Junior Research Seminar: Probability, $L$-Functions, Random Matrix Theory and Ramanujan Graphs

Yakov Sinai$^3$, Steven Miller$^4$, Lior Silberman$^5$

May 4, 2003

$^1$Homepage: http://www.math.princeton.edu/~mathlab/
$^2$Wednesdays 8:15 – 9:45am Fine 314 and Thursday 8 – 8:50pm Fine 214
$^3$E-mail: sinai@math.princeton.edu
$^4$E-mail:sjmiller@math.princeton.edu
$^5$E-mail:lior@math.princeton.edu
## Contents

1 Spacings of Eigenvalues of Real Symmetric Matrices; Semi-Circle Law 7
   1.1 Joint density function of eigenvalues of real symmetric matrices ('GOE') 7
   1.1.1 Dirac Notation 7
   1.1.2 $2 \times 2$ Gaussian Orthogonal Ensemble (GOE) 8
   1.1.3 Transformation to diagonal representation 10
   1.1.4 Generalization to $n \times n$ case 12
   1.2 Eigenvalue spacing distribution in $2 \times 2$ real symmetric matrices 13
   1.2.1 Reminder: Integral of the Gaussian 13
   1.2.2 Spacing distribution 14
   1.3 Delta Function(al) 15
   1.4 Definition of the Semi-Circle Density 16
   1.5 Semi-Circle Rule: Preliminaries 16
   1.6 Sketch of Proof of the Semi-Circle Law 17
      1.6.1 Calculation of Moments via Trace Formula 17
      1.6.2 Calculation of Moments from the Semi-Circle 20

2 Random Matrix Theory and the Mehta-Gaudin Theorem 22
   2.1 Gaudin’s Method 22
      2.1.1 Introduction 22
      2.1.2 Vandermonde Determinants 23
      2.1.3 Orthonormal Polynomials 24
      2.1.4 Kernel $K_N(x, y)$ 25
      2.1.5 Gaudin-Mehta Theorem 25
      2.1.6 Example 27
3 Rebecca C. Lehman’s JP: First-Order Spacings of Random Matrix Eigenvalues

3.1 A Little Motivation ................................................. 30
  3.1.1 Classical Theory ............................................ 30
  3.1.2 Quantum mechanics and Random Matrix Theory ....... 32
3.2 Facts about Random Matrix Eigenvalues .................... 33
  3.2.1 Joint Probability Density ................................. 33
  3.2.2 Semicircle Law ............................................ 34
3.3 Gaussian .......................................................... 37
  3.3.1 Local Spacings ............................................. 39
3.4 Computations ..................................................... 42
3.5 Results ............................................................. 43
3.6 Bibliography ....................................................... 44

4 Yi-Kai Liu’s JP: Statistical Behavior of the Eigenvalues of Random Matrices

4.1 Introduction ....................................................... 45
4.2 The Theory of Random Matrices .............................. 46
  4.2.1 Application to Nuclear Physics ......................... 46
  4.2.2 The Semicircle Rule ....................................... 47
  4.2.3 The Gaussian Orthogonal Ensemble ..................... 49
  4.2.4 Probability Distribution of the Eigenvalues .......... 53
  4.2.5 Eigenvalue Spacings and the Wigner Surmise ........ 56
4.3 Computer Experiments ......................................... 58
  4.3.1 Notes on Implementation .................................. 58
  4.3.2 Sparse Matrices .......................................... 59
  4.3.3 Band-Diagonal Matrices ................................. 60
  4.3.4 Conclusions ............................................... 60
4.4 References ....................................................... 60

5 Graphs .................................................................. 62
5.1 Definition .......................................................... 62
5.2 Counting Games .................................................. 64
5.3 Adjacency Matrix and Laplacian .............................. 65
5.4 Spectrum ........................................................... 66
  5.4.1 1 an Eigenvalue ............................................ 66
  5.4.2 −1 an Eigenvalue ......................................... 66

2
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Graphs - II</td>
<td>68</td>
</tr>
<tr>
<td>6.1</td>
<td>Problems from Professor Sinai</td>
<td>68</td>
</tr>
<tr>
<td>6.2</td>
<td>Connectivity Problem</td>
<td>68</td>
</tr>
<tr>
<td>7</td>
<td>Graphs III</td>
<td>71</td>
</tr>
<tr>
<td>7.1</td>
<td>Review</td>
<td>71</td>
</tr>
<tr>
<td>7.2</td>
<td>First Problem</td>
<td>71</td>
</tr>
<tr>
<td>7.3</td>
<td>Model of Random Regular Graphs on $n$ Vertices</td>
<td>72</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Definition of the 1-Point Function (Density of States)</td>
<td>73</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Calculating the 1-Point Function for Random Regular Graphs</td>
<td>74</td>
</tr>
<tr>
<td>7.4</td>
<td>Comments on Random Graphs (Sinai)</td>
<td>76</td>
</tr>
<tr>
<td>8</td>
<td>Introduction to Dynamics Problems</td>
<td>78</td>
</tr>
<tr>
<td>8.1</td>
<td>Introduction</td>
<td>78</td>
</tr>
<tr>
<td>8.2</td>
<td>Poincare Recurrence Theorem</td>
<td>79</td>
</tr>
<tr>
<td>8.3</td>
<td>Interval Exchange</td>
<td>79</td>
</tr>
<tr>
<td>9</td>
<td>More on Random Graphs</td>
<td>81</td>
</tr>
<tr>
<td>9.1</td>
<td>$n(G)$</td>
<td>81</td>
</tr>
<tr>
<td>9.2</td>
<td>More on Traces</td>
<td>81</td>
</tr>
<tr>
<td>10</td>
<td>Dynamical Systems</td>
<td>82</td>
</tr>
<tr>
<td>10.1</td>
<td>Problem of the Adiabatic Piston</td>
<td>82</td>
</tr>
<tr>
<td>10.2</td>
<td>Probability Distribution on Particles</td>
<td>83</td>
</tr>
<tr>
<td>10.2.1</td>
<td>Probability Distribution for Coordinates</td>
<td>83</td>
</tr>
<tr>
<td>10.2.2</td>
<td>Probability Distribution for Velocities</td>
<td>83</td>
</tr>
<tr>
<td>10.3</td>
<td>Bouncing Ball and Moving Wall</td>
<td>85</td>
</tr>
<tr>
<td>11</td>
<td>More on Dynamical Piston</td>
<td>89</td>
</tr>
<tr>
<td>11.1</td>
<td>Review</td>
<td>89</td>
</tr>
<tr>
<td>11.2</td>
<td>Solution</td>
<td>90</td>
</tr>
<tr>
<td>11.3</td>
<td>Generalizations</td>
<td>90</td>
</tr>
<tr>
<td>11.4</td>
<td>Three Particles</td>
<td>91</td>
</tr>
<tr>
<td>12</td>
<td>$3x + 1$ Problem</td>
<td>92</td>
</tr>
<tr>
<td>12.1</td>
<td>Introduction</td>
<td>92</td>
</tr>
<tr>
<td>12.2</td>
<td>Statistical $3x + 1$ Problem</td>
<td>93</td>
</tr>
</tbody>
</table>
13 More on $3x + 1$

13.1 Review .................................................. 95
13.2 Structure Theorem ....................................... 95
13.3 Density Arguments ...................................... 97
13.4 Moivre-Laplace Theorem .............................. 97
13.5 Proof of the Structure Theorem ...................... 99

A Linear Algebra Review 100

A.1 Transposition ........................................... 100
A.2 Matrix Multiplication .................................. 101
A.3 Transformations of a Matrix ......................... 102
A.4 Orthogonal Matrices ................................ 102
A.5 Definitions ............................................. 104
A.6 Trace ....................................................... 105
A.7 Spectral Theorem for Real Symmetric Matrices: Easy Case . 106
  A.7.1 Preliminary Lemmas ................................ 107
  A.7.2 Proof of the Spectral Theorem: Easy Case ....... 108
A.8 Spectral Theorem for Real Symmetric Matrices: Harder Case . 110

B Fourier Analysis and the Equi-Distribution of $\{n\alpha\}$ 112

B.1 Inner Product of Functions ......................... 112
B.2 Fourier Series and $\{n\alpha\}$ ....................... 114
  B.2.1 Fourier Series .................................... 114
  B.2.2 Weighted partial sums ......................... 115
  B.2.3 Equidistribution ................................ 118

C Introduction to Probability 121

C.1 Probabilities of Discrete Events ................. 121
  C.1.1 Introduction ..................................... 121
  C.1.2 Conditional Probabilities .................... 122
  C.1.3 Independent Events ............................ 124
  C.1.4 Means ............................................ 126
  C.1.5 Variances ....................................... 129
  C.1.6 Disjoint Events ................................ 132
C.2 Standard Processes ................................. 132
  C.2.1 Bernoulli Process .............................. 132
  C.2.2 Poisson Distribution ......................... 134
  C.2.3 Continuous Poisson Distribution ............ 135
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>F.1</td>
<td>Germain Primes</td>
<td>175</td>
</tr>
<tr>
<td>F.2</td>
<td>Preliminaries</td>
<td>177</td>
</tr>
<tr>
<td>F.2.1</td>
<td>Definitions</td>
<td>177</td>
</tr>
<tr>
<td>F.2.2</td>
<td>Partial Summation</td>
<td>177</td>
</tr>
<tr>
<td>F.2.3</td>
<td>Siegel-Walfisz</td>
<td>178</td>
</tr>
<tr>
<td>F.2.4</td>
<td>Germain Integral</td>
<td>178</td>
</tr>
<tr>
<td>F.2.5</td>
<td>Major and Minor Arcs</td>
<td>179</td>
</tr>
<tr>
<td>F.2.6</td>
<td>Formula for Number of Germain Prime Pairs</td>
<td>180</td>
</tr>
<tr>
<td>F.2.7</td>
<td>Reformulation of Germain Integral</td>
<td>181</td>
</tr>
<tr>
<td>F.3</td>
<td>$f_N(x)$ and $u(x)$</td>
<td>182</td>
</tr>
<tr>
<td>F.3.1</td>
<td>$f \left( \frac{a}{q} \right)$</td>
<td>182</td>
</tr>
<tr>
<td>F.3.2</td>
<td>$u(x)$</td>
<td>183</td>
</tr>
<tr>
<td>F.4</td>
<td>$f_N(\alpha) - \frac{c_q(a)c_q(-2a)}{\phi^2(q)} u(\alpha - \frac{a}{q}), \alpha \in \mathcal{M}_{a,q}$</td>
<td>184</td>
</tr>
<tr>
<td>F.4.1</td>
<td>Setup</td>
<td>184</td>
</tr>
<tr>
<td>F.4.2</td>
<td>$S_1 \sum_{\alpha, q}$</td>
<td>185</td>
</tr>
<tr>
<td>F.4.3</td>
<td>$S_1 f_{\alpha, q}$</td>
<td>187</td>
</tr>
<tr>
<td>F.5</td>
<td>Integrals of $u(x)$</td>
<td>188</td>
</tr>
<tr>
<td>F.5.1</td>
<td>Formulations</td>
<td>188</td>
</tr>
<tr>
<td>F.5.2</td>
<td>$\int_{-\frac{1}{2}}^{\frac{1}{2}} u(x)e(-x)dx$</td>
<td>189</td>
</tr>
<tr>
<td>F.5.3</td>
<td>$\int_{-\frac{2}{q}}^{\frac{2}{q}} u(x)e(-x)dx + \int_{\frac{1}{2}}^{\frac{1}{2}} u(x)e(-x)dx$</td>
<td>189</td>
</tr>
<tr>
<td>F.5.4</td>
<td>Integral over $I_2, I_3$</td>
<td>190</td>
</tr>
<tr>
<td>F.5.5</td>
<td>Integral over $I_1, I_4$</td>
<td>190</td>
</tr>
<tr>
<td>F.5.6</td>
<td>Collecting the Pieces</td>
<td>193</td>
</tr>
<tr>
<td>F.6</td>
<td>Determination of the Main Term</td>
<td>193</td>
</tr>
<tr>
<td>F.6.1</td>
<td>Properties of $C_q(a)$ and $\rho_q$</td>
<td>195</td>
</tr>
<tr>
<td>F.6.2</td>
<td>Determination of $\varnothing_N$ and $\bar{\varnothing}$</td>
<td>201</td>
</tr>
<tr>
<td>F.6.3</td>
<td>Number of Germain Primes and Weighted Sums</td>
<td>202</td>
</tr>
</tbody>
</table>
Chapter 1

Spacings of Eigenvalues of Real Symmetric Matrices; Semi-Circle Law

Joint Density Function for eigenvalues of real symmetric matrices; spacing of eigenvalues for $2 \times 2$ real symmetric matrices; Semi-Circle Rule. Lecture by Steven J. Miller; notes by Steven J. Miller and Alex Barnett.

1.1 Joint density function of eigenvalues of real symmetric matrices (‘GOE’)

1.1.1 Dirac Notation

The derivation handed out in lecture used physics notation which should be explained. The matrix is called the ‘Hamiltonian’ (meaning that it happened to arise in a quantum physics problem). Vectors are often called states (referring to quantum states), however they can be thought of as your usual vectors. (Quantum mechanics is just linear algebra, amazingly). A general vector in 2D is written

$$|u\rangle \text{ equivalent to } u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix},$$

the latter being its coordinate representation in some basis. The unit vectors are

$$|1\rangle, |2\rangle \text{ equivalent to } \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$
The $|u\rangle$ is a column vector, and $\langle u| \equiv |u\rangle^T$ is a row vector. Inner product can be written as $\langle v|u\rangle = v^T \cdot u$. General bilinear product can be written $\langle v|M|u\rangle = v^T \cdot M \cdot u$.

1.1.2 2 × 2 Gaussian Orthogonal Ensemble (GOE)

We consider 2 × 2 real symmetric matrices,

$$A \equiv \begin{pmatrix} x & y \\ y & z \end{pmatrix}. \quad (1.3)$$

Understanding this case is vital to building intuition about Random Matrix Theory for $N \times N$ matrices.

$A$ can always be diagonalized by an orthogonal matrix $Q$ as follows,

$$Q^T \begin{pmatrix} x & y \\ y & z \end{pmatrix} Q = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \equiv D. \quad (1.4)$$

In 2 × 2 case, the characteristic equation $\det(A - \lambda I) = 0$ is quadratic:

$$\lambda^2 - \text{Tr}(A)\lambda + \det(A) = 0, \quad (1.5)$$

where

$$\text{Tr}(A) = x + z, \quad \det(A) = xz - y^2. \quad (1.6)$$

Solutions are

$$\lambda_{1,2} = \frac{x + z}{2} \pm \sqrt{\left(\frac{x - z}{2}\right)^2 + y^2}, \quad (1.7)$$

where 1 is the + case, 2 the −.

If the two eigenvalues are equal, we say the matrix is degenerate. Initially we are in a three-dimensional space (as $x$, $y$ and $z$ are arbitrary). Degeneracy requires that $x$, $y$ and $z$ satisfy

$$\left(\frac{x - z}{2}\right)^2 + y^2 = 0, \quad (1.8)$$

or, equivalently,

$$x - z = 0, \quad y = 0. \quad (1.9)$$

Thus, we lose two degrees of freedom, because there are two equations which must be satisfied. The set of solutions is $\{(x, y, z) = (x, 0, x)\}$.
Exercise 1.1.1. Show that $\lambda_1 - \lambda_2$ is twice the distance from the origin in this 2D subspace.

Corresponding eigenvectors are,

$$v_1 = \begin{pmatrix} c \\ s \end{pmatrix}, \quad v_2 = \begin{pmatrix} -s \\ c \end{pmatrix}. \tag{1.10}$$

We use abbreviations $c \equiv \cos \theta$ and $s \equiv \sin \theta$.

Why can we write the eigenvectors as above? We can always normalize the eigenvector attached to a given eigenvalue to have length 1. We have previously shown that, if the eigenvalues are distinct, then the eigenvectors of a real symmetric matrix are perpendicular. This forces the above form for the two eigenvectors, at least when $\lambda_1 \neq \lambda_2$.

One rotation angle $\theta$ defines the orthogonal matrix,

$$Q = Q(\theta) = \left( v_1 \ v_2 \right) = \begin{pmatrix} c & -s \\ s & c \end{pmatrix}. \tag{1.11}$$

The structure of the eigenvectors is actually quite rich.

Exercise 1.1.2. Find $\theta$ in terms of $x, y, z$. Hint: use trigonometric identities to simplify the resulting form. Hint: solve $(A - \lambda_1 v_1) = 0$.

Exercise 1.1.3. Show that a general $A$ can be written

$$A = \alpha \begin{pmatrix} \cos \beta & \sin \beta \\ \sin \beta & -\cos \beta \end{pmatrix} + \gamma \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{1.12}$$

Exercise 1.1.4. Find $\lambda_{1,2}$ in terms of $\alpha, \beta, \gamma$. Show that the eigenvector angle is given by $\theta = \beta/2$. This result is quite deep: for instance notice that taking a complete $2\pi$ cycle in $\beta$ reverses the signs of the eigenvectors! This isn’t that relevant for the rest of this lecture.

We adopt two assumptions about the joint distribution over $A$, called $p(A) \equiv p(x, y, z)$:

1. Invariance of $p$ under orthogonal transformations (aka ‘basis-invariance’), $p(M^T A M) = p(A)$ for all orthogonal $M$.

2. Independence of distributions of individual matrix elements, $p(x, y, z) = p_x(x)p_y(y)p_z(z)$. 

9
Section 2C-1 of the handout reminds us that these two assumptions taken together demand a unique form of distribution,

\[ p(x, y, z) \propto e^{-C \text{Tr}(A^2)}, \quad (1.13) \]

depending on only one parameter \( C \); we choose \( C = 1 \). Note \( \propto \) means proportional to; the constant of proportionality is what is needed to make \( P(x, y, z) \) a probability distribution (ie, the integral of \( \int \int \int p(x, y, z)dx dy dz = 1 \)).

This corresponds to Gaussian distributions of matrix elements,

\[ p_y(y) = \sqrt{\frac{2}{\pi}} e^{-2y^2} \quad \text{off-diag} \]
\[ p_x(x) = p_z(x) = \frac{1}{\sqrt{\pi}} e^{-x^2} \quad \text{diag} . \quad (1.14) \]

Note that the diag elements have variance \( \frac{1}{2} \), the off-diag variance \( \frac{1}{4} \). We show how to compute the normalization prefactors later on. This form (for general \( C \)) is the so-called GOE. The \( n \times n \) case is derived in Miller’s handout of 9/25/02.

### 1.1.3 Transformation to diagonal representation

The operation of diagonalizing \( A \) can be viewed as the transformation from one 3D space to another 3D space,

\[ \mathbf{r} \equiv (x, y, z) \quad \langle A \rangle \quad \longleftrightarrow \quad \mathbf{r}' \equiv (\lambda_1, \lambda_2, \theta) \quad \langle D, Q \rangle . \quad (1.15) \]

This is 1-to-1 apart from the set of measure zero (ie, a lower dimensional subspace) corresponding to degenerate eigenvalues. Looking at Eq. 1.4 we can see the transformation is linear in the eigenvalues, nonlinear in \( \theta \). We are interested in the marginal distribution of the eigenvalues,

\[ p' (\lambda_1, \lambda_2) \equiv \int d\theta \, p'(\lambda_1, \lambda_2, \theta), \quad (1.16) \]
in other words we don’t care what \( \theta \) is. We use primes to signify distributions over final \((Q, D)\) variables.

We want to know how to transform probability density from \( \mathbf{r} \) space to \( \mathbf{r}' \) space. In general this must follow the law,

\[ p(\mathbf{r}) d\mathbf{r} = p'(\mathbf{r'}) d\mathbf{r'}, \quad (1.17) \]
giving

$$p'(r') = \det(J)p(r). \quad (1.18)$$

The ratio of the volume elements is $|\det J|$ where $J$ is the $3 \times 3$ Jacobian matrix of the transformation. $J$ has elements $J_{ij} = \partial r_j / \partial r_i'$.

Inverting Eq. 1.4 we can write $A(r')$ as

$$\begin{pmatrix} x & y & z \\ y & z & \end{pmatrix} = QDQ^T = \begin{pmatrix} c & -s & 0 \\ s & c & 0 \\ \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \\ \end{pmatrix} \begin{pmatrix} c & s \\ -s & c \\ \end{pmatrix} = \begin{pmatrix} \lambda_1 c^2 + \lambda_2 s^2 & (\lambda_1 - \lambda_2)sc \\ (\lambda_1 - \lambda_2)sc & \lambda_1 s^2 + \lambda_2 c^2 \\ \end{pmatrix} \quad (1.19)$$

We evaluate $J$ for this case,

$$J \equiv \begin{pmatrix} \frac{\partial x}{\partial \lambda_1} & \frac{\partial y}{\partial \lambda_1} & \frac{\partial z}{\partial \lambda_1} \\ \frac{\partial x}{\partial \lambda_2} & \frac{\partial y}{\partial \lambda_2} & \frac{\partial z}{\partial \lambda_2} \\ \end{pmatrix}. \quad (1.20)$$

We see $\lambda$’s only appear in the bottom three entries, and furthermore they only appear as factors $(\lambda_1 - \lambda_2)$ in each entry.

**Exercise 1.1.5.** *Evaluate the bottom row of $J$ to prove the above.*

Therefore this factor of a row of $J$ can be brought out in evaluating the determinant:

$$\det(J) = \begin{vmatrix} \text{messy } \theta \text{-dep } 3 \times 3 \text{ matrix} \end{vmatrix} \cdot (\lambda_1 - \lambda_2) = g(\theta)(\lambda_1 - \lambda_2). \quad (1.21)$$

**Warning!** The Jacobian is the absolute value of the determinant. Thus, we need $|\lambda_1 - \lambda_2|$ above, or we need to adopt the convention that we label the eigenvalues so that $\lambda_1 \geq \lambda_2$.

The only dependence on the $\lambda$’s is given by the second factor. Plugging into Eq. 1.18 and marginalizing over $\theta$ gives,

$$p'(\lambda_1, \lambda_2) = \int d\theta g(\theta) (\lambda_1 - \lambda_2)e^{-(\lambda_1^2 + \lambda_2^2)} \propto (\lambda_1 - \lambda_2)e^{-(\lambda_1^2 + \lambda_2^2)}. \quad (1.22)$$

Note that we do not need the absolute value sign around $(\lambda_1 - \lambda_2)$ because we chose $\lambda_1 > \lambda_2$. This is the joint density of the eigenvalues in $2 \times 2$ GOE.
1.1.4 Generalization to $n \times n$ case

The above generalizes quite easily, with the dimension of the two spaces being $N = \frac{1}{2}n(n + 1)$. The $\frac{1}{2}n(n + 1)$ degrees of freedom in $A$ equal $n$ degrees of freedom in $D$ (namely the eigenvalues $\{\lambda_i\}$) plus $\frac{1}{2}n(n - 1)$ degrees of freedom in $Q$ (namely the generalized angles $\Omega$). We sort the eigenvalues such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

**Theorem 1.1.6.** If $\lambda_i = \lambda_j$ for some $1 \leq i < j \leq n$, then the Jacobean of the transformation $A \leftrightarrow (D, Q)$ vanishes, that is $\det(J) = 0$.

The proof relies on realising that the two eigenvectors $v_i$ and $v_j$ span a 2D subspace, invariant under $A$. (Recall here we refer to a subspace of the $n$-dim vector space upon which $A$ operates by multiplication). The invariance means that the choices of directions of the eigenvectors is arbitrary in this 2D plane. Therefore there is one angle degree of freedom in $\Omega$ which in not constrained by $A$, that is, it is independent of the elements of $A$. Now think of the inverse transformation from $(D, Q) \rightarrow A$. An infinitesimal volume element is transformed as

$$dr = \det(J)dr'.$$

(1.23)

Changes of eigenvector angle within the 2D subspace have no effect on $A$, so the volume element $dr$ is collapsed to zero. (Another way of putting this is that $J$ acquires a null-space of dimension 1). Therefore $\det(J) = 0$. □

This vanishing of the Jacobean at degeneracies renders the non-uniqueness of the forward map $A \rightarrow (D, Q)$ at these points harmless in the following.

The upper $n$ rows of $J$ are messy functions of angles $\Omega$, and the bottom $\frac{1}{2}n(n - 1)$ rows contain entries each which is linear in the eigenvalues. Therefore $\det(J)$ is a polynomial of degree $\frac{1}{2}n(n - 1)$ in the eigenvalues $\lambda_i$. Further, $\det(J) = 0$ if any two eigenvalues are equal.

Consider the polynomial $\prod_{1 \leq i < j \leq n}(\lambda_i - \lambda_j)$. First, note that this polynomial vanishes whenever two eigenvalues are the same. We claim it is a polynomial of degree $\frac{1}{2}n(n - 1)$ in the eigenvalues. For each $j$, there are $j - 1$ choices for $i$. Thus, the degree is

$$\sum_{j=2}^{n} j - 1 = \sum_{k=1}^{n-1} k = \frac{(n - 1)(n - 1 + 1)}{2} = \frac{n(n - 1)}{2}. \quad (1.24)$$

Thus, $\det(J)$ and $\prod_{1 \leq i < j \leq n}(\lambda_i - \lambda_j)$ both vanish whenever two eigenvalues are equal, and they have the same degree. Therefore, they must be scalar multiples of each other.
So,

\[ \det(J) \propto \prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j). \]  

(1.25)

Combining with the GOE form of \( p(A) \) gives, after marginalizing over \( \Omega \) as before,

\[ p(\{\lambda_i\}) = \prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j) \cdot e^{-\sum_{i=1}^{n} \lambda_i^2}. \]  

(1.26)

The vanishing of this probability density as any two eigenvalues come close is called \textit{level repulsion}.

1.2 Eigenvalue spacing distribution in \( 2 \times 2 \) real symmetric matrices

1.2.1 Reminder: Integral of the Gaussian

We want

\[ I = \int_{-\infty}^{\infty} e^{-x^2} dx. \]  

(1.27)

Square it and rearrange the summation over area by using polar coordinates:

\[ I^2 = \int_{-\infty}^{\infty} e^{-x^2} dx \cdot \int_{-\infty}^{\infty} e^{-y^2} dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-x^2-y^2} dx dy \]

\[ = \int_{0}^{2\pi} d\theta \int_{0}^{\infty} r dr e^{-r^2} = 2\pi \cdot \left[ \frac{1}{2} e^{-r^2} \right]_0^{\infty} = \pi. \]  

(1.28)

Introduction of the radius factor \( r \) produced \( re^{-r^2} \), a known differential. So,

\[ I = \sqrt{\pi}. \]  

(1.29)

Changing the variable in the above, and rearranging, gives

\[ \frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx = 1. \]  

(1.30)

This is therefore the correct normalization for a 1D Gaussian probability density, of variance \( \sigma^2 \).
1.2.2 Spacing distribution

Figure 1.1: Change of coordinates to get spacing distribution in $E$. Light shading suggests form of density across the 2D plane. Dark shading shows a graph of its projection onto the $E$ axis.

Here for convenience we present a slightly simpler derivation than in lecture. Given the 2D density $p'(\lambda_1, \lambda_2)$ we want the 1D density of the difference $E \equiv \lambda_1 - \lambda_2$. This will require marginalizing again, since there is a reduction in dimensionality. We define $S \equiv \lambda_1 + \lambda_2$. The linear transformation $(\lambda_1, \lambda_2) \to (E, S)$ has fixed Jacobian (it is a rotation by $-45^\circ$ and a compression by $\sqrt{2}$ in each axis). See Fig. 1.1.

Therefore, substituting in $\lambda_1 = (S + E)/2$ and $\lambda_1 = (S - E)/2$ into Eq. 1.22 gives

\[
p'(E, S) \propto p'(\lambda_1(E, S), \lambda_2(E, S)) = E e^{-\frac{1}{2}[(S+E)^2+(S-E)^2]} = E e^{-E^2/2} e^{-S^2/2},
\]

which is separable. Therefore integrating over $S$ gives an $E$-independent number, and

\[
p'(E) \propto E e^{-E^2/2}.
\]
This is the so-called ‘Wigner Surmise’ for the eigenvalue spacing density. Remarkably, in the $n \times n$ case, even for large $n$, this density is very close to the true spacing distribution of adjacent eigenvalues. The limiting powerlaw $\lim_{E \to 0} p'(E) \propto E^\beta$ with $\beta = 1$ is intimately related to the matrix symmetry class GOE that we close. It is also possible to achieve $\beta = 2$ and $\beta = 4$ by choosing different symmetry classes.

Finally, let’s say you couldn’t be bothered to construct your second variable $S$. Instead you could derive the above using the Dirac delta-function (see below) to marginalize:

$$p'(E) = \int_{-\infty}^{\infty} d\lambda_1 \int_{-\infty}^{\lambda_1} d\lambda_2 \, p'(\lambda_1, \lambda_2) \, \delta(E - (\lambda_1 - \lambda_2)).$$

(1.33)

Apart from the unusual limits (due to ordering of eigenvalues), this is the standard procedure to extract a marginal density.

**Exercise 1.2.1.** Simplify the above to arrive at $p'(E)$.

### 1.3 Delta Function(al)

Let $f(x)$ be a nice function; for example, let $f(x)$ be an infinitely differentiable function whose Taylor Series converges to $f(x)$:

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \cdots$$

(1.34)

Let

$$\delta_n(x) = \begin{cases} n & \text{if } x \in \left[-\frac{1}{2n}, \frac{1}{2n}\right] \\ 0 & \text{otherwise} \end{cases}$$

(1.35)

**Exercise 1.3.1.** Show that

$$\int_{-\infty}^{\infty} f(x)\delta_n(x)dx = f(0) + O\left(\frac{1}{n}\right).$$

(1.36)

Let $\delta$ be the limit as $n \to \infty$ of $\delta_n$. We find / define

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} f(x)\delta_n(x)dx = \int_{-\infty}^{\infty} f(x)\delta(x)dx = f(0).$$

(1.37)
Exercise 1.3.2. Show that
\[ \int_{-\infty}^{\infty} f(x) \delta(x - a) \, dx = f(a). \] (1.38)

A good analogy for the $\delta$ functional is a point mass. A point mass has no extension (no length, width or height) but finite mass. Therefore, a point mass has infinite density.

A probability density must integrate to one. This corresponds to $\int 1 \cdot \delta(x) \, dx = 1$. We often refer to $\delta(x)$ as a point mass at the origin, and $\delta(x - a)$ as a point mass at $a$.

1.4 Definition of the Semi-Circle Density

Consider
\[ P(x) = \begin{cases} \frac{2}{\pi} \sqrt{1 - x^2} & \text{if } |x| \leq 1 \\ 0 & \text{otherwise} \end{cases} \] (1.39)

Exercise 1.4.1. Show that $P(x)$ is a probability density. IE, show that it is non-negative and integrates to 1. Graph $P(x)$.

We call $P(x)$ the semi-circle density.

1.5 Semi-Circle Rule: Preliminaries

Let $\lambda_j$ be the eigenvalues of a real, symmetric $N \times N$ matrix $A$. We normalize the eigenvalues of $A$ by dividing by $2\sqrt{N}$.

Define
\[ \mu_{A,N}(x) = \frac{1}{N} \sum_{j=1}^{N} \delta \left( x - \frac{\lambda_j(A)}{2\sqrt{N}} \right). \] (1.40)

$\delta \left( x - \frac{\lambda_j(A)}{2\sqrt{N}} \right)$ is a point mass at $\frac{\lambda_j(A)}{2\sqrt{N}}$. By summing these point masses and dividing by $N$, we have a probability distribution. For example,
\[ \int_{-\infty}^{\infty} f(x) \mu_{A,N}(x) \, dx = \sum_{j=1}^{N} f \left( \frac{\lambda_j(A)}{2\sqrt{N}} \right). \] (1.41)
We will show that, as \( N \to \infty \), the above converges to the integral of \( f \) against the semi-circle density:

\[
\int_{-\infty}^{\infty} f(x) P(x) \, dx.
\] (1.42)

What does this mean?

\[
\sum_{j=1}^{N} f\left( \frac{\lambda_j(A)}{2\sqrt{N}} \right)
\] (1.43)

looks like a Riemann Sum. The statement that, for nice \( f(x) \),

\[
\sum_{j=1}^{N} f\left( \frac{\lambda_j(A)}{2\sqrt{N}} \right) \to \int_{-\infty}^{\infty} f(x) P(x) \, dx
\] (1.44)

means that as \( N \to \infty \), the number of eigenvalues of a random \( A \) in \([a, b]\) equals

\[
\int_{a}^{b} P(x) \, dx.
\] (1.45)

**Theorem 1.5.1.** Choose the entries \( a_{ij} \) of a real, symmetric matrix independently from a fixed probability distribution \( p \) with mean zero, variance one, and finite higher moments. For each \( A \), form the probability measure \( \mu_{A,N} \). As \( N \to \infty \), with probability one the measures \( \mu_{A,n}(x) \, dx \) converge to the semi-circle probability \( P(x) \, dx \).

This is not the most general version; however, it is rich enough for our purposes.

### 1.6 Sketch of Proof of the Semi-Circle Law

#### 1.6.1 Calculation of Moments via Trace Formula

We will show that the expected value of the moments of the \( \mu_{A,N}(x) \) equal the moments of the semi-circle.
Definition 1.6.1. $M_{A,N}(k)$ is the $k$th moment of the probability measure attached to $\mu_{A,N}(x)dx$:

$$M_{A,N}(k) = \int x^k \mu_{A,N}(x)dx = \frac{1}{N} \sum_{j=1}^{N} \left( \frac{\lambda_j(A)}{2\sqrt{N}} \right)^k. \quad (1.46)$$

Note that $\sum \lambda_j(A)^k = \text{Trace}(A^k)$. Thus, we have

$$M_{A,N}(k) = \frac{1}{2^k N^{1+k/2}} \text{Trace}(A^k). \quad (1.47)$$

We now calculate the expected values of the first few moments ($k = 0, 1, 2$ and $3$).

Lemma 1.6.2. The expected value of $M_{A,N}(0) = 1$.

Proof:

$$E[M_{A,N}(0)] = \frac{1}{N} E[\text{Trace}(I)] = 1. \quad (1.48)$$

Note that summing the eigenvalues to the zeroth power is the same as taking the trace of the identity matrix. □

Lemma 1.6.3. The expected value of $M_{A,N}(1) = 0$.

Proof:

$$E[M_{A,N}(1)] = \frac{1}{2N^{3/2}} E[\text{Trace}(A)] = \frac{1}{2N^{3/2}} E[\sum_i a_{ii}] = \frac{1}{2N^{3/2}} \sum_i E[a_{ii}] = 0, \quad (1.49)$$

because we have assumed that each $a_{ij}$ is drawn from a probability distribution with mean zero. □

Lemma 1.6.4. The expected value of $M_{A,N}(2) = \frac{1}{2}$. 18
Proof: Note that
\[ \text{Trace}(A^2) = \sum_i \sum_j a_{ij}a_{ji}. \]  
(1.50)

As our matrix is symmetric, \( a_{ij} = a_{ji} \). Thus, the trace is \( \sum_i \sum_j a_{ij}^2 \).

Now
\[
E\left[M_{A,N}(2)\right] = \frac{1}{4N^2} E\left[\text{Trace}(A^2)\right] \\
= \frac{1}{4N^2} E\left[\sum_i \sum_j a_{ij}^2\right] \\
= \frac{1}{4N^2} \sum_i \sum_j E[a_{ij}^2] = \frac{1}{4},
\]  
(1.51)

where the last line follows from each \( a_{ij} \) has variance 1. As their means are zero, the variance \( E[a_{ij}^2] - E[a_{ij}]^2 = 1 \) implies \( E[a_{ij}^2] = 1 \). There are \( N^2 \) pairs \( (i, j) \). Thus, we have \( \frac{1}{4N^2} \cdot (N \cdot 1) = \frac{1}{4} \). \( \square \)

**Lemma 1.6.5.** The expected value of \( M_{A,N}(3) = 0 \) as \( N \to \infty \).

We need
\[ \text{Trace}(A^3) = \sum_i \sum_j \sum_k a_{ij}a_{jk}a_{ki}. \]  
(1.52)

We find
\[
E\left[M_{A,N}(3)\right] = \frac{1}{8N^{2.5}} E\left[\text{Trace}(A^3)\right] \\
= \frac{1}{8N^{2.5}} E\left[\sum_i \sum_j \sum_k a_{ij}a_{jk}a_{ki}\right] \\
= \frac{1}{8N^{2.5}} \sum_i \sum_j \sum_k E[a_{ij}a_{jk}a_{ki}].
\]  
(1.53)

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. Hence
\[ E[a_{ij}a_{jk}a_{ki}] = E[a_{ij}] \cdot E[a_{jk}] \cdot E[a_{ki}] = 0. \]  
(1.54)
If two of the subscripts are the same (say \( i = j \)) and the third is distinct, we have

\[
E[a_{ii}a_{ik}a_{ki}] = E[a_{ii}] \cdot E[a_{ik}^2] = 0 \cdot 1 = 0.
\]  

(1.55)

If all three subscripts are the same, we have

\[
E[a_{ii}^3]
\]  

(1.56)

This is the third moment of \( a_{ii} \). It is the same for all variables \( a_{ii} \), and is finite by assumption. There are \( N \) triples where \( i = j = k \).

Thus,

\[
E\left[M_{A,N}(3)\right] = \frac{1}{8N^{2.5}} \cdot NE[a_{11}^3] = \frac{E[a_{ii}^3]}{8} \cdot \frac{1}{N^{1.5}}.
\]  

(1.57)

Thus, as \( N \to \infty \), the expected value of the third moment is zero. \( \square \)

To calculate the higher moments requires significantly more delicate combinatorial arguments.

### 1.6.2 Calculation of Moments from the Semi-Circle

We now calculate the moments of the semi-circle. 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 \( N \to \infty \).

\[
C(k) = \int_{-\infty}^{\infty} x^k P(x)dx = \frac{2}{\pi} \int_{-1}^{1} x^k \sqrt{1-x^2}dx.
\]  

(1.58)

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 \( x = \sin \theta \).

\[
C(2m) = \frac{2}{\pi} \int_{-\pi/2}^{\pi/2} \sin^{2m} \theta \cdot \cos^2 \theta d\theta.
\]  

(1.59)

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

\[
C(2m) = \frac{2}{\pi} \int_{-\pi/2}^{\pi/2} \sin^{2m} \theta d\theta - \frac{2}{\pi} \int_{-\pi/2}^{\pi/2} \sin^{2m+2} \theta d\theta.
\]  

(1.60)
The above integrals can be evaluated exactly. We constantly use

\[
\begin{align*}
\cos^2(\phi) &= \frac{1}{2} + \frac{1}{2} \cos(2\phi) \\
\sin^2(\phi) &= \frac{1}{2} - \frac{1}{2} \cos(2\phi).
\end{align*}
\]  

(1.61)

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

\[
n!! = \begin{cases} 
n \cdot (n-2) \cdots 2 & \text{if } n \text{ is even} \\
n \cdot (n-2) \cdots 1 & \text{if } n \text{ is odd}
\end{cases} 
\]  

(1.62)

We find (either prove directly or by induction) that

\[
\frac{2}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2m} \theta d\theta = 2 \frac{(2m-1)!!}{(2m)!!}.
\]  

(1.63)

**Exercise 1.6.6.** Show the above gives

\[
C(2m) = 2 \frac{(k-1)!!}{(k+2)!!}.
\]  

(1.64)

Also, show \(C(2)\) agrees with our earlier calculation.
Chapter 2

Random Matrix Theory and the Mehta-Gaudin Theorem

Lecture by Steven J. Miller; notes by Steven J. Miller and Alex Barnett.

2.1 Gaudin’s Method

2.1.1 Introduction

From Random Matrix Theory: we have a probability distribution on $\mathbb{R}^n$:

$$p_{\beta}(x_1, \ldots, x_N) = c_N(\beta) \prod_{j<k} |x_j - x_k|^\beta e^{-\sum_{j=1}^{N} x_j^2} dx_1 \cdots dx_N. \quad (2.1)$$

Start off with a real $N \times N$ matrix, diagonalize with eigenvalues $x_1, \ldots, x_N$. If you choose the matrix at random, you get $N$ numbers, and you have a probability distribution on the eigenvalues.

We’ve derived the probability above for $N \times N$ matrices. For convenience, we order the eigenvalues.

If $\beta = 1$ we call the ensemble GOE (Gaussian Orthogonal Ensembles); if $\beta = 2$ we have GUE (Unitary) and if $\beta = 4$ we have GSE (Symplectic).

What is the correlation between two eigenvalues? What is the probability of observing a given spacing between two eigenvalues? We’ve done this in the $2 \times 2$ case.

In the $N \times N$ case, we would need to integrate out most of the eigenvalues. The difficulty is $\prod |x_j - x_k|^\beta$.  

22
For $\beta = 1, 2$ or $4$, we can evaluate these integrals; we are fortunate that these values are the ones that arise in practice.

In fact, even just determining $c_N(\beta)$ is difficult. This is called the Selberg Integral, which A. Selberg solved in high school!

We will only consider $\beta = 2$, and will be interested in the limit as $N \to \infty$ (under appropriate re-scaling).

$$R_N(x_1, \ldots, x_N) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} p_2(x_1, x_2, \ldots, x_n, x_{n+1}, \ldots, x_N) dx_{n+1} \cdots dx_N.$$  \hspace{1cm} (2.2)

This will be a symmetric function of the first $n$ variables. If we integrate all but 1 variable we get the density of eigenvalues; if we integrate all but two we get information on pairs of eigenvalues.

**Remark 2.1.1.** $\beta = 0$ is Poissonian.

### 2.1.2 Vandermonde Determinants

Notation: $dp$ means

$$dp(\theta_1, \ldots, \theta_N) = c_N \prod_{j<k} \left| e^{i\theta_j} - e^{i\theta_k} \right|^2 d\theta_1 \cdots d\theta_n. \hspace{1cm} (2.3)$$

We are now working on the $N$-torus $[0, 2\pi] \times \cdots \times [0, 2\pi]$. This goes under the name CUE (Circular Unitary Ensemble).

Remember the group

$$U(N) = \{ N \times N \text{ matrices } A \text{ with } AA^* = I \}. \hspace{1cm} (2.4)$$

Similar to the diagonalization of symmetric matrices, for any unitary matrix $U$ there is a unitary matrix $V$ such that $V^{-1}UV$ is diagonal; further, the eigenvalues have absolute value 1, and hence can be written as $e^{i\theta}$.

Suppose we have $f_1, \ldots, f_N$. We form the Vandermonde of the $N$-variables

$$\text{Van}(f_1, \ldots, f_N) = \prod_{i<j} (f_i - f_j). \hspace{1cm} (2.5)$$

Today we will only use the square, so we don’t worry about ordering so that $f_i < f_j$. 
Exercise 2.1.2.

\[ \text{Van}(f_1, \ldots, f_N) = \det \left( f_i^{j-1} \right)_{1 \leq i, j \leq N}. \quad (2.6) \]

Thus, we have

\[
\begin{vmatrix}
1 & 1 & \cdots & 1 \\
f_1 & f_2 & \cdots & f_N \\
\vdots & \vdots & \ddots & \vdots \\
f_1^{N-1} & f_2^{N-1} & \cdots & f_N^{N-1}
\end{vmatrix}
\]  

(2.7)

2.1.3 Orthonormal Polynomials

On the unit circle \( T \), we have the measure

\[ d\mu(t) = \frac{dt}{2\pi}. \quad (2.8) \]

Let \( f(t) \) be a function such that

\[ \int_T f(t) d\mu(t) = 0, \quad \int_T |f(t)|^2 d\mu(t) = 1. \quad (2.9) \]

Define a sequence of monic polynomials \( P_n(x) \) for \( n \in \mathbb{N} \) and \( \phi_n(t) \) with

\[ \phi_n(t) = P_n(f(t)), \quad \phi_0(t) = \frac{1}{\sqrt{\mu(T)}}, \quad \int_T \phi_i(t) \overline{\phi}_j(t) d\mu(t) = \delta_{ij}. \quad (2.10) \]

This is Gramm-Schmidt, where the inner product between two functions \( f \) and \( g \) is given by

\[ \langle f, g \rangle = \int_T f(t) \overline{g}(t) d\mu(t), \quad (2.11) \]

and the ‘Kronecker delta’ symbol (the discrete analog of the continuous delta ‘function’ \( \delta(\cdot) \)) is defined by

\[ \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (2.12) \]

We introduce orthogonal polynomials to handle the integral. The above process (constructing the \( P_n \)s and the \( \phi_n \)s) gives an orthonormal sequence of polynomials.
2.1.4 Kernel $K_N(x, y)$

Define the kernel

$$K_N(x, y) = \sum_{j=0}^{N-1} \phi_j(x)\bar{\phi}_j(y). \quad (2.13)$$

Exercise 2.1.3. Prove the following:

1. $\int_T K_N(x, x)d\mu(x) = N.$
2. $\int_T K_N(x, y)K_N(y, z)d\mu(y) = K_N(x, z).$

Remark 2.1.4.

$$\int_T K_N(x, y)g(y)d\mu(y) = \sum_{j=0}^{N-1} \left[ \int_T \bar{\phi}_j(y)g(y)d\mu(y) \right] \phi_j(x). \quad (2.14)$$

Thus, integrating $g$ against $K_N$ projects $g$ onto the first $N$ vectors.

Define, for $1 \leq n \leq N$,

$$D_{n,N}(t_1, \ldots, t_n) = \det \left( K_N(t_j, t_k) \right)_{1 \leq j,k \leq n}. \quad (2.15)$$

For example,

$$D_{1,N} = K_N(t_1, t_1) \quad (2.16)$$

and

$$D_{2,N} = \begin{vmatrix} K_N(t_1, t_1) & K_N(t_1, t_2) \\ K_N(t_2, t_1) & K_N(t_2, t_2) \end{vmatrix} \quad (2.17)$$

2.1.5 Gaudin-Mehta Theorem

Theorem 2.1.5 (Gaudin-Mehta). We have

1. 

$$\frac{1}{\mu(T)} \text{Van}\left( f(t_1), \ldots, f(t_N) \right) = \det_{N \times N} \left( \phi_{i-1}(t_j) \right)_{1 \leq i,j \leq N}. \quad (2.18)$$
2. \[ \frac{1}{\mu(T)} \left| \text{Van} \left( f(t_1), \ldots, f(t_N) \right) \right|^2 = D_{N,N}(t_1, \ldots, t_N). \] (2.19)

3. For \( 1 \leq n \leq N \),

\[ \int_T D_{n,N}(t_1, \ldots, t_n) d\mu(t_n) = (N + 1 - n) D_{n-1,N}(t_1, \ldots, t_{n-1}). \] (2.20)

The third statement is the beef, allowing us to integrate out one variable at a time by induction.

Remember

\[ D_{n,N}(t_1, \ldots, t_n) = \det_{n \times n} \left( K_N(t_j, t_k) \right)_{1 \leq j, k \leq n}. \] (2.21)

**Corollary 2.1.6.** Let \( F \) be a symmetric function of \( t_1, \ldots, t_n \), and define

\[
F_{n,N}(t_1, \ldots, t_N) = \sum_{1 \leq i_1 < i_2 < \cdots < i_n < N} F(t_{i_1}, \ldots, t_{i_n})
\]

\[ d\mu_{n,N}(t_1, \ldots, t_N) = \frac{1}{n!} D_{n,N}(t_1, \ldots, t_N) d\mu(t_1) \cdots d\mu(t_N). \] (2.22)

Then

\[
\int_{T^N} F_{N,N}(t_1, \ldots, t_N) d\mu_{n,N}(t_1, \ldots, t_N) = \int_{T^n} F(t_1, \ldots, t_n) d\mu_{n,N}(t_1, \ldots, t_n).
\] (2.23)

How might we use the above? For example, consider for \( 1 \leq j, k \leq N \), and consider for \( f \) even

\[
\sum_{1 \leq j < k \leq N} f(x_j - x_k).
\] (2.24)

What is the expectation of the above? In this case, \( F \) is a function of two variables, and \( F(x_1, x_2) = f(|x_1 - x_2|) \) and we now just need to integrate \( f(|x_1 - x_2|) \) against the determinant of a \( 2 \times 2 \) matrix, and this is the only place where \( N \) will arise.
Suppose we had
\[ dp(x_1, \ldots, x_N) = e^{-\sum x_j^2} \prod_{j<k} |x_j - x_k|^2 dx_1 \cdots dx_N. \] (2.25)

Consider the expectation of
\[ \sum_{1 \leq j < k \leq N} f(|x_j - x_k|). \] (2.26)

According to the corollary, the answer is just
\[ \int_{\mathbb{R}} \int_{\mathbb{R}} f(|x_1 - x_2|) \frac{1}{2!} \begin{vmatrix} K_N(x_1, x_1) & K_N(x_1, x_2) \\ K_N(x_2, x_1) & K_N(x_2, x_2) \end{vmatrix} e^{-x_1^2 - x_2^2} dx_1 dx_2. \] (2.27)

This is enormous progress – we started with \( N \) variables; we now have 2 variables. We need to take the \( N \to \infty \) limit of the determinant, a much easier question.

