\left(-\frac{\pi^2(b-a)}{8aby^2}\right)}{1-\exp\left(-\frac{\pi^2(b-a)}{8aby^2}\right)},$$

whence it is readily seen that if b is very near to 1, Δ is negligible. Observe that the first factor of the main term depends only on a, the second only on b; this fact simplifies the computation to a great extent; namely, because we can obtain the first factor from the table of $L(z)$, the second from the table of the normal distribution function.

INSTITUTE FOR APPLIED MATHEMATICS
OF THE HUNGARIAN ACADEMY OF SCIENCES.

(Received 30 October 1953)

Bibliography

[1] K. Pearson, Note on Francis Galton's problem, *Biometrika*, **1** (1902), pp. 390—399.

[2] L. v. Bortkiewicz, Variationsbreite und mittlere Fehler, *Sitzungsber. Berl. Math. Ges.*, **21** (1922), pp. 3—11.

[3] E. L. Dood, The greatest and the least variate under general laws of error, *Trans. Amer. Math. Soc.*, **25** (1923), pp. 525—539.

[4] L. H. C. Tippet, On the extreme individuals and the range of samples taken from a normal population, *Biometrika*, **17** (1925), pp. 264—387.

[5] M. Fréchet, Sur la loi de probabilité de l'écart maximum, *Ann. Soc. Polon. Math.*, **6** (1928), pp. 92—116.

[6] a) A. N. Kolmogorov, Sulla determinazione empirico di una legge di distribuzione, *Giornale dell'Istituto Italiano d. Attuari*, **4** (1933), pp. 83—91.

b) A. N. Kolmogorov, Confidence limits for an unknown distribution function, *Ann. Math. Stat.*, **12** (1941), pp. 461—463.

[7] V. I. Glivenko, Sulla determinazione empirica delle leggi di probabilità, *Giorn. Ist. Ital. Attuari*, **4** (1933), pp. 92—99.

[8] a) N. V. Smirnov, Über die Verteilung des allgemeinen Gliedes in der Variationsreihe, *Metron*, **12** (1935), pp. 59—81.

b) Н. В. Смирнов, Sur la distribution de ω^2, *Comptes Rendus Ac. Sci., Paris*, **202** (1936), pp. 449—452.

c) Н. В. Смирнов, О распределении ω^2-критерия Мизеса, Матем. Сборник, **2** (44) (1937), pp. 973—994.

d) Н. В. Смирнов, Об уклонениях эмпирической кривой распределения, Матем. Сборник, **6** (44) (1939), pp. 3—26.

e) Н. В. Смирнов, Оценка расхождения между эмпирическими кривыми распределения в двух независимых выборках, Бюлл. Моск. Унив., **2** (1939), pp. 3—14.

f) Н. В. Смирнов, Приближение законов распределения случайных величин по эмпирическим данным, Успехи Мат. Наук, **10** (1944), pp. 176—206.

g) Н. В. Смирнов, Предельные законы распределения для членов вариационного ряда, Труды Мат. Инст. Стеклова, **25** (1949), pp. 1—60.

[9] B. V. Gnedenko, Limit theorems for sums of independent random variables, *Transactions Amer. Math. Soc.*, **45** (1951), pp.

[10] Б. В. Гнеденко и В. С. Королюк, О максимальном расхождение двух эмпирических распределений, Докл. Акад. Наук СССР, **80** (1951), pp. 525—528.

[11] Б. В. Гнеденко и Е. Л. Рвачева, Об одной задаче сравнения двух эмпирических распределений, Докл. Акад. Наук СССР, **82** (1952), pp. 513—516.

[12] Б. В. Гнеденко и В. С. Михалевич, Две теоремы о поведений эмпирических функций распределения, Докл. Акад. Наук СССР, **82** (1952), pp. 841—843 and pp. 25—27.

[13] В. С. Михалевич, О взаимном расположений двух эмпирических функций распределения, Докл. Акад. Наук СССР, **85** (1952), pp. 485—488.

[14] И. Д. Квит, О теореме Н. В. Смирнова относительно сравнения двух выборок, Докл. Акад. Наук СССР, **71** (1950), pp. 229—231.

[15] Г. М. Мания, Обобщение критерия А. Н. Колмогорова для оценки закона распределения по эмпирическим данным, Докл. Акад. Наук СССР, **69** (1949), pp. 495—497.

[16] И. И. Гихман, Об эмпирической функций распределения в случае группировки данных, Докл. Акад. Наук СССР, **82** (1952), pp. 837—840.

[17] W. Feller, On the Kolmogorov—Smirnov limit theorems for empirical distributions, *Ann. Math. Statistics*, **19** (1948), pp. 177—180.

