Consecutive–Sum Representations: A Random–Block Construction with Deterministic Repair

Przemyslaw Chojecki

February 11, 2026

Abstract

Let $A = \{a_1 < a_2 < \cdots\} \subset \mathbb{N}$ and let

$$f(n) := \#\Big\{(u,v) :\ 1 \le u \le v,\quad n = \sum_{i=u}^{v} a_i\Big\}.$$

We give a probabilistic construction of an infinite increasing A for which

$$f(n)\ \ge\ c\log n \qquad \text{for all sufficiently large } n,$$

in particular $f(n) \to \infty$ as $n \to \infty$ (Erdős Problem 358) The construction is "red blocks" (random) plus "blue blocks" (deterministic repairs). A key technical ingredient is a single–block lower bound (Proposition 1) proved via local limit theorems for linear forms in independent geometric variables (standard results; we cite references and include the reduction).

1 Preliminaries

Lemma 1 (At most one representation per length)**.** *Fix a length $L \ge 1$. Then the function*

$$u \ \longmapsto\ \sum_{i=u}^{u+L-1} a_i$$

is strictly increasing in u (because $a_{u+L} - a_u > 0$), hence each n has at most one *representation of length L.*

Thus, to ensure $f(n) \ge K$ we must produce representations of *distinct lengths.*

2 The construction

Fix a parameter $\alpha \in (1/2, 1)$; for concreteness one may take $\alpha = 3/4$. We partition $\mathbb{N}$ into contiguous blocks $I_1, I_2, \ldots$ defined by:

- $x_1 := 2$.
- Given x_k, set $N_k := \lfloor x_k^{\alpha} \rfloor$ and define

$$I_k := [x_k,\ x_k + N_k) \cap \mathbb{N}, \qquad x_{k+1} := x_k + N_k.$$

So $|I_k| = N_k \asymp x_k^{\alpha}$ and consecutive blocks have comparable sizes.

Colors (deterministic alternation)

We color blocks in an alternating pattern:

$$I_k \text{ is } \textit{red} \text{ if } k \text{ is odd, and } \textit{blue} \text{ if } k \text{ is even.}$$

(Any fixed positive density pattern works; alternation keeps the bookkeeping simple and ensures $\asymp 1/2$ of integers lie in blue blocks.)

Definition of A

1. **Red blocks:** If I_k is red, include each integer of I_k in A independently with probability $1/2$. Write $R_k \subseteq I_k$ for the resulting random set.
2. **Blue blocks:** If I_k is blue, initially include nothing. After the red randomness is realized, we will deterministically choose $U_k \subseteq I_k$ to repair the remaining exceptional integers.

Finally set

$$A := \Big(\bigcup_{k \text{ odd}} R_k \Big) \cup \Big(\bigcup_{k \text{ even}} U_k \Big),$$

listed in increasing order as $\{a_1 < a_2 < \cdots\}$.

Safe red representations

A representation $n = \sum_{i=u}^{v} a_i$ is called *red–internal* if $a_u, \dots, a_v$ all lie in a *single* red block. Let $f_{\text{red}}(n)$ be the number of red–internal representations of n. Since blue blocks are disjoint value intervals, adding elements inside blue blocks never inserts between two elements of the same red block, hence every red–internal representation survives in the final A:

$$f(n) \;\geq\; f_{\text{red}}(n) \qquad \text{for all } n. \tag{1}$$

3 Red analysis: logarithmically many representations for most n

For a block $I = [x, x+N)$ (here $N = \lfloor x^\alpha \rfloor$) define the indicator

$$\mathbf{1}_I(n) = \begin{cases} 1, & \text{if } n \text{ has at least one red–internal representation using only elements of } A \cap I, \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$f_{\text{red}}(n) \;\geq\; \sum_{I_k \text{ red}} \mathbf{1}_{I_k}(n),$$

and the indicators $\mathbf{1}_{I_k}(n)$ are independent over distinct red blocks.

The next proposition is the crucial single–block input.

