ON THE CENTRAL LIMIT THEOREM FOR SAMPLES FROM A FINITE POPULATION

by

PAUL ERDŐS and ALFRÉD RÉNYI

Let $a_1, a_2, \ldots, a_n$ be arbitrary real numbers. Let us consider all possible $\binom{n}{s}$ sums

$$a_{i_1} + a_{i_2} + \ldots + a_{i_s},\ 1 \leqq i_1 < i_2 < \ldots < i_s \leqq n$$

formed by choosing s arbitrary different elements of the sequence $a_1, a_2, \ldots, a_n$. Let us put

$$M_n = \sum_{k=1}^{n} a_k$$

and

$$D_n = \sqrt{\sum_{k=1}^{n} \left(a_k - \frac{M_n}{n}\right)^2},$$

further

$$D_{n,s} = \sqrt{\frac{s}{n}\left(1 - \frac{s}{n}\right)} \cdot D_n .$$

Let $N_{n,s}(x)$ denote the number of those sums $a_{i_1} + a_{i_2} + \ldots + a_{i_s}$ which do not exceed

$$\frac{sM_n}{n} + x D_{n,s}$$

and put

$$F_{n,s}(x) = \frac{N_{n,s}(x)}{\binom{n}{s}} .$$

We ask about conditions concerning the sequence $\{a_1, \ldots, a_n\}$ and s under which $F_{n,s}(x)$ will be approximately equal to the normal distribution function $\Phi(x)$ defined by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt . \tag{1}$$

In what follows we shall suppose throughout that

$$M_n = 0 .$$

This is no essential restriction, because if $M_n \neq 0$ we may consider instead of the sequence $\{a_1, \ldots, a_n\}$ the sequence $\{a'_1, \ldots, a'_n\}$ where $a'_k = a_k - M_n/n$, and for the new sequence a'_k we have evidently

$$M'_n = \sum_{k=1}^{n} a'_k = 0 .$$

We may also suppose $s \leq n/2$, because evidently if $M_n = 0$,

$$\sum_{i=1}^{s} a_{i_k} = - \sum_{l=1}^{n-s} a_{j_l}$$

where $j_1, j_2, \ldots, j_{n-s}$ are those among the numbers $1, 2, \ldots, n$ which are not contained in the sequence $i_1, i_2, \ldots, i_s$ and therefore

$$F_{n,n-s}(x) = 1 - F_{n,s}(-x) .$$

The problem can be interpreted statistically as follows: We choose a sample $\{a_{i_1}, a_{i_2}, \ldots, a_{i_s}\}$ of size s from the finite population $\{a_1, a_2, \ldots, a_n\}$ and ask under which conditions will the mean

$$\frac{1}{s} \sum_{j=1}^{s} a_{i_j}$$

of the sample be distributed asymptotically normally about the mean $\frac{M_n}{n}$ of the population. We shall prove that $F_{n,s}(x)$ tends to $\Phi(x)$, e. g. when $n \to +\infty$ and $s = s_n \to +\infty$ in such a manner that putting

$$L_n = \frac{\left(\frac{1}{n} \sum_{k=1}^{n} |a_k|^3\right)^2}{\left(\frac{1}{n} \sum_{k=1}^{n} a_k^2\right)^3}$$

we have

$$\lim_{n \to +\infty} \frac{L_n}{s_n} = 0 ;$$

or more generally, if putting

$$d_{n,s}(\varepsilon) = \frac{1}{D_n^2} \sum_{\substack{1 \leq k \leq n \\ |a_k| > \varepsilon D_{n s}}} a_k^2$$

we have for any $\varepsilon > 0$

$$\lim_{n \to \infty} d_{n,s_n}(\varepsilon) = 0 .$$

Clearly our question asks about the distribution of *a sum of weakly dependent random variables*. As a matter of fact if we imagine the elements of our sample as chosen successively without replacement and ξ_k is the element of the sequence which has been chosen at the k-th choice, then we have to investigate the distribution of

$$\frac{1}{s}\sum_{k=1}^{s}\xi_k ,$$

the random variables $\xi_1, \ldots, \xi_s$ being not independent. (They are however equivalent.)

