JOURNAL DE THÉORIE DES NOMBRES DE BORDEAUX

EDGARDO UGALDE

An alternative construction of normal numbers

Journal de Théorie des Nombres de Bordeaux, tome 12, nº 1 (2000), p. 165-177

<http://www.numdam.org/item?id=JTNB_2000__12_1_165_0>

Journal de Théorie des Nombres
de Bordeaux **12** (2000), 165–177

An alternative construction of normal numbers

par EDGARDO UGALDE

RÉSUMÉ. Nous construisons une nouvelle classe de nombres normaux en base b de manière récursive en utilisant des chemins eulériens dans une suite de digraphes de de Bruijn. Dans cette construction chaque chemin est fabriqué comme une extension du chemin précédent, de telle manière que le bloc b-adique déterminé par le chemin contienne le nombre maximal de sous-blocs b-adiques distincts de longueurs consécutives dans l'arrangement le plus compact. Toute source de redondance est évitée à chaque étape. Notre construction récursive est une alternative à plusieurs constructions par concaténation à la Champernowne qui sont bien connues.

ABSTRACT. A new class of b–adic normal numbers is built recursively by using Eulerian paths in a sequence of de Bruijn digraphs. In this recursion, a path is constructed as an extension of the previous one, in such way that the b–adic block determined by the path contains the maximal number of different b–adic subblocks of consecutive lengths in the most compact arrangement. Any source of redundancy is avoided at every step. Our recursive construction is an alternative to the several well–known concatenative constructions *à la* Champernowne.

1. INTRODUCTION

Let b be a fixed integer. A number $x \in [0,1]$ is a b–adic normal if each block $q(1)q(2)\cdots q(k)$ on b symbols appears in the b–adic expansion of x with frequency b^{-n}. In [Bo] Borel proved that the set of all b–adic normals is a set of Lebesgue measure 1, but it was only 35 years after Borel's proof that Champernowne [Ch] presented the first explicit normal number in base 10. He proceeded to concatenate the decimal expressions of all natural numbers to yield the decimal expansion $0.123456789101112131415161718\cdots$.

Several other constructions of normal numbers have been proposed following Champernowne's idea, i.e., a concatenation of blocks of digits of increasing length. Besicovitch [Be] proved that concatenating the sequence of

Manuscrit reçu le 7 septembre 1999.

squares of all the natural numbers produces the normal number 0.149162536 496481100121 ⋯ . Copeland and Erdös [CE] proved that the concatenation of all prime numbers gives the normal 0.23571113171923293137 ⋯ . More recently Schiffer [Sc], and Nakai and Shiokawa [NS] proved that for a non–constant, eventually increasing polynomial p, the number $0.[p(1)][p(2)][p(3)]$ $[p(4)]\cdots$ is also a normal number. Here $[x]$ stands for the b–adic expression of the integral part of x. This construction produces both Besicovitch and Champernowne numbers.

In this paper we present a recursive construction of normal numbers which is not *à la* Champernowne. Our b–adic expansions are produced through Eulerian paths in a sequence of de Bruijn digraphs of increasing size. Thus, we call the numbers generated by our algorithm Eulerian normal numbers.

The paper is organized as follows. In Section 2 we give the main definitions. The algorithm to generate Eulerian normal numbers is presented in Section 3. This section is mainly combinatorial. Section 4 is more of a probabilistic nature. There we prove that Eulerian numbers are indeed normal and we give an estimate of their discrepancy. Finally, the rate of convergence to equidistribution for the Eulerian numbers is discussed in Section 5.

2. Definitions and notation

2.1. **Normal numbers.** Except for the rationals of the kind m/b^n, a number $x \in [0,1]$ has a unique *b-adic expansion* $x(1)x(2)\cdots x(n)\cdots$, which is the sequence on $\mathbb{Z}_b$ such that $x = \sum_{j=1}^{\infty} x(j)/b^j$. The b–adic block from site n to site $n+k$ in the b–adic expansion of x is denoted by $x(n:n+k)$.

A number $x \in [0,1]$ is said to be a *b-adic normal* number if for each finite b–adic block $q(1:k) \in \mathbb{Z}_b^k$,

$$\lim_{N\to\infty} f(x(1:N), q(1:k)) = b^{-k},$$

where $f(x(1:N), q(1:k))$ is the frequency of $q(1:k)$ as subblock of $x(1:N)$.

