Gap distribution of $\sqrt{n}$ modulo 1.

(Joint work with Niclas Technau)

Gap distribution

Given a sequence u_n with $n < N$ we reduce this sequence modulo 1.

- We then obtain N points in the interval $[0, 1]$
- We order the points in increasing order, and call them

$$u_{1,N} < u_{2,N} < \ldots < u_{N,N}$$

- Therefore $u_{i,N}$ is just a permutation of $u_n \pmod 1$.

Gap distribution

Question

Does the gap distribution of $u_n \pmod 1$ exist? In other words, given $0 < \alpha < \beta$, does

$$\frac{1}{N}\#\Big\{n < N : N(u_{i+1,N} - u_{i,N}) \in (\alpha, \beta)\Big\}$$

converge? (where $u_{i,N}$ is an order of $u_n \pmod 1$ with $n < N$ in increasing order)

Poisson gap distribution

- What should we expect generically for the gap distribution?
- In a generic situation we can pretend that the points $u_n \pmod 1$ are random variables X_n uniformly distributed in $(0, 1)$.
- In that situation the probability of spacings being greater than t/N would converge to

$$\int_t^\infty e^{-u} du$$

- In that case we say that the gap distribution is *Poisson*.

The Elkies-McMullen result.

A priori, the following conjecture makes perfect sense.

Conjecture

Let β be positive and non-integer. Then the gap distribution of the sequence $n^\beta \pmod 1$ exists and is Poisson.

The Elkies-McMullen result.

- As it turns out this common sense conjecture is false when $\beta = \frac{1}{2}$.
- Perversely, $\beta = \frac{1}{2}$ is also the only exponent for which the gap distribution is known to exist.
- In all other cases the gap distribution is still conjectured to be Poisson (and the rest of this talk will provide good reason why we believe this).

The Elkies-McMullen result

Theorem

Consider the sequence $\sqrt{n} \pmod 1$ *with* $n < N$ *sorted in order as*

$$u_{1,N} < u_{2,N} < \ldots < u_{N,N}.$$

Then for each t*, the limit*

$$\frac{1}{N}\#\Big\{n < N : N(u_{i+1,N} - u_{i,N}) > t\Big\}$$

exists, and decays like $\asymp t^{-3}$*.*

The Elkies-McMullen result

The proof of Elkies-McMullen uses techniques from dynamics, specifically Ratner's theorem.

- We would like to understand what is the actual mechanism that makes the gap distribution non-Poissonian. It's hard to believe that Ratner's theorem or the equidistribution of horocycle flow is central to this.

Correlations

- Before launching an attack on this problem we make a simple observation. Instead of proving that the gap distribution exists, it's enough to show instead that the so-called void statistic exists.
- The void statistic is the probability that an interval $(x, x + t/N)$ with x taken uniform at random in $(0, 1)$ contains no element u_k.
- In other words, it's enough to show the convergence of

$$\frac{1}{N}\int_0^1 \mathbf{1}\Big(\sum_{x<u_n<x+t/N} 1 = 0\Big)dx$$

as $N \to \infty$.

Correlations

- The moments associated to the void-statistic are,

$$\int_0^1 \Big(\sum_{x < u_n < x + t/N} 1 \Big)^k dx$$

- As is well known convergence of moments implies convergence of the probability distribution.

A mindless attempt

- A first mindless attempt at proving the Elkies-McMullen theorem would attempt to compute straight away the moments of the void statistic (or equivalently correlation functions).
- This is doomed to fail for two reasons.

Why it fails

- First, Elkies-McMullen claim that the gap distribution has heavy tails, i.e the probability of gaps $> t/N$ is $\asymp t^{-3}$.
- Therefore we expect already the third correlation to be divergent.

Why it fails

- Second, even if we ignore this warning and attempt to compute,

$$\int_0^1 \Big(\sum_{x-t/N \le u_n \le x+t/N} 1 \Big)^k dx$$

  we notice another source of divergence.

In our case $u_n := \sqrt{n} \pmod 1$.

- As it turns $u_{k^2} = \sqrt{k^2} \pmod 1 = 0$.
- Therefore in a $O(1/N)$ neighborhood of zero we always have a cluster of at least $\sqrt{N}$ values.

