Chapter Fifteen

From Nuclear Physics to L-Functions

In attempting to describe the energy levels of heavy nuclei ([Wig1, Wig3, Po, BFFMPW]), researchers were confronted with daunting calculations for a many bodied system with extremely complicated interaction forces. Unable to explicitly calculate the energy levels, physicists developed Random Matrix Theory to predict general properties of the systems. Surprisingly, similar behavior is seen in studying the zeros of L-functions!

In this chapter we give a brief introduction to classical Random Matrix Theory, Random Graphs and L-Functions. Our goal is to show how diverse systems exhibit similar universal behaviors, and introduce the techniques used in the proofs. In some sense, this is a continuation of the Poissonian behavior investigations of Chapter 12. The survey below is meant to only show the broad brush strokes of this rich landscape; detailed proofs will follow in later chapters. We assume familiarity with the basic concepts of L-functions (Chapter 3), probability theory (Chapter 8) and linear algebra (a quick review of the needed background is provided in Appendix B).

While we assume the reader has some familiarity with the basic concepts in physics for the historical introduction in §15.1, no knowledge of physics is required for the detailed expositions. After describing the physics problems, we describe several statistics of eigenvalues of sets of matrices. It turns out that the spacing properties of these eigenvalues is a good model for the spacings between energy levels of heavy nuclei and zeros of L-functions; exactly why this is so is still an open question. For those interested in learning more (as well as a review of recent developments), we conclude this chapter with a brief summary of the literature.

15.1 HISTORICAL INTRODUCTION

A central question in mathematical physics is the following: given some system with observables $t_1 \le t_2 \le t_3 \le \dots$, describe how the t_i are spaced. For example, we could take the t_i to be the energy levels of a heavy nuclei, or the prime numbers, or zeros of L-functions, or eigenvalues of real symmetric or complex Hermitian matrices (or as in Chapter 12 the fractional parts $\{n^k\alpha\}$ arranged in increasing order). If we completely understood the system, we would know exactly where all the t_i are; in practice we try and go from knowledge of how the t_i are spaced to knowledge of the underlying system.

15.1.1 Nuclear Physics

In classical mechanics it is possible to write down closed form solutions to the two body problem: given two points with masses m_1 and m_2 and initial velocities $\vec{v}_1$ and $\vec{v}_2$ and located at $\vec{r}_1$ and $\vec{r}_2$, describe how the system evolves in time given that gravity is the only force in play. The three body problem, however, defies closed form solutions (though there are known solutions for special arrangements of special masses, three bodies in general position is still open; see [Wh] for more details). From physical grounds we know of course a solution must exist; however, for our solar system we cannot analyze the solution well enough to determine whether or not billions of years from now Pluto will escape from the sun's influence! In some sense this is similar to the problems with the formula for counting primes in Exercise 2.3.18.

Imagine how much harder the problems are in understanding the behavior of heavy nuclei. Uranium, for instance, has over 200 protons and neutrons in its nucleus, each subject to and contributing to complex forces. If the nucleus were completely understood, one would know the energy levels of the nucleus. Physicists were able to gain some insights into the nuclear structure by shooting high-energy neutrons into the nucleus, and analyzing the results; however, a complete understanding of the nucleus was, and still is, lacking. Later, when we study zeros of L-functions from number theory, we will find analogues of high-energy neutrons!

One powerful formulation of physics is through infinite dimensional linear algebra. The fundamental equation for a system becomes

$$H\psi_n \;=\; E_n\psi_n, \tag{15.1}$$

where H is an operator (called the **Hamiltonian**) whose entries depend on the physical system and the ψ_n are the energy eigenfunctions with eigenvalues E_n. Unfortunately for nuclear physics, H is too complicated to write down and solve; however, a powerful analogy with Statistical Mechanics leads to great insights.

15.1.2 Statistical Mechanics

For simplicity consider N particles in a box where the particles can only move left or right and each particle's speed is v; see Figure 15.1.

If we want to calculate the pressure on the left wall, we need to know how many particles strike the wall in an infinitesimal time. Thus we need to know how many particles are close to the left wall and moving towards it. Without going into all of the physics (see for example [Re]), we can get a rough idea of what is happening. The complexity, the enormous number of configurations of positions of the molecules, actually helps us. For each configuration we can calculate the pressure due to that configuration. We then *average* over all configurations, and hope that a generic configuration is, in some sense, close to the system average.

Wigner's great insight for nuclear physics was that similar tools could yield useful predictions for heavy nuclei. He modeled the nuclear systems as follows: instead of the infinite dimensional operator H whose entries are given by the physical laws, he considered collections of $N \times N$ matrices where the entries were independently chosen from some probability distribution p. The eigenvalues of these

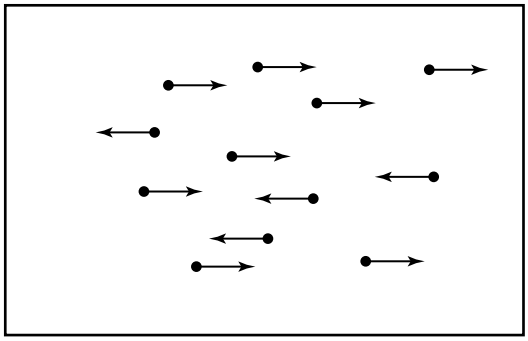

Figure 15.1 Molecules in a box

matrices correspond to the energy levels of the physical system. Depending on physical symmetries, we consider different collections of matrices (real symmetric, complex Hermitian). For any given finite matrix we can calculate statistics of the eigenvalues. We then average over all such matrices, and look at the limits as $N \to \infty$. The main result is that ***the behavior of the eigenvalues of an arbitrary matrix is often well approximated by the behavior obtained by averaging over all matrices, and this is a good model for the energy levels of heavy nuclei***. This is reminiscent of the Central Limit Theorem (§8.4). For example, if we average over all sequences of tossing a fair coin $2N$ times, we obtain N heads, and *most* sequences of $2N$ tosses will have approximately N heads.

Exercise 15.1.1. *Consider $2N$ identical, indistinguishable particles, which are in the left (resp., right) half of the box with probability $\frac{1}{2}$. What is the expected number of particles in each half? What is the probability that one half has more than $(2N)^{\frac{3}{4}}$ particles than the other half? As $(2N)^{\frac{3}{4}} \ll N$, most systems will have similar behavior although of course some will not. The point is that a* typical *system will be close to the system average.*

Exercise 15.1.2. *Consider $4N$ identical, indistinguishable particles, which are in the left (resp., right) half of the box with probability $\frac{1}{2}$; each particle is moving left (resp., right) with probability $\frac{1}{2}$. Thus there are four possibilities for each particle, and each of the 4^{4N} configurations of the $4N$ particles is equally likely. What is the expected number of particles in each possibility (left-left, left-right, right-left, right-right)? What is the probability that one possibility has more than $(4N)^{\frac{3}{4}}$ particles than the others? As $(4N)^{\frac{3}{4}} \ll N$, most systems will have similar behavior.*

15.1.3 Random Matrix Ensembles

The first collection of matrices we study are $N \times N$ real symmetric matrices, with the entries independently chosen from a fixed probability distribution p on $\mathbb{R}$. Given

such a matrix A,

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1N} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & a_{N3} & \cdots & a_{NN} \end{pmatrix} = A^T \tag{15.2}$$

(so $a_{ij} = a_{ji}$), the probability density of observing A is

$$\text{Prob}(A)dA = \prod_{1 \le i \le j \le N} p(a_{ij})da_{ij}. \tag{15.3}$$

We may interpret this as giving the probability of observing a real symmetric matrix where the probability of the ij^{th} entry lying in $[a_{ij}, a_{ij} + da_{ij}]$ is $p(a_{ij})da_{ij}$. More explicitly,

$$\text{Prob}(A : a_{ij} \in [\alpha_{ij}, \beta_{ij}]) = \prod_{1 \le i \le j \le N} \int_{\alpha_{ij}}^{\beta_{ij}} p(a_{ij})da_{ij}. \tag{15.4}$$

Example 15.1.3. *For a 2×2 real symmetric matrix we would have*

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}, \quad \text{Prob}(A) = p(a_{11})p(a_{12})p(a_{22})da_{11}da_{12}da_{22}. \tag{15.5}$$

An $N \times N$ real symmetric matrix is determined by specifying $\frac{N(N+1)}{2}$ entries: there are N entries on the main diagonal, and $N^2 - N$ off-diagonal entries (for these entries, only half are needed, as the other half are determined by symmetry). We say such a matrix has $\frac{N(N+1)}{2}$ **degrees of freedom**. Because p is a probability density, it integrates to 1. Thus

$$\int \text{Prob}(A)dA = \prod_{1 \le i \le j \le N} \int_{a_{ij}=-\infty}^{\infty} p(a_{ij})da_{ij} = 1; \tag{15.6}$$

this corresponds to the fact that we must choose some matrix.

For convergence reasons we often assume that the moments of p are finite. We mostly study $p(x)$ satisfying

$$\begin{aligned} p(x) &\ge 0 \\ \int_{-\infty}^{\infty} p(x)dx &= 1 \\ \int_{-\infty}^{\infty} |x|^k p(x)dx &< \infty. \end{aligned} \tag{15.7}$$

The last condition ensures that the probability distribution is not too spread out (i.e., there is not too much probability near infinity). Many times we normalize p so that the mean (first moment) is 0 and the variance (second moment if the mean is zero) is 1.

Exercise 15.1.4. *For the k^{th} moment $\int_{\mathbb{R}} x^k p(x)dx$ to exist, we require $\int_{\mathbb{R}} |x|^k p(x)dx < \infty$; if this does not hold, the value of the integral could depend on how we approach infinity. Find a probability function $p(x)$ and an integer k such that*

$$\lim_{A \to \infty} \int_{-A}^{A} x^k p(x)dx = 0 \quad \textit{but} \quad \lim_{A \to \infty} \int_{-A}^{2A} x^k p(x)dx = \infty. \tag{15.8}$$

Exercise 15.1.5. *Let p be a probability density such that all of its moments exist. If p is an even (resp., odd) function, show all the odd (resp., even) moments vanish.*

Exercise 15.1.6. *Let p be a continuous probability density on $\mathbb{R}$. Show there exist constants a, b such that $q(x) = a \cdot p(ax + b)$ has mean 0 and variance 1. Thus in some sense the third and the fourth moments are the first "free" moments as the above transformation is equivalent to translating and rescaling the initial scale.*

Exercise 15.1.7. *It is not necessary to choose each entry from the same probability distribution. Let the ij^{th} entry be chosen from a probability distribution p_{ij}. What is the probability density of observing A? Show this also integrates to 1.*

Definition 15.1.8 (Ensembles)**.** *A collection of matrices, along with a probability density describing how likely it is to observe a given matrix, is called an **ensemble** of matrices (or a **random matrix ensemble**).*

Example 15.1.9. *Consider the ensemble of 2×2 real symmetric matrices A where for a matrix $A = \left(\begin{smallmatrix} x & y \\ y & z \end{smallmatrix} \right)$,*

$$p(A) = \begin{cases} \frac{3}{4\pi} & \text{if } x^2 + y^2 + z^2 \leq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{15.9}$$

Note the entries are not independent. We can parametrize these matrices by using spherical coordinates. For a sphere of radius r we have

$$\begin{aligned} x = x(r, \theta, \phi) &= r\cos(\theta)\sin(\phi) \\ y = y(r, \theta, \phi) &= r\sin(\theta)\sin(\phi) \\ z = z(r, \theta, \phi) &= r\cos(\phi), \end{aligned} \tag{15.10}$$

where $\theta \in [0, 2\pi]$ is the azimuthal angle, $\phi \in [0, \pi]$ is the polar angle and the volume of the sphere is $\frac{4}{3}\pi r^3$.

