Mathematica Slovaca

Peter J. Grabner; Robert Franz Tichy

Remarks on statistical independence of sequences

Mathematica Slovaca, Vol. 44 (1994), No. 1, 91--94

Persistent URL: http://dml.cz/dmlcz/136601

Math. Slovaca, **44** (1994), No. 1, 91–94

Mathematica Slovaca

REMARKS ON STATISTICAL INDEPENDENCE OF SEQUENCES

PETER J. GRABNER — ROBERT F. TICHY [1]

(*Communicated by Oto Strauch*)

ABSTRACT. We show that an adequate quantitative measure for statistical independence of sequences is the so-called L^2-discrepancy, whereas the usual extremal is not suitable for this purpose.

DEFINITION. *Two sequences x_n, y_n in the unit interval $U = [0,1]$ are called statistically independent if*

$$\lim_{N\to\infty}\left(\frac{1}{N}\sum_{n=1}^{N} f(x_n)g(y_n) - \frac{1}{N^2}\sum_{n=1}^{N} f(x_n)\sum_{n=1}^{N} g(y_n)\right) = 0 \tag{1}$$

for all continuous real functions f, g.

This notion was studied extensively by several authors (cf. [C-L], [Li], [Ra]). Obviously, two sequences x_n, y_n are statistically independent provided that the two-dimensional sequence (x_n, y_n) is uniformly distributed with respect to the measure $\mu_1 \times \mu_2$, where μ_1 and μ_2 are the distributions of x_n, y_n respectively. As a general reference for the theory of uniformly distributed sequences, we give the classical monograph [K-N]. In this case, the limit relation (1) is also true for characteristic functions. This motivates the definition of the extremal discrepancy

$$D_N(x_n, y_n) = \sup_{I,J}\left|\frac{1}{N}\sum_{n=1}^{N}\chi_I(x_n)\chi_J(y_n) - \frac{1}{N^2}\sum_{n=1}^{N}\chi_I(x_n)\sum_{n=1}^{N}\chi_J(y_n)\right|, \tag{2}$$

AMS Subject Classification (1991): Primary 11K06.
Key words: Statistical Independence, Discrepancy, L^2-Discrepancy.
[1] The authors are supported by the Austrian Science Foundation project Nr. P8274-PHY

where the supremum is taken over all intervals $I, J \subset U$.

The following example shows that this discrepancy does not necessarily converge to 0 if the sequences are statistically independent.

Example. Take u_n, a sequence in $I = \left[0, \frac{1}{2}\right)$ converging to $\frac{1}{2}$, and $v_n = 1 - u_n$. Let

$$x_n = \begin{cases} u_k & \text{for } n = 2k\,, \\ v_k & \text{for } n = 2k-1\,, \end{cases}$$

and y_n, any sequence with $y_{2n} \geq \frac{1}{2}$ and $y_{2n-1} < \frac{1}{2}$. Let f and g be continuous functions and $\varepsilon > 0$, arbitrary. Thus, for $n > N_0 = N_0(\varepsilon)$ we have $\left|f(x_n) - f\left(\frac{1}{2}\right)\right| < \varepsilon$. Hence,

$$\begin{aligned}&\left|\frac{1}{N}\sum_{n=1}^{N} f(x_n)g(y_n) - \frac{1}{N^2}\sum_{n=1}^{N} f(x_n)\sum_{n=1}^{N} g(y_n)\right| \\ &\qquad \leq \|f\|_\infty \|g\|_\infty \left(\frac{N_0}{N} + \frac{N_0^2}{N^2}\right) + 2\varepsilon\|g\|_\infty\,.\end{aligned}$$

Letting $N \to \infty$ yields the statistical independence of x_n and y_n.

Now we consider $f = g = \chi_I$.

$$\left|\frac{1}{N}\sum_{n=1}^{N} f(x_n)g(y_n) - \frac{1}{N^2}\sum_{n=1}^{N} f(x_n)\sum_{n=1}^{N} g(y_n)\right| = \frac{1}{2N}\sum_{n=1}^{N} \chi_I(y_n) \to \frac{1}{4}\,.$$

Thus, $D_N \geq \frac{1}{4}$.

