Random matrix theory and the distribution of the zeros of the Riemann zeta function

LMS Summer School
Manchester

Christopher Hughes

July 2017

Contents

Disclaimer

These notes follow the same structure as my three lectures at the Summer School, and pretty much contain everything I spoke about. My aim was to keep the focus strongly on developing enough mathematics to understand the distribution of the zeros of zeta, and consequently many interesting (but tangential) results were ignored.

These notes were mainly written from scratch and therefore may contain some errors. If you notice of any typos, please email me at `christopher.hughes@york.ac.uk`.

Chapter 1

Lecture 1: The Riemann zeta function

The Riemann zeta function is one of the most significant functions in modern analytic number theory, although we won't have sufficient time in these notes to fully discover the complete reasons why. In essence, in order to understand prime numbers one needs to know about the zeros of the Riemann zeta function.

1.1 Definition of zeta as Dirichlet series and Euler product

The Riemann zeta function is defined by the Dirichlet series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

Remark. Since Riemann (1859) it is customary to use the notation $s = \sigma + \mathrm{i}t$, with $\sigma, t \in \mathbb{R}$.

It is easy to see that for $\mathrm{Re}(s) = \sigma > 1$

$$\begin{aligned} \left| \sum_{n=1}^{\infty} \frac{1}{n^{\sigma+\mathrm{i}t}} \right| &\leq \sum_{n=1}^{\infty} \left| \frac{1}{n^{\sigma+\mathrm{i}t}} \right| \\ &= \sum_{n=1}^{\infty} \frac{1}{n^{\sigma}} \\ &\leq 1 + \int_1^{\infty} x^{-\sigma} \mathrm{d}x \\ &= 1 + \frac{1}{\sigma - 1} < \infty. \end{aligned}$$

Thus the sum converges uniformly on any compact set in the half-plane $\mathrm{Re}(s) > 1$ and, since each function n^{-s} is analytic in this half-plane, the Dirichlet series defines an analytic function, denoted by $\zeta(s)$, on $\mathrm{Re}(s) > 1$.

At the point $s = 1$ we can see that the sum does not converge, because it "equals" the harmonic series $\sum_n \frac{1}{n}$, which diverges.

We shall later see that this function has a meromorphic continuation into the whole complex plane with the only simple pole at $s = 1$ (i.e. it is analytic everywhere except the point $s = 1$).

Lemma 1.1. *For* $\mathrm{Re}(s) > 1$*, the Riemann zeta function satisfies the equation*

$$\zeta(s) = \prod_{p} (1 - p^{-s})^{-1}$$

where the product is taken over all primes. This is known as the Euler product *for the Riemann zeta function.*

Remark. Euler discovered this relation between the Dirichlet series and the product representation for the Riemann zeta function in 1737, 89 years before Riemann was even born!

Proof. This is essentially the fundamental theorem of arithmetic, that says that every number greater than 1 can be written uniquely as the product of primes (up to the order of factors).

Here's the rough proof that gives the essential idea of what's going on:

Recall the Taylor expansion for $1/(1 - x)$ with $|x| < 1$:

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots$$

Hence, for each prime p in the Euler product

$$(1 - p^{-s})^{-1} = 1 + p^{-s} + p^{-2s} + p^{-3s} + \dots$$

since $|p^{-s}| < 1$ whenever $\mathrm{Re}(s) > 0$ (and we are assuming $\mathrm{Re}(s) > 1$).

Hence the Euler product is

$$(1 + 2^{-s} + 4^{-s} + 8^{-s} + \dots) \times (1 + 3^{-s} + 9^{-s} + 27^{-s} + \dots) \times \\ \times (1 + 5^{-s} + 25^{-s} + 125^{-s} + \dots) \times \cdots$$

and expanding out the brackets we see that since every integer n can be written as a product of primes powers p^k in essentially exactly one way, n^{-s} appears once and only once. Therefore the Euler product equals the Dirichlet series definition of the Riemann zeta function. □

Remark. This proof ignores all issues of convergence, but can be made rigorous by considering the limit of truncations:

$$\lim_{P \to \infty} \left| \zeta(s) - \prod_{p \le P} (1 - p^{-s})^{-1} \right| = 0$$

The fact that the Riemann zeta function has an Euler product means that its logarithm can be written in terms of a Dirichlet series:

Lemma 1.2. *For* $\mathrm{Re}(s) > 1$,

$$\log \zeta(s) = \sum_{n=2}^{\infty} \frac{\Lambda(n)}{\log n} \frac{1}{n^s}$$

where $\Lambda(n)$ *is the von-Mangoldt function*

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^k \text{ is a power of a prime} \\ 0 & \text{otherwise} \end{cases}$$

Proof. Starting from the Euler product,

$$\log \zeta(s) = \log\left(\prod_p (1 - p^{-s})^{-1}\right) = -\sum_p \log(1 - p^{-s}) = \sum_p \sum_{k=1}^{\infty} \frac{1}{k} p^{-ks}$$

where we use the Taylor expansion

$$\log(1 - x) = -\sum_{k=1}^{\infty} \frac{1}{k} x^k$$

valid for $|x| < 1$. Taking $x = p^{-s}$ and noting that $\frac{\Lambda(n)}{\log n} = \frac{1}{k}$ when $n = p^k$ completes the proof. □

Lemma 1.3. *For* $\mathrm{Re}(s) > 1$*,*

$$\frac{\zeta'}{\zeta}(s) = -\sum_{n=2}^{\infty} \frac{\Lambda(n)}{n^s}$$

where $\Lambda(n)$ *is the von-Mangoldt function given above.*

Proof. Differentiate w.r.t. s the result of the previous lemma, that for $\mathrm{Re}(s) > 1$

$$\log \zeta(s) = \sum_{n=2}^{\infty} \frac{\Lambda(n)}{\log n} \frac{1}{n^s},$$

and note that

$$\frac{\mathrm{d}}{\mathrm{d}s} \log \zeta(s) = \frac{\zeta'(s)}{\zeta(s)}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}s} n^{-s} = -\log(n) n^{-s}$$

and the fact that one can interchange differentiation and summation (since $\mathrm{Re}(s)$ is big enough to have absolute convergence). □

1.2 Analytic continuation and functional equation

Theorem 1.4. *The Riemann zeta function can be analytically continued into* $\mathbb{C} \setminus \{1\}$*, and satisfies the following functional equation, relating values at the point s with values at the point* $1 - s$*,*

$$\zeta(s) = \chi(s)\zeta(1 - s)$$

where

$$\begin{aligned} \chi(s) &= 2^s \pi^{s-1} \sin(\tfrac{1}{2}\pi s)\Gamma(1 - s) \\ &= \pi^{s-1/2} \frac{\Gamma(\frac{1}{2} - \frac{1}{2}s)}{\Gamma(\frac{1}{2}s)} \end{aligned}$$

Remark. Riemann's paper has two different proofs of this result. In Titchmarsh's book *The Theory of the Riemann Zeta-Function* there are seven distinct proofs!

We will prove this theorem following one of Riemann's original proofs, via Poisson summation.

1.2.1 Poisson summation

Lemma 1.5 (Poisson summation)**.** *If $f \in L^1(\mathbb{R})$ then*

$$\sum_{k=-\infty}^{\infty} f(k) = \sum_{h=-\infty}^{\infty} \widehat{f}(h)$$

where

$$\widehat{f}(u) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x u} \mathrm{d}x$$

is the Fourier transform of f.

Proof. Let

$$F(\alpha) = \sum_{k=-\infty}^{\infty} f(k+\alpha)$$

Note that $F(\alpha)$ is a periodic function of α with period 1, and the conditions on f imply that the series converges in $L^1([0,1])$ and also almost everywhere. Because of the L^1 convergence, the Fourier coefficients of F can be calculated by integrating term by term, so

$$F(\alpha) = \sum_{h=-\infty}^{\infty} \widehat{F}(h) e^{2\pi i h\alpha} \tag{1.1}$$

where

$$\begin{aligned}
\widehat{F}(h) &= \int_0^1 F(\alpha) e^{-2\pi i h\alpha} \mathrm{d}\alpha \\
&= \int_0^1 \sum_{k=-\infty}^{\infty} f(k+\alpha) e^{-2\pi i h\alpha} \mathrm{d}\alpha \\
&= \sum_{k=-\infty}^{\infty} \int_0^1 f(k+\alpha) e^{-2\pi i h\alpha} \mathrm{d}\alpha \\
&= \sum_{k=-\infty}^{\infty} \int_k^{k+1} f(x) e^{-2\pi i h(x-k)} \mathrm{d}x
\end{aligned}$$

by a simple change of variables $x = k+\alpha$. Note that $e^{-2\pi i h(x-k)} = e^{-2\pi i h x}$, so the integrand is independent of k. Furthermore note that the sum now adds up the integral over distinct unit intervals, and hence we have

$$\begin{aligned}
\widehat{F}(h) &= \int_{-\infty}^{\infty} f(x) e^{-2\pi i h x} \mathrm{d}x \\
&= \widehat{f}(h)
\end{aligned}$$

Therefore, substituting back into (1.1) we have shown that

$$\sum_{k=-\infty}^{\infty} f(k+\alpha) = F(\alpha) = \sum_{h=-\infty}^{\infty} \widehat{f}(h) e^{2\pi i h\alpha}$$

Setting $\alpha = 0$ completes the proof. □

Lemma 1.6. *For $u > 0$, let*

$$\theta(u) = \sum_{k=-\infty}^{\infty} e^{-\pi k^2 u}$$

then

$$\theta(u) = \frac{1}{\sqrt{u}}\theta\left(\frac{1}{u}\right)$$

Proof. For $u > 0$, let

$$f(x) = e^{-\pi x^2 u}$$

so that the Fourier transform is

$$\widehat{f}(h) = \int_{-\infty}^{\infty} e^{-\pi x^2 u} e^{-2\pi i x h} \mathrm{d}x = \int_{-\infty}^{\infty} e^{-\pi u (x + ih/u)^2 - \pi h^2/u} \mathrm{d}x$$

(from completing the square). Shifting the contour to make $t = x + ih/u$ real, which can be justified by the rapid decay of the integrand, we have

$$\widehat{f}(h) = e^{-\pi h^2/u} \int_{-\infty}^{\infty} e^{-\pi u t^2} \mathrm{d}t = e^{-\pi h^2/u} \frac{1}{\sqrt{u}}$$

with the last step coming from evaluating the Gaussian integral.

Therefore, we see that

$$\theta(u) = \sum_{k=-\infty}^{\infty} f(k)$$

and by Poisson summation we have

$$\begin{aligned}\theta(u) &= \sum_{h=-\infty}^{\infty} \widehat{f}(h) = \frac{1}{\sqrt{u}} \sum_{h=-\infty}^{\infty} e^{-\pi h^2/u} \\ &= \frac{1}{\sqrt{u}}\theta\left(\frac{1}{u}\right).\end{aligned}$$

We will use this result in our proof of the functional equation for zeta. □

1.2.2 The Gamma function

For $\mathrm{Re}(z) > 0$ the Gamma function is defined as follows:

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} \mathrm{d}x\,.$$

The integral has 2 singularities: at 0 and at ∞, but we can easily see that it is absolutely convergent for $\mathrm{Re}(z) > 0$.

