Spectral Statistics of Random Matrices

A Thesis
Presented to
The Division of Mathematics and Natural Sciences
Reed College

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Arts

Ali Taqi

May 2021

Approved for the Division
(Mathematics)

Jonathan M. Wells

# Table of Contents

# Abstract

On their own, random variables exude deterministic properties regarding their uncertainty. The same generalization can be made for random matrices, which are matrices whose entries are random variables. One particular statistic worth investigating is the distribution of a matrix ensemble's eigenvalues, or its spectrum. In this thesis, there will be an exploration of various classes of random matrices and relevant spectral statistics like their spectra and mixing times.

# Dedication

For my mother.

# Introduction

So, what are *spectral statistics*? Do they have to do with rainbows? Sceptres? No, they don't, but they're almost as colorful and regal. The word spectral is borrowed from the spectral-like patterns observed in statistical physics - whether it may be atomic spectra or other quantum mechanical phenomena. The borrowing is loose and not literal, but still somewhat well founded.

The field of Random Matrix Theory was extensively developed in the 1930s by the nuclear physicist Eugene Wigner. He found connections between the deterministic properties of atomic nuclei and their random and stochastic behaviors. The link? Random matrices.

So in the context of this thesis, *Spectral statistics* will be an umbrella term for random matrix statistics that somehow involve that matrix's eigenvalues and eigenvectors.

To explore these spectral statistics, this thesis will use the **RMAT** package. This package was developed alongside this thesis in order to facilitate the simulation of these random matrices and spectral statistics. As such, there is a large simulation component to this thesis. To showcase the methodology of the simulations, code snippets will be sprinkled about the thesis. It will use code derived from the package RMAT which can be found on GitHub; minimal source code will also be available in the appendix, as well as pre-simulated data available on the Reed database.

# Chapter 1

# Random Matrices

As discussed in the introduction, this thesis will be an exploration of spectral statistics of random matrices. This means that we must first be able to understand what random matrices are. At a fundamental level, random matrices are simply matrices whose entries are randomly distributed in accordance to some distribution or method. To formalize all these notions, we will define what random matrices are and what it means for them to be $\mathcal{D}$-distributed.

Prior to beginning the discussion on $\mathcal{D}$-distributions, the reader should be familiar or at least accquainted with the notion of random variables and what they are. A summary is available in the appendix in A.x.

When it comes to random simulation, there must always be a rule to which our randomness must conform, regardless of complexity. For example, sampling a vector from a distribution is a rudimentary example of this. For random matrices, there will be a few methods of generating their entries that are not just sampling from theoretical distributions. As such, we motivate the $\mathcal{D}$-distribution.

## 1.1 $\mathcal{D}$-Distributions

Now, we motivate the $D$-distribution. A formalization on how to initialize a random matrix.

**Definition 1.1.1** ($\mathcal{D}$-distribution)**.** *When we define a random matrix that is $\mathcal{D}$-distributed, we say that $P \sim \mathcal{D}$. In the simplest of terms, $\mathcal{D}$ is essentially the algorithm that generates the random matrix $P$. We define two primary methods of distribution: explicit distribution and implicit distribution. If $\mathcal{D}$ is an explicit distribution, then every entry of $P$ is homogenously sampled from that distribution. Otherwise, if it is implicit, we utilize an algorithm that imposes an implicit distribution on the entries.*

### 1.1.1 Explicit Distributions

The simplest case is homogenous, explicitly distributed random matrices. If $\mathcal{D}$ is an expxlicit distribution, then we overload the notation $\mathcal{D}$ to mean a probability distribution in the classical sense (see Appendix A.x). So, if $\mathcal{D}$ is a probability distribution, the matrix $P \sim \mathcal{D}$ when $p_{ij} \sim \mathcal{D}$. In otherwords, we simply perform entry-wise sampling from that distribution.

Take for example the following explicit distributions which we cover in this thesis.

1. If $\mathcal{D} = \mathcal{N}(0,1)$, then $p_{ij} \sim \mathcal{N}(0,1)$.
2. If $\mathcal{D} = \text{Unif}(0,1)$, then $p_{ij} \sim \text{Unif}(0,1)$.

**Remark 1.1.1** (Formalization). *Explicit distributions can be formalized in scope of the standard notation in probability theory. At the heart of the formalization is the usage of index hacking to collapse the array's indices from two dimensions to one. This way, our random matrix has a representation as a random vector, which we commonly encounter in probability theory as a (i.i.d) sequence of random variables! Generally, an $N \times N$ random matrix that is explicity and homogenously $\mathcal{D}$-distributed is essentially a sequence of $N^2$ i.i.d random variables sampled from $\mathcal{D}$.*

**Example 1.1.1** (Formalization). *For example, suppose $P$ is a $2 \times 2$ random matrix with $\mathcal{D} = \mathcal{N}(0,1)$. Then, we have four random variables to initialize. In the random matrix representation, we need to initialize $P_{11}, P_{12}, P_{21},$ and $P_{22}$ by sampling them from $\mathcal{D}$. In the vector representation, we just say we are sampling four i.i.d random variables from $\mathcal{D}$. The matrix indexing is intrinsic and preservable by using index hacking with some modular math.*

### 1.1.2 Implicit Distributions

In the latter case, we are concerned less about the distribution of the matrix entries and moreso about its holisitic properties.

Consider for example, the following implicit distributions.

1. If $\mathcal{D}$ = Stochastic, then the matrix is a row of random stochastic rows. (See Algorithm B.x)
2. If $\mathcal{D}$ is any distribution (implicit or explicit), then $\mathcal{D}^\dagger$ is the Symmetric/Hermitian version of $\mathcal{D}$. (See Algorithm B.x)

### 1.1.3 Random Matrices

**Definition 1.1.2** (Random Matrix). *Assuming $\mathcal{D}$ is an explicit distribution, a random matrix is any matrix over the field $\mathbb{F}$ is a matrix $M \in \mathbb{F}^{N \times N}$ is a matrix whose entries are i.i.d random variables. So, if a random matrix $M = (m_{ij})$ is $\mathcal{D}$-distributed, then we say $m_{ij} \sim \mathcal{D}$. In the scope of this thesis, assume every random matrix to be homogenously distributed. Otherwise, if $\mathcal{D}$ is an implicit distribution, then $P$ is a matrix whose entries are determined by the algorithm imposed by $\mathcal{D}$.*

Sometimes, we want our matrix to have complex entries. We notate this by specifying $\mathbb{F} = \mathbb{C}$.

**Remark 1.1.2** (Complex Entries)**.** *To say that a random matrix is explicitly $\mathcal{D}$-distributed over $\mathbb{C}$ would mean that its entries take the form $a + bi$ where $a, b \sim \mathcal{D}$ are random variables. In other words, if we allow the matrix to have complex entries by setting $\mathbb{F} = \mathbb{C}$, then we must sample the real and imaginary component as $\mathcal{D}$-distributed i.i.d. random variables.*

Below, we can see code on how to generate a standard normal random matrix using the **RMAT** package.

**Code Example 1.1.1** (Standard Normal Matrix)**.** *Let $\mathcal{D} = \mathcal{N}(0, 1)$. We can generate $P \sim \mathcal{D}$, a $4 \times 4$ standard normal matrix, as such:*

```
# Using the RMAT package
library(RMAT)
P <- RM_norm(N = 4, mean = 0, sd = 1)

# Outputs the following
P

           [,1]       [,2]       [,3]        [,4]
[1,]  0.1058257 -1.0835598 -0.7031727  1.01608625
[2,] -0.2170453  1.8206070 -0.4539230  0.06828296
[3,]  1.3002145  0.1254992 -0.5214005 -0.61516174
[4,] -1.0398587  0.1975445 -0.8511950  0.86366082
```

**Common Matrix $\mathcal{D}$-Distributions**

| Table of Random Matrix Distributions | | | |
|---|---|---|---|
| Distribution | Notation ($\mathcal{D}$) | Parameters | Class |
| Normal | $\mathcal{N}(\mu, \sigma)$ | $\mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$ | Explicit |
| Uniform | $\text{Unif}(a, b)$ | $a, b \in \mathbb{R}$ | Explicit |
| Hermite-$\beta$ | $\mathcal{H}(\beta)$ | $\beta \in \mathbb{N}$ | Implicit |
| Erdos-$p$ | $\text{ER}(p)$ | $p \in [0, 1]$ | Implicit |

## 1.2 The Crew: Ensembles

With a random matrix well defined, we may now motivate one of the most important ideas - the random matrix ensemble. One common theme in this thesis will be that random matrices on their own provide little information. When we consider them at the ensemble level, we start to obtain more fruitful results. Without further ado, we motivate the random matrix ensemble.

