Analytic Number Theory: Part III
Modern Knowledge

Henry Twiss

March 7, 2023

Prerequisites

The purpose of this text is to serve as an overview to important techniques and major philosophes one should be familar with in analytic number theory. Accordingly, we assume that the reader has a decent understanding of the theory already. For reference, everything discussed in *Analytic Number Theory: Part I An Introduction to L-functions* and *Analytic Number Theory: Part II Trace Formulas & Multiple Dirichlet Series* will be assumed. Unlike the previous texts, the following is an overview of the subject and many proofs will be omitted.

Contents

Contents

Chapter 1

Preliminaries

There is quite a bit of additional material that readers should be familar with before atempting to understand the material. A good selection is the following:

- Todo: [XXX]

Many of these topis are not directly mentioned in classical number theory, so in the interest of not being too overwhelming, this chapter is dedicated to the basics of these topics. Anyone well-versed in these topics is still encouraged to read through this chapter thoroughly. We will use the results presented here without reference in the following chapters unless it is a matter of clarity.

1.1   Todo: [XXX]

Chapter 2

Well-known Results

2.1 Stastical Information About $\zeta(s)$

We collect some important stastical information about the Riemann zeta function $\zeta(s)$ with citations whenever possible:

- At least 40% o the non-trivial zeros of $\zeta(s)$ lie on the critial line (see Todo: [cite]).

- The first $1.5 \times 10^9 + 1$ zeros of $\zeta(s)$ lie on the critial line (see Todo: [cite]).

Chapter 3

Techniques

3.1 Todo: [The Approximate Funcational Equation]

If $L(s, f)$ is a Selberg class L-function then the **approximate functional equation** is a compromise between the functional equation for $L(s, f)$ and expressing $L(s, f)$ as a Dirichlet series which is valid inside the cirital strip:

Theorem 3.1.1 (The approximate functional equation). *Let $L(s)$ be a Selberg class L-function and $g(u)$ be any even holomorphic function bounded on the strip $|\mathrm{Re}(u)| < 4$ and such that $g(0) = 1$. Let $X > 0$. Then for s in the critial strip,*

$$L(s, f) = \sum_{n \geq 1} \frac{a_f(n)}{n^s} V_s\left(\frac{n}{X\sqrt{q(f)}}\right) + \varepsilon(f, s) \sum_{n \geq 1} \frac{a_{\cong f}(n)}{n^{1-s}} V_{1-s}\left(\frac{nX}{\sqrt{q(f)}}\right) + R$$

, where $V_s(y)$ is a smooth function defined by

$$V_s(y) = \frac{1}{2\pi i} \int_{\mathrm{Re}(u)=3} y^{-u} g(u) \frac{\gamma(f, s+u)}{\gamma(f, s)} \frac{du}{u},$$

and

$$\varepsilon(f, s) = \varepsilon(f) q(f)^{\frac{1}{2} - s} \frac{\gamma(f, 1-s)}{\gamma(f, s)}.$$

Moreover, the remainder R is zero if $\Lambda(f, s)$ is entire and otherwise

$$R = \left(\operatorname*{Res}_{u=1-s} + \operatorname*{Res}_{u=-s}\right) \frac{\Lambda(f, s+u)g(u)}{q^{\frac{s}{2}}\gamma(f, s)} X^u.$$

The $g(u)$ is a test function which so that $V_s(y)$ has the effect of smoothing out the two sums on the right-hand side. Alternatively, it can be a cutoff function by choosing Todo: [$g(u) = ...$]. This technique was first introduced by Hardy and Littlewood in Todo: [year] (see Todo: [cite]) to obtain asymptotics for the second moment of the zeta function.

Chapter 4

Philosophies

4.1 The Katz-Sarnak Philosophy

The **Katz-Sarnak philosophy** is the idea that stastics of various kinds of L-functions should, in the limit, match stastics for random matrices coming from some particular classical compact group. One starts with some class of zeros to look at, say high zeros of an individal L-function or low zeros for some family (this notion of family will become more precise) of related L-functions. Next, one normalizes the nontrivial zeros being inspected. For example, the high zeros of an individual L-function on the critial line tend to cluster so a normalization is chosen such that the average spacing is 1. It is this rescaled zeros that one works with when estimating the stastics. Then some class of test functions are introduced in order to carry out the stastical calculations in order to reveal similarity with some class of random matrices.

