Combinatorics, Probability and Computing (2025), **34**, pp. 283–297
doi:10.1017/S0963548324000361

ARTICLE

On the smallest gap in a sequence with Poisson pair correlations

Daniel Altman[1] and Zachary Chase[2,*]

[1]Department of Mathematics, Stanford University, Stanford, CA, USA and [2]Department of Computer Science and Engineering, University of California, San Diego, CA, USA
Corresponding author: Daniel Altman; Email: daniel.h.altman@gmail.com

(Received 8 January 2023; revised 18 September 2024; accepted 27 September 2024; first published online 21 November 2024)

Abstract
We prove that any increasing sequence of real numbers with average gap 1 and Poisson pair correlations has some gap that is at least $3/2 + 10^{-9}$. This improves upon a result of Aistleitner, Blomer, and Radziwiłł.

Keyword: Poisson pair correlation

2020 MSC Codes: Primary: 11K36

1. Introduction

Let $\lambda = (\lambda_n)_{n=1}^{\infty}$ be an increasing sequence of real numbers. Often, for number-theoretic sequences λ, the average gap $\lambda_{n+1} - \lambda_n$ is well-understood, while little is known about the distribution function of the gaps. Sometimes, however, statistical information about the collection of gaps $\lambda_{n+k} - \lambda_n$ is of importance.

For example, letting $(\gamma_n)_{n=1}^{\infty}$ denote the imaginary parts of the zeroes of the Riemann zeta function in the critical strip in increasing order, we know

$$\#\{\gamma_n \le T\} \sim \frac{T \log T}{2\pi}$$

as $T \to \infty$, and Montgomery's pair-correlation conjecture predicts that

$$\frac{2\pi}{T\log T}\#\left\{(n,m) : \gamma_n, \gamma_m \le T, \frac{2\pi a}{\log T} \le \gamma_m - \gamma_n \le \frac{2\pi b}{\log T}\right\} \to \int_a^b \left(1 - \operatorname{sinc}^2(\pi t)\right) dt$$

as $T \to \infty$, for any fixed $0 < a < b$, where $\operatorname{sinc}(x) := \frac{\sin x}{x}$.

Henceforth, let $\lambda = (\lambda_n)_{n=1}^{\infty}$ denote an increasing sequence with average gap 1:

$$\frac{1}{N}\sum_{n \le N} (\lambda_{n+1} - \lambda_n) \to 1.$$

With this normalization, we may define the *pair correlation function* R_λ by

$$R_\lambda(I, N) := \frac{1}{N}\#\left\{(i,j) : 1 \le i \ne j \le N : \lambda_j - \lambda_i \in I\right\},$$

*Zachary Chase is partially supported by Ben Green's Simons Investigator Grant 376201 and gratefully acknowledges the support of the Simons Foundation.

where $I \subseteq \mathbb{R}$ is a bounded interval and N a positive integer.

For example, up to some normalization technicalities, Montgomery's pair-correlation conjecture asserts that $R_{(\gamma_n)_n}$ converges (in distribution) to a distribution with cumulative distribution function $1 - \text{sinc}^2(\pi t)$.

Motivated by the fact that the pair correlation function of a random sequence generated by a Poisson point process converges in distribution to the uniform distribution, an increasing sequence of real numbers $(\lambda_n)_{n=1}^{\infty}$ with average gap 1 is said to have *Poisson pair correlations* (PPC) if $R_{(\lambda_n)_n}$ converges to the uniform distribution:

$$\lim_{N\to\infty} \frac{1}{N} \left|\left\{(i,j) : 1 \le i \neq j \le N : \lambda_j - \lambda_i \in I\right\}\right| = |I| \tag{1}$$

for all intervals $I \subseteq \mathbb{R}$, where $|\cdot|$ denotes the Lebesgue measure.

Despite Montgomery's pair-correlation conjecture concerning increasing sequences of real numbers, most research on properties of general sequences with Poisson pair correlation concerns sequences in the torus (see, for example, [1–12]), with little investigated about sequences of real numbers.

Some specific number-theoretic sequences of real numbers have been shown to have PPC. Sarnak [10] showed that almost every positive definite binary quadratic form (in a suitable sense) gives rise to a sequence with PPC, by ordering the values it takes on (pairs of) positive integers and appropriately normalizing. Concretely, a consequence of the work of Eskin, Margulis, and Mozes [4] is that the ordered sequence of values of $x^2 + \sqrt{2}y^2$ for $x, y \in \mathbb{N}$ has PPC.

Aistleitner, Blomer, and Radziwiłł [1] studied the related triple correlation function of certain number-theoretic sequences, while also initiating a study of general sequences of real numbers with Poisson pair (and triple) correlations. They asked the following.

Question. *Let* $\lambda_1 < \lambda_2 < \ldots$ *be an increasing sequence of real numbers with average gap* 1 *and with Poisson pair correlations. How small can* $\limsup_{n\to\infty} \lambda_{n+1} - \lambda_n$ *be?*

Among increasing sequences of real numbers with average gap 1 and PPC, Aistleitner, Blomer, and Radziwiłł exhibited one with maximum gap 2, and proved that any such sequence must have a gap of size at least $3/2 - \epsilon$, for any $\epsilon > 0$. They asked in their paper [1] as well as at Oberwolfach 2019 (communicated to the authors by Ben Green) to improve either bound. Our main theorem is an improved lower bound.

Theorem 1.1. *Let* $\lambda_1 < \lambda_2 < \ldots$ *be an increasing sequence of real numbers with average gap* 1 *and Poisson pair correlations. Then* $\limsup_{n\to\infty} \lambda_{n+1} - \lambda_n > \frac{3}{2} + 10^{-9}$.

We leave open the question of how small the largest gap can be; in light of Theorem 1.1, it lies between $\frac{3}{2} + 10^{-9}$ and 2, inclusive.

2. Motivation and proof sketch of Theorem 1.1

In this section we motivate the proof of Theorem 1.1, overlooking some technical complications and emphasizing the main ideas.

Let us begin by sketching the proof given in [1] that any strictly increasing sequence of real numbers with mean gap 1 and PPC has a gap at least $\frac{3}{2} - \epsilon$, for any $\epsilon > 0$. An interested reader may wish to consult the proof sketch of this result in [1, Section 1.5], or the proof itself in [1, Section 7].

One begins by observing that for any sequence $(\lambda_n)_n$ that has PPC, the distribution function of the gaps $\lambda_{n+1} - \lambda_n$, which we will denote by F, can grow at most linearly (indeed, if it grows more than linearly in any small interval, say, this will contradict the PPC condition for this interval). Now, if $\lambda_{n+1} - \lambda_n \le 3/2 - \epsilon$ for each n (that is, $F(3/2 - \epsilon) = 1$), then the linear growth condition implies that the distribution function, when plotted, lies on or above the straight line between $(1/2-\epsilon, 0)$ and $(3/2-\epsilon, 1)$. This implies that that the mean gap size is strictly smaller than 1, a contradiction.

