Poisson generic real numbers

Verónica Becher

Universidad de Buenos Aires

joint work with Nicolás Álvarez, Martín Mereb and joint with Gabriel Sac Himelfarb

One World Numeration Seminar - May 31, 2022.

Poisson generic real numbers

This talk is about expansions of real numbers in an integer base, hence sequences of symbols in a finite alphabet.

Years ago Zeev Rudnick defined the Poisson generic real numbers as those where the distribution of long blocks of digits in the initial segments of the fractional expansions is the Poisson distribution. He called them supernormal.

The Poisson distribution

Consider the random allocation of N balls in K bins.

If N is smaller than K, a lot of bins will be empty and we expect some with exactly one ball, fewer with exactly two, still fewer with exactly three. . . .

The Poisson distribution

Consider N balls and K bins.
The probability p that a bin is allocated is $1/K$.
The expected proportion of bins with exactly i balls, for $i = 0, 1, 2, \ldots$

$$\chi(i) = \binom{N}{i} p^i (1-p)^{N-i}.$$

When N and K go to infinity but $N/K = \lambda$ is a fixed constant

$$\chi(i) \text{ converges to } e^{-\lambda} \frac{\lambda^i}{i!},$$

the Poisson probability mass function with paramenter λ.

Sequences as random events

Ω	alphabet with b symbols, $b \geq 2$.
$\Omega^{\mathbb{N}}$	infinite sequences of symbols in Ω
Ω^k	words of length k, for each $k \geq 1$

Let μ be the probability on Ω, with equal probability to each symbol.

Random allocation of N balls in $K = b^k$ bins

In the probability space (Ω^k, μ^k) the initial segment of length N of an infinite sequence can be seen as N almost independent events of placing words of length k, each one with equal probability b^{-k}.

Poisson generic sequences

For $x \in \Omega^{\mathbb{N}}$ and $\omega \in \Omega^k$, let the indicator function

$$I_j(x,\omega) = \mathbb{1}_{\{x[j,j+k)=\omega\}}.$$

Let $Z^{\lambda}_{i,k}(x)$ be the proportion of words of length k that occur exactly i times in the first $\lfloor \lambda b^k \rfloor$ symbols of $x \in \Omega^{\mathbb{N}}$,

$$Z^{\lambda}_{i,k}(x) = \mu^k \left(\omega \in \Omega^k : \sum_{1 \le j \le \lambda b^k} I_j(x,\omega) = i \right).$$

Example for $\Omega = \{0,1\}$, $b = 2$, $\lambda = 1$, $k = 3$, $b^k = \lambda b^k = 8$,

$$x = 10000100\ldots$$

For $i = 0$, $Z^{\lambda}_{i,k}(x) = 4/8$ (witnesses $011, 101, 110, 111$)
For $i = 1$, $Z^{\lambda}_{i,k}(x) = 2/8$ (witnesses $001, 010$)
For $i = 2$, $Z^{\lambda}_{i,k}(x) = 2/8$ (witnesses $100, 000$)
For $i \ge 3$, $Z^{\lambda}_{i,k}(x) = 0$

Poisson generic sequences

Definition (Zeev Rudnick)

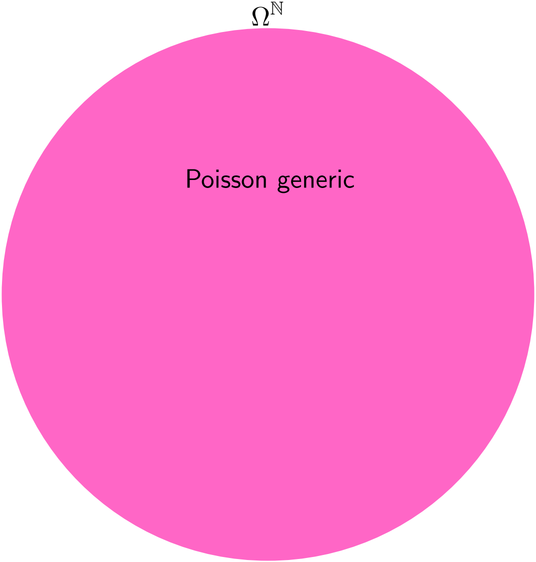
Let λ be a positive real number. A sequence x in $\Omega^{\mathbb{N}}$ is λ-Poisson generic if for every non-negative integer i,