Properties:

1. $\Gamma(1) = 1$;
2. $\Gamma(z + 1) = z\Gamma(z)$.
3. The Gamma function can be meromorphically extended to $\mathbb{C}$ using the above functional equation.

4. Γ has simple poles at non-positive integers $z = -n$ ($n = 0, 1, 2, \dots$) with residue $(-1)^n/n!$.
5. The reflection formula says that
$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}.$$
6. The duplication formula says that
$$\Gamma(z)\Gamma(z + \tfrac{1}{2}) = 2^{1-2z}\sqrt{\pi}\Gamma(2z)$$
7. Stirling's Formula says that for large z with $|\arg(z)| < \pi - \delta$ (for fixed $\delta > 0$) we have
$$\log\Gamma(z) = z\log z - z - \tfrac{1}{2}\log z + \tfrac{1}{2}\log(2\pi) + O\left(\frac{1}{z}\right)$$

1.2.3 Proof of the functional equation and analytic continuation

Let
$$\zeta^*(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$$
and let
$$\psi(u) = \sum_{n=1}^{\infty} e^{-\pi n^2 u}$$

Lemma 1.7. *We have*
$$\psi(u) = \frac{1}{\sqrt{u}}\psi\left(\frac{1}{u}\right) + \frac{1}{2\sqrt{u}} - \frac{1}{2}.$$

Proof. Since the summand in $\theta(u)$ is even in n, we have
$$\theta(u) = 1 + 2\sum_{n=1}^{\infty} e^{-\pi u n^2} = 1 + 2\psi(u)$$
Lemma 1.6, from the Poisson summation section, says, $\theta(u) = \theta(1/u)/\sqrt{u}$, so
$$\begin{aligned}\psi(u) &= \frac{1}{2}\theta(u) - \frac{1}{2}\\ &= \frac{\theta(1/u)}{2\sqrt{u}} - \frac{1}{2}\\ &= \frac{\psi(1/u)}{\sqrt{u}} + \frac{1}{2\sqrt{u}} - \frac{1}{2}\end{aligned}$$
□

Lemma 1.8. *For* $\mathrm{Re}(s) > 1$,
$$\zeta^*(s) = \int_0^{\infty} u^{s/2-1}\psi(u)\mathrm{d}u.$$

Proof. We have for $\mathrm{Re}(s) > 1$,
$$\zeta(s) = \sum_{n=1}^{\infty}\frac{1}{n^s}$$

and

$$\Gamma(z) = \int_0^\infty e^{-x} x^z \frac{\mathrm{d}x}{x},$$

so

$$\zeta^*(s) = \pi^{-s/2} \int_0^\infty e^{-x} x^{s/2} \frac{\mathrm{d}u}{x} \sum_{n=1}^\infty \frac{1}{n^s}$$

Switching the order of summation and integration (justified by absolute convergence)

$$\zeta^*(s) = \sum_{n=1}^\infty \int_0^\infty e^{-x} \left(\frac{x}{\pi n^2}\right)^{s/2} \frac{\mathrm{d}x}{x}$$

and then changing variables to $u = \frac{x}{\pi n^2}$ (and note that $\frac{\mathrm{d}x}{x} = \frac{\mathrm{d}u}{u}$) we have

$$\zeta^*(s) = \sum_{n=1}^\infty \int_0^\infty e^{-\pi u n^2} u^{s/2} \frac{\mathrm{d}u}{u}.$$

Now pull the sum inside the integral (justified by absolute convergence in the half-plane under consideration) and we obtain the desired result. □

Lemma 1.9. *We have*

$$\zeta^*(s) = \frac{1}{s(s-1)} + \int_1^\infty \left(u^{-s/2-1/2} + u^{s/2-1}\right) \psi(u)\mathrm{d}u.$$

Proof. We just shown that for $\mathrm{Re}(s) > 1$,

$$\zeta^*(s) = \int_0^\infty u^{s/2-1} \psi(u)\mathrm{d}u.$$

Split the integral into a integral from 0 to 1 and an integral from 1 to ∞, which yields

$$\zeta^*(s) = \int_0^1 \psi(u) u^{s/2} \frac{\mathrm{d}u}{u} + \int_1^\infty \psi(u) u^{s/2} \frac{\mathrm{d}u}{u} \tag{1.2}$$

and letting $x = 1/u$, and noticing that $\frac{\mathrm{d}x}{x} = -\frac{\mathrm{d}u}{u}$ we see that the first integral equals

$$\int_0^1 \psi(u) u^{s/2} \frac{\mathrm{d}u}{u} = \int_1^\infty \psi(1/x) x^{-s/2} \frac{\mathrm{d}x}{x}$$

We have just shown in Lemma 1.7 that

$$\psi(1/x) = \sqrt{x}\psi(x) + \frac{\sqrt{x}}{2} - \frac{1}{2}$$

which means this integral equals

$$\int_1^\infty \left(\sqrt{x}\psi(x) + \frac{\sqrt{x}}{2} - \frac{1}{2}\right) x^{-s/2} \frac{\mathrm{d}x}{x} = \int_1^\infty \psi(x) x^{-s/2-1/2} \mathrm{d}x + \frac{1}{2}\int_1^\infty x^{-s/2-1/2}\mathrm{d}x - \frac{1}{2}\int_1^\infty x^{-s/2-1}\mathrm{d}x$$

Since $\mathrm{Re}(s) > 1$ the last two integrals exist and equal

$$\frac{1}{s-1} - \frac{1}{s} = \frac{1}{s(s-1)}$$

and so we have
$$\int_0^1 \psi(u)u^{s/2}\frac{\mathrm{d}u}{u} = \int_1^\infty \psi(x)x^{-s/2-1/2}\mathrm{d}x + \frac{1}{s(s-1)}.$$
Substituting this into (1.2) we have

$$\zeta^*(s) = \int_1^\infty \psi(x)x^{-s/2-1/2}\mathrm{d}x + \frac{1}{s(s-1)} + \int_1^\infty \psi(u)u^{s/2-1}\mathrm{d}u$$

which is the required formula. □

Proof of Theorem 1.4. When u gets large, $\psi(u)$ decays exponentially quickly. Therefore the integral exists (and is analytic) for all s since exponential decay beats polynomial growth. The only poles are at $s = 0$ and $s = 1$, and come from the last term, $\frac{1}{s(s-1)}$

Note that if you replace s with $1-s$, the RHS is unchanged. Therefore, we have $\zeta^*(s) = \zeta^*(1-s)$.

Substituting back for the definition of $\zeta^*(s)$ in terms of $\zeta(s)$ this means

$$\pi^{-s/2}\Gamma(s/2)\zeta(s) = \pi^{-(1-s)/2}\Gamma((1-s)/2)\zeta(1-s)$$

that is,
$$\zeta(s) = \chi(s)\zeta(1-s)$$
where

$$\begin{aligned}\chi(s) &= \pi^{s-1/2}\frac{\Gamma(\frac{1}{2}-\frac{1}{2}s)}{\Gamma(\frac{1}{2}s)} \\ &= 2^s\pi^{s-1}\sin(\tfrac{1}{2}\pi s)\Gamma(1-s)\end{aligned} \tag{1.3}$$

where the last equality uses the reflection formula for the Gamma function. □

1.3 The zeros of the Riemann zeta function

Recall the functional equation which says that

$$\zeta(s) = \chi(s)\zeta(1-s)$$

where
$$\chi(s) = \pi^{s-1/2}\frac{\Gamma(\frac{1}{2}-\frac{1}{2}s)}{\Gamma(\frac{1}{2}s)}$$
Since for $\mathrm{Re}(s) > 1$, $\zeta(s)$ can be written as an absolutely convergent Euler product, none of whose terms vanish, ζ does not vanish for $\mathrm{Re}(s) > 1$. Hence, by the functional equation, zeta vanishes for $\mathrm{Re}(s) < 0$ only when $\chi(s)$ vanishes, which occurs at the poles of $\Gamma(s/2)$, which are at $0, -2, -4, -6, \ldots$. However, the pole of $\Gamma(s/2)$ at $s = 0$ actually cancels the pole of zeta at $s = 1$, so $\zeta(s) \neq 0$ at $s = 0$.

Therefore, we have shown that zeta has no zeros for $\mathrm{Re}(s) > 1$ and no zeros other than $-2, -4, -6, \ldots$ for $\mathrm{Re}(s) < 0$ (these are called the trivial zeros).

This leaves the critical strip $0 \leq \mathrm{Re}(s) \leq 1$ to be studied.

Theorem 1.10. *Let*

$$N(T) = \#\{\rho\ ,\ \zeta(\rho) = 0\ ,\ 0 \leq \mathrm{Re}(\rho) \leq 1\ ,\ 0 < \mathrm{Im}(\rho) \leq T\}$$

then for large T

$$N(T) = \frac{T}{2\pi}\log\frac{T}{2\pi e} + \frac{7}{8} + S(T) + O\left(\frac{1}{T}\right)$$

where

$$S(T) = \frac{1}{\pi}\operatorname{Im}\log\zeta(\tfrac{1}{2} + \mathrm{i}T)$$

defined by continuous variation along the straight lines joining 2 *to* $2 + \mathrm{i}T$ *to* $1/2 + \mathrm{i}T$*, starting with the value* 0*.*

Furthermore the following result is known (this is due to von Mangoldt in 1905, though we won't prove it here)

$$S(T) = O(\log T)$$

Hence,

Corollary.

$$N(T) = \frac{T}{2\pi}\log\frac{T}{2\pi e} + O(\log T)$$

We won't prove the theorem concerning $S(T)$ here, but assuming this result, we will provide a proof for the formula for $N(T)$. (This is essentially what Riemann did in 1859, since he gave no hint of a proof for bounding $S(T)$).

Proof of formula for $N(T)$. Define

$$\xi(s) = \frac{1}{2}s(s-1)\Gamma(s/2)\pi^{-s/2}\zeta(s)$$

and note that this is an entire function which vanishes only at the zeros of zeta inside the critical strip. (The trivial zeros are canceled by the poles of the Gamma function; the pole of gamma at $s = 0$ is canceled by the s term; the pole of zeta at $s = 1$ is canceled by the $s - 1$ term).

Using the general formula for zeros and poles inside a closed contour, and the fact that $\xi(s)$ has no poles and no zeros outside the critical strip, we have

$$N(T) = \frac{1}{2\pi\mathrm{i}}\oint_R \frac{\xi'}{\xi}(s)\mathrm{d}s = \frac{1}{2\pi}\{\text{change in argument around } R \text{ of } \xi(s)\}$$

where R is the rectangle with vertices 2, $2 + \mathrm{i}T$, $-1 + \mathrm{i}T$ and -1. Other than the horizontal parts, note that this contour is all outside the critical strip. Furthermore, note that $\xi(s)$ does not vanish on the real axis (it is always positive). Thus if T is such that $T \neq \operatorname{Im}(\rho)$ for any zero ρ, then there are no zeros on the contour R.

