Introduction to Random Matrix Theory

from *An Invitation to Modern Number Theory*

Steven J. Miller and
Ramin Takloo-Bighash

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

Contents

Chapter One

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 **??**. 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 **??**), probability theory (Chapter **??**) and linear algebra (a quick review of the needed background is provided in Appendix **??**).

While we assume the reader has some familiarity with the basic concepts in physics for the historical introduction in §1.1, 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.

1.1 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 **??** 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.

1.1.1 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 **??**.

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

$$H\psi_n \;=\; E_n\psi_n, \tag{1.1}$$

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.

1.1.2 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 1.1.

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 matri-

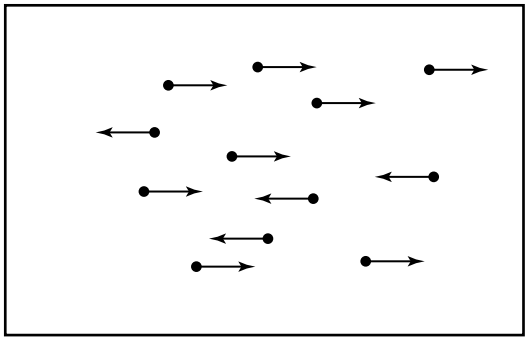

Figure 1.1 Molecules in a box

ces 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 (§**??**). 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 1.1.1. *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 1.1.2. *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.*

1.1.3 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,

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1N} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & a_{N3} & \cdots & a_{NN} \end{pmatrix} = A^T \tag{1.2}$$

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

$$\text{Prob}(A)dA = \prod_{1 \le i \le j \le N} p(a_{ij})da_{ij}. \tag{1.3}$$

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,

$$\text{Prob}(A : a_{ij} \in [\alpha_{ij}, \beta_{ij}]) = \prod_{1 \le i \le j \le N} \int_{\alpha_{ij}}^{\beta_{ij}} p(a_{ij})da_{ij}. \tag{1.4}$$

Example 1.1.3. *For a 2×2 real symmetric matrix we would have*

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}, \quad \text{Prob}(A)dA = p(a_{11})p(a_{12})p(a_{22})da_{11}da_{12}da_{22}. \tag{1.5}$$

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

$$\int \text{Prob}(A)dA = \prod_{1 \le i \le j \le N} \int_{a_{ij}=-\infty}^{\infty} p(a_{ij})da_{ij} = 1; \tag{1.6}$$

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

$$\begin{aligned} p(x) &\ge 0 \\ \int_{-\infty}^{\infty} p(x)dx &= 1 \\ \int_{-\infty}^{\infty} |x|^k p(x)dx &< \infty. \end{aligned} \tag{1.7}$$

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 1.1.4. *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*

$$\lim_{A \to \infty} \int_{-A}^{A} x^k p(x)dx = 0 \quad \textit{but} \quad \lim_{A \to \infty} \int_{-A}^{2A} x^k p(x)dx = \infty. \tag{1.8}$$

Exercise 1.1.5. *Let p be a probability density such that all of its moments exist. If p is an even function ($p(-x) = p(x)$), show all the odd moments vanish.*

Exercise 1.1.6. *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 1.1.7. *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 1.1.8 (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 1.1.9. *Consider the ensemble of 2×2 real symmetric matrices A where for a matrix $A = \left(\begin{smallmatrix} x & y \\ y & z \end{smallmatrix} \right)$,*

$$p(A) \;=\; \begin{cases} \frac{3}{4\pi} & \text{if } x^2 + y^2 + z^2 \le 1 \\ 0 & \text{otherwise.} \end{cases} \tag{1.9}$$

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

$$\begin{aligned} x \;=\; x(r,\theta,\phi) \;&=\; r\cos(\theta)\sin(\phi) \\ y \;=\; y(r,\theta,\phi) \;&=\; r\sin(\theta)\sin(\phi) \\ z \;=\; z(r,\theta,\phi) \;&=\; r\cos(\phi), \end{aligned} \tag{1.10}$$

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 **??** 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 1.1.10 (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 1.1.11. *Prove the eigenvalues of real symmetric and complex Hermitian matrices are real.*

Exercise 1.1.12. *How many degrees of freedom does a complex Hermitian matrix have?*

1.2 EIGENVALUE PRELIMINARIES

1.2.1 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:

$$\mathrm{Trace}(A) \;=\; a_{11} + \cdots + a_{NN}. \tag{1.11}$$

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

$$A \;=\; \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{1.12}$$

has

$$\mathrm{Trace}(A^2) \;=\; a_{11}a_{11} + a_{12}a_{21} + a_{12}a_{21} + a_{22}a_{22} \;=\; \sum_{i=1}^{2}\sum_{j=1}^{2} a_{ij}a_{ji}. \tag{1.13}$$

In general we have

Theorem 1.2.1. *Let A be an $N \times N$ matrix. Then*

$$\mathrm{Trace}(A^k) \;=\; \sum_{i_1=1}^{N} \cdots \sum_{i_k=1}^{N} a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{k-1} i_k} a_{i_k i_1}. \tag{1.14}$$

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 1.2.2. *Show* (1.13) *is consistent with Theorem 1.2.1.*

Exercise 1.2.3. *Prove Theorem 1.2.1.*

Theorem 1.2.4 (Eigenvalue Trace Formula)**.** *For any non-negative integer k, if A is an $N \times N$ matrix with eigenvalues $\lambda_i(A)$, then*

$$\mathrm{Trace}(A^k) \;=\; \sum_{i=1}^{N} \lambda_i(A)^k. \tag{1.15}$$

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:

$$\mathrm{Trace}(A^k) \;=\; \sum_{i=1}^{N} \lambda_i(A^k) \;=\; \sum_{i=1}^{N} \lambda_i(A)^k. \tag{1.16}$$

□

Exercise 1.2.5. *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 $\overrightarrow{v_1}$ 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.*

1.2.2 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 **??** for an exact statement of the error term) states that $\pi(x)$, the number of primes less than x, satisfies

$$\pi(x) \;=\; \frac{x}{\log x} \;+\; \text{lower order terms}. \tag{1.17}$$

Remark 1.2.6. 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 **??**) 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 **??** 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 1.2.1, for $A = A^T$ we have that

$$\text{Trace}(A^2) \;=\; \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}a_{ji} \;=\; \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2. \tag{1.18}$$

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

$$\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2 \;\sim\; N^2 \cdot 1, \tag{1.19}$$

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

$$\sum_{i=1}^{N} \lambda_i(A)^2 \;\sim\; N^2, \tag{1.20}$$

which yields

$$N \cdot \text{Ave}(\lambda_i(A)^2) \;\sim\; N^2. \tag{1.21}$$

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

$$|\text{Ave}(\lambda_i(A))| \;\sim\; \sqrt{N}. \tag{1.22}$$

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 1.2.7. *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 1.2.8. *Use Chebyshev's inequality (Exercise **??**) 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.*

1.2.3 Eigenvalue Distribution

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

$$\delta_{x_0}(f) \ := \ \int_{-\infty}^{\infty} f(x)\delta(x-x_0)dx \ = \ f(x_0). \tag{1.23}$$

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**):

$$\mu_{A,N}(x)dx \ = \ \frac{1}{N}\sum_{i=1}^{N}\delta\left(x-\frac{\lambda_i(A)}{2\sqrt{N}}\right)dx. \tag{1.24}$$

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 fraction of normalized eigenvalues in $[a,b]$:

$$\int_a^b \mu_{A,N}(x)dx \ = \ \frac{\#\{i : \frac{\lambda_i(A)}{2\sqrt{N}} \in [a,b]\}}{N}. \tag{1.25}$$

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

Definition 1.2.9. *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 1.3.15 why it is so useful.

Lemma 1.2.10. $M_{N,k}(A) = \frac{\text{Trace}(A^k)}{2^k N^{\frac{k}{2}+1}}$.

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

$$\begin{aligned} M_{N,k}(A) \ = \ \mathbb{E}[x^k]_A \ &= \ \int x^k \mu_{A,N}(x)dx \\ &= \ \frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}} x^k \delta\left(x-\frac{\lambda_i(A)}{2\sqrt{N}}\right)dx \\ &= \ \frac{1}{N}\sum_{i=1}^{N}\frac{\lambda_i(A)^k}{(2\sqrt{N})^k} \\ &= \ \frac{\text{Trace}(A^k)}{2^k N^{\frac{k}{2}+1}}. \end{aligned} \tag{1.26}$$

□

Exercise 1.2.11. *Let A be an $N\times N$ real symmetric matrix with $|a_{ij}| \le B$. In terms of B, N and k bound $|\text{Trace}(A^k)|$ and $M_{N,k}(A)$. How large can $\max_i |\lambda_i(A)|$ be?*

1.3 SEMI-CIRCLE LAW

1.3.1 Statement

A natural question to ask concerning the eigenvalues of a matrix is: *What fraction 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

$$\int_a^b \mu_{A,N}(x)dx. \tag{1.27}$$

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 1.3.1 (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 fraction of normalized eigenvalues of A in $[a, b] \subset [-1, 1]$ for a typical A as $N \to \infty$ is

$$\int_a^b \frac{2}{\pi}\sqrt{1-x^2}dx. \tag{1.28}$$

Later in §1.3.4 we discuss what happens if the higher moments are infinite.

1.3.2 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 1.3.2) exists in an interval and the distribution is uniquely determined.

Exercise 1.3.2 (Non-uniqueness of moments)**.** *For $x \in [0, \infty)$, consider*

$$\begin{aligned} f_1(x) &= \frac{1}{\sqrt{2\pi}x}e^{-(\log x)^2/2} \\ f_2(x) &= f_1(x)\left[1 + \sin(2\pi \log x)\right]. \end{aligned} \tag{1.29}$$

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***

$$M_f(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx \tag{1.30}$$

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

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

$$M_{N,k} \;=\; \int_A M_{N,k}(A)\text{Prob}(A)dA, \quad m_k \;=\; \lim_{N\to\infty} M_{N,k}. \tag{1.31}$$

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 1.3.3 (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 1.3.5).

Remark 1.3.4. While it is not true that every sequence of numbers m_k that grow sufficiently slowly determines a continuous probability density (see Exercise 1.3.8), 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 **??**.

Exercise 1.3.5. *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 $[-\infty, x]$:

$$C_f(x) \;=\; \int_{-\infty}^x f(x)dx. \tag{1.32}$$

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

$$C_{\mu_{A,N}}(x) \;=\; \frac{1}{N} \sum_{\frac{\lambda_i(A)}{2\sqrt{N}} < x} 1, \tag{1.33}$$

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 fraction 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 1.3.6. 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 1.3.7. *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 1.3.8. *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 1.3.9. *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 1.3.10) grow sufficiently rapidly so that the Gaussian has unbounded support. If $\lim_{j\to\infty} m_{2j}^{1/2j} < \infty$ must the distribution have finite support?*

Exercise 1.3.10 (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 1.3.11. *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 1.3.12. *Do there exist two probability distributions that have a large Kolmogoroff-Smirnov discrepancy but are close pointwise?*

1.3.3 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 2. We use Moment Method from §1.3.2.

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 §2.1. By Lemma 1.2.10,

$$\begin{aligned} M_{N,2} &= \int_A M_{N,k}(A)\text{Prob}(A)dA \\ &= \frac{1}{2^2 N^{\frac{2}{2}+1}} \int_A \text{Trace}(A^2)\text{Prob}(A)dA. \end{aligned} \tag{1.34}$$

We use Theorem 1.2.1 to expand $\text{Trace}(A^2)$ and find

$$M_{N,2} = \frac{1}{2^2 N^2} \int_A \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2 \, \text{Prob}(A)dA. \tag{1.35}$$

We now expand $\text{Prob}(A)dA$ by (1.3):

$$\begin{aligned} & M_{N,2} \\ &= \frac{1}{2^2 N^2} \int_{a_{11}=-\infty}^{\infty} \cdots \int_{a_{NN}=-\infty}^{\infty} \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2 \cdot p(a_{11})da_{11}\cdots p(a_{NN})da_{NN} \\ &= \frac{1}{2^2 N^2} \sum_{i=1}^{N}\sum_{j=1}^{N} \int_{a_{11}=-\infty}^{\infty} \cdots \int_{a_{NN}=-\infty}^{\infty} a_{ij}^2 \cdot p(a_{11})da_{11}\cdots p(a_{NN})da_{NN}; \end{aligned} \tag{1.36}$$

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

$$\int_{a_{ij}=-\infty}^{\infty} a_{ij}^2 p(a_{ij})da_{ij} \cdot \prod_{\substack{(k,l)\neq(ij)\\ k\leq l}} \int_{a_{kl}=-\infty}^{\infty} p(a_{kl})da_{kl}. \tag{1.37}$$

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 (1.36), 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.

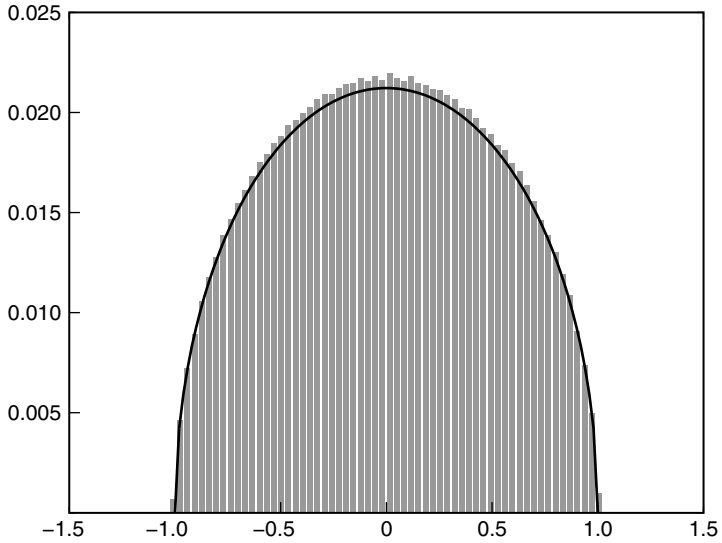

Figure 1.2 Distribution of eigenvalues: 500 Gaussian matrices (400×400)

Exercise 1.3.13. *Show the second moment of the semi-circle is* $\frac{1}{4}$.

Exercise 1.3.14. *Calculate the third and fourth moments, and compare them to those of the semi-circle.*

Remark 1.3.15 (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 1.2.4.

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 1.3.16. 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.

1.3.4 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 1.2 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

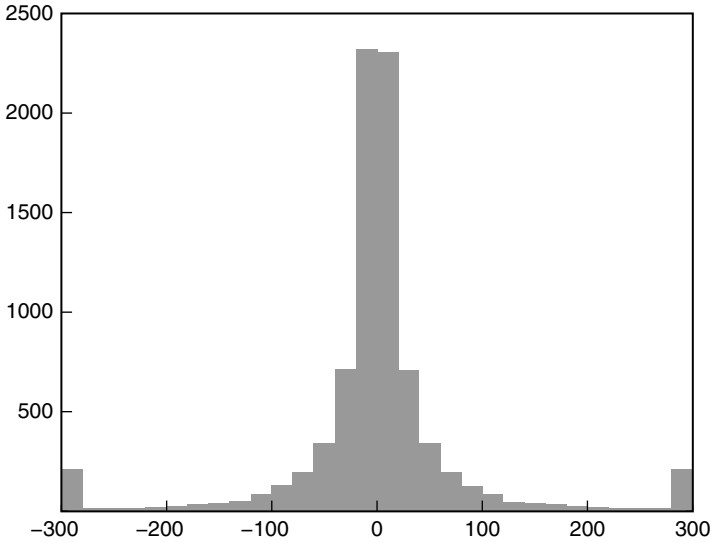

Figure 1.3 Distribution of eigenvalues: 5000 Cauchy matrices (300×300)

Cauchy distribution,

$$p(x) \ = \ \frac{1}{\pi(1+x^2)}. \tag{1.38}$$

The behavior is clearly not semi-circular (see Figure 1.3). 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 1.3.17. *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.*

1.3.5 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 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].