[18] J. L. Doob, Heuristic approach to the Kolmogorov—Smirnov theorems, *Ann. Math. Statistics*, **20** (1949), pp. 393—403.

[19] a) F. J. Massey, A note on the estimation of a distribution function by confidence limits, *Ann. Math. Statistics*, **21** (1950), pp. 116—119.

b) F. J. Massey, A note on the power of a non-parametric test, *Ann. Math. Statistics*, **21** (1950), pp. 440—443.

c) F. J. Massey, Distribution table for the deviation between two sample cumulatives. *Ann. Math. Statistics*, **23** (1952), pp. 435—441.

[20] a) M. D. Donsker, Justification and extension of Doob's heuristic approach to the Kolmogorov—Smirnov theorems, *Ann. Math. Statistics*, **23** (1952), pp. 277—281,

b) M. D. Donsker, An invariance principle for certain probability limit theorems, *Memoirs Amer. Math. Soc.*, **6** (1951), pp. 1—12.

[21] T. W. Anderson—D. A. Darling, Asymptotic theory of certain „goodness of fit" criteria based on stochastic processes, *Ann. Math. Statistics*, **23** (1952), pp. 193—212

[22] P. Erdős—M. Kac, On certain limit theorems of the theory of probability, *Bull. Amer Math. Soc.*, **52** (1946), pp. 292—302.

[23] S. Malmquist, On a property of order statistics from a rectangular distribution, *Skand. Aktuerietidsskrift*, **33** (1950), pp. 214—222.

[24] G. Hajós and A. Rényi, Elementary proofs of some basic facts concerning order statistics, *Acta Math. Acad. Sci. Hung.*, **5** (1954) (under press).

[25] H. Cramér, *Mathematical methods of statistics* (Princeton, 1946).

[26] Л. И. Брагински, Оперативный статистический контроль качества в машиностроении (Москва, 1951), Машгиз.

[27] A. Wald, Limit distribution of the maximum and minimum of successive cumulative sums of random variables, *Bull. Amer. Math. Soc.*, **53** (1947), pp. 142—153.

[28] A. Wald, On the distribution of the maximum of successive cumulative sums of independently, but not indentically distributed chance variables, *Bull. Amer. Math. Soc.*, **54** (1948), pp. 422—430.

[29] K. L. Chung, Asymptotic distribution of the maximum cumulative sum of independent random variables, *Bull. Amer. Math. Soc.*, **54** (1948), pp. 1162—1170.

[30] S. S. Wilks, Order statistics, *Bull. Amer. Math. Soc.*, **54** (1948), pp. 6—50.

