Sums of the digits in bases 2 and 3

Shanta Laishram

Stat Math Unit, Indian Statistical Institute
New Delhi, India

Prime Numbers and Automatic Sequences: Determinism and Randomness
CIRM, Luminy

May 26, 2017

Joint work with

J.-M. Deshouillers, L. Habsieger and B. Landreau

Supported by MuDeRa and CEFIPRA

Sum of Digits

- Let $b \geq 2$ and $n \geq 1$, let $s_b(n)$ denote the sum of digits of n in base b.
- If we write $n = \sum_{j \geq 0} \varepsilon_j b^j$ with $0 \leq \varepsilon_j < b$, then $s_b(n) = \sum_{j \geq 0} \varepsilon_j$.

Sum of Digits

- Let $b \geq 2$ and $n \geq 1$, let $s_b(n)$ denote the sum of digits of n in base b.
- If we write $n = \sum_{j\geq 0} \varepsilon_j b^j$ with $0 \leq \varepsilon_j < b$, then $s_b(n) = \sum_{j\geq 0} \varepsilon_j$.
- When $b = 2$, $s_b(n)$ is the number of $1'$s in the binary representation of n.

Sum of Digits

- Let $b \geq 2$ and $n \geq 1$, let $s_b(n)$ denote the sum of digits of n in base b.
- If we write $n = \sum_{j \geq 0} \varepsilon_j b^j$ with $0 \leq \varepsilon_j < b$, then $s_b(n) = \sum_{j \geq 0} \varepsilon_j$.
- When $b = 2$, $s_b(n)$ is the number of $1'$s in the binary representation of n.
- $s_b(n)$ is well studied in different variants.

Sum of Digits

- One of the interesting question in the distribution of digits is the following:

Sum of Digits

- One of the interesting question in the distribution of digits is the following:
- Which digits appear infinitely often as the last nonzero digit of $n!$ in base b?

Sum of Digits

- One of the interesting question in the distribution of digits is the following:
- Which digits appear infinitely often as the last nonzero digit of $n!$ in base b?
- **Legendre:** $\nu_p(n!) = \frac{n - s_p(n)}{p-1}$ for any prime p where $p^{\nu_p(n!)} || n!$.

Sum of Digits

- One of the interesting question in the distribution of digits is the following:
- Which digits appear infinitely often as the last nonzero digit of $n!$ in base b?
- **Legendre:** $\nu_p(n!) = \frac{n - s_p(n)}{p-1}$ for any prime p where $p^{\nu_p(n!)} || n!$.
- Let $b = \prod_{p^a || n} p^a$. Then the number of zeroes in base at the end of $n!$ is

$$\min_{p^a || b} \left\lfloor \frac{n - s_p(n)}{a(p-1)} \right\rfloor$$

Sum of Digits

- One of the interesting question in the distribution of digits is the following:
- Which digits appear infinitely often as the last nonzero digit of $n!$ in base b?
- **Legendre:** $\nu_p(n!) = \frac{n-s_p(n)}{p-1}$ for any prime p where $p^{\nu_p(n!)}||n!$.
- Let $b = \prod_{p^a||n} p^a$. Then the number of zeroes in base at the end of $n!$ is

$$\min_{p^a||b} \left\lfloor \frac{n - s_p(n)}{a(p-1)} \right\rfloor$$

- Write $b = p_1^{a_1} \cdots p_r^{a_r}$ with $a_1(p_1 - 1) \leq \cdots \leq a_r(p_r - 1)$.

Sum of Digits

- One of the interesting question in the distribution of digits is the following:
- Which digits appear infinitely often as the last nonzero digit of $n!$ in base b?
- **Legendre:** $\nu_p(n!) = \frac{n-s_p(n)}{p-1}$ for any prime p where $p^{\nu_p(n!)}||n!$.
- Let $b = \prod_{p^a||n} p^a$. Then the number of zeroes in base at the end of $n!$ is

$$\min_{p^a||b} \left\lfloor \frac{n-s_p(n)}{a(p-1)} \right\rfloor$$

- Write $b = p_1^{a_1} \cdots p_r^{a_r}$ with $a_1(p_1-1) \le \cdots \le a_r(p_r-1)$.
- When $a_1(p_1-1) < a_2(p_2-1)$, multiples of p_1 appear as the last nonzero digit of $n!$ in base b.

