SELF-NORMALIZATION UNDER DEPENDENCE AND THE SUBSEQUENCE PRINCIPLE

Salem and Zygmund (1947) Assume $n_{k+1}/n_k \geq q > 1$, then

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{N} \sin n_k x \xrightarrow{d} N(0,1)$$

with respect to the probability space $((0,2\pi), \mathcal{B}, \frac{1}{2\pi}\lambda)$

Idea: **multiplicative orthogonality**

For $n_k = 2^k$

$$\int_0^{2\pi} \sin n_{k_1} x \cdots \sin n_{k_r} x \, dx = 0 \qquad k_1 < \ldots < k_r$$

Morgenthaler (1955) Given any uniformly bounded orthonormal system (f_n), there exists a (f_{n_k}) and a bounded $g \geq 0$ such that

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{N} f_{n_k} \xrightarrow{d} N(0, g)$$

where the characteristic function of g is $\int_{-\infty}^{\infty} e^{-\lambda^2 g(x)/2} dx$ (**mixed normal**)

Example: $f_n(x) = H_n(x)$, $L^2(\mathbb{R}, e^{-x^2/2})$, $g(x) = ce^{x^2/2}$

Orthogonality is not needed, only

Riemann property: $\int f_n h dx \to 0$ for all $h \in L^2$ $\quad f_n \xrightarrow{w} 0$

$|\int f_{n_1} dx| < 1/2, \quad |\int f_{n_1} f_{n_2} dx| < 1/2, \dots$

What is g? $\int f_{n_k}^2 h \longrightarrow \int gh$ for all $h \in L^2$ $\qquad f_{n_k}^2 \xrightarrow{w} g$

Gaposhkin (1966) Given any sequence (X_n) of random variables with $\sup_n E|X_n|^2 < \infty$, there exists a subsequence (X_{n_k}) and random variables X, Y, $Y \geq 0$ such that

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{N} (X_{n_k} - X) \xrightarrow{d} N(0, Y)$$

and

$$\limsup_{N \to \infty} \frac{1}{\sqrt{2N \log \log N}} \sum_{k=1}^{N} (X_{n_k} - X) = Y^{1/2} \qquad \text{a.s.}$$

Construction of X, Y: $X_{n_k} \xrightarrow{w} X, \quad (X_{n_k} - X)^2 \xrightarrow{w} Y$

Banach-Saks (1930):

$$\lim_{N\to\infty} \frac{(X_{n_1}-X)^2+\ldots+(X_{n_N}-X)^2}{N} = Y \text{ in } L^1 \text{ norm}$$

$$\frac{1}{\sqrt{NY}} \sum_{k=1}^{N} (X_{n_k} - X) \xrightarrow{d} N(0,1)$$

Self-normalized CLT

Giné, Götze, Mason (1997) Let $X, X_1, X_2, \ldots$ be i.i.d. random variables with $EX = 0$, $EX^2 I\{|X| \le T\}$ slowly varying as $T \to \infty$. Let $S_n = \sum_{k=1}^n X_k$, $V_n^2 = \sum_{k=1}^n X_k^2$. Then

$$S_n/V_n \xrightarrow{d} N(0,1)$$

Komlós (1967) Given any sequence (X_n) of random variables with $\sup_n E|X_n| < \infty$, there exists a subsequence (X_{n_k}) and an integrable random variable Y such that

$$\frac{1}{N}\sum_{k=1}^{N} X_{n_k} \longrightarrow Y \qquad \text{a.s.}$$

Chatterji (1971) Given $0 < p < 2$ and any sequence (X_n) of random variables with $\sup_n E|X_n|^p < \infty$, there exists a subsequence (X_{n_k}) and a random variable Y such that

$$\frac{1}{N^{1/p}}\sum_{k=1}^{N} (X_{n_k} - X) \longrightarrow 0 \qquad \text{a.s.}$$

Subsequence Principle (Chatterji 1972). *Let T be a limit theorem valid for all sequences of i.i.d.r.v.'s belonging to an integrability class L defined by the finiteness of a norm $\| \cdot \|_L$. Then if (X_n) is an arbitrary (dependent) sequence of r.v.'s with* $\sup_n \|X_n\|_L < +\infty$ *then there exists a subsequence (X_{n_k}) satisfying T in a mixed form.*