### 2.1.6 Example

\( T = [0, 2\pi] \), \( d\mu(x) = \frac{dx}{2\pi} \), \( f(x) = e^{ix} \), \( f^n(x) = e^{inx} \), and \( P_n(x) = x^n \). These \( P_n \)s are monic, \( \phi_n(x) = P_n(f(x)) \) is orthonormal, which gives \( \phi_n(x) = e^{inx} \), and clearly
\[ \int_0^{2\pi} e^{inx} e^{-imx} dx = \delta_{ij}. \] (2.28)

Finally, we obtain a geometric progression
\[ K_N(x, y) = \sum_{n=0}^{N-1} e^{in(x-y)} = \frac{1 - e^{iN(x-y)}}{1 - e^{i(x-y)}}. \] (2.29)

We will symmetrize (and go from \(-N\) to \( N\)), and when we take the \( 2 \times 2 \) determinant, we get something like
\[
\sin \left( \frac{N(x-y)}{2} \right) \over \sin \left( \frac{x-y}{2} \right). \tag{2.30}
\]

The most famous pair correlation: we have \( N \) eigenvalues so that the mean spacing is 1. The *pair correlation* is

\[
1 - \left[ \frac{\sin \left( \pi(x-y) \right)}{\pi(x-y)} \right]^2. \tag{2.31}
\]
Chapter 3

Rebecca C. Lehman’s JP: First-Order Spacings of Random Matrix Eigenvalues

The eigenvalues of large random matrices are useful in many contexts, particularly statistical physics. For the Gaussian Orthogonal Ensemble, we present the known distribution of their local spacings, an analogue of the Central Limit Theorem for eigenvalues. We then investigate the local spacings of eigenvalues from other distributions: in particular the Uniform, Cauchy and Poisson, and show evidence that the distribution from the Gaussian may in fact be universal.
3.1 A Little Motivation

3.1.1 Classical Theory

In statistical mechanics, large and complicated systems can often be modeled as ensembles of random numbers. But random numbers are only meaningful if their probability distribution is known, and in physics it is sometimes difficult to guess the appropriate distribution.

Classical probability theory gives us the Weak Law of Large Numbers and the Central Limit Theorem, which essentially state that large sums of random numbers seem to behave in a certain way regardless of the particular distribution. However, not everything is independent of the distribution. Some things can only be proved for certain nice distributions. For instance, the first-order spacings of an ordered set are exponential for the uniform distribution, but the proof does not work for other distributions.

Theorem 3.1.1 (Weak Law of Large Numbers). If \( S_n = \frac{1}{n} \sum_{i=1}^{n} x_i \), where the \( x_i \) are independently randomly distributed over any distribution with mean 0 and variance normalized to 1, then \( E(S_n) = 0 \), and \( \lim_{n \to \infty} E(S_n^2) = 0 \). So for all \( \epsilon \geq 0 \) the probability that \( |S_n| \geq \epsilon \) goes to 0.

Proof:

\[
E(S_n) = \frac{1}{n} \sum_{i=1}^{n} E(x_i) = E(x_i) = 0
\]  

\[
E(S_n^2) = \frac{1}{n^2} \sum_{i,j} E(x_i x_j)
\]

\[
\frac{1}{n^2} \sum_{i,j} E(x_i^2) = \frac{E(x_i^2)}{n}
\]

\[
\frac{1}{n^2} \sum_{i \neq j} E(x_i x_j) = \frac{(n^2 - n)E(x_i)^2}{n^2}
\]

But \( E(x_i^2) = 1 \) and \( E(x_i)^2 = 0 \). So

\[
E(S_n^2) = \frac{1}{n} + 0 = \frac{1}{n}
\]
The probability of $|S_n| \geq \epsilon$ is the probability of $|S_n|^2 \geq \epsilon^2$. By Chebyshev’s inequality,

$$\int_{|x| \geq \lambda} \frac{|x|}{\lambda} \, dP \leq \int_{|x| \geq \lambda} \frac{x}{\lambda} \, dP,$$

$$\int_{|x| \geq \lambda} \frac{x}{\lambda} \, dP \leq \frac{1}{\lambda} \int |x| \, dP$$

$$\text{Prob}(S_n^2 \geq \epsilon^2) \leq \frac{1}{\epsilon^2} \int |S_n^2| \, dP$$

$$= \frac{1}{n \ast \epsilon^2} \quad (3.5)$$

(3.6)

If we hold $\epsilon$ constant and allow $n$ to get large, $\text{Prob}(S_n^2 \geq \epsilon^2)$ approaches 0.

**Theorem 3.1.2 (Central Limit Theorem).** Let $\Omega_n = \frac{1}{\sqrt{n}} \sum_{j} x_j$, where the $x_j$ are independently randomly distributed according to some probability distribution with mean 0 and variance 1, and suitably integrable. Then as $n \to \infty$, $\Omega_n$ has a Gaussian limiting probability distribution $e^{-2\pi \xi^2}$, appropriately normalized.

Proof: We use Fourier analysis. If $x$ and $y$ have probability distributions $\mu$ and $\nu$, then $x + y$ has distribution $\mu \ast \nu$. So $\Omega_n$ has distribution $\mu \ast \ldots \ast \mu$, where $\mu$ is the distribution evaluated at $x \sqrt{n}$, and normalized appropriately by multiplying by $\sqrt{n}$. The Fourier transform converts convolution to multiplication, and $\hat{f}(cx) = \frac{1}{c} \hat{f}(\xi)$, so the Fourier transform of the distribution is $(\hat{\mu})^n$ evaluated at $\frac{\xi}{\sqrt{n}}$, with the normalizations canceling out.

$$\hat{\mu}(\xi) = \int e^{2\pi i x \xi} d\mu$$

(3.7)

$$\hat{\mu}(0) = 1$$

(3.8)

$$\frac{d\hat{\mu}}{d\xi}|_{\xi=0} = 2i \pi E(x) = 0$$

(3.9)

$$\frac{d^2\hat{\mu}}{d\xi^2}|_{\xi=0} = 4\pi^2 E(x^2) = 4\pi^2$$

(3.10)

(3.11)
We do a Taylor expansion of \( \hat{\mu}(t) \) about the origin, then substitute \( t = \frac{\xi}{\sqrt{n}} \).

\[
(\hat{\mu})^n(t) = (1 - 2\pi^2 t^2 + \cdots)^n
\]

\[
(\hat{\mu})^n \left( \frac{\xi}{\sqrt{n}} \right) = \left( 1 - \frac{2\pi^2 \xi}{n} \right)^n
\]

(3.12)

(3.13)

As \( n \to \infty \), \( (\hat{\mu})^n \to e^{-2\pi^2 \xi^2} \), a Gaussian. Since Gaussians are only renormalized by Fourier transform, the original distribution \( \mu \ast \cdots \ast \mu \) must also be Gaussian.

**Theorem 3.1.3 (First Order Spacings).** Let \( x_0, \cdots, x_n \) be random numbers drawn from the uniform distribution on \([0,1)\), ordered by size. Then the probability distribution of the first-order spacings \( x_j - x_{j-1} \), normalized by dividing by the average spacing \( \frac{1}{n} \), approaches \( e^{-x} \).

**Proof:** First we note that the uniform distribution on \([0,1)\) can be considered as the uniform distribution on \( S^1 = \mathbb{R}/\mathbb{Z} \). Since the distribution is uniform and therefore invariant under translations, without loss of generality we can relabel the zero so that \( x_j \) is \( x_1 \) and \( x_{j-1} = x_0 = 0 \). Since \( x_1 \) is the first non-zero value, the probability that the first value \( x_1 \) is greater than \( a \) is the probability that all the \( x_j \) except \( x_0 = 0 \) are greater than \( a \), which is \( (1 - a)^n \). The probability that \( nx_1 \) is between \( t \) and \( t + \Delta \) is thus

\[
\left( 1 - \frac{t}{n} \right)^n - \left( 1 - \frac{t + \Delta}{n} \right)^n
\]

(3.14)

In the limit, as \( n \to \infty \), this goes to \( e^{-t} - e^{-(t+\Delta)} \).

Dividing by \( \Delta \) and taking the limit as \( \Delta \) shrinks to 0, the probability distribution of \( x_1 \) is \( e^{-x} \).

The proof as given applies only to the uniform distribution. However, it can be extended to other suitably integrable distributions by scaling local density to 1, using the cumulative distribution function.

### 3.1.2 Quantum mechanics and Random Matrix Theory

In quantum mechanics, many properties of a system in a given state can be represented by the eigenvalues of a symmetric or Hermitian linear operator; for instance the energy levels of a system in state \( \Psi \) are the eigenvalues solving the equation...
\[ H \Psi = E \Psi. \] Unfortunately, in practice, for systems of any reasonable complexity the size of the matrix is usually impractically large, if it is even known to be finite. Computing the actual entries of such a matrix is usually impossible. Hence it is often useful to treat most systems as random matrices of size \( N \) approaching infinity. Just as classical statistical mechanics treats positions and velocities as random variables in order to study their aggregate properties (for instance, the frequency of their collisions or their total energy), so in the quantum framework the analogous assumption would be to treat the linear operators defining the system as random matrices, and the individual properties as their eigenvalues.

It would therefore be useful to have analogues of the Law of Large Numbers and the Central Limit Theorem, revealing universal properties of the eigenvalues of unknown matrix distributions. Wigner's Semicircle is one such property. The local eigenvalue spacings, unlike the spacings of a classical random number distribution, are conjectured to be another, but as yet this is unproved.

### 3.2 Facts about Random Matrix Eigenvalues

#### 3.2.1 Joint Probability Density

**Theorem 3.2.1 (Joint Probability Density).** If an \( n \times n \) random symmetric matrix \( X \) is drawn from a probability density \( g(\lambda_1, \cdots, \lambda_n) \) with respect to the Lebesgue measure \( \prod_{i \leq j} dX_{ij} \) on \( R_{(n)(n-1)/2} \), where \( g \) is expressed in terms of the increasingly ordered eigenvalues \( \lambda_1, \cdots, \lambda_n \), then the joint probability density function of the eigenvalues on the eigenvalue space \( R^n | \lambda_1 < \cdots < \lambda_n \) is

\[
C_n g(\lambda_1, \cdots, \lambda_n) \prod_{i<j} |\lambda_i - \lambda_j|
\]

(3.1)

**Proof:** We construct the measure on eigenvalue space corresponding to the standard Lebesgue measure. With probability 1, \( X = ODO^T \), where \( D \) is a diagonal matrix with distinct eigenvalues, and \( O \) is an orthogonal matrix: \( O^T = I \), so \( dO^T \cdot O = -O^T \cdot dO \). Differentiating \( X \), and substituting \( dM = \)
\( dO^T * O \), we get

\[
dX = dO * D * O^T + O * dD * O^T + O * dO^T * O * D * O^T
\]

\[
dO * D * O^T = O * O^T * dO * D * O^T
\]

\[
= -O * dO^T * O * D * O^T
\]

\[
= -O * dM * D * O^T
\]

\[
O * D * dO^T = O * D * dO^T * O * O^T
\]

\[
= O * D * dM * O^T
\]

(3.2)

Therefore we find

\[
dX = O(dD - dM * D + D * dM)O^T
\]

(3.3)

Since the measure \( dX \) is invariant under orthogonal transformations, we get

\[
\prod_{i \leq j} dX_{ij} = \prod_{i \leq j} (dD - dM * D + D * dM)_{ij}
\]

\[
= \prod_{i<j} (\lambda_j - \lambda_i)dM_{ij} \prod_{i=1}^n d\lambda_i
\]

\[
= \prod_{i<j} |\lambda_i - \lambda_j| \prod_{i=1}^n d\lambda_i \prod_{i<j} dM_{ij}
\]

(3.4)

We can integrate out the \( dM_{ij} \) to find that the induced measure is

\[
C_n \prod_{i<j} |\lambda_i - \lambda_j| \prod_{i=1}^n d\lambda_i
\]

It follows that if \( X \) has a density \( g(\lambda_1, \cdots, \lambda_n) \prod_{i<j} dX_{ij} \), then the probability density in eigenvalue space is

\[
C_n g(\lambda_1, \cdots, \lambda_n) \prod_{i<j} |\lambda_i - \lambda_j| \prod_{i=1}^n d\lambda_i.
\]

### 3.2.2 Semicircle Law

**Theorem 3.2.2 (Wigner’s Semicircle Law).** If \( X \) is an \( n \times n \) symmetric matrix from some probability distribution \( \mu \) such that the elements \( \xi_{ij} \), up to the symmetric condition, are independently randomly distributed with mean 0, variance 1
and as $n \to \infty$, $C_k(n) = \sup_{1 \leq i < j \leq n} E(|\xi_{ij}|^k) = O(1)$, then the mean eigenvalue distribution of the matrix $\frac{X}{\sqrt{n}}$ tends to the semicircle distribution $\frac{1}{\pi n} \sqrt{4-x^2}$ as $n \to \infty$.

Mehta [5] proves this in his discussion of the Gaussian ensembles, relying on the joint probability distribution. Hiai and Petz [2] prove the theorem by a more conceptual combinatorial argument citing Voiculescu, which does not rely on the messy joint probability distribution, and we will follow their argument here.

The proof is by the method of moments: we can write the moments of the mean eigenvalue distribution in a combinatorial form, and show that the same form characterizes the sequence of moments of the semicircle.

The $k^{th}$ moment of the mean eigenvalue distribution is

$$E(Tr(X^k)) = \frac{1}{n^{k^2+1}} \sum_{1 \leq m_1, m_2, \ldots, m_k \leq n} E(\xi_{m_1m_2} \cdots \xi_{m_km_1})$$

(3.5)

**Definition 3.2.3.** Non-Crossing Partitions We define a non-crossing partition $P$ of a set $S$ to be a partition into pairs $S_j = \{s_{j_1}, s_{j_2}\}$ such that $s_{j_1} < s_{k_1} < s_{j_2}$ iff $s_{j_1} < s_{k_2} < s_{j_2}$.

**Lemma 3.2.4.** The $k^{th}$ moment $E(Tr(X^k))$ approaches 0 if $k$ is odd, and the number of non-crossing partitions of $[k]$ if $k$ is even, as $n \to \infty$.

Proof: If any $\xi_{m_{i},m_{i+1}}$ appears without repetition, its expectation value is 0, so by independence the expectation value of the product containing it is 0. In particular, any term containing more than $\frac{k^2}{2} + 1$ distinct terms contributes nothing to the sum.

There are at most $\binom{n}{l}^k$ possible terms where $l$ of the $m_j$ are distinct, since each of the $l$ distinct $m_j$ can take one of $n$ values, and each of the $k$ factors is chosen from the $l$ distinct values. Since

$$|E(\xi_{m_1m_2} \cdots \xi_{m_km_1})| \leq E(|\xi_{m_1m_2}|^k)^{\frac{1}{k}} \cdots E(|\xi_{m_km_1}|^k)^{\frac{1}{k}} \leq C_k(n),$$

(3.6)

the sum over all such terms is $\binom{n}{l}^k \frac{C_k(n)}{n^{l+1}}$ which vanishes as $n \to \infty$ if $l \leq \frac{k^2}{2} + 1$.

So the only possible sum that doesn’t go to zero is over the terms with $l = \frac{k^2}{2} + 1$ distinct factors. If $k$ is odd, the moment is 0. If $k$ is even, we replace $k$ by $2k'$.

Then we are interested in

$$\frac{1}{n^{k^2+1}} \sum (E(\xi_{m_1m_2} \cdots \xi_{m_km_1}))$$

(3.7)
where the sum is over sequences \( \{m_j\} \) such that exactly \( k' + 1 \) of the \( m_j \) are distinct, and every consecutive pair \( \{m_j, m_{j+1}\} \) (considering \( j \) modulo \( 2k' \)) appears at least twice.

By induction, to any non-crossing partition of \([2k']\) we can associate \( n(n-1)\cdots(n-k') \) terms in this sum: if \( k' = 1 \) there is a single partition \( \{1, 2\} \) to which we assign the \( n(n-1) \) terms defined by \( \xi_{m_1m_2}\xi_{m_2m_1} \). Any non-crossing partition of \([2k' + 2]\) must contain some pair of form \( s_{j_1}, s_{j_1} \) appearing at least twice. To each of the \( n(n-1)\cdots(n-k') \) terms corresponding to this partition, we associate \( (n-k')^2k' + 2 \) terms by inserting one of the \( (n-k'-1) \) terms not yet used, in the \( s_{j_1} \) and \( s_{j_2} \) positions.

Conversely, any sequence satisfying the conditions has each pair \( \{m_j, m_{j+1}\} \) appearing exactly twice, and defines a non-crossing partition by \( \{i, j\} \in P \) iff \( \{m_i, m_{i+1}\} = \{m_j, m_{j+1}\} \). We prove this inductively: for \( k' = 1 \) it is trivial. Assume it holds for \( k' - 1 \). There must be some \( r \) such that \( m_r \) appears only once in the \( m \) sequence. Then \( m_{r-1} = m_{r+1} \neq m_r \). Removing \( m_{r-1} \) and \( m_{r+1} \) we get a sequence with \( k' \) distinct elements. It defines some partition. Combine this non-crossing partition with the additional pair \( \{r-1, r\} \). The result is still non-crossing, so the result holds.

So the sum
\[
\frac{1}{n^{k'+1}} \sum (E(\xi_{m_1m_2}\cdots\xi_{m_n}m_1)) = \frac{n(n-1)\cdots(n-k')}{n^{k'+1}}s_{k'} \quad (3.8)
\]

where \( s_{k'} \) is the number of non-crossing partitions of \([2k']\). As \( n \to \infty \) the coefficient goes to 1, and our lemma is proved.

**Lemma 3.2.5.** The number \( s_k \) of non-crossing partitions of \([2k]\) is the \( k \)-th Catalan number \( c_k = \frac{1}{k+1} \binom{2k}{k} \).

Any non-crossing partition pairs 1 with some even element \( 2m \), since any element \( s_{j_1} \) between 1 and its pair partner must also have \( s_{j_2} \) between 1 and its pair partner. The number of pair partitions containing \( \{1, 2m\} \) is \( s_{m-1}s_{k-m} \) it is determined by a non-crossing partition of the numbers inside \( \{1, 2m\} \) and one of those outside \( \{1, 2m\} \). This gives us the recursion relation \( s_k = \sum_{i=0}^{k-1} s_is_{k-1-i} \) for \( k \geq 2 \).

The function
\[
g(x) = \frac{1}{2}(1 - \sqrt{1 - 4x}) = \sum_{k=0}^{\infty} x^{k+1} \binom{2k}{k} \quad (3.9)
\]
is a generator function of the Catalan numbers. Since \( g(x) \) satisfies the functional equation \( g(x)^2 = g(x) - x \), its Taylor coefficients satisfy \( c_n = \sum_{i=0}^{n-1} c_i c_{n-1-i} \) for \( n \geq 2 \), the same recursion as the \( s_n \), with the same initialization: \( c_0 = s_0 = c_1 = s_1 = 1 \). So \( c_n = s_n \) by induction.

**Lemma 3.2.6.** The \((2k + 1)\text{th}\) moment of the semicircle distribution is 0, and the \(2k\text{th}\) is \(c_k\).

The odd moments are 0 because the semicircle distribution is an even function, so the integral vanishes by symmetry. For the evens, we integrate by parts to get

\[
m_{2k} = \frac{1}{2\pi} \int_{-2}^{2} \sqrt{4 - x^2} (x^{2k-1} (4 - x^2))' dx = 4(2k - 1)(m_{2k-2}) - (2k - 1)m_{2k} \]

\[
= \frac{2(2k - 1)}{k - 1} m_{2k-2} \quad (3.10)
\]

The Catalan sequence clearly satisfies this recursion relation, and \(m_0 = c_0 = 1\), so the lemma is proved.

Thus the moments of the semicircle \( \frac{1}{2\pi} \sqrt{4 - x^2} \) and the moments of the mean eigenvalue distribution are both equal to 0 if \( k \) is odd and the Catalan numbers if \( k \) is even, so since a function is completely determined by its moments, the eigenvalue distribution must be semicircular.

### 3.3 Gaussian

The Gaussian Orthogonal Ensemble, the probability distribution defined on real symmetric matrices by choosing \( x_{ij} \) from a Gaussian distribution \( C e^{-ax_{ij}^2} \), where \( C \) and \( a \) are appropriate normalization constants chosen such that the variance of the trace \( E(Tr(X^2)) = 1 \), is a particularly nice distribution both physically and mathematically. It is invariant under the orthogonal group, which makes it suitable for modeling physical spaces, and also makes the critical properties of the eigenvalues relatively easy to compute.

The joint distribution for the GOE is \( C' e^{-a(\sum_{i=1}^{n} \lambda_i^2)} \prod_{i<j} |\lambda_i - \lambda_j| \) (see Theorem 2.1).

**Theorem 3.3.1 (Characterizing the Gaussian).** The probability distributions on real-symmetric matrices which are independent of choice of basis (i.e. \( P(Q^T Q) \)...
\( X \ast Q \) = \( P(X) \) for \( Q \) orthogonal) and have all entries independently randomly distributed are precisely those of form \( e^{-a \text{Tr}(X)^2 + b \text{Tr}(X) + c} \) for some constants \( a, b, c, a \geq 0 \).

Proof: We follow Mehta [5]. Let \( P = \prod_{i \leq j} f_{ij}(X_{ij}) \): \( X \) has entries independently distributed, and suppose \( P \) is invariant under the Orthogonal group. In particular, if \( X' = O^T \ast X \ast O \), where

\[
O = \begin{pmatrix}
\cos \theta & \sin \theta & 0 & 0 & \cdots & 0 \\
-\sin \theta & \cos \theta & 0 & 0 & \cdots & 0 \\
0 & 0 & 1 & 0 & \cdots & 0 \\
& & & & \ddots & \\
& & & & & \\
& & & & & \\
\end{pmatrix}
\]

then

\[
\frac{\partial X}{\partial \theta} = \frac{\partial O^T}{\partial \theta} X' O + O^T X' \frac{\partial O}{\partial \theta} \\
= \frac{\partial O^T}{\partial \theta} OXO^T O + O^T OXO^T \frac{\partial O}{\partial \theta} \\
= AX + XA^T
\] (3.1)

where

\[
A = \frac{\partial O^T}{\partial \theta} O = \begin{bmatrix}
0 & -1 & 0 & 0 & \cdots & 0 \\
1 & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 1 & 0 & \cdots & 0 \\
& & & \ddots & \cdots & \\
& & & & \\
& & & & & \\
\end{bmatrix}
\]

Since \( P \) is invariant under orthogonal transformations, the logarithmic derivative \( \frac{\partial}{\partial \theta} \sum \log(f_{ij}(X_{ij})) \) should vanish:

\[
\sum_{i,j} \frac{1}{f_{ij}} \frac{\partial f_{ij}}{\partial X_{ij}} \frac{\partial X_{ij}}{\partial \theta} = 0
\] (3.2)

Substituting for \( \frac{\partial X_{ij}}{\partial \theta} \) and expanding, we get

\[
\left( \frac{1}{f_{11}} \frac{\partial f_{11}}{\partial X_{11}} + \frac{1}{f_{22}} \frac{\partial f_{22}}{\partial X_{22}} \right) (2X_{12}) + \left( \frac{1}{f_{12}} \frac{\partial f_{12}}{\partial X_{12}} \right) (X_{11} - X_{22}) \\
+ \sum_{k=3}^{n} \left( -\frac{1}{f_{1k}} \frac{\partial f_{1k}}{\partial X_{1k}} X_{2k} + \frac{1}{f_{2k}} \frac{\partial f_{2k}}{\partial X_{2k}} X_{1k} \right)
\] (3.3)

38
Since each term of the sum depends on independent variables, each individually must be constant. Dividing by $X_{1k}X_{2k}$ we get

$$-\frac{1}{X_{1k}f_{1k}} \frac{\partial f_{1k}}{\partial X_{1k}} + \frac{1}{X_{2k}f_{2k}} \frac{\partial f_{2k}}{\partial X_{2k}} = \frac{C_k}{X_{1k}X_{2k}}$$  \hspace{1cm} (3.4)

This equation is of form $f(x_1) + g(x_2) = h(x_1,x_2)$, which can only be solved by functions of form $a + b \ln x$. So $C_k = 0$ and

$$\frac{1}{X_{1k}f_{1k}} \frac{\partial f_{1k}}{\partial X_{1k}} = \frac{1}{X_{2k}f_{2k}} \frac{\partial f_{2k}}{\partial X_{2k}} = c$$  \hspace{1cm} (3.5)

Integrating, we get $f_{1k}(X_{1k}) = e^{\frac{c}{2}X_{1k}^2}$. We can do the same for $f_{jk}$ for any $j \neq k$. Since all invariants can be expressed in terms of traces of powers of $X$, and the off-diagonal elements appear as squares in the exponent, $P(X)$ can be expressed as an exponential in $Tr(X)$ and $Tr(X^2)$.

3.3.1 Local Spacings

The Gaussian is also mathematically nice in that it is possible calculate its local eigenvalue spacings. Wigner (of the Wigner Semicircle Law) surmised that the local nearest neighbor distribution, on small intervals normalized to have density 1, would be $Axe^{-Bx^2}$, with constants chosen so as to set the integral and the mean to 1. Remarkably, Mehta [4] and Gaudin [1] have proved that the actual spacings are extremely close to Wigner’s surmise.

**Theorem 3.3.2 (First-Order Spacings).** As $n \to \infty$, the probability density $P(S)$ of the distance between two consecutive eigenvalues of an $n \times n$ matrix from the Gaussian distribution (in the region of constant density, normalized by their mean) approaches $\frac{1}{\pi} \frac{2 d^2 \Psi}{dt^2}$ where $\Psi(t)$ is (up to several constants) the Fredholm determinant corresponding to the linear convolution operator $f \to f^\ast - t K * f$, for the kernel $K = \frac{1}{\pi} \left( \frac{\sin(\xi-\eta)}{\xi-\eta} + \frac{\sin(\xi+\eta)}{\xi+\eta} \right)$

The proof of this theorem is very technical. Essentially the idea is to fix $n$ and rewrite $P_n(S)$ by repeatedly converting from product to determinant form and using determinant operations. Eventually $P_n(S)$ is written as a finite Fredholm determinant whose kernels, fortunately, have a known limit as $n \to \infty$. We give only the barest sketch of the key points:
Since we are interested in what happens as $n \rightarrow \infty$ it suffices to consider even $n = 2m$. Fixing $m$, and writing $S = 2\theta$, the spacing distribution $P_m(S)$ for a matrix of size $2m$ is derived from the 2-point correlation function

$$P_m(-\theta, \theta) = \frac{(2m - 2)!}{\mu_0} \int P(-\theta, \theta, \theta_1 \cdots \theta_{2m-2}) d\theta_1 \cdots d\theta_{2m-2}$$  \hspace{1cm} (3.6)

where

$$P(\theta_1 \cdots \theta_n) = e^{-(\theta_1^2 + \cdots + \theta_n^2)} \prod_{i<j} |\theta_i - \theta_j|$$  \hspace{1cm} (3.7)

and the integral is taken over $\theta_i$ ordered in increasing size and with no $\theta_i$ in the interval $(-\theta, \theta)$.

Expanding the absolute value product and factoring out a $2\theta$, we find

$$P_m(-\theta, \theta) = \frac{(2m - 2)!}{\mu_0} 2\theta e^{-2\theta^2} R(\theta)$$  \hspace{1cm} (3.8)

where

$$\mu_0 = (2m)! \frac{2^{2m(2m-1)}}{4} \prod_{k=1}^{\infty} \Gamma \left( \frac{k}{2} \right)$$  \hspace{1cm} (3.9)

$$R(\theta) = \int e^{-(\theta_1^2 + \cdots + \theta_{2m-2}^2)} \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ -\theta & \theta & \theta_1 & \cdots & \theta_{2m-2} \\ \theta^2 & \theta^2 & \theta_1^2 & \cdots & \theta_{2m-2}^2 \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ \theta_{2m-2}^2 & \xi_{2m-2} & \xi_{2m} & \cdots & \xi_{4m-4} \end{vmatrix} d\theta_1 \cdots d\theta_{2m-2}$$  \hspace{1cm} (3.10)

By integrating over the odd variables and applying column operations to the determinant, we get $R(\theta)$ as a symmetric integral over the even variables. This allows us to integrate independently over all the variables and divide by $(m-1)!$.

Expanding the determinant by minors and changing variables several times, it is possible to derive

$$R(\theta) = -\left\{ R_1(\theta) - \frac{1}{4\theta} \frac{d}{d\theta} R_1(\theta) \right\}$$  \hspace{1cm} (3.11)

where if $\xi_{2i} = 2 \int_0^\infty e^{-y^2} y^{2i} dy$,

$$R(1) = \begin{vmatrix} 0 & 1 & \theta^2 & \cdots & \theta^{2m-2} \\ 1 & \xi_0 & \xi_2 & \cdots & \xi_{2m} \\ \theta^{2m-2} & \xi_{2m-2} & \xi_{2m} & \cdots & \xi_{4m-4} \end{vmatrix}$$
Then \( P(-\theta, \theta) = \frac{2m-2}{2\mu_0} d \left( e^{-2\theta^2} R_1(\theta) \right) \) so the probability of an arbitrary spacing being at \((-\theta, \theta)\) is \( 2m(2m-1)P_m(-\theta, \theta) \).

\( R_1 \) can also be written out explicitly as an integral \( \theta \) and \( dy \), and if we define

\[
\phi_m(\theta) = \int_{\theta}^{\infty} \cdots \int_{\theta}^{\infty} e^{-2(y_1^2+\cdots+y_m^2)} \prod_{i<j}(y_i^2-y_j^2)^2 dy_1 \cdots dy_m, \tag{3.12}
\]

then it is possible to substitute in for \( R_1 \) to get

\[
P(S) = \frac{2m!2^{m-1}}{m!\mu_0} \frac{d^2}{d\theta^2} \phi_m(\theta) \tag{3.13}
\]

Let \( \Psi_m \) be the normalization \( \frac{\phi_m(\theta)}{\phi_m(0)} \).

We want the limit of \( \Psi_m \) when \( m \) increases but \( \theta \) is normalized to the magnitude of the mean spacing: set \( t = 2\theta \sqrt{2m} \).

Let \( \Psi(t) = \lim_{m \to \infty} \Psi_m(t) \); then \( P(S) \) is a constant times \( \frac{d^2 \Psi}{dt^2} \). The constant can be found to be \( \frac{\pi}{4} \) by calculating the moments. But we still need to find \( \Psi \).

Changing variables one more time, we renormalize \( t \) to \( \tau \) and \( y_i \) to \( z_i \) by dividing by \( \sqrt{2} \):

\[
\phi_m(\theta) = \int_\theta^\infty \cdots \int_\theta^\infty e^{-2(y_1^2+\cdots+y_m^2)} \prod_{i<j}(y_i^2-y_j^2)^2 dy_1 \cdots dy_m \tag{3.14}
\]

\[
= C \int_{\tau}^\infty \cdots \int_{\tau}^\infty e^{-2(z_1^2+\cdots+z_m^2)} \begin{vmatrix} 1 & z_1^2 & \cdots & z_1^{2m-2} \\ 1 & z_2^2 & \cdots & z_2^{2m-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_m^2 & \cdots & z_m^{2m-2} \end{vmatrix} \tag{3.15}
\]

Rewriting the determinant with column operations, one can write it in terms of the Hermite polynomials and then the harmonic oscillator functions; finally the integration is brought inside the determinant to get

\[
\Psi_m(t) = det \left[ \delta_{ij} - \int_{-\tau}^{\tau} u_{2i}(z) u_{2j}(z) dz \right] \tag{3.16}
\]

which is the Fredholm determinant for the kernel \( K_m = \sum_{0}^{m-1} u_{2k}(x)u_{2k}(y) \).

This kernel can be rewritten more suggestively as

\[
\sqrt{m} \frac{1}{2} \left( \frac{u_{2m}(x)u_{2m-1}(y) - u_{2m}(y)u_{2m-1}(x)}{x-y} + \frac{u_{2m}(x)u_{2m-1}(y) + u_{2m}(y)u_{2m-1}(x)}{x+y} \right)
\]

41
Finally changing variables and letting $m \to \infty$ we get $K = \frac{1}{2\pi} \left( \frac{\sin(\xi - \eta)}{\xi - \eta} + \frac{\sin(\xi + \eta)}{\xi + \eta} \right)$.

Despite its ugly infinite determinant form, $\Psi$ is actually rapidly convergent and $P(S)$ can be computed to reasonable accuracy. It is approximated within 5% [1] by the Wigner surmise, which we will use in our plots because it is much faster to compute.

Note that the proof of the spacings distribution relied crucially on the normal distribution: with any other distribution it would not be possible to write $P(-\theta, \theta)$ as a determinant and integrate out the odd terms. Most distributions are thought to be completely intractable at this time. Nevertheless, the level spacing distribution is conjectured to be a robust universal property: in fact, our data suggests that it may be even more universal than the semicircle law. This is the conjecture which we investigate numerically.

3.4 Computations

It is conjectured that the known first-order distributions for the Gaussian are in fact universal, like the Central Limit Theorem and the Semicircle Law, so that they hold for any suitably normalized, sufficiently integrable function, or and even beyond. We investigate this conjecture by constructing large matrices using the software package Matlab [3] and plotting their first-order spacings against the Wigner surmise.

To construct matrix entries according to an arbitrary distribution $\mu$, one can draw random numbers from the uniform distribution on $[0, 1]$, and then invert the cumulative distribution function $\int_0^t d\mu$. In some cases this integral and its inverse can be computed in closed form. If not, it is necessary to construct a lookup table of the CDF values for small increments. Matlab contains efficient built-in random matrix generators for the uniform, Gaussian and various other distributions. We found that to generate a symmetric matrix, it was significantly more efficient to call the random matrix generator (generating $\frac{n^2}{2}$ entries) then and replace the lower half by symmetry, than to make $\frac{n(n-1)}{2}$ calls to the random number generator.

We look at matrix elements identically distributed according to the following continuous and discrete distributions:

Continuous:
Gaussian (to test the programs) $P(t) = \frac{1}{\sqrt{2\pi}}$,
Uniform on $[-1,1]$ $P(t) = \frac{1}{2}$,
Symmetric Cauchy on $(-\infty, \infty)$ $P(t) = \frac{1}{1+t^2}$,
Discrete:
Sign: $P(1) = \frac{1}{2}, \quad P(-1) = \frac{1}{2},$

Poisson: $P(n) = e^{-\lambda} \frac{\lambda^n}{n!}.$

Note that the Poisson and Cauchy distributions do not have mean zero. The Poisson distribution has variable mean $\lambda$. The Cauchy distribution does not even have finite moments, and therefore does not obey a semicircle law at all.

3.5 Results

The program could process about 700 300x300 matrices per hour, varying slightly according to the distribution. Our data seems consistent with the conjecture for all the test distributions, for sets of up to 5000 matrices of size up to 300x300. The Cauchy distribution (which is not semicircular and does not provide a bound to the entries) converges more slowly to the surmise than the others, but as the size of the matrix grows, all do seem to be approaching the surmise.

We then examined the minima from sets of normalized spacings. The minima from sets of 20 and 100 were predicted to look like $e^{-x}$, but our data appear like a compressed front end of the Wigner curve. More work must be done in this area.
3.6 Bibliography


Chapter 4

Yi-Kai Liu’s JP: Statistical Behavior of the Eigenvalues of Random Matrices

4.1 Introduction

This paper will investigate the statistical behavior of the eigenvalues of real symmetric random matrices. In particular, we shall be interested in the spacings \( s \) between adjacent eigenvalues. Let \( P(s) \) be the distribution of these spacings, in the limit of matrices of large dimension. Empirical evidence suggests that, for a large class of random matrices, \( P(s) \) is given approximately by the “Wigner surmise”:

\[
P(s) \approx A s \exp(-B s^2)
\]  \hspace{1cm} (4.1)

However, the validity of this approximation has only been proved for matrices whose elements are i.i.d. Gaussian (the “Gaussian Ensembles”). This paper will summarize the theoretical understanding of the “Gaussian Orthogonal Ensemble” (GOE), and then describe some computer experiments which suggest that the eigenvalues of sparse and band-diagonal random matrices behave in a similar way.
4.2 The Theory of Random Matrices

4.2.1 Application to Nuclear Physics

To put the theory in context, we begin with some applications of random matrices to nuclear physics. Indeed, this was one of the original motivations for the study of random matrices. Loosely speaking, a quantum mechanical system is described by an eigenvalue problem

\[ H \psi_n = E_n \psi_n \]  

(4.2)

where \( H \) is a Hermitian operator on function-space, \( \psi_n \) is an eigenfunction, and \( E_n \) is the corresponding (scalar) eigenvalue. The operator \( H \) stands for some physical measurement or observation, which can distinguish among different “states” of the system. Each state is represented by an eigenfunction, and the corresponding eigenvalue is the value that would be measured if the system were in that state. (Because \( H \) is Hermitian, its eigenvalues are real.) In the case of an atomic nucleus, \( H \) is the “Hamiltonian”, and the eigenvalue \( E_n \) denotes the \( n \)-th energy level.

Most nuclei have thousands of states and energy levels, and are too complex to be described exactly. Instead, one must settle for a model that captures the statistical properties of the energy spectrum. Instead of dealing with the actual operator \( H \), one can consider a family of random matrices, and compute the distribution of the eigenvalues of these matrices. This is a rather crude model, since it replaces the infinite-dimensional operator \( H \) with a set of matrices of finite dimension \( N \). To make up for this shortcoming, one should look for the asymptotic behavior as \( N \to \infty \). Also, each random matrix has only a finite spectrum of eigenvalues; but one might reasonably hope that the middle part of this spectrum, away from the edges, will still have similar properties to the actual infinite spectrum of the nucleus.

It is worth mentioning that some properties of the physical system are reflected in the choices of random matrices. For instance, if a physical system is invariant under time reversal, its Hamiltonian must be self-dual. The Gaussian Orthogonal Ensemble describes the special case of time-reversal invariance with integral total angular momentum. These issues are discussed fully in Chapter 2 of [Mehta, 1991]; see also the introduction in [Statistical Theories of Spectra].

We will now describe some results from random matrix theory on the distribution of eigenvalues, and the distribution of eigenvalue spacings.
4.2.2 The Semicircle Rule

Take a family of symmetric random matrices, of dimension $N$, chosen from some distribution $D$. Let $P_N(x)$ be the distribution of the eigenvalues, normalized so that the eigenvalues lie in the interval [-1,1], and the total area under the distribution is 1. Then, for suitable distributions $D$, the eigenvalues obey a “semicircle rule”: As $N \to \infty$, $P_N(x)$ converges to a semicircular distribution $P(x)$. That is,

$$P_N(x) \to P(x) = \frac{2}{\pi} \sqrt{1 - x^2}$$  \hspace{1cm} (4.3)

In particular, if the distribution $D$ has finite moments of all orders, then we can prove the semicircle rule using the “method of moments”. Here is an example:

**Theorem 4.2.1. [Semicircle Rule]** Let $D$ be a distribution having finite moments of all orders; specifically, if $X$ is chosen from distribution $D$, then $E(X) = 0$, $E(X^2) = 1$, and, for all $k \geq 3$, $E(X^k)$ is finite.

Let us construct a family of real symmetric random matrices, of dimension $N$, as follows: To get a random matrix $A$, choose its elements $A_{ij}$ (with $i \leq j$) independently from distribution $D$; the remaining elements are then determined by symmetry.

Let $\lambda_j/2\sqrt{N}$ be the normalized eigenvalues of $A$, and define their distribution

$$\mu_{A,N}(x) = \frac{1}{N} \sum_{j=1}^{N} \delta\left(x - \frac{\lambda_j}{2\sqrt{N}}\right)$$  \hspace{1cm} (4.4)

We claim that, as $N \to \infty$,

$$\mu_{A,N}(x) \to P(x) = \frac{2}{\pi} \sqrt{1 - x^2}$$  \hspace{1cm} (4.5)

with probability $\to 1$.

**Proof:** Let $U(x^k)$ be the $k$-th moment of $\mu_{A,N}$:

$$U(x^k) = \int_{-\infty}^{+\infty} x^k \mu_{A,N}(x)dx$$

$$= \frac{1}{N} \sum_{j=1}^{N} \left(\frac{\lambda_j}{2\sqrt{N}}\right)^k$$  \hspace{1cm} (4.6)
Calculate the expected value of each moment:

\[ E(U(x^0)) = E(1) = 1 \] (4.7)

\[ E(U(x^1)) = E \left( \frac{1}{N} \sum_{j=1}^{N} \frac{\lambda_j}{2\sqrt{N}} \right) = \frac{1}{2N^{3/2}} E(\text{Tr } A) \]

\[ = \frac{1}{2N^{3/2}} E \left( \sum_{j=1}^{N} A_{jj} \right) \] (4.8)

\[ = \frac{1}{2N^{3/2}} \sum_{j=1}^{N} E(A_{jj}) = 0 \]

For \( k = 2 \), use the fact that \((A^2)_{jj} = \sum_{k=1}^{N} A_{jk}A_{kj} = \sum_{k=1}^{N}(A_{jk})^2\).

\[ E(U(x^2)) = E \left( \frac{1}{N} \sum_{j=1}^{N} \left( \frac{\lambda_j}{2\sqrt{N}} \right)^2 \right) = \frac{1}{4N^2} E(\text{Tr}(A^2)) \]

\[ = \frac{1}{4N^2} E \left( \sum_{j=1}^{N} (A^2)_{jj} \right) \]

\[ = \frac{1}{4N^2} E \left( \sum_{j=1}^{N} \sum_{k=1}^{N} (A_{jk})^2 \right) \]

\[ = \frac{1}{4N^2} \sum_{j=1}^{N} \sum_{k=1}^{N} E((A_{jk})^2) = \frac{N^2}{4N^2} = \frac{1}{4} \] (4.9)

And so on, up to higher-order moments. The actual distribution is determined by its moments, provided that those moments do not increase too rapidly with \( k \). Thus it will suffice to compare the moments of the eigenvalue distribution with the moments of the “semicircle”.

Let \( C(x^k) \) be the \( k \)-th moment of the semicircle, \( P(x) \):

\[ C(x^k) = \int_{-1}^{+1} x^k P(x) \, dx = \int_{-1}^{+1} 2x^k \sqrt{1-x^2} \, dx \] (4.10)

Make the substitution \( x = \sin \theta \):

\[ C(x^k) = \int_{-\pi/2}^{+\pi/2} 2 \frac{\sin^k \theta \cos^2 \theta}{\pi} \, d\theta \] (4.11)
For $k$ odd, $C(x^k)$ vanishes by symmetry. For $k$ even, we substitute in $\cos^2 \theta = 1 - \sin^2 \theta$, and find that:

$$C(x^k) = \int_{-\pi/2}^{\pi/2} \frac{2}{\pi} \sin^k \theta \, d\theta - \int_{-\pi/2}^{\pi/2} \frac{2}{\pi} \sin^{k+2} \theta \, d\theta \quad (4.12)$$

These integrals can be evaluated analytically (see Gradshteyn and Ryzhik, *Table of Integrals, Series, and Products*, 5th ed., p.412). Let us define some notation: For $n$ even, $n!! = 2 \cdot 4 \cdots n$; and for $n$ odd, $n!! = 1 \cdot 3 \cdots n$. Then we can write:

$$C(x^k) = 2 \left( \frac{(k - 1)!!}{k!!} - \frac{(k + 1)!!}{(k + 2)!!} \right) \left( 1 - \frac{k + 1}{k + 2} \right) \quad (4.13)$$

In particular, we have $C(x^0) = 1$, $C(x^1) = 0$, $C(x^2) = 1/4$ and $C(x^3) = 0$. These are equal to the corresponding moments (calculated above) for the eigenvalue distribution. By extending this approach to include higher moments, we can prove that the eigenvalue distribution goes asymptotically to the semicircle. \( \Xi \)

### 4.2.3 The Gaussian Orthogonal Ensemble

>From this point on, we will concentrate specifically on random matrices belonging to the “Gaussian Orthogonal Ensemble” (GOE). First, consider the family of real symmetric random matrices, of dimension $N$. For any matrix in this family, the $N(N + 1)/2$ matrix elements which lie on or above the diagonal can be chosen freely; the remaining elements are then determined by symmetry. So a random matrix $H$ depends on $N(N + 1)/2$ random variables, namely, the elements $H_{ij}$ with $i \leq j$. We assume that these random variables are independent and identically distributed (i.i.d.).

Moreover, we seek to define a probability measure $P(H)$ on this family of matrices, subject to the following conditions:

1. Orthogonal invariance: For any real orthogonal matrix $Q$, we have:

$$P(Q^T H Q) = P(H) \quad (4.14)$$
2. The random matrix elements $H_{ij}$ (with $i \leq j$) are statistically independent. Thus $P(H)$ can be written as a product:

$$P(H) = \prod_{i \leq j} f_{ij}(H_{ij})$$

(4.15)

where $f_{ij}$ is the probability distribution of $H_{ij}$.

The second condition is intended mainly to simplify things, but the first is absolutely essential, if we are to deal with random matrix eigenvalues. $P(H)$ must depend on abstract linear transforms, irrespective of basis, because these abstract transforms determine the eigenvalues.

So far we have not specified a probability distribution for the matrix elements $H_{ij}$. However, we shall prove the following result:

**Theorem 4.2.2.** Suppose $H$ is a real symmetric random matrix, of dimension $N$; suppose the matrix elements $H_{ij}$ are i.i.d., as described above; and suppose there exists a probability measure $P(H)$ satisfying both condition 1 (orthogonal invariance) and condition 2 (independence of matrix elements).

Then the matrix elements $H_{ij}$ must be Gaussian distributed.

The resulting family of real symmetric matrices, whose elements are i.i.d. Gaussian, is called the Gaussian Orthogonal Ensemble (GOE).

Theorem 4.2.2 tells us that, by assuming conditions 1 and 2 above, we are restricting ourselves to work with Gaussian ensembles. Since our analysis of random matrix eigenvalues will rely heavily on these two conditions, it will be valid only for Gaussian ensembles; it cannot be generalized to other kinds of random matrices. This is one of the chief obstacles to any general theory of random matrices.

Proof of Theorem 4.2.2: The following argument is from [Porter and Rosenzweig, 1960]; see also Chapter 2 in [Mehta, 1991].