In this introduction we confine ourselves to real symmetric matrices, although many other ensembles of matrices are important. Complex Hermitian matrices (the generalization of real symmetric matrices) also play a fundamental role in the theory. Both of these types of matrices have a very important property: *their eigenvalues are real*; this is what allows us to ask questions such as how are the spacings between eigenvalues distributed.

In constructing our real symmetric matrices, we have not said much about the probability density p. In Chapter 17 we show for that some physical problems, additional assumptions about the physical systems force p to be a Gaussian. For many of the statistics we investigate, it is either known or conjectured that the answers should be independent of the specific choice of p; however, in this method of constructing random matrix ensembles, there is often no unique choice of p. Thus, for this method, there is no unique answer to what it means to choose a matrix *at random*.

Remark 15.1.10 (Advanced)**.** We would be remiss if we did not mention another notion of randomness, which leads to a more natural method of choosing a matrix at random. Let $U(N)$ be the space of $N \times N$ unitary matrices, and consider

its compact subgroups (for example, the $N \times N$ orthogonal matrices). There is a natural (canonical) measure, called the **Haar measure**, attached to each of these compact groups, and we can use this measure to choose matrices *at random*. Further, the eigenvalues of unitary matrices have modulus 1. They can be written as $e^{i\theta_j}$, with the θ_j real. We again obtain a sequence of real numbers, and can again ask many questions about spacings and distributions. This is the notion of random matrix ensemble which has proven the most useful for number theory.

Exercise 15.1.11. *Prove the eigenvalues of real symmetric and complex Hermitian matrices are real.*

Exercise 15.1.12. *How many degrees of freedom does a complex Hermitian matrix have?*

15.2 EIGENVALUE PRELIMINARIES

15.2.1 Eigenvalue Trace Formula

Our main tool to investigate the eigenvalues of matrices will be the Eigenvalue Trace Formula. Recall the trace of a matrix is the sum of its diagonal entries:

$$\mathrm{Trace}(A) \ = \ a_{11} + \cdots + a_{NN}. \tag{15.11}$$

We will also need the trace of powers of matrices. For example, a 2×2 matrices

$$A \ = \ \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \tag{15.12}$$

we have

$$\mathrm{Trace}(A^2) \ = \ a_{11}a_{11} + a_{12}a_{21} + a_{12}a_{21} + a_{22}a_{22} \ = \ \sum_{i=1}^{2}\sum_{j=1}^{2} a_{ij}a_{ji}. \tag{15.13}$$

In general we have

Theorem 15.2.1. *Let A be an $N \times N$ matrix. Then*

$$\mathrm{Trace}(A^k) \ = \ \sum_{i_1=1}^{N} \cdots \sum_{i_k=1}^{N} a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{k-1} i_k} a_{i_k i_1}. \tag{15.14}$$

For small values of k, instead of using $i_1, i_2, i_3, \ldots$ we often use $i, j, k, \ldots$. For example, $\mathrm{Trace}(A^3) = \sum_i \sum_j \sum_k a_{ij}a_{jk}a_{ki}$.

Exercise 15.2.2. *Show* (15.13) *is consistent with Theorem 15.2.1.*

Exercise 15.2.3. *Prove Theorem 15.2.1.*

Theorem 15.2.4 (Eigenvalue Trace Formula)**.** *For any non-negative integer k, if A is an $N \times N$ matrix with eigenvalues $\lambda_i(A)$, then*

$$\mathrm{Trace}(A^k) \ = \ \sum_{i=1}^{N} \lambda_i(A)^k. \tag{15.15}$$

The importance of this formula is that it relates the *eigenvalues* of a matrix (which is what we *want* to study) to the *entries* of A (which is what we *choose* at random). The importance of this formula cannot be understated – it is what makes the whole subject possible.

Sketch of the proof. The case $k = 1$ follows from looking at the characteristic polynomial $\det(A - \lambda I) = 0$. For higher k, note any matrix A can be conjugated to an upper triangular matrix: $U^{-1}AU = T$ where T is upper triangular and U is unitary. The eigenvalues of A equal those of T and are given by the diagonal entries of T. Further the eigenvalues of A^k equal those of T^k. If $\lambda_i(A)$ and $\lambda_i(A^k)$ are the eigenvalues of A and A^k, note $\lambda_i(A^k) = \lambda_i(A)^k$. The claim now follows by applying the $k = 1$ result to the matrix A^k:

$$\text{Trace}(A^k) \;=\; \sum_{i=1}^{N} \lambda_i(A^k) \;=\; \sum_{i=1}^{N} \lambda_i(A)^k. \tag{15.16}$$

□

Exercise 15.2.5. *Prove all the claims used in the proof of the Eigenvalue Trace Formula. If A is real symmetric, one can use the diagonalizability of A. To show any matrix can be triangularized, start with every matrix has at least one eigenvalue-eigenvector pair. Letting $\vec{v_1}$ be the eigenvector, using Gram-Schmidt one can find an orthonormal basis. Let these be the columns of U_1, which will be a unitary matrix. Continue by induction.*

15.2.2 Normalizations

Before we can begin to study fine properties of the eigenvalues, we first need to figure out what is the correct scale to use in our investigations. For example, the celebrated Prime Number Theorem (see Theorem 13.2.10 for an exact statement of the error term) states that $\pi(x)$, the number of primes less than x, satisfies

$$\pi(x) \;=\; \frac{x}{\log x} \;+\; \text{lower order terms.} \tag{15.17}$$

Remark 15.2.6. If we do not specify exactly how much smaller the error terms are, we do not need the full strength of the Prime Number Theorem; Chebyshev's arguments (Theorem 2.3.9) are sufficient to get the order of magnitude of the scale.

The average spacing between primes less than x is about $\frac{x}{x/\log x} = \log x$, which tends to infinity as $x \to \infty$. Asking for primes that differ by 2 is a very hard question: as $x \to \infty$, this becomes insignificant on the "natural" scale. Instead, a more natural question is to inquire how often two primes are twice the average spacing apart. This is similar to our investigations in Chapter 12 where we needed to find the correct scale.

If we fix a probability density p, how do we expect the sizes of the eigenvalues $\lambda_i(A)$ to depend on N as we vary A? A good estimate falls out immediately from the Eigenvalue Trace Formula; this formula will be exploited numerous times in the arguments below, and is essential for all investigations in the subject.

We give a heuristic for the eigenvalues of our $N \times N$ ensembles of matrices being roughly of size $\sqrt{N}$. Fix a matrix A whose entries a_{ij} are randomly and independently chosen from a fixed probability distribution p with mean 0 and variance 1. By Theorem 15.2.1, for $A = A^T$ we have that

$$\text{Trace}(A^2) \;=\; \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}a_{ji} \;=\; \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2. \tag{15.18}$$

From our assumptions on p, we expect each a_{ij}^2 to be of size 1. By the Central Limit Theorem (Theorem 8.4.1) or Chebyshev's Theorem (Exercise 8.1.55), we expect with high probability

$$\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2 \sim N^2 \cdot 1, \tag{15.19}$$

with an error of size $\sqrt{N^2} = N$ (as each a_{ij}^2 is approximately of size 1 and there are N^2 of them, with high probability their sum should be approximately of size N^2). Thus

$$\sum_{i=1}^{N} \lambda_i(A)^2 \sim N^2, \tag{15.20}$$

which yields

$$N \cdot \text{Ave}(\lambda_i(A)^2) \;\sim\; N^2. \tag{15.21}$$

For heuristic purposes we shall pass the square root through to get

$$|\text{Ave}(\lambda_i(A))| \;\sim\; \sqrt{N}. \tag{15.22}$$

In general the square root of an average need not be the same as the average of the square root; however, our purpose here is merely to give a heuristic as to the correct scale. Later in our investigations we shall see that $\sqrt{N}$ is the correct normalization.

Thus it is natural to guess that the correct scale to study the eigenvalues of an $N \times N$ real symmetric matrix is $c\sqrt{N}$, where c is some constant independent of N. This yields normalized eigenvalues $\widetilde{\lambda}_1(A) = \frac{\lambda_i(A)}{c\sqrt{N}}$; choosing $c = 2$ leads to clean formulas. One could of course normalize the eigenvalues by $f(N)$, with f an undetermined function, and see which choices of f give good results; eventually one would find $f(N) = c\sqrt{N}$.

Exercise 15.2.7. *Consider real $N \times N$ matrices with entries independently chosen from a probability distribution with mean 0 and variance 1. How large do you expect the average eigenvalue to be?*

Exercise 15.2.8. *Use Chebyshev's Theorem (Exercise 8.1.55) to bound the probability that $|\sum_i \sum_j a_{ij}^2 - N^2| \geq N \log N$. Conclude that with high probability that the sum of the squares of the eigenvalues is of size N^2 for large N.*

15.2.3 Eigenvalue Distribution

We quickly review the theory of point masses and induced probability distributions (see §11.2.2 and §12.5). Let δ_{x_0} represent a unit point mass at x_0. We define its action on functions by

$$\delta_{x_0}(f) \ := \ \int_{-\infty}^{\infty} f(x)\delta(x-x_0)dx \ = \ f(x_0). \tag{15.23}$$

δ_{x_0}, called the **Dirac delta functional** at x_0, is similar to our approximations to the identity. There is finite mass (its integral is 1), the density is 0 outside x_0 and infinite at x_0. As its argument is a function and not a complex number, δ_{x_0} is a **functional** and not a function. To each A, we attach a probability measure (the **eigenvalue probability distribution**):

$$\mu_{A,N}(x)dx \ = \ \frac{1}{N}\sum_{i=1}^{N} \delta\left(x - \frac{\lambda_i(A)}{2\sqrt{N}}\right)dx. \tag{15.24}$$

At each normalized eigenvalue $\frac{\lambda_i(A)}{2\sqrt{N}}$ we have placed a mass of weight $\frac{1}{N}$; there are N masses, thus we have a probability distribution. If $p(x)$ is a probability distribution then $\int_a^b p(x)dx$ is the probability of observing a value in $[a,b]$. For us, $\int_a^b \mu_{A,N}(x)dx$ is the percentage of normalized eigenvalues in $[a,b]$:

$$\int_a^b \mu_{A,N}(x)dx \ = \ \frac{\#\{i : \frac{\lambda_i(A)}{2\sqrt{N}} \in [a,b]\}}{N}. \tag{15.25}$$

We can calculate the moments of $\mu_{A,N}(x)$.

Definition 15.2.9. *Let $\mathbb{E}[x^k]_A$ denote the k^{th} moment of $\mu_{A,N}(x)$. We often denote this $M_{N,k}(A)$.*

The following corollary of the Eigenvalue Trace formula is the starting point of many of our investigations; we see in Remark 15.3.15 why it is so useful.