For a more concrete case, suppose that the stastical information we are after is a moment of a family of related L-functions. A classical argument would invoke the approxiate functional equation for these L-functions and then average the Dirichlet coefficients over the family. This averaging process is not unlike a harmonic detection device. There are usually two pieces, a simple one and more complicated one. The simple piece is usually enough to obtain the main term (or first two main terms) of the asymptotic expansion. The terms in this piece are usually referred to as the **diagonal terms**. Unforuntely, with higher moments the harmonic detection devices becomes increasingly more complicated and more terms, even off diagonal ones, are necessary in order to obtain the main term.

The Work of Montgomery & Dyson

The start of the connection between random matrix theory and analytic number theory was at Princeton in the 1907s via discussions between H.L. Montgomery and F.J. Dyson. In particular there were similarities between stastical information about the distribution of the zeros of $\zeta(s)$ and calculations in random matrix theory about unitary matrices (see Todo: [cite]). More precisely if $\frac{1}{2} + it$ is a zero of $\zeta(s)$ on the critial line, then we define the **unfolded zero** to be

$$w = t\frac{1}{2\pi}\log\left(\frac{t}{2\pi}\right).$$

If we let t_n denote the height of the n-th zero and similarly for w_n, then we can state the following well-known result (see Todo: [cite] for a proof):

Theorem 4.1.1.
$$\lim_{W\to\infty}\frac{1}{W}\#\left\{w_n \le W\right\} = 1.$$

We interpret Theorem 4.1.1 as saying that the zeros of $\zeta(s)$ are logarithmically dense on the critial line. That is, they tend to accumulate as one moves up the crtical line. The purpose of using the unfolding the zeros is to dispense with this clustering. The effect is that the unfolded zeros have unit spacing between each other on average since the limit in the theorem is 1. Now we can consider the **two-point correlation function**

$$F_\zeta(\alpha, \beta; W) = \frac{1}{W} \#\{w_n, w_m \in [0, W] : \alpha \le w_n - w_m \le \beta\},$$

for any real $\alpha < \beta$ and $W > 0$. What this function measures is the probability of how close pairs of zeros tend to be with respect to some fixed distance $(\beta - \alpha)$ be up to some height (W). That is, the correlation be distances of zeros. We want to understand if the limiting distribution

$$F_\zeta(\alpha, \beta;) = \lim_{W \to \infty} F_\zeta(\alpha, \beta; W),$$

exists and what we can say about it. Now Montgomery conjectured the following (see Todo: [cite]):

Conjecture 4.1.1. $F_\zeta(\alpha, \beta;)$ *exists and moreover,*

$$F_\zeta(\alpha, \beta;) = \int_\alpha^\beta 1 - \left(\frac{\sin(\pi x)}{\pi x} \right)^2 + \delta(x)\, dx,$$

where $\delta(x)$ is the Dirac delta distribution.

Suprisingly, Conjecture 4.1.1 mimics a similar situation in random matrix theory. One can consider and $N \times N$ unitary matrix $A \in U(N)$ with eigenvalues $e^{i\theta_n}$ for $1 \le n \le N$ on the unit circle. The average density is cleary $\frac{N}{2\pi}$, but we can construct a two-point correlation function with resct to the eigenphases θ_n:

$$F(\alpha, \beta; A, N) = \frac{1}{N} \#\{\theta_n, \theta_m \in [0, W] : \alpha \le \theta_n - t_m \le \beta\}.$$

Since $U(N)$ has a natrual Haar measure dA we can compute the average of $F(\alpha, \beta; A, N)$ over the group:

$$F_U(\alpha, \beta; N) = \int_{U(N)} F(\alpha, \beta; A, N)\, dA.$$

If we take the limit as $N \to \infty$, F.J. Dyson showed the following:

Theorem 4.1.2.
$$F_U(\alpha, b) = \int_\alpha^\beta 1 - \left(\frac{\sin(\pi x)}{\pi x} \right)^2 + \delta(x)\, dx.$$

The right-hand side of Theorem 4.1.2 is exactly Montgomery's prediction in Conjecture 4.1.1 for the two-point correlation function for $\zeta(s)$. We can think of this as follows: the two-point correlation of the zeros of $\zeta(s)$ in the limit as we move up the critical line exactly match the two-point correlation of the eigenphases of unitary matrices in the limit of the size of the matrices. In short, stastical information about $\zeta(s)$ agrees with stastical information about the eigenvalues of unitary matrices.

