Acta Academiae Paedagogicae Agriensis, Sectio Mathematicae **30** *(2003) 109–114*

A NOTE ON THE CORRELATION COEFFICIENT OF ARITHMETIC FUNCTIONS

Milan Păstéka, Robert F. Tichy (Bratislava, Slovakia – Graz, Austria)

Dedicated to the memory of Professor Péter Kiss

1. Introduction

The statistical independence was studied by G. Rauzy [9], and later in the papers [3], [5]. We remark that two arithmetical functions F, G with values in $[0,1]$ are called statistically independent if and only if

$$\frac{1}{N}\sum_{n=1}^{N} F(f(n))G(g(n)) - \frac{1}{N^2}\sum_{n=1}^{N} F(f(n))\sum_{n=1}^{N} G(g(n)) \to 0,$$

as $N \to \infty$ for all continuous real valued functions f, g defined on $[0,1]$ (cf. [9]). In the papers [3], [5] a characterization of this type of independence is given in terms of the L^p-discrepancy.

The aim of the present note is to give a "statistical" condition of linear dependence of some type of functions. We consider two polyadically continuous functions f and g. Such functions can be uniformly approximated by the periodic functions (cf. [8]). Let Ω be the space of polyadic integers, constructed as a completion of positive integers with respect to the metric $d(x,y) = \sum_{n=1}^{\infty} \frac{\varphi_n(x-y)}{2^n}$, where $\varphi_n(z) = 0$ if $n|z$ and $\varphi_n(z) = 1$ otherwise, (see the paper [7]). For a survay on the properties of this metric ring we refer also to the monograph [8]. The functions f, g can be extended to uniformly continuous functions $\tilde{f}, \tilde{g}$ defined on Ω. The space Ω is equipped with a Haar probability measure P, thus $\tilde{f}, \tilde{g}$ can be considered as random variables on Ω. Put

$$\tilde{\rho} = \frac{|E(\tilde{f}\cdot\tilde{g}) - E(\tilde{f})\cdot E(\tilde{g})|}{D^2(\tilde{f})\cdot D^2(\tilde{g})},$$

where $E(\cdot)$ is the mean value and $D^2(\cdot)$ is the dispersion (variance) (cf. [1], [10]). The value $\tilde{\rho}$ is called the correlation coefficient of $\tilde{f}, \tilde{g}$, thus if $\tilde{\rho} = 1$ then $\tilde{g} = A\tilde{f} + B$ for some constants A, B. In the following we will prove a similar result for a greater class of functions.

2. Correlation on a set with valuation

Let $\mathbf{M}$ be a set with valuation

$$|\cdot|: \mathbf{M} \to [\mathbf{0}, \infty)$$

such that

(i) The set $\mathbf{M(x)} = \{\mathbf{a} \in \mathbf{M} : |\mathbf{a}| \leq \mathbf{x}\}$ is finite for every $x \in [0, \infty)$,

(ii) If $N(x) = \operatorname{card} \mathbf{M(x)}$, then $N(x) \to \infty$ as $x \to \infty$.

Let $S \subseteq \mathbf{M}$ and put for $x > 0$

$$\gamma_x(S) = \frac{\operatorname{card}(S \cap \mathbf{M(x)})}{N(x)}.$$

Then γ_x is an atomic probability measure with atoms $\mathbf{M(x)}$. If for some $S \subseteq \mathbf{M}$ there exists the limit

$$\lim_{x\to\infty} \gamma_x(S) := \gamma(S), \tag{2.1}$$

then the value $\gamma(S)$ will be called the asymptotic density of S.

If h is a real-valued function defined on $\mathbf{M}$, then it can be considered as a random variable with respect to γ_x for $x > 0$ with mean value

$$E_x(h) := \frac{1}{N(x)} \sum_{|a|\leq x} h(a)$$

and dispersion

$$D_x^2(h) = \frac{1}{N(x)} \sum_{|a|\leq x} (h(a) - E_x(h))^2 = \frac{1}{N(x)} \sum_{|a|\leq x} h^2(a) - (E_x(h))^2$$

(cf. [1]).

Remark. In the case $\mathbf{M} = \mathbf{N}$ (the set of positive integers) we obtain by (2.1) the well known asymptotic density. Various examples of such sets $\mathbf{M}$ with valuations satisfying (i),(ii) are special arithmetical semigroups equipped with absolute value $|\cdot|$ in the sense of Knopfmacher [6].

Let f, g be two real-valued functions defined on $\mathbf{M}$ and $D_x^2(f) > 0, D_x^2(g) > 0$ for sufficiently large x. Consider their correlation coefficient with respect to γ_x given as follows

$$\rho_x = \rho_x(f, g) = \frac{|E_x(f \cdot g) - E_x(f)E_x(g)|}{D_x(f) \cdot D_x(g)}. \tag{2.2}$$

Clearly, if $\rho_x = 1$, then for every $\alpha \in \mathbf{M}(\mathbf{x})$ we have

$$g(\alpha) = A_x f(\alpha) + B_x,$$

where

$$A_x = \frac{E_x(f \cdot g) - E_x(f)E_x(g)}{D_x^2(f)},$$

and

$$B_x = E_x(g) - A_x E_x(f)$$

(cf. [1], [10]).