Why it fails

- This means that the kth moment will grow at least like $N^{k/2-1}$. So it blows up as soon as $k > 2$.
- One attempt to remediate this problem is to subtract the contribution of the squares.
- This was done by El-Baz, Markloff and Vinogradov, who suspiciously find that the pair correlation is then Poisson. But the third correlation still blows up.

Why it fails

- This situations leads one to suspect that perhaps such blow ups happen at every point x that is close to a fraction with small denominators.
- Even though we didn't quite manage to construct such examples this was a useful intuition.

Understanding

- By a theorem of Dirichlet that every $\alpha \in [0,1]$ has an approximation of the form

$$\left|\alpha - \frac{a}{q}\right| \leq \frac{1}{qQ} \ , \ q < Q$$

- Let's choose $Q := \Delta\sqrt{N}$ with Δ some fixed constant.
- The motivation for this choice is simple: We want access to neighborhoods that are smaller than $1/N$, so we need $Q > \sqrt{N}$. Since we are optimistic we pick it just a little bit larger.
- We then notice that our optimism pays off: the majority of real numbers fit in the cases with $\sqrt{N} < q < \Delta\sqrt{N}$ of Dirichlet's theorem.
- In the circle method parlance we call this area "minor arcs", and the complements we call "major arcs".

Change of measure

Since the majority of real numbers are in the minor arcs

$$\bigcup_{\substack{(a,q)=1 \\ \sqrt{N} \le q \le \Delta\sqrt{N}}} \left(\frac{a}{q} - \frac{1}{\Delta N}, \frac{a}{q} + \frac{1}{\Delta N}\right)$$

it's actually enough to restrict the void statistic,

$$\sum_{x < u_n < x + t/N} 1$$

to those x that belong to the minor arcs.

Approximation

Pick now

$$x \in \left(\frac{a}{q} - \frac{1}{\Delta N}, \frac{a}{q} + \frac{1}{\Delta N}\right)$$

- For such x we have,

$$\sum_{x < u_n < x + t/N} 1 = \sum_{a/q < u_n < a/q + t/N + O(1/\Delta N)} 1$$

- It's not too difficult to believe that the perturbation by $1/\Delta N$ won't matter in the long run especially if we let Δ go to infinity later.
- So let's pretend that,

$$\sum_{x < u_n < x + t/N} 1 \approx \sum_{a/q < u_n < a/q + t/N} 1$$

Change of measure

Thus it's enough to understand

$$\sum_{\substack{(a,q)=1 \\ \sqrt{N}\le q\le \Delta\sqrt{N}}} \Big(\sum_{a/q<u_n<a/q+t/N} 1 \Big)^k$$

- This has some fighting chance, as now we avoid all the regions where the blow-ups happen.
- For example 0 does not belong to the minor arcs; there is no $(a, q) = 1$ and $\sqrt{N} < q < \Delta\sqrt{N}$ such that,

$$\left|0 - \frac{a}{q}\right| \le \frac{1}{\Delta N}$$

Jutila's approximation

- In reality we do not have to use Dirichlet's theorem. If we only care about approximating most real numbers, it was observed by Jutila that it's enough to just have a large enough set of denominators (regardless of what they are).
- Jutila's approach allows us to restrict our minor arcs to be

$$\bigcup_{\substack{(a,q)=1 \\ q\in\mathcal{Q}}} \left(\frac{a}{q} - \frac{1}{\Delta N}, \frac{a}{q} + \frac{1}{\Delta N}\right)$$

  no matter what $\mathcal{Q} \subset [\Delta\sqrt{N}, 2\Delta\sqrt{N}]$ is, as long as it's dense enough (say $\gg N/\log\Delta$ elements).

Moments

- Thus instead of computing,

$$\int_0^1 \Big(\sum_{x < u_n < x + t/N} 1 \Big)^k dx$$

for each $k \in \mathbb{N}$ which is divergent, we have managed to reduce ourselves to the problem of computing,

$$\frac{1}{N} \sum_{\substack{(a,q)=1 \\ q \in \mathcal{Q}}} \Big(\sum_{a/q < u_n < a/q + t/N} 1 \Big)^k$$

for any set $\mathcal{Q} \subset [\Delta\sqrt{N}, 2\Delta\sqrt{N}]$ that is dense enough (say $\gg N/\log \Delta$ elements).

