Primes and almost primes in short intervals: Statistics over $\mathbb{Z}$ and $\mathbb{F}_q[T]$

Brad Rodgers

Department of Mathematics
University of Zürich

Statistics and Number Theory
CRM Montréal 2014

Primes at mean spacing

Theorem (Prime number theorem)

$$\pi(X) = |\mathcal{P} \cap [0, X]| \sim \frac{X}{\log X}$$

Primes at mean spacing

Theorem (Prime number theorem)

$$\pi(X) = |\mathcal{P} \cap [0, X]| \sim \frac{X}{\log X}$$

$$\Updownarrow$$

On average, $p_{n+1} - p_n \sim \log p_n$.

$$\Updownarrow$$

$$\mathbb{E}_{n \approx X} \left| \mathcal{P} \cap [n, n + \log n] \right| \sim 1.$$

Primes at mean spacing

Theorem (Prime number theorem)

$$\pi(X) = |\mathcal{P} \cap [0, X]| \sim \frac{X}{\log X}$$

$$\Updownarrow$$

On average, $p_{n+1} - p_n \sim \log p_n$.

$$\Updownarrow$$

$$\mathbb{E}_{n \approx X} \left| \mathcal{P} \cap [n, n + \log n] \right| \sim 1.$$

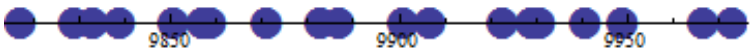
Conjecture (Distribution of primes at mean spacing)

For $I_\lambda := [n, n + \lambda \log n]$,

$$\mathbb{P}_{n \approx X}(\#\{p \in I_\lambda\} = k) := \frac{1}{X} \sum_{n \leq X} \mathbf{1}[\#\{p \in I_\lambda\} = k\}] \sim \frac{e^{-\lambda} \lambda^k}{k!}.$$

Poisson processes

Meaning of the conjecture: in a short interval $[n, n + \log n]$, the primes look like a Poisson process:



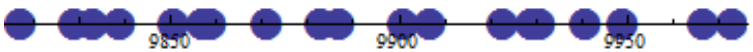
Primes

Poisson

Poisson processes are placements of random points that don't 'see' each other.

Poisson processes

Meaning of the conjecture: in a short interval $[n, n + \log n]$, the primes look like a Poisson process:



Primes

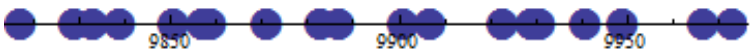
Poisson

Poisson processes are placements of random points that don't 'see' each other. Some examples:

- The timing of calls made into a telethon.
- The timing of horse-kicks to soldiers in the Prussian army.

Poisson processes

Meaning of the conjecture: in a short interval $[n, n + \log n]$, the primes look like a Poisson process:



Primes

Poisson

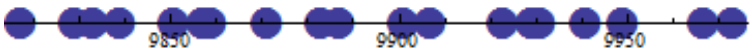
Poisson processes are placements of random points that don't 'see' each other. Some examples:

- The timing of calls made into a telethon.
- The timing of horse-kicks to soldiers in the Prussian army.
- Prime factors of a random number n have a distribution closely related to a Poisson process. In particular (Landau):

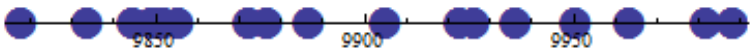
$$\frac{\pi_k(X)}{X} = \mathbb{P}_{n\approx X}(\Omega(n) = k) \sim e^{-\log\log X}\frac{(\log\log X)^{k-1}}{(k-1)!}.$$

Poisson processes

Meaning of the conjecture: in a short interval $[n, n + \log n]$, the primes look like a Poisson process:



Primes



Poisson

Poisson processes are placements of random points that don't 'see' each other. Some examples:

- The timing of calls made into a telethon.
- The timing of horse-kicks to soldiers in the Prussian army.
- Prime factors of a random number n have a distribution closely related to a Poisson process. In particular (Landau):

$$\frac{\pi_k(X)}{X} = \mathbb{P}_{n\approx X}(\Omega(n) = k) \sim e^{-\log\log X}\frac{(\log\log X)^{k-1}}{(k-1)!}.$$

Moreover (Erdős-Kac):

$$\mathbb{P}_{n\approx X}\left(\frac{\Omega(n) - \log\log X}{\sqrt{\log\log X}} \in [\alpha, \beta]\right) \sim \int_\alpha^\beta \frac{e^{-t^2/2}}{\sqrt{2\pi}}\, dt.$$

(Very) Short intervals

Conjecture (Distribution of primes in very short intervals)

$$\mathbb{P}_{n\approx X}\left(\frac{\pi(n+H)-\pi(n)-H/\log X}{\sqrt{H/\log X}}\in[\alpha,\beta]\right)\sim\int_{\alpha}^{\beta}\frac{e^{-t^2/2}}{\sqrt{2\pi}}\,dt$$

for H growing such that

$$H/\log X\to\infty,\quad \text{yet}\quad \log H/\log X\to 0.$$

(Very) Short intervals

Conjecture (Distribution of primes in very short intervals)

$$\mathbb{P}_{n\approx X}\left(\frac{\psi(n+H)-\psi(n)-H}{\sqrt{H\log X}}\in[\alpha,\beta]\right)\sim\int_{\alpha}^{\beta}\frac{e^{-t^2/2}}{\sqrt{2\pi}}\,dt$$

for H growing such that

$$H/\log X\to\infty,\quad \text{yet}\quad \log H/\log X\to 0.$$

(Very) Short intervals

Conjecture (Distribution of primes in very short intervals)

$$\mathbb{P}_{n\approx X}\left(\frac{\psi(n+H)-\psi(n)-H}{\sqrt{H\log X}}\in[\alpha,\beta]\right)\sim\int_{\alpha}^{\beta}\frac{e^{-t^2/2}}{\sqrt{2\pi}}\,dt$$

for H growing such that

$$H/\log X\to\infty,\quad \text{yet}\quad \log H/\log X\to 0.$$

Cramer Model:

*For a random integer n, think of the event "n is prime" as the flip of a biased coin, that turns up '***prime***' $\frac{1}{\log n}$ of the time, and '***not prime***' otherwise*

(Very) Short intervals

Conjecture (Distribution of primes in very short intervals)

$$\mathbb{P}_{n\approx X}\left(\frac{\psi(n+H)-\psi(n)-H}{\sqrt{H\log X}}\in[\alpha,\beta]\right)\sim\int_{\alpha}^{\beta}\frac{e^{-t^2/2}}{\sqrt{2\pi}}\,dt$$

for H growing such that

$$H/\log X\to\infty,\quad \text{yet}\quad \log H/\log X\to 0.$$

Cramer Model:

*For a random integer n, think of the event "n is prime" as the flip of a biased coin, that turns up '**prime**' $\frac{1}{\log n}$ of the time, and '**not prime**' otherwise*

- The Cramer model is consistent with the conjecture above.
- Why $\log H/\log X\to 0$? **The Cramer model would not require this.**

Exceptions to the Cramer model

- The Cramer model does not suggest any $\log H/\log X \to 0$ cutoff for the distribution of primes in very short intervals. But based in part on Hardy-Littlewood heuristics:

Conjecture (Goldston-Montgomery)

For $\epsilon > 0$ and $X^\epsilon < H < X^{1-\epsilon}$

$$\frac{1}{X}\int_0^X (\psi(n+H)-\psi(n)-H)^2\,dn \sim H(\log X - \log H).$$

Conjecture (Montgomery-Soundararajan)

For random $n \approx X$,

$$\frac{\psi(n+H)-\psi(n)-H}{\sqrt{H(\log X - \log H)}} \Rightarrow N(0,1).$$

for $X^\epsilon < H < X^{1-\epsilon}$.

