Normal numbers with digit dependencies

Verónica Becher

Universidad de Buenos Aires & CONICET

Joint work with Christoph Aistleitner and Olivier Carton

Equidistribution: Arithmetic, Computational and Probabilistic Aspects
Institute for Mathematical Sciences, National University Singapore
29 April-May 17, 2019

Expansion of a real number in an integer base

For a real number x, its fractional expansion in an integer base $b \geq 2$ is a sequence of integers $a_1, a_2 \ldots$, where $0 \leq a_j < b$ for every j, such that

$$x = \lfloor x \rfloor + \sum_{j=1}^{\infty} a_j b^{-j} = \lfloor x \rfloor + 0.a_1 a_2 a_3 \ldots$$

We require that $a_j < b - 1$ infinitely often to ensure that every number has a unique representation.

Borel normal numbers

A real number x is simply normal to base b if every digit in $\{0, \dots, b-1\}$ occurs in the b-ary expansion of x with the same asymptotic frequency (that is, with frequency $1/b$).

A real number x is normal to base b if it is simply normal to all the bases $b, b^2, b^3, \dots$ (Pillai 1940)

Borel normal numbers

A real number x is simply normal to base b if every digit in $\{0,\ldots,b-1\}$ occurs in the b-ary expansion of x with the same asymptotic frequency (that is, with frequency $1/b$).

A real number x is normal to base b if it is simply normal to all the bases $b, b^2, b^3, \ldots$ (Pillai 1940)

For example, $0.0101010101010\ldots$ is simply normal to base 2 but not to base 2^2 nor to base 2^3, etc.

Borel normal numbers

A real number x is simply normal to base b if every digit in $\{0, \ldots, b-1\}$ occurs in the b-ary expansion of x with the same asymptotic frequency (that is, with frequency $1/b$).

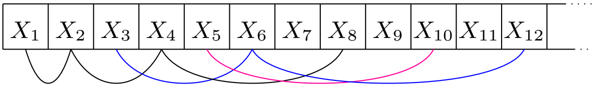
A real number x is normal to base b if it is simply normal to all the bases $b, b^2, b^3, \ldots$ (Pillai 1940)

For example, $0.0101010101010\ldots$ is simply normal to base 2 but not to base 2^2 nor to base 2^3, etc.

Émile Borel proved that almost all numbers with respect to Lebesgue measure are normal to all integer bases.

The question

How much digit dependence can be allowed so that, still, almost all real numbers are normal?



First result

Our first theorem counts how many consecutive digits have to be independent, in order to keep the property that almost all numbers are normal.

First result

Our first theorem counts how many consecutive digits have to be independent, in order to keep the property that almost all numbers are normal.

Almost all real numbers whose base b-expansion is such that for every sufficiently large position n, slightly more than $\log \log n$ consecutive positions have mutually independent digits, are normal to base b.

Independence of just $\log \log n$ consecutive digits is not sufficient.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space as $([0,1], \mathcal{B}(0,1), \lambda)$. Let integer $b \geq 2$. Let $X_1, X_2, \ldots$ be a sequence of random variables into $\{0, \ldots, b-1\}$.

Theorem 1

Assume that for every n the random variable X_n is uniformly distributed on $\{0, \ldots, b-1\}$. Assume that there exists a function $g: \mathbb{N} \mapsto \mathbb{R}$ unbounded and monotonically increasing such that for all sufficiently large n the random variables

$$X_n, \; X_{n+1}, \; \ldots, \; X_{n+\lceil g(n) \log\log n \rceil}$$

are mutually independent. Let x be the real number whose base-b expansion is given by $x = 0.X_1X_2\ldots$. Then $\mathbb{P}$-almost surely x is normal to base b.

Theorem 1, continued

On the other hand, for every base b and every positive constant K there is an example where for every $n \geq 1$ the random variable X_n is uniformly distributed on $\{0, \ldots, b-1\}$ and where for all sufficiently large n the random variables

$$X_n,\ X_{n+1},\ \ldots,\ X_{n+\lceil K \log \log n \rceil}$$

are mutually independent but $\mathbb{P}$-almost surely the number $x = 0.X_1X_2\ldots$ fails to be simply normal.