**Definition 1.2.1** (Random Matrix Ensemble). *A $\mathcal{D}$-distributed random matrix ensemble $\mathcal{E}$ over $\mathbb{F}^{N\times N}$ of size $K$ is defined as a set of $\mathcal{D}$-distributed random matrices $\mathcal{E} = \{P_i \sim \mathcal{D} \mid P_i \in \mathbb{F}^{N\times N}\}_{i=1}^{K}$. In simple words, it is simply a collection of $K$ iterations of a specified class of random matrix.*

So, for example, we could compute a simple ensemble of matrices as follows.

**Code Example 1.2.1** (Standard Normal Hermitian Ensemble). *Let $\mathcal{D} = \mathcal{N}(0,1)^\dagger$. We can generate $\mathcal{E} \sim \mathcal{D}$ over $\mathbb{C}$, an ensemble of $4\times 4$ complex Hermitian standard normal matrices of size 10 as such:*

```
# Using the RMAT package
library(RMAT)
# Note that RM_norm takes mean = 0 and sd = 1 as default values.
ensemble <- RME_norm(N = 4, cplx = TRUE, herm = TRUE, size = 10)

# Outputs the following
ensemble

...
[[10]]
           [,1]       [,2]       [,3]        [,4]
[1,]  0.1058257 -1.0835598 -0.7031727  1.01608625
[2,] -0.2170453  1.8206070 -0.4539230  0.06828296
[3,]  1.3002145  0.1254992 -0.5214005 -0.61516174
[4,] -1.0398587  0.1975445 -0.8511950  0.86366082
```

With this in mind, we will gloss over, characterize, and briefly discuss a few special recurring ensembles in this thesis.

### 1.2.1 Hermite $\beta$-Ensembles

The Hermite $\beta$-Ensembles will be one of the primary ensembles discussed in this thesis. This ensemble will be characterized, motivated, and defined more thoroughly in **Chapter 4**. However, we will give a brief introduction to the ensemble.

At a practical level, all one would need to know is that the matrices are generated in accordance to an algorithm found in the appendix (Algorithm B.x) and in Chapter 4.

### 1.2.2 Erdos-Renyi $p$-Ensembles

The Hermite $\beta$-ensembles are a normal-like class of random matrices. Now, we will veer away from the normal distribution as a whole and switch to a different class of matrices: stochastic matrices. Stochastic matrices, in short, are matrices that represent Markov Chains (see A.x). We can also think of them as the matrix representation of a specific setup of a walk on a random graph.

A particular class of random graphs that we will consider are the Erdos-Renyi random graphs. Essentially, these are graphs whose vertices are connected with a uniform probability $p$. We can interpret this as saying an Erdos-Renyi graph is a simple random walk on a graph with parameterized sparsity (given by $p$). Without further ado, we motivate the Erdos-Renyi graph:

**Definition 1.2.2** (Erdos-Renyi Graph)**.** *An Erdos-Renyi graph is a graph $G = (V, E)$ with a set of vertices $V = \{1, \dots, N\}$ and edges $E = \mathbb{1}_{i,j \in V} \sim Bern(p_{ij})$. It is homogenous if $p_{ij} = p$ is fixed for all $i, j$.*

Essentially, an Erdos-Renyi graph is a graph whose 'connectedness' is parameterized by a probability $p$ (assuming it's homogenous, which this document will unless otherwise noted). As $p \to 0$, we say that graph becomes more sparse; analogously, as $p \to 1$ the graph becomes more connected.

Recall from probability theory that a sum of i.i.d Bernoulli random variables is a Binomial variable. As such, we may alternatively say that the degree of each vertex $v$ is distributed as $deg(v) \sim Bin(N, p)$ where $N$ is the number of vertices. This makes simulating the graphs much easier.

**Code Example 1.2.2** (Erdos-Renyi p = 0.5 Ensemble)**.** *Let $\mathcal{D} = Erdos(p = 0.5)$. We can generate $\mathcal{E} \sim \mathcal{D}$, an ensemble of $4 \times 4$ Erdos-Renyi matrices ($p = 0.5$) of size 10 as such:*

```
# Using the RMAT package
library(RMAT)
ensemble <- RME_erdos(N = 4, p = 0.5, size = 10)

# Outputs the following
ensemble


...
[[10]]
          [,1]       [,2]      [,3]      [,4]
[1,] 0.0000000 0.37154735 0.6284527 0.0000000
[2,] 0.0000000 0.08954572 0.3362090 0.5742453
[3,] 0.6058502 0.00000000 0.3941498 0.0000000
[4,] 0.0000000 0.64146116 0.0000000 0.3585388
```

## 1.3 Analytical Results

# Chapter 2

# Spectra

## 2.1 Introduction

So, what are *spectral statistics*? Do they have to do with rainbows? Sceptres? No, they don't, but they're almost as colorful and regal. The word spectral is borrowed from the spectral-like patterns observed in statistical physics - whether it may be atomic spectra or other quantum mechanical phenomena. The borrowing is loose and not literal, but still somewhat well founded. In fact, the field of Random Matrix Theory was extensively developed in the 1930s by the nuclear physicist Eugene Wigner. He found connections between the deterministic properties of atomic nuclei and their random and stochastic behaviors. The link? Random matrices.

So in the context of this thesis, *spectral statistics* will be an umbrella term for random matrix statistics that somehow involve that matrix's eigenvalues and eigenvectors. That being said, if we fix a *random matrix*, we can study its features by studying its eigenvalues - fundemental numbers that tell us a lot about the matrix. They are quite important for many reasons. For instance in statistical physics, many processes are represented by operators or matrices, and as such, their behaviours could be partially determined by the eigenvalues of their corrosponding matrices. The study of eigenvalues and eigenvectors primarily falls in the scope of Linear Algebra, but their utility is far-reaching. So, what exactly are *eigenvalues* exactly?

### 2.1.1 The Quintessential Spectral Statistic: the Eigenvalue

Given any standard square matrix $P \in \mathbb{F}^{N \times N}$, its *eigenvalues* are simply the roots of the characteristic polynomial $\mathrm{char}_P(\lambda) = \det(P - \lambda I)$. By the Fundamental Theorem of Algebra, we know that there is always have as many complex eigenvalues $\lambda \in \mathbb{C}$ as the dimension of the matrix.

That being said, when our random matrix has a specified distribution (say, standard normal), we can see patterns in the eigenvalue distributions. So, an eigenvalue is a **spectral statistic** of a random matrix! To talk about a matrix's eigenvalues in a more formal and concise manner, we motivate what is the eigenvalue spectrum.

**Definition 2.1.1** (Spectrum). *Suppose $P \in \mathbb{F}^{N\times N}$ is a square matrix of size $N$ over $\mathbb{F}$. Then, the (eigenvalue) spectrum of $P$ is defined as the multiset of its eigenvalues and it is denoted $\sigma(P) = \{\lambda_i \in \mathbb{C}\}_{i=1}^{N}$. Note that it is important to specify that a spectrum is a multiset and not just a set; eigenvalues could be repeated due to algebraic multiplicity and we opt to always have $N$ eigenvalues.*

While our definition for spectrum is nice and clean, there are a few caveats that we must take care of. First, how are spectra statistics? So we can even call them statistics, so there needs to be some formalization as to why we consider eigenvalues statistics. In this dialogue, we will call back on the same sort of formalization dialogue in the Random Matrices section. Before beginning, the reader is encouraged to review what a statistic is in the review appendix (A.x).

Recall that when we defined and motivated the $\mathcal{D}$-distribution framework of simulating random matrices, we always had one thing - a vector representation of random variables. Using this framework, the formalization is trivial.

