# Density of sets defined by sum-of-digits function in base 2

J. Emme, A. Prikhodko, P. Hubert

25 janvier 2017

# Summary

# The problem and its motivation

Let $n \in \mathbb{N}$, there exists $\{n_0, ..., n_m\} \in \{0,1\}^{m+1}$ such that

$$n = \sum_{k=0}^{m} n_k 2^k.$$

Define the sum-of-digit function in base 2, $s_2$ :

$$\forall n \in \mathbb{N}, s_2(n) = \sum_{k=0}^{m} n_k,$$

denote

$$\underline{n}_2 = n_m ... n_0 \in \{0,1\}^*$$

we are interested, for any parameters $a \in \mathbb{N}$ and $d \in \mathbb{Z}$, in the following equation :

$$\boxed{s_2(n+a) - s_2(n) = d.}$$

# The problem and its motivation

This quantity naturally appears in Bésineau's work on the autocorrelation of some arithmetic functions :

$$\begin{array}{rccc} f : & \mathbb{N} & \to & \mathbb{C} \\ & n & \mapsto & e^{2i\pi\alpha s_2(n)} \end{array}$$

where $\alpha$ is a real parameter. The autocorrelation of $f$ is the function $\gamma_f$ defined by :

$$\forall a \in \mathbb{N},\ \gamma_f(a) = \lim_{N\to+\infty} \frac{1}{N} \sum_{n=0}^{N-1} \overline{f(n)} f(n+a).$$

# The problem and its motivation

**Theorem (Bésineau, 1971)**

*For any integer a, $\mathbb{N}$ can be partitioned in arithmetic progressions such that $n \mapsto s_2(n+a) - s_2(n)$ is constant on every element of this partition.*

**Corollary**

*For any $a \in \mathbb{N}$ and $d \in \mathbb{Z}$, the set $E_{a,d} := \{n \in \mathbb{N} \mid s_2(n+a) - s_2(n) = d\}$ admits an asymptotic density named $\mu_a(d)$. Namely :*

$$\boxed{\mu_a(d) = \lim_{N \to +\infty} \frac{\#E_{a,d} \cap \{0, \ldots, N-1\}}{N}.}$$

Remark that $\mu_a$ is a probability measure on $\mathbb{Z}$.

# The problem and its motivation

We state a conjecture by Cusick. For any $a \in \mathbb{N}$, denote $c_a := \sum_{d \in \mathbb{N}} \mu_a(d)$.

It is conjectured that:

$$\forall a \in \mathbb{N}, c_a > \frac{1}{2}$$

and

$$\liminf_{a \in \mathbb{N}} c_a = \frac{1}{2}.$$

The results we expose here give only a partial answer.

# Combinatorial study of the main equation

## Proposition

- $\forall a \in \mathbb{N}, \quad \forall d \in \mathbb{Z}, \quad \exists S \subset \{0,1\}^* \quad$ *such that*

$$s_2(n+a) - s_2(n) = d \Leftrightarrow \exists w \in \{0,1\}^*,\ \exists s \in S,\ \underline{n}_2 = ws.$$

- *For all* $a \in \mathbb{N}$*, there exists an infinite tree* $\tau_a$ *with edges labelled in* $\{0,1\}$ *such that the prefixes can be read as paths on this tree.*

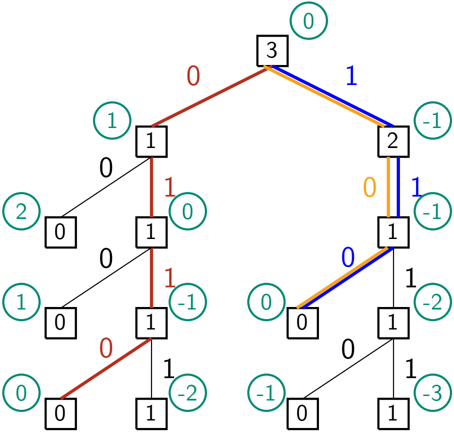
*Remark :*

- We can prove the first point by induction on $a$.
- Any tree $\tau_a$ is computed via a simple inductive process.

