Spectral Statistics of Random Matrices

A Thesis
Presented to
The Division of Mathematical 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

Acknowledgements

Nate, thank you for being an amazing mentor, and for being kind and insightful every step of the way. Also, thank you for introducing me to the wonderful worlds of probability theory and random matrices.

Kelly, taking introductory statistics with you was such an instrumental experience in my journey as a statistician. Thank you for teaching me R and being my wonderful (re)introduction to the world of statistics. I owe all these awesome visualizations to you.

Angélica, I really appreciate your encouragement and the values of persistence you tried to instill in me. In my darkest days, I always tried to fall back on and recall what you told me. By and large, I really do believe that was the reason I kept going.

Samuel, thank you for bringing the love and joy I needed to get through this year. No matter what or when, I know I can always count on our memories to fill me with glee and elation. I just know you are going to do amazing things and I cannot wait to hear all about it.

Lucy, I cannot explain the odds of the circumstances in which our paths crossed. I guess sometimes we just have to marvel at the serendipity life brings us. You will always be my truest friend.

Candice, I do not think I can overstate the level with which you shaped me both academically and personally. I still remember the day you picked me from the Portland airport. Though I had just arrived from a 7000 mile journey, you made it such that it felt like I was already home.

Juce, I will always treasure our countless memories of joy, love, and friendship. No friendship have I ever had was as genuine and consequential as ours. I cannot wait to see what we both grow to become.

And to all my friends who know themselves – thank you for all your love and support!

List of Abbreviations

r.v	Random Variable
i.i.d	Independent and Identically Distributed
pdf	Probability Distribution Function
cdf	Cumulative Distribution Function

Table of Contents

List of Tables

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 (H)
Uniform	Unif(a, b)	$a, b \in \mathbb{R}$	Explicit (H)
Hermite-β	$\mathcal{H}(\beta)$	$\beta \in \mathbb{N}$	Explicit (NH)
Stochastic	Stoch	-	Implicit
Erdos-p	ER(p)	$p \in [0, 1]$	Implicit

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

Table of Dispersion Metrics				
Metric*	Notation	Formula	Symmetric	Parameters
Standard Norm	δ_n	$\lvert z' - z\rvert$	True	-
β-Norm	δ_β	$\lvert z' - z\rvert^\beta$	True	$\beta \in \mathbb{N}$
Difference of Absolutes	δ_{abs}	$\lvert z'\rvert - \lvert z\rvert$	False	-
Identity Difference	δ_{id}	$z' - z$	False	-

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	Π_C	$\{(i, j) \mid i = j + 1 \text{ for } i, j \in \mathbb{N}_N\}$
All	Π_0	$\{(i, j) \mid i, j \in \mathbb{N}_N\}$

Abstract

On their own, random variables exhibit deterministic properties regarding their uncertainty. The same generalization applies to random matrices, which are matrices where some or all of its entries are random variables. One particular random matrix statistic worth investigating is the distribution of its eigenvalues or its spectrum. In this thesis, we explore various classes of random matrices and their corresponding spectral statistics. Namely, we observe their spectra and dispersions. That will be demonstrated with simulations, in which we survey several results in random matrix theory and related fields.

Dedication

For my mother *Eman*, who I am eternally indebted to.

Introduction

You begin to read the title of this thesis. The words random and matrix are familiar, and odds are you correctly surmise what random matrices are. If you cannot wait to know — they are matrices where some or all of its entries are random variables. But, what are *spectral statistics*? Do they have to do with rainbows? Sceptres? Well, no they don't, but they're almost as colorful and regal.

At first, one might surmise that the origin of the word has to do with the Spectral theorem commonly taught in a first-year linear algebra or functional analysis course. This would be partially correct. For context, the Spectral theorem is concerned with the diagonalization of a matrix, and in turn, its eigenvalues and eigenvectors. However, this is not precisely where the term originates. The term originally comes from David Hilbert, who coined the term 'spectral theory' to describe his original formulation of Hilbert space theory. So, the initial Spectral theorem actually was formulated as an infinite-dimensional variant of the theorem on principal axes of an ellipsoid. It was thus serendipitous that 'spectral' theory could explain features of *atomic spectra* later in quantum mechanics – this application surprised even Hilbert himself! Hilbert himself says, "I developed my theory of infinitely many variables from purely mathematical interests, and even called it 'spectral analysis' without any presentiment that it would later find application to the actual spectrum of physics." [Steen (1973)]

That being said, our field of study, aptly called *random matrix theory*, has surprising connections to the world of statistical physics. To put it briefly, we can say that random matrix theory is a field at the intersection of linear algebra (matrices) and probability theory (random variables). For a primer or review of these subjects, **Appendix A** may be a useful resource. In any case, the field of random matrix theory was extensively developed in the early 20th century, and its development can be accredited to two individuals: John Wishart and Eugene Wigner. Surprisingly, they did not work together and were not interested in random matrices for the same reason.

Namely, Wishart was interested in applying random matrices in the realm of covariance matrices. Today, both Wishart matrices and the Wishart distribution are named after him. Wigner, on the other hand, was interested in finding a model of atoms with heavy nuclei. In turn, Wigner found connections between the deterministic properties of atomic nuclei and their random and stochastic behaviors [Wigner (1955)]. The link? Random matrices. Named after Wigner, one of the results that we will discuss in **Section 3.3** is Wigner's Surmise, a result about the eigenvalue

spacings of symmetric random matrices. All that being said, it was not until another mathematician, Freeman Dyson, came along a few decades later and realized both mathematicians were working on similar problems. Named after him is the Dyson index, which is mentioned in **Section 4.1.2**.

So, to answer the question: 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. Namely, we will consider two spectral statistics of random matrices:

1. Their eigenvalues, which we call their **spectra**.
2. The spacings between those eigenvalues, which we called their **dispersion**.

With all the technicalities out of the way, we finally state the intention of this paper. The intention is that a reader with a decent background in the prerequisites mentioned above can come out with a toolkit to study more advanced results in random matrix theory. The field is rich and there are so many connections to other fields. Out of this paper, the reader will hopefully come out knowing many essential results, theorems, and ideas to build on top of. In a way, reading through this, the reader's journey is akin to mine as I was learning the material. We achieve our goals by performing simulations and surveying important results in the field. Additionally, we report some new findings and provide numerous examples of how the parameterization or distribution of random matrices can impact their spectral statistics.

Since it is necessary to achieve our goal, this thesis sets out to provide a standardized language and notation to formalize several objects in random matrix theory. This was done hand-in-hand in developing the **RMAT** package, the package written for the simulation component of this thesis. In fact, the definitions are very much pragmatically motivated; they were written in consideration of the programming maneuvering that took place to simulate the random matrices and their spectral statistics. Again, it is paramount to highlight that there is a large simulation component of this thesis. Generally speaking, to be able to study random objects, we need to be able to simulate them first. So, for all of our simulations, this thesis will utilize the aforementioned **RMAT** package.

The RMAT Package

> Tell me and I forget; teach me and I may remember; involve me and I learn.
>
> *Confucius*

The **R**andom **M**atrix **A**nalysis **T**oolkit package was written with two purposes in mind: *interactivity and reproducibility*. As the quote above implies, interactivity is a critical part of learning. For this reason, all the necessary source code for this thesis will be made available. You are strongly encouraged to reproduce these simulations yourself! There will be instructions below for how to obtain the **RMAT** package. The code appendix **(Appendix C)** contains a minimalist version of the code needed to perform these simulations.

As mentioned before, this package was developed alongside this thesis in order to facilitate the simulation of these random matrices and spectral statistics. To showcase the methodology of the simulations, code snippets will be sprinkled throughout the thesis. All code examples in this thesis are reproducible by setting the seed using `set.seed(23)`.

In any case, with the package explained, the honest truth is that the formalizations and definitions provided in this thesis were written after the code was. As mentioned earlier, a large and very important part of the thesis is developing intuitive definitions of random matrix objects that are consistent with the way the code **RMAT** package is implemented. In other words, the thesis in many ways formalizes the programming maneuvers used in **RMAT** after the fact.

Package Installation

There are three ways to get the RMAT package. In order of convenience:

1. CRAN: simply run `install.packages("RMAT")` in R.
2. Github: either run `devtools::install_github(repo = "ataqi23/RMAT")` or clone the repository found here[1].
3. Source code: reproduce the code available in **Appendix C**.

homogeneous

[1] https://www.github.com/ataqi23/RMAT

Chapter 1

Random Matrices

> Unfortunately, no one can be told what The Matrix is. You'll have to see it for yourself.
>
> *Morpheus*
> *The Matrix*

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 which have *some or all* entries as random variables that are distributed in accordance to either some explicit distribution or algorithm. So, to define a random matrix, our approach will be to do so by formalizing and defining what it means for one to be $\mathcal{D}$-distributed. This way, the notation encapsulates and completely characterizes the random matrix.

Prior to beginning the discussion on $\mathcal{D}$-distributions, the reader should be familiar or at least acquainted with the notion of random variables and what they are. For a review, see **Appendix A.2.1**.

1.1 $\mathcal{D}$-Distributions

As a general rule, when it comes to random simulation, there is usually a rule or constraint to which the randomness must conform. For example, sampling a vector of random variables from a distribution is a rudimentary example of this. For random matrices, there are various techniques for generating their entries that are not just sampling from theoretical distributions. As such, we motivate the $\mathcal{D}$-distribution: a generalized matrix entry distribution framework.

Definition 1.1.1 ($\mathcal{D}$-distribution)**.** *Suppose P is a $\mathcal{D}$-distributed random matrix. Then, we notate this $P \sim \mathcal{D}$. In the simplest of terms, $\mathcal{D}$ is essentially the algorithm that generates the entries of P. We define two primary methods of distribution:* ***explicit*** *distribution, and* ***implicit*** *distribution. If $\mathcal{D}$ is an explicit distribution, then*

some or all the entries of P are independent random variables with a given distribution. Otherwise, if $\mathcal{D}$ is implicit, then the matrix has dependent entries imposed by the algorithm that generates it.

1.1.1 Explicit Distributions

Homogeneous Explicit $\mathcal{D}$-distributions

The simplest type of $\mathcal{D}$-distribution is one that is explicitly and homogeneously distributed. From thereon after, this will be shortened as e.h.d. Suppose $\mathcal{D}$ is a probability distribution for random variables in the classical sense. The simplest way to think of e.h.d distributions is to use the concept of notational overload. For example, it is unambiguous to say that an r.v. $X \sim \mathcal{N}(0,1)$. However, the same cannot be said if we said a matrix $P \sim \mathcal{N}(0,1)$.

That being said, we can define e.h $\mathcal{D}$-distributions as an notational extension that means **every** entry of the matrix is an i.i.d random variable with that same distribution! In other words, we simply perform entry-wise sampling from the corresponding r.v. distribution. Note that this means by construction, $\mathcal{D}$ can only be e.h.d. if it has a corresponding probability distribution for random variables.

Definition 1.1.2 (Homogeneous Explicit $\mathcal{D}$-distribution)**.** *Suppose $P \sim \mathcal{D}$ where $\mathcal{D}$ is a homogeneous and explicit distribution. Additionally, let $\mathcal{D}^*$ denote the corresponding random variable analogue of $\mathcal{D}$. Then, every single entry of P is an i.i.d random variable with the corresponding distribution. That is,*

$$P \sim \mathcal{D} \iff \forall i, j \mid p_{ij} \sim \mathcal{D}^*$$

Example. *Suppose $P \sim \mathcal{N}(0,1)$ and that P is a 2×2 matrix. Then, $p_{11}, p_{12}, p_{21}, p_{22}$ are independent, identically distributed random variables with the standard normal distribution.*

Algorithm 1.1.1 (Homogeneous Explicit $\mathcal{D}$-Matrix)**.**

1. *To simulate a $\mathcal{D}$-distributed square matrix P of size N, fix $N \in \mathbb{N}$.*
2. *Sample a vector $\vec{X}$ with N i.i.d entries from $\mathcal{D}$.*
3. *Assign the vector $\vec{X}$ as a row of the matrix P. Repeat for every other row.*
4. *Return the $\mathcal{D}$-distributed matrix P.*

Formalization. *Explicit homogeneous distributions can be formalized as overloading the standard notation of random variable distribution as seen in probability theory. 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! Suppose P is an $N \times N$ random matrix that is explicitly and homogeneously $\mathcal{D}$-distributed. Then, this would mean that P is an array of N^2 i.i.d random variables sampled from $\mathcal{D}$.*

Non-Homogeneous Explicit $\mathcal{D}$-distributions

There are a few instances where we encounter the need to only initialize a subset of a matrix's entries as random variables. In this case, we say that a matrix has a non-homogeneous explicit $\mathcal{D}$-distribution. This type of distribution is similar to e.h. $\mathcal{D}$-distributions, but we no longer have the ability to overload random variable notation and indicate that only some entries have such distribution. This would be confusing because not every entry has the same distribution anymore. As such, we have to define new notation to do so.

Simply put, a non-homogeneous explicit $\mathcal{D}$-distribution can be completely characterized by listing the distributions of every entry. In this thesis, there is only one application of this type of $\mathcal{D}$-distribution; however, it describes the entry distributions by its diagonals — this is the Hermite β-ensemble matrix. Since its definition is characterized by distributions on the diagonals, we may avoid full entry-wise generality for the purpose of being succinct. As such, we will only describe non-homogeneous explicit $\mathcal{D}$-distributions as schemes where we assign diagonal bands a specific vector of i.i.d random variables.

Definition 1.1.3 (Diagonal Bands)**.** *Suppose $P = (p_{ij})$ is an $N \times N$ matrix. Then, P may be partitioned into $2n-1$ rows called diagonal bands. Each band is denoted $[\rho]_P$ where $[\rho]_P = \{p_{ij} \mid \rho = i - j\}$. We have $\rho \in \{-(N-1), \ldots, -1, 0, 1, \ldots, N-1\}$.*

Here's an example of using the diagonal bands constructor notation.

Example. *Suppose P is an $N \times N$ matrix. Then, $[0]_P$ is the main diagonal of P since $p_{ii} \Rightarrow i = j \Rightarrow i - j = 0 \Rightarrow p_{ii} \in [0]_P$. Similarly, the main off-diagonal in the upper triangle is $[-1]_P$ since $p_{12} \in [-1]_P$. Likewise, the main off-diagonal in the lower triangle is $[1]_P$. The entry in the top-right corner of the matrix, p_{1N} solely comprises $[1-N]_P$.*

Code Example. *Here is an example utilizing diagonal bands constructors in R.*

```
# Set the dimension of the matrix
N <- 5
# Generate an example matrix of zeros
P <- matrix(rep(0, N^2), nrow = N)
# Assign the upper main off-diagonal band as a vector
rho <- -1
P[row(P) - col(P) == rho] <- rnorm(n = 4, mean = 0, sd = 1)
# Assign the main diagonal as a vector
rho <- 0
P[row(P) - col(P) == rho] <- rep(10, N)
# Returns the following
P
```

```
     [,1]        [,2]        [,3]        [,4]       [,5]
[1,]   10  0.1932123   0.0000000   0.0000000   0.000000
[2,]    0 10.0000000  -0.4346821   0.0000000   0.000000
[3,]    0  0.0000000  10.0000000   0.9132671   0.000000
[4,]    0  0.0000000   0.0000000  10.0000000   1.793388
[5,]    0  0.0000000   0.0000000   0.0000000  10.000000
```

For the distribution using diagonal bands, consider the definition of the β-matrix below. The definition utilizes a generative algorithm which tells us the distribution on the diagonals of the matrix, shown below.

Algorithm 1.1.2 (Dumitriu's Beta Matrix)**.**

1. *To simulate an $N \times N$ beta matrix, 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 $\chi(df = c_j)$ where $c_j = \beta \cdot j$ for columns spanning $j = 1, \ldots, N-1$.*
4. *Normalize the entries by dividing by $\sqrt{2}$.*

Definition 1.1.4 (Hermite-β Matrix)**.** *Suppose $P \sim \mathcal{H}(\beta)$ is an $N \times N$ matrix. Then, the main diagonal $[0]_P \sim \mathcal{N}(0, 2)$. Additionally, both the main off-diagonals are equal and they are given by $[1]_P = [-1]_P = \vec{X} = (X_k)_{k=1}^{N-1}$ where $X_k \sim \chi(df = \beta k)$. Lastly, after normalizing by $\frac{1}{\sqrt{2}}$, we obtain a Hermite-β distributed matrix. Note that this is a symmetric tridiagonal matrix.*

That being said, one must be careful to not say that the entries are identically distributed, even within the diagonal bands.

