Some applications of distribution functions of sequences

Oto Strauch, Bratislava

Contents

In this lecture we characterize some known classes of sequences x_n, originally defined by some properties of x_n, by using the set $G(x_n)$ of all distribution functions of x_n, :

- (λ, λ')-distribution.
- Statistically independent sequences.
- Statistically convergent sequences.
- Statistical limit points.
- Uniform maldistributed sequences.
- $\xi(3/2)^n \bmod 1$, $n = 1, 2, \dots$

Then we give some new results:

- Benford's law.
- Copulas.

Definitions

In the uniform distribution theory a random variable is replaced by

- a sequence x_n, $n = 1, 2, \ldots$, $x_n \in [0, 1)$.

- Define step d.f. of x_n as

$$F_N(x) = \frac{\#\{n \le N; x_n \in [0, x)\}}{N}.$$

- A function $g : [0, 1] \to [0, 1]$ is d.f. of x_n if there exists a sequence of indices $N_1 < N_2 < \ldots$ such that $F_{N_k}(x) \to g(x)$ for all continuity points x of $g(x)$ as $k \to \infty$.

- The set of all such $g(x)$ we shall denote by $G(x_n)$ and the notion of the distribution of x_n we shall identify with $G(x_n)$, i.e. the distribution of x_n is known if we know the set $G(x_n)$.

Example

Let $\{x\}$ be the fractional part of x. For $x_n = \{\log n\}$, $n = 1, 2, \dots,$ we have the set of d.f.s

$$G(x_n) = \left\{ g_u(x) = \frac{e^{\min(x,u)} - 1}{e^u} + \frac{1}{e^u}\frac{e^x - 1}{e - 1}; u \in [0,1] \right\},$$

and $\{\log N_k\} \to u$ implies $F_{N_k}(x) \to g_u(x)$. The lower and upper d.f. of $\log n \bmod 1$ are

$$\underline{g}(x) = \frac{e^x - 1}{e - 1}, \qquad \overline{g}(x) = \frac{1 - e^{-x}}{1 - e^{-1}},$$

and $\underline{g} \in G(x_n)$ but $\overline{g} \notin G(x_n)$. This set $G(x_n)$ was found by A. Wintner [1935].

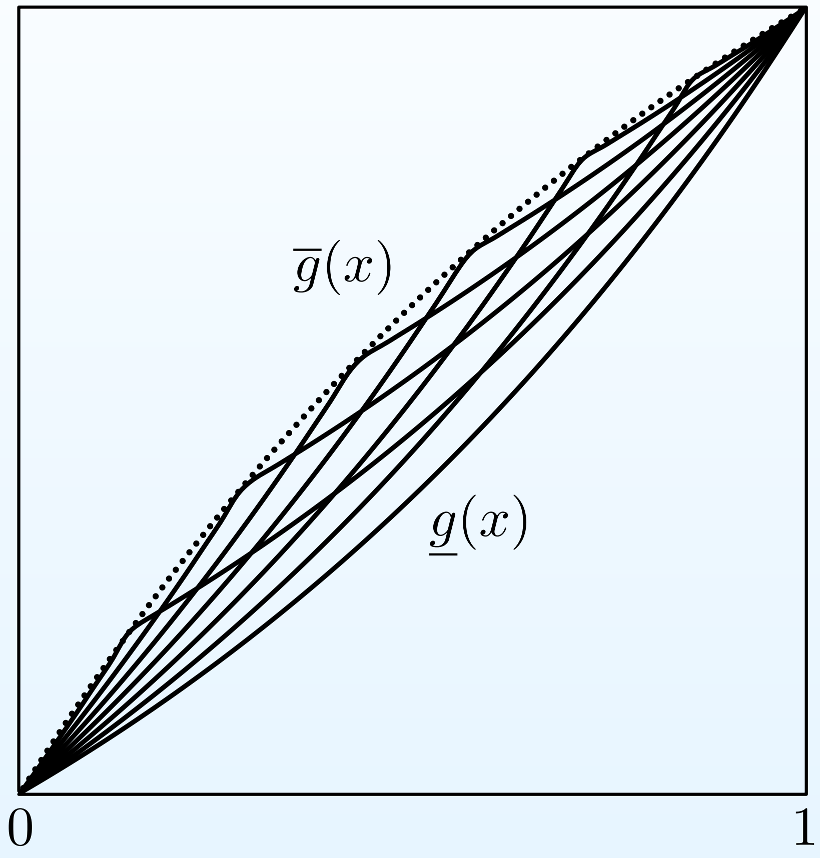


Figure 1: Distribution functions of $\log n \mod 1$

(λ, λ')-distribution

J. Chauvineau [1968] introduced:

Definition. Let λ and λ' be two real numbers such that $0 < \lambda \le 1 \le \lambda'$. The sequence $x_n \in [0, 1)$ is said to be (λ, λ')–*distributed* if, for every non–empty proper subinterval $I \subset [0, 1]$, we have both

(i) $\liminf_{N\to\infty} \frac{\#\{n\le N; x_n\in I\}}{N} \ge \lambda|I|$, and

(ii) $\limsup_{N\to\infty} \frac{\#\{n\le N; x_n\in I\}}{N} \le \lambda'|I|$.

Theorem. A sequence $x_n \bmod 1$ is (λ, λ')–distributed if and only if every $g(x) \in G(x_n \bmod 1)$ has the lower derivative $\ge \lambda$ and the upper derivative $\le \lambda'$ at every point $x \in (0, 1)$.