Proof of Theorem 1, simple normality to base b

Fix base b. Fix ε.
Partition the set of all positive integers in $N_1, N_2, N_3, \ldots$ such that each N_j has consecutive integers and the cardinality of N_j grows exponentially in j.
Let j be large enough.
Partition of N_j in $S_1, S_2, \ldots S_r$ each of $\lceil (\log j)/\varepsilon^2 \rceil$ consecutive integers.
By hypothesis, the variables with indices in each S are mutually independent.
Fix a digit d in base b. By Hoeffding's inequality, for each S

$$\mathbb{P}\left(\Big| \frac{1}{\#S} \sum_{n \in S} \mathbf{1}(X_n = d) - \frac{1}{b} \Big| > \varepsilon \right) \leq 2e^{-2\varepsilon^2 \#S}$$

Proof of Theorem 1, simple normality to base b

Fix base b. Fix ε.
Partition the set of all positive integers in $N_1, N_2, N_3, \ldots$ such that each N_j has consecutive integers and the cardinality of N_j grows exponentially in j.
Let j be large enough.
Partition of N_j in $S_1, S_2, \ldots S_r$ each of $\lceil (\log j)/\varepsilon^2 \rceil$ consecutive integers.
By hypothesis, the variables with indices in each S are mutually independent.
Fix a digit d in base b. By Hoeffding's inequality, for each S

$$\mathbb{P}\left(\Big| \frac{1}{\#S} \sum_{n \in S} \mathbf{1}(X_n = d) - \frac{1}{b} \Big| > \varepsilon \right) \leq 2e^{-2\varepsilon^2 \#S} < \frac{2}{j^2}.$$

Proof of Theorem 1, simple normality to base b

Fix base b. Fix ε.
Partition the set of all positive integers in $N_1, N_2, N_3, \ldots$ such that each N_j has consecutive integers and the cardinality of N_j grows exponentially in j.
Let j be large enough.
Partition of N_j in $S_1, S_2, \ldots S_r$ each of $\lceil (\log j)/\varepsilon^2 \rceil$ consecutive integers.
By hypothesis, the variables with indices in each S are mutually independent.
Fix a digit d in base b. By Hoeffding's inequality, for each S

$$\mathbb{P}\left(\Big| \frac{1}{\#S} \sum_{n \in S} \mathbf{1}(X_n = d) - \frac{1}{b} \Big| > \varepsilon \right) \leq 2e^{-2\varepsilon^2 \#S} < \frac{2}{j^2}.$$

Let Z_S be the random variable for $\displaystyle\frac{1}{\#S} \sum_{n \in S} \mathbf{1}(X_n = d) - 1/b$, we obtain

$$\mathbb{P}\left(\sum_{S \in \{S_1, \ldots, S_r\}} |Z_S| > 2\varepsilon r \right) < \frac{2}{\varepsilon j^2}.$$

Proof of Theorem 1, simple normality to base b

Fix base b. Fix ε.
Partition the set of all positive integers in $N_1, N_2, N_3, \ldots$ such that each N_j has consecutive integers and the cardinality of N_j grows exponentially in j.
Let j be large enough.
Partition of N_j in $S_1, S_2, \ldots S_r$ each of $\lceil (\log j)/\varepsilon^2 \rceil$ consecutive integers.
By hypothesis, the variables with indices in each S are mutually independent.
Fix a digit d in base b. By Hoeffding's inequality, for each S

$$\mathbb{P}\left(\left| \frac{1}{\#S} \sum_{n \in S} \mathbf{1}(X_n = d) - \frac{1}{b} \right| > \varepsilon \right) \leq 2e^{-2\varepsilon^2 \#S} < \frac{2}{j^2}.$$

Let Z_S be the random variable for $\dfrac{1}{\#S} \displaystyle\sum_{n \in S} \mathbf{1}(X_n = d) - 1/b$, we obtain

$$\mathbb{P}\left(\sum_{S \in \{S_1, \ldots, S_r\}} |Z_S| > 2\varepsilon r \right) < \frac{2}{\varepsilon j^2}.$$

Thus, $\mathbb{P}$-almost surely $\quad \left| \dfrac{1}{\#N_j} \#\{n \in N_j : X_n = d\} - \dfrac{1}{b} \right| \leq 2\varepsilon,$