Now, the functional equation implies that $\xi(s) = \xi(1-s)$, and since $\xi(s) = \overline{\xi(\overline{s})}$, which together imply that

$$\xi(\tfrac{1}{2} + x + \mathrm{i}t) = \overline{\xi(\tfrac{1}{2} - x + \mathrm{i}t)}$$

In other words, if we let $\xi(\frac{1}{2}+x+\mathrm{i}t) = re^{\mathrm{i}\theta}$ then $\xi(\frac{1}{2}-x+\mathrm{i}t) = re^{-\mathrm{i}\theta}$, and so the change in the imaginary part of the logarithm from 2 to $2 + \mathrm{i}T$ to $1/2 + \mathrm{i}T$, is exactly equal to the change in the imaginary part of the logarithm from $1/2 + \mathrm{i}T$ to $-1 + \mathrm{i}T$ to -1.

Since $\xi(s)$ is positive on the real axis, the imaginary part of the logarithm (that is, the argument) of ξ does not change along that line (in fact, the argument of ξ can be taken to equal 0 all along the real axis). Hence we have shown that

$$N(T) = \frac{1}{\pi}\{\text{change along straight lines joining } 2 \text{ to } 2 + \mathrm{i}T \text{ to } 1/2 + \mathrm{i}T \text{ of } \operatorname{Im}\log\xi(s)\}$$

Now,

$$\operatorname{Im}\log\xi(s) = \operatorname{Im}\log(\tfrac{1}{2}s(s-1)) + \operatorname{Im}\log\left(\Gamma(\tfrac{1}{2}s)\pi^{-s/2}\right) + \operatorname{Im}\log\zeta(s)$$

By definition, the change argument of $\operatorname{Im}\log\zeta(s)$ as s varies from 2 to $2+\mathrm{i}T$ to $1/2+\mathrm{i}T$ exactly equals the definition of $\pi S(T)$. Furthermore, the change in $\operatorname{Im}\log(\tfrac{1}{2}s(s-1))$ as s moves from 2 to $2+\mathrm{i}T$ to $1/2+\mathrm{i}T$ equals

$$\operatorname{Im}\log(-\tfrac{1}{8}-\tfrac{1}{2}T^2) - \operatorname{Im}\log(1) = \pi$$

Therefore if we define

$$\begin{aligned}\theta(T) &= \operatorname{Im}\log\left(\Gamma(\tfrac{1}{2}s)\pi^{-s/2}\right)\Big|_{s=1/2+\mathrm{i}T}\\ &= \operatorname{Im}\log\Gamma(\tfrac{1}{4}+\tfrac{1}{2}\mathrm{i}T) - \tfrac{1}{2}T\log\pi \end{aligned}\tag{1.4}$$

then we have shown that

$$N(T) = \frac{1}{\pi}\theta(T) + 1 + S(T) \tag{1.5}$$

We now deal with $\theta(T)$ using Stirling's formula. We have

$$\begin{aligned}&\operatorname{Im}\log\Gamma(\tfrac{1}{4}+\tfrac{1}{2}\mathrm{i}T)\\ &\quad= \operatorname{Im}\left\{(\tfrac{1}{4}+\tfrac{1}{2}\mathrm{i}T)\log(\tfrac{1}{4}+\tfrac{1}{2}\mathrm{i}T) - (\tfrac{1}{4}+\tfrac{1}{2}\mathrm{i}T) - \tfrac{1}{2}\log(\tfrac{1}{4}+\tfrac{1}{2}\mathrm{i}T) + \tfrac{1}{2}\log(2\pi) + O(1/T)\right\}\\ &\quad= \tfrac{1}{4}\operatorname{Im}\log\left(\tfrac{1}{4}+\tfrac{1}{2}\mathrm{i}T\right) + \tfrac{1}{2}T\operatorname{Re}\log\left(\tfrac{1}{4}+\tfrac{1}{2}\mathrm{i}T\right) - \tfrac{1}{2}T - \tfrac{1}{2}\operatorname{Im}\log\left(\tfrac{1}{4}+\tfrac{1}{2}\mathrm{i}T\right) + O(1/T)\end{aligned}$$

Taking these terms separately, we have

- The first term is $\tfrac{1}{4}\operatorname{Im}\log(\tfrac{1}{4}+\tfrac{1}{2}\mathrm{i}T) = \tfrac{\pi}{8} + O\left(\tfrac{1}{T}\right)$
- Since $\operatorname{Re}\log(\tfrac{1}{4}+\tfrac{1}{2}\mathrm{i}T) = \log\left|\tfrac{1}{4}+\tfrac{1}{2}\mathrm{i}T\right|$ we have that

$$\begin{aligned}\tfrac{1}{2}T\operatorname{Re}\log(\tfrac{1}{4}+\tfrac{1}{2}\mathrm{i}T) &= \tfrac{1}{2}T\log\left(\sqrt{\frac{T^2}{4}+\frac{1}{16}}\right)\\ &= \tfrac{1}{2}T\log\left(\frac{T}{2}\sqrt{1+\frac{1}{4T^2}}\right)\\ &= \tfrac{1}{2}T\log\frac{T}{2} + \tfrac{1}{2}T\log\left(1+\frac{1}{4T^2}\right)\\ &= \tfrac{1}{2}T\log\frac{T}{2} + O(1/T)\end{aligned}$$

- The last term is $-\tfrac{1}{2}\operatorname{Im}\log\left(\tfrac{1}{4}+\tfrac{1}{2}\mathrm{i}T\right) = -\tfrac{\pi}{4} + O(1/T)$

Hence, the change of argument of gamma equals

$$\frac{\pi}{8} + \frac{1}{2}T\log\frac{T}{2} - \frac{1}{2}T - \frac{\pi}{4} + O\left(\frac{1}{T}\right) = \frac{1}{2}T\log\frac{T}{2e} - \frac{\pi}{8} + O\left(\frac{1}{T}\right)$$

where we write $-\tfrac{T}{2}$ as $-\tfrac{T}{2}\log e$. Substituting this into (1.4) we have

$$\theta(T) = \frac{T}{2}\log\left(\frac{T}{2\pi e}\right) - \frac{\pi}{8} + O\left(\frac{1}{T}\right)$$

Dividing by π, and substituting into (1.5) we see that

$$N(T) = \frac{T}{2\pi}\log\frac{T}{2\pi e} + S(T) + \frac{7}{8} + O\left(\frac{1}{T}\right)$$

Using the fact (which we won't prove in this course) that $S(T) = O(\log T)$ we have

$$N(T) = \frac{T}{2\pi}\log\frac{T}{2\pi e} + O(\log T)$$

as required. □

1.3.1 The Riemann Hypothesis

Using the fact that $\zeta(s) = \overline{\zeta(\overline{s})}$ and the functional equation, we see that if $\zeta(\rho) = 0$ with ρ in the critical strip (that is, not a non-trivial zero), then $1 - \rho$, $\overline{\rho}$ and $1 - \overline{\rho}$ are also zeros.

That is, there is a symmetry about the real axis and about the line $\mathrm{Re}(s) = 1/2$. This line is called the critical line, and in 1859 Riemann conjectured that all the non-trivial zeros of zeta lie on the critical line.

Conjecture (Riemann Hypothesis)**.** *All the zeros of the Riemann zeta function which lie in the critical strip have real part equal to* $\frac{1}{2}$.

Remark. It is traditional to write the zeros of zeta as ρ, and assuming the Riemann Hypothesis we have $\rho = \frac{1}{2} + \mathrm{i}\gamma$ with $\gamma \in \mathbb{R}$.

In the next lecture we will look at ways of calculating the zeros of zeta, and find out some known results concerning them. In the third lecture we will discover how one can model these zeros by random matrix theory, and some of the consequences of that model.

Chapter 2

Lecture 2: The zeros of zeta

2.1 History of counting zeros on the line

In the previous lecture we saw that there are about $\frac{T}{2\pi}\log\frac{T}{2\pi e}$ zeros of the zeta function in the critical strip, with the Riemann Hypothesis asserting that they all lie on the critical line. Let $N_0(T)$ denote the number of zeros which lie on the critical line. In 1914 Hardy showed that an infinite number of zeros of zeta lie on the critical line. Later in 1921 Hardy and Littlewood showed that $N_0(T) > AT$ for some constant A.

The next great breakthrough occurred in 1942 when Selberg showed that there is a positive constant c such that $N_0(T) > cT\log T$, which means a positive proportion of zeros of the zeta function lie on the line. In his 1947 PhD thesis, Szu-Hoa Min explicitly calculated the constant implicit in Selberg's work, and found it is very small.

By a different method, called "mollifiying", Levinson showed in 1974 that that $N_0(T) > \frac{1}{3}N(T)$. In 1989 Conrey optimised that approach to show that more than two-fifths of all the zeros lie on the line (his actual number is 0.4088). This estimate was not improved upon until 2010 when Bui, Conrey, and Young increased it to 0.4105. Currently the record is that at least 41.28% of all zeros lie on the critical line (Shaoji Feng, 2011).

2.2 Numerical calculations of zeros

In order to verify the Riemann Hypothesis numerically, one needs to check that the zeros lie *exactly* on the critical line. This is hard to do for complex-valued functions (see Figures 2.1 and 2.2).

Last lecture we saw the Riemann xi function, defined as

$$\xi(s) = \frac{1}{2}s(s-1)\Gamma(s/2)\pi^{-s/2}\zeta(s)$$

The functional equation is $\xi(s) = \xi(1-s)$, and the fact that $\xi(s) = \overline{\xi(\overline{s})}$ means that $\xi(\frac{1}{2}+\mathrm{i}t)$ is a real function for $t \in \mathbb{R}$.

Therefore sign changes of $\xi(\frac{1}{2}+\mathrm{i}t)$ yield zeros of zeta (since the only zeros of ξ are the non-trivial zeros of zeta).

Unfortunately Stirling's formula implies that $\xi(\frac{1}{2}+\mathrm{i}t)$ decays exponentially quickly as $t \to \infty$, making numerical observation of sign changes difficult (see Figure 2.3).

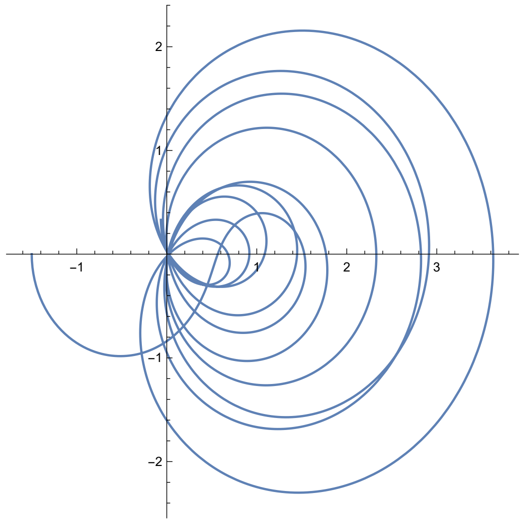


Figure 2.1: Plot of the complex values of $\zeta(\frac{1}{2} + \mathrm{i}t)$ for $0 \le t \le 50$. The graph appears to go through the origin (that is, there are values of t that appear to yield zeros of zeta on the critical line).