Remark (Off-Diagonal Distributions)**.** *Note that the off-diagonal distributions are chi variables with varying degrees of freedom; in fact it is an increasing linear sequence that is a multiple of β. For this reason, we cannot say that the off-diagonal entries are identically distributed despite them being within the same diagonal band. However, they do remain independent. On the other hand, the diagonal is a vector of N i.i.d $\mathcal{N}(0, 2)$ variables.*

The Hermite-β matrix is a matrix model for the β-ensembles. This notation might clash with the random variable distribution called the Beta distribution.

Warning (β-Notation)**.** *Please note that the Hermite β-ensemble matrices are* ***not*** *related to the Beta distribution. Instead, $\beta \in \mathbb{N}$ is a natural number that determines the nature of the matrix ensemble (more later). The Beta(a, b) distribution is a r.v. distribution that takes in two parameters. For this reason, it is possible to have a homogeneous explicit distribution $P \sim$ Beta(a, b), but they do not mean the same thing.*

1.1.2 Implicit Distributions

Now, we are done with defining the explicit $\mathcal{D}$-distributions. The other type of $\mathcal{D}$-distribution is the implicit $\mathcal{D}$-distribution. As a general rule, when it comes to implicit $\mathcal{D}$-distributions, we are concerned less about the distribution of the actual matrix entries and more so about its holistic properties.

In this thesis, we will cover one type of implicit $\mathcal{D}$-distributed matrix: the stochastic matrix. One might ask, what are stochastic matrices? Stochastic matrices, in short, are matrices that represent random walks on Markov Chains (see **Appendix A.3**). So, for a fixed graph, there exists a matrix that represents a random walk on that graph. We call that either the stochastic matrix or transition matrix of the graph.

We will consider two variations of stochastic matrices: fully connected stochastic matrices and sparse stochastic matrices. The sparse stochastic matrices will be referred as Erdos-Renyi matrices, they are essentially the same as stochastic matrices, except they have a parameter p that determines how connected the graph it represents is. They will be discussed in more detail in **Section 1.2.1**.

Stochastic Matrices

Stochastic matrices will serve as our canonical implicitly distributed $\mathcal{D}$-distributed matrices. How are they distributions?

Conventionally, stochastic matrices represent random walks on fixed graphs - so what does it mean to sample a random stochastic matrix? Well, the algorithm tells us that sampling a random stochastic matrix represents a random walk on a fixed random graph with **randomized weights**. In other words, we sample a fully connected graph, and randomize the weights of each edge. As mentioned previously, in **Section 1.2.1**, we explore walks on Erdos-Renyi graphs, which are graphs with introduced sparsity (have edges with weight 0) as opposed to fully connected graphs.

Consider the following generating algorithms below. We first start by generating a stochastic row - a row that sums to one. Afterwards, we simply take N stochastic rows to generate a stochastic matrix.

Algorithm 1.1.3 (Stochastic Row)**.**

1. *To sample a row r of size N, fix $N \in \mathbb{N}$.*
2. *Sample a vector $\vec{X}$ with N i.i.d entries between $[0, 1]$. So, sample $\vec{X} \sim$ Unif$(0, 1)$.*
3. *Assign $r \leftarrow \vec{X}$, and then normalize the row by dividing each entry by the row sum; so assign $(r) \leftarrow (r)/\sum_{j=1}^{N} r_j$.*
4. *Return the stochastic row r.*

As such, we define the implicit $\mathcal{D}$-distribution corresponding to a transition matrix on a complete graph with randomized weights as $\mathcal{D} =$ Stochastic as the distribution that is characterized by the generating algorithm below.

Algorithm 1.1.4 (Stochastic Matrix)**.**

1. *To generate a stochastic square matrix P of size $N \times N$, fix $N \in \mathbb{N}$.*
2. *Then, for every row of P, randomly sample a stochastic row of size N and assign it.*
3. *Return the stochastic matrix P.*

Remark (Implicit Distribution)**.** *Recall that the reason we call stochastic matrices implicitly $\mathcal{D}$-distributed is because their entries are not explicitly sampled from a probability distribution. This is due to our method of generating stochastic rows in* ***Algorithm 1.1.3****, which normalizes the vector of random variables by its sum. This procedure of normalizing by the sum causes complications, since the entries all become dependent on each other. That being said, to avoid the process of finding the distribution of these dependent entries, we will encapsulate the distribution and abstract it away by calling it an implicit distribution!*

If the reader is interested, we do know that the relevant distributions to consider are the Dirichlet distribution and the related Beta distributions.

Remark (Stochastic Row Distributions)**.** *The distribution of a stochastic row of size N is in fact known to be related to the Dirichlet distributions. First, we will cover a neat geometric representation of a stochastic row. Consider the standard N-simplex, a polytope which represents the generalization of an equilateral triangle/tetrahedron to N dimensions. The standard N-simplex is denoted $\Delta^N = \{x \in \mathbb{R}^N \mid \sum_{i=1}^N x_i = 1\}$. Realize that by definition, the standard N-simplex is precisely the space of all N-dimensional stochastic rows! So, the relevant distribution to consider is the Dirichlet distribution, whose support is precisely the standard N-simplex. As such, sampling a stochastic row can be represented as uniformly sampling from the support of the Dirichlet distribution.*

1.1.3 Random Matrices

With the various types of $\mathcal{D}$-distributions defined, the definition of a random matrix is quite simple.

Definition 1.1.5 (Random Matrix)**.** *Let $P \sim \mathcal{D}$ be an $N \times N$ matrix over $\mathbb{F}$. Then, P is represented as a vector of K random variables over $\mathbb{F}$, which we denote $\vec{X} = (X_1, X_2, \ldots, X_K)$ for some $K \leq N^2$. In other words, some or all the entries of P are random variables, and their distributions are precisely determined by the $\mathcal{D}$-distribution. Additionally, if $\mathcal{D}$ is an explicit distribution, $\mathcal{D}^\dagger$ represents the symmetric/hermitian version of $\mathcal{D}$.*

Here is a clarification on symmetric/hermitian versions of distributions.

Remark (Symmetric/Hermitian Matrices)**.** *As mentioned in the definition, we automatically have a class of derivative $\mathcal{D}$-distributions given that they are explicitly distributed denoted by $\mathcal{D}^\dagger$. To make a matrix symmetric (or Hermitian if $\mathbb{F} = \mathbb{C}$), then we simply set the elements in the upper triangle to be equal to those in the lower triangle.*

As mentioned in the definition, a random matrix defined over a field $\mathbb{F}$ has entries from $\mathbb{F}$ that are determined by the $\mathcal{D}$-distribution. Sometimes, we may specify a matrix to have complex entries. We notate this by specifying $\mathbb{F} = \mathbb{C}$, and interpret it as described below.

Remark (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. Note that this means we cannot set the field for implicit $\mathcal{D}$ distributions, as the field is automatically chosen by the generative algorithm.*

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

Code Example (Standard Normal Matrix)**.** *Let $P \sim \mathcal{N}(0, 1)$ be a 4×4 random matrix. Then, we may generate P as such:*

```
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
```

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. Simply put, a random matrix ensemble is essentially a collection of various matrices distributed the same way. One common theme in this thesis is that we will find that on their own, random matrices provide little information. However, when we consider them at the ensemble level, we start to obtain more fruitful results. Without further ado, we define the random matrix ensemble.

Definition 1.2.1 (Random Matrix Ensemble)**.** *A $\mathcal{D}$-distributed ensemble $\mathcal{E}$ of $N \times N$ random matrices over $\mathbb{F}$ of size K is defined as a set of K $\mathcal{D}$-distributed random matrices, and it is denoted:*

$$\mathcal{E} = \bigcup_{i=1}^{K} P_i \text{ where } P_i \sim \mathcal{D} \text{ and } P_i \in \mathbb{F}^{N \times N}$$

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

Code Example (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 3×3 complex Hermitian standard normal matrices of size 10 as such:*

```
library(RMAT)
# By default, mean = 0 and sd = 1.
ensemble <- RME_norm(N = 3, cplx = TRUE, herm = TRUE, size = 10)
# Outputs the following
ensemble
```

```
[[1]]
                    [,1]                 [,2]                [,3]
[1,]  0.19321+1.57578i -0.43468-0.21829i 0.91327+1.04654i
[2,] -0.43468+0.21829i  0.99661+0.48155i 1.10749+1.21638i
[3,]  0.91327-1.04654i  1.10749-1.21638i 0.04544-0.44231i

...

[[10]]
                    [,1]                 [,2]                [,3]
[1,] -0.59931+1.24286i  1.29457+0.66058i  0.83539-0.16662i
[2,]  1.29457-0.66058i  0.78841+0.09818i -1.16592+1.14666i
[3,]  0.83539+0.16662i -1.16592-1.14666i -0.51256+0.17750i
```

Now, we are ready to survey, characterize, and briefly discuss a few special recurring ensembles in this thesis.

1.2.1 Erdos-Renyi p-Ensembles

As mentioned in **Section 1.1.3**, Erdos-Renyi transition matrices are a specific class of stochastic matrices. Again, a stochastic matrix represents a random walk on a fixed graph. For $\mathcal{D}$ = Stochastic, we generate matrices that represent walks on fully connected graphs with randomized weights. For the Erdos-Renyi matrices, we will alternatively be considering a different graph to represent — the p-Erdos-Renyi graphs.

Essentially, these are graphs whose vertices are connected with a uniform probability p. This way, we add more variation since the connectivity of a graph is one feature of interest. 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 homogeneous 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. We will assume every edge (i, j) has a non-zero weight with probability p (so $p_{ij} \neq 0$). As $p \to 0$, we say that graph becomes more *sparse*; analogously, as $p \to 1$ the graph becomes more *connected*.

Remark (Homogeneity)**.** *In this thesis, we will assume that every edge has a uniform probability p of being connected. That is, we will assume p_{ij} is non-zero with probability p for all i, j. We say that the edge probabilities are homogeneous - akin to our description of homogeneous $\mathcal{D}$-distributions.*

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 (corresponding to a row) is distributed as $deg(v) \sim Bin(N, p)$ where N is the number of vertices. This makes simulating the graphs much easier. To achieve this effect, we first sample the vertex degree, and then uniformly randomly sever edges by setting the weights of $N - deg(v)$ edges to 0.

All, that being said, we now motivate the defintion of an p-Erdos-Renyi transition matrix as an implicit $\mathcal{D}$-distribution by the generating algorithm below.

Algorithm 1.2.1 (Transition Matrix for an Erdos-Renyi Graph)**.**

1. *Fix $N \in \mathbb{N}$ and $p \in [0, 1]$.*
2. *Generate a matrix $Q \sim Unif(0, 1)$, i.e. with randomly uniform entries on [0,1].*
3. *For each row r in $\{1, \dots, N\}$, generate $deg(r) \sim Bin(N - 1, p)$.*
4. *Randomly choose $N - deg(r)$ vertices, then set the entries r_j in the j columns to 0 to sever them.*
5. *Renormalize the matrix by dividing each row by its sum; let $(r) \leftarrow (r)/\sum_{j=1}^{N}(r_j)$.*

We can interpret this as saying an Erdos-Renyi graph is a simple random walk on a graph with parameterized sparsity (given by p).

Warning. *Note that we are not considering the adjacency matrix of an Erdos-Renyi graph. Adjacency matrices are a commonly studied class of matrices that also represent graphs. However, we are simulating a transition matrix, which* ***represents a random walk on one****.*

Consider the code example below, where we generate an Erdos-Renyi ensemble of matrices.

Code Example (Erdos-Renyi p = 0.5 Ensemble)**.** *Let* $\mathcal{D} = ER(p = 0.5)$*. We can generate* $\mathcal{E} \sim \mathcal{D}$*, an ensemble of* 4×4 *Erdos-Renyi matrices* ($p = 0.5$) *of size 10 as such below. Notice how half of all the entries are* 0*, which is what we expect since* $p = 0.5$ *implies that we expect half the edges to not be connected (have weight 0).*

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

```
[[1]]
          [,1]      [,2]      [,3]      [,4]
[1,] 0.0000000 1.0000000 0.0000000 0.0000000
[2,] 0.0000000 0.5350731 0.4649269 0.0000000
[3,] 0.1287541 0.0000000 0.0000000 0.8712459
[4,] 0.1525212 0.0000000 0.0000000 0.8474788

...

[[10]]
          [,1]      [,2]      [,3]     [,4]
[1,] 0.0000000 0.1729581 0.8270419 0.000000
[2,] 0.0000000 0.0000000 1.0000000 0.000000
[3,] 0.2557890 0.3766740 0.0000000 0.367537
[4,] 0.2151029 0.3929580 0.3919391 0.000000
```

1.2.2 Hermite β-Ensembles

The Hermite β-Ensembles will be one of the primary ensembles discussed in this thesis. The Hermite β-ensembles are a normal-like class of random matrices. This ensemble will be characterized, motivated, and defined more thoroughly in **Chapter 4**. However, we will give a brief introduction to the ensemble.

As mentioned briefly in **Section 1.1.1**, the β-ensemble is an ensemble whose matrix model has a special unique generative algorithm. There is a reason we care about the β-matrices. This is because their original characterization is not in this matrix model with a non-homogeneous $\mathcal{D}$-distribution. Rather, it is a matrix model that is characterized by its joint density of eigenvalues!

That is, given a parameter $\beta \in \mathbb{N}$, the Hermite β-ensemble represents a class of matrices that have an explicit joint p.d.f of eigenvalues as a function of β. This is ultimately what characterizes the ensemble. Speaking of which, this leads in quite nicely to our discussion of our first spectral statistic: the spectrum of a matrix, which is concerned all about the distribution of eigenvalues.

Summary Table of $\mathcal{D}$-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 (H)
Uniform	Unif(a, b)	$a, b \in \mathbb{R}$	Explicit (H)
Hermite-β	$\mathcal{H}(\beta)$	$\beta \in \mathbb{N}$	Explicit (NH)
Stochastic	Stoch	-	Implicit
Erdos-p	ER(p)	$p \in [0, 1]$	Implicit

Chapter 2

Spectra

Life is like a box of crayons.

Unknown

2.1 Introduction

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 - fundamental 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 corresponding matrices. The study of eigenvalues and eigenvectors primarily falls in the scope of linear algebra, but their utility is far-reaching. So, what are *eigenvalues* exactly?

2.1.1 The Quintessential Spectral Statistic: the Eigenvalue

Given any standard square matrix $N \times N$ matrix P over some field $\mathbb{F}$, its *eigenvalues* are simply the roots of the characteristic polynomial, $\text{char}_P(\lambda) = \det(P - \lambda I)$. The characteristic polynomial is just like any other polynomial. What is important to know is that char_P will have a degree of N (i.e. equal to the dimension of the matrix). By the Fundamental Theorem of Algebra, this implies that there are always 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*

$$\sigma(P) = \{\lambda_i \in \mathbb{C} \mid char_P(\lambda_i) = 0\}_{i=1}^{N}$$

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.

For example, consider the following code example from the **RMAT** package.

Code Example (Spectrum of a Standard Normal Matrix)**.** *Let $P \sim \mathcal{N}(0,1)$ be a 5×5 standard normal random matrix. We can generate the spectrum of P, $\sigma(P)$ as follows:*

```
library(RMAT)
P <- RM_norm(N = 5, mean = 0, sd = 1)
spectrum_P <- spectrum(P)
# Outputs the following
spectrum_P
...
        Re       Im    Norm Order
 1 -0.5434   1.3539 1.4589     1
 2 -0.5434  -1.3539 1.4589     2
 3  0.2255   1.4250 1.4427     3
 4  0.2255  -1.4250 1.4427     4
 5 -0.8678   0.0000 0.8678     5
```

Notice, in the table above, we obtain the spectrum of the matrix P as returned by the spectrum wrapper function in **RMAT**. In this tidy[1] table, each row represents one eigenvalue and the output is composed of four columns. The first two columns represent the real and imaginary components of the eigenvalue, respectively. The third column denotes its norm. Lastly, the fourth column denotes the eigenvalue's relative order within the matrix's spectrum; the order is with respect to the default order scheme the function chooses, which is by norm. Order schemes will be discussed in more depth in **Section 2.2.1**.