Consider a particular orthogonal transformation, namely the two-dimensional rotation through angle $\theta$:

$$Q = \begin{pmatrix}
\cos \theta & \sin \theta & 0 & \ldots & 0 \\
-\sin \theta & \cos \theta & 0 & \ldots & 0 \\
0 & 0 & 1 & & \\
\vdots & \vdots & & \ddots & \\
0 & 0 & & & 1 
\end{pmatrix}$$

(4.16)
The rotation $Q$ acts as follows:

$$H = Q^T H' Q$$  \hfill (4.17)

It is straightforward to calculate the relations between the matrix elements $H_{ij}$ and $H'_{ij}$.

$$
\begin{align*}
H_{11} &= \frac{H'_{11} + H'_{22}}{2} + \frac{H'_{11} - H'_{22}}{2} \cos 2\theta - H'_{12} \sin 2\theta \\
H_{12} &= \frac{H'_{11} - H'_{22}}{2} \sin 2\theta + H'_{12} \cos 2\theta \\
H_{22} &= \frac{H'_{11} + H'_{22}}{2} - \frac{H'_{11} - H'_{22}}{2} \cos 2\theta + H'_{12} \sin 2\theta \\
H_{ij} &= H'_{ij} \text{ for all other } i, j
\end{align*}
$$  \hfill (4.18)

In the product formula for $P(H)$, notice that the factors which depend on $\theta$ are $f_{11}, f_{12}$ and $f_{22}$.

$$P(H) = f_{11}(H_{11}) f_{12}(H_{12}) f_{22}(H_{22}) \prod_{\text{all other } i \leq j} f_{ij}(H_{ij})$$  \hfill (4.19)

Orthogonal invariance of $P$ implies that:

$$\frac{dP}{d\theta} = 0$$  \hfill (4.20)

Therefore

$$\frac{dP}{d\theta} = \frac{f'_{11}}{f_{11}} \frac{dH_{11}}{d\theta} P + \frac{f'_{12}}{f_{12}} \frac{dH_{12}}{d\theta} P + \frac{f'_{22}}{f_{22}} \frac{dH_{22}}{d\theta} P = 0$$  \hfill (4.21)

However, from equations (4.18) we can deduce:

$$
\begin{align*}
\frac{dH_{11}}{d\theta} &= -(H'_{11} - H'_{22}) \sin 2\theta - 2H'_{12} \cos 2\theta \\
&= -2H_{12} \\
\frac{dH_{12}}{d\theta} &= (H'_{11} - H'_{22}) \cos 2\theta - 2H'_{12} \sin 2\theta \\
&= H_{11} - H_{22} \\
\frac{dH_{22}}{d\theta} &= (H'_{11} - H'_{22}) \sin 2\theta + 2H'_{12} \cos 2\theta \\
&= 2H_{12}
\end{align*}
$$  \hfill (4.22)
Substitute these into (4.21) to get:

\[
\frac{f'_{11}}{f_{11}}(-2H_{12}) + \frac{f'_{12}}{f_{12}}(H_{11} - H_{22}) + \frac{f'_{22}}{f_{22}}(2H_{12}) = 0 \quad (4.23)
\]

Divide across by \(-H_{12}(H_{11} - H_{22})\) to get:

\[
\frac{f'_{11}}{f_{11}}\frac{2}{H_{11} - H_{22}} - \frac{f'_{22}}{f_{22}}\frac{2}{H_{11} - H_{22}} = \frac{f'_{12}}{f_{12}}\frac{1}{H_{12}} = -C \quad (4.24)
\]

where the constant \(C\) is introduced because the two sides of the equation depend on different variables. This allows us to solve for \(f_{12}\):

\[
f'_{12} = -CH_{12}f_{12} \text{ has a solution } f_{12} = B_{12} \exp(-CH_{12}^2/2) \quad (4.25)
\]

We can rewrite equation (4.24) as:

\[
\frac{f'_{11}}{f_{11}} - \frac{f'_{22}}{f_{22}} = \frac{C}{2}(H_{11} - H_{22}) \quad (4.26)
\]

\[
\frac{f'_{11}}{f_{11}} + \frac{C}{2}H_{11} = \frac{f'_{22}}{f_{22}} + \frac{C}{2}H_{22} = K \quad (4.27)
\]

The second step introduced another constant \(K\) because the two sides of the equation depend on different variables. Then we solve for \(f_{11}\) and \(f_{22}\):

\[
f_{11} = B_{11} \exp(-CH_{11}^2/4 + KH_{11}) \quad (4.28)
\]

\[
f_{22} = B_{22} \exp(-CH_{22}^2/4 + KH_{22}) \quad (4.29)
\]

If \(H_{11}\) and \(H_{22}\) have mean 0, then \(K = 0\) in the above solutions.

By choosing different rotations \(Q\), the above argument can be made to cover all the elements of the matrix \(H\). Hence, the matrix elements \(H_{ij}\) must follow a Gaussian distribution. So we have arrived at the Gaussian Orthogonal Ensemble. \(\Xi\)

We can derive many interesting results about the GOE. Using our solutions for the \(f_{ij}\), we can re-write the probability density (see equation (4.15)) as:

\[
P(H) = C \exp\left(-\frac{1}{4\sigma^2}\left(\sum_j H_{jj}^2 + 2\sum_{i<j} H_{ij}^2\right)\right)
\]

\[
= C \exp\left(-\frac{1}{4\sigma^2}\sum_{i,j} H_{ij}^2\right) \quad (4.30)
\]

52
This is the expression for $P(H)$ in the case where the matrix elements all have mean 0, the diagonal elements have variance $2\sigma^2$ and the off-diagonal elements have variance $\sigma^2$. When we discuss the theory of the GOE random matrices, we will usually be dealing with this case. (Note, however, that the off-diagonal elements greatly outnumber the diagonal elements. For purposes of numerical experimentation, it does not matter much if the variances of the diagonal and off-diagonal elements are not the same.)

We can write equation (4.30) very neatly, by observing that $(H^2)_{ii} = \sum_j H_{ij}H_{ji} = \sum_j (H_{ij})^2$.

$$P(H) = C \exp \left( -\frac{1}{4\sigma^2} \sum_i (H^2)_{ii} \right)$$

(4.31)

In this form, the orthogonal invariance of $P(H)$ is plainly visible.

### 4.2.4 Probability Distribution of the Eigenvalues

Using (4.30), we can derive the probability distribution of the eigenvalues, for random matrices belonging to the GOE. We will follow the argument given in [Porter and Rosenzweig, 1960]. Let $E_1, E_2, ..., E_N$ be the eigenvalues, and for each $E_k$, let $V_k = (a_{1k}, a_{2k}, ..., a_{Nk})$ be the corresponding eigenvector. The eigenvectors make up the columns of an $N \times N$ orthogonal matrix:

$$A = \begin{pmatrix}
    a_{11} & a_{12} & \cdots & a_{1N} \\
    a_{21} & a_{22} & \cdots & a_{2N} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{N1} & a_{N2} & \cdots & a_{NN}
\end{pmatrix}$$

(4.32)

The matrix $A$ must satisfy $N$ normalization constraints (one for each column), as well as $N(N-1)/2$ constraints to ensure that it is an orthogonal matrix. Thus, the matrix $A$ can be determined by at most $N(N-1)/2$ independent parameters; call them $\alpha_1, \alpha_2, ..., \alpha_{N(N-1)/2}$.

We will take the probability measure $P(H)$ in equation (4.30), and rewrite it in terms of the $N$ eigenvalues $E_k$ and the $N(N-1)/2$ parameters $\alpha$. This will involve substituting for the matrix elements $H_{ij}$, and calculating the Jacobian $J$. 

53
The eigenvectors $V_k$ and the eigenvalues $E_k$ obey relations of the form $HV_k = E_k V_k$. We can combine these into a single matrix equation:

$$
\begin{pmatrix}
  H_{11} & \cdots & H_{1N} \\
  \vdots & \ddots & \vdots \\
  H_{N1} & \cdots & H_{NN}
\end{pmatrix}
\begin{pmatrix}
  a_{11} & \cdots & a_{1N} \\
  \vdots & \ddots & \vdots \\
  a_{N1} & \cdots & a_{NN}
\end{pmatrix}
= 
\begin{pmatrix}
  a_{11} & \cdots & a_{1N} \\
  \vdots & \ddots & \vdots \\
  a_{N1} & \cdots & a_{NN}
\end{pmatrix}
\begin{pmatrix}
  E_1 & \cdots & 0 \\
  \vdots & \ddots & \vdots \\
  0 & \cdots & E_N
\end{pmatrix}
$$

This can be written more succinctly as

$$HA = AE \quad (4.34)$$

where $E = \text{diag}(E_1, E_2, \ldots, E_N)$ is a diagonal matrix containing the eigenvalues. Now multiply by $A^T$, keeping in mind that $A$ is orthogonal, to get:

$$H = AE A^T \quad (4.35)$$

which gives the following expression for the matrix elements:

$$H_{ij} = \sum_k E_k a_{ik} a_{jk} \quad (4.36)$$

Also observe that, because $A$ is orthogonal, $A^T A = I$, which implies the following:

$$\sum_k a_{ki} a_{kj} = \delta_{ij} \quad (4.37)$$

We use the above results to derive an expression which can be substituted into $P(H)$:

$$\sum_{i,j} H_{ij}^2 = \sum_{i,j} \sum_k E_k a_{ik} a_{kj} a_{\ell k} a_{\ell j}$$

$$\quad = \sum_k E_k^2 \quad (4.38)$$

The Jacobian for this change of variables takes the form

$$J =
\begin{vmatrix}
  \frac{\partial H_{11}}{\partial E_1} & \cdots & \frac{\partial H_{11}}{\partial E_N} & \frac{\partial H_{11}}{\partial \alpha_1} & \cdots & \frac{\partial H_{11}}{\partial \alpha_{N(N-1)/2}} \\
  \frac{\partial H_{12}}{\partial E_1} & \cdots & \frac{\partial H_{12}}{\partial E_N} & \frac{\partial H_{12}}{\partial \alpha_1} & \cdots & \frac{\partial H_{12}}{\partial \alpha_{N(N-1)/2}} \\
  \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\
  \frac{\partial H_{NN}}{\partial E_1} & \cdots & \frac{\partial H_{NN}}{\partial E_N} & \frac{\partial H_{NN}}{\partial \alpha_1} & \cdots & \frac{\partial H_{NN}}{\partial \alpha_{N(N-1)/2}}
\end{vmatrix} \quad (4.39)$$
Equation (4.36) shows that each $H_{ij}$ is a linear function of all the eigenvalues $E_k$. So every entry of the form $\partial H_{ij} / \partial \alpha$ is linear in the eigenvalues. It follows that $J$ is a polynomial of degree $N(N - 1)/2$ in each of the eigenvalues.

Now suppose that two eigenvalues $E_i$ and $E_j$ are equal. Then their corresponding eigenvectors $V_i$ and $V_j$ are not uniquely determined. As a result, the inverse of the transformation in (4.36) is not unique. So the Jacobian $J$ must vanish at $E_i = E_j$; therefore, $J$ must contain the factor $(E_i - E_j)$.

This reasoning applies for any $E_i$ and $E_j$, so $J$ must contain all factors of the form $(E_i - E_j)$. Since there are exactly $N(N - 1)/2$ such factors, and we know $J$ is a polynomial of degree $N(N - 1)/2$ in the eigenvalues, it follows that we have completely accounted for $J$’s dependence on the eigenvalues. $J$ can thus be written as:

$$J = \prod_{i<j} (E_i - E_j) \cdot h(\alpha_1, \ldots, \alpha_{N(N-1)/2}) \quad (4.40)$$

Equation (4.30) gives the probability distribution $P(H)$ on the matrix elements $H_{ij}$. Using (4.38) and (4.40), we transform it to a distribution on the eigenvalues $E_k$ and the parameters $\alpha$.

$$P(E_1, \ldots, E_N, \alpha_1, \ldots, \alpha_{N(N-1)/2}) = P(H) \cdot |J| = C \exp \left( -\frac{1}{4\sigma^2} \sum_{i,j} H_{ij}^2 \right) \cdot |J|$$

$$= C \exp \left( -\frac{1}{4\sigma^2} \sum_k E_k^2 \right) \left| \prod_{i<j} (E_i - E_j) \cdot h(\alpha_1, \ldots, \alpha_{N(N-1)/2}) \right| \quad (4.41)$$

Then integrate with respect to the $\alpha$. This gives us the distribution of the eigenvalues $E_k$:

$$P(E_1, \ldots, E_N) = K \cdot \prod_{i<j} |E_i - E_j| \cdot \exp \left( -\frac{1}{4\sigma^2} \sum_k E_k^2 \right) \quad (4.42)$$

Using the above result, one can rigorously derive the “semicircle law” stated earlier. Essentially, one takes equation (4.42), and integrates out each of the variables $E_k$ except for one:

$$P(E_1) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} P(E_1, \ldots, E_N) \, dE_2 \cdots dE_N \quad (4.43)$$
The “semicircle law” arises as we let $N \rightarrow \infty$. The above integration was first carried out by Mehta and Gaudin, using a number of remarkable techniques, including an integration over alternating variables, rewriting the integrand as a determinant, and eventually expressing it in terms of orthogonal Hermite polynomials. We will not describe the details, but refer to [Mehta and Gaudin, 1960].

4.2.5 Eigenvalue Spacings and the Wigner Surmise

We can sort the eigenvalues in ascending order, and compute the spacings $s$ between adjacent eigenvalues. Let $P(s)$ be the distribution of these spacings, in the limit of matrices of large dimension. Wigner postulated that $P(s)$ can be approximated by:

$$P(s) \approx As \exp(-Bs^2)$$  \hspace{1cm} (4.44)

For the GOE, the Wigner surmise becomes:

$$P(s) \approx \frac{\pi}{2} s \exp\left(-\frac{\pi}{4}s^2\right)$$  \hspace{1cm} (4.45)

This distribution is quite different from the Poisson density $\rho(s) = \exp(-s)$ (which gives the intervals between random events). Wigner’s surmise approaches zero for small values of $s$, implying that very small spacings are unlikely, and that the eigenvalues somehow “repel” each other.

Wigner guessed equation (4.44) by assuming that the energy levels of a nucleus behave like a modified Poisson process. Specifically, he postulated that the probability of finding a level in the interval $dx$ is proportional to $f(x) \, dx$ given a level at $x = 0$, where $f(x)$ is a linear function. The physical interpretation is described in [Wigner, 1956].

Using the probability distribution on the eigenvalues (equation (4.42)), Gaudin was able to derive an exact expression for the spacing distribution $P(s)$. The derivation is lengthy, and utilizes many novel techniques; we will only summarize it here, and refer the reader to [Gaudin, 1961] and [Mehta, 1960] for details. Essentially, we start with the following integral, which describes random matrices of dimension $N$:

$$P(E_1, E_2) = \int \cdots \int P(E_1, ..., E_N) \, dE_3 \cdots dE_N$$  \hspace{1cm} (4.46)

The integral is taken over all values of $E_3...E_N$ outside the interval between $E_1$ and $E_2$ (so that $E_1$ and $E_2$ are adjacent eigenvalues). Set $E_1 = -\theta$ and $E_2 = \theta$;
and for convenience, write $N = 2m$. Mehta proves the following result:

$$2m(2m - 1)P(-\theta, \theta) = C \frac{2m!2^{m-1}}{m!} \frac{d^2}{d\theta^2} \phi_m(\theta)$$ (4.47)

where

$$\phi_m(\theta) = \int_0^\infty \cdots \int_0^\infty \exp(-2(y_1^2 + \cdots + y_m^2)) \cdot \prod_{i<j} (y_i^2 - y_j^2) dy_1 \cdots dy_m$$ (4.48)

Define $\Psi_m(\theta) = \phi_m(\theta)/\phi_m(0)$. Let $D = 2\sqrt{2m}/\pi$, and fix:

$$s = \frac{2\theta}{D}, \quad t = \frac{2t}{\pi}$$ (4.49)

Let $\Psi(t)$ be the asymptotic form of $\Psi_m(\theta)$, as $m \to \infty$, under suitable conditions. It can then be deduced that:

$$P(s) = \pi^2 \frac{d^2}{4} \frac{d^2 \Psi}{dt^2}$$ (4.50)

To evaluate (4.48), Gaudin rewrites the integrand using a Vandermonde determinant, then expresses it as a determinant involving orthonormal wavefunctions of the harmonic oscillator, and finally applies Gram’s theorem. In this way, $\Psi_m(\theta)$ becomes a Fredholm determinant of a certain integral operator.

Letting $m \to \infty$, we ultimately express $\Psi(t)$ as the Fredholm determinant of the operator

$$Tf(x) = \int_{-t}^{+t} Q(x, y) f(y) dy$$ (4.51)

whose kernel is

$$Q(x, y) = \frac{1}{2\pi} \left( \frac{\sin(x-y)}{x-y} + \frac{\sin(x+y)}{x+y} \right)$$ (4.52)

$\Psi(t)$ can thus be written as an infinite product:

$$\Psi(t) = \prod_{q=0}^{\infty} \left( 1 - \frac{t}{2\pi} \gamma_{2q}^2 \right)$$ (4.53)

where the $\gamma_{2q}$ are constants which can be computed numerically. In practice, this infinite product converges quite rapidly. As it turns out, Wigner’s surmise is a very good approximation to the exact spacing distribution $P(s)$; in fact, the approximation is accurate to within 5% in the region where $P(s)$ is not too small.
4.3 Computer Experiments

Wigner’s surmise that the spacing distribution $P(s)$ can be approximated by

$$P(s) \approx As \exp(-Bs^2)$$  \hspace{1cm} (4.54)

has only been proven for Gaussian ensembles. However, empirical evidence suggests that it is almost universally applicable—for instance, it appears to hold for random matrices constructed with distributions other than the Gaussian. I performed some computer experiments with sparse and band-diagonal matrices, and the results support the Wigner surmise in these cases as well.

4.3.1 Notes on Implementation

The experiments were carried out using the software package MATLAB, which has built-in functions to do numerical linear algebra. In particular, I used MATLAB’s functions to generate random numbers, and to compute matrix eigenvalues.

There were two performance-related issues which arose: First, MATLAB provides special data structures and functions which are optimized for sparse matrices. These sparse-matrix functions appear to save memory at the expense of speed. For extremely large sparse matrices, one may have no choice but to use the sparse-matrix data structures. However, I found that the regular (non-sparse) data structures worked for matrices as large as $500 \times 500$ (the largest in my experiments); and the regular matrix functions were two to three times faster than their sparse-matrix counterparts.

Secondly, MATLAB’s built-in functions are significantly faster than MATLAB code written by the user. This is partly because the built-in functions work on vectors, calculating their components “in parallel”, whereas user code usually works with scalars, using a loop to calculate them in sequence.\footnote{This is not entirely true, though. On most microprocessors, a built-in vector function is not really computed “in parallel”—it usually contains a small loop. It is a peculiarity of MATLAB that a built-in loop is so much faster than a user-defined loop.}

In any case, many computations will run faster if they can be posed using vectors and the built-in MATLAB functions. For instance, the following two code fragments both generate a real symmetric matrix $A$ whose elements are Gaussian distributed:

\begin{verbatim}
This is not entirely true, though. On most microprocessors, a built-in vector function is not really computed “in parallel”—it usually contains a small loop. It is a peculiarity of MATLAB that a built-in loop is so much faster than a user-defined loop.

58
% Generate random matrix (method 1)
for i = 1 : N
    for j = 1 : i-1
        A(i,j) = A(j,i);
        % Always have j < i
        % (copy from upper triangle to lower triangle)
    end
    for j = i : N
        A(i,j) = randn; % Gaussian distribution
    end
end

% Generate random matrix (method 2--using vectors)
A = triu(randn(N), 0); % Generate upper triangle and diagonal
A = A + (triu(A,1))'; % Transpose the upper triangle to form
                      % the lower triangle

Using method 1, MATLAB will spend as much time generating the matrices as it will computing their eigenvalues! By contrast, method 2 has a negligible running time compared to the eigenvalue computation.

### 4.3.2 Sparse Matrices

I generated sparse random matrices according to the following rule: given a small constant $p$, each element is chosen independently from a distribution which returns 1 with probability $p/2$, -1 with probability $p/2$, and 0 otherwise. I tested values of $p$ ranging from 0.05 to 0.25. The resulting eigenvalues obeyed the semicircle rule, and their normalized spacings agreed with the Wigner surmise. I did not observe any unexpected behavior.

For future work, there is another class of matrices which are “more sparse” than these, namely the adjacency matrices of sparse graphs. For instance, the adjacency matrix of a $k$-regular graph has $O(kN)$ nonzero elements, as compared to $O(pN^2)$ nonzero elements in the matrices that I tested. These adjacency matrices also suggest a possible relationship to the theory of random graphs.
4.3.3 Band-Diagonal Matrices

I also tested band-diagonal matrices: matrices whose nonzero elements lie only in a “band” above and below the main diagonal. Given a parameter $r$, I constructed matrices with exactly $r$ nonzero diagonals above the main diagonal, and $r$ nonzero diagonals below it; so $r = 0$ would give a diagonal matrix, and $r = N - 1$ would give a full matrix. In my experiments, the elements within the “band” were either i.i.d. Gaussian, or i.i.d. uniform (in the interval $[-1,1]$).

The results are consistent with what is predicted by the theory. By restricting ourselves to band-diagonal matrices, we impose a fairly strong condition on the structure of the matrices. The behavior of the eigenvalues depends on the value of $r$. At one extreme, $r = 0$, the matrix is diagonal, and the eigenvalues are simply the nonzero matrix elements; so the eigenvalues have the same distribution as the matrix elements, and the spacings follow the Poisson density $\rho(s) = \exp(-s)$. At the other extreme, $r = N - 1$, the matrix is full, and so the eigenvalues obey the semicircle law, and their spacings follow Wigner’s surmise.

It is of some interest to see when the behavior of the spacings changes from the Poisson density to the Wigner surmise, because the two are qualitatively so different. According to Professor Sarnak, some recent work suggests that the change-over occurs around $r = \sqrt{N}$. My own data are roughly consistent with this; but without measuring quantitatively the change in the distributions, it is difficult to draw any reliable conclusions.

4.3.4 Conclusions

These computer experiments appear to support the Wigner surmise for sparse and band-diagonal matrices. Unfortunately, they do not provide any clues about why the Wigner surmise is so universally valid, or how one might develop a more general theory to explain it. One intriguing point, however, is the behavior of band-diagonal matrices, which varies in a natural way between the opposite extremes of diagonal and full random matrices.

4.4 References

It is a pleasure to thank Professor Peter Sarnak and Steve Miller for helping me learn this material, and for their comments on my work. In addition, Rebecca Lehman did related work on random matrices with Gaussian, uniform and Cauchy
distributions; her paper is also available at Princeton.


Tracy, Craig A., and Harold Widom. *Introduction to Random Matrices.* (A collection of notes.)


Papers marked with an asterisk (*) are included in *Statistical Theories of Spectra: Fluctuations*, which is also listed above.
Chapter 5

Graphs

Lecture by Lior Silberman

5.1 Definition

**Definition 5.1.1 (Simple Graph).** A simple undirected graph is pair \( G = (V, E) \) whose

1. \( V \) is a set (elements called vertices);
2. \( E \subset V \times V \) (elements called the edges);
3. \( (v, v) \not\in E \) for any \( v \in V \);
4. \( (u, v) \in E \) if and only if \( (v, u) \in E \).

**Definition 5.1.2 (Complete Graph).** A graph is complete if between any two vertices there is an edge.

**Definition 5.1.3 (Subgraph).** Take a subset of the vertices and a subset of the edges on those vertices. If \( V' \subset V \) and \( E' \subset E \cap (V' \times V') \).

**Definition 5.1.4 (Induced Subgraph).** Same as above, except \( E' = E \cap (V' \times V') \).

**Definition 5.1.5 (Path).** A path in \( G \) is a sequence of edges \( (e_1, \ldots, e_k) \) such that if \( e_i = (u_i, v_i) \) then \( v_i = u_i + 1 \) for all \( 1 \leq i \leq k - 1 \). We say this path is from \( u_1 \) to \( u_k \).
Definition 5.1.6 (Connected). A graph is connected if there is a path between any two vertices.

Path connectedness is an equivalence relation – if there is a path from \( u \) to \( v \), then there is a path from \( v \) to \( u \) (if we allow paths of length zero, ie a path from \( u \) to itself). Denote by \( u \sim v \); this breaks up \( G \) into equivalence classes. We will concentrate on connected graphs (if a graph is not connected, just analyze each piece separately).

Definition 5.1.7 (\( G \) is \( k \) edge-connected). We say \( G \) is \( k \)-edge connected if after removing any \( k - 1 \) edges from \( G \), \( G \) is still connected.

Theorem 5.1.8. If \( G \) is \( k \) edge-connected, then for any \( u \neq v \in V \) there exists \( k \) edge-disjoint paths from \( u \) to \( v \) (ie, no edge is in two paths).

Definition 5.1.9 (Multigraph). We generalize our notion of graph. We have a pair \((V, E)\) and an edge function \( E : V \times V \to \mathbb{Z}_{\geq 0} \) such that

1. the graph is undirected: \( E(u, v) = E(v, u) \);  
2. \( E(u, u) \) is even for all \( u \in V \)

If \( A \) and \( B \) are sets of vertices, let

\[
e(A, B) = \sum_{(a, b) \in A \times B} E(a, b).
\]

(5.1)

Definition 5.1.10 (Degree of a vertex). For \( v \in V \), the degree of \( v \in V \) is the total number of edges incident on \( V \):

\[
d_v = \deg(v) = e(\{v\}, V).
\]

(5.2)

Definition 5.1.11 (d-Regular). A graph is \( d \)-regular if \( d_v = d \) for all \( v \in V \).

Definition 5.1.12 (n). Standard notation: \( n = |V| \).

Definition 5.1.13 (Isomorphic). Two graphs are isomorphic if there is an identification of the vertices such that the two are the same.
5.2 Counting Games

How many edges are there in a graph?

\[ \sum_{v \in V} d_v = 2|E| = e(V, V), \]  

(5.3)

as we count each edge twice. This is another reason why we define loops at a vertex with weight two and not weight one.

Clearly, a graph with \( n \) vertices has at most \( \binom{n}{2} \) edges. If \( G \) is \( d \)-regular, then \( nd = 2e \). So, at most \( O(n^2) \) edges; if \( d \)-regular, have \( O(n) \) (call this sparse).

Let \( V = A \cup B \) be a disjoint partition of \( V \). If the graph is \( k \) edge-connected, there are at least \( k \) edges crossing the partition.

Without loss of generality, assume \( A \) has fewer edges then \( B \). As we increase the size of \( A \), how many edges are there leaving \( A \)? IE, how many edges are leaving \( A \) to its complement in \( V \), \( B \)?

Let \( e(A) = e(A, A) \), \( e(B) = e(B, B) \), ie, the number of internal edges of each. We assume \( e(A, V) \leq e(B, V) \). Note \( e(A, V) + e(B, V) = e(V, V) \).

Study the ratio

\[ \frac{e(A, B)}{e(A, V)}, \]

the proportion of edges from \( A \) crossing the boundary.

If \( d \)-regular, then \( e(A, V) \) is of size \( d|A| \); think of \( e(A, B) \) as the boundary. In the plane, a disk of radius \( r \) has boundary \( 2\pi r \) and area \( \pi r^2 \); we cannot guarantee that the boundary is a fixed proportion of the area.

**Definition 5.2.1 (\( \alpha \)-expanding).** A graph is \( \alpha \)-expanding if

\[ \frac{e(A, B)}{e(A, V)} \geq \alpha \]  

(5.4)

for all \( A \) such that \( e(A, V) \leq e(B, V) \) (where \( B \) is the complement of \( A \) in \( V \)). Call the largest such \( \alpha \) the Cheeger constant of \( G \).

In \( n \)-dimensional space, have seen

\[ \Delta = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}. \]  

(5.5)

We call this the Cheeger constant because of the similarity to \( n \)-dimensional case.
5.3 Adjacency Matrix and Laplacian

The vector space of \( G = \mathbb{C}^V \): for \( f : V \to \mathbb{C} \), set

\[
(A_G f)(u) = \frac{1}{d_u} \sum_{(u,v) \in E} f(v). \tag{5.6}
\]

What is the matrix of this operator? It acts on an \( n \)-dimensional space. The matrix is basically the adjacency matrix. For a regular graph, it is exactly the adjacency matrix up to a constant.

For \( f, g : V \to \mathbb{C} \), set

\[
\langle f, g \rangle_G = \frac{1}{2|E|} \sum_v d_v f(v) \overline{g(v)}. \tag{5.7}
\]

Set

\[
\langle A_G f, g \rangle_G = \langle f, A_G g \rangle_G. \tag{5.8}
\]

Assume \( G \) is \( d \)-regular. Then

\[
(A_G f)(u) = \frac{1}{d} \sum_{(u,v) \in E} f(v)
\]

\[
\langle f, g \rangle = \frac{1}{n} \sum_v f(v) \overline{g(v)}. \tag{5.9}
\]

**Exercise 5.3.1.**

\[
|\langle Af, f \rangle| \leq \langle f, f \rangle. \tag{5.10}
\]

This implies \( ||A|| \leq 1 \), or each eigenvalue is at most 1 in absolute value.

**Definition 5.3.2 (Laplacian).**

\[
\Delta_G = I - A_G. \tag{5.11}
\]

The above is self-adjoint, positive semi-definite, and if we are \( d \)-regular, the above is real symmetric.
5.4 Spectrum

Definition 5.4.1 (Spectrum). The spectrum of \( G \) is the spectrum of \( A_G \).

If \( G \) is not connected, expansion properties are analyzed component by component. For example, say \( G \) is the disjoint union of \( G_1 \) and \( G_2 \). Then

\[
\mathbb{C}^V = \mathbb{C}^{V_1} \oplus \mathbb{C}^{V_2}.
\]

Moreover, if \( f = (f_1, f_2) \), then

\[
A_G f = (A_{G_1} f_1, A_{G_2} f_2).
\]

Thus, the matrix looks like

\[
\begin{pmatrix}
A_{G_1} & 0 \\
0 & A_{G_2}
\end{pmatrix}
\]

5.4.1 1 an Eigenvalue

Question 5.4.2. When is 1 an eigenvalue? Assume \( Af = f \) or \( f(u) = \frac{1}{d} \sum_{(u,v) \in E} f(v) \).

For each connected component of \( G \), set \( f_i = \sqrt{\frac{1}{|G_i|}} 1_{V_i} \), where \( 1_{V_i} \) is one on \( V_i \) and 0 elsewhere. This function will have length 1.

No other functions work. Say \( f \) takes its maximum value at \( u_0 \). What do we know about the neighbors of \( u_0 \)? The values at the neighbors is at most \( f(u_0) \), so the average at the neighbors is at most \( f(u_0) \). Thus, each neighbor must also have the value \( f(u_0) \). Propagating, this gives that \( f \) is constant on each connected component.

5.4.2 −1 an Eigenvalue

For now on, we assume the graph is connected.

Question 5.4.3. When is −1 an eigenvalue?

\[
|f(u)| \leq \frac{1}{d} \sum_{(u,v) \in E} |f(v)|.
\]
Say $|f|$ takes its maximum at some point. Then all the neighbors must have the same absolute value. (Recall the maximum modulus principle from complex analysis). Thus, $|f|$ is constant. Therefore, we can write $f$ as $1_A - 1_B$, where $1_X$ is the characteristic function of the set $X$, and $A$ and $B$ are a disjoint partition of $G$.

Let $a \in A$, say $f(a) = 1$. Some of the neighbors of $f$ are 1, some are $-1$. But, as $f$ is an eigenfunction with eigenvalue $-1$, the average over the neighbors must be $-1$. Thus, the only way to have an average of $-1$ is if every neighbor has average $-1$, ie, in $B$. By the same logic, if $b \in B$, all of its neighbors are in $A$. We call such a graph bi-partite.
Chapter 6

Graphs - II

6.1 Problems from Professor Sinai

Exercise 6.1.1. Consider a \( d \)-regular graph. Let \( Z_d(n) \) be the number of \( d \)-regular graphs. Prove that \( Z_d(n) \geq \lambda^n \) for some \( \lambda > 1 \).

Consider \( n = 2^m \).

Exercise 6.1.2. What is

\[
\lim_{n \to \infty} \frac{\log Z_d(n)}{n}.
\] (6.1)

6.2 Connectivity Problem

We will let \( h \) represent the Cheeger constant now.

\[
h(G) \leq \inf \left\{ \frac{e(A, B)}{e(a, V)} : A \bigcup B = V, |A| \leq |B| \right\}.
\] (6.2)

This is the Cheeger constant (iso-perimetric: how large is the boundary relative to the set). If the graph is \( d \)-regular, \( e(A, v) = d|A| \).

Definition 6.2.1 (Spectral Gap).

\[
\text{Spectral Gap} : 1 - \lambda_1(G).
\] (6.3)

Definition 6.2.2 (\( u \sim v \)). \( u \sim v \) means all vertices \( v \) such that \( (u, v) \) is an edge.
Definition 6.2.3 (Neighborhood). \( N_u = \{(u, v) \in E : v \in V\} \).

Theorem 6.2.4. Let

\[ \lambda_1(G) = \max\{\lambda : \lambda \neq 1 \text{ an eigenvalue of } G\}. \]

Then

\[ \frac{1 - \lambda_1(G)}{2} \leq h(G) \leq \sqrt{2(1 - \lambda_1(G))}. \]

The second largest eigenvalue has combinatorial implications; the other eigenvalues do not have known combinatorial applications.

Every graph has a Spectral Gap (if connected, only one eigenvalue at 1). So, every graph is an expander. What is more interesting is a family of expanders:

Definition 6.2.5 (Family of Expanders). Consider a family of graphs \( \{G_i\} \) with \( |G_i| \to \infty \) such that \( h(G_i) \geq h_0 \) (implying \( \lambda_1(G) \leq \lambda_1 \)). Will often consider a family of \( d \)-regular graphs.

Miracle: there exists infinite families of \( d \)-regular graphs! Thus, can find an \( h_0 \) such that arbitrarily large graphs have that expander.

Proof: the second inequality is much harder, and we will refer to the notes.

Other direction: without loss of generality, assume \( G \) is connected, say \( V = A \sqcup B \). Consider

\[ f = a1_A + b1_B, \quad 1_X \text{ the characteristic function of } X. \]

What is \( Af \)? Say \( u \in V \), then

\[ Af(u) = \frac{1}{d} \sum_{u \sim v} f(u) \]

\[ = \frac{1}{d} [a|N_u \cap A| + b|N_u \cap B|]. \]

Say \( u \in A \). We have

\[ \sum_{u \in A} Af(u) f(u) = \sum_{u \in A} \frac{1}{d} a^2 e(\{u\}, A) + \frac{ab}{d} e(\{u\}, B) \]

\[ = \frac{1}{d} a^2 e(A, A) + \frac{ab}{d} e(A, B). \]
Do a similar sum for \( u \in B \) and add, obtaining

\[
\langle Af, f \rangle_G = \sum_{u \in V} A f(u) f(u) = \frac{a^2}{d} e(A, A) + \frac{2ab}{d} e(A, B) + \frac{b^2}{d} e(B, B). \tag{6.9}
\]

Let \( f = \sum_{i=0}^{n-1} \alpha_i \psi_i \), where \( \{ \psi_i \} \) is an orthonormal basis of eigenfunctions of \( A \) with eigenvalues \( \lambda_0 = 1 \geq \lambda_1 \geq \cdots \geq \lambda_{n-1} \).

Now (we will explain later) we may assume that \( \alpha_i = 0 \), i.e., that \( f \) is orthogonal to the eigenfunction associated to \( \lambda_0 = 1 \).

\[
\langle Af, f \rangle_G = \sum_{i=0}^{n-1} |\alpha_i|^2 \lambda_i \leq \lambda_1 \sum_{i=1}^{n-1} |\alpha_i|^2 = \lambda_1 \langle f, f \rangle, \tag{6.10}\]

which is the desired result. All we need to do is make sure that \( f \) averages to zero. If it does, then \( f \) is orthogonal to the constant eigenfunction. Choosing \( a = |B| \) and \( b = |A| \) will give \( \langle f, 1_V \rangle = 0 \).

We have

\[
\frac{|B|^2}{d} (d|A| - e(A, B)) - \frac{2|A||B|}{d} e(A, B) + \frac{|A|^2}{d} (d|B| - e(A, B)) \leq \lambda_1 \left( |B|^2 |A| + |A|^2 |B| \right). \tag{6.11}\]

Shifting yields, noting \( |A| + |B| = |V| = n \) and \( \frac{n}{|B|} \leq 2 \) as \( |A| \leq |B| \).

\[
\left[ |B|^2 |A| + |A|^2 |B| \right] \left[ 1 - \lambda_1(G) \right] \leq \frac{e(A, B)}{d} \left[ 2|A||B| + |B|^2 + |A|^2 \right] \\
1 - \lambda_1(G) \leq \frac{e(A, B)n}{d|A||B|} \leq 2 \frac{e(A, B)}{d|A|} \tag{6.12}\]
Chapter 7

Graphs III

7.1 Review

Let $G$ be a finite $d$-regular graph. Then

$$\frac{1 - \lambda(G)}{2} \leq h(G) \leq \sqrt{\frac{1 - \lambda(G)}{2}}.$$  (7.1)

Thus, the Spectral Gap, the distance between 1 and the second largest eigenvalue, $\lambda(G)$, contains important combinatorial notation. Note we have normalized the eigenvalues.

**Definition 7.1.1 (Ramanujan).** For a $d$-regular graph $G$, if the eigenvalues are either $\pm 1$ or in the interval $[-\sqrt{d-1}^2, 2\sqrt{d-1}]$.

We won’t explain why, in the limit, one can’t do better than $\frac{\sqrt{d-1}}{2}$.

*IF* you pick a random 3-regular graph, then with high probability the graph will be Ramanujan. The first construction of an infinite family of Ramanujan graphs (a family where the number of vertices tends to infinity) is due to Lubotzky, Phillips and Sarnak, and uses the Ramanujan bounds for certain modular forms.

7.2 First Problem

**Conjecture 7.2.1.** For all $\epsilon > 0$, a random $d$-regular graph has all eigenvalues (except possibly $\pm 1$) in the interval $[-2\sqrt{d-1}^2 - \epsilon, 2\sqrt{d-1}^2 + \epsilon]$ almost surely.
Definition 7.2.2. Let $\mathcal{P}$ be a property of graphs. We say $\mathcal{P}$ holds almost surely for the random regular graphs if for

$$p_n = \text{Prob}(\{G : G \text{ does not have property } \mathcal{P}, \text{ G has } n \text{ vertices}\}), \quad (7.2)$$

then $\lim_{n \to \infty} p_n = 0$.

Definition 7.2.3 (Random Lift Model (Linial-Amit)). Fix a graph $G = (V, E)$ – this will be the base graph with vertex set $V$ and edges $E$. For each $n \in \mathbb{N}$, the $n^{th}$ model is constructed by

1. Take $V \times [n]$ as the vertex set, where $[n] = \{1, \ldots, n\}$.

2. For each edge $e = (u, v) \in E$, choose a permutation $\sigma_e \in S_n$ independently and uniformly at random, and take the edges $((u, i), (v, \sigma_e(i)))$.

Above a vertex $v$, we now have $n$ points above:

$$(v_1, n) \quad (v_2, n) \cdots$$

$$(v_1, n - 1) \quad (v_2, n - 1) \cdots$$

$$\vdots \quad \vdots$$

$$(v_1, 1) \quad (v_2, 1) \cdots \quad (7.3)$$

The projection map $V \times [n]$ is a covering map for these graphs. Look at coverings of $K_8$ (the complete graph with 8 vertices) and covers of $K_{3,3}$ (the complete bi-partite graph). They look at thousands of covers of these, and everything one expects seems to hold.

Say originally $G$ has eigenvalues $1$ and $\lambda(G)$. The spectral gap cannot get better, as $\lambda(G)$ stays an eigenvalue.

With high probability, a graph with exactly one cycle will have a non-connected lift; if there are at least two cycles, almost surely the lift will be connected.

7.3 Model of Random Regular Graphs on $n$ Vertices

A model of random regular graphs on $n$ vertices can be looked at as a model of sparse $n \times x$ symmetric matrices. These matrices are very sparse.
In the Random Matrix Models, we chose entries independently; in the random regular graph model, all but at most $d$ entries in each row (or column) is zero. The entries of the matrix (ie, the adjacency matrix) are not independent. We can ask about the distribution of eigenvalues in either case.

This is an very thin subset of all random $n \times n$ matrices. Is it so thin that we see different properties? To some extent yes, to some extent no.

### 7.3.1 Definition of the 1-Point Function (Density of States)

Given a symmetric matrix $A \in M_{n \times n}$, set

$$\nu_A(x) = \frac{1}{n} \delta(x - \lambda_i), \quad (7.4)$$

where the $\lambda_i$ are the eigenvalues. This is the same as building a probability function by putting $n$ masses of size $\frac{1}{n}$ at the eigenvalues $\lambda_i$.

If we have a model of random matrices (say choosing uniformly from regular graphs, or choosing entries independently, et cetera), we can consider the ensemble average

$$\nu_n = \int_A \nu_A d\text{Prob}(A). \quad (7.5)$$

**Example 7.3.1.** Let $\mathcal{N}$ be the set of all $d$-regular graphs on $n$ vertices. Then

$$\nu_n = \frac{1}{|\mathcal{N}|} \sum_{G \in \mathcal{N}} \nu_G. \quad (7.6)$$

Here $\nu_n$ is some discrete measure on $[-1, 1]$. We have $\nu_n(1) = \frac{1}{n}$.

As $n \to \infty$, does $\nu_n$ converge (in some sense) to a limit $\nu$?

For more information, see [JMRR].

**Example 7.3.2.** In the GOE case (choosing $n \times n$ real symmetric matrices with entries independent Gaussians), then if we look at the normalized eigenvalues, $\nu_n$ converges to $\frac{2}{\pi} \sqrt{1 - x^2} dx$. What this means is for any nice $f$,

$$\int f(x) d\nu_n(x) \to \frac{2}{\pi} \int_{-1}^{1} f(x) \sqrt{1 - x^2} dx. \quad (7.7)$$

This is called Wigner’s Semi-Circle Law.
Remark 7.3.3. If $p(x)dx$ has mean zero and finite variance, then choosing the entries $a_{ij}$ independently and symmetrically with distribution $p(x)dx$, will also obey the Semi-Circle Law (up to scaling).

7.3.2 Calculating the 1-Point Function for Random Regular Graphs

Definition 7.3.4 ($n^{th}$ moment). Let $\mu$ be a probability measure on $\mathbb{R}$. If $x^n$ is $\mu$-integrable, call $\int x^n d\mu(x)$ the $n^{th}$ moment of $\mu$.

More generally, if we have a probability space $(\Omega, \mathcal{B}, \mu)$, a random variable $X : \Omega \rightarrow \mathbb{C}$, and $X^n$ is $\mu$-integrable, then $E[X^n]$ is the $n^{th}$ moment of $X$.

Remark 7.3.5. If the moments all exist and don’t grow to fast, then they determine $\mu$.

The moments of $e^{-|x|^4}$ has moments growing too fast (we’re not going to quantify what too fast means), and there is not a unique distribution having these moments.

Method: For each $m$, we will compute the $m^{th}$ moment of $\nu_n$, and let $n \rightarrow \infty$. Remember,

$$\nu_A = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}.$$ (7.8)

Thus, we must calculate $\int x^m \nu_A(x)$. Now, $\int f(x) \delta_a(x)dx = f(a)$. Therefore,

$$\int x^m d\nu_A(x) = \frac{1}{n} \sum_{i=1}^{n} \lambda(A)_i^m$$
$$= \frac{1}{n} \sum_{i=1}^{n} Tr(A^m).$$ (7.9)

Eigenvalues of a matrix are hard to extract; however, the trace of a matrix is easy to get a handle on – it comes directly from the diagonal entries.

Thus, we must now calculate the expected value of the trace of $A^m$. We will assume $A$ is the normalized adjacency matrix for a regular graph $G$. If the entries of $A^m$ are $A_{ij}^m$, then
The proof is by induction: trivial for \( m = 0 \) or 1. Note

\[
A_{ij}^{m+1} = \sum_k A_{ik} A_{kj}^m.
\]  

(7.11)

Basically, just conditioning on the first step.

Therefore,

\[
A_{ii}^m = \frac{1}{d^m} \text{ (Number of closed paths starting at } i). 
\]  

(7.12)

For a general graph, this seems hopeless. For a specific matrix, this is very hard; however, we are averaging over many graphs.

As \( n \to \infty \), almost surely at radius \( m \) the graph \( G \) looks like a \( d \)-regular tree. In other words, locally there are no loops. There are a few small cycles, but with high probability our graph will locally look like a tree.

**Theorem 7.3.6.**

\[
\lim_{n \to \infty} \int dA \frac{1}{n} Tr(A^m) \to \frac{1}{d^m} \left( \text{Number of paths of length } m \text{ in the } d\text{-regular tree starting at root} \right). 
\]  

(7.13)

Given a sequence \( a_n \), its generating function is

\[
F(x) = \sum_{n=0}^{\infty} a_n x^n. 
\]  

(7.14)

This is a formal power series – if the sequence doesn’t grow too fast, this defines an analytic function.

If \( n \) is odd, there are no closed paths on the \( d \)-regular tree, so the odd moments vanish.

Thus, we need only study the even moments. Mark one vertex on the tree \( x \), and call this the root.

Define the following

\[
a_n = \text{Number of closed paths of length } 2n \text{ starting at } x \text{ without backtracking}
\]

\[
b_n = \text{Number of paths of length } 2n \text{ starting at } x \text{ without backtracking whose first return is at time } \]

75
Set \( b_0 = 0 \) and \( a_0 = 1 \).

\[
\begin{align*}
  b_n &= (d-1) \cdot a_{n-1} \cdot 1 \\
  a_n &= \sum_{k=1}^{n} b_k a_{n-k} = \sum_{k=0}^{n} b_k a_{n-k}.
\end{align*}
\]  
(7.16)

Set \( A(x) = \sum_{n=0}^{\infty} a_n x^n \) and \( B(x) = \sum_{n=0}^{\infty} b_n x^n \). Multiplying (and using the first property) gives

\[
B(x) = (d-1)x A(x).
\]  
(7.17)

Using the second property gives

\[
\sum_{n=1}^{\infty} a_n x^n = \sum_{n=1}^{\infty} \sum_{k=0}^{n} b_k x^k a_{n-k} x^{n-k}
\]

\[
A(x) = 1 + B(x) A(x).
\]  
(7.18)

We now have two formulas for \( A(x) \) and \( B(x) \). Solving for \( A(x) \) gives

\[
A(x) = \frac{1 - \sqrt{1 - 4x(d-1)}}{2x(d-1)}.
\]  
(7.19)

### 7.4 Comments on Random Graphs (Sinai)

Last time, we showed that if we considered a \( d \)-regular graph \( G = (V, E) \), we can attach an \( n \times n \) adjacency matrix \( A(G) \), where \( a_{ij} = \frac{1}{d} \) if vertices \( i \) and \( j \) are connected by an edge (here, we assume our graph has no loops and no multiple edges).

We can study the spectrum (the eigenvalues) of \( A(G) \). There is always an eigenvalue at 1 (and if the graph is bi-partite, an eigenvalue at \(-1\)). Let \( \lambda_1(G) \) be the second largest eigenvalue. If the graph is connected, there is a gap between \( \lambda_1(G) \) and 1; this is called the Spectral Gap.

Let \( s(d) = \frac{2\sqrt{d-1}}{d} \). As \( n \to \infty \), for any fixed epsilon, do all but the \( \pm1 \) eigenvalues lie in \([s(d) - \epsilon, s(d) + \epsilon]\)?

Consider the trace of \( A^m \):
\[ \text{Tr}(A^m) = 1 + \sum_{i \geq 1} \lambda_i^m(A). \quad (7.20) \]

Assume not bipartite. As \( m \to \infty \), all eigenvalues but the \( \lambda = 1 \) will have \( \lambda_i^m \to 0 \).

Let \( m \sim c \ln n \). Is it true that

\[ \text{Tr}(A^m) - 1 \to 0 \quad (7.21) \]

almost surely?

Comments: is there a nice function such that, dividing the above by it, the limit behaves nicely?
Chapter 8

Introduction to Dynamics Problems

8.1 Introduction

Studying motion, change in the phase space $X$.

Topological Space: can state what it means to be a continuous function (or convergence of points to a limit). In finite dimensional cases, no problem.

In the metrical theory, $X$ is a measure space. Thus, we have a $\sigma$-algebra $\Omega$ (usually a Borel $\sigma$-algebra) and a measure $\mu$ (generalizes the concept of mass / volume / probability). Thus, for each $C \subset \Omega$, we have its measure $\mu(C)$, and the following hold:

1. $\mu(C) \geq 0$.
2. $\mu(X) = 1$.
3. $\sigma$-additivity: $\mu(\bigcup_i C_i) = \sum_i \mu(C_i)$, where the $C_i$ are a disjoint collection of subsets of $\Omega$.