As a quantitative measure for statistical independence we suggest the so-called L^2-discrepancy

$$\begin{aligned}&\tilde{D}_N(x_n, y_n)^2 \\ &= \int_0^1\int_0^1 \left(\frac{1}{N}\sum_{n=1}^{N} \chi_{[0,x)}(x_n)\chi_{[0,y)}(y_n) - \frac{1}{N^2}\sum_{n=1}^{N} \chi_{[0,x)}(x_n)\sum_{n=1}^{N} \chi_{[0,y)}(y_n)\right)^2 \mathrm{d}x\,\mathrm{d}y\,.\end{aligned} \tag{3}$$

THEOREM. *The sequences x_n, y_n are statistically independent if and only if*

$$\tilde{D}_N(x_n, y_n) \to 0 \qquad \textit{for} \quad N \to \infty\,.$$

The proof is mainly based on the following proposition:

PROPOSITION.

$$\tilde{D}_N(x_n, y_n)^2 = \frac{1}{16\pi^4} \sum_{\substack{k,l=-\infty \\ k,l\neq 0}}^{\infty} \frac{1}{k^2 l^2} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i(kx_n + ly_n)} - \frac{1}{N^2} \sum_{n=1}^{N} \sum_{m=1}^{N} e^{2\pi i(kx_n + ly_m)} \right|^2 .$$

P r o o f. We expand the function $f(x,y) = \chi_{[0,x)}(x_n)\chi_{[0,y)}(y_m)$ into its two-dimensional Fourier series and obtain for the coefficients

$$a_{k,l} = \int_0^1 \int_0^1 f(x,y)\, e^{2\pi i(kx+ly)}\, dx\, dy = -\frac{1}{4\pi^2 kl}\left(1 - e^{2k\pi i x_n}\right)\left(1 - e^{2l\pi i y_m}\right)$$

for $k, l \neq 0$. Applying Parseval's equation in (3) and observing that all terms only depending on one variable cancel out yield the desired result.

The proof of the theorem follows immediately from the proposition applying that any continuous function can be uniformly approximated by trigonometric polynomials. This is essentially the crucial point for the proof of Weyl's criterion in the theory of uniformly distributed sequences. Of course, this criterion is also true for statistical independence.

R e m a r k 1. Obviously, two sequences are statistically independent if $D_N \to 0$.

R e m a r k 2. The extremal discrepancy (2) satisfies a law of iterated logarithm (for almost all sequences in the sense of product measure). This can be shown, applying the general method of P h i l i p p [Ph].

P r o b l e m . Define for $1 < p < \infty$ the L^p-discrepancy of two sequences by

$$D_N^{(p)}(x_n, y_n)^p = \int_0^1 \int_0^1 \left(\frac{1}{N} \sum_{n=1}^{N} \chi_{[0,x)}(x_n)\chi_{[0,y)}(y_n) - \frac{1}{N^2} \sum_{n=1}^{N} \chi_{[0,x)}(x_n) \sum_{n=1}^{N} \chi_{[0,y)}(y_n) \right)^p dx\, dy .$$

Are the sequences x_n, y_n statistically independent if and only if $D_N^{(p)} \to 0$ for $N \to \infty$?

REFERENCES

[C-L] COQUET, J.—LIARDET, P.: *Répartitions uniformes des suites et indépendance statistique*, Compositio Math. **51** (1984), 215–236.

[K-N] KUIPERS, L.—NIEDERREITER, H.: *Uniform Distribution of Sequences*, J. Wiley and Sons, New York, 1974.

[Li] LIARDET, P.: *Some metric properties of subsequences*, Acta Arith. **55** (1990), 119–135.

[Ph] PHILIPP, W.: *Mixing Sequences of Random Variables and Probabilistic Number Theory. Mem. Amer. Math. Soc. 114*, Amer. Math. Soc., Providence R.I., 1971.

[Ra] RAUZY, G.: *Propriétés statistique de suites arithmétiques. Le Mathématicien, No. 15. Collection SUP*, Presses Universitaires de France, Paris, 1976.

Received October 19, 1992

Institut für Mathematik
Technische Universität Graz
Steyrergasse 30
A-8010 Graz
Austria