

A METRIC THEORY OF MINIMAL GAPS

ZÉÉV RUDNICK

Abstract. We study the minimal gap statistic for fractional parts of sequences of the form $\mathcal{A}^\alpha = \{\alpha a(n)\}$, where $\mathcal{A} = \{a(n)\}$ is a sequence of distinct integers. Assuming that the additive energy of the sequence is close to its minimal possible value, we show that for almost all α , the minimal gap $\delta_{\min}^\alpha(N) = \min\{\alpha a(m) - \alpha a(n) \bmod 1 : 1 \leq m \neq n \leq N\}$ is close to that of a random sequence.

We start with a sequence of points $\mathcal{X} = \{x_n : n = 1, 2, \dots\} \subset \mathbb{R}/\mathbb{Z}$ in the unit interval/circle, which we assume is asymptotically uniformly distributed: for any subinterval $I \subset \mathbb{R}/\mathbb{Z}$, we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \#\{n \leq N : x_n \in I\} = |I|. \quad (1)$$

In particular, the mean spacing between the points lying in any subinterval is $1/N$. Our goal is to understand the *minimal gap*

$$\delta_{\min}(\mathcal{X}, N) = \min(|x_n - x_m| : n, m \leq N, n \neq m)$$

(with a suitable modification for wrapping around).

For *random* points, namely N independent uniform points in the unit interval (Poisson process), the minimal gap is almost surely of size $1/N^2$ [9]. In this note we study the metric theory of the minimal gap statistic for a class of deterministic sequences of fractional parts, such as fractional parts of polynomials. The case of quadratic polynomials $x_n = \alpha n^2$ has its roots in the recent paper [3], which studies the more complicated case of the minimal gap statistic for the sequence of eigenvalues of the Laplacian on a rectangular billiard, namely the points $\{\alpha m^2 + n^2 : m, n \geq 1\}$ on the real line.

We fix a sequence $\mathcal{A} = \{a(n) : n = 1, 2, \dots\} \subset \mathbb{Z}$ of distinct integers ($a(n) \neq a(m)$ if $m \neq n$), and study the minimal gap statistic of fractional parts of the set

$$\mathcal{A}^\alpha = \{\alpha a(n) \bmod 1 : n = 1, 2, \dots\} \subset \mathbb{R}/\mathbb{Z}.$$

(it is an old result of Weyl that \mathcal{A}^α satisfies (1) for almost all α). We want to know under which conditions we can show that for *almost all* α , the minimal gap statistics

$$\delta_{\min}^\alpha(N) = \delta_{\min}(\mathcal{A}^\alpha, N)$$

follows that of the random case, that is of size about $1/N^2$ for almost all α . It is easy to see that we cannot have much smaller minimal gaps.

Received 18 October 2017.
MSC (2010): 11K06 (primary).

THEOREM 1. *Assume that \mathcal{A} consists of distinct integers. Then for all $\eta > 0$, for almost all α ,*

$$\delta_{\min}^\alpha(N) > \frac{1}{N^{2+\eta}} \text{ for all } N > N_0(\alpha).$$

To make the minimal gap small, we give a criterion in terms of the “additive energy” $E(\mathcal{A}, N)$ of the sequence

$$E(\mathcal{A}, N) := \#\{(n_1, n_2, n_3, n_4) \in [1, N]^4 : a(n_1) + a(n_2) = a(n_3) + a(n_4)\}.$$

Note that $N^2 \leq E(\mathcal{A}, N) \leq N^3$. The result is the following.

THEOREM 2. *Assume that \mathcal{A} consists of distinct integers, and that the additive energy satisfies*

$$E(\mathcal{A}, N) \ll N^{2+o(1)} \text{ for all } N \gg 1.$$

Then for all $\eta > 0$, for almost all α ,

$$\delta_{\min}^\alpha(N) < \frac{1}{N^{2-\eta}} \text{ for all } N > N_0(\alpha).$$

Examples: for $a(n) = n^d$, $d \geq 2$, it is shown in [11] that the $E(\mathcal{A}, N) \ll N^{2+o(1)}$. For lacunary sequences, we have $E(\mathcal{A}, N) \ll N^2$ [12]. Hence, Theorem 2 applies to these sequences.