**Discrete Time:** Have a transition from $x$ to $y$, usually write $y = Tx$. Thus, in a unit of time, $x$ goes to $y = Tx$. Sometimes say $T$ gives the Law of Motion

**Continuous Time:** For any $x$ and any time $t$, can say where the particle is. Thus, we have the transformation $x \rightarrow x_t, x_t = T^t x$. The transformation satisfies the following: $T^{t_2} \circ T^{t_1} = T^{t_1 + t_2}$.

Dynamics described by the group of transformations of the Phase Space. In the Metrical Theory,
\begin{align*}
\text{Discrete Time:} & \quad \int f(x) d\mu(x) = \int f(Tx) d\mu(x) \\
\text{Continuous Time:} & \quad \int f(x) d\mu(x) = \int f(T^t x) d\mu(x) dx.
\end{align*}

\section*{8.2 Poincare Recurrence Theorem}

\textbf{Theorem 8.2.1 (Poincare).} Let $\mu(C) > 0$. For almost all $x \in C$, one can find $n(x)$ such that $T^{n(x)} x \in C$, and can do this infinitely often (ie, have an infinite sequences of $n(x)$).

Thus, for almost all $x \in C$ (ie, up to a set of measure 0), can find a time in which it returns to $C$ (hence the term recurrence).

\textbf{Remark 8.2.2.} Suppose $X$ is the two-dimensional plane, and suppose each trajectory is a spiral to the origin (say a Coates spiral). The origin is called the focus here; each trajectory heads there. Clearly, no point returns to a neighborhood of its initial point. The solution is the measure: we are supposed to have an invariant measure (a measure preserved under dynamics); in this case, all the mass is concentrated at the origin.

\section*{8.3 Interval Exchange}

Take the unit interval $[0, 1]$, and cut it into a number of pieces of different lengths, say $\lambda_1, \ldots, \lambda_n$. Permute (rearrange) these pieces.

If only two intervals, only rearrangements are $\lambda_1, \lambda_2$ and $\lambda_2, \lambda_1$.

This transformation is called an interval exchange transformation. In the above, we have exchanged two intervals.

Suppose we have a small interval. Apply the map several times. Eventually, the interval will lie in two different sub-intervals (it will split). However, the sum of the lengths of the two split intervals equals the length of the original interval (this is what we mean by measure preserving). As we repeat, it could split further and further, but the sum of the lengths of the resulting intervals will equal the length of the initial interval.

\textbf{Exercise 8.3.1.} Analyze this transformation. Study its dynamics, check numerically. See the recent work of Maxim Kontsevich and Anton Zorich (1998).
So, we have an orbit of an $x$:

$$x \rightarrow Tx \rightarrow T^2x \rightarrow T^3x \rightarrow \cdots \quad (8.2)$$

We could have a closed loop, say $x \rightarrow Tx \rightarrow T^2x \rightarrow x$.

Transitivity: given any subinterval of $(a, b) \subset (0, 1)$ and any $x \in [0, 1]$, there exists $n = n(a, b, x)$ such that $T^nx \in (a, b)$. Simply put, eventually get everywhere.

Consider a rotation of a circle by an irrational angle. Eventually the orbit of any point becomes dense.

**Exercise 8.3.2.** An interval exchange (an exchange of two intervals) is the same thing as a rotation of a circle.

If we have an interval, we can take a smaller interval and use the Poincare Recurrence Theorem. Then every point of the smaller interval will return to the smaller interval. Take an $x$. It jumps around, and then eventually returns, say at $T^{2003}x$. Here is a new map: the first return time (also called the induced transformation).

So, for each $x$, what is the time it returns? It will turn out this is also an interval exchange.

**Theorem 8.3.3.** Ergodic Theorem of Birkoff and Khinchine Let $f : X \rightarrow \mathbb{R}$ be such that $\int fdu < \infty$. Then for lamost any $x$,

$$\frac{f(x) + f(Tx) + \cdots + f(T^{n-1}x)}{n}$$

converges.
Chapter 9

More on Random Graphs

9.1 \( n(G) \)

Suppose that we have a \( d \)-regular graph \( G = (V, E) \). Let

\[
n(G) = \max_{v \in G} \max_{v' \in G} \min \text{Length}(p(v, v')).
\]  

(9.1)

In words, \( n(v) \) is the smallest number such that we can get from \( v \) to any other \( v' \) in at most \( n(v) \) steps. \( p(v, v') \) is a path from \( v \) to \( v' \).

Almost surely, we expect \( n(G) \leq c \ln n \). Start with a vertex, can go to \( d \) vertices, can go at most to another \( d \) for each, and so on. Gives \( d^m = n \), or \( m = \log_d n \). There are, of course, repetitions; by increasing by a constant, we should be safe.

9.2 More on Traces

Consider \( A \), we have \( Tr(A^p) \approx 1 + \lambda_1^p n \). This formula follows from the normalized largest eigenvalue is 1, and the next largest eigenvalue is less than 1.

Now \( Tr(A^p) = \sum_v a_{vv}^{(p)} \), where \( a_{vv}^{(p)} \) is the number of paths of length \( p \) from \( v \) to \( v \).

From probability theory, \( a_{vv}^{(p)} \to \frac{1}{n} \) as \( p \to \infty \) (probability theorem for Markov chains). Thus, the trace converges to 1 plus a correction.
Chapter 10
Dynamical Systems

10.1 Problem of the Adiabatic Piston

Consider a two dimensional rectangle with dimensions $L_1$ and $L_2$. Suppose we have a vertical wall with mass $M$; we call this separating wall a piston, separating two sub-volumes. To each side of the piston we have some number of particles, say $N_1$ and $N_2$. The particles move and interact, and hit the piston. The piston moves under the action.

We call the system adiabatic, as the only reason the piston moves is due to mechanical interactions.

Suppose the density of the particles on the left (where we have $N_1$) is less than on the right (where we have $N_2$). Then the system will evolve so that the pressures are equal, so that the density on the left will equal the density on the right.

Assume the following:

1. All particles the same mass, $m$.

2. $M = m$, ie, the piston is a very lite wall, and is very sensitive to each collision.

3. We assume the particles on each side are ideal gases (ie, each particle moves independently of each other, and particles do not interact with particles, only with the piston and the walls).

Remark 10.1.1. Note only the horizontal components of the velocities of the particles matter, thus, it is enough to study the one-dimensional analogue.
Thus, we may assume for simplicity that we have the interval \([0, L]\), with the piston initially at \(\frac{L}{2}\), with \(N_1\) particles in \([0, \frac{L}{2}]\), and \(N_2\) particles in \([\frac{L}{2}, L]\). Let \(\rho_i\) be the density in either the first or second sub-interval: \(\rho_1 = \frac{N_1}{L/2}, \rho_2 = \frac{N_2}{L/2}\).

Let’s assume \(N_1 = 2N_2\), so \(\rho_1 = 2\rho_2\).

Each particles has some velocity, assume they collide elastically.

So, consider \((x_i, v_i, m)\) and \((x_j, v_j, m)\). We expect in the end the piston will be such that the densities are equalized.

Take a large room, look at a cubic centimeter. How many molecules are in that cubic centimeter? What is the average number of molecules? The precise statement is the Law of Large Numbers.

See the Appendix at the end of the notes.

10.2 Probability Distribution on Particles

10.2.1 Probability Distribution for Coordinates

Consider two intervals \([0, \frac{L}{2}]\) and \([\frac{L}{2}, L]\). We randomly (from the uniform distribution) and independently put down \(N_1\) points in the first sub-interval and \(N_2\) in the second sub-interval.

Independent means each time we pick a new number, in no way influenced by the previous numbers, and in no way influencing future numbers.

As before, \(\rho_i = \frac{N_i}{L/2}\).

Suppose that we have a small interval \(dx\) in \([0, \frac{L}{2}]\). What is the probability it has \((1 - \frac{dx}{L/2})\) is the probability that a given particle is not in the interval \(dx\). Thus, the probability that the interval \(dx\) has no particles is \((1 - \frac{dx}{L/2})^{N_i}\), so the probability it has a particle is

\[
1 - (1 - \frac{dx}{L/2})^{N_i} = 1 - 1 + \frac{N_1 dx}{L/2} + O((dx)^2) \approx \frac{N_1 dx}{L/2} = \rho_1 dx. \quad (10.1)
\]

10.2.2 Probability Distribution for Velocities

Each particle has mass \(m\), coordinate \(x_i\) (and we know the distribution of coordinates). What is the distribution of the velocities?

Assume we have some distribution \(p\) for the velocities \(v\). Then
\[ \text{Prob}(v \in [a, b]) = p(a)(b - a) + o(b - a), \quad b - a \ll 1. \quad (10.2) \]

As \( p \) is a probability distribution, \( \int_{\mathbb{R}} p(v) dv = 1 \). We will additionally assume \( p \) is continuous and \( p(v) \to 0 \) as \( |v| \to \infty \). We will often assume compact support, or sometimes a normal or Gaussian distribution \( p(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} \).

Let \( N(X, t) \) be the number of particles whose coordinates at time \( t \) are less than or equal to \( X \). If \( X \) is the location of the piston, then \( N(X, t) = N_1 + 1 \) (as the piston counts as a particle). Thus, we must solve \( N(X, t) = N_1 + 1 \).

We will replace \( N(X, t) \) with its expectation, \( E[N(X, t)] \).

What happens when a particle with velocity \( v \) collides with another particle? Then the new particle will move with velocity \( v \). Thus, the trajectory of the velocity will be periodic motion, propagating down. It’s period will be \( \frac{L}{v} \).

We will consider time that is large relative to \( L \), say time \( tL \). In other words, we want to allow many reflections on the boundary.

Consider \([0, L]\), sub-interval of size \( dy \). Have some particle \( y \) in the interval \( dy \) with velocity \( v \). To go from \( y \) to \( L \), it needs a time \( \frac{L - y}{v} \). It needs \( \frac{L}{v} \) time to return to \( 0 \), and then \( \frac{L}{v} \) to return to \( y \).

It’s speeds are + when it moves to the right, negative when it moves to the left. Let \( t_1 \) equal the time when it hits the right wall the first time. Let \( t_2 = tL - n \frac{2L}{v} - \frac{L - y}{v} \).

When is the particle in the interval \([0, X]\)? Clear that for each period, it will be of size \( \frac{2X}{v} \). So

\[ \frac{L}{v} - \frac{X}{v} < t_2 < \frac{L}{v} + \frac{X}{v}. \quad (10.3) \]

We find, by linearity,

\[ L - X \leq vtL - 2Ln - (L - y) \leq L + X. \quad (10.4) \]

Thus

\[ 2Ln + (L - y) + (L - X) \leq vtL \leq 2Ln + (L - y) + (L + X). \quad (10.5) \]

Dividing gives

\[ \frac{2n}{t} + \frac{2L - X - y}{tL} \leq v \leq \frac{2n}{t} + \frac{2L + X - y}{tl} \quad (10.6) \]

or
\[
\frac{2n + 2}{t} - \frac{X + y}{tL} \leq v \leq \frac{2n + 2}{t} + \frac{X - y}{tl}.
\] (10.7)

This is a small interval near \(\frac{2n+2}{t}\). Thus,

\[
P_n = p \left( \frac{2n + 2}{t} \right) \cdot \frac{2X}{Lt},
\] (10.8)

where as \(Lt \gg t\), we can evaluate \(p\) at \(\frac{2n+2}{t}\) and ignore the correction of \(-\frac{y}{tL}\).

We have

\[
\frac{X}{L} \sum_{n=-\infty}^{\infty} p \left( \frac{2n + 2}{t} \right) \cdot \frac{2}{t}.
\] (10.9)

As we change \(n\), we are taking steps of size \(\frac{2}{t}\). For \(t\) large, this will look like a definite integral, and we will get \(\frac{X}{L} (1 + o(1))\).

### 10.3 Bouncing Ball and Moving Wall

**Bouncing Ball and Moving Wall**

Steven J. Miller, January 2003

**Question:** A ball starts moving at \(v_0\) meters per second between two walls initially \(l\) meters apart. One wall is stationary, the other moves at a constant speed of \(u\) meters per second. Find the average force exerted by the ball on stationary wall when the two walls are separated by \(x\) meters. Assume all collisions are elastic, and \(u \ll v_0\)

**Solution:** If \(u = 0\), a straightforward calculation shows the average force is \(\frac{mv_0^2}{l}\). Quick proof: conservation of energy shows (for an elastic collision, basically think the walls have infinite mass) the ball leaves the wall with equal but opposite speed. Thus, the change in momentum is \(2mv_0\). Thus, the impulse is \(F_{ave} \Delta t = 2mv_0\), with the force \(0\) past time \(\Delta t\). The time between collisions is \(\frac{2l}{v_0}\) (as the ball must travel \(2l\) meters, and it is travelling at \(v_0\) meters per second). Thus, the average force is

\[
F_{ave} = \frac{1}{\frac{2}{v_0}} \cdot \frac{2mv_0 \Delta t}{\Delta t} = \frac{mv_0^2}{l}.
\] (10.10)
If \( u \neq 0 \), every time the ball hits the moving wall, it rebounds with its speed increased by \( 2u \). Quick proof: Say the ball is traveling at a speed of \( v \). An observer standing on the moving wall sees the ball approaching at a speed of \( v + u \), thus by conservation of energy, an observer on the wall sees the ball leaving at a speed of \( v + u \).

An observer at rest (ie, moving back \( u \) meters per second from the moving wall) will see the same rate of approach \( (v + u) / (v + u) \) between ball and wall as an observer on the wall. Thus, as a person on the wall sees the ball approaching and leaving at \( v + u \), the person at rest sees the ball approaching the moving wall at \( v \), and departing at \( v + 2u \).

We now consider the problem at hand, proving a useful lemma.

**Lemma 10.3.1.** Assume the moving wall and stationary wall are separated by \( d \) meters, the moving wall is travelling at speed \( u \) and the ball is travelling at speed \( v \). Assume at time 0 the ball is just leaving the moving wall. Then when the ball hits the moving wall next, the distance between the walls is, to first order, \( d \cdot \left[ 1 - \frac{2u}{v} \right] \).

Proof: Assume the moving wall travels a distance \( y \) in the time it takes the ball to leave the moving wall, hit the stationary wall, bounce off and return and hit the moving wall. Since the two walls were initially \( d \) units apart, the ball travels a distance \( 2d - y \).

As the moving wall travels at speed \( u \), the time it spends in transit is \( \frac{y}{u} \); similarly, the time the ball spends in transit is \( \frac{2d - y}{v} \), and these times must be equal.

Therefore, \( \frac{y}{u} = \frac{2d - y}{v} \). Algebra gives

\[
y = 2d \cdot \left( 1 + \frac{u}{v} \right)^{-1} \cdot \frac{u}{v}.
\]

(10.11)

The new separation between the walls is \( d - y \), which is

\[
d - y = d - 2d \cdot \left( 1 + \frac{u}{v} \right)^{-1} \cdot \frac{u}{v}
\]

\[
= d \cdot \left[ 1 - 2 \left( 1 + \frac{u}{v} \right)^{-1} \cdot \frac{u}{v} \right]
\]

\[
= d \cdot \left[ 1 - \frac{2u}{v} + O \left( \frac{u^2}{v^2} \right) \right]
\]

\[
= d \cdot \left[ 1 - \frac{2u}{v} \right] + O \left( \frac{u^2}{v^2} \right).
\]

(10.12)
As \( u \ll v_0 \), and clearly \( v \geq v_0 \), we may ignore the \((\frac{u}{v})^2\) terms.

Thus, if we start with the walls separated by \( d \) units and the ball travelling at \( v \), then after one loop the walls are (to first order) separated by \( d \cdot \left[ 1 - \frac{2u}{v} \right] \) and the ball is now travelling at \( v + 2u \). This completes the proof.

We now calculate the number of times the ball hits the moving wall. Let \( N \) be this number. The wall has moved a distance of \( x \), starting with a separation of \( l \) and the ball at a speed of \( v_0 \). Thus,

\[
x = l \cdot \left[ 1 - \frac{2u}{v_0} \right] \cdot \left[ 1 - \frac{2u}{v_0 + 2u} \right] \cdots \left[ 1 - \frac{2u}{v_0 + N2u} \right].
\]

(10.13)

Taking logarithms converts the product to a sum. As \( u \ll v_0 + k2u \), we use the first order Taylor expansion, namely \( \log(1 - z) \approx -z \), up to errors of size \( z^2 \). Thus,

\[
\log \left( \frac{x}{l} \right) = \sum_{k=0}^{N} \log \left( 1 - \frac{2u}{v_0 + k2u} \right)
\]

\[
\approx - \sum_{k=0}^{N} \frac{2u}{v_0 + k2u}
\]

\[
= - \sum_{k=0}^{N} \frac{1}{1 + \frac{2u}{v_0}} \cdot \frac{2u}{v_0}.
\]

(10.14)

The last expression, again up to lower order corrections of size \((\frac{u}{v_0})^2\), will look like a nice Riemann integral. The step size is \( \Delta z = \frac{2u}{v_0} \); \( z \) ranges from 0 to \( N\frac{2u}{v_0} \), with \( z_k = k\frac{2u}{v_0} \). Note the step size, \( \frac{2u}{v_0} \ll 1 \).

Thus (moving the minus sign to the other side) we have

\[
\log \left( \frac{l}{x} \right) = \int_{0}^{N\frac{2u}{v_0}} \frac{1}{1 + z} \, dz
\]

\[
= \log \left( 1 + N\frac{2u}{v_0} \right).
\]

(10.15)

Therefore
\[
\frac{l}{x} = 1 + N \frac{2u}{v_0} \quad (10.16)
\]

or

\[
v_0 + 2u = \frac{l}{x}v_0. \quad (10.17)
\]

Note \(v_0 + 2u\) is the speed of the ball when the walls are separated by \(x\) meters.

This is now the same as the first case considered; if we just care about the average force on the stationary wall when the separation between the walls is \(x\) and the ball is moving at approximately \(\frac{l}{x}v_0\), then for a few loops, the two walls don’t really get much closer and the ball doesn’t really increase much in speed.

Thus, we use the earlier result (from \(u = 0\)), using now as the separation between the walls as \(x\) (instead of \(l\)) and the speed of the ball as \(\frac{l}{x}v_0\) (instead of \(v_0\)).

Hence, instead of \(\frac{mv_0^2}{I}\) we now have

\[
\frac{m \cdot \left(\frac{l}{x}v_0\right)^2}{x} = \frac{mv_0^2l^2}{x^3} = \frac{mv_0^2}{l} \cdot \left(\frac{l}{x}\right)^3. \quad (10.18)
\]

We see the average force has increased by a factor of \(\left(\frac{l}{x}\right)^3\); note that the way we wrote the average force above makes such a comparison easy. We quickly check extreme cases. If \(x = l\), we recover the \(u = 0\) case; as \(x \to 0\), the force becomes infinite (as the speed of the ball becomes infinite because of the great number of interactions with the moving wall).
Chapter 11

More on Dynamical Piston

11.1 Review

Have the interval \([0, L]\), and at the midpoint \(\frac{L}{2}\) have a marked point (also called the piston). We have \(N_1\) points in \([0, \frac{L}{2}]\) and \(N_2\) points in \([\frac{L}{2}, N]\). Thus, if \(\rho_i\) denotes the two densities, if \(N_1 = 2N_2\), then \(\rho_1 = 2\rho_2\). In the limit, we expect the two densities to be equal; thus, in this case, we expect the piston to end up at \(\frac{2L}{3}\).

We will consider the limiting case when \(L, N_1, N_2 \to \infty\) but \(\rho_1, \rho_2\) are fixed.

Let \(dx\) denote a small interval. It is extremely unlikely that there will be two particles in \(dx\). In the first sub-interval \([0, \frac{L}{2}]\), the mean distance between particles is \(\frac{L/2}{N_1} = \frac{1}{\rho_1}\), which is fixed and finite in the limit. Hence, the mean distance is \(O(1)\).

Therefore, the number of particles in \(dx\) is about \(\rho_1 dx + o(dx)\).

A probability distribution always exists, the density distribution exists only in some cases. The density distribution means that if we have an interval \([a, b]\), the probability that we have values in \([a, b]\) is the integral over \([a, b]\). For small intervals, the probability is proportional to \(b - a\). For example, if we take two values, each with probability \(\frac{1}{2}\), there is no density (the probability in a small interval containing one of the points will not be proportional to the size of the interval).

Let \(N(X, t)\) equal the number of particles in \([0, X]\). If \(N(X, t) = N_1\) or \(N(X, t) = N_1 + 1\) will give the location of the piston (depending on whether or not we count the piston as a particle).

We replace \(N(X, t)\) with its expectation. We then solve \(E[N(X, t)] = N_1\) for \(\overline{X}(t)\). Thus,
\[
\frac{X(t)}{L} \xrightarrow{L \to \infty} \frac{2}{3}.
\] (11.1)

### 11.2 Solution

Decompose the interval \([0, L]\) into arbitrarily small sub-intervals \(dx\). Let \(N_i(X, t)\) be the number of particles in the \(i^{th}\) interval at most \(X\) units. Then

\[
N(X, t) = \sum_i N_i(X, t).
\] (11.2)

Further,

\[
P_i \xrightarrow{t \to \infty} \frac{X}{L}.
\] (11.3)

We rescale time and look at \(tL\), a large time. Then

\[
E[N(X, tL)] \xrightarrow{t \to \infty} \sum_{i}^{\text{left}} \rho_1 dx \frac{X}{L} + \sum_{i}^{\text{right}} \rho_2 dx \frac{X}{L} + \cdots
\]

\[
= \rho_1 \frac{L X}{2} + \rho_2 \frac{L X}{2}
\]

\[
= \frac{X}{2} (\rho_1 + \rho_2)
\]

\[
= \frac{X \cdot 3 \rho_2}{2}.
\] (11.4)

Thus, having

\[
\frac{X \cdot 3 \rho_2}{2} = N_1,
\] (11.5)

and recalling the relations between \(N_1, N_2, \rho_1\) and \(\rho_2\), we obtain after some algebra that \(\frac{X}{L} = \frac{2}{3}\). Use \(\rho_i = \frac{N_i}{L/2}\) and \(\rho_1 = 2\rho_2\).

### 11.3 Generalizations

Assume now all particles on the left have mass \(m_1\), the piston has mass \(m\), and all particles on the right have mass \(m_2\). The collisions are now different, and it is not known what the dynamics of the piston are in this case.
Denote the piston’s data by $M, X, V$. For the particles on the left, we write $m_i^{(l)}, x_i^{(l)}, v_i^{(l)}$; for the right, we use $m_i^{(r)}, x_i^{(r)}, v_i^{(r)}$.

We assume that initially $V(0) = 0$, i.e., that the piston starts at rest.

What is the energy of the system? As the piston starts at rest, there is no contribution from the piston at time $t = 0$. At a later time $t$ the energy is

$$H = \frac{MV^2(t)}{2} + \sum_i \frac{m_i^{(l)} v_i^{(l)}(t)^2}{2} + \sum_j \frac{m_j^{(r)} v_j^{(r)}(t)^2}{2}. \quad (11.6)$$

By conservation of energy, as the energy does not initially depend on $M$, it never depends on $M$. Thus,

$$\frac{MV^2(t)}{2} \leq H, \quad \text{or} \quad |V(t)| \leq \sqrt{\frac{2H}{M}}. \quad (11.7)$$

Consider slow time $t = \tau \sqrt{M}$, $0 \leq \tau \leq 1$. We have

$$X(t) = X(\tau \sqrt{M}) = x_M(\tau) \quad (11.8)$$

and

$$V(t) = V(\tau \sqrt{M}) = v_M(\tau) \quad (11.9)$$

as $M \to \infty$.

**Theorem 11.3.1.** $x_M(\tau) \to x(\tau)$ and $v_M(\tau) \to v(\tau)$ as $M \to \infty$. We have

$$\frac{v^2(\tau)}{2} + \frac{c_1}{x^2(\tau)} + \frac{c_2}{(1 - x(\tau))^2} = \mathcal{H} \quad (11.10)$$

for some constants $c_1, c_2$. Thus, the piston will oscillate.

### 11.4 Three Particles

One particle on left, one piston, one particle on the right. Say initially only piston is moving.
Chapter 12

3x + 1 Problem

12.1 Introduction

Also called the Collatz, Hasse, Katani, Ulum, . . . problem.

Let \( x \) be odd. Then \( 3x + 1 \) is even. Let \( k = k(x) \) be the largest power of 2 dividing \( 3x + 1 \). Let

\[
y = \frac{3x + 1}{2^k}.
\]

(12.1)

We denote this by

\[
T_x = y.
\]

(12.2)

Note \( y \) is not divisible by 2 or 3. We consider \( x \in (-\infty, \infty) \). Note if \( x > 0 \) \((< 0)\) then \( y > 0 \) \((< 0)\).

Let \( \Pi \) be the set of numbers not divisible by 2 or 3. Then

\[
\begin{align*}
\Pi &= \Pi^+ \cup \Pi^- \\
\Pi^+ &= \{ n : n = 6p + 1, p \in \mathbb{Z} \} \\
\Pi^- &= \{ n : n = 6p - 1, p \in \mathbb{Z} \}.
\end{align*}
\]

(12.3)

\( T \) takes points from one subset to the other.

We have some special points which give rise to periodic sequences.

1. \( x = 1 \).
2. $x = -1$.

3. $x = -5$: giving the sequence $-5, -7, -5$.

4. $x = -17$: periodic with period 11.

**Problem 12.1.1** ($3x + 1$). *Is it true that for every $x \in \Pi$ one can find $n(x)$ such that $T^{n(x)}x \in \{1, -1, -5, -17\}$?*

### 12.2 Statistical $3x + 1$ Problem

Let $A$ be a set of integers. Consider the number of points in $A$ and in $[-R, R]$, denoted by $|A \cap [-R, R]|$.

Consider $\pi(A) = \lim_{R \to \infty} \frac{|A \cap [-R, R]|}{2R + 1}$. (12.4)

**Problem 12.2.1.** *Is it true that*

$$\pi\left(\{x : 3x + 1 \text{ problem is true}\}\right) = 1.$$ (12.5)

**Question 12.2.2.** *Consider $\{k_1, k_2, \ldots, k_m, \epsilon\}$, $k_j \geq 1$, $\epsilon = \pm 1$. Let $T = T^{(k)}$ if in the definition of $T$ we divide by $2^k$.

What is the set of $x \in \Pi'$ to which we can apply $T^{(k_1)}T^{(k_2)}\ldots T^{(k_m)}$?*

**Theorem 12.2.3 (Structure Theorem).**

$$\sum_{(k_1, \ldots, k_m, \epsilon)} = \left\{ 6\left(q_{m}(k_1, \ldots, k_m, \epsilon) + 2^{k_1+\cdots+k_m}p\right) + \epsilon \right\},$$ (12.6)

where $q_{m}$ is a special number depending on $k_1, \ldots, k_m, \epsilon$ and $p \in \mathbb{Z}$. *Thus, we have an arithmetic progression. Further, $0 \leq q_{m} \leq 2^{k_1+\cdots+k_m}$.*

Let $x_0 \in \Pi$ and $x_k \in Y^k , y_0 = \ln x_0, y_k = \ln x_k$, and $z_k = y_k - y_0 = \ln \frac{x_k}{x_0}$.

We can graph in the $k, z$ plane. It will be in the fourth quadratic, spiking around a line.

Now $z = -(2 \ln 2 - \ln 3)k$ is the straight line.

Let
\[ A = \left\{ x_0 : a \leq \frac{z_m + m(2\ln 2 - \ln 3)}{\sqrt{m}} \leq b \right\}. \] (12.7)

Then

\[ \lim_{k \to \infty} \pi(A_0) = \frac{1}{\sqrt{2\pi \sigma}} \int_a^b e^{-\frac{u^2}{2 \sigma}} du, \] (12.8)

where \( \sigma \) is some computable constant.

**Lemma 12.2.4.** Consider \( k_1 + \cdots + k_m = k \), where each \( k_i \geq 1 \) and \( k \geq m \). The number of solutions is \( \binom{k-1}{m-1} \).

Now

\[ \frac{1}{2^n} \binom{s}{n} \sim \frac{1}{\sqrt{2\pi \frac{1}{4}s}} e^{-\frac{1}{4}s}. \] (12.9)
Chapter 13

More on $3x + 1$

13.1 Review

Let $x$ be odd, and $k(x)$ the power such that $\frac{3x+1}{2^{k(x)}} = y$ is odd. Thus, we have $Tx = y$.

We consider numbers of the form

$$\Pi = \Pi^+ \cup \Pi^-$$

$$\Pi^+ = \{6p \pm 1 : p \in \mathbb{Z}\}.$$  \hspace{1cm} (13.1)

More generally, we could consider

$$\frac{\tau x + 1}{2^{k(x)}}, \quad \tau \text{ odd.}$$ \hspace{1cm} (13.2)

13.2 Structure Theorem

Given numbers $k_1, k_2, \ldots, k_n$ integers at least 1, what is the set $A$ of $x \in \Pi$ to which one can apply $T^{k_1} \cdots T^{k_n}$?

Theorem 13.2.1 (Structure Theorem). For all quantities as above, with $\epsilon = \pm 1$,

$$A = \Sigma^{k_1, \ldots, k_n, \epsilon} = \left\{ 6\left( g_m(k_1, \ldots, k_m, \epsilon) + 2^{k_1 + \cdots + k_m, \epsilon} \right) + \epsilon \right\},$$ \hspace{1cm} (13.3)
where $0 \leq q_m \leq 2^{k_1 + \cdots + k_m}$, $p \in \mathbb{Z}$.

More generally,

$$
\Lambda_{\tau_m, \delta_m} = \left\{ 6 \left( \tau_m(k_1, \ldots, k_m, \epsilon) + 3^m p \right) + \delta_m \right\},
$$

(13.4)

with $0 \leq \delta_m \leq 3^m$ and $p \in \mathbb{Z}$.

Consequences of the Structure Theorem:

Given $x_0, x_m = T^m x_0$, $y_m = \ln x_m$, $z_m = y_m - y_0 = \ln \frac{x_m}{x_0}$.

Consider the line $-\ln \frac{2}{3} m = z$. Then our plot oscillates around this line.

Assume $x_0$ is large and $x_0 \in \Sigma^{k_1, \ldots, k_m, \epsilon}$. Let $k = k_1 + \cdots + k_m$. For such $x_0$,

$$
x_0 = 6(q_m + 2^k p) + \epsilon
\quad = 6 \cdot 2^k \cdot p \left( 1 + \frac{q_m}{2^k p} + \frac{\epsilon}{6 \cdot 2^k \cdot p} \right),
$$

$$
x_m = 6(\tau + 3^m p) + \delta_m
\quad = 6 \cdot 3 \cdot p \left( 1 + \frac{\tau_m}{3^m p} + \frac{\delta_m}{6 \cdot 3^m p} \right).
$$

(13.5)

Therefore

$$
z_m = \ln \frac{x_m}{x_0}
\quad = \ln \frac{3^m}{2^k} + \alpha(x, m)
\quad = m \ln 3 - k \ln 2 + \alpha(x, m).
$$

(13.6)

For a majority of $k_1, \ldots, k_m$, the sum $k_1 + \cdots + k_m = k$ is close to $2m$.

Therefore, we can rewrite as

$$
z_m = m \ln 3 - 2m \ln 2 - (k - 2m) \ln 2 + \alpha(x, m).
$$

(13.7)

We will show that $k - 2m = O(\sqrt{m})$, as well as $\alpha(x, m)$ is small.
13.3 Density Arguments

Suppose we have an infinite set $A \subset \mathbb{Z}^1$. Consider nice sets such that the limit

$$\frac{|A \cap [1, R]|}{R} \xrightarrow{R \to \infty} \pi(A)$$

exists.

For example, if

$$A = \{6(q + 2^k p) + \epsilon\}, \quad p \in \mathbb{Z}$$

(13.9)

Then $\pi(A) = \frac{1}{6} \cdot 2^k$.

Remembering the definition of $\Pi$ and $\Pi^\pm$, we have $\pi(\Pi^\pm) = \frac{1}{6}$ and $\pi(\Pi) = \frac{1}{3}$.

We define (relative) probabilities: for $A \subset \Pi$,

$$P(A) = \frac{\pi(A)}{\pi(\Pi)} = 3\pi(A).$$

(13.10)

Thus, for the $A$ given above, $P(A) = \frac{1}{2 \cdot 2^k}$.

**Lemma 13.3.1.** Let $k > m$. Then the number of ways of writing $k = k_1 + \cdots + k_m$, each $k_i \geq 1$ an integer, is $\binom{k-1}{m-1}$.

The (relative) probability of numbers which belong to $\Sigma^{k_1,\ldots,k_m,\epsilon}$ with a given $k$ is

$$P = \frac{1}{2 \cdot 2^k} \cdot \binom{k-1}{m-1} \cdot 2 = \frac{1}{2^k} \binom{k-1}{m-1}.$$ 

(13.11)

13.4 Moivre-Laplace Theorem

**Theorem 13.4.1 (Moivre-Laplace Theorem).** Consider $\binom{n}{s} p^s (1 - p)^{n-s}$, $0 \leq s \leq n$. Then

$$\binom{n}{s} p^s (1 - p)^{n-s} \sim \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(s-np)^2}{2np(1-p)}},$$

(13.12)

where $|s - np| \leq C\sqrt{n}$, $n \to \infty$, $p$ and $C$ are fixed.
Thus, for us we have

\[
\frac{1}{2^k} \binom{k-1}{m-1} = \frac{1}{2^{k-1}} \binom{k-1}{m-1} = \frac{1}{2} \cdot \frac{1}{2^{m-1}} 2^{k-1-(m-1)} \binom{k-1}{m-1}.
\] (13.13)

Letting \( n = k - 1, s = m - 1, p = \frac{1}{2} \), the above is equivalent to

\[
\frac{1}{\sqrt{2\pi k}} \cdot \frac{1}{2} e^{-\frac{(m - k)^2}{2k}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(2m-k)^2}{2k}}.
\] (13.14)

Thus, as claimed, for the majority of \( k \), the sum \( k_1 + \cdots + k_m \) is close to \( 2m \).

What is

\[
P \left( \left\{ x_0 : a \leq z_m + m \ln \frac{4}{3} \leq b \right\} \right).
\] (13.15)

Remember

\[
z_m = m \ln 3 - 2m \ln 2 - (k - 2m) \ln 2 + \alpha(x, m)
\]
\[
z_m + m \ln \frac{4}{3} = -(k - 2m) \ln 2 + \alpha(x, m).
\] (13.16)

The above arguments give us that, on average, \( k - 2m \) is very small: \( |k - 2m| \) will usually be of size at most \( C \sqrt{m} \).

Thus, we need the measure of the set

\[
a \leq -\frac{(k - 2m) \ln 2}{\sqrt{m}} + \frac{\alpha(x, m)}{\sqrt{m}} \leq b.
\] (13.17)

For simplicity, we will forget about the \( \alpha(x, m) \). We are reduced to finding the measure of the set

\[
a \leq -\frac{(k - 2m) \ln 2}{\sqrt{m}} \leq b.
\] (13.18)

Thus, we have a (restricted as above) sum over \( k \) of

98
\[
\sqrt{\frac{m}{k \ln 2}} a \leq \frac{k - 2m}{\sqrt{k}} \leq \sqrt{\frac{m}{k \ln 2}}.
\] (13.19)

We now have

\[
\sum \frac{1}{\sqrt{2\pi k}} e^{-\frac{(k-2m)^2}{4k}} \rightarrow \int_{\sqrt{\frac{a}{2\pi \ln 2}}}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du,
\] (13.20)

where we have made the substitution \( u = \frac{k - 2m}{\sqrt{m}} \), and we are dropping in the \( dk\cdot du \) terms the error of size \( O\left(\frac{m}{k^3}\right) \).

### 13.5 Proof of the Structure Theorem

Consider \( \Pi^\epsilon, \epsilon = \pm 1 \). Consider the set of \( x \in \Pi^\epsilon \) where we can apply \( T^k \) to \( x \).

Then \( x = 6p + \epsilon \), and \( 2^k \) must divide \( 3(6p + \epsilon) + 1 \). Therefore,

\[
3(6p + \epsilon) + 1 = 2^k \cdot (6s + \delta).
\] (13.21)

Rewriting yields

\[
6(3p - 2^k s) = 2^k \delta - 3\epsilon - 1.
\] (13.22)

This yields the condition

\[
2^k \delta \equiv 1 \mod 3.
\] (13.23)

If \( k \) is odd, then \( \delta = -1 \); if \( k \) is even, \( \delta = +1 \). Let \( \delta \) satisfy the above congruence. Then

\[
3p - 2^k s = \frac{2^k \delta - 3\epsilon - 1}{6} = c.
\] (13.24)

Then the solutions are

\[
p = p_0 + 2^k t, \quad s = s_0 + 3t,
\] (13.25)

where \( (p_0, s_0) \) is the smallest solution to \( 3p_0 - 2^k s_0 = c \). We are solving a linear equation \( Ax + By = c \); see for example Davenport’s *The Higher Arithmetic*. 
Appendix A

Linear Algebra Review

Matrices can either be thought of as rectangular (often square) arrays of numbers, or as linear transformations from one space to another (or possibly to the same space). The former picture is the simplest starting point, but it is the latter, geometric view that gives a deeper understanding.

To connect with the simpler vector case, a vector can be thought of as a list of real numbers which change in a certain way when the coordinate system changes, or as a geometric object with length and direction. The latter object is coordinate-independent, and has different representations in different choices of coordinate axes. Try to keep the geometric picture in mind for matrices.

A.1 Transposition

Given an $n \times m$ matrix $A$ (where $n$ is the number of rows and $m$ is the number of columns), the transpose of $A$, denoted $A^T$, is the $m \times n$ matrix where the rows of $A^T$ are the columns of $A$ (or, equivalently, the columns of $A^T$ are the rows of $A$).


We leave the proof to the reader.

If an $n \times n$ matrix (also called a square matrix) $A$ satisfies $A^T = A$, then we say $A$ is symmetric.
Example A.1.2. Let $A$ be the matrix

\[
\begin{pmatrix}
2 & 2 & 4 & 2 \\
1 & 1 & -2 & 2 \\
-2 & 0 & 0 & 1 \\
1 & 1 & 2 & 1 \\
\end{pmatrix}
\]  

(A.1)

Then $A^T$ is

\[
\begin{pmatrix}
2 & 1 & -2 & 1 \\
2 & 1 & 0 & 1 \\
4 & -2 & 0 & 2 \\
2 & 2 & 1 & 1 \\
\end{pmatrix}
\]  

(A.2)

Note the above matrix is not symmetric.

The number of degrees of freedom in a symmetric matrix (i.e. independent real numbers needed to completely specify the matrix) is $n(n+1)/2$. Why? There are $n^2$ entries, $n$ on the diagonal. If you specify all entries above the diagonal and all entries on the diagonal, then you know the symmetric matrix.

There are $n^2 - n$ non-diagonal entries (half above the diagonal, half below). Thus, one needs to specify $\frac{n^2-n}{2} + n = \frac{n^2+n}{2}$ entries.

Exercise A.1.3. If $A$ and $B$ are symmetric, show $AB$ is symmetric.

A.2 Matrix Multiplication

We call the element in the $i^{th}$ row and $j^{th}$ column $a_{ij}$. Think of $i = 1 \cdots n$ going down the left side, and $j = 1 \cdots M$ going across the top. A vector $v$ we represent as a column of elements with the $i^{th}$ being $v_i$. A nice way to see matrix-vector multiplication is that the $v_i$ give the coefficients by which the columns of $A$ are linearly mixed together. For the product $w = Av$ to make sense, the length (dimension) of $v$ must equal $m$, and the dimension of $w$ will be $n$. $A$ is therefore a linear transform from $m$-dim space to $n$-dim space.

Multiple transformations appear written backwards: if we apply $A$ then $B$ then $C$ to a vector, we write

\[ w = CBAv. \]  

(A.3)

Note that taking the product of two $n \times n$ matrices requires $O(n^3)$ effort.
Exercise A.2.1. Show that there are two ways to evaluate triple matrix products of the type $CBA$. The slow way involves $O(n^4)$ effort. How about the fast way? How do these results scale for the case of a product of $k$ matrices?

Definition A.2.2 (Invertible Matrices). A is invertible if a matrix $B$ can be found such that $BA = AB = I$. The inverse is then written $B = A^{-1}$. Invertibility requires $A$ to be square.

A.3 Transformations of a Matrix

Just as with vectors, we can find out how the components of a square matrix $A$ change under transformation. Say we have a scalar quantity $x = w^T Av$. We transform our coordinate system linearly such that the vector $v$ has components $v' = Mv$, where $M$ is some invertible matrix representing the transformation. Therefore also $w' = Mw$. The only way that $x$ can remain unchanged by the transformation (as any scalar must), for all choices of $v$ and $w$, is if the transformed matrix is written $A' = M^{-T}AM^{-1}$. Check this via

$$x' = w'^T A' v' = (Mw)^T(M^{-T}AM^{-1})(Mv) = w^T IA Iv = w^T Av = x. \quad (A.4)$$

This is called a similarity transformation, or a conjugation. Really we have one object, the transformation $A$, but it may have different representations by a matrix of numbers, depending on the choice of basis.

A.4 Orthogonal Matrices

Definition A.4.1 (Orthogonal Matrices). $Q$ is an orthogonal $n \times n$ matrix if it has real entries and $Q^T Q = QQ^T = I$.

$Q$ is invertible, with inverse $Q^T$. The geometric meaning of $Q$ is a rotation: the vector $w = Qv$ is just $v$ rotated (about the origin).

The number of degrees of freedom in an orthogonal matrix is $n(n-1)/2$.

Exercise A.4.2. In 3 dimensions a general rotation involves 3 angles (for example, azimuth, elevation, and ‘roll’). How many angles are needed in 4 dimensions? In 3d you rotate about a line-like axis (the set of points which do not move under rotation); what object do you rotate about in 4d?
Exercise A.4.3. Show that the identity matrix $I$, always has representation $I_{ij} = \delta_{ij}$ regardless of the choice of basis. Hint: perform orthogonal transformation on the matrix $\delta_{ij}$.

The set of orthogonal matrices of order $n$ forms a continuous (or topological) group, which we call $O(n)$. (Not to be confused with “of order N”). Group properties:

- Associativity follows from that of matrix multiplication.
- The identity matrix acts as an identity element, since it is in the group.
- Inverse is the transpose (see above): $Q^{-1} = Q^T$.
- Closure is satisfied because any product $QR$ of orthogonal matrices is itself orthogonal.

Exercise A.4.4. Prove the last assertion.

However, not all the elements of $O(n)$ can ‘talk’ to each other, i.e. you cannot reach all the elements by continuous transformation from the identity $I$.

Example for $n = 2$: a general order-2 orthogonal matrix can be written

\[
\begin{pmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{pmatrix}
\quad \text{or} \quad
\begin{pmatrix}
\cos \theta & -\sin \theta \\
-\sin \theta & -\cos \theta
\end{pmatrix},
\]  

(A.5)

where $0 \leq \theta < 2\pi$ is a real angle. The first has determinant $+1$ and defines the ‘special’ (i.e. unit determinant) group $SO(2)$ which is a subgroup of $O(2)$ with identity $I$. The second has determinant $-1$ and corresponds to rotations with a reflection; this subgroup is disjoint from $SO(2)$, and has the weird (reflecting) identity can be written in some basis as

\[
\begin{pmatrix}
1 & 0 \\
0 & -1
\end{pmatrix}.
\]  

(A.6)

Note that $SO(2)$, alternatively written as the family of planar rotations $R(\theta)$, is isomorphic to the unit length complex numbers under the multiplication operation:

\[
R(\theta) \longleftrightarrow e^{i\theta}.
\]  

(A.7)

Therefore we have $R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2)$. This commutativity relation does not hold in higher $n > 2$. 

103
**Orthogonal transformations.** If an orthogonal matrix $Q$ is used for conjugation of a general square matrix $A$, then the rule Eq. A.4 for transformation becomes,

$$A' = QAQ^T.$$  \hspace{1cm} (A.8)

This tells you how to ‘rotate’ a (square) matrix.

### A.5 Definitions

**Definition A.5.1 (Complex Conjugate Transpose).** Let $A$ be an $n \times m$ matrix. Then the complex conjugate transpose of $A$, denoted $A^*$, is obtained by the following: (1) take the complex conjugate of $A$; ie, replace every entry $a_{jk} = x_{jk} + iy_{jk}$ with $\overline{a_{jk}} = x_{jk} - iy_{jk}$, and call this matrix $A_1$; (2) take the transpose of $A_1$.

**Exercise A.5.2.** Prove that $(AB)^* = B^*A^*$.

**Definition A.5.3 (Dot or Inner Product).** The dot (or inner) product of two real vectors $v$ and $w$ is defined as $v^Tw$; if the vectors are complex, we instead use $v^*w$.

**Exercise A.5.4.** Show that the dot product is invariant under orthogonal transformation. That is, show that given two vectors, transforming them using the same orthogonal matrix leaves their dot product unchanged.

**Definition A.5.5 (Length of a vector).** The length of a real vector $v$ is $|v|^2 = v^Tv$; for a complex vector, we have $|v|^2 = v^*v$.

**Definition A.5.6 (Orthogonality).** Two real vectors are orthogonal (also called perpendicular) if $v^Tw = 0$; for two complex vectors, the equivalent condition is $v^*w = 0$.

**Definition A.5.7 (Eigenvalue, Eigenvector).** Recall $\lambda$ is an eigenvalue and $v$ is an eigenvector if $Av = \lambda v$ and $v$ is not the zero vector.

**Exercise A.5.8.** If $v$ is an eigenvector of $A$ with eigenvalue $\lambda$, show $w = av$, $a \in \mathbb{C}$, is also an eigenvector of $A$ with eigenvalue $\lambda$.

**Exercise A.5.9.** Show that given an eigenvalue $\lambda$ and an eigenvector $v$, you can always find an eigenvector $w$ with the same eigenvalue, but $|w| = 1$.

To find the eigenvalues, we solve the equation $\det(\lambda I - A) = 0$. This gives a polynomial $p(\lambda) = 0$. We call $p(\lambda)$ the **characteristic polynomial**.
A.6 Trace

The trace of a matrix $A$, denote $\text{Tr}(A)$ is the sum of the diagonal entries of $A$:

$$\text{Tr}(A) = \sum_{i=1}^{n} a_{ii}. \quad (A.9)$$

Lemma A.6.1. $\text{Tr}(A) = \sum_{i=1}^{n} \lambda_i$.

The proof relies on writing out the characteristic equation and comparing powers of $\lambda$ with the factorized version. By the fact that the polynomial has roots $\lambda_i$, we can write

$$\det(\lambda I - A) = p(\lambda) = \prod_{i=1}^{n} (\lambda - \lambda_i). \quad (A.10)$$

Note the coefficient of $\lambda^n$ is 1, thus we have $\prod_{i} (\lambda - \lambda_i)$ and not $c \prod_{i} (\lambda - \lambda_i)$ for some constant $c$.

By expanding out the RHS, the coefficient of $\lambda^{n-1}$ is $-\sum_{i=1}^{n} \lambda_i$, which we will show is $-\text{Tr}(A)$. Expanding the LHS, we want to find the corresponding coefficient in

$$\det(\lambda I - A) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \ddots & \ddots & \cdots \\ -a_{n1} & \cdots & \cdots & \lambda - a_{nn} \end{vmatrix}.$$ 

We have to remember the expansion of the determinant. Taking the top-left-most $2 \times 2$ block, we see its determinant is $$(\lambda - a_{11})(\lambda - a_{22}) - a_{12}a_{21} = \lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}).$$ The determinant of the top-left-most $3 \times 3$ block is then formed by $(\lambda - a_{33})$ times the above $2 \times 2$ determinant, plus two other multiples of determinants which can give only a highest power of $\lambda$ of $\lambda^3$. Thus we see that the coefficient in $\lambda^2$ is $-(a_{11} + a_{22} + a_{33})$. Repeating this argument for $4 \times 4$ block up to $n \times n$ gives us the coefficient of $\lambda^{n-1}$ in the full determinant is $-\sum_{i=1}^{n} a_{ii}$. Since the LHS and RHS must be equal $\forall \lambda$, the LHS and RHS coefficients in $\lambda^{n-1}$ are equal. \square

Corollary A.6.2. $\text{Tr}(A)$ is invariant under rotation of basis.