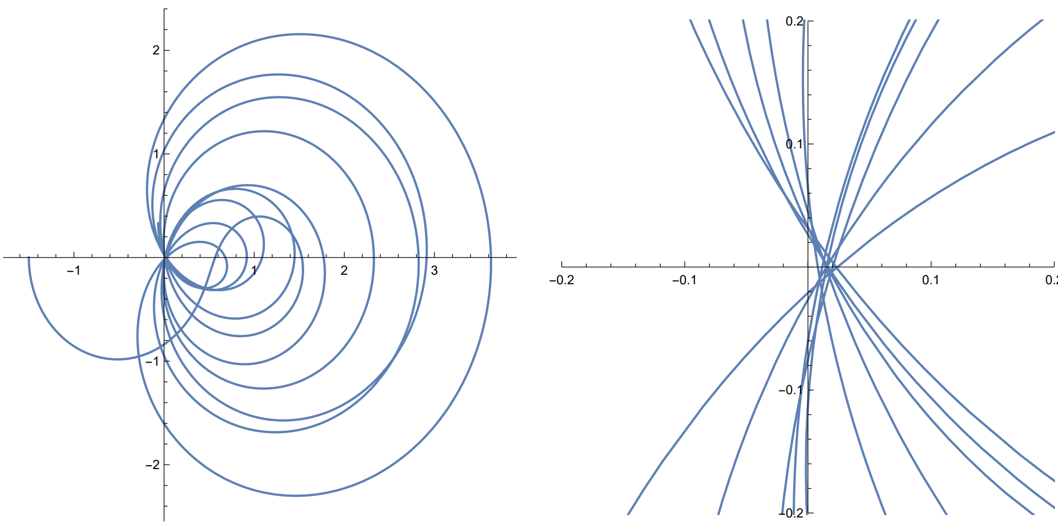


Figure 2.2: Plot of the complex values of $\zeta(\frac{1}{2} + 0.01 + \mathrm{i}t)$ for $0 \le t \le 50$. Zooming in, it is clear there are no zeros in this range.

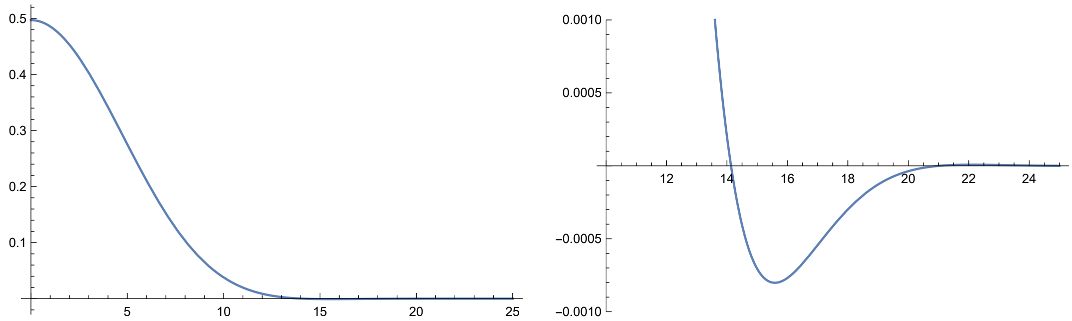


Figure 2.3: Plot of $\xi(\frac{1}{2} + \mathrm{i}t)$ for $0 \le t \le 25$ but it's hard to see any sign changes. Zooming in, the zero at $14.13\ldots$ is now clearly visible, but it's hard to see the second zero at $21.02\ldots$.

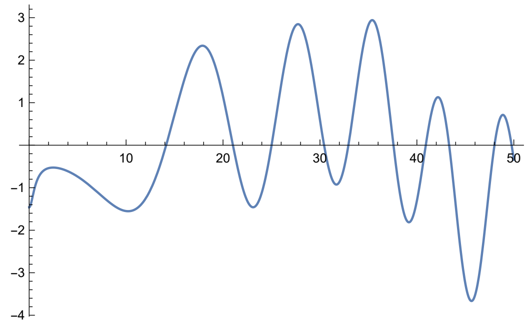


Figure 2.4: Plot of $Z(t)$ for $0 \le t \le 50$. The sign changes are now obvious, and these are the zeros of $\zeta(\frac{1}{2} + it)$ on the critical line.

Recall the functional equation

$$\zeta(s) = \chi(s)\zeta(1-s)$$

implies that $\chi(s)\chi(1-s) = 1$, and so $|\chi(\frac{1}{2} - it)| = 1$ for real t. That is, we can define a real function $\theta(t)$ such that

$$\chi(\tfrac{1}{2} - it) = e^{2i\theta(t)}$$

(The reason for the factor of 2 becomes evident later). Equation (1.3) tells us

$$\chi(s) = \pi^{s-1/2}\frac{\Gamma(\frac{1}{2} - \frac{1}{2}s)}{\Gamma(\frac{1}{2}s)}$$

and so we have

$$\chi(\tfrac{1}{2} - it) = \pi^{-it}\frac{\Gamma(\frac{1}{4} + \frac{1}{2}it)}{\Gamma(\frac{1}{4} - \frac{1}{2}it)}$$

This means

$$\begin{aligned} 2\theta(t) &= -t\log(\pi) + 2\operatorname{Im}\log\Gamma(\tfrac{1}{4} + \tfrac{1}{2}it) \\ &= t\log\left(\frac{t}{2\pi e}\right) - \frac{\pi}{4} + O\left(\frac{1}{t}\right) \end{aligned} \tag{2.1}$$

(This is a very similar analysis using Stirling's formula to that which we did to find $N(T)$ in the previous lecture).

This enables us to remove the exponential decay of $\xi(\frac{1}{2} + it)$ by considering the function

$$Z(t) = \sqrt{\chi(\tfrac{1}{2} - it)}\zeta(\tfrac{1}{2} + it) = e^{i\theta(t)}\zeta(\tfrac{1}{2} + it)$$

From the functional equation, one can see that $Z(t)$ is an even *real* function for real t, and clearly $|Z(t)| = |\zeta(\frac{1}{2} + it)|$, and so if one can find t_1 and t_2 such that $Z(t_1) > 0$ and $Z(t_2) < 0$ then there must be a zero of zeta on the critical line between t_1 and t_2. (See Figure 2.4).

To verify the Riemann Hypothesis up to a certain height, one numerically calculates all the zeros that lie on the line (by counting sign changes), and then compares the answer with $N(T)$, which is the number of zeros inside the strip. If the two are equal, then you've found all the zeros, and they all lie on the line. This idea was made practicable by Turing in Manchester in 1952 (in his last paper before he died), and in practice it involves an integral of $S(t)$.

Using either a high-powered computers (eg a Cray supercomputer) or a distributed computing method, the first 10^{13} zeros are known to lie on the critical line, as are the 175 million zeros around zero number 10^{20}. (See Table 2.1).

Year	n	Author
c.1859	10?	B. Riemann
1903	15	J. P. Gram
1914	79	R. J. Backlund
1925	138	J. I. Hutchinson
1935	1,041	E. C. Titchmarsh
1953	1,104	A. M. Turing
1956	15,000	D. H. Lehmer
1956	25,000	D. H. Lehmer
1958	35,337	N. A. Meller
1966	250,000	R. S. Lehman
1968	3,500,000	J. B. Rosser, J. M. Yohe, L. Schoenfeld
1977	40,000,000	R. P. Brent
1979	81,000,001	R. P. Brent
1982	200,000,001	R. P. Brent, J. van de Lune, H. J. J. te Riele, D. T. Winter
1983	300,000,001	J. van de Lune, H. J. J. te Riele
1986	1,500,000,001	J. van de Lune, H. J. J. te Riele, D. T. Winter
2004	900,000,000,000	S. Wedeniwski
2004	10,000,000,000,000	X. Gourdon and Patrick Demichel

Table 2.1: Historic records for verification of the Riemann Hypothesis

2.3 Euler-Maclaurin summation

Lemma 2.1 (Euler-Maclaurin formula)**.** *Let $f \in C^1$. Then*

$$\sum_{n=M}^{N} f(n) = \int_M^N f(x)\,\mathrm{d}x + \tfrac{1}{2}f(M) + \tfrac{1}{2}f(N) + \int_M^N (x - \lfloor x \rfloor - \tfrac{1}{2})f'(x)\,\mathrm{d}x\,.$$

Quick proof. Apply integration by parts to the following Riemann-Stieltjes integral:

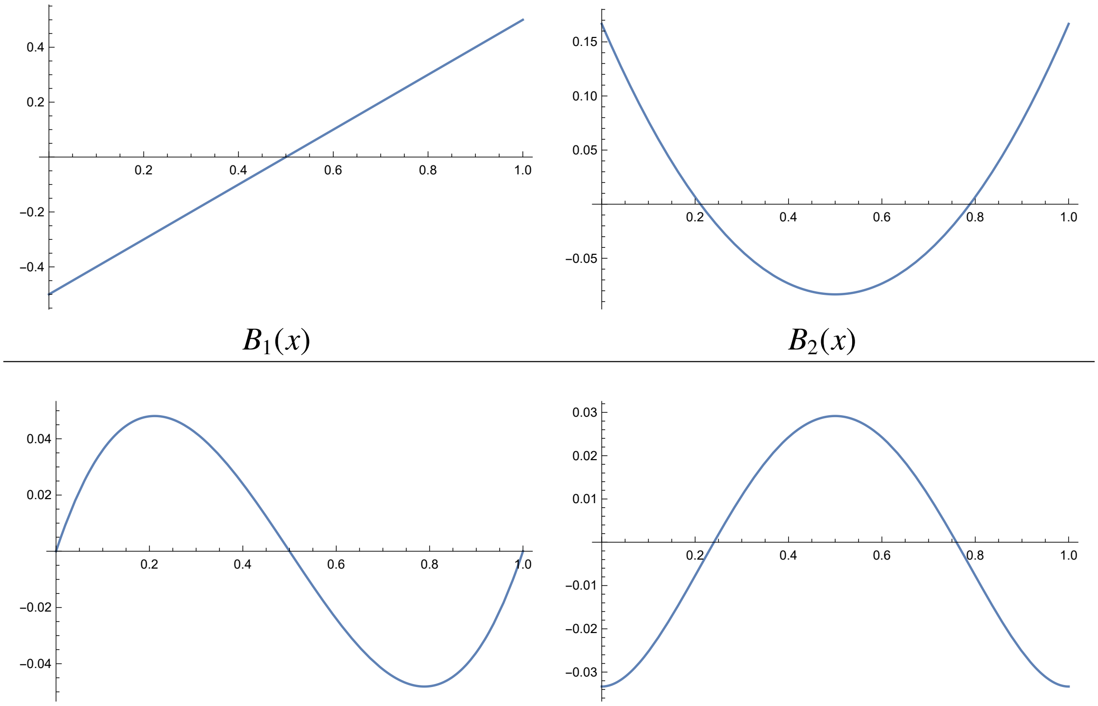
$$\begin{aligned}\sum_{n=M}^{N} f(n) &= \int_{M^-}^{N^+} f(x)\mathrm{d}(\lfloor x \rfloor)\\ &= \int_{M^-}^{N^+} f(x)\mathrm{d}(\lfloor x \rfloor + \tfrac{1}{2} - x) + \int_M^N f(x)\mathrm{d}x\end{aligned}$$

