



HAL
open science

A central limit theorem for the variation of the sum of digits

Yohan Hosten, Élise Janvresse, Thierry de la Rue

► **To cite this version:**

Yohan Hosten, Élise Janvresse, Thierry de la Rue. A central limit theorem for the variation of the sum of digits. 2021. hal-03416309v1

HAL Id: hal-03416309

<https://hal.science/hal-03416309v1>

Preprint submitted on 8 Nov 2021 (v1), last revised 31 Aug 2023 (v2)

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.

A central limit theorem for the variation of the sum of digits

Thierry DE LA RUE¹, Élise JANVRESSE², and Yohan HOSTEN²

¹LMRS, Université de Rouen Normandie, CNRS UMR 6085,
Avenue de l'Université, 76801, Saint Étienne du Rouvray, France,
Thierry.de-la-Rue@univ-rouen.fr

²LAMFA, Université de Picardie Jules Verne, CNRS UMR 7352,
33, rue Saint-Leu, 80000, Amiens, France,
elise.janvresse@u-picardie.fr and yohan.hosten@u-picardie.fr

Abstract. We prove a Central Limit Theorem for probability measures defined via the variation of the sum-of-digits function, in base $b \geq 2$. For $r \geq 0$ and $d \in \mathbb{Z}$, we consider $\mu^{(r)}(d)$ as the density of integers $n \in \mathbb{N}$ for which the sum of digits increases by d when we add r to n . We give a probabilistic interpretation of $\mu^{(r)}$ on the probability space given by the group of b -adic integers equipped with the normalized Haar measure. We split the base- b expansion of the integer r into so-called “blocks”, and we consider the asymptotic behaviour of $\mu^{(r)}$ as the number of blocks goes to infinity. We show that, up to renormalization, $\mu^{(r)}$ converges to the standard normal law as the number of blocks of r grows to infinity. We provide an estimate of the speed of convergence. The proof relies, in particular, on a ϕ -mixing process defined on the b -adic integers.

Keywords. Sum of digits, Central Limit Theorem, b -adic odometer, ϕ -mixing.

MSC (2020). 11A63, 37A44 and 60F05.

1 Introduction

Throughout this article, *integers* mean elements of the set $\mathbb{N} := \{0, 1, 2, \dots\}$, and b denotes a fixed integer, $b \geq 2$.

For an integer n , we consider the associated sequence of *digits* $(n_k) \in \{0, \dots, b-1\}^{\mathbb{N}}$, finitely many of them being strictly positive, such that

$$n = \sum_{k \geq 0} n_k b^k.$$

For $n \neq 0$ and $\ell := \max\{k : n_k \neq 0\}$, we introduce the notation $\overline{n_\ell \cdots n_0} := n$, which generalises the usual way we write numbers in base 10, and which we refer to as the (*base b*) *expansion* of n . By convention, we set $\overline{0} := 0$. Then, we define the *sum-of-digits* function (in base b) as

$$s(n) := \sum_{k \geq 0} n_k.$$

A central object in our paper is the variation of the sum of digits when we add a fixed integer r to n : for $r, n \in \mathbb{N}$, we set

$$\Delta^{(r)}(n) := s(n+r) - s(n). \quad (1)$$

An interesting feature of $\Delta^{(r)}$ is that it gives the number of carries created during the addition $n+r$ in base b . To be more precise, if c is the number of carries then

$$\Delta^{(r)}(n) = s(r) - c(b-1). \quad (2)$$

Bésineau [1] proves that, for every $d \in \mathbb{Z}$, the following asymptotic density exists

$$\mu^{(r)}(d) := \lim_{N \rightarrow +\infty} \frac{1}{N} \left| \left\{ n < N : \Delta^{(r)}(n) = d \right\} \right|$$

and he studies these asymptotic densities through their correlation function. Actually, since $\sum_{d \in \mathbb{Z}} \mu^{(r)}(d) = 1$, the function $\mu^{(r)}$ defines a probability measure on \mathbb{Z} .

Morgenbesser and Spiegelhofer [6] show an amazing property: the measure $\mu^{(r)}$ remains the same if we reverse the order of the digits in the expansion of r . They call it the *reverse property*.

In the particular case $b = 2$, Emme and Prikhod'ko [4] show that the variance of $\mu^{(r)}$ is bounded from above and below by a constant multiplied by the number of blocks of 1's in the binary expansion of r . In Section 3, we extend this result to each $b \geq 2$.

Also in the binary case, Emme and Hubert [3] show that, for almost every sequence of integers $(r_n)_{n \in \mathbb{N}}$ written in binary and defined via a balanced Bernoulli process, the sequence of measures $(\mu^{(r_n)})_{n \in \mathbb{N}}$, after renormalization, converges in distribution to the standard normal law. The proof is done by computing all the moments of $\mu^{(r_n)}$ and by showing that, after renormalization, they converge to the moments of the standard normal law. In our paper, we prove a more accurate and more general Central Limit Theorem (CLT).

To do so, we study the variations of the sum of digits in the context of an appropriate probability space. We consider the compact additive group $(\mathbb{X}, +)$ of *b-adic integers*. The space \mathbb{X} is endowed with the Borel σ -algebra and its normalized Haar measure \mathbb{P} .

We extend $\Delta^{(r)}$ almost everywhere on \mathbb{X} and show in Section 2 (Proposition 2.1) that, for every $d \in \mathbb{Z}$

$$\mu^{(r)}(d) = \mathbb{P} \left(\{x \in \mathbb{X} : \Delta^{(r)}(x) = d\} \right).$$

To state our main result, we have to define the notion of *blocks* in the base b expansion of an integer r .

Definition 1.1. A block in the expansion $\overline{r_\ell \cdots r_0}$ of an integer $r \in \mathbb{N}$ is defined as follows: it is either

1. a maximal sequence of consecutive digits equal to 0 (“block of 0’s”) or
2. a maximal sequence of consecutive digits equal to $b-1$ (“block of $(b-1)$ ’s”) or
3. when $b \geq 3$, a digit between 1 and $b-2$ (“single-digit block”).

We also define the quantity $\rho(r)$ as the number of blocks in the base b expansion of r .

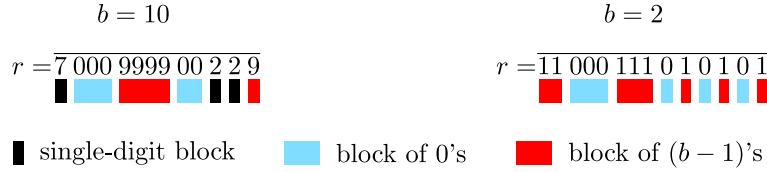


Figure 1: Examples of the decomposition in blocks in decimal and binary bases. On the left-hand side, $\rho(r) = 7$. On the right-hand side, $\rho(r) = 9$.

We specify that a block of 0’s or of $(b-1)$ ’s of length 1 is not considered, in this paper, as a single-digit block.

The following theorem, which generalizes Emme and Prikhod’ko’s result, states that the number of blocks $\rho(r)$ controls the variance of $\mu^{(r)}$.

Theorem 1.2. For every integer $r \geq 1$

$$\frac{b}{4}\rho(r) \leq \text{Var}(\mu^{(r)}) \leq 2b^2\rho(r).$$

Now, we need to introduce, for an integer $r \geq 1$, the standard deviation $\sigma_r := \sqrt{\text{Var}(\mu^{(r)})} > 0$ and the renormalized measure $\tilde{\mu}^{(r)}$ which is the measure on \mathbb{R} concentrated on the points of the form $\frac{d}{\sigma_r}$ ($d \in \mathbb{Z}$) and which satisfies

$$\forall d \in \mathbb{Z}, \quad \tilde{\mu}^{(r)}\left(\frac{d}{\sigma_r}\right) := \mu^{(r)}(d).$$

Our main result states that, for an integer $r \geq 1$, the renormalized measure $\tilde{\mu}^{(r)}$ converges in distribution to the standard normal law as the number of blocks tends to infinity.

Theorem 1.3. We have the convergence

$$\tilde{\mu}^{(r)} \xrightarrow[\rho(r) \rightarrow +\infty]{d} \mathcal{N}(0, 1).$$

Theorem 1.3 can be seen as a direct consequence of the following theorem, which furthermore provides an estimation of the speed of convergence.

Theorem 1.4. *Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a thrice differentiable function with $\|h'''\|_\infty < \infty$. Let Z be a random variable following $\tilde{\mu}^{(r)}$ and Y a standard normal random variable. Then*

$$\left| \mathbb{E}(h(Z)) - \mathbb{E}(h(Y)) \right| = O_{\rho(r) \rightarrow \infty} \left(\frac{1}{\sqrt{\rho(r)}} \right). \quad (3)$$

Furthermore, if we denote by F_r (respectively F) the cumulative distribution function of $\tilde{\mu}^{(r)}$ (respectively $\mathcal{N}(0, 1)$), then there exists $\tilde{K} > 0$ such that for every integer $r \geq 1$

$$\sup_{t \in \mathbb{R}} |F_r(t) - F(t)| \leq \frac{\tilde{K}}{\rho(r)^{\frac{1}{8}}}. \quad (4)$$

A result in the same spirit has recently been published by Spiegelhofer and Wallner [7]. In the case of the base 2, they give a very accurate estimation of the measure $\mu^{(r)}(d)$ for every $d \in \mathbb{Z}$

$$\mu^{(r)}(d) = \frac{1}{\sigma_r \sqrt{2\pi}} e^{-\frac{d^2}{2\sigma_r^2}} + O_{\rho(r) \rightarrow \infty} (\rho(r)^{-1} (\log(\rho(r)))^4). \quad (5)$$

Their result is proved using a combination of several techniques such as recurrence relations, cumulant generating functions, and integral representations. It seems possible but extremely difficult to generalize (5) to other bases. It also implies a CLT when $\rho(r)$ tends to infinity: using (5), it is possible to show that for every real numbers $a < b$

$$\tilde{\mu}^{(r)}([a, b]) \xrightarrow{\rho(r) \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt$$

with a speed of convergence of $\frac{\log^4(\rho(r))}{\rho(r)^{\frac{1}{2}}}$. However, it is not clear how we can get a speed of convergence of $F_r(t)$ to $F(t)$. On our side, we use a drastically different approach which applies directly in any base, relying on the concept of ϕ -mixing process and on a result from Sunklodas [8].

Roadmap

Section 2 is devoted to placing the study of the measures $\mu^{(r)}$ in the context of the odometer on the set \mathbb{X} of b -adic integers. We extend $\Delta^{(r)}$ almost everywhere on \mathbb{X} and we show that the convergence

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n < N} f(\Delta^{(r)}(n)) = \int_{\mathbb{X}} f(\Delta^{(r)}(x)) d\mathbb{P}(x)$$

(where \mathbb{P} is the normalized Haar measure on \mathbb{X}) is satisfied for functions $f : \mathbb{Z} \rightarrow \mathbb{C}$ of polynomial growth (Proposition 2.1) and, more generally, for functions f such that $f \circ \Delta^{(r)}$ is integrable (Proposition 2.4). We deduce from Proposition 2.1 that $\mu^{(r)}(d) = \mathbb{P}(\{x \in \mathbb{X} : \Delta^{(r)}(x) = d\})$ and that $\mu^{(r)}$ has finite moments. In particular, we show that $\mu^{(r)}$ is of zero-mean.

In Section 3, we focus on the second moment of $\mu^{(r)}$. First, we establish in Proposition 3.1 an inductive relation between the measures in the spirit of Bésineau’s result [1, p.13]. From that, we deduce an inductive relation on the variance of the measures (Lemma 3.3). Then, we prove the estimation of the variance stated in Theorem 1.2.

In Section 4, we build a finite sequence of random variables associated to the addition of some integer r that will be used to prove Theorem 1.4. We estimate the ϕ -mixing coefficients for this sequence.

The last section is devoted to the proof of Theorem 1.4. We show how we can apply a result from Sunklodas [8, Theorem 1] giving a speed of convergence in the CLT for ϕ -mixing process.

Acknowledgment

We thank M. El Machkouri for fruitful discussions and references about the speed of convergence in CLT for mixing sequences.

