Mathematical Publications
DOI: 10.2478/tmmp-2020-0031
Tatra Mt. Math. Publ. **77** (2020), 43–52

CENTRAL LIMIT THEOREM AND THE DISTRIBUTION OF SEQUENCES

Milan Paštéka

Department of Mathematics and Informatics Faculty of Education University of Trnava,
Institute of Mathematics Slovak Academy of Sciences Bratislava
SLOVAKIA

ABSTRACT. The paper deals with independent sequences with continuous asymptotic distribution functions. We construct a compact metric space with Borel probability measure. We use its properties to prove the central limit theorem for independent sequences with continuous distribution functions.

Let $\mathbb{N}$ be the set of positive integers. We say that a set $A \subset \mathbb{N}$ *has an asymptotic density* if the limit

$$\lim_{N\to\infty} \frac{1}{N}|A \cap [1, N]| := d(A) \text{ exists.}$$

($|B|$ denotes the cardinality of the set B.) In this case, the value $d(A)$ will be called the *asymptotic density* of the set A. We shall denote $\mathcal{D}$ the system of all subsets of $\mathbb{N}$ having an asymptotic density. If $\{v(n)\}$ is a sequence, then for each set S, we denote $v^{-1}(S) = \{n \in \mathbb{N}; v(n) \in S\}$. We say that real valued sequence $\{v(n)\}$ has an asymptotic distribution function if for each real number x the set $v^{-1}((-\infty, x))$ belongs to $\mathcal{D}$. In this case, the function

$$F(x) = d\big(v^{-1}((-\infty, x))\big)$$

is called the *asymptotic distribution function* of the sequence $\{v(n)\}$ (see [9], [22] and [23]).

2010 Mathematics Subject Classification: 11B05.
Keywords: asymptotic density, distribution function, probability, measure density.
The research is supported by grant VEGA 02/0109/18.

Consider $N \in \mathbb{N}$ and a real sequence $\{v(n)\}$. Put

$$E_N(v) = \frac{1}{N}\sum_{n=1}^{N} v(n).$$

We say that $\{v(n)\}$ has a *mean value* if the proper limit

$$E(v) := \lim_{N\to\infty} E_N(v) \text{ exists.}$$

$E(v)$ is in this case called the *mean value* of $\{v(n)\}$. If $\{v(n)\}$ is a bounded sequence having asymptotic distribution function F, then the Weyl criterion provides that $\{v(n)\}$ has a mean value and

$$E(v) = \int_{x_1}^{x_2} x \mathrm{d}F(x).$$

Analogously in this case, there exists the *dispersion* of $\{v(n)\}$

$$D^2(v) = E\big((v - E(v))^2\big) = \int_{x_1}^{x_2} x^2 \mathrm{d}F(x) - E(v)^2.$$

The sequences $\{v_1(n)\}, \{v_2(n)\}, \{v_3(n)\}, \dots$ of elements of some interval $[x_1, x_2]$ are called *statistically independent* if for each $k \in \mathbb{N}$ we have

$$E_N\big(f_1(v_1)\dots f_k(v_k)\big) - E_N\big(f_1(v_1)\big)\dots E_N\big(f_k(v_k)\big) \to 0$$

for $N \to \infty$ for every real function $f_1, \dots, f_k$ continuous on $[x_1, x_2]$, (see [21], [23]).

The aim of this paper is to prove the following

Theorem 1. *Let $\{v_i(n)\}, i = 1, 2, 3, \dots$ be statistical independent sequences of elements from the interval $[0, 1]$. Suppose that each of these sequences has continuous asymptotic distribution function F. Let E be the mean value of $\{v_i(n)\}$, $i = 1, 2, 3, \dots$ and D^2 be the dispersion of these sequences. Denote*

$$S_N(x) = \left\{ n \in \mathbb{N}; \frac{\sum_{i=1}^{N} v_i(n) - NE}{\sqrt{N}D} \le x \right\}, \quad \textit{for } N = 1, 2, 3, \dots \tag{1}$$