$$\lim_{k\to\infty} Z^{\lambda}_{i,k}(x) = e^{-\lambda}\frac{\lambda^i}{i!}.$$

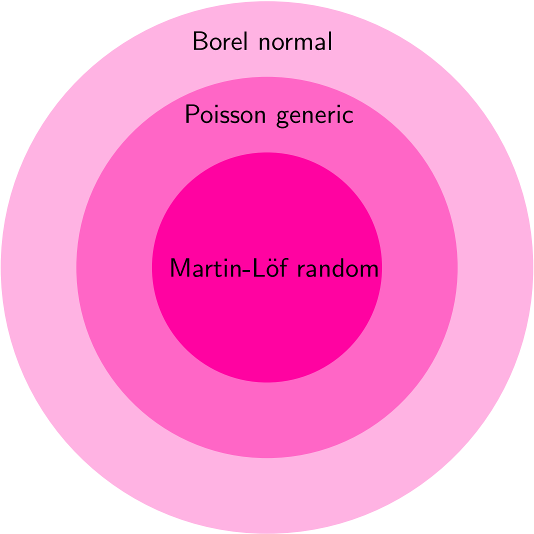
A sequence is Poisson generic if it is λ-Poisson generic, for all positive λ.

Theorem (Peres and Weiss)

$\mu^{\mathbb{N}}$-almost all x in $\Omega^{\mathbb{N}}$ are Poisson generic.



Strict inclusions



Benjamin Weiss. Random-like behavior in deterministic systems, Institute for Advanced Study Princeton University USA, 16 June 2010.

Benjamin Weiss. Poisson generic points.
Jean-Morlet Chair conference on Diophantine Problems, Determinism and Randomness. Centre International de Rencontres Mathématiques, 23-27 November 2020. Audio-visual resource: doi:10.24350/CIRM.V.19690103.

All Poisson generic are Borel normal

A sequence x in $\Omega^{\mathbb{N}}$ is Borel normal if all words of the same length occur in x with the same limiting frequency.

Theorem (Peres and Weiss)

Let λ be a positive real number. Every λ-Poisson generic sequence is Borel normal, but the two notions do not coincide.

Champernowne's sequence is not 1-Poisson generic

$$1234567891011121314151617\ldots$$

The proof technique was used in:

Teturo Kamae and Yu-Mei Xue. An easy criterion for randomness (2015).
and
Teture Kamae, Dong Han Kim, and Yu-Mei Xue. Randomness criterion Σ and its applications (2018).

Infinite de Bruijn sequences

A cyclic de Bruijn word of order n is a word v such that each word of length n occurs exactly once in the circular word determined by v.

Thus, at each position of a de Bruijn word of order n starts a different word of length n.

Lemma (Becher and Heiber 2011)

Every de Bruijn word of order n over an alphabet of at least three symbols can be extended to a de Bruijn word of order $n+1$. In case the alphabet has exactly two symbols, the extension is only guaranteed from order n to order $n+2$ or higher.

012110022 is de Bruijn of order 2 for alphabet $\{0,1,2\}$

$012110022101020001112021222$ is de Bruijn of order 3.

Definition

Let Ω be a b-symbol alphabet. A sequence $x \in \Omega^{\mathbb{N}}$ is infinite de Bruijn if for every $n \geq 1$, $x[1, b^k]$ is a cyclic de Bruijn word of order k.

Infinite de Bruijn sequences

Fact

Infinite de Bruijn sequences are Borel normal but not Poisson generic.