The proof follows immediately from the invariance of the eigenvalues under rotation of basis. We need the following:

105
Lemma A.6.3. \( \det(AB) = \det(A) \det(B) \). Further, by induction one can show \( \det(AB \cdots Z) = \det(A) \det(B) \cdots \det(Z) \). Further, \( \det(I) = 1 \).

Proof of Corollary: Let \( A = Q^T B Q \). We show \( A \) and \( B \) have the same trace by showing \( A \) and \( B \) have the same eigenvalues. To find the eigenvalues of \( A \) we must solve:

\[
\det(\lambda I - A) = \det(\lambda I - Q^T B Q)
= \det(\lambda Q^T Q - Q^T B Q)
= \det(Q^T \lambda I Q - Q^T B Q)
= \det \left( Q^T (\lambda I - B) Q \right)
= \det(Q^T) \det(\lambda I - B) \det(Q)
= \det(Q^T) \det(Q) \det(\lambda I - B)
= \det(Q^T Q) \det(\lambda I - B) = \det(I) \det(\lambda I - B) = \det(\lambda I - B).
\]

(A.11)

As the eigenvalues of \( A \) and \( B \) satisfy the same equation, they are equal. \( \square \)

### A.7 Spectral Theorem for Real Symmetric Matrices: Easy Case

The main theorem we will prove is

**Theorem A.7.1 (Spectral Theorem).** Let \( A \) be a real symmetric \( n \times n \) matrix. Then there exists an orthogonal \( n \times n \) matrix \( Q \) and a diagonal matrix \( \Lambda \) such that \( Q^T A Q = \Lambda \). Moreover, the \( n \) eigenvalues of \( A \) are the diagonal entries of \( \Lambda \).

This result is remarkable: it tells you that any real, symmetric matrix is diagonal when rotated into an appropriate basis (recall the rotation effect of conjugation using \( Q \)). In other words, the operation of matrix \( A \) on a vector \( v \) can be broken down into three steps:

\[ Av = Q \Lambda Q^T v = \text{(undo the rotation)(stretch along coord axes)(rotation)} v. \]

(A.12)

Recall the ordering of transformations is read like Hebrew, right to left. The rotation is just the rotation into the basis of eigenvectors.
Furthermore, the eigenvalues $\lambda_i (= \text{diags of } \Lambda)$ are a set of numbers invariant under rotations of $A$. In other words, if $A' = PAP^T$ is an orthogonally-conjugated (i.e. $P$ is orthogonal) version of $A$, then $A'$ has the same $\{\lambda_i\}$ as $A$. Of course the ordering of the $\lambda_i$ has to be chosen the same.

A.7.1 Preliminary Lemmas

For the Spectral Theorem we prove a sequence of needed lemmas:

Lemma A.7.2. The eigenvalues of a real symmetric matrix are real.

Let $A$ be a real symmetric matrix with eigenvalue $\lambda$ and eigenvector $v$. Note that we do not yet know that $v$ has only real coordinates!

Therefore, $Av = \lambda v$. Take the dot (or inner) product of both sides with the vector $v^*$, the complex conjugate transpose of $v$:

$$v^* A v = \lambda v^* v. \quad (A.13)$$

But the left hand side is real. The two sides are clearly complex numbers (ie, 1-dimensional matrices). Taking the complex conjugate transpose of the LHS gives

$$\left(v^*(Av)\right)^* = (Av)^*(v^*)^* = v^* Av. \quad (A.14)$$

Therefore, the LHS is real, implying the RHS is real. But clearly $v^*v$ is real (similar calculation). Thus, $\lambda$ is real. $\square$

We will only prove the Spectral Theorem when all the eigenvalues are distinct. Henceforth, we shall always assume $A$ is a real symmetric matrix.

Lemma A.7.3. The eigenvectors of a real symmetric matrix are real.

The eigenvalues solve the equation $(\lambda I - A)v = 0$. Let $\lambda$ be an eigenvalue. Then $\det(\lambda I - A) = 0$. Therefore the matrix $B = \lambda I - A$ is not invertible. Therefore it send a vector to 0 (standard linear algebra calculation).

Lemma A.7.4. If $\lambda_1$ and $\lambda_2$ are two distinct eigenvalues of a real symmetric matrix $A$, then their corresponding eigenvectors are perpendicular.
We study $v_1^T A v_2$. Now

$$v_1^T A v_2 = v_1^T (A v_2) = v_1^T (\lambda_2 v_2) = \lambda_2 v_1^T v_2. \quad \text{(A.15)}$$

Also,

$$v_1^T A v_2 = v_1^T A^T v_2 = (v_1^T A^T) v_2 = (A v_1)^T v_2 = (\lambda_1 v_1)^T v_2 = \lambda_1 v_1^T v_2. \quad \text{(A.16)}$$

Therefore

$$\lambda_2 v_1^T v_2 = \lambda_1 v_1^T v_2 \quad \text{or} \quad (\lambda_1 - \lambda_2) v_1^T v_2 = 0. \quad \text{(A.17)}$$

As $\lambda_1 \neq \lambda_2$, $v_1^T v_2 = 0$. Thus, the eigenvectors $v_1$ and $v_2$ are perpendicular. □

### A.7.2 Proof of the Spectral Theorem: Easy Case

We can now prove the Spectral Theorem for real symmetric matrices if there are $n$ distinct eigenvectors.

Let $\lambda_1$ to $\lambda_n$ be the $n$ distinct eigenvectors, and let $v_1$ to $v_n$ be the corresponding eigenvectors chosen so that each $v_i$ has length 1.

Consider the matrix $Q$, where the first column of $Q$ is $v_1$, the second column of $Q$ is $v_2$, all the way to the last column of $Q$ which is $v_n$:

$$Q = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} \quad \text{(A.18)}$$

The transpose of $Q$ is

$$Q^T = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \quad \text{(A.19)}$$

**Exercise A.7.5.** Show that $Q$ is an orthogonal matrix. Use the fact that the $v_i$ all have length one, and are orthogonal (perpendicular) to each other.

Consider $Q^T A Q$. This is a matrix, call it $B$. To find its entry in the $i^{th}$ row and $j^{th}$ column, we look at

$$c_i^T B c_j \quad \text{(A.20)}$$

108
where the $e_k$ are column vectors which are 1 in the $k^{th}$ position and 0 elsewhere:

$$e_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (A.21)$$

Thus, we need only show that $e_i^T B e_j = 0$ if $i \neq j$ and equals $\lambda_j$ if $i = j$.

**Exercise A.7.6.** Show $Q e_i = v_i$ and $Q^T v_i = e_i$.

We calculate

$$
e_i^T B e_j = e_i^T Q^T A Q e_j
= (e_i^T Q^T) A (Q e_j)
= (Q e_i)^T A (Q e_j)
= v_i^T A v_j
= v_i^T (A v_j)
= v_i^T \lambda_j v_j = \lambda_j v_i^T v_j. \quad (A.22)$$

As $v_i^T v_j$ equals 0 if $i \neq j$ and 1 if $i = j$, this proves the claim.

Thus, the off-diagonal entries of $Q^T A Q$ are zero, and the diagonal entries are the eigenvalues $\lambda_j$. This shows that $Q^T A Q$ is a diagonal matrix whose entries are the $n$ eigenvalues of $A$. $\square$

Note that, in the case of $n$ distinct eigenvalues, not only can we write down the diagonal matrix, we can easily write down what $Q$ should be. Further, by reordering the columns of $Q$, we see we reorder the positioning of the eigenvalues on the diagonal.
A.8 Spectral Theorem for Real Symmetric Matrices: Harder Case

Let $A$ be a real symmetric matrix acting on $\mathbb{R}^n$. Then $A$ has an orthonormal basis $v_1, \ldots, v_n$ such that $Av_j = \lambda_j v_j$.

Write the inner or dot product $\langle v, w \rangle = v^t w$. As $A$ is symmetric, $\langle Av, w \rangle = \langle v, Aw \rangle$.

Definition A.8.1. $V^\perp = \{w : \forall v \in V, \langle w, v \rangle = 0\}$.

Lemma A.8.2. Suppose $V \subset \mathbb{R}^n$ is an invariant vector subspace under $A$ (if $v \in V$, then $Av \in V$). Then $V^\perp$ is also $A$-invariant: $A(V^\perp) \subset V^\perp$.

This proves the spectral theorem. Suppose we find a $v_0 \neq 0$ such that $Av_0 = \lambda_0 v_0$. Take $V = \{\mu v_0 : \mu \in \mathbb{R}\}$ for the invariant subspace.

By Lemma A.8.2, $V^\perp$ is left invariant under $A$, and is one dimension less. Thus, by whatever method we used to find an eigenvector, we apply the same method on $V^\perp$.

Thus, all we must show is given an $A$-invariant subspace, there is an eigenvector.

Consider

$$\max_{v \text{ with } \langle v, v \rangle = 1} \{\langle Av, v \rangle\}.$$  \hfill (A.23)

Standard fact: every continuous function on a compact set attains its maximum (not necessarily uniquely). See, for example, W. Rudin, Principles of Mathematical Analysis.

Let $v_0$ be a vector giving the maximum value, and denote this maximum value by $\lambda_0$. As $\langle v_0, v_0 \rangle = 1$, $v_0$ is not the zero vector.

Lemma A.8.3. $Av_0 = \lambda_0 v_0$.

Clearly, if $Av_0$ is a multiple of $v_0$ it has to be $\lambda_0$ (from the definition of $v_0$ and $\lambda_0$). Thus, it is sufficient to show

Lemma A.8.4. $\{\mu v_0 : \mu \in \mathbb{R}\}$ is an $A$-invariant subspace.

Proof: let $w$ be an arbitrary vector perpendicular to $v_0$, and $\epsilon$ be an arbitrary small real number. Consider
\langle A(v_0 + \epsilon w), v_0 + \epsilon w \rangle \quad (A.24)

We need to renormalize, as \( v_0 + \epsilon w \) is not unit length; it has length \( 1 + \epsilon^2 \langle w, w \rangle \).

As \( v_0 \) was chosen to maximize \( \langle Av, v \rangle \) subject to \( \langle v, v \rangle = 1 \), after normalizing the above cannot be larger. Thus,

\begin{align*}
\langle A(v_0 + \epsilon w), v_0 + \epsilon w \rangle &= \langle Av_0, v_0 \rangle + 2\epsilon \langle Av_0, w \rangle + \epsilon^2 \langle w, w \rangle. \quad (A.25)
\end{align*}

Normalizing the vector \( v_0 + \epsilon w \) by its length, we see that in Equation A.25, the order \( \epsilon \) terms must be zero. Thus,

\[ \langle Av_0, w \rangle = 0; \quad (A.26) \]

however, this implies \( Av_0 \) is in the space spanned by \( v_0 \) (as \( w \) was an arbitrary vector perpendicular to \( v_0 \)), completing our proof. \( \square \)

**Corollary A.8.5.** Any local maximum will lead to an eigenvalue-eigenvector pair.

The second largest eigenvector (denoted \( \lambda_1 \)) is

\[ \lambda_1 = \max_{\langle v, v_0 \rangle = 0} \frac{\langle Av, v \rangle}{\langle v, v \rangle}. \quad (A.27) \]

We can either divide by \( \langle v, v \rangle \), or restrict to unit length vectors.

**Remark A.8.6.** The proof given above generalizes to a wide variety of spaces. For \( n \times n \) matrices, a more explicit proof can be given. We just need to find one eigenvector \( v_0 \), as clearly scalar multiples of that vector form an \( A \)-invariant space. Consider the characteristic polynomial of \( A \), \( \det(A - \lambda I) \). By the Fundamental Theorem of Algebra, this has at least one root, say \( \lambda_0 \). As \( A - \lambda_0 I \) has determinant 0, we can find a non-zero vector \( v_0 \) such that \( (A - \lambda_0 I)v_0 \) is the zero vector; in other words, we can find the desired eigenvector.
Appendix B

Fourier Analysis and the Equi-Distribution of \( \{ n\alpha \} \)

### B.1 Inner Product of Functions

We define the exponential function by means of the series

\[
e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \tag{B.1}
\]

which converges everywhere. Given the Taylor series expansion of \(\sin x\) and \(\cos x\), we can verify the identity

\[
e^{ix} = \cos x + i\sin x. \tag{B.2}
\]

**Exercise B.1.1.** Prove \(e^x\) converges for all \(x \in \mathbb{R}\) (even better, for all \(x \in \mathbb{C}\). Show the series for \(e^x\) also equals

\[
\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n, \tag{B.3}
\]

which you may remember from compound interest problems.

**Exercise B.1.2.** Prove, using the series definition, that \(e^{x+y} = e^x e^y\). Use this fact to calculate the derivative of \(e^x\). If instead you try to differentiate the series directly, you must justify the derivative of the infinite sum is the infinite sum of the derivatives.
Remember the definition of **inner or dot product**: for two vectors  \( \vec{v} = (v_1, \cdots, v_n) \),  \( \vec{w} = (w_1, \cdots, w_n) \), we take the inner product  \( \vec{v} \cdot \vec{w} \) (also denoted  \( \langle v, w \rangle \)) to mean
\[
\vec{v} \cdot \vec{w} = \langle v, w \rangle = \sum_i v_i w_i. \tag{B.4}
\]

Further, the length of a vector  \( v \) is
\[
|v| = \langle v, v \rangle. \tag{B.5}
\]

We generalize this for functions. For definiteness, assume  \( f \) and  \( g \) are functions from  \([0, 1]\) to \( \mathbb{C} \). Divide the interval  \([0, 1]\) into  \( n \) equal pieces. Then we can represent the functions by
\[
f(x) \longleftrightarrow \left( f(0), f \left( \frac{1}{n} \right), \ldots, f \left( \frac{n - 1}{n} \right) \right), \tag{B.6}
\]
and similarly for  \( g \). Call these vectors  \( f_n \) and  \( g_n \). As before, we consider
\[
\langle f_n, g_n \rangle = \sum_{i=0}^{n-1} f \left( \frac{i}{n} \right) \cdot \overline{g} \left( \frac{i}{n} \right). \tag{B.7}
\]

In general, as we continue to divide the interval (\( n \to \infty \)), the above sum diverges. For example, if  \( f \) and  \( g \) are identically 1, the above sum is  \( n \).

There is a natural rescaling: we multiply each term in the sum by  \( \frac{1}{n} \), the size of the sub-interval. Note for the constant function, the sum is now independent of  \( n \).

Thus, for good  \( f \) and  \( g \) we are led to
\[
\langle f, g \rangle = \lim_{n \to \infty} \sum_{i=0}^{n-1} f \left( \frac{i}{n} \right) \cdot \overline{g} \left( \frac{i}{n} \right) \frac{1}{n} = \int_0^1 f(x) \overline{g(x)} \, dx. \tag{B.8}
\]

The last result follows by Riemann Integration.

**Definition B.1.3.** We say two continuous functions on  \([0, 1]\) are orthogonal (or perpendicular) if their dot product equals zero.

**Exercise B.1.4.** Prove  \( x^n \) and  \( x^m \) are not perpendicular on  \([0, 1]\) for  \( n \neq m \).
We will see that the exponential function behaves very nicely under the inner product. Define
\[ e_n(x) = e^{2\pi i nx} \quad \text{for } n \in \mathbb{Z}. \] (B.9)

Then a straightforward calculation shows
\[ \langle e_n(x), e_m(x) \rangle = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{otherwise.} \end{cases} \] (B.10)

Thus \( e_0(x), e_1(x), e_2(x), \ldots \) are an orthogonal set of functions, which means they are pairwise perpendicular. As each function has length 1, we say the functions \( e_n(x) \) are an orthonormal set of functions.

**Exercise B.1.5.** Prove \( \langle e_n(x), e_m(x) \rangle = 1 \) if \( n = m \) and 0 otherwise.

### B.2 Fourier Series and \( \{n\alpha\} \)

#### B.2.1 Fourier Series

Let \( f \) be continuous and periodic on \( \mathbb{R} \) with period one. Define the \( n \)th Fourier coefficient \( \hat{f}(n) \) of \( f \) to be
\[ \hat{f}(n) = a_n = \langle f(x), e_n(x) \rangle = \int_0^1 f(x)e^{-2\pi nx} \, dx. \] (B.11)

Returning to the intuition of \( \mathbb{R}^m \), we can think of the \( e_n(x) \)’s as an infinite set of perpendicular directions. The above is simply the projection of \( f \) in the direction of \( e_n(x) \).

**Exercise B.2.1.** Show
\[ \langle f(x) - \hat{f}(n)e_n(x), e_n(x) \rangle = 0. \] (B.12)

This agrees with our intuition, namely, that if you remove the projection in a certain direction, what is left is perpendicular to that direction.

The \( N \)th partial Fourier series of \( f \) is
\[ s_N(x) = \sum_{n=-N}^{N} \hat{f}(n)e_n(x). \] (B.13)

114
Exercise B.2.2. Prove

1. \( \langle f(x) - s_N(x), e_n(x) \rangle = 0 \) if \( |n| \leq N \).

2. \( |\hat{f}(n)| \leq \int_0^1 |f(x)| \, dx \).

3. If \( \langle f, f \rangle < \infty \), then \( \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2 \leq \langle f, f \rangle \).

4. If \( \langle f, f \rangle < \infty \), then \( \lim_{|n| \to \infty} \hat{f}(n) = 0 \).

As \( \langle f(x) - s_N(x), e_n(x) \rangle = 0 \) if \( |n| \leq N \), we might think that we just have to let \( N \) go to infinity to obtain a series \( s_\infty \) such that

\[
\langle f(x) - s_\infty(x), e_n(x) \rangle = 0.
\]  \( \text{(B.14)} \)

Assume that for a periodic function \( g(x) \) to be orthogonal to \( e_n(x) \) for every \( n \) it must be zero for every \( x \). Then \( f(x) - s_\infty(x) = 0 \), and hence \( f = s_\infty \). Voilá – an expression for \( f \) as a sum of exponentials! Be careful, however. We have just glossed over the two central issues – completeness and, even worse, convergence.

We will now see a way of avoiding some of our problems.

B.2.2 Weighted partial sums

Define

\[
D_N(x) = \sum_{n=-N}^{N} e_n(x) = \frac{\sin((2N + 1)\pi x)}{\sin \pi x},
\]

\[
F_N(x) = \frac{\sin^2(N\pi x)}{N \sin^2 \pi x} = \frac{1}{N} \sum_{n=0}^{N-1} D_n(x).
\]  \( \text{(B.15)} \)

Here \( F \) stands for Féjer, \( D \) for Dirichlet. In general, functions which we are interested in taking their inner product against \( f \) are called kernels; thus, the Dirichlet kernel, the Féjer kernel, etc.

Note that, no matter what \( N \) is, \( F_N(x) \) is positive for all \( x \).

We say that a sequence \( f_1(x), f_2(x), f_3(x), \ldots \) of functions is an approximation to the identity if

1. \( f_N(x) \geq 0 \) for all \( x \) and every \( N \);

2. \( \int_0^1 f_N(x) \, dx = 1 \);
3. \( \lim_{N \to \infty} \int_{\delta}^{1-\delta} f_N(x) \, dx = 0 \) if \( 0 < \delta < \frac{1}{2} \).

**Theorem B.2.3.** The Féjer kernels \( F_1(x), F_2(x), F_3(x), \ldots \) are an approximation to the identity.

**Proof:** The first property is immediate. The second follows from the observation that \( F_N(x) \) can be written as

\[
F_N(x) = e_0(x) + \frac{N-1}{N} \left( e_{-1}(x) + e_1(x) \right) + \cdots, \tag{B.16}
\]

and all integrals are zero but the first, which is 1.

To prove the third property, note that \( F_N(x) \leq \frac{1}{N \sin^2 \pi \delta} \) for \( \delta \leq x \leq 1 - \delta \). □

Let \( f \) be a continuous, periodic function on \( \mathbb{R} \) with period one. Thus, we can consider \( f \) as a function on just \([0, 1]\), with \( f(0) = f(1) \). Define

\[
T_N(x) = \int_0^1 f(y) F_N(x - y) \, dy. \tag{B.17}
\]

Recall the following definition and theorem:

**Definition B.2.4 (Uniform Continuity).** A continuous function is uniformly continuous if given an \( \epsilon > 0 \), there exists a \( \delta > 0 \) such that \( |x - y| < \delta \) implies \( |f(x) - f(y)| < \epsilon \). Note that the same \( \delta \) works for all points.

**Theorem B.2.5.** Any continuous function on a closed, compact interval is uniformly continuous.

**Exercise B.2.6.** Show \( x^n \) is uniformly continuous on \([a, b]\) for \( -\infty < a < b < \infty \).

**Theorem B.2.7.** Given \( \epsilon > 0 \), there is an \( N \) such that

\[
|f(x) - T_N(x)| \leq \epsilon \tag{B.18}
\]

for every \( x \in [0, 1] \).

**Proof.** For any positive \( N \),
\[
T_N(x) - f(x) = \int_0^1 f(x - y) F_N(y) dy - f(x) \cdot 1
\]
\[
= \int_0^1 f(x - y) F_N(y) dy - \int_0^1 f(x) F_N(y) dy \quad \text{(property 2 of } F_N) 
\]
\[
= \int_0^\delta \left( f(x - y) - f(x) \right) F_N(y) dy 
+ \int_\delta^{1-\delta} \left( f(x - y) - f(x) \right) F_N(y) dy 
+ \int_{1-\delta}^1 \left( f(x - y) - f(x) \right) F_N(y) dy. 
\] (B.19)

Let \( \delta \in (0, 1/2) \). Then, using the fact that the \( F_N(x) \)'s are an approximation to the identity, we find

\[
\left| \int_\delta^{1-\delta} \left( f(x - y) - f(x) \right) F_N(y) dy \right| \leq 2 \max |f(x)| \cdot \int_\delta^{1-\delta} F_N(y) dy. \quad \text{(B.20)}
\]

Since

\[
\lim_{N \to \infty} \int_\delta^{1-\delta} F_N(y) dy = 0, \quad \text{(B.21)}
\]

we obtain

\[
\lim_{N \to \infty} \int_\delta^{1-\delta} \left(f(x - y) - f(x) \right) F_N(y) dy = 0. \quad \text{(B.22)}
\]

Thus, by choosing \( N \) large enough (where large depends on \( \delta \)), we can insure that this piece is at most \( \frac{\epsilon}{3} \).

It remains to estimate what happens near zero. Since \( f \) is continuous and \([0, 1]\) is compact, \( f \) is uniformly continuous. Thus, we can choose \( \delta \) small enough that \( |f(x - y) - f(x)| < \frac{\epsilon}{3} \) for any \( x \) and any positive \( y < \delta \). Then

\[
\left| \int_0^\delta \left( f(x - y) - f(x) \right) F_N(y) dy \right| \leq \int_0^\delta \frac{\epsilon}{3} F_N(y) dy \leq \frac{\epsilon}{3} \int_0^1 F_N(y) dy \leq \frac{\epsilon}{3}. \] (B.23)
Similarly
\[ \left| \int_{1-\delta}^{1} \left( f(x-y) - f(x) \right) F_N(y) \, dy \right| \leq \frac{\epsilon}{3}. \quad (B.24) \]

Therefore
\[ |T_N(x) - f(x)| \leq \epsilon \quad (B.25) \]
for all \( N \) sufficiently large. \( \square \)

**Definition B.2.8 (Trigonometric Polynomials).** Any finite linear combination of the functions \( e_n(x) \) is called a trigonometric polynomial.

From Theorem B.2.7 we immediately get the Stone-Weierstrass theorem:

**Theorem B.2.9 (Stone-Weierstrass).** Any continuous period function can be uniformly approximated by trigonometric polynomials.

### B.2.3 Equidistribution

We say that a sequence \( \{x_n\}, x_n \in [0, 1] \) is *equidistributed* if
\[
\lim_{N \to \infty} \frac{1}{2N+1} \# \{ n : |n| \leq N, x_n \in (a, b) \} = b - a \quad (B.26)
\]
for all \( (a, b) \subset [0, 1] \).

**Theorem B.2.10 (Weyl).** Let \( \alpha \) be an irrational number in \([0, 1]\). Let \( x_n = \{n\alpha\} \), where \( \{y\} \) denotes the fractional part of \( y \). Then the sequence \( \{x_n\} \) is equidistributed.

**Proof.** We will estimate \( \frac{1}{2N+1} \sum_{n=-N}^{N} \chi_{(a,b)}(x_n) \) as \( N \to \infty \), where \( \chi_{(a,b)} \) is the function taking the value 0 outside \((a,b)\) and 1 inside \((a,b)\). We call \( \chi_{(a,b)} \) the characteristic function of the interval \((a,b)\).

Thus, we must show
\[
\lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \chi_{(a,b)}(x_n) = b - a. \quad (B.27)
\]

Consider \( e_k(x) = e^{2\pi ikx} \). Since \( x_n = \{n\alpha\} = n\alpha - \lfloor n\alpha \rfloor \) and \( e_k(x) = e_k(x + m) \) for every integer \( m \),
\[ e_k(x_n) = e^{2\pi i k \alpha}. \] (B.28)

Hence
\[
\frac{1}{2N+1} \sum_{n=-N}^{N} e_k(x_n) = \frac{1}{2N+1} \sum_{n=-N}^{N} e_k(n \alpha) = \frac{1}{2N+1} \sum_{n=-N}^{N} (e^{2\pi i k \alpha})^n
\]
(B.29)
\[
= \begin{cases} 
1 & \text{if } k = 0 \\
\frac{1}{2N+1} e_k(-N \alpha) - e_k((N+1) \alpha) & \text{if } k > 0.
\end{cases}
\]

Now for a fixed irrational \( \alpha \), \( |1 - e_k(\alpha)| > 0 \). Therefore if \( k \neq 0 \):
\[
\lim_{N \to \infty} \frac{1}{2N+1} \frac{e_k(-N \alpha) - e_k((N+1) \alpha)}{1 - e_k(\alpha)} = 0. \] (B.30)

Let \( P(x) = \sum_k a_k e_k(x) \) be a finite sum (ie, \( P(x) \) is a trigonometric polynomial). By possibly adding some zero coefficients, we can write \( P(x) \) as a sum over a symmetric range: \( P(x) = \sum_{K=-K}^{K} a_k a_k(x) \).

**Exercise B.2.11.** Show \( \int_0^1 P(x) \, dx = a_0 \).

By the above arguments, we have shown that for any (finite) trigonometric polynomial \( P(x) \):
\[
\lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} P(x_n) \to a_0 = \int_0^1 P(x) \, dx. \] (B.31)

Consider two approximations to the characteristic function \( \chi(a,b) \):

1. \( f_{1m} \): \( f_{1m}(x) = 1 \) if \( a + \frac{1}{m} \leq x \leq b - \frac{1}{m} \), drops linearly to 0 at \( a \) and \( b \), and is zero elsewhere.

2. \( f_{2m} \): \( f_{1m}(x) = 1 \) if \( a \leq x \leq b \), drops linearly to 0 at \( a - \frac{1}{m} \) and \( b + \frac{1}{m} \), and is zero elsewhere.
Note there are trivial modifications if \( a = 0 \) or \( b = 1 \). Clearly

\[
f_{1m}(x) \leq \chi_{(a,b)}(x) \leq f_{2m}(x).
\]

Therefore

\[
\frac{1}{2N+1} \sum_{n=-N}^{N} f_{1m}(x_n) \leq \frac{1}{2N+1} \sum_{n=-N}^{N} \chi_{(a,b)}(x_n) \leq \frac{1}{2N+1} \sum_{n=-N}^{N} f_{2m}(x_n).
\] (B.32)

By Theorem B.2.7, for each \( m \), given \( \epsilon > 0 \) we can find trigonometric polynomials \( P_{1m}(x) \) and \( P_{2m}(x) \) such that \( |P_{1m}(x) - f_{1m}(x)| < \epsilon \) and \( |P_{2m}(x) - f_{2m}(x)| < \epsilon \).

As \( f_{1m} \) and \( f_{2m} \) are continuous functions, we can replace

\[
\frac{1}{2N+1} \sum_{n=-N}^{N} f_{1m}(x_n) \quad \text{with} \quad \frac{1}{2N+1} \sum_{n=-N}^{N} P_{1m}(x_n)
\]

at a cost of at most \( \epsilon \).

As \( N \to \infty \),

\[
\frac{1}{2N+1} \sum_{n=-N}^{N} P_{1m}(x_n) \to \int_{0}^{1} P_{1m}(x)dx.
\] (B.35)

But \( \int_{0}^{1} P_{1m}(x)dx = (b - a) - \frac{1}{m} \) and \( \int_{0}^{1} P_{2m}(x)dx = (b - a) + \frac{1}{m} \). Therefore, given \( m \) and \( \epsilon \), we can choose \( N \) large enough so that

\[
(b - a) - \frac{1}{m} - \epsilon \leq \frac{1}{2N+1} \sum_{n=-N}^{N} \chi_{(a,b)}(x_n) \leq (b - a) + \frac{1}{m} + \epsilon.
\] (B.36)

Letting \( m \) tend to \( \infty \) and \( \epsilon \) tend to 0, we see \( \frac{1}{2N+1} \sum_{n=-N}^{N} \chi_{(a,b)}(x_n) \to b - a. \)

\( \square \)

**Exercise B.2.12.** Rigorously do the necessary book-keeping to prove the previous theorem.

**Exercise B.2.13.** Prove

1. If \( \alpha \in \mathbb{Q} \), then \( \{n\alpha\} \) is periodic.

2. If \( \alpha \not\in \mathbb{Q} \), then no two \( \{n\alpha\} \) are equal.
Appendix C

Introduction to Probability

C.1 Probabilities of Discrete Events

C.1.1 Introduction

Let $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}$ be an at most countable set of events; later we will deal with probabilities of continuous events. We call $\Omega$ the sample (or outcome) space. We call the elements $\omega \in \Omega$ the events. Let $x : \Omega \to \mathbb{R}$. That is, for each event $\omega \in \Omega$, we attach a real number $x(\omega)$. We call $x$ a random variable.

Example C.1.1. Flip a fair coin 3 times. The possible outcomes are

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}. \quad \text{(C.1)}$$

One possible random variable is $x(\omega)$ equals the number of heads in $\omega$. Thus, $x(HHT) = 2$ and $x(TTT) = 0$.

Example C.1.2. Let $\Omega$ be the space of all flips of a fair coin where all but the last flip are tails, and the last is a head. Thus, $\Omega = \{H, TH, TTH, TTT, \ldots\}$. One possible random variable is $x(\omega)$ is the number of tails; another is $x(\omega)$ equals the number of the flip which is a head.

We say $p(\omega)$ is a probability function on $\Omega$ if

1. $0 \leq p(\omega_i) \leq 1$ for all $\omega_i \in \Omega$.
2. $p(\omega) = 0$ if $\omega \notin \Omega$. 

121
3. \( \sum_i p(\omega_i) = 1. \)

We call \( p(\omega) \) the probability of event \( \omega \).

Often, we have a random variable where \( x(\omega) = \omega \). In a convenient abuse of notation, we write \( X \) for \( \Omega \) and \( x \) for \( x(\omega) \) and \( \omega \).

For example, consider a roll of a fair die. The outcome space is \( \Omega = \{1, 2, 3, 4, 5, 6\} \), and the probability of each \( \omega \in \Omega \) is \( \frac{1}{6} \). Let \( x(\omega) \) be the value when we roll \( \omega \).

Then \( x(1) = 1, x(2) = 2 \), and so on. In this example, it is very convenient to call the event space the number rolled. Denoting this space by \( X \), instead of writing \( x(x) = x \) we just write \( x \).

We will often talk about the probability that \( x(\omega) = a \). This means the subset of \( \omega \in \Omega \) such that \( x(\omega) = a \), and we denote this by \( \text{Prob}(x(\omega) = a) \). Sometimes we will introduce a new outcome space and a new space of events.

For example, consider the space of three flips of a fair coin. We can define a new outcome space \( X = \{0, 1, 2, 3\} \) (C.2)

with \( p(0) = \frac{1}{8}, p(1) = \frac{3}{8}, p(2) = \frac{3}{8} \) and \( p(3) = \frac{1}{8} \). We will often write \( p(a) \) here instead of \( p(x(\omega) = a) \).

**Definition C.1.3** \( (x(\Omega)) \).

\[
x(\Omega) = \{ r \in \mathbb{R} : \exists \omega \in \Omega \text{ with } x(\omega) = r \}.
\]

(C.3)

In other words, \( x(\Omega) \) is the set of values attained by \( x(\omega) \) as we vary \( \omega \in \Omega \).

Given a set \( S \subset x(\Omega) \), we let \( x^{-1}(S) = \{ \omega \in \Omega : x(\omega) \in S \} \).

We often call a subset \( A \subset \Omega \) an event, and write

\[
\text{Prob}(A) = \sum_{\omega \in A} p(\omega).
\]

(C.4)

**C.1.2 Conditional Probabilities**

More generally, consider two outcome spaces \( \Omega_1 \) and \( \Omega_2 \) with events \( \omega_1 \) and \( \omega_2 \).

We can look a new outcome space

\[
\Omega = \{ \omega = (\omega_1, \omega_2) : \omega_1 \in \Omega_1 \text{ and } \omega_2 \in \Omega_2 \}.
\]

(C.5)
We need to define a probability function \( p(\omega) \), ie, we need to assign probabilities to these events. One natural way is as follows: let \( p_i \) be the probability function for events \( \omega_i \in \Omega_i \). We let
\[
p(\omega) = p_1(\omega_1) \cdot p_2(\omega_2) \text{ if } \omega = (\omega_1, \omega_2).
\] (C.6)

**Exercise C.1.4.** Show the above actually defines a probability function.

Of course, we could also define a probability function \( p : \Omega \to \mathbb{R} \) directly. We again consider two tosses of a fair coin. We have events \( \omega = (x, y) \). Let us define \( p(\omega) = \frac{1}{36} \), ie, each of the 36 outcomes is equally likely. Let \( x(\omega) = \omega_1 \), the roll of the first die; similarly, set \( y(\omega) = \omega_2 \), the roll of the second die.

**Example C.1.5.** What is \( \text{Prob}(x(\omega) = 2) \)?

There are 6 pairs with first roll 2: (2, 1), (2, 2), \ldots, (2, 6). Each pair has probability \( \frac{1}{36} \). Thus, \( \text{Prob}(x = 2) = \frac{6}{36} = \frac{1}{6} \).

More generally, we would have
\[
\text{Prob}(x(\omega) = a) = \sum_{\omega = (\omega_1, \omega_2) \atop x(\omega) = a} p(\omega).
\] (C.7)

The above is a simple recipe to find \( \text{Prob}(x(\omega_1) = a) \): it is the probability of all pairs \((\omega_1, \omega_2)\) such that \(x(\omega) = a\), \(\omega_2\) arbitrary.

Let us consider a third random variable, the sum of the two rolls. Thus, let \( z(\omega) = \omega_1 + \omega_2 \), each pair \(\omega = (\omega_1, \omega_2)\) occurs with probability \( \frac{1}{36} \).

We have just seen that, if we have no information about the second roll, the probability that the first roll is a 2 is \( \frac{1}{6} \) (what we would expect).

What if, however, we know the sum of the two rolls is 2, or 7, or 10? Now what is the probability that the first roll is a 2?

Thus, we are looking for pairs \((\omega_1, \omega_2)\) such that \(\omega_1 = 2\) and \(\omega_1 + \omega_2 = 2, 7,\) or 10. A quick inspection shows there are no pairs with sum 2 or 10. For a sum of 7, only one pair works: (2, 5).

This leads us to the concept of conditional probability: what is the probability of an event \(A\), given event \(B\) has occurred?

We can write
\[
\text{Prob}(A) = \frac{\sum_{\omega \in A} p(\omega)}{\sum_{\omega \in \Omega} p(\omega)}.
\] (C.8)
Note the denominator is 1. For conditional probabilities, we restrict to $\omega \in B$. Thus, we have

$$\text{Prob}(A|B) = \text{Prob}(A, \text{ given } B) = \frac{\sum_{\omega \in A} p(w)}{\sum_{\omega \in B} p(\omega)}.$$  \hspace{1cm} (C.9)

The numerator above may be regarded as the event $A \cap B$. Thus,

**Lemma C.1.6.**

$$\text{Prob}(A|B) = \frac{\text{Prob}(A \cap B)}{\text{Prob}(B)}. \hspace{1cm} (C.10)$$

In the example above, if the sum of the rolls is 7, then

$$A = \{(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6)\}$$

$$B = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$$

$$A \cap B = \{(2, 5)\}. \hspace{1cm} (C.11)$$

Thus,

$$\text{Prob}(A|B) = \frac{\text{Prob}(A \cap B)}{\text{Prob}(B)} = \frac{\frac{1}{36}}{\frac{6}{36}} = \frac{1}{6}. \hspace{1cm} (C.12)$$

### C.1.3 Independent Events

The concept of *independence* is one of the most useful in probability. Simply put, two events are independent if knowledge of one gives no information about the other. Rigorously,

$$\text{Prob}(A|B) = \frac{\text{Prob}(A \cap B)}{\text{Prob}(B)} = \text{Prob}(A). \hspace{1cm} (C.13)$$

Thus, knowing event $B$ occurred gives us no additional information on the probability that event $A$ occurred.

Again, consider two rolls of a fair die with outcome space $\Omega$ consisting of pairs of rolls $\omega = (\omega_1, \omega_2)$. Let $x(\omega) = \omega_1$ (the result of the first roll), $y(\omega) = \omega_2$ (the result of the second roll), and $z(\omega) = \omega_1 + \omega_2$ (the sum of the two rolls).

Let $A$ be the event that the first roll is 2 and $B$ the event that the sum of the two rolls is 7.
We have shown
\[ \Pr(A | B) = \frac{1}{6} = \Pr(A); \quad (C.14) \]
thus, \( A \) and \( B \) are independent events. If, however, we had taken \( B \) to be the event that the sum of the two rolls is 2 (or 10), we would have found
\[ \Pr(A | B) = 0 \neq \Pr(A); \quad (C.15) \]
in this case, the two events are not independent.

**Exercise C.1.7.** Let \( A \) be the event that the first roll equals \( a \); let \( B \) be the event that the second roll equals \( b \). Prove \( A \) and \( B \) are independent for any choices of \( a, b \in \{1, \ldots, 6\} \).

We rewrite the definition of independence in a more useful manner. Since, for two independent events \( A \) and \( B \),
\[ \Pr(A | B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \Pr(A), \quad (C.16) \]
we have
\[ \Pr(A \cap B) = \Pr(A)\Pr(B). \quad (C.17) \]

Note the more symmetric form of the above.
In general, events \( A_1, \ldots, A_n \) are independent if for any subset \( \{i_1, \ldots, i_k\} \) of \( \{1, \ldots, n\} \) we have
\[ \Pr(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \Pr(A_{i_1})\Pr(A_{i_2})\cdots\Pr(A_{i_k}). \quad (C.18) \]

If events \( A_1, \ldots, A_n \) are pairwise independent, it is possible that the events are not independent.

**Exercise C.1.8.** Consider again two flips of a fair coin, each pair occurs with probability \( \frac{1}{36} \). Let \( A \) be the event that the first toss is even, \( B \) the event that the second toss is odd, and \( C \) the event that the sum of the tosses is odd. Prove the events are pairwise independent, but not independent.
More generally, let \( A \) denote the first roll of a fair dice, and \( B \) denote the second roll. As each pair of rolls is equally likely, the probability that the first roll is \( a \) is \( \frac{1}{6} \) (as six of the thirty-six pairs will give a first roll of \( a \)). Thus, for any choices of \( a \) and \( b \), the result of the first roll is independent of the second roll, and we say that the two rolls (or the events \( A \) and \( B \)) are independent.

Consider now event \( C \), the sum of the two rolls. If the sum of the rolls is 7, then the probability that the first roll equals \( a \) is \( \frac{1}{6} \) for all \( a \); however, in general the conditional probabilities for the first roll will depend on the sum. For example, if the sum is 2 then the probability that the first roll is 1 is 1 and the probability that the first roll is 2 or more is 0. Thus, event \( A \) and \( C \) (the first roll and the sum of the rolls) are not independent.

In general, if even \( A \) has possible values \( a_1, a_2, \ldots, a_M \) with probabilities \( p(a_i) \) and event \( B \) has possible values \( b_1, b_2, \ldots, b_N \) with probabilities \( q(b_j) \), then \( A \) and \( B \) are independent if

\[
\text{Prob}(a_i|b_j) = \frac{\text{Prob}(A = a_i \text{ and } B = b_j)}{\text{Prob}(B = b_j)} = \text{Prob}(A = a_i) = p(a_i). \quad (C.19)
\]

Thus, if \( A \) and \( B \) are independent,

\[
\text{Prob}((A = a_i) \cap (B = b_j)) = \text{Prob}(A = a_i)\text{Prob}(B = b_j) = p(a_i)q(b_j).
\]

(C.20)

### C.1.4 Means

If \( x(\omega) = \omega \), the mean (or expected value) of an event \( x \) is defined by

\[
\bar{x} = \sum_i \omega_i p(\omega_i) \quad \text{or} \quad \sum_i x_i p(x_i). \quad (C.21)
\]

More generally, for a sample space \( \Omega \) with events \( \omega \) and a random variable \( x(\omega) \), we have

\[
\bar{x}(\omega) = \sum_i x(\omega_i)p(\omega_i). \quad (C.22)
\]

For example, the mean of one roll of a fair die is 3.5.
Exercise C.1.9. Consider all finite fair tosses of a coin where all but the last toss are tails (and the last toss is heads). We denote the outcome space by either
\[ \Omega = \{H, TH, TTH, TTTH, \ldots \} \]  
(C.23)  
or
\[ X = \{1, 2, 3, 4, \ldots \}. \]  
(C.24)  

Let \( x \) be the random variable equal to the number of tosses before getting the first head. For example, \( x(TTH) = 3 \).
Calculate \( p(x = x_i) \), the probability that the first head is after the \( x_i \)th toss; by a convenient abuse of notation, we will write \( p(x = x_i) \) or \( p(x_i) \).
Calculate \( \bar{x}(\omega) \) (or \( \bar{x} \)).

Instead of writing \( \bar{x} \), we often write \( E[x] \) or \( E[X] \), read as the expected value of \( x \) or \( X \). More generally, we would have \( \bar{x}(\omega) \) and \( E[x(\omega)] \).

The \( k \)th moment of \( X \) is the expected value of \( x^k \):
\[ E[x^k] = \sum_i x_i^k p(x_i) \]  
(C.25)  
or
\[ E[x^k(\omega)] = \sum_i x_i^k(\omega_i)p(\omega_i). \]  
(C.26)

Lemma C.1.10 (Additivity of the Means). Let \( x \) and \( y \) be two events with joint probability
\[ \text{Prob}(x(\omega) = x_i, y(\omega) = y_j) = p(x_i, y_j). \]  
(C.27)  

Let \( z = x + y \). Then \( E[z] = E[x + y] = E[x] + E[y] \).

Proof: First, note
\[ \text{Prob}(x(\omega) = x_i) = \sum_j \text{Prob}(x(\omega) = x_i, y(\omega) = y_j) = \sum_j p(x_i, y_j). \]  
(C.28)

Thus, the expected value of \( x(\omega) \) is
\[ \bar{x}(\omega) = \sum_i x_i \sum_j p(x_i, y_j), \] (C.29)

and similarly for \( y(\omega) \). Thus,

\[ E[x + y] = \sum_{(i,j)} (x_i + y_j)p(x_i, y_j) \]
\[ = \sum_i \sum_j x_i p(x_i, y_j) + \sum_i \sum_j y_j p(x_i, y_j) \]
\[ = \sum_i x_i \sum_j p(x_i, y_j) + \sum_j y_j \sum_i p(x_i, y_j) \]
\[ = E[x] + E[y]. \] (C.30)

The astute reader may notice that some care is needed to interchange the order of summations. If \( \sum_i \sum_j |x_i y_j| p(x_i, y_j) < \infty \), then Fubini’s Theorem is applicable, and we may interchange the summations at will.

Given an outcome space \( X = \{x_1, x_2, \ldots \} \) with probabilities \( p(x_i) \), let \( aX \) be shorthand for the event \( a \) times \( X \) with outcome space \( \{ax_1, ax_2, \ldots \} \) and probabilities \( p_a(ax_i) = p(x_i) \).

**Lemma C.1.11.** Let \( x_1 \) through \( x_N \) be a finite collection of random variables. Let \( a_1 \) through \( a_N \) be real constants. Then

\[ E[a_1 x_1 + \cdots + a_N x_N] = a_1 E[x_1] + \cdots + a_N E[x_N]. \] (C.31)

**Exercise C.1.12.** Prove Lemmas C.1.11.

**Lemma C.1.13.** Let \( X \) and \( Y \) be independent events. Then \( E[xy] = E[x]E[y] \).

As the two events are independent,

\[ \text{Prob}(X = x_i \text{ and } Y = y_j) = \text{Prob}(X = x_i) \text{Prob}(Y = y_j) = p(x_i)q(y_j). \] (C.32)

In words, if \( r(x_i, y_j) \) is the probability that event \( X \) is \( x_i \) and event \( Y \) is \( y_j \), then independence implies \( r(x_i, y_j) = p(x_i)q(y_j) \) for two probability functions \( p \) and \( q \). Thus,
\[ E[xy] = \sum_i \sum_j x_i y_j r(x_i, y_j) \]
\[ = \sum_i sum_j x_i y_j p(x_i) q(y_j) \]
\[ = \sum_i x_i p(x_i) \cdot \sum_j y_j q(y_j) \]
\[ = E[x] \cdot E[y]. \quad (C.33) \]

C.1.5 Variances

The variance \( \sigma_x^2 \) (and its square-root, the standard deviation \( \sigma_x \)) measure how spread out a probability distribution is. Assume \( x(\omega) = \omega \). Given an event \( X \) with mean \( \bar{x} \), we define the standard deviation \( \sigma_x^2 \) by

\[ \sigma_x^2 = \sum_i (x_i - \bar{x})^2 p(x_i). \quad (C.34) \]

More generally, given a sample space \( \Omega \), events \( \omega \), and a random variable \( x : \Omega \to \mathbb{R} \),

\[ \sigma_{x(\omega)}^2 = \sum_i \left( x(\omega_i) - \bar{x}(\omega) \right) p(\omega_i). \quad (C.35) \]

Exercise C.1.14. Let \( X = \{0, 25, 50, 75, 100\} \) with probabilities \{.2, .2, .2, .2, .2\}. Let \( Y \) be the same outcome space, but with probabilities \{.1, .25, .3, .25, .1\}. Calculate the means and the variances of \( X \) and \( Y \).

For computing variances, instead of equation C.34 one often uses

Lemma C.1.15. \( \sigma_x^2 = E[x^2] - E[x]^2 \).

Proof: Recall \( \bar{x} = E[x] \). Then

129
\[
\sigma_x^2 = \sum_i \left( x_i - E[x] \right)^2 p(x_i)
\]
\[
= \sum_i (x_i^2 - 2x_i E[x] + E[x]^2) p(x_i)
\]
\[
= \sum_i x_i^2 p(x_i) - 2E[x] \sum_i x_i p(x_i) + E[x]^2 \sum_i p(x_i)
\]
\[
= E[x^2] - 2E[x]^2 + E[x]^2 = E[x^2] - E[x]^2. \tag{C.36}
\]

The main result on variances is

**Lemma C.1.16 (Variance of a Sum).** Let \( X \) and \( Y \) be two independent events.

Then \( \sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 \).