The Riemann zeta function

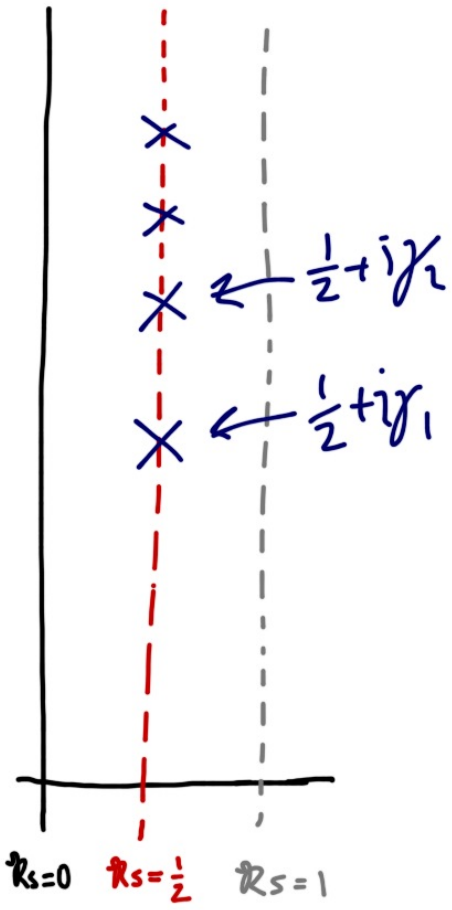


Figure: Zeros of $\zeta(s)$.

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad \Re(s) > 1.$$

- Label non-trivial zeros by $1/2 + i\gamma$.
- Riemann hypothesis: γ is always real.

The Riemann zeta function

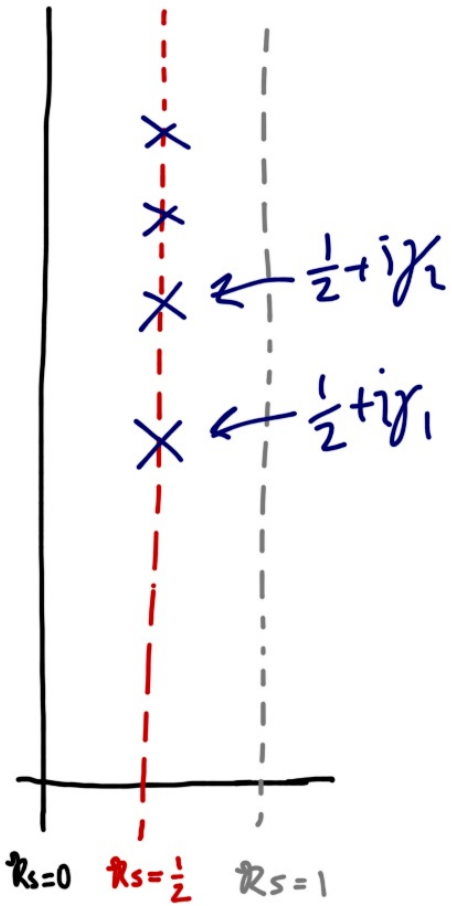


Figure: Zeros of $\zeta(s)$.

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad \Re(s) > 1.$$

- Label non-trivial zeros by $1/2 + i\gamma$.
- Riemann hypothesis: γ is always real.
- Riemann hypothesis is equivalent to the statement:

$$\psi(x) := \sum_{n \leq x} \Lambda(n) = x + O(x^{1/2+\epsilon}).$$

Spacing of zeros

Theorem (Riemann - von Mangoldt)

$$N(T) = \#\{\gamma \in (0, T),\ \zeta(\tfrac{1}{2} + i\gamma) = 0\}$$
$$= \tfrac{T}{2\pi} \log(\tfrac{T}{2\pi}) - \tfrac{T}{2\pi} + O(\log T)$$

- The likelihood of finding a zero $\gamma \in [T, T + dx]$ is roughly $\log T/2\pi\, dx$.

Spacing of zeros

Theorem (Riemann - von Mangoldt)

$$N(T) = \#\{\gamma \in (0, T),\ \zeta(\tfrac{1}{2} + i\gamma) = 0\}$$
$$= \tfrac{T}{2\pi}\log(\tfrac{T}{2\pi}) - \tfrac{T}{2\pi} + O(\log T)$$

- The likelihood of finding a zero $\gamma \in [T, T + dx]$ is roughly $\log T/2\pi\, dx$.

Conjecture (Montgomery's pair correlation conjecture)

For any nice test function η,

$$\frac{1}{N(T)} \sum_{\substack{0 \le \gamma, \gamma' \le T \\ \gamma \neq \gamma'}} \eta\Big(\frac{\log T}{2\pi}(\gamma - \gamma')\Big) \sim \int \eta(x)\Big[1 - \Big(\frac{\sin \pi x}{\pi x}\Big)^2\Big]\, dx$$

- Given a zero γ around height T, the likelihood of finding another zero

$$\gamma' \in [\gamma + \tfrac{2\pi x}{\log T}, \gamma + \tfrac{2\pi x}{\log T} + dx]$$