1.4 ADJACENT NEIGHBOR SPACINGS

1.4.1 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 1.4.1. *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:

$$\lambda_1(A) \leq \lambda_2(A) \leq \cdots \leq \lambda_N(A). \tag{1.39}$$

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

$$\lambda_2(A) - \lambda_1(A),\ \lambda_3(A) - \lambda_2(A),\ \ldots,\ \lambda_N(A) - \lambda_{N-1}(A). \tag{1.40}$$

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

$$\frac{\lambda_2(A)}{2\sqrt{N}} - \frac{\lambda_1(A)}{2\sqrt{N}},\ \ldots,\ \frac{\lambda_N(A)}{2\sqrt{N}} - \frac{\lambda_{N-1}(A)}{2\sqrt{N}}. \tag{1.41}$$

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 1.4.2.

$$\nu_{A,N}(s)ds = \frac{1}{N-1}\sum_{i=2}^{N} \delta\left(s - \frac{\lambda_i(A) - \lambda_{i-1}(A)}{2\sqrt{N}}\right) ds. \tag{1.42}$$

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

Remark 1.4.4. GOE stands for Gaussian Orthogonal Ensemble; the conjecture is known if p is (basically) a Gaussian. We explain the nomenclature in Chapter **??**.

Remark 1.4.5 (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 1.4.6. *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 **??**.

Exercise$^{(hr)}$ 1.4.7 (Wigner's surmise)**.** *In* 1957 *Wigner conjectured that as* $N \to \infty$ *the spacing between adjacent normalized eigenvalues is given by*

$$p_W(s) \;=\; \frac{\pi}{2} s \exp\left(-\frac{\pi s^2}{4}\right). \tag{1.43}$$

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 (1.43)*.*

1.4.2 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 1.4). 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 1.5 and 1.6. We note good agreement with the conjecture, and as N increases the fit improves.

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

$$p(n) = \frac{\lambda^n}{n!}\, e^{-\lambda}. \tag{1.44}$$

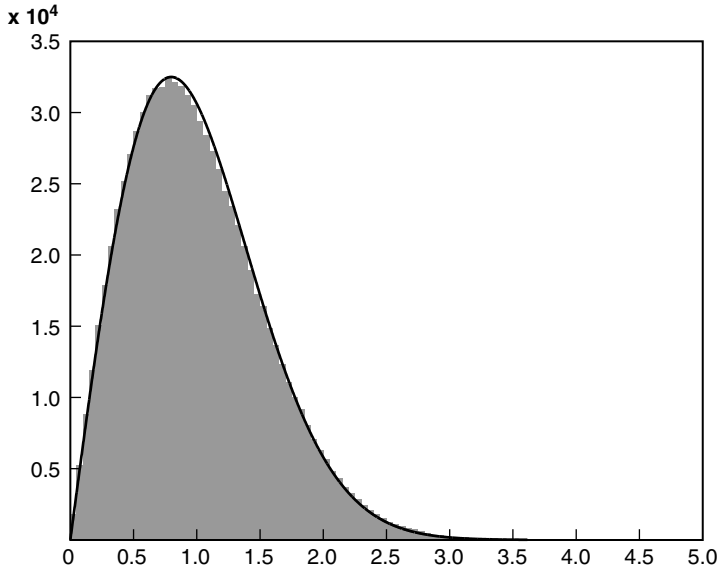

Figure 1.4 The local spacings of the central three-fifths of the eigenvalues of 5000 matrices (300×300) whose entries are drawn from the Uniform distribution on $[-1, 1]$

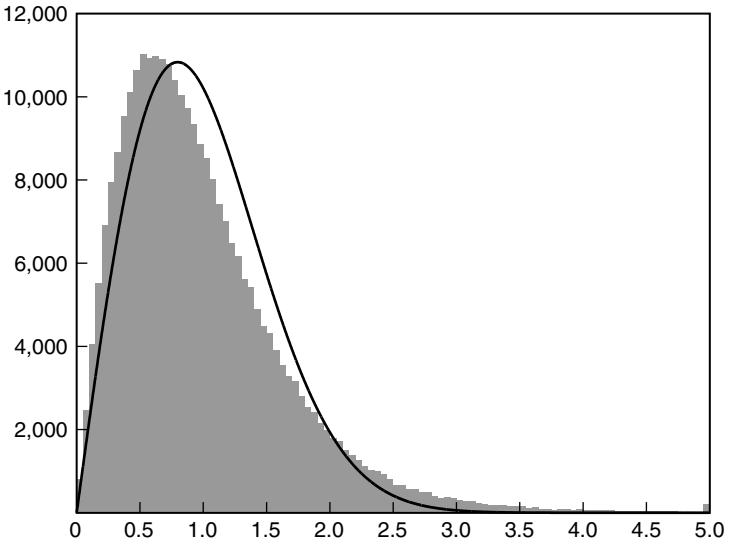

Figure 1.5 The local spacings of the central three-fifths of the eigenvalues of 5000 matrices (100×100) whose entries are drawn from the Cauchy distribution

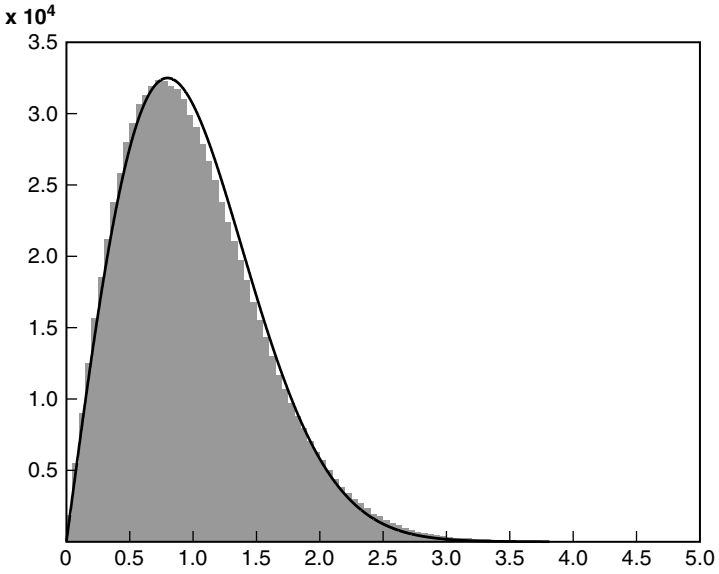

Figure 1.6 The local spacings of the central three-fifths of the eigenvalues of 5000 matrices (300×300) whose entries are drawn from the Cauchy distribution

We 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 1.7 and 1.8):

1.5 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 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 1.5.1. *By appropriately normalizing the even squares, show we obtain a new sequence with a similar distribution of spacings between adjacent elements as*

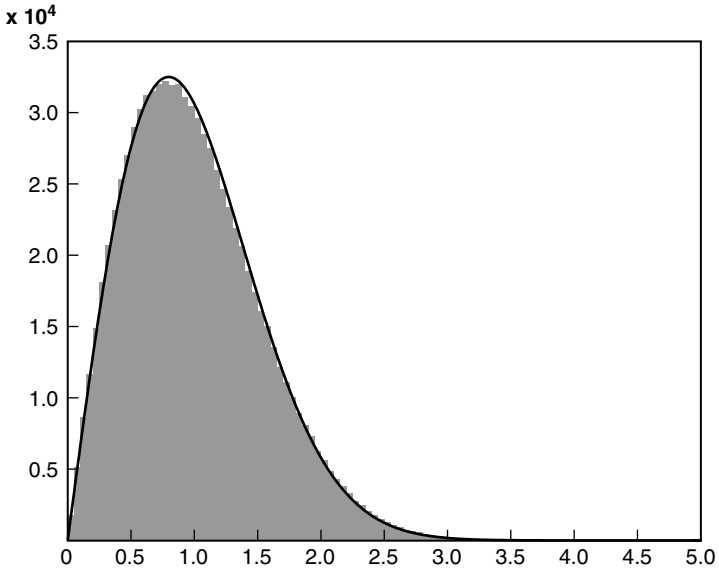

Figure 1.7 The local spacings of the central three-fifths of the eigenvalues of 5000 matrices (300×300) whose entries are drawn from the Poisson distribution ($\lambda = 5$)

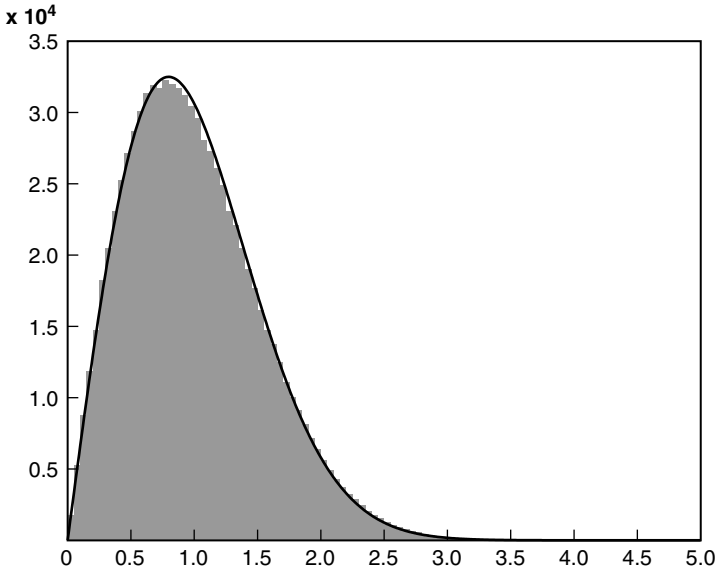

Figure 1.8 The local spacings of the central three-fifths of the eigenvalues of 5000 matrices (300×300) whose entries are drawn from the Poisson distribution ($\lambda = 20$)

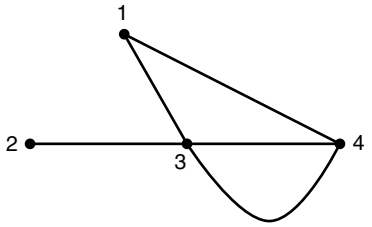

Figure 1.9 A typical graph

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 1.5.2. 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 **??**). 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 **??**.

1.5.1 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 1.9: 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 1.5.3). Let a_{ij} be the number of edges from vertex i to vertex j. For the graph above, we have

$$A = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 2 \\ 1 & 0 & 2 & 0 \end{pmatrix}. \tag{1.45}$$

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 §1.2.3) 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 2.2.13 and §**??**).

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 1.5.3. *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 1.5.4. Fundamental quantities should not depend on presentation. Exercise 1.5.3 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 1.5.5. *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 1.5.6 (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 1.5.7. Exercise 1.5.6 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 1.5.8. *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.

1.5.2 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 1.5.9 (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*

$$f(x) \;=\; \begin{cases} \frac{d}{2\pi(d^2-x^2)}\sqrt{4(d-1)-x^2}, & |x| \le 2\sqrt{d-1} \\ 0 & \text{otherwise.} \end{cases} \tag{1.46}$$

Exercise 1.5.10. *Show that as $d \to \infty$, by changing the scale of x, Kesten's measure converges to the semi-circle distribution.*

Below (Figures 1.10 and 1.11) 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 1.5.11. *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 1.12 shows that, surprisingly, the result *does* appear to be the same as that from all real symmetric matrices. See [JMRR] for more details.

1.6 NUMBER THEORY

We assume the reader is familiar with the material and notation from Chapter **??**. For us an L-**function** is given by a **Dirichlet series** (which converges if $\Re s$ is suffi-

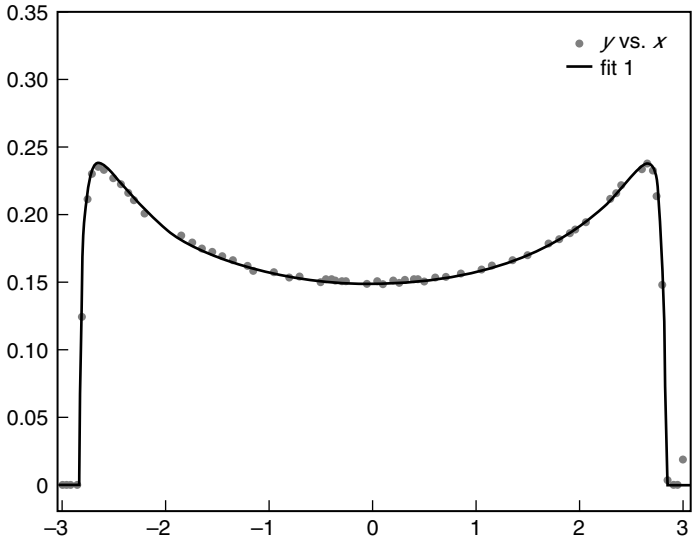

Figure 1.10 Comparison between theory (solid line) and experiment (dots) for 1000 eigenvalues of 3-regular graphs (120 bins in the histogram)

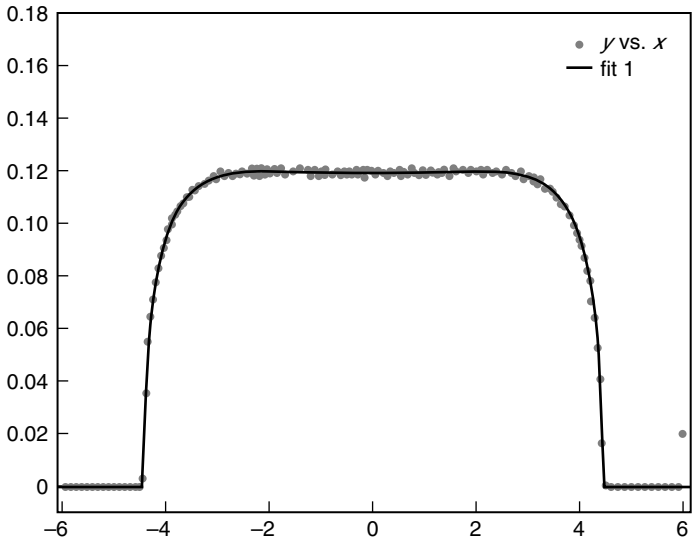

Figure 1.11 Comparison between theory (solid line) and experiment (dots) for 1000 eigenvalues of 6-regular graphs (240 bins in the histogram)

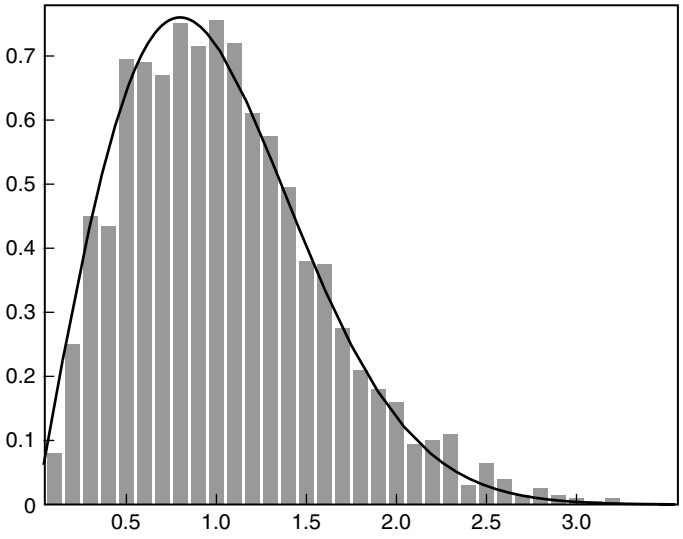

Figure 1.12 3-regular, 2000 vertices (graph courtesy of [JMRR])

ciently large), has an **Euler product**, and the coefficients have arithmetic meaning:

$$L(s,f) \;=\; \sum_{n=1}^{\infty} \frac{a_n(f)}{n^s} \;=\; \prod_p L_p(p^{-s},f)^{-1}, \quad \Re s > s_0. \tag{1.47}$$

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 **??** 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 §**??** 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 1.6.1. 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 **??**. 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 §1.6.1 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.

1.6.1 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

$$\lim_{N\to\infty} \frac{\#\left\{\left(\alpha_{j_1} - \alpha_{j_2}, \dots, \alpha_{j_{n-1}} - \alpha_{j_n}\right) \in B, j_i \le N; j_i \ne j_k\right\}}{N}. \quad (1.48)$$

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

$$I_{a_i,b_i}(x) = \begin{cases} 1 & \text{if } x \in [a_i, b_i], \\ 0 & \text{otherwise.} \end{cases} \tag{1.49}$$

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 1.6.2. *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 §**??**) says that if $E(\mathbb{Q})$ (the group of rational solutions for an elliptic curve E; see §**??**) 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 1.6.3 (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.

1.6.2 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

$$D_f(h) = \sum_j h\left(\gamma_f \frac{\log c_f}{2\pi}\right). \tag{1.50}$$

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 (1.50) 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**:

$$D_{\mathcal{F}}(h) = \frac{1}{|\mathcal{F}|} \sum_{f \in \mathcal{F}} D_f(h). \tag{1.51}$$

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 **??** and give an application as well.