□

Longer proof. Note that for n an integer, for all $n \le x < n+1$ we have $(x - \lfloor x \rfloor - \frac{1}{2}) = x - n - \frac{1}{2}$, so

$$\begin{aligned}\int_n^{n+1} (x - \lfloor x \rfloor - \tfrac{1}{2})f'(x)\,\mathrm{d}x &= \int_n^{n+1} xf'(x)\mathrm{d}x - (n + \tfrac{1}{2})\int_n^{n+1} f'(x)\mathrm{d}x\\ &= \left((n+1)f(n+1) - nf(n) - \int_n^{n+1} f(x)\mathrm{d}x\right) - (n + \tfrac{1}{2})\,(f(n+1) - f(n))\\ &= \tfrac{1}{2}f(n+1) + \tfrac{1}{2}f(n) - \int_n^{n+1} f(x)\mathrm{d}x\end{aligned}$$

and now sum from $n = M$ to $n = N - 1$ and note that each summand, apart from $f(M)$ and $f(N)$ appears twice, and that the integrals combine to be $\int_M^N$. □



$B_1(x)$ $B_2(x)$

$B_3(x)$ $B_4(x)$

Figure 2.5: Plot of the first four Bernoulli polynomials.

Integrating by parts the remainder term, $\int_M^N (x - \lfloor x \rfloor - \frac{1}{2}) f'(x)\, \mathrm{d}x$, leads to an improved formula. To obtain this, we first introduce the Bernoulli polynomials:

Definition. The nth Bernoulli polynomial is a degree n polynomial $B_n(x)$ such that for all t,

$$\int_t^{t+1} B_n(x)\mathrm{d}x = t^n$$

$$\begin{aligned} B_0(x) &= 1 \\ B_1(x) &= x - \tfrac{1}{2} \\ B_2(x) &= x^2 - x + \tfrac{1}{6} \\ B_3(x) &= x^3 - \tfrac{3}{2}x^2 + \tfrac{1}{2}x \\ B_4(x) &= x^4 - 2x^3 + x^2 - \tfrac{1}{30} \end{aligned}$$

We have that $B_n'(x) = nB_{n-1}(x)$ and there is a simple generating function for them

$$\frac{te^{xt}}{e^t - 1} = \sum_{n=0}^{\infty} B_n(x) \frac{t^n}{n!}$$

Apart from the case $n = 1$, we have $B_n(0) = B_n(1)$.

Definition. The Bernoulli numbers are defined as the constant term in the nth Bernoulli polynomial, that is $B_n = B_n(0)$

The Bernoulli numbers initially shrink, but eventually get very large. We have

$$
\begin{array}{llll}
B_0 = 1 & B_1 = -1/2 & B_2 = 1/6 & B_3 = 0 \\
B_4 = -1/30 & B_5 = 0 & B_6 = 1/42 & B_7 = 0 \\
B_8 = -1/30 & B_9 = 0 & B_{10} = 5/66 & B_{11} = 0 \\
\vdots & & \vdots & \\
B_{20} = -174611/330 & \cdots & B_{30} = 8615841276005/14322 &
\end{array}
$$

With the exception of $n = 1$, the odd Bernoulli numbers are zero.

Note that

$$x - \lfloor x \rfloor - \tfrac{1}{2} = B_1(\{x\})$$

where $\{x\} = x - \lfloor x \rfloor$ denotes the fractional part of the real number x.

After repeated integration by parts r times, the Euler-Maclaurin formula can be rewritten as

$$\sum_{n=M}^{N} f(n) = \int_M^N f(x)\,\mathrm{d}x + \tfrac{1}{2}f(M) + \tfrac{1}{2}f(N) + \sum_{k=1}^{\lfloor r/2 \rfloor} \frac{B_{2k}}{(2k)!}\left(f^{(2k-1)}(N) - f^{(2k-1)}(M)\right) + (-1)^{r+1}\int_M^N \frac{1}{r!} B_r(\{x\}) f^{(r)}(x)\,\mathrm{d}x\,.$$

2.3.1 Application to the Riemann zeta function

Let $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 1$, $x \geq 1$ and $f(x) = x^{-s}$. Then using Euler Maclaurin with the trivial error term on the second sum yields

$$
\begin{aligned}
\zeta(s) &= \sum_{n=1}^{M-1} \frac{1}{n^s} + \sum_{n=M}^{\infty} f(n) \\
&= \sum_{n=1}^{M-1} \frac{1}{n^s} + \int_M^\infty \frac{\mathrm{d}x}{x^s} + \tfrac{1}{2}M^{-s} - s\int_M^\infty B_1(\{x\})x^{-s-1}\,\mathrm{d}x \\
&= \sum_{n=1}^{M-1} \frac{1}{n^s} + \left[\frac{u^{-s+1}}{-s+1}\right]_M^\infty + \tfrac{1}{2}M^{-s} - s\int_M^\infty B_1(\{x\})x^{-s-1}\,\mathrm{d}x
\end{aligned}
$$

and so, after a spot of rearranging,

$$\zeta(s) = \sum_{n=1}^{M} \frac{1}{n^s} + \frac{M^{-s+1}}{s-1} - \tfrac{1}{2}M^{-s} - s\int_M^\infty B_1(\{x\})x^{-s-1}\,\mathrm{d}x$$

Now observe that $|B_1(\{x\})| \leq \frac{1}{2}$ and $|x^{-s-1}| = x^{-\mathrm{Re}(s)-1}$, and so the last integral will exist for $\mathrm{Re}(s) > 0$. Substituting $s = 1/2 + \mathrm{i}t$, we therefore have

$$\zeta(\tfrac{1}{2} + \mathrm{i}t) = \sum_{n=1}^{M} \frac{1}{n^{1/2+\mathrm{i}t}} + \frac{M^{1/2-\mathrm{i}t}}{-1/2+\mathrm{i}t} - \frac{1}{2M^{1/2+\mathrm{i}t}} + E$$

where the naive bound would suggest

$$|E| < \sqrt{\tfrac{1}{4} + t^2}\int_M^\infty \tfrac{1}{2}x^{-3/2}\,\mathrm{d}x = \frac{\sqrt{\frac{1}{4} + t^2}}{\sqrt{M}}$$

and thus to have $|E| < 0.01$ one might think we need $M \approx 10,000t^2$.

However, applying the improved Euler Maclaurin summation formula with just one extra integration by parts, we obtain

$$\zeta(\tfrac{1}{2} + \mathrm{i}t) = \sum_{n=1}^{M} \frac{1}{n^{1/2+\mathrm{i}t}} + \frac{M^{1/2-\mathrm{i}t}}{-1/2 + \mathrm{i}t} - \frac{1}{2M^{1/2+\mathrm{i}t}} + \frac{1/2 + \mathrm{i}t}{12} M^{-3/2-\mathrm{i}t} + R$$

where

$$R = -\frac{1}{2}(\tfrac{1}{2} + \mathrm{i}t)(\tfrac{3}{2} + \mathrm{i}t) \int_M^{\infty} B_2(\{x\}) x^{-5/2-\mathrm{i}t} \, \mathrm{d}x$$

and so the true size of E is “only” about $tM^{-3/2}$ (that is, the first omitted term), a considerable reduction. That is, we have

$$\zeta(\tfrac{1}{2} + \mathrm{i}t) = \sum_{n=1}^{M} \frac{1}{n^{1/2+\mathrm{i}t}} + \frac{M^{1/2-\mathrm{i}t}}{-1/2 + \mathrm{i}t} + O\left(\frac{1}{\sqrt{M}}\right) + O\left(\frac{|t|}{M^{3/2}}\right)$$

Note that if $M = t$ both error terms give $O(t^{-1/2})$.

Using $M = 1000$ and $t = 14.1$ we have $1/\sqrt{M} = 0.0316228$ and $t/M^{3/2} = 0.000449043$ so we keep the first error term but drop the second one, and estimating zeta using

$$\zeta(\tfrac{1}{2} + \mathrm{i}t) \approx \sum_{n=1}^{M} \frac{1}{n^{1/2+\mathrm{i}t}} + \frac{M^{1/2-\mathrm{i}t}}{-1/2 + \mathrm{i}t} - \frac{1}{2M^{1/2+\mathrm{i}t}}$$

We find that

$$\zeta(1/2 + 14.1\mathrm{i}) \approx 0.00470009 - 0.0270211\mathrm{i}$$

and since $\theta(14.1) = -1.74272$ we have

$$Z(14.1) \approx e^{-1.74272\mathrm{i}}(0.00470009 - 0.0270211\mathrm{i}) = -0.0274269 - 8.01765 \times 10^{-6}\mathrm{i}$$

(ignore the spurious imaginary part).

Similarly, for $t = 14.2$,

$$\zeta(1/2 + 14.2\mathrm{i}) \approx -0.00679108 + 0.0516247\mathrm{i}$$

and since $\theta(14.2) = -1.70214$ we have

$$Z(14.2) \approx e^{-1.70214\mathrm{i}}(-0.00679108 + 0.0516247\mathrm{i}) = 0.0520695 - 2.85538 \times 10^{-5}\mathrm{i}$$

Thus there is a sign change of the real function $Z(t)$ between $t = 14.1$ and $t = 14.2$, and this is a zero of the zeta function.

2.4 The Riemann-Siegel formula

Another way to calculate $Z(t)$ quickly and accurately in order to find these sign-changes was found by Riemann himself, but never published. Siegel discovered it 50 years after his death in his unpublished notes, which he wrote up for publication in 1932. It is now known as the Riemann-Siegel formula, and is superior to Euler Maclaurin for moderately large values of t.

Theorem 2.2 (Riemann-Siegel formula)**.**

$$Z(t) = 2 \sum_{n=1}^{\lfloor \sqrt{t/2\pi} \rfloor} \frac{\cos(\theta(t) - t \log n)}{\sqrt{n}} + O(t^{-1/4})$$

where $\theta(t)$ is given by equation (2.1).

