Poisson genericity in numeration systems with exponentially mixing probabilities

Eda Cesaratto*

Joint work with
Nicolás Álvarez, Verónica Becher, and Martín Mereb

* Universidad Nac. de General Sarmiento & CONICET, Argentina

Numeration 2024

Universiteit Utrecht

Poisson generic real numbers

This talk is about numerations systems as sequences of symbols of a finite or countable set Ω, called an alphabet.

An overview

Consider a numeration system associated with an invariant exponentially mixing measure.

For almost all infinite sequences of symbols x, the number of times that the words w of length k which are in the initial segment of x follows a Poisson law as $k \to \infty$.

Numeration systems covered by our result

- Integer bases and continued fractions.
- Fibred systems with an invariant and exponentially mixing measure (including the Ostrowski continued fraction algorithm in the plane).
- Stochastic processes as aperiodic and irreducible Markov chains

Motivation

Yuval Peres and Benjamin Weiss proved the result for integer bases b

Poisson for integer bases

For almost all $x \in [0,1]$ with respect to the Lebesgue measure, the number of times that words w of length k are in the base b expansion of x follows a Poisson law as $k \to \infty$.

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.

Álvarez, Becher and Mereb transcribed their proof and related Poisson genericity with the notion randomness from computability theory.

Poisson generic sequences. International Mathematics Research Notices, rnac234, 2022

Our initial question: Are the methods of Peres and Weiss amenable to continued fractions?

The symbols in CF expansions are infinitely correlated.

Notations and examples

Ω	alphabet finite or countable
$x \in \Omega^{\mathbb{N}}$	infinite sequences of symbols in Ω
$w \in \Omega^k$	words of length k, for each $k \geqslant 1$

Example

$\Omega = \{0, \ldots, 9\}$

$x = 414\mathbf{21}35623730950488016887242096\mathbf{69}80785\mathbf{69}671875376\mathbf{69}48\ldots$

$k = 2$, $i = 4$, $x[4, 5] = \mathbf{21}$

The word **69** is three times in the first 50 symbols of x.

Statistics of x with words w of two symbols

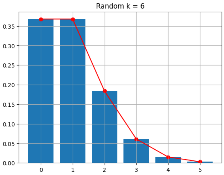
j	# Words of two symbols that are j times in x	Proportion of words of two symbols that are j times in x
0	61	0.61
1	30	0.3
2	8	0.08
3	1	0.01
4 or more	0	0

Examples

Decimal expansions, Lebesgue measure
$\lambda = 1$.
$k = 6$, $w \in \Omega^6$.

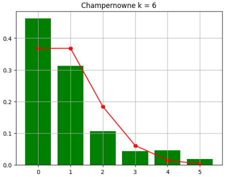
Initial segment of length 10^6.
In red, the Poisson probability mass function.
In blue/green, the histogram of the proportion of words w which appears 0, 1, ... times

Random x



x is the decimal expansion of a number generated at random

Champernowne



x is the Champernowne number
$x = 123456789101112131415\ldots$
The Champernowne number is not Poisson for $\lambda = 1$ (Peres and Weiss)

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 or 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 j balls, for $j = 0, 1, 2, \dots$

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

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

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

the Poisson probability mass function with parameter λ.

Notation:

$\text{Po}(\lambda)$: probability mass function of parameter λ.

$X \sim \text{Po}(\lambda)$: The r.v. X is distributed according to a Poisson of parameter λ

Continued fractions and beyond

$\Omega = \mathbb{N}$
Any irrational number in $[0, 1]$ admits a representation of the form:

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}} = [a_1, a_2, \dots] \quad a_i \in \mathbb{N},\ i \geqslant 1.$$

The natural measure μ associated with CF is the Gauss measure:

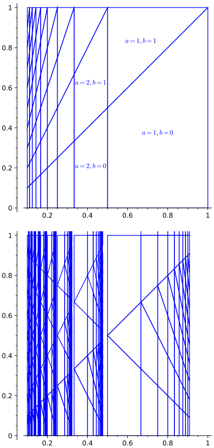
$$d\mu = \frac{dx}{\ln(2)(1 + x)}.$$

Integer bases vs continued fractions

Integer bases	Continued fractions
Finite alphabet	Infinite alphabet
Independence of symbols	infinite correlations between symbols

And beyond

Partitions of $[0,1] \times [0,1]$



The Ostrowski map

Given irrationals $x, y \in [0,1]$ define

$$T(x,y) = (\{1/x\}, \{y/x\}) \ ,$$

where $\{t\} := t - \lfloor t \rfloor$ is the fractional part.

$$\Omega = \{(a,b) \in \mathbb{Z}^2 : \ a \geqslant 1, 0 \leqslant b \leqslant a\}$$

T has an invariant and exponentially mixing probability μ (Berthé and Lee, 2024)

A general setting: sequences and measures

Ω: finite or countable set of symbols

μ: Borel probability measure on $\Omega^{\mathbb{N}}$.