The rate of convergence to equidistribution is measured by the (N,k)–*discrepancy associated to* $x \in [0,1]$. It is denoted by $D_{N,k}$ and is defined as follows:

$$D_{N,k}(x) = \max\{|f(x(1:N), q(1:k)) - b^{-k}| : \ q(1:k) \in \mathbb{Z}_b^k\}.$$

We readily see that $x \in [0,1]$ is normal if and only if $\lim_{N\to\infty} D_{N,k}(x) = 0$ for all $k \in \mathbb{N}$.

2.2. **Digraphs.** A *digraph* is a couple (V,A), where V is a finite set and A a binary relation on that set. The elements of V are called *vertices*, and the ordered pairs in A are called *arrows*. We denote by $v \mapsto v'$ the arrow $(v, v') \in A$.

To a given vertex $v \in V$ we associate the sets $O(v) = \{v' \in V | \ v \mapsto v'\}$ and $I(v) = \{v' \in V | \ v' \mapsto v\}$.

Given $v \mapsto v'$ we call v' *the head* and v *the tail.*

A digraph (V, A) is *b-regular* if for each vertex $v \in V$ $|I(v)| = |O(v)| = b$.

A *path* in (V, A) is a sequence $(v_1, v_2, \dots , v_s)$ of vertices such that for each $1 \leq i < s$, $v_i \mapsto v_{i+1}$.

A path $(v_1, v_2, \dots , v_s)$ in (A, V) such that $v_s \mapsto v_1$ defines a *cycle* in (V, A), which is the equivalence class of all paths of the kind $(v_i, v_{i+1}, \dots , v_s, v_1, \dots , v_{i-1})$ obtained by cyclic permutations of the vertices in the original path.

Since a cycle is determined by one of its possible paths, we use any of its paths to denote it.

The digraph (V, A) is *connected* if for each couple v, v' of vertices there is a cycle $(v, v_1, \dots , v', v'_1, \dots , v'_\ell)$ containing them.

Let (V, A) be a connected digraph. The arrow $v \mapsto v'$ in A is a *bridge* if the digraph $(V, A \setminus \{v \mapsto v'\})$, obtained by erasing that arrow form the original digraph, is no longer connected.

A path where each vertex in V appears exactly once is a *Hamiltonian path*, while a path where each arrow appears exactly once is an *Eulerian path*. Correspondingly, Hamiltonian cycles and Eulerian cycles are cycles made respectively of Hamiltonian and Eulerian paths.

2.3. **The de Bruijn digraphs.** The *(b,k)–de Bruijn digraph*, denoted by $B(b, k)$, has vertices in the set of b–adic blocks of length k, and arrows between overlapping blocks. Thus, $p(1:k) \mapsto q(1:k)$ if and only if $p(2:k) = q(1:k-1)$. Note that $B(b, k)$ is b–regular for all k.

The family $B(b, k)$ was introduced by de Bruijn [Br] and was already used in the framework of normal numbers by Good [Go].

An Eulerian path on $B(b, k)$ defines a Hamiltonian path in $B(b, k+1)$ by identifying the arrows of the first digraph with the vertices of the second digraph. We also associate paths in $B(b, k)$ to b–adic blocks. The path $(a(1:k), a(2:k+1), \dots , a(s:s+k-1))$ in $B(b, k)$ is associated to the block $a(1:s+k-1) \in \mathbb{Z}_b^{s+k-1}$.

A b–adic block $q(1:s)$ of length $s > t$ is an *extension* of the b–adic block $p(1:t)$ if $p(1:t) = q(1:t)$. Similarly, an *extension* of a path $(v_1, v_2, \dots , v_t)$ is a path of the form $(v_1, v_2, \dots , v_t, \dots , v_s)$.

3. Construction of the Eulerian numbers

A b–adic Eulerian number contains in its b–adic expansion all the b–adic blocks of all lengths, but in the most compact possible arrangement.

The main ingredient in the construction of Eulerian numbers is the function Ext which extends b–adic blocks associated to Hamiltonian paths to blocks associated to Eulerian paths. For $b > 2$, given the b–adic block $a(1 : b^k + k - 1)$ defining a Hamiltonian path in $B(b, k)$, $\mathrm{Ext}(a(1 : b^k + k - 1)) = a(1 : b^{k+1} + k)$ is one of the possible extensions of $a(1 : b^k + k - 1)$ whose associated path in $B(b, k)$ is Eulerian.

