Block occurrences in the binary expansion of n and $n + t$

Bartosz Sobolewski
(joint work with Lukas Spiegelhofer)

Jagiellonian University in Kraków
Montanuniversität Leoben

One World Numeration Seminar
February 13, 2024

Cusick's conjecture

Binary expansion of $n \in \mathbb{N} = \{0, 1, \ldots\}$:

$$n = \sum_{j \geq 0} \delta_j 2^j, \qquad \delta_j \in \{0, 1\}.$$

Sum of binary digits:

$$\mathrm{s}(n) = \sum_{j \geq 0} \delta_j.$$

Conjecture (Cusick, 2012)

For each $t \in \mathbb{N}$ the natural density

$$\begin{aligned} c_t &= \mathrm{dens}\{n \in \mathbb{N} : \mathrm{s}(n+t) \geq \mathrm{s}(n)\} \\ &= \lim_{N \to \infty} \frac{1}{N} \#\{0 \leq n < N : \mathrm{s}(n+t) \geq \mathrm{s}(n)\} \end{aligned}$$

satisfies

$$c_t > \frac{1}{2}.$$

Connections to other problems

The conjecture is related to:

- the **Tu–Deng** conjecture (2011);
- the distribution of the 2-adic valuation of binomial coefficients:

$$\mathsf{s}(n+t) - \mathsf{s}(n) = \mathsf{s}(t) - \nu_2\left(\binom{n+t}{t}\right);$$

- hyperbinary expansions.

More details in a paper of **Drmota, Kauers, Spiegelhofer** (2016).

Some results

- **Drmota, Kauers, Spiegelhofer** (2016):
  - The conjecture holds for $t < 2^{30}$;
  - $1/2 < c_t < 1/2 + \varepsilon$ for t in a set of density 1;
- **Emme, Hubert** (2019):
  Central-limit type result on the distribution of $\mathrm{s}(n+t) - \mathrm{s}(n)$ for t random;
- **Spiegelhofer** (2022):
  $c_t > 1/2 - \varepsilon$ for all t having sufficiently many occurrences of 01 in their binary expansion.

Some results

Theorem (Spiegelhofer, Wallner (2023))

We have $c_t > 1/2$ for all t having sufficiently many occurrences of 01 in their binary expansion.

For $j \in \mathbb{Z}$ let $\sigma(t,j) = \mathrm{dens}\{n \in \mathbb{N} : \mathrm{s}(n+t) - \mathrm{s}(n) = j\}$.

Theorem (Spiegelhofer, Wallner (2023))

For $t \geq 1$ let us define

$$\kappa(1) = 2, \qquad \kappa(2t) = \kappa(t), \qquad \kappa(2t+1) = \frac{\kappa(t) + \kappa(t+1)}{2} + 1.$$

Then for all t having N occurrences of 01 in their binary expansion we have

$$\sigma(t,j) = \frac{1}{\sqrt{2\pi\kappa(t)}} \exp\left(-\frac{j^2}{2\kappa(t)}\right) + \mathcal{O}(N^{-1}(\log N)^4).$$

Further research

Cusick: "the hard cases remain".

Some possible generalizations:

- Consider expansions in any base b;
- Multidimensional analogues;
- Count occurrences of other blocks.

Question

Can we prove similar results for the function r, counting the occurrences of 11 in the binary expansion?

$$n = \sum_{j \geq 0} \delta_j 2^j \implies \mathrm{r}(n) = \sum_{j \geq 0} \delta_j \delta_{j+1}.$$

The distribution of $\mathrm{r}(n+t)-\mathrm{r}(n)$

For $k \in \mathbb{Z}$ we let

$$c_t(k) = \mathrm{dens}\{n \in \mathbb{N} : \mathrm{r}(n+t) - \mathrm{r}(n) = k\}.$$

Theorem (S., Spiegelhofer (2024+))

For $t \in \mathbb{N}$ let $v_0 = 0$, $v_1 = 3/2$, and

$$\begin{aligned} v_{4t} &= v_{2t}, \\ v_{4t+2} &= v_{2t+1} + 1, \\ v_{2t+1} &= \frac{v_t + v_{t+1}}{2} + \frac{3}{4}. \end{aligned}$$