μ_k: the Borel probability measure on Ω^k induced by μ, for each $k \in \mathbb{N}$.

Cylinders: $w \in \Omega^k$,

$$C_k(w) = \{x \in \Omega^{\mathbb{N}} : x[1,k) = w\}, \quad \mu_k(w) = \mu(C_k(w))$$

Length of initial segments: $\lfloor \lambda/\mu_k(w) \rfloor$, $\quad \mu_k(w) \neq 0$, for each $\lambda > 0$.

Variable of interest: $M_k(x,w)(\lambda) = \#\{w \text{ is in } x[1, \lfloor \lambda/\mu_k(w) \rfloor]\}$

Example with continued fractions

$x = \pi - 3 = [0; 7, 15, 1, 292, 1, 1, \mathbf{1}, \mathbf{2}, 1, 3, 1, 14, 2, 1 \underbrace{\mathbf{1}}_{14}, 1, 2, 2, 2, 2, \ldots]$

$k = 2$, $w = 12$,

$C_2(12) = \left[\frac{1}{1+\frac{1}{2}}, \frac{1}{1+\frac{1}{3}}\right] = [2/3, 3/4]$, $\mu_2(12) = \int_{2/3}^{3/4} \mu(x)dx = 0,0708\ldots$

$\lambda = 1$, $\lambda/\mu_2(12) = 14.2\ldots$

$$M_2(\pi - 3, 12)(1) = 1$$

Poisson law for point processes on $\mathbb{R}^+$

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

$$I_i(x, w) = \mathbb{1}_{x[i,i+k)=w}$$

For each $k \in \mathbb{N}$, for each $x \in \Omega^{\mathbb{N}}$, on the space Ω^k with measure μ_k,

$$M_k^x(w)(S) = M_k(x, w)(S) = \sum_{i:\, i\mu_k(w)\in S} I_i(x, w), \quad \text{for any Borel set } S \subseteq \mathbb{R}^+.$$

is a integer-valued random measure on $\mathbb{R}^+$.
The sum runs over $i \in \mathbb{N}$, $i\mu_k(w) \in S$.

Choose w at random and assign to S a nonnegative number

$$S \mapsto M_k^x(w)(S)$$

If $S = (0, \lambda]$, the initial segment is $\lfloor \lambda/\mu_k(w) \rfloor$ as before.

Peres and Weiss' Poisson genericity: Point processes on $\mathbb{R}^+$

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

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

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

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

We write $X_k(\cdot) \xrightarrow{(d)} \mathrm{Po}(\cdot)$

Fact

The sequence $(M_k^x(\cdot))_{k\geqslant 1}$ is a sequence of point processes on $\mathbb{R}^+$.
(w chosen at random in Ω^k.)

Poisson generic numbers: Definition

Definition (Poisson genericity)

We say that $x \in \Omega^{\mathbb{N}}$ is Poisson generic if the sequence $(M_k^x(.))_{k \geqslant 1}$ of point processes on $\mathbb{R}^+$ converges in distribution to a Poisson point process on $\mathbb{R}^+$, as k goes to infinity.

This means that, for each fixed x, for every Borel set $S \subseteq \mathbb{R}^+$,

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

or, for each $j \geqslant 0$,

$$\mu_k\left(\left\{w \in \Omega^k : \ M_k^x(w)(S) = j\right\}\right) \to e^{-|S|}|S|^j/j! \text{ as } k \to \infty.$$

$|S|$ is the Lebesgue measure of S.

Our main theorem

Assumptions on the probability measure μ on $\Omega^{\mathbb{N}}$.

- Invariant: $\mu_k(w) = \mu_k(\{x : x[i, i+k) = w\})$ for any $i, k \in \mathbb{N}$ and $w \in \Omega^k$.
- Exponentially mixing (nonindependent "ma non troppo")
  There exists a $0 < \sigma < 1$ such that for any $A, B \subset \Omega^{\mathbb{N}}$ of positive measure with
  A depending on the first i symbols,
  B depending on the symbols from position j, $j \geqslant i + k$,

$$\left| \frac{\mu(A \cap B)}{\mu(A)\mu(B)} - 1 \right| = O(\sigma^{j-i-k}).$$

Our main theorem

For any invariant and exponentially mixing probability measure μ on $\Omega^{\mathbb{N}}$, μ-almost all $x \in \Omega^{\mathbb{N}}$ are Poisson generic.

Warning

The rol of x and w is not symmetric.