Proof: We constantly use the expected value of a sum is the sum of expected values (Lemma C.1.10 and Lemma C.1.11).

\[
\sigma_{x+y}^2 = E[(x+y)^2] - E[(x+y)]^2
\]
\[
= E[x^2 + 2xy + y^2] - \left( E[x] + E[y] \right)^2
\]
\[
= (E[x^2] + 2E[xy] + E[y^2]) - \left( E[x]^2 + 2E[x]E[y] + E[y]^2 \right)
\]
\[
= \left( E[x^2] - E[x]^2 \right) + \left( E[y^2] - E[y]^2 \right) + 2 \left( E[xy] - E[x]E[y] \right)
\]
\[
= \sigma_x^2 + \sigma_y^2 + 2 \left( E[xy] - E[x]E[y] \right). \tag{C.37}
\]

By Lemma C.1.13, \( E[xy] = E[x]E[y] \), completing the proof.

**Lemma C.1.17.** Consider \( n \) independent copies of the same event (for example, \( n \) flips of a coin or \( n \) rolls of a die). Then \( \sigma_{nx}^2 = \sqrt{n} \sigma_x^2 \).

**Exercise C.1.18.** Prove Lemma C.1.17.

**Lemma C.1.19.** Given an outcome space \( X \) with events \( x_i \) with probabilities \( p(x_i) \), consider the new random variable \( ax + b \). Then

\[
\sigma_{ax+b}^2 = a^2 \sigma_x^2. \tag{C.38}
\]
Exercise C.1.20. Prove C.1.19.

Note that, if the event $X$ has units of meters, then the variance $\sigma_x^2$ has units meters-squared, and the standard deviation $\sigma_x$ and the mean $\bar{x}$ have units meters. Thus, it is the standard deviation that gives a good measure of the deviations of an event around the mean.

There are, of course, alternate measures one can use. For example, one could consider

$$\sum_i (x_i - \bar{x})p(x_i). \quad \text{(C.39)}$$

Unfortunately, this is a signed quantity, and large positive deviations can cancel with large negatives. This leads us to consider

$$\sum_i |x_i - \bar{x}|p(x_i). \quad \text{(C.40)}$$

While this has the advantage of avoiding cancellation of errors (as well as having the same units as the events), the absolute value function is not a good function analytically. For example, it is not differentiable. This is primarily why we consider the standard deviation (the square-root of the variance).

Exercise C.1.21. Consider the following set of data: for $i \in \{1, \ldots, n\}$, given $x_i$ one observes $y_i$. Believing that $X$ and $Y$ are linearly related, find the best fit straight line. Namely, determine constants $a$ and $b$ that minimize the error (calculated via the variance)

$$\sum_{i=1}^n (y_i - (ax_i + b))^2 = \sum_{i=1}^n (\text{Observed}_i - \text{Predicted}_i)^2. \quad \text{(C.41)}$$

Hint: use Multi-variable Calculus to find linear equations for $a$ and $b$, and then solve with Linear Algebra.

If instead of measuring total error by the squares of the individual error (for example, using the absolute value), closed form expressions for $a$ and $b$ become significantly harder.

If one requires that $a = 0$, show that the $b$ leading to least error is $b = \bar{y} = \frac{1}{n} \sum_i y_i$. 

131
C.1.6 Disjoint Events

As always, consider an outcome space $\Omega$ with events $\omega$. Let $A_i \subset \Omega$ be an event, and assume that $A_i \cap A_j$ is empty (for $i \neq j$).

The following observations are often very useful:

1. $\text{Prob}(\bigcup_i A_i) = \sum_i \text{Prob}(A_i)$;
2. If $\bigcup_{i=1}^{\infty} A_i = \Omega$, then
   $$\text{Prob}(A_j) = 1 - \sum_{i=1}^{\infty} \text{Prob}(A_i).$$  \hspace{1cm} (C.42)

C.2 Standard Processes

C.2.1 Bernoulli Process

Recall \( \binom{N}{r} = \frac{N!}{r!(N-r)!} \) is the number of ways to choose \( r \) objects from \( N \) objects when order does not matter. Consider \( n \) independent repetitions of an event with only two possible outcomes. We typically call one outcome success and the other failure, the event a Bernoulli Trial, and a collection of independent Bernoulli Trials a Bernoulli Process.

In each Bernoulli Trial, let there be probability \( p \) of success and \( q = 1 - p \) of failure. Often, we represent a success with 1 and a failure with 0.

Exercise C.2.1. *For a Bernoulli Trial, show $\bar{x} = p$, $\sigma^2_x = pq$, and $\sigma_x = \sqrt{pq}$."

Let $Y_N$ be the number of successes in $N$ trials. Clearly, the possible values are $Y_N = \{0, 1, \ldots, N\}$. We analyze $p_N(k)$. Rigorously, the sample space $\Omega$ is all possible sequences of $N$ trials, and the random variable $y_N : \Omega \to \mathbb{R}$ is given by $y_N(\omega)$ equals the number of successes in $\omega$.

If $k \in Y_N$, we need $k$ successes and $N - k$ failures. We don’t care what order we have them (ie, if $k = 4$ and $N = 6$ then SSFSSF and FSSSSF both contribute). Each such string of $k$ successes and $N - k$ failures has probability of $p^k \cdot (1-p)^{N-k}$. There are $\binom{N}{k}$ such strings.

Thus, $p_N(k) = \binom{N}{k} p^k \cdot (1-p)^{N-k}$ if $k \in \{0, 1, \ldots, N\}$ and 0 otherwise.

By clever algebraic manipulations, one can directly evaluate the mean $\overline{Y_N}$ and the variance $\sigma^2_{y_N}$; however, Lemmas C.1.11 and C.1.17 allow one to calculate
both quantities immediately, once one knows the mean and variance for one occurrence.

**Lemma C.2.2.** For a Bernoulli Process with \( N \) trials, each having probability \( p \) of success, the expected number of successes is \( \overline{y_N} = Np \), and the variance is \( \sigma^2_{y_N} = Npq \).

**Exercise C.2.3.** Prove Lemma C.2.2.

Consider the following problem: Let \( Z = \{0, 1, 2, \ldots \} \) be the number of trials before the first success. What is \( \overline{z} \) and \( \sigma^2_z \)?

First, we determine \( p(k) \), the probability that the first success occurs after \( k \) trials. Clearly this probability is non-zero only for \( k \) a positive integer, in which case the string of results must be \( k-1 \) failures followed by 1 success. Therefore,

\[
p(k) = p \cdot (1 - p)^{k-1} \quad \text{if} \quad k \in \{1, 2, \ldots \}, \quad \text{and} \quad 0 \quad \text{otherwise}.
\]

(C.43)

To determine the mean \( \overline{z} \) we must evaluate

\[
\overline{z} = \sum_{k=1}^{\infty} k \cdot p \cdot (1 - p)^{k-1}
\]

\[
= p \sum_{k=1}^{\infty} kq^{k-1}, \quad 0 < q = 1 - p < 1.
\]

(C.44)

Consider the geometric series

\[
f(q) = \sum_{k=0}^{\infty} q^k = \frac{1}{1-q}.
\]

(C.45)

A careful analysis shows we can differentiate term by term if \( 0 \leq q < 1 \). Then

\[
f'(q) = \sum_{k=0}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2}.
\]

(C.46)

Recalling \( q = 1 - p \) and substituting yields
\[ z = p \sum_{k=1}^{\infty} k q^{k-1} \]
\[ = \frac{p}{\left( 1 - (1 - p) \right)^2} = \frac{1}{p}. \]  

(C.47)

Differentiating under the summation sign is a powerful tool in Probability Theory.

**Exercise C.2.4.** Calculate \( \sigma_z^2 \). Hint: differentiate \( f(q) \) twice.

**C.2.2 Poisson Distribution**

Divide the unit interval into \( N \) equal pieces. Consider \( N \) independent Bernoulli Trials, one for each sub-interval. If the probability of a success is \( \frac{\lambda}{N} \), then by Lemma C.2.2 the expected number of successes is \( N \cdot \frac{\lambda}{N} = \lambda \).

We consider the limit as \( N \to \infty \). We still expect \( \lambda \) successes in each unit interval, but what is the probability of \( 3\lambda \) successes? How long do we expect to wait between successes?

We call this a **Poisson process with parameter** \( \lambda \). For example, look at the midpoints of the \( N \) intervals. At each midpoint we have a Bernoulli Trial with probability of success \( \frac{\lambda}{N} \) and failure \( 1 - \frac{\lambda}{N} \).

We determine the \( N \to \infty \) limits. For fixed \( N \), the probability of \( k \) successes in a unit interval is

\[
p_N(k) = \binom{N}{k} \left( \frac{\lambda}{N} \right)^k \left( 1 - \frac{\lambda}{N} \right)^{N-k} = \frac{N!}{k!(N-k)!} \frac{\lambda^k}{N^k} \left( 1 - \frac{\lambda}{N} \right)^{N-k} = \frac{N \cdot (N-1) \cdots (N-k+1) \lambda^k}{N^k N \cdots N} k! \left( 1 - \frac{\lambda}{N} \right)^N \left( 1 - \frac{\lambda}{N} \right)^{-k} = 1 \cdot \left( 1 - \frac{1}{N} \right) \cdots \left( 1 - \frac{k-1}{N} \right) \lambda^k \frac{k!}{N^k} \left( 1 - \frac{\lambda}{N} \right)^N \left( 1 - \frac{\lambda}{N} \right)^{-k} \]  

(C.48)

For fixed, finite \( k \), as \( N \to \infty \), the first \( k \) factors in \( p_N(k) \) tend to \( 1 \), \( \left( 1 - \frac{\lambda}{N} \right)^N \to e^{-\lambda} \), and \( \left( 1 - \frac{\lambda}{N} \right)^{-k} \to 1 \).
Thus, we are led to the **Poisson Distribution**: Given a parameter \( \lambda \) (interpreted as the expected number of occurrences per unit interval), the probability of \( k \) occurrences in a unit interval is 
\[
p(k) = \frac{\lambda^k}{k!} e^{-\lambda}
\]
for \( k \in \{0, 1, 2, \ldots \} \).

**Exercise C.2.5.** Check that \( p(k) \) given above is a probability distribution. Namely, show \( \sum_{k \geq 0} p(k) = 1 \).

**Exercise C.2.6.** Show, for the Poisson Distribution, that the mean \( \bar{x} = \lambda \) and the variance \( \sigma^2_x = \lambda \). Hint: let
\[
f(\lambda) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^\lambda.
\]
(D.49)

Differentiate once to determine the mean, twice to determine the variance.

### C.2.3 Continuous Poisson Distribution

We calculate a very important quantity related to the Poisson Distribution (with parameter \( \lambda \)), namely, how long does one expect to wait between successes?

We’ve discussed that we expect \( \lambda \) successes per unit interval, and we’ve calculated the probability of \( k \) successes per unit interval.

Start counting at 0, and assume the first success is at \( x \). What is \( p_S(x) \)? As before, we divide each unit interval into \( N \) equal pieces, and consider a Bernoulli Trial at the midpoint of each sub-interval, with probability \( \frac{\lambda}{N} \) of success.

We have approximately \( \frac{x - 0}{1/N} = Nx \) midpoints from 0 to \( x \) (with \( N \) midpoints per unit interval). Let \( \lceil y \rceil \) be the smallest integer greater than or equal to \( y \). Then we have \( \lceil Nx \rceil \) midpoints, where the results of the Bernoulli Trials of the first \( \lceil Nx \rceil - 1 \) midpoints are all failures and the last is a success.

Thus, the probability of the first success occurring in an interval of length \( \frac{1}{N} \) containing \( x \) (with \( N \) divisions per unit interval) is
\[
p_{N,S}(x) = \left(1 - \frac{\lambda}{N}\right)^{\lceil Nx \rceil - 1} \cdot \left(\frac{\lambda}{N}\right)^1.
\]
(C.50)

For \( N \) large, the above converges to \( e^{-\lambda x} \frac{\lambda}{N} \).

We say \( p(x) \) is a **continuous probability distribution on** \( \mathbb{R} \) if
1. \( p(x) \geq 0 \) for all \( x \in \mathbb{R} \).

2. \( \int_{\mathbb{R}} p(x)dx = 1 \).

3. Probability \((a \leq x \leq b) = \int_a^b p(x)dx \).

We call \( p(x) \) the **probability density function**; \( p(x)dx \) is basically the probability of the interval \([x, x+dx]\).

Thus, as \( N \to \infty \), we see the probability density function \( p_S(x) = \lambda e^{-\lambda x} \) \((\frac{1}{N} \text{ is like } dx \text{ for } N \text{ large})\). In the special case of \( \lambda = 1 \), we get the standard exponential decay, \( e^{-x} \).

For instance, let \( \pi(M) \) be the number of primes that are at most \( M \). The Prime Number Theorem states \( \pi(M) = \frac{M}{\log M} + \text{lower order terms} \).

Thus, the average spacing between primes around \( M \) is about \( \log M \). We can model the distribution of primes as a Poisson Process, with parameter \( \lambda = \frac{1}{\log M} \). While possible locations of primes (obviously) is discrete (it must be an integer, and in fact the location of primes aren’t independent), a Poisson model often gives very good heuristics.

We can often renormalize so that \( \lambda = 1 \). This is denoted **unit mean spacing**. For example, one can show the \( M^{th} \) prime \( p_M \) is about \( M \log M \), and spacings between primes around \( p_M \) is about \( \log M \). Then the normalized primes, \( q_M \approx \frac{p_M}{\log M} \) will have unit mean spacing and \( \lambda = 1 \).

### C.3 Coin Flipping: Brute Force

Consider the classic example of a drunk, staggering in unit steps. He does \( 2N \) independent flips a fair coin. For every head he takes a step to the right; for each tail, a step to the left. How likely is he to be at various points?

Clearly, we expect the mean distance from the origin to be 0, as heads and tails are equally likely. For notational convenience, we harmlessly modify the problem: for each head he takes two steps to the right, for each tail he stays still. Now, we expect him to be \( N \) units from the origin. What is the variance?

Let \( \omega \) be a sequence of \( 2N \) coin flips, and let \( n(\omega) \) be twice the number of heads in \( \omega \). In other words, for a given sequence \( \omega \) of flips, \( n(\omega) \) is how far to the right the drunk has moved.

There are \( 2^{2N} \) choices for \( \omega \), each equally likely. We want to calculate the mean and the standard deviation of \( n(\omega) \). Thus, we need to determine \( \text{Prob}(n(\omega) = 2n) \) (note the drunk can only be an even distance from the origin).
To be $2n$ units from the origin, we need $n$ heads. As there are $2N$ flips, there must be $2N - n$ tails. There are $\binom{2N}{n}$ ways to choose $\omega$ such that there are $n$ heads in $\omega$ and $2N - n$ tails.

We find

$$\text{Prob}(n(\omega) = 2n) = \frac{\binom{2N}{n}}{2^{2N}}. \quad (C.51)$$

To find the mean, we must evaluate

$$E[2n] = \sum_{n=0}^{2N} 2n \cdot \frac{\binom{2N}{n}}{2^{2N}}, \quad (C.52)$$

while the variance is given by

$$\text{Variance} = \sum_{n=0}^{2N} (2n)^2 \cdot \frac{\binom{2N}{n}}{2^{2N}} - E[2n]^2. \quad (C.53)$$

### C.3.1 Calculation of the Mean

As heads and tails are equally likely, we expect $N$ flips will be heads, moving the drunk $2N$ units. Thus, we predict $E[2n] = 2N$. We give an additional proof, which will be needed for the standard deviation calculation.

Again, we must evaluate

$$E[2n] = \sum_{n=0}^{2N} 2n \cdot \frac{\binom{2N}{n}}{2^{2N}}. \quad (C.54)$$

We will prove this by differentiating combinatorial identities.

$$(p + q)^{2N} = \sum_{n=0}^{2N} \binom{2N}{n} p^n q^{2N-n} \quad (C.55)$$

Taking the derivative with respect to $p$ gives

$$2N(p + q)^{2N-1} = \sum_{n=0}^{2N} \binom{2N}{n} np^{n-1} q^{2N-n}. \quad (C.56)$$

Setting $p = q = 1$ yields
\[ \sum_{n=0}^{2N} 2n \cdot \binom{2N}{n} = 2N \cdot 2^{2N}. \]  
(C.57)

Substituting into Equation C.54 completes the proof.

### C.3.2 Calculation of the Variance

To calculate

\[ \sum_{n=0}^{2N} (2n)^2 \cdot \frac{\binom{2N}{n}}{2^{2N}} - E[2n]^2, \]  
(C.58)

we differentiate Equation C.55 twice with respect to \( p \), obtaining

\[ 2N(2N-1)(p+q)^{2N-2} = \sum_{n=0}^{2N} \binom{2N}{n} n(n-1)p^{n-2}q^{2N-n}. \]  
(C.59)

Setting \( p = q = 1 \) yields

\[ \sum_{n=0}^{2N} n(n-1)\binom{2N}{n} = 2N(2N-1)2^{2N-2}. \]  
(C.60)

Re-writing, we have

\[ \sum_{n=0}^{2N} n^2 \binom{2N}{n} = \sum_{n=0}^{2N} n \binom{2N}{n} + 2N(2N-1)2^{2N-2} \]
\[ = N \cdot 2^{2N} + N(2N-1)2^{2N-1} \]
\[ = N^2 2^{2N} + N^2 2^{2N-1}. \]  
(C.61)

Thus,

\[ \text{Variance} = \sum_{n=0}^{2N} (2n)^2 \cdot \frac{\binom{2N}{n}}{2^{2N}} - E[2n]^2 \]
\[ = \frac{4N^2 2^{2N} + 2N 2^{2N}}{2^{2N}} - (2N)^2 \]
\[ = 4N^2 + 2N - 4N^2 = 2N. \]  
(C.62)
Thus, the variance is $2N$ and the standard deviation is $\sqrt{2N}$, the square-root of the number of tosses.

### C.4 Central Limit Theorem

#### C.4.1 Random Walks

Consider the classical problem of a drunk staggering home from a lamp post late at night. We flip a fair coin $N$ times. With probability $\frac{1}{2}$ we get heads (tails). For each head (tail) the drunk staggers one unit to the right (left). How far do we expect the drunk to be?

It is very unlikely the drunk will be very far to the left or right.

**Exercise C.4.1.** Let $x$ be $+1$ if we flip a head, $-1$ for a tail. For a fair coin, prove $E[x] = 0$, $\sigma^2_x = 1$, $\sigma_x = 1$.

**Exercise C.4.2.** Let $p_N(y)$ be the probability that after $N$ flips of a fair coin, the drunk is $y$ units to the right of the origin (lamp post).

1. Prove $p_N(y) = p_N(-y)$.

2. Consider $N = 2M$. Prove $p_{2M}(2k) = \left(\frac{2M}{M+k}\right)\frac{1}{2^M}$, where $\binom{n}{r} = \frac{n!}{r!(n-r)!}$

3. Use Stirling’s formula ($n! \approx n^n e^{-n} \sqrt{2\pi n} = \sqrt{2\pi n^{n+\frac{1}{2}} e^{-n}}$) to approximate $p_N(y)$.

Label the coin tosses $X_1$ through $X_N$. Let $X$ denote a generic toss of the coin, and $Y_N$ be the distance of the drunkard after $N$ tosses. By Lemma C.1.11, $E[Y_N] = E[x_1 + \cdots + x_N] = E[x_1] + \cdots + E[x_N]$. As each $E[x_i] = E[x] = 0$, $E[Y_N] = 0$.

Thus, we expect the drunkard to be at the lamp post. How spread out is his expected position? By Lemma C.1.17,

$$\sigma_{y_N} = \sigma_{x_N} = \sqrt{N} \sigma_x = \sqrt{N}. \quad (C.63)$$

This means that a typical distance from the origin is $\sqrt{N}$. This is called a diffusion process and is very common in the real world.
C.4.2 Central Limit Theorem

\( X_1, X_2, X_3, \ldots \) are an infinite sequence of random variables such that the \( X_j \) are independent identically distributed random variables (abbreviated i.i.d.r.v.) with \( E[X_j] = X_j = 0 \) (can always renormalize by shifting) and variance \( E[X_j^2] = 1 \).

Let \( S_N = \sum_{j=1}^{N} X_j \).

**Theorem C.4.3 (Central Limit Theorem).** Fix \( -\infty < a \leq b < \infty \). Then as \( N \to \infty \),

\[
\text{Prob}\left( \frac{S_N}{\sqrt{N}} \in [a, b] \right) \to \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{t^2}{2}} dt. \tag{C.64}
\]

The probability function is called the Gaussian or the Normal distribution. This is the universal curve of probability. Note how robust the Central Limit Theorem is: it doesn’t depend on fine properties of the \( X_j \).

C.5 Buffon’s Needle and \( \pi \)

Consider a collection of infinitely long parallel lines in the plane, where the spacing between any two adjacent lines is \( d \). Let the lines be located at \( x = 0, \pm d, \pm 2d, \ldots \). Consider a rod of length \( l \), where for convenience we assume \( l < d \).

If we were to randomly throw the rod on the plane, what is the probability it hits a line? This question was first asked by Buffon in 1733.

Because of the vertical symmetry, we may assume the center of the rod lies on the line \( x = 0 \), as shifting the rod (without rotating it) up or down will not alter the number of intersections. By the horizontal symmetry, we may assume \(-\frac{d}{2} \leq x < \frac{d}{2}\). We posit that all values of \( x \) are equally likely. As \( x \) is continuous distributed, we may add in \( x = \frac{d}{2} \) without changing the probability. The probability density function of \( x \) is \( \frac{dx}{\frac{d}{2}} \).

Let \( \theta \) be the angle the rod makes with the \( x \)-axis. As each angle is equally likely, the probability density function of \( \theta \) is \( \frac{d\theta}{2\pi} \).

We assume that \( x \) and \( \theta \) are chosen independently. Thus, the probability density for \((x, \theta)\) is \( \frac{dx \cdot d\theta}{2\pi} \).

The projection of the rod (making an angle of \( \theta \) with the \( x \)-axis) along the \( x \)-axis is \( l \cdot |\cos \theta| \). If \( |x| \leq l \cdot |\cos \theta| \), then the rod hits exactly one vertical line exactly once; if \( x > l \cdot |\cos \theta| \), the rod does not hit a vertical line. Note that if \( l > d \), a rod could hit multiple lines, making the arguments more involved.
Thus, the probability a rod hits a line is

$$p = \int_{\theta=0}^{2\pi} \int_{x=-l|\cos \theta|}^{l|\cos \theta|} \frac{dx d\theta}{d \cdot 2\pi} = \int_{\theta=0}^{2\pi} \frac{l \cdot |\cos \theta|}{d} \frac{d\theta}{2\pi} = \frac{2l}{\pi d}. \quad (C.65)$$

**Exercise C.5.1.** Show

$$\frac{1}{2\pi} \int_0^{2\pi} |\cos \theta| d\theta = \frac{2}{\pi}. \quad (C.66)$$

Let $A$ be the random variable which is the number of intersections of a rod of length $l$ thrown against parallel vertical lines separated by $d > l$ units. Then

$$A = \begin{cases} 1 & \text{with probability } \frac{2l}{\pi d} \\ 0 & \text{with probability } 1 - \frac{2l}{\pi d}. \end{cases} \quad (C.67)$$

If we were to throw $N$ rods independently, since the expected value of a sum is the sum of the expected values (Lemma C.1.11), we expect to observe

$$N \cdot \frac{2l}{\pi d}. \quad (C.68)$$

intersections.

Turning this around, let us throw $N$ rods, and let $I$ be the number of observed intersections of the rods with the vertical lines. Then

$$I \approx N \cdot \frac{2l}{\pi d} \rightarrow \pi \approx \frac{N \cdot 2l}{I \cdot d}. \quad (C.69)$$

The above is an experimental formula for $\pi$!
Appendix D

Poissonian Behavior and \( \{ n^k \alpha \} \)

D.1 Equidistribution

We say a sequence of number \( x_n \in [0, 1) \) is equidistributed if

\[
\lim_{N \to \infty} \frac{\# \{ n : 1 \leq n \leq N \text{ and } x_n \in [a, b] \}}{N} = b - a \tag{D.1}
\]

for any subinterval \([a, b]\) of \([0, 1]\).

Recall Weyl’s Result, Theorem B.2.10: If \( \alpha \not\in \mathbb{Q} \), then the fractional parts \( \{n\alpha\} \) are equidistributed. Equivalently, \( n\alpha \mod 1 \) is equidistributed.

Similarly, one can show that for any integer \( k \), \( \{n^k \alpha\} \) is equidistributed. See Robert Lipshitz’s paper for more details.

D.2 Point Masses and Induced Probability Measures

Recall from physics the concept of a unit point mass located at \( x = a \). Such a point mass has no length (or, in higher dimensions, width or height), but finite mass. As mass is the integral of the density over space, a finite mass in zero volume (or zero length on the line) implies an infinite density.

We can make this more precise by the notion of an Approximation to the Identity. See also Theorem B.2.3.

**Definition D.2.1 (Approximation to the Identity).** A sequence of functions \( g_n(x) \) is an approximation to the identity (at the origin) if

1. \( g_n(x) \geq 0 \).
2. \( \int g_n(x) dx = 1. \)

3. Given \( \epsilon, \delta > 0 \) there exists \( N > 0 \) such that for all \( n > N, \int_{|x|>\delta} g_n(x) dx < \epsilon. \)

We represent the limit of any such family of \( g_n(x) \)'s by \( \delta(x) \).

If \( f(x) \) is a nice function (say near the origin its Taylor Series converges) then

\[
\int f(x) \delta(x) dx = \lim_{n \to \infty} \int f(x) g_n(x) = f(0). \tag{D.2}
\]

**Exercise D.2.2. Prove Equation D.2.**

Thus, in the limit the functions \( g_n \) are acting like point masses. We can consider the probability densities \( g_n(x) dx \) and \( \delta(x) dx \). For \( g_n(x) dx, \) as \( n \to \infty, \) almost all the probability is concentrated in a narrower and narrower band about the origin; \( \delta(x) dx \) is the limit with all the mass at one point. It is a discrete (as opposed to continuous) probability measure.

Note that \( \delta(x - a) \) acts like a point mass; however, instead of having its mass concentrated at the origin, it is now concentrated at \( a \).

**Exercise D.2.3. Let**

\[
g_n(x) = \begin{cases} 
  n & \text{if } |x| \leq \frac{1}{2n} \\
  0 & \text{otherwise}
\end{cases} \tag{D.3}
\]

Prove \( g_n(x) \) is an approximation to the identity at the origin.

**Exercise D.2.4. Let**

\[
g_n(x) = c \left( \frac{1}{n^2 + x^2} \right). \tag{D.4}
\]

Find \( c \) such that the above is an approximation to the identity at the origin.

Given \( N \) point masses located at \( x_1, x_2, \ldots, x_N \), we can form a probability measure

\[
\mu_N(x) dx = \frac{1}{N} \sum_{n=1}^{N} \delta(x - x_n) dx. \tag{D.5}
\]

Note \( \int \mu_N(x) dx = 1 \), and if \( f(x) \) is a nice function,

\[
\int f(x) \mu_N(x) dx = \frac{1}{N} \sum_{n=1}^{N} f(x_n). \tag{D.6}
\]
**Exercise D.2.5.** Prove Equation D.6 for nice $f(x)$.  

Note the right hand side of Equation D.6 looks like a Riemann sum. Or it would look like a Riemann sum if the $x_n$s were equidistributed. In general the $x_n$s will not be equidistributed, but assume for any interval $[a, b]$ that as $N \to \infty$, the fraction of $x_n$s ($1 \leq n \leq N$) in $[a, b]$ goes to $\int_a^b p(x)dx$ for some nice function $p(x)$:

$$\lim_{N \to \infty} \frac{\#\{n : 1 \leq n \leq N \text{ and } x_n \in [a, b]\}}{N} \to \int_a^b p(x)dx. \quad (D.7)$$

In this case, if $f(x)$ is nice (say twice differentiable, with first derivative uniformly bounded), then

$$\int f(x)\mu_N(x)dx = \frac{1}{N} \sum_{n=1}^{N} f(x_n) \approx \sum_{k=-\infty}^{\infty} f\left(\frac{k}{N}\right) \frac{\#\{n : 1 \leq n \leq N \text{ and } x_n \in \left[\frac{k}{N}, \frac{k+1}{N}\right]\}}{N} \to \int f(x)p(x)dx. \quad (D.8)$$

**Definition D.2.6 (Convergence to $p(x)$).** If the sequence of points $x_n$ satisfies Equation D.7 for some nice function $p(x)$, we say the probability measures $\mu_N(x)dx$ converge to $p(x)dx$.

**D.3 Neighbor Spacings**

We now consider finer questions. Let $\alpha_n$ be a collection of points in $[0, 1)$. We order them by size:

$$0 \leq \alpha_{\sigma(1)} \leq \alpha_{\sigma(2)} \leq \cdots \leq \alpha_{\sigma(N)}, \quad (D.9)$$

where $\sigma$ is a permutation of $123 \cdots N$. Note the ordering depends crucially on $N$. Let $\beta_j = \alpha_{\sigma(j)}$.

We consider how the differences $\beta_{j+1} - \beta_j$ are distributed. We will use a slightly different definition of distance, however.
Recall $[0, 1)$ is equivalent to the unit circle under the map $x 	o e^{2\pi ix}$. Thus, the numbers .999 and .001 are actually very close; however, if we used the standard definition of distance, then $|.999 - .001| = .998$, which is quite large. Wrapping $[0, 1)$ on itself (identifying 0 and 1), we see that .999 and .001 are separated by .002.

**Definition D.3.1 (mod 1 distance).** Let $x, y \in [0, 1)$. We define the mod 1 distance from $x$ to $y$, $||x - y||$, by

$$||x - y|| = \min \left\{ |x - y|, 1 - |x - y| \right\}.$$ (D.10)

**Exercise D.3.2.** Show that the mod 1 distance between any two numbers in $[0, 1)$ is at most $\frac{1}{2}$.

In looking at spacings between the $\beta_j$s, we have $N - 1$ pairs of neighbors:

$$(\beta_2, \beta_1), (\beta_3, \beta_2), \ldots, (\beta_N, \beta_{N-1}).$$ (D.11)

These pairs give rise to spacings $\beta_{j+1} - \beta_j \in [0, 1)$.

We can also consider the pair $(\beta_1, \beta_N)$. This gives rise to the spacing $\beta_1 - \beta_N \in [-1, 0)$; however, as we are studying this sequence mod 1, this is equivalent to $\beta_1 - \beta_N + 1 \in [0, 1)$.

**Henceforth,** whenever we perform any arithmetic operation, we always mean mod 1; thus, our answers always live in $[0, 1)$.

**Definition D.3.3 (Neighbor Spacings).** Given a sequence of numbers $\alpha_n$ in $[0, 1)$, fix an $N$ and arrange the numbers $\alpha_n$ ($n \leq N$) in increasing order. Label the new sequence $\beta_j$; note the ordering will depend on $N$. Let $\beta_{-j} = \beta_{N-j}$ and $\beta_{N+j} = \beta_j$.

1. The nearest neighbor spacings are the numbers $\beta_{j+1} - \beta_j$, $j = 1$ to $N$.

2. The $k^{th}$-neighbor spacings are the numbers $\beta_{j+k} - \beta_j$, $j = 1$ to $N$.

Remember to take the differences $\beta_{j+k} - \beta_j$ mod 1.

**Exercise D.3.4.** Let $\alpha = \sqrt{2}$, and let $\alpha_n = \{n\alpha\}$ or $\{n^2\alpha\}$. Calculate the nearest neighbor and the next-nearest neighbor spacings in each case for $N = 10$.

**Definition D.3.5 (wrapped unit interval).** We call $[0, 1)$, when all arithmetic operations are done mod 1, the wrapped unit interval.
D.4 Poissonian Behavior

Let $\alpha \not\in \mathbb{Q}$. Fix a positive integer $k$, and let $\alpha_n = \{n^k\alpha\}$. As $N \to \infty$, look at the ordered $\alpha_n$s, denoted by $\beta_n$. How are the nearest neighbor spacings of $\beta_n$ distributed? How does this depend on $k$? On $\alpha$? On $N$?

Before discussing this problem, we consider a simpler case. Fix $N$, and consider $N$ independent random variables $x_n$. Each random variable is chosen from the uniform distribution on $[0, 1)$; thus, the probability that $x_n \in [a, b)$ is $b - a$.

Let $y_n$ be the $x_n$s arranged in increasing order. How do the neighbor spacings behave?

First, we need to decide what is the correct scale to use for our investigations. As we have $N$ objects on the wrapped unit interval, we have $N$ nearest neighbor spacings. Thus, we expect the average spacing to be $\frac{1}{N}$.

**Definition D.4.1 (Unfolding).** Let $z_n = Ny_n$. The numbers $z_n = Ny_n$ have unit mean spacing. Thus, while we expect the average spacing between adjacent $y_n$s to be $\frac{1}{N}$ units, we expect the average spacing between adjacent $z_n$s to be 1 unit.

So, the probability of observing a spacing as large as $\frac{1}{2}$ between adjacent $y_n$s becomes negligible as $N \to \infty$. What we should ask is what is the probability of observing a nearest neighbor spacing of adjacent $y_n$s that is half the average spacing. In terms of the $z_n$s, this will correspond to a spacing between adjacent $z_n$s of $\frac{1}{2}$ a unit.

D.4.1 Nearest Neighbor Spacings

By symmetry, on the wrapped unit interval the expected nearest neighbor spacing is independent of $j$. Explicitly, we expect $\beta_j + 1 - \beta_j$ to have the same distribution as $\beta_{j+1} - \beta_j$.

What is the probability that, when we order the $x_n$s in increasing order, the next $x_n$ after $x_1$ is located between $\frac{t}{N}$ and $\frac{t + \Delta t}{N}$? Let the $x_n$s in increasing order be labeled $y_1 \leq y_2 \leq \cdots \leq y_N$, $y_n = x_{\sigma(n)}$.

As we are choosing the $x_n$s independently, there are $\binom{N-1}{1}$ choices of subscript $n$ such that $x_n$ is nearest to $x_1$. This can also be seen by symmetry, as each $x_n$ is equally likely to be the first to the right of $x_1$ (where, of course, .001 is just a little to the right of .999), and we have $N - 1$ choices left for $x_n$.

The probability that $x_n \in \left[\frac{t}{N}, \frac{t + \Delta t}{N}\right]$ is $\frac{\Delta t}{N}$.
For the remaining \( N - 2 \) of the \( x_n \)'s, each must be further than \( \frac{t + \Delta t}{N} \) from \( x_n \). Thus, they must all lie in an interval (or possibly two intervals if we wrap around) of length \( 1 - \frac{t + \Delta t}{N} \). The probability that they all lie in this region is \( \left( 1 - \frac{t + \Delta t}{N} \right)^{N-2} \).

Thus, if \( x_1 = y_l \), we want to calculate the probability that \( ||y_{l+1} - y_l|| \in \left[ \frac{t}{N}, \frac{t + \Delta t}{N} \right] \). This is

\[
\begin{align*}
\text{Prob} \left( ||y_{l+1} - y_l|| \in \left[ \frac{t}{N}, \frac{t + \Delta t}{N} \right] \right) &= \left( \frac{N - 1}{1} \right) \cdot \frac{\Delta t}{N} \cdot \left( 1 - \frac{t + \Delta t}{N} \right)^{N-2} \\
&= \left( 1 - \frac{1}{N} \right) \cdot \left( 1 - \frac{t + \Delta t}{N} \right)^{N-2} \Delta t .
\end{align*}
\]

(D.12)

For \( N \) enormous and \( \Delta t \) small,

\[
\begin{align*}
\left( 1 - \frac{1}{N} \right) &\approx 1 \\
\left( 1 - \frac{t + \Delta t}{N} \right)^{N-2} &\approx e^{-(t + \Delta t)} \approx e^{-t} .
\end{align*}
\]

(D.13)

Thus

\[
\text{Prob} \left( ||y_{l+1} - y_l|| \in \left[ \frac{t}{N}, \frac{t + \Delta t}{N} \right] \right) \to e^{-t} \Delta t .
\]

(D.14)

**Remark D.4.2.** The above argument is infinitesimally wrong. Once we’ve located \( y_{l+1} \), the remaining \( x_n \)'s do not need to be more than \( \frac{t + \Delta t}{N} \) units to the right of \( x_1 = y_l \); they only need to be further to the right than \( y_{l+1} \). As the incremental gain in probabilities for the locations of the remaining \( x_n \)'s is of order \( \Delta t \), these contributions will not influence the large \( N \), small \( \Delta t \) limits. Thus, we ignore these effects.

To rigorously derive the limiting behavior of the nearest neighbor spacings using the above arguments, one would integrate over \( x_m \) ranging from \( \frac{t}{N} \) to \( \frac{t + \Delta t}{N} \), and the remaining events \( x_n \) would be in the a segment of length \( 1 - x_m \). As

\[
\left| \left( 1 - x_m \right) - \left( 1 - \frac{t + \Delta t}{N} \right) \right| \leq \frac{\Delta t}{N} ,
\]

(D.15)

this will lead to corrections of higher order in \( \Delta t \), hence negligible.

We can rigorously avoid this by instead considering the following:
1. Calculate the probability that all the other $x_n$s are at least $\frac{t}{N}$ units to the right of $x_1$. This is

$$p_t = \left(1 - \frac{t}{N}\right)^{N-1} \rightarrow e^{-t}. \quad \text{(D.16)}$$

2. Calculate the probability that all the other $x_n$s are at least $\frac{t + \Delta t}{N}$ units to the right of $x_1$. This is

$$p_{t+\Delta t} = \left(1 - \frac{t + \Delta t}{N}\right)^{N-1} \rightarrow e^{-(t + \Delta t)}. \quad \text{(D.17)}$$

3. The probability that no $x_n$s are within $\frac{t}{N}$ units to the right of $x_1$ but at least one $x_n$ is between $\frac{t}{N}$ and $\frac{t + \Delta t}{N}$ units to the right is $p_{t+\Delta t} - p_t$:

$$p_t - p_{t+\Delta t} \rightarrow e^{-t} - e^{-(t + \Delta t)}$$

$$= e^{-t} \left(1 - e^{-\Delta t}\right)$$

$$= e^{-t} \left(1 - 1 + \Delta t + O\left((\Delta t)^2\right)\right)$$

$$\rightarrow e^{-t} \Delta t. \quad \text{(D.18)}$$

**Definition D.4.3 (Unfolding Spacings).** If $y_{l+1} - y_l \in \left[\frac{t}{N}, \frac{t + \Delta t}{N}\right]$, then $N(y_{l+1} - y_l) \in [t, t + \Delta t]$. The new spacings $z_{l+1} - z_l$ have unit mean spacing. Thus, while we expect the average spacing between adjacent $y_n$s to be $\frac{1}{N}$ units, we expect the average spacing between adjacent $z_n$s to be 1 unit.

**D.4.2 $k$th Neighbor Spacings**

Similarly, one can easily analyze the distribution of the $k$th neighbor spacings when each $x_n$ is chosen independently from the uniform distribution on $[0, 1)$.

Again, consider $x_1 = y_l$. Now we want to calculate the probability that $y_{l+k}$ is between $\frac{t}{N}$ and $\frac{t + \Delta t}{N}$ units to the right of $y_l$.

Therefore, we need exactly $k - 1$ of the $x_n$s to lie between 0 and $\frac{t}{N}$ units to the right of $x_1$, exactly one $x_n$ (which will be $y_{l+k}$) to lie between $\frac{t}{N}$ and $\frac{t + \Delta t}{N}$ units to the right of $x_1$, and the remaining $x_n$s to lie at least $\frac{t + \Delta t}{N}$ units to the right of $y_{l+k}$.
Remark D.4.4. We face the same problem discussed in Remark D.4.2; a similar argument will show that ignoring these affects will not alter the limiting behavior. Therefore, we will make these simplifications.

There are \( \binom{N-1}{k-1} \) ways to choose the \( x_n \)s that are at most \( \frac{t}{N} \) units to the right of \( x_1 \); there is then \( \binom{N-1-(k-1)}{1} \) ways to choose the \( x_n \) between \( \frac{t}{N} \) and \( \frac{t+\Delta t}{N} \) units to the right of \( x_1 \).

Thus,

\[
\text{Prob}\left( \left| y_{l+k} - y_l \right| \in \left[ \frac{t}{N}, \frac{t+\Delta t}{N} \right] \right) = \frac{(N-1)\cdot(t/N)^{k-1} \cdot (N-1-(k-1)) \cdot \Delta t \cdot (1 - t/N)^{N-(k+1)}}{(N-1)\cdot(N-1-(k-2)) \cdot (N-1-(k-1))} \cdot \frac{t^{k-1}}{(k-1)!} \cdot \frac{(1 - t + \Delta t/N)^{N-(k+1)}}{N} \Delta t
\]

\[
= \frac{t^{k-1}}{(k-1)!} e^{-t} \Delta t.
\]

(D.19)

Again, one way to avoid the complications is to integrate over \( x_m \) ranging from \( \frac{t}{N} \) to \( \frac{t+\Delta t}{N} \).

Or, similar to before, we can proceed more rigorously as follows:

1. Calculate the probability that exactly \( k-1 \) of the other \( x_n \)s are at most \( \frac{t}{N} \) units to the right of \( x_1 \), and the remaining \( (N-1)-(k-1) \) of the \( x_n \)s are at least \( \frac{t}{N} \) units to the right of \( x_1 \). As there are \( \binom{N-1}{k-1} \) ways to choose \( k-1 \) of the \( x_n \)s to be at most \( \frac{t}{N} \) units to the right of \( x_1 \), this probability is

\[
p_t = \frac{(N-1)\cdot(t/N)^{k-1} \cdot (1 - t/N)^{(N-1)-(k-1)}}{(N-1)\cdot(N-1-(k-2)) \cdot (N-1-(k-1))} \cdot \frac{t^{k-1}}{(k-1)!} \cdot \frac{(1 - t + \Delta t/N)^{N-(k+1)}}{N} \Delta t
\]

\[
= \frac{t^{k-1}}{(k-1)!} e^{-t} \Delta t.
\]

(D.20)
2. Calculate the probability that exactly \( k - 1 \) of the other \( x_n \)s are at most \( \frac{t}{N} \) units to the right of \( x_1 \), and the remaining \( (N - 1) - (k - 1) \) of the \( x_n \)s are at least \( \frac{t + \Delta t}{N} \) units to the right of \( x_1 \). Similar to the above, this gives

\[
\begin{align*}
 p_t & \quad = \quad \left( \frac{N - 1}{k - 1} \right)^{k-1} \left( 1 - \frac{t + \Delta t}{N} \right)^{(N-1)-(k-1)} \\
 & \quad \rightarrow \quad \frac{N^{k-1}}{(k-1)!} t^{k-1} e^{-(t+\Delta t)} \\
 & \quad \rightarrow \quad \frac{t^{k-1}}{(k-1)!} e^{-(t+\Delta t)}. \quad \text{(D.21)}
\end{align*}
\]

3. The probability that exactly \( k - 1 \) of the \( x_n \)s are within \( \frac{t}{N} \) units to the right of \( x_1 \) and at least one \( x_n \) is between \( \frac{t}{N} \) and \( \frac{t + \Delta t}{N} \) units to the right is \( p_{t+\Delta t} - p_t \):

\[
\begin{align*}
p_t - p_{t+\Delta t} & \quad \rightarrow \quad \frac{t^{k-1}}{(k-1)!} e^{-t} - \frac{t^{k-1}}{(k-1)!} e^{-(t+\Delta t)} \\
& \quad \rightarrow \quad \frac{t^{k-1}}{(k-1)!} e^{-t} \Delta t. \quad \text{(D.22)}
\end{align*}
\]

Note that when \( k = 1 \), we recover the nearest neighbor spacings.

### D.5 Induced Probability Measures

We have proven the following:

**Theorem D.5.1.** Consider \( N \) independent random variables \( x_n \) chosen from the uniform distribution on the wrapped unit interval \([0, 1)\). For fixed \( N \), arrange the \( x_n \)s in increasing order, labeled \( y_1 \leq y_2 \leq \cdots \leq y_N \).

Form the induced probability measure \( \mu_{N,1} \) from the nearest neighbor spacings. Then as \( N \to \infty \) we have

\[
\mu_{N,1}(t)dt = \frac{1}{N} \sum_{n=1}^{N} \delta \left( t - N(y_n - y_{n-1}) \right) dt \rightarrow e^{-t} dt. \quad \text{(D.23)}
\]

Equivalently, using \( z_n = Ny_n \):

\[
\mu_{N,1}(t)dt = \frac{1}{N} \sum_{n=1}^{N} \delta \left( t - (z_n - z_{n-1}) \right) dt \rightarrow e^{-t} dt. \quad \text{(D.24)}
\]
More generally, form the probability measure from the $k^{th}$ nearest neighbor spacings. Then as $N \to \infty$ we have

$$\mu_{N,k}(t)dt = \frac{1}{N} \sum_{n=1}^{N} \delta\left(t - N(y_n - y_{n-k})\right)dt \to \frac{t^{k-1}}{(k-1)!} e^{-t}dt. \quad (D.25)$$

Equivalently, using $z_n = Ny_n$:

$$\mu_{N,k}(t)dt = \frac{1}{N} \sum_{n=1}^{N} \delta\left(t - (z_n - z_{n-k})\right)dt \to \frac{t^{k-1}}{(k-1)!} e^{-t}dt. \quad (D.26)$$

**Definition D.5.2 (Poissonian Behavior).** We say a sequence of points $x_n$ has Poissonian Behavior if in the limit as $N \to \infty$ the induced probability measures $\mu_{N,k}(t)dt$ converge to $\frac{t^{k-1}}{(k-1)!} e^{-t}dt$.

**Exercise D.5.3.** Let $\alpha \in \mathbb{Q}$, and define $\alpha_n = \{n^m \alpha\}$ for some positive integer $m$. Show the sequence of points $\alpha_n$ does not have Poissonian Behavior.

**Exercise D.5.4.** Let $\alpha \not\in \mathbb{Q}$, and define $\alpha_n = \{n\alpha\}$. Show the sequence of points $\alpha_n$ does not have Poissonian Behavior. Hint: for each $N$, show the nearest neighbor spacings take on at most three distinct values (the three values depend on $N$). As only three values are ever assumed for a fixed $N$, $\mu_{N,1}(t)dt$ cannot converge to $e^{-t}dt$.

**D.6 Non-Poissonian Behavior**

**Conjecture D.6.1.** With probability one (with respect to Lebesgue Measure), if $\alpha \not\in \mathbb{Q}$, if $\alpha_n = \{n^2 \alpha\}$ then the sequence of points $\alpha_n$ is Poissonian.

There are constructions which show certain irrationals give rise to non-Poissonian behavior.

**Theorem D.6.2.** Let $\alpha \not\in \mathbb{Q}$ such that $\left|\alpha - \frac{p_n}{q_n}\right| < \frac{a_n}{q_n^2}$ holds infinitely often, with $a_n \to 0$. Then there exist integers $N_j \to \infty$ such that $\mu_{N_j,1}(t)$ does not converge to $e^{-t}dt$.  

151
As $a_n \to 0$, eventually $a_n < \frac{1}{10}$ for all $n$ large. Let $N_n = q_n$, where $\frac{p_n}{q_n}$ is a good rational approximation to $\alpha$:

$$\left| \alpha - \frac{p_n}{q_n} \right| < \frac{a_n}{q_n^3}. \quad (D.27)$$

Remember that all subtractions are performed on the wrapped unit interval. Thus, $||.999 - .001|| = .002$.

We look at $\alpha_k = \{k^2 \alpha\}$, $1 \leq k \leq N_n = q_n$. Let the $\beta_k$s be the $\alpha_k$s arranged in increasing order, and let the $\gamma_k$s be the numbers $\{k^2 \frac{p_n}{q_n}\}$ arranged in increasing order:

$$\beta_1 \leq \beta_2 \leq \cdots \leq \beta_N \quad \gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_N. \quad (D.28)$$

### D.6.1 Preliminaries

**Lemma D.6.3.** If $\beta_l = \alpha_k = \{k^2 \alpha\}$, then $\gamma_l = \{k^2 \frac{p_n}{q_n}\}$. Thus, the same permutation orders both the $\alpha_k$s and the $\gamma_k$s.