Proposition 1 (Single–block success)**.** *There exist absolute constants $c_0 > 0$ and x_0 (depending only on α) such that the following holds. Let $I = [x, x+N)$ with $x \geq x_0$ and $N = \lfloor x^\alpha \rfloor$. Condition on I being red, and include each integer of I independently with probability* $1/2$. *Then for every integer n in the* bulk range

$$2x + 1 \;\leq\; n \;\leq\; \frac{1}{10} x N,$$

we have

$$\mathbb{P}\big(\mathbf{1}_I(n) = 1\big) \;\geq\; c_0 \frac{N}{x} \;=\; c_0\, x^{\alpha - 1}.$$

Remark 1 (Why the lower endpoint is $2x + 1$). A red–internal representation uses at least two distinct selected values, so the smallest possible sum inside $[x, x + N)$ is $(x) + (x + 1) = 2x + 1$. Thus no uniform lower bound can hold at $n = 2x$.

Remark 2 (Why $\alpha > 1/2$ matters). If $\alpha = 1/2$ (block length $\asymp \sqrt{x}$), then $N^2 \asymp x$ and the set of possible sums inside a block occupies only a bounded fraction of the full range $[2x,\, xN]$, and the "constant per dyadic scale" heuristic breaks at the upper scales. Taking $\alpha > 1/2$ ensures $N^2 \gg x$, which is exactly what makes Proposition 1 uniform for all n in the stated bulk range.

3.1 Many contributing red blocks at many scales

Fix a large n. We will count red blocks with starts x in the scale window

$$\mathcal{X}(n) := \Big[C_1\, n^{1/(1+\alpha)},\ C_2\, n^{1/(2-\alpha)}\Big],$$

for suitable constants $C_1, C_2 > 0$. For $x \in \mathcal{X}(n)$ and $N = \lfloor x^\alpha \rfloor$ we have:

- $n \geq 2x + 1$ for all large n (since $x \leq C_2 n^{1/(2-\alpha)} = o(n)$),
- $n \leq \frac{1}{10} xN$ provided $x \geq C_1 n^{1/(1+\alpha)}$ with C_1 large enough (because $xN \asymp x^{1+\alpha}$).

Hence Proposition 1 applies to all such blocks.

Let

$$Z(n) := \sum_{\substack{I_k \text{ red} \\ x_k \in \mathcal{X}(n)}} \mathbf{1}_{I_k}(n).$$

Then $Z(n)$ is a sum of independent Bernoulli random variables. Moreover, in a dyadic subrange $x_k \in [X, 2X] \subseteq \mathcal{X}(n)$, the number of blocks is $\asymp X/N \asymp X^{1-\alpha}$ and each has success probability $\gg N/X \asymp X^{\alpha-1}$, so the mean contribution from each dyadic scale is $\asymp 1$. Since $\mathcal{X}(n)$ spans $\asymp \log n$ dyadic scales (because $\alpha > 1/2$ implies $n^{1/(2-\alpha)}/n^{1/(1+\alpha)} = n^{\Omega(1)}$), we obtain:

Lemma 2 (Logarithmic mean)**.** *There exist constants $c_1, c_2 > 0$ such that for all sufficiently large n,*

$$\mathbb{E}Z(n) \ \geq\ c_1 \log n.$$

A Chernoff bound for Poisson binomials gives:

Lemma 3 (Polynomially small lower tail)**.** *There exist constants $c_3, \gamma > 0$ such that for all sufficiently large n,*

$$\mathbb{P}\Big(Z(n) \leq \frac{c_1}{2} \log n\Big) \ \leq\ n^{-\gamma}.$$

Consequently,

$$\mathbb{P}\Big(f_{\text{red}}(n) \ \geq\ Z(n) \ \geq\ c_3 \log n\Big) \ \geq\ 1 - n^{-\gamma}.$$

Remark 3 (One-line justification of Lemma 3). For a sum Z of independent Bernoullis with mean μ, one has $\mathbb{P}(Z \leq (1-\varepsilon)\mu) \leq \exp(-\varepsilon^2 \mu/2)$; here $\mu \asymp \log n$ and $\varepsilon = 1/2$.