The special case, when all a_k are equal to α or $\beta \neq \alpha$, the number of those a_k which are equal to α being a positive percentage <1 of n, is well-known ; as a matter of fact, in this case our question reduces to the investigation of the limiting distribution of the hypergeometric distribution, and it is known (see e. g. BERNSTEIN [1]) that the Moivre—Laplace theorem can be extended for the hypergeometric (instead of the binomial) distribution.

We shall express our principal result (Theorem 1.) as relating to the nth row of an infinite triangular matrix with each row-sum being equal to 0 ; we deduce from this theorem statements concerning sampling among the first n terms of an infinite sequence (Theorem 2.).

We shall prove first the following.

Theorem 1. *Let us consider an infinite triangular matrix*

$$\begin{array}{llllll} a_{11} & & & & & \\ a_{21} & a_{22} & & & & \\ \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ a_{n1} & a_{n2} & \cdot & \cdot & \cdot & a_{nn} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \; \cdot \; \cdot \end{array}$$

with real elements, such that the sum of elements in each row is zero;

$$\sum_{k=1}^{n} a_{nk} = 0 \; . \tag{2}$$

Let us put

$$D_n = \sqrt{\sum_{k=1}^{n} a_{nk}^2} \tag{3}$$

and

$$D_{n,s} = \sqrt{\frac{s}{n}\left(1-\frac{s}{n}\right)}\, D_n \; . \qquad 1 \leq s \leq n \tag{4}$$

where s is an integer.

Let $N_{n,s}(x)$ for any real x denote the number of those sums

$$a_{ni_1} + a_{ni_2} + \ldots + a_{ni_s} \qquad 1 \leq i_1 < i_2 < \ldots < i_s \leq n$$

the value of which does not exceed $x D_{n,s}$, *i. e. put*

$$N_{n,s}(x) = \sum_{\substack{a_{ni_1}+\dots+a_{ni_s}<xD_{n,s} \\ 1\leq i_1<i_2<\dots<i_s\leq n}} 1 \tag{5}$$

and put

$$F_{n,s}(x) = \frac{N_{n,s}(x)}{\binom{n}{s}} \tag{6}$$

Let us put for any $\varepsilon > 0$,

$$d_{n,s}(\varepsilon) = \frac{1}{D_n^2} \sum_{|a_{n,k}|>\varepsilon D_{n,s}} a_{nk}^2 \tag{7}$$

If $s_n \leq n/2$ *is chosen in such a manner that when* n *tends to infinity we have for any* $\varepsilon > 0$

$$\lim_{n\to\infty} d_{n,s_n}(\varepsilon) = 0, \tag{8}$$

it follows that for any real x *we have*

$$\lim_{n\to+\infty} F_{n,s_n}(x) = \Phi(x) \tag{9}$$

where $\Phi(x)$ *is the normal distribution function defined by* (1).

Remark. Clearly if we modify our problem as follows: we denote by $N^*_{n,s}(x)$ the number of those sums $a_{ni_1} + a_{ni_2} + \dots + a_{ni_s}$ which are $< x D_{n,s}$, where $1 \leq i_k \leq n$ $(k = 1, 2, \dots, n)$, then the question whether

$$\lim_{n\to\infty} \frac{N^*_{n,s_n}(x)}{n^s} = \Phi(x)$$

holds, can be answered according to the following theorem which is a special case of a well-known theorem (see [2], p. 103):

Let ξ_{nk} $(k = 1, 2, \dots, s)$ be independently and identically distributed random variables with the distribution

$$\mathbf{P}\{\xi_{nk} = a_{nj}\} = \frac{1}{n} \qquad (j = 1, 2, \dots, n).$$

Then the sums

$$\zeta_{n,s} = \frac{\sum_{j=1}^{s} \xi_{nj}}{D_{n,s}}$$

are in the limit for $n \to +\infty$ normally distributed if and only if (8) holds for every $\varepsilon > 0$. Thus our theorem shows that *if we take a sample without replacement from a finite population, the distribution of the sample mean will be in the limit normal, under the same condition as that which is known for sampling with replacement.*