Since there is more than one possible extension with the required property, Ext is a non–deterministic function.

In the 2-adic case this extension is not possible, but a similar algorithm can be implemented. Nevertheless, since 2–adic normal numbers can be obtained from 4–adic normals (see [Ma]), in what follows we only examine the case $b > 2$.

To construct and determine a b–adic Eulerian number $e = e(1)e(2)\cdots$, we only have to iterate the function Ext,

- $e(1 : b^2 + 1) = \mathrm{Ext}(012\dots(b-1))$,
- for $j \geq 2$, $e(1 : b^{j+1} + j) = \mathrm{Ext}(e(1 : b^j + j - 1))$.

A possible outcome of this procedure gives a number whose 3–adic expansion starts with

$$0.0210112200012120201002221110200212212110010120011010221000020220122\cdots$$

We call any number whose b–adic expansion is obtained in the way described above, an *Eulerian number.*

3.1. The non–deterministic function Ext. To complete the description of this construction, we have to prove that these extensions exist. In fact we will propose an algorithm to construct the outputs of the function Ext, following the idea in [Tu], where the following result is proved.

Theorem 1. *A digraph (V, A) has an Eulerian cycle $(v_0, v_1, \dots, v_\ell, v_0)$ if and only if (V, A) is connected and for each vertex $v \in V$ $|O(v)| = |I(v)|$.*

Note that the de Bruijn digraphs satisfy the hypotheses of this theorem, but in order to extend a Hamiltonian cycle into an Eulerian one we need a little more than this result. We prove the following.

Theorem 2. *Let (V, A) be a connected digraph with no bridges, and such that for all $v \in V$ $|O(v)| = |I(v)| \equiv N(v)$. Suppose (A, V) admits a Hamiltonian cycle defined by the path $(v_0, v_1, \dots, v_\ell)$. Then there exists an Eulerian cycle determined by a path $(v_0, v_1, \dots, v_\ell, \dots, v_\ell)$ which extends the given Hamiltonian path.*

We divide the proof into two parts. First we just apply the algorithm presented in [Tu], using the given Hamiltonian path $(v_0, v_1, \dots, v_\ell)$ as the skeleton of the construction. After that we modify the Eulerian cycle we obtained to give it the desired form.

Proof. Let us order all the arrows of A in a table.

$$\begin{array}{ccccc}
v_{\ell-1} \mapsto v_{\ell-1,1} & v_{\ell-1} \mapsto v_{\ell-1,2} & \cdots & v_{\ell-1} \mapsto v_{\ell-1,N(v_{\ell-1})-1} & v_{\ell-1} \mapsto v_{\ell} \\
v_{\ell-2} \mapsto v_{\ell-2,1} & v_{\ell-2} \mapsto v_{\ell-2,2} & \cdots & v_{\ell-2} \mapsto v_{\ell-2,N(v_{\ell-2})-1} & v_{\ell-2} \mapsto v_{\ell-1} \\
\vdots & \vdots & \vdots & \vdots & \\
v_0 \mapsto v_{0,1} & v_0 \mapsto v_{0,2} & \cdots & v_0 \mapsto v_{0,N(v_0)-1} & v_0 \mapsto v_1 \\
v_{\ell} \mapsto v_{\ell,1} & v_{\ell} \mapsto v_{\ell,2} & \cdots & v_{\ell} \mapsto v_{\ell,N(v_{\ell})-1} & v_{\ell} \mapsto v_0
\end{array}$$

In this table there is one line for each vertex in V. In the line corresponding to the vertex v we put all the arrows of the set $O(v)$ following an arbitrary order, except for the last entry, where we place the arrow appearing in the Hamiltonian cycle. Since (V, A) is not necessarily regular, different rows may have different lengths.

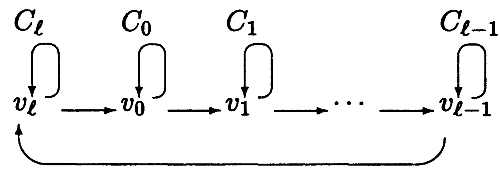
The initial Eulerian cycle: Using this table we determine an Eulerian cycle as follows.