4 Bad sets in dyadic windows

For dyadic $N = 2^k$ define the bad set

$$E_N^{\mathrm{red}} := \Big\{ n \in [N, 2N] : \ f_{\mathrm{red}}(n) < \frac{c_3}{2} \log N \Big\}.$$

By Lemma 3, for each $n \in [N, 2N]$ we have $\mathbb{P}(n \in E_N^{\mathrm{red}}) \le N^{-\gamma}$ for large N, hence

$$\mathbb{E}|E_N^{\mathrm{red}}| \ \le \ N \cdot N^{-\gamma} \ = \ N^{1-\gamma}.$$

Fix $\eta \in (0, \gamma/2)$. By Markov,

$$\mathbb{P}\big(|E_N^{\mathrm{red}}| > N^{1-\eta}\big) \ \le \ N^{-(\gamma-\eta)}.$$

Summing over dyadic N gives a convergent series, so by Borel–Cantelli:

Lemma 4 (Almost sure sparsity of red-bad sets)**.** *With probability* 1 *(over the red randomness), there exists* N_0 *such that for all dyadic* $N \ge N_0$*,*

$$|E_N^{\mathrm{red}}| \ \le \ N^{1-\eta}.$$

Fix a realization of the red randomness satisfying Lemma 4. From now on, all statements are deterministic.

5 Blue repair (deterministic) and completion of the proof

5.1 A deterministic subset–sum lemma

Lemma 5 (Contiguous m–term sums)**.** *Let* $J = [s, s+L) \cap \mathbb{Z}$ *be an interval of* L *consecutive integers and let* $1 \le m \le L$*. Then the set of sums of* m *distinct elements of* J *is exactly the full integer interval*

$$\Big[ms + \frac{m(m-1)}{2}, \quad m(s+L-1) - \frac{m(m-1)}{2} \Big].$$

Proof. Translate so $s = 0$. An m–subset of $\{0, 1, \ldots, L-1\}$ can be written uniquely as

$$\{0 + e_1,\ 1 + e_2,\ \ldots,\ (m-1) + e_m\}$$

with $0 \le e_1 \le \cdots \le e_m \le L - m$. Its sum equals $\frac{m(m-1)}{2} + (e_1 + \cdots + e_m)$. Given any $D \in [0, m(L-m)]$, write $D = qm + r$ with $0 \le r < m$ and $0 \le q \le L - m$, and set

$$e_1 = \cdots = e_{m-r} = q, \qquad e_{m-r+1} = \cdots = e_m = q + 1.$$

Then $\sum e_i = D$. Undo the translation. □

5.2 Repair bands and parameters

Let $N = 2^k$ (dyadic, large). Choose a small constant $\delta > 0$ with

$$2\delta < \eta \qquad \text{and} \qquad \delta < \frac{\alpha}{1+\alpha}. \tag{2}$$

(For example, take $\delta := \min\{\eta/4,\ \alpha/(2+2\alpha)\}$.) Define the *repair band* for window $[N, 2N]$ by

$$B_N := \Big[N^{1-\delta},\ (2N)^{1-\delta}\Big) \cap \mathbb{N}.$$

For dyadic N, these bands are disjoint (up to endpoints), because $(2N)^{1-\delta}$ is exactly the left endpoint of B_{2N}.

We will create repairs for $n \in [N, 2N]$ using only blue blocks inside B_N. Since blocks alternate and lengths vary slowly, there is a constant $c_{\text{blue}} > 0$ such that for all large dyadic N, the number of blue integers in B_N is at least $c_{\text{blue}}|B_N| \asymp c_{\text{blue}} N^{1-\delta}$.

5.3 Which n do we actually repair?

Let A_{pre} denote the set consisting of *all red elements* together with *all blue repair elements already chosen for smaller dyadic windows.* Let $f_{\text{pre}}(n)$ be the representation count with respect to A_{pre}.

Define the *current bad set*

$$E_N := \Big\{n \in [N, 2N] :\ f_{\text{pre}}(n) < \tfrac{c_3}{2}\log N\Big\}.$$

Since $f_{\text{pre}}(n) \geq f_{\text{red}}(n)$, we have $E_N \subseteq E_N^{\text{red}}$, hence by Lemma 4: for all large dyadic N,

$$|E_N| \ \leq\ |E_N^{\text{red}}| \ \leq\ N^{1-\eta}. \tag{3}$$

Thus we only spend blue "budget" on integers that are genuinely under-represented at the time we process $[N, 2N]$.