Next, running the same argument now for the situation when the maximum gap is *equal* to 3/2 yields that in this situation we must have

$$F(x) = 0 \text{ for } x \leq \frac{1}{2} \tag{2}$$

$$F(x) = x - \frac{1}{2} \text{ for } x \in \left[\frac{1}{2}, \frac{3}{2}\right]. \tag{3}$$

Our goal is to show that this is impossible (and to then obtain a small quantitative improvement over 3/2). Suppose that (2) and (3) hold. Then (3) yields that PPC $\left(\frac{1}{2}, \frac{3}{2}\right)$ $\left(\text{i.e., (1) for } I = \left[\frac{1}{2}, \frac{3}{2}\right]\right)$ is already satisfied by the single gaps $\lambda_{n+1} - \lambda_n$, so there cannot be a nontrivial contribution coming from larger gaps $\lambda_{n+m} - \lambda_n$, $m \geq 2$. On the other hand, (2) implies that PPC $\left(0, \frac{1}{2}\right)$ must come entirely from a 0 density part of the sequence, and more specifically only from blocks $[n_1, n_2]$ contained in that 0 density part. Furthermore, for the union of such blocks to nontrivially contribute to the PPC count, the length of the blocks must grow with N.

The natural question then is whether such blocks can satisfy the PPC condition on all subintervals of $\left[0, \frac{1}{2}\right]$. We show that the answer is no. The key is to establish a "bias near 0" of the PPC count on long blocks whose total gap is at most 1/2. A bit more precisely, if $\lambda_1 < \ldots < \lambda_k$ have $\lambda_k - \lambda_1 \leq 1/2$, then $\frac{1}{|J|} \sum_{1 \leq i < j \leq k} 1_{\lambda_j - \lambda_i \in J}$ is larger for intervals $J \subseteq [0, 1/2]$ concentrated near 0, with the bias becoming more pronounced as $k \to \infty$. This would contradict the PPC condition on subintervals of $\left[0, \frac{1}{2}\right]$. A difficulty we encounter in the proof, however, is that the blocks forming the relevant 0 density part of the sequence need not have total gap at most 1/2. We overcome this by suitably partitioning the 0 density part of the sequence. This is implemented in Section 4 where we use a suitable greedy algorithm to decompose.

3. Proof of Theorem 1.1

Let $\epsilon = 10^{-9}$. For this section, we fix an increasing sequence of real numbers $\lambda_1 < \lambda_2 < \ldots$ with average gap 1 and PPC, that has $\lambda_{n+1} - \lambda_n \leq 3/2 + \epsilon$ for sufficiently large n. By truncating the sequence, we may assume that

$$g_n := \lambda_{n+1} - \lambda_n$$

satisfies $g_n \leq 3/2 + \epsilon$ for each $n \geq 1$. As in Section 2, we write PPC(a, b) to denote equation (1) for $I = [a, b]$. We note that a sequence satisfying the PPC condition for all such I necessarily satisfies the same condition for all open or indeed half-open intervals. We may therefore also use PPC(a, b) to refer to equation (1) for the half-open interval $[a, b)$, for example.

Recall that the lower bound of $\frac{3}{2}$ was established in [1], whose proof we sketched in Section 2. We start the proof of Theorem 1.1 by making these arguments quantitative.

We begin with a quantitative extension of equation (2), which said that if the maximum gap is 3/2 then, in the limit $N \to \infty$, the proportion of gaps of size at most 1/2 is 0. We obtain the following when the maximum gap is of size at most $3/2 + \epsilon$.

Proposition 3.1. *For all N sufficiently large, we have*

$$\frac{1}{N} \# \left\{ n \leq N : g_n \leq \frac{1}{2} \right\} \leq 2\sqrt{\epsilon}.$$

https://doi.org/10.1017/S0963548324000361 Published online by Cambridge University Press

Proof. We begin by relating the mean gap size to the gap distribution function under the given restriction on the maximum gap size. To this end, note that

$$\begin{aligned}\frac{1}{N}\sum_{n\le N} g_n &= \int_0^{\frac{3}{2}+\epsilon} \frac{1}{N}\#\left\{n\le N : g_n > x\right\} dx \\ &= \int_0^{\frac{1}{2}+\sqrt{\epsilon}} \frac{1}{N}\#\left\{n\le N : g_n > x\right\} dx + \int_{\frac{1}{2}+\sqrt{\epsilon}}^{\frac{3}{2}+\epsilon} \frac{1}{N}\#\left\{n\le N : g_n > x\right\} dx \\ &= \frac{1}{2}+\sqrt{\epsilon} - \int_0^{\frac{1}{2}+\sqrt{\epsilon}} \frac{1}{N}\#\left\{n\le N : g_n \le x\right\} dx \\ &\quad + \int_{\frac{1}{2}+\sqrt{\epsilon}}^{\frac{3}{2}+\epsilon} \frac{1}{N}\#\left\{n\le N : g_n \in \left(x, \frac{3}{2}+\epsilon\right)\right\} dx. \end{aligned} \tag{4}$$

Now, we claim that

$$\limsup_{N\to\infty} \int_{\frac{1}{2}+\sqrt{\epsilon}}^{\frac{3}{2}+\epsilon} \frac{1}{N}\#\left\{n\le N : g_n \in \left(x, \frac{3}{2}+\epsilon\right)\right\} dx \le \int_{\frac{1}{2}+\sqrt{\epsilon}}^{\frac{3}{2}+\epsilon} \left(\frac{3}{2}+\epsilon - x\right) dx. \tag{5}$$

Indeed, the pointwise upper bound

$$\frac{1}{N}\#\left\{n\le N : g_n \in \left(x, \frac{3}{2}+\epsilon\right)\right\} \le \min\left(1, \frac{1}{N}\sum_{n\le N}\sum_{m\le N-n+1} 1_{g_n+\cdots+g_{n+m-1}\in(x,\frac{3}{2}+\epsilon)}\right)$$

together with the dominated convergence theorem and PPC $\left(x, \frac{3}{2}+\epsilon\right)$ for $x \in \left(\frac{1}{2}+\sqrt{\epsilon}, \frac{3}{2}+\epsilon\right)$, namely,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n\le N}\sum_{m\le N-n+1} 1_{g_n+\cdots+g_{n+m-1}\in\left(x,\frac{3}{2}+\epsilon\right)} = \frac{3}{2}+\epsilon - x,$$

gives (5). That the average gap of $(\lambda_n)_n$ is 1 corresponds to

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n\le N} g_n = 1. \tag{6}$$

Rearranging (4), taking $N\to\infty$, and using (5) and (6) gives

$$\begin{aligned}\limsup_{N\to\infty} \int_0^{\frac{1}{2}+\sqrt{\epsilon}} \frac{1}{N}\#\{n\le N : g_n\le x\}dx &\le -1 + \frac{1}{2} + \sqrt{\epsilon} + \int_{\frac{1}{2}+\sqrt{\epsilon}}^{\frac{3}{2}+\epsilon} \left(\frac{3}{2}+\epsilon - x\right) dx \\ &= \frac{3}{2}\epsilon + \frac{1}{2}\epsilon^2 - \epsilon^{3/2}.\end{aligned}$$

Thus, using the trivial

$$\frac{1}{N}\#\left\{n\le N : g_n\le x\right\} \ge \frac{1}{N}\#\left\{n\le N : g_n \le \frac{1}{2}\right\}$$

for $x\in\left[\frac{1}{2}, \frac{1}{2}+\sqrt{\epsilon}\right]$ and the even more trivial lower bound of 0 when $x\in\left[0, \frac{1}{2}\right]$, yields

$$\limsup_{N\to\infty} \sqrt{\epsilon}\cdot\frac{1}{N}\#\left\{n\le N : g_n\le \frac{1}{2}\right\} \le \frac{3}{2}\epsilon + \frac{1}{2}\epsilon^2 - \epsilon^{3/2}.$$

Dividing by $\sqrt{\epsilon}$, Proposition 3.1 follows. □

https://doi.org/10.1017/S0963548324000361 Published online by Cambridge University Press

We now use the quantitative version of (3) to argue that the PPC($\frac{1}{2}$, $\frac{3}{2}+\epsilon$) contribution comes nearly entirely from single gaps g_n.