Sum of Digits

- One of the interesting question in the distribution of digits is the following:
- Which digits appear infinitely often as the last nonzero digit of $n!$ in base b?
- **Legendre:** $\nu_p(n!) = \frac{n-s_p(n)}{p-1}$ for any prime p where $p^{\nu_p(n!)} || n!$.
- Let $b = \prod_{p^a || n} p^a$. Then the number of zeroes in base at the end of $n!$ is

$$\min_{p^a || b} \left\lfloor \frac{n - s_p(n)}{a(p-1)} \right\rfloor$$

- Write $b = p_1^{a_1} \cdots p_r^{a_r}$ with $a_1(p_1 - 1) \leq \cdots \leq a_r(p_r - 1)$.
- When $a_1(p_1 - 1) < a_2(p_2 - 1)$, multiples of p_1 appear as the last nonzero digit of $n!$ in base b.
- This is the case when b is a prime power or a squarefree integer.

Sum of Digits

- It becomes nontrivial when $a_1(p_1 - 1) = a_2(p_2 - 1)$.

Sum of Digits

- It becomes nontrivial when $a_1(p_1 - 1) = a_2(p_2 - 1)$.
- The first nontrivial case is when $b = 12 = 2^2 \times 3$.

Sum of Digits

- It becomes nontrivial when $a_1(p_1 - 1) = a_2(p_2 - 1)$.
- The first nontrivial case is when $b = 12 = 2^2 \times 3$.
- **Deshouillers and Rusza; Deshouillers**: The digits $3, 4, 6, 8, 9$ appear infinitely often as the last nonzero digit of $n!$ in base 12.

Sum of Digits

- It becomes nontrivial when $a_1(p_1 - 1) = a_2(p_2 - 1)$.
- The first nontrivial case is when $b = 12 = 2^2 \times 3$.
- **Deshouillers and Rusza; Deshouillers**: The digits $3, 4, 6, 8, 9$ appear infinitely often as the last nonzero digit of $n!$ in base 12.
- We still dont know if other digits appear infinitely often as the last nonzero digit of $n!$ in base 12.

Sum of Digits

- It becomes nontrivial when $a_1(p_1 - 1) = a_2(p_2 - 1)$.
- The first nontrivial case is when $b = 12 = 2^2 \times 3$.
- **Deshouillers and Rusza; Deshouillers**: The digits $3, 4, 6, 8, 9$ appear infinitely often as the last nonzero digit of $n!$ in base 12.
- We still dont know if other digits appear infinitely often as the last nonzero digit of $n!$ in base 12.
- It can be easily shown that at least one of the digits $\{1, 5, 7, 11\}$ appear infinitely often if and only if $s_3(n) = s_2(n)$ happens for infinitely many n.

Sum of Digits

- It becomes nontrivial when $a_1(p_1 - 1) = a_2(p_2 - 1)$.
- The first nontrivial case is when $b = 12 = 2^2 \times 3$.
- **Deshouillers and Rusza; Deshouillers**: The digits $3, 4, 6, 8, 9$ appear infinitely often as the last nonzero digit of $n!$ in base 12.
- We still dont know if other digits appear infinitely often as the last nonzero digit of $n!$ in base 12.
- It can be easily shown that at least one of the digits $\{1, 5, 7, 11\}$ appear infinitely often if and only if $s_3(n) = s_2(n)$ happens for infinitely many n.
- However it is a difficult problem to show that $s_3(n) = s_2(n)$ happens for infinitely many n.

Sum of Digits

- It becomes nontrivial when $a_1(p_1 - 1) = a_2(p_2 - 1)$.
- The first nontrivial case is when $b = 12 = 2^2 \times 3$.
- **Deshouillers and Rusza; Deshouillers**: The digits $3, 4, 6, 8, 9$ appear infinitely often as the last nonzero digit of $n!$ in base 12.
- We still dont know if other digits appear infinitely often as the last nonzero digit of $n!$ in base 12.
- It can be easily shown that at least one of the digits $\{1, 5, 7, 11\}$ appear infinitely often if and only if $s_3(n) = s_2(n)$ happens for infinitely many n.
- However it is a difficult problem to show that $s_3(n) = s_2(n)$ happens for infinitely many n.
- Even we dont know if $|s_3(n) - s_2(n)|$ is *significantly small* infinitely often.

