



HAL
open science

Algorithmic construction of a sequence taking values in a countable alphabet with frequency preset

Anne Bertrand Mathis

► To cite this version:

Anne Bertrand Mathis. Algorithmic construction of a sequence taking values in a countable alphabet with frequency preset. 2024. ⟨hal-04375379⟩

HAL Id: hal-04375379

<https://univ-poitiers.hal.science/hal-04375379v1>

Preprint submitted on 5 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Algorithmic construction of a sequence taking values in a countable alphabet with frequency preset.

anne.bertrand@math.univ-poitiers.fr

January 5, 2024

Abstract

Let $A = \{a_n; n \geq 1\}$ be an enumerable alphabet, and let $\{\lambda_n; n \geq 1\}$ be a sequence of positive numbers such that $\sum_{n \geq 1} \lambda_n = 1$. We explain an algorithmic construction of a sequence $(u_n)_{n \geq 1}$ on the alphabet A in which each letter a_k appears with the frequency $\lambda_k = \lambda(a_k)$.

Etant donné un alphabet fini ou dénombrable $A = \{a_n; n \geq 1\}$ et une famille $\{\lambda_n; n \geq 1\}$ de nombres positifs de somme 1 nous construisons de façon simple une suite $(u_n)_{n \geq 1}$ à valeurs dans A dans laquelle chaque lettre a_k apparaît avec la fréquence $\lambda_k = \lambda(a_k)$.

Keywords: Distribution of sequences ; Numeration ; Combinatorics
 $\int \dots \int$ sur un alphabet denombrable dont chaque lettre a une frequence donnee ; Number Theory

Classification AMS : 11K99

1 Presentation of the method.

Method: Let A be a finite or countable alphabet and $(\lambda(a))_{a \in A}$ a family of positive numbers with sum 1. We order the alphabet $A: A = \{a_1, a_2, \dots, a_p, \dots\}$ so that if h is less than k then $\lambda(a_h) \leq \lambda(a_k)$; p is called the index of a_p .

We arbitrarily choose the start $u_1 \dots u_k$ of the sequence $(u_n)_{n \geq 1}$ that we want to construct. It can be a single letter u_1 . Let us assume the first M terms u_1, \dots, u_M already constructed. We call M -frequency of a letter $a \in A$ the number $\lambda_M(a) = \frac{1}{M} \sum_{u_n=a, n \leq M} 1$. We call M -deficit of a the number $D_M(a) = \lambda(a) - \lambda_M(a)$ when positive. If it is strictly positive we say that a is *late at step M* and if $D_M(a) > D_M(b)$ we say that at step M a is more late than b .

If $\lambda(a) - \lambda_M(a)$ is negative we say that a is *ahead* in step M . We then call M -*excess* de a at step M the strictly positive number

$$E_M(a) = |\lambda(a) - \lambda_M(a)| = \lambda_M(a) - \lambda(a).$$

Once u_1, \dots, u_M determined we choose for u_{M+1} that of the letters which is the most late or, if several present the same delay, the one which has the lowest

index (we can also choose at random one of the most late letters, they are in finite numbers). If none of the letters are late we choose a_1 .

Théorème 1. *Let A be a finite or countable alphabet, $(\lambda(a))_{a \in A}$ a family of positive numbers with sum 1, and let $(u_n)_{n \geq 1}$ be a constructed sequence accorded to the method above.*

Then any letter a of the alphabet A appears in the sequence $(u_n)_{n \geq 1}$ with the frequency $\lambda(a)$:

$$\forall a \in A, \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\substack{n \leq N \\ u_n = a}} 1 = \lambda(a) \quad (1).$$

This method is inspired by the work of Ville [3].

The alphabet does not required to be countable, the condition $\sum_{a \in A} (\lambda(a)) = 1$ implies that only a countable amount of letters have a strictly positive frequency.