- *Initialization:* v_{ℓ} is the first vertex of the Eulerian path.
- *Recursion:* At a given step of the construction we already have a sequence $(v_{\ell}, \dots, v)$. Let $v \mapsto v'$ be the first arrow appearing in the line of the vertex v in the table.
  - Increase the sequence $(v_{\ell}, \dots, v)$ by appending v'. So, the new sequence is $(v_{\ell}, \dots, v, v')$.
  - Erase $v \mapsto v'$ from the table.

This procedure stops at a vertex v when all the arrows of $O(v)$ have been erased from the table. When it happens, the resulting path $(v_{\ell}, \dots, v)$ has $N(v)$ arrows with v as tail, and in the last arrow of the path v is head. Since $|O(v)| = |I(v)|$, the only possibility is $v = v_{\ell}$.

Now, let us show that all the arrows of A are included in this path. Suppose $v_k \mapsto v'$ is not in $(v_{\ell}, \dots, v_{\ell})$; then neither is $v_k \mapsto v_{k+1}$, which is the last arrow in the line of v_k in the original table. Because of the choice we made for the last entries of each line of the table, this would imply that none of the arrows $v_{k+1} \mapsto v_{k+2}, v_{k+2} \mapsto v_{k+3}, \dots, v_{\ell-1} \mapsto v_{\ell}$ is in the path, but this is impossible since the line of v_{ℓ} was already erased from the table.

The sequence $(v_{\ell}, \dots, v_{\ell})$ determined by this procedure defines an Eulerian cycle which preserves the order of the arrows in the original Hamiltonian path. This cycle can be drawn as follows.



In this picture, either C_k is empty, or it is a path $(v_{k,1}, v_{k,2}, \cdots, v_{k,n_k})$ which preceded by v_k defines a cycle $(v_k, v_{k,1}, v_{k,2}, \cdots, v_{k,n_k})$. We denote this cycle by $v_k \mapsto C_k$.

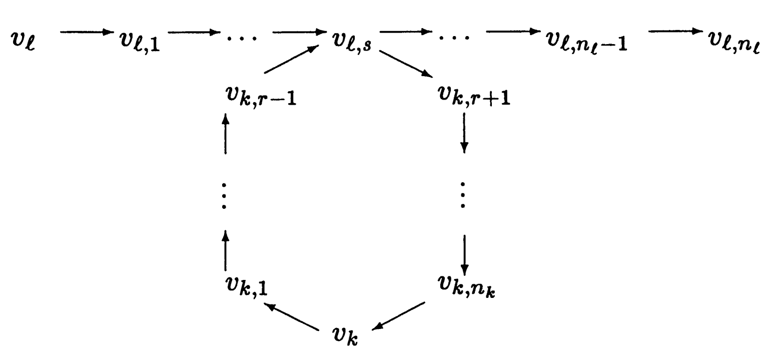
Modifying the Eulerian cycle: The following algorithm allows us to move the cycles $v_k \mapsto C_k$ for $k < \ell$, to include them inside the cycle $v_\ell \mapsto C_\ell$.

Note that for $k < \ell$, if C_k is not empty, then it only contains vertices with index larger than or equal to k. This is because when we first enter C_k, we have already finished all the lines associated to vertices v_j with $0 \leq j < k$.

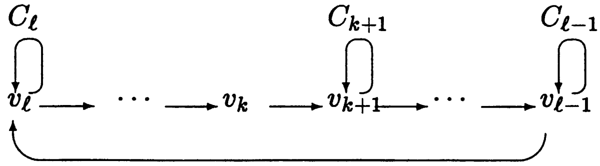
If C_k shares a vertex with C_ℓ then we can move C_k to include all its arrows inside a new C_ℓ.

- *Basic Iteration:* For $k = 0$ to $\ell - 1$, if C_k is not empty, and $v_k \mapsto C_k$ shares one vertex with $v_\ell \mapsto C_\ell$, then modify the Eulerian cycle as indicated below.