Then for every x, N the set $S_N(x)$ has an asymptotic density and

$$\lim_{N\to\infty} d\big(S_N(x)\big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \mathrm{d}t.$$

In the first part, we recall some facts on the uniform distribution in compact spaces. Then, we define a metric on the set $\mathbb{N}$ such that the completion is a compact space equipped with Borel probability measure. This metric will be defined in a manner providing the uniform continuity of sequences $\{v_i(n)\}, i \in \mathbb{N}$. These sequences can be extended to a continuous real function which allows us to study them as random variables. Then, we apply the classical central limit theorem. The definition of mentioned metric is inspired by the Novoselov's construction of the ring of polyadic numbers, (see [10], [11]).

Example 1.

1) If $\{v_i(n)\}, i \in \mathbb{N}$ are van der Corput sequences with relatively prime bases, then they are statistically independent. This result is proved in [20].
2) If 1 and $\theta_i, i \in \mathbb{N}$ are algebraically independent over the field of rational numbers, then the sequences of fractional parts $\{\{\theta_i n\}\}$ are statistically independent. This leads to the special case
3) If θ is a transcendent number, then the sequences $\{\{\theta^i n\}\}$ are statistically independent.

We start by a characterisation of statistical independence. Consider a sequence of vectors $\{(u_1(n), \dots, u_N(n))\}$, where $\{u_i(n)\}$ are sequences of real numbers. We say that this sequence has asymptotic distribution function if for arbitrary real numbers $x_1, \dots, x_N$ the set $\cap_{i=1}^N u_i^{-1}(-\infty, x_i)$ belongs to $\mathcal{D}$ and in this case the function

$$F(x_1, \dots, x_N) = d\left(\bigcap_{i=1}^N u_i^{-1}(-\infty, x_i)\right)$$

is called the *asymptotic distribution function* of the vector sequence

$$\{(u_1(n), \dots, u_N(n))\}.$$

By straight forward transcription of proof of Theorem 1 in [20], we get

Proposition 1. *Assume that $\{u_1(n)\}, \dots, \{u_N(n)\}$ are bounded sequences having continuous asymptotic distribution function $F_i, i = 1, \dots N$, respectively. Then these sequences are statistically independent if the vector sequence*

$$\{(u_1(n), \dots, u_N(n))\}$$

has asymptotic distribution function

$$F(x_1, \dots, x_n), \ \textit{where} \ F(x_1, \dots, x_N) = \prod_{i=1}^N F_i(x_i) \ \textit{for real numbers} \ x_1, \dots, x_N.$$

In the following text, $\{v_i(n)\}, i = 1, 2, 3, \dots$ will be sequences fulfilling the assumption of Theorem 1.

From [20, Theorem 2 and Proposition 15] we can deduce

Proposition 2. *For every $N = 1, 2, 3, \ldots,$ the sequence $\{v_1(n) + \cdots + v_N(n)\}$ has a continuous asymptotic distribution function.*

For the proof, we use the properties of sequences uniformly distributed in compact space.

Uniform distribution on compact spaces

Let $\mathbb{X}$ be a compact metric space equipped with Borel probability measure P. A sequence $\{x(n)\}$ of elements of $\mathbb{X}$ is called *uniformly distributed* with respect to P if and only if

$$\lim_{N\to\infty} E_N\big(f(x)\big) = \int f \mathrm{d}P$$

for each continuous real function $f(x)$ defined on $\mathbb{X}$, (see [9]).

Let $\mathbb{N}$ be the set of positive integers and $(\Omega, \mathfrak{d})$ a compact metric space containing $\mathbb{N}$ as a dense subset. Suppose that P is a Borel probability measure defined on this metric space. Denote for $S \subset \mathbb{N}$

$$\nu^*(S) = P\big(\mathrm{cl}(S)\big),$$

where $\mathrm{cl}(\cdot)$ is the topological closure in $(\Omega, \mathfrak{d})$.