Furthermore we can also start the sequence $(u_n)_{n \geq 1}$ with a Joker J of zero frequency wich will only appears once and has frequency zero, this is what we do if we want the first terms of the sequence $(u_n)_{n \geq 1}$ are not chosen arbitrarily (this will be the case in the study of numeration sequences).

The proof of the theorem is simple when the number of letters is finite but if the letters are in countable quantity the situation becomes complicated. Let us state some lemmas.

Lemma 1. *a) For any integer entier M the sum of the deficits of the letters of A is equal to the sum of the excesses of the letters of A .*

b) If there exists a letter b in advance with $E_M(b)$ strictly positive then there exists a late letter a with $D_M(a)$ strictly positive and vice versa.

c) The algebraic sum of the deficits and excesses of the letters of A is equal to zero:

$$\sum_{a \in A} D_M(a) - \sum_{a \in A} E_M(a) = \sum_{a \in A} \lambda(a) - \lambda_M(a) = 0.$$

Proof of assertion c). Immediate from the definitions of deficits and excesses: in fact the numbers of occurrences of the letters of the alphabet between the step 1 and M is equal to M therefore $\sum_{a \in A} \frac{\lambda_M(a)}{M} = 1 = \sum_{a \in A} (\lambda(a))$ and $\sum_{a \in A} D_M(a) - \sum_{a \in A} E_M(a) = \sum_{a \in A} (\lambda(a) - \lambda_M(a)) = 0$.

Let us call the *upper frequency* of a letter b the number $\bar{\lambda}(b) = \limsup_{N \rightarrow \infty} \frac{1}{N} \sum 1$ where the sum is taken on on the integers $n \leq N$ such that $u_n = b$. We also define the *lower frequency* of a letter. If the sequence $(u_n)_{n \geq 1}$ constructed according to the method indicated does not satisfy condition (1) it is because there exists one or more letters b belonging to A wich do not admit $\lambda(b)$ for frequency. Therefore there exists ar least one letter b whose upper frequency $\bar{\lambda}(b)$ is strictly greater than $\lambda(b)$ or a letter b whoses lower frequency is strictly less than $\underline{\lambda}(b)$. As the number of letters in infinite we are not very sure that the two cases occur simultaneously.

Remark: perhaps the assertions in lemma 1 should only be valid after a certain rank due to the selection of the first terms.

Proposition 1. For any letter of the alphabet A , $\bar{\lambda}(b) \leq \lambda(b)$

Suppose that there exists a letter b which satisfies $\bar{\lambda}(b) > \lambda(b)$.

Let us then write $\bar{\lambda}(b) = \lambda(b) + f_b$ with of course $f_b > 0$. There exists a non-empty subset B of \mathbb{N} such that the frequency of b when N tends to infinity while remaining in B is equal to $\bar{\lambda}(b)$ and we can find an integer N_0 such that:

$$\forall N > N_0, N \in B, \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\substack{n \leq N \\ u_n = b}} 1 > \bar{\lambda}(b) - \frac{f_b}{2} = \lambda(b) + \frac{f_b}{2} \quad (2).$$

As b appears an infinite number of times in the sequence $(u_n)_{n \geq 1}$ there exists an integer n_0 verifying $\frac{1}{n_0-1} < \frac{f_b}{4}$ such that $u_{n_0} = b$.

Let $M > \max(N_0, n_0)$ be an integer belonging to B . We therefore have, according to (2),

$$\frac{1}{M} \sum_{\substack{n \leq M \\ u_n = b}} 1 > \lambda(b) + \frac{f_b}{2} .$$

