Sums of digits in different bases

Second Symposium on Analytic Number Theory
Cetraro, 8-12 July 2019
10/7/2019

Gérald Tenenbaum
Institut Élie Cartan
Université de Lorraine
BP 70239
54506 Vandœuvre-lès-Nancy Cedex
France

gerald.tenenbaum@univ-lorraine.fr

1. Introduction and history

$$q \geqslant 2, \qquad n = \sum_{j\geqslant 0} e_j(n) q^j, \qquad s_q(n) = \sum_{j\geqslant 0} e_j(n).$$

Let $\mathbb{P}_N$ be the uniform probability on $\{1, 2, \ldots, N\}$.

For $N := q^k - 1$, $e_j(n)$ are irv's with $\mathbb{P}_N(e_j = h) = 1/q \quad (0 \leqslant h < q)$.

$\therefore$ if $X_h := |\{0 \leqslant j < k : e_j = h\}|$

$\big(X_0, \ldots, X_{q-1}\big) \in [0, k]^q$ has multinomial distribution:

$$\mathbb{P}(X_0 = h_0, \ldots, X_{q-1} = h_{q-1}) = \frac{1}{q^k}\binom{k}{h_0, \ldots, h_{k-1}},$$

$$\mathbb{P}(s_q = m) = \sum_{h_1 + 2h_2 + \ldots + (q-1)h_{q-1} = m} \frac{1}{q^k}\binom{k}{h_0, \ldots, h_{q-1}},$$

$$\mathbb{E}\big(\mathrm{e}^{its_q}\big) = \frac{\mathrm{e}^{itk(q-1)/2}}{q^k}\left(\frac{\sin(tq/2)}{\sin(t/2)}\right)^k \approx \mathrm{e}^{\frac{1}{2}itk(q-1) - \frac{1}{8}t^2k(q^2-1)} \qquad (t \to 0)$$

$\mathcal{L}(s_q) \approx \mathcal{N}\big(\mu_{N,q}, \sigma_{N,q}\big)$, with $\mu_{N,q} := \frac{1}{2}k(q-1), \quad \sigma^2_{N,q} := \frac{1}{4}k(q^2-1).$

Conjecture : $(\log b)/\log a \in \mathbb{R} \setminus \mathbb{Q} \Rightarrow s_a \perp s_b$.

- ***Independence in congruence classes***

Gelfond (1967/8) conjectures : $(a, b) = 1$, $(r, a-1) = (s, b-1) = 1$,

$\mathbb{P}_N(s_a \equiv u \pmod r,\ s_b \equiv v \pmod s) = 1/rs + O\big(N^{-\delta}\big)$.

→ Proved by Kim (1999), improving a result of Bésineau (1972).

→ Berend-Kolesnik (2016) improve δ, and for $s_a(An+B)$, $s_b(Cn+D)$.

→ Dual result by Dartyge-T. (2005) : $s_q(An+B) \bmod r$, $s_q(Cn+D) \bmod s$, (uniformly in all parameters).

→ Further improvements and generalizations : Aloui (2019).

- ***Regular independence: various approaches***

- Drmota (2001) local laws:

$(a,b)=1,\ d:=(a-1,b-1)$,

$$\mathbb{P}_N\big(s_a(n)=k,\ s_b(n)=\ell\big)=d\,\frac{\mathrm{e}^{-\frac12(k-\mu_{N,a})^2/\sigma_{N,a}-\frac12(\ell-\mu_{N,b})^2/\sigma_{N,b}}}{2\pi\sigma_{N,a}\sigma_{N,b}}+O\Big(\frac{1}{\log N}\Big)$$

$\Rightarrow$ CLT.

- Madritsch-Stoll (2014):

$$\mathbb{E}_N\big(s_b/s_a\big)\sim\mu_{N,b}/\mu_{N,a}\to\tau_0(a,b):=(b-1)(\log a)/\{(a-1)\log b\}$$

& generalizations to polynomial and prime values

- Stewart (1980) :

$$(\log b)/\log a\in\mathbb{R}\smallsetminus\mathbb{Q}\Rightarrow s_a(n)+s_b(n)>(\log_2 n)/(\log_3 n+C)$$

2. Alternative approach: evaluate $\mathbb{P}_N(s_a \sim \tau s_b)$

Note: $\mathbb{P}_N(s_b/s_a \sim \tau_0(a,b)) = 1+o(1)$

$\to \tau_0 =$ trivial accumulation point.