**Remark 2.1.1** (Formalization). *Suppose $P \sim \mathcal{D}$ is an $N \times N$ random matrix. Then, $P$ has a representation as a sequence of $N^2$ random variables, denote it $\vec{X} = \{X_i \mid i = 1, 2, \ldots, N^2 - 1, N^2\}$. Then, the spectrum of the matrix $P$ is simply a function of the vector $\vec{X}$. We can denote this $\sigma(\vec{X})$, where the operator $\sigma$ is overloaded to mean the spectrum of a matrix* ***with respect to the vector representation****. The actual process for $\sigma$ is not necessary to explicitly write out, since characterizing it will be sufficient for now. Essentially, $\sigma$ is a function that would parse the random variables into the array form by index hacking. Then, it must compute the determinant of the matrix $P - \lambda I$, and solve for its roots. To summarize this, consider the flow chart below.*

To simplify the process, here is how we formalize the spectrum as a statistic. Suppose we sample $P \sim \mathcal{D}$. Then, we take its spectrum formally using $\sigma$ as such:

$$P_{[]} \to \vec{P} \to \sigma(\vec{P}) \to \text{index magic} \to \det(P - \lambda I) \to \sigma(P_{[]})$$

### 2.1.2 Interlude: Ensembles

While the spectrum of a matrix provides a good summary of the matrix, a matrix is only considered a single point/observation in random matrix theory. Additionally, simulating large matrices and computing their eigenvalues becomes harder and more computationally expensive as $N \to \infty$. As such, to obtain more eigenvalue statistics efficiently, another dimension is introduced by motiving the *random matrix ensemble*.

[Definition was here]

Now that matrix ensembles are well defined, we can motivate a core object of our study - the spectrum of a random matrix ensemble. From its name, it is indeed what one might expect it to be.

**Definition 2.1.2** (Ensemble Spectrum). *If we have an ensemble $\mathcal{E}$, then we can naturally extend the definition of $\sigma(\mathcal{E})$. To take the spectrum of an ensemble, simply take the union of the spectra of each of its matrices. In other words, if $\mathcal{E} = \{P_i \sim \mathcal{D} \mid P_i \in \mathbb{F}^{N\times N}\}_{i=1}^{K}$, then $\sigma(\mathcal{E}) = \bigcup_{i=1}^{K} \sigma(P_i)$.*

[Plot showing the difference between matrix ensemble and matrix spectrum]

A common theme in this thesis will be that singleton matrices do not provide insightful information on their own. Rather, it is the collective behavior of a $\mathcal{D}$-distributed ensemble that tells us about how $\mathcal{D}$ impacts our spectral statistics. So in a way, ensemble statistics are the engine of this research.

## 2.2 Ordered Spectra

### 2.2.1 Eigenvalue Ordering

When we motivate the idea of matrix dispersion in the next section, we will consider order statistics of that matrix's eigenvalues in tandem with its dispersion. However, to do so presupposes that we have a sense of what *ordered* eigenvalues means. Take a matrix $P$ and its *unordered* spectrum $\sigma(P) = \{\lambda_j\}$. It is paramount to know what ordering scheme $\sigma(P)$ is using, because otherwise, the eigenvalue indices are meaningless! So, to eliminate confusion, we add an index to $\sigma$ that indicates how the spectrum is ordered. Often, the ordering context will be clear and the indexing will be omitted. Consider the two following *ordering schema*:

Standard definitions of an ordered spectrum follows the standard ordering in the reals; denote this as the ordering by the **sign scheme**. Note that because total-ordering is only well-defined on the reals, we can only use this scheme when on a spectrum with real entries. So, we write the *sign-ordered spectrum* as follows:

$$\sigma_S(P) = \{\lambda_j : \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N\}_{j=1}^{N}$$

Alternatively, we can motivate a different scheme that properly handles complex eigenvalues. We could sort the spectrum by the norm of its entries; denote this as ordering using the *norm scheme*. This way, all the eigenvalues are mapped to a real value, in which we could use the sign-scheme of ordering. Without further ado, we write the *norm-ordered spectrum* as follows:

$$\sigma_N(P) = \{\lambda_j : |\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_N|\}_{j=1}^{N}$$

Note that when we take the norms of the eigenvalues, we essentially ignore ”rotational” features of the eigenvalues. Signs of eigenvalues indicate reflection or rotation, so when we take the norm, we essentially become more concerned with scaling.

### 2.2.2 Singular Values

An aternative to using the norm ordering scheme is using the singular values of the matrix. If a matrix is symmetric, the singular values are simply the norm of the eigenvalues. We ignore rotational features and focus solely on scale when we do so.

Suppose $P$ is a random matrix. Then, we can take it singular values as such.

**Definition 2.2.1** (Singular Values)**.** *The singular values of a matrix $P$ are given by the square root of the eigenvalues of the corrosponding product of that matrix and its transpose. That is, $\sigma_+(P) = \sqrt{\sigma(P \cdot P^T)}$.*

**Common Spectrum Schema**

| Table of Spectrum Schema | | | |
|---|---|---|---|
| Scheme | Matrix | Notation | Ordering |
| Sign-Ordered | $P$ | $\sigma_{\pm}(P)$ | $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N$ |
| Norm-Ordered | $P$ | $\sigma_{\lvert\cdot\rvert}(P)$ | $\lvert\lambda_1\rvert \geq \lvert\lambda_2\rvert \geq ... \geq \lvert\lambda_N\rvert$ |
| Singular | $P \cdot P^T$ | $\sigma_{\surd}(P)$ | $\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq ... \geq \sqrt{\lambda_N}$ |

### 2.2.3 Order Statistics

With eigenvalue ordering unambiguous and well-defined, we may proceed to start talking about their order statistics. In short, given a random sample of fixed size, order statistics are random variables defined as the value of an element conditioning on its rank within the sample. (See A.x)

In general, order statistics are quite useful and tell us a lot about how the eigenvalues distribute given a distribution. They tell us how the eigenvalues space themselves and give us useful upper and lower bounds.

For example, the maximum of a sample is an order statistic concerned with the highest ranked element. In our case, this could corrospond the largest eigenvalue of a spectrum. After all, a spectrum is a random sample of fixed size, so this statistic is well-defined.

**Example 2.2.1** (The Largest Eigenvalue)**.** *Suppose we have seek the largest eigenvalue distribution for a ensemble distribution $\mathcal{D}$, we would simulate an ensemble $\mathcal{E}$ and observe $\lambda_1$ for each of its matrices. Then, we can set the distribution of the largest eigenvalue for $\mathcal{D}$ by observing the distribution of $\lambda_1$.*

So, we fill consider the conditional order statistics $\mathbb{E}(\lambda_i \mid i)$ and $\mathrm{Var}(\lambda_i \mid i)$.

### 2.2.4 Case Study: Tracy-Widom Distribution

Distribution of the normalized largest eigenvalue. Let $\lambda_1$ denote the largest eigenvalue. Then, let $< \lambda_j - \lambda_i >$ be the mean eigenvalue spacing (of consecutive eigenvalues). The Tracy-Widom distribution is the distribution of $\frac{(\lambda_1 - 2\sqrt{N})n^{1/6}}{<\lambda_j - \lambda_i>}$.

[Plot]

## 2.3 Symmetric and Hermitian Matrices

### 2.3.1 Introduction

A very important class of matrices in Linear Algebra is that of Symmetric or Hermitian matrices (See A.x). Simply put, those are matrices which are equal to their conjugate transpose.

**Note:** Since real numbers are their own conjugate transpose, every Symmetric matrix is Hermitian. However, we will still delineate the two terms to avoid confusion.

In any case, one critical result in Linear Algebra that will be extensively wielded in this thesis is the fact that a matrix is Symmetric or Hermitian if and only if it has real eigenvalues. In other words:

$$P = \overline{P^T} \iff \sigma(P) = \{\lambda_i \mid \lambda_i \in \mathbb{R}\}$$

Having a complete set of real eigenvalues yields many great properties. For instance, if all eigenvalues are real, we have the option of observing either the sign-ordered spectrum or the norm-ordered spectrum. This way, we can preserve negative signs and we would not lose the rotational aspect of the eigenvalue when we study its statistics. That is just one reason out of many more why having real eigenvalues is quite nice.

### 2.3.2 Case Study: Wigner's Semicircle Distribution

The eigenvalues of Hermitian matrices obey Wigner's Semicircle distribution. Since Hermitian matrices have real eigenvalues, then we can be more precise and generally say that the real component of the eigenvalues follow the semicircle distribution.