Let's go back to the last index m less than or equal to M such that $u_m = b$. This index m is greater than or equal to n_0 . The number of occurrences of b among the first m letters of the sequence $(u_n)_{n \geq 1}$ is equal to the number of occurrences of b among the first M letters. As more $\frac{M}{m}$ is greater or equal to 1 it comes:

$$\frac{1}{m} \sum_{\substack{n \leq m \\ u_n = b}} 1 = \frac{1}{m} \sum_{\substack{u_n = b \\ n \leq M}} 1 = \frac{M}{m} \left(\frac{1}{M} \sum_{\substack{n \leq M \\ u_n = b}} 1 \right) > \lambda(b) + \frac{f_b}{2}$$

Let's look at what happens at step $m-1$. Let us show that at this step b was already in advance. The number of occurrences of b when $n \leq m-1$ is equal to the number of occurrences of b between steps 1 and m , minus 1. Therefore

$$\frac{1}{m-1} \sum_{\substack{n \leq m-1 \\ u_n = b}} 1 = \frac{1}{m-1} \left(\left(\sum_{\substack{n \leq m \\ u_n = b}} 1 \right) - 1 \right) \geq \frac{1}{m} \left(\sum_{\substack{n \leq m \\ u_n = b}} 1 \right) - \frac{1}{m-1} \geq$$

$$\geq \lambda(b) + \frac{f_b}{2} - \frac{1}{m-1}.$$

As m is greater than or equal to $n_0 + 1$, $\frac{1}{m-1}$ is smaller than $\frac{f}{4}$ and at the step $m - 1$ the letter b admits an excess of at least $\frac{f_b}{2} - \frac{f_b}{4} = \frac{f_b}{4}$.

The sum of deficits and excesses being zero at each step according to Lemma 1, it exists at step $m - 1$ a letter a such that $D_{m-1}(a)$ is strictly positive. So the letter b could not be chosen at this step, this is impossible so no letter can have an upper density $\bar{\lambda}(b)$ greater than $\lambda(b)$ \square .

It remains to show that no letter admits a lower frequency strictly lower than its assigned frequency.

Proposition 2 *All letter of the alphabet appears at least once in the sequence $(u_n)_{n \geq 1}$.*

We will assume in the following that the letters of the alphabet are ordered by decreasing frequency: $i < j \Rightarrow \lambda(a_i) \leq \lambda(a_j)$.

Lemma 2. *The deficit at step M of a letter a is always strictly smaller than $\lambda(a)$. Therefore the sum of the deficits of all the letters is at most 1.*

Proof of lemma 2. At step M the deficit $D_M(a)$ of the letter a is equal to $\lambda(a) - \lambda_M(a)$ when his term is positive; as $\lambda_M(a)$ is always positive or zero, this term is therefore always less than or equal to $\lambda(a)$ and so the sum of the deficit is at most 1.

Lemma 3. *Suppose that at a step M the deficit of the letter b is greater or equal to g ; then no letter a with $\lambda(a) < g$ can proceed to this step M .*

Proof of lemma 3. According to lemma 2 the deficit of a would then be less than or equal to $\lambda(a)$, which is strictly less than the deficit of b , a cannot pass.

Proof of Proposition 2.

Suppose there is a letter a_i which never pass. It is this letter that we will consider. It is assumed that $\lambda(a_1) \geq \lambda(a_2) \geq \dots \geq \lambda(a_i) \geq \lambda(a_{i+1}) \dots$.

Suppose that if several letters admit the same frequency as a_i their indices are $i - k, \dots, i - 1, i$. None of the letters a_j such that $\lambda(a_j) < \lambda(a_i)$ can pass, therefore none of the letters a_{i+1}, a_{i+2}, \dots will pass because $\lambda(a_{i+1}) < \lambda(a_i)$ and so tight now. Therefore all their excess are equal to zero and their deficits are all strictly lower than their assigned frequency according to Lemma 2.

Let N_0 be such that if $N \geq N_0$ the sum of the excesses of the letters $a_1, a_2, \dots, a_{i-k}, \dots, a_i$ is less than $\frac{1}{4}\lambda(a_i)$. These letters are finite in number so proposition 1 ensures the existence of N_0 .