Proposition 3.2. *For all N sufficiently large, we have*

$$\frac{1}{N}\sum_{n\leq N}\sum_{2\leq m\leq N-n+1}1_{g_n+\cdots+g_{n+m-1}\in(\frac{1}{2},\frac{3}{2}+\epsilon)}\leq 2\sqrt{\epsilon}.$$

Proof. As in the proof of Proposition 3.1,

$$\frac{1}{N}\sum_{n\leq N}g_n=\frac{1}{2}-\sqrt{\epsilon}-\int_0^{\frac{1}{2}-\sqrt{\epsilon}}\frac{1}{N}\#\left\{n\leq N:g_n\leq x\right\}dx$$
$$+\int_{\frac{1}{2}-\sqrt{\epsilon}}^{\frac{3}{2}+\epsilon}\frac{1}{N}\#\left\{n\leq N:g_n\in\left(x,\frac{3}{2}+\epsilon\right)\right\}dx,$$

which, by merely dropping a (negative) term, gives

$$\frac{1}{N}\sum_{n\leq N}g_n\leq\frac{1}{2}-\sqrt{\epsilon}+\int_{\frac{1}{2}-\sqrt{\epsilon}}^{\frac{3}{2}+\epsilon}\frac{1}{N}\#\left\{n\leq N:g_n\in\left(x,\frac{3}{2}+\epsilon\right)\right\}dx. \tag{7}$$

We write

$$\frac{1}{N}\#\left\{n\leq N:g_n\in\left(x,\frac{3}{2}+\epsilon\right)\right\}=\frac{1}{N}\sum_{n\leq N}\sum_{m\leq N-n+1}1_{g_n+\cdots+g_{n+m-1}\in\left(x,\frac{3}{2}+\epsilon\right)}$$
$$-\frac{1}{N}\sum_{n\leq N}\sum_{2\leq m\leq N-n+1}1_{g_n+\cdots+g_{n+m-1}\in\left(x,\frac{3}{2}+\epsilon\right)}$$

and use the same dominated convergence theorem argument as in the proof of Proposition 3.1 to obtain, from (7), that

$$\limsup_{N\to\infty}\int_{\frac{1}{2}-\sqrt{\epsilon}}^{\frac{3}{2}+\epsilon}\frac{1}{N}\sum_{n\leq N}\sum_{2\leq m\leq N-n+1}1_{g_n+\cdots+g_{n+m-1}\in\left(x,\frac{3}{2}+\epsilon\right)}dx$$
$$\leq -1+\frac{1}{2}-\sqrt{\epsilon}+\int_{\frac{1}{2}-\sqrt{\epsilon}}^{\frac{3}{2}+\epsilon}\left(\frac{3}{2}+\epsilon-x\right)dx$$
$$=\frac{3}{2}\epsilon+\frac{1}{2}\epsilon^2+\epsilon^{3/2},$$

and thus

$$\limsup_{N\to\infty}\sqrt{\epsilon}\cdot\frac{1}{N}\sum_{n\leq N}\sum_{2\leq m\leq N-n+1}1_{g_n+\cdots+g_{n+m-1}\in\left(\frac{1}{2},\frac{3}{2}+\epsilon\right)}\leq\frac{3}{2}\epsilon+\frac{1}{2}\epsilon^2+\epsilon^{3/2}.$$

Dividing by $\sqrt{\epsilon}$, the proposition follows. □

We now exploit the aforementioned "bias" towards 0 exhibited by large intervals with sum of gaps at most 1/2. We will need a technical lemma, proven in the appendix but assumed for now.

Lemma 3.3. *For positive integers a, b, c, L satisfying* $1\leq a\leq b\leq c\leq L$, *we have*

$$(a-1)a+(b-a)(b-a+1)+(c-b)(c-b+1)+(L-c)(L-c+1)$$
$$+(a-1)(b-a)+(b-a)(c-b)+(c-b)(L-c)\geq\frac{5}{12}L^2+\frac{1}{6}L-\frac{7}{12}.$$

https://doi.org/10.1017/S0963548324000361 Published online by Cambridge University Press

Lemma 3.3 allows us to show that, instead of getting the desired $\frac{1}{4}+\frac{1}{8}=\frac{3}{8}$ for PPC $\left(0, \frac{1}{4}\right)+$ PPC $\left(0, \frac{1}{8}\right)$, we get at least $\frac{5}{12}=\frac{3}{8}+\frac{1}{24}$, asymptotically for large intervals.

Proposition 3.4. *Let $L \geq 1$ be a positive integer and $g_1, \ldots, g_L$ be positive reals with $\sum_{i=1}^{L} g_i \leq \frac{1}{2}$. Then,*

$$\sum_{n\leq L}\sum_{m\leq L-n+1} 1_{g_n+\cdots+g_{n+m-1}\leq \frac{2}{8}} + \sum_{n\leq L}\sum_{m\leq L-n+1} 1_{g_n+\cdots+g_{n+m-1}\leq \frac{1}{8}} \geq \frac{5}{6}\binom{L+1}{2} - \frac{5}{6}L.$$

Proof. By scaling, it suffices to prove the proposition when $\sum_{i=1}^{L} g_i = \frac{1}{2}$. Suppose $\sum_{i=1}^{L} g_i = \frac{1}{2}$. Let

$$a = \min\left\{j \leq L : g_1+\cdots+g_j \geq \frac{1}{8}\right\},$$

$$b = \min\left\{j \leq L : g_1+\cdots+g_j \geq \frac{2}{8}\right\},$$

$$c = \min\left\{j \leq L : g_1+\cdots+g_j \geq \frac{3}{8}\right\},$$

and note

$$\sum_{n\leq L}\sum_{m\leq L-n+1} 1_{g_n+\cdots+g_{n+m-1}\leq \frac{1}{8}} \geq \frac{(a-1)a}{2} + \frac{(b-a)(b-a+1)}{2} + \frac{(c-b)(c-b+1)}{2} + \frac{(L-c+1)(L-c+2)}{2}.$$

by doing casework in which intervals n and $n+m-1$ lie (the different intervals are $[1,a)$, $[a,b)$, $[b,c)$, $[c,L]$). Similarly,

$$\sum_{n\leq L}\sum_{m\leq L-n+1} 1_{g_n+\cdots+g_{n+m-1}\leq \frac{2}{8}} \geq \frac{(a-1)a}{2} + \frac{(b-a)(b-a+1)}{2} + \frac{(c-b)(c-b+1)}{2} + \frac{(L-c+1)(L-c+2)}{2} + (a-1)(b-a) + (b-a)(c-b) + (c-b)(L-c+1),$$

Therefore,

$$\sum_{n\leq L}\sum_{m\leq L-n+1} 1_{g_n+\cdots+g_{n+m-1}\leq \frac{2}{8}} + \sum_{n\leq L}\sum_{m\leq L-n+1} 1_{g_n+\cdots+g_{n+m-1}\leq \frac{1}{8}} \geq (a-1)a + (b-a)(b-a+1) + (c-b)(c-b+1) + (L-c+1)(L-c+2) + (a-1)(b-a) + (b-a)(c-b) + (c-b)(L-c+1).$$

Lower bounding $L-c+1$ and $L-c+2$ by $L-c$ and $L-c+1$, respectively, Lemma 3.3 finishes the proof of Proposition 3.4, since $\frac{5}{12}L^2 + \frac{1}{6}L - \frac{7}{12} \geq \frac{5}{6}\binom{L+1}{2} - \frac{5}{6}L$ for $L \geq 1$. □

We now proceed to isolate the relevant "0 density" parts of the sequence on which the gaps are at most 1/2, in order to exploit the bias that Proposition 3.4 illustrates.