2 Odometer and sum-of-digits function

2.1 Unique ergodicity of the b -adic odometer

We define \mathbb{X} as the space of b -adic integers, that is the space $\{0, \dots, b-1\}^{\mathbb{N}}$. Coordinates of a b -adic integer $x \in \mathbb{X}$ are interpreted as digits in base b : elements of \mathbb{X} can be viewed as “generalized integers having possibly infinitely many non zero digits in base b ”. To comply with the usual writing of numbers in base 10, an element $x = (x_n)_{n \in \mathbb{N}} \in \mathbb{X}$ will be represented as a left-infinite sequence (\dots, x_1, x_0) , x_0 being the units digit. The space \mathbb{X} is compact for the product topology. The set of integers \mathbb{N} can be identified with the subset of sequences with finite support. More precisely, using the inclusion function

$$i : \overline{n_\ell \dots n_0} \in \mathbb{N} \mapsto (\dots, 0, n_\ell, \dots, n_1, n_0) \in \mathbb{X}$$

we identify \mathbb{N} and $i(\mathbb{N})$.

\mathbb{X} is equipped with an addition which extends the usual addition on \mathbb{N} and turns \mathbb{X} into an Abelian group. For $x = (\dots, x_0)$ and $y = (\dots, y_0)$ in \mathbb{X} , $(x+y)$ is determined recursively by the following process, in which we generate a sequence of carries $(c_\ell)_{\ell \geq 0} \in \{0, 1\}^{\mathbb{N}}$.

- Initialisation:
 - if $x_0 + y_0 < b$ then we set $(x+y)_0 := x_0 + y_0$ and $c_0 := 0$,
 - else we set $(x+y)_0 := x_0 + y_0 - b$ and $c_0 := 1$.

- Induction step: once, for some $\ell \geq 1$, we have computed $(x + y)_i$ and c_i for $i = 0, \dots, \ell - 1$,
if $x_\ell + y_\ell + c_{\ell-1} < b$ then we set $(x + y)_\ell := x_\ell + y_\ell + c_{\ell-1}$ and $c_\ell := 0$,
else we set $(x + y)_\ell := x_\ell + y_\ell + c_{\ell-1} - b$ and $c_\ell := 1$.

Now, since 1 belongs to $\mathbb{N} \subset \mathbb{X}$, we can consider the application $T : \mathbb{X} \rightarrow \mathbb{X}$

$$T : x \mapsto T(x) := x + 1,$$

which is usually referred to as the *b-adic odometer*. It is well-known that T is a homeomorphism and so (\mathbb{X}, T) is a topological dynamical system [5]. For $\ell \geq 0$ and for integers $r_\ell, \dots, r_0 \in \{0, \dots, b-1\}$, we define the *cylinder* $C_{r_\ell \dots r_0}$ as the set of sequences x such that $x_i = r_i$ for $i = 0, \dots, \ell$. We observe that the image by T of a cylinder is another cylinder : for integers $r_\ell, \dots, r_0 \in \{0, \dots, b-1\}$, if there exists a minimal index $i \in \{0, \dots, \ell\}$ such that $r_i \neq b-1$ then $TC_{r_\ell \dots r_0} = C_{r_\ell \dots r_{i+1}(1+r_i)0^i}$, otherwise $TC_{r_\ell \dots r_0} = C_{0^{\ell+1}}$. Also, we define the *Rokhlin tower of order $\ell \geq 0$* as the family

$$\left(C_{0^{\ell+1}}, TC_{0^{\ell+1}}, \dots, T^{b^{\ell+1}-1}C_{0^{\ell+1}} \right)$$

where $(T^j C_{0^{\ell+1}})_{0 \leq j \leq b^{\ell+1}-1}$ form a partition of \mathbb{X} . We commonly represent this family as a tower as shown in Figure 2.

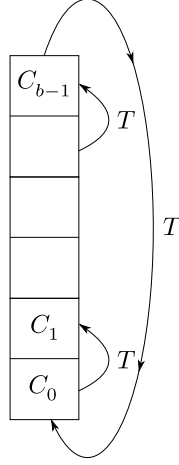


Figure 2: Behavior of T on the Rokhlin tower of order 0.

It is classical that the sequence of towers can be constructed with the so-called *Cut-and-Stack* inductive process, as illustrated in Figure 3.

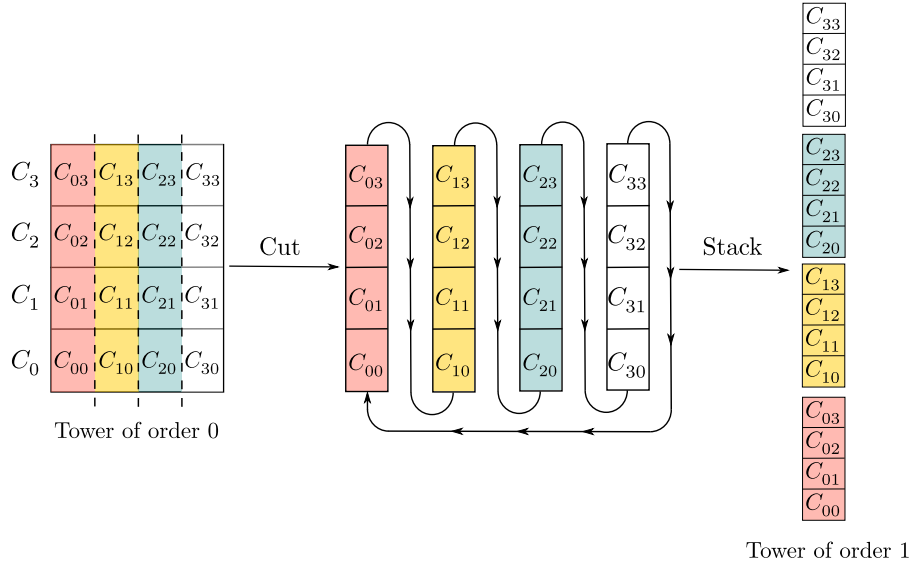


Figure 3: How to construct the tower of order 1 of T from the tower of order 0 in base $b = 4$.

By looking at the behavior of T on these towers, one can show that if \mathbb{P} is a T -invariant probability measure on (\mathbb{X}, T) , then \mathbb{P} gives the same measure to each level in a given tower: for every $\ell \geq 0$ and $r_0, \dots, r_\ell \in \{0, \dots, b-1\}$

$$\mathbb{P}(C_{r_\ell \dots r_0}) = \frac{1}{b^{\ell+1}}.$$

Since the cylinders generate the Borel σ -algebra, \mathbb{P} is uniquely determined by these values on the cylinders, hence (\mathbb{X}, T) is a uniquely ergodic dynamical system. We observe that choosing x in \mathbb{X} according to the unique T -invariant law \mathbb{P} means choosing its digits independently according to the uniform law on $\{0, \dots, b-1\}$. We also note that \mathbb{P} is the normalized Haar measure on \mathbb{X} .

For x in \mathbb{X} , we define the sequence of empirical probability measures along the (beginning of the) orbit of x : for every $N \geq 1$, we set

$$\epsilon_N(x) := \frac{1}{N} \sum_{0 \leq n < N} \delta_{T^n x}$$

(where δ_y denotes the Dirac measure on $y \in \mathbb{X}$).

Since the space of probability measures on \mathbb{X} is compact for the weak- $*$ topology, every subsequential limit of $(\epsilon_N(x))$ is a T -invariant probability measure. By the uniqueness of the T -invariant probability measure, for every $x \in \mathbb{X}$ we have $\epsilon_N(x) \rightarrow \mathbb{P}$. In other words, we have the convergence

$$\forall x \in \mathbb{X}, \quad \forall f \in \mathcal{C}(\mathbb{X}), \quad \frac{1}{N} \sum_{0 \leq n < N} f(T^n x) \xrightarrow{N \rightarrow +\infty} \int_{\mathbb{X}} f d\mathbb{P}. \quad (6)$$

We will be interested here in the special case $x = 0$ because $\mathbb{N} = \{T^n 0 : n \in \mathbb{N}\}$. Then (6) becomes

$$\forall f \in \mathcal{C}(\mathbb{X}), \quad \frac{1}{N} \sum_{0 \leq n < N} f(n) \xrightarrow{N \rightarrow +\infty} \int_{\mathbb{X}} f d\mathbb{P}. \quad (7)$$

Equation (7) shows that, for a continuous function f , averaging f over \mathbb{N} (for the natural density) amounts to averaging over \mathbb{X} (for \mathbb{P}). The next section shows how this convergence can be extended to some non-continuous functions related to the sum-of-digits function.

2.2 Sum of digits on the odometer

For every integer k , we define $s_k : \mathbb{X} \rightarrow \mathbb{Z}$ as the sum of the first $(k + 1)$ digits function, that is to say

$$s_k(x) := x_0 + \cdots + x_k.$$

Let $r \in \mathbb{N}$. We define the functions $\Delta_k^{(r)} : \mathbb{X} \rightarrow \mathbb{Z}$ by

$$\Delta_k^{(r)}(x) := s_k(x + r) - s_k(x).$$

The functions $\Delta_k^{(r)}$ are well-defined, continuous (and bounded) on \mathbb{X} . By (7), we have

$$\frac{1}{N} \sum_{n < N} \Delta_k^{(r)}(n) = \frac{1}{N} \sum_{n < N} \Delta_k^{(r)}(T^n 0) \xrightarrow{N \rightarrow +\infty} \int_{\mathbb{X}} \Delta_k^{(r)} d\mathbb{P}. \quad (8)$$

Although the sum-of-digits function s is not well defined on \mathbb{X} , we can extend the function $\Delta^{(r)}$ defined in (1) on the set of $x \in \mathbb{X}$ for which the number of different digits between x and $x + r$ is finite. This subset contains the b -adic integers x such that there exists an index $k \geq \max(\{\ell : r_\ell \neq 0\})$ such that $x_k \neq b - 1$. So, except for a finite number of b -adic integers, we can define

$$\Delta^{(r)}(x) := \lim_{k \rightarrow \infty} \Delta_k^{(r)}(x).$$

Remark 1. *Let $t \geq u$ be two integers. For every integer k we have the decomposition formula*

$$\Delta_k^{(t+u)} = \Delta_k^{(t)} + \Delta_k^{(u)} \circ T^t. \quad (9)$$

So, taking \mathbb{P} -almost everywhere the limit when k tends to infinity, we get

$$\Delta^{(t+u)} = \Delta^{(t)} + \Delta^{(u)} \circ T^t \quad (\mathbb{P}\text{-a.s.}). \quad (10)$$

Then, by induction on t , we deduce

$$\Delta^{(t)} = \Delta^{(1)} + \Delta^{(1)} \circ T + \cdots + \Delta^{(1)} \circ T^{t-1} \quad (\mathbb{P}\text{-a.s.}). \quad (11)$$

We observe that $\Delta^{(r)}$ also satisfies (2): for each $x \in \mathbb{X}$ for which $\Delta^{(r)}(x)$ is well defined, we have

$$\Delta^{(r)}(x) = s(r) - c(b-1), \quad (12)$$

where $c := \sum_{\ell \geq 0} c_\ell < \infty$ is the total number of carries generated during the computation of $x+r$.

Unfortunately, $\Delta^{(r)}$ is not continuous like the functions $\Delta_k^{(r)}$, it is not even bounded on \mathbb{X} , but we have the following result about functions with polynomial growth.

Proposition 2.1. *Let $r \geq 1$ and $f : \mathbb{Z} \rightarrow \mathbb{C}$. Assume that there exist $\alpha \geq 1$ and C in \mathbb{R}_+^* such that for every $n \in \mathbb{Z}$*

$$|f(n)| \leq C|n|^\alpha + |f(0)|. \quad (13)$$

Then $f \circ \Delta^{(r)} \in L^1(\mathbb{P})$ and we have the convergence

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n < N} f(\Delta^{(r)}(n)) &= \int_{\mathbb{X}} f(\Delta^{(r)}(x)) d\mathbb{P}(x) \\ &= \lim_{k \rightarrow \infty} \int_{\mathbb{X}} f(\Delta_k^{(r)}(x)) d\mathbb{P}(x). \end{aligned}$$

Corollary 2.2. *For every $d \in \mathbb{Z}$*

$$\begin{aligned} \mu^{(r)}(d) &:= \lim_{N \rightarrow \infty} \frac{1}{N} \left| \{n < N : \Delta^{(r)}(n) = d\} \right| \\ &= \mathbb{P} \left(\{x \in \mathbb{X} : \Delta^{(r)}(x) = d\} \right). \end{aligned} \quad (14)$$

Moreover, $\Delta^{(r)}$ has zero-mean and has finite moments.