Lemma 15.2.10. $M_{N,k}(A) = \frac{\mathrm{Trace}(A^k)}{2^k N^{\frac{k}{2}+1}}$.

Proof. As $\mathrm{Trace}(A^k) = \sum_i \lambda_i(A)^k$ we have

$$\begin{aligned} M_{N,k}(A) \ = \ \mathbb{E}[x^k]_A \ &= \ \int x^k \mu_{A,N}(x)dx \\ &= \ \frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}} x^k \delta\left(x - \frac{\lambda_i(A)}{2\sqrt{N}}\right)dx \\ &= \ \frac{1}{N}\sum_{i=1}^{N}\frac{\lambda_i(A)^k}{(2\sqrt{N})^k} \\ &= \ \frac{\mathrm{Trace}(A^k)}{2^k N^{\frac{k}{2}+1}}. \end{aligned} \tag{15.26}$$

□

Exercise 15.2.11. *Let A be an $N \times N$ real symmetric matrix with $|a_{ij}| \le B$. In terms of B, N and k bound $|\mathrm{Trace}(A^k)|$ and $M_{N,k}(A)$. How large can $\max_i |\lambda_i(A)|$ be?*

15.3 SEMI-CIRCLE LAW

15.3.1 Statement

A natural question to ask concerning the eigenvalues of a matrix is: *What percentage of the normalized eigenvalues lie in an interval $[a,b]$?* Let $\mu_{A,N}(x)$ be the eigenvalue probability distribution. For a given A, the answer is

$$\int_a^b \mu_{A,N}(x)dx. \tag{15.27}$$

How does the above behave as we vary A? We have the following startling result, which is almost independent of the underlying probability density p we used to choose the entries of A:

Theorem 15.3.1 (Semi-Circle Law)**.** *Consider the ensemble of $N \times N$ real symmetric matrices with entries independently chosen from a fixed probability density $p(x)$ with mean 0, variance 1, and finite higher moments. As $N \to \infty$, for almost all A, $\mu_{A,N}(x)$ converges to the semi-circle density $\frac{2}{\pi}\sqrt{1-x^2}$.*

Thus the percentage of normalized eigenvalues of A in $[a,b] \subset [-1,1]$ for a typical A as $N \to \infty$ is

$$\int_a^b \frac{2}{\pi}\sqrt{1-x^2}dx. \tag{15.28}$$

Later in §15.3.4 we discuss what happens if the higher moments are infinite.

15.3.2 Moment Problem

We briefly describe a needed result from Probability Theory: the solution to the Moment Problem. See page 110 of [Du] for details; see [ShTa] for a connection between the moment problem and continued fractions!

Let k be a non-negative integer; below we always assume $m_0 = 1$. We are interested in when numbers m_k determine a unique probability distribution P whose k^{th} moment is m_k. If the m_k do not grow too rapidly, there is at most one continuous probability density with these moments (see [Bi, CaBe, Fe]). A sufficient condition is Carleman's Condition that $\sum_{j=1}^{\infty} m_{2j}^{-1/2j} = \infty$. Another is that $\sum_{j=1}^{\infty} \frac{m_j t^j}{j!}$ has a positive radius of convergence. This implies the moment generating function (see Exercise 15.3.2) exists in an interval and the distribution is uniquely determined.

Exercise 15.3.2 (Non-uniqueness of moments)**.** *For $x \in [0,\infty)$, consider*

$$\begin{aligned} f_1(x) &= \frac{1}{\sqrt{2\pi}x}e^{-(\log x)^2/2} \\ f_2(x) &= f_1(x)\left[1+\sin(2\pi\log x)\right]. \end{aligned} \tag{15.29}$$

Show that for $r \in \mathbb{N}$, the r^{th} moment of f_1 and f_2 is $e^{r^2/2}$. The reason for the non-uniqueness of moments is that the ***moment generating function***

$$M_f(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx \tag{15.30}$$

does not converge in a neighborhood of the origin. See [CaBe], Chapter 2. See also Exercise A.2.7.

For us the numbers m_k arise from averaging the moments $M_{N,k}(A)$ of the $\mu_{A,N}(x)$'s and taking the limit as $N \to \infty$. Let

$$M_{N,k} = \int_A M_{N,k}(A)\text{Prob}(A)dA, \quad m_k = \lim_{N\to\infty} M_{N,k}. \tag{15.31}$$

For each N the moments $M_{N,k}$ yield a probability distribution P_N, and $\lim_{N\to\infty} M_{N,k} = m_k$. If the m_k grow sufficiently slowly, there is a unique continuous probability density P with k^{th} moment m_k. It is reasonable to posit that as for each k, $\lim_{N\to\infty} M_{N,k} = m_k$, then "most" $\mu_{A,N}(x)$ converge (in some sense) to the probability density $P(x)$.

Remark 15.3.3 (Warning)**.** For each N, consider N numbers $\{a_{n,N}\}_{n=1}^N$ defined by $a_{n,N} = 1$ if n is even and -1 if N is odd. For N even, note the average of the $a_{n,N}$'s is 0, but each $|a_{n,N}| = 1$; thus, no element is close to the system average. Therefore, it is not always the case that a typical element is close to the system average. What is needed in this case is to consider the variance of the moments (see Exercise 15.3.5).

Remark 15.3.4. While it is not true that every sequence of numbers m_k that grow sufficiently slowly determines a continuous probability density (see Exercise 15.3.8), as our m_k arise from limits of moments of probability distributions, we do obtain a unique limiting probability density. This is similar to determining when a Taylor series converges to a unique function. See also Exercise A.2.7.

Exercise 15.3.5. *Let $\{b_{n,N}\}_{n=1}^N$ be a sequence with mean $\mu(N) = \frac{1}{N}\sum_{n=1}^N b_{n,N}$ and variance $\sigma^2(N) = \frac{1}{N}\sum_{n=1}^N |b_{n,N} - \mu(N)|^2$. Assume that as $N \to \infty$, $\mu(N) \to \mu$ and $\sigma^2(N) \to 0$. Prove for any $\epsilon > 0$ as $N \to \infty$ for a fixed N at most ϵ percent of $b_{n,N}$ are not within ϵ of μ. Therefore,* if *the mean of a sequence converges* and *we have control over the variance,* then *we have control over the limiting behavior of* most *elements.*

In this text we content ourselves with calculating the average moments $m_k = \lim_{N\to\infty}\int_A M_{N,k}(A)dA$. In many cases we derive simple expressions for the probability density P with moments m_k; however, we do not discuss the probability arguments needed to show that as $N \to \infty$, a "typical" matrix A has $\mu_{A,n}(x)$ close to P. The interested reader should see [CB, HM] for an application to moment arguments in random matrix theory.

Some care is needed in formulating what it means for two probability distributions to be close. For us, $\mu_{A,N}(x)$ is the sum of N Dirac delta functionals of mass $\frac{1}{N}$. Note $|P(x) - \mu_{A,N}(x)|$ can be large for individual x. For example, if $P(x)$ is the semi-circle distribution, then $|P(x) - \mu_{A,N}(x)|$ will be of size 1 for almost all $x \in [-1, 1]$. We need to define what it means for two probability distributions to be close.

One natural measure is the Kolmogoroff-Smirnov discrepancy. For a probability distribution $f(x)$ its **Cumulative Distribution Function** $C_f(x)$ is defined to be the

probability of $[-\infty, x]$:

$$C_f(x) \;=\; \int_{-\infty}^{x} f(x)dx. \tag{15.32}$$

If our distribution is continuous, note this is the same as the probability of $[-\infty, x)$; however, for distributions arising from Dirac delta functionals like our $\mu_{A,N}(x)$, there will be finite, non-zero jumps in the cumulative distribution function at the normalized eigenvalues. For example, for $\mu_{A,N}(x)$ we have

$$C_{\mu_{A,N}}(x) \;=\; \frac{1}{N} \sum_{\frac{\lambda_i(A)}{2\sqrt{N}} < x} 1, \tag{15.33}$$

which jumps by at least $\frac{1}{N}$ at each normalized eigenvalue. For two probability distributions f and g we define the **Kolmogoroff-Smirnov discrepency of f and g** to be $\sup_x |C_f(x) - C_g(x)|$. Note as $N \to \infty$ each normalized eigenvalue contributes a smaller percentage of the total probability. Using the Kolmogoroff-Smirnov discrepancy for when two probability distributions are close, one can show that as $N \to \infty$ "most" $\mu_{A,N}(x)$ are close to P.

Remark 15.3.6. It is not true that all matrices A yield $\mu_{A,N}(x)$ that are close to P; for example, consider multiples of the identity matrix. All the normalized eigenvalues are the same, and these real symmetric matrices will clearly not have $\mu_{A,N}(x)$ close to $P(x)$. Of course, as $N \to \infty$ the probability of A being close to a multiple of the identity matrix is zero.

Exercise 15.3.7. *Fix a probability distribution p, and consider $N \times N$ real symmetric matrices with entries independently chosen from p. What is the probability that a matrix in this ensemble has all entries within ϵ of a multiple of the $N \times N$ identity matrix? What happens as $N \to \infty$ for fixed ϵ? How does the answer depend on p?*

Exercise 15.3.8. *Let m_k be the k^{th} moment of a probability density P. Show $m_2 m_0 - m_1^2 \geq 0$. Note this can be interpreted as $\left| \begin{smallmatrix} m_0 & m_1 \\ m_1 & m_2 \end{smallmatrix} \right| \geq 0$. Thus, if $m_2 m_0 - m_1^2 < 0$, the m_k cannot be the moments of a probability distribution. Find a similar relation involving m_0, m_1, m_2, m_3 and m_4 and a determinant. See [Chr] and the references therein for more details, as well as [ShTa, Si] (where the determinant condition is connected to continued fraction expansions!).*

Exercise 15.3.9. *If $p(x) = 0$ for $|x| > R$, show the k^{th} moment satisfies $m_k \leq R^k$. Hence $\lim_{j\to\infty} m_{2j}^{1/2j} < \infty$. Therefore, if a probability distribution has $\lim_{j\to\infty} m_{2j}^{1/2j} = \infty$, then for any R there is a positive probability of observing $|x| > R$. Alternatively, we say such p has unbounded support. Not surprisingly, the Gaussian moments (see Exercise 15.3.10) grow sufficiently rapidly so that the Gaussian has unbounded support.*

Exercise 15.3.10 (Moments of the Gaussian)**.** *Calculate the moments of the Gaussian $g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Prove the odd moments vanish and the even moments are $m_{2k} = (2k-1)!!$, where $n!! = n(n-2)(n-4)\cdots$. This is also the number of*

ways to match $2k$ objects in pairs. Show the moments grow sufficiently slowly to determine a unique continuous probability density.

Exercise 15.3.11. *Consider two probability distributions f and g on $[0,1]$ where $f(x) = 1$ for all x and $g(x) = 1$ for $x \notin \mathbb{Q}$ and 0 otherwise. Note both f and g assign the same probability to any $[a,b]$ with $b \neq a$. Show $\sup_{x\in[0,1]} |f(x) - g(x)| = 1$ but the Kolmogoroff-Smirnov discrepancy is zero. Thus looking at the pointwise difference could incorrectly cause us to conclude that f and g are very different.*

Exercise 15.3.12. *Do there exist two probability distributions that have a large Kolmogoroff-Smirnov discrepancy but are close pointwise?*

15.3.3 Idea of the Proof of the Semi-Circle Law

We give a glimpse of the proof of the Semi-Circle Law below; a more detailed sketch will be provided in Chapter 16. We use Moment Method from §15.3.2.