Relaxing the required bound on the additive energy will give a weaker result on the minimal spacing, basically that $\delta_{\min}^\alpha(N) < E(\mathcal{A}, N)/N^{4-\eta}$ almost surely. For this to be non-trivial, we need the additive energy to be no bigger than $E(\mathcal{A}, N) \ll N^{3-\eta'}$ for some $\eta' > 0$. A notable case where the additive energy is bigger is that of $\mathcal{A} = \mathcal{P}$ being the sequence of primes, where $E(\mathcal{P}, N) \approx N^3/\log N$. In this case, we cannot have gaps much larger than the average gap: a simple argument shows that given any $\varepsilon > 0$, for almost all α , we have $\delta_{\min}(\mathcal{P}^\alpha, N) \gg 1/(N(\log N)^{2+\varepsilon})$; see §3.

§1. *A bilinear statistic.* To study the minimal gap, we introduce statistics counting all possible gaps: we start with a smooth, compactly supported window function $f \in C_c^\infty([-1/2, 1/2])$, which is non-negative: $f \geq 0$, and of unit mass $\int f(x) dx = 1$, and define

$$F_M(x) = \sum_{j \in \mathbb{Z}} f(M(x + j)),$$

which is localized on the scale of $1/M$, and periodic: $F_M(x + 1) = F_M(x)$. We then set

$$D_{\mathcal{A}}(N, M)(\alpha) = \sum_{\substack{1 \leq m, n \leq N \\ m \neq n}} F_M(\alpha a(n) - \alpha a(m)).$$

The expected value of $D_{\mathcal{A}}(N, M)$ is easily seen to equal

$$\int_0^1 D_{\mathcal{A}}(N, M)(\alpha) d\alpha = \frac{N(N - 1)}{M} \sim \frac{N^2}{M}. \tag{2}$$

This already suffices to show that minimal gaps cannot typically be small (Theorem 1); see §2.

We will bound the variance of $D_{\mathcal{A}}(N, M)$, from which Theorem 2 will follow.

PROPOSITION 3.

$$\text{Var } D_{\mathcal{A}}(N, M) \ll \frac{1}{M} N^\epsilon E(\mathcal{A}, N).$$

The statistic $D_{\mathcal{A}}(N, M)(\alpha)$ is related to the pair correlation function of the sequence \mathcal{A}^α , which in our notation is $D_{\mathcal{A}}(N, N)(\alpha)/N$. Pair correlation measures gaps on the scale of the mean spacing, assumed here to be $1/N$, corresponding to $M = N$, here we are looking at much smaller scales of M close to N^2 .

The metric theory of the pair correlation function of fractional parts was initiated in [11], where the sequences $a(n) = \alpha n^d$ were shown to almost surely have Poissonian pair correlation for $d \geq 2$ (see [7, 10] for different proofs of the quadratic case $d = 2$). The problem has since been studied in several other cases and has recently been revived in an abstract setting [1, 4, 8, 14]. In particular, a convenient criterion for almost sure Poisson pair correlation has been formulated by Aistleitner, Larcher and Lewko [1] in terms of the additive energy $E(\mathcal{A}, N)$ of the sequence. The proof of Proposition 3 is close to that of the analogous statement for the pair correlation function in [1], which in turn is based on [11, 12].

Proof. The Fourier expansion of $F_M(x)$ is

$$F_M(x) = \sum_{k \in \mathbb{Z}} \frac{1}{M} \widehat{f}\left(\frac{k}{M}\right) e(kx), \tag{3}$$

where $\widehat{f}(y) = \int_{-\infty}^{\infty} f(x) e^{-2\pi ixy} dx$. Inserting into the definition of $D(N, M)$ gives

$$D(N, M)(\alpha) = \sum_{k \in \mathbb{Z}} \frac{1}{M} \widehat{f}\left(\frac{k}{M}\right) \sum_{\substack{1 \leq m, n \leq N \\ m \neq n}} e(k\alpha(a(m) - a(n))). \tag{4}$$

Integrating over α gives the expected value (we assume $a(m) \neq a(n)$ if $m \neq n$)

$$\int_0^1 D(N, M)(\alpha) d\alpha = \widehat{f}(0) \frac{1}{M} N(N - 1) = \frac{N(N - 1)}{M}.$$