is roughly

$$\Big[1 - \Big(\frac{\sin \pi x}{\pi x}\Big)^2\Big]\, dx.$$

A histogram of the pair correlation conjecture

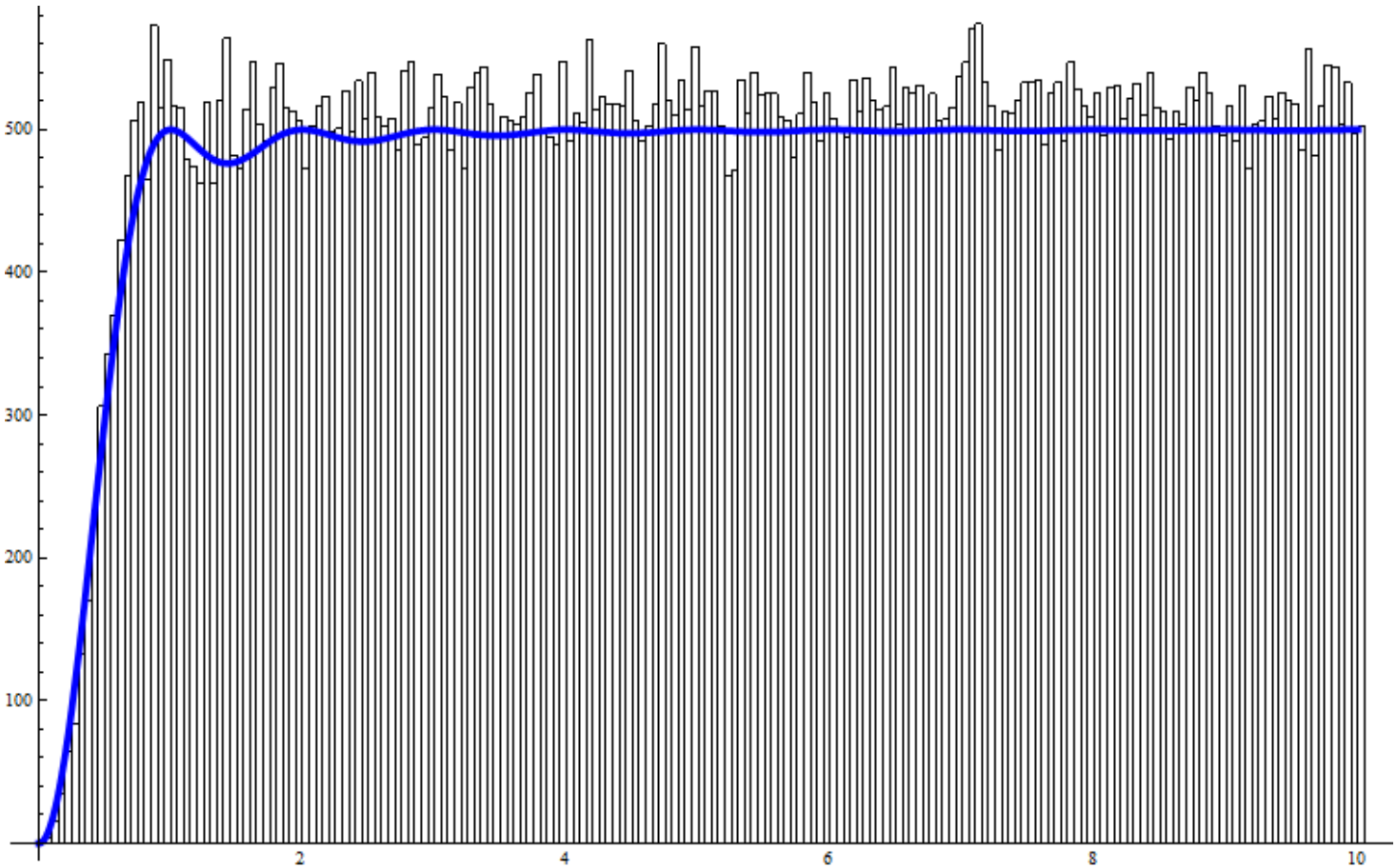


Figure: A histogram of $\frac{\log T}{2\pi}(\gamma - \gamma')$ for the first 10000 zeros, in intervals of size .05, compared to the appropriately scaled prediction $1 - (\frac{\sin \pi x}{\pi x})^2$.

Random matrix statistics

The Institute for Advanced Study

April 7 1972

Dear Atle

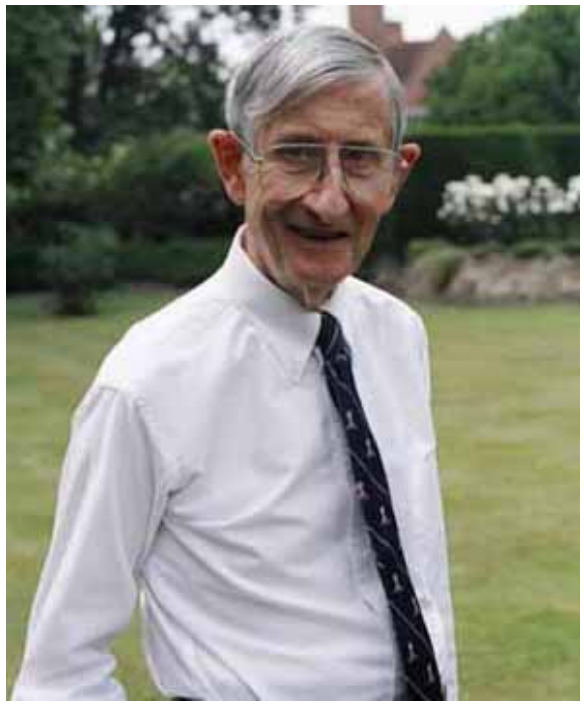
The reference which Dr Montgomery wants is

M. L. Mehta, "Random Matrices"
Academic Press, N.Y. 1967.

Page 76 Equation 6.13
Page 113 Equation 7.61

Showing that the pair-correlation function of zeros of the ζ-function is identical ~~with~~ with that of eigenvalues of a random complex (Hermitian or unitary) matrix of large order.

Freeman Dyson.



This same spacing is seen between the eigenvalues of random matrices!

The GUE Conjecture

- $U(N) := \{g \in \mathcal{M}_{N\times N}(\mathbb{C}) : gg^* = I\}$ with Haar measure.
- Label the eigenvalues of $U(N)$ by

$$\{e^{i2\pi\theta_1}, ..., e^{i2\pi\theta_N}\}, \quad \theta_i \in [-1/2, 1/2)$$

The GUE Conjecture

- $U(N) := \{g \in \mathcal{M}_{N\times N}(\mathbb{C}) : gg^* = I\}$ with Haar measure.
- Label the eigenvalues of $U(N)$ by

$$\{e^{i2\pi\theta_1}, ..., e^{i2\pi\theta_N}\}, \quad \theta_i \in [-1/2, 1/2)$$

- N eigenangles in $[-1/2, 1/2]$, so they have density N.

The GUE Conjecture

- $U(N) := \{g \in \mathcal{M}_{N\times N}(\mathbb{C}) : gg^* = I\}$ with Haar measure.
- Label the eigenvalues of $U(N)$ by

$$\{e^{i2\pi\theta_1}, ..., e^{i2\pi\theta_N}\}, \quad \theta_i \in [-1/2, 1/2)$$

- N eigenangles in $[-1/2, 1/2]$, so they have density N.
- Then the collection of randomly translated and stretched out zeta zeros:

$$\{\frac{\log T}{2\pi}(\gamma - t)\}$$

look like the collection of stretched out eigenangles

$$\{N\theta_i\}$$

of a random unitary matrix from $U(N)$.

Primes in short intervals

Recall, defining,

$$\widetilde{\psi}(x;H) = \psi(x+H) - \psi(x) - H,$$

Conjecture (Goldston-Montgomery)

For $1 \leq H \leq X^{1-\epsilon}$,

$$\frac{1}{X}\int_1^X \widetilde{\psi}(x;H)^2\,dx \sim H\big(\log X - \log H\big).$$

Primes in short intervals

Recall, defining,

$$\widetilde{\psi}(x;H) = \psi(x+H) - \psi(x) - H,$$

Conjecture (Goldston-Montgomery)

For $1 \leq H \leq X^{1-\epsilon}$,

$$\frac{1}{X}\int_1^X \widetilde{\psi}(x;H)^2\,dx \sim H\big(\log X - \log H\big).$$

- Gallagher and Mueller showed that the pair correlation conjecture as before implies an estimate of this sort, with weaker summability method.