The set function ν^* is a measure density, (see [18]), with the algebra of ν^* measurable sets

$$\mathcal{D}_\nu = \{S \subset \mathbb{N}; \nu^*(S) + \nu^*(\mathbb{N} \setminus S) = 1\}.$$

We recall that a sequence of real numbers $\{v(n)\}$ is called *uniformly continuous* with respect to the metric $\mathfrak{d}$ if

$$\forall \varepsilon > 0 \, \exists \, \delta > 0; \quad \mathfrak{d}(n_1, n_2) < \delta \Rightarrow |v(n_1) - v(n_2)| < \varepsilon.$$

In this case the sequence $\{v(n)\}$ can be extended to a continuous real function $\tilde{v}$ defined on Ω in the following way

$$\tilde{v}(\alpha) = \lim_{n\to\infty} v(s_n),$$

where $\{s_n\}$ is a sequence of positive integers where $s_n \to \alpha$ with respect to the metric $\mathfrak{d}$. We obtain from the continuity of $\tilde{v}$ that $\tilde{v}$ is measurable and so it can be considered a random variable in the probability space Ω.

We say that a real valued sequence $\{v(n)\}$ is ν^* *measurable* if and only for each real number x the set $v^{-1}((-\infty, x))$ belongs to $\mathcal{D}_{\nu^*}$. In this case, the function $F(x) = \nu\big(v^{-1}((-\infty, x))\big)$ is called ν^*-*distribution function* of $\{v(n)\}$.

Analogously to the proof of Theorem 6, it can be proved that

Theorem 2. *Let $\{v(n)\}$ be a real valued sequence uniformly continuous with respect to the metric $\mathfrak{d}$ and F a continuous function defined on the real line. The following assertions are equivalent:*

i) *$\{v(n)\}$ is ν^* measurable and its ν^*-distribution function is F.*

ii) *For each real number x we have $\nu^*\big(\{n \in \mathbb{N}; v(n) < x\}\big) = F(x)$.*

iii) *F is the distribution function of the random variable $\tilde{v}$.*

As usual, we denote $\mathcal{X}_S$ the indicator function of a given set S. Analogously, as in [18, Theorem 6 on page 151], it can be proved

Proposition 3. *A set $S \subset \mathbb{N}$ is ν^* measurable if and only if the mean value $E\big(\mathcal{X}_S(k)\big)$ exists for each sequence of positive integers $\{k_n\}$ uniformly distributed in Ω with respect to P. In this case $E\big(\mathcal{X}_S(k)\big) = \nu(S)$.*

We can derive the Weyl criterion from this in a standard way

Proposition 4. *A bounded real sequence $\{v(n)\}$, contained in the interval $[x_1, x_2]$ has ν^*-distribution function F, where F is a continuous function on the real line if for each sequence of positive integers $\{k_n\}$ uniformly distributed in Ω with respect to P we have*

$$\lim_{N\to\infty} E_N\Big(g\big(v(k_n)\big)\Big) = \int_{x_1}^{x_2} g(x)\mathrm{d}F(x)$$

for each continuous real function g defined on $[x_1, x_2]$.

This leads to

Proposition 5. *Let F be a continuous function defined on the real line. If the sequence $\{n\}$ is uniformly distributed in Ω with respect to P, then uniformly continuous real sequence $\{v(n)\}$ has ν^*-distribution function F if and only if it has asymptotic distribution function F.*

Theorem 2 implies

Proposition 6. *Under the assumptions of Proposition 5, we have that the random variable $\tilde{v}$ has distribution function F.*