In particular, we recover Bésineau's result on the existence of the asymptotic density and the fact that $\sum_{d \in \mathbb{Z}} \mu^{(r)}(d) = 1$.

Remark 2. *Using trivial arguments, Proposition 2.1 and Corollary 2.2 are also true when $r = 0$. We observe that $\mu^{(0)} = \delta_0$.*

Before proving this proposition and its corollary, we need the following lemma.

Lemma 2.3. *Let $r \geq 1$. For $N \in \mathbb{N}^*$, for $k \in \mathbb{N}$ and $d, d' \in \mathbb{Z}$, we have the inequality*

$$\begin{aligned} \frac{1}{N} \left| \{n < N : (\Delta^{(r)}(n), \Delta_k^{(r)}(n)) = (d, d')\} \right| \\ \leq rb \mathbb{P} \left(\{x \in \mathbb{X} : (\Delta^{(r)}(x), \Delta_k^{(r)}(x)) = (d, d')\} \right). \end{aligned} \quad (15)$$

In particular, we have

$$\frac{1}{N} \left| \{n < N : \Delta^{(r)}(n) = d\} \right| \leq rb \mathbb{P} \left(\{x \in \mathbb{X} : \Delta^{(r)}(x) = d\} \right). \quad (16)$$

Proof. Of course, (15) implies (16) so we just need to prove (15). We fix $k \in \mathbb{N}$. For every $\ell \in \mathbb{N}$, let V_ℓ be the set of the values reached by the couple $(\Delta^{(r)}, \Delta_k^{(r)})$ on the first $b^{\ell+1} - r$ levels of the Rokhlin tower of order ℓ (see Figure 4). Of course, if $b^{\ell+1} - r \leq 0$ then $V_\ell := \emptyset$. Otherwise, we observe that V_ℓ is a finite set. Indeed, the first $b^{\ell+1} - r$ levels correspond to the b -adic integers x such that, when we add r , the carry propagation does not go beyond the first $\ell + 1$ digits. Since these digits are fixed on a level of the Rokhlin tower of order ℓ , except for the last r levels, $\Delta^{(r)}$ and $\Delta_k^{(r)}$ are constant on each such level. We observe that the sequence $(V_\ell)_{\ell \geq 0}$ is increasing for the inclusion.

Now, for $d, d' \in \mathbb{Z}$, there are 2 cases.

1. If $(d, d') \notin \cup_{\ell \geq 0} V_\ell$, then for each $n \in \mathbb{N}$ we have $(\Delta^{(r)}(n), \Delta_k^{(r)}(n)) \neq (d, d')$. Indeed, for each $n \in \mathbb{N}$, there exists a smallest integer ℓ such that n is in the first $b^{\ell+1} - r$ levels of the tower of order ℓ , hence $(\Delta^{(r)}(n), \Delta_k^{(r)}(n)) \in V_\ell$. In this case, (15) is trivial.
2. If $(d, d') \in \cup_{\ell \geq 0} V_\ell$ then there exists a unique $\ell \geq 0$ such that $(d, d') \in V_\ell \setminus V_{\ell-1}$ (with the convention $V_{-1} := \emptyset$). Since $(\Delta^{(r)}, \Delta_k^{(r)})$ is constant on each of the first $b^{\ell+1} - r$ levels of the tower, it takes the value (d, d') on at least one whole such level of measure $\frac{1}{b^{\ell+1}}$. So, we have

$$\mathbb{P}\left(\{x \in \mathbb{X} : (\Delta^{(r)}(x), \Delta_k^{(r)}(x)) = (d, d')\}\right) \geq \frac{1}{b^{\ell+1}}.$$

Also, since the couple (d, d') does not appear in the first levels of the previous tower $((d, d') \notin V_{\ell-1})$, we claim that, for every $N \geq 1$

$$\frac{1}{N} \left| \{n < N : (\Delta^{(r)}(n), \Delta_k^{(r)}(n)) = (d, d')\} \right| \leq \frac{r}{b^\ell}.$$

Indeed,

- (a) If $r \geq b^\ell$, then the inequality is then trivial.
- (b) If $r < b^\ell$ then, since (d, d') is not in $V_{\ell-1}$, (d, d') can only appear inside the r highest levels of the tower of order $\ell - 1$. So, if we note C the union of these r highest levels, using the fact that 0 is in the first level of the tower, we have

$$\begin{aligned} \frac{1}{N} \left| \{0 \leq n < N : (\Delta^{(r)}(n), \Delta_k^{(r)}(n)) = (d, d')\} \right| \\ \leq \frac{1}{N} |\{0 \leq n < N : T^n 0 \in C\}| \\ \leq \frac{r}{b^\ell}. \end{aligned}$$

Combining both inequalities gives (15).

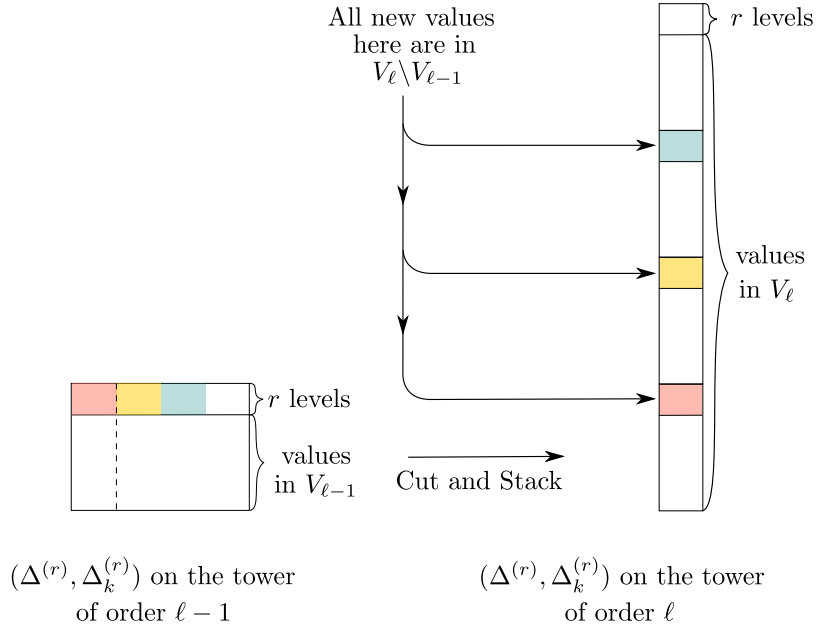


Figure 4: Visual description of $V_{\ell-1}$, V_ℓ and $V_\ell \setminus V_{\ell-1}$.

□

Proof of Proposition 2.1. Let $\varepsilon > 0$. For any integer k , we have

$$\left| \frac{1}{N} \sum_{n < N} f(\Delta^{(r)}(n)) - \int_{\mathbb{X}} f(\Delta^{(r)}(x)) d\mathbb{P}(x) \right| \leq A_1 + A_2 + A_3,$$

where

$$\begin{aligned}
 A_1 &:= \frac{1}{N} \sum_{n < N} \left| f(\Delta^{(r)}(n)) - f(\Delta_k^{(r)}(n)) \right|, \\
 A_2 &:= \left| \frac{1}{N} \sum_{n < N} f(\Delta_k^{(r)}(n)) - \int_{\mathbb{X}} f(\Delta_k^{(r)}(x)) d\mathbb{P}(x) \right|, \\
 A_3 &:= \int_{\mathbb{X}} \left| f(\Delta^{(r)}(x)) - f(\Delta_k^{(r)}(x)) \right| d\mathbb{P}(x).
 \end{aligned}$$

For A_1 , using Lemma 2.3, we get

$$\begin{aligned}
A_1 &= \frac{1}{N} \sum_{n < N} \sum_{j, j' \in \mathbb{Z}} \left| f(j) - f(j') \right| \mathbb{1}_{(j, j')} \left(\Delta^{(r)}(n), \Delta_k^{(r)}(n) \right) \\
&= \sum_{j, j' \in \mathbb{Z}} \left| f(j) - f(j') \right| \frac{1}{N} \sum_{n < N} \mathbb{1}_{(j, j')} \left(\Delta^{(r)}(n), \Delta_k^{(r)}(n) \right) \\
&\leq rb \sum_{j, j' \in \mathbb{Z}} \left| f(j) - f(j') \right| \mathbb{P} \left(\Delta^{(r)}(n) = j, \Delta_k^{(r)}(n) = j' \right) \\
&= rb \int_{\mathbb{X}} \left| f(\Delta^{(r)}(x)) - f(\Delta_k^{(r)}(x)) \right| d\mathbb{P}(x) = rbA_3.
\end{aligned}$$

Hence, controlling A_3 also enables us to control A_1 . For this, we note that the integrand in A_3 converges \mathbb{P} -almost everywhere to 0, therefore it is enough to show that the dominated convergence theorem applies. To find a good dominant function, we write using (13)

$$\left| f \circ \Delta_k^{(r)}(x) \right| \leq C \left| \Delta_k^{(r)}(x) \right|^\alpha + \left| f(0) \right|.$$

We get from (9)

$$\Delta_k^{(r)} = \Delta_k^{(1)} + \Delta_k^{(1)} \circ T + \dots + \Delta_k^{(1)} \circ T^{r-1},$$

then we use the multinomial theorem (we can suppose $\alpha \in \mathbb{N}$) to write

$$\left| \Delta_k^{(1)}(x) + \dots + \Delta_k^{(1)} \circ T^{r-1}(x) \right|^\alpha = \sum_{j_0 + \dots + j_{r-1} = \alpha} \binom{\alpha}{j_0, \dots, j_{r-1}} \prod_{i=0}^{r-1} \left| \Delta_k^{(1)} \circ T^i(x) \right|^{j_i}.$$

We now use Young's inequality to get

$$\prod_{i=0}^{r-1} \left| \Delta_k^{(1)} \circ T^i(x) \right|^{j_i} \leq \frac{1}{r} \sum_{i=0}^{r-1} \left| \Delta_k^{(1)} \circ T^i(x) \right|^{r j_i}.$$

Then, we define, for $i = 0, \dots, r-1$, $g_i(x) := \sup_{k \in \mathbb{N}} \left| \Delta_k^{(1)} \circ T^i(x) \right|$ in order to get the inequality

$$\left| f \circ \Delta_k^{(r)}(x) \right| \leq C \sum_{j_0 + \dots + j_{r-1} = \alpha} \sum_{i=0}^{r-1} \frac{\binom{\alpha}{j_0, \dots, j_{r-1}}}{r} g_i(x)^{r j_i} + \left| f(0) \right|.$$

Moreover, when it is well defined, we have $\Delta^{(r)} \circ T^i = \lim_{k \rightarrow \infty} \Delta_k^{(r)} \circ T^i$ therefore, we also get $\left| \Delta^{(r)} \circ T^i \right| \leq g_i$ which yields the similar inequality

$$\left| f \circ \Delta^{(r)}(x) \right| \leq C \sum_{j_0 + \dots + j_{r-1} = \alpha} \sum_{i=0}^{r-1} \frac{\binom{\alpha}{j_0, \dots, j_{r-1}}}{r} g_i(x)^{r j_i} + \left| f(0) \right|.$$

We just need to show that $g_i^{rj_i}$ is integrable for the measure \mathbb{P} . It is equivalent to show that $\sum_m \mathbb{P}(\{x \in \mathbb{X} : g_i(x)^{rj_i} > m\})$ is a convergent series. We have

$$\begin{aligned} g_i(x) > m^{\frac{1}{rj_i}} &\Leftrightarrow \sup_{k \in \mathbb{N}} |\Delta_k^{(1)} \circ T^i(x)| > m^{\frac{1}{rj_i}} \\ &\Leftrightarrow \exists k \in \mathbb{N}, |\Delta_k^{(1)}(T^i x)| > m^{\frac{1}{rj_i}}. \end{aligned}$$

From (12), this condition is true if and only if the addition $T^i x + 1$ creates sufficiently many carries: strictly more than $\lfloor \frac{m^{\frac{1}{rj_i} + 1}}{b-1} \rfloor$. But, we know that when we add 1 to $T^i x$, the number of carries created by the addition is the number of $(b-1)$'s at the right-hand side of the expansion of $T^i x$. So, because of the T -invariance of \mathbb{P}

$$\mathbb{P}(\{x \in \mathbb{X} : g_i(x) > m^{\frac{1}{rj_i}}\}) \leq \left(\frac{1}{b}\right)^{\lfloor \frac{m^{\frac{1}{rj_i} + 1}}{b-1} \rfloor}.$$

The right quantity is the general term of a convergent series which shows that $g_i^{rj_i} \in L^1(\mathbb{P})$. So the dominated convergence theorem can be applied and, for k large enough, $A_1 + A_3 \leq \frac{\varepsilon}{2}$ for every $N \geq 1$.