Deshouillers, Habsieger, Laishram, Landreau (2017) :

$$\delta := 0.203, \ \exists c < 1 : \ \mathbb{P}_N(|1 - s_2/s_3| \leqslant \delta) \gg N^{-c}$$

$\tau_0(2,3) \approx 1.262$, $\tau_0 - 1 \approx 0.262 > \delta \Rightarrow \exists$ second accumulation point

Theorem. (La Bretèche, Stoll, T., 2019).

$(\log b)/\log a \in \mathbb{R} \smallsetminus \mathbb{Q}$, $\tau > 0$, $\tau \neq \tau_0(a,b)$:

$$N^{-c} \ll \mathbb{P}_N(|s_b/s_a - \tau| \leqslant 1/(\log N)^{\sigma}) \ll N^{-d}$$

with c, d, σ *explicit*, $0 < d < c < 1$.

Theorem. (La Bretèche, Stoll, T., 2019).
$(\log b)/\log a \in \mathbb{R} \smallsetminus \mathbb{Q}$, $\tau > 0$, $\tau \neq \tau_0(a,b)$:

$$N^{-c} \ll \mathbb{P}_N\big(|s_b/s_a - \tau| \leqslant 1/(\log N)^{\sigma}\big) \ll N^{-d}$$

with c, d, σ explicit, $0 < d < c < 1$.

$\sigma \leftrightarrow$ irr. exp. of $\vartheta := (\log a)/\log b$, i.e. $\inf\big\{\gamma : |\vartheta - r/q| \gg 1/q^{\gamma}\big\}$.

For $a = 2$, $b = 3$, $\gamma = 5.117$ (Wu & Wang, 2014).

For $\tau = 1$, we get $c \approx 0.05$, $d \approx 0.006$, $\sigma \approx 0.0013$.

Expected value of c, $d \approx 0.01849$ on independence assumption.

3. Sketch proofs

3·1. Upper bound

s_a and s_b both have a binomial distribution with distinct means

$$\mu_{N,a} = \frac{a-1}{2\log b}\log N, \; \mu_{N,b} = \frac{b-1}{2\log b}\log N$$

and variances $\asymp \log N$.

If $s_b(n) \sim \tau s_a(n)$, $\tau \neq \tau_0(a,b)$, then $s_a(n)$ or $s_b(n)$ must be far from its mean.

Since the distributions are concentrated around their means with standard deviations $\ll \sqrt{\log N} = o\big(\log N\big)$, the corresponding number of integers must be $\leqslant N^{1-d}$.

3·2. Lower bound

Based on Cassels' a construction (1959) of a set of >0 Hausdorff measure whose elements are normal in any base that is not a power of 3.

$a=2$, $b=3$, $\tau=1$, for simplicity.

$N:=3^k-1$. $\mathcal{M}_k:=\{m\leqslant N:(\forall j)\ e_j(m)=0\text{ or }2\}$.

For $\varrho\in]0,1[$, define $p_N(m):=\varrho^{s_3(m)/2}(1-\varrho)^{k-s_3(m)/2}$.

$$\sum_{m\in\mathcal{M}_k}p_N(m)=\sum_{1\leqslant j\leqslant k}\binom{k}{j}\varrho^j(1-\varrho)^{k-j}=1,\qquad E_N(s_3)=2\varrho k.$$

$$V_N(s_3)=4\varrho(1-\varrho)k,\qquad p_N\big(|s_3-2k\varrho|>T\sqrt{k}\big)\leqslant 4\varrho(1-\varrho)/T^2.$$

Select $\varrho:=(\log 3)/\log 4$: $2k\varrho=(\log N)/\log 4=\mu_{N,2}$.

$$\therefore\qquad\qquad p_N\big(s_3\sim\mu_{N,2}\big)=1+o(1).$$

Task: ***show that*** m ***has normal*** 2***-adic expansion with*** p_N***-prob.*** $1+o(1)$.

Indeed, this will imply $p_N(s_2\sim\mu_{N,2})\to 1$.

Let $\mathrm{e}(u) := \mathrm{e}^{2\pi i u}$ $(u \in \mathbb{R})$, $\langle u \rangle := u - \lfloor u \rfloor$.