Example

In G. Pólya and G. Szegõ [1964, Part 2, Ex. 179] it is proved that the derivative (density) $g'(x)$ of any $g(x) \in G(c \log n \bmod 1)$, $c > 0$, has the form

$$g'(x) = \begin{cases} \dfrac{\log q}{q-1}\, q^{x-\alpha+1}, & \text{if } 0 \leq x < \alpha, \\ \dfrac{\log q}{q-1}\, q^{x-\alpha}, & \text{if } \alpha < x \leq 1, \end{cases}$$

where $q = e^{1/c}$ and $\alpha \in (0, 1)$. If $\alpha = 0$ or $\alpha = 1$ then

$$g'(x) = \frac{\log q}{q-1}\, q^{x}$$

and $c \log n \bmod 1$ is (λ, λ')–distributed with $\lambda = \frac{\log q}{q-1}$ and $\lambda' = q\frac{\log q}{q-1}$.

Statistically independent sequences

G. Rauzy [1976, p. 91, 4.1. Def.]:

Definition. Let x_n and y_n be two infinite sequences from the unit interval $[0,1)$. The pair of sequences (x_n, y_n) is called *statistically independent* if

$$\lim_{N\to\infty}\left(\frac{1}{N}\sum_{n=1}^{N} f_1(x_n)f_2(y_n) - \left(\frac{1}{N}\sum_{n=1}^{N} f_1(x_n)\right)\left(\frac{1}{N}\sum_{n=1}^{N} f_2(y_n)\right)\right) = 0$$

for all continuous real functions f_1, f_2 defined on $[0,1]$.

Theorem[G. Rauzy [1976, p. 92, 4.2. par.]] For arbitrary $(x_n, y_n) \in [0,1)^2$, $n = 1, 2, \dots,$ we have (x_n, y_n) is statistically independent if and only if

$$\forall_{g\in G(x_n,y_n)} \; g(x,y) = g(x,1)g(1,y) \quad \text{a.e. on } [0,1]^2.$$

Sketch of proof

For given two-dimensional sequence (x_n, y_n) put

$$F_N(x,y) = \frac{\#\{n \le N; (x_n, y_n) \in [0,x) \times [0,y)\}}{N}$$

and assume that $F_{N_k}(x,y) \to g(x,y)$ for some $N_1 < N_2 < \ldots$. By Riemann-Stieltjes integration and Helly theorem,

$$\frac{1}{N_k}\sum_{n=1}^{N_k} f_1(x_n)f_2(y_n) = \int_0^1\int_0^1 f_1(x)f_2(y)\mathrm{d}F_{N_k}(x,y) \to \int_0^1\int_0^1 f_1(x)f_2(x)\mathrm{d}g(x,y),$$

$$\frac{1}{N_k}\sum_{n=1}^{N_k} f_1(x_n) = \int_0^1 f_1(x)\mathrm{d}F_{N_k}(x,1) \to \int_0^1 f_1(x)\mathrm{d}g(x,1),$$

$$\frac{1}{N_k}\sum_{n=1}^{N_k} f_2(y_n) = \int_0^1 f_2(y)\mathrm{d}F_{N_k}(1,y) \to \int_0^1 f_2(y)\mathrm{d}g(1,y)$$

as $k \to \infty$.

Assuming statistical independence x_n and y_n we have

$$\int_0^1 \int_0^1 f_1(x) f_2(x) \mathsf{d} g(x,y) = \left(\int_0^1 f_1(x) \mathsf{d} g(x,1) \right) \left(\int_0^1 f_2(y) \mathsf{d} g(1,y) \right).$$

Using per-partes we have

$$\int_0^1 \int_0^1 g(x,y) \mathsf{d} f_1(x) \mathsf{d} f_2(y) = \left(\int_0^1 g(x,1) \mathsf{d} f_1(x) \right) \left(\int_0^1 g(1,y) \mathsf{d} f_2(y) \right) \tag{1}$$

for arbitrary differentiable $f_1(x)$ and $f_2(y)$. Now, for a continuity point (x_0, y_0) of $g(x,y)$ we can select $f_1(x)$ and $f_2(y)$ such that (1) implies $g(x_0, y_0) = g(x_0, 1) g(1, y_0)$.

Example

J. Coquet and P. Liardet [1987]: Given an integer $q \geq 2$, a real number θ and a real polynomial $p(x)$, let

(i) $x_n = \theta q^n \bmod 1$,
(ii) $y_n = p(n) \bmod 1$,
(iii) $\mathbf{x}_n = (x_{n+1}, \dots, x_{n+s})$ and $\mathbf{y}_n = (y_{n+1}, \dots, y_{n+s})$.

If x_n is u.d. (i.e. θ is normal in the base q), then for every $s = 1, 2, \dots,$ the sequence

$$(\mathbf{x}_n, \mathbf{y}_n), \quad n = 1, 2, \dots,$$

has d.f.s $g(\mathbf{x}, \mathbf{y}) \in G((\mathbf{x}_n, \mathbf{y}_n))$ of the form $g(\mathbf{x}, \mathbf{y}) = g_1(\mathbf{x}) g_2(\mathbf{y})$ for some $g_1(\mathbf{x}) \in G(\mathbf{x}_n)$ and $g_2(\mathbf{y}) \in G(\mathbf{y}_n)$, i.e. the sequences x_n and y_n are *completely statistically independent*.

Notes

Grabner and Tichy [1994] proved that the extremal discrepancy

$$\sup_{x,y\in[0,1]} |F_N(x,y) - F_N(x,1)F_N(1,y)|$$

does not characterize statistical independence, but the L^2-discrepancy

$$\int_0^1\int_0^1 (F_N(x,y) - F_N(x,1)F_N(1,y))^2 \mathsf{d}x\mathsf{d}y$$

provides a characterization. L^2-discrepancy of statistical independence of x_n and y_n can be computed as

$$\int_{\mathbf{X}}\int_{\mathbf{X}} \left(\frac{1}{N}\sum_{n=1}^{N} f(x_n)g(y_n) - \frac{1}{N}\sum_{n=1}^{N} f(x_n)\frac{1}{N}\sum_{n=1}^{N} g(x_n)\right)^2 \mathsf{d}f\mathsf{d}g,$$

see O. Strauch [1994] with the classical Wiener sheet measure $\mathsf{d}f$.