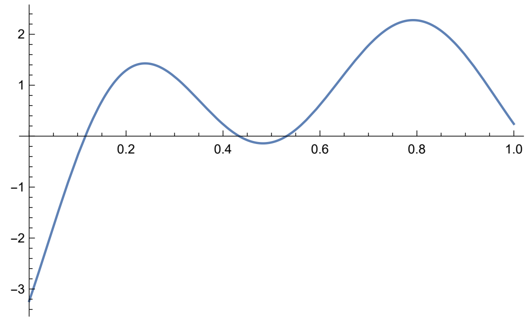


Figure 2.6: Plot of the Riemann-Siegel approximation to $Z(10^9 + x)$ with $0 \le x \le 1$.

This has the advantage of needing fewer summands (only about $\sqrt{t}$) than Euler-Maclaurin to get a decent error term, and also has the advantage of being manifestly real.

Remark. I only intend on giving a heuristic overview of this result. The actual proof by Riemann involves a fair chunk of complex analysis, as is described by Berry and Keating as "a remarkable achievement, because although it was one of the first applications of his method of steepest descent for integrals it was more sophisticated than most applications today, in that the saddle about which the integrand is expanded is accompanied by an infinite string of poles."

The approximate functional equation (which can be derived by Poisson summation — add the details?) tells us that if $xy = \frac{t}{2\pi}$ and if $s = \sigma + it$ is near the critical line, then

$$\zeta(s) = \sum_{n \le x} \frac{1}{n^s} + \chi(s) \sum_{n \le y} \frac{1}{n^{1-s}} + O(t^{-1/4})$$

Putting $\sigma = 1/2$, choosing $x = y = \sqrt{t/2\pi}$ and recalling that $\sqrt{\chi(1/2 - \mathrm{i}t)} = e^{\mathrm{i}\theta(t)}$, so that $\chi(1/2 + \mathrm{i}t) = e^{-2\mathrm{i}\theta(t)}$, we see that

$$\begin{aligned} Z(t) &= e^{\mathrm{i}\theta(t)} \zeta(\tfrac{1}{2} + \mathrm{i}t) \\ &= e^{\mathrm{i}\theta(t)} \sum_{n \le \sqrt{t/2\pi}} \frac{1}{n^{1/2+\mathrm{i}t}} + e^{-\mathrm{i}\theta(t)} \sum_{n \le \sqrt{t/2\pi}} \frac{1}{n^{1/2-\mathrm{i}t}} + O(t^{-1/4}) \\ &= \sum_{n \le \sqrt{t/2\pi}} \frac{e^{\mathrm{i}(\theta(t) - t\log n)}}{\sqrt{n}} + \sum_{n \le \sqrt{t/2\pi}} \frac{e^{\mathrm{i}(-\theta(t) + t\log n)}}{\sqrt{n}} + O(t^{-1/4}) \\ &= 2 \sum_{n \le \sqrt{t/2\pi}} \frac{\cos(\theta(t) - t\log n)}{\sqrt{n}} + O(t^{-1/4}) \end{aligned}$$

as required.

Figure 2.6 shows a plot of the Riemann-Siegel approximation to $Z(10^9 + x)$ with $0 \le x \le 1$. Note the zeros are still found by sign changes easily. At this height, the error in the approximation is about 0.006 and there are just over 12,600 terms in the sum.

2.5 Turing's Theorem (a Manchester connection)

In his last research paper before he died, Turing worked on numerically finding zeros of the zeta function using the Manchester Mark I computer. He wrote

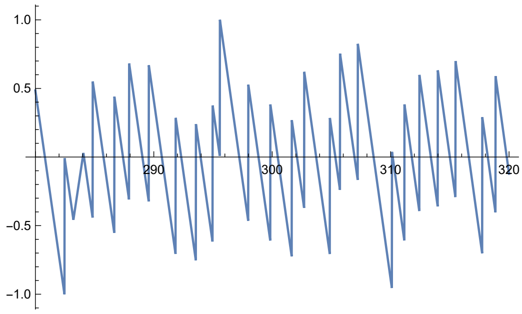


Figure 2.7: Plot of $S(t)$ for $280 \leq t \leq 320$.

> In June 1950 the Manchester University Mark 1 Electronic Computer was used to do some calculations concerned with the distribution of the zeros of the Riemann zeta-function. It was intended in fact to determine whether there are any zeros not on the critical line in certain particular intervals. The calculations had been planned some time in advance, but had in fact to be carried out in great haste. If it had not been for the fact that the computer remained in serviceable condition for an unusually long period from 3 p.m. one afternoon to 8 a.m. the following morning it is probable that the calculations would never have been done at all. As it was, the interval $2\pi 63^2 < t < 2\pi 64^2$ was investigated during that period, and very little more was accomplished.

Whereas his improvement in the number of zeros calculated was, indeed, modest, the method he introduced in this paper (to calculate $N(T)$ exactly and thus *prove* you had found all the zeros) was a vast improvement on the previous ad-hoc methods employed. Indeed, it is still the main method in use today. Turing wrote "The procedure was such that if it had been accurately followed, and if the machine made no errors in the period, then one could be sure that there were no zeros off the critical line in the interval in question.".

Remark. Turing's aim was actually to *disprove* the Riemann Hypothesis. He went on to write "The calculations were done in an optimistic hope that a zero would be found off the critical line, and the calculations were directed more towards finding such zeros than proving that none existed."

Recall that $S(T)$ is the error in the number of zeros lying in the critical strip up to height T. Turing showed that for any $h > 0$ and $T > 168\pi$ one had

$$\left| \int_T^{T+h} S(t) \mathrm{d}t \right| \leq 2.3 + 0.128 \log \frac{T+h}{2\pi}$$

That is, $S(t)$ oscillates very closely around 0 (see Figure 2.7).

If you are counting zeros by sign changes and happen to miss a sign change (so miss two zeros), then a implied value of $S(T)$ will start oscillating around -2 instead (see bottom right in Figure 2.8).

Thus, to prove no zeros have been missed, compute the numerically derived $\int_T^{T+h} S(t)\mathrm{d}t$ and check whether it's bigger than the bound Turing proved. (Thus you will actually need to compute the zeros in $[T, T+h]$ above your stopping point of T, but that is a small price to pay).

If $h \approx \log T$ then that's sufficient to tell whether $S(t)$ is oscillating around 0 or around -2. (This is done by contradiction: Assume a zero has been missed, and throw it in to the numerically derived $N(T)$. If the integral of $S(T)$ gets bigger than its known upper bound, that contradicts the assumption of a missing zero. That is, you have proven you have found *all* the zeros of the zeta function up to height T, and they all lie exactly on the critical line).

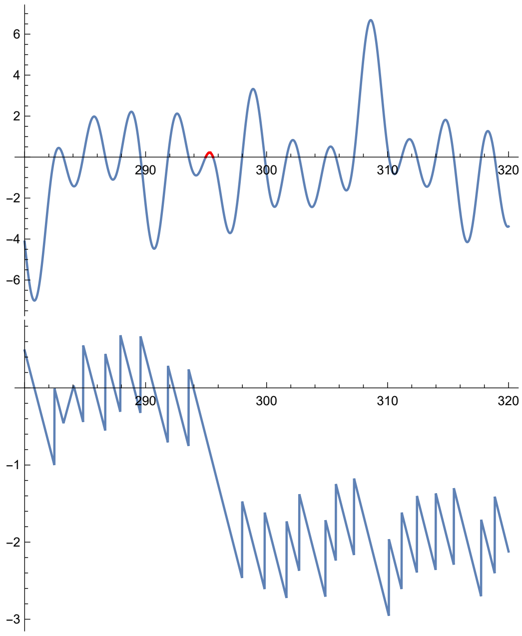


Figure 2.8: Plot of $Z(t)$ for $280 \leq t \leq 320$, and a plot of the deduced $S(t)$ assuming the sign change between two zeros at 294.965 and 294.965 (shown in red) was missed.

Having got to the end of this chapter, we are now able to calculate zeros of the zeta function lying on the critical line, and know we have not missed any. Super-computers and distributed computing has enabled tables, such as `http://www.lmfdb.org/zeros/zeta/` or `http://www.dtc.umn.edu/~odlyzko/zeta_tables/index.html` containing billions of zeros to be collated. In the last lecture we will look at the distribution of these zeros, and in particular in the gaps between them, and see how that is modelled by random matrix theory.

Chapter 3

Lecture 3: Connections between the Riemann zeta function and random matrix theory

Having collected massive sets of zeros, one can then ask questions about their distribution — do they come in clumps, or are they evenly spaced out? (Knowing $N(T) = \frac{T}{2\pi}\log\frac{T}{2\pi e} + O(\log T)$ tells us that the average gap between zeros is $\frac{2\pi}{\log(T/2\pi)}$, but these questions are deeper than that as they concern the distribution of the spacing of the zeros, not just the mean).

In 1972, Hugh Montgomery, then a graduate student at Cambridge, was visiting Atle Selberg at the Institute for Advanced Study at Princeton. Montgomery wanted to discuss his conjecture for how he thought the gaps between zeros of the Riemann zeta function might be distributed. At tea that afternoon, Montgomery was introduced to the physicist Freeman Dyson. After Montgomery told him his conjecture, Dyson quickly recognized Montgomery's results as being the same as the pair correlation of eigenvalues of random Hermitian matrices. In 2006 Montgomery recalled recognising the importance of the connection just made: "Just by chance this conversation took place.... This happened even before I had published the paper. I knew it was important and worth following up."

3.1 Flavours of random matrix theory

Recall that the structure of a matrix is important. For instance, if a matrix is hermitian (that is, $M = M^\dagger$) then it has real eigenvalues (see figure 3.1), whereas if a matrix is unitary (that is, $MM^\dagger = I$), then all its eigenvalues lie on the unit circle (that is, they all have magnitude 1), (see Figure 3.2).

A random matrix is a matrix-valued random variable, where the matrix has certain properties imposed (such as being hermitian, or unitary, or whatever), and with a certain probability measure imposed.