Lemma 1a show that as long as $a_i, a_{i+1} \dots$ have never passed the sum of the excess of letters $a_1, a_2, \dots, a_{i-k}, \dots, a_{i-1}$ is equal to the sum of their deficits increased by the sum of the deficits of the letters a_{i+1}, a_{i+2}, \dots . The sum of these deficits is therefore less than $\frac{1}{4}\lambda(a_i)$, so after step N_0 the deficits of the letters $a_1, a_2, \dots, a_{i-k}, \dots, a_{i-1}$ cannot be greater than $\frac{1}{4}\lambda(a_i)$.

For $N > N_0$ the sum of excesses is less than $\frac{1}{4}\lambda(a_i)$, so no letter can have a deficit greater than $\frac{1}{4}\lambda(a_i)$ at step N . But if a_i never pass his deficit tend to $\lambda(a_i)$ as N goes to infinity. This is impossible so a_i appears in the sequence.

The following lemma is obvious:

Lemma 4 *When the letter a_i passes for the first time to stage N_i all the other letters have a deficit less or equal to $\lambda(a_i)$.*

Proposition 3. No letter a can admit a lower frequency strictly smaller than its assigned frequency : $\forall a \in A, \underline{\lambda}(a) \geq \lambda(a)$.

Lemma 5. For any integer M and any letter b , $|D_{M+1}(b) - D_M(b)| \leq \frac{1}{M+1}$ each time these two quantities are defined.

Proof of lemma 5. When we add a b to step M the deficit of b decreases of $\frac{(\text{number of } b \text{ between } 1 \text{ and } M)+1}{M+1} - \frac{\text{number of } b \text{ between } 1 \text{ and } M}{M} = \frac{M - \text{number of } b \text{ between } 1 \text{ and } M}{M(M+1)}$ wich is smaller than $\frac{1}{M+1}$. When we add a letter different from b the variation of the deficit is less than $\frac{\text{number of } b}{M} - \frac{\text{number of } b}{M+1} = \frac{\text{number of } b}{M(M+1)}$ who is also less than $\frac{1}{M+1}$.

Proof of proposition 3. Let B be the set of numbers γ de $]0, 1]$ such that there exists a_γ whose lower frequency is strictly lower than $\lambda(a_\gamma)$: $\underline{\lambda}(a_\gamma) = \lambda(a_\gamma) - \gamma$. As the sum of the frequencies is equal to 1, the number of letters whose frequency exceed a fixed number d is finite and B cannot admit an accumulation point except perhaps zero 0 : if B is non-empty B admits a maximum element c so there exist at least one letter a_c such that $\underline{\lambda}(a_c) = \lambda(a_c) - c$ and perhaps a finite number of such letters.

Suppose there is only one letter, a_c , such that $\underline{\lambda}(a_c) = \lambda(a_c) - c$. Let P be the (finite) set of letters a such that $\lambda(a) - \underline{\lambda}(a) \in]0, c[$ and let b be the maximum of the quantities $\underline{\lambda}(a)$ for $a \in P$. So $b < c$. Let Q be the set of letters whose frequency is greater or equal to b and such that $\underline{\lambda}(a) = \lambda(a)$ (and is also equal to $\bar{\lambda}(a)$, see Proposition 1); Q is finite because the sum of the frequencies is 1.

There exists an integer N_0 such that for all $N > N_0$, for any letter w in $P \cup Q$, the deficit $D_N(w)$ of the letter w is less than $b + \frac{c-b}{2}$, those of P because their defects have an upper limit at most b and those of the finite set Q because if N is large enough they are closed of their assigned frequency ($\underline{\lambda}(a) = \lambda(a) = \bar{\lambda}(a)$ if $a \in Q$ and Q is finite, use proposition 1).

