sciendo

TATRA MOUNTAINS
Mathematical Publications
DOI: 10.2478/tmmp-2025-0002
Tatra Mt. Math. Publ. **89** (2025), 207–214

ON STATISTICAL INDEPENDENCE AND DENSITY INDEPENDENCE

Milan Paštéka

Institute of Mathematics, Slovak Academy of Sciences, Bratislava, SLOVAKIA

ABSTRACT. In the first part we recall the notion of statistical independence. The second part is devoted to the definition of selective density and its connection with the distribution of sequences. Then we define the independence of sequences with respect to selective density. Finally, we prove that these two types of independence are equivalent.

1. Introduction

The aim of this paper is to characterize statistical independence as the independence of given sequences with respect to a class of selective densities. This is an extension of Theorem 3 of [10]. We apply similar ideas as in the proof of the mentioned theorem.

1.1. Notation

$\mathcal{X}_L$**:** indicator function of the set L

$$v^{-1}(I) = \{n; v(n) \in I\}.$$

$|I|$**:** length of the interval I.

$|A|$**:** cardinality of the finite set A.

$\mathcal{R}([0,1])$**:** space of real functions defined on $[0,1]$

$$\mathcal{R}([0,1])^m = \mathcal{R}([0,1]) \times \cdots \times \mathcal{R}([0,1]) \text{ — } (m \text{ times}).$$

$\mathcal{C}([0,1])$**:** space of continuous real functions defined on $[0,1]$

$$\mathcal{C}([0,1])^m = \mathcal{C}([0,1]) \times \cdots \times \mathcal{C}([0,1]) \text{ — } (m \text{ times}).$$

2020 Mathematics Subject Classification: 11K06.
Keywords: statistical independence, inependence, asymptotic density, density, uniform distribution, distribution function.
Research supported by the Grant VEGA 2/0119/23.

Helly selection principle. The following case of result of Eduard Helly (see [6, page 372]) has crucial role: *Let* $H_N, N = 1, 2, 3, \dots$ *be non-decreasing real functions defined on* $[0, 1]$, *such that* $H_N(0) = 0, H_N(1) = 1$. *Then such increasing sequence of natural numbers* k_N *that the sequence* H_{k_N} *converges to suitable function* H *in each point of* $[0, 1]$.

1.2. Statistical independence

The notion of statistically independent sequences is introduced and studied in the work [11].

Let $v_1, \dots, v_m$ be a sequences of elements of $[0, 1]$. For $\overline{h} \in \mathcal{R}^m([0, 1])$ we define the multi linear forms

$$\Delta_N(\overline{h}) = \frac{1}{N} \sum_{n=1}^{N} h_1\big(v_1(n)\big) \dots h_m\big(v_m(n)\big),$$

$$\delta_N(\overline{h}) = \prod_{i=1}^{m} \frac{1}{N} \sum_{n=1}^{N} h_i\big(v_i(n)\big) = \left(\frac{1}{N}\right)^m \prod_{i=1}^{m} \sum_{n=1}^{N} h_i\big(v_i(n)\big).$$

where $\overline{h} = (h_1, \dots, h_m) \in \mathcal{R}([0, 1])^m$.

We say that these sequences are *statistically independent* if

$$\lim_{N\to\infty} \Delta_N(\overline{f}) - \delta_N(\overline{f}) = 0, \tag{1}$$

for every $\overline{f} = (f_1, \dots, f_m) \in \mathcal{C}([0, 1])^m$.

Let us remark that this notion was later studied in various papers. For example [3], [4].

1.3. Selective density

The distribution of sequences was first studied by Hermann Weyl in his famous paper [15]. We say that a set $L \subset \mathbb{N}$ has asymptotic density if the limit

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

In this case the value $d(L)$ is called *asymptotic density* of the set L. Weyl's definition of uniform distribution is equivalent to: given sequence $\{v(n)\}$ of elements of $[0, 1]$ is *uniformly distributed* mod 1 *if for each interval* $I \subset [0.1]$ *the set* $v^{-1}(I)$ *has asymptotic density and* $d(v^{-1}(I)) = |I|$.

Assume that an increasing sequence of natural numbers $\kappa = \{k_1 < k_2 < \dots \dots < k_N < \dots\}$ *is given. We say that a set* $L \subset \mathbb{N}$ *is* κ *measurable if the limit*

$$\lim_{N\to\infty} \frac{|L \cap [1, k_N]|}{k_N} := d_\kappa(L) \quad \textit{exists.}$$

In this case the value $d_\kappa(L)$ *is called* κ *density of* L. *The system of all* κ*-measurable sets we shall denote* $\mathcal{D}_\kappa$.