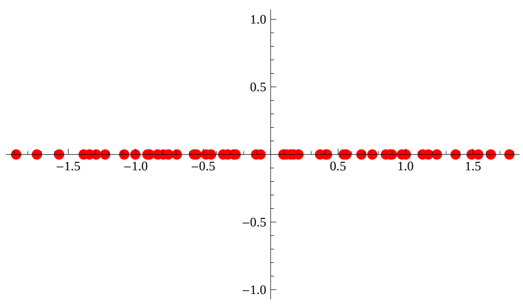


Figure 3.1: The eigenvalues of a hermitian matrix

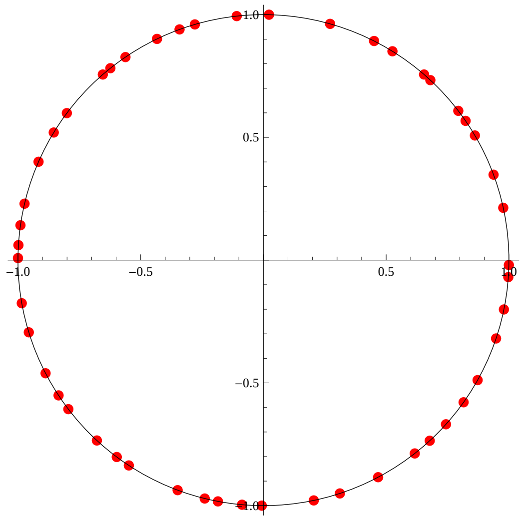


Figure 3.2: The eigenvalues of a unitary matrix

- GUE:
  - M be an $N \times N$ hermitian matrix with entries
$$M_{i,j} = \begin{cases} \mathcal{N}(0, \frac{1}{2N}) + \mathrm{i}\mathcal{N}(0, \frac{1}{2N}) & \text{for } i > j \\ \mathcal{N}(0, \frac{1}{N}) & \text{for } i = j \\ M^*_{j,i} & \text{for } i < j \end{cases}$$
    where $\mathcal{N}(0, \sigma^2)$ is a Gaussian random variable with mean zero and variance σ^2.
  - The measure on this set of matrices is invariant under unitary transforms.
- Unitary with Haar measure
  - The unique measure on the unitary group invariant under the action of any unitary transform, i.e. for any fixed V, $VU \overset{\text{law}}{=} U$
  - Can be realised via QR decomposition of a matrix with complex Gaussian entries

The choice of randomness matters, as can be seen by taking two unitary matrices, one chosen with Haar measure and one chosen by picking the eigenvalues independently, uniformly at random on the unit circle (see Figure 3.3).

Random matrix theory turns out to have many applications

- Nuclear physics (energy spectra of heavy nuclei).
- Quantum Chaos (is a system classically chaotic or integrable?)
- Genetics and population biology
- Correlation matrix of time series of stock prices
- Sea level and atmospheric pressure
- Bus arrival times in Cuernavaca, Mexico & spacing between cars parked in London.
- Longest increasing subsequence and solitaire

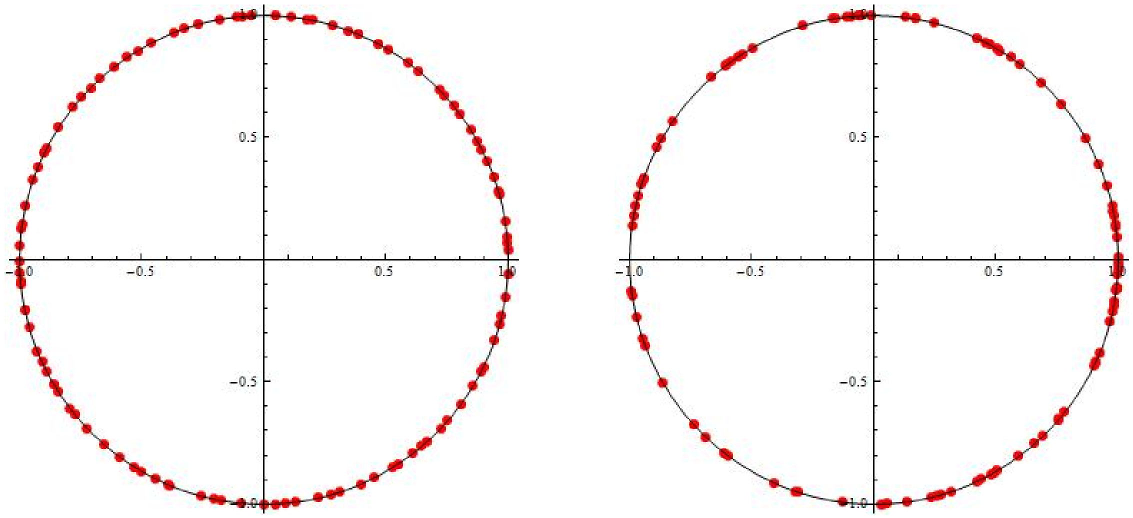


Figure 3.3: The eigenvalues of a 100×100 Haar distributed random unitary matrix, compared with 100 points placed independently at random on the unit circle

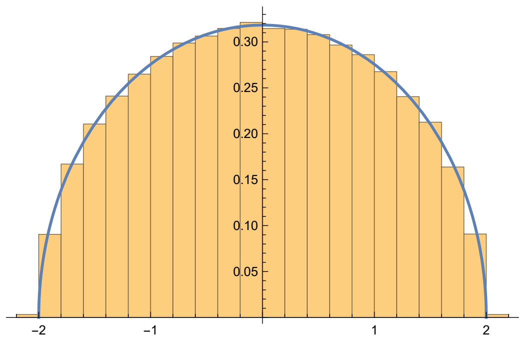


Figure 3.4: Wigner's semi circle law.

- And many, many other applications.

However, we will only concentrate on one application: That to the zeros of the Riemann zeta function.

3.2 Simple questions for GUE matrices

One sample simple question is: For the GUE matrices, where all the eigenvalues lie on the line, how many eigenvalues would you expect to find between a and b? This was one of the original questions asked and solved by Wigner, and it turns out to be a semicircle.

Let $\lambda_1, \ldots, \lambda_N$ be the eigenvalues of a hermitian matrix. Wigner showed that $-2 \le a \le b \le 2$, then

$$\mathbb{E}\left[\#\left\{j : \lambda_j \in [a,b]\right\}\right] \to \frac{1}{2\pi} \int_a^b \sqrt{4 - x^2}\, \mathrm{d}x$$

where the expectation is taken over the GUE. (See Figure 3.4).

Another question one could concerns the gaps between the eigenvalues. Because the eigenvalues are real, they can be ordered: $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_N$. One can then look at the gaps between neighbouring eigenvalues

$$s_i = \lambda_{i+1} - \lambda_i$$

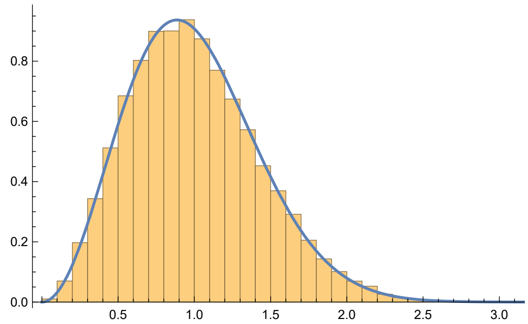


Figure 3.5: The nearest neighbour spacing for eigenvalues 1,000 GUE matrices and the Wigner surmise

One can then ask for the distribution of the s_i (after scaling so the average gap is 1 — this enables large N limits to be taken). The actual result is complicated (and can be given in terms of solutions to certain Painlavé differential equations) but Wigner came up with a good approximation for the distribution of gaps, known as the Wigner surmise for the GUE:

$$p(x) = \frac{32}{\pi^2} x^2 e^{-4/\pi x^2}$$

Figure 3.5 shows the Wigner surmise, plotted against the empirical (scaled) nearest neighbour spacing distribution for a thousand 50×50 GUE matrices.

3.3 Unitary matrices

Denote the eigenvalues of $U \in \mathcal{U}(N)$ by $\exp(\mathrm{i}\theta_1), \ldots, \exp(\mathrm{i}\theta_N)$. Weyl proved that the joint probability density of eigenangles is

$$P_N(\theta_1, \ldots, \theta_N) = \frac{1}{N!(2\pi)^N} \prod_{1 \le j < k \le N} \left| e^{\mathrm{i}\theta_k} - e^{\mathrm{i}\theta_j} \right|^2$$

This density can be reexpressed in terms of determinants, which enables all the useful tools and techniques from matrix analysis to come into play.

We need to introduce Vandermonde determinants: For any set of complex numbers $z_1, \ldots, z_N$ the vandermonde determinant is

$$\det\left(z_k^{j-1}\right)_{1 \le j,k \le N} = \det \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ z_1 & z_2 & z_3 & \ldots & z_N \\ z_1^2 & z_2^2 & z_3^2 & \ldots & z_N^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_1^{N-1} & z_2^{N-1} & z_3^{N-1} & \ldots & z_N^{N-1} \end{pmatrix}$$

Lemma 3.1. *We have*

$$\det\left(z_k^{j-1}\right)_{1 \le j,k \le N} = \prod_{1 \le j < k \le N} \left(z_k - z_j\right)$$

Proof. Observe that both sides of the equation are homogeneous polynomials of degree $N(N-1)/2$. Secondly, note that both sides vanish whenever $z_j = z_k$ (and since there are N variables, this accounts for $N(N-1)/2$ roots). Thus we have identified the two polynomials as the same up to some constant multiplicative factor. Comparing the coefficient of the diagonal term in the matrix fixes that constant as 1. □

By rearranging the Vandermonde matrices enables us to prove

Lemma 3.2. *We have*

$$P_N(\theta_1, \dots, \theta_N) = \frac{1}{N!} \det \left\{ K_N(\theta_k - \theta_j) \right\}_{1 \le j,k \le N}$$

where

$$K_N(\theta) = \frac{1}{2\pi} \frac{\sin(N\theta/2)}{\sin(\theta/2)}$$

Proof. We have

$$\prod_{1 \le j < k \le N} \left| e^{\mathrm{i}\theta_k} - e^{\mathrm{i}\theta_j} \right|^2 = \det \left(e^{\mathrm{i}(n-1)\theta_j} \right)_{1 \le n,j \le N} \det \left(e^{-\mathrm{i}(n-1)\theta_k} \right)_{1 \le n,k \le N}$$

(The dummy variable n is used twice here, once in both determinants, on purpose). Using the fact that $\det A = \det A^t$ and that $\det A \det B = \det(AB)$ and explicitly multiplying out the matrices, we see that this equals

$$\det \left(\sum_{n=1}^{N} e^{\mathrm{i}(n-1)(\theta_j - \theta_k)} \right)_{1 \le j,k \le N}$$

Finally, the inner sum is a simple geometric series, so

$$\sum_{n=1}^{N} e^{\mathrm{i}(n-1)\theta} = \frac{e^{\mathrm{i}N\theta} - 1}{e^{\mathrm{i}\theta} - 1} = e^{\mathrm{i}(N-1)\theta/2} \frac{\sin(N\theta/2)}{\sin(\theta/2)}$$

where the last line comes from factoring $e^{\mathrm{i}N/2\theta}$ out from the numerator and $e^{\mathrm{i}\theta/2}$ from the denominator (the remaining terms being the definition of sine). Thus we see that

$$\frac{1}{(2\pi)^N} \prod_{1 \le j < k \le N} \left| e^{\mathrm{i}\theta_k} - e^{\mathrm{i}\theta_j} \right|^2 = \det \left(e^{\mathrm{i}(N-1)(\theta_j - \theta_k)} K_N(\theta_j - \theta_k) \right)_{1 \le j,k \le N}$$

from pulling the factor of $1/2\pi$ into each row, factoring out $e^{\mathrm{i}(N-1)\theta_j}$ from the jth row, and $e^{-\mathrm{i}(N-1)\theta_k}$ from the kth column, and noting their product equals 1. □

Define

$$R_n^{(N)}(\theta_1, \dots, \theta_n) = \frac{N!}{(N-n)!} \int \cdots \int_{-\pi}^{\pi} P_N(\theta_1, \dots, \theta_N) \, \mathrm{d}\theta_{n+1} \dots \mathrm{d}\theta_N$$

which can be thought of a related to the probability density of finding eigenangles (regardless of labelling) at each of the angles $\theta_1, \theta_2, \dots, \theta_n$, ignoring the position of the remaining $N-n$ eigenangles.