As usual, we say that ν^*-measurable sequences $\{v_1(n)\},\dots,\{v_N(n)\}$ are ν^*-*independent* if and only if for arbitrary real numbers $x_1,\dots,x_N$ we have that

$$v_1^{-1}((-\infty,x_1))\cap\cdots\cap v_N^{-1}((-\infty,x_N))\in\mathcal{D}_\nu$$

and

$$\nu\left(\bigcap_{j=1}^{N} v_j^{-1}((-\infty,x_j))\right)=\prod_{j=1}^{N}\nu(v_j^{-1}((-\infty,x_j)). \tag{2}$$

A ν^* measurable sequence $\{v(n)\}$ such that its ν^* distribution function is continuous is called *continuously distributed.*

Theorem 3. *Let $\{v_j(n)\}, j=1,\dots,N$ be ν^*-continuously distributed sequences uniformly continuous with respect to $\mathfrak{d}$. Then these sequences are ν^*-independent if and only if random variables $\tilde{v_1},\dots,\tilde{v_N}$ are independent.*

Construction of a compact space

Let $\{v_1(n)\},\dots,\{v_j(n)\},\dots$ be statistically independent sequences of elements of $[0,1]$ having continuous asymptotic distribution function F. Without loss of generality, we can assume that it holds for $n,n'\in\mathbb{N}$

$$n=n'\Longleftrightarrow\forall j\in\mathbb{N};\quad v_j(n)=v_j(n'). \tag{3}$$

Denote

$$I_i^{(m)}=\left[\frac{i}{2^m},\frac{i+1}{2^m}\right),\quad i=1,\dots,2^m-2,\quad I_{2^m-1}^{(m)}=\left[\frac{2^m-1}{2^m},1\right],$$

where $m=1,2,3,\dots$

Denote

$$E_{i_1,\dots,i_m}^{(m)}=\bigcap_{k=1}^{m} v_k^{-1}\left(I_{i_k}^{(m)}\right),\quad 0\le i_1,\dots,i_m\le 2^m-1. \tag{4}$$

Clearly, every set $E_{i_1,\dots,i_m}^{(m)}$ has an asymptotic density and

$$d\left(E_{i_1,\dots,i_m}^{(m)}\right)=\prod_{j=1}^{m}\left(F\left(\frac{i_j+1}{2^m}\right)-F\left(\frac{i_j}{2^m}\right)\right),\quad 0\le i_1,\dots,i_m\le 2^m-1. \tag{5}$$

Let $\mathcal{E}_m$ be the system of all sets of the form (4) for $m\in\mathbb{N}$. This system of sets forms a decomposition of $\mathbb{N}$. Put $\psi_m(a,b)=0$ if a,b belong to the same set from

$$\mathcal{E}_m\quad\text{and}\quad\psi_m(a,b)=1,\quad\text{otherwise.}$$

Put

$$\mathfrak{d}(a,b)=\sum_{m=1}^{\infty}\frac{\psi_m(a,b)}{2^n}. \tag{6}$$

P r o o f o f T h e o r e m 1. It is proved in [8] that:

1. $(\mathbb{N}, \mathfrak{d})$ is a totally bounded metric space.

Denote $(\Omega, \mathfrak{d})$ the completion of $(\mathbb{N}, \mathfrak{d})$. Let us remark that we denote the extension of metric $\mathfrak{d}$ by the same symbol. We have

2. The metric space $(\Omega, \mathfrak{d})$ is compact.

Proposition 7. *Each sequence* $\{v_m(n)\}, m \in \mathbb{N}$ *is uniformly continuous with respect to the metric* $\mathfrak{d}$*.*

P r o o f. Let $n > m$. If $\mathfrak{d}(a,b) < \frac{1}{2^n}$, then (6) implies that $a, b \in E$ for suitable $E \in \mathcal{E}_{n+1}$ thus $a, b \in v_m^{-1}(I_j^{(n+1)})$ for some j and so

$$v_m(a), v_m(b) \in I_j^{(n+1)} \quad \text{therefore} \quad |v_m(a) - v_m(b)| \leq \frac{1}{2^{n+1}}. \qquad \square$$