Aldous (1979) Verification of principle for all distributional and strong limit theorems

$$f_k(X_1, X_2, \ldots, \mu) \xrightarrow{d} G_\mu \quad \text{as } k \to \infty, \quad \mu \in S$$

Let (X_n) be tight, then there exists $X_{n_k} \xrightarrow{d} X$:

$P(X_{n_k} < t | X_{n_1}, \ldots, X_{n_{k-1}}) \longrightarrow G(t)$ random distribution function

$\mathrm{dist}(X_{n_k} | X_{n_1}, \ldots, X_{n_{k-1}}) \longrightarrow \mu$ **limit random measure**

Theorem. Let (X_n) be tight, with limit random measure μ and assume that $\mu \in DA(\alpha)$ a.s., $0 < \alpha < 2$.

$$\left(\sum_{k=1}^{n} \xi_k - a_n\right) / b_n \xrightarrow{d} G_\alpha \qquad P(|\xi| \geq t) \sim t^{-\alpha} L(t)$$

$$a_n = nEXI\{|X| \leq c_n\}, \qquad b_n^2 = nEX^2 I\{|X| \leq c_n\}, \quad c_n = \dots$$

Then there exists a subsequence (X_{n_k}) such that

$$\frac{\sum_{k=1}^{N} X_{n_k} - a_N(\mu)}{b_N(\mu)} \xrightarrow{d} G_\alpha$$

Theorem. Let (X_n) be tight with limit random measure μ and assume that $\mu \in MDA(G)$ a.s. for an extremal distribution G

$$(\max_{1 \leq k \leq n} \xi_k - a_k)/b_k \xrightarrow{d} G$$

Then there exists a subsequence (X_{n_k}) satisfying the extremal limit theorem with random centering and norming.

Uniformity

Menshov (1936) From any orthonormal system (f_n) one can select a subsequence (f_{n_k}) which is a convergence system, i.e. $\sum_{k=1}^{\infty} c_k f_{n_k}$ converges a.e. provided $\sum_{k=1}^{\infty} c_k^2 < \infty$.

Follows from theory for fixed $(\mathbf{c_k})$ (Kolmogorov two series theorem), but not in general (subsequence depends on (c_k))

Classical problem from 1930's. Can one choose a subsequence which is a **unconditional** convergence system ? (i.e. for which $\sum_{k=1}^{\infty} c_k f_{n_k}$ converges a.e. after any permutation of its terms provided $\sum_{k=1}^{\infty} c_k^2 < \infty$).

Komlós (1974) Proof of conjecture

Gaposhkin (1966) Let (X_n) be a uniformly bounded r.v. sequence. Then there exists a subsequence (X_{n_k}) and r.v.'s X and $Y \geq 0$ such that for any real sequence (a_k) satisfying

$$\max_{1 \leq k \leq N} |a_k| = o(A_N), \qquad A_N^2 = \sum_{k=1}^{N} a_k^2$$

we have

$$A_N^{-1} \sum_{k \leq N} a_k (X_{n_k} - X) \xrightarrow{d} N(0, Y)$$

and for any real sequence (a_k) satisfying

$$\max_{1 \leq k \leq N} |a_k| = o(A_N / (\log\log A_N)^{1/2})$$

we have

$$(2A_N \log\log A_N)^{-1/2} \sum_{k \leq N} a_k (X_{n_k} - X) = Y^{1/2} \qquad \text{a.s.}$$

Berkes and Tichy (2014) Let (X_n) be tight with limit random measure $\mu \in DNA(p)$ a.s., $0 < p < 2$. Then there exists a subsequence (X_{n_k}) such that for any real coefficients (a_k) satisfying

$$|a_n| = o(A_n), \qquad A_n = \left(\sum_{k=1}^{n} |a_k|^p \right)^{1/p}$$

we have

$$\frac{1}{A_n} \sum_{k=1}^{n} a_k X_{n_k} \xrightarrow{d} G_p$$

Discrepancy and quasi-Monte Carlo integration

Compute $\int_{[0,1]^d} f(\mathbf{x})\mathbf{dx}$, $\quad \mathbf{x} = (x_1, \ldots, x_d)$, d large

$$\left| \int_{[0,1]^d} f(\mathbf{x})\mathbf{dx} - \frac{1}{N}\sum_{k=1}^{N} f(\mathbf{x}_k) \right| \leq C D_N(\mathbf{x}_1, \ldots, \mathbf{x}_N)$$

Discrepancy

$$D_N(x_1, \ldots, x_N) = \sup_{0\leq a<b\leq 1} \left| \frac{N(a,b)}{N} - (b-a) \right| = \sup_x |F_N(x) - x|$$

$N(a,b)$: # of terms of $x_1, \ldots, x_N$ in (a,b)

Kolmogorov-Smirnov statistics of sample $(x_1, \ldots, x_N)$

How to construct low discrepancy sequences?