Primes in short intervals

Recall, defining,

$$\widetilde{\psi}(x;H) = \psi(x+H) - \psi(x) - H,$$

Conjecture (Goldston-Montgomery)

For $1 \leq H \leq X^{1-\epsilon}$,

$$\frac{1}{X}\int_1^X \widetilde{\psi}(x;H)^2\,dx \sim H\big(\log X - \log H\big).$$

- Gallagher and Mueller showed that the pair correlation conjecture as before implies an estimate of this sort, with weaker summability method.
- Goldston and Montgomery showed that a slight strengthening of the pair correlation conjecture is exactly equivalent to this conjecture.

Primes in short intervals

Recall, defining,

$$\widetilde{\psi}(x;H) = \psi(x+H) - \psi(x) - H,$$

Conjecture (Goldston-Montgomery)

For $1 \leq H \leq X^{1-\epsilon}$,

$$\frac{1}{X}\int_1^X \widetilde{\psi}(x;H)^2\,dx \sim H\big(\log X - \log H\big).$$

- Gallagher and Mueller showed that the pair correlation conjecture as before implies an estimate of this sort, with weaker summability method.
- Goldston and Montgomery showed that a slight strengthening of the pair correlation conjecture is exactly equivalent to this conjecture.

Rigorously (from prime number theorem):

$$\frac{1}{X}\sum_{n\leq X}\Lambda(n)^2 \sim \log X.$$

Ideas of the Gallagher-Mueller/Goldston-Montgomery proof:

- **The assumption of the pair correlation conjecture.**
- **The explicit formula:**

$$\Lambda(n) \approx 1 - n^{-1/2} \sum_{\gamma} n^{i\gamma}$$

Ideas of the Gallagher-Mueller/Goldston-Montgomery proof:

- **The assumption of the pair correlation conjecture.**
- **The explicit formula:**

$$\Lambda(n)-1 \approx -n^{-1/2}\sum_{\gamma} n^{i\gamma} \quad \Rightarrow \quad (\Lambda(n)-1)n^{-it} \approx -n^{-1/2}\underbrace{\sum_{\gamma} e\Big(\frac{\log n}{2\pi}(\gamma - t)\Big)}_{\approx \mathrm{Tr}(g^r)}.$$

Ideas of the Gallagher-Mueller/Goldston-Montgomery proof:

- **The assumption of the pair correlation conjecture.**
- **The explicit formula:**

$$\Lambda(n)-1 \approx -n^{-1/2}\sum_{\gamma} n^{i\gamma} \quad \Rightarrow \quad (\Lambda(n)-1)n^{-it} \approx -n^{-1/2}\underbrace{\sum_{\gamma} e\Big(\frac{\log n}{2\pi}(\gamma - t)\Big)}_{\approx \mathrm{Tr}(g^r)}.$$

- **Localization via harmonic analysis:**

$$\frac{1}{T}\int_0^T \Big(\frac{n}{m}\Big)^{it} dt \approx \mathbf{1}[n-m = O(m/T)].$$

Ideas of the Gallagher-Mueller/Goldston-Montgomery proof:

- **The assumption of the pair correlation conjecture.**
- **The explicit formula:**

$$\Lambda(n)-1 \approx -n^{-1/2}\sum_{\gamma} n^{i\gamma} \quad \Rightarrow \quad (\Lambda(n)-1)n^{-it} \approx -n^{-1/2}\underbrace{\sum_{\gamma} e\Big(\frac{\log n}{2\pi}(\gamma - t)\Big)}_{\approx \mathrm{Tr}(g^r)}.$$

- **Localization via harmonic analysis:**

$$\frac{1}{T}\int_0^T \Big(\frac{n}{m}\Big)^{it} dt \approx \mathbf{1}[n-m = O(m/T)].$$

With these tools:

$$\frac{1}{X}\int_0^X \tilde{\psi}(x;\, H)^2\, dx \approx \Big(\frac{H}{X}\Big)^2 \int_0^{X/H} \Big|\sum_{n\le X}(\Lambda(n)-1)n^{-it}\Big|^2 dt$$

$$\approx H\int_{U(N)} |\mathrm{Tr}(g^r)|^2\, dg \approx H(\log X - \log H),$$

For $N \approx \log(X/H)$ and $r \approx \log X$.

Higher correlations?

- Bogomolny and Keating showed (heuristically) that the GUE Conjecture is equivalent to estimating the likelihood that

$$p_1 \cdots p_j \approx q_1 \cdots q_k,$$

where p_i, q_i are primes drawn from a certain region.

Higher correlations?

- Bogomolny and Keating showed (heuristically) that the GUE Conjecture is equivalent to estimating the likelihood that

$$p_1 \cdots p_j \approx q_1 \cdots q_k,$$

  where p_i, q_i are primes drawn from a certain region.
- This may be done more rigorously: Farmer-Gonek-Lee-Lester; R.

Higher correlations?

- Bogomolny and Keating showed (heuristically) that the GUE Conjecture is equivalent to estimating the likelihood that

$$p_1 \cdots p_j \approx q_1 \cdots q_k,$$

  where p_i, q_i are primes drawn from a certain region.
- This may be done more rigorously: Farmer-Gonek-Lee-Lester; R.
- But: the resulting formulas estimating likelihood still are not 'nice' or combinatorially transparent.

Almost-primes

- Almost primes of order k: those natural numbers $n = p_1^{e_1} \cdots p_\ell^{e_\ell}$ with $\ell \leq k$.

Almost-primes

- Almost primes of order k: those natural numbers $n = p_1^{e_1} \cdots p_\ell^{e_\ell}$ with $\ell \leq k$.
- It is natural to weight almost primes by higher order von Mangoldt functions $\Lambda_k(n)$, defined inductively by:

$$\Lambda_1(n) := \Lambda(n),$$

$$\Lambda_k(n) := \sum_{d\delta=n} \Lambda_{k-1}(d)\Lambda(\delta) + \log(n)\Lambda_{k-1}(n).$$

Almost-primes

- Almost primes of order k: those natural numbers $n = p_1^{e_1} \cdots p_\ell^{e_\ell}$ with $\ell \leq k$.
- It is natural to weight almost primes by higher order von Mangoldt functions $\Lambda_k(n)$, defined inductively by:

$$\Lambda_1(n) := \Lambda(n),$$

$$\Lambda_k(n) := \sum_{d\delta=n} \Lambda_{k-1}(d)\Lambda(\delta) + \log(n)\Lambda_{k-1}(n).$$

- Λ_k is supported on almost primes of order k. Famously used by Selberg and Erdős in an elementary proof of the prime number theorem, with $k = 2$.