Proof of Theorem 1. Let us put for any real t

$$\varphi_{n,s}(t) = \int_{-\infty}^{+\infty} e^{itx}\, d\, F_{n,s}(x) = \frac{1}{\binom{n}{s}} \sum_{1\leq i_1<i_2<\ldots<i_s\leq n} e^{it(a_{ni_1}+a_{ni_2}+\ldots+a_{ni_s})}. \tag{10}$$

To prove Theorem 1. we have to show that if (8) holds, we have

$$\lim_{n\to+\infty} \varphi_{n,s_n}\left(\frac{t}{D_{n,s_n}}\right) = e^{-\frac{t^2}{2}}. \tag{11}$$

Now evidently, putting

$$p = \frac{s_n}{n} \tag{12}$$

and

$$B_{n,s_n}(p) = \binom{n}{s_n} p^{s_n}(1-p)^{n-s_n} \tag{13}$$

we have

$$\varphi_{n,s_n}(u) = \frac{1}{2\pi B_{n,s_n}(p)} \int_{-\pi}^{+\pi} \left[\prod_{k=1}^{n} (1-p+pe^{i(\theta+ua_{nk})})\right] e^{-i\theta s_n}\, d\,\theta\ . \tag{14}$$

Taking (2) into account we can write

$$\varphi_{n,s_n}(u) = \frac{1}{2\pi B_{n,s_n}(p)} \int_{-\pi}^{+\pi} \prod_{k=1}^{n} \left((1-p)\, e^{-ip(\theta+ua_{nk})} + pe^{i(1-p)(\theta+ua_{nk})}\right) d\,\theta\ . \tag{15}$$

Now as well-known, if $n\to+\infty$ and $s_n\to+\infty$ (which follows clearly from (8))

$$B_{n,s_n}(p) \sim \frac{1}{\sqrt{2\pi np(1-p)}} \tag{16}$$

and thus we have, substituting $u = \dfrac{t}{D_{n,s_n}}$ and $\theta = \dfrac{\psi}{\sqrt{np(1-p)}}$

$$\varphi_{n,s_n}\left(\frac{t}{D_{n,s_n}}\right) \sim \frac{1}{\sqrt{2\pi}} \int_{-\pi\sqrt{np(1-p)}}^{+\pi\sqrt{np(1-p)}} \left[\prod_{k=1}^{n} \varrho_k(\psi, t)\right] d\,\psi \tag{17}$$

where

$$\varrho_k(\psi, t) = (1-p)\, e^{-ip\left(\frac{\psi}{\sqrt{np(1-p)}}+\frac{ta_{nk}}{D_{n,s_n}}\right)} + pe^{i(1-p)\left(\frac{\psi}{\sqrt{np(1-p)}}+\frac{ta_{nk}}{D_{n,s_n}}\right)}. \tag{18}$$

In what follows $C_1, C_2, \ldots$ are absolute constants. Let us choose an $\varepsilon > 0$. If k is such that

$$\frac{|t|\,|a_{n,k}|}{D_{n,s}} < \varepsilon$$

and $|\psi| < 2\,\varepsilon\sqrt{np(1-p)}$, we have

$$\varrho_k(\psi, t) = 1 - \frac{p(1-p)}{2}\left[\frac{\psi}{\sqrt{np(1-p)}} + \frac{ta_{nk}}{D_{n,s_n}}\right]^2 (1 + \vartheta_1\,\varepsilon) \tag{19}$$

with $|\vartheta_1| \leqq C_1$. On the other hand, we have for an arbitrary k

$$|\varrho_k(\psi, t) - 1| \leqq C_2\left(\frac{p(1-p)\,\psi^2}{2\,n} + \frac{p(1-p)\,|t\,\psi\,a_{nk}|}{2\sqrt{n}\,D_{n,s_n}} + \frac{p(1-p)\,a_{nk}^2}{8\,D_{n,s_n}^2}\right). \tag{20}$$

It follows that for $|\psi| < 2\,\varepsilon\sqrt{np\,(1-p)}$ we have

$$\prod_{k=1}^{n} \varrho_k(\psi, t) = e^{-\frac{\psi^2}{2}-\frac{t^2}{2}} \cdot (1 + \eta_n) \tag{21}$$

where

$$|\eta_n| \leqq C_3\left(d_{n,s_n}\left(\frac{\varepsilon}{|t|}\right) + \frac{\psi^2\,p(1-p)\cdot l_n}{n}\right)$$

where

$$l_n = \sum_{\substack{1 \leq k \leq n \\ |a_{nk}| > \frac{\varepsilon}{|t|} D_{n,s_n}}} 1 .$$