Let us place ourselve at a step N_i with $N_i > N_0$ where the letter a_i with $\lambda(a_i) < \frac{b}{2}$ passes for the first time, and such that the deficit of a_c exceed the value $c - \frac{c-b}{20}$ at a step between steps N_i et N_{i+1} (such N_i exist because the sequence N_i tends toward infinity). The letter a_c admits at step N_i a deficit lower than $\frac{b}{2}$ according to lemma 4. The deficit of a_c must therefore increase between N_i et N_{i+1} , very gradually due to lemma 5.

Suppose that at step $M \in [N_i, N_{i+1}]$ the deficit of a_c is between $c - \frac{c-b}{10}$ and $c - \frac{c-b}{20}$; then the letter a_c must pass because no other letter has a greater deficit (lemma 3), and it passes until the deficit of a_c , wich decrease, returns under $c - \frac{c-b}{20}$; this deficit will therefore never be able to rise to exceed $c - \frac{c-b}{20}$. So the hypothesis is impossible, there cannot be a single letter such that the deficit of this letter a_c admits upper limit c ; so the letter a_c cannot have a lower frequency $\underline{\lambda}(a_c) = \lambda(a_c) - c$.

Now look at the case where the maximum c of $\lambda(a) - \underline{\lambda}(a)$ is reached for several letters, and where all the deficits of the other letters are smaller than a common value $b < c$. The sets B, P, Q and the numbers $b < c$ are defined as above and let R designates the finite set of letters a of the alphabet such that $\underline{\lambda}(a) = \lambda(a) - c$ of R .

We also choose for N_i an integer as large as necessary wich represents the

first passage of a letter a_i of assigned frequency less than the common value b and such that at least one of the letters of R such that $\underline{\lambda}(a) = \lambda(a) - c$ has a deficit exceeding $c - \frac{c-b}{20}$ at a step between steps N_i and N_{i+1} (such N_i must exist). Letters of P and Q always have deficits less than $b + \frac{c-b}{2}$ for all $N > N_i$. At least one of the letters of R , say w , see his deficit starting from very low (lemma 4) and slowly rising to $c - \frac{c-b}{20}$; it passes if no letter has a greater deficit. The only letters wich can have a greater deficit are those of R , with $\underline{\lambda}(a) = \lambda(a) - c$ so $\lambda(a) > c$ and are in number k such that $\sum_{a \in R} \lambda(a) < 1$ and then $kc < 1$. If one letter of R passes his deficit decreases and the deficits of the others letters of R increase (see the proof of lemma 5); when a letter pass his defect decreases of ν and the sum of the other deficits increases but increases less than ν (because all the excesses decrease and the lemma 1). As long as this phenom lasts, no letter acquires excess and all letters in excess see their excess diminish; as the sum of excesses is equal to the sum of deficits, the sum of deficits decreases. No letter can see his deficit decrease much because this cause other letters to pass and this deficit decreases a little each time. The existence of such N_i is impossible and no letter can verify $\underline{\lambda}(a) = \lambda(a) - c$ if c is the maximum conceivable value. But in fact it means that no letter cans verify $\underline{\lambda}(a) < \lambda(a)$. The theorem is proved.

2 Examples and applications.

2.1 Approximation of measures on the torus \mathbb{R}/\mathbb{Z} by measures associated with sequences $(\beta^n)_{n \geq 1}$ modulo one.

Here we shall use only finite alphabets in theorem 1, but the consequences (the very lardge variety of measures on the torus \mathbb{R}/\mathbb{Z} associated with the exponential sequences $\{\beta^n\}_{n \geq 1}$ when β travels $]1, \infty[$) is interesting.

Given a real number x , let $\{x\}$ denotes his fractional part and δ_y denotes the Dirac measure at point y .