Almost-primes

- Almost primes of order k: those natural numbers $n = p_1^{e_1} \cdots p_\ell^{e_\ell}$ with $\ell \leq k$.
- It is natural to weight almost primes by higher order von Mangoldt functions $\Lambda_k(n)$, defined inductively by:

$$\Lambda_1(n) := \Lambda(n),$$

$$\Lambda_k(n) := \sum_{d\delta=n} \Lambda_{k-1}(d)\Lambda(\delta) + \log(n)\Lambda_{k-1}(n).$$

- Λ_k is supported on almost primes of order k. Famously used by Selberg and Erdős in an elementary proof of the prime number theorem, with $k = 2$.
- By the prime number theorem:

$$\psi_k(x) := \sum_{n\leq x} \Lambda_k(n) = xP_{k-1}(\log x) + o(x)$$

$$\sim kx \log^{k-1} x.$$

Almost-primes in short intervals

Define

$$\widetilde{\psi}_k(x) := \psi_k(x) - xP_{k-1}(\log x).$$
$$\widetilde{\psi}_k(x; H) := \widetilde{\psi}_k(x + H) - \widetilde{\psi}_k(x).$$

(This is a count of almost primes in a short interval, minus a regular approximation.)

Almost-primes in short intervals

Define

$$\widetilde{\psi}_k(x) := \psi_k(x) - xP_{k-1}(\log x).$$
$$\widetilde{\psi}_k(x; H) := \widetilde{\psi}_k(x+H) - \widetilde{\psi}_k(x).$$

(This is a count of almost primes in a short interval, minus a regular approximation.)

Conjecture (R.)

For $1 \le H \le X^{1-\epsilon}$,

$$\frac{1}{X}\int_1^X \widetilde{\psi}_j(x; H)\,\widetilde{\psi}_k(x; H)\,dx \sim \frac{jk}{j+k-1} H\big(\log X - \log H\big)^{j+k-1}.$$

Almost-primes in short intervals

Define

$$\widetilde{\psi}_k(x) := \psi_k(x) - xP_{k-1}(\log x).$$
$$\widetilde{\psi}_k(x;H) := \widetilde{\psi}_k(x+H) - \widetilde{\psi}_k(x).$$

(This is a count of almost primes in a short interval, minus a regular approximation.)

Conjecture (R.)

For $1 \le H \le X^{1-\epsilon}$,

$$\frac{1}{X}\int_1^X \widetilde{\psi}_j(x;H)\,\widetilde{\psi}_k(x;H)\,dx \sim \frac{jk}{j+k-1}H\big(\log X - \log H\big)^{j+k-1}.$$

- With a weaker summability method, this conjecture can be shown to follow from a GUE Conjecture for the zeta zeros.

Almost-primes in short intervals

Define

$$\widetilde{\psi}_k(x) := \psi_k(x) - xP_{k-1}(\log x).$$
$$\widetilde{\psi}_k(x;H) := \widetilde{\psi}_k(x+H) - \widetilde{\psi}_k(x).$$

(This is a count of almost primes in a short interval, minus a regular approximation.)

Conjecture (R.)

For $1 \le H \le X^{1-\epsilon}$,

$$\frac{1}{X}\int_1^X \widetilde{\psi}_j(x;H)\,\widetilde{\psi}_k(x;H)\,dx \sim \frac{jk}{j+k-1}H\big(\log X - \log H\big)^{j+k-1}.$$

- With a weaker summability method, this conjecture can be shown to follow from a GUE Conjecture for the zeta zeros.

Rigorously (from the prime number theorem):

$$\frac{1}{X}\sum_{n\le X}\Lambda_j(n)\Lambda_k(n) \sim \frac{jk}{j+k-1}\log^{j+k-1} X.$$

$\mathbb{Z}$ to $\mathbb{F}_q[T]$

One may ask similar questions over $\mathbb{F}_q[T]$, and often one can answer them when $q \to \infty$.

$\mathbb{Z}$ to $\mathbb{F}_q[T]$

One may ask similar questions over $\mathbb{F}_q[T]$, and often one can answer them when $q \to \infty$.

A dictionary:

$$\begin{array}{rcl} \text{integers } \mathbb{Z} & \longleftrightarrow & \text{polynomials } \mathbb{F}_q[T] \\ \text{positive integers } \mathbb{N} & \longleftrightarrow & \mathbb{F}_q[T] \text{ monic} \end{array}$$

$\mathbb{Z}$ to $\mathbb{F}_q[T]$

One may ask similar questions over $\mathbb{F}_q[T]$, and often one can answer them when $q \to \infty$.

A dictionary:

$$
\begin{array}{rcl}
\text{integers } \mathbb{Z} & \longleftrightarrow & \text{polynomials } \mathbb{F}_q[T] \\
\text{positive integers } \mathbb{N} & \longleftrightarrow & \mathbb{F}_q[T] \text{ monic} \\
 & . & \\
\log n & \longleftrightarrow & \deg(f) \\
|n| & \longleftrightarrow & q^{\deg(f)} \\
\{n : \ x \leq n \leq x + H\} & \longleftrightarrow & \{g \in \mathbb{F}_q[T] : \ \deg(f - g) \leq h\}
\end{array}
$$

$\mathbb{Z}$ to $\mathbb{F}_q[T]$

One may ask similar questions over $\mathbb{F}_q[T]$, and often one can answer them when $q \to \infty$.

A dictionary:

$$
\begin{aligned}
\text{integers } \mathbb{Z} &\longleftrightarrow \text{polynomials } \mathbb{F}_q[T] \\
\text{positive integers } \mathbb{N} &\longleftrightarrow \mathbb{F}_q[T] \text{ monic} \\
&\cdot \\
\log n &\longleftrightarrow \deg(f) \\
|n| &\longleftrightarrow q^{\deg(f)} \\
\{n:\ x \le n \le x+H\} &\longleftrightarrow \{g \in \mathbb{F}_q[T]:\ \deg(f-g) \le h\} \\
&\cdot \\
\text{primes} &\longleftrightarrow \text{irreducible polynomials} \\
\Lambda(n) &\longleftrightarrow \Lambda(f) := \begin{cases} \deg P, & \text{if } f = cP^k, \\ 0, & \text{otherwise.} \end{cases} \\
\sum_{d_1 d_2 = n} &\longleftrightarrow \sum_{\substack{g_1 g_2 = f \\ f_1, f_2 \text{ monic}}}
\end{aligned}
$$

Statistics in short intervals

- Prime number theorem:

$$\frac{1}{q^n} \sum_{\substack{\deg(f)=n \\ \text{monic}}} \Lambda(f) = 1$$

- Almost-prime number theorem:

$$\frac{1}{q^n} \sum_{\substack{\deg(f)=n \\ \text{monic}}} \Lambda_k(f) = n^k - (n-1)^k$$

Statistics in short intervals

- Prime number theorem:

$$\frac{1}{q^n} \sum_{\substack{\deg(f)=n \\ \text{monic}}} \Lambda(f) = 1$$

- Almost-prime number theorem:

$$\frac{1}{q^n} \sum_{\substack{\deg(f)=n \\ \text{monic}}} \Lambda_k(f) = n^k - (n-1)^k$$

- Define

$$\widetilde{\Psi}_k(f,h) = \sum_{\substack{g:\, \deg(f-g)\le h \\ \text{monic}}} \Big(\Lambda_k(g) - [n^k - (n-1)^k]\Big).$$

Statistics in short intervals

Theorem (Keating-Rudnick)

For $0 \leq h \leq n-4$, *as* $q \to \infty$,

$$\frac{1}{q^n} \sum_{\substack{\deg(f)=n \\ \text{monic}}} \widetilde{\Psi}(f;h)^2 \sim q^{h+1}(n-h-2).$$

Statistics in short intervals

Theorem (Keating-Rudnick)

For $0 \leq h \leq n-4$, as $q \to \infty$,

$$\frac{1}{q^n} \sum_{\substack{\deg(f)=n \\ \text{monic}}} \widetilde{\Psi}(f;h)^2 \sim q^{h+1}(n-h-2).$$

Theorem (R.)