A statistical limit

H. Fast [1951] and I.J. Schoenberg [1959] defined, independently:

Definition. The sequence x_n is said to be *statistically convergent* to the number α provided that for each $\varepsilon > 0$,

$$\lim_{N\to\infty} \frac{1}{N} \#\{n \le N; |x_n - \alpha| \ge \varepsilon\} = 0.$$

Fast [1951] mentioned: A sequence x_n is statistically convergent to α if and only if there exists a sequence of indices k_n of the asymptotic density $d(k_n) = 1$ such that $\lim_{n\to\infty} x_{k_n} = \alpha$ in the standard sense.

Let us consider *one-jump function* $c_\alpha(x)$ which has a jump of size 1 for α.

I.J. Schoenberg [1959] mentioned: The sequence $x_n \in [0,1)$ is statistically convergent to the number $\alpha \in [0,1]$ if and only if the sequence x_n admits the asymptotic distribution function $c_\alpha(x)$.

Criterion

Theorem[O. Strauch [1995]]. The sequence $x_n \in [0,1)$ possesses a statistical limit if and only if

$$\lim_{M,N\to\infty} \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} |x_m - x_n| = 0.$$

Sketch of proof: Let $F_{M_{i_k}}(x) \to g_1(x)$ and $F_{N_{j_k}}(x) \to g_2(x)$. Applying the Helly-Bray lemma we find

$$\lim_{k\to\infty} \int_0^1 \int_0^1 |x-y| \mathrm{d}F_{M_{i_k}}(x) \mathrm{d}F_{N_{j_k}}(y) = \int_0^1 \int_0^1 |x-y| \mathrm{d}g_1(x) \mathrm{d}g_2(y).$$

Using Riemann-Stieltjes integration we obtain

$$\int_0^1 \int_0^1 |x-y| \mathrm{d}F_{M_{i_k}}(x) \mathrm{d}F_{N_{j_k}}(y) = \frac{1}{M_{i_k} N_{j_k}} \sum_{m=1}^{M_{i_k}} \sum_{n=1}^{N_{j_k}} |x_m - x_n|.$$

Thus

$$\int_0^1 \int_0^1 |x - y| \mathrm{d}g_1(x) \mathrm{d}g_2(y) = 0$$

which gives $g_1(x) = g_2(x) = c_\alpha(x)$ a.e. for some $\alpha \in [0, 1]$.

Statistical limit points

Following the concept od statistical convergence J. A. Fridy [1993] introduced:

Definition. A real number x is said to be *a statistical limit point* of the sequence x_n if there exists a subsequence x_{k_n}, $n = 1, 2, \ldots$, such that $\lim_{n\to\infty} x_{k_n} = x$ and the set of indices k_n has a positive upper asymptotic density.
Fridy studied the set $\Lambda(x_n)$ of all such points. Inspired by I.J. Schoenberg [1959], P. Kostyrko, M. Mačaj, T. Šalát and O. Strauch [2001] was found

Theorem. The set $\Lambda(x_n)$, for $x_n \in [0, 1)$ $n = 1, 2, \ldots$, coincides with the set of all discontinuity points of d.f.s $g(x) \in G(x_n)$.

Example

Let $\alpha = \frac{p}{q}\pi$, where p and q are positive integers and g.c.d. $(p,q) = 1$. It is proved in D. Berend, M. D. Boshernitzan, and G. Kolesnik [1995] that the sequence

$$x_n = n\cos(n\cos n\alpha) \bmod 1, \quad n = 1, 2, \dots$$

has $G(x_n) = \{g(x)\}$, where

$$g(x) = \begin{cases} x & \text{if } q \text{ is odd },\\ \left(1 - \frac{1}{q}\right)x + \frac{1}{q}c_0(x) & \text{if } q \text{ is even },\end{cases}$$

and $c_0(x)$ is the one-jump d.f. with the jump 1 in 0. This implies

$$\Lambda(x_n) = \begin{cases} \emptyset & \text{if } q \text{ is odd },\\ \{0\} & \text{if } q \text{ is even }.\end{cases}$$

Uniform maldistributed sequences

G. Myerson [1991] introduced:

Definition. The sequence $x_n \in [0,1)$ $n = 1, 2, \ldots,$ is said to be *uniformly maldistributed* (u.m.) if for every nonempty proper subinterval $I \subset [0,1]$ we have both

$$\liminf_{N\to\infty} \frac{1}{N}\#\{n \le N; x_n \in I\} = 0 \text{ and } \limsup_{N\to\infty} \frac{1}{N}\#\{n \le N; x_n \in I\} = 1.$$

He mentioned that the first condition is superfluous, and showed

Example. The sequence $x_n = \{\log\log n\}$ of fractional parts of the iterated logarithm is u.m.
O. Strauch [1995] proved: Let $c_\alpha(x)$ be one-step d.f. for which $c_\alpha(x) = 0$ for $x \in [0, \alpha]$ and $c_\alpha(x) = 1$ for $x \in (\alpha, 1]$.

Theorem. The sequence x_n is u.m. if and only if

$$\{c_\alpha(x); \alpha \in [0,1]\} \subset G(x_n).$$

Example

Starting with $x_n = \{\log\log n\}$ all the sequences $\{\log\log\dots\log n\}$, $n = n_0, n_0+1, \dots,$ have

$$G(x_n) = \{c_\alpha(x); \alpha \in [0,1]\} \cup \{h_\alpha(x); \alpha \in [0,1]\}$$

and thus are u.m. Here $h_\alpha(x)$ is a constant d.f. for which $h_\alpha(x) = \alpha$ for $x \in (0,1)$, $h_\alpha(0) = 0$ and $h_\alpha(1) = 1$.