For $\lambda = 1$, $\lim_{k\to\infty} Z^{\lambda}_{1,k}(x) = 1$ and for $i \in \mathbb{N}_0$, $i \neq 1$, $\lim_{k\to\infty} Z^{\lambda}_{i,k}(x) = 0$.

All Martin-Löf random are Poisson generic

Definition (Martin-Löf 1966)

A real number is Martin-Löf random if it belongs to no $G_\delta = \bigcap_n O_n$ set of Lebesgue measure zero such that the sequence of open sets O_n is uniformly computable and the Lebesge measure of each O_n is computably bounded.

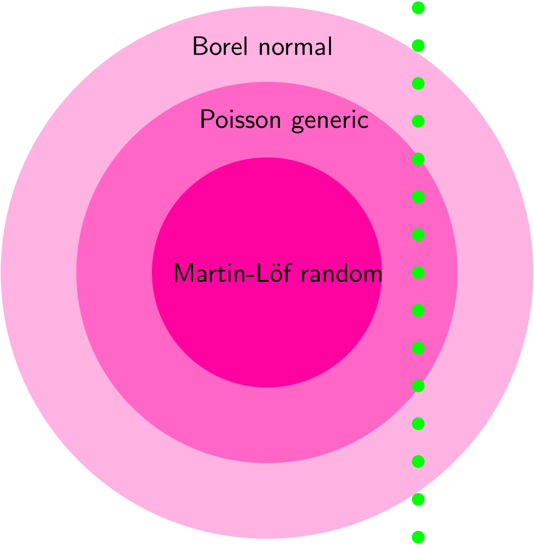
Theorem (Becher and Grigorieff 2021)

A real number x is Martin-Löf random if and only if for any integer $b \geq 2$ the sequence $(b^n x)_{n\geq 1}$ is u.d. modulo one for computably defined open sets.

Theorem (Alvarez, Becher and Mereb 2021)

Every Martin-Löf random real number has a Poisson generic expansion in every integer base $b \geq 2$.

Computable instances



Computable Poisson generic sequences

A sequence x in $\Omega^{\mathbb{N}}$ is computable when there is a computable function $f : \mathbb{N} \to \Omega$ such that $f(n)$ is the n-th symbol of x.

Theorem (Alvarez, Becher and Mereb 2021)

There are countably many computable Poisson generic sequences.

The proof consists in a construction of a Poisson generic sequece, similar to the constructions of absolutely normal numbers by Lebesgue 1916, Sierpinski 1916, Turing 1937, etc.

We based our construction in

Galatolo, Hoyrup and Rojas (2009) A constructive Borel–Cantelli lemma. Constructing orbits with required statistical properties.

It follows that there are Poisson generic instances in every Turing degree.

Theorem (Becher and Sac Himelfarb 2022)

Fix a positive real λ and an alphabet Ω.
There is a construction of an explicit λ-Poisson generic element in $\Omega^{\mathbb{N}}$.

After some experiments

Conjecture

The following are 1*-Poisson generic*

- *Rudin-Shapiro sequence along squares*
- *Thue-Morse sequence along squares*
- *The concatention of the Fibonacci numbers* $0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots$ *(in any base)*

The first two are known to be normal:

Mauduit and Rivat(2018); Müllner(2018), Drmota,Mauduit and Rivat (2019).

Question

Thinking in Borel's conjecture (1951) that all algebraic numbers are absolutely normal,

Are irrational algebraic numbers Poisson generic?

Peres and Weiss' metric theorem

They strengthen the definition of Poisson genericity by considering point processes on $\mathbb{R}^+$. We call it PW-genericity.

Peres and Weiss' metric theorem

A point process $Y(\cdot)$ on $\mathbb{R}^+$ is an integer-valued random measure on $\mathbb{R}^+$.

A point process $Y(\cdot)$ on $\mathbb{R}^+$ is Poisson if

- for all disjoint Borel sets $S_1, \ldots, S_m$ included in $\mathbb{R}^+$, the random variables $Y(S_1), \ldots, Y(S_m)$ are mutually independent;
- for all bounded Borel sets $S \subseteq \mathbb{R}^+$, the random variable $Y(S)$ has the distribution of a Poisson random variable with parameter the Lebesgue measure of S.