Let $\mathcal{Y}$ be the set algebra generated by

$$\bigcup_{m=1}^{\infty} \mathcal{E}_m,$$

then $\mathcal{Y} \subset \mathcal{D}$ and $\Delta = d|_{\mathcal{Y}}$ is a finitely additive probability measure on $\mathcal{Y}$. Put

$$\nu^*(S) = \inf\{\Delta(A); S \subset A \wedge A \in \mathcal{Y}\}.$$

The following is proved in [8] (see also [18])

3. For $S \subset \mathbb{N}$, there exists such Borel probability measure P defined on Ω that

$$\nu^*(S) = P\big(\mathrm{cl}(S)\big).$$

Clearly,

4. $\mathcal{D}_\nu \subset \mathcal{D}$ and $\nu(S) = d(S)$ for $S \in \mathcal{D}_\nu$.

For each set $E \in \mathcal{E}_m, m = 1, 2, 3, \ldots$ the set $\mathrm{cl}(E)$ is closed and open (see [8]) and so

5. $\mathrm{cl}(E) \cap \mathbb{N} = E$.

This yields

6. $E(\mathcal{X}_{\mathrm{cl}(E)}) = d(E)$.

Since every continuous real function $\tilde{v}$ can be uniformly approximated by the step functions

$$\sum_{j=1}^{2^m} c_j \mathcal{X}_{\mathrm{cl}(E_j)}, \quad \text{where } E_j \in \mathcal{E}_m,$$

from 6, we get that

$$\lim_{N \to \infty} E_N(\tilde{v}) = \int \tilde{v} \mathrm{d}P.$$

Thus

7. The sequence $\{n\}$ is uniformly distributed in Ω with respect to P.

Let us define the sequence $\{w_N(n)\}$, where

$$w_N(n) = \frac{\sum_{i=1}^{N} v_i(n) - NE}{\sqrt{N}D} \quad \text{for } N = 1, 2, 3, \ldots$$

The sequences $\{v_i(n)\}$ are uniformly continuous with respect to the metric $\mathfrak{d}$ and so $\{w_N(n)\}$ is uniformly continuous, too. Clearly,

$$\tilde{w}_N = \frac{\sum_{i=1}^{N} \tilde{v}_i - NE}{\sqrt{N}D}$$

The random variables $\tilde{v}_i, i = 1, 2, 3, \ldots$ are independent and have the same distribution function F. The classic central limit theorem gives

$$\lim_{N \to \infty} P(\tilde{w}_N \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \mathrm{d}t. \tag{7}$$

Clearly,

$$S_N(x) = \{n \in \mathbb{N};\ w_N(n) \leq x\}.$$

Proposition 2 yields that the sequence $\{w_N(n)\}$ has continuous asymptotic distribution function for $N = 1, 2, 3, \ldots$ And so, from Proposition 5 we have that the set $S_N(x)$ is ν^*-measurable and its ν^*-distrubution function coincides with its asymptotic distribution function and so

$$\nu\big(S_N(x)\big) = d\big(S_N(x)\big).$$

Theorem 2 yields $d\big(S_N(x)\big) = P\big(\tilde{w}_N \leq x\big)$. And so from (7) we get Theorem 1. □

REFERENCES

[1] BUCK, R. C.: *The measure theoretic approach to density*, Amer. J. Math. **68**, 1946, 560–580.

[2] DRMOTA, M.—TICHY, R. F.: *Sequences, Discrepancies and Applications*. Springer--Verlag, Berlin, 1997.

[3] ERDŐS, P.: *On the density of some sequences of numbers II*, J. London Math. Soc. **12** (1937), 7–11.

[4] GREKOS, G.: *On various definitions of density (survey)*, Tatra Mt. Math. Publ. **31** (2005), 17–27.

[5] GREKOS, G.: *The density set (a survey)*, Tatra Mt. Math. Publ. **31** (2005), 103–111.