Proof of Theorem 1, simple normality to base b

Fix base b. Fix ε.
Partition the set of all positive integers in $N_1, N_2, N_3, \ldots$ such that each N_j has consecutive integers and the cardinality of N_j grows exponentially in j.
Let j be large enough.
Partition of N_j in $S_1, S_2, \ldots S_r$ each of $\lceil (\log j)/\varepsilon^2 \rceil$ consecutive integers.
By hypothesis, the variables with indices in each S are mutually independent.
Fix a digit d in base b. By Hoeffding's inequality, for each S

$$\mathbb{P}\left(\left| \frac{1}{\#S} \sum_{n \in S} \mathbf{1}(X_n = d) - \frac{1}{b} \right| > \varepsilon \right) \leq 2e^{-2\varepsilon^2 \#S} < \frac{2}{j^2}.$$

Let Z_S be the random variable for $\displaystyle\frac{1}{\#S} \sum_{n \in S} \mathbf{1}(X_n = d) - 1/b$, we obtain

$$\mathbb{P}\left(\sum_{S \in \{S_1, \ldots, S_r\}} |Z_S| > 2\varepsilon r \right) < \frac{2}{\varepsilon j^2}.$$

Thus, $\mathbb{P}$-almost surely $\displaystyle\left| \frac{1}{\#N_j} \#\{n \in N_j : X_n = d\} - \frac{1}{b} \right| \leq 2\varepsilon,$

hence, $\displaystyle\left| \frac{1}{N} \#\{n : 1 \leq n \leq N, X_n = d\} - \frac{1}{b} \right| \leq 4\varepsilon.$

Proof of Theorem 1, normality to base b

The same argument yields simple normality to $b^2, b^3, b^4, \ldots$. For b^2 we have

$$(0.X_1X_2X_3X_4\ldots)_b = (0.Y_1Y_2\ldots)_{b^2}$$

where, for each $n \geq 1$,

$$Y_n = bX_{2n-1} + X_{2n}.$$

Mutual independence of

$$X_{2n-1},\ X_{2n},\ \ldots,\ X_{2n-1+\lceil g(2n-1)\log\log(2n-1)\rceil}$$

implies that there is a monotonous increasing function $\hat{g}$ such that for all sufficiently large n,

$$Y_n,\ Y_{n+1},\ \ldots,\ Y_{n+\lceil \hat{g}(n)\log\log n\rceil}$$

are mutually independent. $\square$

Proof of Theorem 1, sharp

To prove that $K \log\log n$ consecutive independent positions is not enough we give an explicit construction that fails simple normality to base b. $\square$

Toeplitz sequences (Jacobs and Keane 1969)

Fix an integer $b \geq 2$, let $A = \{0, \dots, b-1\}$ and let A^ω be the set of all infinite sequences of symbols from A.

For $P = \{2\}$, T_P is the set of all sequences $t_1 t_2 \dots$ such that, for every n,

$$t_n = t_{2n}$$

Thus,

$$t_1 = t_2 \quad = t_{2^2} \quad = \dots$$

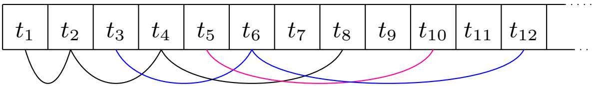
$$t_3 = t_{2\ 3} = t_{2^2\ 3} = \dots$$

$$t_5 = t_{2\ 5} = t_{2^2\ 5} = \dots$$

...

$$t_j = t_{2\ j} = t_{2^2\ j} = \dots$$

for every j that is not a multiple of 2.



Toeplitz sequences

For a positive integer r and a set $P = \{p_1, \ldots, p_r\}$ of r prime numbers, let T_P be the set of all Toeplitz sequences, that is, the set of all sequences $t_1 t_2 t_3 \cdots$ in A^ω such that for every $n \geq 1$ and for every $i = 1, \ldots, r$,

$$t_n = t_{np_i}.$$

Toeplitz transform τ_P

Let $P = \{p_1, \ldots, p_r\}$ a set of r primes. Let $j_1, j_2, j_3, \ldots$ be the enumeration in increasing order of all positive integers that are not divisible by any of the primes $p_1, \ldots, p_r$.