Note that if $\mathbf{M} = \mathbf{N}$ and f, g are statistically independent arithmetic functions, then

$$\rho_x(f, g) \to 0, x \to \infty.$$

The line $\beta = A_x\alpha + B_x$ is well known as the regression line of f, g on $\mathbf{M}(\mathbf{x})$ (cf. [1], [10]). Consider now the function $g - A_x f$. By some calculations we derive

$$E_x(g - A_x f) = B_x,$$

and

$$D_x^2(g - A_x f) = (1 - \rho_x^2)D_x^2(g),$$

where ρ_x is given by (2.2). Thus from Tchebyschev's inequality we get

$$\gamma_x\left(\{a : |g(a) - A_x f(a) - B_x| \geq \varepsilon\}\right) \leq \frac{(1 - \rho_x^2)D_x^2(g)}{\varepsilon^2}. \tag{2.3}$$

Suppose now that there exist some A, B such that $A_x \to A, B_x \to B$.

We have

$$|g(a) - Af(a) - B| \leq |g(a) - A_x f(a) - B_x| + |f(a)||A_x - A| + |B_x - B|.$$

Thus if f is bounded we obtain for $\varepsilon > 0$ and sufficiently large x

$$|g(a) - Af(a) - B| \geq \varepsilon \Rightarrow |g(a) - A_x f(a) - B_x| \geq \frac{\varepsilon}{2},$$

and so (2.3) yields

$$\gamma_x(\{a : |g(a) - Af(a) - B| \geq \varepsilon\}) \leq \frac{4(1 - \rho_x^2)D_x^2(g)}{\varepsilon^2}. \tag{2.4}$$

Now we can state our main result.

Theorem 1. *Let f, g be two bounded real-valued functions on* $\mathbf{M}$.

(1) Suppose that $D_x^2(f) > 0, D_x^2(g) > 0$ *for sufficiently large* x *and* $A_x \to A, B_x \to B$ *and* $\rho_x \to 1$ *(as* $x \to \infty$*). Then for every* $\varepsilon > 0$

$$\gamma(\{a : |g(a) - Af(a) - B| \geq \varepsilon\}) = 0. \tag{2.5}$$

(2) Let $D_x^2(g) > K > 0$ *for some* K *and assume (2.5) for every* $\varepsilon > 0$ *and suitable constants* A, B*. Then* $\rho_x \to 1$ *(as* $x \to \infty$*).*

Proof. If g is bounded, then also $D_x^2(g)$ is bounded and the assertion (1) follows directly from (2.4).

Put $g_1 := Af + B$. The assumptions of (2) imply that $A \neq 0$ and $D_x^2(f) > K_1 > 0, D_x^2(g_1) > K_2 > 0$ for some constants K_1, K_2. Then we have

$$\rho_x(g_1, f) = 1 \tag{2.6}$$

for each x.

Denote for two bounded real-valued functions h_1, h_2:

$$h_1 \sim h_2 \Longleftrightarrow \gamma(\{a : |h_1(a) - h_2(a)| \geq \varepsilon\}) = 0.$$

It can be verified easily that $\sim$ is an equivalence relation compatible with addition and multiplication, moreover for each uniformly continuous function F it follows from (ii)

$$h_1 \sim h_2 \Rightarrow E_x(F(h_1)) - E_x(F(h_2)) \to 0$$

as $x \to \infty$. In the case (2) we have $g \sim g_1$. This yields

$$D_x^2(g) - D_x^2(g_1) \to 0, x \to \infty, \tag{2.7}$$

but (2.6) gives

$$D_x(g_1)D_x(f) = |E_x(g_1 f) - E_x(g_1)E_x(f)|.$$

Hence, observing that $D_x(f)$ is bounded we obtain from (2.7).

$$D_x(g)D_x(f) - |E_x(g_1 f) - E_x(g_1)E_x(f)| \to 0, x \to \infty\ .$$

Therefore

$$D_x(g)D_x(f) - |E_x(gf) - E_x(g)E_x(f)| \to 0, x \to \infty\ ,$$

and the assertion follows.

The Besicovitch functions. Consider now the case $\mathbf{M} = \mathbf{N}$. An arithmetic function h is called almost periodic if for each $\varepsilon > 0$ there exists a periodic function h_ε such that

$$\overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n \leq N} |h(n) - h_\varepsilon(n)| < \varepsilon.$$

(These functions are also called Besicovitch functions). The class of all such arithmetic functions will be denoted by B^1. For a survey of the properties of B^1 we refer to [8] or [2]. For each $h \in B^1$ there exist the limits

$$\lim_{N\to\infty} E_N(h) := E(h)$$

and

$$\lim_{N\to\infty} D_N^2(h) := D^2(h).$$

If $f, g \in B^1$ are bounded then also $f+g, f \cdot g \in B^1$.