Now, once we have fixed such a k , for N large enough, A_2 is bounded by $\frac{\varepsilon}{2}$ because of (7) and the continuity of $\Delta_k^{(r)}$ and of f . The convergence in the statement is thus proved.

Note that the argument of the dominated convergence theorem also proves that $f \circ \Delta^{(r)} \in L^1(\mathbb{P})$ and $\int_{\mathbb{X}} f \circ \Delta^{(r)} d\mathbb{P} = \lim_{k \rightarrow \infty} \int_{\mathbb{X}} f \circ \Delta_k^{(r)} d\mathbb{P}$. \square

Proof of Corollary 2.2. We just apply Proposition 2.1 with particular functions f . First, for $d \in \mathbb{Z}$, we use the function $f = \mathbb{1}_{\{d\}}$. It gives

$$\mu^{(r)}(d) = \mathbb{P}(\{x \in \mathbb{X} : \Delta^{(r)}(x) = d\}).$$

Then, we take f as the identity function on \mathbb{Z} for which (13) is clearly satisfied. Also, for every integer k , we have by T -invariance of \mathbb{P}

$$\int_{\mathbb{X}} \Delta_k^{(r)} d\mathbb{P} = \int_{\mathbb{X}} s_k \circ T^r d\mathbb{P} - \int_{\mathbb{X}} s_k d\mathbb{P} = 0.$$

We then deduce that

$$\sum_{d \in \mathbb{Z}} d \mu^{(r)}(d) = \int_{\mathbb{X}} \Delta^{(r)} d\mathbb{P} = \lim_{k \rightarrow \infty} \int_{\mathbb{X}} \Delta_k^{(r)} d\mathbb{P} = 0. \quad (17)$$

Finally, we use, for every $j \geq 2$, the function $f(n) = n^j$ which satisfies (13). This gives the existence of moments of order j for $\Delta^{(r)}$. \square

More generally, we have the following convergence.

Proposition 2.4. *Let $r \geq 1$ and $f : \mathbb{Z} \rightarrow \mathbb{C}$ be such that $f \circ \Delta^{(r)} \in L^1(\mathbb{P})$. Then*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n < N} f(\Delta^{(r)}(n)) = \int_{\mathbb{X}} f(\Delta^{(r)}(x)) d\mathbb{P}(x).$$

Proof of Proposition 2.4. By (12), the values reached by $\Delta^{(r)}$ are of the form $a_k^{(r)} := s(r) - k(b-1)$, $k \geq 0$, and we have

$$\int_{\mathbb{X}} f(\Delta^{(r)}(x)) d\mathbb{P}(x) = \sum_{k \geq 0} f(a_k^{(r)}) \mathbb{P}(\{x \in \mathbb{X} : \Delta^{(r)}(x) = a_k^{(r)}\}). \quad (18)$$

On the other hand, we write

$$\begin{aligned} \frac{1}{N} \sum_{n < N} f(\Delta^{(r)}(n)) &= \frac{1}{N} \sum_{n < N} f(\Delta^{(r)}(n)) \sum_{k \geq 0} \mathbb{1}_{\{a_k^{(r)}\}}(\Delta^{(r)}(n)) \\ &= \sum_{k \geq 0} f(a_k^{(r)}) \underbrace{\frac{1}{N} \sum_{n < N} \mathbb{1}_{\{a_k^{(r)}\}}(\Delta^{(r)}(n))}_{:= u_{N,k}}. \end{aligned}$$

We conclude by applying the dominated convergence theorem in the metric space $\ell^1(\mathbb{N})$ endowed with the counting measure to show

$$\sum_{k \geq 0} u_{N,k} \xrightarrow{N \rightarrow +\infty} \sum_{k \geq 0} f(a_k^{(r)}) \mathbb{P}(\{x \in \mathbb{X} : \Delta^{(r)}(x) = a_k^{(r)}\}).$$

1. The pointwise limit on \mathbb{N} of $u_{N,k}$ is $f(a_k^{(r)}) \mathbb{P}(\{x \in \mathbb{X} : \Delta^{(r)}(x) = a_k^{(r)}\})$ by (14).
2. A dominant function is given using Lemma 2.3: for all k and N we have

$$u_{N,k} \leq g(k) := \left| f(a_k^{(r)}) \right| r b \mathbb{P}(\{x \in \mathbb{X} : \Delta^{(r)}(x) = a_k^{(r)}\}).$$

By (18) and the hypothesis that $f \circ \Delta^{(r)}$ is integrable, $\sum g(k)$ is convergent.

□

We have shown that, for any integer r , $\Delta^{(r)}$ has finite moments. The next section will focus on the second moment, which is the one we are most interested in for our CLT.

3 Variance of $\mu^{(r)}$

3.1 Inductive relation on the measures

Given $r \in \mathbb{N}$, there exist $\tilde{r} \in \mathbb{N}$ and $0 \leq r_0 \leq b-1$ such that $r = b\tilde{r} + r_0$. The integer r_0 is actually the units digit of the expansion of r and, if $r \geq b$,

\tilde{r} corresponds to the integer whose expansion is obtained by erasing r_0 in the expansion of r . The expansion of \tilde{r} is then one digit shorter than the expansion of r . If $r < b$ then, of course, $r_0 = r$ and $\tilde{r} = 0$. First of all, we have a well known inductive relation on the length of the expansion of r .

Proposition 3.1. *For $\tilde{r} \in \mathbb{N}$, $0 \leq r_0 \leq b - 1$ and $d \in \mathbb{Z}$*

$$\mu^{(b\tilde{r}+r_0)}(d) = \frac{b-r_0}{b}\mu^{(\tilde{r})}(d-r_0) + \frac{r_0}{b}\mu^{(\tilde{r}+1)}(d+b-r_0). \quad (19)$$

Proof. Let $x = (\dots, x_1, x_0) \in \mathbb{X}$. We define $\tilde{x} := (\dots, x_1)$. Let us consider the computation of the digits of $x + r$, where $r := b\tilde{r} + r_0 = \overline{r_\ell \dots r_0}$

$$\begin{array}{rcccc} & \dots & x_\ell & \dots & x_1 & x_0 \\ + & & r_\ell & \dots & r_1 & r_0 \\ \hline = & & & \dots & & (x+r)_0 \end{array}$$

If $x_0 + r_0 < b$, no carry is created and $\Delta^{(b\tilde{r}+r_0)}(x) = r_0 + \Delta^{(\tilde{r})}(\tilde{x})$. Otherwise, $x_0 + r_0 \geq b$ and we have to subtract b from the units digit of the result: we are left with the addition of \tilde{x} and $\tilde{r} + 1$: so $\Delta^{(b\tilde{r}+r_0)}(x) = r_0 - b + \Delta^{(\tilde{r}+1)}(\tilde{x})$. To sum up, we have

$$\Delta^{(b\tilde{r}+r_0)}(x) = \begin{cases} r_0 + \Delta^{(\tilde{r})}(\tilde{x}) & \text{if } x_0 + r_0 < b, \\ r_0 - b + \Delta^{(\tilde{r}+1)}(\tilde{x}) & \text{otherwise.} \end{cases}$$

Now, let $d \in \mathbb{Z}$. We partition the set $\{x \in \mathbb{X} : \Delta^{(b\tilde{r}+r_0)}(x) = d\}$ according to the value of x_0

$$\begin{aligned} \{x \in \mathbb{X} : \Delta^{(b\tilde{r}+r_0)}(x) = d\} &= \bigcup_{j=0}^{b-r_0-1} \{x \in \mathbb{X} : x_0 = j \text{ and } \Delta^{(\tilde{r})}(\tilde{x}) = d - r_0\} \\ &\quad \bigcup_{j=b-r_0}^{b-1} \{x \in \mathbb{X} : x_0 = j \text{ and } \Delta^{(\tilde{r}+1)}(\tilde{x}) = d + b - r_0\}. \end{aligned}$$

We observe that if x is randomly chosen with law \mathbb{P} , then \tilde{x} is independent of x_0 and also follows \mathbb{P} . We just need to take the measure to conclude. \square

If we apply finitely many times (19), we can express $\mu^{(r)}(d)$ as a convex combination of the measures $\mu^{(0)}$ and $\mu^{(1)}$ evaluated on particular points. We recall that $\mu^{(0)} = \delta_0$ (Dirac measure on 0). We can also compute $\mu^{(1)}$.

Lemma 3.2. *For every $d \in \mathbb{Z}$*

$$\mu^{(1)}(d) := \begin{cases} \frac{1}{b^k} - \frac{1}{b^{k+1}} & \text{if } d = 1 - k(b-1) \text{ for some } k \in \mathbb{N} \\ 0 & \text{otherwise.} \end{cases}$$

Proof. We use again the notation $a_k^{(1)} = 1 - k(b-1)$.

By (12), it is trivial that if $d \neq a_k^{(1)}$ for all $k \in \mathbb{N}$ then $\mu^{(1)}(d) = 0$. Otherwise,

we recall again by (12) that for all $k \in \mathbb{N}$, $\Delta^{(1)}(x) = a_k^{(1)}$ if and only if the right-hand side of the expansion of x is exactly a block of $(b-1)$'s of length k . Thus, if $k \geq 1$

$$\begin{aligned} \mu^{(1)}\left(a_k^{(1)}\right) &= \mathbb{P}\left(\{x \in \mathbb{X} : x_0 = \dots = x_{k-1} = b-1 \text{ and } x_k < b-1\}\right) \\ &= \frac{b-1}{b^{k+1}}. \end{aligned}$$

And, if $k = 0$

$$\mu^{(1)}\left(a_k^{(1)}\right) = \mathbb{P}\left(\{x \in \mathbb{X} : x_0 < b-1\}\right) = \frac{b-1}{b}.$$

□

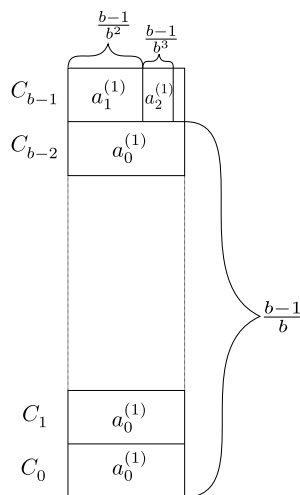


Figure 5: Values of $\Delta^{(1)}$ on the levels of the Rokhlin tower of order 0, and the corresponding \mathbb{P} -measures.

3.2 Inductive relation on the variance, first results

Emme and Hubert [3, Theorem 3.1] give an explicit formula for the variance in base 2. It is possible to adapt their methods in order to find a similar expression in any base. However, here we just need some basic estimations about the variance. We first deduce from Proposition 19 an inductive relation on the variance.

Lemma 3.3. For $\tilde{r} \in \mathbb{N}$ and $0 \leq r_0 \leq b - 1$, we have the relation

$$\text{Var}(\mu^{(b\tilde{r}+r_0)}) = \frac{b-r_0}{b} \text{Var}(\mu^{(\tilde{r})}) + \frac{r_0}{b} \text{Var}(\mu^{(\tilde{r}+1)}) + r_0(b-r_0).$$

Proof. We compute using (17) and (19)

$$\begin{aligned} \text{Var}(\mu^{(b\tilde{r}+r_0)}) &= \sum_{d \in \mathbb{Z}} d^2 \mu^{(b\tilde{r}+r_0)}(d) \\ &= \sum_{d \in \mathbb{Z}} d^2 \left(\frac{b-r_0}{b} \mu^{(\tilde{r})}(d-r_0) + \frac{r_0}{b} \mu^{(\tilde{r}+1)}(d+b-r_0) \right) \\ &= \frac{b-r_0}{b} \sum_{d' \in \mathbb{Z}} (d'+r_0)^2 \mu^{(\tilde{r})}(d') + \frac{r_0}{b} \sum_{d' \in \mathbb{Z}} (d'+b-r_0)^2 \mu^{(\tilde{r}+1)}(d') \end{aligned}$$

Using again (17), we get

$$\begin{aligned} \text{Var}(\mu^{(b\tilde{r}+r_0)}) &= \frac{b-r_0}{b} \left(\text{Var}(\mu^{(\tilde{r})}) + r_0^2 \right) + \frac{r_0}{b} \left(\text{Var}(\mu^{(\tilde{r}+1)}) + (b-r_0)^2 \right) \\ &= \frac{b-r_0}{b} \text{Var}(\mu^{(\tilde{r})}) + \frac{r_0}{b} \text{Var}(\mu^{(\tilde{r}+1)}) + r_0(b-r_0). \end{aligned}$$

□

Even if we do not need an explicit formula of $\text{Var}(\mu^{(r)})$ for a general $r \in \mathbb{N}$, we will need one in some specific cases. First, we are interested in the variance of $\mu^{(r)}$ when the expansion of r is one digit long.