*Example :* The solutions of the equation $s_2(n+3) - s_2(n) = 0$

are, in binary:

- **0110**
- **011**
- **001.**



# Computing $\mu_a$ and first asymptotic properties

## Proposition

*For every $a \in \mathbb{N}$ and $d \in \mathbb{Z}$.*

$$\mu_{2a}(d) = \mu_a(d)$$

$$\mu_{2a+1}(d) = \frac{1}{2}\left(\mu_a(d-1) + \mu_{a+1}(d+1)\right)$$

This can be seen by noticing that:

$$E_{2a,d} = \{2n+b,\ n \in E_{a,d}, b \in \{0,1\}\}$$

and

$$E_{2a+1,d} = \{2n,\ n \in E_{a,d-1}\} \cup \{2n+1,\ n \in E_{a+1,d+1}\}$$

# Computing $\mu_a$ and first asymptotic properties

Denote $\widehat{\mu}_a$ the Fourier transform of the measure $\mu_a$, i.e.

$$\forall \theta \in [0, 2\pi), \ \widehat{\mu}_a(\theta) = \sum_{d \in \mathbb{Z}} e^{id\theta} \mu_a(d).$$

## Proposition

*For any $a \in \mathbb{N}$ and any $\theta \in [0, 2\pi)$ :*

$$\widehat{\mu}_{2a}(\theta) = \widehat{\mu}_a(\theta)$$

$$\widehat{\mu}_{2a+1}(\theta) = \frac{e^{i\theta}}{2} \widehat{\mu}_a(\theta) + \frac{e^{-i\theta}}{2} \widehat{\mu}_{a+1}(\theta)$$

We can easily compute

$$\widehat{\mu}_0(\theta) = 1$$

and

$$\widehat{\mu}_1(\theta) = \frac{e^{i\theta}}{2 - e^{-i\theta}}$$

# Computing $\mu_a$ and first asymptotic properties

We get the following central property:

## Proposition

*Let* $a \in \mathbb{N}$ *s.t.* $a = \sum_{k=0}^{m} a_k 2^k$,

$$\forall \theta \in [0, 2\pi),\ \widehat{\mu_a}(\theta) = \begin{pmatrix} 1 & 0 \end{pmatrix} \hat{A}_{a_0}(\theta) \cdots \hat{A}_{a_{m-1}}(\theta) \hat{A}_{a_m}(\theta) \begin{pmatrix} \widehat{\mu_0}(\theta) \\ \widehat{\mu_1}(\theta) \end{pmatrix}$$

*where*

$$\forall \theta \in [0, 2\pi),\ \hat{A}_0(\theta) := \begin{pmatrix} 1 & 0 \\ \frac{1}{2} e^{i\theta} & \frac{1}{2} e^{-i\theta} \end{pmatrix},\ \hat{A}_1(\theta) := \begin{pmatrix} \frac{1}{2} e^{i\theta} & \frac{1}{2} e^{-i\theta} \\ 0 & 1 \end{pmatrix}.$$

# Computing $\mu_a$ and first asymptotic properties

## Definition

For any $a$ in $\mathbb{N}$, denote $l(a)$ the number of occurrences of 01 in the binary expansion of $a$.

## Theorem (E, Prikhod'ko 2015)

*There exists $C > 0$ s.t.:*

$$\forall a \in \mathbb{N}, \; \|\mu_a\|_{l^2(\mathbb{Z})} \leq C\,(l(a))^{-\frac{1}{4}}$$

*Proof :*

In order to prove this theorem we study the Fourier transform $\widehat{\mu}_a$ of $\mu_a$.

# Computing $\mu_a$ and first asymptotic properties

## Definition

Let's define the algebraic norm $\|\cdot\|_1$ on $n \times n$ matrices :

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{i,j}|.$$

## Lemma

*For all* $\theta \in [-\pi, \pi)$,

$$\|\widehat{A_0}(\theta)\|_1 = \|\widehat{A_1}(\theta)\|_1 = \|\widehat{A_0}(\theta)\widehat{A_1}(\theta)\|_1 = \|\widehat{A_1}(\theta)\widehat{A_0}(\theta)\|_1 = 1$$

$$\|\widehat{A_0}(\theta)\widehat{A_1}(\theta)\widehat{A_0}(\theta)\|_1 = \|\widehat{A_1}(\theta)\widehat{A_0}(\theta)\widehat{A_1}(\theta)\|_1 = \phi(\theta) \leq e^{-\theta^2/15}.$$

*and for any* $k \in \mathbb{N}^*$

$$\|\widehat{A_0}(\theta)\,(\widehat{A_1}(\theta))^k\,\widehat{A_0}(\theta)\|_1 \leq \|\widehat{A_0}(\theta)\,\widehat{A_1}(\theta)\,\widehat{A_0}(\theta)\|_1$$

# Computing $\mu_a$ and first asymptotic properties

Applying the lemma to half of the patterns $\widehat{A_0}(\theta)\,(\widehat{A_1}(\theta))^k\,\widehat{A_0}(\theta)$ encountered in the matricial expression of $\widehat{\mu}_a$ yields :

$$|\widehat{\mu}_a(\theta)| \leq \phi^{\frac{l(a)-1}{2}}(\theta)$$

which in turn implies

$$\|\widehat{\mu}_a\|_2 \leq \left(\frac{15\pi}{l(a)-1}\right)^{1/4}$$

and proves the theorem.