Discrepancy of $\{n_k\alpha\}$

Ostrowski, Khinchin, Hardy & Littlewood (1921–1924) The asymptotic behavior of $D_N(\{n\alpha\})$ is closely connected to the continued fraction digits of α

Khinchin (1924)

$$D_N(\{k\alpha\}) = O\left(\frac{(\log N)^{1+\varepsilon}}{N}\right) \qquad \text{a.e. for } \varepsilon > 0$$

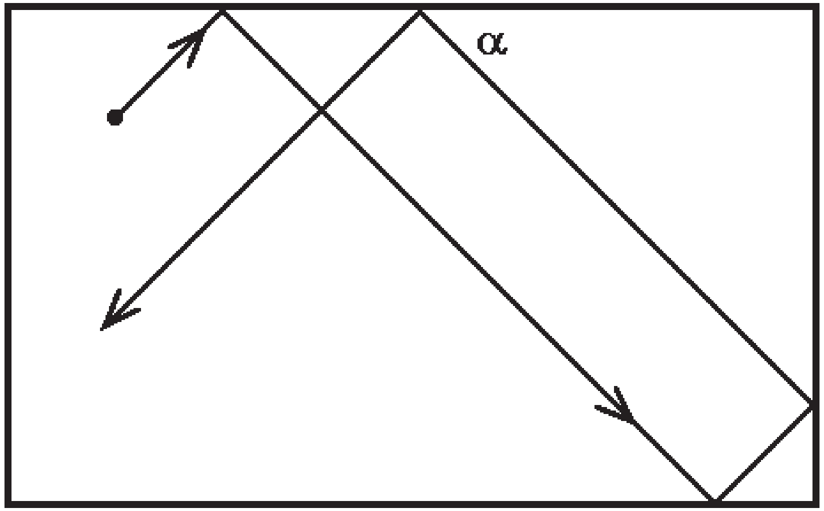
Kesten (1964)

$$D_N(\{k\alpha\}) \sim \frac{2}{\pi^2}\frac{\log N \log\log N}{N} \qquad \text{in measure}$$

Chung & Smirnov (1944, 1948) For i.i.d. sequences (ξ_k)

$$D_N(\{\xi_k\}) = O\left(\sqrt{\frac{\log\log N}{N}}\right) \qquad \text{a.s.}$$

Super-uniformity of the billiard path



Beck (2011) $\qquad A \subset [0,1]^2$ fix, speed $=1$

For $(1-\varepsilon)$-almost all starting positions

$$|\int_0^T I_A(X(t))dt - T\mu(A)| \leq c_\varepsilon \sqrt{\log T}, \qquad T \geq T_0$$

Hardy and Littlewood (1915) $n_k = k^2$

Philipp (1975) If $n_{k+1}/n_k \ge q > 1$, then

$$0 < \limsup_{N\to\infty} \sqrt{\frac{N}{\log\log N}} D_N(\{n_k x\}) < \infty \qquad \text{a.e.}$$

For i.i.d. sequences

$$\limsup_{N\to\infty} \sqrt{\frac{N}{\log\log N}} D_N(\{\xi_k\}) = \frac{1}{2} \qquad \text{a.s.}$$

Open problems:

(a) Value of limsup?
(b) Is the limsup constant a.e.?
(c) What is the limit distribution of $\sqrt{N} D_N(\{n_k x\})$?