Sum of Digits

- It becomes nontrivial when $a_1(p_1 - 1) = a_2(p_2 - 1)$.
- The first nontrivial case is when $b = 12 = 2^2 \times 3$.
- **Deshouillers and Rusza; Deshouillers**: The digits $3, 4, 6, 8, 9$ appear infinitely often as the last nonzero digit of $n!$ in base 12.
- We still dont know if other digits appear infinitely often as the last nonzero digit of $n!$ in base 12.
- It can be easily shown that at least one of the digits $\{1, 5, 7, 11\}$ appear infinitely often if and only if $s_3(n) = s_2(n)$ happens for infinitely many n.
- However it is a difficult problem to show that $s_3(n) = s_2(n)$ happens for infinitely many n.
- Even we dont know if $|s_3(n) - s_2(n)|$ is *significantly small* infinitely often.
- In this talk, we will try to address this question.

N. L. Bassily and I. Kátai

Theorem 1.

Let ψ be a function tending to infinity with its argument. The sequence of natural numbers n for which

$$\begin{aligned}\left(\frac{1}{\log 3}-\frac{1}{\log 4}\right)\log n-\psi(n)\sqrt{\log n}\\ &\leq s_3(n)-s_2(n)\\ &\leq\left(\frac{1}{\log 3}-\frac{1}{\log 4}\right)\log n+\psi(n)\sqrt{\log n}\end{aligned}$$

has asymptotic natural density 1*.*

- In other words, the sequence of natural numbers n for which

$$\left| s_3(n) - s_2(n) - \left(\frac{1}{\log 3} - \frac{1}{\log 4} \right) \log n \right| \leq \psi(n) \sqrt{\log n}$$

has asymptotic natural density 1.

N. L. Bassily and I. Kátai

- In other words, the sequence of natural numbers n for which

$$\left|s_3(n) - s_2(n) - \left(\frac{1}{\log 3} - \frac{1}{\log 4}\right)\log n\right| \leq \psi(n)\sqrt{\log n}$$

has asymptotic natural density 1.

- We prove that there are infinitely many n for which $\frac{|s_3(n)-s_2(n)|}{\log n}$ is *significantly smaller* than $(\frac{1}{\log 3} - \frac{1}{\log 4}) = 0.18889\cdots$.

Theorem 2.

For sufficiently large N, we have

$$\#\{n \le N : |s_3(n) - s_2(n)| \le 0.146 \log n\} > N^{0.97}.$$

Remarks

- For the proof, we use separate (or marginal) distributions of $(s_2(n))_n$ and $(s_3(n))_n$, without using any further information concerning their joint distribution.

Remarks

- For the proof, we use separate (or marginal) distributions of $(s_2(n))_n$ and $(s_3(n))_n$, without using any further information concerning their joint distribution.
- However our present knowledge of the joint distribution of s_2 and s_3 (either the Diophantine approach or the probabilistic approach) does not allow us to improve on Theorem 2.

A Heuristic Approach

- For a positive integer n, let

$$n = \sum_{j=0}^{J(n)} \varepsilon_j(n) b^j \text{ with } J(n) = \left\lfloor \frac{\log n}{\log b} \right\rfloor .$$

A Heuristic Approach

- For a positive integer n, let

$$n = \sum_{j=0}^{J(n)} \varepsilon_j(n) b^j \text{ with } J(n) = \left\lfloor \frac{\log n}{\log b} \right\rfloor .$$

- The smaller is j the more equidistributed are the $\varepsilon_j(n)$'s, and the smaller are the elements of a family $\mathcal{J} = \{j_1 < j_2 < \cdots < j_s\}$ the more independent are the $\varepsilon_j(n)$'s for $j \in \mathcal{J}$.

A Heuristic Approach

- For a positive integer n, let

$$n = \sum_{j=0}^{J(n)} \varepsilon_j(n) b^j \text{ with } J(n) = \left\lfloor \frac{\log n}{\log b} \right\rfloor .$$

- The smaller is j the more equidistributed are the $\varepsilon_j(n)$'s, and the smaller are the elements of a family $\mathcal{J} = \{j_1 < j_2 < \cdots < j_s\}$ the more independent are the $\varepsilon_j(n)$'s for $j \in \mathcal{J}$.
- Given N, a model of $s_b(n)$ for integers n around N is to consider a sum of $\lfloor \frac{\log N}{\log b} \rfloor$ independent random variables uniformly distributed in $\{0, 1, \ldots, b-1\}$.