1.4. κ-measurability

A sequence of $v = \{v(n)\}$ of elements of $[0,1]$ will be called κ-measurable if for each $x \in [0,1]$ the set $v^{-1}([0,x))$ belongs to $\mathcal{D}_\kappa$. In this case the function $F(x) = d_\kappa(v^{-1}([0,x))$ we shall call κ-distribution function of v. The κ measurability of v implies

$$\lim_{N\to\infty} \frac{1}{k_N} \sum_{n=1}^{k_N} f\big(v(n)\big) = \int_0^1 f(x)dF(x) \tag{2}$$

for each Riemann - Stjelties integrable function f defined on the interval $[0,1]$. These types of equalities are known as Weyl criterions. For the proof we refer to [7], [1], [13], [5].

Consider a sequence $v = \{v(n)\}$ of elements of $[0,1]$. Applying Helly selection principle to the sequence of functions

$$H_N(x) = \frac{|v^{-1}([0,x)) \cap [1,N]|}{N}, x \in [0,1] \quad \textit{for } N = 1,2,3,\ldots$$

we get that such increasing sequence $\kappa = \{k_N\}$ exists that v is κ-measurable.

1.5. κ-independence

Suppose that sequences $v_i, i = 1,\ldots,m$ of elements of $[0,1]$ are given. Applying the Helly selection principle step by step to the sequences v_i we get that such increasing sequence of natural numbers κ exists that these sequences are κ-measurable.

We say that these sequences are κ-independent if the set

$$v_1^{-1}([0,x_1)) \cap \cdots \cap v_m^{-1}([0,x_m))$$

belongs to $\mathcal{D}_\kappa$ and

$$d_\kappa\big(v_1^{-1}([0,x_1)) \cap \cdots \cap v_m^{-1}([0,x_m))\big) = F_1(x_1)\cdots F_m(x_m), \tag{3}$$

for every $x_1,\ldots,x_m \in [0,1]$ such that F_i is continuous in the point

$$x_i, i = 1,\ldots,m.$$

1.6. The equivalence

Now we can formulate the main result:

Theorem 1. *Let $\{v_1(n)\}, \{v_2(n)\}, \ldots, \{v_m(n)\}$ be sequences of the interval $[0,1]$. Then the following statements are equivalent:*

(i) *given sequences are statistically independent,*

(ii) *given sequences are κ-independent for each sequence κ such that these sequences are κ measurable.*

We start by some preliminary observations. Put

$$v_1^{-1}\big([0,x_1)\big)\cap\dots\cap v_m^{-1}\big([0,x_m)\big)=L(x_1,\dots,x_m):=L.$$

We see that the indicator functions fulfil the equation

$$\mathcal{X}_{v_1^{-1}([0,x_1))}(n)\dots\mathcal{X}_{v_m^{-1}([0,x_m))}(n)=\mathcal{X}_L(n).$$

Taking into account that $\mathcal{X}_I\big(v(n)\big)=\mathcal{X}_{v^{-1}(I)}(n)$ *for arbitrary interval* I *and real sequence* $\{v(n)\}$ *we get*

$$\mathcal{X}_{v_i^{-1}([0,x_i))}(n)=\mathcal{X}_{[0,x_1)}\big(v_i(n)\big),\quad i=1,\dots,m. \tag{4}$$

This implies

$$\Delta_N\left(\mathcal{X}_{[0,x_1)},\dots,\mathcal{X}_{[0,x_m)}\right)=\frac{1}{N}\sum_{n=1}^{N}\mathcal{X}_L(n). \tag{5}$$

Let us remark that for κ *measurable sequences* $v_i, i=1,\dots,m$ *we have*

$$\lim_{N\to\infty}\delta_{k_N}(h_1,\dots,h_m)=\prod_{i=1}^{m}\int_0^1 h_i(x)dF_i(x), \tag{6}$$

where F_i *is* d_κ*-distribution function of* v_i*, and* h_i *are Riemann Stjelties integrable functions with respect to* F_i*.*

P r o o f o f T h e o r e m 1.