As by the inequality of SCHWARZ we have

$$l_n \leqq \frac{t^2}{\varepsilon^2\,p(1-p)}\, d_{n,s_n}\left(\frac{\varepsilon}{|t|}\right)$$

we obtain

$$\lim_{n\to+\infty} \eta_n = 0 . \tag{22}$$

On the other hand if $|\psi| \geqq 2\varepsilon\sqrt{np(1-p)}$ we use the following estimates:

if $\dfrac{|t|\,|a_{n,k}|}{D_{n,s_n}} > \varepsilon$ then $|\varrho_k(\psi, t)| \leqq 1$

and

if $\dfrac{|a_{n,k}|\,|t|}{D_{n,s_n}} \leqq \varepsilon$, we have $|\varrho_k(\psi, t)| \leqq 1 - p\,(1-p)\,(1-\cos\varepsilon)$

and thus

$$\left| \int\limits_{2\varepsilon\sqrt{np(1-p)} \le |\psi| \le \pi\sqrt{np(1-p)}} \prod_{k=1}^{n} \varrho_k(\psi, t)\, d\psi \right| \le 2\pi \left(1 - \frac{C(\varepsilon)\, s_n}{n}\right)^n \sqrt{s_n} \le 2\pi e^{-C(\varepsilon) s_n} \sqrt{s_n}$$

where $C(\varepsilon) > 0$ is a constant depending only on ε.

Thus (11) holds and therewith our theorem is proved.

Remark. Clearly our condition (8) is a condition of Lindeberg's type. It is easy to see that the following condition of Liapounoff's type implies (8) :

$$\lim_{n \to +\infty} \sqrt{\frac{n}{s_n}} \cdot \frac{\left(\sum\limits_{k=1}^{n} |a_{nk}|^3\right)}{\left(\sum\limits_{k=1}^{n} a_{nk}^2\right)^{3/2}} = 0 . \tag{23}$$

As a matter of fact, if $\frac{s_n}{n} \le \frac{1}{2}$ we have

$$d_{n,s_n}(\varepsilon) \le \frac{1}{\varepsilon D_{n,s_n}} \frac{\sum\limits_{k=1}^{n} |a_{nk}|^3}{\sum\limits_{k=1}^{n} a_{nk}^2} = \frac{1}{\varepsilon} \sqrt{\frac{2n}{s_n}} \cdot \frac{\left(\sum\limits_{k=1}^{n} |a_{nk}|^3\right)}{\left(\sum\limits_{k=1}^{n} a_{nk}^2\right)^{3/2}} .$$

Thus (23) is a sufficient condition for the validity of (9).

For the special case when $a_{nk} = a_k$ does not depend on n ($k = 1, 2, \ldots, n$) we obtain from Theorem 1. the following

Theorem 2. *Let $\{a_n\}$ be an arbitrary sequence of real numbers. Put*

$$M_n = \sum_{k=1}^{n} a_k \tag{24}$$

$$D_n = \sqrt{\sum_{k=1}^{n} \left(a_k - \frac{M_n}{n}\right)^2}$$

and

$$D_{n,s} = \sqrt{\frac{s}{n}\left(1 - \frac{s}{n}\right)} D_n .$$

Let $N_{n,s}(x)$ denote the number of those sums

$$a_{i_1} + a_{i_2} + \ldots + a_{i_s} \qquad 1 \le i_1 < i_2 < \ldots < i_s \le n$$

which do not exceed

$$\frac{s M_n}{n} + x D_{n,s}$$

and put

$$F_{n,s}(x) = \frac{N_{n,s}(x)}{\binom{n}{s}} . \tag{25}$$

If putting $a'_k = a_k - \frac{M_n}{n}$, *and*

$$d_{n,s}(\varepsilon) = \frac{1}{D_n^2} \sum_{\substack{|a'_k| > \varepsilon D_{ns} \\ 1 \leq k \leq n}} |a'_k|^2 \tag{26}$$

if $n \to \infty$, $s = s_n \leqq \frac{n}{2}$ *depends in such a manner on* n *that for any* $\varepsilon > 0$

$$\lim_{n \to +\infty} d_{n,s_n}(\varepsilon) = 0 \tag{27}$$

then we have for any real x

$$\lim_{n \to +\infty} F_{n,s_n}(x) = \Phi(x) . \tag{28}$$

Let us consider some examples.