There exist effective absolute constants C, N_0 such that for any $k \in \mathbb{Z}$ we have

$$\left| c_t(k) - (2\pi v_t)^{-1/2} \exp\left(-\frac{k^2}{2v_t}\right) \right| \leq C \frac{(\log N)^2}{N},$$

for any $t \in \mathbb{N}$ having $N \geq N_0$ occurrences of 01 *in its binary expansion.*

Some remarks

- The main term
$$(2\pi v_t)^{-1/2} \exp\left(-\frac{k^2}{2v_t}\right)$$
dominates the error term $N^{-1}(\log N)^2$ when $|k| < \delta\frac{\sqrt{3}}{2}\sqrt{N \log N}$, where $\delta \in (0,1)$
- We have
$$\sum_{k\geq 0} c_t(k) \geq 1/2 - C_1 N^{-1/2}(\log N)^3,$$
for some effective absosute constant C_1.
- For $t < 2^{20}$ we have
$$\sum_{k\geq 0} c_t(k) > \frac{1}{2}.$$

Sketch of the proof

- Put $\mathrm{e}(\vartheta) = \exp(i\vartheta)$ and consider the characteristic functions

$$\gamma_t(\vartheta) = \sum_{k\in\mathbb{Z}} c_t(k)\,\mathrm{e}(k\vartheta),$$

  so that

$$c_t(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \gamma_t(\vartheta)\,\mathrm{e}(-k\vartheta)\,\mathrm{d}\vartheta.$$

- Main term comes from the approximation of the integral over a suitable subinterval $[-\vartheta_0, \vartheta_0]$ by replacing γ_t with Gaussian characteristic functions

$$\gamma_t^*(\vartheta) = \exp\left(-\frac{v_t}{2}\vartheta^2\right),$$

  where v_t is the variance of $k \mapsto c_t(k)$ (mean $= 0$).
- Bound the tails of the integral from above.

Main ingredients

Let N be the number of occurrences of 01 in the binary expansion of t. We will use the following facts:

1. There exists an absolute constant K such that for $|\vartheta| \leq \pi$ we have
$$|\gamma_t(\vartheta) - \gamma_t^*(\vartheta)| \leq KN|\vartheta|^3.$$
2. For $|\vartheta| \leq \pi$ we have
$$|\gamma_t(\vartheta)| \leq \left(1 - \frac{1}{128}\vartheta^2\right)^{\lfloor N/2 \rfloor}.$$
3. For all $t \in \mathbb{N}$ we have
$$\frac{3}{4}N \leq v_t \leq 5N.$$

Basic properties and recurrences

Initial goals:

- show that $c_t(k)$ exist;
- obtain a recurrence relation for γ_t.

Put

$$d(t,n) = \mathrm{r}(n+t) - \mathrm{r}(n),$$
$$C_t(k) = \{n \in \mathbb{N} : d(t,n) = k\}.$$

We partition $C_t(k)$ depending on the parity of n:

$$A_t(k) = \{n \in \mathbb{N} : d(t,2n) = k\}, \qquad B_t(k) = \{n \in \mathbb{N} : d(t,2n+1) = k\},$$

so that

$$C_t(k) = 2A_t(k) \cup (2B_t(k)+1).$$

Basic properties and recurrences

Starting from $\mathrm{r}(0) = 0$ and the recurrence relations

$$\mathrm{r}(2n) = \mathrm{r}(n), \qquad \mathrm{r}(4n+1) = \mathrm{r}(n), \qquad \mathrm{r}(4n+3) = \mathrm{r}(2n+1) + 1,$$

we get relations for $d(t, n)$, and finally for $A_t(k), B_t(k)$:

$$A_{4t+j}(k) = 2A_{2t+\ell}(k+\sigma) \cup (2B_{2t+\ell}(k+\rho) + 1)$$

for some $\ell \in \{0, 1\}$, $\sigma, \rho \in \{-1, 0, 1\}$ (similarly for $B_{4t+j}(k)$).

Proposition

For all $t \in \mathbb{N}$ and $k \in \mathbb{Z}$ the sets $A_t(k), B_t(k)$ are finite unions of arithmetic progressions of the form $2^q\mathbb{N} + p$, where $0 \leq p < 2^q$. Therefore, their densities (and thus also $c_t(k)$) exist and satisfy a set of recurrence relations.