Thus, in the theory of uniform maldistribution we need not consider d.f.s other than one-jump d.f. $c_\alpha(x)$ which has a jump of size 1 at α. This suggests the definition (see Strauch [1995]):

Definition. The sequence x_n is said to be *uniformly maldistributed in the strict sense* (u.m.s.) if $G(x_n) = \{c_\alpha(x); \alpha \in [0,1]\}$.

Criterion

Theorem[O. Strauch [1995]]. For every sequence $x_n \in [0,1)$ we have

$$G(x_n) \subset \{c_\alpha(x); \alpha \in [0,1]\} \Longleftrightarrow \lim_{N\to\infty} \frac{1}{N^2} \sum_{m,n=1}^{N} |x_m - x_n| = 0.$$

Moreover, if $G(x_n) \subset \{c_\alpha(x); \alpha \in [0,1]\}$, then $G(x_n) = \{c_\alpha(x); \alpha \in I\}$, where I is a closed subinterval of $[0,1]$ which can be found as

$$I = \left[\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x_n,\ \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x_n \right],$$

and the length $|I|$ of I can also be found as

$$|I| = \limsup_{M,N\to\infty} \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} |x_m - x_n|.$$

Criterion

Theorem. The sequence $x_n \in [0,1)$ is u.m.s. if and only if

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{m,n=1}^{N} |x_m - x_n| = 0 \text{ and } \limsup_{M,N\to\infty} \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} |x_m - x_n| = 1,$$

or alternatively $\limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} x_n - \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} x_n = 1.$

Example. Let x_n, $n = 1, 2, \dots$ be defined as

$$x_n = \left\{ 1 + (-1)^{\left[\sqrt{[\sqrt{\log_2 n}]}\right]} \left\{ \sqrt{\left[\sqrt{\log_2 n}\right]} \right\} \right\},$$

were $[x]$ denotes the integral part and $\{x\}$ the fractional part of x. Then $G(x_n) = \{c_\alpha(x); \alpha \in [0,1]\}$.

$\xi(3/2)^n \bmod 1, n = 1, 2, \ldots$

We present a selection of known conjectures:

(i) $(3/2)^n \bmod 1$ is uniformly distributed in $[0, 1]$.

(ii) $(3/2)^n \bmod 1$ is dense in $[0, 1]$.

(iii) (K. Mahler [1968]) There exists no $\xi \in \mathbb{R}^+$ such that $0 \leq \{\xi(3/2)^n\} < 1/2$ for $n = 0, 1, 2, \ldots$.

Few positive results are known, for instance:

(iv) L. Flatto, J. C. Lagarias and A. D. Pollington [1995] stated that

$$\limsup_{n\to\infty}\{\xi(3/2)^n\} - \liminf_{n\to\infty}\{\xi(3/2)^n\} \geq \frac{1}{3}$$

for every $\xi > 0$.

(v) G. Choquet [1980] gave infinitely many $\xi \in \mathbb{R}$ for which

$$\frac{1}{19} \leq \{\xi(3/2)^n\} \leq 1 - \frac{1}{19} \text{ for } n = 0, 1, 2, \ldots.$$

$G(\xi(3/2)^n \bmod 1)$

Some conjectures involving a d.f. $g(x)$ of $\xi(3/2)^n \bmod 1$ may be formulated strongly as in (i)–(iii). For example, the following conjecture implies Mahler's conjecture:

(vi) If $g(x) =$ constant for all $x \in I$, where I is a subinterval of $[0,1]$, then the length $|I| < 1/2$.

Since $\{2\{\xi(3/2)^n\}\} = \{3\{\xi(3/2)^{n-1}\}\}$ then for

$$f(x) = 2x \bmod 1, \text{ and } h(x) = 3x \bmod 1$$

we have $f(\{\xi(3/2)^n\})$ and $h(\{\xi(3/2)^{n-1}\})$ are the same sequences and thus we have

Theorem. For every $g(x) \in G(\xi(3/2)^n \bmod 1)$ and $x \in [0,1]$ we have

$$g(x/2) + g((x+1)/2) - g(1/2) =$$
$$g(x/3) + g((x+1)/3) + g((x+2)/3) - g(1/3) - g(2/3),$$

This functional equation we shall write as $g_f(x) = g_h(x)$.

Results

Theorem. Let g_1, g_2 be any two distribution functions satisfying $g_{i_f}(x) = g_{i_h}(x)$ for $i = 1, 2$ and $x \in [0, 1]$. Denote

$$I_1 = [0, 1/3], \quad I_2 = [1/3, 2/3], \quad I_3 = [2/3, 1].$$

If $g_1(x) = g_2(x)$ for $x \in I_i \cup I_j$, $1 \leq i \neq j \leq 3$, then $g_1(x) = g_2(x)$ for all $x \in [0, 1]$.