Families of L-functions

Katz and Sarnak generalized the connection between stastical information of the zeros of $\zeta(s)$ and random unitary matrices to other L-functions. Moreover, they established a connection between specific families of similar L-functions and other compact matrix groups. The following expands upon this relationship. To each family of L-functions, yet to be explained, there is an associated **symmetry type** coming from a group or quotient of matrices. In this setting, these groups (not the quotients) are refered to as **matrix ensembles**:

Symmetry Type	Matrix Ensemble
Unitary	$U(N)$ the compact group of $N \times N$ unitary matrices
Orthgonal	$SO(N)$ the compact group of $N \times N$ special orthogonal matrices
Symplectic	$USp(2N)$ the compact group of $2N \times 2N$ unitary symplectic matrices
COE	$U(N)/O(N)$
CSE	$U(2N)/USp(2N)$

Now let $G(N)$ be any of the first three compact matrix groups: $U(N)$, $SO(N)$, or $USp(2N+1)$. Let dA denote the Haar measure and for any $A \in G(N)$ lable the eigenphases as θ_n for $1 \leq n \leq N$ in increasing order. One of the stastics that Katz and Sarnak investigated was the k-**th consecutive spacing**:

$$\mu_k(A)[a,b] = \frac{1}{N}\# \left\{ 1 \leq j \leq N : \frac{N}{2\pi}(\theta_{j+k} - \theta_j) \in [a,b] \right\},$$

for any $a < b$. They were able the prove the following result (see Todo: [cite]):

Theorem 4.1.3. *For any $a < b$ and $k \geq 1$,*

$$\lim_{N \to \infty} \int_{G(N)} \mu_k(A)[a,b]\,dA = \mu_k(\mathrm{CUE})[a,b].$$

That is, the average of $\mu_k(A)[a,b]$ for any of the matrix ensembles, in the limit of the size of the matrices, always resembles $\mu_k(\mathrm{CUE})[a,b]$ (it turns out that this is the same as $\mu_k(\mathrm{COE})[a,b]$ too so they all agree). On the other hand, there are stastics that depend on the particular ensemble chosen. For example we can define the distribution of the k-th eigenphase over $G(N)$ as follows:

$$\nu_k(G(N))[a,b] = dA\left(\left\{ A \in G(N) : \frac{\theta_k(A)N}{2\pi} \in [a,b] \right\} \right),$$

where $\theta_k(A)$ is the k-th eigenphase of A. Katz and Sarnak showed that

$$\nu_k(G) = \lim_{N \to \infty} \nu_k(G(N))[a,b],$$

exists for all $a < b$ but that ν_k depends upon the particular ensemble chosen (see Todo: [cite]). Heuristically, the stastic $\mu_k(A)[a,b]$ is a generalization of the two-point correlation function and so should mimic stastics about the distribution of the zeros of a single L-function lying high up on the critial line. That is, the stastics of **high zeros**. On the other hand, the stastic $\nu_k(G(N))[a,b]$ is modeling the distribution of a single zero of a family of related L-functions. That is, the stastics of **low zeros**. Katz and Sarnak then proposed that stastics regarding the distribution of high zeros of and single L-function should mimic those of $U(N)$, while stastics regarding the distribution of low zeros for a family of L-functions should mimic one of the matrix ensembles. For example, below are some well-studied families and their symmetry type:

Symmetry Type	Family
Unitary	$\{L(s+iy) : y \geq 0\}$ ordered by y where $L(s)$ is any Selberg class L-function
	$\{L(s,\chi) : \chi\}$ ordered by q where χ is a Dirichlet character modulo $q \geq 1$
Orthgonal	$\{L(s,f) : f \in \mathcal{S}_k(\Gamma_0(N)), k \geq 4\}$ ordered by k where $N \geq 1$ is fixed
	$\{L(s,f) : f \in \mathcal{S}_k(\Gamma_0(N)), N \geq 1\}$ ordered by N where $k \geq 4$ is fixed
Symplectic	$\{L(s,\chi_d) : d$ a fundamental discriminant$\}$ ordered by $|d|$ where $\chi_d(n) = \left(\dfrac{d}{n}\right)$
	$\{L(s,\text{sym}^2 f) : f \in \mathcal{S}_k(\Gamma_0(1))\}$ ordered by $k \geq 4$

The following two questions we need to address:

(1) What is the precisely meant by a family of L-functions?

(2) Given a family how do we determine its symmetry type?

The first question has an unfortunate answer. There is not yet a precise definition of a family of L-functions. Families are determined by if interesting stastical data arises from their study. However, it is generally believed that the family should be indexed by either a real parameter, such as $y > 0$ for the unitary family $\{L(s+iy) : y \geq 0\}$, or partially ordered by the conductor of the L-function as is the case for every other family in the table above.

As for the second question, the symmetry type can be determined but it is, in general, a difficult question. The method used by Katz and Sarnak is that for some families, one can define an analgous family over finite fields. The L-functions here are polynomials (in q^{-s} with q the order of the field), and the Grothendieck-Lefschetz trace formula implies that the L-functions are characteristic polynomials of matrices in the monodromy group of the family (see Todo: [cite]). By the function field number field analogy, the symmetry type of this monodromy group is then assumed to by the symmetry type of the original family.

Modeling L-functions by Characteristic Polynomials of Unitary Matrices

Another example of important stastical information about $\zeta(s)$ is the following: at a given height t up the critical line, how are the the real and imaginary parts of $\log \zeta(\frac{1}{2} + it)$ distributed? In the limit as we move up the critical line, this question is famously answered by **Selberg's central limit theorem** (see Todo: [cite]):

Theorem 4.1.4 (Selberg's central limit theorem). *For any rectangle $B \in \mathbb{C}$,*

$$\lim_{T \to \infty} \frac{1}{T}\mu\left(\left\{ t : t \in [T, 2T], \frac{\log \zeta(\frac{1}{2}+it)}{\sqrt{\frac{1}{2}\log\log(T)}} \in B \right\}\right) = \frac{1}{2\pi}\iint_B e^{-\frac{1}{2}(x^2+y^2)}\, dx\, dy,$$

where μ is the Lebesgue measure.

We interpret the left-hand side in Selberg's central limit theorem as the limiting value distribution, and the right-hand side as two independent Gaussian distributions with unit variance and zero mean. So as we move up the critical line, the real and imagnary parts of $\frac{\log \zeta(\frac{1}{2}+it)}{\sqrt{\frac{1}{2}\log\log(T)}}$ both independently tend to normal distributions. Interestingly, numerical computations show that when $T \sim t_{10^{20}}$ the computed values are

still quite far from the Gaussian distributions. So the convergence in Selberg's central limit theorem is very slow. This does, however, beg the question of how $\log \zeta(\frac{1}{2}+it)$ should be modeled by random matrices when t is large but finite. The answer is quite natural. Since the zeros (unfolded) of $\zeta(s)$ are distributed like the eigenvalues (eigenphases) of a random unitary matrix, we might expect that $\zeta(s)$ is modeled by a function whose zeros are those eigenvalues. That is, $\zeta(s)$ should be modeled by the characteristic polynomial of a unitary matrix (see Todo: [cite] for details). This should work for any single L-function since they all exhibit the same distribution of zeros in the limit as we move up the critical line.