A Heuristic Approach

- To use the central limit theorem, we can consider a continuous model, representing $s_b(n)$ for n around N by a Gaussian random variable $S_{b,N}$ with expectation and variance given by

$$\mathbb{E}\left(S_{b,N}\right) = \frac{(b-1)\log N}{2\log b} \text{ and } \mathbb{V}\left(S_{b,N}\right) = \frac{(b^2-1)\log N}{12\log b}.$$

A Heuristic Approach

- To use the central limit theorem, we can consider a continuous model, representing $s_b(n)$ for n around N by a Gaussian random variable $S_{b,N}$ with expectation and variance given by

$$\mathbb{E}\left(S_{b,N}\right) = \frac{(b-1)\log N}{2\log b} \text{ and } \mathbb{V}\left(S_{b,N}\right) = \frac{(b^2-1)\log N}{12\log b}.$$

- In particular

$$\mathbb{E}\left(S_{2,N}\right) = \frac{\log N}{\log 4} \text{ and } \mathbb{E}\left(S_{3,N}\right) = \frac{\log N}{\log 3},$$

A Heuristic Approach

- To use the central limit theorem, we can consider a continuous model, representing $s_b(n)$ for n around N by a Gaussian random variable $S_{b,N}$ with expectation and variance given by

$$\mathbb{E}\left(S_{b,N}\right) = \frac{(b-1)\log N}{2\log b} \text{ and } \mathbb{V}\left(S_{b,N}\right) = \frac{(b^2-1)\log N}{12\log b}.$$

- In particular

$$\mathbb{E}\left(S_{2,N}\right) = \frac{\log N}{\log 4} \text{ and } \mathbb{E}\left(S_{3,N}\right) = \frac{\log N}{\log 3},$$

- Their standard deviations are of the order of magnitude $\sqrt{\log N}$.

A Heuristic Approach

- Let $u < \left(\frac{1}{\log 3} - \frac{1}{\log 4}\right)$.
- The
$$\text{Prob}\left\{\left|S_{2,N} - \mathbb{E}\left(S_{2,N}\right)\right| > u \log N\right\}$$
is asymptotically 0.

A Heuristic Approach

- Let $u < \left(\frac{1}{\log 3} - \frac{1}{\log 4}\right)$.
- The
$$\text{Prob}\left\{\left|S_{2,N} - \mathbb{E}\left(S_{2,N}\right)\right| > u \log N\right\}$$
is asymptotically 0.
- However for a given N, it is positive.

- Let $u < \left(\frac{1}{\log 3} - \frac{1}{\log 4}\right)$.
- The
$$\mathrm{Prob}\left\{\left|S_{2,N} - \mathbb{E}\left(S_{2,N}\right)\right| > u \log N\right\}$$
is asymptotically 0.
- However for a given N, it is positive.

Remark

- Since $\mathbb{V}(S_{3,N}) > \mathbb{V}(S_{2,N})$,

$$\begin{aligned} &\text{Prob}\left\{S_{3,N} \text{ is at a distance d from its mean}\right\} \\ >&\text{Prob}\left\{S_{2,N} \text{ is at a distance d from its mean}\right\} \end{aligned}$$

Remark

- Since $\mathbb{V}\left(S_{3,N}\right) > \mathbb{V}\left(S_{2,N}\right)$,

$$\begin{aligned}&\text{Prob}\left\{S_{3,N} \text{ is at a distance d from its mean}\right\}\\ >&\text{Prob}\left\{S_{2,N} \text{ is at a distance d from its mean}\right\}\end{aligned}$$

- Hence we expect to find some $u < \left(\frac{1}{\log 3} - \frac{1}{\log 4}\right)$ such that

$$\begin{aligned}&\text{Prob}\left\{\left|S_{2,N} - \mathbb{E}\left(S_{2,N}\right)\right| > u \log N\right\}\\ <&\text{Prob}\left\{S_{3,N} \text{ is close to } \mathbb{E}\left(S_{2,N}\right)\right\}\end{aligned}$$

Remark

- Since $\mathbb{V}\left(S_{3,N}\right) > \mathbb{V}\left(S_{2,N}\right)$,

$$\begin{aligned}&\text{Prob}\left\{S_{3,N} \text{ is at a distance d from its mean}\right\}\\ >&\text{Prob}\left\{S_{2,N} \text{ is at a distance d from its mean}\right\}\end{aligned}$$

- Hence we expect to find some $u < \left(\frac{1}{\log 3} - \frac{1}{\log 4}\right)$ such that

$$\begin{aligned}&\text{Prob}\left\{\left|S_{2,N} - \mathbb{E}\left(S_{2,N}\right)\right| > u \log N\right\}\\ <&\text{Prob}\left\{S_{3,N} \text{ is close to } \mathbb{E}\left(S_{2,N}\right)\right\}\end{aligned}$$

- For some ω,

$$\left|S_{3,N}(\omega) - S_{2,N}(\omega)\right| \leq u \log N.$$

Proposition 1.