[1] In reference to the tidyverse principles of tidy data frames.

Eigenvalues as Statistics: A Formalization

While our definition for spectrum is nice and clean, there are a few caveats that we must take care of. First, how are spectra considered statistics? Simply put, we know that a statistic is some function of random data. However, before we can even call them so, there needs to be some formalization as to justify considering eigenvalues statistics. In the upcoming dialogue, we will reminisce on the corresponding formalization dialogue in **Section 1.1.3** when we formalized random matrices. Before beginning, the reader is encouraged to review what a statistic is in **Appendix A.2.2**.

Without further ado, recall that when we defined and motivated the $\mathcal{D}$-distribution framework of simulating random matrices, there was always one common factor - a representation using vectors of random variables. Using this framework, the formalization of eigenvalue spectra as statistics is not too difficult.

Formalization. *Suppose $P \sim \mathcal{D}$ is an $N \times N$ random matrix. Then, P has a representation as a sequence of K random variables (for some $K \leq N^2$). Denote it as the vector $\vec{X} = X_1, X_2, \ldots, X_K$. 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 σ is overloaded to mean the spectrum of a matrix **with respect to the vector representation**. In that case, σ is a function that would parse the random variables into the appropriate matrix form by index hacking. Then, using the matrix, the eigenvalues are computed by finding the roots of the characteristic polynomial (or some other algorithm). Sparing the details, this characterization of σ is sufficient to show that the eigenvalues are in one way or another a function of random variables – or in other words, a statistic. To summarize this, consider the flow chart below.*

Note how in the formalization we said that an $N \times N$ matrix P has a representation as a vector of K random variables rather than N^2 variables. Recall that some non-homogenous explicit $\mathcal{D}$-distributions do not initialize every entry of the matrix. Only our homogenous explicit and implicit $\mathcal{D}$-distributions do. That being said, to summarize, here is how we formalize the spectrum as a statistic. Suppose we sample $P \sim \mathcal{D}$. Then, we take its spectrum formally using σ as such:

$$X_1, X_2, \ldots, X_K \xrightarrow{\text{make array}} (\vec{X} \rightsquigarrow P) \xrightarrow{\det(P-\lambda I)} char_P(\lambda) \xrightarrow{\text{solve for roots}} \sigma(P)$$

Remark (Computation)**.** *While eigenvalues are theoretically defined as the roots of the characteristic polynomial, they are not computed using them in practice. This is intractable as N grows; for an 1000×1000 matrix, we would need to solve for one thousand roots. Needless to say, that is unviable. To simulate our random spectra, the **RMAT** package relies on the `eigen()` function in base R. The function relies on techniques such as the QR algorithm and power iteration. [Kublanovskaya (1962)] However, this is not our subject matter, and this does not impact our formalization at all. Whether the eigenvalues are computed by solving for roots or by the QR algorithm, they will always remain a function of random variables — a statistic.*

2.1.2 Interlude: Ensembles

While the spectrum of a matrix provides a good summary of the matrix in and of itself, 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 *spectrum of a random matrix ensemble*. If we have an ensemble $\mathcal{E}$, then we can naturally extend the definition of $\sigma(\mathcal{E})$.

Definition 2.1.2 (Ensemble Spectrum)**.** *Let $\mathcal{E} \sim \mathcal{D}$ be an ensemble of matrices $P_i \in \mathbb{F}^{n \times n}$. To take the spectrum of $\mathcal{E}$, simply take the union of the spectra of each of its matrices. In other words, if $\mathcal{E} = \{P_i \sim \mathcal{D}\}_{i=1}^{K}$, then we denote the spectrum of the ensemble*

$$\sigma(\mathcal{E}) = \bigcup_{i=1}^{K} \sigma(P_i)$$

For example, consider the following code example from the **RMAT** package.

Code Example (Spectrum of a Standard Normal Matrix Ensemble)**.** *Let $\mathcal{E} \sim \mathcal{N}(0, 1)$ be an ensemble of 3×3 standard normal random matrices of size 3. We can generate the spectrum of $\mathcal{E}$, $\sigma(\mathcal{E})$ as follows:*

```
library(RMAT)
ens <- RME_norm(N = 3, mean = 0, sd = 1, size = 3)
spectrum_ens <- spectrum(ens)
# Outputs the following
spectrum_ens
...
```

```
        Re       Im    Norm Order
 1  1.7581   0.0000 1.7581     1
 2 -0.2614   1.0012 1.0347     2
 3 -0.2614  -1.0012 1.0347     3
 4  1.2327   0.4227 1.3032     1
 5  1.2327  -0.4227 1.3032     2
 6 -0.8504   0.0000 0.8504     3
 7 -0.5296   1.0508 1.1767     1
 8 -0.5296  -1.0508 1.1767     2
 9  0.7357   0.0000 0.7357     3
```

Notice, in the array above, we have a table similar to that of a spectrum of a random matrix (like in the code example above). However, we now have multiple values at a given order since we are taking the union of all the spectra. For this reason, ensembles can tell us more about the bounds of the eigenvalues by observing the largest and smallest eigenvalues.

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.

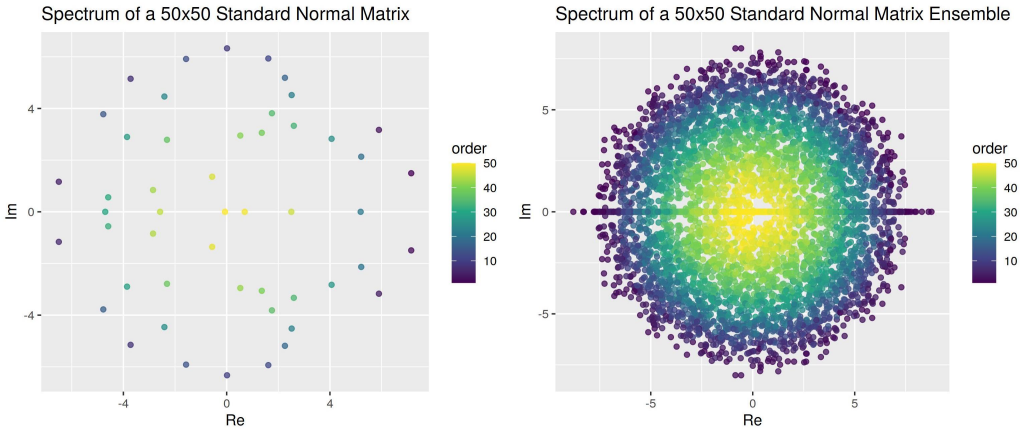


Figure 2.1: Spectrum of a Matrix versus an Ensemble

From the plot above, we can see that the spectrum of the ensemble is much more populated. In both plots, we are simulating the spectrum of the same object: a standard normal 50×50 matrix. For this reason, the Order legends on the right-hand side are equivalent across the plots. However, since we are only considering the spectrum of a singleton matrix on the left-hand side, it means that every Order corresponds to exactly one point. On the right-hand side, each Order is populated with K points (where K is the size of the ensemble). As such, from just taking numerous iterations of the matrix, we are able to "complete the picture" and see that the eigenvalues seem to be uniformly distributed on a complex disk!

2.2 Spectrum Analysis

2.2.1 Ordered Spectra

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 σ that indicates how the spectrum is ordered. Often, the ordering context will be clear and the indexing will be omitted.

Order Schemes: How to Order Eigenvalues

In this thesis, we will discuss three order schemes. However, when it comes to practice, we will most often use the two primary order schemes listed below.

In the relevant literature, researchers often use the standard ordering on $\mathbb{R}$ to define an ordered spectrum. We denote this as the ordering by the **sign scheme**. However, because total-ordering is only well-defined on $\mathbb{R}$, we can only use this scheme when none of our eigenvalues lives in $\mathbb{C}$. 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 could sort the spectrum by the norm of its entries; denote this as ordering using the **norm scheme**. We are forced to use this scheme if some of our eigenvalues live in $\mathbb{C}$, due to the reasons mentioned above. By taking their norms, each eigenvalue is mapped to a real value. From which, we could then sort them using the regular ordering on $\mathbb{R}$. Essentially, we are simply using the sign ordering scheme on the norms of the eigenvalues. 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.

That being said, consider the following plots, showing the difference in using the sign and norm ordering schemes for the same spectrum.

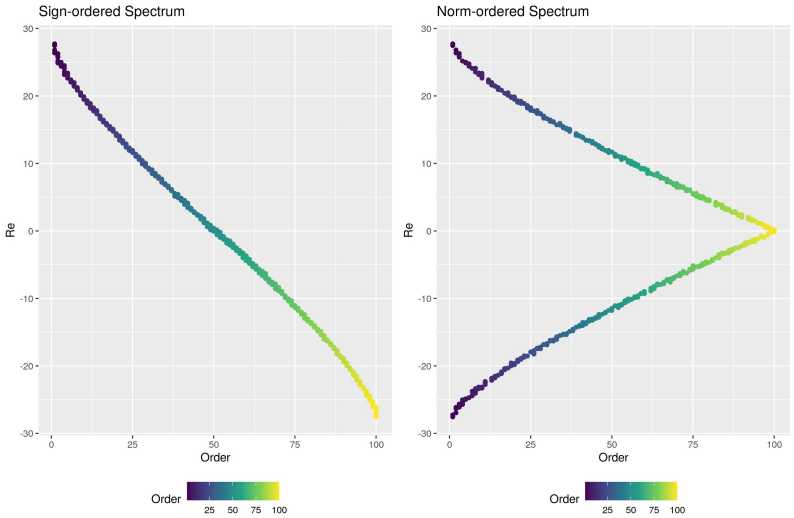


Figure 2.2: Spectrum of an Ensemble Using Two Different Ordering Schemes

Singular Values

Our last ordering scheme is moreso a different method of defining the spectrum of a matrix than one of sorting the eigenvalues. However, for simplicity, we will call it an ordering scheme. Suppose P is a random matrix. Then, we can take its 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 corresponding product of that matrix and its transpose. That is,*

$$\sigma_+(P) = \sqrt{\sigma(P \cdot P^T)}$$

Because of the way they are defined, singular values are always non-negative, even if the eigenvalues are negative. As a result, unlike the eigenvalues, singular values in a sense "ignore" rotational features of the matrix. Instead, they only tell us about the "scaling" features of the matrix.

There are a few reasons why singular values are important. One enormous advantage of using singular values is that it opens the door for studying spectral statistics for **non-square** matrices. This is because any matrix $Q \in \mathbb{F}^{n\times m}$ multiplied by its transpose $Q^T \in \mathbb{F}^{m\times n}$ yields a square matrix $S = QQ^T \in \mathbb{F}^{n\times n}$.

Additionally, it is worth noting that because of the way singular values are defined, taking the singular values of a symmetric matrix is equivalent to taking its spectrum with respect to the norm ordering scheme. This follows because if a matrix is symmetric, then its singular values are simply the norm of the eigenvalues.

Summary Table of Spectrum Schema

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

2.2.2 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 **Appendix A.2.2**)

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 correspond to the largest eigenvalue of a spectrum. After all, a spectrum is a random sample of fixed size, so this statistic is well-defined.

Remark (Indices)**.** *It is a common convention when notating ordering statistics to say that $X_i > X_j$ when $i > j$. However, in this thesis, we will oppose this convention. This is because we are often interested in notating the largest eigenvalue λ_1 rather than λ_N where N is the dimension of the matrix. So, to make the larger eigenvalues intrinsic (independent of N), we use the opposite convention and say that $X_i < X_j$ when $i > j$*

That being said, then we can notate the value λ_1 the largest eigenvalue in the spectrum, for instance. Given an $N \times N$ matrix P, the eigenvalue λ_N represents the smallest eigenvalue.

Warning (Order Schema & Indices)**.** *Notice how when we are discussing summary statistics below, we make sure to say value or size (and generalize by saying quantity) because the i^{th} rank has a different interpretation under different order schema. In the norm-ordered scheme, λ_1 means the largest eigenvalue whereas in the sign-ordered scheme, λ_1 means the "most positive" eigenvalue.*

All that being said, one framework of studying order statistics will be conditioning on their values. So, we will consider the following summary statistics that condition on the order statistics.

Expectation. $\mathbb{E}(\lambda_i)$ One useful summary statistic to consider when analyzing a spectrum is the expected norm or value (which we will just call quantity hereinafter) of the eigenvalue at the i^{th} rank. When considered at the ensemble level, the mean quantity order statistic can tell us a lot about the bounds on the eigenvalue sizes or values for a given $\mathcal{D}$-distribution. For instance, given the norm-ordered scheme the expected extreme values (assuming $N \times N$ matrices), given by $\mathbb{E}(\lambda_1)$ and $\mathbb{E}(\lambda_N)$ respectively can tell us the expected range our spectrum might take.

Variance. $\mathrm{Var}(\lambda_i)$ Similarly, the variance of the eigenvalue quantity at a given order i can tell us a lot about an ensemble. For example, we may observe that variances are heteroskedastic (not equal in every level) across all the levels i. This insight can tell us how a $\mathcal{D}$-distribution disperses its eigenvalues or perhaps how the eigenvalues "repel" each other.

2.2.3 Case Study: Perron-Frobenius Theorem

Definition 2.2.2 (Spectral Radius)**.** *Let P be any matrix and $\sigma(P)$ be its ordered spectrum. Then, the spectral radius of P is defined as $\rho(P) = ||\sup \sigma(P)||$ is the norm of its largest eigenvalue.*

Detour: The Perron-Frobenius Theorem

Using the toolkit we have now acquired, we can now discuss an elegant, visual representation of the Perron-Frobenius theorem (see **Section A.1.2**). To put it shortly, the Perron-Frobenius theorem, applied to stochastic matrices is a result that guarantees the existence of a stationary distribution to an ergodic Markov Chain (see **Section A.3**). That being said, the following facts give us a heuristic demonstration of the Perron-Frobenius theorem.

Theorem 2.2.1 (Perron-Frobenius Theorem)**.**
Consider the two following facts:

1. *The largest eigenvalue of stochastic matrices is 1.*
2. *When multiplied by P^K, any point v asymptotically enters the eigenspace of the matrix's largest eigenvalue as $K \to \infty$. That is, as K grows, vP^K approaches an eigenvector of P of λ_1.*

So, there is an eigenvector of the largest eigenvalue – for an irreducible Markov Chain, this is a stationary distribution.

Note. *Note that this is only part of the results of the theorem. The theorem is more general and has a wider scope and applies to matrices in a more general fashion. We only demonstrate this because we have a case with a constant largest eigenvalue and an eigenvector with a unique interpretation (a stationary distribution).*

Again, those two statements were not formally proven in this thesis. However, there are large amounts of computational evidence for both of these results.

Consider the first statement. Below we have a stochastic ensemble spectrum. Notice how the largest eigenvalue is 1, living far from the island of **complex** eigenvalues.

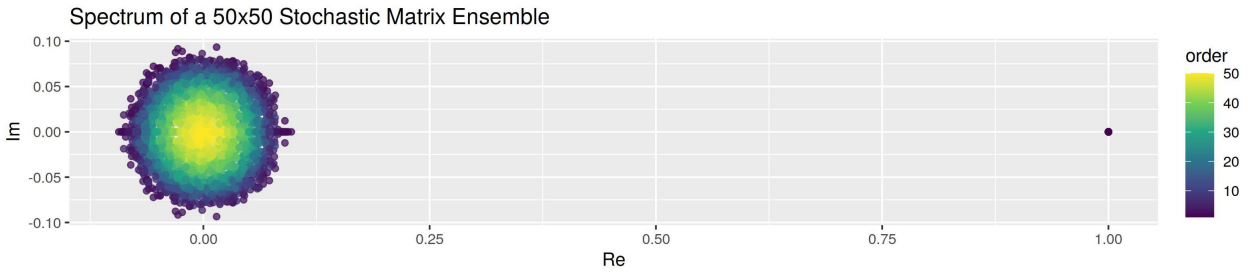


Figure 2.3: Spectrum of a Stochastic Matrix ensemble

This shows that the spectral radius of stochastic matrices is 1! What is left is the other fact. For the second statement, again, there is plenty of computational evidence to back this up. The reader may refer to **Appendix D** for an empirical demonstration of fact 2. Because the discussion of eigenvectors is too far afield from the current topic, so we will not include it in the current section.