Thus, if $D^2(f), D^2(g) > 0$ then the limits $\lim_{x\to\infty} A_x$, $\lim_{x\to\infty} B_x$ and $\lim_{x\to\infty} \rho_x$ always exist.

The relation $h \sim L$ for an arithmetic function h and some constant L, used in the proof of Theorem 1, is defined in [4] as the statistical convergence of h to L. Šalát [11] gives the following characterisation of the statistical convergence:

Theorem 2. *Let h be an arithmetic function, and L a constant. Then $h \sim L$ if and only if there exists a subset $K \subset \mathbf{N}$ such that the asymptotic density of K is 1 and $\lim_{n\to\infty, n\in K} h(n) = L$.*

Denote by B^2 the set of all Besicovitch functions of h, such that h is bonded and $D^2(h) > 0$. Thus for two functions $f, g \in B^2$ there exists the limit $\rho(f,g) := \lim_{n\to\infty} \rho_N(f,g)$. Theorem 1 and Theorem 2 immediately imply:

Theorem 3. *Let $f, g \in B^2$. Then $\rho(f,g) = 1$ if and only if there exist some constants A, B and a set $K \subset \mathbf{N}$ of asymptotic density 1 such that*

$$\lim_{n\to\infty, n\in K} f(n) - Ag(n) - B = 0.$$

Let us conclude this note by the remarking that the statistical convergence of the real valued function on $\mathbf{M}$ can be characterized analogously as in the paper [11], using the same ideas. Let h be a real valued function on $\mathbf{M}$ and L a real constant. Consider $K \subset \mathbf{M}$, then we write

$$\lim_{a\in K} h(a) = L \Leftrightarrow \forall\varepsilon > 0 \exists x_0 \forall a \in K : |a| > x_0 \Longrightarrow |h(a) - L| < \varepsilon.$$

Theorem 4. *Let h be a real valued function on $\mathbf{M}$ and L a constant. Then $h \sim L$ if and only if there exists a set $K \subset \mathbf{M}$ such that $\gamma(K) = 1$ and $\lim_{a\in K} h(a) = L$.*

Sketch of proof. Put $K_n = \{a \in \mathbf{M} : |h(a) - L| < \frac{1}{n}\}$ for $n \in \mathbf{N}$. Clearly it holds that $\gamma(K_n) = 1, n = 1, 2, \ldots$. Thus it can be selected such an increasing sequence of positive integers $\{x_n\}$ that for $x > x_n$ we have

$$\gamma_x(K_n) > \left(1 - \frac{1}{n}\right), \quad n = 1, 2, \ldots.$$

Put

$$K = \bigcup_{n=1}^{\infty} K_n \cap \Big(M(x_{n+1}) \setminus M(x_n)\Big).$$

Using the fact that the sequence of sets K_n is non increasing it can be proved that $\gamma(K) = 1$, and $\lim_{a\in K} h(a) = L$, by a similary way as in [11].

References

[1] BILLINGSLEY, P., *Measure and Probability*, John Willey.

[2] BESICOVITCH, A. S., *Almost periodic functions,* Dover, New York, 1954.

[3] GRABNER, P. J. and TICHY, R. F., Remarks on statistical independence of sequences, *Math. Slovaca*, **44**, No. 1, (1994), 91–94.

[4] FAST, H., Sur la convergence statistique, *Coll. Math.*, **2** (1951), 241–244.

[5] GRABNER, P. J., STRAUCH, O. and TICHY, R. F., L^p-discrepancy and statistical independence of sequences, *Czech. Math. Journal*, **49**, (124), (1999), 97–110.

[6] KNOPFMACHER, J., *Abstract analytic number theory,* Dover Publications, INC, New York, 1975.

[7] NOVOSELOV, E. V., Topological theory of divisibility, (Russian), *Uchen. Zap. Elabuz.* PI **8** (1960), 3–23.

[8] POSTNIKOV, R. G., *Introduction to analytic number theory,* Nauka, Moscow, 1971, (in Russian), english translation Amer. Math. Soc., Providence, R. I., 1988.

[9] RAUZY, G., *Propriétés statistiques de suites arithmetiques,* Le Mathematicien, **15**, Collection SVP, Presses Universitaires de France, Paris, 1976.

[10] RÉNYI, A., *Wahrscheinlichkeitstheorie,* Teubner Verlag, VEB, Berlin.

[11] ŠALÁT, T., On statistically convergent sequences of real numbers, *Math. Slovaca*, **30**, (1980), No. 2, 139–150.

[12] WAERDEN, V. D. B. L., *Mathematische Statistik,* Springer-Verlag, Berlin–Göttingen–Heidelberg, 1957.

Milan Păstéka
Slovac Academy of Sciences
Stefanikova 49, SK-814 73 Bratislava
Slovakia
e-mail: pasteka@mat.savba.sk

Robert F. Tichy
Institut für Mathematik TU Graz
Steyrergasse 30, A-801 Graz
Austria
e-mail: tichy@tugraz.at