Lemma 3.4. If $0 \leq r \leq b - 1$ then

$$\text{Var}(\mu^{(r)}) = r(1+b-r).$$

Proof. We recall that $\mu^{(0)} = \delta_0$ so $\text{Var}(\mu^{(0)}) = 0$. Then, if $r = 1$, Lemma 3.3 gives

$$\text{Var}(\mu^{(1)}) = \frac{b-1}{b} \text{Var}(\mu^{(0)}) + \frac{1}{b} \text{Var}(\mu^{(1)}) + (b-1).$$

It follows

$$\text{Var}(\mu^{(1)}) = b.$$

Finally, if $2 \leq r \leq b - 1$, again from Lemma 3.3, we have

$$\text{Var}(\mu^{(r)}) = \frac{b-r}{b} \text{Var}(\mu^{(0)}) + \frac{r}{b} \text{Var}(\mu^{(1)}) + r(b-r) = r(1+b-r).$$

□

Now, we are interested in the variance of $\mu^{(r)}$ when the expansion of r has a rightmost block of $(b-1)$'s of length $m \geq 1$ that is to say when there exists $\hat{r} \in \mathbb{N}$ such that $r = b^m \hat{r} + b^m - 1$.

Lemma 3.5. For $\hat{r} \in \mathbb{N}$ and $m \geq 1$, the variance of $\mu^{(b^m \hat{r} + b^m - 1)}$ is

$$\text{Var}(\mu^{(b^m \hat{r} + b^m - 1)}) = \frac{1}{b^m} \text{Var}(\mu^{(\hat{r})}) + \left(1 - \frac{1}{b^m}\right) \text{Var}(\mu^{(\hat{r}+1)}) + b - \frac{1}{b^{m-1}}.$$

Remark 3. We observe that, if we take $\hat{r} = 0$, then we find the variance of $\mu^{(r)}$ where the expansion of r is composed of only one block of $(b-1)$'s. Thus, with Lemma 3.4, we now have the exact value of the variance of $\mu^{(r)}$ when the expansion of r is composed of exactly 1 non-zero block:

$$\text{Var}(\mu^{(r)}) = \begin{cases} r(1+b-r) & \text{if } r = 1, \dots, b-2, \\ 2b - \frac{2}{b^{m-1}} & \text{if } r = b^m - 1 \text{ (} m \geq 1 \text{)}. \end{cases}$$

Proof. We prove the lemma by induction on $m \geq 1$.

1. If $m = 1$ then, from Lemma 3.3

$$\text{Var}(\mu^{(b\hat{r}+b-1)}) = \frac{1}{b} \text{Var}(\mu^{(\hat{r})}) + \frac{b-1}{b} \text{Var}(\mu^{(\hat{r}+1)}) + (b-1).$$

That is what we want.

2. If we assume that the lemma is true for $m-1$ then we consider the integer $b^m \hat{r} + b^m - 1$. We observe the trivial identity

$$b^m \hat{r} + b^m - 1 = b \times (b^{m-1} \hat{r} + (b^{m-1} - 1)) + (b-1)$$

It follows from Lemma 3.3

$$\begin{aligned} \text{Var}(\mu^{(b^m \hat{r} + b^m - 1)}) &= \frac{1}{b} \text{Var}(\mu^{(b^{m-1} \hat{r} + (b^{m-1} - 1))}) + \frac{b-1}{b} \text{Var}(\mu^{(b^{m-1}(\hat{r}+1))}) \\ &\quad + (b-1). \end{aligned}$$

We use the induction hypothesis and the fact that $\mu^{(b^{m-1}(\hat{r}+1))} = \mu^{(\hat{r}+1)}$ (Lemma 19)

$$\begin{aligned} \text{Var}(\mu^{(b^{m-1} \hat{r} + (b^{m-1} - 1))}) &= \frac{1}{b^{m-1}} \text{Var}(\mu^{(\hat{r})}) + \left(1 - \frac{1}{b^{m-1}}\right) \text{Var}(\mu^{(\hat{r}+1)}) \\ &\quad + b - \frac{1}{b^{m-2}}. \end{aligned}$$

Combining both gives the result for m . □

Finally, we consider the case where the expansion of r has a units digit 1, possibly with a block of 0's on its left. This corresponds to the existence of $\hat{r} \in \mathbb{N}$ and $m \geq 1$ such that $r = b^m \hat{r} + 1$.

Lemma 3.6. For $\hat{r} \in \mathbb{N}$ and $m \geq 1$, the variance of $\mu^{(b^m \hat{r} + 1)}$ is

$$\text{Var}(\mu^{(b^m \hat{r} + 1)}) = \left(1 - \frac{1}{b^m}\right) \text{Var}(\mu^{(\hat{r})}) + \frac{1}{b^m} \text{Var}(\mu^{(\hat{r}+1)}) + b - \frac{1}{b^{m-1}}.$$

Proof. We show by induction on $m \geq 1$.

1. If $m = 1$ then, from Lemma 3.3

$$\text{Var}(\mu^{(b\hat{r}+1)}) = \frac{b-1}{b}\text{Var}(\mu^{(\hat{r})}) + \frac{1}{b}\text{Var}(\mu^{(\hat{r}+1)}) + (b-1).$$

That is exactly what we want.

2. If we assume that the formula is true for $m-1$, then we consider the integer $b^m\hat{r}+1$. We have again by Lemma 3.3

$$\begin{aligned} \text{Var}(\mu^{(b^m\hat{r}+1)}) &= \text{Var}(\mu^{(b \times b^{m-1}\hat{r}+1)}) \\ &= \frac{b-1}{b}\text{Var}(\mu^{(b^{m-1}\hat{r})}) + \frac{1}{b}\text{Var}(\mu^{(b^{m-1}\hat{r}+1)}) + (b-1). \end{aligned}$$

We use the induction hypothesis.

$$\text{Var}(\mu^{(b^m\hat{r}+1)}) = \left(1 - \frac{1}{b^{m-1}}\right)\text{Var}(\mu^{(\hat{r})}) + \frac{1}{b^{m-1}}\text{Var}(\mu^{(\hat{r}+1)}) + b - \frac{1}{b^{m-2}}.$$

Combining both gives the result for m . □

3.3 Upper and lower bound of the variance

Since $\text{Var}(\mu^{(0)}) = 0$, the case $r = 0$ is irrelevant and we suppose $r \geq 1$. We wish to find an upper and a lower bound of the variance of $\mu^{(r)}$ depending on $\rho(r)$, the number of blocks of r defined in Definition 1.1. For convenient reasons, it is better to think in terms of non-zero blocks. We define, for an integer r , the quantity $\lambda(r)$ which corresponds to the number of non-zero blocks.

Example 3.7. We use again the example of Figure 1.

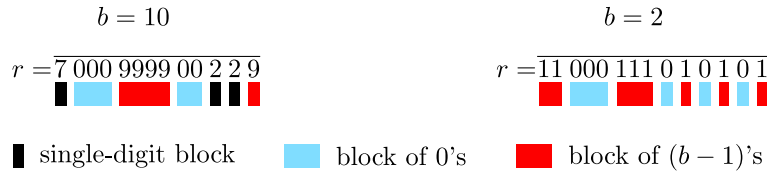


Figure 6: On the left-hand side, $\rho(r) = 7$ and $\lambda(r) = 5$. On the right-hand side, $\rho(r) = 9$ and $\lambda(r) = 5$.

Of course, there is a relation between $\lambda(r)$ and $\rho(r)$ (we recall that $r \geq 1$):

$$\lambda(r) \leq \rho(r) \leq 2\lambda(r). \tag{20}$$

We first give an upper bound of $\text{Var}(\mu^{(r)})$ depending on $\lambda(r)$.

Proposition 3.8. For any $r \geq 1$

$$\text{Var}(\mu^{(r)}) \leq b^2 \lambda(r). \quad (21)$$

We need the following lemma.

Lemma 3.9. For $r \geq 1$, we have the following inequality

$$|\text{Var}(\mu^{(r+1)}) - \text{Var}(\mu^{(r)})| \leq b.$$

Proof. We recall that $r_0 \in \{0, \dots, b-1\}$ is the units digit of r and the relation $r = b\tilde{r} + r_0$. Let us prove

$$|\text{Var}(\mu^{(r+1)}) - \text{Var}(\mu^{(r)})| \leq \frac{|\text{Var}(\mu^{(\tilde{r}+1)}) - \text{Var}(\mu^{(\tilde{r})})|}{b} + b - 1.$$

Indeed, there are two cases.

1. If $r_0 = b - 1$, then $r = b\tilde{r} + (b - 1)$ and using twice Lemma 3.3, we get

$$\begin{aligned} \text{Var}(\mu^{(r+1)}) - \text{Var}(\mu^{(r)}) &= \text{Var}(\mu^{(b(\tilde{r}+1))}) - \text{Var}(\mu^{(b\tilde{r}+b-1)}) \\ &= \frac{\text{Var}(\mu^{(\tilde{r}+1)}) - \text{Var}(\mu^{(\tilde{r})})}{b} - (b - 1). \end{aligned}$$

2. If $r_0 \in \{0, \dots, b - 2\}$ then, using the same tools, we compute

$$\begin{aligned} \text{Var}(\mu^{(r+1)}) - \text{Var}(\mu^{(r)}) &= \text{Var}(\mu^{(b\tilde{r}+r_0+1)}) - \text{Var}(\mu^{(b\tilde{r}+r_0)}) \\ &= \frac{\text{Var}(\mu^{(\tilde{r}+1)}) - \text{Var}(\mu^{(\tilde{r})})}{b} + b - 2r_0 - 1. \end{aligned}$$

We observe that $|b - 2r_0 - 1| \leq b - 1$.

Then, we conclude by an easy induction on the number of digits of r and by checking that

$$\begin{aligned} |\text{Var}(\mu^{(r+1)}) - \text{Var}(\mu^{(r)})| &\leq \frac{|\text{Var}(\mu^{(\tilde{r}+1)}) - \text{Var}(\mu^{(\tilde{r})})|}{b} + b - 1 \\ &\leq \frac{b}{b} + b - 1 = b. \end{aligned}$$

□

Proof of Proposition 3.8. Observe that we have

$$b^2 \geq j(1 + b - j) \quad \text{for all } j = 1, \dots, b - 1, \quad (C_0)$$

$$b^2 \geq 2b, \quad (C_1)$$

$$b^2 \geq jb \quad \text{for all } j = 0, \dots, b - 1. \quad (C_2)$$

We proceed by induction on $\lambda(r) \geq 1$.

1. Initialisation: if $\lambda(r) = 1$ then we have two cases depending on the type of blocks we are considering. However, from Lemma 3.4, Lemma 3.5 and Conditions (C_0) and (C_1) , we can deduce that in both cases we have $\text{Var}(\mu^{(r)}) \leq b^2$.
2. Inductive step: we let $n > 1$ and we assume that if $\lambda(r) \leq n$ then $\text{Var}(\mu^{(r)}) \leq b^2\lambda(r)$. We now assume that our $r \in \mathbb{N}$ satisfies $\lambda(r) = n + 1$. Let $\ell \geq 0$, $\widehat{r} \in \mathbb{N}$ and B_1 a non-zero block such that we can write the expansion of r as follow

$$\overline{\widehat{r} B_1 0^\ell}.$$

We can assume $\ell = 0$ but \widehat{r} may have a rightmost block composed of 0's. We discuss on the type of B_1 .