For each $\mu_{A,N}(x)$, we calculate its k^{th}-moment, $M_{N,k}(A) = \mathbb{E}[x^k]_A$. Let $M_{N,k}$ be the average of $M_{N,k}(A)$ over all A. We must show as $N \to \infty$, $M_{N,k}$ converges to the k^{th} moment of the semi-circle. We content ourselves with just the second moment below, and save the rest for §16.1. By Lemma 15.2.10,

$$\begin{aligned} M_{N,2} &= \int_A M_{N,k}(A)\mathrm{Prob}(A)dA \\ &= \frac{1}{2^2 N^{\frac{2}{2}+1}} \int_A \mathrm{Trace}(A^2)\mathrm{Prob}(A)dA. \end{aligned} \tag{15.34}$$

We use Theorem 15.2.1 to expand $\mathrm{Trace}(A^2)$ and find

$$M_{N,2} = \frac{1}{2^2 N^2} \int_A \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2 \,\mathrm{Prob}(A)dA. \tag{15.35}$$

We now expand $\mathrm{Prob}(A)dA$ by (15.3):

$$\begin{aligned} & M_{N,2} \\ &= \frac{1}{2^2 N^2} \int_{a_{11}=-\infty}^{\infty} \cdots \int_{a_{NN}=-\infty}^{\infty} \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2 \cdot p(a_{11})da_{11}\cdots p(a_{NN})da_{NN} \\ &= \frac{1}{2^2 N^2} \sum_{i=1}^{N}\sum_{j=1}^{N} \int_{a_{11}=-\infty}^{\infty} \cdots \int_{a_{NN}=-\infty}^{\infty} a_{ij}^2 \cdot p(a_{11})da_{11}\cdots p(a_{NN})da_{NN}; \end{aligned} \tag{15.36}$$

we may interchange the summations and the integrations as there are finitely many sums. For each of the N^2 pairs (i,j), we have terms like

$$\int_{a_{ij}=-\infty}^{\infty} a_{ij}^2 p(a_{ij})da_{ij} \cdot \prod_{\substack{(k,l)\neq(ij)\\ k\leq l}} \int_{a_{kl}=-\infty}^{\infty} p(a_{kl})da_{kl}. \tag{15.37}$$

The above equals 1. The first factor is 1 because it is the variance of a_{ij}, which was assumed to be 1. The second factor is a product of integrals where each integral

is 1 (because p is a probability density). As there are N^2 summands in (15.36), we find $M_{N,2} = \frac{1}{4}$ (so $\lim_{N\to\infty} M_{N,2} = \frac{1}{4}$), which is the second moment of the semi-circle.

Exercise 15.3.13. *Show the second moment of the semi-circle is $\frac{1}{4}$.*

Exercise 15.3.14. *Calculate the third and fourth moments, and compare them to those of the semi-circle.*

Remark 15.3.15 (Important)**.** Two features of the above proof are worth highlighting, as they appear again and again below. First, note that we want to answer a question about the *eigenvalues* of A; however, our notion of randomness gives us information on the *entries* of A. The key to converting information on the entries to knowledge about the eigenvalues is having some type of Trace Formula, like Theorem 15.2.4.

The second point is the Trace Formula would be useless, merely converting us from one hard problem to another, if we did not have a good Averaging Formula, some way to average over all random A. In this problem, the averaging is easy because of how we defined randomness.

Remark 15.3.16. While the higher moments of p are not needed for calculating $M_{N,2} = \mathbb{E}[x^2]$, their finiteness comes into play when we study higher moments.

15.3.4 Examples of the Semi-Circle Law

First we look at the density of eigenvalues when p is the standard Gaussian, $p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. In Figure 15.2 we calculate the density of eigenvalues for 500 such matrices (400×400), and note a great agreement with the semi-circle.

What about a density where the higher moments are infinite? Consider the Cauchy distribution,

$$p(x) \;=\; \frac{1}{\pi(1+x^2)}. \tag{15.38}$$

The behavior is clearly not semi-circular (see Figure 15.3). The eigenvalues are unbounded; for graphing purposes, we have put all eigenvalues greater than 300 in the last bin, and less than -300 in the first bin.

Exercise 15.3.17. *Prove the Cauchy distribution is a probability distribution by showing it integrates to 1. While the distribution is symmetric, one cannot say the mean is 0, as the integral $\int |x|p(x)dx = \infty$. Regardless, show the second moment is infinite.*

15.3.5 Summary

Note the universal behavior: though the proof is not given here, the Semi-Circle Law holds for all mean zero, finite moment distributions. The independence of the behavior on the exact nature of the underlying probability density p is a common feature of Random Matrix Theory statements, as is the fact that as $N \to \infty$ most

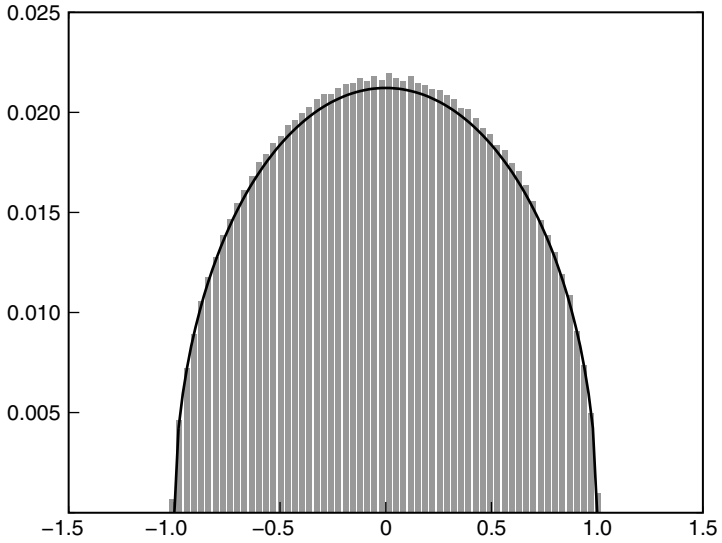

Figure 15.2 Distribution of eigenvalues: 500 Gaussian matrices (400×400)

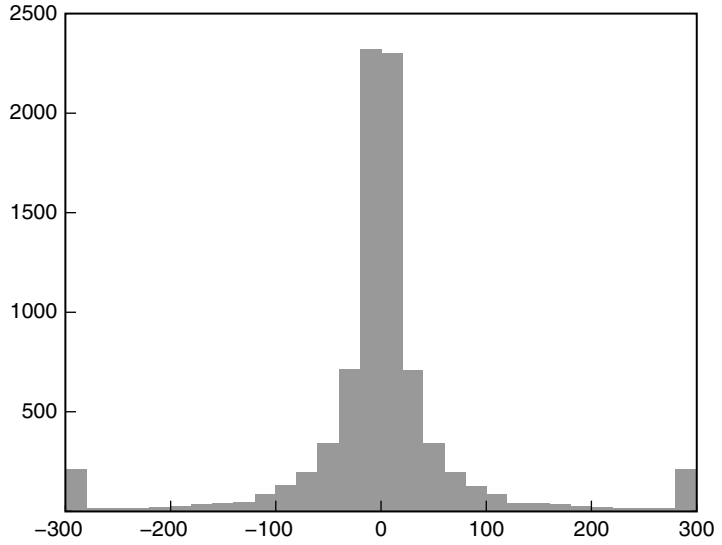

Figure 15.3 Distribution of eigenvalues: 5000 Cauchy matrices (300×300)

A yield $\mu_{A,N}(x)$ that are close (in the sense of the Kolmogoroff-Smirnov discrepancy) to P (where P is determined by the limit of the average of the moments $M_{N,k}(A)$). For more on the Semi-Circle Law, see [Bai, BK].

15.4 ADJACENT NEIGHBOR SPACINGS

15.4.1 GOE Distribution

The Semi-Circle Law (when the conditions are met) tells us about the density of eigenvalues. We now ask a more refined question:

Question 15.4.1. *How are the spacings between adjacent eigenvalues distributed?*

For example, let us write the eigenvalues of A in increasing order; as A is real symmetric, the eigenvalues will be real:

$$\lambda_1(A) \le \lambda_2(A) \le \cdots \le \lambda_N(A). \tag{15.39}$$

The spacings between adjacent eigenvalues are the $N-1$ numbers

$$\lambda_2(A) - \lambda_1(A),\ \lambda_3(A) - \lambda_2(A),\ \ldots,\ \lambda_N(A) - \lambda_{N-1}(A). \tag{15.40}$$

As before (see Chapter 12), it is more natural to study the spacings between adjacent normalized eigenvalues; thus, we have

$$\frac{\lambda_2(A)}{2\sqrt{N}} - \frac{\lambda_1(A)}{2\sqrt{N}},\ \ldots,\ \frac{\lambda_N(A)}{2\sqrt{N}} - \frac{\lambda_{N-1}(A)}{2\sqrt{N}}. \tag{15.41}$$

Similar to the probability distribution $\mu_{A,N}(x)$, we can form another probability distribution $\nu_{A,N}(s)$ to measure spacings between adjacent normalized eigenvalues.

Definition 15.4.2.

$$\nu_{A,N}(s)ds = \frac{1}{N-1}\sum_{i=2}^{N} \delta\left(s - \frac{\lambda_i(A) - \lambda_{i-1}(A)}{2\sqrt{N}}\right) ds. \tag{15.42}$$

Based on experimental evidence and some heuristical arguments, it was conjectured that as $N \to \infty$, the limiting behavior of $\nu_{A,N}(s)$ is independent of the probability density p used in randomly choosing the $N \times N$ matrices A.

Conjecture 15.4.3 (GOE Conjecture:)**.** *As $N \to \infty$, $\nu_{A,N}(s)$ approaches a universal distribution that is independent of p.*

Remark 15.4.4. GOE stands for Gaussian Orthogonal Ensemble; the conjecture is known if p is (basically) a Gaussian. We explain the nomenclature in Chapter 17.

Remark 15.4.5 (Advanced)**.** The universal distribution is $\frac{\pi^2}{4}\frac{d^2\Psi}{dt^2}$, where $\Psi(t)$ is (up to constants) the Fredholm determinant of the operator $f \to \int_{-t}^{t} K * f$ with kernel $K = \frac{1}{2\pi}\left(\frac{\sin(\xi-\eta)}{\xi-\eta} + \frac{\sin(\xi+\eta)}{\xi+\eta}\right)$. This distribution is well approximated by $p_W(s) = \frac{\pi}{2} s \exp\left(-\frac{\pi s^2}{4}\right)$.

Exercise 15.4.6. *Prove $p_W(s) = \frac{\pi}{2} s \exp\left(-\frac{\pi s^2}{4}\right)$ is a probability distribution with mean 1. What is its variance?*

We study the case of $N = 2$ and p a Gaussian in detail in Chapter 17.