From the simulations, we find that there is overwhelming evidence that most matrices, if not all, satisfy the second statement. The "almost all" part here is an artifact of random matrices, since it is unlikely they will have duplicate eigenvalues, or eigenvalues corresponding to pure rotations (with no scaling). These are the types of eigenvalues that lead to the second statement not holding.

Conclusions. So, to conclude, we can say that the computational evidence for these two facts supports and provides an alternative method of empirically demonstrating that the Perron-Frobenius theorem is true.

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 **Appendix A.1.1**). Simply put, those are matrices which are equal to their conjugate transpose.

Remark (Symmetric versus Hermitian)**.** *Since real numbers are their own conjugate transpose, every symmetric matrix is hermitian. However, we will still delineate the two terms to avoid confusion and indicate what field* $\mathbb{F}$ *we are working with.*

In any case, one critical result in linear algebra that will be extensively wielded in this thesis is the fact that if a matrix is symmetric or hermitian, then it has real eigenvalues [Horn (2012)]. In other words:

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

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.

Example: Stochastic Matrices

One very pleasing example to look at is stochastic matrices. As seen previously in **Section 2.2.2**, stochastic matrices tend to have two components: a complex disk of eigenvalues about the origin and isolated point that is the largest eigenvalue.

Below we have a **symmetric** stochastic ensemble spectrum. Notice how the largest eigenvalue is 1, living far from the island of **real** eigenvalues. This can be compared to the spectrum of non-symmetric stochastic matrices in the previous figure. The patterns are similar, but now, imposing symmetry on the matrices forces the eigenvalues to "collapse" onto the real line!

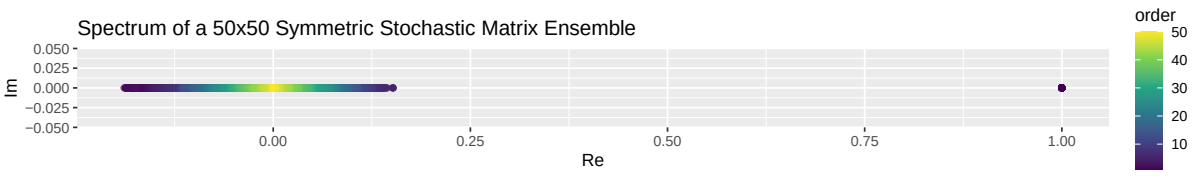


Figure 2.4: Spectrum of a Symmetric Stochastic Matrix ensemble

2.3.2 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. The distribution is named after physicist Eugene Wigner.

Definition 2.3.1 (Semicircle Distribution)**.** *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:*

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

As it turns out, the dimension of the matrix (N) impacts the radius of the eigenvalue distribution [Tao (2012)].

Remark (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 approximately semicircle distributed with radius $R = 2\sqrt{N}$; the approximation improves as N gets larger. That is, $P^\dagger$ has a spectrum $\sigma(P) \sim SC(R = 2\sqrt{N})$ as $N \to \infty$.*

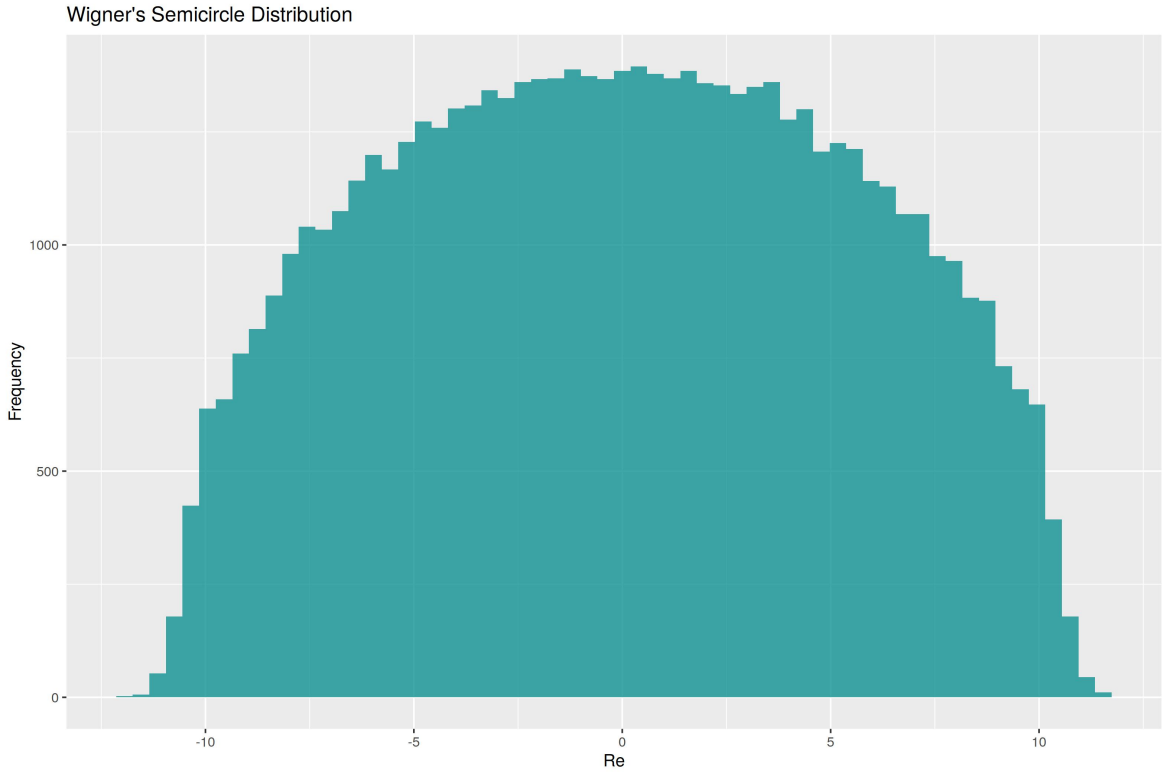


Figure 2.5: Eigenvalues of a Symmetric Matrix displaying the Semicircle Distribution

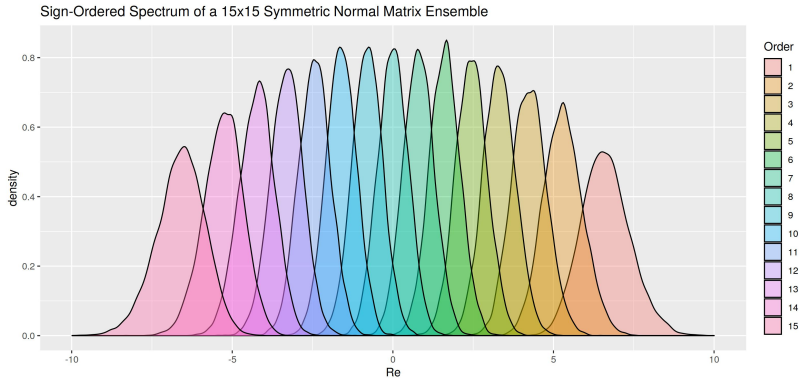
2.3.3 Spectra

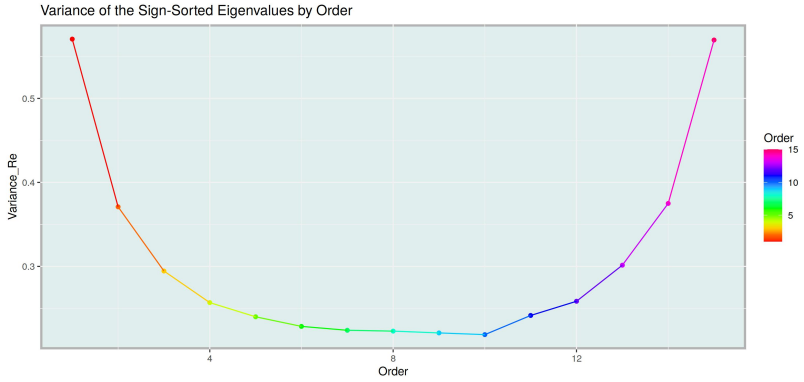
In this section, we will be carefully analyzing an ensemble of symmetric matrices to showcase the special properties of symmetric and hermitian matrices. Namely, we will be considering an ensemble $\mathcal{E} \sim \mathcal{N}(0,1)^{\dagger}$ of 15×15 matrices over $\mathbb{R}$.

Sign-Ordered

Consider the plot of the eigenvalues in our sign-ordered spectrum of our ensemble $\mathcal{E}$ below. To understand why our eigenvalues assume this shape of distribution, we need to recall that symmetric matrices (approximately) obey the Semicircle distribution. What we are observing is the "echo" of that being the underlying distribution.

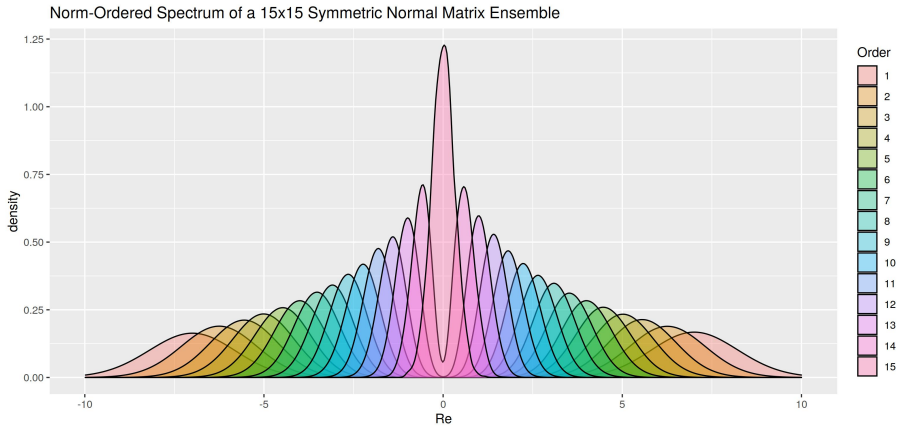
What is more interesting, on the other hand, is the peaked variances we observe at the boundaries in the variance plot. Since this is the sign-ordered scheme, the boundaries correspond to the largest eigenvalues. In fact, if we consider symmetry about the origin, we could deduce a very clear trend that as the eigenvalue gets larger (closer to the semicircle boundary), the more variance we expect to observe in the quantities there. As such, we now switch to the *norm-ordering scheme* for a different perspective.

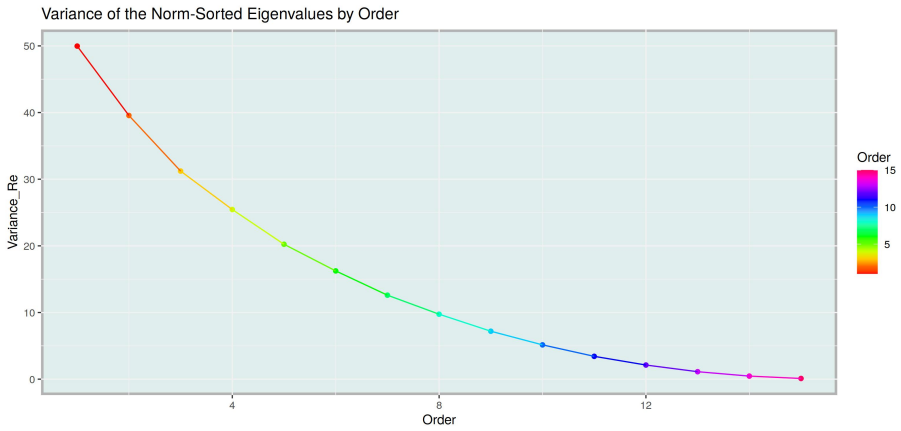




Norm-Ordered

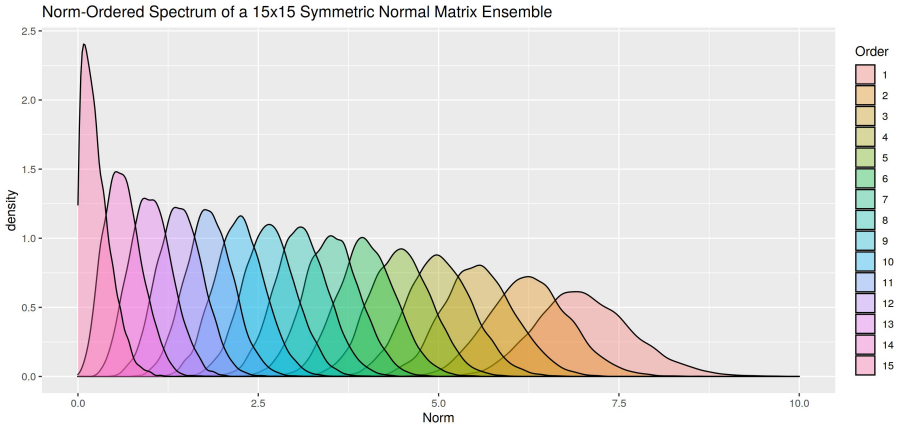
Consider the plot of the **same** eigenvalues in our norm-ordered spectrum of our ensemble $\mathcal{E}$ below. Now, consider the norm component of our norm-ordered spectrum of our ensemble $\mathcal{E}$ below. One of the very first things to note about the distribution of the eigenvalues is that when conditioning on its order, we see a very clear trend in the variance of the eigenvalue. Namely, we notice that as i grows, $\text{Var}(\lambda_i) \to 0$. In simple words, this means that the smallest eigenvalues have less "freedom" in their distribution. This is opposed to the largest eigenvalue, which has the highest variance or the most "freedom".

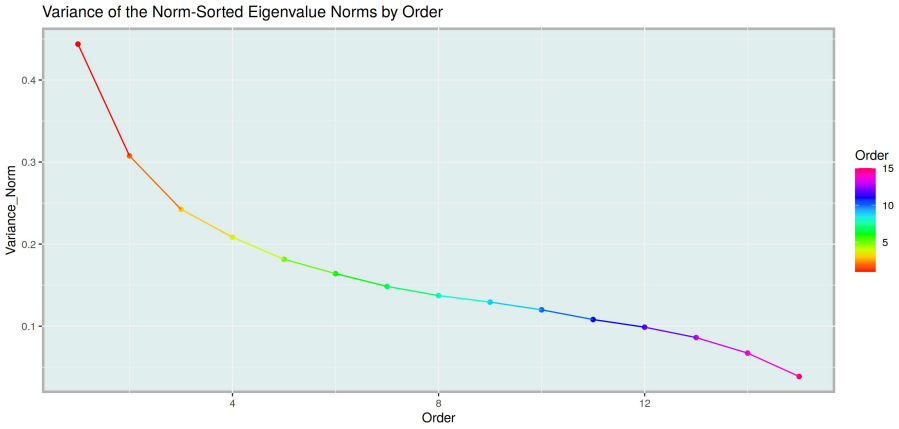




Lastly, consider the distribution of the **eigenvalue norms**. As mentioned previously, the variance seems to increase closer to the boundaries. Since we are now using the norm-ordering scheme, the fact that the distribution is symmetric means we are observing twice as many "occupants" in a statistical observation. Imagine this as "folding the semicircle in half". This explains why the variance is "increasing" more rapidly as our eigenvalues shrink.

That being said, the variance plot shows that the variance of the eigenvalue norms seems to have an interesting, non-linear shape. Namely, we observe a rapidly decreasing variance initially, a levelling off, and then a slight dip to the minimum variance. We could trace this back as an artifact of the c.d.f of the semicircle distribution. Since the p.d.f. is the derivative of the c.d.f., we know that the flattest part of the semicircle occurs closer to the origin. As such, we observe the most "stability" there.



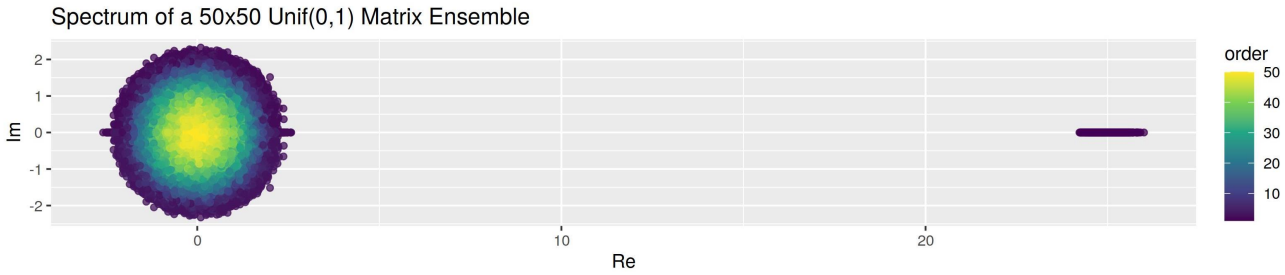


2.4 A Survey of Spectra

In this section, we will briefly survey the spectra for a variety of $\mathcal{D}$-distributions and characterize their properties.