The variance is the second moment of the sum over non-zero frequencies:

$$\text{Var } D(N, M) = \int_0^1 \left| \sum_{0 \neq k \in \mathbb{Z}} \frac{1}{M} \widehat{f}\left(\frac{k}{M}\right) \sum_{\substack{1 \leq m, n \leq N \\ m \neq n}} e(k\alpha(a(m) - a(n))) \right|^2 d\alpha.$$

Squaring out and integrating gives

$$\begin{aligned} \text{Var } D(N, M) &= \sum_{k_1, k_2 \neq 0} \frac{1}{M^2} \widehat{f}\left(\frac{k_1}{M}\right) \widehat{f}\left(\frac{k_2}{M}\right) \\ &\quad \times \#\{m_1 \neq n_1, m_2 \neq n_2 : k_1(a(m_1) - a(n_1)) = k_2(a(m_2) - a(n_2))\}. \end{aligned}$$

We now follow [1, Lemma 3] to convert this to “greatest common divisor (GCD) sums”. Let

$$R(v) = \#\{1 \leq m \neq n \leq N : a(m) - a(n) = v\}.$$

Then

$$\text{Var } D(N, M) = \sum_{v_1, v_2 \neq 0} \sum_{k_1, k_2 \neq 0} c(k_1)c(k_2)R(v_1)R(v_2)\delta(k_1v_1 = k_2v_2),$$

where we set

$$c(k) := \frac{1}{M} \widehat{f}\left(\frac{k}{M}\right).$$

LEMMA 4. Let $f \in C_c^\infty(\mathbb{R})$. For any non-zero integers $v_1, v_2 \neq 0$,

$$\sum_{k_1, k_2 \neq 0} c(k_1)c(k_2)\delta(k_1v_1 = k_2v_2) \ll_f \frac{1}{M} \frac{\text{gcd}(v_1, v_2)}{\sqrt{|v_1v_2|}}. \tag{5}$$

Proof. For $k_1, k_2 \neq 0$, we have $k_1v_1 = k_2v_2$ if and only if

$$(k_1, k_2) = \ell \left(\frac{v_2}{\text{gcd}(v_1, v_2)}, \frac{v_1}{\text{gcd}(v_1, v_2)} \right)$$

for some non-zero integer $0 \neq \ell \in \mathbb{Z}$. Abbreviating $a_i = v_i/(M \text{gcd}(v_1, v_2))$, we find

$$\begin{aligned} \sum_{k_1, k_2 \neq 0} c(k_1)c(k_2)\delta(k_1v_1 = k_2v_2) &= \sum_{0 \neq \ell \in \mathbb{Z}} c\left(\ell \frac{v_2}{\text{gcd}(v_1, v_2)}\right) c\left(\ell \frac{v_1}{\text{gcd}(v_1, v_2)}\right) \\ &= \frac{1}{M^2} \sum_{0 \neq \ell \in \mathbb{Z}} \widehat{f}(a_2\ell) \widehat{f}(a_1\ell) \end{aligned}$$

so that it suffices to show that

$$\sum_{0 \neq \ell \in \mathbb{Z}} \widehat{f}(a_1\ell) \widehat{f}(a_2\ell) \ll_f \frac{1}{\sqrt{|a_1a_2|}} = \frac{M \text{gcd}(v_1, v_2)}{\sqrt{|v_1v_2|}}. \tag{6}$$

Applying Cauchy–Schwarz we get

$$\sum_{0 \neq \ell \in \mathbb{Z}} \widehat{f}(\ell a_1) \widehat{f}(\ell a_2) \leq \left(\sum_{\ell \neq 0} \widehat{f}(a_1 \ell)^2 \right)^{1/2} \left(\sum_{\ell \neq 0} \widehat{f}(a_2 \ell)^2 \right)^{1/2}.$$

We will obtain (6) if we show that for any $a > 0$,

$$\sum_{\ell \neq 0} \widehat{f}(a\ell)^2 \ll_f \frac{1}{a}.$$

Indeed, if $0 < a \ll 1$ then we get a Riemann sum for $(\widehat{f})^2$:

$$\sum_{\ell \neq 0} \widehat{f}(a\ell)^2 \sim \frac{1}{a} \int_{-\infty}^{\infty} \widehat{f}(y)^2 dy = \frac{1}{a} \int_{-\infty}^{\infty} f(x)^2 dx.$$