5.4 A one-gadget hitting lemma (now fully explicit)

Lemma 6 (A universal valid start for $L = 4m$)**.** *Fix integers $n \geq 1$ and $m \geq 2$. Let*

$$L := 4m, \qquad s_0 := \Big\lfloor \frac{n}{m} \Big\rfloor - 2m.$$

Then n lies in the m–term sum interval of $[s_0, s_0+L)$, i.e. there exist m distinct integers $t_1, \ldots, t_m \in [s_0, s_0+L)$ with $t_1 + \cdots + t_m = n$.

Proof. By Lemma 5, the set of m–term sums from $[s, s+L)$ is the full interval

$$\Big[ms + \tfrac{m(m-1)}{2},\ \ m(s+L-1) - \tfrac{m(m-1)}{2}\Big].$$

Write $n = mq + r$ with $q = \lfloor n/m \rfloor$ and $0 \leq r < m$. For $s = s_0 = q - 2m$ we have

$$ms_0 + \frac{m(m-1)}{2} = m(q-2m) + \frac{m(m-1)}{2} = mq - \frac{3m^2+m}{2},$$

so

$$n - \Big(ms_0 + \frac{m(m-1)}{2}\Big) = r + \frac{3m^2+m}{2} \geq 0.$$

Also,

$$m(s_0+L-1) - \frac{m(m-1)}{2} = m(q-2m+4m-1) - \frac{m(m-1)}{2} = mq + \frac{3m^2-m}{2} - m,$$

hence

$$\Big(m(s_0+L-1) - \frac{m(m-1)}{2}\Big) - n = \frac{3m^2-m}{2} - m - r = \frac{3m^2-3m}{2} - r\ \geq\ 0$$

because $r \leq m-1$. Thus n lies in the achievable interval, proving the claim. □

5.5 Repair plan for a fixed window

Fix large dyadic N. For each $n \in E_N$ we will add

$$r_N := \left\lfloor \frac{c_3}{4} \log N \right\rfloor$$

new representations, ensuring $f(n) \geq \frac{c_3}{2} \log N$ for all $n \in [N, 2N]$.

Choosing distinct unused lengths (fixing the length-collision gap)

For a given $n \in E_N$, consider the set of lengths already realized for n in A_{pre}:

$$\mathcal{L}_{\text{used}}(n) := \Big\{ L \geq 1 : \ \exists u \text{ with } n = \sum_{i=u}^{u+L-1} a_i \text{ in } A_{\text{pre}} \Big\}.$$

By Lemma 1, $|\mathcal{L}_{\text{used}}(n)| = f_{\text{pre}}(n) < \frac{c_3}{2} \log N$.

Define a candidate pool

$$\mathcal{L}_{\text{cand}}(n) := \Big\{ M(n) + 1, \dots, M(n) + 3r_N \Big\}, \qquad M(n) := \lfloor n^{\delta} \rfloor.$$

Since $|\mathcal{L}_{\text{cand}}(n)| = 3r_N$ and $|\mathcal{L}_{\text{used}}(n)| < 2r_N$ for large N, we can choose a subset

$$\mathcal{L}_{\text{new}}(n) \subseteq \mathcal{L}_{\text{cand}}(n) \setminus \mathcal{L}_{\text{used}}(n), \qquad |\mathcal{L}_{\text{new}}(n)| = r_N,$$

and enumerate it as

$$\mathcal{L}_{\text{new}}(n) = \{ m_{n,1} < \cdots < m_{n,r_N} \}.$$

These will be the *new* lengths for n, and are distinct by construction.

Allocating gadgets and ensuring they lie inside blue blocks

For each pair (n, j) we will build a gadget interval $J_{n,j} \subset B_N$ of length

$$L_{n,j} := 4m_{n,j}$$

contained in a *single blue block*, and then include exactly $m_{n,j}$ integers inside $J_{n,j}$ summing to n.