- (a) If B_1 is a block of $(b - 1)$'s of length m then $r = b^m\widehat{r} + b^m - 1$ and we have the trivial equality

$$\begin{aligned} \text{Var}(\mu^{(r)}) &= \text{Var}(\mu^{(\widehat{r})}) + \left(\text{Var}(\mu^{(r)}) - \text{Var}(\mu^{(r+1)}) \right) \\ &\quad - \left(\text{Var}(\mu^{(\widehat{r})}) - \text{Var}(\mu^{(\widehat{r}+1)}) \right). \end{aligned}$$

Using Lemma 3.9, the inductive hypothesis and Condition (C_1)

$$\text{Var}(\mu^{(r)}) \leq b^2n + b + b \leq b^2(n + 1).$$

- (b) If B_1 is a single-digit block r_0 then $r = b\widehat{r} + r_0$ and with we have another trivial equality

$$\text{Var}(\mu^{(r)}) = \text{Var}(\mu^{(\widehat{r})}) + \sum_{k=1}^{r_0} \text{Var}(\mu^{(b\widehat{r}+k)}) - \text{Var}(\mu^{(b\widehat{r}+k-1)}).$$

Then, using Lemma 3.9, the inductive hypothesis and Condition (C_2) , we find

$$\text{Var}(\mu^{(r)}) \leq b^2n + r_0b \leq b^2(n + 1).$$

This concludes the inductive step and the proof. □

We also have a lower bound depending on $\lambda(r)$.

Proposition 3.10. *For any $r \geq 1$*

$$\text{Var}(\mu^{(r)}) \geq \frac{b}{4}\lambda(r). \tag{23}$$

Proof. Observe that

$$\frac{b}{4} \leq \min\{j(b-j) : j = 1, \dots, b-1\}, \quad (C_3)$$

$$\frac{b}{4} \leq (b-1), \quad (C_4)$$

$$\frac{b}{4} \leq \min\{j(b-j-\frac{1}{2}) : j = 1, \dots, b-1\}, \quad (C_5)$$

Of course, some of these conditions are redundant but we keep them all for simplicity because each one will be used in the proof. We prove the result using induction on $\lambda(r) \geq 1$.

1. Initialisation: if $\lambda(r) = 1$ then we have two cases depending on the type of blocks we are considering. However, from Lemma 3.4, Lemma 3.5 and Conditions (C₃) and (C₄), we can deduce $\text{Var}(\mu^{(r)}) \geq \frac{b}{4}$.
2. Inductive step: we let $n > 1$ and we assume that if $\lambda(r) \leq n$ then $\text{Var}(\mu^{(r)}) \geq \frac{b}{4}\lambda(r)$. We now take $r \in \mathbb{N}$ such that $\lambda(r) = n+1$, and we write

$$r = b\tilde{r} + r_0.$$

Without loss of generality, we can assume that $r_0 \neq 0$. Indeed, if $r_0 = 0$ then $\text{Var}(\mu^{(r)}) = \text{Var}(\mu^{(\tilde{r})})$ and $\lambda(r) = \lambda(\tilde{r})$.

From Lemma 3.3, we get

$$\text{Var}(\mu^{(r)}) = \frac{b-r_0}{b}\text{Var}(\mu^{(\tilde{r})}) + \frac{r_0}{b}\text{Var}(\mu^{(\tilde{r}+1)}) + r_0(b-r_0). \quad (25)$$

We discuss about the value of $\lambda(\tilde{r})$ which can be either n or $n+1$.

- (a) If $\lambda(\tilde{r}) = n$, which means that $r_0 \neq b-1$ or $r_1 \neq b-1$: we use the induction hypothesis to get $\text{Var}(\mu^{(\tilde{r})}) \geq \frac{b}{4}\lambda(\tilde{r}) = \frac{bn}{4}$, and it follows that

$$\text{Var}(\mu^{(r)}) \geq \frac{b-r_0}{4}n + \frac{r_0}{b}\text{Var}(\mu^{(\tilde{r}+1)}) + r_0(b-r_0).$$

We now observe that $n-2 \leq \lambda(\tilde{r}+1) \leq n+1$. The reader is referred to Figure 7 (with \tilde{r} instead of r) for more details. We consider two cases.

- i. If $\lambda(\tilde{r}+1) = n+1$ then we cannot apply the induction hypothesis. In this case \tilde{r} is a multiple of b if $b \geq 3$ and even b^2 when $b = 2$ (see Figure 7). So, there exists $\hat{r} \in \mathbb{N}$ and $m \geq 1$ (or 2 if $b = 2$) such that

$$\tilde{r} = b^m \hat{r} \quad \text{and} \quad b \nmid \hat{r}.$$

We can apply Lemma 3.6

$$\text{Var}(\mu^{(\tilde{r}+1)}) = \left(1 - \frac{1}{b^m}\right) \text{Var}(\mu^{(\tilde{r})}) + \frac{1}{b^m} \text{Var}(\mu^{(\tilde{r}+1)}) + b - \frac{1}{b^{m-1}}.$$

Since $\mu^{(\tilde{r})} = \mu^{(\widehat{r})}$ and $\lambda(\tilde{r}) = n$, we deduce from the induction hypothesis and (25) that

$$\begin{aligned}\text{Var}(\mu^{(r)}) &\geq \frac{b-r_0}{4}n + \frac{r_0}{b} \left[\frac{b}{4}n + b - \frac{3}{2b^{m-1}} \right] + \frac{b}{4} \\ &\geq \frac{b}{4}(n+1) + r_0 \left(1 - \frac{3}{2b^{m-1}} \right) \\ &\geq \frac{b}{4}(n+1).\end{aligned}$$

ii. If $\lambda(\tilde{r}+1) \neq n+1$ then we can apply the induction hypothesis.

$$\begin{aligned}\text{Var}(\mu^{(r)}) &\geq \frac{b-r_0}{4}n + \frac{r_0}{2}(n-2) + r_0(b-r_0) \\ &\geq \frac{b}{4}n - \frac{r_0}{2} + r_0(b-r_0).\end{aligned}$$

Thanks to (C_5)

$$r_0(b-r_0) - \frac{r_0}{2} \geq \frac{b}{4}.$$

And so

$$\text{Var}(\mu^{(r)}) \geq \frac{b}{4}n + \frac{b}{4} = \frac{b}{4}(n+1).$$

We conclude the case $\lambda(\tilde{r}) = n$.

(b) If $\lambda(\tilde{r}) = n+1$: it means that the rightmost block in the expansion of r is a block of $(b-1)$'s of length $m \geq 2$. So, there exists $\widehat{r} \in \mathbb{N}$ such that $r = b^m \widehat{r} + b^m - 1$ and the rightmost digit in the expansion of \widehat{r} is not $(b-1)$, that is to say $b \nmid \widehat{r} + 1$. We are in the context of Lemma 3.5, we have

$$\text{Var}(\mu^{(r)}) = \frac{1}{b^m} \text{Var}(\mu^{(\widehat{r})}) + \left(1 - \frac{1}{b^m} \right) \text{Var}(\mu^{(\widehat{r}+1)}) + b - \frac{1}{b^{m-1}}.$$

Since $\lambda(\widehat{r}) = n$ and $n-1 \leq \lambda(\widehat{r}+1) \leq n+1$ (\widehat{r} does not start with a block of $(b-1)$'s so $\lambda(\widehat{r}+1) \neq n-2$, see Figure 7), we have now

$$\text{Var}(\mu^{(r)}) \geq \frac{n}{4b^{m-1}} + \left(1 - \frac{1}{b^m} \right) \text{Var}(\mu^{(\widehat{r}+1)}) + b - \frac{1}{b^{m-1}}. \quad (26)$$

We discuss about the possible values of $\lambda(\widehat{r}+1)$.

i. If $\lambda(\widehat{r}+1) = n+1$ then it is just as in the point (a)i. of this proof, it means that \widehat{r} starts with a block of 0's. So we let $m' \geq 1$ (2 if $b=2$) and $\widehat{\widehat{r}} \in \mathbb{N}$ such that $\widehat{r} = b^{m'} \widehat{\widehat{r}}$ and $b \nmid \widehat{\widehat{r}}$. We are again in the context of Lemma 3.6

$$\text{Var}(\mu^{(\widehat{r}+1)}) = \left(1 - \frac{1}{b^{m'}} \right) \text{Var}(\mu^{(\widehat{\widehat{r}})}) + \frac{1}{b^{m'}} \text{Var}(\mu^{(\widehat{\widehat{r}}+1)}) + b - \frac{1}{b^{m'-1}}.$$

We observe $\lambda(\widehat{r}) = \lambda(\widehat{r}) = n$ and $n - 2 \leq \lambda(\widehat{r} + 1) \leq n$ (again, it cannot be $n + 1$ because \widehat{r} does not start with a block of 0's). So we have

$$\begin{aligned}\text{Var}(\mu^{(\widehat{r}+1)}) &\geq \left(1 - \frac{1}{b^{m'}}\right) \frac{b}{4}n + \frac{1}{4b^{m'-1}}(n-2) + b - \frac{1}{b^{m'-1}} \\ &= \frac{b}{4}n + b - \frac{3}{2b^{m'-1}}.\end{aligned}$$

From (26) we deduce

$$\text{Var}(\mu^{(r)}) \geq \frac{b}{4}n + \left(1 - \frac{1}{b^m}\right) \left(b - \frac{3}{2b^{m'-1}}\right) + b - \frac{1}{b^{m-1}}.$$

We observe that

$$b - \frac{3}{2b^{m'-1}} \geq 0$$

as well as

$$b - \frac{1}{b^{m-1}} \geq \frac{b}{4}.$$

So we can write

$$\text{Var}(\mu^{(r)}) \geq \frac{b}{4}(n+1).$$

ii. If $\lambda(\widehat{r} + 1) \neq n + 1$ then we can apply the induction hypothesis.

$$\begin{aligned}\text{Var}(\mu^{(r)}) &\geq \frac{n}{4b^{m-1}} + \left(1 - \frac{1}{b^m}\right) \frac{b}{4}(n-1) + b - \frac{1}{b^{m-1}} \\ &\geq \frac{b}{4}n - \left(1 - \frac{1}{b^m}\right) \frac{b}{4} + b - \frac{1}{b^{m-1}}\end{aligned}$$

We observe that







$$b - \frac{1}{b^{m-1}} - \left(1 - \frac{1}{b^m}\right) \frac{b}{4} \geq \frac{b}{4},$$











so

$$\text{Var}(\mu^{(r)}) \geq \frac{b}{4}n + \frac{b}{4} = \frac{b}{4}(n+1).$$

It concludes the case $\lambda(\widehat{r}) = n + 1$. The statement is thus true when $\lambda(r) = n + 1$. □

The following figure shows how the number of blocks behaves of an integer when we add 1 to it.

$b = 2$	
Rightmost block(s) of r	$\lambda(r+1) - \lambda(r)$
	1 if length() ≥ 2 0 if length() = 1
	0 if length() ≥ 2 -1 if length() = 1

$b \geq 3$	
Rightmost block(s) of r	$\lambda(r+1) - \lambda(r)$
	1
	0
	-1 if  = $b - 2$ 0 otherwise
	0
	0
	-1
	-1
	-2 if  = $b - 2$ -1 otherwise




 block of 0's
 single digit between 1 and $b - 2$
 block of $(b - 1)$'s

Figure 7: Variations of the number of non-zero blocks when we add 1.

Proof of Theorem 1.2. We use Proposition 3.8, Proposition 3.10 and (20) to get the result. \square

4 A ϕ -mixing process

We work on the probability space $(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mathbb{P})$. For a given integer r , $\Delta^{(r)}$ is viewed as a random variable with law $\mu^{(r)}$ by Corollary 2.2 (the randomness comes from the argument x of $\Delta^{(r)}$, considered as a random outcome in \mathbb{X} with law \mathbb{P}). Our purpose in this section is to study the asymptotic behaviour of $\mu^{(r)}$ as the number of blocks $\rho(r)$ goes to infinity. For this, we will decompose $\Delta^{(r)}$ as a sum

$$\Delta^{(r)} = \sum_{i=1}^{\lambda(r)} X_i^{(r)}$$

where $\lambda(r)$ is the number of non-zero blocks in the base- b expansion of r (see Section 3.3), and $(X_1^{(r)}, \dots, X_{\lambda(r)}^{(r)})$ is a finite process defined in the next section. We will prove and use some mixing properties of this process to get our result.