A formula

It's not clear that this is much of a gain unless we have a way of computing,

$$\sum_{a/q < u_n < a/q + t/N} 1$$

- As it turns out that's exactly the case.

A formula

Recall that $u_n = \sqrt{n} \pmod 1$. We now expand the condition

$$\frac{a}{q} < \sqrt{n} \pmod 1 < \frac{a}{q} + \frac{t}{N}$$

using Fourier series.

- We find that

$$\sum_{\substack{n<N \\ a/q<\{\sqrt{n}\}<a/q+t/N}} \approx t + \frac{1}{N} \sum_{\ell \sim N} \sum_{n \leq N} e\Big(\ell\Big(\sqrt{n} - \frac{a}{q}\Big)\Big).$$

Transforming

- We now transform this approximation by applying Poisson summation twice.
- Executing Poisson summation in n we get

$$t + \frac{1}{N^{3/4}} \sum_{\ell \sim N} \sum_{v \sim \sqrt{N}} e\Big(\frac{\ell^2}{4v} - \frac{\ell a}{q}\Big)$$

- The sum over ℓ is now almost a complete exponential sum. Executing Poisson summation in ℓ and computing the Gauss sum leads us to ...

Transforming

the following main formula:

$$\sum_{\substack{n<N \\ a/q \le \{\sqrt{n}\} < a/q+t/N}} \approx t + \sum_{\substack{2va \equiv u \pmod q \\ v \le \sqrt{N}, |u| \le \Delta}} e\Big(-\frac{\overline{q^2}u^2}{4v}\Big)$$

where $\overline{x}$ denotes the inverse of x modulo $4v$.

- Notice that the formula contains on average only $O(1)$ terms. This already explains why $\sqrt{n}$ will be special and non-Poisson.
- Here we get a dual sum of length $O(\Delta)$ a factor of q from the congruence condition and $\sqrt{v} \approx N^{1/4}$ from the normalization of the Gauss sum.

Transforming

- In fact it's worth reflecting on the miracle that happened here because it's the key as to why $\sqrt{n} \pmod 1$ is special. Two things happened here:
- First of all, after the first Poisson we get exponential sums that can become complete exponential sums modulo an integer. Other exponents where this can happen are $n^{1-1/k}$ with $k \geq 2$ integers. For all other exponents we get exponential sums that cannot be related to anything arithmetic.
- The second miracle that happens, is that after applying the second Poisson the dual length becomes essentially $O(1)$. Out of the exponents $n^{1-1/k}$ with $k \geq 2$ integer, this only happens for $k = 2$. This explain why the exponent $\frac{1}{2}$ is so special.

Transforming

- In effect we also achieved something else here: We have shown that if x is in the minor arcs, then we have an essentially $O(1)$ time algorithm to determine the number of $\sqrt{n} \pmod 1$ with $n < N$ lying in the interval $(x, x + t/N)$.
- Thus the remainder of the proof consists in analyzing this "algorithm".

Computing moments.

- The problem is therefore reduced to computing the moments,

$$\sum_{\substack{(a,q)=1\\ q\in\mathcal{Q}}}\Big(\sum_{\substack{2va\equiv u \pmod q\\ v\leq\sqrt{N},|u|\leq\Delta}} e\Big(-\frac{\overline{q^2}u^2}{4v}\Big)\Big)^k$$

- We will show that these moments grow basically like Δ^{k-3}. This is consistent with the decay rate of the gaps being $\asymp t^{-3}$.
- It also leads us to the following interpretation: Restricting to minor arcs of width $1/(\Delta N)$ is morally equivalent to working with truncated void statistics of the form $X\mathbf{1}_{|X|\leq\Delta}$ where X is the void statistic.

Computing moments.

- Expanding everything out, computing the moments is equivalent to obtaining a power-saving in

$$\sum_{\substack{q\in\mathcal{Q},|v_i|\le\sqrt{N}\\ v_j u_i\equiv u_j v_i \pmod q}} e\Big(-\sum_{i=1}^{k}\frac{\overline{q}^2 u_i^2}{4v_i}\Big)$$

- Unfortunately this looks hard! The reason is that the combined modulus is $4v_1\dots v_k \asymp N^{k/2}$, while the total length of summation that we have at our disposal is $\asymp N$.