i)⇒(ii) Suppose that given sequences are statistically independent. Let x_i is a point of continuity of F_i, $i=1,\dots,m$. Then for given $\varepsilon>0$ such continuous functions

$$0\le f_i\le f_i',\quad i=1,\dots,m$$

exist that

$$f_i\le\mathcal{X}_{[0,x_i)}\le f_i' \tag{7}$$

and

$$\int_0^1 f_i'(x)dF_i(x)-\int_0^1 f_i(x)dF_i(x)\le\varepsilon. \tag{8}$$

Taking into account (4) we get from (7)

$$f_i\big(v_i(n)\big)<\mathcal{X}_{[0,x_i)}(n)<f_i'\big(v_i(n)\big)\quad\text{for } n\in\mathbb{N}\quad\text{and}\quad i=1,\dots,m.$$

This yields

$$\Delta_N(f_1,\dots,f_m)\le\Delta_N\left(\mathcal{X}_{[0,x_1)},\dots,\mathcal{X}_{[0,x_m)}\right)\le\Delta_N(f_1',\dots,f_m')$$

and so (5) yields

$$\Delta_N(f_1,\dots,f_m) \le \frac{1}{N}\sum_{n=1}^{N}\mathcal{X}_L(n) \le \Delta_N(f_1',\dots,f_m'). \tag{9}$$

Moreover, it holds

$$\delta_N(f_1,\dots,f_m) \le \delta_N\left(\mathcal{X}_{[0,x_1)},\dots,\mathcal{X}_{[0,x_m)}\right) \le \delta_N(f_1',\dots,f_m'). \tag{10}$$

Suppose that $\kappa = \{k_N\}$ is such increasing sequence of natural numbers that the sequences $v_i, i = 1,\dots,m$ are κ measurable. Then

$$\lim_{N\to\infty}\delta_{k_N}\left(\mathcal{X}_{[0,x_1)},\dots,\mathcal{X}_{[0,x_m)}\right) = \prod_{i=1}^{m} d_\kappa\bigl(v_i^{-1}([0,x_i))\bigr).$$

We suppose that the sequences $\{v_1(n)\},\dots,\{v_m(n)\}$ are statistically independent and so we can apply (1) for k_N and we get

$$\lim_{N\to\infty}\Delta_{k_N}(f_1,\dots,f_m) = \prod_{i=1}^{m}\int_0^1 f_i(x)dF_i(x)$$

and

$$\lim_{N\to\infty}\Delta_{k_N}(f_1',\dots,f_m') = \prod_{i=1}^{m}\int_0^1 f_i'(x)dF_i(x).$$

And so

$$\prod_{i=1}^{m}\int_0^1 f_i(x)dF_i(x) \le \varliminf_{N\to\infty}\frac{1}{k_N}\sum_{n=1}^{k_N}\mathcal{X}_L(n) \le \prod_{i=1}^{m}\int_0^1 f_i'(x)dF_i(x)$$

and

$$\prod_{i=1}^{m}\int_0^1 f_i(x)dF_i(x) \le \varlimsup_{N\to\infty}\frac{1}{k_N}\sum_{n=1}^{k_N}\mathcal{X}_L(n) \le \prod_{i=1}^{m}\int_0^1 f_i'(x)dF_i(x).$$

Clearly,

$$\prod_{i=1}^{m}\int_0^1 f_i(x)dF_i(x) \le \prod_{i=1}^{m} d_\kappa\bigl(v_i^{-1}([0,x_i))\bigr) \le \prod_{i=1}^{m}\int_0^1 f_i'(x)dF_i(x).$$

And so taking into account (8) and $\varepsilon \to 0^+$ we get

$$\prod_{i=1}^{m} d_\kappa\bigl(v_i^{-1}([0,x_i))\bigr) = \lim_{N\to\infty}\frac{1}{k_N}\sum_{n=1}^{k_N}\mathcal{X}_L(n) = d_\kappa(L).$$

(ii)$\Rightarrow$ (i).
Choose now the sequence κ in such manner that the sequences $v_1, \dots, v_m$ are κ-measurable and there are κ-independent.

Let $\mathcal{S}$ be the set of all step functions $s = a_1\mathcal{X}_{I_1} + \dots + a_\ell \mathcal{X}_{I_\ell}$, where the end-points of the intervals I_j are points of continuity of $F_1, \dots, F_m$. Then for every $g_1, \dots, g_m \in \mathcal{S}$ we have

$$\lim_{N\to\infty} \Delta_{k_N}(g_1, \dots, g_m) - \delta_{k_N}(g_1, \dots, g_m) = 0.$$

And so from (6) we get

$$\lim_{N\to\infty} \Delta_{k_N}(g_1, \dots, g_m) = \int_0^1 g_1(x)dF_1(x) \cdots \int_0^1 g_m(x)dF_m(x). \tag{11}$$