As a first insight to the model, the characteristic polynomial of unitary matrices have strikingly similar resemblance to L-functions. If $A \in U(N)$, then A is diagionalizable with eigenvalues $e^{i\theta_n}$ and eigenphases θ_n for $1 \leq n \leq N$. Let

$$\Lambda_A(s) = A \det(I - sA) = \prod_{n \leq N}(1 - se^{i\theta_n}),$$

be the characteristic polynomial of A. It turns out that $\Lambda_A(s)$ has strikingly similar properties to an L-function. Indeed, if we exapand the product expression, we obtain

$$\Lambda_A(s) = \sum_{0 \leq n \leq N} a_n s^n,$$

for some coefficients a_n. This is the analogue to a the Dirichlet series representation of an L-function. Of course, as $\Lambda_A(s)$ is a polynomial it admits analytic continuation to $\mathbb{C}$. It also has a functional equation of shape $s \to \frac{1}{s}$. To see this, the characteristic polynomial of A^{-1} is the conjugate reciprocial polynomial of $\Lambda_A(s)$ up to the constant term so that

$$\Lambda_{A^{-1}}(s) = \frac{s^N}{\Lambda_A(0)}\Lambda_A\left(\frac{1}{s}\right).$$

But since $\Lambda_A(0) = (-1)^N \det(A)$ and A is unitary so that $\Lambda_{A^{-1}}(s) = \Lambda_{A^*}(s) = \overline{\Lambda_A(s)}$, upon sending $s \to \frac{1}{s}$, we can write the previous expression by isolating the right-hand side as

$$\Lambda_A(s) = (-1)^N \det(A)s^N \overline{\Lambda_A\left(\frac{1}{s}\right)}.$$

This is the functional equation for $\Lambda_A(s)$ and it is of shape $s \to \frac{1}{s}$. We identify the root number as $(-1)^n \det(A)$ and the gamma factor as s^N. The invariant subspace is the unit circle and this plays the role of the critial line for $\Lambda_A(s)$ with the critial value being the symmetric point under $s \to \frac{1}{s}$ which is $s = 1$. In fact, under the change of variables $s \to \frac{1}{s+1}$ the functional equation is of shape $s \to 1 - s$, the unit circle is sent to the critical line, and $s = 1$ is sent to $s = \frac{1}{2}$. Viewing the conductor the absolute value of the derivative of gamma factor evaulated at the critial value $s = 1$, we see that it is N. We also have an approxiate functional equation. By substituting the polynomial representation of $\Lambda_A(s)$ into the functional equation, we obtain

$$\sum_{0 \leq n \leq N} a_n s^n = (-1)^N \det(A)s^N \sum_{0 \leq n \leq N} \overline{a_n}s^{-n} = (-1)^N \det(A) \sum_{0 \leq n \leq N} \overline{a_n}s^{N-n}.$$

Comparing coefficients,

$$a_n = (-1)^N \det(A)\overline{a_{N-n}}.$$

So that for odd N,

$$\Lambda_A(s) = \sum_{0 \leq n \leq \frac{N-1}{2}} a_n s^n + (-1)^N \det(A)s^N \sum_{0 \leq n \leq \frac{N-1}{2}} \overline{a_n}s^{-n},$$

and for even N,

$$\Lambda_A(s) = a_{\frac{N}{2}} s^{\frac{N}{2}} + \sum_{0 \leq n \leq \frac{N}{2}-1} a_n s^n + (-1)^N \det(A) s^N \sum_{0 \leq n \leq \frac{N}{2}-1} \overline{a_n} s^{-n}.$$

These are the approximate functional equations for $\Lambda_A(s)$.

All of this is suggestive of the fact that characteristic polynomials of unitary matrices model L-functions at a theoritical level. As for stastical modeling, we state one last result. This is due to Keating and Sanith and it's an analgous version of Selberg's central limit theorem for $\Lambda_A(s)$ (see Todo: [cite] for a proof):

Theorem 4.1.5. *Let* $|s| = 1$. *For any rectangle* $B \in \mathbb{C}$,

$$\lim_{N \to \infty} \frac{1}{T} dA \left(\left\{ A : A \in U(N), \frac{\log \Lambda_A(s)}{\sqrt{\frac{1}{2} \log(N)}} \in B \right\} \right) = \frac{1}{2\pi} \iint_B e^{-\frac{1}{2}(x^2+y^2)} \, dx \, dy,$$

where μ *is the Lebesgue measure.*

Index