Theorem. Denote

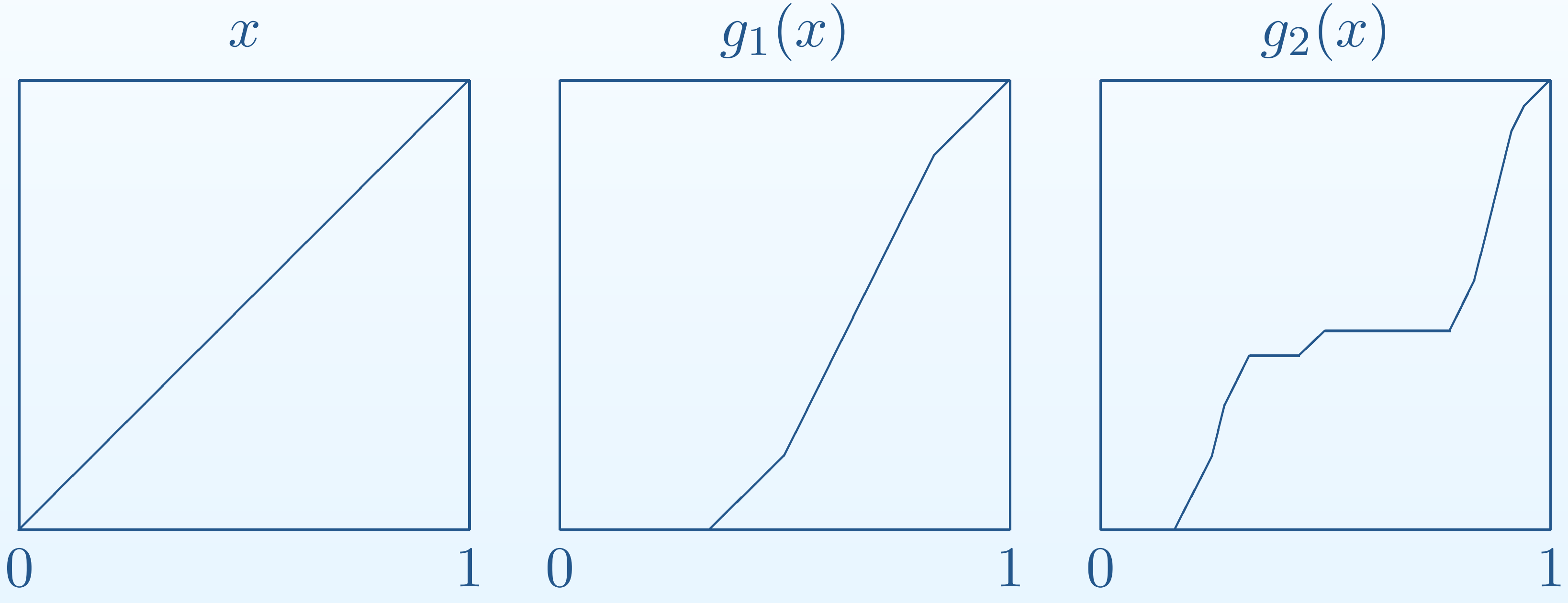
$$F(x, y) = |\{2x\} - \{3y\}| + |\{2y\} - \{3x\}| - |\{2x\} - \{2y\}| - |\{3x\} - \{3y\}|.$$

The continuous distribution function $g(x)$ satisfies $g_f(x) = g_h(x)$ for $x \in [0, 1]$ if and only if

$$\int_0^1 \int_0^1 F(x, y) \mathrm{d}g(x) \mathrm{d}g(y) = 0.$$

Example

In O. Strauch [1997, Th. 4] there is described a method that starting with a given absolutely continuous solution $g_f(x) = g_h(x)$ we derive a new solution $g_{1_f}(x) = g_{1_h}(x)$ as follows



Benford's law

The first digit problem:

Definition. An infinite sequence $x_n \geq 1$ of real numbers satisfies *Benford's law*, if the frequency (the asymptotic density) of occurrences of a given first digit a, when x_n is expressed in the decimal form is given by $\log_{10}\left(1 + \frac{1}{a}\right)$ for every $a = 1, 2, \ldots, 9$ (0 as a first possible digit is not admitted).

It was S. Newcomb (1881) who firstly noted "*That the ten digits do not occur with equal frequency must be evident to anyone making use of logarithm tables*".

F. Benford (1938) compared the empirical frequency of occurrences of a with $\log_{10}((a+1)/a)$ in twenty different tables having lengths running from 91 entries (atomic weights) to 5000 entries in a mathematical handbook which led him to the conclusion that "the logarithmic law applies particularly to those outlaw numbers that are without known relationships ..."

Criterions

Since x_n has the first digit a if and only if

$$\log_{10} x_n \bmod 1 \in [\log_{10} a, \log_{10}(a+1)),$$

Benford's law for x_n follows from the uniform distribution (u.d.) of $\log_{10} x_n \bmod 1$.

Definition. A sequence x_n is defined to satisfy strong Benford's law (strong B.L.) if for every block of digits $a_1 a_2 \cdots a_r$ the density of x_n having this initial block $a_1 a_2 \cdots a_r$ is $\log_{10}\left(1 + \frac{1}{a_1 a_2 \cdots a_r}\right)$ (initial zero digits are omitted).

Theorem. A sequence x_n, $x_n > 0$, $n = 1, 2, \ldots$, satisfies strong B.L. in the base b if and only if the sequence $\log_b x_n \bmod 1$ is u.d. in $[0, 1)$.

In the following we study d.f.s of a sequence $x_n \in (0, 1)$ which satisfy strong B.L.

Theorem

V. Baláž, K. Nagasaka and O. Strauch [2010] proved

Theorem. Let x_n, $n = 1, 2, \ldots,$ be a sequence in $(0, 1)$ and $G(x_n)$ be the set of all d.f.s of x_n. Assume that every d.f. $g(x) \in G(x_n)$ is continuous at $x = 0$. Then the sequence x_n satisfies strong B.L. in the base b if and only if for every $g(x) \in G(x_n)$ we have

$$x = \sum_{i=0}^{\infty} \left(g\left(\frac{1}{b^i}\right) - g\left(\frac{1}{b^{i+x}}\right)\right) \text{ for } x \in [0, 1]. \qquad (2)$$