Remark 1.6.4. 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$^{(h)}$ 1.6.5. *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 1.6.6 (Advanced)**.** In practice, one takes h in (1.50) to be a Schwartz function whose Fourier transform has finite support (see §**??**). 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 **??**. Then for suitable test functions h, we prove

$$\begin{aligned}\lim_{m\to\infty} D_{\mathcal{F}_m}(h) &= \lim_{m\to\infty} \frac{1}{|\mathcal{F}_m|} \sum_{\chi\in\mathcal{F}_m} \sum_{\gamma_\chi} h\left(\gamma_\chi \frac{\log c_\chi}{2\pi}\right) \\ &= \int_{-\infty}^{\infty} h(x) W_{G(\mathcal{F})}(x)dx. \end{aligned} \tag{1.52}$$

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 1.6.8).

Theorem 1.6.7. *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*

$$\sum_p \frac{\log p}{p\log N} F\left(\frac{\log p}{\log N}\right) = \int_0^\infty F(x)dx + O\left(\frac{1}{\log N}\right). \tag{1.53}$$

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

$$\sum_{p\le x} \log p = x + O(x^{\frac{1}{2}}\log^2(x)). \tag{1.54}$$

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

$$\sum_{p\le x} \log p \cdot \frac{1}{p}. \tag{1.55}$$

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

$$\sum_{p\le x} \frac{\log p}{p} \;=\; O(1) - \int_2^x (u + O(u^{\frac{1}{2}}\log^2 u))\frac{-1}{u^2}du \;=\; \log x + O(1). \quad (1.56)$$

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

$$\begin{aligned}
\sum_{p\ge 2} \frac{\log p}{p} F\left(\frac{\log p}{\log N}\right) \;&=\; -\int_2^\infty (\log x + O(1))\,\frac{d}{dx}F\left(\frac{\log x}{\log N}\right)dx \\
&=\; \int_2^\infty \left[\frac{1}{x}F\left(\frac{\log x}{\log N}\right) + O\left(\frac{1}{x\log N}\left|F'\left(\frac{\log x}{\log N}\right)\right|\right)\right]dx \\
&=\; \log N\int_{u_0}^\infty \left[F(u) + O\left(\frac{1}{\log N}\,|F'(u)|\right)\right]du \\
&=\; \log N\int_0^\infty \left[F(u) + O\left(\frac{|F'(u)|}{\log N}\right)\right]du + O(u_0\log N \max_{t\in[0,u_0]} F(t)) \\
&=\; \log N\int_0^\infty F(u)du + O\left(\int_0^\infty |F'(u)|du\right) + O\left(u_0\log N \max_{t\in[0,u_0]} F(t)\right) \\
&=\; \log N\int_0^\infty F(u)du + O(1), \qquad (1.57)
\end{aligned}$$

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 1.6.8. *Redo the above arguments using the bounds from §**??**, 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 §**??** for more on distribution of points.

Exercise 1.6.9. *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 1.6.10. *Why is the natural scale for Theorem 1.6.7 $\log N$ (i.e., why is it natural to evaluate the test function at $\frac{\log p}{\log N}$ and not p)?*

Exercise 1.6.11. *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 §**??**. Modify the statement and proof of Theorem 1.6.7 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 **??** *for the conjectured value of T_2.*

1.7 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 §**??**), we pass from sums over zeros in (1.50) 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 **??**). In §**??** 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 **??**, 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.

1.8 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 [Ha2] 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 2 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 **??** 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 **??** 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, McW, 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.

Chapter Two

Random Matrix Theory: Eigenvalue Densities

In this chapter we study the eigenvalue densities for many collections of random matrices. We concentrate on the density of normalized eigenvalues, though we mention a few questions regarding the spacings between normalized eigenvalues (which we investigate further in Chapter **??**). We use the notation of Chapter 1.

2.1 SEMI-CIRCLE LAW

Consider an ensemble of $N \times N$ real symmetric matrices, where for simplicity we choose the entries independently from some fixed probability distribution p. One very important question we can ask is: given an interval $[a, b]$, how many eigenvalues do we expect to lie in this interval? We must be careful, however, in phrasing such questions. We have seen in §1.2.2 that the average size of the eigenvalues grows like $\sqrt{N}$. Hence it is natural to look at the density of normalized eigenvalues.

For example, the Prime Number Theorem states that the number of primes $p \le x$ is $\frac{x}{\log x}$ plus lower order terms; see Theorem **??** for an exact statement. Thus the average spacing between primes $p \le x$ is $\frac{x}{x/\log x} = \log x$. Consider two intervals $[10^5, 10^5 + 1000]$ and $[10^{200}, 10^{200} + 1000]$. The average spacing between primes in the first is about 11.5; the average spacing between primes in the second is about 460.5. We expect to find about 87 primes in the first interval, and about 2 in the second. In order to obtain a universal answer, we instead look at the density of normalized primes.

The appropriate question to ask is not what is the density of eigenvalues or primes in an interval $[a, b]$, but rather in an interval $[a \cdot (\text{Ave Spacing}), b \cdot (\text{Ave Spacing})]$.

Exercise 2.1.1. *As $x \to \infty$ how many numbers at most x are square-free (m is **square-free** if $n^2|m$ implies $n = \pm 1$)? What is the average spacing between square-free numbers?*

2.1.1 Moments of the Semi-Circle Density

Consider

$$P(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2} & \text{if } |x| \le 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

Exercise 2.1.2. *Show that $P(x)$ is a probability density (show that it is non-negative and integrates to 1). Graph $P(x)$.*

We call $P(x)$ the semi-circle density. We calculate the moments of the semi-circle. We prove that for $k \leq 3$, the k^{th} moment of the semi-circle $C(k)$ equals the expected k^{th} moment of $\mu_{A,N}(x)$ as $N \to \infty$, and sketch the proof for higher moments; see §1.2.3 for the definition of $\mu_{A,N}(x)$. We have

$$C(k) \;=\; \int_{-\infty}^{\infty} x^k P(x)dx \;=\; \frac{2}{\pi}\int_{-1}^{1} x^k\sqrt{1-x^2}dx. \tag{2.2}$$

We note that, by symmetry, $C(k) = 0$ for k odd, and $C(0) = 1$ as $P(x)$ is a probability density. For $k = 2m$ even, we change variables. Letting $x = \sin\theta$,

$$C(2m) \;=\; \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2m}(\theta)\cdot\cos^2(\theta)d\theta. \tag{2.3}$$

Using $\sin^2(\theta) = 1 - \cos^2(\theta)$ gives

$$C(2m) \;=\; \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2m}(\theta)d\theta \;-\; \frac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2m+2}(\theta)d\theta. \tag{2.4}$$

The above integrals can be evaluated exactly. We repeatedly use

$$\begin{aligned}\cos^2(\phi) \;&=\; \frac{1}{2} + \frac{1}{2}\cos(2\phi)\\ \sin^2(\phi) \;&=\; \frac{1}{2} - \frac{1}{2}\cos(2\phi).\end{aligned} \tag{2.5}$$

Repeated applications of the above allow us to write $\sin^{2m}(\theta)$ as a linear combination of $1, \cos(2\theta), \ldots, \cos(2m\theta)$. Let

$$n!! \;=\; \begin{cases} n\cdot(n-2)\cdots 2 & \text{if } n \text{ is even}\\ n\cdot(n-2)\cdots 1 & \text{if } n \text{ is odd.}\end{cases} \tag{2.6}$$

We find (either prove directly or by induction) that

$$\frac{2}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2m}(\theta)d\theta \;=\; 2\frac{(2m-1)!!}{(2m)!!}. \tag{2.7}$$

Exercise 2.1.3. *Calculate $C(2)$ and $C(4)$ and show that in general*

$$C(2m) \;=\; 2\frac{(2m-1)!!}{(2m+2)!!}. \tag{2.8}$$

To each $N \times N$ real symmetric matrix A we associate a probability distribution $\mu_{A,N}(x)$ (see §1.2.3). We now sketch the proof that as $N \to \infty$ most of the $\mu_{A,N}(x)$ are close to $P(x)$, the semi-circle density.

2.1.2 Moment Preliminaries

Definition 2.1.4. *$M_{N,k}(A)$ is the k^{th} moment of the probability measure attached to $\mu_{A,N}(x)$:*

$$M_{N,k}(A) \;=\; \int x^k \mu_{A,N}(x)dx \;=\; \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\lambda_i(A)}{2\sqrt{N}}\right)^k. \tag{2.9}$$

As always, the starting point for our investigations is Theorem 1.2.4, which says $\sum \lambda_i(A)^k = \mathrm{Trace}(A^k)$. By Lemma 1.2.10,

$$M_{N,k}(A) \;=\; \frac{1}{2^k N^{1+\frac{k}{2}}}\;\mathrm{Trace}(A^k). \tag{2.10}$$

We show that as $N \to \infty$ the expected value of the moments $M_{N,k}(A)$ of the $\mu_{A,N}(x)$ converge to the moments of the semi-circle. This does not prove Wigner's Semi-Circle Law; we need some results from Probability Theory to complete the proof (see §1.3.2 for an explanation of the needed technical arguments, and [CB] for a rigorous derivation of the Semi-Circle Law).

See §1.3.3 for a review of notation. Let $M_{N,k} = \mathbb{E}[M_{N,k}(A)]$ be the average over all A (appropriately weighted by the probability density) of $M_{N,k}(A)$. Explicitly, the probability density of observing a matrix A with entries a_{ij} is $P(A)dA = \prod_{1\le i\le j\le N} p(a_{ij})da_{ij}$, and averaging over all matrices gives the expected value of $M_{N,k}(A)$ is

$$M_{N,k} \;=\; \int_{a_{11}=-\infty}^{\infty}\cdots\int_{a_{NN}=-\infty}^{\infty} M_{N,k}(A)\prod_{1\le i\le j\le N} p(a_{ij})da_{ij}. \tag{2.11}$$

From Theorem 1.2.1

$$\mathrm{Trace}(A^k) \;=\; \sum_{1\le i_1,\dots,i_k\le N} a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}. \tag{2.12}$$

This and (2.10) yield

$$M_{N,k} \;=\; \frac{1}{2^k N^{1+\frac{k}{2}}}\sum_{1\le i_1,\dots,i_k\le N}\mathbb{E}[a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}], \tag{2.13}$$

where

$$\begin{aligned}&\mathbb{E}[a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}]\\ &\;=\; \int_{a_{11}=-\infty}^{\infty}\cdots\int_{a_{NN}=-\infty}^{\infty} a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}\prod_{1\le i\le j\le N} p(a_{ij})da_{ij}.\end{aligned} \tag{2.14}$$

There are N^k terms in $\mathrm{Trace}(A^k)$, each term is a product of k factors a_{ij}. We use the notation $\mathbb{E}[a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}]$ as we integrate each term $a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}$ against $P(A)dA$, and this gives the expected value of the term.

We can write (2.14) in a more useful manner. While the above formula is correct, by grouping terms together we can rewrite it in such a way that it will be easier to evaluate. For small k, by brute force one can determine these integrals; however, as k increases, the computation becomes significantly harder and good combinatorics are needed, and the reformulation emphasizes the important parts of the calculation. Further, writing out the integrals each time leads to long formulas; by using expectation we have more concise formulas (though of course they convey the same information).

In the product $a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}$, group a_{ij} together that have the same subscripts; as the matrices are symmetric, $a_{ij} = a_{ji}$ and we consider the pairs (i,j) and (j,i) equal. Say we can write

$$a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1} \;=\; a_{x_1y_1}^{r_1}\cdots a_{x_\ell y_\ell}^{r_\ell}, \tag{2.15}$$

where all pairs (x_j, y_j) are distinct (remember, we consider the pairs (x, y) and (y, x) equal). For example,

$$a_{13}a_{34}a_{45}a_{53}a_{31}a_{14}a_{43}a_{31} \;=\; a_{13}^3 a_{34}^2 a_{14}a_{45}a_{35}.$$

As $a_{ij} = a_{ji}$, we have chosen to write the lower subscript first, especially as $P(A)dA = \prod_{1\le i\le j\le N} p(a_{ij})da_{ij}$ has $i \le j$. We then obtain

$$\begin{aligned}&\mathbb{E}[a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}]\\ &\quad= \int_{a_{11}=-\infty}^{\infty}\cdots\int_{a_{NN}=-\infty}^{\infty} a_{x_1y_1}^{r_1}\cdots a_{x_\ell y_\ell}^{r_\ell}\prod_{1\le i\le j\le N} p(a_{ij})da_{ij}. \qquad (2.16)\end{aligned}$$

As all entries are *independently* drawn from the *same* distribution, this integral greatly simplifies. Let p_k be the k^{th} moment of p:

$$p_k \;=\; \int_{a=-\infty}^{\infty} a^k p(a)da. \qquad (2.17)$$

Then (2.16) becomes

Lemma 2.1.5. *Let $a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1} = a_{x_1y_1}^{r_1}\cdots a_{x_\ell y_\ell}^{r_\ell}$, where all pairs (x_j, y_j) are distinct, remembering that we consider (i, j) the same as (j, i). Then*

$$\mathbb{E}[a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}] \;=\; p_{r_1}\cdots p_{r_\ell}. \qquad (2.18)$$

Note we could also write

$$\mathbb{E}[a_{i_1i_2}a_{i_2i_3}\cdots a_{i_ki_1}] \;=\; \mathbb{E}[a_{x_1y_1}^{r_1}]\cdots\mathbb{E}[a_{x_\ell y_\ell}^{r_\ell}] \;=\; p_{r_1}\cdots p_{r_\ell}. \qquad (2.19)$$

As we assume p has mean 0, variance 1 and finite higher moments, if any $r_j = 1$ then the above product vanishes. If each $r_j = 2$ the above product is 1.

Instead of proving Lemma 2.1.5, we handle the contribution from one of the N^k terms; the general proof proceeds similarly. Let us calculate the contribution from the term in (2.16), assuming $N > 5$. Let

$$\begin{aligned}\mathcal{S} \;&=\; \{(1,3),(3,4),(1,4),(4,5),(3,5)\}\\ \mathcal{T} \;&=\; \{(i,j) : 1\le i\le j\le N, (i,j)\notin\mathcal{S}\}. \qquad (2.20)\end{aligned}$$

For each $(i, j) \in \mathcal{S}$, let $r(i, j)$ be the exponent of $a_{i,j}$ in (2.16):

$$r(1,3) = 3,\;\; r(3,4) = 2,\;\; r(1,4) = r(4,5) = r(3,5) = 1. \qquad (2.21)$$

We have $\frac{N(N+1)}{2}$ integrals over a_{ij}, with $1 \le i \le j \le N$. Thus the integral in (2.16) from the term in (2.16) becomes

$$\prod_{(i,j)\in\mathcal{S}}\int_{a_{ij}=-\infty}^{\infty} a_{ij}^{r(i,j)}p(a_{ij})da_{ij}\prod_{(i,j)\in\mathcal{T}}\int_{a_{ij}=-\infty}^{\infty} p(a_{ij})da_{ij}. \qquad (2.22)$$

Each integral over an $(i, j) \in \mathcal{T}$ gives 1, and the integrals over $(i, j) \in \mathcal{S}$ give $p_{r(i,j)}$. Explicitly,

$$\begin{aligned}&\prod_{(i,j)\in\mathcal{S}}\int_{a_{ij}=-\infty}^{\infty} a_{ij}^{r(i,j)}p(a_{ij})da_{ij}\\ &\quad= \int_{a_{13}=-\infty}^{\infty} a_{13}^3 p(a_{13})da_{13}\int_{a_{34}=-\infty}^{\infty} a_{34}^2 p(a_{34})da_{34}\int_{a_{14}=-\infty}^{\infty} a_{14}p(a_{14})da_{14}\\ &\qquad\cdot\int_{a_{45}=-\infty}^{\infty} a_{45}p(a_{45})da_{45}\int_{a_{35}=-\infty}^{\infty} a_{35}p(a_{35})da_{35}\\ &\quad= p_3p_2p_1p_1p_1. \qquad (2.23)\end{aligned}$$

Therefore, the contribution from the term in (2.16) is $p_3 p_2 p_1^3 \cdot 1^{\frac{N(N+1)}{2}-5}$; the exponent of 1 is $|\mathcal{T}| = \frac{N(N+1)}{2} - 5$. This is zero as p has mean zero, implying $p_1 = 0$.