[6] GRABNER, P. J.—STRAUCH, O.—TICHY, R. F.: *L_p-discrepancy and statistical independence of sequences*, Czechoslovak Math. J. **49** (1999), no. 1, 97–110.

[7] GRABNER, P. J.—TICHY, R. F.: *Remark on statistical independence of sequences*, Math. Slovaca **44** (1994), no. 1, 91–94.

[8] IACO, M. R.—PAŠTÉKA, M.—TICHY R. F.: *Measure density for set decompositions and uniform distribution*, Rend. Circ. Math. Palermo **64** (2015), no. 2, 323–339.

[9] KUIPERS, L.—NIEDERREITER, H.: *Uniform Distribution of Sequences*. John Wiley and Sons, N. Y. London, Sydney, Toronto, 1974.

[10] NOVOSELOV, E. V.: *Topological theory of polyadic numbers*,Trudy Tbilis. Mat. Inst. **27** (1960), 61–69, (In Russian)

[11] NOVOSELOV, E. V.: *New method in the probabilistic number theory*, Izv. Doklady Akad. Nauk. Ser. Mat. **28** (1964), no. 2, 307–364. (In Russian)

[12] NARKIEWICZ, W.: *Teoria Liczb*, PWN, Warszawa, 1991. (In Polish)

[13] NIVEN, I.: *Uniform distribution of sequences of integers*, Trans. Amer. Math. Soc. **98** (1961), 52–61.

[14] PAŠTÉKA, M.: *Some properties of Buck's measure density*, Math. Slovaca **42** (1992), no. 1, 15–32.

[15] PAŠTÉKA, M.: *Remarks on one type of uniform distribution*, Unif. Distrib. Theory **2** (2007), no. 1, 79–92.

[16] PAŠTÉKA, M.: *On Four Approaches to Density*. Spectrum Slovakia Vol. 3. Peter Lang, Frankfurt am Main; VEDA, Publishing House of the Slovak Academy of Sciences, Bratislava, 2014.

[17] PAŠTÉKA, M.: *Remarks on Buck's measure density*, Tatra Mt. Math. Publ. **3** (1993), 191–200.

[18] PAŠTÉKA, M.: *Density and Related Topics*. Veda, Bratislava; Academia, Praha, 2017.

[19] PAŠTÉKA, M.—TICHY, R.: *A note on the correlation coefficient of arithmetic functions*, Acta Acad. Paedagog. Agriensis Sect. Math. (N.S.) **30** (2003), 109–114.

[20] PAŠTÉKA, M.—TICHY, R.: *Measurable sequences*. Riv. Mat. Univ. Parma (N.S.) **10** (2019), no. 1, 63–84.

[21] RAUZY, G.: *Proprietes Statistique de Suites Arithmetiques*. Le Mathématicien Vol. l5, Collection SUP, Presses Universitaires de France, Paris, 1976.

[22] SCHOENBERG, I.: *Über die asymptotische Verteilung reeler Zahlen mod 1*, Math. Z. **28** (1928), 171–199. `https://link.springer.com/article/10.1007/BF01181156`

[23] STRAUCH, O.: *Distribution of Sequences: A Theory*. VEDA, Bratislava; Academia, Prague, 2019.

[24] STRAUCH, O.—PORUBSKÝ, Š.: *Distribution of Sequences a Sampler*. In: *Series of the Slovak Academy of Sciences*. Vol. 1. Peter Lang, Frankfurt am Main, 2005.

[25] WEYL, H.: *Über die Gleichverteilung von Zahlen mod. Eins*, Math. Ann **77** (1916), 313–352.

Received April 14, 2020

Department of Mathematics and Informatics
Faculty of Education
University of Trnava
Priemyselna 4
918 43 Trnava
P. O. BOX 9
SLOVAKIA

Institute of Mathematics
Slovak Academy of Sciences
Štefánikova 49
814 73 Bratislava
SLOVAKIA

E-mail: milan.pasteka@truni.sk
milan.pasteka@mat.savba.sk