A sequence $(u_n)_{n \geq 1}$ of reals numbers is said to be distributed according to the probability measure ν on the torus if the sequence of measures $\nu_N = \frac{1}{N} (\delta_{\{u_1\}} + \dots + \delta_{\{u_N\}})$ weakly converge toward ν . In general a sequences does not admit a distribution measure but there always exists a set of measures that we call be associated with this sequence: *we say that the measure γ is associated with the sequence $(u_n)_{n \geq 1}$ if it exists an infinite subset B of \mathbb{N} such that $\gamma = \lim_{N \rightarrow \infty} \frac{1}{N} (\delta_{\{u_1\}} + \dots + \delta_{\{u_n\}})$.*

$$N \in B$$

Théorem 2. *Let μ be a probability measure on the torus. then there exists a sequences of real numbers $\{\beta_k\}_{k \geq 1}$ and for all k a measure μ_k associated with the sequence $(\{\beta_k^n\})_{n \geq 1}$ such that μ is the weak limit of the sequence of measures μ_k .*

In other words the set of associated measure of sequences $(\{\beta^n\})_{n \geq 1}$ is dense in the set of probabilities measures on the torus (in the sense of weak conver-

gence).

Remark: For almost all $\beta > 1$ the sequence $(\beta)_{n>0}^n$ is uniformly distributed modulo one. The incriminated β_k are generally part of the exceptional set.

Proof : We will use the following idea:

Proposition 4 [1] *Given a sequence $(I_m)_{m \geq 1}$ of closed intervals I_m of the of the torus \mathbb{R}/\mathbb{Z} , of the same length (or of length greater than a positive number c), there exists a real number β such that for all $m \geq 1$, β^m belongs to I_m .*

Proof of the Theorem. Let us fix a measure μ on the torus; if she has atoms let $(y_n)_{n \geq 1}$ be these atoms; they are finite or countably in number.

Fix an integer k , then let us divide the torus into a finite numbers of intervals $I_1^k, I_2^k, \dots, I_{p_k}^k$ whose union is the entire torus and whose possible intersections two by two are reduced at their ends. We choose their lengths greater than a number $c_k > 0$ and such that no atom y_i , if any, is one of the ends of the intervals. For any integer i less than or equal to p_k we set $\lambda^k(i) = \mu(I_i^k)$. The choice of the $(I_i^k)_{i=1, \dots, p_k}$ guarantees that the sum of the $\lambda^k(i)$ is equal to 1.

Still for this fixed k apply the theorem 1 to the alphabet $I_1^k, I_2^k, \dots, I_{p_k}^k$, assigning to I_i^k the frequency $\lambda^k(i) = \mu(I_i^k)$; construct with the Theorem 1 a sequence of intervals $(u_n^k)_{n \geq 1}$ where u_n^k is an interval belonging to the alphabet $I_1^k, I_2^k, \dots, I_{p_k}^k$; u_n^k take the value I_i^k with the frequency $\lambda^k(i)$.

According to the proposition 4 there exists a real number β_k such that, for all n , $\{\beta_k^n\}$ falls in the interval u_n^k . The sequence $(\{\beta_k^n\})_{n \geq 1}$ does not always admit a distribution measure but she always has an associated measure ν_k ; as $(u_n) = I_i^k$ with the frequency $\lambda^k(i) = \mu(I_i^k)$, for all interval $I_1^k, I_2^k, \dots, I_{p_k}^k$, $\nu_k(I_i^k) = \mu(I_i^k)$ (recall that the intervals $I_1^k, I_2^k, \dots, I_{p_k}^k$ are disjoint relatively to the measure μ).

Let us increase k and choose the $I_1^k, I_2^k, \dots, I_{p_k}^k$ so that the limit of the σ -algebra generated by the set $\{I_1^k, I_2^k, \dots, I_{p_k}^k\}$ give rise to the Borelians of the torus, it then comes that $\lim_{k \rightarrow \infty} \nu_k = \mu$.