Fukuyama (2008) $n_k = \theta^k$, $\theta > 1$

Let $f_a = I_{(0,a)} - a$, then

$$\limsup_{N\to\infty} \frac{\sum_{k=1}^{N} f_a(n_k x)}{(2N \log\log N)^{1/2}} = \sigma_a \qquad \text{a.s.}$$

where

$$\sigma_a^2 := \int_0^1 f_a^2(x)dx + 2\sum_{k=1}^{\infty}\int_0^1 f_a(x) f_a(\theta^k x)dx$$

$$\limsup_{N\to\infty} \sqrt{\frac{N}{\log\log N}} D_N(\{\theta^k x\}) = \sup_{0\le a<1} \sigma_a \qquad \text{a.e.}$$

$$
\begin{aligned}
\Sigma &= \sqrt{42}/9 && \text{if } \theta = 2 \\
\Sigma &= \frac{\sqrt{(\theta+1)\theta(\theta-2)}}{2\sqrt{(\theta-1)^3}} && \text{if } \theta \geq 4 \text{ is an even integer,} \\
\Sigma &= \frac{\sqrt{\theta+1}}{2\sqrt{\theta-1}} && \text{if } \theta \geq 3 \text{ is an odd integer} \\
\Sigma &= 1/2 && \text{if } \theta^r \text{ is irrational for } r = 1, 2, \ldots
\end{aligned}
$$

First values: $\sqrt{42}/9,\ 1/\sqrt{2},\ \sqrt{10/27},\ \sqrt{6}/4,\ \sqrt{42}/10$

General n_k:

$\limsup_{N\to\infty} \frac{\sum_{k=1}^{N} f(n_k x)}{(2N \log\log N)^{1/2}}$ is generally not a constant a.s.

$f(x) = \sin 2\pi x + \sin 4\pi x, \quad n_k = 2^k - 1$ (Erdős and Fortet (1949))

$\limsup = \sqrt{2} \cos \pi x$

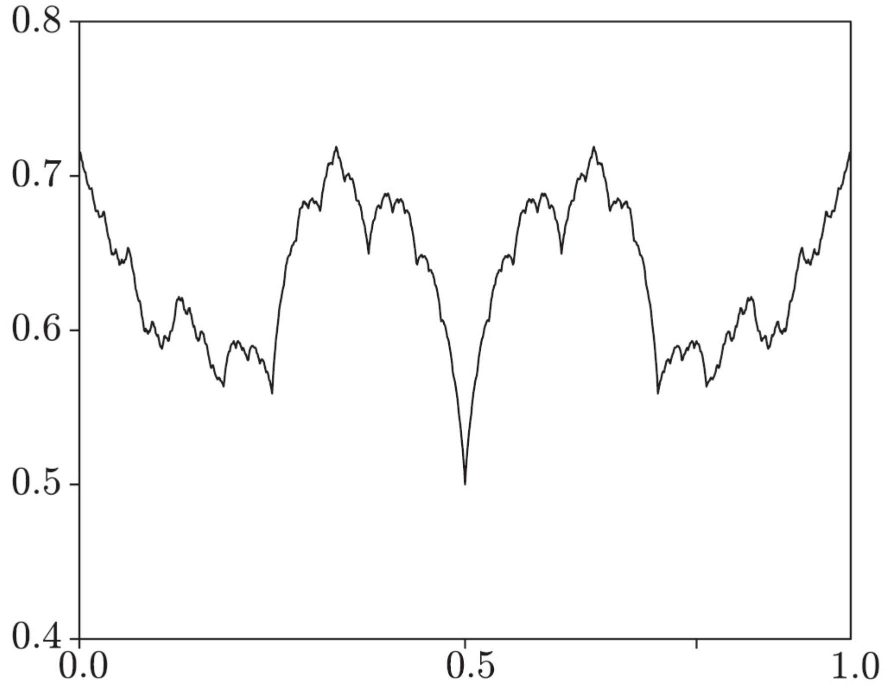
But self-normalizing LIL holds:

$$\limsup_{N\to\infty} \frac{\sum_{k=1}^{N} f(n_k x)}{(2V_N \log\log V_N)^{1/2}} = 1 \text{ a.s.}, \qquad V_N/N \to \sigma^2(x)$$

$$\limsup_{N\to\infty} \frac{\sum_{k=1}^{N} f_a(n_k x)}{(2N \log\log N)^{1/2}} = \sigma(x)$$