Here and henceforth, we let $[N] := \{1, 2, \ldots, N\}$.

https://doi.org/10.1017/S0963548324000361 Published online by Cambridge University Press

Definition 3.5. *For a nonempty interval $J \subseteq [N]$, let $L(J), R(J)$ denote the left and right endpoints of J, respectively, and let* $\mathrm{sum}(J) = \sum_{n \in J} g_n$.

Definition 3.6. *For a interval $J \subseteq [N]$, we denote*

$$\mathrm{PPC}^J(0, a) := \sum_{n \le n' \in J} 1_{g_n + \cdots + g_{n'} < a}.$$

For intervals $J_1, J_2 \subseteq [N]$, with $R(J_1) < L(J_2)$, we denote

$$\mathrm{PPC}^{J_1, J_2}(0, a) := \sum_{(n,n') \in J_1 \times J_2} 1_{g_n + \cdots + g_{n'} < a}.$$

Take a large N. Let $\mathcal{I}^N$ denote the collection of all maximal intervals on which the gaps g_n are at most $\frac{1}{2}$. More formally, we define $\mathcal{I}^N$ to be the collection of all intervals $I \subseteq [N]$ such that (1) $g_n \le \frac{1}{2}$ for each $n \in I$, (2) $L(I) = 1$ or $g_{L(I)-1} > \frac{1}{2}$, and (3) $R(I) = N$ or $g_{R(I)+1} > \frac{1}{2}$.

We begin by noting the following.

Lemma 3.7. *For all large N, we have*

$$\sum_{I \in \mathcal{I}^N} |I| \le 2\sqrt{\epsilon} N.$$

Furthermore, as $N \to \infty$, we have

$$\sum_{I \in \mathcal{I}^N} \binom{|I|+1}{2} \ge N\left(\frac{1}{2} + o(1)\right).$$

Proof. By definition we have

$$\sum_{I \in \mathcal{I}^N} |I| = \frac{1}{N} \# \left\{ n \le N : g_n \le \frac{1}{2} \right\}.$$

Proposition 3.1 then gives the first inequality. For the second inequality, note, by the maximality of the intervals comprising $\mathcal{I}^N$, that

$$\left(\frac{1}{2} + o(1)\right) N = \mathrm{PPC}^{[N]}\left(0, \frac{1}{2}\right) = \sum_{I \in \mathcal{I}^N} \mathrm{PPC}^I\left(0, \frac{1}{2}\right) \le \sum_{I \in \mathcal{I}^N} \binom{|I|+1}{2}.$$

□

We provide a quick remark on motivation.

Remark 3.8. Observe that, if it were the case that $\mathrm{sum}(I) \le \frac{1}{2}$ for each $I \in \mathcal{I}^N$, then we could conclude the proof of Theorem 1.1 as follows. By Proposition 3.4, one has

$$\begin{aligned} \left(\frac{3}{8} + o(1)\right) N &= \mathrm{PPC}^{[N]}\left(0, \frac{1}{8}\right) + \mathrm{PPC}^{[N]}\left(0, \frac{1}{4}\right) \\ &\ge \frac{5}{6} \sum_I \binom{|I|+1}{2} - \frac{5}{6} \sum_I |I| \\ &\ge \left(\frac{5}{12} - \frac{5}{3}\sqrt{\epsilon} + o(1)\right) N \end{aligned}$$

as $N \to \infty$, which would give our desired contradiction by taking N sufficiently large.

However, it need not be the case that $\mathrm{sum}(I) \le \frac{1}{2}$ for each $I \in \mathcal{I}^N$. In light of Remark 3.8, therefore, the strategy is to partition each $I \in \mathcal{I}^N$ into subintervals $\{J_k^I\}_k$ with sum $\left(J_k^I\right) \le 1/2$ for each k and such that, for all $a \in (0, 1/2]$ and all k, the contribution to PPC$(0, a)$ from windows that overlap with J_k^I comes nearly entirely from windows that lie entirely inside J_k^I. The existence of such a

https://doi.org/10.1017/S0963548324000361 Published online by Cambridge University Press

partition is not at all immediate. We provide in Proposition 3.9 below a precise statement of what is needed.

Proposition 3.9. *Let $\mathcal{I}^N$ be as above. There exists a partition of each $I \in \mathcal{I}^N$ into subintervals $\{J_k^I\}_{k=1}^{r_I}$ such that the following two hold.*

1. $\operatorname{sum}\left(J_k^I\right) \leq \frac{1}{2}$ *for each $I \in \mathcal{I}^N$ and $k \in \{1, \ldots, r_I\}$, and*
2. $\sum_{I \in \mathcal{I}^N} \sum_{k=1}^{r_I} \binom{|J_k^I|+1}{2} \geq \left(\frac{1}{2} - 4\sqrt{2}\epsilon^{1/4}\right) N.$

To quickly conclude the proof of Theorem 1.1, we postpone the proof of Proposition 3.9 (and the description of the partition) to the following section, and assume it for now.

Proof of Theorem 1.1. We will proceed along the lines of Remark 3.8, but now use Proposition 3.9 to partition into subintervals on which we can apply Proposition 3.4. This yields the following computations, where Proposition 3.4 is used in the penultimate line, and we use Lemma 3.7 in the final line:

$$\begin{aligned}
\left(\frac{3}{8} + o(1)\right) N &= \mathrm{PPC}^{[N]}\left(0, \frac{1}{8}\right) + \mathrm{PPC}^{[N]}\left(0, \frac{1}{4}\right) \\
&\geq \sum_{I \in \mathcal{I}^N} \sum_{k} \mathrm{PPC}^{J_k^I}\left(0, \frac{1}{8}\right) + \mathrm{PPC}^{J_k^I}\left(0, \frac{1}{4}\right) \\
&\geq \frac{5}{6} \sum_{I \in \mathcal{I}^N} \sum_{k} \binom{|J_k^I| + 1}{2} - \frac{5}{6} \sum_{I \in \mathcal{I}^N} \sum_{k} |J_k^I| \\
&\geq \frac{5}{6}\left(\frac{1}{2} - 4\sqrt{2}\epsilon^{1/4}\right) N - \frac{5}{3}\sqrt{\epsilon} N.
\end{aligned}$$