2.2 Numeration.

Let β be an integer; we shall take $\beta = 10$ but β could be any integer. Let $(u_n)_{n \geq 1}$ be a sequence wich begins with a Joker $u_1 = J$ (J is a symbol wich is only encoutered there, it correspond to zero: no digit has yet grow but zero is part of the sequence of numeration in base 10). Then we look at the sequence of expansions of the integers in base 10 and we write $u_N = a_i$ if to go from $N - 1$ to N we increase of 1 the i -th digit starting from the right (that of the 10^{i-1}) or we make it appears if there was none 10^{i-1} before (and all the digits on the right of the $(i - 1)$ th digit goes to zero). The beginning of the sequence is

$u_1 u_2 u_3 u_4 u_5 u_6 u_7 u_8 u_9 u_{10} u_{11} u_{12} u_{13} u_{14} u_{15} u_{16} u_{17} u_{18} u_{19} u_{20} u_{21} u_{22} \dots u_{101} \dots u_{111} u_{112} u_{113} \dots =$
 $J a_1 a_1 a_1 a_1 a_1 a_1 a_1 a_1 a_1 a_2 a_1 a_1 a_1 a_1 a_1 a_1 a_1 a_2 a_1 \dots a_3 a_1 a_1 a_1 a_1 a_1 a_1 a_1 a_2 a_1 a_1 \dots$

We obtain the same sequence by constructing the democratic sequence starting with a Joker (step 1) and such that the democratic process begin at step 2 with the alphabet $(a_n)_{n \geq 1}$ with frequency $\frac{9}{10}$ for the symbol a_1 , $(\frac{9}{10})\frac{1}{10}$ for a_2 , $(\frac{9}{10})\frac{1}{10^2}$ for a_3, \dots , $(\frac{9}{10})\frac{1}{10^{n-1}}$ for a_n and so on; the n -th digit appears 10 time less

than the $(n - 1)$ th.

This is also an automatic sequence: let $(a_n)_{n \geq 1}$ an alphabet, let σ be the morphism $a_1 \rightarrow a_1 \dots a_1 a_2$, $a_2 \rightarrow a_1 \dots a_1 a_3$, and more generally $a_n \rightarrow a_1 \dots a_1 a_{n+1}$ (nine a_1 at each time). Let $(u_n)_{n \geq 1}$ be the fix point of the substitution σ starting with a_1 and let a_i represent “the i -th digit is growing at step n if and only if $u_n = a_i$ ”. The democratic sequence is equal to the sequence beginning with the Joker followed by the fixed point.

Proposition 4. Let $A = (a_1, a_2, \dots, a_n, \dots)$ an enumerable alphabet, and let $\frac{1}{k}(1 - \frac{1}{k}), \frac{1}{k^2}(1 - \frac{1}{k}), \dots, \frac{1}{k^n}(1 - \frac{1}{k}), \dots$ a sequence of numbers with sum 1. Let $(u_n)_{n \geq 1}$ be the associated democratic sequence beginning by a Joker. If $u_n = i$, the i -th digit is growing at the step n .

You can easily check it in the case where β is an integer; we conjecture it for all $\beta > 1$ (see [2] for numeration in non integer base).

In the case where this would be true for all $\beta > 1$, maybe it would help to prove another conjecture: does the β -shift is the symbolic system with entropy $\log \beta$ having the minimum of n -words? (this is true for $\beta \in \mathbb{N}$).

There is also an interesting problem : is the convergence toward the right frequencies fast? That might have an impact on the previous question.

Bibliographie

[1]A. Bertrand Mathis. Sur les parties fractionnaires des suites $(\beta^n)_{n \geq 1}$, *Uniform Distribution Theory* **14** n°2 (2019),69 -72.

[2]A. Bertrand-Mathis. Comment écrire les nombres entiers dans une base qui n'est pas entière. *Acta Math.Hungar.* **54** (1989) n°3-4 237-241.

[3]J.A. Ville. Etude critique de la notion de collectif, Thèse, Faculté des Sciences de Paris, 1939.