Characteristic functions

Define the densities

$$a_t(k) = \operatorname{dens} A_t(k), \qquad b_t(k) = \operatorname{dens} B_t(k),$$

so that

$$c_t(k) = \frac{a_t(k) + b_t(k)}{2}$$

Define characteristic functions of the probability distributions $k \mapsto a_t(k)$ and $k \mapsto b_t$:

$$\alpha_t(\vartheta) = \sum_{k \in \mathbb{Z}} a_t(k) \operatorname{e}(k\vartheta),$$
$$\beta_t(\vartheta) = \sum_{k \in \mathbb{Z}} b_t(k) \operatorname{e}(k\vartheta).$$

We have

$$\gamma_t(\vartheta) = \frac{\alpha_t(\vartheta) + \beta_t(\vartheta)}{2}.$$

Characteristic functions

The functions α_t, β_t satisfy the relations

$$\alpha_{2t} = \frac{1}{2}(\alpha_t + \beta_t), \qquad \beta_{4t} = \frac{1}{2}(\alpha_{2t} + \beta_{2t}),$$

$$\alpha_{2t+1} = \frac{1}{2}(\alpha_t + \mathrm{e}(\vartheta)\beta_t), \qquad \beta_{4t+2} = \frac{1}{2}(\mathrm{e}(\vartheta)\alpha_{2t+1} + \mathrm{e}(-\vartheta)\beta_{2t+1}),$$

$$\beta_{2t+1} = \frac{1}{2}(\alpha_{t+1} + \mathrm{e}(-\vartheta)\beta_{t+1}).$$

We can rewrite them using column vectors

$$S_t(\vartheta) = \begin{pmatrix} \alpha_{2t}(\vartheta) & \beta_{2t}(\vartheta) & \alpha_{2t+1}(\vartheta) & \beta_{2t+1}(\vartheta) & \alpha_{2t+2}(\vartheta) & \beta_{2t+2}(\vartheta) \end{pmatrix}^T,$$

and 6×6 coefficient matrices $D_0(\vartheta), D_1(\vartheta)$.

Proposition

For all $t \in \mathbb{N}$ we have the recurrence relations

$$S_{2t}(\vartheta) = D_0(\vartheta)S_t(\vartheta),$$
$$S_{2t+1}(\vartheta) = D_1(\vartheta)S_t(\vartheta).$$

Mean and variance

Let $m_t, m_t^\alpha, m_t^\beta$ denote the mean and $v_t, v_t^\alpha, v_t^\beta$ the variance of $\gamma_t, \alpha_t, \beta_t$.

Proposition

We have

$$m_{2t}^\alpha = m_{2t}^\beta = 0, \qquad m_{2t}^\alpha = -m_{2t}^\beta = 1/2,$$

and

$$v_{2t}^\alpha = \frac{1}{2}(v_t^\alpha + v_t^\beta) + q_{2t}^\alpha, \qquad v_{2t}^\beta = \frac{1}{2}(v_t^\alpha + v_t^\beta) + q_{2t}^\beta,$$
$$v_{2t+1}^\alpha = \frac{1}{2}(v_t^\alpha + v_t^\beta) + q_{2t+1}^\alpha, \quad v_{2t+1}^\beta = \frac{1}{2}(v_{t+1}^\alpha + v_{t+1}^\beta) + q_{2t+1}^\beta,$$

where q_t^α, q_t^β repeat with period 4.

Corollary

We have $m_t = 0$ and

$$v_{4t} = v_{2t}, \qquad v_{4t+2} = v_{2t+1} + 1, \qquad v_{2t+1} = \frac{1}{2}(v_t + v_{t+1}) + \frac{3}{4}.$$

1st ingredient: Gaussian approximation

Recall: we want to approximate γ_t using

$$\gamma_t^*(\vartheta) = \exp\left(-\frac{v_t}{2}\vartheta^2\right).$$

First, approximate α_t, β_t by

$$\alpha_t^*(\vartheta) := \exp\left(m_t^\alpha i\vartheta - \frac{1}{2}v_t^\alpha\vartheta^2\right), \qquad \beta_t^*(\vartheta) := \exp\left(m_t^\beta i\vartheta - \frac{1}{2}v_t^\beta\vartheta^2\right).$$