Rearranging, dividing by N, and sending $N \to \infty$, we obtain

$$\frac{10\sqrt{2}}{3}\epsilon^{1/4} + \frac{5}{3}\sqrt{\epsilon} - \frac{1}{24} \geq 0,$$

which is indeed false for $\epsilon = 10^{-9}$ (but not for $\epsilon = 10^{-8}$). This gives the desired contradiction to our assumption that a sequence $(\lambda_n)_n$ with PPC, average gap 1, and maximum gap $3/2 + \epsilon$ exists. □

4. Partitioning, and a proof of Proposition 3.9

The following examples are helpful to keep in mind to explain the need for care when choosing the partition of a given $I \in \mathcal{I}^N$, and to help motivate the partition we will use.

$$\begin{aligned}
I_1 &= \left\{\frac{2}{5}, 0, 0, 0, \ldots, 0, 0, 0, \frac{1}{3}, 0, 0, 0, \ldots, 0, 0, 0, \frac{2}{5}\right\} \\
I_2 &= \left\{\frac{1}{4}, 0, 0, 0, \ldots, 0, 0, 0, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0, 0, 0, \ldots, 0, 0, 0, \frac{1}{4}\right\} \\
I_3 &= \left\{0, 0, \ldots, 0, 0, \frac{1}{3}, 0, 0, \ldots, 0, 0, \frac{1}{3}, 0, 0, \ldots, 0, 0\right\}.
\end{aligned}$$

The first example I_1 shows that a "greedy division", in which one goes from left to right, dividing immediately before the sum first exceeds $1/2$, will not work. Indeed, that division is

$$\left\{\frac{2}{5}, 0, 0, 0, \ldots, 0, 0, 0\right\}, \left\{\frac{1}{3}, 0, 0, 0, \ldots, 0, 0, 0\right\}, \left\{\frac{2}{5}\right\},$$

https://doi.org/10.1017/S0963548324000361 Published online by Cambridge University Press

which is problematic as there will be much contribution to the PPC$(0, \frac{1}{2})$ count coming from different subintervals; specifically, any index besides the first in the first subinterval and any index in the second subinterval would prove a nontrivial contribution.

Similarly, another "greedy division" in which the largest numbers successively "claim" the largest subinterval they can, will not work. For the case of I_1, the division is

$$\left\{\frac{2}{5}, 0, 0, 0, \ldots, 0, 0, 0\right\}, \left\{\frac{1}{3}\right\}, \left\{0, 0, 0, \ldots, 0, 0, 0, \frac{2}{5}\right\},$$

which has a large contribution to PPC $\left(0, \frac{1}{2}\right)$ coming from any index in the first subinterval besides the first and any index in the last subinterval besides the last. Note thus that even "non-adjacent" subintervals can cause issues.

A division that does work for I_1 is

$$\left\{\frac{2}{5}\right\}, \left\{0, 0, 0, \ldots, 0, 0, 0, \frac{1}{3}, 0, 0, 0, \ldots, 0, 0, 0\right\}, \left\{\frac{2}{5}\right\},$$

as there is only a minor contribution (namely, linear in the size of the interval rather than quadratic) coming from different subintervals.

For I_2, essentially any reasonable division is permissible, but we draw attention to it as it shows that sometimes the *reason* for negligible contribution from different subintervals are subintervals in between. For example, if we decompose as

$$\left\{\frac{1}{4}, 0, 0, 0, \ldots, 0, 0, 0\right\}, \left\{\frac{1}{2}\right\}, \left\{\frac{1}{2}\right\}, \left\{\frac{1}{2}\right\}, \left\{\frac{1}{2}\right\}, \left\{\frac{1}{2}\right\}, \left\{0, 0, 0, \ldots, 0, 0, 0, \frac{1}{4}\right\},$$

then the reason that there is no contribution to PPC $\left(0, \frac{1}{2}\right)$ from the subintervals $\{\frac{1}{4}, 0, 0, \ldots, 0, 0\}$ and $\{0, 0, \ldots, 0, 0, \frac{1}{4}\}$ are the five subintervals $\{\frac{1}{2}\}$ in between.

For I_3, any division will admit a large PPC contribution from different subintervals. This would of course be harmful, but we make use of the fact that it won't exist often in our situation, since it provides a nontrivial contribution to PPC$(\frac{1}{2}, 1)$ (which we already know comes nearly entirely from single gaps).

With the above examples in mind, we now choose the partition we use to prove Proposition 3.9.

Fix $I \in \mathcal{I}^N$. In the following definition, ties may be broken arbitrarily. Let J_1 be the largest subinterval of I with sum$(J_1) \le \frac{1}{2}$. With $J_1, \ldots, J_r$ already defined, if $\cup_{k=1}^{r} J_k \neq I$, let J_{r+1} be the largest subinterval of $I \setminus \cup_{k=1}^{r} J_k$ with sum$(J_{r+1}) \le \frac{1}{2}$. Let $J_1, \ldots, J_s$ be all the subintervals resulting from this process. Of course $s \le |I| < +\infty$.

Clearly $I = \sqcup_{k=1}^{s} J_k$ and sum$(J_k) \le 1/2$ for each k, establishing the first requirement of Proposition 3.9. We now begin to proceed to establish the second.

Hopefully not confusing the reader, we renumber now so that J_1 is the leftmost interval, with J_2 to the immediate right of J_1, J_3 to the immediate right of J_2, etc.. For $1 \le k \le s - 1$, let $g_1(k) \in \{k, k+1\}$ and $b_1(k) \in \{k, k+1\} \setminus \{g_1(k)\}$ be such that $J_{g_1(k)}$ was chosen before $J_{b_1(k)}$. Note, in particular, that $|J_{g_1(k)}| \ge |J_{b_1(k)}|$.

We quickly pin down exactly which different subintervals need to be considered with regards to their contribution to PPC$(0, a)$, for $a \le 1/2$.

https://doi.org/10.1017/S0963548324000361 Published online by Cambridge University Press

Definition 4.1. *Call $k \in [2, s-1]$ sandwiched if it was chosen after each of its neighbouring subintervals, i.e., if $b_1(k-1) = k$ and $b_1(k) = k$. For a sandwiched k, let $g_2(k) \in \{k-1, k+1\}$ and $b_2(k) \in \{k-1, k+1\} \setminus \{g_2(k)\}$ be such that $J_{g_2(k)}$ was chosen before $J_{b_2(k)}$. In particular, $|J_{g_2(k)}| \geq |J_{b_2(k)}|$.*

Lemma 4.2. *If $n \leq n' \in I$ have $g_n + \cdots + g_{n'} \leq \frac{1}{2}$, then either*

1. *$(n, n') \in J_k \times J_k$ for some k,*
2. *$(n, n') \in J_k \times J_{k+1}$ for some k, or*
3. *$(n, n') \in J_{k-1} \times J_{k+1}$ for some sandwiched k.*

Proof. First note that if $n \in J_k$ and $n' \in \cup_{\Delta \geq 3} J_{k+\Delta}$, then $g_n + \cdots + g_{n'} > \frac{1}{2}$, since $\operatorname{sum}(J_{k+1} \cup J_{k+2}) > \frac{1}{2}$ (for otherwise whichever of J_{k+1}, J_{k+2} was chosen first would have "engulfed" the other). Now suppose $n \in J_{k-1}$ and $n' \in J_{k+1}$ for some k. If $b_1(k-1) = k-1$, then $\operatorname{sum}(R(J_{k-1}) \cup J_k) > \frac{1}{2}$, so $g_n + \cdots + g_{n'} > \frac{1}{2}$. Similarly, if $b_1(k) = k+1$, then $\operatorname{sum}(J_k \cup L(J_{k+1})) > \frac{1}{2}$ also yields $g_n + \cdots + g_{n'} > \frac{1}{2}$. Hence, k is sandwiched. □

We now proceed to argue that the PPC$\left(0, \frac{1}{2}\right)$ contribution coming from cases (2) or (3) in Lemma 4.2 is small. We begin with case (2).