Exercise[(hr)] 15.4.7 (Wigner's surmise)**.** *In 1957 Wigner conjectured that as $N \to \infty$ the spacing between adjacent normalized eigenvalues is given by*

$$p_W(s) \;=\; \frac{\pi}{2} s \exp\left(-\frac{\pi s^2}{4}\right). \tag{15.43}$$

He was led to this formula from the following assumptions:

- *Given an eigenvalue at x, the probability that another one lies s units to its right is proportional to s.*
- *Given an eigenvalue at x and $I_1, I_2, I_3, \ldots$ any disjoint intervals to the right of x, then the events of observing an eigenvalue in I_j are independent for all j.*
- *The mean spacing between consecutive eigenvalues is 1.*

Show these assumptions imply (15.43)*.*

15.4.2 Numerical Evidence

We provide some numerical support for the GOE Conjecture. In all the experiments below, we consider a large number of $N \times N$ matrices, where for each matrix we look at a small (small relative to N) number of eigenvalues in the **bulk of the eigenvalue spectrum** (eigenvalues near 0), not near the **edge** (for the semi-circle, eigenvalues near ± 1). We do not look at all the eigenvalues, as the average spacing changes over such a large range, nor do we consider the interesting case of the largest or smallest eigenvalues. We study a region where the average spacing is approximately constant, and as we are in the middle of the eigenvalue spectrum, there are no edge effects. These edge effects lead to fascinating questions (for random graphs, the distribution of eigenvalues near the edge is related to constructing good networks to rapidly transmit information; see for example [DSV, Sar]).

First we consider 5000 300×300 matrices with entries independently chosen from the uniform distribution on $[-1, 1]$ (see Figure 15.4). Notice that even with N as low as 300, we are seeing a good fit between conjecture and experiment.

What if we take p to be the Cauchy distribution? In this case, the second moment of p is infinite, and the alluded to argument for semi-circle behavior is not applicable. Simulations showed the density of eigenvalues did not follow the Semi-Circle Law, which does not contradict the theory as the conditions of the theorem were not met. What about the spacings between adjacent normalized eigenvalues of real symmetric matrices, with the entries drawn from the Cauchy distribution?

We study 5000 100×100 and then 5000 300×300 Cauchy matrices (see Figures 15.5 and 15.6. We note good agreement with the conjecture, and as N increases the fit improves.

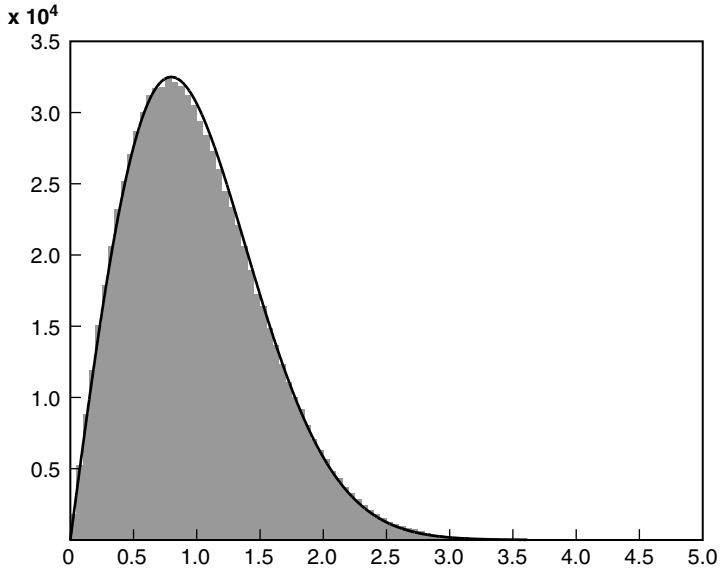

Figure 15.4 The local spacings of the central three-fifths of the eigenvalues of 5000 matrices (300×300) whose entries are drawn from the Uniform distribution on $[-1, 1]$

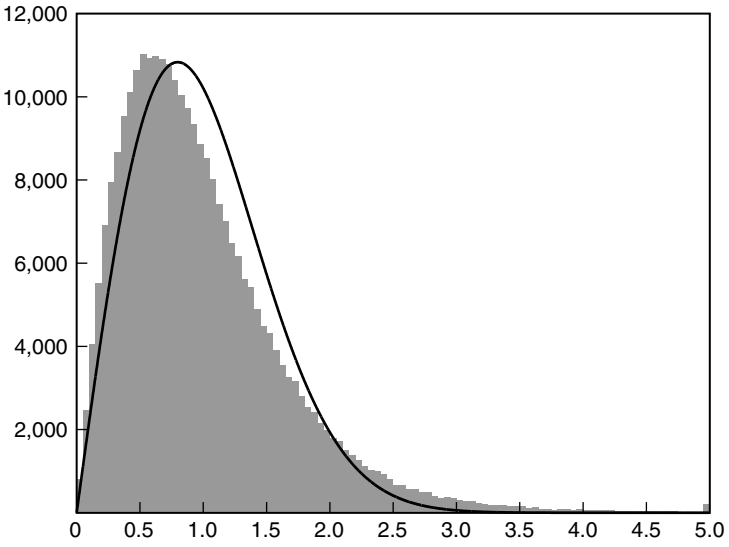

Figure 15.5 The local spacings of the central three-fifths of the eigenvalues of 5000 matrices (100×100) whose entries are drawn from the Cauchy distribution

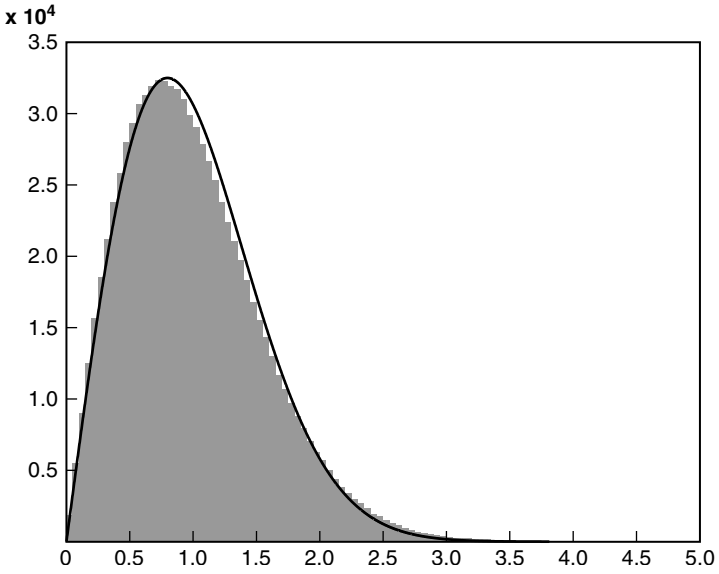

Figure 15.6 The local spacings of the central three-fifths of the eigenvalues of 5000 matrices (300×300) whose entries are drawn from the Cauchy distribution

We give one last example. Instead of using continuous probability distribution, we investigate a discrete case. Consider the Poisson Distribution:

$$p(n) = \frac{\lambda^n}{n!}\, e^{-\lambda}. \tag{15.44}$$

We investigate 5000 300×300 such matrices, first with $\lambda = 5$, and then with $\lambda = 20$, noting again excellent agreement with the GOE Conjecture (see Figures 15.7 and 15.8):

15.5 THIN SUB-FAMILIES

Before moving on to connections with number theory, we mention some very important subsets of real symmetric matrices. The subsets will be large enough so that there are averaging formulas at our disposal, but thin enough so that sometimes we see new behavior. Similar phenomena will resurface when we study zeros of Dirichlet L-functions.

As motivation, consider as our initial set all even integers. Let $N_2(x)$ denote the number of even integers at most x. We see $N_2(x) \sim \frac{x}{2}$, and the spacing between adjacent integers is 2. If we look at *normalized* even integers, we would have $y_i = \frac{2i}{2}$, and now the spacing between adjacent normalized even integers is 1.

Now consider the subset of even squares. If $N_\square(x)$ is the number of even squares at most x, then $N_\square(x) \sim \frac{\sqrt{x}}{2}$. For even squares of size x, say $x = (2m)^2$, the next

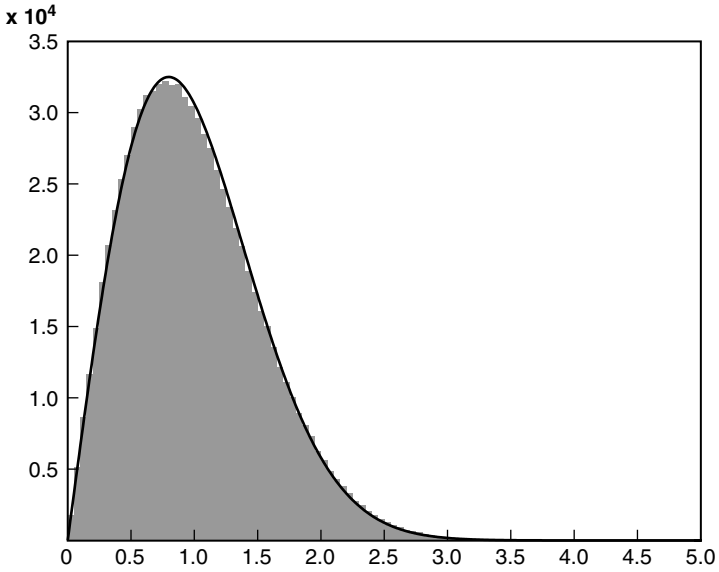

Figure 15.7 The local spacings of the central three-fifths of the eigenvalues of 5000 matrices (300×300) whose entries are drawn from the Poisson distribution ($\lambda = 5$)

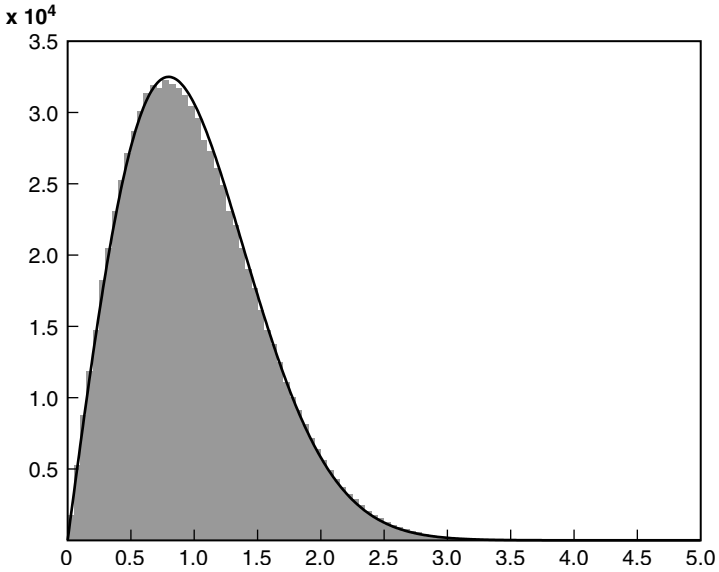

Figure 15.8 The local spacings of the central three-fifths of the eigenvalues of 5000 matrices (300×300) whose entries are drawn from the Poisson distribution ($\lambda = 20$)

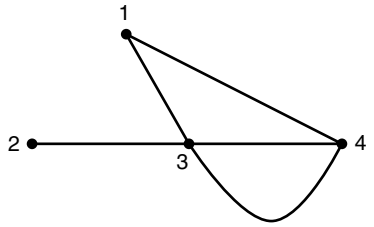

Figure 15.9 A typical graph

even square is at $(2m+2)^2 = x + 8m + 4$. Note the spacing between adjacent even squares is about $8m \sim 4\sqrt{x}$ for m large.