  Let $v_k \mapsto C_k \equiv v_k \mapsto v_{k,1} \mapsto \cdots \mapsto v_{k,r} \mapsto \cdots \mapsto v_{k,n_k}$ and $v_\ell \mapsto C_\ell \equiv v_\ell \mapsto v_{\ell,1} \mapsto \cdots \mapsto v_{\ell,s} \mapsto \cdots \mapsto v_{\ell,n_\ell}$ be such that $v_{k,r} = v_{\ell,s}$. Then erase C_k from the original Eulerian cycle and replace $v_\ell \mapsto C_\ell$ by the path



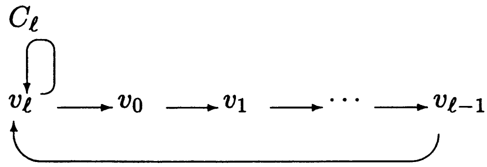
  In this way we obtain a new Eulerian cycle



where C_k is empty and the new C_ℓ has more vertices than the original one.

- *Recursion:* Repeat the previous step until none of the remaining cycles $v_k \mapsto C_k$ shares a vertex with $v_\ell \mapsto C_\ell$.

At the end of the recursion we obtain a cycle



which is defined by the sequence $(v_0, v_1, \dots, v_\ell, C_\ell, v_\ell)$. This is the desired extension.

In order to complete the proof we have to ensure that at the end of the recursion, which takes no more than $\ell + 1$ steps, we have eliminated all the cycles $v_k \to C_k$ for $0 \le k < \ell$. Suppose on the contrary that it remains a non empty path C_k, with k minimal. In that case all the vertices between the first v_k and the the first v_ℓ, have index larger or equal to k. On the other hand, between the last v_ℓ and the first v_k, all vertices have index smaller than k. In this situation all paths from v_j with $j < k$ to v_k contain the arrow $v_{k-1} \mapsto v_k$, making this arrow a bridge. This is a contradiction to one of the hypotheses of the theorem. □

Extension procedure for the de Bruijn digraphs.

Theorem 3. *Let $b \ge 3$ and $e(1 : b^k + k - 1)$ be a b–adic block defining a Hamiltonian cycle in $B(b,k)$. There exists an extension $e(1 : b^{k+1} + k)$ of $e(1 : b^k + k - 1)$, which defines an Eulerian cycle in $B(b,k)$.*

Proof. We simply apply Theorem 2. Since $B(b,k)$ is b–regular, it immediately satisfies the requirement $|O(v)| = |I(v)|$. On the other hand, it is easy to check that $B(b,k)$ is connected. The only nontrivial requirement is the lack of bridges. Let us prove that for $b \ge 3$, $B(b,k)$ does not contain any bridge.

Let us eliminate the arrow $a_1 a_2 \cdots a_k \mapsto a_2 a_3 \cdots a_{k+1}$ from $B(b,k)$. To this arrow there corresponds the b–adic block $a_1 a_2 \cdots a_{k+1} \in \mathbb{Z}_b^{k+1}$. Now take any pair of vertices $p_1p_2\cdots p_k$ and $q_1q_2\cdots q_k$ in $B(b,k)$. For any $a \in \mathbb{Z}_b \setminus \{a_1, a_{k+1}\}$, the b-adic block $p_1p_2\cdots p_k aa\cdots aq_1q_2\cdots q_k$ in $\mathbb{Z}_b^{3k}$, with a in the middle repeated k times, determines a path in $B(b,k)$ which never passes through the arrow $a_1a_2\cdots a_k \mapsto a_2a_3\cdots a_{k+1}$. Note that in the path defined by $p_1p_2\cdots p_k aa\cdots aq_1q_2\cdots q_k$, the arrows correspond to subblocks of length $k+1$, but the choice of a ensures that all those subblocks are different from $a_1a_2\cdots a_{k+1}$. Note also that the choice of a may be impossible in the case $b=2$. □

The procedure described in the proof of Theorem 2 is not adequate for computer implementation. Instead we use the Tutte's algorithm:

- Start with the block $e(1:b) = 01\cdots(b-1)$.
- At a given step of the construction we have a b-adic block $e(1:b^k+k-1)$ which defines a Hamiltonian path in $B(b,k)$.
  - Erase from $B(b,k)$ all the arrows $e(j:j+k-1)\mapsto e(j+1:j+k)$ appearing in that path.
  - Order the arrows of the new digraph in a table similar to the one we used in Theorem 2, but now, at the end of each row place the arrows appearing on a spanning tree converging to the vertex $e(b^k:b^k+k-1)$ (see [Tu] for details).
  - Following the order in this table, as we did in the proof of Theorem 2, we finally determine an Eulerian path starting at $e(b^k:b^k+k-1)$ and ending at $e(1:k)$.
  - The b-adic block associated to this path, appended to $e(1:b^k+k-1)$, is the desired extension.