We shall argue that the PPC $\left(0, \frac{1}{2}\right)$ contribution coming from adjacent subintervals J_k, J_{k+1} is small by arguing that $|J_{b_1(k)}|$ is (usually) small. We do this by arguing that we would otherwise have too large of a contribution to PPC $\left(\frac{1}{2}, \frac{3}{2}\right)$ coming from gaps $\lambda_{n+m} - \lambda_n$ with $m \geq 2$ (contradicting Proposition 3.2).

Proposition 4.3. *For any $k \in [s-1]$, one has*

$$\sum_{(n,n') \in J_k \times J_{k+1}} 1_{g_n + \cdots + g_{n'} > \frac{1}{2}} \geq \frac{1}{2} |J_{b_1(k)}|^2.$$

Proof. Without loss of generality, by symmetry we may assume $b_1(k) = k+1$.

For $y \in J_{k+1}$ and $x \in J_k$, if $y - x + 1 > |J_k|$, then $g_x + \cdots + g_y > \frac{1}{2}$, since otherwise $[x, y]$ would have been chosen as J_k (in the greedy process defining the partition) instead of J_k. Therefore,

$$\begin{aligned}
\sum_{(n,n') \in J_k \times J_{k+1}} 1_{g_n + \cdots + g_{n'} > \frac{1}{2}} &\geq \sum_{y = L(J_{k+1})}^{R(J_{k+1})} \sum_{x = L(J_k)}^{y - R(J_k) + L(J_k) - 1} 1 \\
&= \left(\frac{L(J_{k+1}) + R(J_{k+1})}{2} \right) |J_{k+1}| - R(J_k) |J_{k+1}| \\
&= \frac{1}{2} |J_{k+1}|^2 + \frac{1}{2} |J_{k+1}|,
\end{aligned}$$

with the last equality using $R(J_k) = L(J_{k+1}) - 1$. □

Next we proceed to bound the contribution from intervals J_{k-1}, J_{k+1} for k sandwiched. For such k, we argue that $|J_{b_2(k)}|$ is (usually) small.

Proposition 4.4. *For a sandwiched k, one has*

$$\sum_{(n,n') \in J_{k-1} \times J_{k+1}} 1_{g_n + \cdots + g_{n'} > \frac{1}{2}} \geq \frac{1}{2} |J_{b_2(k)}|^2.$$

Proof. Without loss of generality, by symmetry we may assume $b_2(k) = k+1$.

https://doi.org/10.1017/S0963548324000361 Published online by Cambridge University Press

For $y \in J_{k+1}$ and $x \in J_{k-1}$, if $y - x + 1 > |J_{k-1}|$, then $g_x + \cdots + g_y > \frac{1}{2}$, since otherwise $[x, y]$ would have been chosen instead of J_{k-1} (recall J_k was also chosen after J_{k-1}, since k is sandwiched). Therefore,

$$\sum_{(n,n')\in J_{k-1}\times J_{k+1}} 1_{g_n+\cdots+g_{n'}>\frac{1}{2}} \geq \sum_{y=L(J_{k+1})}^{R(J_{k+1})} \sum_{x=L(J_{k-1})}^{\min(R(J_{k-1}),y-|J_{k-1}|)} 1.$$

If $R(J_{k+1}) - |J_{k-1}| \leq R(J_{k-1})$, then we obtain

$$\begin{aligned}\sum_{(n,n')\in J_{k-1}\times J_{k+1}} 1_{g_n+\cdots+g_{n'}>\frac{1}{2}} &\geq \sum_{y=L(J_{k+1})}^{R(J_{k+1})} \sum_{x=L(J_{k-1})}^{y-|J_{k-1}|} 1 \\ &= |J_{k+1}|\left(\frac{R(J_{k+1})+L(J_{k+1})}{2} - |J_{k-1}| - L(J_{k-1}) + 1\right) \\ &= |J_{k+1}|\left(\frac{R(J_{k+1})+L(J_{k+1})}{2} - R(J_{k-1})\right) \\ &= |J_{k+1}|\left(\frac{R(J_{k+1})+L(J_{k+1})}{2} - L(J_{k+1}) + |J_k| + 1\right) \\ &= |J_{k+1}|\left(\frac{|J_{k+1}|-1}{2} + |J_k| + 1\right),\end{aligned}$$

and conclude by observing that $|J_k| \geq 0$ and $\frac{|J_{k+1}|-1}{2} + 1 \geq \frac{1}{2}|J_{k+1}|$.
If, instead, $R(J_{k+1}) - |J_{k-1}| \geq R(J_{k-1})$, we obtain

$$\sum_{(n,n')\in J_{k-1}\times J_{k+1}} 1_{g_n+\cdots+g_{n'}>\frac{1}{2}} \geq \sum_{y=L(J_{k+1})}^{|J_{k-1}|+R(J_{k-1})} \sum_{x=L(J_{k-1})}^{y-|J_{k-1}|} 1 + \sum_{y=|J_{k-1}|+R(J_{k-1})+1}^{R(J_{k+1})} \sum_{x=L(J_{k-1})}^{R(J_{k-1})} 1. \tag{8}$$

The first double sum on the RHS of (8) is equal to

$$(|J_{k-1}| + R(J_{k-1}) - L(J_{k+1}) + 1)\left(\frac{|J_{k-1}| + R(J_{k-1}) + L(J_{k+1})}{2} - R(J_{k-1})\right),$$

while the second double sum is equal to

$$\big(R(J_{k+1}) - |J_{k-1}| - R(J_{k-1})\big)\,|J_{k-1}|.$$

Adding these two sums and simplifying we obtain,

$$\begin{aligned}(|J_{k-1}| - |J_k|)\left(\frac{|J_{k-1}| + |J_k| + 1}{2}\right) &+ |J_{k-1}|(|J_k| + |J_{k+1}| - |J_{k-1}|) \\ &= -\frac{1}{2}(|J_{k-1}| - |J_k|)^2 + \frac{1}{2}(|J_{k-1}| - |J_k|) + |J_{k-1}||J_{k+1}|.\end{aligned}$$

Now recall that we have $|J_k| \leq |J_{k+1}| \leq |J_{k-1}|$ and furthermore, by our assumption in the second case, we have $|J_{k-1}| \leq |J_k| + |J_{k+1}|$. Thus we may obtain

$$-\frac{1}{2}(|J_{k-1}| - |J_k|)^2 + \frac{1}{2}(|J_{k-1}| - |J_k|) + |J_{k-1}||J_{k+1}| \geq -\frac{1}{2}|J_{k+1}|^2 + 0 + |J_{k+1}|^2 = \frac{1}{2}|J_{k+1}|^2.$$

This completes the proof. □

https://doi.org/10.1017/S0963548324000361 Published online by Cambridge University Press

We now cease referring to a specific $I \in \mathcal{I}^N$. To denote dependence on $I \in \mathcal{I}^N$, we denote $I = \sqcup_{k=1}^r J_k^I$ its decomposition.