Define the vector of approximations

$$S_t^*(\vartheta) = \begin{pmatrix}\alpha_{2t}^*(\vartheta) & \beta_{2t}^*(\vartheta) & \alpha_{2t+1}^*(\vartheta) & \beta_{2t+1}^*(\vartheta) & \alpha_{2t+2}^*(\vartheta) & \beta_{2t+2}^*(\vartheta)\end{pmatrix}^T$$

By definition,

$$S_t(\vartheta) - S_t^*(\vartheta) = O(\vartheta^3),$$

where the implied constant depends on t.

1st ingredient: Gaussian approximation

We show that for any $k \in \mathbb{N}$ and $|\vartheta| \leq \pi$ we have

$$\begin{aligned} \|S_{2^k t} - S^*_{2^k t}\|_\infty &\leq \|D_0^k(S_t - S^*_t)\|_\infty + \|D_0^k S^*_t - S^*_{2^k t}\|_\infty \\ &\leq \|S_t - S^*_t\|_\infty + K|\vartheta|^3, \\ \|S_{2^k t + 2^k - 1} - S^*_{2^k t + 2^k - 1}\|_\infty &\leq \|S_t - S^*_t\|_\infty + K|\vartheta|^3, \end{aligned}$$

for some absolute constant K.
By induction on the number N of occurrences of 01 in the binary expansion of t:

$$\|S_t - S^*_t\|_\infty \leq \|S_0 - S^*_0\|_\infty + 2KN|\vartheta|^3 \leq K'N|\vartheta|^3.$$

Finally,

$$|\gamma_t - \gamma^*_t| \leq \frac{1}{2}|\alpha_t + \beta_t - \alpha^*_t - \beta^*_t| + \left|\frac{\alpha^*_t + \beta^*_t}{2} - \gamma^*_t\right| \leq \left(K' + K''\right) N|\vartheta|^3.$$

2nd ingredient: an upper bound on γ_t

Let t have binary expansion $\delta_n\delta_{n-1}\cdots\delta_1\delta_0$ and let N be the number of occurrences of 01. We want to bound $\|\cdot\|_\infty$ of

$$S_t = D_{\delta_0} D_{\delta_1} \cdots D_{\delta_{n-1}} D_{\delta_n} S_0.$$

Observe the following:

- $\|D_0\|_\infty = \|D_1\|_\infty = 1$;
- there occur at least $\lfloor N/2 \rfloor$ nonoverlapping strings from the set $\{0001, 0101, 1001, 1101\}$ in the expansion;
- each row of $D_1 D_0 D_0 D_0, D_1 D_0 D_1 D_0, D_1 D_0 D_0 D_1, D_1 D_0 D_1 D_1$ has an entry containing $(\mathrm{e}(k\vartheta) + \mathrm{e}((k+1)\vartheta))/16$.

2nd ingredient: an upper bound on γ_t

We have

$$|\,\mathrm{e}(k\vartheta)+\mathrm{e}((k+1)\vartheta)| = |1+\exp(i\vartheta)| = \sqrt{2(1+\cos\vartheta)} = 2\left|\cos\frac{\vartheta}{2}\right| \leq 2-\frac{\vartheta^2}{8}.$$

The norm of each row of $D_1D_0D_0D_0, D_1D_0D_1D_0, D_1D_0D_0D_1, D_1D_0D_1D_1$ is at most

$$\frac{1}{16}\left(16-\frac{\vartheta^2}{8}\right) = 1-\frac{\vartheta^2}{128}.$$

By submultiplicativity

$$\|S_t\|_\infty \leq \left(1-\frac{\vartheta^2}{128}\right)^{\lfloor N/2\rfloor}\|S_0\|_\infty = \left(1-\frac{\vartheta^2}{128}\right)^{\lfloor N/2\rfloor}.$$

3rd ingredient: bounds on the variance v_t

Recall: $v_0 = 0$, $v_1 = 3/2$, and

$$v_{4t} = v_{2t}, \qquad v_{4t+2} = v_{2t+1} + 1, \qquad v_{2t+1} = \frac{v_t + v_{t+1}}{2} + \frac{3}{4}.$$