For $0 \leq h \leq n-4$, as $q \to \infty$,

$$\frac{1}{q^n} \sum_{\substack{\deg(f)=n \\ \text{monic}}} \widetilde{\Psi}_j(f;h)\widetilde{\Psi}_k(f;h)$$

$$\sim q^{h+1} \sum_{d=1}^{n-h-2} (d^j-(d-1)^j)(d^k-(d-1)^k).$$

Statistics in short intervals

Theorem (Keating-Rudnick)

For $0 \leq h \leq n-4$, as $q \to \infty$,

$$\frac{1}{q^n} \sum_{\substack{\deg(f)=n \\ \text{monic}}} \widetilde{\Psi}(f;h)^2 \sim q^{h+1}(n-h-2).$$

Theorem (R.)

For $0 \leq h \leq n-4$, as $q \to \infty$,

$$\frac{1}{q^n} \sum_{\substack{\deg(f)=n \\ \text{monic}}} \widetilde{\Psi}_j(f;h)\widetilde{\Psi}_k(f;h)$$

$$\sim q^{h+1} \sum_{d=1}^{n-h-2} (d^j-(d-1)^j)(d^k-(d-1)^k).$$

Note:

$$X \leftrightarrow q^n, \quad H \leftrightarrow q^{h+1} \quad \log X \leftrightarrow n, \quad \log H \leftrightarrow h+1$$

and

$$\sum_{d=1}^{n-h-2} (d^j-(d-1)^j)(d^k-(d-1)^k) \approx \frac{jk}{j+k-1}(n-h)^{j+k-1}.$$

A naive idea: the zeta function of $\mathbb{F}_q[T]$

$$\mathcal{Z}(u) := \sum_{f \text{ monic}} u^{\deg(f)}$$

$$\left(= \sum_{f \text{ monic}} \frac{1}{|f|^s} \quad \text{for} \quad u = q^{-s} \right)$$

$$= \prod_{P \text{ monic}} \frac{1}{1 - u^{\deg P}}$$

$$= \frac{1}{1 - qu}$$

A better idea: Dirichlet characters over $\mathbb{F}_q[T]$

- Defined in exact analogy with $\mathbb{Z}$.
- For $Q \in \mathbb{F}_q[T]$, and $(g, Q) = (f, Q) = 1$,

$$\frac{1}{\Phi(Q)} \sum_{\chi \pmod{Q}} \overline{\chi(f)}\chi(g) = \begin{cases} 1, & \text{if } f \equiv g \pmod{Q} \\ 0, & \text{otherwise.} \end{cases}$$

A better idea: Dirichlet characters over $\mathbb{F}_q[T]$

- Defined in exact analogy with $\mathbb{Z}$.
- For $Q \in \mathbb{F}_q[T]$, and $(g, Q) = (f, Q) = 1$,

$$\frac{1}{\Phi(Q)} \sum_{\chi \pmod{Q}} \overline{\chi(f)}\chi(g) = \begin{cases} 1, & \text{if } f \equiv g \pmod{Q} \\ 0, & \text{otherwise.} \end{cases}$$

- L-functions are defined in analogously:

$$\begin{aligned} \mathcal{L}(u, \chi) :&= \sum_{f \text{ monic}} \chi(f) u^{\deg(f)} \\ &= \sum_{f \text{ monic}} \frac{\chi(f)}{|f|^s} \quad \text{for} \quad u := q^{-s} \end{aligned}$$

A better idea: Dirichlet characters over $\mathbb{F}_q[T]$

- Defined in exact analogy with $\mathbb{Z}$.
- For $Q \in \mathbb{F}_q[T]$, and $(g, Q) = (f, Q) = 1$,

$$\frac{1}{\Phi(Q)} \sum_{\chi \pmod{Q}} \overline{\chi(f)}\chi(g) = \begin{cases} 1, & \text{if } f \equiv g \pmod{Q} \\ 0, & \text{otherwise.} \end{cases}$$

- L-functions are defined in analogously:

$$\begin{aligned} \mathcal{L}(u, \chi) :&= \sum_{f \text{ monic}} \chi(f) u^{\deg(f)} \\ &= \sum_{f \text{ monic}} \frac{\chi(f)}{|f|^s} \quad \text{for} \quad u := q^{-s} \end{aligned}$$

- Here the Riemann Hypothesis is a theorem: All nontrivial zeros of $\mathcal{L}(u, \chi)$ lie on the circle $|u| = q^{-1/2}$ (or $\Re s = 1/2$).

Ideas of the proof

- **A passage to random matrices:** Let
$$\{q^{-1/2}e^{i2\pi\vartheta_1}, ..., q^{-1/2}e^{i2\pi\vartheta_N}\}$$
be the non-trivial zeros of $\mathcal{L}(u, \chi)$ and define a matrix,
$$\Theta_\chi := \mathrm{diag}(e^{i2\pi\vartheta_1}, ..., e^{i2\pi\vartheta_N}).$$
Roughly: *Averaging over χ, the matrix Θ_χ looks like a random unitary matrix.* (Katz)

Ideas of the proof

- **A passage to random matrices:** Let
$$\{q^{-1/2}e^{i2\pi\vartheta_1}, ..., q^{-1/2}e^{i2\pi\vartheta_N}\}$$
be the non-trivial zeros of $\mathcal{L}(u, \chi)$ and define a matrix,
$$\Theta_\chi := \mathrm{diag}(e^{i2\pi\vartheta_1}, ..., e^{i2\pi\vartheta_N}).$$
Roughly: *Averaging over χ, the matrix Θ_χ looks like a random unitary matrix.* (Katz)
- **Localization via harmonic analysis:** Replace the characters n^{-it} used for $\mathbb{Z}$ with $\chi(f)$. (An important role is played by an involution of Keating and Rudnick).