Exercise 15.5.1. *By appropriately normalizing the even squares, show we obtain a new sequence with a similar distribution of spacings between adjacent elements as in the case of normalized even integers. Explicitly, look at the spacings between N consecutive even squares with each square of size x and $N \ll x$.*

Remark 15.5.2. A far more interesting example concerns prime numbers. For the first set, consider all prime numbers. For the subset, fix an integer m and consider all prime numbers p such that $p + 2m$ is also prime; if $m = 1$ we say p and $p + 2$ are a twin prime pair. It is unknown if there are infinitely many elements in the second set for any m, though there are conjectural formulas (using the techniques of Chapter 13). It is fascinating to compare these two sets; for example, what is the spacing distribution between adjacent (normalized) primes look like, and is that the same for normalized twin prime pairs? See Research Project 12.9.5.

15.5.1 Random Graphs: Theory

A **graph** G is a collection of points (the **vertices** V) and lines connecting pairs of points (the **edges** E). While it is possible to have an edge from a vertex to itself (called a **self-loop**), we study the subset of graphs where this does not occur. We will allow multiple edges to connect the same two vertices (if there are no multiple edges, the graph is **simple**). The **degree of a vertex** is the number of edges leaving (or arriving at) that vertex. A graph is d-**regular** if every vertex has exactly d edges leaving (or arriving).

For example, consider the graph in Figure 15.9: The degrees of vertices are 2, 1, 4 and 3, and vertices 3 and 4 are connected with two edges.

To each graph with N vertices we can associate an $N \times N$ real symmetric matrix, called the **adjacency matrix**, as follows: First, label the vertices of the graph from 1 to N (see Exercise 15.5.3). Let a_{ij} be the number of edges from vertex i to vertex j. For the graph above, we have

$$A = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 2 \\ 1 & 0 & 2 & 0 \end{pmatrix}. \tag{15.45}$$

For each N, consider the space of all d-regular graphs. To each graph G we associate its adjacency matrix $A(G)$. We can build the eigenvalue probability distributions (see §15.2.3) as before. We can investigate the density of the eigenvalues and spacings between adjacent eigenvalues. We are no longer choosing the matrix elements at random; once we have chosen a graph, the entries are determined. Thus we have a more combinatorial type of averaging to perform: we average over all graphs, not over matrix elements. Even though these matrices are all real symmetric and hence a subset of the earlier ensembles, the probability density for these matrices are very different, and lead to different behavior (see also Remark 16.2.13 and §7.10).

One application of knowledge of eigenvalues of graphs is to network theory. For example, let the vertices of a graph represent various computers. We can transmit information between any two vertices that are connected by an edge. We desire a well connected graph so that we can transmit information rapidly through the system. One solution, of course, is to connect all the vertices and obtain the **complete graph**. In general, there is a cost for each edge; if there are N vertices in a simple graph, there are $\frac{N(N-1)}{2}$ possible edges; thus the complete graph quickly becomes very expensive. For N vertices, d-regular graphs have only $\frac{dN}{2}$ edges; now the cost is linear in the number of vertices. The distribution of eigenvalues (actually, the second largest eigenvalue) of such graphs provide information on how well connected it is. For more information, as well as specific constructions of such well connected graphs, see [DSV, Sar].

Exercise 15.5.3. *For a graph with N vertices, show there are $N!$ ways to label the vertices. Each labeling gives rise to an adjacency matrix. While a graph could potentially have $N!$ different adjacency matrices, show all adjacency matrices have the same eigenvalues, and therefore the same eigenvalue probability distribution.*

Remark 15.5.4. Fundamental quantities should not depend on presentation. Exercise 15.5.3 shows that the eigenvalues of a graph do not depend on how we label the graph. This is similar to the eigenvalues of an operator $T : \mathbb{C}^n \to \mathbb{C}^n$ do not depend on the basis used to represent T. Of course, the eigenvectors *will* depend on the basis.

Exercise 15.5.5. *If a graph has N labeled vertices and E labeled edges, how many ways are there to place the E edges so that each edge connects two distinct vertices? What if the edges are not labeled?*

Exercise 15.5.6 (Bipartite graphs)**.** *A graph is bipartite if the vertices V can be split into two distinct sets, A_1 and A_2, such that no vertices in an A_i are connected by an edge. We can construct a d-regular bipartite graph with $\#A_1 = \#A_2 = N$. Let A_1 be vertices $1, \dots, N$ and A_2 be vertices $N+1, \dots, 2N$. Let $\sigma_1, \dots, \sigma_d$ be permutations of $\{1, \dots, N\}$. For each σ_j and $i \in \{1, \dots, N\}$, connect vertex $i \in A_1$ to vertex $N + \sigma_j(i) \in A_2$. Prove this graph is bipartite and d-regular. If $d = 3$, what is the probability (as $N \to \infty$) that two vertices have two or more edges connecting them? What is the probability if $d = 4$?*

Remark 15.5.7. Exercise 15.5.6 provides a method for sampling the space of bipartite d-regular graphs, but does this construction sample the space uniformly (i.e., is every d-regular bipartite graph equally likely to be chosen by this method)? Further, is the behavior of eigenvalues of d-regular bipartite graphs the same as the behavior of eigenvalues of d-regular graphs? See [Bol], pages 50–57 for methods to sample spaces of graphs uniformly.

Exercise 15.5.8. *The **coloring number** of a graph is the minimum number of colors needed so that no two vertices connected by an edge are colored the same. What is the coloring number for the complete graph on N? For a bipartite graph with N vertices in each set?*

Consider now the following graphs. For any integer N let G_N be the graph with vertices the integers $2, 3, \dots, N$, and two vertices are joined if and only if they have a common divisor greater than 1. Prove the coloring number of G_{10000} is at least 13. Give good upper and lower bounds as functions of N for the coloring number of G_N.

15.5.2 Random Graphs: Results

The first result, due to McKay [McK], is that while the density of states is *not* the semi-circle there is a universal density for each d.

Theorem 15.5.9 (McKay's Law)**.** *Consider the ensemble of all d-regular graphs with N vertices. As $N \to \infty$, for almost all such graphs G, $\mu_{A(G),N}(x)$ converges to Kesten's measure*

$$f(x) \;=\; \begin{cases} \frac{d}{2\pi(d^2-x^2)}\sqrt{4(d-1)-x^2}, & |x| \le 2\sqrt{d-1} \\ 0 & \text{otherwise.} \end{cases} \tag{15.46}$$

Exercise 15.5.10. *Show that as $d \to \infty$, by changing the scale of x, Kesten's measure converges to the semi-circle distribution.*

Below (Figures 15.10 and 15.11) we see excellent agreement between theory and experiment for $d = 3$ and 6; the data is taken from [QS2].

The idea of the proof is that locally almost all of the graphs almost always look like trees (connected graphs with no loops), and for trees it is easy to calculate the eigenvalues. One then does a careful book-keeping. Thus, this sub-family is thin enough so that a new, universal answer arises. Even though all of these adjacency matrices are real symmetric, it is a very thin subset. It is *because* it is such a thin subset that we are able to see new behavior.

Exercise 15.5.11. *Show a general real symmetric matrix has $\frac{N(N+1)}{2}$ independent entries, while a d-regular graph's adjacency matrix has $\frac{dN}{2}$ non-zero entries.*

What about spacings between normalized eigenvalues? Figure 15.12 shows that, surprisingly, the result *does* appear to be the same as that from all real symmetric matrices. See [JMRR] for more details.

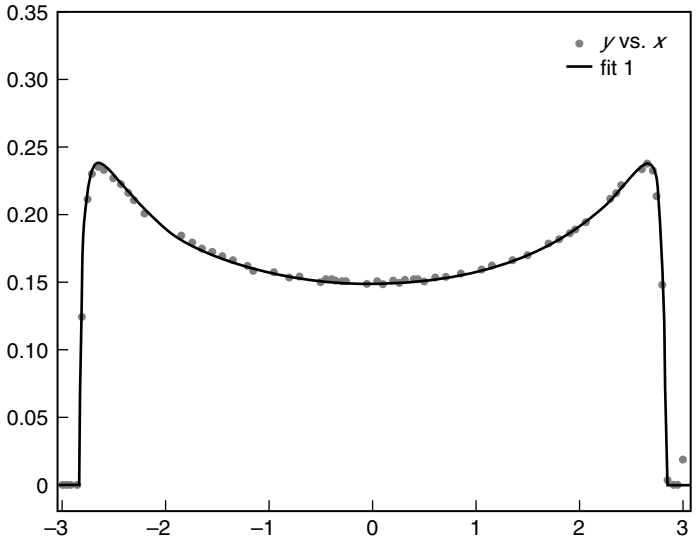

Figure 15.10 Comparison between theory (solid line) and experiment (dots) for 1000 eigenvalues of 3-regular graphs (120 bins in the histogram)

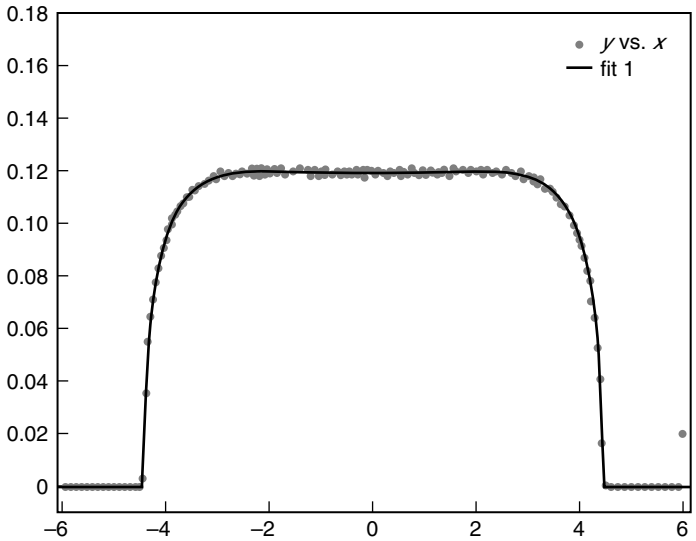

Figure 15.11 Comparison between theory (solid line) and experiment (dots) for 1000 eigenvalues of 6-regular graphs (240 bins in the histogram)

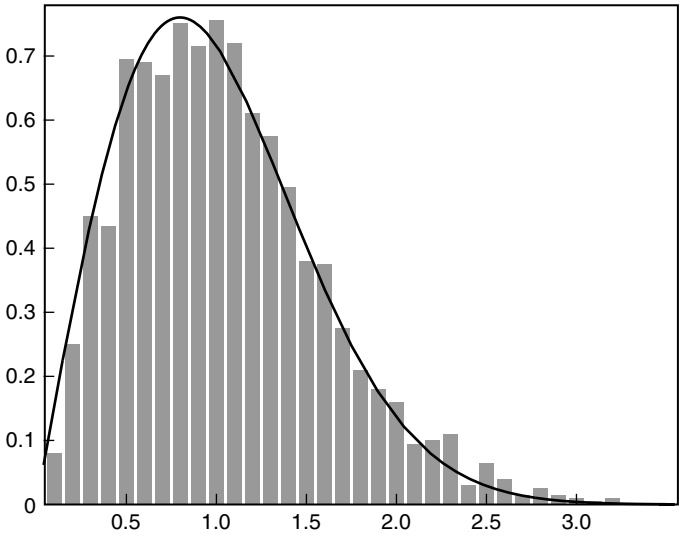

Figure 15.12 3-regular, 2000 vertices (graph courtesy of [JMRR])

15.6 NUMBER THEORY

We assume the reader is familiar with the material and notation from Chapter 3. For us an L-**function** is given by a **Dirichlet series** (which converges if $\Re s$ is sufficiently large), has an **Euler product**, and the coefficients have arithmetic meaning:

$$L(s,f) \;=\; \sum_{n=1}^{\infty} \frac{a_n(f)}{n^s} \;=\; \prod_p L_p(p^{-s},f)^{-1}, \quad \Re s > s_0. \tag{15.47}$$

The **Generalized Riemann Hypothesis** asserts that all non-trivial zeros have $\Re s = \frac{1}{2}$; i.e., they are on the **critical line** $\Re s = \frac{1}{2}$ and can be written as $\frac{1}{2} + i\gamma$, $\gamma \in \mathbb{R}$.