Sketch of proof:
For ith inverse $f_i^{-1}(x)$ of $-\log_b x \bmod 1$ we have

$$f_i^{-1}([0, x)) = \left(\frac{1}{b^{i+x}}, \frac{1}{b^i}\right].$$

Examples

We present some solutions of (2)

$$g(x) = \begin{cases} x & \text{if } x \in \left[0, \frac{1}{b}\right], \\ 1 + \frac{\log x}{\log b} + (1-x)\frac{1}{b-1} & \text{if } x \in \left[\frac{1}{b}, 1\right]. \end{cases}$$

$$g^*(x) = \begin{cases} 0 & \text{if } x \in \left[0, \frac{1}{b^2}\right], \\ 2 + \frac{\log x}{\log b} & \text{if } x \in \left[\frac{1}{b^2}, \frac{1}{b}\right], \\ 1 & \text{if } x \in \left[\frac{1}{b}, 1\right] \end{cases}$$

$$g^{**}(x) = \begin{cases} 0 & \text{if } x \in \left[0, \frac{1}{b^3}\right], \\ 3 + \frac{\log x}{\log b} & \text{if } x \in \left[\frac{1}{b^3}, \frac{1}{b^2}\right], \\ 1 & \text{if } x \in \left[\frac{1}{b^2}, 1\right] \end{cases}$$

If H is the set of all $t_1 g(x) + t_2 g^*(x) + t_3 g^{**}(x)$, $t_1 + t_2 + t_3 = 1$, $t_1, t_2, t_3 \geq 0$, then there exists a sequence x_n such that $G(x_n) = H$, and this x_n satisfies strong B.L.

Theorem

V. Baláž, K. Nagasaka and O. Strauch [2010] proved

Theorem. For a sequence $x_n \in (0,1)$, $n = 1, 2, \dots$, assume that every d.f. $g(x) \in G(x_n)$ is continuous at $x = 0$. Then there exist only finitely many different integer bases b for which the sequence x_n satisfies strong B.L. simultaneously. Moreover, if the sequence x_n satisfies strong B.L. in base b, and for some $k = 1, 2, \dots$ there exists kth integer root $\sqrt[k]{b}$, then x_n satisfies strong B.L. also in base $\sqrt[k]{b}$.

For an integer sequence it does not hold:

Example. In the Sampler of O. Strauch and Š. Porubský [2005, p. 2–117, 2.12.14] we see that the sequence

$$\alpha n \log^\tau n \bmod 1, \quad \alpha \neq 0, 0 < \tau \leq 1,$$

is u.d. From this it follows that $x_n = n^n$ satisfies strong B.L. for an arbitrary integer base b, because $\log_b n^n = n \log n \frac{1}{\log b}$.

Example

As it is well known that the increasing sequence of all positive integers $1, 2, 3, \ldots$ does not satisfy strong B.L. (simple B.L. also) in every base $b \geq 2$. It follows from the fact that $\log_b n \bmod 1$ is not u.d. For a density of n for which r initial digits are $K = k_1 k_2 \ldots k_r$, A.I. Pavlov [1981] proved that

$$\liminf_{N\to\infty} \frac{\#\{n \leq N; n \text{ has the first } r \text{ digits } = K\}}{N} = \frac{1}{K(b-1)},$$

$$\limsup_{N\to\infty} \frac{\#\{n \leq N; n \text{ has the first } r \text{ digits } = K\}}{N} = \frac{b}{(K+1)(b-1)}.$$

We give the following extension:

By G. Pólya and G. Szegö [1964] the d.f.s of $\log_b n \mod 1$ is of the form

$$g_u(x) = \frac{b^{\min(x,u)} - 1}{b^u} + \frac{1}{b^u}\frac{b^x - 1}{b - 1},$$

where the parameter u runs $[0, 1]$. By R. Giuliano and O. Strauch [2008], for increasing sequence N_i, $i = 1, 2, \dots$, we have

$$\log_b N_i \bmod 1 \to u \Longrightarrow F_{N_i}(x) \to g_u(x)$$

and thus

$$\frac{\#\{n \le N_i; n \text{ has the first } r \text{ digits } = K\}}{N_i} \to (g_u(x_2) - g_u(x_1))$$

as $i \to \infty$, $x_1 = \log_b(k_1.k_2k_3\dots k_r)$, $x_2 = \log_b(k_1.k_2k_3\dots(k_r + 1))$, and the minimum appears in $u = x_1$ and maximum in $u = x_2$.

Copulas

by F. Pillichshammer and S. Steinerberger [2009]. They proved:

Theorem. Let x_n and y_n be two uniformly distributed sequences in $[0,1)$. Then

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |x_n - y_n| \le \frac{1}{2}$$

and in particular $\limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |x_{n+1} - x_n| \le \frac{1}{2}$ and this result is best possible.

They also found

Example.
$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |x_{n+1} - x_n| = \frac{2(b-1)}{b^2}$ for van der Corput sequence x_n in the base b and
$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |x_{n+1} - x_n| = 2\{\alpha\}(1-\{\alpha\})$ for $x_n = n\alpha \bmod 1$, where α is irrational.

Generalization

Let $F(x,y)$ be a Riemann integrable function defined on $[0,1]^2$ and x_n, y_n, $n=1,2,\dots$, be two u.d. sequences in $[0,1)$. A problem is to find limit points of the sequence

$$\frac{1}{N}\sum_{n=1}^{N} F(x_n,y_n), \qquad N=1,2,\dots \tag{1}$$

Applying Helly theorems we have that limit points of (1) form the set

$$\left\{\int_0^1\int_0^1 F(x,y)\mathrm{d}_x\mathrm{d}_y g(x,y); g(x,y)\in G((x_n,y_n))\right\}, \tag{2}$$

where $G(x_n,y_n)$ is the set of all d.f.s of the two-dimensional sequence (x_n,y_n), $n=1,2,\dots$

Definition

If x_n and y_n are u.d. sequences, then two-dimensional sequence (x_n, y_n) doesn't need to be u.d. but every d.f. $g(x,y) \in G((x_n, y_n))$ satisfies
(i) $g(x,1) = x$ for $x \in [0,1]$ and
(ii) $g(1,y) = y$ for $y \in [0,1]$.