The Toeplitz transform $\tau_P : A^\omega \to T_P$ is defined as

$$\tau_P(a_1 a_2 a_3 \ldots) = t_1 t_2 t_3 \ldots$$

$$\text{where } t_n = a_k \quad \text{when } n \text{ has the decomposition} \quad n = j_k p_1^{e_1} \cdots p_r^{e_r}.$$

Since elements of A^ω can be identified with real numbers in $[0, 1]$ via their expansion, the transform τ_P induces a transform $[0, 1] \mapsto T_P$, which we denote by τ_P as well.

Uniform probability measure μ on T_P

Let λ be the uniform probability measure on A^ω (the infinite product measure generated by the uniform measure on $\{0, \ldots, b-1\}$).

We endow T_P with a probability measure μ, which is the forward-push by τ_P of the uniorm measure λ.

For any measurable set $X \subseteq T_P$,

$$\mu(X) = \lambda(\tau_P^{-1}(X)).$$

By identifying infinite sequences with real numbers, the measure μ on A^ω also induces a measure on $[0, 1]$, which we denote by μ as well.

Independence

The Toeplitz transform τ_P also induces a function $\delta : \mathbb{N} \mapsto \mathbb{N}$ where

$$t_1 t_2 t_3 \cdots = \tau_P(a_1 a_2 a_3 \cdots) = a_{\delta(1)} a_{\delta(2)} a_{\delta(3)} \cdots .$$

The n-th symbol $t_n(x)$ of $\tau_P(x)$, $t_n(x)$, is a random variable on the space $([0,1], \mathcal{B}(0,1), \lambda)$. That is, it is a measurable function $[0,1] \mapsto \{0, \ldots, b-1\}$.

Since $t_n(x) = a_{\delta(n)}$ for all n, two random variables t_m and t_n are independent (with respect to both measures λ and μ) if and only if $\delta(m) \neq \delta(n)$, that is, if they do not originate in the same digit of x by the Toeplitz transform.

Theorem 2

Let $b \geq 2$ be an integer, and let P be a finite set of primes. Let μ be the uniform probability measure on the set T_P. Then, μ-almost all elements of T_P are the expansion in base b of a normal number.

Proof of Theorem 2

We prove Theorem 2 by showing that it is a consequence of Theorem 1, together with a gap condition that follows from Tijdeman's results in 1973.

Let $P = \{p_1, \ldots, p_r\}$ a set of primes. Define an equivalence classs over the integers, $n \sim m$ if there are non-negative integers $k, n_1, \ldots n_r, m_1, \ldots, m_r$ such that k is not a multiple of p_1 nor p_2 nor $\ldots p_r$.

$$n = kp_1^{n_1} p_2^{n_2} \ldots p_r^{n_r} \text{ and } m = kp_1^{m_1} p_2^{m_2} \ldots p_r^{m_r}$$

There is n_0 such that if $n > m > n_0$ and $n \sim m$ then $n - m > 2\sqrt{m}$.

Fix an integer base b. Almost all real numbers whose base-b expansion, $t_1 t_2 \ldots$, is such that for every n, $t_n = t_{2n}$, are normal to every integer base.

Theorem 3

Let $b \geq 2$ be an integer, let $P = \{2\}$ and let μ be the uniform probability measure on T_P. Then, μ-almost all elements of T_P are the expansion in base b of an absolutely normal number.

About the proof of Theorem 3

A real number x is normal to base b when the fractional parts of $x, bx, b^2x, b^3x, \ldots$ are equidistributed in the unit interval. By Weyl's criterion, this amounts to bound some exponential sums.

We adapt the work of Cassels 1959 and Schmidt 1961. Our argument is also based on giving upper bounds for certain Riesz products.

About the proof of Theorem 3

A real number x is normal to base b when the fractional parts of $x, bx, b^2x, b^3x, \ldots$ are equidistributed in the unit interval. By Weyl's criterion, this amounts to bound some exponential sums.

We adapt the work of Cassels 1959 and Schmidt 1961. Our argument is also based on giving upper bounds for certain Riesz products.