4.1 The process

Again, the case $r = 0$ is irrelevant so we assume $r \geq 1$. For $1 \leq i \leq \lambda(r)$, we will write B_i as the i^{th} non-zero block present in the expansion of r , starting from the left-hand side of the expansion and ending at the units digit. We now define $r[i]$ as the integer whose base- b expansion is obtained as follows: for $k = i + 1, \dots, \lambda(r)$, we replace the block B_k by a block of 0's of the same length (see Figure 8 below). We observe that $r[\lambda(r)] = r$.

$$\begin{aligned}
 r &= \overline{\underbrace{7}_{B_1} \ 000 \ \underbrace{999}_{B_2} \ 900 \ \underbrace{229}_{B_3 B_4 B_5}} \\
 r[1] &= \overline{7 \ 000 \ 000 \ 000 \ 000} \\
 r[2] &= \overline{7 \ 000 \ 999 \ 900 \ 000} \\
 r[3] &= \overline{7 \ 000 \ 999 \ 900 \ 200} \\
 r[4] &= \overline{7 \ 000 \ 999 \ 900 \ 220} \\
 r[5] &= r
 \end{aligned}$$

Figure 8: Example in base $b = 10$.

With the convention $r[0] := 0$, we observe the trivial equality

$$r = \sum_{i=1}^{\lambda(r)} r[i] - r[i-1].$$

For $1 \leq i \leq \lambda(r)$, we define almost everywhere on \mathbb{X} (see Subsection 2.2)

$$X_i^{(r)} := \Delta^{(r[i]-r[i-1])} \circ T^{r[i-1]}.$$

Since $r[i] - r[i-1] = \overline{B_i 0 \dots 0}$, the function $X_i^{(r)}$ is a random variable corresponding to the action of the i^{th} block B_i once the previous blocks have already been taken into consideration. From (10), we deduce

$$\Delta^{(r)} = \sum_{i=1}^{\lambda(r)} X_i^{(r)}.$$

In particular, if $x \in \mathbb{X}$ is randomly chosen with law \mathbb{P} , then $\sum_{i=1}^{\lambda(r)} X_i^{(r)}(x)$ follows the law $\mu^{(r)}$. Hence, the standard deviation σ_r of $\mu^{(r)}$ defined in Theorem 1.4 satisfies

$$\sigma_r^2 = \text{Var} \left(\sum_{i=1}^{\lambda(r)} X_i^{(r)} \right).$$

We first show that every moment of $X_i^{(r)}$ is bounded from above by a constant independent of r and i .

Lemma 4.1. *For every $k \in \mathbb{N}$, there exists a constant $C_k > 0$ such that*

$$\forall r \in \mathbb{N}, \forall 1 \leq i \leq \lambda(r), \quad \mathbb{E} \left(\left| X_i^{(r)} \right|^k \right) \leq C_k.$$

Proof. Let $k \in \mathbb{N}$. If $X_i^{(r)}$ corresponds to the action of a single-digit block, that is the action of one digit α between 1 and $b-2$, then its law is given by $\mu^{(\alpha)}$ whose moments are all finite (see Corollary 2.2). So, we have in this case

$$\mathbb{E} \left(\left| X_i^{(r)} \right|^k \right) \leq \max \left\{ \mathbb{E} \left(\left| \Delta^{(\alpha)} \right|^k \right) : 1 \leq \alpha \leq b-2 \right\}.$$

Now, if $X_i^{(r)}$ corresponds to the action of a block of $(b-1)$'s, there exist two integers $n \leq m$ such that $r[i] - r[i-1] = b^m - b^n$. It follows from the definition of $X_i^{(r)}$ and (10) that

$$X_i^{(r)} = \Delta^{(b^m - b^n)} \circ T^{b^n} \stackrel{d}{=} \Delta^{(b^m - b^n)} = \Delta^{(b^m)} - \Delta^{(b^n)} \circ T^{b^m - b^n}$$

where $\stackrel{d}{=}$ means the equality in distribution. Since $\mu^{(b^n)} = \mu^{(b^m)} = \mu^{(1)}$, we can write $X_i^{(r)}$ as the difference of two dependent random variables following the law $\mu^{(1)}$ which has finite moments. So there exists a constant that depends only on k such that $\mathbb{E} \left(\left| X_i^{(r)} \right|^k \right)$ is bounded by this constant. \square

The next part is devoted to the estimation of the so-called ϕ -mixing coefficients for the finite sequence $(X_i^{(r)})_{1 \leq i \leq \lambda(r)}$.

4.2 The ϕ -mixing coefficients

There exist many types of mixing coefficients (see e.g. the survey [2] by Bradley). Those we are working with are commonly called “ ϕ -mixing coefficients”.

Definition 4.2. *Let $(X_i)_{i \geq 1}$ be a (finite or infinite) sequence of random variables. The associated ϕ -mixing coefficients $\phi(k)$, $k \geq 1$, are defined by*

$$\phi(k) := \sup_{p \geq 1} \sup_{A, B} |\mathbb{P}_A(B) - \mathbb{P}(B)|$$

where the second supremum is taken over all events A and B such that

- $A \in \sigma(X_i : 1 \leq i \leq p)$,
- $\mathbb{P}(A) > 0$ and
- $B \in \sigma(X_i : i \geq k + p)$.

By convention, if X_i is not defined when $i \geq k + p$ then the σ -algebra is trivial.

In the case of a finite sequence (X_1, \dots, X_n) , the convention implies that $\phi(k) = 0$ for $k \geq n$. We now give an upper bound on the ϕ -mixing coefficients for the process $(X_i^{(r)})$ defined in Section 4.1.

Lemma 4.3. *For $r \geq 1$, the mixing coefficients of $(X_i^{(r)})_{1 \leq i \leq \lambda(r)}$ satisfy*

$$\forall k \geq 1, \quad \phi(k) \leq 2 \left(\frac{b-1}{b} \right)^{\frac{k}{2}-1}.$$

Proof. Let k and p be two integers. We observe that if $k = 1, 2$, the inequality is trivial so we assume that $k \geq 3$. We call *buffer strip* the set of indices corresponding to the positions of the digits between B_p and B_{k+p} (both excluded). It depends on r, p and k so we denote it by $\mathcal{I}_{r,p,k}$. We consider the event

$$C := \{x \in \mathbb{X} : \exists j \in \mathcal{I}_{r,p,k} \text{ with } r_j < b-1 \text{ such that } x_j = 0\}.$$

We are going to show that, for k large enough, C is a high-probability event and that, conditioned to C , two events $A \in \sigma(X_i^{(r)} : 1 \leq i \leq p)$ and $B \in \sigma(X_i^{(r)} : i \geq k+p)$ are always independent.

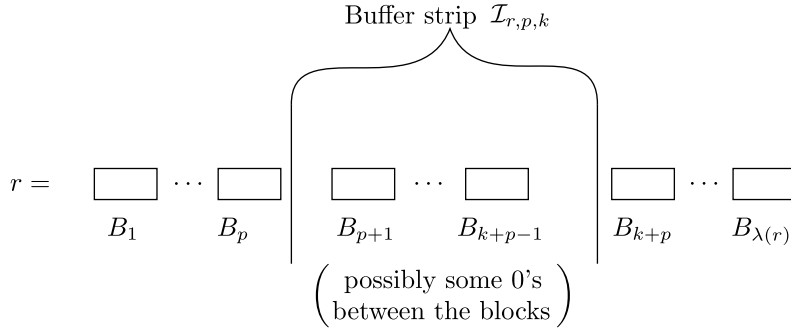


Figure 9: Visualization of the buffer strip.

Denoting by \overline{C} the complement of C in \mathbb{X} , we have

$$\mathbb{P}(\overline{C}) = \left(\frac{b-1}{b} \right)^t \tag{27}$$

where $t := |\{j \in \mathcal{I}_{r,p,k} : r_j \neq b-1\}|$. We are going to show that

$$t \geq \frac{k}{2} - 1. \tag{28}$$

Indeed, there are $k-1$ non-zero blocks in the buffer strip. Let $\ell \in \mathbb{N}$ be the number of blocks of $(b-1)$'s. Then there are $k-1-\ell$ single-digit blocks. There are two cases.

1. If $\ell \leq \frac{k}{2}$, then $k - 1 - \ell \geq \frac{k}{2} - 1$ and we get (28).
2. Otherwise, $\ell > \frac{k}{2}$. Since the blocks of $(b - 1)$'s are separated using blocks of zeros or single-digit blocks, there are at least $\frac{k}{2} - 1$ blocks of zeros of single-digit blocks, which also yields (28).

So, we get from (27) and (28) that

$$\mathbb{P}(C) \geq 1 - \left(\frac{b-1}{b}\right)^{\frac{k}{2}-1} > 0.$$

Now, let $A \in \sigma\left(X_i^{(r)} : 1 \leq i \leq p\right)$ with $\mathbb{P}(A) > 0$ and $B \in \sigma\left(X_i^{(r)} : i \geq k+p\right)$.

Observe that A and C are independent. Indeed, C only depends on indices in $\mathcal{I}_{r,p,k}$ while A , by construction of the random variables $(X_i^{(r)})_{1 \leq i \leq \lambda(r)}$, only depends on the subset of indices on the left-hand side of the buffer strip. We deduce

$$\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C) > 0. \quad (29)$$

Observe also that, conditioned to C , A and B are independent. Indeed, when C is realized, there exists an index j in $\mathcal{I}_{r,p,k}$ such that $r_j \neq b - 1$ and $x_j = 0$. At this position, a carry cannot be created and, furthermore, a carry propagation coming from the right-hand side will be stopped at this index. In other words, when C is realized, the carries created by the blocks B_i , $i \geq k+p$, never spread on the left-hand side of the buffer strip. Moreover, we deduce that A and $B \cap C$ are independent. Indeed, we compute

$$\begin{aligned} \mathbb{P}(A \cap B \cap C) &= \mathbb{P}_C(A \cap B)\mathbb{P}(C) \\ &= \mathbb{P}_C(A)\mathbb{P}_C(B)\mathbb{P}(C) \\ &= \mathbb{P}(A)\mathbb{P}(B \cap C). \end{aligned} \quad (30)$$

Now, we have

$$|\mathbb{P}_A(B) - \mathbb{P}(B)| \leq |\mathbb{P}_A(B) - \mathbb{P}_{A \cap C}(B)| + |\mathbb{P}_{A \cap C}(B) - \mathbb{P}(B)|. \quad (31)$$

But, using (29) and (30), we obtain

$$\mathbb{P}_{A \cap C}(B) = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(A \cap C)} = \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(C)} = \mathbb{P}_C(B).$$

So, (31) becomes

$$|\mathbb{P}_A(B) - \mathbb{P}(B)| \leq |\mathbb{P}_A(B) - \mathbb{P}_{A \cap C}(B)| + |\mathbb{P}_C(B) - \mathbb{P}(B)|. \quad (32)$$

Now, observe that for any event D , by the law of total probability,

$$\begin{aligned} |\mathbb{P}(D) - \mathbb{P}_C(D)| &= |\mathbb{P}(C)\mathbb{P}_C(D) + \mathbb{P}(\overline{C})\mathbb{P}_{\overline{C}}(D) - \mathbb{P}_C(D)| \\ &= \mathbb{P}(\overline{C})|\mathbb{P}_{\overline{C}}(D) - \mathbb{P}_C(D)|. \end{aligned}$$

Since both terms are non negative and less than 1, we have

$$|\mathbb{P}_{\overline{C}}(D) - \mathbb{P}_C(D)| \leq 1.$$

It follows

$$|\mathbb{P}(D) - \mathbb{P}_C(D)| \leq \mathbb{P}(\overline{C}).$$

These computations are also true replacing \mathbb{P} by \mathbb{P}_A . We observe that the measure \mathbb{P}_A conditioned to C is the measure $\mathbb{P}_{A \cap C}$. So, coming back to (32), we get

$$|\mathbb{P}_A(B) - \mathbb{P}(B)| \leq \mathbb{P}_A(\overline{C}) + \mathbb{P}(\overline{C}) = 2\mathbb{P}(\overline{C}) \leq 2 \left(\frac{b-1}{b} \right)^{\frac{k}{2}-1}.$$

□

5 Proof of Theorem 1.4

5.1 A result from Sunklodas and first step of the proof

In this section, we state a result by Sunklodas [8] about the speed of convergence in the Central Limit Theorem for ϕ -mixing sequences. Actually, our formulation is new but the proof is an immediate consequence of [8, Theorem 1].