Let $\lambda \in (0,1)$. *For any*

$$\nu > 1 - ((1-\lambda)\log(1-\lambda) + (1+\lambda)\log(1+\lambda)) \,/\, \log 4$$

and any sufficiently large integer H, we have

$$\#\{n < 2^{2H} : |s_2(n) - H| \geq \lambda H\} \leq 2^{2H\nu}.$$

- This proved by using

$$\#\left\{0 \leq n < 2^{2H} \;:\; s_2(n) = m\right\} = \binom{2H}{m}.$$

Proposition 2.

Let L be sufficiently large an integer. We have

$$\#\left\{n < 3^L : s_3(n) = \left\lfloor \frac{L \log 3}{\log 4} \right\rfloor\right\} \geq 3^{0.970359238L}.$$

- This proved by using

$$\#\left\{0 \leq n < 3^L \ : \ s_3(n) = m\right\} = \sum_{\substack{l_0+l_1+l_2=L \\ l_1+2l_2=m}} \frac{L!}{l_0!l_1!l_2!}.$$

- Let N be sufficiently large and put

$$K = \left\lfloor \frac{\log N}{\log 3} \right\rfloor - 2 \text{ and } H = \left\lfloor \frac{(K-1)\log 3}{\log 4} \right\rfloor + 2.$$

- Then

$$\frac{N}{81} \leq 3^{K-1} < 3^K < 2^{2H} \leq N.$$

Proof of Theorem 2

- Let N be sufficiently large and put

$$K = \left\lfloor \frac{\log N}{\log 3} \right\rfloor - 2 \text{ and } H = \left\lfloor \frac{(K-1)\log 3}{\log 4} \right\rfloor + 2.$$

- Then

$$\frac{N}{81} \leq 3^{K-1} < 3^K < 2^{2H} \leq N.$$

- Use Proposition 1 with $\lambda = 0.14572049 \log 4$ gives

(1)
$$\#\{n \leq 2^{2H} : |s_2(n) - H| \geq \lambda H\} \leq 2^{0.970359 \times 2H} \leq N^{0.970359}.$$

Proof of Theorem 2

- For $n \in [2 \cdot 3^{K-1}, 3^K)$ we have $s_3(n) = 2 + s_3(n - 2 \cdot 3^{K-1})$. Hence by Proposition 2,

$$\begin{aligned} &\#\{n \in [2 \cdot 3^{K-1}, 3^K) : s_3(n) = H\} \\ =&\#\{n < 3^{K-1}) : s_3(n) = H - 2\} \\ =&\#\{n < 3^{K-1}) : s_3(n) = \lfloor \frac{(K-1)\log 3}{\log 4} \rfloor\} \\ \geq& 3^{0.970359238(K-1)} \geq N^{0.970359237} \end{aligned}$$

which gives

(2) $$\#\{n \leq 2^{2H} : s_3(n) = H\} \geq N^{0.970359237}.$$

- The assertion follows by combining (1) and (2).

Remarks

- We believe that our result is the best possible using the methods used.

Remarks

- We believe that our result is the best possible using the methods used.
- It is possible to extend the ideas of our result and proof to extend to any pair of bases (b_1, b_2).

Questions

- We showed that there is a limit point of $\frac{s_3(n)-s_2(n)}{\log n}$ in an interval of length 0.146 around 0.
- We can ask what is the smallest interval around 0 for which $\frac{s_3(n)-s_2(n)}{\log n}$ has a limit point?

Questions

- We showed that there is a limit point of $\frac{s_3(n)-s_2(n)}{\log n}$ in an interval of length 0.146 around 0.
- We can ask what is the smallest interval around 0 for which $\frac{s_3(n)-s_2(n)}{\log n}$ has a limit point?
- Though we expect that it is true for any interval, we are not able to prove it.

Questions

- We showed that there is a limit point of $\frac{s_3(n)-s_2(n)}{\log n}$ in an interval of length 0.146 around 0.
- We can ask what is the smallest interval around 0 for which $\frac{s_3(n)-s_2(n)}{\log n}$ has a limit point?
- Though we expect that it is true for any interval, we are not able to prove it.
- One specific problem is to show that existence of a $\delta > 0$ such that there are infinitely many n with $\frac{s_3(n)-s_2(n)}{\log n} < -\delta$? And what is the largest possible δ?

Thank you for your attention