Computing moments

- Therefore another miracle needs to happen in the transformation of this phase.
- We use the fact that the variables u_i and v_j are entangled through $v_j u_i \equiv u_j v_i \pmod{q}$.
- Using a few rounds of the Chinese remainder theorem in the form
$$\frac{\overline{a}}{b} + \frac{\overline{b}}{a} \equiv \frac{1}{ab} \pmod{1}$$
and the deep fact that
$$\overline{1 + qa} \equiv 1 - qa \pmod{q^2}$$
we arrive at a massive simplification ...

Computing moments

we find that,

$$\sum_{i=1}^{k} \frac{\overline{q}^2 u_i^2}{4v_i} \equiv \frac{u_1 \overline{4v_1} \sum_{i=1}^{k} u_i}{q^2} - \frac{u_1 \overline{4v_1^2} \sum_{i=1}^{k} \ell_i}{q} + \mathcal{E} \pmod 1$$

where $\mathcal{E} \ll 1/N$ is negligible and $u_1 v_i = v_1 u_i + q\ell_i$ with $\ell_i \ll \Delta$.

- Therefore the entire problem reduces to bounding non-trivially exponential sums of the form

$$\sum_{v \sim \sqrt{N}} e\Big(\frac{\kappa \overline{4v}}{q^2} - \frac{\eta \overline{4v^2}}{q}\Big).$$

- Still difficult! The issue is that this is exactly a little bit past the so-called Polya-Vingoradov range: if we apply Poisson summation we get a dual sum of length $q^2/\sqrt{N} = \Delta\sqrt{N}$.

Using the moduli

- This is where we use the freedom to choose the moduli $q \in \mathcal{Q}$.
- We choose $\mathcal{Q}$ to be a set of moduli in $[\Delta\sqrt{N}, 2\Delta\sqrt{N}]$ that factor as ab with $a \sim Q^{\delta}$ and $b \sim Q^{1-\delta}$ with $Q = \Delta\sqrt{N}$.
- The point is that now we can use Weyl differencing with shift a multiple of a. This leads to a conductor drop, after which Poisson summation wins the game.

Conductor drop

- The idea of the conductor drop is that if we do usual Weyl differencing and then apply Poisson to estimate,

$$\sum_{n<N} \Big| \sum_{h<H} e\Big(\frac{(n+h)^2}{q}\Big)\Big|^2$$

then the dual sum in Poisson is q/N.

- But if $q = ab$ and we apply Weyl differencing with shift ha we get,

$$\sum_{n<N} \Big| \sum_{h<H} e\Big(\frac{(n+ah)^2}{ab}\Big)\Big|^2$$

we get after Poisson a dual sum of length b/N (instead of ab/N) this is shorter!

Concluding the proof.

Let us recapitulate what we have done.

- We wish to show the existence of

$$\mathcal{F}_N(t) := \mathbb{P}_N\Big(\sum_{x<u_n<x+t/N} 1 = 0\Big)$$

  when $N \to \infty$.

- We have shown that for each $\Delta > 10$ there exists an auxiliary probability measure, call it $\mathcal{F}_{\Delta,N}(t)$ such that,

$$\Big|\mathcal{F}_N(t) - \mathcal{F}_{\Delta,N}(t)\Big| \leq \frac{1}{\Delta^{1/4}}$$

- Moreover we can show that each of these auxiliary measures $\mathcal{F}_{\Delta,N}$ converges to a limit $\mathcal{F}_\Delta(t)$ as N goes to infinity. Incidentally $\mathcal{F}_\Delta(t)$ is supported in $|t| \leq \Delta$.

Concluding the proof

- Thus letting N to infinity we have show that

$$\left|\limsup_{N\to\infty} \mathcal{F}_N(t) - \mathcal{F}_\Delta(t)\right| \leq \frac{1}{\Delta^{1/4}}$$

and similarly

$$\left|\liminf_{N\to\infty} \mathcal{F}_N(t) - \mathcal{F}_\Delta(t)\right| \leq \frac{1}{\Delta^{1/4}}$$

- So we have shown

$$\left|\limsup_{N\to\infty} \mathcal{F}_N(t) - \liminf_{N\to\infty} \mathcal{F}_N(t)\right| \leq \frac{1}{\Delta^{1/4}}$$

- Letting $\Delta \to \infty$ the existence of the gap distribution follows.