Ideas of the proof

- **A passage to random matrices:** Let
$$\{q^{-1/2}e^{i2\pi\vartheta_1}, ..., q^{-1/2}e^{i2\pi\vartheta_N}\}$$
be the non-trivial zeros of $\mathcal{L}(u, \chi)$ and define a matrix,
$$\Theta_\chi := \mathrm{diag}(e^{i2\pi\vartheta_1}, ..., e^{i2\pi\vartheta_N}).$$
Roughly: *Averaging over χ, the matrix Θ_χ looks like a random unitary matrix.* (Katz)
- **Localization via harmonic analysis:** Replace the characters n^{-it} used for $\mathbb{Z}$ with $\chi(f)$. (An important role is played by an involution of Keating and Rudnick).
- **The explicit formula:**

Theorem (Weil)

$$\sum_{\substack{\deg(f)=n \\ \text{monic}}} \Lambda(f)\chi(f) = -q^{n/2}\mathrm{Tr}\ \Theta_\chi^n + O(1),$$

A variant explicit formula

- An example:

$$\sum_{\substack{\deg(f)=n \\ \text{monic}}} \Lambda \star \Lambda(f)\chi(f) = \sum_{\substack{\deg(g_1)+\deg(g_2)=n \\ \text{monic}}} \Lambda(g_1)\Lambda(g_2)\chi(g_1 g_2)$$

$$= \sum_{\substack{\ell+\lambda=n \\ \ell,\lambda\geq 1}} \left(\sum_{\substack{\deg(g_1)=\ell \\ \deg(g_2)=\lambda \\ \text{monic}}} \Lambda(g_1)\Lambda(g_2)\chi(g_1)\chi(g_2) \right)$$

$$= q^{n/2} \sum_{\substack{\ell+\lambda=n \\ \ell,\lambda\geq 1}} (-\mathrm{Tr}\ \Theta_\chi^\ell)(-\mathrm{Tr}\ \Theta_\chi^\lambda) + O(q^{(n-1)/2}).$$

A variant explicit formula

- An example:

$$\sum_{\substack{\deg(f)=n \\ \text{monic}}} \Lambda \star \Lambda(f)\chi(f) = \sum_{\substack{\ell+\lambda=n \\ \ell,\lambda\geq 1}} \left(\sum_{\substack{\deg(g_1)=\ell \\ \deg(g_2)=\lambda \\ \text{monic}}} \Lambda(g_1)\Lambda(g_2)\chi(g_1)\chi(g_2) \right)$$

$$= q^{n/2} \sum_{\substack{\ell+\lambda=n \\ \ell,\lambda\geq 1}} (-\mathrm{Tr}\ \Theta_\chi^\ell)(-\mathrm{Tr}\ \Theta_\chi^\lambda) + O(q^{(n-1)/2}).$$

- Define inductively:

$$H_1^{(n)} := -\mathrm{Tr}(g^n),$$

$$H_j^{(n)} = \sum_{\substack{\ell+\lambda=n \\ \ell,\lambda\geq 1}} H_1^{(\lambda)} H_{j-1}^{(\ell)} + nH_{j-1}^{(n)}.$$

Lemma (A higher order explicit formula)

$$\sum_{\substack{\deg(f)=n \\ \text{monic}}} \Lambda_j(f)\chi(f) = q^{n/2} H_j^{(n)}(\Theta_\chi) + O(q^{(n-1)/2}).$$

A short interval dictionary

$$\begin{aligned}
\text{Arithmetic} &\longleftrightarrow \text{Random matrix} \\
\Lambda(m) &\longleftrightarrow -\mathrm{Tr}(g^r) \\
\Lambda_j(m) &\longleftrightarrow H_j^{(r)}(g) \\
\mu(m) &\longleftrightarrow h_j(g) = \sum_{i_1 \le i_2 \cdots \le i_k} \omega_{i_1}\omega_{i_2}\cdots\omega_{i_k}
\end{aligned}$$

(where $\{\omega_i\}$ are the eigenvalues of g)

A short interval dictionary

$$\begin{array}{rcl} \text{Arithmetic} & \longleftrightarrow & \text{Random matrix} \\ \Lambda(m) & \longleftrightarrow & -\mathrm{Tr}(g^r) \\ \Lambda_j(m) & \longleftrightarrow & H_j^{(r)}(g) \\ \mu(m) & \longleftrightarrow & h_j(g) = \sum_{i_1 \le i_2 \cdots \le i_k} \omega_{i_1}\omega_{i_2}\cdots\omega_{i_k} \end{array}$$

(where $\{\omega_i\}$ are the eigenvalues of g)

- $$\frac{1}{q^n} \sum_{\substack{\deg(f)=n \\ \text{monic}}} \tilde{\Psi}(f;h)^2 \sim q^{h+1} \int_{U(n-h-2)} |\mathrm{Tr}(g^n)|^2 \, dg$$

- $$\frac{1}{q^n} \sum_{\substack{\deg(f)=n \\ \text{monic}}} \tilde{\Psi}_j(f;h)\tilde{\Psi}_k(f;h) \sim q^{h+1} \int_{U(n-h-2)} \overline{H_j^{(r)}} H_k^{(r)} \, dg.$$

- Variance of Möbius sums in short intervals (Keating – Rudnick)
- Variance of divisor sums (Keating – Roditty-Gershon – Rudnick)

A short interval dictionary

$$\begin{aligned} \text{Arithmetic} &\longleftrightarrow \text{Random matrix} \\ \Lambda(m) &\longleftrightarrow -\mathrm{Tr}(g^r) \\ \Lambda_j(m) &\longleftrightarrow H_j^{(r)}(g) \\ \mu(m) &\longleftrightarrow h_r(g) = \sum_{i_1 \le i_2 \cdots \le i_r} \omega_{i_1}\omega_{i_2}\cdots\omega_{i_r} \end{aligned}$$

(where $\{\omega_i\}$ are the eigenvalues of g)

- $$\frac{1}{q^n} \sum_{\substack{\deg(f)=n \\ \text{monic}}} \tilde{\Psi}(f;h)^2 \sim q^{h+1} \int_{U(n-h-2)} |\mathrm{Tr}(g^n)|^2 \, dg$$

- $$\frac{1}{q^n} \sum_{\substack{\deg(f)=n \\ \text{monic}}} \tilde{\Psi}_j(f;h)\tilde{\Psi}_k(f;h) \sim q^{h+1} \int_{U(n-h-2)} \overline{H_j^{(r)}} H_k^{(r)} \, dg.$$

- Variance of Möbius sums in short intervals (Keating – Rudnick)
- Variance of divisor sums (Keating – Roditty-Gershon – Rudnick)