y \ a	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1	0,2	0,3	0,4	0,5
0,1													0,0000	0,0000
0,5											0,0000	0,0000	0,0008	0,0092
1,0						0,0000	0,0000	0,0000	0,0000	0,0000	0,0092	0,0716	0,2001	0,3708
1,5			0,0000	0,0000	0,0000	0,0002	0,0009	0,0023	0,0050	0,0092	0,1420	0,3543	0,5591	0,7328
2,0		0,0000	0,0001	0,0008	0,0036	0,0101	0,0212	0,0367	0,0563	0,0791	0,3708	0,6193	0,7951	0,9082
2,5		0,0001	0,0022	0,0112	0,0299	0,0578	0,0925	0,1320	0,1730	0,2155	0,5778	0,7966	0,9714	0,9751
3,0	0,0000	0,0015	0,0157	0,0474	0,0941	0,1487	0,2061	0,2632	0.3184	0,3708	0,7328	0,9009	0,9915	0,9954
3,5	0,0001	0,0092	0,0491	0,1135	0,1879	0,2629	0,3341	0,3994	0,4598	0,5140	0,8398	0,9561	0,9978	0,9991
4,0	0,0006	0,0291	0,1052	0,2001	0,2942	0,3804	0,4570	0,5244	0,5835	0,6353	0,9082	0,9823	0,9995	0,9999
4,5	0,0031	0,0643	0,1776	0,2950	0,4001	0,4902	0,5665	0,6311	0,6860	0,7328	0,9511	0,9936	0,9999	1,0000
5,0	0,0096	0,1135	0,2582	0,3895	0,4985	0,5873	0,6594	0,7193	0,7683	0,8088	0,9751	0,9979	1,0000	
5,5	0,0225	0,1726	0,3511	0,4784	0,5863	0,6723	0.7374	0,7903	0,8326	0,8665	0,9887	0,9994		
6,0	0,0428	0,2375	0,4204	0,5591	0,6627	0,7409	0,8006	0,8463	0,8817	0,9081	0,9954	0,9999		
6,5	0,0707	0,3045	0,4952	0,6310	0,7282	0,7989	0,8509	0,8895	0,9181	0,9395	0,9977	1,0000		
7,0	0,1053	0,3708	0,5639	0,6939	0,7834	0,8461	0,8904	0,9220	0,9446	0,9607	0,9991			
7,5	0,1452	0,4347	0,6193	0,7484	0,8294	0,8839	0,9207	0,9460	0,9633	0,9752	0,9996			
8,0	0,1889	0,4959	0,6811	0,7951	0,8671	0,9135	0,9436	0,9634	0,9763	0,9847	0,9999			
8,5	0,2348	0,5513	0,7301	0,8345	0,8977	0,9365	0,9606	0,9756	0,9850	0,9908	1,0000			
9,0	0,2819	0,6032	0,7731	0,8696	0,9221	0,9540	0,9729	0,9849	0,9907	0,9946				
9,5	0,3290	0,6510	0,8104	0,8950	0,9410	0,9713	0,9817	0,9898	0,9944	0,9969				
10,0	0,3754	0,6938	0,8427	0,9175	0,9564	0,9770	0,9878	0,9936	0,9967	0,9983				
10,5	0,4205	0,7328	0,8704	0,9358	0,9680	0,9840	0,9921	0,9961	0,9981	0,9991				
11,0	0,4640	0,7678	0,8939	0,9505	0,9768	0,9891	0,9949	0,9976	0,9989	0,9995				
11,5	0,5055	0,7992	0,9137	0,9622	0,9833	0,9927	0,9968	0,9986	0,9994	0,9997				
12,0	0,5450	0,8271	0,9303	0,9713	0,9882	0,9951	0,9980	0,9992	0,9997	0,9999				
12,5	0,5824	0,8517	0,9441	0,9784	0,9917	0,9968	0,9988	0,9995	0,9998	0,9999				
13,0	0,6174	0,8734	0,9555	0,9841	0,9943	0,9980	0,9993	0,9997	0,9999	1,0000				
13,5	0,6509	0,8924	0,9648	0,9883	0,9961	0,9987	0,9996	0,9999	1,0000					
14,0	0,6812	0,9090	0,9724	0.9915	0,9973	0,9992	0,9998	0,9999						
14,5	0,7099	0,9234	0,9780	0.9938	0,9982	0,9995	0,9999	1,0000						
15,0	0,7367	0,9358	0,9833	0,9956	0,9988	0,9997	0,9999							
15,5	0,7615	0,9464	0,9872	0,9969	0,9992	0,9998	1,0000							
16,0	0,7844	0,9555	0,9902	0,9978	0,9995	0,9999								
16,5	0,8055	0,9631	0,9927	0,9985	0,9997	0,9999								
17,0	0,8249	0,9697	0,9944	0,9990	0,9998	1,0000								

17,5	0,8428	0,9752	0,9958	0,9993	0,9999								
18,0	0,8591	0,9797	0,9969	0,9995	0,9999								
18,5	0,8740	0,9836	0,9977	0,9997	1,0000								
19,0	0,8876	0,9867	0,9983	0,9998									
19,5	0,9000	0,9893	0,9988	0,9999									
20,0	0,9112	0,9915	0,9991	0,9999									
20,5	0,9213	0,9932	0,9994	0,9999									
21,0	0,9304	0,9946	0,9996	1,0000									
21,5	0,9386	0,9957	0,9997										
22,0	0,9460	0,9967	0,9998										
22,5	0,9526	0,9974	0,9998										
23,0	0,9590	0,9980	0,9999										
23,5	0,9636	0,9984	0,9999										
24,0	0,9696	0,9988	1,0000										
24,5	0,9724	0,9991											
25,0	0,9760	0,9993											
26	0,9821	0,9996											
27	0,9867	0,9998											
28	0,9902	0,9999											
29	0,9929	0,9999											
30	0,9949	1,0000											
35	0,9991												
40	0,9999												
43	1,0000												

Values of the function $L\left(y\sqrt{\frac{a}{1-a}}\right)$.

К ТЕОРИИ ВАРИАЦИОННЫХ РЯДОВ

А. РЕНЬИ (Будапешт)

(Резюме)

Цель настоящей статьи — изложение нового метода, с помощью которого можно простым и систематическим образом построить теорию вариационных рядов и доказать ряд новых теорем. Сущность метода состоит в том, что исследование предельных распределений величин, зависящих от членов вариационного ряда сводится к исследовании распределении функций от сумм независимых случайных величин. Этот метод исходит из факта, который первым заметил А. Н. Колмогоров в своей работе [6a], что члены вариационного ряда образуют цепь Маркова. Более того, как доказал S. Malmquist, если ξ_k^* $(k=1,2,\ldots,n)$—расположенные в возрастающем порядке члени элементов $\xi_k (k=1,2,\ldots,n)$ выборки объема n из статистической совокупности с непрерывной функцией распределение $F(x)$ и $\eta_k^* = F(\xi_k^*)$ $(k=1,2,\ldots,n)$, то величины $\left(\frac{\eta_k^*}{\eta_{k+1}^*}\right)^k$ $(k=1,2,\ldots,n)$ являются вполне независимыми и в интервале $(0,1)$ равномерно распределенными случайными величинами, и поэтому величини $\zeta_k = \log \eta_k^*$ образуют аддитивную цепь Маркова. Простое доказательство этого факта дано в § 1. § 2 содержит изложение применения этого факта к простого доказательства некоторых известных теорем теорий вариационных рядов. В § 3 сформулированы следующие новые результаты, полученные с помощью нового метода.