Example 1. If $0 < c \leqq |a_k| \leqq C$, then (27) is satisfied if $s_n \to +\infty$ arbitrarily slowly.

Example 2. (27) is satisfied, provided that $s_n \to +\infty$ arbitrary slowly if $0 < c\,k^\alpha \leqq |a_k| \leqq Ck^\alpha$ with $\alpha > -1/2$.

Example 3. If

$$a_k = \frac{(-1)^k}{\sqrt{k}} ,$$

then (27) is satisfied if

$$\lim_{n \to +\infty} \frac{\log s_n}{\log n} = 1 ,$$

e. g. for

$$s_n \sim \frac{n}{(\log n)^A}$$

with arbitrary large $A > 0$.

Example 4. Evidently it is impossible to satisfy (8) if

$$\sum a_{nk}^2 \leqq \alpha$$

and

$$\max_k |a_{nk}| \geqq \beta > 0 .$$

The most far reaching previous result known to us is that of W. G. MADOW [3]. His theorem, which can be compared with our result can be stated with our notations as follows:

If

$$\lambda_{nr} = \frac{\left|\frac{1}{n}\sum_{k=1}^{n} a_{nk}^{r}\right|}{\left(\frac{1}{n}\sum_{k=1}^{n} a_{nk}^{2}\right)^{\frac{r}{2}}} \leq C \text{ for all } n \text{ and } r = 3, 4, \ldots \tag{29}$$

then the assertion of Theorem 1. holds, provided that $S_n \to +\infty$. Clearly Madow's theorem follows from ours because if (29) holds for $r = 4$, we have by Hölder's inequality

$$\sqrt{\frac{n}{s_n}}\frac{\sum_{k=1}^{n}|a_{nk}|^3}{\left(\sum_{k=1}^{n} a_{nk}^2\right)^{3/2}} \leq \frac{\lambda_{n4}^{3/4}}{\sqrt{s_n}} \leq \frac{C^{3/4}}{\sqrt{s_n}} \tag{30}$$

and thus if $s_n \to +\infty$, (23) is satisfied, which as remarked above ensures the validity of (9). Thus if (29) holds for the single value $r = 4$, our theorem can be applied, which implies that Madow's result is much weaker than ours.

Madow's result includes as a special case a previous result of F. N. DAVID [4]. Thus our theorem contains as a special case that of the paper [4] too.

(Received 1. December, 1958.)

REFERENCES

[1] БЕРНШТЕЙН, С. Н.: *Теория вероятностей.* Москва, Гостехиздат, 1945.
[2] ГНЕДЕНКО, Б. В и КОЛМОГОРОВ, А. Н: *Предельные распределения для зумм независимых случайных величин.* Москва, Гостехиздат, 1949.
[3] MADOW, W. G.: „On the limiting distributions of estimates based on samples from finite universes." *Annals of Math. Stat.* 19 (1948) 535—545.
[4] DAVID, F. N.: „Limiting distributions connected with certain methods of sampling human population." *Stat. Res. Mem.* (1938) pp. 69—99. especially p. 77.

VÉGES SOKASÁGBÓL VETT MINTÁKRA VONATKOZÓ CENTRÁLIS HATÁRELOSZLÁSTÉTEL

ERDŐS P. és RÉNYI A.