If $a \gg 1$ then use the decay rate of the Fourier transform: for $y \neq 0$,

$$|\widehat{f}(y)| \leq \frac{1}{2\pi|y|} \int_{-\infty}^{\infty} |f'(x)| dx,$$

to obtain

$$\sum_{\ell \neq 0} \widehat{f}(a\ell)^2 \ll \sum_{\ell \neq 0} \left(\frac{\int_{-\infty}^{\infty} |f'(x)| dx}{|a\ell|} \right)^2 \ll_f \frac{1}{a^2},$$

which for $a \gg 1$ is $\ll 1/a$.

Hence,

$$\text{Var } D(N, M) \ll \frac{1}{M} \sum_{v_1, v_2 \neq 0} R(v_1) R(v_2) \frac{\text{gcd}(v_1, v_2)}{\sqrt{|v_1 v_2|}}.$$

According to the GCD bounds of [6],

$$\sum_{v_1, v_2 \neq 0} R(v_1) R(v_2) \frac{\text{gcd}(v_1, v_2)}{\sqrt{|v_1 v_2|}} \ll \exp\left(\frac{10 \log N}{\log \log N}\right) \sum_v R(v)^2$$

(see [5] for an essentially optimal refinement). Now

$$\sum R(v)^2 = \#\{m_i, n_j \leq N, m_1 \neq n_1, m_2 \neq n_2 : a(m_1) - a(n_1) = a(m_2) - a(n_2)\}$$

is at most the additive energy $E(\mathcal{A}, N)$. Thus,

$$\text{Var } D(N, M) \ll \frac{1}{M} N^\epsilon E(\mathcal{A}, N)$$

as claimed. □

COROLLARY 5. Assume that the additive energy satisfies $E(\mathcal{A}, N) < N^{2+o(1)}$. If $M < N^{2-\eta}$, then for almost all α ,

$$D_{\mathcal{A}}(N, M)(\alpha) \sim \frac{N^2}{M}.$$

Proof. Take $N_k = \lfloor k^{4/\eta} \rfloor$, so that $\sum_k N_k^{-\eta/2} < \infty$, and pick any $M_k < N_k^{2-\eta}$, then we find from Proposition 3,

$$\begin{aligned} \sum_k \int_0^1 \left| \frac{D(N_k, M_k)(\alpha)}{N_k(N_k - 1)/M_k} - 1 \right|^2 d\alpha &= \sum_k \frac{\text{Var } D(N_k, M_k)}{(N_k(N_k - 1)/M_k)^2} \\ &< \sum_k N_k^{o(1)} \frac{E(N_k)M_k}{N_k^4} \ll \sum_k \frac{1}{N_k^{\eta/2}} < \infty, \end{aligned}$$

and so for almost all α ,

$$D(N_k, M_k)(\alpha) \sim \frac{N_k^2}{M_k} \quad \text{for all } k > k_0(\alpha). \tag{7}$$

A priori the set depends on the test function f , but that can be taken care of by a standard diagonalization procedure; for our purposes we only need one test function.

Given $N \gg 1$, there is a unique value of k so that $N_k \leq N < N_{k+1}$. Note that $N/N_k = 1 + O(N^{-\eta/4})$. Since $M < N^{2-\eta} \sim N_k^{2-\eta} < N_{k+1}^{2-\eta}$, we know from (7) that almost surely $D(N_k, M)/(N_k^2/M) \rightarrow 1$.

Note that

$$D(N_k, M) \leq D(N, M) \leq D(N_{k+1}, M).$$

This is because the sums $D(N, M)$ consist of non-negative terms, and hence,

$$\begin{aligned} D(N, M) &= \sum_{1 \leq m \neq n \leq N} F_M(\alpha(a(m) - a(n))) \\ &\geq \sum_{1 \leq m \neq n \leq N_k} F_M(\alpha(a(m) - a(n))) = D(N_k, M) \end{aligned}$$

(we dropped all pairs (m, n) where $\max(m, n) > N_k$).