Define $\sigma_h(m,n) := \dfrac{1}{n} \sum_{\nu \leqslant n} \mathrm{e}\big(hm/2^\nu\big) \qquad (m \in \mathcal{M}_k,\ n \geqslant 1,\ h \in \mathbb{Z})$,

$$\Delta_n(m) := \frac{1}{H+1} + \sum_{1 \leqslant h \leqslant H} \frac{|\sigma_h(m,n)|}{h}.$$

Erdős-Turán (+Rivat-T. 05): $\left| \dfrac{1}{n} \sum_{\substack{1 \leqslant \nu \leqslant n \\ \langle m/2^\nu \rangle \in I}} 1 - |I| \right| \leqslant \Delta_n(m) \quad (m \in \mathcal{M}_k,\ n \geqslant 1)$.

However, if $m = \sum_{r \geqslant 0} e_r(m) 2^r$, we have

$$e_{r-1}(m) = j \Leftrightarrow \langle m/2^r \rangle \in \big[\tfrac{1}{2}j, \tfrac{1}{2}(j+1)\big[\quad (m \geqslant 1,\ j = 0,1).$$

For $n := \lfloor (\log N)/\log 2 \rfloor \approx k(\log 3)/\log 2$, we get

$$s_2(m) = \frac{k \log 3}{\log 4} + O\big(k\Delta_n(m)\big) = \mu_{N,2} + O\big(k\Delta_n(m)\big).$$

Thus, we need $p_N(\Delta_n \to 0) \to 1$.

$$E_N\left(|\sigma_h(\cdot,n)|^2\right) = \sum_{m\in\mathcal{M}_k} r_k(m)|\sigma_h(m,n)|^2$$

$$= \frac{1}{n^2}\sum_{1\leqslant\mu,\nu\leqslant n}\sum_{m\in\mathcal{M}_k} r_k(m)\mathrm{e}\Big(hm\Big(\frac{1}{2^\nu}-\frac{1}{2^\mu}\Big)\Big).$$

Inner sum may be expanded using binomial formula.

It is $\leqslant \mathrm{e}^{-cS_h(\mu,\nu)}$ with

$$S_h(\mu,\nu) := \sum_{0\leqslant j<k} \left\|2h3^j\left(2^{-\nu}-2^{-\mu}\right)\right\|^2 \geqslant \sum_{L<\ell\leqslant L+R} \|3^\ell\alpha\|^2,$$

$\alpha = 2(1-1/2^\delta)3^{\langle\varphi\log h\rangle-\langle\nu\vartheta\rangle}$,

$\delta := \mu-\nu$, $\varphi := 1/\log 2$, $\vartheta := (\log 3)/\log 2$, $L := \lfloor\varphi\log h\rfloor$, $R \ll \sqrt{k}$

Need to show $S_h(\mu,\nu)$ large for most pairs (μ,ν).

$$S_h(\mu,\nu) \geqslant \sum_{L<\ell\leqslant L+R} \|3^\ell \alpha\|^2, \qquad \alpha = 2(1-1/2^\delta)3^{\langle \varphi \log h\rangle - \langle \nu\vartheta\rangle}$$

Expand $\alpha = \sum_{t\in\mathbb{Z}} \varepsilon_t(\alpha)/3^t$ and note : $\varepsilon_\ell(\alpha) = 1 \Rightarrow \|3^\ell\alpha\| \geqslant 1/3$.

$$\therefore \qquad S_h(\mu,\nu) \gg \sigma_3(\alpha) := \sum_{\substack{L<\ell\leqslant L+R\\ \varepsilon_\ell(\alpha)=1}} 1$$

Select $s < R/2$: we have $\mathcal{A} := \{\alpha : \sigma_3(\alpha) \leqslant s\} \subset [0,6]$, and

$\sigma_3(\alpha) = j \leqslant s \Rightarrow \alpha \in$ union of $\ll 3^L\binom{R}{j}2^{R-j}$ intervals of length $\leqslant 3^{-L-R}$.

$\Rightarrow$ $\operatorname{meas}\mathcal{A} \ll \mathrm{e}^{-cs}$ $(s = R/6)$.

Second application of Erdős-Turán : $\sum_{\substack{\nu\leqslant n\\ \sigma_3(\alpha)\leqslant s}} 1 \leqslant n\mathrm{e}^{-cs} + nD_n 3^{R+L}\mathrm{e}^{-cs}$

where $D_n :=$ discrepancy of $\langle\vartheta\nu\rangle$ $(\nu \leqslant n)$.

If $|\vartheta - r/q| \gg 1/q^\gamma$, we have (Dirichlet) $D_n \ll n^{-1/\gamma}$.

Optimizing parameters we get $S_h(\mu,\nu) \geqslant s/9$ for most (μ,ν), and so

$s_2(m) \sim \mu_{N,2}$ for p_N-almost all m.

Grazie a tutti !