The d.f. $g(x,y)$ which satisfies (i) and (ii) is called *copula* and a basic theory of copulas can be found in R.B. Nelsen [1999].

Open problem is to find extreme values of

$$\int_0^1 \int_0^1 F(x,y) \mathrm{d}_x \mathrm{d}_y g(x,y)$$

over copulas $g(x,y)$.

Results

J. Fialová and O. Strauch [2010] proved:

Theorem. Let $F(x,y)$ be a Riemann integrable function defined on $[0,1]^2$. For differential of $F(x,y)$ let us assume that $\mathsf{d}_x\mathsf{d}_y F(x,y) > 0$ for every $(x,y) \in [0,1]^2$. Then

$$\max_{g(x,y)\text{-copula}} \int_0^1\int_0^1 F(x,y)\mathsf{d}_x\mathsf{d}_y g(x,y) = \int_0^1 F(x,x)\mathsf{d}x,$$

$$\min_{g(x,y)\text{-copula}} \int_0^1\int_0^1 F(x,y)\mathsf{d}_x\mathsf{d}_y g(x,y) = \int_0^1 F(x,1-x)\mathsf{d}x,$$

where, precisely, $\max$ is attained in $g(x,y) = \min(x,y)$ and $\min$ in $g(x,y) = \max(x+y-1,0)$, uniquely.

Shortened proof

In proof is used

$$\int_0^1\int_0^1 F(x,y)\mathsf{d}_x\mathsf{d}_y g(x,y) = F(1,1) - \int_0^1 g(1,y)\mathsf{d}_y F(1,y)$$
$$- \int_0^1 g(x,1)\mathsf{d}_x F(x,1) + \int_0^1\int_0^1 g(x,y)\mathsf{d}_x\mathsf{d}_y F(x,y)$$

which holds for every Riemann integrable $F(x,y)$ and d.f. $g(x,y)$ which doesn't have any common discontinuity points. And then Fréchet-Hoeffding bounds (see Nelsen [1999, p. 9])

$$\max(x+y-1,0) \leq g(x,y) \leq \min(x,y)$$

which holds for every $(x,y) \in [0,1]^2$ and for every copula $g(x,y)$.

Example

For $y = x$, $\mathsf{d}y = \mathsf{d}x$ we have $\mathsf{d}_x\mathsf{d}_y|y-x| = -2\mathsf{d}x$ then we have

$$\int_0^1\int_0^1 |x-y|\mathsf{d}_x\mathsf{d}_y g(x,y) = 1 - 2\int_0^1 g(x,x)\mathsf{d}x.$$

For van der Corput sequence $\gamma_q(n)$, $n = 0, 1, \dots,$ in base q every point $(\gamma_q(n), \gamma_q(n+1))$, $n = 0, 1, 2, \dots,$ lie on line segments $Y = X - 1 + 1/q^k + 1/q^{k+1}$ which gives

$$g(x,x) = \begin{cases} 0 & \text{if } x \in \left[0, \frac{1}{q}\right], \\ x - \frac{1}{q} & \text{if } x \in \left[\frac{1}{q}, 1 - \frac{1}{q}\right], \\ 2x - 1 & \text{if } x \in \left[1 - \frac{1}{q}, 1\right] \end{cases}$$

and

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} |\gamma_q(n) - \gamma_q(n+1)| = 1 - 2\int_0^1 g(x,x)\mathsf{d}x = \frac{2(q-1)}{q^2}.$$

Final theorem

Now in this part sequence x_n and y_n are not u.d., but arbitrary.

Theorem. Let $x_n \in [0,1)$ be a sequence with an a.d.f. $g_1(x)$ and $y_n \in [0,1)$ with an a.d.f. $g_2(x)$. Assume that $F(x,y)$ is a continuous function such that $\mathsf{d}_x\mathsf{d}_y F(x,y) \geq 0$ for every $(x,y) \in [0,1]^2$. Then we have

$$\max_{g(x,1)=g_1(x)} \int_0^1 \int_0^1 F(x,y)\mathsf{d}_x\mathsf{d}_y g(x,y) = \int_0^1 \int_0^1 F(x,y)\mathsf{d}_x\mathsf{d}_y \min(g_1(x), g_2(y)),$$

$$\min_{g(1,y)=g_2(y)} \int_0^1 \int_0^1 F(x,y)\mathsf{d}_x\mathsf{d}_y g(x,y) = \int_0^1 \int_0^1 F(x,y)\mathsf{d}_x\mathsf{d}_y \max(g_1(x)+g_2(y)-1, 0).$$

Sketch of proof: By Sklar's theorem to $g(x,y)$ there exists copula $c(x,y)$ such that $g(x,y) = c(g_1(x), g_2(y))$ for every $(x,y) \in [0,1]^2$.
Applying Fréchet-Hoeffding bounds we find

$$\max(g_1(x)+g_2(y)-1, 0) \leq g(x,y) \leq \min(g_1(x), g_2(y)).$$

References

[2010] Balá̌ž, V. – Nagasaka, K. – Strauch, O.: *Benford's law and distribution functions of sequences in* $(0, 1)$, Mathematical Notes (2010) (sending).

[1995] D. Berend – M.D. Boshernitzan – G. Kolesnik: *Distribution modulo* 1 *of some oscillating sequences, II.*, Israel J. Math. **92** (1995), 125–147.

[1968] J. Chauvineau: *Sur la rpartition dans* R *et dans* Q_p, Acta Arith. **14** (1967/68), 225–313 (MR **39**#6835).

[1980] G. Choquet: *Construction effective de suites* $(k(3/2)^n)$*. Étude des mesures* $(3/2)$*-stables*, C. R. Acad. Sci. Paris. Sér. A-B **291** (1980), A69–A74 (MR 82h:10062a-d).