Since $N/N_k = 1 + O(N^{-\eta/4})$, we have

$$\frac{D(N_k, M)}{N_k^2/M} \leq \frac{D(N, M)}{N^2/M(1 + O(N^{-\eta/4}))} = \frac{D(N, M)}{N^2/M} (1 + O(N^{-\eta/4}))$$

and likewise

$$\frac{D(N, M)}{N^2/M} \leq \frac{D(N_{k+1}, M)}{N_{k+1}^2/M} (1 + O(N^{-\eta/4})).$$

Since we know that almost surely $D(N_k, M)/(N_k^2/M) \rightarrow 1$, we deduce that almost surely also $D(N, M)/(N^2/M) \rightarrow 1$. □

COROLLARY 6. *Theorem 2 holds.*

Proof. Fix $\eta > 0$, and let $M = 1/2N^{2-\eta}$. Since $D(N, M)(\alpha) \sim N^2/M > 1$ by Corollary 5, we have a gap of size at most $1/(2M) = 1/N^{2-\eta}$, that is $\delta_{\min}^\alpha(N) < 1/N^{2-\eta}$ almost surely. □

§2. *Lower bounds: proof of Theorem 1.* We take any sequence of integers $\mathcal{A} = \{a(n)\}$ with distinct elements. We want to show that for any $\eta > 0$, almost surely,

$$\delta^\alpha(N) > 1/N^{2+\eta} \quad \text{for all } N > N_0(\alpha).$$

Let $N_k = \lfloor k \rfloor^{2/\eta}$. We claim that it suffices to show that, for almost all α ,

$$\delta_{\min}^\alpha(N_k) > 2/N_k^{2+\eta} \quad \text{for all } k > k_0(\alpha). \tag{8}$$

Indeed, note that if $N_k \leq N < N_{k+1}$ then $\delta_{\min}^\alpha(N) \geq \delta_{\min}^\alpha(N_{k+1})$. Since $N_{k+1} \sim N$, by (8) we have, for almost all α

$$\delta_{\min}^\alpha(N) \geq \delta_{\min}^\alpha(N_{k+1}) > 2/N_{k+1}^{2+\eta} > 1/N^{2+\eta}$$

for $N > N_0(\alpha)$.

To prove (8), it suffices, by the Borel–Cantelli lemma, to show that

$$\sum_k \text{Prob}(\delta_{\min}^\alpha(N_k) \leq 2/N_k^{2+\eta}) < \infty. \tag{9}$$

In the definition of $D(N, M) = D_f(N, M)$, choose f so that $f(x) \geq 1$ if $|x| \leq 1/4$ (and in addition, $f \geq 0$ is non-negative, $\int_{-\infty}^\infty f(x) dx = 1$, f is smooth and supported in $[-1/2, 1/2]$). Now note that for such f , if $D_f(N, M) < 1$ then $\delta_{\min}^\alpha(N) > 1/(4M)$. This is because $D_f(N, M)$ is a sum of non-negative terms, and if there is one gap of size $\leq 1/(4M)$ then the corresponding term $F_M(\alpha(a(m) - a(n))) = \sum_j f(M(\alpha(a(m) - a(n)) + j)) \geq 1$ by the choice of f , so that $D_f(N, M) \geq 1$. Thus, we find that

$$\delta_{\min}^\alpha(N) \leq \frac{1}{4M} \quad \Rightarrow \quad D_f(N, M) \geq 1,$$

and hence,

$$\text{Prob}\left(\delta_{\min}^\alpha(N) \leq \frac{1}{4M}\right) \leq \text{Prob}(D_f(N, M) \geq 1). \tag{10}$$

Now since $D_f \geq 0$,

$$\text{Prob}(D_f(N, M) \geq 1) \leq \int_0^1 D_f(N, M)(\alpha) d\alpha$$

so that by (2), for $M_k = \frac{1}{8}N_k^{2+\eta}$,

$$\int_0^1 D(N_k, M_k)(\alpha) d\alpha \sim 8N_k^{-\eta} \ll \frac{1}{k^2},$$

which together with (10) proves (9), and hence, (8). This proves Theorem 1.