$\square$

# Computing $\mu_a$ and first asymptotic properties

**Theorem (E, Prikhod'ko 2015)**

*For any $a$, $\mu_a$ is centered and its variance denoted $Var(\mu_a)$ satisfies:*

$$l(a) - 1 \leq Var(\mu_a) \leq 4l(a) + 2$$

Question : For a given sequence $(a_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to+\infty} l(a_n) = +\infty$, what can we say about $\left(\frac{Var(\mu_{a_n})}{l(a_n)}\right)_{n\in\mathbb{N}}$?

# A central limit theorem

Let $X \in \{0,1\}^{\mathbb{N}}$. Define the associated sequence $(a_X(n))_{n\in\mathbb{N}}$ by:

$$\forall n \in \mathbb{N},\ a_X(n) = \sum_{k=0}^{n-1} X_k 2^k.$$

Let $\mathbb{P}$ be the symmetric Bernoulli measure $\{0,1\}^{\mathbb{N}}$. We have the following:

**Theorem (E, Hubert 2016)**

*There exists $\mathcal{U} \subset \{0,1\}^{\mathbb{N}}$ such that $\mathbb{P}(\mathcal{U}) = 1$ and, defining for every $X \in \mathcal{U}$, $n \in \mathbb{N}$, $\delta \in \sqrt{\frac{2}{n}}\mathbb{Z}$, the renormalised measure $\widetilde{\mu}_{a_X(n)}(\delta) = \mu_{a_X(n)}(\sqrt{\frac{n}{2}}\delta)$, we get:*

$$\widetilde{\mu}_{a_X(n)} \underset{n\to\infty}{\rightharpoonup} \varphi$$

*where*

$$\forall t \in \mathbb{R},\ \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}.$$

# A central limit theorem

First we compute $\mathrm{Var}(\mu_a)$:

## Proposition

*For any* $a \in \mathbb{N}$ *with* $\underline{a}_2 = a_n ... a_0$*, define, for any* $j \in \{0, ..., n\}$, $b_j = (-1)^{a_j+1}$. *Then the variance of* $\mu_a$ *is:*

$$Var(\mu_a) = \frac{n+3}{2} - \frac{1}{2^{n+1}} - \frac{1}{2}\sum_{i=1}^{n}\sum_{k=0}^{n-i}\frac{b_{k+i}b_k}{2^i} + \sum_{k=0}^{n}\frac{b_k + b_{n-k}}{2^{k+1}}.$$

So for a generic $X$ for the measure $\mathbb{P}$, we see that:

$$Var\left(\mu_{a_X(n)}\right) \sim \frac{n}{2}$$

*Proof:*

We start by computing the Taylor expansion of $\widehat{\mu_a}$ near 0.

# A central limit theorem

Indeed, we have:

$$\hat{A}_j(\theta) = I_j + i\theta\alpha_j - \frac{1}{2}\theta^2\beta_j + O(\theta^3).$$

with $j \in \{0,1\}$ and the matrices playing a role in the Taylor expansion of $\widehat{\mu_a}$ near 0 being:

$$I_0 = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \alpha_0 = \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix}, \quad \beta_0 = \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix},$$

$$I_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{pmatrix}, \quad \alpha_1 = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}, \quad \beta_1 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

Notice that the Taylor expansion near 0 of $\theta \mapsto \frac{e^{-i\theta}}{2-e^{i\theta}}$ is:

$$\frac{e^{i\theta}}{2 - e^{-i\theta}} = 1 - \theta^2 + O(\theta^3).$$

# A central limit theorem

The variance is given by:

$$\mathrm{Var}(\mu_a) = \begin{pmatrix} 1 & 0 \end{pmatrix} \left( (\beta_{a_0} + I_{a_0}\beta_{a_1} + \ldots + I_{a_0}\ldots I_{a_{n-1}}\beta_{a_n}) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + I_{a_0}\ldots I_{a_n} \begin{pmatrix} 0 \\ 2 \end{pmatrix} \right).$$

since

$$\alpha_j \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0$$

and

$$I_j \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

# A central limit theorem

We apply a change of basis to simultanously trigonalize the matrices $I_0$ and $I_1$.
Let us note

$$P := \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \quad P^{-1} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

and compute

$$\forall j \in \{0,1\},\ PI_jP^{-1} = \begin{pmatrix} 1 & \frac{(-1)^{j+1}}{2} \\ 0 & \frac{1}{2} \end{pmatrix}, \quad P\beta_jP^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ -\frac{(-1)^{j+1}}{2} & 0 \end{pmatrix}$$

and

$$P\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad P\begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \end{pmatrix} P^{-1} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

# A central limit theorem

Idea of proof for the Central Limit Theorem:

- Do a preliminary study of which terms might contribute after the renormalization
- Compute explicitly the important terms after renormalization
- Estimate the correlations almost everywhere in order to get a limit

Thank you for your attention.