Exercise 2.1.6. *Prove* (2.12), (2.13) *and Lemma 2.1.5.*

2.1.3 The First Few Moments

We use the expansions from §2.1.2 to calculate the first few moments. See §1.3 for a review of the formulation of Wigner's Semi-Circle Law. We must show that $\lim_{N\to\infty} M_{N,k} = C(k)$, where $C(k)$ is the k^{th} moment of the semi-circle distribution.

Lemma 2.1.7. *The expected value of* $M_{N,0}(A) = 1$, *thus* $\lim_{N\to\infty} M_{N,0} = C(0)$.

Proof. We have

$$M_{N,0} \;=\; \mathbb{E}\left[M_{N,0}(A)\right] \;=\; \frac{1}{N}\mathbb{E}\left[\text{Trace}(I)\right] \;=\; \frac{1}{N}\mathbb{E}[N] \;=\; \frac{N}{N}\mathbb{E}[1] \;=\; 1. \tag{2.24}$$

□

Lemma 2.1.8. *The expected value of* $M_{N,1}(A) = 0$, *thus* $\lim_{N\to\infty} M_{N,1} = C(1)$.

Proof. We have

$$\begin{aligned} M_{N,1} \;=\; \mathbb{E}\left[M_{N,1}(A)\right] \;&=\; \frac{1}{2N^{3/2}}\mathbb{E}\left[\text{Trace}(A)\right] \\ &=\; \frac{1}{2N^{3/2}}\mathbb{E}\left[\sum_{i=1}^{N} a_{ii}\right] \\ &=\; \frac{1}{2N^{3/2}}\sum_{i=1}^{N}\mathbb{E}[a_{ii}]. \end{aligned} \tag{2.25}$$

As each a_{ij} is drawn from a probability distribution with mean zero, each $\mathbb{E}[a_{ii}] = 0$. □

Lemma 2.1.9. *The expected value of* $M_{N,2}(A) = \frac{1}{4}$, *thus* $\lim_{N\to\infty} M_{N,2} = C(2)$.

Proof. By Theorem 1.2.1

$$\text{Trace}(A^2) \;=\; \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}a_{ji}. \tag{2.26}$$

As A is symmetric, $a_{ij} = a_{ji}$. Thus, the trace is $\sum_i \sum_j a_{ij}^2$. Now

$$\begin{aligned} M_{N,2} \;=\; \mathbb{E}\left[M_{N,2}(A)\right] \;&=\; \frac{1}{4N^2}\mathbb{E}\left[\text{Trace}(A^2)\right] \\ &=\; \frac{1}{4N^2}\mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2\right] \\ &=\; \frac{1}{4N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}[a_{ij}^2]. \end{aligned} \tag{2.27}$$

Each $\mathbb{E}[a_{ij}^2] = 1$ because we have assumed p has mean 0 and variance 1 (which implies the second moment of p is 1). There are N^2 pairs (i,j). Thus, we have $\frac{1}{4N^2} \cdot (N^2 \cdot 1) = \frac{1}{4}$. □

Lemma 2.1.10. $\lim_{N\to\infty} M_{N,3} = C(3) = 0$.

Proof. By Theorem 1.2.1

$$\text{Trace}(A^3) \;=\; \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} a_{ij}a_{jk}a_{ki}. \tag{2.28}$$

Therefore

$$\begin{aligned} M_{N,3} \;=\; \mathbb{E}\left[M_{N,3}(A)\right] \;&=\; \frac{1}{8N^{2.5}}\mathbb{E}\left[\text{Trace}(A^3)\right] \\ &=\; \frac{1}{8N^{2.5}}\mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} a_{ij}a_{jk}a_{ki}\right] \\ &=\; \frac{1}{8N^{2.5}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\mathbb{E}[a_{ij}a_{jk}a_{ki}]. \end{aligned} \tag{2.29}$$

There are three cases. If the subscripts i, j and k are all distinct, then a_{ij}, a_{jk}, and a_{ki} are three independent variables (in other words, these are three distinct pairs). As p has mean zero, by Lemma 2.1.5

$$\mathbb{E}[a_{ij}a_{jk}a_{ki}] \;=\; \mathbb{E}[a_{ij}]\cdot\mathbb{E}[a_{jk}]\cdot\mathbb{E}[a_{ki}] \;=\; 0. \tag{2.30}$$

If two of the subscripts are the same (say $i = j$) and the third is distinct, by Lemma 2.1.5

$$\mathbb{E}[a_{ii}a_{ik}a_{ki}] \;=\; \mathbb{E}[a_{ii}]\cdot\mathbb{E}[a_{ik}^2] \;=\; 0\cdot 1 \;=\; 0 \tag{2.31}$$

because p has mean zero and variance 1. If all three subscripts are the same, we have

$$\mathbb{E}[a_{ii}^3]. \tag{2.32}$$

This is the third moment of p. It is the same for all pairs (i,i), equal to p_3 by Lemma 2.1.5. This is where we use the assumption that the higher moments of p are finite. There are N triples where $i = j = k$. Thus,

$$M_{N,3} \;=\; \mathbb{E}\left[M_{N,3}(A)\right] \;=\; \frac{1}{8N^{2.5}}\cdot Np_3 \;=\; \frac{p_3}{8N^{1.5}}. \tag{2.33}$$

Letting $N \to \infty$ we see that the expected value of the third moment is zero in the limit. □

Remark 2.1.11. Many of the above calculations are unnecessary. We are dividing by $N^{2.5}$. There are N^3 triples $a_{ij}a_{jk}a_{ki}$. If i, j and k are distinct, we showed by Lemma 2.1.5 the contribution is zero. If the indices are *not* distinct, there are at most $3N^2$ such triples, and as all moments of p are finite, by Lemma 2.1.5 each such triple contributes a bounded amount (independent of N). As we divide by $N^{2.5}$, the total contribution is at most some universal constant times $\frac{1}{\sqrt{N}}$, which tends to zero as $N \to \infty$. This illustrates a general principle: often order of magnitude calculations are sufficient to show certain terms do not contribute in the limit.

2.1.4 The Higher Moments

Lemma 2.1.12. *For odd k, the expected value of $M_{N,k}(A)$ as $N \to \infty$ is zero.*

Exercise[(hr)] 2.1.13. *Prove Lemma 2.1.12.*

We are left with calculating the limit of the averages of $M_{N,k}(A)$ for $k = 2m$ even.

Lemma 2.1.14. *Notation as before, the only tuples which contribute as $N \to \infty$ to the main term of the average of $M_{N,2m}(A)$ are those where each $r_j = 2$.*

Exercise[(hr)] 2.1.15. *Prove Lemma 2.1.14.*

We are reduced to calculating the contributions to the average of $M_{N,2m}(A)$ from tuples with each $r_j = 2$. By Lemma 2.1.5, a tuple

$$a_{i_1 i_2} \cdots a_{i_{2m} i_1} \;=\; a^2_{x_1 y_1} \cdots a^2_{x_m y_m} \tag{2.34}$$

contributes 1^m (because we have a product of m second moments of p, and the second moment of p is 1). The above arguments and (2.13) yield, up to lower order terms,

$$M_{N,2m} \;=\; \mathbb{E}[M_{N,2m}(A)] \;=\; \frac{1}{2^m N^{1+m}} \sideset{}{^*}\sum_{1 \le 1_i, \ldots, i_{2m} \le N} 1, \tag{2.35}$$

where $\sum^*$ means we restrict to tuples $(i_1, \ldots, i_{2m})$ such that the corresponding r_j's are all 2. The determination of the limits of the even moments is completed by showing

$$\frac{1}{2^m N^{1+m}} \sideset{}{^*}\sum_{1 \le 1_i, \ldots, i_{2m} \le N} 1 \;=\; C(2m) \;=\; 2\frac{(2m-1)!!}{(2m+2)!!}. \tag{2.36}$$

The solution of this counting problem involves the Catalan number (see [CG]) $c_k = \frac{1}{k+1}\binom{2k}{k}$. See [Leh] for details on these calculations.

Exercise 2.1.16. *Compute the fourth and sixth moments directly, and compare them to $C(4)$ and $C(6)$.*

Exercise 2.1.17. *For each m show there exists a constant $c_m > 0$ such that*

$$\sideset{}{^*}\sum_{1 \le 1_i, \ldots, i_{2m} \le N} 1 \;\ge\; c_m N^{1+m}. \tag{2.37}$$

This implies that the even moments do not vanish.

Exercise 2.1.18. *For each m show there exists a constant d_m such that*

$$\sideset{}{^*}\sum_{1 \le 1_i, \ldots, i_{2m} \le N} 1 \;\le\; d_m N^{1+M}. \tag{2.38}$$

This implies that the even moments are finite.

Exercise[(h)] 2.1.19. *Prove* (2.36)*.*

2.2 NON-SEMI-CIRCLE BEHAVIOR

In our investigations of random d-regular graphs, we showed the density of normalized eigenvalues do not converge to the semi-circle (Theorem 1.5.9). We give several more examples of ensembles of matrices where the density of eigenvalues is provably *not* given by the Semi-Circle Law. The d-regular graphs are combinatorial objects, and we are not constructing our matrices by choosing entries at random from a fixed probability distribution p. Now we give several examples where we do choose the entries randomly, but with additional structure (otherwise we would of course just have the ensemble of all real symmetric matrices). A generic real symmetric matrix has $\frac{N(N+1)}{2}$ independent entries. We now consider subsets with far fewer independent entries, often of size N. The hope is that these thin ensembles might exhibit new, interesting behavior.

2.2.1 Band Matrices

Definition 2.2.1 (Band Matrix (of width r))**.** *We say a real symmetric matrix is a band matrix (of width r) if $a_{ij} = 0$ whenever $|i - j| > r$.*

A band matrix of width 0 is a diagonal matrix and of width 1 has non-zero entries only along the main diagonal and the diagonals immediately above and below. In general the number of independent entries is of size $(2r + 1)N$.

Exercise 2.2.2. *Calculate exactly how many entries can be non-zero if the band width is r.*

While band matrices are a subset of real symmetric matrices, they are a very thin subset for $r \ll N$. Do they obey the Semi-Circle Law? Is the spacing between adjacent eigenvalues the GOE?

If the band width $r = N - 1$, then the matrix is "full"; in other words, every entry can be non-zero and the density of normalized eigenvalues converges to the semi-circle. What about the opposite extreme, when $r = 0$? Consider $N \times N$ real symmetric band matrices of width 0, each entry which can be non-zero is chosen randomly and independently from some fixed probability distribution p. For $r = 0$, we do not need to assume anything about the moments of p.

Theorem 2.2.3. *The normalized eigenvalue density is* not *the semi-circle; it is just p.*

Proof. There is no need to normalize the eigenvalues. As we have a diagonal matrix, the entries *are* the eigenvalues! Asking how many eigenvalues are in $[a, b]$ is equivalent to calculating the probability that an $a_{ii} \in [a, b]$, which is given by $\int_a^b p(x)dx$. □

Exercise 2.2.4. *Let A be an $N \times N$ band matrix of width 1 with real entries, but not necessarily symmetric. Which entries can be non-zero in $A^T A$?*

2.2.2 Toeplitz Matrices

We consider another ensemble of random matrices with far fewer independent entries than the ensemble of all real symmetric matrices.

Definition 2.2.5. *A Toeplitz matrix A is of the form*

$$A = \begin{pmatrix} b_0 & b_1 & b_2 & b_3 & \cdots \\ b_{-1} & b_0 & b_1 & b_2 & \cdots \\ b_{-2} & b_{-1} & b_0 & b_1 & \cdots \\ b_{-3} & b_{-2} & b_{-1} & b_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{2.39}$$

That is, A is constant along its diagonals. Note $a_{ij} = b_{j-i}$.

We consider real symmetric Toeplitz matrices whose entries are chosen according to some distribution p with mean 0 and variance 1. Thus $b_{i-j} = b_{j-i}$. It is convenient to normalize the eigenvalues of these Toeplitz matrices by $\frac{1}{\sqrt{N}}$ rather than $\frac{1}{2\sqrt{N}}$. Thus

$$\mu_{A,N}(x) = \frac{1}{N}\sum_{i=1}^{N} \delta\left(x - \frac{\lambda_i(A)}{\sqrt{N}}\right). \tag{2.40}$$

Remark 2.2.6. As the main diagonal is constant, the effect of the main diagonal being b_0 is simply to shift all eigenvalues by b_0. For simplicity, we take $b_0 = 0$. Note there are $N-1$ independent entries $b_1, \dots, b_{N-1}$.

Exercise 2.2.7. *If $B = A + mI$, prove the eigenvalues of B are m plus the eigenvalues of A.*

The eigenvalue distribution is again not the semi-circle. As long as the probability distribution p has mean 0, variance 1 and finite higher moments, the answer is universal (i.e., independent of all other properties of p). It is *almost* the standard Gaussian. Its moments are bounded by the moments of the standard Gaussian. Its fourth moment is $2\frac{2}{3}$, while the standard Gaussian's is 3.

Exercise 2.2.8. *Show $M_{N,1} = 0$ and $M_{N,2} = 1 - \frac{1}{N}$. Thus as $N \to \infty$ the expected value of the first two moments are 0 and 1, respectively. Recall the second moment of the semi-circle is $\frac{1}{4}$.*

Just because $\lim_{N\to\infty} M_{N,2} \neq \frac{1}{4}$ does not imply that the eigenvalue probability distribution does not converge to a semi-circle; it only implies it does not converge to the *standard* semi-circle — we need to examine the fourth moment. See Exercise 1.1.6.

It turns out that it is not the semi-circle that this distribution is trying to approach, but rather the Gaussian. The odd moments of the Gaussian vanish, and the even moments are $G(2m) = (2m-1)!!$. The limits of the average of the moments want to be $G(2m)$; however, to calculate these moments involves solving a system of Diophantine equations (see Chapter **??**). Obstructions to these equations arise due

to the fact that the indices must be in $\{1, \dots, N\}$, and this prevents the limits from equalling the Gaussian's moments.

The fourth moment calculation highlights the Diophantine obstructions, which bound the moments away from the Gaussian. As $a_{ij} = b_{j-i} = b_{i-j}$, the trace expansion becomes

$$M_{N,4}(A) \;=\; \frac{1}{N^3} \sum_{1 \le i_1 i_2 i_3, i_4 \le N} \mathbb{E}(b_{i_1-i_2} b_{i_2-i_3} b_{i_3-i_4} b_{i_4-i_1}). \tag{2.41}$$

Let $x_j = |i_j - i_{j+1}|$. If any b_{x_j} occurs to the first power, its expected value is zero (since the mean of p is zero, and the b's are drawn from p), and these tuples do not contribute. Thus either the x_j's are matched in pairs (with different values), or all four are equal (in which case they are still matched in pairs). There are 3 possible matchings; however, by symmetry (simply relabel), we see the contribution from $x_1 = x_2$, $x_3 = x_4$ is the same as the contribution from $x_1 = x_4$, $x_2 = x_3$.

If $x_1 = x_2$, $x_3 = x_4$, we have

$$i_1 - i_2 \;=\; \pm(i_2 - i_3) \quad \text{and} \quad i_3 - i_4 \;=\; \pm(i_4 - i_1). \tag{2.42}$$

Exercise 2.2.9. *Show the number of tuples (i_1, i_2, i_3, i_4) satisfying the pair of equations in* (2.42) *is $O(N^2)$ if a $+$ sign holds in either equation. As we divide by N^3, in the limit these terms do not contribute and the main contribution arises when both equations have the minus sign.*

If both signs are negative in (2.42), then $i_1 = i_3$ and i_2 and i_4 are arbitrary. We see there are N^3 such tuples. Almost all of these will have $x_1 \neq x_3$, and contribute 1; the rest will contribute a smaller term. Explicitly, let p_4 denote the fourth moment of p. Given i_1 and i_2, $N-1$ choices of i_4 yield $x_1 \neq x_3$ (contributing $\mathbb{E}[b_{x_1}^2 b_{x_3}^2] = 1$), and one choice yields the two equal (contributing $\mathbb{E}[b_{x_1}^4] = p_4$). Therefore this case contributes

$$\frac{1}{N^3}\left(N^2(N-1)\cdot 1 + N^2(1)\cdot p_4\right) \;=\; 1 - \frac{1}{N} + \frac{p_4}{N} \;=\; 1 + O\left(\frac{1}{N}\right). \tag{2.43}$$

The case of $x_1 = x_4$ and $x_2 = x_3$ is handled identically, and contributes $1 + O\left(\frac{1}{N}\right)$.