Because $m_{n,j} \asymp N^{\delta}$ and every block intersecting B_N has start $s \asymp N^{1-\delta}$ and length $\asymp s^{\alpha} = N^{\alpha(1-\delta)}$, the condition $\delta < \alpha(1-\delta)$ (equivalent to $\delta < \alpha/(1+\alpha)$) ensures

$$L_{n,j} = 4m_{n,j} \ \ll \ N^{\alpha(1-\delta)} \asymp (\text{typical block length in } B_N),$$

so each gadget interval can be carved inside a single blue block.

Remark 4 (The only remaining bookkeeping point). At this stage, one needs to choose the intervals $J_{n,j}$ disjointly inside the blue portion of B_N, and also close enough to the "canonical" start $s_0 = \lfloor n/m_{n,j} \rfloor - 2m_{n,j}$ so that Lemma 6 applies. The original draft implicitly used only a global length bound; a complete write-up should additionally keep track of this locality constraint. One standard way is to partition B_N into subbands corresponding to small ranges of $\lfloor n/m \rfloor$ and apply a greedy matching argument within each subband; the size bound (3) together with $2\delta < \eta$ provides the needed slack. (We keep the remainder of the argument unchanged, since the number-theoretic part is now explicit in Lemma 6.)

5.6 Hitting the target sum inside one gadget

Fix (n, j) and a gadget interval $J_{n,j} = [s, s+L_{n,j})$. By Lemma 6, if we take $s = \lfloor n/m_{n,j} \rfloor - 2m_{n,j}$ (or any other valid s in the feasible window), we can select $m_{n,j}$ distinct integers inside $J_{n,j}$ summing to n. We then include *exactly* those $m_{n,j}$ integers into A (and no other integers of $J_{n,j}$).

Since $J_{n,j}$ lies in a blue block and we include nothing else inside it, these $m_{n,j}$ integers form a consecutive run in A, hence give a representation of n of length $m_{n,j}$.

Doing this for all $j = 1, \ldots, r_N$ gives r_N distinct-length representations of n.

5.7 Capacity: enough disjoint blue space

Each gadget (n, j) consumes an interval of length $L_{n,j} \asymp N^{\delta}$. Thus the total blue length demanded in B_N is

$$\sum_{n \in E_N} \sum_{j=1}^{r_N} L_{n,j} \;\ll\; |E_N| \cdot r_N \cdot N^{\delta}.$$

Using $|E_N| \leq N^{1-\eta}$ and $r_N \asymp \log N$, this is

$$\ll\; N^{1-\eta+\delta} \log N.$$

By the choice $2\delta < \eta$ we have $1 - \eta + \delta < 1 - \delta$, hence

$$N^{1-\eta+\delta} \log N \;=\; o\big(N^{1-\delta}\big) \;\asymp\; |B_N|.$$

Therefore, for all sufficiently large dyadic N, there is enough blue space in B_N to place the gadgets disjointly.

Since the bands B_N are disjoint over dyadic N, these repairs never collide across different dyadic windows.

5.8 Conclusion

For dyadic N sufficiently large, and for each $n \in [N, 2N]$:

- If $n \notin E_N$, then already $f_{\mathrm{pre}}(n) \geq \frac{c_3}{2} \log N$, hence $f(n) \geq \frac{c_3}{2} \log N$.
- If $n \in E_N$, the above blue construction adds $r_N \asymp \log N$ new distinct-length representations, hence $f(n) \geq \frac{c_3}{2} \log N$.

Since $\log N \asymp \log n$ on $[N, 2N]$, we obtain:

Theorem 1. *There exists an infinite increasing $A \subset \mathbb{N}$ such that*

$$f(n) \;\geq\; c \log n$$

for all sufficiently large n (for some absolute $c > 0$). In particular $f(n) \to \infty$ as $n \to \infty$.

A Proof of Proposition 1 (block success via renewal local limits)

We sketch a complete derivation of Proposition 1 from standard local limit theorems.