The simplest example is $\zeta(s)$, where $a_n(\zeta) = 1$ for all n; in Chapter 3 we saw how information about the distribution of zeros of $\zeta(s)$ yielded insights into the behavior of primes. The next example we considered were Dirichlet L-functions, the L-functions from Dirichlet characters χ of some conductor m. Here $a_n(\chi) = \chi(n)$, and these functions were useful in studying primes in arithmetic progressions.

For a fixed m, there are $\phi(m)$ Dirichlet L-functions modulo m. This provides our first example of a **family** of L-functions. We will not rigorously define a family, but content ourselves with saying a family of L-functions is a collection of "similar" L-functions.

The following examples will be considered families: (1) all Dirichlet L-functions with conductor m; (2) all Dirichlet L-functions with conductor $m \in [N, 2N]$; (3) all Dirichlet L-functions arising from quadratic characters with prime conductor $p \in [N, 2N]$. In each of the cases, each L-function has the same conductor, similar functional equations, and so on. It is not unreasonable to think they might share other properties.

Another example comes from elliptic curves. We commented in §4.2.2 that given a cubic equation $y^2 = x^3 + A_f x + B_f$, if $a_p(f) = p - N_p$ (where N_p is the number of solutions to $y^2 \equiv x^3 + A_f x + B_f \bmod p$), we can construct an L-function using the $a_p(f)$'s. We construct a family as follows. Let $A(T), B(T)$ be polynomials with integer coefficients in T. For each $t \in \mathbb{Z}$, we get an elliptic curve E_t (given by $y^2 = x^3 + A(t)x + B(t)$), and can construct an L-function $L(s, E_t)$. We can consider the family where $t \in [N, 2N]$.

Remark 15.6.1. Why are we considering "restricted" families, for example Dirichlet L-functions with a fixed conductor m, or $m \in [N, 2N]$, or elliptic curves with $t \in [N, 2N]$? The reason is similar to our random matrix ensembles: we do not consider infinite dimensional matrices: we study $N \times N$ matrices, and take the limit as $N \to \infty$. Similarly in number theory, it is easier to study finite sets, and then investigate the limiting behavior.

Assuming the zeros all lie on the line $\Re s = \frac{1}{2}$, similar to the case of real symmetric or complex Hermitian matrices, we can study spacings between zeros. We now describe some results about the distribution of zeros of L-functions. Two classical ensembles of random matrices play a central role: the Gaussian Orthogonal Ensemble **GOE** (resp., Gaussian Unitary Ensemble **GUE**), the space of real symmetric (complex Hermitian) matrices where the entries are chosen independently from Gaussians; see Chapter 17. It was observed that the spacings of energy levels of heavy nuclei are in excellent agreement with those of eigenvalues of real symmetric matrices; thus, the GOE became a common model for the energy levels. In §15.6.1 we see there is excellent agreement between the spacings of normalized zeros of L-functions and those of eigenvalues of complex Hermitian matrices; this led to the belief that the GUE is a good model for these zeros.

15.6.1 n-Level Correlations

In an amazing set of computations starting at the $10^{20\text{th}}$ zero, Odlyzko [Od1, Od2] observed phenomenal agreement between the spacings between adjacent normalized zeros of $\zeta(s)$ and spacings between adjacent normalized eigenvalues of complex Hermitian matrices. Specifically, consider the set of $N \times N$ random Hermitian matrices with entries chosen from the Gaussian distribution (the GUE). As $N \to \infty$ the limiting distribution of spacings between adjacent eigenvalues is indistinguishable from what Odlyzko observed in zeros of $\zeta(s)$!

His work was inspired by Montgomery [Mon2], who showed that for suitable test functions the pair correlation of the normalized zeros of $\zeta(s)$ agree with that of normalized eigenvalues of complex Hermitian matrices. Let $\{\alpha_j\}$ be an increasing sequence of real numbers, $B \subset \mathbb{R}^{n-1}$ a compact box. Define the n-**level correlation** by

$$\lim_{N\to\infty} \frac{\#\left\{\left(\alpha_{j_1} - \alpha_{j_2}, \dots, \alpha_{j_{n-1}} - \alpha_{j_n}\right) \in B, j_i \le N; j_i \ne j_k\right\}}{N}. \quad (15.48)$$

For example, the 2-level (or pair) correlation provides information on how often two normalized zeros (not necessarily adjacent zeros) have a difference in a given

interval. One can show that if all the n-level correlations could be computed, then we would know the spacings between adjacent zeros.

We can regard the box B as a product of $n-1$ characteristic functions of intervals (or binary indicator variables). Let

$$I_{a_i,b_i}(x) = \begin{cases} 1 & \text{if } x \in [a_i, b_i], \\ 0 & \text{otherwise.} \end{cases} \tag{15.49}$$

We can represent the condition $x \in B$ by $I_B(x) = \prod_{i=1}^{n} I_{a_i,b_i}(x_i)$. Instead of using a box B and its function I_B, it is more convenient to use an infinitely differentiable test function (see [RS] for details). In addition to the pair correlation and the numerics on adjacent spacings, Hejhal [Hej] showed for suitable test functions the 3-level (or triple) correlation for $\zeta(s)$ agrees with that of complex Hermitian matrices, and Rudnick-Sarnak [RS] proved (again for suitable test functions) that the n-level correlations of *any* "nice" L-function agree with those of complex Hermitian matrices.

The above work leads to the **GUE conjecture**: in the limit (as one looks at zeros with larger and larger imaginary part, or $N \times N$ matrices with larger and larger N), the spacing between zeros of L-functions is the same as that between eigenvalues of complex Hermitian matrices. In other words, the GUE is a good model of zeros of L-functions.

Even if true, however, the above cannot be the complete story.

Exercise 15.6.2. *Assume that the imaginary parts of the zeros of $\zeta(s)$ are unbounded. Show that if one removes any finite set of zeros, the n-level correlations are unchanged. Thus this statistic is insensitive to finitely many zeros.*

The above exercise shows that the n-level correlations are not sufficient to capture all of number theory. For many L-functions, there is reason to believe that there is different behavior near the central point $s = \frac{1}{2}$ (the center of the critical strip) than higher up. For example, the **Birch and Swinnerton-Dyer conjecture** (see §4.2.2) says that if $E(\mathbb{Q})$ (the group of rational solutions for an elliptic curve E; see §4.2.1) has rank r, then there are r zeros at the central point, and we might expect different behavior if there are more zeros.

Katz and Sarnak [KS1, KS2] proved that the n-level correlations of complex Hermitian matrices are also equal to the n-level correlations of the **classical compact groups**: unitary matrices (and its subgroups of symplectic and orthogonal matrices) with respect to Haar measure. Haar measure is the analogue of fixing a probability distribution p and choosing the entries of our matrices randomly from p; it should be thought of as specifying how we "randomly" chose a matrix from these groups. As a unitary matrix U satisfies $U^*U = I$ (where U^* is the complex conjugate transpose of U), we see each entry of U is at most 1 in absolute value, which shows unitary matrices are a compact group. A similar argument shows the set of orthogonal matrices Q such that $Q^TQ = I$ is compact.

What this means is that *many* different ensembles of matrices have the same n-level correlations – there is not one unique ensemble with these values. This led to a new statistic which is different for different ensembles, and allows us to "determine" which matrix ensemble the zeros follow.

Remark 15.6.3 (Advanced)**.** Consider the following classical compact groups: $U(N)$, $USp(2N)$, SO, SO(even) and SO(odd) with their Haar measure. Fix a group and choose a generic matrix element. Calculating the n-level correlations of its eigenvalues, integrating over the group, and taking the limit as $N \to \infty$, Katz and Sarnak prove the resulting answer is universal, independent of the particular group chosen. In particular, we cannot use the n-level correlations to distinguish the other classical compact groups from each other.

15.6.2 1-Level Density

In the n-level correlations, given an L-function we studied differences between zeros. It can be shown that any "nice" L-function has infinitely many zeros on the line $\Re s = \frac{1}{2}$; thus, if we want to study "high" zeros (zeros very far above the central point $s = \frac{1}{2}$), each L-function has enough zeros to average over.

The situation is completely different if we study "low" zeros, zeros near the central point. Now each L-function only has a few zeros nearby, and there is nothing to average: wherever the zeros are, that's where they are! This led to the introduction of families of L-functions. For example, consider Dirichlet L-functions with characters of conductor m. There are $\phi(m)$ such L-functions. For each L-function we can calculate properties of zeros near the central point and then we can *average* over the $\phi(m)$ L-functions, taking the limit as $m \to \infty$.

Explicitly, let $h(x)$ be a continuous function of rapid decay. For an L-function $L(s, f)$ with non-trivial zeros $\frac{1}{2} + i\gamma_f$ (assuming GRH, each $\gamma_f \in \mathbb{R}$), consider

$$D_f(h) = \sum_j h\left(\gamma_f \frac{\log c_f}{2\pi}\right). \tag{15.50}$$

Here c_f is the **analytic conductor**; basically, it rescales the zeros near the central point. As h is of rapid decay, almost all of the contribution to (15.50) will come from zeros very close to the central point. We then average over all f in a family $\mathcal{F}$. We call this statistic the 1-**level density**:

$$D_{\mathcal{F}}(h) = \frac{1}{|\mathcal{F}|} \sum_{f \in \mathcal{F}} D_f(h). \tag{15.51}$$

Katz and Sarnak conjecture that the distribution of zeros near the **central point** $s = \frac{1}{2}$ in a family of L-functions should agree (in the limit) with the distribution of eigenvalues near 1 of a classical compact group (unitary, symplectic, orthogonal); *which* group depends on underlying symmetries of the family. The important point to note is that the GUE is not the entire story: other ensembles of matrices naturally arise. These conjectures, for suitable test functions, have been verified for a variety of families: we sketch the proof for Dirichlet L-functions in Chapter 18 and give an application as well.

Remark 15.6.4. Why does the central point $s = \frac{1}{2}$ correspond to the eigenvalue 1? As the classical compact groups are subsets of the unitary matrices, their eigenvalues can be written $e^{i\theta}$, $\theta \in (-\pi, \pi]$. Here $\theta = 0$ (corresponding to an eigenvalue of 1) is the center of the "critical line." Note certain such matrices have a forced

eigenvalue at 1 (for example, any $N \times N$ orthogonal matrix with N odd); this is expected to be similar to L-functions with a forced zeros at the central point. The situation with multiple forced zeros at the central point is very interesting; while in some cases the corresponding random matrix models are known, other cases are still very much open. See [Mil6, Sn] for more details.