Cassels worked on a Cantor-type set of real numbers whose ternary expansion avoids the digit 2 (and which therefore cannot be normal to base 3), and he established certain regularity properties of the uniform measure supported on this fractal set. In contrast, we deal with the measure μ which is the uniform measure on the set of real numbers which respect the digit dependencies.

Proof of Theorem 3

As usual, we write $e(x)$ to denote $\mathrm{e}^{2\pi i x}$.
We show that for all integers $r \geq 2$ which are multiplicatively independent of b and for all non-zero integers h,

$$\lim_{N\to\infty} \frac{1}{N} \left| \sum_{n=1}^{N} e(r^n h x) \right| = 0, \qquad \mu\text{-almost surely.}$$

Proof of Theorem 3

Lemma

Let $r \geq 2$ be multiplicatively independent to b. Then for all integers $h \geq 1$ there exist constants $c > 0$ and $k_0 > 0$, depending only on b, r and h but not on m and k, such that

$$\int_0^1 \left| \sum_{j=m+1}^{m+k} e(r^j hx) \right|^2 d\mu(x) \leq k^{2-c}$$

holds for all positive integers k, m satisfying $m \geq k + 1 + 2\log_r b \geq k_0$.

Proof of Theorem 3

Lemma (adapted form Schmidt's Hilfssatz 5, 1961)

Let r and b be multiplicatively independent bases. There is a constant $c > 0$, depending only on r and b, such that for all positive integers K and L with $L \geq b^K$,

$$\sum_{n=0}^{N-1} \prod_{\substack{k=K+1 \\ k \text{ odd}}}^{\infty} \left(\frac{1}{b} + \frac{b-1}{b} \left| \cos\left(\pi r^n L b^{-k} \right) \right| \right) \leq 2N^{1-c}.$$

The proof of Schmidt's Hilfssatz, uses that the function $|\cos(\pi x)|$ is periodic, the fact that $|\cos(\pi x)| \leq 1$ and that $|\cos(\pi / b^2)| < 1$. All these properties also hold for the function $\frac{1}{b} + \frac{b-1}{b} \left|\cos\left(\pi x\right)\right|$.

Theorem (Becher, Carton and Heiber 2016)

We construct a normal sequence in T_P for $P = \{2\}$.

Theorem (Becher, Carton and Heiber 2016)

We construct a normal sequence in T_P *for* $P = \{2\}$.

Problem

Construct a normal sequence in T_P *for* $P = \{2, 3\}$.

A conclusion

Imposing digit dependencies does not destroy the fact that almost all numbers are normal.

C. Aistleitner, V. Becher and O. Carton. Normal numbers with digit dependencies, Transactions of American Mathematical Society, in press 2019.

C. Aistleitner. Metric number theory, lacunary series and systems of dilated functions. In Uniform distribution and quasi-Monte Carlo methods, volume 15 of Radon Ser. Comput. Appl. Math., pages 1–16. De Gruyter, Berlin, 2014.

V. Becher and O. Carton. Normal numbers and computer science. In V. Berthé and M. Rigó, editors, Sequences, Groups, and Number Theory, Trends in Mathematics Series. Birkhauser/Springer, 2018.

V. Becher, O. Carton, and P. A. Heiber. Finite-state independence. Theory of Computing Systems, Theory of Computing Systems 62(7):1555–1572, 2018.

E. Borel. Les probabilités dénombrables et leurs applications arithmétiques. Rendiconti del Circolo Matematico di Palermo, 27:247–271, 1909.

Y. Bugeaud. Distribution Modulo One and Diophantine Approximation. Series: Cambridge Tracts in Mathematics 193. Cambridge University Press, 2012.

J. W. S. Cassels. On a problem of Steinhaus about normal numbers. Colloquium Mathematicum, 7:95–101, 1959.

K. Jacobs and M. Keane. 0-1 sequences of Toeplitz type. Z. Wahrsheinlichkeitstheorie verw. Geb., 13:123–131, 1969.

L. Kuipers and H. Niederreiter. Uniform distribution of sequences. Wiley-Interscience, New York, 1974.

W. M. Schmidt. Über die Normalität von Zahlen zu verschiedenen Basen. Acta Arithmetica, 7:299–309, 1961/1962.

R. Tijdeman. On integers with many small prime factors. Compositio Mathematica, 26(3):319–330, 1973.