**Definition 2.3.1** (Semicircle Distribtion). *If a random variable $X$ is semicircle distributed with radius $R \in \mathbb{R}^+$, then we say $X \sim SC(R)$. $X$ has the following probability density function:*

$$P(X = x) = \frac{2}{\pi R^2}\sqrt{R^2 - x^2} \text{ for } x \in -R, R$$

**Remark 2.3.1** (Radius and Matrix Dimension). *The dimension of the matrix determines the radius of the eigenvalues. Namely, if a Hermitian matrix $P$ is $N \times N$, then its eigenvalues are semicircle distributed with radius $R = 2\sqrt{N}$. That is, $P^\dagger$ has a spectrum $P \sim SC(2\sqrt{R})$.*

## 2.4 Findings

# Chapter 3

# Dispersions

## 3.1 Introduction

In this section, we define the final spectral statistic studied in this chapter: eigenvalue dispersions. As the name suggests, these statistics are concerned with the distirbution of the spacings between the eigenvalues. Interestingly, this is almost as literal as it gets when we use the word "spectral". In physics and chemistry, atomic spectra are essentially differences between energy levels or quanta, so the translation is close.

In any case, we will begin this chapter by first motivating a few definitions and formalisms in this section. Then, once our setup is ready, we will motivate the definition of a matrix's eigenvalue dispersion and formalize it as a statistic. To outline the section, we will first define two things: the dispersion metric and the pairing scheme. In simple terms, we formalize **what** "eigenvalue spacings" are and **which** eigenvalue pairs' spacings to consider. Without further ado, we motivate the dispersion metric.

### 3.1.1 Dispersion Metrics

Before we may even start to consider studying dispersions of eigenvalues, we must first formalize and make clear what "metric" of spacing we are using. To do so, we motivate the dispersion metric.

**Definition 3.1.1** (Dispersion Metric)**.** *A dispersion metric* $\delta : \mathbb{C} \times \mathbb{C} \to \mathbb{R}^+$ *is defined as a function from the space of pairs of complex numbers to the positive reals. In simple terms, it is a way of measuring "space" between two complex numbers - our eigenvalues.*

Consider the following dispersion metrics below.

1. The standard norm: $\delta(z, z') = |z' - z|$
2. The $\beta$-norm: $\delta_\beta(z, z') = |z' - z|^\beta$
3. The difference of absolutes: $\delta_{\text{abs}}(z, z') = |z'| - |z|$

**Remark 3.1.1** (Symmetric Metrics)**.** *Note that the standard norm is a symmetric operation. The $\beta$-norm can also be a symmetric operation when $\beta$ is even. Otherwise, the difference of absolutes metric is not symmetric.*

While we have defined disperison metrics to be functions from $\mathbb{C}^2$, there is one special case where we can make an exception so that the domain of $\delta$ is not $\mathbb{R}^+$.

**Remark 3.1.2** (Identity Difference Heuristic)**.** *Suppose we take the arithmetic difference of two complex numbers. Then, the range of $\delta$ is $\mathbb{C}$. For this reason, we won't consider the arithmetic difference a formal dispersion metric, but we will honor it as a dispersion huerestic. As such, we will denote this as the "identity difference" huerestic and call it $\delta_{id}$. So, we define $\delta_{id} : (z, z') \mapsto z' - z$*

### Common Dispersion Metrics

For the formula, assume the order of arguments is $\delta(z, z')$.

| Table of Dispersion Metrics | | | | |
|---|---|---|---|---|
| Metric* | Notation | Formula | Symmetric | Parameters |
| Standard Norm | $\delta$ | $\|z' - z\|$ | True | - |
| $\beta$-Norm | $\delta_\beta$ | $\|z' - z\|^\beta$ | $\beta$ Even | $\beta \in \mathbb{N}$ |
| Difference of Absolutes | $\delta_{\text{abs}}$ | $\|z'\| - \|z\|$ | False | - |
| Identity Difference | $\delta_{\text{id}}$ | $z' - z$ | False | - |

*Note that the identity difference is a heurestic, and not a formal metric.

### 3.1.2 Pairing Schema

The next thing we need to motivate before talking about eigenvalue dispersions are pairing schema. In simple terms, pairing schema are templates (for some $N \in \mathbb{N}$) for which eigenvalue pairs to pick. There are many subtle reasons why this is important, which will be covered in detail later. Without further ado, we motivate the pairing schema.

**Definition 3.1.2** (Pairing Scheme)**.** *Suppose $P$ is any $N \times N$ matrix and $\sigma(P)$ is its corrosponding ordered spectrum. A pairing scheme for the matrix $P$ is simply a subset of $\mathbb{N}_N \times \mathbb{N}_N$ - a subset of pairs of numbers from $\mathbb{N}_N = \{1, \ldots, N\}$. In other words, it is a subset of pair indices for $N$ objects - in our case, eigenvalues. We denote a pairing scheme as some set $\Pi = \{(\alpha, \beta) \mid \alpha, \beta \in \mathbb{N}_N\} \subseteq \mathbb{N}_N \times \mathbb{N}_N$. To take the matrix's corrosponding eigenvalue pairs with respect to $\Pi$, we simply take the set of eigenvalue pairs with the corrosponding indices, $\tilde{\sigma}(P \mid \Pi) = \{(\lambda_\alpha, \lambda_\beta) \mid (\alpha, \beta) \in \Pi\}$.*

The reader might find this definition slightly obscure, and rightfully so. The definition is a mere formality, as we will usually only consider a few specific pairing schema. Seeing the explicit examples will hopefully make things more clear. With all the technical details aside, we can just say that a pairing scheme tells us which subset of eigenvalue pairs to consider. If one visualizes an array with $N$ objects on two axes, we are simply choosing a subset of that plane. In fact, consider the following.

**Remark 3.1.3** (Spectrum Subset)**.** *If we denote $\mathbb{S} = \sigma(P)$ as the **ordered** spectrum of some matrix $P$, then for any valid pairing scheme $\Pi$, we will find that $\tilde{\sigma}(P \mid \Pi) \subseteq \mathbb{S}^2$ by definition. Note the emphasis on ordered - otherwise, the indices are meaningless.*

### Common Pairing Schema

Suppose $P$ in an $N \times N$ square matrix, and $\sigma(P)$ is its ordered spectrum.

1. The unique pair combinations schema are two complementary pair schema. By specifying **either** $i > j$ or $i < j$, we characterize this scheme to entail all unique pair combinations of eigenvalue without repeats. The reason we call them upper and lower pair combinations is an allusion to the indices of the upper and lower triangular matrices.

   (a) Let $\Pi_>$ be the lower-pair combinations of ordered eigenvalues. This will be the standard unique pair combination scheme used in lieu of the argument orders of our dispersion metrics (more later). In this pairing scheme, the eigenvalue with the lower rank is always listed first, and the higher rank second.
   $$\sigma(P \mid \Pi_>) = \{\pi_{ij} = (\lambda_i, \lambda_j) \mid i > j\}_{i=1}^{N-1}$$

   (b) For completeness, we will also define the upper-pair scheme. $\Pi_<$ is the set of upper-pair unique combinations of ordered eigenvalues. Wont be used because we want bigger - smaller to make positive definite.
   $$\sigma(P \mid \Pi_<) = \{\pi_{ij} = (\lambda_i, \lambda_j) \mid i < j\}_{i=1}^{N-1}$$

   **Benefits:** Solves the issue of repeated pairs for symmetric dispersion metrics.

2. Let $\Pi_C$ be the consecutive pairs of eigenvalues in a spectrum.
   $$\sigma(P \mid \Pi_C) = \{\pi_j = (\lambda_{j+1}, \lambda_j)\}_{j=1}^{N-1}$$

   **Benefits:** This pairing scheme gives us the minimal information needed to express important bounds and spacings in terms of its elements.

3. Let $\Pi_0$ be all the pairs in a spectrum. For completeness, we define this pairing scheme as a standard.
   $$\sigma(P \mid \Pi_0) = \{\pi_{ij} = (\lambda_i, \lambda_j) \mid i, j \in \mathbb{N}_N\}$$

**Remark 3.1.4** (Order Statistics)**.** *Since the pairing scheme assumes the eigenvalues are given in an ordered spectrum, our analysis will be mostly leveraging this fact by using the indices. However, as notated above, some pairing schemes are indexed with $i$ and $j$, and some like the consecutive pairs only with $j$. These indices are interchangable with "order statistic", so in the next section, we will devise a synthetic order statistic that uses two indices.*

So, if a pair $\pi_j$ has only one index $j$, the pair will be a consecutive pair denoted by $\pi_j = (\lambda_{j+1}, \lambda_j)$. Otherwise, $\pi_{ij}$ will be given by $(\lambda_i, \lambda_j)$. The ranking difference class will group pairs into equivalence classes given by an integer $\rho = i - j$.

### Special pairs

There are a few pairs that are so special, it may be worth highlighting them explicity.