2.4.1 Uniform Ensembles

Here we have an ensemble of $\mathcal{E} \sim \text{Unif}(0, 1)$ matrices. We see a resemblance to stochastic matrices.



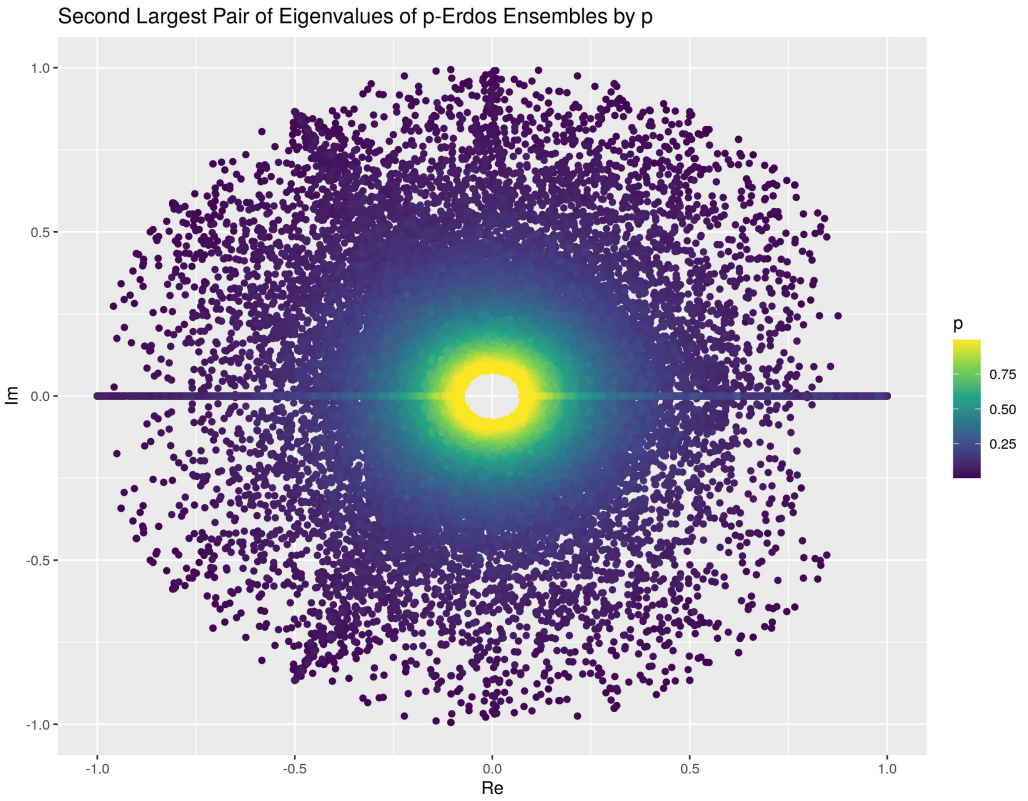
The first thing that stands out about the spectrum of the $\text{Unif}(0, 1)$ ensemble spectrum is that it very much resembles the spectrum of the stochastic matrix ensemble! While surprising at first, it starts to make sense once we consider what the Stochastic $\mathcal{D}$-distribution fundamentally represents. Namely, it represents a matrix with normalized, uniformly random weights. The algorithm for generating the Stochastic matrices itself samples from the $\text{Unif}(0, 1)$ distribution. Additionally, the general shape of the spectrum of the uniform ensemble is similar to that of the stochastic matrix ensemble. This is in the sense that they both have two "islands" of eigenvalues, one complex disk, and one "island" of real eigenvalues.

However, despite these similarities, there is a notable difference between the two $\mathcal{D}$-distributions. This lies in the distribution of the largest eigenvalues which live in the aforementioned "real island". For the uniform ensemble, the largest eigenvalues have **non-zero variance**, we can see them being scattered around roughly the same place. On the other hand, stochastic matrices have a constant largest eigenvalue of one, meaning the largest eigenvalue has **zero variance**.

2.4.2 Erdos p-Ensembles

For this section, we will simulate an Erdos-Renyi ensemble for various values of $p \in [0, 1]$ and find the second largest pair of eigenvalues. That is, we will simulate the distribution of $\lambda_2, \lambda_3 \in \sigma(\mathcal{E})$ where $\mathcal{E} \sim \text{ER}(p)$ for many values of p in hopes of finding a trend. The reason we are doing this is to find the radius of the complex disk portion of the Stochastic matrix ensemble but for matrices with parameterized sparsity.

Remark (Conjugate Pairs)**.** *The reader might ask, if we are interested in the radius of the complex disk, why is the distribution of λ_2 not sufficient? Well, the issue lies in the fact that eigenvalues quite often come in conjugate pairs. For this reason, R must arbitrarily select which eigenvalue in a conjugate pair to select, and it selects the values with a positive imaginary component. So, to generate the outer rim of the complex disk, we will need to find the* ***second largest pair*** *of eigenvalues to capture any conjugate pairs.*



At first glance, we can see a resounding pattern in the plot. It seems like the value of p is inversely related to the radius of the complex disk. So, we conclude that the more connected a graph is, the smaller is the magnitude of the second largest eigenvalue. On the ensemble level, this means that the more connected an ensemble of graphs is, the smaller will be the radius of the complex disk. Additionally, it is worth noting that the values of p close to 0 have **higher variance** roughly speaking.

Also, the maximum radius of the eigenvalues is less than 1, which makes sense since the largest eigenvalue must be 1 for an arbitrary stochastic matrix. As a result, this also demonstrates that the second largest eigenvalue is bound by the **unit complex disk**.

2.4.3 Normal Ensembles

We have observed in Figure 2.1 in **Section 2.1.2** that a standard normal matrix ensemble tends to have a spectrum that is uniformly distributed about a complex disk. This happens to be the case for both real-valued and complex-valued standard normal matrices. Similarly, the result regarding symmetric and hermitian matrices in **Section 2.3.1** also applies to complex-valued standard normal matrix ensembles. Consider the plot below.

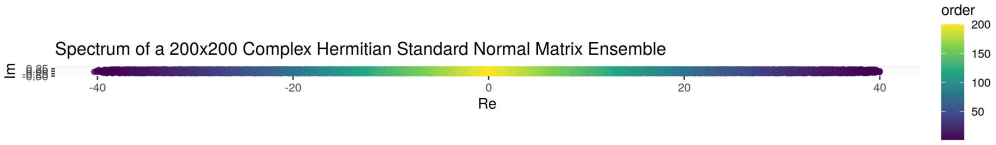


Figure 2.6: Spectrum of a Complex Hermitian Standard Normal Matrix ensemble

As we can see, despite having purely complex entries, this ensemble of Hermitian standard normal matrices over $\mathbb{C}$ has a spectrum of real eigenvalues! However, we do observe a slight imaginary component.

Remark (Floating Point Errors)**.** *The* `eigen()` *function used to complete eigenvalues in* ***RMAT*** *uses a numerical algorithm as opposed to solving for the roots of the characteristic polynomial. For this reason, floating point errors and algorithmic systematic error will yield small, but negligible imaginary components.*

Chapter 3

Dispersions

> Distance doesn't exist, in fact, and neither does time. Vibrations from love or music can be felt everywhere, at all times.
>
> *Yoko Ono*

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 distribution 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. So, to begin, we first formally define an eigenvalue pair.

Definition 3.1.1 (Eigenvalue Pair)**.** *Suppose P is a matrix and $\sigma(P)$ is its ordered spectrum. Then, an eigenvalue pair with respect to this ordered spectrum is denoted π_{ij}. It is defined as the ordered pair $\pi_{ij} = (\lambda_i, \lambda_j)$.*

Consecutive Pairs

In the future section where we define pairing schemes, we will introduce a new notation of eigenvalue pairs with one index when defining the consecutive pairing sch Π_C, and it takes the form $\tilde{\pi}_j$. This notation is used to denote the consecutive eigenvalue pair for a given matrix. The consecutive eigenvalue pairs are so special that

we denote them with a unique notation for convenience. It also makes the discussion regarding order statistics more intrinsic.

Definition 3.1.2 (Consecutive Pairs)**.** *Suppose P is a matrix and $\sigma(P)$ is its ordered spectrum. Then, let $\tilde{\pi}_j$ denote a pair of consecutive eigenvalues, the largest of the two being the j^{th} largest eigenvalue. So, $\tilde{\pi}_j = (\lambda_{j-1}, \lambda_j)$.*

Since the consecutive pair scheme can be sufficiently indexed by one index (j), we will use the convention of omitting the index of the smaller eigenvalue to be more concise and idiomatic. So, whenever one sees the notation $\tilde{\pi}_j$, one should think of the j^{th} largest eigenvalue and its smaller neighbour.

Example (The Largest Eigenvalues)**.** *With this notation at hand, we say that $\tilde{\pi}_1$ represents the pair of the two largest eigenvalues in an ordered spectrum.*

$$\tilde{\pi}_1 = (\lambda_2, \lambda_1)$$

Without further ado, we now 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. In simple words, a dispersion metric is a function that takes in a pair of eigenvalues and returns a positive real number that represents some notion of dispersion or spacing.

Definition 3.1.3 (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.*

In the scope of this thesis, we consider the following dispersion metrics below.

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

Remark (Symmetric Metrics)**.** *Note that the standard norm and the β-norm are symmetric operations, compared to the difference of absolutes metric, which is not. For a metric to be symmetric means that switching the order of arguments with respect to that metric will have no effect on the function's output. So, a metric δ^* is symmetric if it satisfies:*

$$\delta^*(\pi_{ij}) = \delta^*(\pi_{ji})$$

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

Remark (Identity Difference Heuristic)**.** *Suppose we take the arithmetic difference of two complex numbers. Then, the range of δ 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 heuristic. As such, we will denote this as the "identity difference" heuristic and call it δ_{id}. So, we define $\delta_{id} : (z, z') \mapsto z' - z$*

Summary Table of Dispersion Metrics

For every dispersion metric, assume the functions' order of arguments is $\delta(z, z')$.

Table of Dispersion Metrics				
Metric*	Notation	Formula	Symmetric	Parameters
Standard Norm	δ_n	$\lvert z' - z \rvert$	True	-
β-Norm	δ_β	$\lvert z' - z \rvert^\beta$	True	$\beta \in \mathbb{N}$
Difference of Absolutes	δ_{abs}	$\lvert z' \rvert - \lvert z \rvert$	False	-
Identity Difference	δ_{id}	$z' - z$	False	-

*Note that the identity difference is a heuristic, 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.

Before we may begin talking about pairing schema we define an auxiliary object, the spectral pairs of a matrix, which we denote $\sigma^{(2)}$. Since we are now talking about eigenvalue pairs, it is helpful to define this object before proceeding.

Definition 3.1.4 (Spectral Pairs)**.** *Suppose P is an $N \times N$ random matrix. Then, taking the spectral pair of P, denoted $\sigma^{(2)}(P)$ is equivalent to taking the Cartesian product of its ordered spectrum $\sigma(P)$. That is, $\sigma^{(2)}(P) = \sigma(P) \times \sigma(P) = \{\pi_{ij} = (\lambda_i, \lambda_j) \mid i, j \in \mathbb{N}_N\}$.*

Now, to select eigenvalue pairs, we motivate the pairing scheme - this is what tells us which indices to select.

Definition 3.1.5 (Pairing Scheme)**.** *Suppose P is any $N \times N$ matrix and $\sigma^{(2)}(P)$ are its spectral pairs. A pairing scheme for the matrix P is 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 a set $\Pi = \{(\alpha, \beta) \mid \alpha, \beta \in \mathbb{N}_N\} \subseteq \mathbb{N}_N \times \mathbb{N}_N$. To take a matrix's spectral pairs with respect to Π, we simply take the set of eigenvalue pairs with the matching indices, $\sigma^{(2)}(P \mid \Pi) = \{(\lambda_\alpha, \lambda_\beta) \mid (\alpha, \beta) \in \Pi\}$.*

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 (Proper Subset)**.** *If $\sigma^{(2)}(P)$ is the spectral pairs of the matrix P, then for any pairing scheme Π, we will find that $\sigma^{(2)}(P \mid \Pi) \subseteq \sigma^{(2)}(P)$. By definition, taking a spectral pair with respect to some pairing scheme constricts which pairs to select - meaning it is a proper subset of the general spectral pairs.*

Selected Pairing Schema

Suppose P in an $N \times N$ square matrix, and $\sigma^{(2)}(P)$ are its spectral pairs. For every pairing scheme, assume $i, j \in \mathbb{N}_N$.

1. Let Π_C be the consecutive pairs of eigenvalues in a spectrum.

$$\sigma^{(2)}(P \mid \Pi_C) = \{\tilde{\pi}_j = (\lambda_{j+1}, \lambda_j)\}_{j=1}^{N-1}$$

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

2. 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 eigenvalues without repeats. They are named in allusion to indexing upper and lower triangular matrices.

   (a) Let $\Pi_>$ be the lower-pair combinations of ordered eigenvalues. This will be the preferred unique pair combination scheme used in lieu of the argument orders because our dispersion metrics expect the eigenvalue with the lower rank first and the higher rank second.

$$\sigma^{(2)}(P \mid \Pi_>) = \{\pi_{ij} = (\lambda_i, \lambda_j) \mid i > j\}_{i=1}^{N-1}$$

   (b) For completeness, we will also define the upper-pair scheme. $\Pi_<$ is the set of upper-pair unique combinations of ordered eigenvalues. However, it won't be used because we want dispersions to be positive-definite.

$$\sigma^{(2)}(P \mid \Pi_<) = \{\pi_{ij} = (\lambda_i, \lambda_j) \mid i < j\}_{i=1}^{N-1}$$

   Benefits: Solves the issue of repeated pairs for symmetric dispersion metrics. Also, represents the set of pairs obtained from randomly sampling two distinct eigenvalues.

3. Let Π_0 be all the pairs in a spectrum. For completeness, we define this pairing scheme as the implicit pairing scheme for the spectral pairs of a matrix.

$$\sigma^{(2)}(P \mid \Pi_0) = \sigma^{(2)}(P) = \{\pi_{ij} = (\lambda_i, \lambda_j) \mid i, j \in \mathbb{N}_N\}$$

Consider some examples of generating pairing schemes in R. These methods are not exported in the **RMAT** package.

Code Example (Consecutive Pairing Scheme)**.** *We generate the pairing scheme (indices) for what would be a* 5×5 *matrix.*

```
# Helper function in the source code
pair_indices <- .consecutive_pairs(N = 5)
# Outputs the following
pair_indices
...
     i j
[1,] 2 1
[2,] 3 2
[3,] 4 3
[4,] 5 4
```

Code Example (Lower Pairing Scheme)**.** *We generate the pairing scheme (indices) for what would be a* 4×4 *matrix.*

```
# Helper function in the source code
pair_indices <- .unique_pairs_lower(N = 4)
# Outputs the following
pair_indices
...
     i j
[1,] 2 1
[2,] 3 1
[3,] 3 2
[4,] 4 1
[5,] 4 2
[6,] 4 3
```

Think of the pairing schemes as a conditional argument when considering dispersions. You ask yourself: which pairs do I want to look at? There will be a discussion in **Section 3.2** about why we have to formalize the notions of dispersion metrics and pairing schemes when we eventually talk about dispersions.

Additionally, making the pairing schemes independent from the matrices (taking only their dimension, $N \in \mathbb{N}$) was a difficult decision. It may cloud understanding initially but the payoff comes if you consider the implementation of the functions in **RMAT**. As a whole, you should not be bothered by the details of formalizing the pairing schemes but you should definitely understand what they do. In fact, it's totally fair that you do. The **RMAT** implementation abstracts this all away in the dispersion function. The user would input a string denoting which pairing scheme and the function would handle it. Similarly, you should do the same!

Summary Table of Pairing Schema

Suppose P is an $N \times N$ matrix. Then, its spectral pairs $\sigma^{(2)}(P)$ may take on the following pairing schemes. Assume that $i, j \in \mathbb{N}_N$.