§3. *Minimal gaps for the primes.* Let $a(n) = p_n$, the n th prime. By Khinchin’s theorem, for all $\varepsilon > 0$ there is a set of full measure of the α so that $\|q\alpha\| \geq 1/(q(\log q)^{1+\varepsilon})$ for any integer $q \geq q_0(\alpha)$. In particular, for such α , the gap between fractional parts of $\alpha p_n \bmod 1$ are

$$\|\alpha(p_m - p_n)\| \gg \frac{1}{|p_m - p_n|(\log |p_m - p_n|)^{1+\varepsilon}} \geq \frac{1}{N(\log N)^{2+\varepsilon}},$$

since $|p_m - p_n| \leq p_N \sim N \log N$ for $m < n \leq N$. Hence, for such α , the minimal gap satisfies $\delta_{\min}^\alpha(N) > 1/N(\log N)^{2+\varepsilon}$.

A similar argument applies to other dense cases, such as the sequence of square-free integers. An extreme case is that when $\mathcal{A} = \mathbb{N}$ is the sequence of all natural numbers. The argument above gives the minimal gap here is, for almost all α , at least $\delta_{\min}^\alpha(N) \gg 1/(N(\log N)^{1+\varepsilon})$. Note that in this case the “three-gap” theorem shows that there are at most three distinct gaps between the fractional parts $\{\alpha n \bmod 1 : n \leq N\}$. Concerning other “dense” sequences, it is known that for any sequence of integers $\mathcal{A} \subset [1, M]$, the fractional parts $\alpha a(m) \bmod 1$ have at most $O(\sqrt{M})$ distinct gaps [2, 13].

Acknowledgements. We thank Andrew Granville and Niclas Technau for their comments. The work was supported by the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 320755.

References

1. C. Aistleitner, G. Larcher and M. Lewko, Additive energy and the Hausdorff dimension of the exceptional set in metric pair correlation problems, with an appendix by J. Bourgain. *Israel J. Math.* **222**(1) (2017), 463–485.
2. A. Balog, A. Granville and J. Solymosi, Gaps between fractional parts, and additive combinatorics. *Q. J. Math.* **68** (2017), 1–12.
3. V. Blomer, J. Bourgain, M. Radziwiłł and Z. Rudnick, Small gaps in the spectrum of the rectangular billiard. *Ann. Sci. Éc. Norm. Supér.* (4) **50**(5) (2017), 1283–1300.
4. T. F. Bloom, S. Chow, A. Gafni and A. Walker, Additive energy and the metric Poissonian property. *Preprint*, 2017, [arXiv:1709.02634](https://arxiv.org/abs/1709.02634) [math.NT].
5. A. Bondarenko and K. Seip, GCD sums and complete sets of square-free numbers. *Bull. Lond. Math. Soc.* **47**(1) (2015), 29–41.
6. T. Dyer and G. Harman, Sums involving common divisors. *J. Lond. Math. Soc.* **34** (1986), 1–11.
7. D. R. Heath-Brown, Pair correlation for fractional parts of an^2 . *Math. Proc. Cambridge Philos. Soc.* **148**(3) (2010), 385–407.
8. T. Lachmann and N. Technau, On exceptional sets in the metric Poissonian pair correlations problem, *Preprint*, 2017, [arXiv:1708.08599](https://arxiv.org/abs/1708.08599).
9. P. Lévy, Sur la division d’un segment par des points choisis au hasard. *C. R. Acad. Sci. Paris* **208** (1939), 147–149.

10. J. Marklof and A. Strömbergsson, Equidistribution of Kronecker sequences along closed horocycles. *Geom. Funct. Anal.* **13** (2003), 1239–1280.
11. Z. Rudnick and P. Sarnak, The pair correlation function of fractional parts of polynomials. *Comm. Math. Phys.* **194**(1) (1998), 61–70.
12. Z. Rudnick and A. Zaharescu, A metric result on the pair correlation of fractional parts of sequences. *Acta Arith.* **89**(3) (1999), 283–293.
13. M. Vâjâitu and A. Zaharescu, Distinct gaps between fractional parts of sequences. *Proc. Amer. Math. Soc.* **130** (2002), 3447–3452.
14. A. Walker, The primes are not metric Poissonian. *Mathematika* **64**(1) (2018), 230–236.

Zeév Rudnick,
Raymond and Beverly Sackler School
of Mathematical Sciences,
Tel Aviv University,
Tel Aviv 69978,
Israel
E-mail: rudnick@post.tau.ac.il