4. Discrepancy and equidistribution

Our aim in this section is to prove the normality of the Eulerian numbers. The structure of the proof is the classical one given in [Ch, Be]. We use the entropic bound defined below, to estimate the (N,k)-discrepancy. Unfortunately this estimate is not the best possible one. We discuss this topic in more detail in the next section.

For all $q(1:k)\in\mathbb{Z}_b^k$ and $a(1:N)\in\mathbb{Z}_b^N$, with $k\le N\in\mathbb{N}$, let $f(a(1:N),q(1:k))$ be the frequency of $q(1:k)$ as subblock of $a(1:N)$, and

$$B_{\epsilon,k}^{(N)} = \left\{a(1:N)\in\mathbb{Z}_b^N:\ \left|f(a(1:n),q(1:k))-b^{-k}\right|\ge\epsilon,\ \forall\, q(1:k)\in\mathbb{Z}_b^k\right\}$$

the set of (ϵ,k)-bad blocks of length N.

Let

$$h_k(\epsilon) = \frac{(1+b^k\epsilon)^{1/k}\log(1+b^k\epsilon)}{bk} + \frac{b-(1+b^k\epsilon)^{1/k}}{b}\log\left(\frac{b-(1+b^k\epsilon)^{1/k}}{b-1}\right)$$

be the (ϵ, k)–entropic rate of convergence.

Theorem 4 (Entropic bound). *There are constants α and C depending only on b and k, such that*

$$\textit{for all } \epsilon > 0 \textit{ and } N \geq k, \ \left|B^{(N)}_{\epsilon,k}\right| \leq CN^\alpha b^N \exp(-Nh_k(\epsilon)).$$

Proof. Let ν be the probability distribution on $\mathbb{Z}_b^k$ generated by the b–adic block $a(1:N)$ as follows:

$$\nu : \mathbb{Z}_b^k \to [0,1] \text{ such that } \nu(q(1:k)) = f(a(1:N), q(1:k)).$$

The number of b–adic blocks generating the same distribution can be computed as follows.

Let $B_\nu(b, k-1)$ the "multi–digraph" (where an arrow may appear several times repeated) obtained from $B(b, k-1)$, by repeating each arrow $q(1:k-1) \mapsto q(2:k)$, $(N-k+1)\nu(q(1:k))$ times.

By construction, this multi–digraph allows Eulerian paths, in particular the path defined by the b–adic block of length N which originally determined ν. As in the proof of Theorem 2, to each Eulerian path there corresponds a table containing all the arrows in a given order. There are $\prod_{q(1:k-1)}\left(\sum_{q(k)}(N-s+1)\nu(q(1:k))\right)!$ possible tables of that kind, not all of them yielding an Eulerian path.

Each Eulerian path in $B_\nu(b, k-1)$ corresponds to a b–adic block generating the measure ν, but the converse is not true. All the repetitions of one arrow in $B(b, k-1)$ that we included in $B_\nu(b, k-1)$ give rise to different Eulerian paths when they are permuted in the table, but all those paths define the same b–adic block. Because of this, the number of b–adic blocks of length N generating the same distribution ν is less than

$$\prod_{q(1:k-1)} \frac{\left(\left(\sum_{q(k)}(N-k+1)\nu(q(1:k))\right)!\right)}{\left(\prod_{q(k)}\left((N-k+1)\nu(q(1:k))\right)!\right)}.$$

The Stirling approximation gives us the more useful exponential bound $C_1 b^N \exp(-Nh_k(\nu))$, with

$$h_k(\nu) = \sum_{q(1:k)} \nu(q(1:k)) \log\left(b\,\nu(q(1:k))/(\sum_{q(k)}\nu(q(1:k)))\right),$$

and C_1 a constant depending only on b and k.

On the other hand, the number of probability distributions in $\mathbb{Z}_b^k$ that can be generated by a b–adic block of length N is bounded by $C_2 N^\alpha$, where C_2 and α are constants depending only on b and k.