1. Let $\pi_1$ be the pair of the two largest eigenvalues in an ordered spectrum.

$$\pi_1 = \{(\lambda_2, \lambda_1)\}$$

Another way we can define the pairing scheme is defining an auxilliary object: the eigenvalue (pair) matrix.

**Definition 3.1.3** (Eigenvalue Matrix)**.** *Suppose $P$ is an $N \times N$ square matrix. Then, the eigenvalue matrix of $P$, given by $\Lambda(P)$ is the matrix with entries $\pi_{ij} = (\lambda_i, \lambda_j)$. It is assumed that the eigenvalues, $\lambda_i$ come from some* ***ordered*** *spectrum $\sigma(P)$.*

### Common Pairing Schema

Suppose $P$ is an $N \times N$ matrix.

| Table of Pairing Schema | | |
|---|---|---|
| Scheme | Notation | Formula |
| Lower | $\Pi_<$ | $\{(i,j) \mid i < j \text{ for } i,j \in \mathbb{N}_N\}$ |
| Upper | $\Pi_>$ | $\{(i,j) \mid i > j \text{ for } i,j \in \mathbb{N}_N\}$ |
| Consecutive | $\Pi_C$ | $\{(i,j) \mid i = j+1 \text{ for } i,j \in \mathbb{N}_N\}$ |
| All | $\Pi_0$ | $\{(i,j) \mid i,j \in \mathbb{N}_N\}$ |

**Pairing Scheme Diagrams for a $5 \times 5$ Matrix**

All Pairs

$$\begin{bmatrix} O & O & O & O & O \\ O & O & O & O & O \\ O & O & O & O & O \\ O & O & O & O & O \\ O & O & O & O & O \end{bmatrix}$$

The Lower Pair Combinations

$$\begin{bmatrix} - & - & - & - & - \\ O & - & - & - & - \\ O & O & - & - & - \\ O & O & O & - & - \\ O & O & O & O & - \end{bmatrix}$$

The Upper Pair Combinations

$$\begin{bmatrix} - & O & O & O & O \\ - & - & O & O & O \\ - & - & - & O & O \\ - & - & - & - & O \\ - & - & - & - & - \end{bmatrix}$$

The Consecutive Pairs

$$\begin{bmatrix} - & - & - & - & - \\ O & - & - & - & - \\ - & O & - & - & - \\ - & - & O & - & - \\ - & - & - & O & - \end{bmatrix}$$

### 3.1.3 Dispersions

Finally, we are able to motivate the definition of a matrix dispersion! Suppose we have a $\mathcal{D}$-distributed random matrix $P \in \mathbb{F}^{N\times N}$ or a random matrix ensemble $\mathcal{E} = \{P_i \sim \mathcal{D} \mid P_i \in \mathbb{F}^{N\times N}\}$. Then we define their dispersion as follows.

**Definition 3.1.4** (Dispersion)**.** *The dispersion of a matrix $P \in \mathbb{F}^{N\times N}$ with respect to a dispersion metric $d : \mathbb{C}\times\mathbb{C} \to \mathbb{F}$ and pairing scheme $\Pi$, call it $\Delta_d(P, \Pi)$, is defined as follows. Suppose $\sigma(P) := \mathbb{S}$ is the ordered spectrum of $P$ where $\sigma(P) = \{\lambda_1, \ldots, \lambda_N\}$. Then, let $\Pi = \{\pi_{ij} = (\lambda_i, \lambda_j)\} \subseteq \mathbb{S}^2$ be a subset of eigenvalue ordered pairs. Then, the dispersion of $P$ with respect to $d$ is simply the set $\Delta_d(P, \Pi) = \{\delta_{ij} = d(\pi_{ij}) \mid \pi_{ij} = (\lambda_i, \lambda_j) \in \Pi\}$.*

As we usually do, we define the dispersion of an ensemble in a similar fashion.

**Definition 3.1.5** (Ensemble Dispersion)**.** *If we have an ensemble $\mathcal{E}$, then we can naturally extend the definition of $\Delta_d(\mathcal{E}, \Pi)$. To take the dispersion of an ensemble, simply take the union of the dispersions of each of its matrices. In other words, if $\mathcal{E} = \{P_i \sim \mathcal{D} \mid P_i \in \mathbb{F}^{N\times N}\}_{i=1}^{K}$, then $\Delta_d(\mathcal{E}, \Pi) = \bigcup_{i=1}^{K} \Delta_d(P_i, \Pi)$.*

With our spectral statistics defined, we are prepared to discuss prominent results in Random Matrix Theory alongside our new findings from the simulations.

## 3.2 Order Statistics

### 3.2.1 Introduction

With eigenvalue dispersions and eigenvalue orderings well-defined, we may proceed to start talking about their order statistics.

In addition to these simple order statistics, we introduce a new variant statistic called the **ranking difference class**. Instead of observing a single eigenvalue at a given rank, we will now observe a pair of eigenvalues at a time. To standardize the process, we introduce a new eqivalence class called the *ranking difference*. As suggested, it is precisely the integer difference of the eigenvalue ranks.

**Definition 3.2.1** (Ranking Difference)**.** *The ranking difference is a function $\delta : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ which takes the index of two eigenvalues (from an ordered spectrum) and returns their difference. In other words, $\delta : (\lambda_i, \lambda_j) \mapsto (i - j)$.*

With the function $\delta$, we may take the set of unique eigenvalue pairs $(i > j)$ and partition it into equivalence classes. To do so, we define the equivalence relation $\sim_\delta$ which says $(\lambda_a, \lambda_b) \sim_\delta (\lambda_c, \lambda_d) \iff (a - b) = (c - d)$. These equivalence classes then naturally corrospond to pairs a set distance $\rho = i - j$ apart. So, for an $N \times N$ matrix, $\delta$ assumes a range $\rho \in \{1, \ldots, N-1\}$.

In summary, $\sim_\delta$ takes the set $\{(\lambda_i, \lambda_j) \mid \lambda_i, \lambda_j \in \sigma(P) \text{ and } i > j\}$ and surjectively partitions it onto the equivalence classes $[(\lambda_i, \lambda_j)]_\rho$ for $\rho \in \{1, \ldots, N-1\}$. For example, if we consider $\rho = 1$, then we are considering all pairs of eigenvalue neighbors.

Note that the sizes of each equivalence class are **never equal**. With this partition in mind, we can consider various statistics conditioning on the value of $\rho$. Conditioning on $\rho$ will be especially useful in the cases where we are considering matrices like the Hermite-$\beta$ matrices; the eigenvalues of those matrices tend to *repel*, so to speak, and we can observe these patterns using $\rho$.

### 3.2.2 Conditional Statistics

We will considering the conditional statistics $\mathbb{E}(\delta_{ij} \mid \rho)$ and $\mathrm{Var}(\delta_{ij} \mid \rho)$.

## 3.3 Analytical Results

### 3.3.1 Case Study: Wigmer's Surmise

Limiting distribution of the eigenvalue spacings of symmetric matrices.

[Plot]

## 3.4 Findings

# Chapter 4

# $\beta$-Ensembles

## 4.1 Introduction

In this chapter, we will talk about the Hermite $\beta$-ensembles more in depth. The beta ensembles have wide applications in statistical physics, eningeering, and many other places. They are defined by the joint density of their eigenvalues, and have a special characterization discussed in the next subsection.

### 4.1.1 Hermite $\beta$-Ensembles

The Hermite $\beta$-ensembles, also called the Gaussian ensembles, are an important class of random matrix ensembles studied in engineering, statistical physics, and probability theory. Parameterized by $\beta \in \mathbb{N}$ through the Dyson index, this ensemble is charecterized by a few things.

- The Dyson index $\beta$ corrosponds to the number of real number of components the subject matrices have.
- The subject matrices are classically defined for $\beta = 1, 2, 4$ and they corrospond to matrices with real, complex, and quaternionic entries. The corrosponding fields are $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{H}$.
- The matrices in this ensemble, most importantly, have a feature called conjugation invariance. With respect to the conjugation by the respective group of matrices.
- Most importantly, the eigenvalues are determined by the joint probability density function given below.