We first need to introduce some notations. Let Y be a standard normal random variable. Consider ξ_1, \dots, ξ_n a finite sequence of n random variables with ϕ -mixing coefficients $\phi(k)$ for $k = 1, 2, \dots$. Write

$$V := \sqrt{\text{Var} \left(\sum_{i=1}^n \xi_i \right)} \quad \text{and} \quad Z := \sum_{i=1}^n \frac{\xi_i}{V}.$$

We need to add, for technical reason, some other notations.

$$\Phi_{1/2} := \sum_{k \geq 1} k \sqrt{\phi(k)} \quad \text{and} \quad \overline{\Phi}_{1/2} := \max \left\{ \sqrt{\Phi_{1/2}}, \Phi_{1/2}^2 \right\}.$$

We observe that $\Phi_{1/2}$ is defined by a sum on, actually, a finite number of non-zero terms because we are in the case of a finite sequence of random variables. Our formulation of [8, Theorem 1] is the following.

Theorem 5.1. *Assume for $i = 1, \dots, n$ that $\mathbb{E}(\xi_i) = 0$ and $\mathbb{E}(\xi_i^4) < \infty$. Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a thrice differentiable function such that $\|h'''\|_\infty < \infty$. Then*

$$\begin{aligned} \left| \mathbb{E}(h(Z)) - h(Y) \right| &\leq \|h'''\|_\infty \left(\frac{5}{2} + 28\overline{\Phi}_{1/2} \right) \sum_{i=1}^n \mathbb{E} \left(\left| \frac{\xi_i}{V} \right|^3 \right) \\ &\quad + 120 \|h'''\|_\infty \overline{\Phi}_{1/2} \sqrt{\sum_{i=1}^n \mathbb{E} \left(\left| \frac{\xi_i}{V} \right|^2 \right)} \sqrt{\sum_{i=1}^n \mathbb{E} \left(\left| \frac{\xi_i}{V} \right|^4 \right)}. \end{aligned}$$

We are going to show that we can apply this result to prove Theorem 1.4. We start with (3).

Proof of (3). Let $r \in \mathbb{N}^*$. The variables $(X_i^{(r)})_{1 \leq i \leq \lambda(r)}$ are of zero-mean and have a finite moment of order 4. Lemma 4.3 gives a universal upper bound for $\Phi_{1/2}$. So we can apply Theorem 5.1 to the sequence $(X_i^{(r)})_{1 \leq i \leq \lambda(r)}$. We observe that, from (20), Theorem 1.2 and Lemma 4.1, for $j = 2, 3, 4$

$$\sum_{i=1}^{\lambda(r)} \mathbb{E} \left(\left| \frac{X_i^{(r)}}{\sigma_r} \right|^j \right) = \frac{C_j \lambda(r)}{\sigma_r^j} \leq \frac{2^j C_j}{b^{\frac{j}{2}}} \rho(r)^{1-\frac{j}{2}}.$$

So, by Theorem 5.1, there exists $K > 0$ such that

$$\left| \mathbb{E} \left(h \left(\sum_{i=1}^{\lambda(r)} \frac{X_i^{(r)}}{\sigma_r} \right) - h(Y) \right) \right| \leq \frac{K \|h'''\|_\infty}{\sqrt{\rho(r)}} \quad (33)$$

□

It remains to prove (4).

5.2 Speed of convergence of the cumulative distribution functions

Observe that for all $t \in \mathbb{R}$

$$F_r(t) - F(t) = \mathbb{E} \left(\mathbb{1}_{]-\infty, t]} \left(\sum_{i=1}^{\lambda(r)} \frac{X_i^{(r)}}{\sigma_r} \right) - \mathbb{1}_{]-\infty, t]}(Y) \right).$$

The idea is thus to find, for $t \in \mathbb{R}$, a family of thrice differentiable function $(h_{t,\varepsilon})_{\varepsilon > 0}$ with, for every $\varepsilon > 0$, $\|h_{t,\varepsilon}'''\|_\infty < \infty$ and which converges pointwise to the indicator function $\mathbb{1}_{]-\infty, t]}$ when ε tends to 0.

5.2.1 Approximation of the indicator function

There exists a function $f : \mathbb{R} \rightarrow \mathbb{R} \in \mathcal{C}^3(\mathbb{R})$ satisfying the following conditions $f'(0) = f'(1) = f''(0) = f''(1) = f'''(0) = f'''(1) = 0$, $f(t) = 1$ if $t \leq 0$, $f(t) = 0$ if $t \geq 1$ and $0 \leq f(t) \leq 1$, for all real t . Then we define the linear function $\theta_{t,\varepsilon} : [t - \varepsilon, t + \varepsilon] \rightarrow [0, 1]$, $u \mapsto \frac{1}{2\varepsilon}(\varepsilon - t + u)$. Finally, we get our approximation by

$$h_{t,\varepsilon}(u) := \begin{cases} 1 & \text{if } u \leq t - \varepsilon, \\ f \circ \theta_{t,\varepsilon}(u) & \text{if } t - \varepsilon \leq u \leq t + \varepsilon, \\ 0 & \text{otherwise.} \end{cases}$$

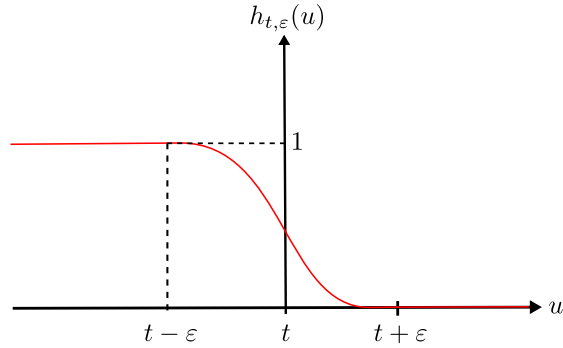


Figure 10: Graph of $h_{t,\epsilon}$.

We have the following properties satisfy by $h_{t,\epsilon}$.

Lemma 5.2. *Let $r \geq 1$, let Y be a standard normal random variable.*

1. $\forall t \in \mathbb{R}$, the sequence of $\mathcal{C}^3(\mathbb{R})$ functions $(h_{t,\epsilon})_\epsilon$ converges pointwise to the indicator function $\mathbb{1}_{]-\infty; t]}$ when ϵ tends to 0.
2. $\forall \epsilon > 0, \forall t \in \mathbb{R}, \|h_{t,\epsilon}'''\|_\infty = \frac{\|f'''\|_\infty}{8\epsilon^3}$ and, in particular, the upper bound is independent of t .
3. $\forall \epsilon > 0$, we have, for any random variable X

$$\begin{aligned} \sup_{t \in \mathbb{R}} \left| \mathbb{P}(X \leq t) - \mathbb{P}(Y \leq t) \right| &\leq \sup_{t \in \mathbb{R}} \left| \mathbb{E}(h_{t,\epsilon}(X) - h_{t,\epsilon}(Y)) \right| \\ &+ \frac{4\epsilon}{\sqrt{2\pi}}. \end{aligned} \quad (34)$$

5.2.2 Last step of the proof of (4)

Proof of (4). Let $\epsilon > 0$. From (33) and (34), we obtain

$$\begin{aligned} \sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\frac{\Delta^{(r)}}{\sigma_r} \leq t\right) - \mathbb{P}(Y \leq t) \right| &\leq \sup_{t \in \mathbb{R}} \left| \mathbb{E}\left(h_{t,\epsilon}\left(\frac{\Delta^{(r)}}{\sigma_r}\right) - h_{t,\epsilon}(Y)\right) \right| + \frac{4\epsilon}{\sqrt{2\pi}} \\ &\leq \sup_{t \in \mathbb{R}} \left(\frac{K \|h_{t,\epsilon}'''\|_\infty}{\sqrt{\rho(r)}} \right) + \frac{4\epsilon}{\sqrt{2\pi}} \end{aligned}$$

Lemma 5.2 gives $\|h_{t,\epsilon}'''\|_\infty = \frac{\|f'''\|_\infty}{8\epsilon^3}$. So, we obtain

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\frac{\Delta^{(r)}}{\sigma_r} \leq t\right) - \mathbb{P}(Y \leq t) \right| \leq \frac{\|f'''\|_\infty K}{8\epsilon^3 \sqrt{\rho(r)}} + \frac{4\epsilon}{\sqrt{2\pi}}$$

Now, we choose $\varepsilon > 0$ such that $\frac{1}{\varepsilon^3 \sqrt{\rho(r)}} = \frac{4\varepsilon}{\sqrt{2\pi}}$ and get the existence of a constant $\tilde{K} > 0$ such that

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P} \left(\frac{\Delta^{(r)}}{\sigma_r} \leq t \right) - \mathbb{P}(Y \leq t) \right| \leq \frac{\tilde{K}}{\rho(r)^{\frac{1}{8}}}.$$

□

It only remains to prove Lemma 5.2.

5.2.3 Proof of Lemma 5.2

Proof of Lemma 5.2. Let $\varepsilon > 0$. The first point is trivial by construction of $h_{t,\varepsilon}$. The second point is also quite simple to show. We write

$$\begin{aligned} \sup_{u \in \mathbb{R}} |h_{t,\varepsilon}'''(u)| &= \sup_{t-\varepsilon \leq u \leq t+\varepsilon} |h_{t,\varepsilon}'''(u)| \\ &= \sup_{t-\varepsilon \leq u \leq t+\varepsilon} \left| \theta'_{t,\varepsilon}(u) \right|^3 \left| f''' \circ \theta_{t,\varepsilon}(u) \right| \\ &= \frac{1}{8\varepsilon^3} \sup_{0 \leq u \leq 1} |f'''(u)|. \end{aligned}$$

For the last point, let t and $x \in \mathbb{R}$. Since

$$h_{t-\varepsilon,\varepsilon}(x) \leq \mathbf{1}_{]-\infty, t]}(x) \leq h_{t+\varepsilon,\varepsilon}(x),$$

we deduce that for any random variable X

$$\mathbb{E}(h_{t-\varepsilon,\varepsilon}(X)) \leq \mathbb{P}(X \leq t) \leq \mathbb{E}(h_{t+\varepsilon,\varepsilon}(X)). \quad (35)$$

Moreover,

$$\mathbb{E}(h_{t+\varepsilon,\varepsilon}(Y)) - \mathbb{E}(h_{t-\varepsilon,\varepsilon}(Y)) = \int_{t-2\varepsilon}^{t+2\varepsilon} \underbrace{(h_{t+\varepsilon,\varepsilon}(y) - h_{t-\varepsilon,\varepsilon}(y))}_{\leq 1} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy \leq \frac{4\varepsilon}{\sqrt{2\pi}}.$$

Hence, we get that

$$\mathbb{E}(h_{t+\varepsilon,\varepsilon}(Y)) - \frac{4\varepsilon}{\sqrt{2\pi}} \leq \mathbb{P}(Y \leq t) \leq \mathbb{E}(h_{t-\varepsilon,\varepsilon}(Y)) + \frac{4\varepsilon}{\sqrt{2\pi}}. \quad (36)$$

Then, we subtract (36) from (35) and we take the supremum over $t \in \mathbb{R}$, observing that

$$\sup_{t \in \mathbb{R}} \left| \mathbb{E} \left(h_{t-\varepsilon,\varepsilon}(X) - h_{t-\varepsilon,\varepsilon}(Y) \right) \right| = \sup_{t \in \mathbb{R}} \left| \mathbb{E} \left(h_{t+\varepsilon,\varepsilon}(X) - h_{t+\varepsilon,\varepsilon}(Y) \right) \right|.$$

We get (34). □

References

- [1] Jean Bésineau. Indépendance statistique d'ensembles liés à la fonction "somme des chiffres". In *Séminaire de Théorie des Nombres, 1970-1971 (Univ. Bordeaux I, Talence), Exp. No. 21*, page 20 pp. Lab. Théorie des Nombres. 1971.
- [2] Richard C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.*, 2:107–144, 2005. Update of, and a supplement to, the 1986 original.
- [3] Jordan Emme and Pascal Hubert. Central limit theorem for probability measures defined by sum-of-digits function in base 2. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 19(2):757–780, 2019.
- [4] Jordan Emme and Alexander Prikhod'ko. On the asymptotic behavior of density of sets defined by sum-of-digits function in base 2. *Integers*, 17:Paper No. A58, 28, 2017.
- [5] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.
- [6] Johannes F. Morgenbesser and Lukas Spiegelhofer. A reverse order property of correlation measures of the sum-of-digits function. *Integers*, 12:Paper No. A47, 5, 2012.
- [7] Lukas Spiegelhofer and Michael Wallner. The digits of $n+t$. arXiv : 2005.07167, 2021.
- [8] Jonas Kazys Sunklodas. Some estimates of the normal approximation for ϕ -mixing random variables. *Lith. Math. J.*, 51(2):260–273, 2011.