The other possibility is for $x_1 = x_3$ and $x_2 = x_4$. Non-adjacent pairing is what leads to Diophantine obstructions, which decreases the contribution to the moment. We call this a non-adjacent pairing as the neighbors of x_1 are x_2 and x_4, but x_1 is paired with x_3. Now we have

$$i_1 - i_2 \;=\; \pm(i_3 - i_4) \quad \text{and} \quad i_2 - i_3 \;=\; \pm(i_4 - i_1). \tag{2.44}$$

Exercise 2.2.10. *Show the number of tuples (i_1, i_2, i_3, i_4) satisfying the pair of equations in* (2.44) *is $O(N^2)$ if a $+$ sign holds in either equation. As we divide by N^3, in the limit these terms do not contribute and the main contribution arises when both equations have the minus sign.*

If both signs are negative in (2.44), then we have

$$i_1 \;=\; i_2 + i_4 - i_3, \quad i_1, i_2, i_3, i_4 \in \{1, \dots, N\}. \tag{2.45}$$

The fact that each $i_j \in \{1, \dots, N\}$ is what leads to the Diophantine obstructions. In the first case (when $x_1 = x_2$ and $x_3 = x_4$), we saw we had three independent variables and $N^3 + O(N^2)$ choices that were mutually consistent. Now it is possible for choices of i_2, i_3 and i_4 to lead to impossible values for i_1. For example, if $i_2, i_4 \geq \frac{2N}{3}$ and $i_3 < \frac{N}{3}$, this forces $i_1 > N$, which is not allowed, which implies there are at most $(1 - \frac{1}{27})N^3$ valid choices. This is enough to show the Gaussian's moment is strictly greater; we have lost a positive fraction of solutions. The following lemma shows this case contributes $\frac{2}{3}$ to the fourth moment.

Lemma 2.2.11. *Let $I_N = \{1, \dots, N\}$. Then $\#\{x, y, z \in I_N : 1 \leq x + y - z \leq N\} = \frac{2}{3}N^3 + \frac{1}{3}N$.*

Proof. Say $x + y = S \in \{2, \dots, 2N\}$. For $2 \leq S \leq N$, there are $S - 1$ choices of z, and for $S \geq N + 1$, there are $2N - S + 1$. Similarly, the number of $x, y \in I_N$ with $x + y = S$ is $S - 1$ if $S \leq N + 1$ and $2N - S + 1$ otherwise. The number of triples is therefore

$$\sum_{S=2}^{N}(S-1)^2 + \sum_{S=N+1}^{2N}(2N-S+1)^2 \;=\; \frac{2}{3}N^3 + \frac{1}{3}N. \tag{2.46}$$

□

Collecting all the pieces, we have shown

Theorem 2.2.12 (Fourth Moment)**.** $M_{N,4} = 2\frac{2}{3} + O\left(\frac{1}{N}\right)$.

In [BDJ, HM] the Toeplitz ensemble is investigated and shown to be non-Semi-Circular and non-Gaussian. See [HM] for upper and lower bounds for the moments of the new distribution that the densities $\mu_{A,N}(x)$ converge to.

Remark 2.2.13. Similar to our previous arguments, one can show that the odd moments vanish and the main contribution to the even moments occur when the b_x's are matched in pairs. For $2m$ objects there are $(2m-1)!!$ ways to match in pairs. Each matching wants to contribute 1 (and if they all did, then we would have the standard Gaussian's moments); however, not all matchings contribute 1. For some matchings, a positive fraction of tuples are inaccessible. Explicitly, for each matching we divide by N^{m+1}. It turns out that of the $2m$ indices $i_1, \dots, i_{2m}$, once $m + 1$ are specified the others are determined. If we could choose $m + 1$ indices freely, we would have N^{m+1} tuples for each matching, and a contribution of 1. It is here that the loss of a positive percent is felt. Interestingly, if we add additional symmetries, all the moments are Gaussian. Explicitly, assume the first row is a palindrome; for $N = 2M$ this means the first row is

$$(0\ b_1\ b_2\ b_3\ \dots\ b_{M-2}\ b_{M-1}\ b_{M-1}\ b_{M-2}\ \dots\ b_3\ b_2\ b_1\ 0). \tag{2.47}$$

Instead of $N - 1$ free variables, there are now just $\frac{N-2}{2}$. Similar to the density of states of d-regular graphs (§1.5.1), we have a sub-ensemble with different behavior. See [MMS] for a proof that the moments are Gaussian.

2.2.3 Truncated Cauchy Distribution

In §1.3.4 we saw that numerical simulations of eigenvalues of matrices with entries independently chosen from the Cauchy distribution appeared not to satisfy the Semi-Circle Law. For $N \times N$ matrices, instead of choosing the entries from the Cauchy distribution, choose the entries from a *truncated* Cauchy distribution, where the truncation depends on N. Explicitly, let

$$p_N(x) \ = \ \begin{cases} A_N \frac{1}{\pi(1+x^2)} & \text{if } |x| \leq f(N) \\ 0 & \text{otherwise,} \end{cases} \tag{2.48}$$

where A_N is chosen to make $\int_{\mathbb{R}} p_N(x)dx = 1$. By appropriately choosing the cutoff $f(N)$ and normalizing the eigenvalues, one obtains a new distribution. See [Za] for complete statements and proofs, as well as generalizations to other distributions.

2.3 SPARSE MATRICES

A common theme of some of the above problems (band matrices, random graphs) is that we are considering **sparse matrices**: real symmetric matrices where most entries are zero. Such matrices open up fascinating possibilities to see new behavior. In general, the following heuristic principle is a good guide: if you consider a very small subset of objects, you can see very special behavior. However, in mathematical proofs, we need to average over many similar objects. Thus, if we have too few objects, we cannot perform the averaging; if we have too many objects, non-standard behavior (which occurs rarely) could be washed away.

For example, as most matrices are not band symmetric of small width, even though they have different eigenvalue statistics, this difference will not be noticed when we look at all symmetric matrices. The goal, therefore, is to find an ensemble that is large enough so that we can do the averaging, yet small enough so that new interesting behavior is still present.

The **generalized coin toss matrices** provide another candidate. For $q_N \in [0, \frac{1}{2}]$, let $p_N(1) = \frac{q_N}{2}$, $p_N(-1) = \frac{q_N}{2}$, and $p_N(0) = 1 - q_N$. We use the probability function p_N to construct real symmetric matrices A by choosing the independent entries from p_N. We expect to have about $q_N \cdot \frac{N(N+1)}{2}$ non-zero entries in a typical A. If q_N is small relative to N, these matrices are sparse, and there is the possibility for new behavior. Note, of course, that if q_N is independent of N then the standard proof of the Semi-Circle Law is applicable. See [Liu] for more details.

2.4 RESEARCH PROJECTS

For more on connections between random graphs and number theory, see [DSV] and Chapter 3 of [Sar].

Research Project 2.4.1 (Band Matrices)**.** Investigate how the eigenvalue density depends on the band width. When do we observe the transition from p to the semi-circle? In other words, how large must r be in order to see semi-circle behavior.

Does this critical r depend on p? It has been observed for many systems that transitions occur around $r = \sqrt{N}$.

Research Project 2.4.2 (Band, Sparse, d-Regular)**.** Compare the eigenvalue distributions and spacing distributions (see Chapter **??**) of band matrices of width r, generalized coin toss matrices, and d-regular random graphs. If we choose r, q and d so that

$$\frac{(r+1)(2N-r)}{2} \sim \frac{qN(N+1)}{2} \sim \frac{dN}{2}, \tag{2.49}$$

are the distributions similar? All three ensembles have approximately the same number of non-zero entries, but they differ greatly in *where* the non-zero entries may lie.

Research Project 2.4.3. In the above project we considered sparse matrices with entries in $\{-1, 0, 1\}$. As the probability distribution depends on N, the arguments used to prove Wigner's Semi-Circle Law are not applicable. The adjacency matrix of a simple d-regular graph with no self-loops has $\frac{dN}{2}$ of the a_{ij} (with $1 \le i < j \le N$) equal to 1 (and the rest are zero). Let now

$$p_{N,d}(x) = \begin{cases} \frac{d}{N-1} & \text{if } x = 1 \\ 1 - \frac{d}{N-1} & \text{if } x = 0. \end{cases} \tag{2.50}$$

If we choose the entries a_{ij} (with $1 \le i < j \le N$) from $p_{N,d}$ and consider the graph of such a matrix, the expected number of edges from each vertex is d. Thus it is natural to see whether or not such an ensemble approximates d-regular graphs. How are the eigenvalues distributed? See also Remark 1.5.7.

Research Project 2.4.4 (Self-Adjoint Matrices)**.** Fix a probability distribution p and choose *all* the entries of A randomly and independently from p. Consider the matrix $A^T A$. This matrix is real symmetric, but has N^2 degrees of freedom. What is the density of its eigenvalues, or at least what are the first few moments? Are the eigenvalues real? Are they non-negative? What is the density of the *square root* of the eigenvalues? Matrices of this form, called Wishart matrices, are well studied by other techniques. See for example [SM, Wis].

Research Project 2.4.5 (Weighted Random Graphs)**.** Consider the space of d-regular graphs. To each graph we attach an adjacency matrix, and we can study the distribution of the eigenvalues. Consider the following generalization: fix a probability distribution p. Let A be the adjacency matrix of a d-regular graph G. Construct a matrix B as follows: if $a_{ij} = 1$, choose b_{ij} randomly from p; if $a_{ij} = 0$, set $b_{ij} = 0$. How does the distribution of eigenvalues of B depend on p? The density of eigenvalues of d-regular graphs is not the semi-circle; however, is there a choice of p that leads to semi-circular behavior? These are called weighted graphs; one can regard these weights (especially if p is positive) as encoding different information about the system (for example, how far apart different vertices are, or how long or how expensive it is to transmit information between vertices). See [Gold, QS2] for more details.

Research Project 2.4.6 (Complex Hermitian)**.** Investigate the eigenvalue densities for some of the ensembles for complex Hermitian rather than real symmetric matrices. For example, consider complex Hermitian Toeplitz matrices.

Research Project 2.4.7 (Spherical Ensemble: Non-Independent Entries)**.** In the spirit of Example 1.1.9, consider the ensemble of $N \times N$ real symmetric matrices where

$$\sum_{i=1}^{N}\sum_{j=i}^{N} a_{ij}^2 \;=\; \frac{N(N+1)}{2}. \tag{2.51}$$

Note the entries are not chosen independently from a fixed probability distribution, but rather we choose a point on a sphere of radius $\sqrt{N(N+1)/2}$; we do this so each a_{ij} is of size 1. What is the density of eigenvalues? Warning: the integrals will probably be too difficult to directly evaluate (except possibly for low N), though one can numerically investigate the eigenvalues. If we let $x_1 = a_{11}$, ..., $x_{N(N+1)/2} = a_{NN}$, then we have

$$x_1^2 + \cdots + x_{N(N+1)/2}^2 \;=\; R^2, \tag{2.52}$$

where $R = \sqrt{N(N+1)/2}$ is the radius of the $\frac{N(N+1)}{2}$-dimensional sphere. The following coordinate transformations are useful to generate points on an n-sphere of radius r:

$$\begin{array}{lcll}
x_1 & = & x_1(r,\phi_1,\ldots,\phi_{n-1}) & = & r\cos(\phi_1) \\
x_2 & = & x_2(r,\phi_1,\ldots,\phi_{n-1}) & = & r\sin(\phi_1)\cos(\phi_2) \\
x_3 & = & x_3(r,\phi_1,\ldots,\phi_{n-1}) & = & r\sin(\phi_1)\sin(\phi_2)\cos(\phi_3) \\
 & \vdots & & & \\
x_{n-1} & = & x_{n-1}(r,\phi_1,\ldots,\phi_{n-1}) & = & r\sin(\phi_1)\cdots\sin(\phi_{n-2})\cos(\phi_{n-1}) \\
x_n & = & x_n(r,\phi_1,\ldots,\phi_{n-1}) & = & r\sin(\phi_1)\cdots\sin(\phi_{n-2})\sin(\phi_{n-1}),
\end{array}$$

where $\phi_1, \ldots, \phi_{n-2} \in [0,\pi]$, $\phi_{n-1} \in [0,2\pi]$ and the volume is $\frac{\pi^{n/2} r^n}{\Gamma\left(\frac{n}{2}+1\right)}$. One can also consider other ensembles where the entries are not chosen independently; the point is to find ensembles that are easy to work with (either in determining the eigenvalues or in generating the matrices).

Bibliography

Links to many of the references below are available online at
http://www.math.princeton.edu/mathlab/book/index.html

[Acz] A. Aczel, *Fermat's Last Theorem: Unlocking the Secret of an Ancient Mathematical Problem*, Four Walls Eight Windows, New York, 1996.

[AKS] R. Adler, M. Keane, and M. Smorodinsky, *A construction of a normal number for the continued fraction transformation*, J. Number Theory **13** (1981), no. 1, 95–105.

[AgKaSa] M. Agrawal, N. Kayal and N. Saxena, *PRIMES is in P*, Ann. of Math. (2) **160** (2004), no. 2, 781–793.

[Al] L. Ahlfors, *Complex Analysis*, 3rd edition, McGraw-Hill, New York, 1979.

[AZ] M. Aigner and G. M. Ziegler, *Proofs from THE BOOK*, Springer-Verlag, Berlin, 1998.

[AGP] W. R. Alford, A. Granville, and C. Pomerance, *There are infinitely many Carmichael numbers*, Ann. Math. **139** (1994), 703–722.

[AMS] AMS MathSciNet, http://www.ams.org/msnmain?screen=Review

[AB] U. Andrews IV and J. Blatz, *Distribution of digits in the continued fraction representations of seventh degree algebraic irrationals*, Junior Thesis, Princeton University, Fall 2002.

[Ap] R. Apéry, *Irrationalité de $\zeta(2)$ et $\zeta(3)$*, Astérisque **61** (1979) 11–13.

[Apo] T. Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag, New York, 1998.

[ALM] S. Arms, A. Lozano-Robledo and S. J. Miller, *Constructing One-Parameter Families of Elliptic Curves over $\mathbb{Q}(T)$ with Moderate Rank*, to appear in the Journal of Number Theory.

[Art] M. Artin, *Algebra*, Prentice-Hall, Englewood Cliffs, NJ, 1991.

[Ay] R. Ayoub, *Introduction to the Analytic Theory of Numbers*, AMS, Providence, RI, 1963.

[Bai] Z. Bai, *Methodologies in spectral analysis of large-dimensional random matrices, a review*, Statist. Sinica **9** (1999), no. 3, 611–677.

[B] A. Baker, *Transcendental Number Theory*, Cambridge University Press, Cambridge, 1990.

[BM] R. Balasubramanian and C. J. Mozzochi, *Siegel zeros and the Goldbach problem*, J. Number Theory **16** (1983), no. 3, 311–332.

[BR] K. Ball and T. Rivoal, *Irrationalité d'une infinité valeurs de la fonction zeta aux entiers impairs*, Invent. Math. **146** (2001), 193–207.

[BT] V. V. Batyrev and Yu. Tschinkel, *Tamagawa numbers of polarized algebraic varieties*, Nombre et répartition de points de hauteur bornée (Paris, 1996), Astérisque (1998), No. 251, 299–340.

[BL] P. Baxandall and H. Liebeck, *Vector Calculus*, Clarendon Press, Oxford, 1986.

[Be] R. Beals, *Notes on Fourier series*, Lecture Notes, Yale University, 1994.

[Bec] M. Beceanu, *Period of the continued fraction of* $\sqrt{n}$, Junior Thesis, Princeton University, 2003.

[Ben] F. Benford, *The law of anomalous numbers*, Proceedings of the American Philosophical Society **78** (1938) 551–572.

[BBH] A. Berger, Leonid A. Bunimovich, and T. Hill, *One-dimensional dynamical systems and Benford's Law*, Trans. Amer. Math. Soc. **357** (2005), no. 1, 197–219.

[BEW] B. Berndt, R. Evans, and K. Williams, *Gauss and Jacobi Sums*, Canadian Mathematical Society Series of Monographs and Advanced Texts, Vol. 21, Wiley-Interscience Publications, John Wiley & Sons, New York, 1998.

[Ber] M. Bernstein, *Games, hats, and codes*, lecture at the SUMS 2005 Conference.

[BD] P. Bickel and K. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, San Francisco, 1977.

[Bi] P. Billingsley, *Probability and Measure*, 3rd edition, Wiley, New York, 1995.

[Bl1] P. Bleher, *The energy level spacing for two harmonic oscillators with golden mean ratio of frequencies*, J. Stat. Phys. **61** (1990) 869–876.