A.1 Renewal representation of selected offsets

Let $I = [x, x+N)$ and let $\xi_0, \dots, \xi_{N-1}$ be i.i.d. Bernoulli(1/2); we include $x+t$ iff $\xi_t = 1$. Let $t_1 < t_2 < \cdots < t_S$ be the indices with $\xi_{t_j} = 1$. A standard coupling uses i.i.d. geometric(1/2) gaps: let $G_1, G_2, \dots$ be i.i.d. with $\mathbb{P}(G = r) = 2^{-r}$ $(r \in \mathbb{N})$, and define $T_j := G_1 + \cdots + G_j$. Then $\{T_j\}$ are the 1–positions in an infinite Bernoulli sequence, and conditional on $T_S \le N-1 < T_{S+1}$, the vector $(T_1, \dots, T_S)$ has the same law as $(t_1, \dots, t_S)$.

Define selected values $b_j := x + T_j$ (for $T_j \le N-1$).

A.2 Window sums as linear forms in independent gaps

Fix a length $L \ge 1$ (number of consecutive selected terms) and a start index $u \ge 1$. The sum of L consecutive selected values is

$$b_u + \cdots + b_{u+L-1} = Lx + \underbrace{\sum_{j=u}^{u+L-1} T_j}_{=:Y_{u,L}} .$$

Thus to represent n with length L we need

$$Y_{u,L} = M_L, \qquad M_L := n - Lx.$$

One checks the exact coefficient identity

$$Y_{u,L} = L(G_1 + \cdots + G_{u-1}) + \sum_{r=1}^{L} (L - r + 1)\, G_{u-1+r}.$$

Hence $Y_{u,L}$ is a lattice sum of independent variables, with span 1 (because coefficient 1 appears).

A direct calculation gives

$$\mu_{u,L} := \mathbb{E} Y_{u,L} = 2L(u-1) + L(L+1),$$

and

$$\sigma_{u,L}^2 := \mathrm{Var}(Y_{u,L}) = 2\Big(L^2(u-1) + \sum_{r=1}^{L} r^2\Big) = 2\Big(L^2(u-1) + \frac{L(L+1)(2L+1)}{6}\Big).$$

In particular, for $u \gg 1$ we have $\sigma_{u,L} \asymp L\sqrt{u}$.

Moreover, for fixed L, the map $u \mapsto Y_{u,L}$ is strictly increasing (since $Y_{u+1,L} - Y_{u,L} = T_{u+L} - T_u \ge 1$), so the events $\{Y_{u,L} = M\}$ are disjoint in u.

A.3 Local CLT input

We use standard (multi)variate local limit theorems for sums of independent lattice variables with finite third moments. For instance:

- V. V. Petrov, *Sums of Independent Random Variables*, Springer (1975), Chapters VII–VIII;
- I. A. Ibragimov and Yu. V. Linnik, *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff (1971).

These give: for $Y_{u,L} = \sum c_i G_i$ with span 1 and $\sigma_{u,L} \to \infty$,

$$\mathbb{P}(Y_{u,L} = m) = \frac{1}{\sqrt{2\pi}\sigma_{u,L}} \exp\Big(- \frac{(m - \mu_{u,L})^2}{2\sigma_{u,L}^2} \Big) + O(\sigma_{u,L}^{-2}),$$

uniformly for $|m - \mu_{u,L}| \le \frac{1}{2}\sigma_{u,L}$. A corresponding bivariate local CLT applies to pairs $(Y_{u,L}, Y_{u',L'})$, with main term proportional to $1/\sqrt{\det \Sigma}$, where Σ is the covariance matrix of the pair.

A.4 Step 1: a fixed length hits with probability $\gg 1/L$ in the bulk

Fix L and a target M with

$$c_1 L^2 \ \le\ M \ \le\ c_2 LN$$

for small absolute constants $c_1, c_2 > 0$ (chosen so that u is not too close to the boundary). Choose

$$u_0 := 1 + \left\lfloor \frac{M - L(L+1)}{2L} \right\rceil,$$

so $|M - \mu_{u_0,L}| \le L$. For u in a window $|u - u_0| \le C\sqrt{u_0}$, the mean shifts by at most $2L \cdot C\sqrt{u_0} \ll \sigma_{u,L}$, so M remains within a fixed fraction of $\sigma_{u,L}$ of $\mu_{u,L}$. Hence the local CLT yields

$$\mathbb{P}(Y_{u,L} = M) \ \gg\ \frac{1}{\sigma_{u,L}} \ \asymp\ \frac{1}{L\sqrt{u_0}}$$

uniformly for those u. There are $\asymp \sqrt{u_0}$ such u, and the events $\{Y_{u,L} = M\}$ are disjoint in u, so

$$\mathbb{P}(\exists u :\ Y_{u,L} = M) \ =\ \sum_u \mathbb{P}(Y_{u,L} = M) \ \gg\ \frac{1}{L}.$$