**Definition 4.1.1** (Hermite $\beta$-ensembles)**.** *The Hermite $\beta$-ensembles, commonly known as the Gaussian ensembles, are an ensemble of random matrices parameterized by $\beta$, and their eigenvalues have the joint probability density function:*

$$f_\beta(\Lambda) = c_H^\beta \prod_{i<j} |\lambda_i - \lambda_j|^\beta e^{-1/2 \sum_i \lambda_i^2}$$

*where the normalization constant $c_H^\beta$ is given by:*

$$c_H^\beta = (2\pi)^{-n/2} \prod_{j=1}^{n} \frac{\Gamma(1+\beta/2)}{\Gamma(1+\beta j/2)}$$

To simulate matrices from the $\beta$-ensemble, we will be using a recent result published in "Matrix Models for Beta Ensembles" Dumitriu (2018). This makes the $\beta$-ensemble a canonical example of an implicity distributed matrix; we do not care about the actual distribution of the entries, but rather the effect they have on the eigenvalues (trace) of the matrices. The algorithm used is directly cited from the results of Dumitriu's paper, and can be found in Algorithms B.1.3.

**Code Example 4.1.1** (Hermite Beta = 2 Ensemble)**.** *Let $\mathcal{D} = \mathcal{H}(\beta = 2)$. We can generate $\mathcal{E} \sim \mathcal{D}$, an ensemble of $4 \times 4$ Hermite matrices ($\beta = 2$) of size 10 as such:*

```
# Using the RMAT package
library(RMAT)
ensemble <- RME_beta(N = 4, beta = 2, size = 10)

# Outputs the following
ensemble

...
[[10]]
          [,1]     [,2]      [,3]       [,4]
[1,] 0.3812855 2.592124 0.0000000  0.0000000
[2,] 2.5921244 1.362211 1.4197438  0.0000000
[3,] 0.0000000 1.419744 0.8220259  0.3917667
[4,] 0.0000000 0.000000 0.3917667 -0.9740052
```

### 4.1.2 Dimitriu's Matrix Model

To generate Hermite $\beta$ matrices, we consider the result of Dimitriu's paper. We obtain the following algorithm.

**Algorithm 4.1.1** (Dimitriu's Beta Matrix)**.**

1. *Fix $N \in \mathbb{N}$.*

2. *Start by taking a diagonal of $\mathcal{N}(0,2)$ variables.*

3. *Set both of the nearest off-diagonals to the row that samples from a $\chi(df = df_i)$ where $df_i = \beta * i$ for columns going from $i = 1$ to $i = n - 1$.*

## 4.2 Spectra

We know that $\beta$-matrices must be symmetric. So, their eigenvalues must be real. Any imaginary component we observe is simply computational error, and we may safely ignore it. It is good to see that the error is uniform and small.

**Remark 4.2.1** (Implicit Distribution)**.** *Note that the beta ensemble is an ensemble characterized by some joint density function on its eigenvalues. So, this is a specific instance of an implcitly distributed matrix. There are many things to consider about* ***identifiability*** *that are subtle, but important. Recall that in our formalization of a specturm as a formal statistic, we vectorized the matrix then tooks the determinant of the characteristic polynomial that came about it. For this reason, we can see that while Dimitriu's model provides an explicit formula, there is an issue of identifiability. That is, given some eigenvalues, there is no injective function to the characteristic polynomial that produced it. Rather, there are infinitly many equivalence classes of characteristic polynomials (and such, random matrices) that surjectively produce a given multiset of eigenvalues.*

**Remark 4.2.2** (Floating Point Errors)**.** *Because the model involves diving by* $\sqrt{2}$*, floating point errors and algorithmic systemic error will yield small, but negligible imaginary components.*

## 4.3 Dispersions

Consider the following plot of the sign-sorted eigenvalue dispersions below.

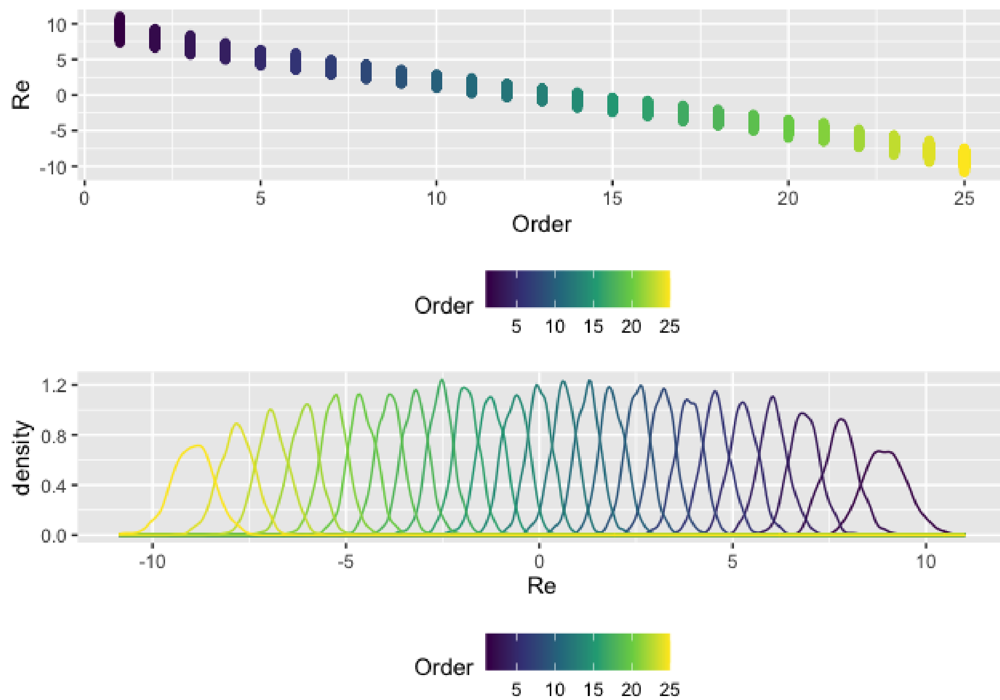


Figure 4.1: Dispersions of a $\beta = 4$ matrix with respect to ranking difference

# Appendix A

# Math Review

## A.1 Linear Algebra

### A.1.1 Introduction

**Definition A.1.1** (Eigenvalue)**.**

**Definition A.1.2** (Eigenvector)**.**

**Definition A.1.3** (Transpose Matrix)**.**

**Definition A.1.4** (Symmetric Matrix)**.**

**Definition A.1.5** (Hermitian Matrix)**.**

### A.1.2 Proof: Real Symmetric Matrices have Real Eigenvectors

**Notation**. For notational convenience, for any $N \in \mathbb{N}$, let $\widetilde{N} = \{1, \ldots, N\}$.

In this document, we prove that for any $M \times M$ real symmetric matrix, $S_M$, there exists for some eigenvalue $\lambda$, a corrosponding **real** eigenvector $\vec{v} \in \mathbb{R}^M$. Prior to starting the main proof, we begin with a lemma.

**Lemma**. Suppose we have a $M \times M$ real symmetric matrix with a some eigenvalue $\lambda$. If there we have a corrosponding eigenvector $v \in \mathbb{C}^M$, then every entry of $v$, say $v_i$ is equal to a **real** linear combination of the other entries $v_j \mid j \neq i$.

So, we will show that:

$$\forall i \in \widetilde{M} : v_i = \sum_{j \neq i} c_j v_j \quad (c_j \in \mathbb{R})$$

**Proof of Lemma**. Begin by taking a real symmetric matrix $S_M$ for some $M \in \mathbb{N}$. Suppose we have an eigenvalue $\lambda$. Then, if we have some eigenvector $v$, we know that:

$$(1) : \forall i \in \widetilde{M} : a_1 v_1 + \cdots + d_i v_i + \cdots + a_{m-1} v_m = \lambda v_i \quad (a_j \in \mathbb{R})$$

We obtain (1) by expanding the equality $Av = \lambda v$ and noticing that every row of $Av$ is expressible as the sum of the non-diagonal entries multiplied by $v_j \mid j \neq i$ plus $d_i v_i$. Note that since our matrix is symmetric, for some rows, some of the constants $a_j$ are not distinct but this should not raise any issues. Next, we collect the terms:

$$\forall i \in \widetilde{M} : a_1 v_1 + \cdots + a_{m-1} v_m = v_i(\lambda - d_i)$$

Since $S_M$ is a real symmetric matrix, the $a_j$ terms are real so we can say:

$$\forall i \in \widetilde{M} : v_i(\lambda - d_i) = \sum_{j \neq i} a_j v_j \quad (a_j \in \mathbb{R})$$

Finally, divide both sides by $(\lambda - d_i)$. Since $S_M$ is a real symmetric matrix, we know $\lambda \in \mathbb{R}$ then also $(\lambda - d_i) \in \mathbb{R}$. On the right hand side, the coefficients of the $v_j$ become $\frac{a_j}{(\lambda - d_i)}$. Since $a_j \in \mathbb{R}$, then also $\frac{a_j}{(\lambda - d_i)} \in \mathbb{R}$. Letting $c_j = \frac{a_j}{(\lambda - d_i)}$, we obtain:

$$\forall i \in \widetilde{M} : v_i = \sum_{j \neq i} c_j v_j \quad (\forall j : c_j \in \mathbb{R})$$

Thus, for any $M \in \mathbb{N}$, a real symmetric matrix with eigenvalue $\lambda$ must have a corrosponding eigenvector $v$ such that each of its entries is expressible as a real linear combination of the other entries. □

Now, we will prove the main theorem.