[Bl2] P. Bleher, *The energy level spacing for two harmonic oscillators with generic ratio of frequencies*, J. Stat. Phys. **63** (1991), 261–283.

[Bob] J. Bober, *On the randomness of modular inverse mappings*, Undergraduate Mathematics Laboratory report, Courant Institute, NYU, 2002.

[Bol] B. Bollobás, *Random Graphs*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2001.

[BoLa] E. Bombieri and J. Lagarias, *Complements to Li's criterion for the Riemann Hypothesis*, J. Number Theory **77** (1999), no. 2, 274–287.

[BG] E. Bombieri and W. Gubler, *Heights in Diophantine Geometry*, Cambridge University Press, Cambridge, UK, 2006.

[BP] E. Bombieri and A. van der Poorten, *Continued fractions of algebraic numbers*. Pages 137–152 in *Computational Algebra and Number Theory (Sydney, 1992)*, Mathematical Applications, Vol. 325, Kluwer Academic, Dordrecht, 1995.

[Bon] D. Boneh, *Twenty years of attacks on the RSA cryptosystem*, Notices of the American Mathematical Society **46** (1999), no. 2, 203–213.

[BS] Z. Borevich and I. Shafarevich, *Number Theory*, Academic Press, New York, 1968.

[BB] J. Borwein and P. Borwein, *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity*, John Wiley and Sons, New York, 1987.

[BK] A. Boutet de Monvel and A. Khorunzhy, *Some elementary results around the Wigner semicircle law*, lecture notes.

[BoDi] W. Boyce and R. DiPrima, *Elementary differential equations and boundary value problems*, 7th edition, John Wiley & Sons, New York, 2000.

[Bre1] R. Brent, *The distribution of small gaps between successive primes*, Math. Comp. **28** (1974), 315–324.

[Bre2] R. Brent, *Irregularities in the distribution of primes and twin primes*, Collection of articles dedicated to Derrick Henry Lehmer on the occasion of his seventieth birthday, Math. Comp. **29** (1975), 43–56.

[BPR] R. Brent, A. van der Poorten, and H. te Riele, *A comparative study of algorithms for computing continued fractions of algebraic numbers*. Pages 35–47 in *Algorithmic number theory (Talence, 1996)*, Lecture Notes in Computer Science, Vol. 1122, Springer, Berlin, 1996.

[deBr] R. de la Bretèche, *Sur le nombre de points de hauteur bornée d'une certaine surface cubique singulière*. Pages 51–77 in *Nombre et répartition de points de hauteur bornée (Paris, 1996)*, Astérisque, (1998) no. 251, 51–77.

[BBD] R. de la Bretèche, T. D. Browning, and U. Derenthal, *On Manin's conjecture for a certain singular cubic surface*, preprint.

[BPPW] B. Brindza, A. Pintér, A. van der Poorten, and M. Waldschmidt, *On the distribution of solutions of Thue's equation*. Pages 35–46 in *Number theory in progress (Zakopane-Koscielisko, 1997)*, Vol. 1, de Gruyter, Berlin, 1999.

[BFFMPW] T. Brody, J. Flores, J. French, P. Mello, A. Pandey, and S. Wong, *Random-matrix physics: spectrum and strength fluctuations*, Rev. Mod. Phys. **53** (1981), no. 3, 385–479.

[BrDu] J. Brown and R. Duncan, *Modulo one uniform distribution of the sequence of logarithms of certain recursive sequences*, Fibonacci Quarterly **8** (1970) 482–486.

[Bro] T. Browning, *The density of rational points on a certain singular cubic surface*, preprint.

[BDJ] W. Bryc, A. Dembo, T. Jiang, *Spectral measure of large random Hankel, Markov and Toeplitz matrices*, Ann. Probab. **34** (2006), no. 1, 1–38.

[Bry] A. Bryuno, *Continued frations of some algebraic numbers*, U.S.S.R. Comput. Math. & Math. Phys. **4** (1972), 1–15.

[Bur] E. Burger, *Exploring the Number Jungle: A Journey into Diophantine Analysis*, AMS, Providence, RI, 2000.

[BuP] E. Burger and A. van der Poorten, *On periods of elements from real quadratic number fields*. Pages 35–43 in *Constructive, Experimental, and Nonlinear Analysis (Limoges, 1999)*, CMS Conf. Proc., **27**, AMS, Providence, RI, 2000.

[CaBe] G. Casella and R. Berger, *Statistical Inference*, 2nd edition, Duxbury Advanced Series, Pacific Grove, CA, 2002.

[CGI] G. Casati, I. Guarneri, and F. M. Izrailev, *Statistical properties of the quasi-energy spectrum of a simple integrable system*, Phys. Lett. A **124** (1987), 263–266.

[Car] L. Carleson, *On the convergence and growth of partial sums of Fourier series*, Acta Math. **116** (1966), 135–157.

[Ca] J. W. S. Cassels, *An Introduction to Diophantine Approximation*, Cambridge University Press, London 1957.

[Ch] D. Champernowne, *The construction of decimals normal in the scale of ten*, J. London Math. Soc. 8 (1933), 254–260.

[Cha] K. Chang, *An experimental approach to understanding Ramanujan graphs*, Junior Thesis, Princeton University, Spring 2001.

[ChWa] J. R. Chen and T. Z. Wang, *On the Goldbach problem*, Acta Math. Sinica **32** (1989), 702–718.

[Chr] J. Christiansen, *An introduction to the moment problem*, lecture notes.

[Ci] J. Cisneros, *Waring's problem*, Junior Thesis, Princeton University, Spring 2001.

[CW] J. Coates and A. Wiles, *On the conjecture of Birch and Swinnterton-Dyer*, Invent. Math. **39** (1977), 43–67.

[CB] S. Chatterjee and A. Bose, *A new method for bounding rates of convergence of empirical spectral distributions*, J. Theoret. Probab. **17** (2004), no. 4, 1003–1019.

[Cof1] M. Coffey, *Toward verification of the Riemann hypothesis: Application of the Li criterion*, to appear in Math. Physics, Analysis and Geometry. http://arxiv.org/pdf/math-ph/0505052.

[Cof2] M. Coffey, *On the coefficients of the Baez-Duarte criterion for the Riemann hypothesis and their extensions*, preprint. http://arxiv.org/pdf/math-ph/0608050.

[CL1] H. Cohen and H. W. Lenstra, Jr., *Heuristics on class groups of number fields*. Pages 33–62 in *Number Theory*, Lecture Notes in Mathematics, Vol. 1068, Springer-Verlag, Berlin, 33–62.

[CL2] H. Cohen and H. W. Lenstra, Jr., *Heuristics on class groups*, in *Number Theory*, Lecture Notes in Mathematics, Vol. 1052, Springer-Verlag, Berlin, 26–36.

[Coh] P. Cohen, *The independence of the continuum hypothesis*, Proc. Nat. Acad. Sci. U.S.A, **50** (1963), 1143–1148; **51** (1964), 105–110.

[Cohn] J. Cohn, *The length of the period of simple continued fractions*, Pacific Journal of Mathematics, **71** (1977), no. 1, 21–32.

[Con1] J. B. Conrey, *L-Functions and random matrices*. Pages 331–352 in *Mathematics unlimited — 2001 and Beyond*, Springer-Verlag, Berlin, 2001.

[Con2] J. B. Conrey, *The Riemann hypothesis*, Notices of the AMS, **50** (2003), no. 3, 341–353.

[CFKRS] B. Conrey, D. Farmer, P. Keating, M. Rubinstein and N. Snaith, *Integral moments of L-functions*, Proc. London Math. Soc. (3) **91** (2005), no. 1, 33–104.

[Conw] J. H. Conway, *The weird and wonderful chemistry of audioactive decay*. Pages 173–178 in *Open Problems in Communications and Computation*, ed. T. M. Cover and B. Gopinath, Springer-Verlag, New York, 1987.

[CG] J. H. Conway and R. Guy, *The Book of Numbers*, Springer-Verlag, Berlin, 1996.

[CS] J. H. Conway and N. J. A. Sloane, *Lexicographic Codes: Error-Correcting Codes from Game Theory*, IEEE Trans. Inform. Theory, **32** (1986), no. 3, 219–235.

[Corl] R. M. Corless,*Continued fractions and chaos*. Amer. Math. Monthly **99** (1992), no. 3, 203–215.

[Cor1] Cornell University, *arXiv*, http://arxiv.org

[Cor2] Cornell University, *Project Euclid*, http://projecteuclid.org/

[CFS] I. P. Cornfeld, S. V. Fomin, and I. G. Sinai, *Ergodic Theory*, Grundlehren Der Mathematischen Wissenschaften, Springer-Verlag, Berlin, 1982.

[Da1] H. Davenport, *The Higher Arithmetic: An Introduction to the Theory of Numbers*, 7th edition, Cambridge University Press, Cambridge, 1999.

[Da2] H. Davenport, *Multiplicative Number Theory*, 2nd edition, revised by H. Montgomery, Graduate Texts in Mathematics, Vol. 74, Springer-Verlag, New York, 1980.

[Da3] H. Davenport, *On the distribution of quadratic residues (mod p)*, London Math. Soc. **6** (1931), 49–54.

[Da4] H. Davenport, *On character sums in finite fields*, Acta Math. **71** (1939), 99–121.

[DN] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd edition, Wiley Interscience, Hoboken, NJ, 2003.

[DSV] G. Davidoff, P. Sarnak, and A. Valette, *Elementary Number Theory, Group Theory, and Ramanujan Graphs*, London Mathematical Society, Student Texts, Vol. 55, Cambridge University Press, Cambridge 2003.

[Dev] R. Devaney, *An Introduction to Chaotic Dynamical Systems*, 2nd edition, Westview Press, Cambridge, MA, 2003.

[Dia] P. Diaconis, *Patterns in eigenvalues: the* 70^{th} *Josiah Williard Gibbs lecture*, Bulletin of the American Mathematical Society **40** (2003), no. 2, 155–178.

[Di] T. Dimofte, *Rational shifts of linearly periodic continued fractions*, Junior Thesis, Princeton University, 2003.

[DM] E. Dueñez and S. J. Miller, *The Low Lying Zeros of a* $\mathrm{GL}(4)$ *and a* $\mathrm{GL}(6)$ *family of L-functions*, to appear in Compositio Mathematica.

[Du] R. Durrett, *Probability: Theory and Examples*, 2nd edition, Duxbury Press, 1996.

[Dy1] F. Dyson, *Statistical theory of the energy levels of complex systems: I, II, III*, J. Mathematical Phys. **3** (1962) 140–156, 157–165, 166–175.

[Dy2] F. Dyson, *The threefold way. Algebraic structure of symmetry groups and ensembles in quantum mechanics*, J. Mathematical Phys., **3** (1962) 1199–1215.

[Edg] G. Edgar, *Measure, Topology, and Fractal Geometry*, 2nd edition, Springer-Verlag, 1990.

[Ed] H. M. Edwards, *Riemann's Zeta Function*, Academic Press, New York, 1974.

[EST] B. Elias, L. Silberman and R. Takloo-Bighash, *On Cayley's theorem*, preprint.

[EE] W. J. Ellison and F. Ellison, *Prime Numbers*, John Wiley & Sons, New York, 1985.

[Est1] T. Estermann, *On Goldbach's problem: Proof that almost all even positive integers are sums of two primes*, Proc. London Math. Soc. Ser. 2 **44** (1938) 307–314.

[Est2] T. Estermann, *Introduction to Modern Prime Number Theory*, Cambridge University Press, Cambridge, 1961.

[Fal] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, 2nd edition, John Wiley & Sons, New York, 2003.

[Fef] C. Fefferman, *Pointwise convergence of Fourier series*, Ann. of Math. Ser. 2 **98** (1973), 551–571.

[Fe] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd edition, Vol. II, John Wiley & Sons, New York, 1971.

[Fi] D. Fishman, *Closed form continued fraction expansions of special quadratic irrationals*, Junior Thesis, Princeton University, 2003.

[Fol] G. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd edition, Pure and Applied Mathematics, Wiley-Interscience, New York, 1999.

[For] P. Forrester, *Log-gases and random matrices*, book in progress.

[Fou] E. Fouvry, *Sur la hauteur des points d'une certaine surface cubique singulière*. In *Nombre et répartition de points de hauteur bornée (Paris, 1996)*, Astérisque, (1999) no. 251, 31–49.

[FSV] P. J. Forrester, N. C. Snaith, and J. J. M. Verbaarschot, *Developments in Random Matrix Theory*. In *Random matrix theory*, J. Phys. A **36** (2003), no. 12, R1–R10.

[Fr] J. Franklin, *Mathematical Methods of Economics: Linear and Nonlinear Programming, Fixed-Point Theorem*, Springer-Verlag, New York, 1980.

[Ga] P. Garrett, *Making, Breaking Codes: An Introduction to Cryptography*, Prentice-Hall, Englewood Cliffs, NJ, 2000.

[Gau] M. Gaudin, *Sur la loi limite de l'espacement des valeurs propres d'une matrice aléatoire*, Nucl. Phys. **25** (1961) 447–458.

[Gel] A. O. Gelfond, *Transcendental and Algebraic Numbers*, Dover, New York, 1960.

[Gl] A. Gliga, *On continued fractions of the square root of prime numbers*, Junior Thesis, Princeton University, 2003.

[Gö] K. Gödel, *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*, Dover, New York, 1992.

[Gol1] D. Goldfeld, *The class number of quadratic fields and the conjectures of Birch and Swinnerton-Dyer*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. 3, **4** (1976), 624–663.

[Gol2] D. Goldfeld, *The Elementary proof of the Prime Number Theorem, An Historical Perspective*. Pages 179–192 in *Number Theory, New York Seminar 2003*, eds. D. and G. Chudnovsky, M. Nathanson, Springer-Verlag, New York, 2004.

[Gold] L. Goldmakher, *On the limiting distribution of eigenvalues of large random regular graphs with weighted edges*, American Institute of Mathematics Summer REU, 2003.

[GV] D. A. Goldston and R. C. Vaughan, *On the Montgomery-Hooley asymptotic formula*. Pages 117–142 in *Sieve Methods, Exponential Sums and their Applications in Number Theory*, ed. G. R. H. Greaves, G. Harman, and M. N. Huxley, Cambridge University Press, Cambridge, 1996.

[GG] M. Golubitsky and V. Guillemin, *Stable Mappings and Their Singularities*, Graduate Texts in Mathematics, Vol. 14, Springer-Verlag, New York, 1973.

[Gou] X. Gourdon, *The 10^{13} first zeros of the Riemann zeta function, and zeros computation at very large height*, preprint. http://numbers.computation.free.fr/Constants/Miscellaneous/zetazeros1e13-1e24.pdf

[GKP] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, Addison-Wesley, Reading, MA, 1988.

[GK] A. Granville and P. Kurlberg, *Poisson statistics via the Chinese remainder theorem*, preprint.

[GT] A. Granville and T. Tucker, *It's as easy as abc*, Notices of the AMS **49** (2002), no. 10, 224–1231.

[GZ] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320.

[Guy] R. Guy, *Unsolved Problems in Number Theory (Problem Books in Mathematics)*, 2nd edition, Springer-Verlag, New York, 1994.

[HM] C. Hammond and S. J. Miller, *Eigenvalue spacing distribution for the ensemble of real symmetric Toeplitz matrices*, Journal of Theoretical Probability **18** (2005), no. 3, 537–566.

[HL1] G. H. Hardy and J. E. Littlewood, *A new solution of Waring's problem*, Q. J. Math. **48** (1919), 272–293.

[HL2] G. H. Hardy and J. E. Littlewood, *Some problems of "Partitio Numerorum." A new solution of Waring's problem*, Göttingen Nach. (1920), 33–54.

[HL3] G. H. Hardy and J. E. Littlewood, *Some problems of "Partitio Numerorum." III. On the expression of a number as a sum of primes*, Acta Math. **44** (1923), 1–70.

[HL4] G. H. Hardy and J. E. Littlewood, *Some problems of "Partitio Numerorum." IV. Further researches in Waring's problem*, Math. Z. **23** (1925) 1–37.

[HR] G. H. Hardy and S. Ramanujan, *Asymptotic formulae in combinatorial analysis*, Proc. London Math. Soc. **17** (1918), 75–115.

[HW] G. H. Hardy and E. Wright, *An Introduction to the Theory of Numbers*, 5th edition, Oxford Science Publications, Clarendon Press, Oxford, 1995.