Table of Pairing Schema		
Scheme	Notation	Formula
Lower	$\Pi_<$	$\{(i,j) \mid i < j\}$
Upper	$\Pi_>$	$\{(i,j) \mid i > j\}$
Consecutive	Π_C	$\{(i,j) \mid i = j+1\}$
All	Π_0	$\{(i,j)\}$

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

An Alternative Representation: Eigenvalue Matrix

Definition 3.1.6 (Eigenvalue Matrix)**.** *Suppose P is an $N \times N$ square matrix and $\sigma^{(2)}(P)$ are its spectral pairs. Then, the eigenvalue matrix of P, given by $\Lambda(P)$ is the matrix with entries $\pi_{ij} = (\lambda_i, \lambda_j)$ for $\pi_{ij} \in \sigma^{(2)}(P)$. Again, it is given that the eigenvalues λ_i come from some **ordered** spectrum $\sigma(P)$ as per the definition of spectral pairs.*

With the matrix analogy, describing the pairing schemes can be much simpler. For instance, the **lower pairs** are given by the indices of the entries in the lower triangle of the matrix. Analogously, the **upper pairs** by those of the upper triangle of the matrix. The **consecutive pairs**, given our default lower pair scheme, are given by the lower main off-diagonal band of the matrix.

Example (Eigenvalue Matrix for a 5×5 Matrix)**.** *Suppose we have a 5×5 matrix P and its corresponding spectral pairs $\sigma^{(2)}(P)$. Then, its eigenvalue matrix has the following structure:*

$$\begin{bmatrix} - & \pi_{12} & \pi_{13} & \pi_{14} & \pi_{15} \\ \tilde{\pi}_1 & - & \pi_{23} & \pi_{24} & \pi_{25} \\ \pi_{31} & \tilde{\pi}_2 & - & \pi_{34} & \pi_{35} \\ \pi_{41} & \pi_{42} & \tilde{\pi}_3 & - & \pi_{45} \\ \pi_{51} & \pi_{52} & \pi_{53} & \tilde{\pi}_4 & - \end{bmatrix}$$

Remark (Main Diagonal)**.** *Note that in the matrix example above, we omit the entries in the main diagonal. We do so because those pairs given by $\{\pi_{ii} \mid i \in \mathbb{N}_N\}$ are redundant to include. We are only interested in comparing eigenvalues with other eigenvalues. This is because every the dispersion of eigenvalue with itself is zero, meaning $\forall \delta : \delta(\pi_{ii}) = 0$. So, as a convention, we omit these pairs.*

3.1.3 Dispersions

Now, with dispersion metrics and pairing schemes defined, we are finally able to motivate the definition of a matrix dispersion for both singleton matrices and their ensemble counterparts. In simple words, we define a dispersion of a matrix as a function that takes in a matrix along with a pairing scheme and dispersion metric. Then, it outputs the mapping of that dispersion metric onto the subset of spectral pairs specified by the pairing scheme. To formalize this, consider the following definition of a matrix dispersion:

Definition 3.1.7 (Dispersion)**.** *Suppose P is an $N \times N$ matrix, and $\sigma^{(2)}(P)$ are its spectral pairs. The dispersion of P with respect to the pairing scheme Π and dispersion metric δ_M is denoted by $\Delta_M(P \mid \Pi)$ and it is given by the following:*

$$\Delta_M(P \mid \Pi) = \{\delta_M(\pi_{ij}) \mid \pi_{ij} \in \sigma^{(2)}(P \mid \Pi)\}$$

Consider the following code example, generating the dispersion of standard normal 5×5 matrix with respect to the consecutive pairing scheme.

Code Example (Consecutive Pair Dispersion of a Standard Normal Matrix)**.** *In our notation, we are simulating the dispersion of $P \sim \mathcal{N}(0, 1)$ where $P \in \mathbb{R}^{5\times5}$. Specifically, we are simulating $\Delta(P \mid \Pi_C)$. In the code implementation, we obtain an array for every dispersion metric δ.*

```
library(RMAT)
P <- RM_norm(N = 5, mean = 0, sd = 1)
disp_P <- dispersion(P, pairs = "consecutive")
# Outputs the following
disp_P
...
i  j  eig_i        eig_j        id_diff     iddiff_norm abs_diff rank_diff
2  1  -0.54-1.35i  -0.54+1.35i   0.00+2.71i 2.71          0.00     1
3  2   0.23+1.43i  -0.54-1.35i  -0.77-2.78i 2.88          0.02     1
4  3   0.23-1.43i   0.23+1.43i   0.00+2.85i 2.85          0.00     1
5  4  -0.87+0.00i   0.23-1.43i   1.09-1.43i 1.80          0.57     1
```

Next, we extend the definition of dispersion for an ensemble as we usually do.

Definition 3.1.8 (Ensemble Dispersion)**.** *If we have an ensemble $\mathcal{E}$, then we can naturally extend the definition of $\Delta_M(\mathcal{E} \mid \Pi)$. To take the dispersion of an ensemble, simply take the union of the dispersions of each of its matrices. In other words, if $\mathcal{E} = \{P_i \sim \mathcal{D}\}_{i=1}^{K}$, then its dispersion is given by:*

$$\Delta_M(\mathcal{E} \mid \Pi) = \bigcup_{i=1}^{K} \Delta_M(P_i \mid \Pi)$$

Consider the following code example, generating the dispersion of a beta ensemble with respect to the consecutive pairing scheme.

Code Example (Consecutive Pair Dispersions of a Beta Ensemble)**.** *In our notation, we are simulating the dispersion of* $\mathcal{E} \sim \mathcal{H}(\beta = 4)$ *where* $P \in \mathbb{R}^{5\times 5}$*. Specifically, we are simulating* $\Delta(P \mid \Pi_C)$*. In the code implementation, we obtain an array for every dispersion metric* δ*.*

```
library(RMAT)
ens <- RME_beta(N = 4, beta = 4, size = 3)
disp_ens <- dispersion(ens, pairs = "consecutive")
# Outputs the following
disp_ens
...
i j eig_i     eig_j     id_diff id_diff_norm abs_diff rank_diff
2 1 -3.78+0i 4.00+0i  7.78+0i 7.78          0.22     1
3 2 2.06+0i -3.78+0i -5.84+0i 5.84          1.72     1
4 3 0.19+0i  2.06+0i  1.88+0i 1.88          1.88     1
2 1 3.80+0i -4.00+0i -7.80+0i 7.80          0.20     1
3 2 -1.80+0i 3.80+0i  5.60+0i 5.60          2.00     1
4 3 0.89+0i -1.80+0i -2.69+0i 2.69          0.92     1
2 1 3.51+0i -3.53+0i -7.04+0i 7.04          0.03     1
3 2 1.35+0i  3.51+0i  2.16+0i 2.16          2.16     1
4 3 -0.67+0i 1.35+0i  2.02+0i 2.02          0.68     1
```

3.2 Dispersion Analysis

With dispersions well defined, we provide a few guidelines for analyzing a dispersion of a matrix. We will synthesize all the techniques, notations, and remarks in the previous section to provide a comprehensive manual on how to analyze the dispersion of a matrix.

3.2.1 Considerations

After an extensive amount of setup, possibly too many definitions, the reader should be proud! This is the last of it. (I promise.) Now, in the future sections, we will utilize our toolkit to analyze and study dispersions.

However, before we may proceed, this section will be a primer prior to starting analyzing dispersions. Specifically, this section will discuss various subtle caveats and general considerations to take into account before proceeding. Essentially, we will be justifying the truck-load of definitions that were provided by showing why it is useful to have short-hand for them and use them concisely. Without further ado, here are some considerations that should be taken into account when studying dispersions.

Generally speaking, when we consider a dispersion set, what we care to obtain from it are insights and patterns in the eigenvalue spacings. The dispersion of a matrix P, $\Delta_M(P)$ could tell us quite a few things. One thing it could tell us is the

distribution of the dispersion of eigenvalues given that we uniformly and randomly select a pair of eigenvalues; this works when δ is a symmetric function. However, if we are concerned about spacings, then we must take into consideration a few facts.

Linear Combinations. The identity difference metric is not necessarily a "bad metric". In fact, it is one of the few metrics which allow transitivity under composition. This brings up an important detail. In the all-pairs dispersion $\Delta_{id}(P \mid \Pi_0)$, there is some redundant information if we are totally concerned with spacings. Namely, consider the following fact:

$$\lambda_i - \lambda_j = (\lambda_i - \lambda_k) - (\lambda_j - \lambda_k)$$

In other words, if $\delta = \delta_{\text{id}}$, then $\delta(\pi_{ij}) = \delta(\pi_{ik}) - \delta(\pi_{jk})$. We could say that λ_k is a pivot in which we could perform right-cancellation. Because of this, using the all-pairs scheme Π_0 leads to finding various linear combinations in our dispersions.

Triangle Inequality. However, as we said previously, only the identity difference metric gives us transitive properties. On the other hand, the other metrics provide us bounds when we compose them. They are given by a simple application of the reverse triangle inequality. Namely, consider the following fact, which applies to any numbers x, y in either $\mathbb{R}$ or $\mathbb{C}$.

$$||y| - |x|| \leq |y - x|$$

Realize that two eigenvalues are just a pair of complex numbers, so we may rephrase that inequality as saying $|\delta_{abs}(\pi_{ij})| \leq \delta_n(\pi_{ij})$. However, if we choose the appropriate pairing scheme, we can remove the absolute value to obtain a more meaningful upper bound.

Consider the following. The value of $\delta_{abs}(\pi_{ij})$ can be interpreted as the difference between the sizes of λ_j and λ_i. However, if we use the **lower** pair combinations $\Pi_<$, then we know that $i > j$, making the value $|\lambda_j|$ always greater than $|\lambda_i|$ by construction. As such, the left-hand value would always be positive, allowing us to drop the absolute value. So, we could say:

$$\delta_{abs}(\pi) \leq \delta_n(\pi) \text{ given } \pi \in \sigma^{(2)}(P \mid \Pi_>)$$

Sufficiency of Consecutive Pairs. Recall from previously that with respect to the identity difference, some elements of the dispersion $\Delta_{id}(P)$ can be expressed as a linear combination of other elements. To avoid this issue of linear combinations, we simply take the dispersion with respect to the consecutive pairing scheme. In other words, we consider only consecutive eigenvalues. By imposing this condition, we get that none of the values are a linear combination of the other in $\Delta_{id}(P \mid \Pi_C)$. So, we can say that the consecutive pairs are sufficient in the sense that they give us almost all the information we need.

3.2.2 Order Statistics

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

In **Section 2.2.3**, we observed simple order statistics regarding the eigenvalues. This time around, we have two eigenvalues being compared at once, so there needs to be further setup with regards to dispersion order statistics.

However, prior to introducing any new concepts, there is one scenario where using one index is sufficient in terms of discussing dispersion order statistics.

Remark (Consecutive Pair Dispersions)**.** *When we consider the dispersion with repsect to the consecutive pairs, things are much simpler since our pairs are intrinsically defined by one index (j). This is because we are observing the j^{th} eigenvalue and its smaller neighbour. As such, we will consider dispersion statistics in the form of $\mathbb{E}(\tilde{\pi}_j \mid j)$ and $Var(\tilde{\pi}_j \mid j)$.*

Otherwise, we synthesize a new order statistic that takes in two indices (from a given pair) called the **ranking difference class**. Since we are no longer observing a single eigenvalue at a given rank, we will need a way to standardize observing a pair of eigenvalues at a time. To do so, we introduce a new **equivalence relation** called the **ranking difference**. As the name suggests, it is precisely the integer difference of the eigenvalue orders.

Ranking Difference

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

With the ranking difference, we may now take the spectral pairs of some random matrix with respect to either the lower ($\Pi_<$) or upper ($\Pi_>$) pairing scheme and partition the eigenvalues into distinct equivalence classes.

Remark (Equivalence Relation)**.** *Formally speaking, we may define $\sim_\rho$ as an equivalence relation. It satisfies all three properties of reflexivity, symmetry, and transitivity. More precisely, it is an equivalence class which partitions $\mathbb{N}_N \times \mathbb{N}_N$ into $N - 1$ equivalence classes given by $[r]_\rho$ for $r \in \mathbb{N}_{N-1}$. We would define the equivalence relation $\sim_\rho$ as follows:*

$$(\lambda_a, \lambda_b) \sim_\rho (\lambda_c, \lambda_d) \iff (a - b) = (c - d)$$

So, two eigenvalue pairs would belong to the same class $[r]_\rho$ if their ranking difference is the same. That is,

$$[r]_\rho = \{(\lambda_\alpha, \lambda_\beta) \mid \alpha - \beta = r\}$$

Remark (Matrix Analogy)**.** *Using the eigenvalue matrix analogy, $\sim_\rho$ partitions the matrix into diagonal bands. The diagonal represents $[0]_\rho$, the first off-diagonal represents $[1]_\rho$, and so on... Notice that this means every equivalence class has a different size.*

Usage. With this diagonal band partition in mind, we can consider various statistics conditioning on the value of r. Conditioning on r will be especially useful in the cases where we are considering matrices like the Hermite-β matrices; we will find that the eigenvalues of those matrices tend to *repel*, so to speak, and we can observe those patterns using r.

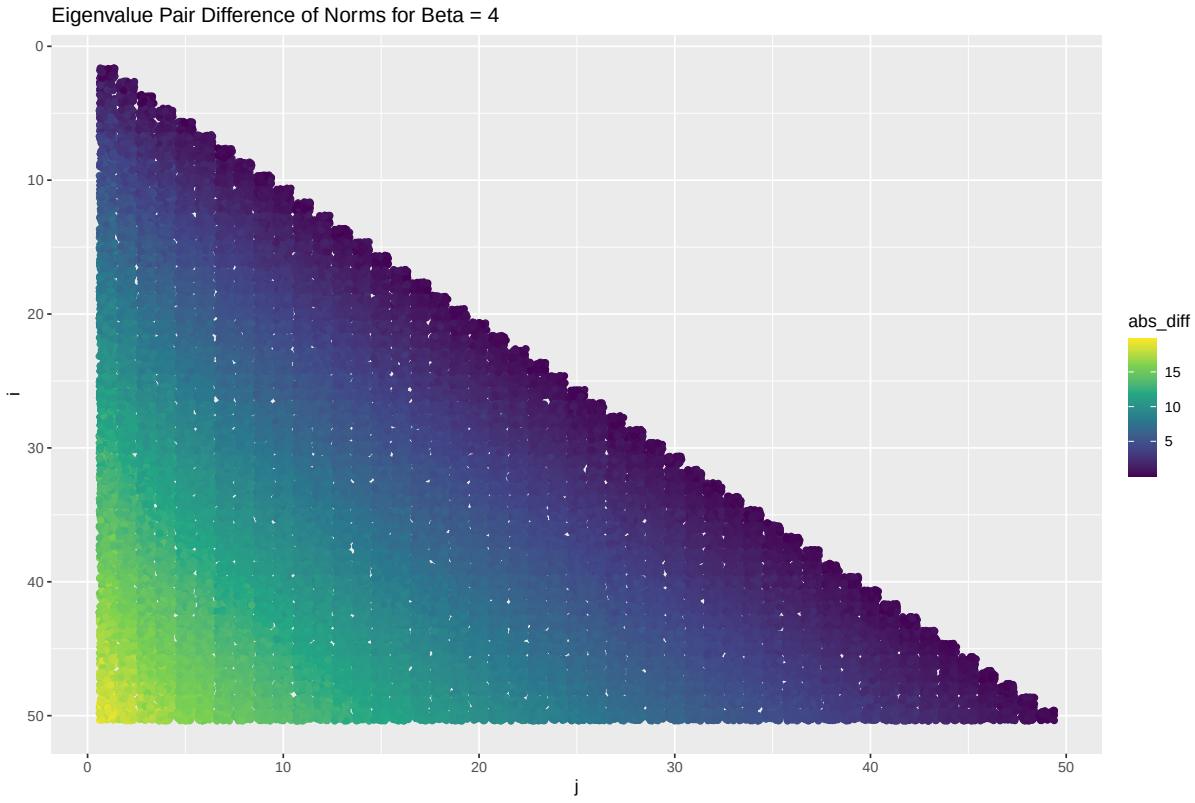
Consider the following example utilizing and identifying ranking differences.