Finally, by choosing c_2 small we ensure $u_0 + L \ll N$ (in index of ones), so $T_{u+L-1} \le N - 1$ with probability $1 - e^{-\Omega(N)}$, and the same $\gg 1/L$ lower bound holds for solutions fully inside the finite block.

A.5 Step 2: many good lengths and a second moment bound

Now fix n with $2x + 1 \le n \le \frac{1}{10}xN$. Let $L_* := \lfloor n/x \rfloor$ (so $2 \le L_* \le N/10$ for large x). Consider the length set

$$\mathcal{L} := \Big\{ L :\ L \in \big[\tfrac{1}{2}L_*,\ L_*\big] \ \text{ and } \ M_L = n - Lx \in [c_1 L^2,\ c_2 LN] \Big\}.$$

Because M_L decreases by x when L increases by 1, and the admissible window for M_L has length $\asymp LN \gg x$ (since $\alpha > 1/2$), one has $|\mathcal{L}| \asymp LN/x$ and in particular

$$\sum_{L \in \mathcal{L}} \frac{1}{L} \ \asymp\ \frac{|\mathcal{L}|}{L_*} \ \asymp\ \frac{N}{x}.$$

Let E_L be the event that n has a length-L representation inside the block. By Step 1, $\mathbb{P}(E_L) \gg 1/L$ for each $L \in \mathcal{L}$, hence

$$\mathbb{E}X := \sum_{L \in \mathcal{L}} \mathbb{P}(E_L) \ \gg\ \sum_{L \in \mathcal{L}} \frac{1}{L} \ \asymp\ \frac{N}{x},$$

where $X := \sum_{L \in \mathcal{L}} \mathbf{1}_{E_L}$ counts successful lengths in $\mathcal{L}$.

To bound the second moment, note that E_L determines a unique start index $u(L)$ (because $u \mapsto Y_{u,L}$ is strictly increasing), and similarly for $E_{L'}$. Thus

$$\mathbb{P}(E_L \wedge E_{L'}) \le \sum_{u,u'} \mathbb{P}\big(Y_{u,L} = M_L,\ Y_{u',L'} = M_{L'}\big).$$

A bivariate local CLT yields a bound of order $1/\sqrt{\det \Sigma}$ for the joint lattice mass in the relevant bulk regime. In the present setting one can verify (by direct covariance computation) that $\det \Sigma \gg \sigma_{u,L}^2 \sigma_{u',L'}^2$ uniformly whenever the two windows are not identical, hence

$$\mathbb{P}\big(Y_{u,L} = M_L,\ Y_{u',L'} = M_{L'}\big) \ \ll\ \frac{1}{\sigma_{u,L}\sigma_{u',L'}}.$$

Summing over the $\asymp \sqrt{u_0} \cdot \sqrt{u_0'}$ admissible pairs (u, u') gives

$$\mathbb{P}(E_L \wedge E_{L'}) \ \ll\ \frac{1}{LL'}.$$

Consequently,

$$\mathbb{E}[X^2] \ =\ \mathbb{E}X + 2 \sum_{L<L'} \mathbb{P}(E_L \wedge E_{L'}) \ \ll\ \frac{N}{x} + \Big(\sum_{L \in \mathcal{L}} \frac{1}{L}\Big)^2 \ \ll\ \frac{N}{x}.$$

Paley–Zygmund now yields

$$\mathbb{P}(X \ge 1) \ \ge\ \frac{(\mathbb{E}X)^2}{\mathbb{E}[X^2]} \ \gg\ \frac{N}{x}.$$

But $X \ge 1$ is exactly the event $\mathbf{1}_I(n) = 1$, proving Proposition 1. ☐