**Proof.** Multiplying both sides of Equation D.27 by $k^2 \leq q_n^2$ yields

$$\left| k^2 \alpha - k^2 \frac{p_n}{q_n} \right| < k^2 \frac{a_n}{q_n^3} \leq \frac{a_n}{q_n^3} < \frac{1}{2q_n}. \quad (D.29)$$

Thus, $k^2 \alpha$ and $k^2 \frac{p_n}{q_n}$ differ by at most $\frac{1}{2q_n}$. Therefore

$$\left| \left\{ k^2 \alpha \right\} - \left\{ k^2 \frac{p_n}{q_n} \right\} \right| < \frac{1}{2q_n}. \quad (D.30)$$

As the numbers $\{m^2 \frac{p_n}{q_n}\}$ all have denominators of size at most $\frac{1}{q_n}$, we see that $\{k^2 \frac{p_n}{q_n}\}$ is the closest of the $\{m^2 \frac{p_n}{q_n}\}$ to $\{k^2 \alpha\}$.

This implies that if $\beta_l = \{k^2 \alpha\}$, then $\gamma_l = \{k^2 \frac{p_n}{q_n}\}$, completing the proof.

**Exercise D.6.4.** Prove the ordering is as claimed. Hint: about each $\beta_l = \{k^2 \alpha\}$, the closest number of the form $\{c^2 \frac{p_n}{q_n}\}$ is $\{k^2 \frac{p_n}{q_n}\}$. 

152
D.6.2 Proof of Theorem D.6.2

Exercise D.6.5. Assume $||a - b||, ||c - d|| \leq \frac{1}{10}$. Show

\begin{equation}
||(a - b) - (c - d)|| < ||a - b|| + ||c - d||.
\end{equation}

Proof of Theorem D.6.2: We have shown

\begin{equation}
||\beta_l - \gamma_l|| < a_n.
\end{equation}

Thus, as $N_n = q_n$:

\begin{equation}
\left|\left| N_n(\beta_l - \gamma_l) \right|\right| < a_n,
\end{equation}

and the same result holds with $l$ replaced by $l - 1$.

By Exercise D.6.5,

\begin{equation}
\left|\left| N_n(\beta_l - \gamma_l) - N_n(\beta_{l-1} - \gamma_{l-1}) \right|\right| < 2a_n.
\end{equation}

Rearranging gives

\begin{equation}
\left|\left| N_n(\beta_l - \beta_{l-1}) - N_n(\gamma_l - \gamma_{l-1}) \right|\right| < 2a_n.
\end{equation}

As $a_n \to 0$, this implies the difference between $\left|\left| N_n(\beta_l - \beta_{l-1}) \right|\right|$ and $\left|\left| N_n(\gamma_l - \gamma_{l-1}) \right|\right|$ goes to zero.

The above distance calculations were done mod 1. The actual differences will differ by an integer. Thus,

\begin{equation}
\mu^S_{N_n,1}(t)dt = \frac{1}{N_n} \sum_{l=1}^{N_n} \delta\left(t - N_n(\beta_l - \beta_{l-1})\right)
\end{equation}

and

\begin{equation}
\mu^\mu_{N_n,1}(t)dt = \frac{1}{N_n} \sum_{l=1}^{N_n} \delta\left(t - N_n(\gamma_l - \gamma_{l-1})\right)
\end{equation}

are extremely close to one another; each point mass from the difference between adjacent $\beta_l$s is either within $a_n$ units of a point mass from the difference between adjacent $\gamma_l$s, or is within $a_n$ units of a point mass an integer number of units from a point mass from the difference between adjacent $\gamma_l$s. Further, $a_n \to 0$. 

153
Note, however, that if $\gamma_l = \{ k^2 p_n \}_{q_n}$, then

$$\mathcal{N}_n \gamma_l = q_n \left\{ k^2 p_n \right\} \in \mathbb{N}.$$

(D.38)

Thus, the induced probability measure $\mu_{\mathcal{N}_n} p_n q_n(t) dt$ formed from the $\gamma_l$s is supported on the integers! Thus, it is impossible for $\mu_{\mathcal{N}_n} p_n q_n(t) dt$ to converge to $e^{-t} dt$.

As $\mu_{\mathcal{N}_n} p_n q_n(t) dt$, modulo some possible integer shifts, is arbitrarily close to $\mu_{\mathcal{N}_n} p_n q_n(t) dt$, the sequence $\{ k^2 \alpha \}$ is not Poissonian along the subsequence of $N$s given by $\mathcal{N}_n$, where $N_n = q_n$, $q_n$ is a denominator in a good rational approximation to $\alpha$. □

### D.6.3 Measure of $\alpha \notin \mathbb{Q}$ with Non-Poissonian Behavior along a sequence $\mathcal{N}_n$

What is the (Lebesgue) measure of $\alpha \notin \mathbb{Q}$ such that there are infinitely many $n$ with

$$\left| \alpha - \frac{p_n}{q_n} \right| < \frac{\alpha_n}{q_n^2}, \alpha_n \to 0.$$

(D.39)

If the above holds, then for any constant $k(\alpha)$, for $n$ large (large depends on both $\alpha$ and $k(\alpha)$) we have

$$\left| \alpha - \frac{p_n}{q_n} \right| < \frac{k(\alpha)}{q_n^2 + \epsilon}.$$

(D.40)

By Theorem ??, this set has (Lebesgue) measure or size 0. Thus, almost no irrational numbers satisfy the conditions of Theorem D.6.2, where *almost no* is relative to the (Lebesgue) measure.

**Exercise D.6.6.** In a topological sense, how many algebraic numbers satisfy the conditions of Theorem D.6.2? How many transcendental numbers satisfy the conditions?

**Exercise D.6.7.** Let $\alpha$ satisfy the conditions of Theorem D.6.2. Consider the sequence $\mathcal{N}_n$, where $\mathcal{N}_n = q_n$, $q_n$ the denominator of a good approximation to $\alpha$. We know the induced probability measures $\mu_{\mathcal{N}_{n,1}} p_n q_n(t) dt$ and $\mu_{\mathcal{N}_{n,1}} p_n q_n(t) dt$ do not converge to $e^{-t} dt$. Do these measures converge to anything?
Remark D.6.8. In \cite{?} it is shown that for most $\alpha$ satisfying the conditions of Theorem D.6.2, there is a sequence $N_j$ along which $\mu_{N_j,1}^\alpha(t)dt$ does converge to $e^{-t}dt$. 
Appendix E

Introduction to the Circle Method

E.1 Origins

The Circle Method is a beautiful idea which makes many problems in additive number theory tractable, and gives excellent heuristics for many more. It originated in investigations by Hardy and Ramanujan (1918) on the partition function, $P(n)$.

For $n \in \mathbb{N}$, $P(n)$ is the number of ways of writing $n$ as a sum of positive integers, where we do not distinguish re-orderings.

For example, if $n = 5$ then

\[
5 = 5 \\
= 4 + 1 \\
= 3 + 2 \\
= 3 + 1 + 1 \\
= 2 + 2 + 1 \\
= 2 + 1 + 1 + 1 \\
= 1 + 1 + 1 + 1 + 1, \tag{E.1}
\]

and $P(5) = 7$. Note we do not distinguish between $3 + 2$ and $2 + 3$.

Hardy and Ramanujan showed that

\[
P(n) \sim \frac{e^\pi}{4\pi n^{3/4}}. \tag{E.2}
\]
The circle method deals with additive problems of the following nature: given some subset \( A \subset \mathbb{N} \) and a positive integer \( s \), what numbers can be written as a sum of \( s \) elements of \( A \)?

Explicitly, what is \( \{a_1 + \cdots + a_s : a_i \in A\} \cap \mathbb{N} \). \hspace{1cm} (E.3)

More generally, we could have a collection of set \( A_1, \ldots, A_s \subset \mathbb{N} \) and study

\[ \{a_1 + \cdots + a_s : a_i \in A_i\} \cap \mathbb{N}. \] \hspace{1cm} (E.4)

We give several problems where the Circle Method is useful. We will consider two choices for \( A \).

The first choice is \( P \), the set of primes; \( P = \{2, 3, 5, 7, 11, \ldots\} \). We will denote elements of \( P \) by \( p \).

The second choice is \( K \), the set of \( k \)-powers of non-negative integers; \( K = \{0, 1, 2^k, 3^k, 4^k, \ldots\} \). We will denote elements of \( K \) by \( n^k \).

1. Consider \( A = P \) and \( s = 2 \). Thus, we are investigating

\[ \{p_1 + p_2 : p_i \text{ a prime}\} \cap \mathbb{N}. \] \hspace{1cm} (E.5)

Goldbach in a letter to Euler (June 7, 1742) conjectured that every integer is the sum of three primes. Euler reformulated this conjecture to every even integer is the sum of two primes. Thus, if Goldbach is correct, the above intersection must contain all even integers.

To date, Goldbach’s conjecture has been verified up to \( 2 \cdot 10^{16} \) (see http://www.ieeta.pt/~tos/goldbach.html).

2. Again let \( A = P \), but now let \( s = 3 \). Thus, we are investigating

\[ \{p_1 + p_2 + p_3 : p_i \text{ a prime}\} \cap \mathbb{N}. \] \hspace{1cm} (E.6)

Vinogradov proved every sufficiently large odd number is the sum of three primes; thus, the above set contains all odd \( n \) from some \( n_0 \) on. Sufficiently large may be taken to be \( e^{11.503} \) (see Chen, J. R. and Wang, T. Z. "On the Goldbach Problem." Acta Math. Sinica 32, 702 – 718, 1989).
3. Let $A = K$ and $s$ is arbitrary for now. We are studying

$$\{n_1^k + \cdots + n_s^k : n_i \in \mathbb{N}\} \cap \mathbb{N}. \quad (E.7)$$

Waring (1770) conjectured that for each $k$, there is an $s = s(k)$ such that every positive integer is the sum of at most $s$ $k^{th}$ powers. First solved by Hilbert (1909); using the Circle Method one can show that we may take $s = 2^k + 1$.

**Exercise E.1.1.** Show that Euler's reformulation is equivalent to Goldbach's original problem.

### E.2 Outline of the Circle Method

#### E.2.1 Without Worrying about Convergence

Let us consider the following problem: *Given $A \subset \mathbb{N}$ and a positive integer $s$, what integers can be written as a sum of $s$ elements of $A$?*

The circle method will often not only tell us *which* numbers are the sum of $s$ elements of $A$, but will give an estimate for how many ways it is possible.

We quickly review some needed preliminaries:

**Definition E.2.1 ($e(z)$).** We define $e(z) = e^{2\pi iz}$.

**Exercise E.2.2.** Let $m, n \in \mathbb{Z}$. Prove

$$\int_0^1 e(nx)e(-mx)dx = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{otherwise} \end{cases} \quad (E.8)$$

Consider the generating function

$$f_A(x) = \sum_{a \in A} e(ax). \quad (E.9)$$

We will ignore convergence problems for now.

Now
\[
(f_A(x))^s = \left( \sum_{a \in A} e(ax) \right)^s \\
= \prod_{i=1}^s \sum_{a_i \in A} e(a_i x) \\
= \sum_{(a_1, \ldots, a_s) \in A^s} e \left( (a_1 + \cdots + a_s)x \right) \\
= \sum_{m \in \mathbb{N}} r_{A,s}(m)e(mx). \quad \text{(E.10)}
\]

The last line follows by collecting terms. Clearly, \(f_A(x)^s\) can be written as \(\sum r_{A,s}(m)e(mx)\) for some coefficients \(r_{A,s}(m)\), as each \(e((a_1 + \cdots + a_s)x)\) equals \(e(mx)\) for some \(m\). How do we get a term \(e(mx)\) above? Every time we have \(a_1 + \cdots + a_s = m\), we obtain an \(e(mx)\) term. Thus, \(r_{A,s}(m)\) is simply the number of ways of writing \(m\) as a sum of \(s\) elements of \(A\)!

The orthogonality relations of the exponentials allow us to pick off these coefficients. Again, ignoring convergence problems,

\[
\int_0^1 f_A^s(x)e(-nx)dx = \int_0^1 \sum_{m \in \mathbb{N}} r_{A,s}(m)e(mx)e(-nx)dx \\
= \sum_{m \in \mathbb{N}} r_{A,s}(m) \int_0^1 e(mx)e(-nx)dx = r_{A,s}(n). \quad \text{(E.11)}
\]

Thus, if we can evaluate the above integral, not only will we know which \(n\) can be written as the sum of \(s\) elements of \(A\), but we will know in how many ways.

### E.2.2 Worrying about Convergence

Such additive problems are interesting only if \(A\) is infinite – otherwise we can just enumerate \(a_1 + \cdots + a_s\) in finite time. And, if \(A\) is infinite, the defining sum for the generating function \(f_A(x)\) need not converge.

For each \(N\), define

\[
A_N = \{a \in A : a \leq N\} = A \cap \{0, 1, \ldots, N\}. \quad \text{(E.12)}
\]
Note the $A_N$s are an increasing sequence of subsets: $A_N \subset A_{N+1}$, and
$$\lim_{N \to \infty} A_N = A.$$ For each $N$, let
$$f_N(x) = \sum_{a \in A_N} e(ax). \quad (E.13)$$

As $f_N(x)$ is a finite sum, all the convergence issues vanish. A similar argument as before yields

$$f_N^s(x) = \prod_{i=1}^{s} \sum_{a_i \in A} e(a_i x)$$

$$= \sum_{(a_1, \ldots, a_s) \in A^s} \sum_{a_i \leq N} e \left( (a_1 + \cdots + a_s) x \right)$$

$$= \sum_{m \leq sN} r_{A,N,s}(m) e(mx). \quad (E.14)$$

Note $r_{A,N,s}(m)$ is the number of ways of writing $m$ as the sum of $s$ elements of $A$, with each element at most $N$. If $m \leq N$, then $r_{A,N,s}(m) = r_{A,s}(m)$, the number of ways of writing $m$ as the sum of $s$ elements of $A$.

For example, if $A = P$ (the set of primes), $N = 10$ and $s = 2$, then $A_N = P_N = \{2, 3, 5, 7\}$. An easy calculation gives $r_{P,7,2}(8) = r_{P,2}(8) = 2$. However, $r_{P,7,2}(14) = 1 (7 + 7)$ but $r_{P,2}(14) = 2 (7 + 7$ and $3 + 11)$.

We have shown

**Lemma E.2.3.** If $m \leq N$:

$$r_A(m) = r_{A,N,s}(m) = \int_0^1 \left( (f_N(x))^s e(-mx) \right) dx. \quad (E.15)$$

Note that, just because we have a closed form expression for $r_{A,N,s}(m)$, this does not mean we can actually evaluate the above integral. Recall, for example, the inclusion - exclusion formula for the number of primes at most $N$. This is an exact formula, but very hard to evaluate.

The difficult part of the Circle Method is evaluating the above. All (successful) applications of the Circle Method proceed in the same way:
1. First, given a set $A$, we construct a generating function $f_N(x)$ for $A_N$. As $f_N(x)$ is a sum of exponentials, we expect there will often be significant cancellation.

2. Second, we split the unit interval into two disjoint pieces, called (for reasons to be explained below) the Major Arcs $\mathcal{M}$ and the Minor Arcs $m$. Then

$$r_{A,s}(m) = r_{A,N,s}(m) = \int_{\mathcal{M}} f_N^s(x) e(-mx) dx + \int_{m} f_N^s(x) e(-mx) dx. \quad (E.16)$$

3. On the Major Arcs $\mathcal{M}$, we find a function which, up to lower order terms, agrees with $f_N^s(x)$ and is easily integrable. We then perform the integration, and are left with a contribution over the Major Arcs which is bounded away from zero.

4. If we can show that, as $N \to \infty$, the Minor Arcs’ contribution is of lower order than the Major Arcs’ contribution, we will have shown that $r_{A,N}(m) > 0$, proving that large $m$ can be represented as a sum of $s$ elements of $A$.

The last is the most difficult step. It is often highly non-trivial to obtain the required cancellation over the Minor Arcs.

For the problems mentioned, we are able to obtain the needed cancellation for $A = P$ and $s = 3$ (every large odd number is the sum of three primes), but not $A = P$ and $s = 2$ (we will give some heuristics later). For $A = K$ (the set of $k^{th}$ powers of integers), we can obtain the desired cancellation for $s = 2k + 1$; however, in general we expect the result to hold for smaller $s$.

### E.3 Sums of Two and Three Primes

We consider the following problems: Let $A = P$ be the set of primes, and let $A_N = P_N = \{p \in P : p \leq N\}$. Is every odd number the sum of three primes ($s = 3$)? Is every even number the sum of two primes ($s = 2$)?

We will not prove either claim (especially the second, which is believed but not yet known). Instead, we will provide a sketch of how the Circle Method is applied to such problems: what are the Major and Minor Arcs? How do we obtain the necessary cancellation when $s = 3$ but not when $s = 2$?
These examples are well known in the literature, and we will content ourselves with a very brief introduction. In the next chapter, we will give a very thorough treatment of another Circle Method problem, one not as well known.

For analytical reasons, it turns out that it is more convenient to analyze the weighted function

$$F_N(x) = \sum_{p \leq N} \log p \cdot e(px)$$  \hspace{1cm} (E.17)

instead of $f_N(x)$. Working analogously as before, we are led to

$$R_{N,s}(m) = \int_0^1 F^*_N(x)e(-mx)dx,$$  \hspace{1cm} (E.18)

where now

$$R_{N,s}(m) = \sum_{\sum_{p_i \leq N} p_i = m} \log p_1 \cdots \log p_s.$$  \hspace{1cm} (E.19)

By partial summation, it is very easy to go from $R_{N,s}(m)$ to $r_{N,s}(m)$.

**Exercise E.3.1.** Using partial summation, relate $R_{N,s}(m)$ and $r_{N,s}(m)$.

Thus, if we can show $R_{N,s}(m)$ is non-zero for $N$ and $m$ sufficiently large, then $r_s(m) = r_{P,s}(m)$ will also be non-zero.

**E.3.1 Needed Number Theory Results**

We will use the following statements freely:

**Theorem E.3.2 (Prime Number Theorem).** Let $\pi(x)$ denote the number of primes at most $x$. Then

$$\pi(x) = \sum_{p \leq x} 1 = \frac{x}{\log x} + \text{smaller}. \hspace{1cm} (E.20)$$

Upon applying Partial Summation, we may rewrite the above as

$$\sum_{p \leq x} \log p = x + \text{smaller}. \hspace{1cm} (E.21)$$
Theorem E.3.3 (Siegel-Walfisz). Let $C, B > 0$, and let $a$ and $q$ be relatively prime. Then

$$\sum_{\substack{p \leq x \atop p \equiv a(q)}} \log p = \frac{x}{\phi(q)} + O\left(\frac{x}{\log^C x}\right) \quad (E.22)$$

for $q \leq \log^B x$, and the constant above does not depend on $x$, $q$ or $a$ (i.e., it only depends on $C$ and $B$).

E.3.2 Average Sizes of $|f_N(x)|$

Recall

$$F_N(x) = \sum_{p \leq N} \log p \cdot e(px) \quad (E.23)$$

Lemma E.3.4. $|F_N(x)| \leq N$ plus lower order terms.

Proof: The number of primes $p \leq N$ is $\pi(x) = \frac{x}{\log x}$ plus lower order terms. Hence

$$|F_N(x)| = \left| \sum_{p \leq N} \log p \cdot e(px) \right| \leq \sum_{p \leq N} \log N = \log N \cdot \pi(x), \quad (E.24)$$

which is of size $\log N \cdot \frac{N}{\log N}$ plus lower order terms.

Alternatively, we could have argued that

$$|F_N(x)| \leq \sum_{p \leq N} \log p = N + \text{smaller.} \quad (E.25)$$

Both approaches return the same bound. □

Lemma E.3.5. $F_N(0) = F_N(1) = N + \text{smaller, and } F_N(\frac{1}{2}) = -N + \text{smaller.}$

Proof: $F_N(0)$ and $F_N(1)$ are immediate, as $e(p \cdot 0) = 1$ for all $p$. For $F_N(\frac{1}{2})$, note
\[ e \left( p \cdot \frac{1}{2} \right) = e^{\pi i p} = \begin{cases} -1 & \text{if } p \text{ is odd} \\ +1 & \text{if } p \text{ is even} \end{cases} \quad (E.26) \]

As there is only one even prime,

\[ F_N \left( \frac{1}{2} \right) = \log 2 - \sum_{p \leq N} \log p, \quad (E.27) \]

and the argument proceeds as before.

**Exercise E.3.6.** How large is \( F_N \left( \frac{1}{4} \right) \) and \( F_N \left( \frac{3}{4} \right) \)? We will calculate this exactly in the next chapter.

Thus, \( F_N(x) \) is occasionally as large as \( N \). We can, however, show that the average square of \( F_N(x) \) is significantly smaller:

**Lemma E.3.7.** The average value of \( |F_N(x)|^2 \) is \( N \log N \).

Proof: The following trivial observation will be extremely useful in our arguments. Let \( g(x) \) be a complex-valued function, and let \( \overline{g}(x) \) be its complex conjugate. Then \( |g(x)|^2 = g(x)\overline{g}(x) \).

In our case, as \( F_N(x) = F_N(-x) \) we have

\[
\int_0^1 |F_N(x)|^2 dx = \int_0^1 F_N(x)F_N(-x)dx
= \int_0^1 \sum_{p \leq N} \log p \cdot e(px) \sum_{q \leq N} \log q \cdot e(-qx)dx
= \sum_{p \leq N} \sum_{q \leq N} \log p \log q \int_0^1 e((p-q)x) dx
= \sum_{p \leq N} \log^2 p. \quad (E.28)
\]

Using \( \sum_{p \leq N} \log p = N + \text{small} \) and Partial Summation, we can show

\[ \sum_{p \leq N} \log^2 p = N \log N + \text{smaller}. \quad (E.29) \]
Thus,

$$\int_0^1 |F_N(x)|^2 dx = N \log N + \text{smaller.} \quad \text{(E.30)}$$

Taking square-roots, we see on average $|F_N(x)|^2$ is $N \log N$, significantly smaller than the maximum possible value ($N^2$). Thus, we see we are getting almost square-root cancellation on average.  

\textbf{Exercise E.3.8.} Using the Prime Number Theorem and Partial Summation, prove

$$\sum_{p \leq N} \log^2 p = N \log N + \text{smaller.} \quad \text{(E.31)}$$

\textbf{E.3.3 Definition of the Major and Minor Arcs}

We split the unit interval $[0, 1)$ into two disjoint parts, the Major and the Minor arcs.

Roughly, the Major arcs will be a union of very small intervals centered at rationals with small denominator (relative to $N$). Near these rationals, we will be able to approximate $F_N(x)$ very well, and $F_N(x)$ will be of size $N$.

The minor arcs will be the rest of $[0, 1)$; we will show that $F_N(x)$ is significantly smaller than $N$ here.

In all applications of the Circle Method, the choice of the Major and Minor arcs will depend on the problem. The idea is as follows: the Major arcs will have small length; however, the function will be well understood here, and we will be able to evaluate the integral over the Major Arcs up to a small error. On the Minor Arcs, which have most of the length, there is good cancellation in the series expansion of our function; thus, the integral over the minor arcs will be smaller than the contribution from the Major Arcs (even though most of the length of $[0, 1]$ is in the Minor Arcs).

Obtaining such cancellation in the series expansion is not easy – this is the hardest part of the problem. In many cases, we are not able to prove the integral over the Minor Arcs is smaller than the contribution from the Major Arcs, though we often believe this will be the case.

\textbf{Major Arcs}

Let $B > 0$, and let $Q = (\log N)^B \ll N$. 

165
For each \( q \in \{1, 2, \ldots, Q\} \) and \( a \in \{1, 2, \ldots, q\} \) with \( a \) and \( q \) relatively prime, consider the set

\[
\mathcal{M}_{a,q} = \left\{ x \in [0,1) : \left| x - \frac{a}{q} \right| < \frac{Q}{N} \right\}.
\]  

(E.32)

We also add in one interval centered at either 0 or 1, i.e., the "interval" (or wrapped-around interval)

\[
\left[ 0, \frac{Q}{N} \right] \cup \left[ 1 - \frac{Q}{N}, 1 \right].
\]  

(E.33)

**Exercise E.3.9.** Show, if \( N \) is large, that the major arcs \( \mathcal{M}_{a,q} \) are disjoint for \( q \leq Q \) and \( a \leq q \), \( a \) and \( q \) relatively prime.

We define the Major Arcs to be the union of each arc \( \mathcal{M}_{a,q} \):

\[
\mathcal{M} = \bigcup_{q=1}^{Q} \bigcup_{\substack{a=1 \\ (a,q)=1}}^{\min(a,q)} \mathcal{M}_{a,q},
\]  

(E.34)

where \((a, q)\) is the greatest common divisor of \( a \) and \( q \).

**Exercise E.3.10.** Show \(|\mathcal{M}| < \frac{2Q^3}{N}\). As \( Q = \log^B N \), this implies as \( N \to \infty \), the major arcs are zero percent of the unit interval.

We really should write \( \mathcal{M}_{a,q}(N) \), as the arc about \( \frac{a}{q} \) depends on \( N \). Standard practice, however, is to drop this \( N \), as \( N \) is usually fixed throughout the problem, and sent to infinity only at the end.

There are two measures of small. First, we require the denominator \( q \) to be small relative to \( N \): \( q \leq Q = \log^B N \). Note once a denominator is small for some \( N \), it is small for all larger integers.

The second is the size of the arc: we want a small interval about \( \frac{a}{q} \). The interval has length \( \frac{2Q}{N} = \frac{2 \log^B N}{N} \). Thus, for a fixed \( \frac{a}{q} \), the size of the arc about it tends to zero as \( N \) tends to infinity.

**Minor Arcs**

The Minor Arcs, \( m \), are whatever is not in the Major Arcs. Thus,

\[
m = [0,1) - \mathcal{M}.
\]  

(E.35)
Clearly, as $N \to \infty$, almost all of $[0, 1)$ is in the Minor Arcs. The hope is that, staying away from rationals with small denominator, that we will be able to obtain significant cancellation in $F_N(x)$.

**E.3.4 Contribution from the Major Arcs**

Fix a $q \leq Q$ and an $a \leq q$ with $a$ and $q$ relatively prime. We evaluate $F\left(\frac{a}{q}\right)$.

\[
F\left(\frac{a}{q}\right) = \sum_{p \leq N} \log p \cdot e^{2\pi i \cdot \frac{a}{q} \cdot \frac{p}{q}}
\]

\[
= \sum_{r=1}^{q} \sum_{p \equiv r (q) \ p \leq N} \log p \cdot e^{2\pi i \cdot \frac{aq}{q}}
\]

\[
= \sum_{r=1}^{q} e^{2\pi i \cdot \frac{ar}{q}} \sum_{p \equiv r (q) \ p \leq N} \log p
\]  

(E.36)

Note the beauty of the above. The dependence on $p$ in the original sums is very weak – there is a $\log p$ factor, and there is $e^{\left(\frac{aq}{q}\right)}$. In the exponential, we only need to know $p \mod q$. Now, $p$ runs from 2 to $N$, and $q$ is at most $\log B^N$. Thus, in general $p \gg q$.

We use the Siegel-Walfisz Theorem. We first remark that we may assume $r$ and $q$ are relatively prime. Why? If $p \equiv r \mod q$, this means $p = \alpha q + r$ for some $\alpha \in \mathbb{N}$. If $r$ and $q$ have a common factor, there can be at most one prime $p$ (namely $r$) such that $p \equiv r \mod q$, and this can easily be shown to give a negligible contribution.

For any $C > 0$, by the Siegel-Walfisz Theorem

\[
\sum_{p \equiv r (q) \ p \leq N} \log p = \frac{N}{\phi(q)} + O\left(\frac{N}{\log^C N}\right).
\]  

(E.37)

Now, as $\phi(q)$ is at most $q$ which is at most $\log B^N$, we see that if we take $C > B$, the main term is significantly greater than the error term.
Note the Siegel-Walfisz Theorem would be useless if \( q \) were "large", say \( q \approx N^\epsilon \). Then the main term would be like \( N^{1-\epsilon} \), which would be smaller than the error term.

This is one reason why, in constructing the major arcs, we take the denominators to be small (of size \( \log^B N \)). Theorems such as Siegel-Walfisz are our main tools for evaluating the necessary prime sums, and they are useful only when the error term is less than the main term.

Thus, we find

\[
F\left(\frac{a}{q}\right) = \sum_{r=1}^{q} e^{2\pi i \frac{ar}{q}} \frac{N}{\phi(q)} + \text{smaller}
= \frac{N}{\phi(q)} \sum_{r=1}^{q} e^{2\pi i \frac{ar}{q}} + \text{smaller.}
\]  

(E.38)

We merely sketch what happens now.

First, one shows that for \( x \in \mathcal{M}_{a,q} \) that \( F_N(x) \) is very close to \( F\left(\frac{a}{q}\right) \). This is a standard analysis, similar to a Taylor Series Expansion – the constant term is a good approximation if you are sufficiently close. In practice, it is technically easier to find a function which is non-constant and agrees with \( F_N\left(\frac{a}{q}\right) \).

As the major arcs are distinct,

\[
\int_{\mathcal{M}} F_N^3(x) e(-mx) dx = \sum_{q=1}^{Q} \sum_{(a,q)=1} \int_{\mathcal{M}_{a,q}} F_N^3(x) e(-mx) dx.
\]

(E.39)

For heuristic purposes, we approximate \( F_N^3(x) \) by \( F\left(\frac{a}{q}\right) \); integrating a constant gives the constant times the length of the interval. Each of the major arcs has length \( \frac{2Q^3}{N} \). Thus we find that, up to a smaller correction term, the contribution from the Major Arcs is
\[
\int_{\mathcal{M}} F_N^3(x)e(-mx)dx = \frac{2Q^3}{N} \sum_{q=1}^{Q} \sum_{a=1}^{\phi(q)} \left( \frac{N}{\phi(q)} \sum_{r=1}^{q} e^{2\pi i \frac{ar}{q}} \right)^3 e\left(- \frac{2\pi ima}{q}\right)
\]

\[
= N^2 \cdot 2Q^3 \sum_{q=1}^{Q} \frac{1}{\phi(q)^3} \sum_{a=1}^{\phi(q)} \left( \sum_{r=1}^{q} e^{2\pi i \frac{ar}{q}} \right)^3 e\left(- \frac{2\pi ima}{q}\right).
\]

(E.40)

To complete the proof, we need to show that what is multiplying \(N^2\) is non-negative, and not too small. An elementary analysis often bounds the factor away from 0 and infinity. We will perform such an analysis in the next chapter for a different problem.

Note that, up to factors of \(\log N\) (which are important!), the contribution from the Major Arcs is of size \(N^2\). A more careful analysis (where we don’t just replace \(f_N(x)\) with \(f_N(\frac{x}{q})\) on \(\mathcal{M}_{a,q}\)) would show that the Major Arcs contribute a positive quantity of size \(N^2\).

Explicitly, the Major Arcs contribute

\[
\mathfrak{S}(N) \frac{N^2}{2} + \text{smaller},
\]

where \(\exists c_1, c_2 > 0\) such that, for all \(N, c_1 < \mathfrak{S}(N) < c_2\).

\(\mathfrak{S}(N)\) is called the Singular Series, and will be discussed in more detail in the next chapter.

### E.3.5 Contribution from the Minor Arcs

We bound the contribution from the minor arcs to \(r_N(m)\):

\[
\left| \int_{\mathcal{M}} F_N^3(x)e(-mx)dx \right| \leq \int_{\mathcal{M}} |F_N(x)|^2dx
\]

\[
\leq \left( \max_{x \in \mathcal{M}} |F_N(x)| \right) \int_{\mathcal{M}} |F_N(x)|^2dx
\]

\[
\leq \left( \max_{x \in \mathcal{M}} |F_N(x)| \right) \int_{0}^{1} F_N(x)F_N(-x)dx
\]

\[
\leq \left( \max_{x \in \mathcal{M}} |F_N(x)| \right) N \log N. \quad \text{(E.42)}
\]
As the minor arcs are most of the unit interval, replacing \( \int_m \) with \( \int_0^1 \) doesn’t introduce much of an over-estimation.

**In order for the Circle Method to succeed, we need a non-trivial, good bound** for

\[
\max_{x \in m} |F_N(x)|
\]  

(E.43)

This is where most of the difficulty arises, showing that there is good cancellation in \( F_N(x) \) if we stay away from rationals with small denominator.

Thus, we need an estimate such as

\[
\max_{x \in m} |F_N(x)| \leq \frac{N}{\log^{1+\epsilon} N},
\]  

(E.44)

or even

\[
\max_{x \in m} |F_N(x)| \ll o \left( \frac{N}{\log N} \right).
\]  

(E.45)

Relative to the average size of \( |F_N(x)|^2 \), which is \( N \log N \), this is significantly larger. Unfortunately, as we have inserted absolute values, it is not enough to bound \( |F_N(x)| \) on average – we need to obtain a good bound uniformly in \( x \). This is a significantly more delicate question. We know such a bound cannot be true for all \( x \in [0, 1) \) (see below, and not that \( F_N(0) = N \)). The hope is that if \( x \) is not near a rational with small denominator, we will get moderate cancellation.

While this is very reasonable to expect, it is not easy to prove.

Of course, we do not need to prove such a bound – all we need to do is prove that the Minor Arcs contribute less than the Major Arcs. This would suffice to prove \( R_{N,3}(m) > 0 \) for \( m \) large.

Unfortunately, it is very hard to evaluate the Minor Arc contribution. As we can evaluate \( \int_0^1 |f_N^2(x)| dx \) exactly, this seems a very natural way to proceed.

**Exercise E.3.11.** Using the definition of the minor arcs, bound

\[
\left| \int_0^1 |f_N^2(x)| dx - \int_m |f_N^2(x)| dx \right|.
\]  

(E.46)

Show, therefore, that there is little harm in extending the integral of \( |f_N^2(x)| \) to all of \([0, 1]\).
E.4 Why Goldbach is Hard

Using

\[ F_N(x) = \sum_{p \leq N} \log p \cdot e^{2\pi ipx}, \]  

(E.47)

we find we must study

\[ \int_0^1 F_N^s(x) dx, \]  

(E.48)

where \( s = 3 \) if we are looking at \( p_1 + p_2 + p_3 = 2n + 1 \) and \( s = 2 \) if we are looking at \( p_1 + p_2 = 2n \). Why does the circle method work for \( s = 3 \) but fail for \( s = 2 \)?

E.4.1 \( s = 3 \) Sketch

Let us recall briefly the \( s = 3 \) case. Near rationals \( \frac{a}{q} \) with small denominator (small means \( q \leq \log^B N \)), we can evaluate \( F_N(\frac{a}{q}) \). Using Taylor, if \( x \) is very close to \( \frac{a}{q} \), we expect \( F_N(x) \) to be close to \( F_N(\frac{a}{q}) \).

The Major Arcs have size \( \log^B N \). As \( F_N(x) \) is around \( N \) near such rationals, we expect the integral of \( F_N^s(x)e(-mx) \) to be \( N^2 \) times a power of \( \log N \). Doing a careful analysis of the singular series shows that the contribution is actually \( \Theta(N)N^2 \), where there exist constants independent of \( N \) such that \( 0 < c_1 < \Theta(N) < c_2 < \infty \).

A direct calculation shows that

\[ \int_0^1 |F_N(x)|^2 dx = \int_0^1 F_N(x)F_N(-x) dx = N \log N. \]  

(E.49)

Thus, if \( m \) denotes the minor arcs,

\[ \left| \int_m F_N^3(x)e(-mx) dx \right| \leq \max_{x \in m} |F_N(x)| \int_0^1 |F_N(x)|^2 dx \leq N \log N \cdot \max_{x \in m} |F_N(x)|, \]  

(E.50)

As the major arcs contribute \( \Theta(N)N^2 \), we need to show
\[ \max_{x \in \mathbb{m}} |F_N(x)| \ll \frac{N}{\log D N}. \tag{E.51} \]

Actually, we just need to show the above is \( \ll o\left(\frac{N}{\log N}\right) \). This is the main difficulty – the trivial bound is \( |F_N(x)| \leq N \). As \( F_N(0) = N \) plus lower order terms, we cannot do better in general.

The key observation is that, if we stay away from rationals with small denominator, we can prove there is cancellation in \( F_N(x) \). While we don’t go into details here (see, for example, Nathanson’s Additive Number Theory: The Classical Bases, Chapter 7), the savings we obtain is small. We show

\[ \max_{x \in \mathbb{m}} |F_N(x)| \ll \frac{N}{\log D N}. \tag{E.52} \]

Note that Equation E.49 gives us significantly better cancellation on average, telling us that \( |F_N(x)|^2 \) is usually of size \( N \).

Thus, it is our dream to be so lucky as to see \( \left| \int_I |F_N(x)|^2 dx \right| \) for any \( I \subset [0,1) \), as we can evaluate this extremely well.

**E.4.2 \( s = 2 \) Sketch**

What goes wrong when \( s = 2 \)? As a first approximation, if \( s = 3 \) has the Major Arcs contributing a constant times \( N^2 \) (and \( F_N(x) \) was of size \( N \) on the Major Arcs), one might guess that the Major Arcs for \( s = 2 \) will contribute a constant times \( N \).

How should we estimate the contribution from the Minor Arcs? We have \( F_N^2(x) \). If we just throw in absolute values we get

\[ \left| \int_m F_N^2(x)e(-mx) dx \right| \leq \int_0^1 |F_N(x)|^2 dx = N. \tag{E.53} \]

Note, unfortunately, that this is the same size as the expected contribution from the Major Arcs!

We could try pulling a \( \max_{x \in \mathbb{m}} |F_N(x)| \) outside the integral, and hope to get a good savings. The problem is this leaves us with \( \int_m |F_N(x)| dx \), which we now analyze.

Recall the Cauchy-Schwartz Inequality:
Lemma E.4.1.

\[
\int_0^1 |f(x)g(x)| dx \leq \left( \int_0^1 |f(x)|^2 dx \right)^{\frac{1}{2}} \cdot \left( \int_0^1 |g(x)|^2 dx \right)^{\frac{1}{2}}.
\] (E.54)

For a proof, see Lemma ??.

Thus,

\[
\left| \int_m F_N^2(x)e(-mx)dx \right| \leq \max_{x \in m} |F_N(x)| \int_0^1 |F_N(x)| dx
\]

\[
\leq \max_{x \in m} |F_N(x)| \left( \int_0^1 |F_N(x)|^2 dx \right)^{\frac{1}{2}} \cdot \left( \int_0^1 1^2 dx \right)^{\frac{1}{2}}
\]

\[
\leq \max_{x \in m} |F_N(x)| \cdot (N \log N)^{\frac{1}{2}} \cdot 1.
\] (E.55)

As the Major Arcs contribute something of size \( N \), we would need

\[
\max_{x \in m} |F_N(x)| \ll o \left( \sqrt{\frac{N}{\log N}} \right).
\] (E.56)

There is almost no chance of such cancellation (this is better than square-root cancellation!). We know

\[
\int_0^1 |F_N(x)|^2 dx = N \log N \text{ plus lower order terms.} \] (E.57)

Thus, the average size of \(|F_N(x)|^2\) is \( N \), so we expect \(|F_N(x)|\) to be about \( \sqrt{N} \). To get \( o(\sqrt{N}) \) would be unbelievably good fortune!

While the above sketch shows the Circle Method is not, at present, powerful enough to handle the Minor Arc contributions, all is not lost. The quantity we need to bound is

\[
\left| \int_m F_N^2(x)e(-mx)dx \right|.
\] (E.58)

However, we have instead been studying

\[
\int_m |F_N(x)|^2 dx
\] (E.59)

and

173
Thus, we are ignoring the probable oscillation / cancellation in the integral $\int F_N(x)e(-mx)dx$. It is this cancellation that will lead to the Minor Arcs contributing significantly less than the Major Arcs.

However, showing there is cancellation in the above integral is very difficult. It is a lot easier to work with absolute values.

Further, just because we cannot prove that the Minor Arc contribution is small, does not mean the Circle Method is not useful. Numerical simulations confirm, for many problems, that the Minor Arcs do not contribute for many $N$. For example, for $N \leq 10^9$, the observed values are in excellent agreement with the Major Arc predictions.
Appendix F

Circle Method: Heuristics for Germain Primes

F.1 Germain Primes

We apply the Circle Method to the following problem:

Consider an odd prime \( p \). Clearly, \( p - 1 \) cannot be prime, as it is even; however, \( \frac{p-1}{2} \) could be prime.

We say \( p \) is a Germain prime (or \( p \) and \( \frac{p-1}{2} \) are a Germain prime pair) if both \( p \) and \( \frac{p-1}{2} \) are prime. Equivalently, we often re-write and ask for primes \( p \leq N \) such that \( 2p + 1 \) is prime.

Germain primes have many wonderful properties. Around 1825, Sophie Germain proved that if \( p \) is a Germain prime, then Fermat’s Last Theorem is true for exponent \( p \) (the only integer solutions of \( x^p + y^p = z^p \) have \( xyz = 0 \)). Recent advances in cryptography are known to run fast if there are many Germain primes.

Germain primes are just one example of the following type of problem: for \( p \leq N \), how often are \( p \) and \( ap + b \) prime for fixed \( a \) and \( b \).

One well known example is the famous Twin Prime Conjecture, that there are infinitely many primes \( p \) such that \( p + 2 \) is also prime. It is not known if this is true; however, Brun has shown that the sum of the reciprocals of the twin primes converges (unlike the sum of the reciprocals of the primes, which diverges). Therefore, if there are infinitely many twin primes, there are in some sense "less" twin primes than primes.

Explicitly, Brun proved that there exists an \( N_0 \) such that for all \( N > N_0 \), the number of twin primes less than \( N \) is at most \( \frac{100N}{\log^2 N} \). Compare this to the number
of primes less than \( N \), which is of size \( \frac{N}{\log N} \).

More generally, one can consider how many primes \( p \leq N \) there are with \( p + 2m \) also prime.

The Circle Method is a powerful tool for such investigations. In all such problems, numerical experiments agree with the Major Arc calculations; unfortunately, in all these problems we are unable to prove that the Minor Arcs contribute less than the Major Arcs.

We have chosen to go through the calculation of the number of Germain primes less than \( N \) rather than twin primes (or even numbers being the sum of two primes or odd numbers the sum of three primes) as these other problems are well documented in the literature.

Note the Germain problem is slightly different from the original formulations of the Circle Method. Here, we are investigating how often \( p_1 - 2p_2 = 1 \), with \( p_2 < p_1 < N \).

It is easy to re-write this in the original formulation. Let us choose two sets

\[
A_1 = \{ p : p \text{ a prime} \} \\
A_2 = \{ -2p : p \text{ a prime} \}.
\]  

(F.1)

To have our sums converge, let

\[
A_{1N} = \{ p : p \text{ a prime, } p \leq N \} \\
A_{2N} = \{ -2p : p \text{ a prime, } p \leq N \}.
\]  

(F.2)

We are interested in

\[
\{ a_1 + a_2 : a_i \in A_{iN} \} \cap \{1\}.
\]  

(F.3)

In the original applications of the Circle Method, we were just interested in whether or not a number \( m \) was in \( A_{1N} \cap A_{2N} \). In order to show \( m \) was in the intersection, we calculated \( r_{A,2}(m) \), showing it was bounded away from 0.

Now, we know \( 1 \in A_{1N} \cap A_{2N} \) (for example, 11 is a Germain prime, as \( \frac{11-1}{2} = 5 \) is prime). We are now interested in the size of \( r_{A,N,2}(m) \) (which will tell us how many Germain primes there are less than \( N \), instead of just \( r_{A,N,2}(m) > 0 \).
F.2 Preliminaries

We will apply the Circle Method and calculate the contribution from the Major Arcs for the Germain problem, namely, how many primes $p$ are there such that $\frac{p-1}{2}$ is also prime?

F.2.1 Definitions

Let

$$e(x) = e^{2\pi ix} \quad (F.4)$$

and

$$\lambda(n) = \begin{cases} \log p & \text{if } n = p \text{ is prime} \\ 0 & \text{otherwise} \end{cases} \quad (F.5)$$

Finally, define

$$c_q(a) = \sum_{r=1}^{q} e\left(\frac{ra}{q}\right). \quad (F.6)$$

F.2.2 Partial Summation

Recall

Lemma F.2.1 (Partial Summation: Discrete Version). Let $A_N = \sum_{n=1}^{N} a_n$. Then

$$\sum_{n=M}^{N} a_n b_n = A_N b_N - A_{M-1} b_M + \sum_{n=M}^{N-1} A_n (b_n - b_{n+1}) \quad (F.7)$$

Proof. Since $A_n - A_{n-1} = a_n$,
\[
\sum_{n=M}^{N} a_n b_n = \sum_{n=M}^{N} (A_n - A_{n-1}) b_n
\]
\[
= (A_N - A_{N-1}) b_N + (A_{N-1} - A_{N-2}) b_{N-1} + \cdots + (A_M - A_{M-1}) b_M
\]
\[
= A_N b_N + (-A_{N-1} b_N + A_{N-1} b_{N-1}) + \cdots + (-A_M b_{M+1} + A_M b_M) - a_{M-1} b_M
\]
\[
= A_N b_N - a_M b_M + \sum_{n=M}^{N-1} A_n (b_n - b_{n+1}). \quad \text{(F.8)}
\]

Recall

**Lemma F.2.2 (Abel’s Summation Formula - Integral Version).** Let \( h(x) \) be a continuously differentiable function. Let \( A(x) = \sum_{n \leq x} a_n. \) Then

\[
\sum_{n \leq x} a_n h(n) = A(x) h(x) - \int_{1}^{x} A(u) h'(u) du \quad \text{(F.9)}
\]

See, for example, W. Rudin, *Principles of Mathematical Analysis*, page 70.

**F.2.3 Siegel-Walfisz**

**Theorem F.2.3. [Siegel-Walfisz]** Let \( C, B > 0, \) and let \( a \) and \( q \) be relatively prime. Then

\[
\sum_{p \leq x \atop p \equiv a \pmod{q}} \log p = \frac{x}{\phi(q)} + O\left(\frac{x}{\log^C x}\right) \quad \text{(F.10)}
\]

for \( q \leq \log^B x, \) and the constant above does not depend on \( x, q \) or \( a \) (ie, it only depends on \( C \) and \( B \)).

**F.2.4 Germain Integral**

Define
\[
f_{1N}(x) = \sum_{p_1 \leq N} \log p_1 \cdot e(p_1 x)
\]
\[
f_{2N}(x) = \sum_{p_2 \leq N} \log p_2 \cdot e(-2p_2 x)
\]
\[
f_N(x) = \sum_{p_1 \leq N} \sum_{p_2 \leq N} \log p_1 \log p_2 \cdot e\left((p_1 - 2p_2)x\right). \quad (F.11)
\]

Consider

\[
\int_{-\frac{1}{2}}^{\frac{1}{2}} f_N(x)e(-x)dx = \sum_{p_1 \leq N} \sum_{p_2 \leq N} \log p_1 \log p_2 \int_{-\frac{1}{2}}^{\frac{1}{2}} e\left((p_1 - 2p_2 - 1)x\right)dx. \quad (F.12)
\]

Note

\[
\int_{-\frac{1}{2}}^{\frac{1}{2}} e\left((p_1 - 2p_2 - 1)x\right)dx = \begin{cases} 
1 & \text{if } p_1 - 2p_2 - 1 = 0 \\
0 & \text{if } p_1 - 2p_2 - 1 \neq 0
\end{cases} \quad (F.13)
\]

Thus, we get a contribution of \( \log p_1 \log p_2 \) if \( p_1 \) and \( p_2 = \frac{p_1 - 1}{2} \) are both prime. Thus,

\[
\int_{-\frac{1}{2}}^{\frac{1}{2}} f_N(x)e(-x)dx = \sum_{p_1 \leq N} \log p_1 \log p_2. \quad (F.14)
\]

The above is a weighted counting of Germain primes.