A sequence $(X_k(\cdot))_{k\geq 1}$ of point processes converges in distribution to a point process $Y(\cdot)$ if, for every Borel set $S \subseteq \mathbb{R}^+$, the random variables $X_k(S)$ converge in distribution to $Y(S)$ as k goes to infinity.

The point processes $M_k^x(\cdot)$ on $\mathbb{R}^+$

For each $x \in \Omega^{\mathbb{N}}$ and for each $k \in \mathbb{N}$, on the probability space (Ω^k, μ^k) define the integer-valued random measure $M_k^x = M_k^x(\omega)$ on $\mathbb{R}^+$ by setting for all Borel sets $S \subseteq \mathbb{R}^+$,

$$M_k^x(S)(\omega) = \sum_{j \in \mathbb{N} \cap b^k S} I_j(x, \omega)$$

where $\mathbb{N} \cap b^k S$ denotes the set of integer values in $\{b^k s : s \in S\}$.

Since each $x \in \Omega^{\mathbb{N}}$ is fixed, the I_j's are random variables on (Ω^k, μ^k).

For each $x \in \Omega^{\mathbb{N}}$ and each $k \in \mathbb{N}$, $M_k^x(\cdot)$ is a point process on $\mathbb{R}^+$.

PW-generic sequences

Definition

A sequence x in $\Omega^{\mathbb{N}}$ is PW-generic if the point prosesses $M_k^x(.)$ converge in distribution to a Poisson point process on $\mathbb{R}^+$ as k goes to ∞.

Theorem (Peres and Weiss)

$\mu^{\mathbb{N}}$-almost all x in $\Omega^{\mathbb{N}}$ are PW-generic.

PW-generic implies Poisson generic

Restricting $M_k^x(S)$ to sets $S = (0, \lambda]$ we obtain

$$Z_{i,k}^{\lambda}(x) = \mu^k \left(\omega \in \Omega^k : M_k^x((0,\lambda])(\omega) = i\}\right)$$

Corollary (Peres and Weiss' metric theorem for plain Poisson genericity)

$\mu^{\mathbb{N}}$-almost all x in $\Omega^{\mathbb{N}}$ are Poisson generic.

Peres and Weiss' proof

1. **The annealed result** (randomize x in $\Omega^{\mathbb{N}}$)

   Probability spaces $(\Omega^{\mathbb{N}} \times \Omega^k, \mu^{\mathbb{N}} \times \mu^k)$, $k \geq 1$

   Define point process $M_k(\cdot)$ on $\mathbb{R}^+$.

   Prove $M_k(\cdot) \xrightarrow{(d)} Y(\cdot)$, as $k \to \infty$,
   where $Y(\cdot)$ is Poisson point process on $\mathbb{R}^+$.

$$\mu^{\mathbb{N}} \times \mu^k \left((x,\omega) : M_k(S)(x,\omega) = i\right) = \boxed{\mathbb{E}\left[\mu^k\left(\omega : M_k^x(S)(\omega) = i\right)\right]}$$

2. **The quenched result** (apply a concentration inequality)

   Prove for all k large enough and for $\mu^{\mathbb{N}}$-almost all $x \in \Omega^{\mathbb{N}}$,

$\boxed{\mathbb{E}\left[\mu^k\left(\omega : M_k^x(S)(\omega) = i\right)\right]}$ and $\mu^k\left(\omega : M_k^x(S)(\omega) = i\right)$ are within $1/k$

Then, for $\mu^{\mathbb{N}}$-almost all $x \in \Omega^{\mathbb{N}}$, $M_k^x(\cdot) \xrightarrow{(d)} Y(\cdot)$, as $k \to \infty$.