Kivonat

A dolgozat a valószínűségszámítás centrális határeloszlástételének bizonyos gyengén függő változók esetére való kiterjesztésével foglalkozik, amenyiben kimutatja, hogy a Lindeberg-féle feltétellel analóg feltétel teljesülése esetén egy véges sokaságból vett minta elemeinek középértéke közelítőleg normális eloszlású. Más szóval a szerzők a következő tételt bizonyítják be :

1. tétel: *Tekintsünk egy valós elemű*

$$\begin{matrix} a_{11} \\ a_{21} & a_{22} \\ . & . & . \\ . & . & . & . \\ . & . & . & . & . \\ a_{n1} & a_{n2} & . & . & . & a_{nn} \\ . & . & . & . & . & . & . \end{matrix}$$

háromszög alakú mátrixot, melyben az egy sorban álló elemek összege mindenütt nulla :

$$\sum_{k=1}^{n} a_{nk} = 0 \,. \tag{2}$$

Vezessük be a

$$D_n = \sqrt{\sum_{k=1}^{n} a_{nk}^2} \tag{3}$$

és

$$D_{n,s} = \sqrt{\frac{s}{n}\left(1 - \frac{s}{n}\right)} D_n \tag{4}$$

jelöléseket.

Legyen tetszőleges valós x-re $N_{n,s}(x)$ azon

$$a_{ni_1} + a_{ni_2} + \ldots + a_{ni_s} \qquad 1 \leqq i_1 < i_2 < \ldots < i_s \leqq n$$

alakú összegek száma, melyek kisebbek $x\,D_{n,s}$-nél, azaz

$$N_{n,s}(x) = \sum_{\substack{a_{ni_1} + \ldots + a_{ni_s} < x D_{n,s} \\ 1 \leqq i_1 < \ldots < i_s \leqq n}} 1 \tag{5}$$

és legyen

$$F_{n,s}(x) = \frac{N_{n,s}(x)}{\binom{n}{s}} \,, \tag{6}$$

Vezessük be a

$$d_{n,s}(\varepsilon) = \frac{1}{D_n^2} \sum_{|a_{nk}| > \varepsilon D_{n,s}} a_{nk}^2 \qquad 1 \leqq s \leqq n \tag{7}$$

jelölést. Ha az $\{s_n\}$ $(s_n \leqq n/2)$ sorozatot úgy választjuk, hogy minden pozitív ε-ra

$$\lim_{n \to \infty} d_{n,s_n}(\varepsilon) = 0 \tag{8}$$

legyen, akkor minden valós x-re

$$\lim_{n \to \infty} F_{n,s_n}(x) = \Phi(x) \,,$$

ahol $\Phi(x)$ a normális eloszlásfüggvény.

Az 1. tétel speciális eseteként nyerhető az alábbi

2. tétel. *Legyen $\{a_n\}$ tetszőleges valós számsorozat. Vezessük be az*

$$M_n = \sum_{k=1}^{n} a_k$$

$$D_n = \sqrt{\sum_{k=1}^{n} \left(a_k - \frac{M_n}{n}\right)^2} \tag{24}$$

$$D_{n,s} = \sqrt{\frac{s}{n}\left(1 - \frac{s}{n}\right)} D_n \, .$$

jelöléseket. Jelölje $N_{n,s}(x)$ azon

$$a_{i_1} + a_{i_2} + \ldots + a_{i_s} \qquad 1 \leqq i_1 < i_2 < \ldots < i_s \leqq n$$

alakú összegek számát, melyek kisebbek, mint

$$s\frac{M_n}{n} + x D_{n,s}$$

és legyen

$$F_{n,s}(x) = \frac{M_{n,s}(x)}{\binom{n}{s}} \, . \tag{25}$$

Vezessük be az $a'_k = a_k - \frac{M_n}{n}$, és

$$d_{n,s}(\varepsilon) = \frac{1}{D_n^2} \sum_{\substack{|a'_k| > \varepsilon D_{n,s} \\ 1 \leq k \leq n}} a'^2_k \tag{26}$$

jelölést. Ha az $\{s_n\}$ $(s_n \leqq n/2)$ sorozatot úgy választjuk, hogy minden pozitív ε-ra

$$\lim_{n\to\infty} d_{n,s}(\varepsilon) = 0 \tag{27}$$

akkor minden valós x-re

$$\lim_{n\to\infty} F_{n,s_n}(x) = \Phi(x) \, . \tag{28}$$

ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА ДЛЯ ВЫБОРКАХ ВЗЯТЫХ ИЗ КОНЕЧНЫХ МНОЖЕСТВ

P. ERDŐS и A. RÉNYI

Резюме

Работа занимается распространением центральной предельной теоремы вероятностей на случай некоторых слабо зависящих случайных величин. Авторы доказывают, что в случае выполнения условия, аналогичного условию Lindeberg-a, среднее значение элементов выборки взятой из конечного множества, имеет приблизительно нормальное распределение. Иначе говоря доказывается следующая теорема :