For fixed $w \in \Omega^k$, it is feasible to prove the estimate

$$\mathbb{E}_\mu[M_k^w(S)] \approx |S| \quad \text{as } k \to \infty$$

for any $S \subset \mathbb{R}^+$ which is a finite union of bounded intervals.

For fixed $x \in \Omega^{\mathbb{N}}$, to obtain estimates of

$$\mathbb{E}_{\mu_k}[M_k^x(S)] \quad \text{as } k \to \infty$$

is not immediate.

Adaptation of Peres and Weiss' general strategy

Annealed result. Integrate on $\Omega^{\mathbb{N}} \times \Omega^k$

- Fix $w \in \Omega^k$ and integrate with respect to $x \in \Omega^{\mathbb{N}}$. Only finite union of bounded intervals S.
  Use the Chen-Stein method (only for invariant and exponentially mixing probabilities). Bound the total variation distance between $M_k^w(S)$ and Po$(|S|)$.
- Integrate with respect to $w \in \Omega^k$.
- Use Kallenberg's criterion of convergence for point processes:
  $$M_k(\cdot) \xrightarrow{(d)} \text{Po}(\cdot) \quad \text{as } k \to \infty.$$

Quenched result (almost all $x \in \Omega^{\mathbb{N}}$, integrate on Ω^k)

- With 'high probability", for $x \in \Omega^{\mathbb{N}}$. Use a concentration result
  $$M_k^x(\cdot) \sim M_k(\cdot) \sim \text{Po}(\cdot) \quad \text{as } k \to \infty.$$
- From "high probability" to almost all x: Use Borel Cantelli's lemma
  Only finite union of bounded intervals S.
- Use Kallenberg's criterion of convergence for point processes and conclude:

Poisson genericity for almost all $x \in \Omega^{\mathbb{N}}$

Sketch of the proof: Annealed

Annealed result. Integrate on $\Omega^{\mathbb{N}} \times \Omega^k$

- Fix $w \in \Omega^k$ and integrate with respect to $x \in \Omega^{\mathbb{N}}$. Only finite union of bounded intervals S:

  $$\mathbb{E}_\mu[M_k^w(S)] \approx |S| \quad \text{and} \quad \mathbb{V}_\mu[M_k^w(S)] \approx |S| + \text{error}(w)$$

  Use the Chen-Stein method: invariant and exponentially mixing probabilities

  If X is a sum of indicators and its expectation is λ, the total variation distance between X and $\mathrm{Po}(\lambda)$ is controlled by $|\mathbb{V}[X] - \lambda|$.

- Integrate with respect to $w \in \Omega^k$:

  $$\mathbb{E}_{\mu_k}[\text{error}(w)] \to 0 \quad \text{as} k \to \infty.$$

- Use Kallenberg's criterion of convergence for point processes: From finite union of bounded intervals with rationals points to Borel sets. For every Borel set S,

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

  The integration is done with respect to the measure $d\mu \times d\mu_k$.

Sketch of the proof: Quenched result

Use a concentration result

For some prescribed conditions on $\varphi : \Omega^{\mathbb{N}} \to \mathbb{R}$, φ is to close to its mean:

$$\mu(\{x : |\varphi(x) - \mathbb{E}_\mu[\varphi]| \geqslant t\}) \to 0 \quad \text{as } t \to +\infty.$$

Kontorovich and Ramanan (2007- 2008)

A concentration results holds if φ depends on a finite and fixed number of symbols and one of the following conditions holds

- Ω is finite or
- Ω is countable and φ satisfies the constant weighted Hamming distance property.

Our "functions" $M_k(x, w)(S)$ depend of all symbols of x as $k \to \infty$.

Our concentration

A concentration result holds for $\varphi : \Omega^{\mathbb{N}} \mapsto \mathbb{R}^+$ if

- φ is a "strong" limit of a sequence of functions $(\varphi_N)_{N \geqslant 1}$, each φ_N depends on N symbols.
- Each φ_N satisfies a concentration result á la Kontorovich-Ramanan.

Sketch of the proof: Quenched result

Quenched result (almost all $x \in \Omega^{\mathbb{N}}$, integrate on Ω^k)

- Use a concentration result: with 'high probability" on $x \in \Omega^{\mathbb{N}}$,
  $$M_k^x(S) \approx \mathbb{E}_\mu[M_k(S)] \approx \text{Po}(|S|) \quad \text{as } k \to \infty.$$
- Use Borel Cantelli's lemma: from "high probability" to almost all x.
  For almost all $x \in \Omega^{\mathbb{N}}$, for every finite union of bounded intervals of rationals endpoints S,
  $$M_k^x(S) \xrightarrow{(d)} \text{Po}(|S|) \quad \text{as } k \to \infty.$$
  (Integrate with respect to $w \in \Omega^k$)
- Use Kallenberg's criterion of convergence for point processes.
  From finite union of bounded intervals of rationals end points to Borel sets.
  For almost all $x \in \Omega^{\mathbb{N}}$,
  $$M_k^x(\cdot) \xrightarrow{(d)} \text{Po}(\cdot) \text{ as } k \to \infty.$$

Poisson genericity for almost all $x \in \Omega^{\mathbb{N}}$

Poisson limit law in Dynamical Systems

Return time: the number of visits of a given orbit to a set.