Point process $M_k(\cdot)$ on $\mathbb{R}^+$

For each $k \in \mathbb{N}$, on the probability space

$$(\Omega^{\mathbb{N}} \times \Omega^k, \mu^{\mathbb{N}} \times \mu^k)$$

consider the random variables given by the indicator functions

$$I_j = I_j(x, \omega)$$

and define the integer-valued random measure $M_k = M_k(x, \omega)$ on $\mathbb{R}^+$ by setting for all Borel sets $S \subseteq \mathbb{R}^+$,

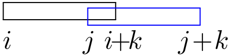
$$M_k(S)(x, \omega) = \sum_{j \in \mathbb{N} \cap b^k S} I_j(x, \omega)$$

where $\mathbb{N} \cap b^k S$ denotes the set of integer values in $\{b^k s : s \in S\}$.

Then, $M_k(\cdot)$ is a point process on $\mathbb{R}^+$.

The indicator variables are almost independent

Consider $i, j \in \mathbb{N}$ such that $i < j$ and $j - i < k$.
If $\omega \in \Omega^k$ is at positions i and j of some x the two occurrences overlap.



The probability that $\omega \in \Omega^k$ has prefix and suffix of the length of the overlap

$$\mu^k \left(\omega \in \Omega^k : \omega(j-i, k] = \omega[1, k-(j-i)]\right) = \boxed{b^{-\text{overlap}}},$$

The probability that $x \in \Omega^{\mathbb{N}}$ contains one of these ω's at positions i and j

$$\mu^{\mathbb{N}}\Big(x \in \Omega^{\mathbb{N}} : x[i, i+k) = x[j, j+k) = \omega\Big) = \boxed{b^{-(2k-\text{overlap})}},$$

Hence,

$$\mu^{\mathbb{N}} \times \mu^k \left((x,\omega) \in \Omega^{\mathbb{N}} \times \Omega^k : I_i(x,\omega) I_j(x,\omega) = 1\right) = \boxed{b^{-2k}},$$

the same as if I_i and I_j were independent.

For all $i, j \in \mathbb{N}$, $\mathbb{E}\left[I_j\right] = b^{-k}$, and $\mathbb{E}\left[I_i I_j\right] = b^{-2k}$.

Dependency graph

The dependency graph for a family of J random variables is a graph L with underlying vertex set J such that for any pair of disjoint subsets $A, B \subseteq J$ of vertices with no edge $e = (a, b)$, $a \in A$, $b \in B$ connecting them, the subfamilies $\{I_i\}_{i \in A}$ and $\{I_j\}_{j \in B}$ are mutually independent.

For each k, on probability space $(\Omega^{\mathbb{N}} \times \Omega^k, \mu^{\mathbb{N}} \times \mu^k)$ the dependency graph L is for a family I_j is:

$$(i, j) \in \operatorname{edges}(L) \text{ if and only if } |i - j| < k.$$

A criterion to converge to Poisson

Proposition (Kallenberg)

Let $(X_k(\cdot))_{k\geq 1}$ *be a sequence of point processes on* $\mathbb{R}^+$ *and let* $Y(\cdot)$ *be a Poisson point process on* $\mathbb{R}^+$*. If for any* $S \subseteq \mathbb{R}^+$ *that is a finite union of disjoint intervals with rational endpoints we have*

1. $\limsup_{k\to\infty} \mathbb{E}[X_k(S)] \leq \mathbb{E}[Y(S)]$ *and*
2. $\lim_{k\to\infty} \Pr\left(X_k(S) = 0\right) = \Pr\left(Y(S) = 0\right)$

then $X_k(\cdot) \xrightarrow{(d)} Y(\cdot)$ *as* $k \to \infty$*.*

Total variation distance

Proposition (Janson, Łuczak and Ruzinski Theorem 6.23)

Let $\mathrm{Po}(\lambda)$ *be a Poisson random variable with mean* λ.
Let $(I_j)_{j\in J}$ *be a family of random variables on a given probability space.*
Let L *be its dependency graph with underlying vertex set* J.