Variance of traces

Theorem (Dyson)

$$\int_{U(N)} |\mathrm{Tr}(g^r)|^2 \, dg = r \wedge N \quad for \quad r \neq 0$$

Variance of traces

Theorem (Dyson)

$$\int_{U(N)} |\mathrm{Tr}(g^r)|^2 \, dg = r \wedge N \quad for \quad r \neq 0$$

$$\int_{U(N)} \overline{\mathrm{Tr}(g^r)}\mathrm{Tr}(g^r) \, dg$$

$$= \int_{U(N)} \sum_i e^{-i2\pi r\vartheta_i} \sum_j e^{i2\pi r\vartheta_j} \, dg$$

$$= \int_{U(N)} \sum_{i=j} e^{-i2\pi r\vartheta_i} e^{i2\pi\vartheta_j} \, dg + \int_{U(N)} \sum_{i\neq j} e^{-i2\pi r\vartheta_i} e^{i2\pi r\vartheta_j} \, dg$$

$$= N + \int_{\mathbb{T}^2} e^{i2\pi r(x_1-x_2)} \left(1 - \left(\frac{\sin \pi N(x_1-x_2)}{\sin \pi(x_1-x_2)}\right)^2\right) dx_1 dx_2$$

$$= N - (N-r)_+$$

Variance of traces

Theorem (Dyson)

$$\int_{U(N)} |\mathrm{Tr}(g^r)|^2 \, dg = r \wedge N \quad for \quad r \neq 0$$

$$\int_{U(N)} \overline{\mathrm{Tr}(g^r)}\mathrm{Tr}(g^r) \, dg$$

$$= \int_{U(N)} \sum_i e^{-i2\pi r\vartheta_i} \sum_j e^{i2\pi r\vartheta_j} \, dg$$

$$= \int_{U(N)} \sum_{i=j} e^{-i2\pi r\vartheta_i} e^{i2\pi\vartheta_j} \, dg + \int_{U(N)} \sum_{i\neq j} e^{-i2\pi r\vartheta_i} e^{i2\pi r\vartheta_j} \, dg$$

$$= N + \int_{\mathbb{T}^2} e^{i2\pi r(x_1-x_2)} \left(1 - \left(\frac{\sin \pi N(x_1-x_2)}{\sin \pi(x_1-x_2)}\right)^2\right) dx_1 dx_2$$

$$= N - (N-r)_+$$

Variance of traces

Theorem (Dyson)

$$\int_{U(N)} |\mathrm{Tr}(g^r)|^2 \, dg = r \wedge N \quad for \quad r \neq 0$$

$$\int_{U(N)} \overline{\mathrm{Tr}(g^r)}\mathrm{Tr}(g^r) \, dg$$

$$= \int_{U(N)} \sum_i e^{-i2\pi r\vartheta_i} \sum_j e^{i2\pi r\vartheta_j} \, dg$$

$$= \int_{U(N)} \sum_{i=j} e^{-i2\pi r\vartheta_i} e^{i2\pi\vartheta_j} \, dg + \int_{U(N)} \sum_{i\neq j} e^{-i2\pi r\vartheta_i} e^{i2\pi r\vartheta_j} \, dg$$

$$= N + \int_{\mathbb{T}^2} e^{i2\pi r(x_1-x_2)} \left(1 - \left(\frac{\sin \pi N(x_1-x_2)}{\sin \pi(x_1-x_2)}\right)^2\right) dx_1 dx_2$$

$$= N - (N-r)_+$$

Variance of traces

Theorem (Dyson)

$$\int_{U(N)} |\mathrm{Tr}(g^r)|^2 \, dg = r \wedge N \quad for \quad r \neq 0$$

$$\int_{U(N)} \overline{\mathrm{Tr}(g^r)}\mathrm{Tr}(g^r) \, dg$$

$$= \int_{U(N)} \sum_i e^{-i2\pi r\vartheta_i} \sum_j e^{i2\pi r\vartheta_j} \, dg$$

$$= \int_{U(N)} \sum_{i=j} e^{-i2\pi r\vartheta_i} e^{i2\pi\vartheta_j} \, dg + \int_{U(N)} \sum_{i\neq j} e^{-i2\pi r\vartheta_i} e^{i2\pi r\vartheta_j} \, dg$$

$$= N + \int_{\mathbb{T}^2} e^{i2\pi r(x_1-x_2)} \left(1 - \left(\frac{\sin \pi N(x_1-x_2)}{\sin \pi(x_1-x_2)}\right)^2\right) dx_1 dx_2$$

$$= N - (N-r)_+ = r \wedge N$$

Covariance of $H_j^{(r)}$

Lemma (R.)

$$\int_{U(N)} \overline{H_j^{(r)}} H_k^{(r)} \, dg = \sum_{\nu=1}^{r \wedge N} (\nu^j - (\nu-1)^j)(\nu^k - (\nu-1)^k) \quad for \quad r \neq 0$$

Covariance of $H_j^{(r)}$

Lemma (R.)

$$\int_{U(N)} \overline{H_j^{(r)}} H_k^{(r)}\, dg = \sum_{\nu=1}^{r\wedge N} (\nu^j - (\nu-1)^j)(\nu^k - (\nu-1)^k) \quad for \quad r \neq 0$$

$$H_j^{(r)} = \sum_{\nu=1}^{r\wedge N} (-1)^\nu (\nu^j - (\nu-1)^j)\, s_{(r-\nu+1,1^{\nu-1})}$$

Here $s_{(a,1^b)}$ are hook Schur functions for the partition $(a, \underbrace{1, ..., 1}_{b})$.

Covariance of $H_j^{(r)}$

Lemma (R.)

$$\int_{U(N)} \overline{H_j^{(r)}} H_k^{(r)}\, dg = \sum_{\nu=1}^{r\wedge N} (\nu^j - (\nu-1)^j)(\nu^k - (\nu-1)^k) \quad \textit{for} \quad r \neq 0$$

$$H_j^{(r)} = \sum_{\nu=1}^{r\wedge N} (-1)^\nu (\nu^j - (\nu-1)^j)\, s_{(r-\nu+1,1^{\nu-1})}$$

Here $s_{(a,1^b)}$ are hook Schur functions for the partition $(a, \underbrace{1, \ldots, 1}_{b})$.

Orthogonality relation:

$$\int_{U(N)} \overline{s_{(a,1^b)}} s_{(c,1^d)}\, dg = \begin{cases} 1, & \text{if } (a,1^b) = (c,1^d) \\ 0, & \text{otherwise.} \end{cases}$$

A short interval dictionary

$$
\begin{array}{rcl}
\text{Arithmetic} & \longleftrightarrow & \text{Random matrix} \\
\Lambda(m) & \longleftrightarrow & -\mathrm{Tr}(g^r) \\
\Lambda_j(m) & \longleftrightarrow & H_j^{(r)}(g) \\
\mu(m) & \longleftrightarrow & h_r(g) \\
\sum_{\substack{d|m \\ d<K}} \mu(d) & \longleftrightarrow & s_{(a,1^b)}(g)
\end{array}
$$

A short interval dictionary

$$
\begin{array}{rcl}
\text{Arithmetic} & \longleftrightarrow & \text{Random matrix} \\
\Lambda(m) & \longleftrightarrow & -\mathrm{Tr}(g^r) \\
\Lambda_j(m) & \longleftrightarrow & H_j^{(r)}(g) \\
\mu(m) & \longleftrightarrow & h_r(g) \\
\sum_{\substack{d|m \\ d<K}} \mu(d) & \longleftrightarrow & s_{(a,1^b)}(g)
\end{array}
$$

for $g \in U(N)$, with $r \approx \log n$, $a \approx \log K$, $b \approx \log(m/K)$ and $N \approx \log(m/H)$, for H the size of the short interval.

Thanks:

Thanks!