[Hata] R. Hata, *Improvement in the irrationality measures of π and π^2*, Proc. Japan. Acad. Ser. A Math. Sci. **68** (1992), 283–286.

[Ha1] B. Hayes, *Third Base: Three cheers for base 3!*, American Scientist **89** (2001), no. 6, 490–494.

[Ha2] B. Hayes, *The spectrum of Riemannium*, American Scientist **91** (2003), no. 4, 296–300.

[He] R. Heath-Brown, *The density of rational points on Cayley's cubic surface*, preprint.

[Hei] H. Heillbronn, *On the average length of a class of finite continued fractions.* In *Number Theory and Analysis (A collection of papers in honor of E. Landau)*, VEB Deutscher Verlag, Berlin, 1968.

[Hej] D. Hejhal, *On the triple correlation of zeros of the zeta function*, Internat. Math. Res. Notices (1994), no. 7, 294–302.

[Hil] D. Hilbert, *Beweis für die Darstellbarkeit der ganzen zahlen durch eine feste Anzahl n^{ter} Potenzen (Waringsches Problem)*, Mat. Annalen **67** (1909), 281–300.

[Hi1] T. Hill, *The first-digit phenomenon*, American Scientist **86** (1996), 358–363.

[Hi2] T. Hill, *A statistical derivation of the significant-digit law*, Statistical Science **10** (1996), 354–363.

[HS] M. Hindry and J. Silverman, *Diophantine Geometry: An Introduction*, Graduate Texts in Mathematics, Vol. 201, Springer-Verlag, New York, 2000.

[HSP] J. Hoffstein, J. H. Silverman and J. Pipher, *An Introduction to Mathematical Cryptography*.

[HJ] K. Hrbacek and T. Jech, *Introduction to Set Theory*, Pure and Applied Mathematics, Marcel Dekker, New York, 1984.

[Hua] Hua Loo Keng, *Introduction to Number Theory*, Springer-Verlag, New York, 1982.

[HuRu] C. Hughes and Z. Rudnick, *Mock Gaussian behaviour for linear statistics of classical compact groups*, J. Phys. A **36** (2003) 2919–2932.

[Hu] J. Hull, *Options, Futures, and Other Derivatives*, 5th edition, Prentice-Hall, Englewood Cliffs, NJ, 2002.

[IR] K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, Graduate Texts in Mathematics, Vol. 84, Springer-Verlag, New York, 1990.

[Iw] H. Iwaniec, *Topics in Classical Automorphic Forms*, Graduate Studies in Mathematics, Vol. 17, AMS, Providence, RI, 1997.

[IK] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, AMS Colloquium Publications, Vol. 53, AMS, Providence, RI, 2004.

[ILS] H. Iwaniec, W. Luo, and P. Sarnak, *Low lying zeros of families of L-functions*, Inst. Hautes Études Sci. Publ. Math. **91** (2000), 55–131.

[IS1] H. Iwaniec and P. Sarnak, *Dirichlet L-functions at the central point*. Pages 941–952 in *Number Theory in Progress, (Zakopane-Kościelisko, 1997)*, Vol. 2, de Gruyter, Berlin, 1999.

[IS2] H. Iwaniec and P. Sarnak, *The non-vanishing of central values of automorphic L-functions and Landau-Siegel zeros*, Israel J. Math. **120** (2000), 155–177.

[JMRR] D. Jakobson, S. D. Miller, I. Rivin, and Z. Rudnick, *Eigenvalue spacings for regular graphs*. Pages 317–327 in *Emerging Applications of Number Theory (Minneapolis, 1996)*, The IMA Volumes in Mathematics and its Applications, Vol. 109, Springer, New York, 1999.

[J] N. Jacobson, *Basic Algebra I*, 2nd edition, W H Freeman & Co, San Francisco, 1985.

[Je] R. Jeffrey, *Formal Logic: Its Scope and Limits*, McGraw-Hill, New York, 1989.

[Ka] S. Kapnick, *Continued fraction of cubed roots of primes*, Junior Thesis, Princeton University, Fall 2002.

[Kar] A. Karlsson, *Applications of heat kernels on Abelian groups:* $\zeta(2n)$*, quadratic reciprocity, Bessel integral*, preprint.

[KS1] N. Katz and P. Sarnak, *Random Matrices, Frobenius Eigenvalues and Monodromy*, AMS Colloquium Publications, Vol. 45, AMS, Providence, RI, 1999.

[KS2] N. Katz and P. Sarnak, *Zeros of zeta functions and symmetries*, Bull. AMS **36** (1999), 1–26.

[KeSn] J. P. Keating and N. C. Snaith, *Random matrices and* L*-functions*. In *Random Matrix Theory*, J. Phys. A **36** (2003), no. 12, 2859–2881.

[Kei] J. B. Keiper, *Power series expansions of Riemann's* ξ *function*, Math. Comp. **58**(1992), 765-773.

[Kel] D. Kelley, *Introduction to Probability*, Macmillan Publishing Company, London, 1994.

[Kh] A. Y. Khinchin, *Continued Fractions*, 3rd edition, University of Chicago Press, Chicago, 1964.

[KSS] D. Kleinbock, N. Shah, and A. Starkov,*Dynamics of subgroup actions on homogeneous spaces of Lie groups and applications to number theory*. Pages 813–930 in *Handbook of Dynamical Systems*, Vol. 1A, North-Holland, Amsterdam, 2002.

[Kn] A. Knapp, *Elliptic Curves*, Princeton University Press, Princeton, NJ, 1992.

[Knu] D. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, 3rd edition, Addison-Wesley, MA, 1997.

[Kob1] N. Koblitz, *Why study equations over finite fields?*, Math. Mag. **55** (1982), no. 3, 144–149.

[Kob2] N. Koblitz, *Elliptic curve cryptosystems*, Math. Comp. **48** (1987), no. 177, 203–209.

[Kob3] N. Koblitz, *A survey of number theory and cryptography*. Pages 217-239 in *Number Theory*, Trends in Mathematics, Birkhäuser, Basel, 2000.

[Ko] V. Kolyvagin, *On the Mordell-Weil group and the Shafarevich-Tate group of modular elliptic curves*. Pages 429-436 in *Proceedings of the International Congress of Mathematicians (Kyoto, 1990)*, vols. I and II, Math. Soc. Japan, Tokyo, 1991.

[KonMi] A. Kontorovich and S. J. Miller, *Benford's law, values of* L*-functions and the* $3x+1$ *problem*, Acta Arith. **120** (2005), 269–297.

[KonSi] A. Kontorovich and Ya. G. Sinai, *Structure theorem for* (d, g, h)*-maps*, Bull. Braz. Math. Soc. (N.S.) 33 (2002), no. 2, 213–224.

[Kor] A. Korselt, *Probléme chinois*, L'intermédiaire math. **6** (1899), 143–143.

[Kos] T. Koshy, *Fibonacci and Lucas Numbers with Applications*, Wiley-Interscience, New York, 2001

[Kua] F. Kuan, *Digit distribution in the continued fraction of* $\zeta(n)$, Junior Thesis, Princeton University, Fall 2002.

[KN] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, John Wiley & Sons, New York, 1974.

[KR] P. Kurlberg and Z. Rudnick, *The distribution of spacings between quadratic residues*, Duke Math. J. **100** (1999), no. 2, 211–242.

[Ku] R. Kuzmin, *Ob odnoi zadache Gaussa*, Doklady Akad. Nauk, Ser. A (1928), 375–380.

[Lag1] J. Lagarias, *The* $3x + 1$ *problem and its generalizations*. Pages 305-334 in *Organic mathematics (Burnaby, BC, 1995)*, CMS Conf. Proc., vol. 20, AMS, Providence, RI, 1997.

[Lag2] J. Lagarias, *The 3x+1 problem: An annotated bibliography*, preprint.

[LaSo] J. Lagarias and K. Soundararajan, *Benford's Law for the* $3x + 1$ *function*, preprint.

[La1] S. Lang, *Diophantine Geometry*, Interscience Publishers, New York, 1962.

[La2] S. Lang, *Introduction to Diophantine Approximations*, Addison-Wesley, Reading, MA, 1966.

[La3] S. Lang, *Undergraduate Algebra*, 2nd edition, Springer-Verlag, New York, 1986.

[La4] S. Lang, *Calculus of Several Variables*, Springer-Verlag, New York, 1987.

[La5] S. Lang, *Undergraduate Analysis*, 2nd edition, Springer-Verlag, New York, 1997.

[La6] S. Lang, *Complex Analysis*, Graduate Texts in Mathematics, Vol. 103, Springer-Verlag, New York, 1999.

[LT] S. Lang and H. Trotter, *Continued fractions for some algebraic numbers*, J. Reine Angew. Math. **255** (1972), 112–134.

[LF] R. Larson and B. Farber, *Elementary Statistics: Picturing the World*, Prentice-Hall, Englewood Cliffs, NJ, 2003.

[LP] R. Laubenbacher and D. Pengelley, *Gauss, Eisenstein, and the "third" proof of the quadratic reciprocity theorem: Ein kleines Schauspiel*, Math. Intelligencer 16 (1994), no. 2, 67–72.

[Law1] J. Law, *Kuzmin's theorem on algebraic numbers*, Junior Thesis, Princeton University, Fall 2002.

[Law2] J. Law, *The circle method on the binary Goldbach conjecture*, Junior Thesis, Princeton University, Spring 2003.

[Leh] R. Lehman, *First order spacings of random matrix eigenvalues*, Junior Thesis, Princeton University, Spring 2000.

[LS] H. Lenstra and G. Seroussi, *On hats and other covers*, 2002, preprint.

[Le] P. Lévy, *Sur les lois de probabilité dont dependent les quotients complets et incomplets d'une fraction continue*, Bull. Soc. Math. **57** (1929), 178–194.

[XLi] X.-J. Li, *The positivity of a sequence of numbers and Riemann hypothesis*, J. Number Theory **65** (1997), 325-333.

[LU] C. Liaw and H. Úlfarsson, *Transcendence of e and π*, class notes for Math 252 (Graduate Algebra), Brown University, Spring 2006.

[Lidl] R. Lidl, *Mathematical aspects of cryptanalysis*. Pages 86–97 in *Number Theory and Cryptography (Sydney, 1989)*, London Mathematical Society Lecture Note Series, vol. 154, Cambridge University Press, Cambridge, 1990.

[Li] R. Lipshitz, *Numerical results concerning the distribution of $\{n^2\alpha\}$*, Junior Thesis, Princeton University, Spring 2000.

[Liu] Y. Liu, *Statistical behavior of the eigenvalues of random matrices*, Junior Thesis, Princeton University, Spring 2000.

[Mah] K. Mahler, *Arithmetische Eigenschaften einer Klasse von Dezimalbrüchen*, Amsterdam Proc. Konin. Neder. Akad. Wet. **40** (1937), 421–428.

[Ma] E. S. Mahmoodian, *Mathematical Olympiads in Iran*, Vol. I, Sharif University Press, Tehran, Iran, 2002.

[Man] B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman, New York, 1982.

[Mar] J. Marklof, *Almost modular functions and the distribution of n^2x modulo one*, Int. Math. Res. Not. (2003), no. 39, 2131–2151.

[MaMc] R. Martin and W. McMillen, *An elliptic curve over $\mathbb{Q}$ with rank at least 24*, Number Theory Listserver, May 2000.

[MMS] A. Massey, S. J. Miller, and J. Sinsheimer, *Eigenvalue spacing distribution for the ensemble of real symmetric palindromic Toeplitz matrices*, to appear in the Journal of Theoretical Probability.

[Maz1] B. Mazur, *Modular curves and the Eisenstein ideal*, IHES Publ. Math. **47** (1977), 33–186.

[Maz2] B. Mazur, *Rational isogenies of prime degree (with an appendix by D. Goldfeld)*, Invent. Math. **44** (1978), no. 2, 129–162.

[Maz3] B. Mazur, *Number Theory as Gadfly*, Amer. Math. Monthly, **98** (1991), 593–610.

[McK] B. McKay, *The expected eigenvalue distribution of a large regular graph*, Linear Algebra Appl. **40** (1981), 203–216.

[McW] B. McKay and N. Wormald, *The degree sequence of a random graph. I. The models*, Random Structures Algorithms **11** (1997), no. 2, 97–117.

[Meh1] M. Mehta, *On the statistical properties of level spacings in nuclear spectra*, Nucl. Phys. **18** (1960), 395–419.

[Meh2] M. Mehta, *Random Matrices*, 2nd edition, Academic Press, Boston, 1991.

[Met] N. Metropolis, *The beginning of the Monte Carlo method*, Los Alamos Science, No. 15, Special Issue (1987), 125–130.

[MU] N. Metropolis and S. Ulam, *The Monte Carlo method*, J. Amer. Statist. Assoc. **44** (1949), 335–341.

[Mic1] M. Michelini, *Independence of the digits of continued fractions*, Junior Thesis, Princeton University, Fall 2002.

[Mic2] M. Michelini, *Kuzmin's extraordinaty zero measure set*, Senior Thesis, Princeton University, Spring 2004.

[Mi1] N. Miller, *Various tendencies of non-Poissonian distributions along subsequences of certain transcendental numbers*, Junior Thesis, Princeton University, Fall 2002.

[Mi2] N. Miller, *Distribution of eigenvalue spacings for band-diagonal matrices*, Junior Thesis, Princeton University, Spring 2003.

[Mill] S. D. Miller, *A simpler way to show $\zeta(3)$ is irrational*, preprint.

[Mil1] S. J. Miller, *1- and 2-level densities for families of elliptic curves: Evidence for the underlying group symmetries*, Compositio Mathematica **140** (2004), no. 4, 952–992.

[Mil2] S. J. Miller, *Density functions for families of Dirichlet characters*, preprint.

[Mil3] S. J. Miller, *The arithmetic mean and geometric inequality*, Class Notes from Math 187/487, The Ohio State University, Fall 2003.

[Mil4] S. J. Miller, *Differentiating identities*, Class Notes from Math 162: Statistics, Brown University, Spring 2005.

[Mil5] S. J. Miller, *The Pythagorean won-loss formula in baseball*, preprint.

[Mil6] S. J. Miller, *Investigations of zeros near the central point of elliptic curve L-functions*, to appear in Experimental Mathematics.

[Mil7] S. J. Miller, *Die battles and order statistics*, Class Notes from Math 162: Statistics, Brown University, Spring 2006.

[Mil8] S. J. Miller, *Beyond the Pigeon-Hole Principle: Many pigeons in the same box*, Class Notes from Math 162: Statistics, Brown University, Spring 2006.

[MN] S. J. Miller and M. Nigrini, *Order Statistics and Shifted Almost Benford Behavior*, preprint.

[M] V. Miller, *Use of elliptic curves in cryptography*. Pages 417–426 in *Advances in cryptology – CRYPTO '85 (Santa Barbara, CA, 1985)*, Lecture Notes in Computer Science, Vol. 218, Springer-Verlag, Berlin, 1986.

[Milne] J. S. Milne, *Elliptic Curves*, course notes.

[Min] S. Minteer, *Analysis of Benford's law applied to the $3x+1$ problem*, Number Theory Working Group, The Ohio State University, 2004.

[Mon1] H. Montgomery, *Primes in arithmetic progression*, Michigan Math. J. **17** (1970), 33–39.

[Mon2] H. Montgomery, *The pair correlation of zeros of the zeta function*. Pages 181–193 in *Analytic Number Theory*, Proceedings of Symposia in Pure Mathematics, vol. 24, AMS, Providence, RI, 1973.

[MoMc] D. Moore and G. McCabe, *Introduction to the Practice of Statistics*, W. H. Freeman and Co., London, 2003.

[MS] H. Montgomery and K. Soundararajan, *Beyond pair correlation*. Pages 507–514 in *Paul Erdös and His Mathematics, I (Budapest, 1999)*, Bolyai Society Mathematical Studies, Vol. 11, János Bolyai Math. Soc., Budapest, 2002.

[MW] C. J. Moreno and S. S. Wagstaff, Jr., *Sums of Squares of Integers*, Chapman and Hall, 2006.

[Moz1] C. J. Mozzochi, *An analytic sufficiency condition for Goldbach's conjecture with minimal redundancy*, Kyungpook Math. J. **20** (1980), no. 1, 1–9.

[Moz2] C. J. Mozzochi, *The Fermat Diary*, AMS, Providence, RI, 2000.

[Moz3] C. J. Mozzochi, *The Fermat Proof*, Trafford Publishing, Victoria, 2004.

[Mu1] R. Murty, *Primes in certain arithmetic progressions*, Journal of the Madras University, (1988), 161–169.