**Theorem** (**Taqi**). Suppose we have a $M \times M$ real symmetric matrix, $S_M$. Then, we will show that there exists for some eigenvalue $\lambda$, a corrosponding **real** eigenvector $\vec{v} \in \mathbb{R}^M$.

**Proof**. For this proof we will induct on the dimension of the matrix, $M$. So let the inductive statement be

$$f(M) : S_M \text{ has a real eigenvector } v \text{ corrosponding to an eigenvalue } \lambda$$

**Base Case**. Take the base case $M = 2$. Then by **Zoom Meeting 11.12**, we know $f(2)$ is true.

**Inductive Step**. For our inductive step, we need to show that $f(M) \Rightarrow f(M+1)$. So, let us assume $f(M)$. This means that we can assume any real symmetric matrix $S_M$ has a real eigenvector $v \in \mathbb{R}^M$ corrosponding to $\lambda$.

Next, we will write $S_{M+1}$ as the matrix $S_M$ augmented by some $u \in \mathbb{R}^M$ as follows:

$$S_{M+1} = \left[\begin{array}{c|c} S_M & u \\ \hline u^T & d_{M+1} \end{array}\right]$$

From our lemma, we use the fact that $S_{M+1}$ is symmetric and our assumption of $f(M)$ to obtain:

$$(1) : \forall i \in \{1, \ldots, m+1\} : v_i = \sum_{j \neq i} c_j v_j \quad (c_j \in \mathbb{R})$$

$$(2) : \forall i \in \tilde{M} : v_i \in \mathbb{R}$$

In particular for (2), we know that $v_i = \left(\sum_{j \neq i} \frac{a_j}{d_i - \lambda} v_j\right)$.

From (1), we know that for row $i = m+1$: $v_{m+1} = \sum_{j \neq m+1} c_j v_j \quad (c_j \in \mathbb{R})$ By (2), this is a linear combination of real entries $v_i$. Since $v_{m+1} \in \mathbb{R}$, it follows that:

$$\forall i \in \{1, \ldots, m+1\} : v_i \in \mathbb{R}$$

So, we have established that $f(m) \Rightarrow f(M+1)$.

By the induction, the theorem is proved. $\square$.

# A.2 Probability Theory

## A.2.1 Introduction

**Definition A.2.1** (Random Variable)**.** *A random variable* $X : \Omega \to \mathbb{R}$ *is a function from some sample space* $\Omega = \{s_i\}_{i=1}^n$ *to the real numbers* $\mathbb{R}$*. The sample space is taken to be any set of events such that the probability function corrosponding to the random variable,* $p_X$ *exhausts over all the events in* $\Omega$*. In other words, we expect* $\int_\Omega p_X(s) = 1$*.*

**Definition A.2.2** (Order Statistic)**.** *The ith order statistic*

## A.3 Markov Chains

**Definition A.3.1** (Markov Chain). *Say a set of random variables $X_i$ each take a value in a set, called the state space, $S_M = \{1, 2, \ldots, M\}$. Then, a sequence of such random variables $X_0, X_1, \ldots, X_n$ is called a Markov Chain if the following conditions are satisifed:*

- $\forall X_i : X_i$ has support and range $S_M = \{1, 2, ..., M\}$.
- *(Markov Property)* The transition probability from state $i \to j$, $\mathrm{P}(X_{n+1} = j \mid X_n = i)$ is conditionally independent from all past events in the sequence $X_{n-1} = i', X_{n-2} = i'', \ldots, X_0 = i^{(n-1)}$, excluding the present/last event in the sequence. In other words, given the present, the past and the future are conditionally independent.

$$\forall i, j \in S_M : \mathrm{P}(X_{n+1} = j \mid X_n = i) = \mathrm{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i', \ldots, X_0 = i^{(n-1)})$$

**Definition A.3.2** (Transition Matrixx). *Let $X_0, X_1, \ldots, X_M$ be a Markov Chain with state space $S_M$. Letting $q_{ij} = P(X_{n+1} = j \mid X_n = i)$ be the transition probability from $i \to j$, then the matrix $Q \in \mathcal{M}_{\mathbb{R}^+}[M \times M] : Q = (q_{ij})$ is the transition matrix of the chain. $Q$ must satisfy the following conditions to be a valid transition matrix:*

**Definition A.3.3** (Transition Matrix). *Take a Markov Chain with states $1, \ldots, M$. Letting $q_{ij} = P(X_{n+1} = j \mid X_n = i)$ be the transition probability from $i \to j$, then the matrix $Q = (q_{ij})$ is the transition matrix of the chain. For this transiton matrix to be valid, its rows have to be stochastic, meaning their entries sum to 1; $\forall i \in 1, \ldots, M$ : $\sum_{j \in 1,\ldots,M} q_{ij} = 1$.*

- $Q$ is a non-negative matrix. That is, note that $Q \in \mathcal{M}_{\mathbb{R}^+}[M \times M]$ so every $q_{ij} \in \mathbb{R}^+$. This follows because probabilities are necessarily non-negative values.
- The entries of every row $i$ of $Q$ must sum up to 1. This may be understood as applying the law of total probability to the event of transitioning from any given state $\forall i \in S_M$. In other words, the chain has to go somewhere with probability 1.
$$\forall i \in S_M : \sum_{j \in S_M} q_{ij} = 1$$
- Note, it is NOT necessary that the converse holds. The columns of our transition matrix need not sum to 1 for it to be a valid transition matrix.

**Definition A.3.4** (n-Step Transition Probability). *The $n-$step transition probability of $i \to j$ is the probability of being at $j$ exactly $n$ steps after being at $i$. We denote this value $q_{ij}^{(n)}$ :*

$$q_{ij}^{(n)} : \mathrm{P}(X_n = j \mid X_0 = i)$$

Realize:

$$q_{ij}^{(2)} = \sum_{k \in S_M} q_{ik} \cdot q_{kj}$$

Because by definition, a Markov Chain is closed under a support/range of $S_M$ so the event $i \to j$ may have taken any intermediate step $k \in S_M$. Realize by notational equivalence, $Q^2 = (q_{ij}^{(2)})$. Inducting over $n$, we then obtain that:

$$q_{ij}^{(n)} \text{ is the } (i, j) \text{ entry of } Q^n$$

**Definition A.3.5** (Marginal Distribution of Xn)**.** *Let* $\boldsymbol{t} = (t_1, t_2, \ldots, t_M)$ *such that* $\forall i \in S_M : t_i = P(X_0 = i)$*So,* $\boldsymbol{t} \in \mathcal{M}_{\mathbb{R}}[1, M]$*. Then, the marginal distribution of* $X_n$ *is given by the product of the vector* $\boldsymbol{t}Q^n \in \mathcal{M}_{\mathbb{R}}[1, M]$*. That is, the* $j^{th}$ *component of that vector is* $P(X_n = j)$ *for any* $j \in S_M$*. We may call* $\boldsymbol{t}$ *an initial state distribution.*

### A.3.1 Classification of states

- A state $i \in S_M$ is said to be **recurrent** if starting from $i$, the probability is 1 that the chain will *eventually* return to $i$. If the chain is not recurrent, it is **transient**, meaning that if it starts at $i$, there is a non-zero probability that it never returns to $i$.
- Caveat: As we let $n \to \infty$, our Markov chain will gurantee that all transient states will be left forever, no matter how small the probability is. This can be proven by letting the probability be some $\varepsilon$, then realizing that by the support of Geom($\varepsilon$) is always some finite value, then the equivalence between the Markov property and independent Geometric trials gurantees the existence of some finite value such that there is a success of never returning to $i$.