Пусть $F_n(x)$ означает эмпирическую функцию распределения выборки, т. е. положим $F_n(x) = \frac{k}{n}$ для $\xi_k^* \leqq x < \xi_{k+1}^*$ $(k=1,2,\ldots,n-1)$, $F_n(x)=0$ для $x<\xi_1^*$ и $F_n(x)=1$ для $\xi_n^* \leqq x$. Тогда имеем

Теорема 5.

$$\lim_{n\to\infty} \mathbf{P}\left(\sqrt{n} \sup_{0<a\leqq F(x)\leqq 1} \left\{\frac{F_n(x)-F(x)}{F(x)}\right\} < y\right) = \begin{cases} \sqrt{\frac{2}{\pi}} \int\limits_0^{y\sqrt{\frac{a}{1-a}}} e^{-\frac{t^2}{2}}\,dt & \text{для } y<0, \\ 0 & \text{для } y \leqq 0. \end{cases}$$

Теорема 6.

$$\lim_{n\to\infty} \mathbf{P}\left(\sqrt{n} \sup_{0<a\leqq F(x)\leqq 1} \left|\frac{F_n(x)-F(x)}{F(x)}\right| < y\right) = \begin{cases} \frac{4}{\pi} \sum\limits_{k=0}^{\infty} (-1)^k \frac{e^{-\frac{(2k+1)^2\pi^2(1-a)}{8y^2 a}}}{2k+1} & \text{для } y>0, \\ 0 & \text{для } y \leqq 0. \end{cases}$$

Теорема 7.

$$\lim_{n\to\infty} \mathbf{P}\left(\sqrt{n} \sup_{0<a\leqq F(x)\leqq b\leqq 1} \left\{\frac{F_n(x)-F(x)}{F(x)}\right\} < y\right) = \frac{1}{\pi} \int\limits_{-\infty}^{y\sqrt{\frac{b}{1-b}}} e^{-\frac{u^2}{2}} \left[\int\limits_0^{\left(\sqrt{\frac{b}{1-b}}y-u\right)\sqrt{\frac{a(1-b)}{b-a}}} e^{-\frac{t^2}{2}}\,dt\right] du$$

$$-\infty < y < +\infty.$$

Теорема 8.

$$\lim_{n\to 0} \mathbf{P}\left(\sqrt{n} \sup_{0<a\leqq F(x)\leqq b<1} \left|\frac{F_n(x)-F(x)}{F(x)}\right| < y\right) = \frac{4}{\pi}\sum_{k=0}^{\infty}(-1)^k \frac{e^{-\frac{(2k+1)^2\pi^2(1-a)}{8y^2a}}}{2k+1} E_k$$

для $y \geqq 0$,

$$\text{где} \quad E_k = 1 - \frac{2}{\sqrt{2\pi}} \int_{y\sqrt{\frac{b}{1-b}}}^{\infty} e^{-\frac{u^2}{2}} du + \frac{2e^{-\frac{by^2}{2(1-b)}}}{\sqrt{2\pi\left(\frac{a}{1-a}\right)}\, y} \int_0^{(2k+1)\frac{\pi}{2}} e^{\frac{(1-b)u^2}{2by^2}} \sin u\, du.$$

Эти теоремы, которые аналогичны известным теоремам Н. В. Смирнова и А. Н. Колмогорова, дают критерии для гипотез относительно $F(x)$, соотв. дают доверительные границы для неизвестной функций $F(x)$.

Доказательство этих теорем содержится в § 5, и опирается, кроме упомянутого метода, на некоторых новых предельных теорем, изложенных в § 4, относительно максимума частных сумм последовательностей независимых случайних величин. § 6 содержит некоторые замечания относительно вычисления предельных функций распределения, фигурирующие в теоремах 5—8. В конце стати дана таблица значений функций

$$L(z) = \frac{4}{\pi}\sum_{k=0}^{\infty}(-1)^k e^{-\frac{(2k+1)^2\pi^2}{8z^2}} \quad \text{где} \quad z = y\sqrt{\frac{a}{1-a}}$$

для различных значений от y и a.