[Mu2] R. Murty, *Problems in Analytic Number Theory*, Springer-Verlag, New York, 2001.

[MM] M. R. Murty and V. K. Murty, *Non-Vanishing of L-Functions and Applications*, Progress in Mathematics, vol. 157, Birkhäuser, Basel, 1997.

[NS] K. Nagasaka and J. S. Shiue, *Benford's law for linear recurrence sequences*, Tsukuba J. Math. **11** (1987), 341–351.

[Nar] W. Narkiewicz, *The Development of Prime Number Theory*, Springer Monographs in Mathematics, Springer-Verlag, New York, 2000.

[Na] M. Nathanson, *Additive Number Theory: The Classical Bases*, Graduate Texts in Mathematics, Springer-Verlag, New York, 1996.

[NT] J. von Neumann and B. Tuckerman, *Continued fraction expansion of* $2^{1/3}$, Math. Tables Aids Comput. **9** (1955), 23–24.

[Ni1] T. Nicely, *The pentium bug*, http://www.trnicely.net/pentbug/pentbug.html

[Ni2] T. Nicely, *Enumeration to* 10^{14} *of the Twin Primes and Brun's Constant*, Virginia J. Sci. **46** (1996), 195–204.

[Nig1] M. Nigrini, *Digital Analysis and the Reduction of Auditor Litigation Risk.* Pages 69–81 in *Proceedings of the 1996 Deloitte & Touche / University of Kansas Symposium on Auditing Problems*, ed. M. Ettredge, University of Kansas, Lawrence, KS, 1996.

[Nig2] M. Nigrini, *The Use of Benford's Law as an Aid in Analytical Procedures*, Auditing: A Journal of Practice & Theory, **16** (1997), no. 2, 52–67.

[NZM] I. Niven, H. Zuckerman, and H. Montgomery, *An Introduction to the Theory of Numbers*, 5th edition, John Wiley & Sons, New York, 1991.

[Nov] T. Novikoff, *Asymptotic behavior of the random* 3*-regular bipartite graph*, Undergraduate Mathematics Laboratory report, Courant Institute, NYU, 2002.

[Ny] J. E. Nymann, *An Application of Diophantine Approximation*, The American Mathematical Monthly **76** (1969), no. 6, 668–671.

[Od1] A. Odlyzko, *On the distribution of spacings between zeros of the zeta function*, Math. Comp. **48** (1987), no. 177, 273–308.

[Od2] A. Odlyzko, *The* 10^{22}*-nd zero of the Riemann zeta function*. Pages 139–144 in *Proceedings of the Conference on Dynamical, Spectral and Arithmetic Zeta Functions*, ed. M. van Frankenhuysen and M. L. Lapidus, Contemporary Mathematics Series, AMS, Providence, RI, 2001.

[Ok] T. Okano, *A note on the transcendental continued fractions*, Tokyo J. Math **10** (1987), no. 1, 151–156.

[Ol] T. Oliveira e Silva, *Verification of the Goldbach conjecture up to* $6 \cdot 10^{16}$, NMBRTHRY@listserv.nodak.edu mailing list, Oct. 3, 2003, http://listserv.nodak.edu/scripts/wa.exe?A2=ind0310&L=nmbrthry&P=168 and http://www.ieeta.pt/∼tos/goldbach.html

[Ols] L. Olsen, *Extremely non-normal continued fractions*, Acta Arith. **108** (2003), no. 2, 191–202.

[Pi] R. G. E. Pinch, *The Carmichael numbers up to* 10^{18}, preprint, http://arxiv.org/abs/math.NT/0604376.

[Pol] G. Polya, *Heuristic reasoning in the theory of numbers*, Amer. Math. Monthly **66** (1959) 375–384.

[vdP1] A. van der Poorten, *An introduction to continued fractions*. Pages 99-138 in *Diophantine Analysis (Kensington, 1985)*, London Mathematical Society Lecture Note Series, Vol. 109, Cambridge University Press, Cambridge, 1986.

[vdP2] A. van der Poorten, *Notes on continued fractions and recurrence sequences*. Pages 86–97 in *Number theory and cryptography (Sydney, 1989)*, London Mathematical Society Lecture Note Series, Vol. 154, Cambridge University Press, Cambridge, 1990.

[vdP3] A. van der Poorten, *Continued fractions of formal power series*. Pages 453–466 in *Advances in Number Theory (Kingston, ON, 1991)*, Oxford Science Publications, Oxford University Press, New York, 1993.

[vdP4] A. van der Poorten, *Fractions of the period of the continued fraction expansion of quadratic integers*, Bull. Austral. Math. Soc. **44** (1991), no. 1, 155–169.

[vdP5] A. van der Poorten, *Continued fraction expansions of values of the exponential function and related fun with continued fractions*, Nieuw Arch. Wisk. (4) **14** (1996), no. 2, 221–230.

[vdP6] A. van der Poorten, *Notes on Fermat's Last Theorem*, Canadian Mathematical Society Series of Monographs and Advanced Texts, Wiley-Interscience, New York, 1996.

[PS1] A. van der Poorten and J. Shallit, *Folded continued fractions*, J. Number Theory **40** (1992), no. 2, 237–250.

[PS2] A. van der Poorten and J. Shallit, *A specialised continued fraction*, Canad. J. Math. **45** (1993), no. 5, 1067–1079.

[Po] C. Porter (editor), *Statistical Theories of Spectra: Fluctuations*, Academic Press, New York, 1965.

[Py] R. Pyke, *Spacings*, J. Roy. Statist. Soc. Ser. B **27** (1965), 395–449.

[QS1] R. Qian and D. Steinhauer, *Rational relation conjectures*, Junior Thesis, Princeton University, Fall 2003.

[QS2] R. Qian and D. Steinhauer, *Eigenvalues of weighted random graphs*, Junior Thesis, Princeton University, Spring 2003.

[Rai] R. A. Raimi, *The first digit problem*, Amer. Math. Monthly **83** (1976), no. 7, 521–538.

[Ra] K. Ramachandra, *Lectures on Transcendental Numbers*, Ramanujan Institute, Madras, 1969.

[Re] F. Reif, *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill, New York, 1965.

[Ric] P. Richter, *An investigation of expanders and ramanujan graphs along random walks of cubic bipartite graphs*, Junior Thesis, Princeton University, Spring 2001.

[RDM] R. D. Richtmyer, M. Devaney, and N. Metropolis, *Continued fraction of algebraic numbers*, Numer. Math. **4** (1962), 68–84.

[Rie] H. J. J. te Riele, *On the sign of the difference* $\pi(x) - \mathrm{Li}(x)$, Mathematics of Computation **48** (1987), no. 177, 323–328.

[Ri] G. F. B. Riemann, *Über die Anzahl der Primzahlen unter einer gegebenen Grösse*, Monatsber. Königl. Preuss. Akad. Wiss. Berlin, Nov. 1859, 671–680 (see [Ed] for an English translation).

[RSA] R. Rivest, A. Shamir, and L. Adleman, *A method for obtaining digital signatures and public key cryptosystems*, Comm. ACM **21** (1978), 120–126.

[Roc] D. Rockmore, *Stalking the Riemann Hypothesis: The Quest to Find the Hidden Law of Prime Numbers*, Pantheon, New York, 2005.

[Ro] K. Roth, *Rational approximations to algebraic numbers*, Mathematika **2** (1955), 1–20.

[Rub1] M. Rubinstein, *A simple heuristic proof of Hardy and Littlewood's conjecture B*, Amer. Math. Monthly **100** (1993), no. 5, 456–460.

[Rub2] M. Rubinstein, *Low-lying zeros of L-functions and random matrix theory*, Duke Math. J. **109** (2001), no. 1, 147–181.

[RubSa] M. Rubinstein and P. Sarnak, *Chebyshev's bias*, Experiment. Math. **3** (1994), no. 3, 173–197.

[Rud] W. Rudin, *Principles of Mathematical Analysis*, 3rd edition, International Series in Pure and Applied Mathematics, McGraw-Hill, New York, 1976.

[RS] Z. Rudnick and P. Sarnak, *Zeros of principal L-functions and random matrix theory*, Duke J. of Math. **81** (1996), 269–322.

[RS2] Z. Rudnick and P. Sarnak, *The pair correlation function of fractional parts of polynomials*, Comm. Math. Phys. **194** (1998), no. 1, 61–70.

[RSZ] Z. Rudnick, P. Sarnak, and A. Zaharescu, *The distribution of spacings between the fractional parts of* $n^2\alpha$, Invent. Math. **145** (2001), no. 1, 37–57.

[RZ1] Z. Rudnick and A. Zaharescu, *A metric result on the pair correlation of fractional parts of sequences*, Acta Arith. **89** (1999), no. 3, 283–293.

[RZ2] Z. Rudnick and A. Zaharescu, *The distribution of spacings between fractional parts of lacunary sequences*, Forum Math. **14** (2002), no. 5, 691–712.

[Sar] P. Sarnak *Some applications of modular forms*, Cambridge Trusts in Mathemetics, Vol. 99, Cambridge University Press, Cambridge, 1990.

[Sch] D. Schmidt, *Prime Spacing and the Hardy-Littlewood Conjecture B*, Junior Thesis, Princeton University, Spring 2001.

[Sc] P. Schumer, *Mathematical Journeys*, Wiley-Interscience, John Wiley & Sons, New York, 2004.

[Se] J. P. Serre, *A Course in Arithmetic*, Springer-Verlag, New York, 1996.

[Sh] A. Shidlovskii, *Transcendental Numbers*, Walter de Gruyter & Co., New York, 1989.

[ShTa] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, AMS, Providence, RI, 1943.

[Sil1] J. Silverman, *The Arithmetic of Elliptic Curves*, Graduate Texts in Mathematics, Vol. 106, Springer-Verlag, New York, 1986.

[Sil2] J. Silverman, *A Friendly Introduction to Number Theory*, 2nd edition, Prentice-Hall, Englewood Cliffs, NJ, 2001.

[ST] J. Silverman and J. Tate, *Rational Points on Elliptic Curves*, Springer-Verlag, New York, 1992.

[Si] B. Simon, *The classical moment problem as a self-adjoint finite difference operator*, Adv. Math. **137** (1998), no. 1, 82–203.

[SM] S. Simon and A. Moustakas, *Eigenvalue density of correlated complex random Wishart matrices*, Bell Labs Technical Memo, 2004.

[Sk] S. Skewes, *On the difference* $\pi(x) - \mathrm{Li}(x)$, J. London Math. Soc. **8** (1933), 277–283.

[Sl] N. Sloane, *On-Line Encyclopedia of Integer Sequences*, http://www.research.att.com/∼njas/sequences/Seis.html

[Sn] N. Snaith, *Derivatives of random matrix characteristic polynomials with applications to elliptic curves*, J. Phys. A **38** (2005), no. 48, 10345–10360.

[SS1] E. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton University Press, Princeton, NJ, 2003.

[SS2] E. Stein and R. Shakarchi, *Complex Analysis*, Princeton University Press, Princeton, NJ, 2003.

[SS3] E. Stein and R. Shakarchi, *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*, Princeton University Press, Princeton, NJ, 2005.

[StTa] I. Stewart and D. Tall, *Algebraic Number Theory*, 2nd edition, Chapman & Hall, London, 1987.

[St] Strang, *Linear Algebra and Its Applications*, 3rd edition, Wellesley-Cambridge Press, Wellesley, MA 1998.

[Str] K. Stromberg, *The Banach-Tarski paradox*, Amer. Math. Monthly **86** (1979), no. 3, 151–161.

[Sz] P. Szüsz, *On the length of continued fractions representing a rational number with given denominator*, Acta Arithmetica **37** (1980), 55–59.

[Ta] C. Taylor, *The Gamma function and Kuzmin's theorem*, Junior Thesis, Princeton University, Fall 2002.

[TW] R. Taylor and A. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. Math. **141** (1995), 553–572.

[TrWi] C. Tracy and H. Widom, *Correlation functions, cluster functions, and spacing distributions for random matrices*, J. Statist. Phys. **92** (1998), no. 5–6, 809–835.

[Te] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, Cambridge University Press, Cambridge, 1995.

[Ti] E. C. Titchmarsh, *The Theory of the Riemann Zeta-function*, revised by D. R. Heath-Brown, Oxford University Press, Oxford, 1986.

[Va] R. C. Vaughan, *On a variance associated with the distribution of primes in arithmetic progression*, Proc. London Math. Soc. (3) **82** (2001), 533–553.

[VW] R. C. Vaughan and T. D. Wooley, *Waring's problem: a survey*. Pages 301–340 in *Number Theory for the Millennium, III (Urbana, IL, 2000)*, A. K. Peters, Natick, MA, 2002.

[Vin1] I. Vinogradov, *Representation of an odd number as the sum of three primes*, Doklady Akad. Nauk SSSR **15** (1937), no. 6–7, 291–294.

[Vin2] I. Vinogradov, *Some theorems concerning the theory of primes*, Mat. Sbornik **2** (1937), no. 44, 179–195.

[Vo] A. Voros, *A sharpening of Li's criterion for the Riemann hypothesis*, preprint.

[VG] W. Voxman and R. Goetschel, Jr., *Advanced Calculus*, Mercer Dekker, New York, 1981.

[Wa] L. Washington, *Elliptic Curves: Number Theory and Cryptography*, Chapman & Hall / CRC, New York, 2003.

[Wed] S. Wedeniwski, *ZetaGrid*, http://www.zetagrid.net

[Wei1] A. Weil, *Numbers of Solutions of Equations in Finite Fields*, Bull. Amer. Math. Soc. **14** (1949), 497–508.

[Wei2] A. Weil, *Prehistory of the zeta-function*. Pages 1–9 in *Number Theory, Trace Formulas and Discrete Groups (Oslo, 1987)*, Academic Press, Boston, 1989.

[Weir] B. Weir, *The local behavior of Germain primes*, Undergraduate Mathematics Laboratory report, Courant Institute, NYU, 2002.

[We] E. Weisstein, *MathWorld — A Wolfram Web Resource*, http://mathworld.wolfram.com

[Weyl] H. Weyl, *The Classical Groups: Their Invariants and Representations*, Princeton University Press, Princeton, NJ, 1946.

[Wh] E. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies: With an Introduction to the Problem of Three Bodies*, Dover, New York, 1944.

[WW] E. Whittaker and G. Watson, *A Course of Modern Analysis*, 4th edition, Cambridge University Press, Cambridge, 1996.

[Wig1] E. Wigner, *On the statistical distribution of the widths and spacings of nuclear resonance levels*, Proc. Cambridge Philo. Soc. **47** (1951), 790–798.

[Wig2] E. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. of Math. **2** (1955), no. 62, 548–564.

[Wig3] E. Wigner, *Statistical Properties of real symmetric matrices*. Pages 174–184 in *Canadian Mathematical Congress Proceedings*, University of Toronto Press, Toronto, 1957.

[Wig4] E. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions. II*, Ann. of Math. Ser. 2 **65** (1957), 203–207.

[Wig5] E. Wigner, *On the distribution of the roots of certain symmetric matrices*, Ann. of Math. Ser. 2 **67** (1958), 325–327.

[Wi] A. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. Math. **141** (1995), 443–551.

[Wilf] H. Wilf, *Algorithms and Complexity*, 2nd edition, A. K. Peters, Natick, MA, 2002.

[Wir] E. Wirsing, *On the theorem of Gauss-Kuzmin-Lévy and a Frobenius-type theorem for function spaces*, Acta Arith. **24** (1974) 507–528.

[Wis] J. Wishart, *The generalized product moment distribution in samples from a normal multivariate population*, Biometrika **20 A** (1928), 32–52.

[Wor] N. C. Wormald, *Models of random regular graphs*. Pages 239–298 in *Surveys in combinatorics, 1999 (Canterbury)* London Mathematical Society Lecture Note Series, vol. 267, Cambridge University Press, Cambridge, 1999.

[Wo] T. Wooley, *Large improvements in Waring's problem*, Ann. of Math. (2), **135** (1992), no. 1, 131–164.

[Za] I. Zakharevich, *A generalization of Wigner's law*, preprint.

[Zu] W. Zudilin, *One of the numbers* $\zeta(5), \zeta(7), \zeta(9), \zeta(11)$ *is irrational*, Uspekhi Mat. Nauk **56** (2001), 149-150.

[Zy] A. Zygmund, *Trigonometrical Series*, vols. I and II, Cambridge University Press, Cambridge, 1968.

Index