**Definition A.3.6** (Reducibility)**.** *A Markov chain is said to be* ***irreducible*** *if for any* $i, j \in S_M$*, it is possible to go from* $i \to j$ *in a finite number of steps with positive probability. In other words:*

$$\forall i, j \in S_M : \exists n \in \mathbb{N} : q_{ij}^{(n)} > 0$$

- From our quantifier formulation of irreducible Markov chains, note that we can equivalently say that a chain is irreducible if there is integer $n \in \mathbb{N}$ such that the $(i, j)$ entry of $Q^n$ is positive for any $i, j$.
- A Markov chain is **reducible** if it is not **irreducible**. Using our quantifier formulation, it means that it suffices to find transient states so that:

$$\exists i, j \in S_M : \nexists n \in \mathbb{N} : q_{ij}^{(n)} > 0$$

# Appendix B

# Algorithm Appendix

## B.1 Matrix Simulation

### B.1.1 Stochastic Matrices

**Algorithm B.1.1** (Stochastic Matrix)**.**

1. *Fix* $N \in \mathbb{N}$*.*

**Algorithm B.1.2** (Symmetric Stochastic Matrix)**.**

1. *Fix* $N \in \mathbb{N}$*.*

**Algorithm B.1.3** (Transition Matrix of an Erdos-Renyi Graph)**.**

1. *Fix* $N \in \mathbb{N}$ *and* $p \in [0, 1]$*.*
2. *Generate a matrix* $Q$ *such that every entry* $i, j \in 1, \ldots, N$ *is* $x_{ij} \sim$ *Unif*$(0, 1)$*.*
3. *For each* $v_i$ *in* $\{1, \ldots, N\}$*, generate* $deg(v_i) \sim Bin(N, p)$*.*
4. *Randomly chose* $(1 - deg(v_i)$ *vertices, set the entries* $x_{ij}$ *in the* $j$ *columns to 0.*
5. *Renormalize the matrix by dividing each row by its sum; let* $(x_i) \leftarrow (x_i)/\sum_j (x_i)$*.*

### B.1.2 Normal Matrices

**Algorithm B.1.4** (Hermite $\beta$-Matrix)**.**

1. *Fix* $N \in \mathbb{N}$*. Use L's result.*

**Algorithm B.1.5** (Matrix)**.**

1. *Fix* $N \in \mathbb{N}$*.*

**Algorithm B.1.6** (Matrix)**.**

1. *Fix* $N \in \mathbb{N}$*.*

**Algorithm B.1.7** (Matrix)**.**

1. *Fix* $N \in \mathbb{N}$*.*

# Appendix C

# Code Appendix

# Appendix D

# Mixing Time Simulations

## D.1 Introduction

In this chapter, we'll talk about ratio-mixing time simulations. Essentially, these simulations are a method of approximating the distribution of a random transition matrix's mixing time. There will be a fun exploration of the Erdos-Renyi matrix ensembles and we will computationally show that the parameterized ensemble has a mixing time inversely proportional to graph sparsity.

## D.2 Mixing Time Simulations

With the Erdos-Renyi graph defined, we may now motivate the simulation of random walks on them. First, however, we need to generate their corrosponding transition matrices. An algorithm for this is outlined below.

Suppose we have simulated a transition matrix for an Erdos-Renyi graph called $Q$. Now, fixing some initial probability distribution $\vec{x} \in \mathbb{R}^M$, we may consider the evolution sequence of a random walk on this Erdos-Renyi graph by taking its evolution sequence $\mathcal{S}(Q, x)$.

**Definition D.2.1** (Random Batches)**.** *Let $\mathbb{F}$ be a field, and fix some $M \in \mathbb{N}$. Let $\mathcal{B}_\lambda \subset \mathbb{F}^M$ be a uniformly random batch of points in the $M$-hypercube of length $\lambda$. That is,*

$$\mathcal{B}_\lambda = \{\vec{x} \mid x_i \sim \textit{Unif}(-\lambda, \lambda) \textit{ for } i = 1, \ldots, M\}$$

***Note:*** *If $\mathbb{F} = \mathbb{C}$, then take $\vec{x} \in \mathcal{B}_\lambda$ to mean $\vec{x} = a + bi$ where $a, b \sim \textit{Unif}(-\lambda, \lambda)$.*

**Definition D.2.2** (Evolution Sequence)**.** *An evolution sequence of a vector $\vec{\pi}$ and a transition matrix $Q$ is defined as the sequence $\mathcal{S}(Q, \pi) = (\pi'_n)_{n=1}^N$ where $\pi'_n = \pi Q^n$*

**Definition D.2.3** (Finite Evolution Sequences)**.** *Suppose we sample a random point from $\mathcal{B}_\lambda$, emulating a random point $\vec{v} \in \mathbb{F}^M$. Additionally, let $Q \in \mathbb{F}^{M \times M}$ be a transition matrix over $\mathbb{F}$. Fixing a maximum power ('time') $T \in \mathbb{N}$, define the evolution sequence of $\vec{v}$ as follows:*

$$\mathcal{S}(v, Q, T) = (\alpha_n)_{n=1}^T \textit{ where } \alpha_k = vQ^k$$

*If we do not impose a finiteness constraint on the sequence, we consider powers for $n \in \mathbb{N}$ or $t = \infty$*

**Definition D.2.4** (Consecutive Ratio Sequences). *Accordingly, define the consecutive ratio sequence (CST) of $\vec{v}$ as follows:*

$$\mathcal{R}(v, Q, T) = (r_n)_{n=2}^{T} \text{ where } (r_n)_j = \frac{(\alpha_n)_j}{(\alpha_{n-1})_j} \text{ for } j = 1, \dots, M$$

*In other words, the consecutive ratio sequence of $v$ can be obtained by performing* ***component-wise division*** *on consecutive elements of the evolution sequence of $v$.*

**Definition D.2.5** (Near Convergence). *Because these sequences may never truly converge to eigenvectors of the matrix, we formalize a notion of "near convergence". As a prelimenary, we first define $\varepsilon$-equivalence. Let $\mathbb{F}$ be a field, and fix $\varepsilon \in \mathbb{R}^+$. Suppose we have vectors $v, v' \in \mathbb{F}^M$. Then, $v \sim_\varepsilon v'$ if $||v - v'|| < \varepsilon$ where $|| \cdot ||$ is the norm on $\mathbb{F}$.*

Let $\varepsilon \in \mathbb{R}^+$, and suppose we have an evolution sequence $(a[\vec{v}])_n$. Then, $a_n$ $\varepsilon$-converges at $N \in \mathbb{N}$ if:

$$\forall n \geq N \mid a_N \sim_\varepsilon a_n$$

## D.3 Erdos-Renyi Ensemble Simulations

**Definition D.3.1** (Erdos-Renyi Graph). *An Erdos-Renyi graph is a graph $G = (V, E)$ with a set of vertices $V = 1, \dots, M$ and edges $E = \mathbb{1}_{i,j \in V} \sim Bern(p_{ij})$. It is homogenous if $p_{ij} = p$ is fixed for all $i, j$.*

Essentially, an Erdos-Renyi graph is a graph whose 'connectedness' is parameterized by a probability $p$ (assuming it's homogenous, which this document will unless otherwise noted). As $p \to 0$, we say that graph becomes more sparse; analogously, as $p \to 1$ the graph becomes more connected.

Recall from probability theory that a sum of i.i.d Bernoulli random variables is a Binomial variable. As such, we may alternatively say that the degree of each vertex $v$ is distributed as $deg(v) \sim Bin(M, p)$. This is helpful to know because the process of simulating graphs becomes much simpler.

## D.4 Questions

1. How are the entries of the CRS distributed? Are they normal, and if so, what is its mean?

2. Are the entries of the CRS i.i.d as $t \to \infty$?

3. For an Erdos-Renyi matrix, is the mixing time $t$ dependent on the parameter $p$?

4. What impact does the running time parameter $T$ have on $\sigma$ (the variance of the distribution of the CRS entries)?

### D.4.1 Cauchy Distributed Ratios

It seems to be the case that the **log-transformed** entries of the CRS are Cauchy distributed about $\log \lambda_1$ where $\lambda_1 = \max(\sigma(Q))$, the largest eigenvalue of $Q$. That is,

$$r_i \sim \text{Cauchy}(\ln \lambda_1) \text{ for } i = 1, \dots, M$$

# References

Dumitriu, I. (2018). Matrix models of beta ensembles. *Journal of Mathematical Physics 43, 5830*, (pp. 1–5).