If the random variable $X = \sum_{j\in J} I_j$ *satisfies* $\lambda = \mathbb{E}[X] = \sum_{j\in J} \mathbb{E}[I_j]$ *then,*

$$d_{TV}(X, \mathrm{Po}(\lambda)) \leq \min\left\{1, \lambda^{-1}\right\} \left(\sum_{j\in J} \mathbb{E}[I_j]^2 + \sum_{i,j:(i,j)\in \mathrm{edges}(L)} \mathbb{E}[I_i I_j] + \mathbb{E}[I_i]\,\mathbb{E}[I_j] \right).$$

Apply is to $M_k(S)$ and $\lambda = |S|$

Our dependency graph L is: $(i,j) \in \mathrm{edges}(L)$ if and only if $|i-j| < k$.

$\mathbb{E}[I_j] = b^{-k}$, and $\mathbb{E}[I_i I_j] = b^{-2k}$.

We proved
$M_k(\cdot) \xrightarrow{(d)} Y(\cdot)$ as $k \to \infty$, where $Y(\cdot)$ is Poisson point process on $\mathbb{R}^+$.

McDiarmid's concentration inequality

Proposition (McDiarmid's inequality)

Assume $f : \Omega^N \to \mathbb{R}$ *satisfies that there is a positive* $\epsilon = \epsilon(N)$ *such that for any two tuples* $x, x' \in \Omega^N$ *which differ only in a single coordinate,*

$$|f(x) - f(x')| \leq \epsilon.$$

Let $X_1, \ldots, X_N$ *be independent random variables with values in* Ω*. Let* $\Pr$ *denote the probability on the underlying domain. Then for any* $\delta \geq 0$,

$$\Pr\left(|f(X_1 \ldots X_N) - \mathbb{E}[f(X_1 \ldots X_N)]| > \delta\right) \leq 2\exp\left(\frac{-2\delta^2}{N\epsilon^2}\right).$$

Application of McDiarmid inequality

A one-coordinate change in $x \in \Omega^{\mathbb{N}}$ affects k many words of length k

For each $i \geq 0$ and $S \subseteq \mathbb{R}^+$ finite union of intervals with rational endpoints, let $f_k : \Omega^N \to \mathbb{R}^+$, where $N = \lfloor |S| b^k +$ small $\rfloor$

$$f_k(x) = \mu^k \left(\omega : M_k^x(S)(\omega) = i \right)$$

Then,

$$\mu^{\mathbb{N}} \left(x : \, |f_k(x) - \mathbb{E}[f_k(x)]| > 1/k \right)$$

has exponential decay in b^k.

$\mu^{\mathbb{N}}$- almost all $x \in \Omega^{\mathbb{N}}$ are PW-generic

By Borel–Cantelli lemma this set has $\mu^{\mathbb{N}}$ zero,

$$\left\{ x \in \Omega^{\mathbb{N}} : |f_k(x) - \mathbb{E}[f_k(x)]| > 1/k \text{ for infinitely many } k \right\}$$

where

$$f_k(x) = \mu^k \left(\omega : M_k^x(S)(\omega) = i\right)$$

$$\mathbb{E}[f_k(x)] = \mu^{\mathbb{N}} \times \mu^k \left((x,\omega) \,:\, M_k(S)(x,\omega) = i\right).$$

$$M_k(S) \xrightarrow{(d)} \mathrm{Po}(|S|), \text{ as } k \text{ to } \infty.$$

This holds for every $i \geq 0$ and every $S \subseteq \mathbb{R}^+$ finite union of intervals with rational endpoints, we conclude

$$\text{for } \mu^{\mathbb{N}}\text{-almost all } x \in \Omega^{\mathbb{N}},\ M_k^x(S) \xrightarrow{(d)} \mathrm{Po}(|S|) \text{ as } k \text{ goes to infinity}$$

By Kallenberg's Proposition, for $\mu^{\mathbb{N}}$-almost all $x \in \Omega^{\mathbb{N}}$,
$M_k^x(\cdot) \xrightarrow{(d)} Y(\cdot)$ as $k \to \infty$, where $Y(\cdot)$ is Poisson point process on $\mathbb{R}^+$.