[1987] J. Coquet – P. Liardet: *A metric study involving independent sequences*, J. Analyse Math. **49** (1987), 15–53 (MR 89e:11043).

[1997] Drmota, M. – Tichy, R. F.: *Sequences, Discrepancies and Applications*, Lecture Notes in Mathematics 1651, Springer-Verlag, Berlin, Heidelberg, 1997.

[2010] J. Fialová – O. Strauch *On two-dimensional sequences composed by one-dimensional uniformly distributed seqiuences*, (Prepared).

[1995] L. Flatto – J. C. Lagarias – A. D. Pollington: *On the range of fractional parts* $\{\xi(p/q)^n\}$, Acta Arith. **70** (1995), 125–147.

[1993] J.A. FRIDY: *Statistical limit points*, Proc. Amer. Math. Soc. **118** (1993), 1187–1192.

[2008] GIULIANO ANTONINI, R. – STRAUCH, O.: *On weighted distribution functions of sequences*, Uniform distribution Theory **3** (2008), no. 1, 1–18.

[1994] P.J. GRABNER – R.F. TICHY: *Remarks on statistical independence of sequences*, Math. Slovaca **44** (1994), 91–94 (MR 95k:11098).

[2001] P. KOSTYRKO – M. MAČAJ – T. ŁALT – O. STRAUCH: *On statistical limit points*, Proc. Amer. Math. Soc. **129** (2001), no. 9, 2647–2654 (MR 2002b:40003).

[1974] KUIPERS, L.– NIEDERREITER, H.: *Uniform Distribution of Sequences*, John Wiley & Sons, New York 1974, reprint: Dover Publications, Inc. Mineola, New York, 2006.

[1968] K. MAHLER: *An unsolved problem on the powers of* $3/2$, J. Austral. Math. Soc. **8** (1968), 313–321 (MR **37**# 2694).

[1991] G. MYERSON: *A sampler of recent developments in the distribution of sequences*, in: Number theory with an emphasis on the Markoff spectrum (Provo, UT, 1991), Lecture Notes in Pure and App. Math. **147** Marcel Dekker, New York, 1993, 163–190.

[1999] NELSEN, R.B.: *An Introduction to Copulas. Properties and Applications*, Lecture Notes in Statistics **139** Springer-Verlag, New York, 1999.

[1981] A.I. PAVLOV: *On distribution of fractional parts and Benford's law* (Russian), Izv. Aka. Nauk SSSR Ser. Mat. **45** (1981), no. 4, 760–774.

[2009] PILLICHSHAMMER, F. – STEINERBERGER, S.: *Average distance between consecutive points of uniformly distributed sequences*, Uniform Distribution Theory **4** (2009), no. 1, 51–67.

[1964] G. PÓLYA – G. SZEGŐ: *Aufgaben und Lehrsätze aus der Analysis*, Vol.1,2; 3rd corr. ed. Grundlehren d. math. Wiss., Band 19, Springer Verlag, Berlin, Göttingen, Heidelberg, New York, 1964.

[1976] G. RAUZY *Proprits statistiques de suites arithmtiques*, *Le Mathmaticien* Vol **15**, Collection SUP, Presses Universitaires de France, Paris, 1976, 133 pp. (MR **53**#13152).

[1959] I.J. SCHOENBERG:. *The integrability of certain functions and related summability methods*, Amer. Math. Monthly **66** (1959), 361–375.

[2009] STEINERBERGER, S.: *Uniform distribution preserving mappings and variational problems*, Uniform Distribution Theory **4** (2009), no. 1, 117–145.

[1982] O. STRAUCH: *Duffin – Schaeffer conjecture and some new types of real sequences*, Acta Math. Univ. Comenian. **40–41** (1982), 233–265 (MR 84f:10065).

[1994] O. STRAUCH: L^2 *discrepancy*, Math. Slovaca **44** (1994), 601–632 (MR 96c:11085).

[1995] O. STRAUCH: *Uniformly maldistributed sequences in a strict sense*, Monatsh. Math. **120** (1995), 153–164 (MR 96g:11095).

[1997] O. STRAUCH: *On set of distribution functions of a sequence*, in: *Proc. Conf. Analytic and Elementary Number Theory*, In Honor of E. Hlawka's 80th Birthday, Vienna, July 18-20, 1996, Universität Wien and Universität für Bodenkultur, (W. G. Nowak and J. Schoißengeier, eds.) Vienna, 1997, 214–229 (Zbl 886.11044).

[1997] O. STRAUCH: *On distribution functions of* $\xi(3/2)^n \bmod 1$, Acta Arith. **LXXXI** (1997), no. 1, 25–35 (MR 98c:11075).

[2010] STRAUCH, O. – NAIR, K.: *Unsolved Problems*, in: Unsolved Problems Section on the homepage of Uniform Distribution Theory, http://udt.mat.savba.sk/udt_unsolv.htm

[2005] O. STRAUCH – Š. PORUBSKÝ: *Distribution of Sequences: A Sampler*, Peter Lang, Frankfurt am Main, 2005.

[1972] R. TIJDEMAN: *Note on Mahler's* $3/2$*-problem*, Norske Vid. Selsk. Skr. **16** (1972), 1–4.

[1938] J.G. VAN DER CORPUT: *Verteilungsfunktionen I–VIII*, Proc. Akad. Amsterdam **38** (1935), 813–821, 1058–1066
textbf39 (1936), 10–19, 19–26, 149–153, 339–344, 489–494, 579–590.

[1940] T. VIJAYARAGHAVAN: *On the fractional parts of the powers of a number, I*, J. London Math. Soc. **15** (1940), 159–160.

[1935] A. WINTNER: *On the cyclical distribution of the logarithms of the prime numbers*, Quart. J. Math. Oxford (1)**6** (1935), 65–68 (Zbl 11, 149).

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