Proposition 4.5. *For any $a \in \left(0, \frac{1}{2}\right)$, it holds that*

$$\sum_{I\in\mathcal{I}^N}\sum_k \mathrm{PPC}^{J_k^I,J_{k+1}^I}(0,a) \le 2\sqrt{2}\epsilon^{1/4}\left(\sum_{I\in\mathcal{I}^N}\sum_k |J_k^I|^2\right)^{1/2}\sqrt{N}$$

and

$$\sum_{I\in\mathcal{I}^N}\sum_{k \text{ sandwiched}} \mathrm{PPC}^{J_{k-1}^I,J_{k+1}^I}(0,a) \le 2\sqrt{2}\epsilon^{1/4}\left(\sum_{I\in\mathcal{I}^N}\sum_k |J_k^I|^2\right)^{1/2}\sqrt{N}.$$

Proof. By trivially bounding $\mathrm{PPC}^{J_k^I,J_{k+1}^I}(0,a) \le |J_{g_1(k)}^I|\,|J_{b_1(k)}^I|$ and Cauchy-Schwarz,

$$\begin{aligned}\sum_{I\in\mathcal{I}^N}\sum_k \mathrm{PPC}^{J_k^I,J_{k+1}^I}(0,a) &\le \sum_{I\in\mathcal{I}^N}\sum_k |J_{g_1(k)}^I|\,|J_{b_1(k)}^I|\\ &\le \left(\sum_{I,k}|J_{g_1(k)}^I|^2\right)^{1/2}\left(\sum_{I,k}|J_{b_1(k)}^I|^2\right)^{1/2}\\ &\le \left(2\sum_{I\in\mathcal{I}^N}\sum_k |J_k^I|^2\right)^{1/2}\left(\sum_{I\in\mathcal{I}^N}\sum_k 2\mathrm{PPC}^{J_k^I,J_{k+1}^I}\left(\frac{1}{2},\frac{3}{2}\right)\right)^{1/2},\end{aligned}$$

where the last inequality used Proposition 4.3 together with the fact that sum $\left(J_k^I \cup J_{k+1}^I\right) \le 1 \le \frac{3}{2}$. Now just observe

$$\sum_{I\in\mathcal{I}^N}\sum_k 2\mathrm{PPC}^{J_k^I,J_{k+1}^I}\left(\frac{1}{2},\frac{3}{2}\right) \le 2\sum_{n\le N}\sum_{2\le m\le N-n+1} 1_{g_n+\cdots+g_{n+m-1}\in\left(\frac{1}{2},\frac{3}{2}\right)},$$

which, by Proposition 3.2, is at most $4\sqrt{\epsilon}N$. The first inequality of the lemma follows.

For the second inequality of the lemma, we argue as above, except this time using Proposition 4.4:

$$\begin{aligned}&\sum_{I\in\mathcal{I}^N}\sum_{k \text{ sandwiched}} \mathrm{PPC}^{J_{k-1}^I,J_{k+1}^I}(0,a)\\ &\quad\le \sum_{I\in\mathcal{I}^N}\sum_{k \text{ sandwiched}} |J_{g_2(k)}^I|\,|J_{b_2(k)}^I|\\ &\quad\le \left(\sum_{I\in\mathcal{I}^N}\sum_{k \text{ sandwiched}} |J_{g_2(k)}^I|^2\right)^{1/2}\left(\sum_{I\in\mathcal{I}^N}\sum_{k \text{ sandwiched}} |J_{b_2(k)}^I|^2\right)^{1/2}\\ &\quad\le \left(2\sum_{I\in\mathcal{I}^N}\sum_k |J_k^I|^2\right)^{1/2}\left(\sum_{I\in\mathcal{I}^N}\sum_k 2\mathrm{PPC}^{J_{k-1}^I,J_{k+1}^I}\left(\frac{1}{2},\frac{3}{2}\right)\right)^{1/2},\end{aligned}$$

https://doi.org/10.1017/S0963548324000361 Published online by Cambridge University Press

where the last inequality used $\mathrm{sum}(J_{k-1} \cup J_k \cup J_{k+1}) \le \frac{3}{2}$. Now just observe

$$\sum_{I \in \mathcal{I}^N} \sum_k 2\mathrm{PPC}^{J^I_k, J^I_{k+1}}\left(\frac{1}{2}, \frac{3}{2}\right) \le 2 \sum_{n \le N} \sum_{2 \le m \le N-n+1} 1_{g_n + \cdots + g_{n+m-1} \in (\frac{1}{2}, \frac{3}{2})},$$

which, by Proposition 3.2, is at most $4\sqrt{\epsilon}N$. The second inequality of the lemma follows. □

We are ready to complete the proof of Proposition 3.9.

Proposition 4.6. *For all N large,*

$$\sum_{I \in \mathcal{I}^N} \sum_k \binom{|J^I_k| + 1}{2} \ge \left(\frac{1}{2} - 4\sqrt{2}\epsilon^{1/4}\right) N.$$

Proof. Using Lemma 4.2 we have that

$$\left(\frac{1}{2} + o(1)\right) N = \sum_{I \in \mathcal{I}^N} \mathrm{PPC}^I\left(0, \frac{1}{2}\right) = \sum_I \sum_{k=1}^{r_I} \mathrm{PPC}^{J^I_k}\left(0, \frac{1}{2}\right) + \sum_{k=1}^{r_I - 1} \mathrm{PPC}^{J^I_k, J^I_{k+1}}\left(0, \frac{1}{2}\right)$$
$$+ \sum_{k \text{ sandwiched}} \mathrm{PPC}^{J_{k-1}, J_{k+1}}\left(0, \frac{1}{2}\right).$$

Invoking Proposition 4.5 in the first line we have:

$$\left(\frac{1}{2} + o(1)\right) N - \sum_{I \in \mathcal{I}^N} \sum_k \binom{|J^I_k| + 1}{2} \le 4\sqrt{2}\epsilon^{1/4} \left(\sum_{I \in \mathcal{I}^N} \sum_k |J^I_k|^2\right)^{1/2} \sqrt{N}$$
$$\le 8\epsilon^{1/4} \left(\sum_{I \in \mathcal{I}^N} \sum_k \binom{|J^I_k| + 1}{2}\right)^{1/2} \sqrt{N}.$$

Writing

$$\sum_{I \in \mathcal{I}^N} \sum_k \binom{|J^I_k| + 1}{2} = \left(\frac{1}{2} - \delta\right) N,$$

we see

$$\delta + o(1) \le 8\epsilon^{1/4} \left(\frac{1}{2} - \delta\right)^{1/2},$$

which yields (after a little computation) $\delta \le 4\sqrt{2}\epsilon^{1/4}$ for N large enough. □

5. Appendix: proof of Lemma 3.3

We restate Lemma 3.3 for the reader's convenience.