Example (Ranking Differences for a 4×4 Matrix)**.** *Suppose P is a 4×4 random matrix. Then, if we take the spectral pairs with respect to the lower pairs $\sigma^{(2)}(P \mid \Pi_<)$ and partition it by ρ, we get the following classes:*

1. $[1]_\rho = \{\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3\}$
2. $[2]_\rho = \{\pi_{31}, \pi_{42}\}$
3. $[3]_\rho = \{\pi_{41}\}$

Example: Dispersion Matrix

As to showcase this dispersion technology, consider the following example, where we analyze the dispersion for the ensemble $\mathcal{E} \sim \mathcal{H}(\beta = 4)$. Specifically, we are considering $\Delta_{\text{abs}}(\mathcal{E} \mid \Pi_<)$.



Takeaways

In the plot above, we are considering the absolute differences dispersion metric $\delta = \delta_{\text{abs}}$. That being considered, we can use the ranking difference construction provided earlier to parse this graph. Recall that the ranking difference partitions the eigenvalue pair into separate classes based on how many "ranks" apart two eigenvalues are. With this interpretation, this tells us that roughly speaking, each diagonal band $[\rho]$ has a uniform difference of absolutes given by the color streak that diagonal occupies.

For instance, we could say that any pairs occupying the midnight blue diagonal $[\rho \approx 10]$ has an approximate dispersion of 5. So, this means that we expect $\delta_{\text{abs}}(\lambda_{j+10}, \lambda_j) \approx 5$. For another example, we could say that any pairs occupying the lime diagonal $[\rho \approx 40]$ has an approximate dispersion of 15. So, this means that we expect $\delta_{\text{abs}}(\lambda_{j+40}, \lambda_j) \approx 15$.

3.3 Wigner's Surmise

Wigner's surmise is a result found by Eugene Wigner regarding the limiting distribution of eigenvalue spacings of symmetric matrices [Mehta (2004)]. To start talking about this, we must talk about normalized spacings, which are the precise items considered in the distribution. Before, we can talk about the normalized spacing, we define the mean spacing.

Definition 3.3.1 (Mean Spacing)**.** *Suppose P is an $N \times N$ symmetric matrix, and $\sigma(P)$ are its real, sign-ordered eigenvalues. Then, the mean (eigenvalue) spacing, denoted $\langle s \rangle$ is the average distance between two consecutive eigenvalues. That is,*

$$\langle s \rangle = \mathbb{E}[\Delta_\delta(P \mid \Pi_C)] = \mathbb{E}[\delta(\tilde{\pi}_j)]_{j=1}^{N-1}$$

So, with the mean spacing defined, we now define the normalized spacing between a pair of consecutive eigenvalues below.

Definition 3.3.2 (Normalized Spacing)**.** *Suppose P is an $N \times N$ symmetric matrix, and $\sigma(P)$ are its real, sign-ordered eigenvalues. Then, the normalized spacing of the j^{th} pair of eigenvalues, denoted s_j is given by the following formula.*

$$s_j = \frac{(\lambda_j - \lambda_{j+1})}{\langle s \rangle} = \frac{\delta(\tilde{\pi}_j)}{\langle s \rangle}$$

Finally, we define the Wigner dispersion, which we may reconstruct using our notation. This way, we can formalize Wigner's Surmise as an observation of the Wigner dispersion for symmetric matrices.

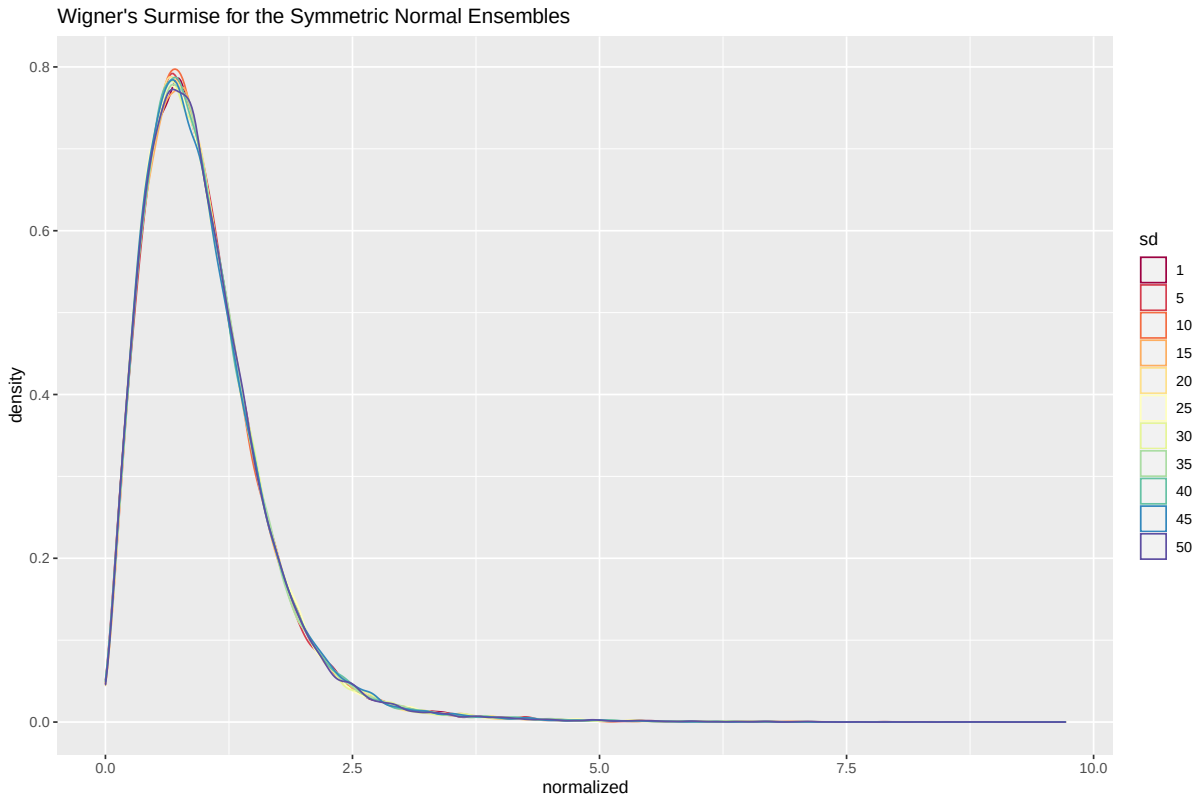
Definition 3.3.3 (Wigner Dispersion)**.** *Suppose P is an $N \times N$ symmetric matrix, and $\sigma(P)$ are its real, sign-ordered eigenvalues. Then, the Wigner dispersion denoted $\Delta_W(P)$ is given by the set of normalized consecutive eigenvalue spacings of P. That is,*

$$\Delta_W(P) = \left\{ \frac{\delta_n(\pi)}{\langle s \rangle} \mid \pi \in \sigma^{(2)}(P \mid \Pi_C) \right\}$$

The extension for ensembles is trivial; it inherits the same notation for matrices and the definition is extended similarly to how we did so for the spectrum and dispersion of an ensemble. That being said, consider the following simulation of Wigner's surmise.

3.3.1 Symmetric Normal Matrices

We simulate the Wigner dispersion distribution (Wigner's surmise) for several $\mathcal{E} \sim \mathcal{N}(0, \sigma^2)^\dagger$ matrices over $\mathbb{R}$. Specifically, we are varying the variance of our sampling distribution σ^2 and taking several values $\sigma \in [1, 50]$.



Takeaways

As we can see, it seems as though the distribution is independent of the variance of the distribution. This means that the variance of our entries have **no bearing** on the mean dispersion of consecutive eigenvalues, which is initially surprising! So, if two ensembles $\mathcal{E} \sim \mathcal{N}(0, 1)^\dagger$ and $\mathcal{E} \sim \mathcal{N}(0, 2500)^\dagger$ behave the same way, how can we vary the distribution of Wigner dispersions? The answer is β-ensembles.

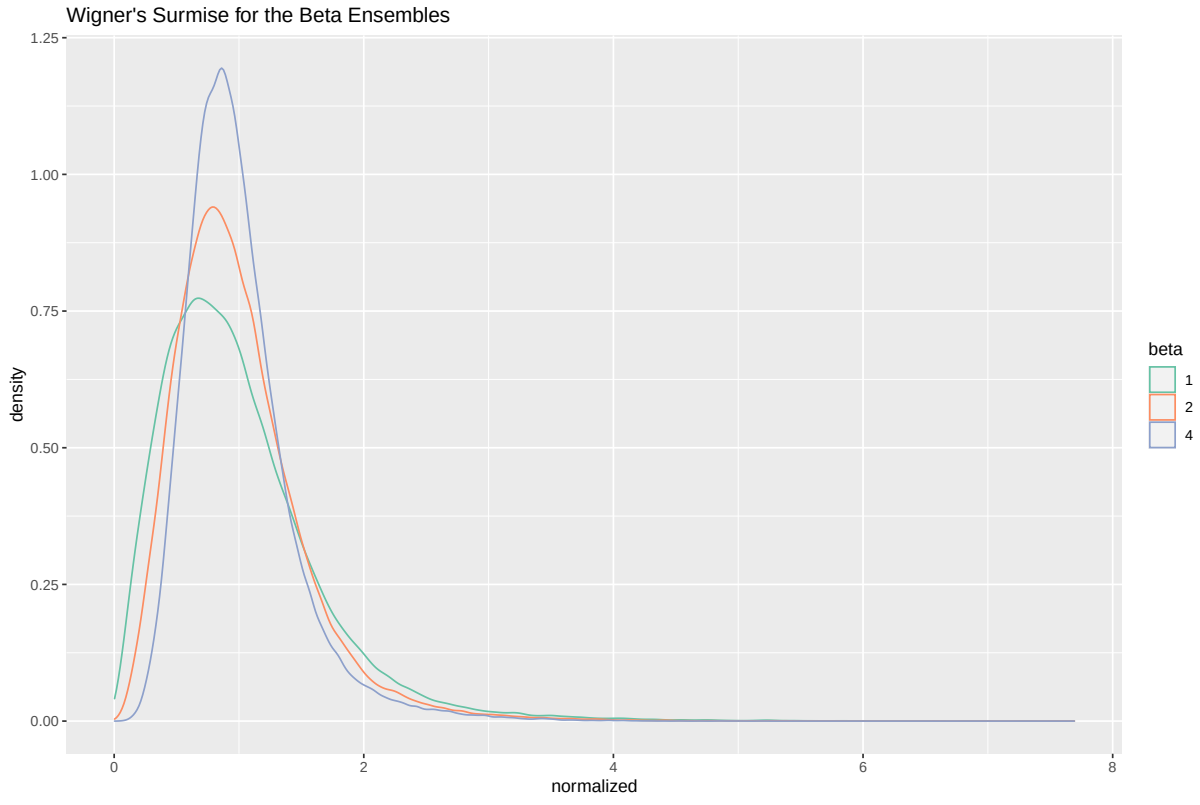
It turns out that the class of $\mathcal{N}(\mu, \sigma^2)$ matrices for any value of μ or σ are represented by the β-ensemble for $\beta = 1$; in other words, they have the same eigenvalue distribution. Curiously enough, if we sampled the same matrices (Hermitian Normal) over $\mathbb{C}$, we would have found an analogous result, where all the Hermitian Normal matrices are represented by the β-ensemble with $\beta = 2$! This will be discussed more in depth in **Chapter 4**.

3.3.2 β-Ensembles

As we just mentioned, it turns out that Symmetric Normal matrices and Hermitian Normal matrices have an analogous representation as β-ensembles for $\beta = 1$ and $\beta = 2$ respectively. Although this will be discussed later, it is helpful to know that there are three "standard" β-ensembles with special properties. We have already talked about two of them. The last special β-ensemble is that corresponding to $\beta = 4$.

That being said, there are analytical results for the Wigner dispersion distributions for the Hermite β-ensembles for $\beta = 1, 2, 4$. We will denote the density of each ensemble by w_β. These are the asymptotic densities for the normalized spacings between consecutive eigenvalues for each of the matrices for $\beta = 1, 2, 4$. Additionally, a simulation plot for each β-ensembles is shown below.

$$w_1(s) = \frac{\pi}{2}\exp\left(-\frac{\pi}{4}s^2\right)$$
$$w_2(s) = \frac{32}{\pi^2}s^2\exp\left(-\frac{4}{\pi}s^2\right)$$
$$w_4(s) = \frac{2^{18}}{3^6\pi^3}s^4\exp\left(-\frac{64}{9\pi}s^2\right)$$



Takeaways

As we can see, given just our three values of β, there are some patterns to consider. The first thing that jumps out is that the variance of the dispersion distribution seems to decrease as β increases. This will be justified in **Chapter 4** in addition to being generalized for $\beta \in \mathbb{N}$. In fact, this observation regarding the variance seems to go hand in hand with the analytical results for the asymptotic densities. Observe that the formulas for the asymptotic densities contain first, second, and fourth degree terms of the normalized spacing. As such, this has an impact on the moments of the p.d.f. and may explain why the variance seems to decrease.

Checkpoint

All things considered, we are now ready to switch topics and start to focus solely on the β-ensembles in the next chapter. At this point, the reader has acquired a vast toolkit and considerable knowledge to study random matrix theory in more depth. However, we have only scratched the surface, and you are encouraged to keep learning! As such, we will sample the β-ensembles, which we are already familiar with, to serve as our "case study" in this thesis.

Chapter 4

β-Ensembles

> With thermodynamics, one can calculate almost everything crudely; with kinetic theory, one can calculate fewer things, but more accurately; and with statistical mechanics, one can calculate almost nothing exactly.
>
> *Eugene Wigner*

4.1 Introduction

In this chapter, we will talk about the Hermite β-ensembles, also commonly known as the *Gaussian ensembles*, more in depth. The β-ensembles have wide applications in statistical physics, engineering, and many other places. Canonically, there are three "standard" β-ensembles, corresponding to the values of $\beta = 1, 2$, and 4. It was briefly mentioned in a previous section, but what makes these ensembles special is how they are characterized. Fundamentally, the ensembles are defined by the joint density of their eigenvalues, dependent on the parameter β of course. While β-ensemble model parameter β has a support of $\mathbb{N}$, the three standard values give us very special properties that we discuss in this chapter. Extending the model beyond those standard values (which we do using the β-matrix model) as we do in the previous sections in turn means that the following special characterizations that will be discussed do not apply anymore.

4.1.1 Hermite β-Ensemble

To begin, we will write down the joint eigenvalue probability density function of a Hermite β-matrix.

Definition 4.1.1 (β-ensemble)**.** *A (Hermite) β-ensemble is an ensemble of random matrices parameterized by β, which determines the joint eigenvalue p.d.f that characterizes it. So, given an observed set of eigenvalues $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$, the joint p.d.f. of Λ is as follows:*

$$f_\beta(\Lambda) = C_\beta \prod_{i<j} |\lambda_i - \lambda_j|^\beta e^{-\frac{1}{2}\sum_{i=1}^N \lambda_i^2}$$

where the normalization constant C_β is given by:

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

Breaking Down the Behemoth

Now, if the reader's reaction to the statement of the β-ensemble was shock or trepidation, do not worry. Everything will become clear soon as we start to understand the components of the density function and focus on characterizing the components rather than try to parse through the dense notation. Speaking of which, let us start to rewrite the density function with friendlier, more readable notation. To begin, let us refer to the two primary components as the "blue term" and the "red term" as shown below.

$$f_\beta(\Lambda) = C_\beta \boxed{\prod_{i<j} |\lambda_i - \lambda_j|^\beta} \quad \boxed{\exp\left(-\tfrac{1}{2}\textstyle\sum_{i=1}^N \lambda_i^2\right)}$$

Normalization Constant. First things first, let us consider the normalization constant C_β. This constant is simply a constant needed by f_β to normalize the p.d.f such that it sums to 1. The term is expressed as a product of gamma terms involving β. Additionally, this term is related to an integral called the *Selberg integral*. While the constant is interesting in terms of its relationship with that integral, it is not under our purview. So, the important takeaway is that it just normalizes our p.d.f, because every p.d.f needs to sum to 1 at the end of the day.

The Blue Term. The first big component, which we will call "the blue term" is the term $\prod_{i<j} |\lambda_i - \lambda_j|^\beta$. Let us break this term down further.