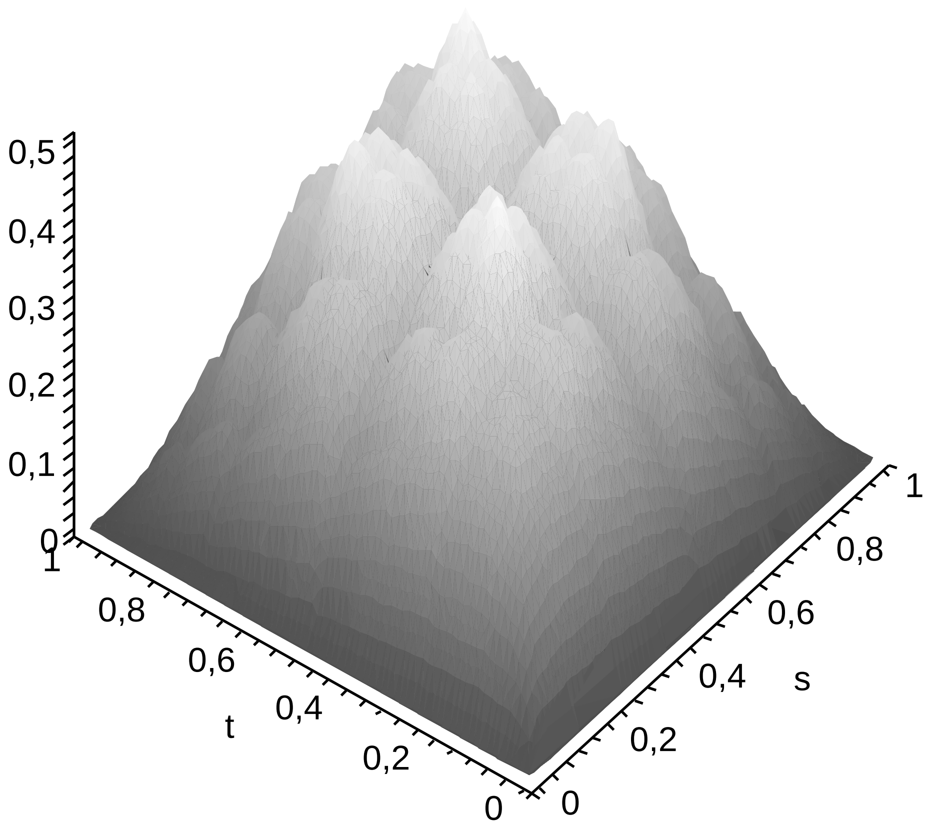
$$\limsup_{N\to\infty} \sqrt{\frac{N}{\log\log N}} D_N(\{n_k x\}) = \sup_{0\le a<1} \sigma_a(x) \qquad \text{a.e.}$$

Fukuyama (2012)



Limsup function for $n_k = 2^k - 1$

Aistleitner and Berkes (2012): Limit distribution of $\sqrt{N} D_N(\{\theta^k x\})$
Gaussian process with "fractal" covariance



Berkes and Raseta (2014) Let (n_k) be an increasing random walk, i.e. let $n_{k+1} - n_k$ be i.i.d. positive bounded random variables with bounded density. Then with probability 1 in the space of the (n_k) we have

$$\limsup_{N\to\infty} \sqrt{\frac{N}{\log\log N}} D_N(\{n_k x\}) = \sigma(x) \qquad \text{a.s.}$$

where $\sigma(x)$ is a nonconstant function explicitly computable from a RKHS model

Applications for Banach space theory

Which Banach spaces contain ℓ^p isomorphically?

$(x_1, x_2, \ldots) : \sum_{k=1}^{\infty} |x_k|^p < \infty \qquad \|\mathbf{x}\| = (\sum_{k=1}^{\infty} |x_k|^p)^{1/p}$

ξ_n i.i.d. p-stable. Then for $q < p$:

$$\|a_1\xi_1 + \ldots + a_k\xi_k\|_q \asymp (a_1^p + \ldots + a_k^p)^{1/p}$$

Let $(f_n) \in B$, find a subsequence whose limit random measure is p-stable.

Aldous (1981) Every infinite dimensional closed subspace H of L^1 contains a further subspace isomorphic to some ℓ^p, $1 \leq p \leq 2$

Kadec and Pelczynski (1962) Every normed weakly null sequence (f_n) in L^p, $p > 2$ has a subsequence (f_{n_k}) spanning ℓ^2 or ℓ^p isomorphically.

Berkes and Tichy (2014) Let $1 \leq p \leq 2$ and let (X_n) be a bounded sequence in L^p with limit random measure μ. Then there exists a subsequence (X_{n_k}) equivalent to the unit vector basis of ℓ^2 iff

$$\int_{\mathbb{R}} x^2 d\mu(x) \in L^{p/2}$$