$\square$

Research program on Poisson genericity

Basic

- Do aligned and non-aligned definitions Poisson genericity coincide?
  Pillai (1940); Niven and Zuckerman (1951).

Research program on Poisson genericity

Discrepancy

- What is the minimal discrepancy of $(b^n x)_{n\geq 1}$ for x Poisson generic?
  Gál and Gal (1964); Schiffer (1986); Levin (1999); Larcher (2022)

Research program on Poisson genericity

Numeration systems

- Is Poisson genericity base invariant?

  Cassels (1950); Schmidt (1968); Stoneham numbers of Bailey and Crandall (2003), Bugeaud (2012).

- Poisson generic continued fractions (asked by Eda Cesaratto)

  Postnikov and Pyatetskii (1957); Adler, Keane and Smorodinsky (1981); Vandehey (2016)

- Poisson genericity for other numerations systems such as as Pisot numbers, Cantor series expansions

  Erdös and Rényi (1959); Turán (1976), Mance (2011); Airey, Mance and Vandehey (2016): Madritsch, Scheerer and Tichy (2018)

Research program on Poisson genericity

Modifying numbers

- What selection functions preserve Poisson genericity?
  Wall (1940); Agafonoff (1968); Kamae and Weiss (1975); Vandehey (2017)
- Sums
  Rényi (1957); Aistleitner (2011)
- Insertion in base b lifting the property to base $b+1$.
  Zylber (2021), Becher (2022)
- Transformation by transducers
  Carton and Orduna (2018)

Research program on Poisson genericity

Poisson genericity and other matematical properties

- Liouville numbers
- Prescribed irrationality exponent
- Mendès-France question on x and $1/x$

Bugeaud (2002); Becher, Heiber and Slaman (2015); Becher and Madritsch (2022)

Research program on Poisson genericity

Descriptive complexity

- Is the set Poisson generic reals Π^0_3 in the effective Borel hierarchy?

  Ki and Linton (1994); Becher, Heiber and Slaman (2014);

  Jackson, Mance and Vandehey (2021), Airey, Jackson and Kwietniak (2022)

Verónica Becher and Serge Grigorieff. Randomness and uniform distribution modulo one. Information and Computation, 104857, 2021.

Nicolás Alvarez, Verónica Becher, and Martín Mereb. Poisson generic sequences. *Submitted*. https://arxiv.org/abs/2202.01632v2

Verónica Becher and Gabriel Sac Himelfarb. A construction of a λ-Poisson generic sequence *Submitted*. https://arxiv.org/abs/2205.03981

Svante Janson, Tomasz Luczak, and Andrzej Rucinski. *Random graphs*, volume 45. John Wiley & Sons, 2000.

Olav Kallenberg. Random measures, theory and applications. Springer Cham, 2017.

J. F. C. Kingman. Poisson Processes. Clarendon Press, 1992. Günter Last and Mathew Penrose. Lectures on the Poisson process. Institute of Mathematical Statistics textbooks. Cambridge University Press, 2017.

Clemens Müllner. The Rudin-Shapiro sequence and similar sequences are normal along squares. *Canadian Journal of Mathematics*, 70(5):1096 – 1129, 2018.

Christian Mauduit and Joël Rivat. Rudin–Shapiro sequences along squares. *Transactions of the American Mathematical Society*, 370(11):7899–7921, 2018.

Zeev Rudnick and Alexandru Zaharescu. The distribution of spacings between fractional parts of lacunary sequences. *Forum Mathematicum*, 14(5):691–712, 2002.

Benjamin Weiss. Random-like behavior in deterministic systems, 16 June 2010. Institute for Advanced Study Princeton University USA.

Benjamin Weiss. Poisson generic points, 23-27 November 2020. Conference Diophantine Problems, Determinism and Randomness - CIRM- organized by Joël Rivat and Robert Tichy. https://library.cirm-math.fr/Record.htm?idlist=13&record=19287810124910050929