**F.2.5 Major and Minor Arcs**

Let \( B \) be a positive integer, \( Q = \log^R N \). Define the Major Arc \( M_{a,q} \) for each pair \((a, q)\) with \( a \) and \( q \) relatively prime and \( 2 \leq q \leq Q \) by

\[
M_{a,q} = \left\{ x \in (0, 1) : \left| x - \frac{a}{q} \right| < \frac{Q}{N} \right\}. \quad (F.15)
\]

We also add in one interval centered at either 0 or 1, ie, the "interval" (or wrapped-around interval)
\[ \left[ \frac{0}{N}, \frac{1}{N} \right] \cup \left[ 1 - \frac{Q}{N}, 1 \right]. \]  
\text{(F.16)}

For convenience, we often use the interval \([-\frac{1}{2}, \frac{1}{2}]\) instead of \([0, 1]\), in which case we would have
\[ \left[ -\frac{1}{2}, -\frac{1}{2} + \frac{Q}{N} \right] \cup \left[ \frac{1}{2} - \frac{Q}{N}, \frac{1}{2} \right]. \]  
\text{(F.17)}

For functions that are periodic of period one, we could instead consider
\[ \left[ \frac{1}{2} - \frac{Q}{N}, \frac{1}{2} + \frac{Q}{N} \right]. \]  
\text{(F.18)}

The Major Arcs are defined by
\[ \mathcal{M} = \bigcup_{q \leq Q} \bigcup_{a=1}^{\phi(q)} \mathcal{M}_{a,q}. \]  
\text{(F.19)}

The Minor Arcs, \(m\), are whatever is \textit{not} in the Major Arcs. Thus, the Major Arcs are the subset of \([-\frac{1}{2}, \frac{1}{2}]\) or \([0, 1]\) near rationals with \textit{small} denominators, and the Minor Arcs are what is left. Here \textit{near} and \textit{small} are relative to \(N\).

Then
\[ \int_{-\frac{1}{2}}^{\frac{1}{2}} f_N(x)e(-x)dx = \int_{\mathcal{M}} f_N(x)e(-x)dx + \int_{\mathcal{M}} f_N(x)e(-x)dx. \]  
\text{(F.20)}

\textbf{F.2.6 Formula for Number of Germain Prime Pairs}

The number of Germain prime pairs \((p_1, p_2)\) with \(p_2 = \frac{p_1 - 1}{2}\) and \(p_1 \leq N\) is given by
\[ \sum_{\substack{p_1 \leq N \\quad \text{and} \quad p_2 = \frac{p_1 - 1}{2}}} 1. \]  
\text{(F.21)}

We will calculate
\[
\sum_{\substack{p_1 \leq N \\ p_2 = \frac{p_1 - 1}{2}}} \log p_1 \log p_2 = \int_{-\log f_N}^{\frac{1}{2}} f_N(x) e(-x) dx
\]
\[
= \int_M f_N(x) e(-x) dx + \int_m f_N(x) e(-x) dx. \quad \text{(F.22)}
\]

By partial summation, it is easy to pass from knowledge of the weighted sum (with the \( \log p_1 \log p_2 \) factors) to the unweighted sum.

There is a lot of oscillation in the generating function \( f_N(x) \). For generic \( x \) in the interval, we expect a lot of cancellation in the size of \( f_N(x) \). Unfortunately, at present no one is able to prove that the Minor Arcs contribute less than the Major Arcs. In this respect, the Germain problem is like Goldbach and many other two-prime problems, such as for fixed \( 2m \), how many prime pairs \( (p, p+2m) \) are there with \( p \leq x \)?

Numerical simulations (for \( x \) up to \( 10^9 \) and higher) support the conjecture that the Minor Arcs do not contribute for the Germain problem. Explicitly, the observed number of Germain prime pairs in this range agrees with the prediction from the Major Arcs.

We will content ourselves with calculating the contribution from the Major Arcs. Thus, in the sequel we investigate
\[
\int_M f_N(x) e(-x) dx. \quad \text{(F.23)}
\]

F.2.7 Reformulation of Germain Integral

\[
f_{1N}(x) = \sum_{m_1 \leq N} \lambda(m_1) \cdot e(m_1 x)
\]
\[
f_{2N}(x) = \sum_{m_2 \leq N} \lambda(m_2) \cdot e(-2m_2 x)
\]
\[
f_N(x) = \sum_{m_1 \leq N} \sum_{m_2 \leq N} \lambda(m_1) \lambda(m_2) \cdot e\left((m_1 - 2m_2)x\right). \quad \text{(F.24)}
\]

We investigate
\[ \int_{\mathcal{M}} f_N(x)e(-x)dx. \] (F.25)

We will show the Major Arcs contribute, up to lower order terms, \( T_2 N \), where \( T_2 \) is a constant independent of \( N \). The length of the Major Arc \( \mathcal{M}_{a,q} \) is \( \frac{Q}{N} \). We sum over \((a, q) = 1 \) and \( q \leq Q \). Thus, the total length is bounded by

\[
\sum_{q \leq Q} q \cdot \frac{Q}{N} \ll \frac{Q^3}{N} \ll \frac{\log B}{N}. \tag{F.26}
\]

By choosing \( B \) sufficiently large, we will be able to make all the errors from the Major Arc calculations less than the main term from the Major Arcs. Of course, we have absolutely no control over what happens on the Minor Arcs, and we will simply assume there is no contribution from the minor arcs.

Thus, on the Major Arc \( \mathcal{M}_{a,q} \), success will be in finding a function of size \( N^2 \) such that the error from this function to \( f_N(x) \) on \( \mathcal{M}_{a,q} \) is much smaller than \( N^2 \), say \( N^2 \) divided by a large power of \( \log N \).

Similarly, when we integrate over the Major Arcs, we will find the main terms will be of size \( N \); again, success will be in showing the errors in the approximations are much smaller than \( N \), say \( N \) divided by a large power of \( \log N \).

We are able to do this because of the Siegel-Walfisz Theorem (Theorem F.2.3). Given any \( B > 0 \), we can find a \( C > 0 \) such that, if \( q \leq \log^B N \), then

\[
\sum_{p \leq N \atop p \equiv r \pmod{q}} \log p = \frac{N}{\phi(q)} + O\left( \frac{N}{\log C} \right), \tag{F.27}
\]

\((r, q) = 1\). Thus, we can take \( C \) enormous, large enough so that even when we multiply by the length of the Major Arcs (of size \( \frac{\log^B N}{N} \)), we still have something small.

**F.3** \( f_N(x) \) and \( u(x) \)

**F.3.1** \( f\left(\frac{a}{q}\right) \)

We now calculate \( f_N\left(\frac{a}{q}\right) \) for \( q \leq \log^B N \).

Up to lower order terms,
\[ f_N \left( \frac{a}{q} \right) = \sum_{p_1 \leq N} \log p_1 \cdot e \left( \frac{p_1 a}{q} \right) \sum_{p_2 \leq N} \log p_2 \cdot e \left( -2 \frac{a}{p_2 q} \right) \]

\[ = \sum_{q} \sum_{r_1 = 1}^{q} \log p_1 \cdot e \left( \frac{p_1 a}{q} \right) \sum_{r_2 = 1}^{q} \sum_{p_2 \leq N \atop p_2 \equiv r_1 \mod q} \log p_2 \cdot e \left( -2 \frac{a}{p_2 q} \right) \]

\[ = \sum_{r_1 = 1}^{q} e \left( \frac{r_1 a}{q} \right) \sum_{r_2 = 1}^{q} e \left( \frac{r_2 a}{q} \right) \sum_{p_1 \leq N \atop p_1 \equiv r_1} \log p_1 \sum_{p_2 \leq N \atop p_2 \equiv r_2} \log p_2 \]

\[ = \frac{N^2}{\phi^2(q)} \sum_{r_1 = 1}^{q} \sum_{r_2 = 1}^{q} e \left( \frac{r_1 a}{q} \right) e \left( \frac{r_2 a}{q} \right) \left( \frac{1}{r_1, q} \right) \left( \frac{1}{r_2, q} \right) \]

where the second to last line follows from the Siegel-Walfisz Theorem (Theorem F.2.3). We restrict to \((r_i, q) = 1\) because if \((r_i, q) > 1\), there is at most one prime \(p_i \equiv r_i \mod q\), and one prime will give a negligible contribution as \(N \to \infty\).

**Exercise F.3.1.** *Show that the contribution from one prime may safely be absorbed by the error term.*

**F.3.2 \( u(x) \)**

Let

\[ u(x) = \sum_{m_1 \leq N} \sum_{m_2 \leq N} e \left( (m_1 - 2m_2)x \right). \]  

(F.29)

We will often look at

\[ \frac{c_q(a)c_q(-2a)}{\phi^2(q)} u(x). \]  

(F.30)

Note

\[ u(0) = N^2 \]  

(F.31)
and up to lower order terms,
\[ f_N \left( \frac{a}{q} \right) = \frac{c_q(a)c_q(-2a)}{\phi^2(q)} u(0). \] (F.32)

The function \( u(x) \) is a lot easier to analyze than \( F_N(x) \). We will show there is negligible error, for \( x \in \mathcal{M}_{a,q} \), in replacing \( f_N(x) \) with \( \frac{c_q(a)c_q(-2a)}{\phi^2(q)} u(x) \). We will then integrate over \( \mathcal{M}_{a,q} \), and then sum over all Major Arcs.

\[ F.4 \quad f_N(\alpha) - \frac{c_q(a)c_q(-2a)}{\phi^2(q)} u(\alpha - \frac{a}{q}), \alpha \in \mathcal{M}_{a,q} \]

Let
\[ C_q(a) = \frac{c_q(a)c_q(-2a)}{\phi^2(q)}. \] (F.33)

We write \( \alpha \) as \( \beta + \frac{a}{q}, \beta \in \left[ -\frac{Q}{N}, \frac{Q}{N} \right], Q = \log^* N. \) As always, we ignore lower order terms.

Note \( f_N(x) \) is approximately \( C_q(a)N^2 \) for \( x \) near \( \frac{a}{q} \). We now expand and show \( f_N(\alpha) \) is \( C_q(a)u\left(\alpha - \frac{a}{q}\right) \) plus errors of size \( \frac{N^2}{\log^{*} 2N} \) for \( \alpha \in \mathcal{M}_{a,q} \).

\[ F.4.1 \quad \text{Setup} \]

\[ S_{a,q}(\alpha) = f_N(\alpha) - C_q(a)u\left(\alpha - \frac{a}{q}\right) \]
\[ = \sum_{m_1, m_2 \leq N} \lambda(m_1)\lambda(m_2)e\left((m_1 - 2m_2)a\right) - C_q(a) \sum_{m_1, m_2 \leq N} e\left((m_1 - 2m_2)\beta\right) \]
\[ = \sum_{m_1, m_2 \leq N} \left[ \lambda(m_1)\lambda(m_2)e\left((m_1 - 2m_2)\frac{a}{q}\right) - C_q(a) \right] e\left((m_1 - 2m_2)\beta\right) \]
\[ = \sum_{m_1 \leq N} \sum_{m_2 \leq N} \left[ \lambda(m_1)\lambda(m_2)e\left((m_1 - 2m_2)\frac{a}{q}\right) - C_q(a) \right] e(-2m_2\beta) e(m_1\beta) \] (F.34)

We now apply Partial Summation multiple times. First, we apply Partial Summation to the \( m_2 \)-sum:
\[ S_{2,a,q} = \sum_{m_2 \leq N} \left[ \lambda(m_1) \lambda(m_2) e \left( (m_1 - 2m_2) \frac{a}{q} \right) - C_q(a) \right] e(-2m_2 \beta) \]
\[ = \sum_{m_2 \leq N} a_{m_2} b_{m_2} \]
\[ = A_2(N) e(-2N \beta) + 4 \pi i \beta \int_0^N \sum_{m_2 \leq u} a_{m_2} e(-u \beta) du. \tag{F.35} \]

We hit the above with \( e(m_1 \beta) \), and sum from \( m_1 = 1 \) to \( N \). We get two pieces:

\[ S_1 \sum_{a,q} = \sum_{m_1 \leq N} A_2(N) e(-2N \beta) \cdot e(m_1 \beta) \]
\[ S_1 \int_{a,q} = \sum_{m_1 \leq N} 4 \pi i \beta \int_0^N \sum_{m_2 \leq u} a_{m_2} e(-u \beta) du \cdot e(m_1 \beta) \]
\[ S_{a,q} = S_1 \sum_{a,q} + S_1 \int_{a,q}. \tag{F.36} \]

**F.4.2 \( S_1 \sum_{a,q} \)**

\[ S_1 \sum_{a,q} = \sum_{m_1 \leq N} A_2(N) e(-2N \beta) \cdot e(m_1 \beta) \]
\[ = e(-2N \beta) \sum_{m_1 \leq N} A_2(N) e(m_1 \beta) \]
\[ = e(-2N \beta) \sum_{m_1 \leq N} \sum_{m_2 \leq N} \left[ \lambda(m_1) \lambda(m_2) e \left( (m_1 - 2m_2) \frac{a}{q} \right) - C_q(a) \right] e(m_1 \beta) \]
\[ = e(-2N \beta) \left[ A_1(N) e(N \beta) \right. \]
\[ \left. - 2 \pi i \beta \int_0^N \sum_{m_1 \leq t} \sum_{m_2 \leq N} \left[ \lambda(m_1) \lambda(m_2) e \left( (m_1 - 2m_2) \frac{a}{q} \right) - C_q(a) \right] e(t \beta) dt. \right] \tag{F.37} \]
First Piece

The first piece, the $A_1(N)e(N\beta)$ term, is small for $q \leq Q$. Why? We have (up to lower order terms)

$$A_1(N)e(N\beta) = \sum_{m_1, m_2 \leq N} \lambda(m_1)\lambda(m_2)e\left(\frac{(m_1 - 2m_2)a}{q}\right) - \sum_{m_1, m_2 \leq N} C_q(a)$$

$$= C_q(a)N^2 - N^2C_q(a) = 0.$$ (F.38)

Thus, because of our choice of functions, the leading terms vanish, and the remaining term is small.

Second Piece

We now study the second piece. Note $|\beta| \leq \frac{Q}{N} = \frac{\log^2 B}{N}$, and $C_q(a) = \frac{c_q(a) c_q(-2a)}{\phi(q) \phi(q)}$.

Up to lower order terms, the $m_2$-sum will leave us with

$$\beta \frac{c_q(-2a)N}{\phi(q)} \int_0^N \sum_{m_1 \leq t} \left[ \lambda(m_1)e\left(\frac{m_1a}{q}\right) - \frac{c_q(a)}{\phi(q)} \right] e(t\beta)dt.$$ (F.39)

Note $f_N(x)$ is a multiple of $N^2$ for $x$ near $\frac{a}{q}$. Thus, we want to make sure the above is well dominated by $N^2$.

For $t \leq \sqrt{N}$, this is immediate. For $t \geq \sqrt{N}$, using Siegel-Walfisz (Theorem F.2.3), we can make the bracketed quantity in the integrant dominated by $\frac{N}{\log^{C'} N}$ for any $C'$ when $q \leq \log^B N$. Thus, we integrate a quantity that is at most $\frac{N^2}{\log^{C'} N}$ over an interval of length $N$, we multiply by $N\beta \ll Q = \log^B N$.

Thus, choosing $C'$ appropriately, the integral contributes $\frac{N^2}{\log^{C'-\eta} N}$, and hence is negligible.

Remark F.4.1. Note, of course, that the contribution is only negligible while $|\beta| \leq \frac{Q}{N}$.

Lemma F.4.2. $S_1\sum_{a,q}$ is a lower order correction.
We must evaluate

\[ S_{1f \cdot a,q} = \sum_{m_1 \leq N} 4\pi i \beta \int_0^N \sum_{m_2 \leq u} a_{m_2} e(-u\beta) du \cdot e(m_1\beta), \]  

where

\[ a_{m_2} = \left[ \lambda(m_1) \lambda(m_2) e\left( (m_1 - 2m_2) \frac{a}{q} \right) - C_q(a) \right]. \]

We bring the sum over \( m_1 \) inside the integral and again use Partial Summation. We will ignore the integration and \( \beta \) for now, as these will contribute \( \beta N \ll Q = \log^B N \) times the maximum value of the integrand. We will leave the \( e(-u\beta) du \) with this integration.

When \( u \leq \sqrt{N} \), we can immediately show the above is a lower order correction. Thus, below we always assume \( u \geq \sqrt{N} \).

**First Piece**

We have

\[ S_{1f \cdot \Sigma \cdot a,q} = \sum_{m_1 \leq N \atop m_2 \leq u} \left[ \lambda(m_1) \lambda(m_2) e\left( (m_1 - 2m_2) \frac{a}{q} \right) - C_q(a) \right] e(N\beta) \]

\[ = e(N\beta) \left[ \sum_{m_1 \leq N \atop m_2 \leq u} \lambda(m_1) \lambda(m_2) e\left( (m_1 - 2m_2) \frac{a}{q} \right) - C_q(a) \sum_{m_1 \leq N \atop m_2 \leq u} 1 \right]. \]

\[ = e(N\beta) \left[ C_q(a) u N - C_q(a) u N + \text{Lower Order Terms} \right], \]

where by the Siegel-Walfisz Theorem (Theorem F.2.3), the error in the bracketed quantity is of size \( \frac{uN}{\log^C N} \).

We then integrate from \( u = \sqrt{N} \) to \( N \) and multiply by \( \beta \), giving a contribution bounded by

\[ \beta N \cdot \frac{N^2}{\log^C N} \ll \frac{\log^B N}{N} \frac{N^3}{\log^C N} \ll \frac{N^2}{\log^{C-B} N}, \]
again getting a lower order correction to \( f_N(x) \) for \( x \) near \( \frac{a}{q} \) (remember \( f_N(x) \) is of size \( N^2 \)).

**Second Piece**

Again, \( u \geq \sqrt{N} \), and we have

\[
2\pi i \beta \int_0^N \sum_{m_1 \leq t} \left[ \sum_{m_2 \leq u} \left[ \lambda(m_1) \lambda(m_2)e\left((m_1 - 2m_2)\frac{a}{q}\right) - C_q(a) \right] \right] e(t\beta) dt. \tag{F.44}
\]

Again, for \( t \leq \sqrt{N} \), the contribution will be a lower order correction. For \( t, u \geq \sqrt{N} \),

Again, executing the sum over \( m_1 \) and \( m_2 \) will give us

\[ C_q(a)ut - C_q(a)ut + \text{Lower Order Terms}, \tag{F.45} \]

with the lower order terms of size \( \frac{ut}{\log C N} \).

Integrating over \( t \) (from \( \sqrt{N} \) to \( N \)), then integrating over \( u \) (from \( \sqrt{N} \) to \( N \)) and then multiplying by \( \beta^2 \) gives an error bounded by

\[
\beta^2 N^2 \cdot \frac{N^2}{\log C N} \ll \frac{\log^{2B} N}{N^2} \cdot \frac{N^4}{\log C N} \ll \frac{N^2}{\log^{C-2B} N}, \tag{F.46}
\]

again a lower order correction.

**F.5 Integrals of \( u(x) \)**

**F.5.1 Formulations**

Remember

\[
u(x) = \sum_{m_1, m_2 \leq N} e\left((m_1 - 2m_2)x\right). \tag{F.47}\]

We need to study \( \int_{-\frac{1}{2}}^{\frac{1}{2}} f_N(x)e(-x) dx \). We have shown that

\[
f_N(\alpha) = C_q(a)u\left(\alpha - \frac{a}{q}\right) + O\left(\frac{N^2}{\log^{C-2B} N}\right), \quad \alpha \in \mathcal{M}_{a,q}. \tag{F.48}\]
Thus, we must evaluate

\[ \int_{M_{a,q}} u(\alpha - \frac{a}{q}) \cdot e(-\alpha) d\alpha = \int_{\frac{a}{q} - \frac{Q}{N}}^{\frac{a}{q} + \frac{Q}{N}} u(\alpha - \frac{a}{q}) \cdot e(-\alpha) d\alpha = \int_{-\frac{Q}{N}}^{\frac{Q}{N}} u(\beta) \cdot e\left(-\frac{q}{q} - \beta\right) d\beta = e\left(-\frac{a}{q}\right) \int_{-\frac{Q}{N}}^{\frac{Q}{N}} u(\beta)e(-\beta) d\beta. \] (F.49)

F.5.2 \[ \int_{-\frac{1}{2}}^{\frac{1}{2}} u(x)e(-x) dx \]

\[ \int_{-\frac{1}{2}}^{\frac{1}{2}} u(x)e(-x) dx = \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{m_1, m_2 \leq N} e\left((m_1 - 2m_2)x\right) \cdot e(-x) dx = \sum_{m_1, m_2 \leq N} \int_{-\frac{1}{2}}^{\frac{1}{2}} e\left((m_1 - 2m_2 - 1)x\right) dx. \] (F.50)

If \( m_1 - 2m_2 - 1 = 0 \), the integral gives 1. There are approximately \( \frac{N}{2} \) ways to choose \( m_1, m_2 \leq N \) such that \( m_1 - 2m_2 - 1 = 0 \).

Assume now \( m_1 - 2m_2 - 1 \neq 0 \). Then the integral vanishes.

Hence,

Lemma F.5.1.

\[ \int_{-\frac{1}{2}}^{\frac{1}{2}} u(x)e(-x) dx = \frac{N}{2} + O(1). \] (F.51)

F.5.3 \[ \int_{-\frac{Q}{N}}^{\frac{1}{2}} + \int_{\frac{1}{2}}^{\frac{1}{2}} u(x)e(-x) dx \]

Define
\[ \begin{align*}
I_1 &= \left[ -\frac{1}{2}, \frac{1}{2} + \frac{Q}{N} \right] \\
I_2 &= \left[ -\frac{1}{2} + \frac{Q}{N}, -\frac{Q}{N} \right] \\
I_3 &= \left[ \frac{Q}{N}, \frac{1}{2} - \frac{Q}{N} \right] \\
I_4 &= \left[ \frac{1}{2} - \frac{Q}{N}, \frac{1}{2} \right] \\
I &= I_1 \cup I_2 \cup I_3 \cup I_4.
\end{align*} \]

(F.52)

**F.5.4 Integral over \( I_2, I_3 \)**

We have

\[
\int_{I_i} u(x)e(-x)dx = \int_{I_i} \sum_{m_1, m_2 \leq N} e\left( (m_1 - 2m_2 - 1)x \right) dx \\
= \int_{I_i} \sum_{m_1 \leq N} e(m_1 x) \sum_{m_2 \leq N} e(-2m_2 x) \cdot e(-x) dx \\
= \int_{I_i} e(x) - e((N + 1)x) e(-2x) - e(-2(N + 1)x) \frac{1 - e(-2x)}{1 - e(x)} e(-x) dx.
\]

(F.53)

On \( I_2 \) and \( I_3 \), the integral is

\[
\ll \int_{I_i} \frac{2}{x} dx \ll \frac{N}{Q} = \frac{N}{\log^B N},
\]

see, for example, Nathanson (Additive Number Theory: The Classical Bases, Chapter 8).

**F.5.5 Integral over \( I_1, I_4 \)**

Each of these intervals has length \( \frac{Q}{N} = \frac{\log B}{N} \). There are \( \frac{N}{2} + O(1) \) pairs such that \( m_1 - 2m_2 - 1 = 0 \). Each of these pairs will contribute (bound the integrand by 1) \( \frac{Q}{N} \). As there are at most \( \frac{N}{2} \) pairs, these contribute at most \( \frac{N \cdot Q}{2N} \ll \log^B N \).

Henceforth we assume \( m_1 - 2m_2 - 1 \neq 0 \). We write
\[ I_1 \cup I_4 = \left[ \frac{1}{2} - \frac{Q}{N}, \frac{1}{2} + \frac{Q}{N} \right] = I'. \quad \text{(F.55)} \]

We have
\[
\sum_{m_1, m_2 \leq N \atop m_1 - 2m_2 - 1 \neq 0} \int_{-Q}^{Q} e \left( (m_1 - 2m_2 - 1)x \right) dx \\
= e \left( -\frac{1}{2} \right) \sum_{m_1, m_2 \leq N \atop m_1 - 2m_2 - 1 \neq 0} (-1)^{m_1} \int_{-Q}^{Q} e \left( (m_1 - 2m_2 - 1)x \right) dx \\
= e \left( -\frac{1}{2} \right) \frac{1}{2\pi i} \sum_{m_1, m_2 \leq N \atop m_1 - 2m_2 - 1 \neq 0} (-1)^{m_1} \frac{2\sin \left( \frac{(m_1 - 2m_2 - 1)Q}{m_1 - 2m_2 - 1} \right)}{m_1 - 2m_2 - 1}, \quad \text{(F.56)}
\]

because, changing variables by sending \( x \) to \( (x - \frac{1}{2}) + \frac{1}{2} \) gives factors of \( e \left( (m_1 - 2m_2 - 1)\frac{1}{2} \right) = e(-\frac{1}{2})e(m_1/2)e(-m_2), \) and \( e(m_1/2) = (-1)^{m_1}. \)

\[ 0 < |m_1 - 2m_2 - 1| \leq N^{1-\epsilon} \]

Let \( w = m_1 - 2m_2 - 1. \) We will do the case \( 0 < w \leq N^{1-\epsilon}, \) the case with \( -N^{1-\epsilon} > w > 0 \) being handled similarly.

For each \( w, \) there are at most \( N \) pairs of \( m_1, m_2 \) giving rise to such a \( w. \) For such \( w, \) \[ \frac{\sin(wQ)}{w} \ll \frac{Q}{N} \]
(because we are taking the sin of a quantity very close to zero).

Thus, these pairs contribute at most
\[ \ll N \cdot \frac{Q}{N} \ll Q = \log^B N. \quad \text{(F.57)} \]

Inserting absolute values in Equation F.56 gives a contribution of at most \( \log^B N \) for such \( w, \) \( 0 < w < N^{1-\epsilon}. \)

\[ N^{1-\epsilon} < |m_1 - 2m_2 - 1| \leq N \]

Again, let \( w = m_1 - 2m_2 - 1 \) and assume \( N^{1-\epsilon} < |w| \leq N. \) We will only consider \( w > 0; \) \( w < 0 \) is handled similarly.
The cancellation is due to the presence of the factor \((-1)^{m_1}\); note that for the pair \((m_1, m_2)\) we only care about the parity of \(m_1\).

Consider \(w\) and \(w - 1\).

For \(m_1 - 2m_2 - 1 = w\), the solutions are

\[
\begin{align*}
m_1 &= w + 3, & m_2 &= 1 \\
m_1 &= w + 5, & m_2 &= 2 \\
m_1 &= w + 7, & m_2 &= 3
\end{align*}
\]

and so on; thus there are about \(\frac{N-w}{2}\) pairs, all with parity \(-(-1)^w\).

For \(m_1 - 2m_2 - 1 = w - 1\), we again have about \(\frac{N-w}{2}\) pairs, but now the parity is \((-1)^w\). Thus, each of the \(\frac{N-w}{2}\) pairs with \(m_1 - 2m_2 - 1 = w\) is matched with one of the \(\frac{N-w}{2}\) pairs with \(m_1 - 2m_2 - 1 = w - 1\), and we are off by at most \(O(1)\) pairs, which will contribute

\[
1 \leq \sum_{w=N^{1-\epsilon}}^{N} \frac{1}{w} \leq \log N.
\]  

(F.59)

For the remaining terms, we subtract in pairs, using the first order Taylor Expansion of \(\sin(x)\). We have

\[
\sum_{w=N^{1-\epsilon}}^{N} \left[ \frac{\sin\left(w \frac{Q}{N}\right)}{w} - \frac{\sin\left(w \frac{Q}{N} - \frac{Q}{N}\right)}{w - 1} \right].
\]  

(F.60)

The Main Term of the Taylor Expansion gives \(1 \leq \frac{1}{w}\), which when summed over \(w\) gives \(\frac{1}{N^{1-\epsilon}}\). As we have about \(\frac{N-w}{2} \ll N\) pairs for each \(w\), this contributes at most \(N \cdot \frac{1}{N^{1-\epsilon}} \ll N^\epsilon\).

We also have the first order term from the Taylor Expansion:

\[
\sin\left(w \frac{Q}{N} - \frac{Q}{N}\right) = \sin\left(w \frac{Q}{N}\right) + O\left(\frac{Q}{N}\right).
\]  

(F.61)

This error leads to (remembering there are \(\frac{N-w}{2} \ll N\) pairs for each \(w\))

\[
1 \leq N \sum_{w=N^{1-\epsilon}}^{N} \frac{Q}{w - 1} \ll Q \log N^\epsilon \ll \log B + 1 N.
\]  

(F.62)
F.5.6 Collecting the Pieces

We have shown

\[
\int_{[-\frac{1}{2}, \frac{1}{2}]} u(x)e(-x)dx = \frac{N}{2} + O(1)
\]
\[
\int_{[-\frac{1}{2}, \frac{1}{2}]} u(x)e(-x)dx = O\left(\frac{N}{\log B N}\right).
\]  

(F.63)

Therefore

**Lemma F.5.2.**

\[
\int_{-\frac{Q}{N}}^{\frac{Q}{N}} u(x)e(-x)dx = \frac{N}{2} + O\left(\frac{N}{\log B N}\right).
\]  

(F.64)

Remembering that we had

\[
\int_{\mathcal{M}_{a,q}} u\left(\alpha - \frac{a}{q}\right) \cdot e(-\alpha)d\alpha = \int_{\frac{a}{q}}^{\frac{a}{q} + \frac{Q}{N}} u\left(\alpha - \frac{a}{q}\right) \cdot e(-\alpha)d\alpha \\
= \int_{\frac{a}{q} - \frac{Q}{N}}^{\frac{a}{q}} u(\beta) \cdot e\left(-\frac{q}{q} - \beta\right)d\beta \\
= e\left(-\frac{a}{q}\right) \int_{-\frac{Q}{N}}^{\frac{Q}{N}} u(\beta)e(-\beta)d\beta,
\]  

(F.65)

we see that

**Lemma F.5.3.**

\[
\int_{\mathcal{M}_{a,q}} u\left(\alpha - \frac{a}{q}\right) \cdot e(-\alpha)d\alpha = e\left(-\frac{a}{q}\right)\frac{N}{2}.
\]  

(F.66)

F.6 Determination of the Main Term

We now calculate the contribution from the Major Arcs. Up to lower order terms,
\[
\int_{M} f_N(x) e(-x) dx = \sum_{q \leq Q} \sum_{a=1}^{q} \int_{\frac{a}{q} - \frac{Q}{N}}^{\frac{a+Q}{q}} f_N(\alpha) e(-\alpha) d\alpha
\]

\[
= \sum_{q \leq Q} \sum_{a=1}^{q} \int_{\frac{a}{q} - \frac{Q}{N}}^{\frac{a+Q}{q}} C_q(a) u\left(\alpha - \frac{a}{q}\right) e(-\alpha) d\alpha
\]

\[
= \sum_{q \leq Q} \sum_{a=1}^{q} e\left(-\frac{a}{q}\right) \int_{-\frac{Q}{N}}^{\frac{Q}{N}} C_q(a) u(\beta) e(-\beta)
\]

\[
= \sum_{q \leq Q} \sum_{a=1}^{q} C_q(a) e\left(-\frac{a}{q}\right) \frac{N}{2}
\]

\[
= \frac{N}{2} \sum_{q \leq Q} \sum_{a=1}^{q} \frac{c_q(a) c_q(-2a)}{\phi^2(q)} \cdot e\left(-\frac{a}{q}\right)
\]

\[
= \frac{N}{2} \sum_{q=1}^{Q} \left[ \sum_{a=1}^{q} C_q(a) e\left(-\frac{a}{q}\right) \right]
\]

\[
= \frac{N}{2} \sum_{q=1}^{Q} \rho_q
\]

\[
= \sigma_N \frac{N}{2}, \quad (F.67)
\]

where we have defined
\[
c_q(a) = \sum_{r=1}^{q} e\left(\frac{ra}{q}\right)
\]

\[
C_q(a) = \frac{c_q(a)c_q(-2a)}{\phi^2(q)}
\]

\[
\rho_q = \sum_{a=1}^{q} C_q(a)e\left(-\frac{a}{q}\right)
\]

\[
\mathcal{S}_N = \sum_{a=1}^{q} \rho_q.
\]  

(F.68)

**F.6.1 Properties of \(C_q(a)\) and \(\rho_q\)**

We will follow the presentation of Nathanson (Additive Number Theory: The Classical Bases, Chapter 8 and Appendix A).

\(c_q(a)\) is Multiplicative

We follow Nathanson, Pages 320 – 321, Theorem A.23. Note that we are labeling by \(r\) what he labels \(a\), and we are labeling by \(a\) what he labels \(n\).

**Lemma F.6.1.** \(c_q(a)\) is multiplicative; ie, if \((q, q') = 1\), then \(c_{qq'}(a) = c_q(a)c_{q'}(a)\).

Proof: We have

\[
\sum_{(\widetilde{r},qq')}^{qq'} e\left(\frac{\widetilde{r}a}{qq'}\right).
\]  

(F.69)

**Exercise F.6.2.** Show that we can write the \(\widetilde{r}\)s above as \(\widetilde{r} \equiv rq' + r'q \mod qq'\), where 1 \(\leq r \leq q\), 1 \(\leq r' \leq q'\), and \((r, q) = (r', q') = 1\).

Thus
\[ c_q(a)c_{q'}(a) = \sum_{r=1}^{q} e\left( \frac{ra}{q} \right) \sum_{r'=1}^{q'} e\left( \frac{r'a}{q'} \right) \]

\[ = \sum_{r=1}^{q} \sum_{r'=1}^{q'} e\left( \frac{rq' + r'q}{qq'} \right) \]

\[ = \sum_{r=1}^{qq'} e\left( \frac{r}{qq'} \right) = c_{qq'}(a). \quad (F.70) \]

\( c_q(a) \text{ for } (a, q) = 1 \)

**Exercise F.6.3.** Show that

\[ h_d(a) = \sum_{r=1}^{d} e\left( \frac{ra}{d} \right) = \begin{cases} d & \text{if } d | a \\ 0 & \text{otherwise} \end{cases} \quad (F.71) \]

Recall the moebius function:

\[ \mu(d) = \begin{cases} (-1)^r & \text{if } d \text{ is the product of } r \text{ distinct primes} \\ 0 & \text{otherwise} \end{cases} \quad (F.72) \]

**Exercise F.6.4.** Prove

\[ \sum_{d|\gcd(r, q)} \mu(d) = \begin{cases} 1 & \text{if } (r, q) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (F.73) \]

Then
\[ c_q(a) = \sum_{r=1}^{q} e\left(\frac{r \cdot a}{q}\right) \]
\[ = \sum_{r=1}^{q} e\left(\frac{r \cdot a}{q}\right) \sum_{d|\langle r, q \rangle} \mu(d) \]
\[ = \sum_{d|q} \mu(d) \sum_{r=1}^{q} e\left(\frac{r \cdot a}{q}\right) \]
\[ = \sum_{d|q} \mu(d) \sum_{l=1}^{\frac{q}{d}} e\left(\frac{l \cdot a}{q}\right) \]
\[ = \sum_{d|q} \mu(d) h_\frac{q}{d}(a) \]
\[ = \sum_{d|q} \mu\left(\frac{q}{d}\right) h_d(a) \]
\[ = \sum_{d|q} \mu\left(\frac{q}{d}\right) \cdot d \]
\[ = \sum_{d|(a, q)} \mu\left(\frac{q}{d}\right) \cdot d. \quad (F.74) \]

Note that if \((a, q) = 1\), then there is only one term above, namely \(d = 1\), which yields
\[ c_q(a) = \mu(q) \quad \text{if} \quad (a, q) = 1. \quad (F.75) \]

**Corollary F.6.5.** If \(q = p^k, k \geq 2\) and \((a, q) = 1\), then \(c_q(a) = 0\).

**C_q(a) is Multiplicative**

We have shown \(c_{qq'}(a) = c_q(a)c_{q'}(a)\) if \((q, q') = 1\). Recall the Euler phi-function, \(\phi(q)\), is the number of numbers less than \(q\) which are relatively prime to \(q\).

**Exercise F.6.6.** Prove that \(\phi(q)\) is multiplicative; ie, if \((q, q') = 1\), then \(\phi(qq') = \phi(q)\phi(q')\).
We now have

**Lemma F.6.7.** \( C_q(a) \) is multiplicative.

Proof: Assume \((q, q') = 1\). We have

\[
C_{qq'}(a) = \frac{c_{qq'}(a)c_{qq'}(-2a)}{\phi^2(qq')}
= \frac{c_q(a)c_{q'}(a)c_q(-2a)c_{q'}(-2a)}{\phi^2(q)\phi^2(q')}
= \frac{c_q(a)c_q(-2a)}{\phi^2(q)} \cdot \frac{c_{q'}(a)c_{q'}(-2a)}{\phi^2(q')}
= C_q(a)C_{q'}(a). \quad (F.76)
\]

\( \rho_q \) is Multiplicative

We first prove a needed lemma.

**Lemma F.6.8.** Consider \( C_{q_1}(a_1q_2) \). Then

\[
C_{q_1}(a_1q_2) = C_{q_1}(a_1) \quad (F.77)
\]

if \((q_1, q_2) = 1\).

Proof:

\[
C_{q_1}(a_1q_2) = \sum_{r_{q_1}=1}^{q_1} e\left(\frac{r_{q_1}a_1q_2}{q_1}\right)
= \sum_{r_{q_1}=1}^{q_1} e\left(\frac{r_{q_1}a_1}{q_1}\right)
= \sum_{r=1}^{q_1} e\left(\frac{r}{q_1}\right) \cdot C_{q_1}(a), \quad (F.78)
\]

because \((q_1, q_2) = 1\) implies that as \(r_{q_1}\) goes through all residue classes that are relatively prime to \(q_1\), so too does \(r = r_{q_1}q_2\). □
Lemma F.6.9. $\rho_q$ is multiplicative.

Recall

$$\rho_q = \sum_{a=1}^{q} C_q(a)e\left(-\frac{a}{q}\right). \quad (F.79)$$

Assume $(q_1, q_2) = 1$. Then we can write the congruence classes mod $q_1q_2$ as

$$a_1q_2 + a_2q_1, \text{ with } 1 \leq a_1 \leq q_1, 1 \leq a_2 \leq q_2 \text{ and } (a_1, q_1) = (a_2, q_2) = 1.$$

$$\rho_{q_1q_2} = \sum_{a=1}^{q_1q_2} C_{q_1q_2}(a)e\left(-\frac{a}{q_1q_2}\right)$$

$$= \sum_{a=1}^{q_1} \sum_{a_2=1}^{q_2} C_{q_1}(a)C_{q_2}(a)e\left(-\frac{a}{q_1q_2}\right)$$

$$= \sum_{a_1=1}^{q_1} \sum_{a_2=1}^{q_2} C_{q_1}(a_1q_2 + a_2q_1)C_{q_2}(a_1q_2 + a_2q_1)e\left(-\frac{a_1q_2 + a_2q_1}{q_1q_2}\right). \quad (F.80)$$

Exercise F.6.10. With $a_1, a_2, q_1, q_2$ as above,

$$C_{q_1}(a_1q_2 + a_2q_1) = C_{q_1}(a_1q_2) \text{ and } C_{q_2}(a_1q_2 + a_2q_1) = C_{q_2}(a_2q_1). \quad (F.81)$$

Thus, we have

$$\rho_{q_1q_2} = \sum_{a_1=1}^{q_1} \sum_{a_2=1}^{q_2} C_{q_1}(a_1q_2)e\left(-\frac{a_1}{q_1}\right) \sum_{a_2=1}^{q_2} C_{q_2}(a_2q_1)e\left(-\frac{a_2}{q_2}\right)$$

$$= \sum_{a_1=1}^{q_1} C_{q_1}(a_1)e\left(-\frac{a_1}{q_1}\right) \sum_{a_2=1}^{q_2} C_{q_2}(a_2)e\left(-\frac{a_2}{q_2}\right)$$

$$= \rho_{q_1} \cdot \rho_{q_2}. \quad (F.82)$$

Thus, $\rho_q$ is multiplicative. □
Calculation of $\rho_q$

Lemma F.6.11. $\rho_{p^k} = 0$ if $k \geq 2$ and $p$ is a prime.

Proof: This follows immediately from $C_{p^k}(a) = 0$. □

Lemma F.6.12. If $p > 2$ is prime, $\rho_p = -\frac{1}{(p-1)^2}$.

Proof:

$$\rho_p = \sum_{a=1}^{p} C_p(a)e\left(-\frac{a}{p}\right)$$

$$= \sum_{a=1}^{p-1} \frac{c_p(a)c_p(-2a)}{\phi^2(p)}e\left(-\frac{a}{p}\right). \quad \text{(F.83)}$$

But as $p > 2$, $c_p(a) = c_p(-2a) = \mu(p)$ as $(a, p) = 1$. As $\mu^2(p) = 1$ and $\phi(p) = p - 1$ we have

$$\rho_p = \sum_{a=1}^{p-1} \frac{1}{(p-1)^2}e\left(-\frac{a}{p}\right)$$

$$= \frac{1}{(p-1)^2} \left[-e\left(-\frac{0}{p}\right) + \sum_{a=0}^{p-1} e\left(-\frac{a}{p}\right)\right]$$

$$= -\frac{1}{(p-1)^2}. \quad \text{(F.84)}$$

Lemma F.6.13. If $p = 2$, then $\rho_2 = 1$.

Proof:

$$\rho_2 = \sum_{a=1}^{2} C_2(a)e\left(-\frac{a}{2}\right)$$

$$= C_2(1)e\left(-\frac{1}{2}\right)$$

$$= c_2(1)c_2(-2) \cdot e^{-\pi i}$$

$$= e^{\pi i}e^{-2\pi i} \cdot e^{-\pi i} = 1, \quad \text{(F.85)}$$
where we have used $c_2(1) = e^{\pi i}$ and $c_2(-2) = e^{-2\pi i}$.

**Exercise F.6.14.** Prove $c_2(1) = e^{\pi i}$ and $c_2(-2) = e^{-2\pi i}$.

### F.6.2 Determination of $\mathcal{S}_N$ and $\mathcal{S}$

Recall

$$\mathcal{S}_N = \sum_{q \leq Q} \rho_q. \quad \text{(F.86)}$$

We define

$$\mathcal{S} = \sum_q \rho_q. \quad \text{(F.87)}$$

**Exercise F.6.15.** Let $h_q$ be any multiplicative sequence (with whatever growth conditions are necessary to ensure the convergence of all sums below). Then

$$\sum_q h_q = \prod_{p \text{ prime}} \left(1 + \sum_{k=1}^{\infty} h_{p^k}\right). \quad \text{(F.88)}$$

$

\mathcal{S}$

We have

$$\mathcal{S} = \sum_q \rho_q$$

$$= \prod_{p \text{ prime}} \left(1 + \sum_{k=1}^{\infty} \rho_{p^k}\right)$$

$$= \prod_p \left(1 + \rho_p\right) \quad \text{(F.89)}$$

because $\rho_{p^k} = 0$ for $k \geq 2$ and $p$ prime by Lemma F.6.11. We have previously shown (see Lemmas F.6.12 and F.6.13) that $\rho_2 = 1$ and $\rho_p = -\frac{1}{(p-1)}$ for $p > 2$ prime. Therefore
\[ S = \prod_p \left( 1 + \rho_p \right) \]
\[ = (1 + \rho_2) \prod_{p > 2} (1 + \rho_p) \]
\[ = 2 \prod_{p > 2} \left[ 1 - \frac{1}{(p - 1)^2} \right] \]
\[ = 2T_2, \quad \text{(F.90)} \]

where

**Definition F.6.16 (Twin Prime Constant).**

\[ T_2 = \prod_{p > 2} \left[ 1 - \frac{1}{(p - 1)^2} \right] \approx .6601618158 \quad \text{(F.91)} \]

is the twin prime constant.

\[ \mathcal{S}_N \]

We need to estimate \(|S - \mathcal{S}_N|\). As \(\rho_q\) is multiplicative and zero if \(q = y^k \ (k \geq 2)\), we see we need only look at sums of \(\rho_p\). As \(\rho_p = -\frac{1}{(p-1)^2}\), one can show that the difference between \(S\) and \(\mathcal{S}_N\) tends to zero as \(N \to \infty\).

Thus,

**Lemma F.6.17.**

\[ S = 2T_2. \quad \text{(F.92)} \]

### F.6.3 Number of Germain Primes and Weighted Sums

Combining the above arguments, we have shown that, up to lower order terms,

\[ \sum_{p \leq N \text{ prime}} \log(p) \cdot \log\left(\frac{p - 1}{2}\right) = \mathcal{S} \frac{N}{2} \]
\[ = 2T_2 \frac{N}{2} \]
\[ = T_2 N. \quad \text{(F.93)} \]
Note that we are counting Germain prime pairs by \( \left( \frac{p-1}{2}, p \right) \) and not \( (p, 2p+1) \). Such a difference in counting will introduce a factor of 2.

We can pass from this weighted sum to a count of the number of Germain prime pairs \( \left( \frac{p-1}{2}, p \right) \) with \( p \leq N \).

Again we follow Nathanson, Chapter 8. Define

\[
\pi_G(N) = \sum_{\substack{p \leq N \\ p, \frac{p-1}{2} \text{ prime}}} 1
\]

\[
G(N) = \sum_{\substack{p \leq N \\ p, \frac{p-1}{2} \text{ prime}}} \log(p) \cdot \log \left( \frac{p-1}{2} \right). \tag{F.94}
\]

Clearly

\[
G(N) \leq \log^2 N \cdot \pi_G(N). \tag{F.95}
\]

Therefore,

**Lemma F.6.18. Up to lower order terms,**

\[
\pi_G(N) \geq \frac{G(N)}{\log^2 N} = \frac{T_2 N}{\log^2 N}. \tag{F.96}
\]

We now provide a bound in the opposite direction.

\[
\pi_G(N^{1-\delta}) = \sum_{\substack{p \leq N^{1-\delta} \\ p, \frac{p-1}{2} \text{ prime}}} 1 \ll \frac{N^{1-\delta}}{\log N}. \tag{F.97}
\]

Then
\[ G(N) \geq \sum_{p \leq N^{1-\delta}, p \text{ prime}} \log p \cdot \log \left( \frac{p-1}{2} \right) \]

\[ = (1 - \delta)^2 \log^2 N \sum_{p \leq N^{1-\delta}, \text{ prime}} 1 \]

\[ = (1 - \delta)^2 \log^2 N \left( \pi_G(N) - \pi_G(N^{1-\delta}) \right) \]

\[ \geq (1 - \delta)^2 \log^2 N \pi_G(N) + O \left( (1 - \delta)^2 \log^2 N \cdot \frac{N^{1-\delta}}{\log N} \right). \quad \text{(F.98)} \]

Therefore

\[ \log^2 N \cdot \pi_G(N) \leq (1 - \delta)^{-2} \cdot G(N) + O \left( \log^2 N \cdot \frac{N^{1-\delta}}{\log N} \right) \]

\[ 0 \leq \log^2 N \cdot \pi_G(N) - G(N) \leq \left[ (1 - \delta)^{-2} - 1 \right] G(N) + O \left( \log N \cdot N^{\delta} \right). \quad \text{(F.99)} \]

If \( 0 < \delta < \frac{1}{2} \), then \((1 - \delta)^{-2} - 1 \ll \delta\). We thus have

\[ 0 \leq \log^2 N \cdot \pi_G(N) - G(N) \ll N \left[ \delta + O \left( \frac{\log N}{N^{\delta}} \right) \right]. \quad \text{(F.100)} \]

Choose \( \delta = \frac{2 \log \log N}{\log N} \). Then we get

\[ 0 \leq \log^2 N \cdot \pi_G(N) - G(N) \leq O \left( N \frac{\log \log N}{\log N} \right). \quad \text{(F.101)} \]

Recalling \( G(N) \approx T_2 N \) gives

Lemma F.6.19.

\[ \pi_G(N) \leq \frac{T_2 N}{\log^2 N}. \quad \text{(F.102)} \]

Combining with the other bound we have finally shown
Theorem F.6.20. Assuming there is no contribution to the main term from the Minor Arcs, up to lower order terms we have

\[ \pi_G(N) = \frac{T_2 N}{\log^2 N}, \]  

where \( T_2 \) is the twin prime constant

\[ T_2 = \prod_{p > 2} \left(1 - \frac{1}{(p - 1)^2}\right) \approx 0.6601618158. \]
Bibliography