It can be shown that

$$R_n^{(N)}(\theta_1, \dots, \theta_n) = \det \left\{ K_N(\theta_j - \theta_k) \right\}_{1 \le j,k \le n}$$

$R_1^{(N)}(\theta_1) = \frac{N}{2\pi}$ is just the density of eigenvalues on the unit circle. Note it is independent of position, θ_1. Therefore to get a non-trivial limit as $N \to \infty$ the eigenangles must be scaled by their density. Writing $x_j = \frac{N}{2\pi}\theta_j$, then

$$\begin{aligned} R_n(x_1, \dots, x_n) &= \lim_{N \to \infty} R_n^{(N)} \left(\frac{2\pi}{N} x_1, \dots, \frac{2\pi}{N} x_n \right) \\ &= \det \left\{ K(x_j - x_k) \right\}_{1 \le j,k \le n} \end{aligned}$$

where

$$K(x) = \frac{\sin(\pi x)}{\pi x}$$

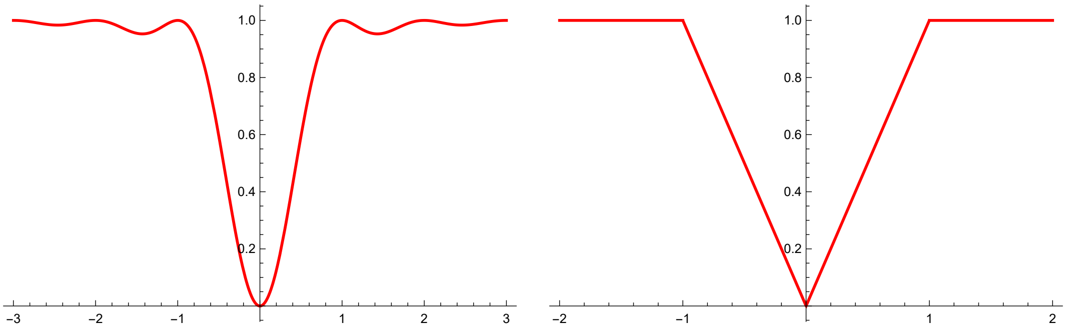


Figure 3.6: The pair correlation function and its Fourier transform

In particular, for $\alpha < \beta$

$$\lim_{N\to\infty} \mathbb{E}_N \frac{1}{N} \# \left\{ \theta_m, \theta_n \ : \ \alpha \leq (\theta_m - \theta_n) \frac{N}{2\pi} \leq \beta \right\} = \int_\alpha^\beta R_2(x,0)\, \mathrm{d}x$$

$$= \int_\alpha^\beta \left(1 - \left(\frac{\sin(\pi x)}{\pi x} \right)^2 \right) \mathrm{d}x + \delta(\alpha, \beta)$$

where $\mathbb{E}_N$ denotes expectation with respect to the Haar measure on $\mathcal{U}(N)$, and where $\delta(\alpha,\beta) = 1$ if $\alpha \leq 0 \leq \beta$ and equals 0 otherwise. $R_2(x,0)$ is called the two-point correlation function or pair correlation function.

Another way to consider this result is to consider smooth function f. We have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j,k} f\left((\theta_j - \theta_k) \tfrac{N}{2\pi} \right) = \int_{-\infty}^{\infty} f(x) \left(1 - \left(\frac{\sin(\pi x)}{\pi x} \right)^2 + \delta(x) \right) \mathrm{d}x$$

$$= \int_{-\infty}^{\infty} \hat{f}(u)\, (\min(1,|u|) + \delta(u))\, \mathrm{d}u$$

where the second line follows from the Plancherel's identity about Fourier transforms, $\int \overline{f} g = \int \overline{\hat{f}} \hat{g}$. See figure 3.6 for a plot of this function. It is that that was the first connection made with the distribution of the zeros of Riemann zeta function.

3.4 Montgomery's pair correlation conjecture

In 1973 Hugh Montgomery was considering the distribution of the zeros of the Riemann zeta function, in particular in trying to prove that there are a positive proportion of zeros which are closer together than 1/2 times their average spacing, as this would have significant consequences concerning the class number problem (which is an old problem from around the time of Gauss, determining the negative quadratic fields $\mathbb{Q}(\sqrt{d})$ with $d < 0$ have class number $h(d) = n$).

The $h = 1$ case was solved by Baker, Stark, Heegner in 1966, 1967 and 1952. Imaginary quadratic fields with class number 1 (i.e. those with unique factorisation) are $d = -1, -2, -3, -7, -11, -19, -43, -67$, and -163 only.

Montgomery and Weinberger (1973) solved the problem for $h = 2$ (and independently Stark, too, by a different method). The Montgomery-Weinberger result is based on showing that if $h(d)$ is small then the low zeros of related L-functions lie on the critical line and are very regularly spaced. Finding zeros closer together than they "should be" rules out $h(d)$ being small for that d.

Montgomery defined

$$F(\alpha) = F(\alpha, T) = \frac{1}{\frac{T}{2\pi}\log T} \sum_{0\le\gamma,\gamma'\le T} T^{\mathrm{i}\alpha(\gamma-\gamma')} w(\gamma-\gamma')$$

where

$$w(u) = \frac{4}{4+u^2}$$

is a weight function (which turns up naturally in the analysis).

He proved:

Theorem 3.3 (Montgomery)**.** *If the RH is true then* $F(\alpha)$ *is a real, even, non-negative function, and uniformly for* $0 \le \alpha \le 1$

$$F(\alpha) = \alpha + o(1) + (1 + o(1))T^{-2\alpha}\log T$$

(Actually, he proved it for the case $\alpha \le 1-\epsilon$, the gap being filled in jointly with Goldston).

This is clearly useful for evaluating sums over differences of zeros. If we let $r \in L^1$ and set the Fourier transform to be

$$\hat{r}(\alpha) = \int_{-\infty}^{\infty} r(x) e^{2\pi\mathrm{i}\alpha x}\mathrm{d}x$$

then multiplying $F(\alpha)$ by $\hat{r}(\alpha)$ and integrating we get

$$\frac{1}{\frac{T}{2\pi}\log T} \sum_{0<\gamma,\gamma'\le T} r\left((\gamma-\gamma')\frac{\log T}{2\pi}\right) w(\gamma-\gamma') = \int_{-\infty}^{\infty} \hat{r}(\alpha)F(\alpha)\mathrm{d}\alpha$$

By choosing a suitable function r, whose Fourier transform was supported in $[-1,1]$ Montgomery was able to prove

Theorem 3.4. *Let*

$$\lambda = \liminf_{n\to\infty}(\gamma_{n+1}-\gamma_n)\frac{\log\gamma_n}{2\pi}$$

then $\lambda \le 0.6072$.

Finally, based on a heuristic analysis using the Hardy-Littlewood k-tuple conjecture, he conjectured

Conjecture 3.5. *For any fixed bounded* $M > 1$*, then uniformly for* $1 \le \alpha \le M$*,*

$$F(\alpha) = 1 + o(1)$$

By taking the Fourier transform, this then implies

Conjecture 3.6 (Pair Correlation Conjecture)**.**

$$\frac{1}{\frac{T}{2\pi}\log T} \sum_{\substack{0<\gamma,\gamma'\le T\\ 0<\gamma'-\gamma\le 2\pi\beta/\log T}} 1 \sim \int_0^\beta \left(1 - \left(\frac{\sin(\pi x)}{\pi x}\right)^2\right)\mathrm{d}x$$

This, you will recall is exactly the same as for unitary matrices chosen with Haar measure (or for GUE matrices). (See Figure 3.7).

Since that connection was established, it has been checked extensively — numerically (initially and foremost by Odlyzko in the 1990's, who provided graphs such as Figures 3.7 and 3.8), heuristically and theoretically.

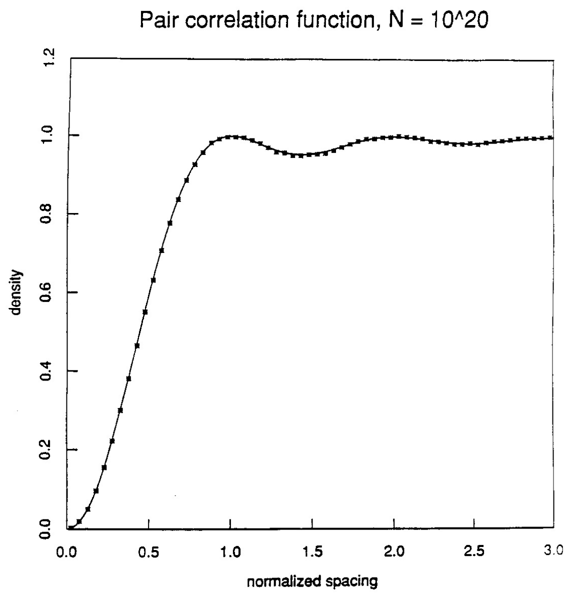


Figure 3.7: The empirical pair correlation for 8×10^6 zeros of the Riemann zeta function near the 10^{20}th zero

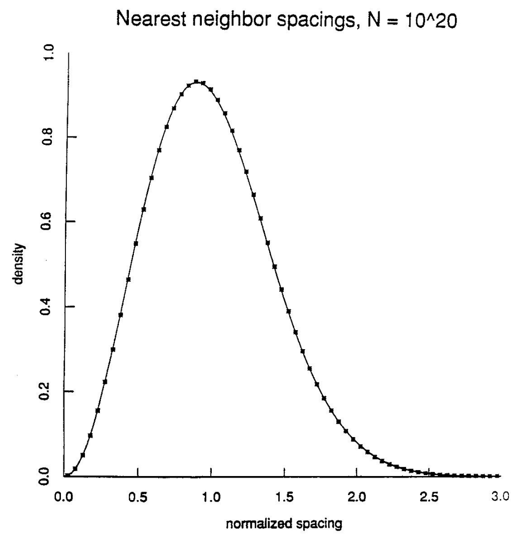


Figure 3.8: The empirical nearest neighbour spacing for the zeros of the Riemann zeta function near the 10^{20}th zero

The connection was further developed by Jon Keating and Nina Snaith who discovered that the Riemann zeta function itself (not just its zeros) could be modelled using random matrix theory, and their results paved the way for the first reasonable conjecture on the moments of the zeta function to be created, something number theorists had been searching for since the 1910's. In another direction, there are other L-functions, not just the Riemann zeta function, all sharing similar properties. Considering their low lying zeros, and averaging over "families" of L-functions, Katz and Sarnak showed they could be modelled by other types of random matrices, the particular type depending upon the family.