Теорема 1: *Рассмотрим треугольную матрицу*

$$\begin{matrix} a_{11} \\ a_{21} & a_{22} \\ \cdot & \cdot \\ \cdot & & \cdot \\ \cdot & & & \cdot \\ a_{n1} & a_{n2} & \cdot & \cdot & \cdot & a_{nn} \\ \cdot & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & & \cdot \end{matrix}$$

из вещественных элементов, в которой сумма элементов каждой строки равна нулю

$$\sum_{k=1}^{n} a_{nk} = 0 \, , \tag{2}$$

Введем обозначения

$$D_n = \sqrt{\sum_{k=1}^{n} a_{nk}^2} \tag{3}$$

и

$$D_{n,s} = \sqrt{\frac{s}{n}\left(1 - \frac{s}{n}\right)}\, D_n \qquad 1 \leqq s \leqq n \, . \tag{4}$$

Пусть для любого вещественного x $N_{n,s}(x)$ обозначает число тех сумм вида $a_{ni_1} + a_{ni_2} + \ldots + a_{ni_s}$ где $1 \leqq i_1 < i_2 < \ldots < i_s \leqq n$, которые меньше чем $x D_{n,s}$ т. е.

$$N_{n,s}(x) = \sum_{\substack{a_{ni_1} + \ldots + a_{ni_s} < x D_{n,s} \\ 1 \leqq i_1 < i_2 < \ldots < i_s \leqq n}} 1 \tag{5}$$

и пусть

$$F_{n,s}(x) = \frac{N_{n,s}(x)}{\binom{n}{s}} \, . \tag{6}$$

Введем обозначение

$$d_{n,s}(\varepsilon) = \frac{1}{D_n^2} \sum_{|a_{nk}| > \varepsilon D_{n,s}} a_{nk}^2 \tag{7}$$

Если последовательность s_n ($\leqq n/2$) выбрана так, что для каждого положительного ε выполняется условие

$$\lim_{n \to \infty} d_{n,s_n}(\varepsilon) = 0 \tag{8}$$

то для всех вещественных x

$$\lim_{n \to +\infty} F_{n,s_n}(x) = \Phi(x) \, .$$

где $\Phi(x)$ нормальная функция распределения.

В качестве специального случая теоремы 1 получается следующая

Теорема 2. *Пусть* $\{a_n\}$ *любая последовательность вещественных чисел. Введем следующие обозначения*

$$M_n = \sum_{k=1}^{n} a_k$$

$$D_n = \sqrt{\sum_{k=1}^{n} \left(a_k - \frac{M_n}{n}\right)^2} \tag{24}$$

$$D_{n,s} = \sqrt{\frac{s}{n}\left(1 - \frac{s}{n}\right)} D_n$$

Обозначим через $N_{n,s}(x)$ *число тех сумм вида*

$$a_{i_1} + a_{i_2} + \ldots + a_{i_s} \qquad (1 \leqq i_1 < \ldots < i_s \leqq n)$$

которые меньше, чем

$$s\frac{M_n}{n} + xD_{n,s} ,$$

и пусть

$$F_{n,s}(x) = \frac{N_{n,s}(x)}{\binom{n}{s}} . \tag{25}$$

Введём обозначение $a'_k = a_k - \dfrac{M_n}{n}$, *и*

$$d_{n,s}(\varepsilon) = \frac{1}{D_n^2} \sum_{\substack{|a'_k| > \varepsilon D_{n,s} \\ 1 \leq k \leq n}} a'^2_k \tag{26}$$

Если последовательность s_n $(\leqq n/2)$ *выбрана так, что для каждого положительного* ε *выполняется условие*

$$\lim_{n\to\infty} d_{n,s_n}(\varepsilon) = 0 \tag{27}$$

то для всех вещественных x

$$\lim_{n\to\infty} F_{n,s_n}(x) = \Phi(x) . \tag{28}$$