Lemma 3.3 (Lemma 3.3). *For positive integers a, b, c, L satisfying* $1 \le a \le b \le c \le L$, *we have*

$$(a-1)a + (b-a)(b-a+1) + (c-b)(c-b+1) + (L-c)(L-c+1)$$
$$+ (a-1)(b-a) + (b-a)(c-b) + (c-b)(L-c) \ge \frac{5}{12}L^2 + \frac{1}{6}L - \frac{7}{12}.$$

Proof. Fix $L \ge 1$. It clearly suffices to prove the inequality for all real numbers a, b, c satisfying $1 \le a \le b \le c \le L$. By compactness, we may work with a triple (a, b, c) that achieves the minimum value of the left hand side (which we denote LHS) minus the right hand side (which we denote RHS). We divide into three cases.

Case 1: $c = L$.

As one may compute, $\frac{\partial}{\partial c}[\text{LHS} - \text{RHS}] = -a + 2c - L = L - a$.

Subcase 1: $a = L$. Then $b = L$, which gives $\text{LHS} - \text{RHS} = \frac{28L^2 - 51L + 28}{48}$, which is non-negative, since it is equal to 5 at $L = 1$ and has derivative $56L - 51$, which is positive for $L \geq 1$.

Subcase 2: $a \neq L$. Then $\frac{\partial}{\partial c}[\text{LHS} - \text{RHS}] > 0$, so since (a, b, c) is a minimizer, we must have $b = c$, for otherwise we can decrease c a bit to decrease LHS − RHS. Thus, $\text{LHS} - \text{RHS} = a^2 - (L + 1)a + \frac{28L^2 - 3L + 28}{48}$, which has minimum occurring at $a = \frac{L+1}{2}$, which gives $\text{LHS} - \text{RHS} = \frac{1}{3}L^2 - \frac{9}{16}L + \frac{1}{3}$, which is always non-negative, since it is at $L = 1$ and the derivative is $\frac{2}{3}L - \frac{9}{16}$, which is non-negative for $L \geq 1$.

Case 2: $c \neq L$ and $b = c$.

Since $c \neq L$ and (a, b, c) is a minimizer, we must have $\frac{\partial}{\partial c}[\text{LHS} - \text{RHS}] \geq 0$, for otherwise we could increase b, c a bit to decrease LHS − RHS. Recall $\frac{\partial}{\partial c}[\text{LHS} - \text{RHS}] = -a + 2c - L$; so, $2c - L \geq a$.

Subcase 1: $a = b$. In this case, $\text{LHS} - \text{RHS} = 2c^2 - (2L + 2)c + \frac{7L^2 + 10L + 7}{12}$, which has minimum at $c = \frac{L+1}{2}$, which yields $\text{LHS} - \text{RHS} = \frac{(L-1)^2}{12}$, which is non-negative.

Subcase 2: $a \neq b$. In this case, we must have $\frac{\partial}{\partial b}[\text{LHS} - \text{RHS}] \leq 0$, since otherwise we could decrease b a bit to decrease LHS − RHS. Note $\frac{\partial}{\partial b}[\text{LHS} - \text{RHS}] = -1 + 2b - L$, so since $2b - L \geq a$, we must have $a = 1$ and $2b - L = 1$. So, we have $a = 1, b = \frac{L+1}{2}, c = \frac{L+1}{2}$, which indeed has $\text{LHS} - \text{RHS} \geq 0$.

Case 3: $c \neq L$ and $b \neq c$.

In this case, $\frac{\partial}{\partial c}[\text{LHS} - \text{RHS}]$ must be 0, for otherwise we could perturb c a bit to decrease LHS − RHS. So, $-a + 2c - L = 0$.

Subcase 1: $a = 1$. Having $a = 1$ and $-a + 2c - L = 0$, i.e., $c = \frac{L+1}{2}$, yields $\text{LHS} - \text{RHS} = \frac{1}{4}[4b^2 - 4(L + 1)b + 3L^2 + 2L - 1]$, which is minimized as $b = \frac{L+1}{2}$, which was dealt with in Subcase 2 of Case 2.

Subcase 2: $a = b$. Then, $48(\text{LHS} - \text{RHS}) = 84b^2 - (72L + 96)b + 16L^2 + 45L + 28$, which has minimum at $b = \frac{72L+96}{84}$, which gives $48(\text{LHS} - \text{RHS}) = 16L^2 + 45L + 28$, which is clearly non-negative.

Subcase 3: $a \notin \{1, b\}$. Then $\frac{\partial}{\partial a}[\text{LHS} - \text{RHS}] = 0$, for otherwise we could perturb a a bit to decrease LHS − RHS. So, $-1 + 2a - c = 0$. Together with $-a + 2c - L = 0$ yields $a = \frac{L+2}{3}, c = \frac{2L+1}{3}$, which yields $\text{LHS} - \text{RHS} = b^2 - (L + 1)b + \frac{12L^2 + 29L + 12}{48}$, which is minimized at $b = \frac{L+1}{2}$, which yields $\frac{5L}{48}$, which is clearly non-negative. □

Acknowledgments

We would like to thank our advisor Ben Green for suggesting this problem to us. We would also like to thank Zach Hunter for providing feedback on an earlier version of this document.

References

[1] Aistleitner, C., Blomer, V. and Radziwiłł, M. (2024) Triple correlation and long gaps in the spectrum of flat tori. *J. Eur. Math. Soc.* **26**(1) 41–74.

[2] Aistleitner, C., Lachmann, T. and Pausinger, F. (2018) Pair correlations and equidistribution. *J. Num. Theory* **182** 206–220.

[3] Aistleitner, C., Larcher, G. and Lewko, M. (2017) Additive energy and the Hausdorff dimension of the exceptional set in metric pair correlation problems, with an Appendix by Jean Bourgain. *Israel J. Math.* **222**(1) 463–485.

[4] Eskin, A., Margulis, G. and Mozes, S. (2005) Quadratic forms of signature (2, 2) and eigenvalue spacings on rectangular 2-tori. *Ann. Math.* **161** 679–725.

[5] Heath-Brown, D. R. (2010) Pair correlation for fractional parts of αn^2, Math. *Proc. Cambridge Philos. Soc.* **148** 385–407.

[6] Kuipers, L. and Niederreiter, H. (2006) *Uniform distribution of sequences*. Dover Publications, Mineola, NY

[7] Rudnick, Z. and Sarnak, P. (1998) The pair correlation function of fractional parts of polynomials. *Comm. Math. Phys.* **194**(1) 61–70.

[8] Rudnick, Z. and Zaharescu, A. (1999) A metric result on the pair correlation of fractional parts of sequences,. *Acta Arith.* **89**(3) 283–293.

[9] Rudnick, Z. and Zaharescu, A. (2002) The distribution of spacings between fractional parts of lacunary sequences. *Forum Math.* **14**(5) 691–712.

[10] Sarnak, P. (1997) Values at integers of binary quadratic forms. Harmonic analysis and number theory (Montreal 1996), 181-203, CMS Conf. Providence, RI.: Amer. Math. Soc., Proc. 21.

[11] Steinerberger, S. (2018) Poissonian pair correlation and discrepancy,. *Indag. Math.* **29** 1167–1178.

[12] Walker, A. (2018) The Primes are not metric Poissonian. *Mathematika* **64**, 230–236.

Cite this article: Altman D and Chase Z (2025). On the smallest gap in a sequence with Poisson pair correlations. *Combinatorics, Probability and Computing* **34**, 283–297. https://doi.org/10.1017/S0963548324000361