From these two bounds we finally deduce that

$$\left|B_{\epsilon,k}^{(N)}\right| \leq C N^\alpha b^N \exp(-N h_k(\epsilon)),$$

where $h_k(\epsilon) = \min\{h_k(\nu)|\ \max_{q(1:k)} |\nu(q(1:k)) - b^{-k}| \geq \epsilon\}$ and $C = C_1 C_2$ depends only on b and k.

It only remains to compute $h_k(\epsilon)$. We can do this by standard variational procedures, taking into account some properties satisfied by h_k that are described in [El]. Those properties allow us to restrict our search of the minimum to the set of product probability distributions in $\mathbb{Z}_b^k$. We obtain

$$h_k(\epsilon) = \frac{(1+b^k\epsilon)^{1/k}\log(1+b^k\epsilon)}{bk} + \frac{b-(1+b^k\epsilon)^{1/k}}{b}\log\left(\frac{b-(1+b^k\epsilon)^{1/k}}{b-1}\right).$$

□

Remark: A similar result can be derived from general techniques of Large Deviations Theory [El]. Here we use a combinatorial proof which follows the ideas presented in [Cs], where they derive entropic rates from the counting of possible Eulerian paths given by Tutte's theorem.

Lemma 1. *For all k, b,* $\displaystyle\lim_{\epsilon\to 0}\frac{\left(b^k\epsilon\right)^2}{2(b-1)k^2 h_k(\epsilon)} = 1.$

Proof. The entropic rate $h_k : [0, 1-b^{-k}] \to \mathbb{R}$ is smooth, convex, and increasing. Taylor's Theorem applies, and gives us the quadratic behavior

$$h_k(\epsilon) = \frac{\left(b^k\epsilon/k\right)^2}{2(b-1)} + O(\epsilon^3).$$

From this we readily obtain the result. In fact, since both $h_k(\epsilon)$ and $p_k(\epsilon) \equiv \left(b^k\epsilon/k\right)^2/(2(b-1))$ are convex, they intersect at most in two points. It is easy to verify that they intersect only at $\epsilon = 0$, and for all $\epsilon \in (0, 1-b^{-k}]$, $p_k(\epsilon) > h_k(\epsilon)$. □

4.1. Normality of the Eulerian numbers.

Theorem 5. *Let $x \in [0,1]$ be a b–adic Eulerian number, then for all k*

$$D_{N,k}(x) \equiv \max\{|f(x(1:N), q(1:k)) - b^{-k}| : \ q(1:k) \in \mathbb{Z}_b^k\} = O\left(\sqrt{\frac{\log(\log(N))}{\log(N)}}\right).$$

The implicit constant in O depends only on b and k.

Proof. Let $n \in \mathbb{N}$ be the maximal integer such that $b^n \leq N - n + 1$. By construction, $x(1 : b^n + n - 1)$ contains all the b–adic blocks of length n as subblocks, and there are b^{n-k} of those subblocks having a given $q(1:k) \in \mathbb{Z}_b^k$ as prefix. Then,

$$f(x(1 : b^n + k - 1), q(1:k)) = b^{-k} \ \forall q(1:k) \in \mathbb{Z}_b^k.$$

On the other hand, the frequency of a given $q(1:k)$ in $x(b^n + k : N)$ can be estimated from the frequency of $q(1:k)$ inside the blocks of lengths $n+1$ appearing in $x(b^n + k : N)$.

Theorem 4 ensures that if we take ϵ_n such that $h_k(\epsilon_n) \propto \log(n+2)/(n+1)$, then the number of (ϵ_n, k)–bad blocks of length $n+1$ is at most $b^{n+1}/(n+1)$. Under this condition we have

$$\left|f(x(b^n + k : N), q(1:k)) - b^{-k}\right| \leq \frac{C_1}{n+1} + \epsilon_n + \frac{\min\left(b^{n+1}/(n+1), N - b^n\right)}{N - b^n},$$

which gives

$$\left|f(x(1 : N), q(1:k)) - b^{-k}\right| \leq \frac{C_1}{n+1} + \epsilon_n + \frac{\min\left(b^{n+1}/(n+1), N - b^n\right)}{N}.$$

The last term in the right hand side is never larger than $b/(n+1)$. Note also that Lemma 1 implies that $\lim_{n\to\infty} \epsilon_n = C_2\sqrt{\log(n+2)/(n+1)}$, for some C_2 depending only on b and k. Then, for n large we can choose a constant C, depending only on k and b, such that

$$\left|f(x(1 : N), q(1:k)) - b^{-k}\right| \leq C\sqrt{\frac{\log(n+2)}{n+1}} = O\left(\sqrt{\frac{\log(\log(N))}{\log(N)}}\right),$$

for all $q(1:k) \in \mathbb{Z}_b^k$. $\square$

As an immediate corollary of this result we obtain the normality of the Eulerian numbers.