Let us consider real functions $f_1, \dots, f_m$ continuous on the interval $[0,1]$. These functions are bounded, thus we can suppose to be positive. For each $\varepsilon > 0$ such step function $s_1, \dots, s_m,\ S_1, \dots, S_m \in \mathcal{S}$ exists that

$$s_i \le f_i \le S_i, \tag{12}$$

$$\int_0^1 S_i(x)dF_i(x) - \int_0^1 s_i(x)dF_i(x) < \varepsilon, \tag{13}$$

for $i = 1, \dots, m$. This yields

$$\int_0^1 s_i(x)dF_i(x) \le \int_0^1 f_i(x)dF_i(x) \le \int_0^1 S_i(x)dF_i(x).$$

Denote for simplicity

$$\Delta_N = \Delta_N(f_1, \dots, f_m), \qquad \delta_N = \delta_N(f_1, \dots, f_m).$$

Then we have

$$\Delta_N(s_1, \dots, s_m) \le \Delta_N \le \Delta_N(S_1, \dots, S_m),$$

and

$$\delta_N(s_1, \dots, s_m) \le \delta_N \le \delta_N(S_1, \dots, S_m).$$

Applying (11) to the step functions

$$s_1, \dots, s_m \quad \text{and} \quad S_1, \dots, S_m$$

we get

$$\prod_{i=1}^m \int_0^1 s_i(x)dF_i(x) \le \underline{\lim}_{N\to\infty} \Delta_{k_N} \le \overline{\lim}_{N\to\infty} \Delta_{k_N} \le \prod_{i=1}^m \int_0^1 S_i(x)dF_i(x).$$

Clearly,

$$\prod_{i=1}^{m} \int_0^1 s_i(x) dF_i(x) \le \lim_{N\to\infty} \delta_{k_N} \le \prod_{i=1}^{m} \int_0^1 S_i(x) dF_i(x).$$

This yields

$$\left| \underline{\lim}_{N\to\infty} \Delta_{k_N} - \lim_{N\to\infty} \delta_{k_N} \right| \le \prod_{i=1}^{m} \int_0^1 S_i(x) dF_i(x) - \prod_{i=1}^{m} \int_0^1 s_i(x) dF_i(x).$$

Analogously,

$$\left| \overline{\lim}_{N\to\infty} \Delta_{k_N} - \lim_{N\to\infty} \delta_{k_N} \right| \le \prod_{i=1}^{m} \int_0^1 S_i(x) dF_i(x) - \prod_{i=1}^{m} \int_0^1 s_i(x) dF_i(x).$$

Taking into account the inequalities (13) and $\varepsilon \to 0^+$ we get

$$\lim_{N\to\infty} \Delta_{k_N} - \delta_{k_N} = 0. \tag{14}$$

Each sequence of natural numbers contains by Helly's selection principle such subsequence κ that the sequences $\{v_1(n)\}, \dots, \{v_m(n)\}$ are κ measurable. Therefore from (14) we get that

$$\overline{\lim}_{N\to\infty} |\Delta_N - \delta_N| = 0. \qquad \square$$

REFERENCES

[1] DRMOTA, M.—TICHY, R. F.: *Sequences, Discrepancies and Applications.* Springer, Berlin, Heidelberg, 1997.

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

[3] 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.

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

[5] HLAWKA, E.: *Theorie der Gleichverteilung.* [Theory of Uniform Distributiion] Bibliographisches Institut, Mannheim, 1979.

[6] KOLOMOGOROV, A. N.—FOMIN, S. V.: *Introductory Real Analysis.* Dover Publications Inc, New York, 1970.

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

[8] 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.

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

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

[11] RAUZY, G.: *Propriétés Statistique de Suites arithmétiques*. Collection SUP: Le Mathématicien, Vol. 15, Presses Universitaires de France, Paris, 1976.

[12] SCHOENBERG, I.: *Über die asymptotische Verteilung reeler Zahlen mod 1*, Math. Z. **28** (1928), 171–199.

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

[14] STRAUCH, O.—PORUBSKÝ, Š.: *Distribution of Sequences a Sampler*. Peter Lang, SAV, Frankfurt am Main, 2005.

[15] WEYL, H.: *Über die Gleichverteilung von Zahlen mod. Eins*, Math. Ann. **77** (1916), no. 3, 313–352.

Received March 11, 2025
Revised March 12, 2025
Accepted March 12, 2025
Publ. online June, 27, 2025

Institute of Mathematics
Slovak Academy of Sciences
Štefánikova 49
SK–814 73 Bratislava
SLOVAKIA
E-mail: milan.pasteka@mat.savba.sk