Exercise$^{(h)}$ 15.6.5. *U is a unitary matrix if $U^*U = I$, where U^* is the complex conjugate transpose of U. Prove the eigenvalues of unitary matrices can be written as $e^{i\theta_j}$ for $\theta_j \in \mathbb{R}$. An orthogonal matrix is a real unitary matrix; thus $Q^TQ = I$ where Q^T is the transpose of Q. Must the eigenvalues of an orthogonal matrix be real?*

Remark 15.6.6 (Advanced)**.** In practice, one takes h in (15.50) to be a Schwartz function whose Fourier transform has finite support (see §11.4.1). Similar to the n-level correlations, one can generalize the above and study n-level densities. The determination of which classical compact group can sometimes be calculated by studying the monodromy groups of function field analogues.

We sketch an interpretation of the 1-level density. Again, the philosophy is that to each family of L-functions $\mathcal{F}$ there is an ensemble of random matrices $G(\mathcal{F})$ (where $G(\mathcal{F})$ is one of the classical compact groups), and to each $G(\mathcal{F})$ is attached a density function $W_{G(\mathcal{F})}$. Explicitly, consider the family of all non-trivial Dirichlet L-functions with prime conductor m, denoted by $\mathcal{F}_m$. We study this family in detail in Chapter 18. Then for suitable test functions h, we prove

$$\begin{aligned}\lim_{m\to\infty} D_{\mathcal{F}_m}(h) &= \lim_{m\to\infty} \frac{1}{|\mathcal{F}_m|} \sum_{\chi\in\mathcal{F}_m} \sum_{\gamma_\chi} h\left(\gamma_\chi \frac{\log c_\chi}{2\pi}\right) \\ &= \int_{-\infty}^{\infty} h(x) W_{G(\mathcal{F})}(x)dx. \end{aligned} \tag{15.52}$$

We see that summing a test function of rapid decay over the scaled zeros is equivalent to integrating that test function against a family-dependent density function. We can see a similar phenomenon if we study sums of test functions at primes. For simplicity of presentation, we assume the Riemann Hypothesis to obtain better error estimates, though it is not needed (see Exercise 15.6.8).

Theorem 15.6.7. *Let F and its derivative F' be continuously differentiable functions of rapid decay; it suffices to assume $\int |F(x)|dx$ and $\int |F'(x)|dx$ are finite. Then*

$$\sum_p \frac{\log p}{p \log N} F\left(\frac{\log p}{\log N}\right) = \int_0^\infty F(x)dx + O\left(\frac{1}{\log N}\right). \tag{15.53}$$

Sketch of the proof. By the Riemann Hypothesis and partial summation (Theorem 2.2.6), we have

$$\sum_{p\le x} \log p = x + O(x^{\frac{1}{2}} \log^2(x)). \tag{15.54}$$

See [Da2] for how this bound follows from RH. We apply the integral version of partial summation (Theorem 2.2.6) to

$$\sum_{p\le x} \log p \cdot \frac{1}{p}. \tag{15.55}$$

In the notation of Theorem 2.2.6, $a_n = \log p$ if p is prime and 0 otherwise, and $h(x) = \frac{1}{x}$. We find

$$\sum_{p\le x} \frac{\log p}{p} \;=\; O(1) - \int_2^x (u + O(u^{\frac12}\log^2 u))\frac{-1}{u^2}du \;=\; \log x + O(1). \tag{15.56}$$

We again use the integral version of partial summation, but now on $\frac{\log p}{p}\cdot F\left(\frac{\log p}{\log N}\right)$ where $a_n = \frac{\log p}{p}$ for p prime and $h(x) = F\left(\frac{\log x}{\log N}\right)$. Let $u_0 = \frac{\log 2}{\log N}$. Then

$$\begin{aligned}
&\sum_{p\ge 2} \frac{\log p}{p} F\left(\frac{\log p}{\log N}\right) \;=\; -\int_2^\infty (\log x + O(1))\,\frac{d}{dx}F\left(\frac{\log x}{\log N}\right)dx \\
&= \int_2^\infty \left[\frac{1}{x}F\left(\frac{\log x}{\log N}\right) + O\left(\frac{1}{x\log N}\left|F'\left(\frac{\log x}{\log N}\right)\right|\right)\right]dx \\
&= \log N \int_{u_0}^\infty \left[F(u) + O\left(\frac{1}{\log N}\,|F'(u)|\right)\right]du \\
&= \log N \int_0^\infty \left[F(u) + O\left(\frac{|F'(u)|}{\log N}\right)\right]du + O(u_0 \log N \max_{t\in[0,u_0]} F(t)) \\
&= \log N \int_0^\infty F(u)du + O\left(\int_0^\infty |F'(u)|du\right) + O\left(u_0\log N \max_{t\in[0,u_0]} F(t)\right) \\
&= \log N \int_0^\infty F(u)du + O(1),
\end{aligned} \tag{15.57}$$

as $u_0 = \frac{\log 2}{\log N}$ and our assumption that F' is of rapid decay ensures that the F' integral is $O(1)$. Dividing by $\log N$ yields the theorem. Using the Prime Number Theorem instead of RH yields the same result, but with a worse error term. □

Exercise 15.6.8. *Redo the above arguments using the bounds from §13.2.6, which eliminate the need to assume the Riemann Hypothesis.*

The above shows that summing a nice test function at the primes is related to integrating that function against a density; here the density is just 1. The 1-level density is a generalization of this to summing weighted zeros of L-functions, and the density we integrate against depends on properties of the family of L-functions. See §4.3.3 for more on distribution of points.

Exercise 15.6.9. *How rapidly must F decay as $x \to \infty$ to justify the arguments above? Clearly if F has compact support (i.e., if $F(x)$ is zero if $|x| > R$ for some R), F decays sufficiently rapidly, and this is often the case of interest.*

Exercise 15.6.10. *Why is the natural scale for Theorem 15.6.7 $\log N$ (i.e., why is it natural to evaluate the test function at $\frac{\log p}{\log N}$ and not p)?*

Exercise 15.6.11. *Instead of studying all primes, fix m and b with $(b, m) = 1$, and consider the set of primes $p \equiv b \bmod m$ (recall such p are called **primes in an arithmetic progression**); see §3.3. Modify the statement and proof of Theorem 15.6.7 to calculate the density for primes in arithmetic progression. If instead we consider twin primes, and we assume the number of twin primes at most x satisfies $\pi_2(x) = T_2 \frac{x}{\log^2 x} + O(x^{\frac{1}{2}+\epsilon})$ for some constant T_2, what is the appropriate normalization and density? See Definition 14.6.17 for the conjectured value of T_2.*

15.7 SIMILARITIES BETWEEN RANDOM MATRIX THEORY AND L-FUNCTIONS

The following (conjectural) correspondence has led to many fruitful predictions: in some sense, the zeros of L-functions behave like the eigenvalues of matrices which in turn behave like the energy levels of heavy nuclei. To study the energy levels of heavy nuclei, physicists bombard them with neutrons and study what happens; however, physical constraints prevent them from using neutrons of arbitrary energy. Similarly, we want to study zeros of L-functions. We "bombard" the zeros with a test function, but not an arbitrary one (*advanced:* the technical condition is the support of the Fourier transform of the test function must be small; the test function's support corresponds to the neutron's energy). To evaluate the sums of the test function at the zeros, similar to physicists restricting the neutrons they can use, number theorists can evaluate the sums for only a small class of test functions.

Similar to our proofs of the Semi-Circle Law, we again have three key ingredients. The first is we average over a collection of objects. Before it was the probability measures built from the normalized eigenvalues, now it is the $D_f(h)$ for each L-function f in the family for a fixed test function h. Second, we need some type of Trace Formula, which tells us what the correct scale is to study our problem and allows us to pass from knowledge of what we can sum to knowledge about what we want to understand. For matrices, we passed from sums over eigenvalues (which we wanted to understand) to sums over the matrix elements (which we were given and could execute). For number theory, using what are known as Explicit Formulas (see §18.1), we pass from sums over zeros in (15.50) to sums over the coefficients $a_n(f)$ in the L-functions. Finally, the Trace Formula is useless if we do not have some type of Averaging Formula. For matrices, because of how we generated matrices at random, we were able to average over the matrix elements; for number theory, one needs powerful theorem concerning averages of $a_n(f)$ as f ranges over a family. We have already seen a special case where there is an averaging relation: the orthogonality relations for Dirichlet characters (see Lemma 3.3.15). In §18.3 we summarize the similarities between Random Matrix Theory and Number Theory calculations. We give an application of the 1-level density to number theory in Theorem 18.2.7, namely bounding the number of characters χ such that $L(s, \chi)$ is non-zero at the central point. See [IS1, IS2] for more on non-vanishing of L-functions at the central point and applications of such results.

15.8 SUGGESTIONS FOR FURTHER READING

In addition to the references in this and subsequent chapters, we provide a few starting points to the vast literature; the interested reader should consult the bibliographies of the references for additional reading.

A terrific introduction to classical random matrix theory is [Meh2], whose exposition has motivated our approach and many others; see also [For]. We recommend reading at least some of the original papers of Wigner [Wig1, Wig2, Wig3, Wig4, Wig5] and Dyson [Dy1, Dy2]. For a more modern treatment via Haar measure, see [KS2]. Many of the properties of the classical compact groups can be found in [Weyl]. See [Ha] for an entertaining account of the first meeting of Random Matrix Theory and Number Theory, and [Roc] for an accessible tour of connections between $\zeta(s)$ and much of mathematics.

In Chapter 16 we sketch a proof of the Semi-Circle Law. See [CB] for a rigorous treatment (including convergence issues and weaker conditions on the distribution p). For more information, we refer the reader to [Bai, BK]. In Chapter 17 we investigate the spacings of eigenvalues of 2×2 matrices. See [Gau, Meh1, Meh2] for the spacings of $N \times N$ matrices as $N \to \infty$.

In Chapter 18 we study the 1-level density for all Dirichlet characters with conductor m, and state that as $m \to \infty$ the answer agrees with the similar statistic for unitary matrices (see [HuRu, Mil2]). If we look just at quadratic Dirichlet characters (Legendre symbols), then instead of seeing unitary symmetry one finds agreement with eigenvalues of symplectic matrices (see [Rub2]). This is similar to the behavior of eigenvalues of adjacency matrices of d-regular graphs, which are a very special subset of real symmetry matrices but have different behavior. For more on connections between random graphs and number theory, see [DSV] and Chapter 3 of [Sar]; see [Bol, McK, MW, Wor] and the student reports [Cha, Gold, Nov, Ric, QS2] for more on random graphs.

The 1-level density (see also [ILS, Mil1]) and n-level correlations [Hej, Mon2, RS] are but two of many statistics where random matrices behave similarly as L-functions. We refer the reader to the survey articles [Con1, Dia, FSV, KS2, KeSn], Chapter 25 of [IK] and to the research works [CFKRS, DM, FSV, KS1, Mil6, Od1, Od2, Sn, TrWi] for more information.