5. Concluding remarks

The estimate on the (N, k)–discrepancy given by the previous theorem is not the best possible one. It was shown in [Sc] that for a huge class of normal numbers including the Champernowne number 0.12345678910111213141516···, the best bound for the convergence rate to the equidistribution is $O(1/\log(N))$. Our numerical experiments show that the discrepancy for the Eulerian numbers should be even $o(1/\log(N))$, but at the moment we are unable to prove it.

We also did numerical experiments using the generalized Champernowne numbers (see [DK] for details), and we also found a convergence slightly faster than $O(1/\log(N))$.

Both the generalized Champernowne numbers and the Eulerian numbers have in common a random disposition of all the blocks of a given length. We think this is the source of a faster convergence. So, it is not the efficient packing of blocks in the Eulerian numbers, but the random nature of the extension function, that produces a convergence faster than $O(1/\log(N))$. We may think in a probabilistic result of the kind: "for almost all Eulerian number x, $D_{N,k}(x) = o(1/\log(N))$". In order to prove such a statement we should probably have to compute the distribution of $D_{N,k}$, considered as a random variable in $\mathbb{Z}_b^N$, and apply a kind of central limit theorem.

Acknowledgements. During the preparation of this paper I was fellow of the *Programa de Repatriación de CONACyT México* at IICO-UASLP. I also received financial support from the *Universidad Autónoma de San Luis Potosí* through the *Convenio FAI-UASLP No. C97-FAI-03-2.15.* I thank Gelasio Salazar and Jesús Urías for their careful reading and comments.

References

[Bo] E. Borel, *Sur les probabilités dénombrables et leurs applications arithmétiques.* Circ. Mat. d. Palermo **29** (1909), 247-271.

[Br] N.G. de Bruijn, *A combinatorial problem.* Konink. Nederl. Akad. Wetersh. Afd. Naturuk. Eerste Reelss, **A49** (1946), 758–764.

[Be] A.S. Besicovitch, *The asymtotic distribution of the numerals in the decimal representation of the squares of the natural numbers.* Math. Z. **39** (1935), 146–156.

[Ch] D.G. Champernowne, *The Construction of The Decimals Normal in base Ten.* J. Lond. Math. Soc. **8** (1933), 254–260.

[CE] A. Copeland, P. Erdös, *Note on normal numbers.* Bull. Amer. Math. Soc. **52** (1946), 857–860.

[Cs] I. Csiszar, T. M. Coven, B.-S. Choi, *Conditional limit theorem under Markov conditioning.* IEEE Trans. Inform. Theory **IT-33** 6 (1987), 788-801.

[DK] M. Denker, K. F. Krämer, *Upper and lower class results for subsequences of the Champernowne number.* Ergodic theory and related topics III, **LNM 1541** Springer (1992), 83–89.

[El] R.S. Ellis, *Entropy, large deviations and statistical mechanics.* Springer–Verlag, 1985.

[Go] I. J. Good, *Normal Recurring Decimals.* J. Lond. Math. Soc. **21** (1946), 167–169.

[Ma] J.E. Maxfield, *Normal k-tuples.* Pacif. J. Math. **3** (1957), 189–196.

[Sc] J. Schiffer, *Discrepancy of normal numbers.* Acta Arith. **74** (1986), 175–186.

[NS] Y.–N. Nakai&I. Shiokawa, *Discrepancy estimates for a class of normal numbers.* Acta Arith. **62** (1992), 271–284.
[Tu] R. Tutte, *Graph Theory.* Encyclopedia of Mathematics and its Applications **Vol 21** Addison Wesley, 1984.

Edgardo UGALDE
IICO–UASLP,
Av. Karakorum 1470,
Lomas 4ta
San Luis Potosí S.L.P.,
78210 MEXICO
E-mail : ugalde@cactus.iico.uaslp.mx.