We want to show:

$$\frac{3}{4}N \leq v_t \leq 5N.$$

First, we claim that $|v_{t+1} - v_t| \leq 3/2$. If $w_t = v_{t+1} - v_t$, then

$$w_{4t} = \frac{1}{2}w_{2t} + \frac{3}{4}, \qquad w_{4t+1} = \frac{1}{2}w_{2t} + \frac{1}{4},$$
$$w_{4t+2} = \frac{1}{2}w_{2t+1} - \frac{1}{4}, \qquad w_{4t+3} = \frac{1}{2}w_{2t+1} - \frac{3}{4},$$

and our claim follows by induction

3rd ingredient: bounds on the variance v_t

$$v_{4t} = v_{2t}, \qquad v_{4t+2} = v_{2t+1} + 1, \qquad v_{2t+1} = \frac{v_t + v_{t+1}}{2} + \frac{3}{4}.$$

Lower bound:

$$\begin{aligned} v_{2t} - v_t &\geq 0, \\ v_{2t+1} - v_t &= \frac{1}{2}(v_{t+1} - v_t) + \frac{3}{4} \geq 0, \\ v_{4t+1} - v_t &= \frac{1}{2}(v_{2t} - v_t) + \frac{1}{2}(v_{2t+1} - v_t) + \frac{3}{4} \geq \frac{3}{4}. \end{aligned}$$

Adding a digit to t does not decrease v_t; adding 01 increases v_t by $\geq 3/4$.

Upper bound:

$$\begin{aligned} |v_{2^k t} - v_t| &= |v_{2t} - v_t| \leq 1, \\ |v_{2^k t + 2^k - 1} - v_t| &\leq |v_{2^k(t+1)-1} - v_{2^k(t+1)}| + |v_{2^k(t+1)} - v_{t+1}| + |v_{t+1} - v_t| \\ &\leq \frac{3}{2} + 1 + \frac{3}{2} = 4. \end{aligned}$$

Adding $0^{k_1}1^{k_2}$ increases v_t by ≤ 5.

Generalization to any block

Let w be a block of binary digits of length $|w| = \ell$. Define $|n|_w$, to be the number the occurrences of w in the binary expansion of n. In particular,

$$\mathrm{s}(n) = |n|_1, \qquad \mathrm{r}(n) = |n|_{11}.$$

Work in progress: prove that the distribution of $|n + t|_w - |n|_w$ is approximately Gaussian when the binary expansion of t has many occurrences of 01.

Some remarks on the generalization

- The result should follow from the same main ingredients;
- If $w \in \{1^\ell, 0^\ell\}$, then the sets

$$C_t(k) = \{n \in \mathbb{N} : |n+t|_w - |n|_w = k\}$$

  are finite unions of arithmetic progressions of the form $2^q\mathbb{N} + p$ and are nonempty for infinitely many k;
  If $w \notin \{1^\ell, 0^\ell\}$, only finitely many $C_t(k)$ are nonempty but the unions may be infinite;
- To get recurrence relations for $c_t(k) = \operatorname{dens} C_t(k)$ and associated characteristic functions γ_t, we condition on the last $\ell - 1$ digits;
- The mean of γ_t is 0, while the conditional means are periodic in t and bounded by $1 - 1/2^{\ell-1}$.

Some remarks on the generalization

- The (unconditional) variance v_t satisfies

$$v_{2t} = v_t + q_{2t}, \qquad v_{2t+1} = \frac{1}{2}(v_t + v_{t+1}) + q_{2t+1},$$

where q_t are periodic in t (we had $q_{4t} = 0, q_{4t+2} = 1, q_{2t+1} = 3/4$ for $w = 11$).
When $w \in \{0^{\ell-1}1, 1^{\ell-1}0\}$, some q_t are negative!

- It seems that v_t grows the fastest with $|t|_{01}$ for $w \in \{0^\ell, 1^\ell\}$ (approximation error is the least), and the slowest for $w \in \{10^{\ell-1}, 01^{\ell-1}\}$.
- Obtaining an upper bound on γ_t requires analyzing long products of $2^{\ell-1} \times 2^{\ell-1}$ coefficient matrices.

Thank you for your attention!