General goal

Consider a discrete dynamical system with an invariant mixing probability and a sequence of sets shrinking to a point (satisfying good properties).

The distribution of return times is asymptotically Poisson

as the measure of the sets goes to zero.

Early works (1940–1990): Doeblin-Iosifescu (CF), Pitskell (MC)
Poisson law of rare events (1990–): Collet, Coelho, Galves, Hirata, Schmitt.
Followed by (2000–):Abadi, Lacroix, Paccaut, Vaienti, Zweimüller, etc.

Poisson limit law in Dynamical Systems

Some differences with respect to our work

- The role of w and x is reversed.
- Many works deal with for $S = (0, \lambda)$.
- In many words the exceptional sets (the sets where the limit does not hold) depend on λ .

Dynamical results:

- error terms,
- periodic orbits (not Poisson).
- different families visited sets (not only cylinders).
- Many different notions of mixing.

Dynamical system methodology

Generating series, transfer operators, Chen-Stein method, dynamical properties of the measure.

A problem

Almost all numbers are decimal Poisson generic with respect to the Lebesgue measure.

Almost all numbers satisfy Lochs' theorem (1964):
Given n decimal digits $d_1, d_2, \ldots, d_n$ of $x \in [0, 1]$, and $L_n(x)$ continued fraction digits (partial quotients)

$$\frac{L_n(x)}{n} \to \frac{6 \ln 10 \ln 2}{\pi^2} \approx 0,97 \quad \text{a.e. } x \quad \text{(Lebesgue measure)}$$

when $n \to \infty$.

Let's say that x is Lochs typical.

Question: Is Poisson genericity (normality) Lochs' invariant?

If a given number x is Poisson generic in decimal and it is Lochs' typical, is it Poisson generic for continued fractions?

Lochs theorem for positive entropy numeration systems: Dajani and Fieldsteel (2001)

M O D L U X V J Y I P C Z F C C A Q F K G H Y V E M J B K D G I
E J G I P L D G Q S X R N T V D H S R Y V Y R G F A M K C H D V
Q C H Z T W B H N M X N A B M O D L U X V J Y I P C Z F C C A
Q F K G H Y V E M J B K D G I E J G H S G L U K C A Q R T C W Z
O I C B K U I U Q O A L **T H A N K Y O U** L R Z I P L D G Q S A
T M O D L U X V J I P C Z F C C A Q F K G H Y V E M J B L D G Q
S A T Z O I C B K N A B **F O R Y O U R** K D G I E J G I P H Y V E
M J B K D G I E J G H S G L U X R N T V D H S R Y V Y R G F A M
K C H D V Q C H Z T W **A T T E N T I O N** A L T H Y V E M J B
K D G I E J G H S G L U M O D L U X V J Y I P C Z F C C A Q F K G
H Y V E M J B K D G I E J G I P L D G Q S A T Z O I C B K U I U Q
O A L T H Y V E M J B K D G I E J G H S G L U U I U Q O A L T M
O D L U X V J Y I P C Z F C C A Q F K G H Y V E M J B K D G I E J
G H S G L U K C A Q R T C W L R Z I P L D G Q S A T Z O I C B K
U I U Q O A L T B I V D X W M X A T Z O I C B K . . .

Some bibliography

Poisson point processes for integer bases.

- N. Alvarez, V. Becher, and M. Mereb. Poisson generic sequences. *International Mathematics Research Notices*, rnac234, 2022.
- B. Weiss. Poisson generic points, 23-27 November 2020. Jean-Morlet Chair conference on Diophantine Problems, Determinism and Randomness. Centre International de Rencontres Math ̀lematiques. Audio-visual resource: doi:10.24350/CIRM.V.19690103.

Poisson Law in dynamical systems: two papers with an historical account:

- M. Hirata, B. Saussol, and Sandro Vaienti. Statistics of return times: A general framework and new applications. *Communications in Mathematical Physics*, 206:33â55, 1999
- M. Iosifescu and C. Kraaikamp. *Metrical Theory of Continued Fractions*. Mathematics and Its Applications. Springer Netherlands, 2002.
- R. Zweimüller. Hitting times and positions in rare events. *Annales Henri Lebesgue*, 5:1361â1415, 2022.