1. The first thing to note is that this term is a product of terms in the form $|\lambda_j - \lambda_i|^\beta$.

2. Secondly, note that this product term runs over the indices $1 \leq i < j \leq n$.

Here's the great part: the reader has already seen and is already acquianted with those two items! Realize,

1. The term $|\lambda_j - \lambda_i|^\beta$ is simply the dispersion of π_{ij} with respect to the β-norm dispersion metric δ_β. So, we can rewrite $\delta_\beta(\pi_{ij}) = |\lambda_j - \lambda_i|^\beta$.

2. The indices $\{i < j \mid i, j \in \mathbb{N}_N\}$ may look familiar. This is the lower pairing scheme!

As such, when we consider these two facts, the interpretation hopefully becomes a little simpler. We can say that the blue term is a product of the dispersion of all eigenvalue pairs in the lower pairing scheme w.r.t the β-norm dispersion metric. Now, we know that for all $\beta \in \mathbb{N}$ that the β-norm is a positive definite metric. As such, this means that the term $\delta_\beta(\pi)$ is a positive term that grows when the dispersion between the eigenvalues in the pair π grows larger. So, the important takeaway here is that the blue term tells us that the eigenvalues in the β-ensemble tend to "repel" each other. The farther all the eigenvalues in Λ are from each other, the more likely we are to observe such Λ.

Additionally, this also means there is **zero probability** of observing any pair of identical eigenvalues. This is because $\delta_\beta(z, z') = 0$ if $z = z'$, meaning that if we observed two identical eigenvalues, then the probability density would be zero. This effectively places a bound on how close our eigenvalues can be to each other.

The Red Term. The second big component, which we will call "the red term" is the term $\exp\Big(-\frac{1}{2}\sum_{i=1}^N \lambda_i^2\Big)$. This term is slightly simpler to understand. Fundamentally, this term is simply an exponential term with an input that is a **negative value**.

To see this, first label the term $S = \sum_{i=1}^N \lambda_i^2$. This simplifies the red term to become $\exp\big(-\frac{1}{2}S\big)$. We know the term S to be positive since the eigenvalues are **real** and that the sum of squares of several real values is strictly positive. As such, the term, multiplied by a negative number $(-\frac{1}{2})$ is guaranteed to be negative. So, to summarize, we know that $-\frac{1}{2}S$ is a negative value. Then, from what we know about the exponential function, we know that:

$$\lim_{x \to -\infty} \exp(x) = 0$$

So, this means that as $S \to \infty$, $\exp\big(-\frac{1}{2}S\big) \to 0$. Since S is simply the sum of magnitudes of the eigenvalues, this implies that the larger our eigenvalues are, the smaller is our red term. In turn, this implies that the joint p.d.f. of eigenvalues is smaller overall. To summarize, this means that our red term tells us that (relatively speaking) the larger the eigenvalues in Λ, the less likely we are to observe those eigenvalues in the β-ensemble. This effectively places a bound on the sizes of our eigenvalues.

Takeaways. Altogether, here is what we can say about the β-ensemble joint eigenvalue p.d.f just from observing the terms.

1. When λ_i is large, then $\mathbb{P}(\Lambda)$ is small.

2. When $\delta(\lambda_i, \lambda_j)$ is small, then $\mathbb{P}(\Lambda)$ is small.

4.1.2 The Invariance Criterion

The Three Musketeers (Associative Algebras)

We know from Abstract Algebra that there are only three associative algebras over the real numbers $\mathbb{R}$ (i.e. algebras with a robust multiplication structure) [Dummit & Foote (2003)]. They are the following:

1. The real numbers, $\mathbb{R}$
2. The complex numbers, $\mathbb{C} = \{z = x + yi \mid x, y \in \mathbb{R}\}$
3. The quaternionic numbers, $\mathbb{H} = \{\alpha = a + bi + cj + dk \mid a, b, c, d \in \mathbb{R}\}$

Diagonalization & Conjugation Invariance

One of the most elegant results in random matrix theory is the equivalence between the β-ensembles and their normal matrix counterparts over the three associative algebras. Without further ado, consider the following.

Real. Suppose $M \sim \mathcal{N}(\mu, \sigma^2)^\dagger$ over $\mathbb{R}$ is a real symmetric matrix. Then, its eigenvalues have a joint p.d.f. of $f_{\beta=1}(\Lambda)$, meaning it represents the $\beta = 1$ ensemble.

Additionally, we know that symmetric matrices are diagonalizable under conjugation with an orthogonal matrix. In other words, we could express M as $P^{-1}DP$ where D is a diagonal matrix and $P \in O(n)$ is an orthogonal matrix. For this reason, we call the $\beta = 1$ ensemble the *Gaussian Orthogonal Ensemble*, abbreviated as GOE.

Complex. Suppose $M \sim \mathcal{N}(\mu, \sigma^2)^\dagger$ over $\mathbb{C}$ is a complex hermitian matrix. Then, its eigenvalues have a joint p.d.f. of $f_{\beta=2}(\Lambda)$, meaning it represents the $\beta = 2$ ensemble.

Additionally, we know that hermitian matrices are diagonalizable under conjugation with a unitary matrix. In other words, we could express M as $P^{-1}DP$ where D is a diagonal matrix and $P \in U(n)$ is an unitary matrix. For this reason, we call the $\beta = 1$ ensemble the *Gaussian Unitary Ensemble*, abbreviated as GUE.

Quaternionic. Suppose $M \sim \mathcal{N}(\mu, \sigma^2)^*$ over $\mathbb{H}$ is a quaternionic self-dual matrix. Then, its eigenvalues have a joint p.d.f. of $f_{\beta=4}(\Lambda)$, meaning it represents the $\beta = 4$ ensemble.

Additionally, we know that self-dual quaternionic matrices are diagonalizable under conjugation with a symplectic matrix. In other words, we could express M as $P^{-1}DP$ where D is a diagonal matrix and $P \in Sp(n)$ is a symplectic matrix. For this reason, we call the $\beta = 1$ ensemble the *Gaussian Symplectic Ensemble*, abbreviated as GSE.

All that being said, we can say that with respect to each matrix group, each β-ensemble matrix has a feature called **conjugation invariance**. What this entails is that any of the matrices can be characterized by two things: their eigenvalues, and an arbitrary group element (from the respective algebra's matrix group) which determines its basis.

Dyson Index

So, all that being said, why do we use the value β and where do the numbers come from? The answer is the Dyson index [Tao (2012)]. The Dyson index β corresponds to the number of real components in the field the conjugation matrix groups cover. Refer to the definitions of the associative algebras above, and you will find that for $\beta = 1, 2, 4$, you have the representation of a matrix over a field with β real dimensions.

4.1.3 A Physical Interpretation

Alongside these great algebraic properties, the β-ensembles also have a interpretation as a physical model. This is because as mentioned previously, these ensembles show up frequently in statistical physics.

So, we will cover one physical model that the β-ensemble represents. Suppose that $P \sim \mathcal{H}(\beta)$ is an $N \times N$ matrix. Then, the eigenvalues of P have a representation as a model of charged point particles.

Charged Particle Model

Suppose N particles are in a line about the origin. Additionally, suppose there is a quadratic field potential $V(x) = x^2$. Then, the eigenvalues of P, $\sigma(P)$ represents a stochastic system of such particles. As discussed previously in the breakdown of the density function, these particles tend to repel each other. Specifically, their repulsion factor is β itself! By observing the red term and blue term at extreme values of β, we can interpret the physical model as follows:

Low Repulsion, High Temperature. As $\beta \to 0$, the temperature of the system $T \to \infty$. At these values, the model starts to behave like an ideal gas. Additionally, there is no interaction between the particles anymore as the power of the blue term implies that the dispersion has no effect anymore. This returns a **fully stochastic** model of particles trying to align themselves along the field potential. In this model, the field potential matters a lot, as it determines the positioning of the particles.

High Repulsion, Low Temperature. As $\beta \to \infty$, the temperature of the system $T \to 0$. At these values, the model loses its stochastic properties and starts to become **deterministic**. Additionally, there is maximal interaction between the particles as the power of the blue term starts to become more prominent. This returns a fully deterministic* model of particles that align themselves equidistantly. In this model, the field potential doesn't matter anymore.

Remark (Deterministic System)**.** *While the second model is effectively deterministic, there does remain the randomness in the indices of the particles which are uniform.*

4.1.4 The Matrix Model

Now, we have thoroughly characterized, interpreted, and explained the β-ensemble. How do we simulate such an ensemble given that we only know the joint p.d.f. of the eigenvalues? Luckily, we do not need to delve deep into this. To simulate matrices from the β-ensemble, we will be using the result published in "Matrix Models for Beta Ensembles" by Dr. Ioana Dumitriu [Dumitriu & Edelman (2018)]. This gives us a matrix model that generates a matrix that would be observed in a β-ensemble given any $\beta \in \mathbb{N}$.

Dumitriu's Matrix Model of β-Ensembles

This result was actually described previously in **Section 1.1.1**! We first described the β-matrices as a non-homogenous $\mathcal{D}$-distribution that we denoted $\mathcal{D} = \mathcal{H}(\beta)$. This is a symmetric, tridiagonal matrix model with entries sampling from the normal and chi distributions. What we get in turn, is a matrix model whose eigenvalues **implicitly** have the joint p.d.f. of the Hermite-β ensemble described in **Section 4.1.1**. The algorithm used is directly cited from the results of Dumitriu's paper, and can be found below.

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

1. *To simulate an $N \times N$ beta matrix, 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 $\chi(df = c_j)$ where $c_j = \beta \cdot j$ for columns spanning $j = 1, \ldots, N-1$.*
4. *Normalize the entries by dividing by $\sqrt{2}$.*

In turn, we can now easily simulate β-ensembles using **RMAT** as such.

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

```
library(RMAT)
ensemble <- RME_beta(N = 4, beta = 2, size = 10)
# Outputs the following
ensemble
...
[[10]]
           [,1]       [,2]         [,3]      [,4]
[1,] 0.7246302 1.8893868  0.00000000 0.000000
[2,] 1.8893868 1.5278221  0.68840045 0.000000
[3,] 0.0000000 0.6884004 -0.03876104 1.944495
[4,] 0.0000000 0.0000000  1.94449533 1.042741
```

4.2 Spectra

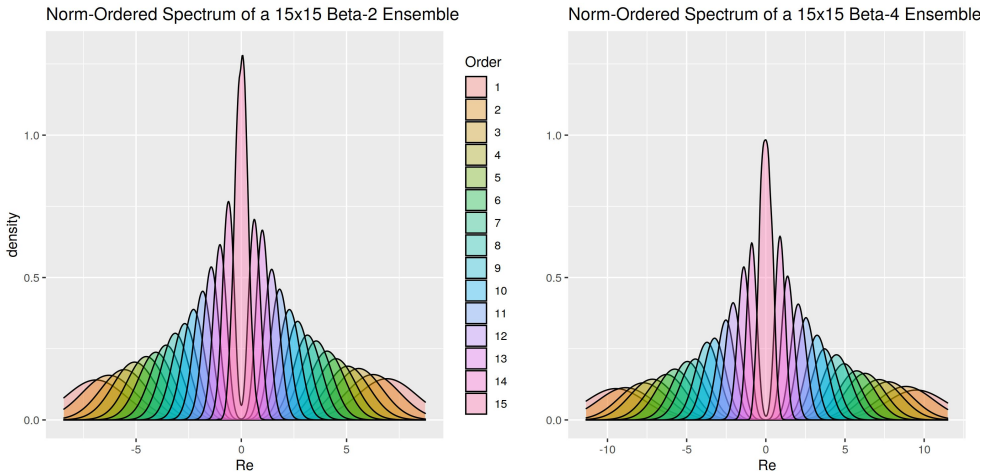
In this section, we will continue to survey the spectra of β-ensembles. The reason we are continuing is because we have already started this work in **Section 2.3.3**! To remind the reader, in that section, we studied the symmetric normal ensemble over $\mathbb{R}$. As we now know, this means that we have already studied the GOE ($\beta = 1$) ensemble. However, before we proceed, here is an important consideration regarding **identifiability** to account for when simulating random eigenvalues from the β-ensembles.

Remark (Implicit Distribution)**.** *Note that the β-ensemble is an ensemble characterized by its joint eigenvalue p.d.f. As such, there are some subtle but important considerations to take into account regarding identifiability. Recall that we formalized the spectrum of a matrix as a function which takes in as an input a vector of random variables. So, while Dumitriu's model provides an explicit formula for the matrix entries, for the eigenvalues, there is the issue of identifiability. That is, given some eigenvalues, there is no injective function to the characteristic polynomial that produced it. Rather, there are infinitely many equivalence classes of characteristic polynomials (and such, random matrices) that surjectively produce a given multiset of eigenvalues.*

Without further ado, we will now take a look at the properties of the spectra of β-ensembles.

4.2.1 Standard Ensembles

As mentioned previously, we have already analyzed the case of $\beta = 1$. As it turns out, all three ensembles share the properties listed in **Section 2.3.3**. Namely, we find that all three ensembles have a semicircle distribution (with potentially differing radii). Additionally, we also observe that the variance is higher towards the edges where the larger eigenvalues are. That being said, consider the spectra of the $\beta = 2$ and $\beta = 4$ ensembles.

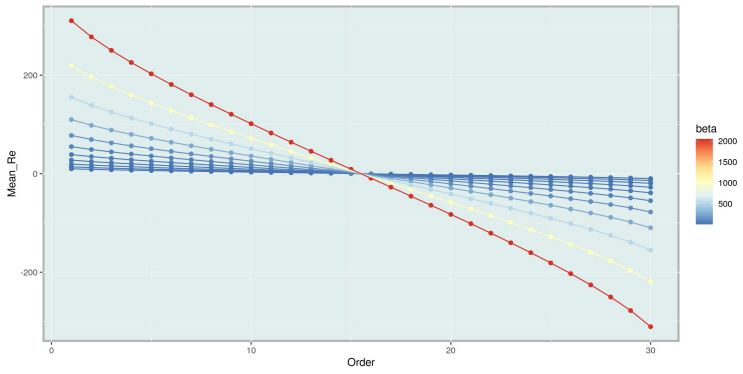


The most notable difference between the two ensembles is that the peaks of the $\beta = 4$ ensemble seem to be more "dull" than those of the $\beta = 2$ ensemble. Recall from previous sections in this chapter, we state that the higher the value of β, the more likely is the ensemble to have equidistant eigenvalues. As such, this aligns with that fact since less peaked densities for the values implies that there is more intersection between them at a given order. As such, this plot second-handedly demonstrates this phenomenon.

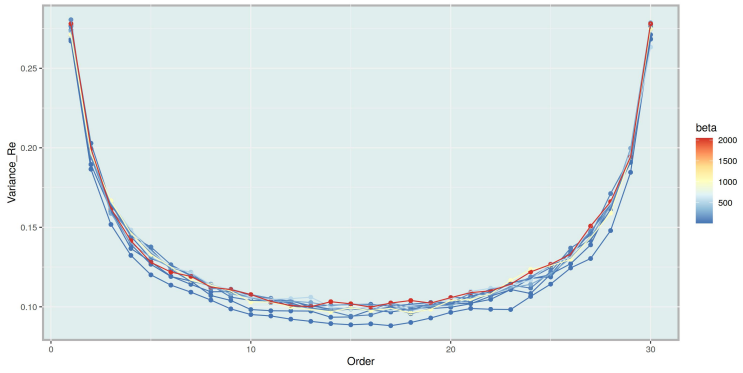
4.2.2 Extending the β-Ensembles

Now, we may consider extending the β-ensembles and observe their spectra. In this section, we observe various β-ensembles for various values of β in the first eleven powers of 2. Additionally, we are using the sign-ordered scheme on the eigenvalues.

In the plot below, we observe the statistic $\mathbb{E}(\lambda_i \mid i)$. Immediately, we start to notice that the larger the value of β, the larger the eigenvalues are (fixing N). This is consistent with what we expect about the limiting behaviour of the eigenvalues as $\beta \to \infty$.



In this plot, we now observe the statistic $\mathrm{Var}(\lambda_i \mid i)$. At a first glance, there doesn't seem to be any resounding patterns. There is roughly a tendency for the larger value of β to have a higher variance towards the center, but this may as well be random noise. This in turn shows that the variance $\mathrm{Var}(\lambda_i \mid i)$ is roughly uniform for any $\beta \in \mathbb{N}$.

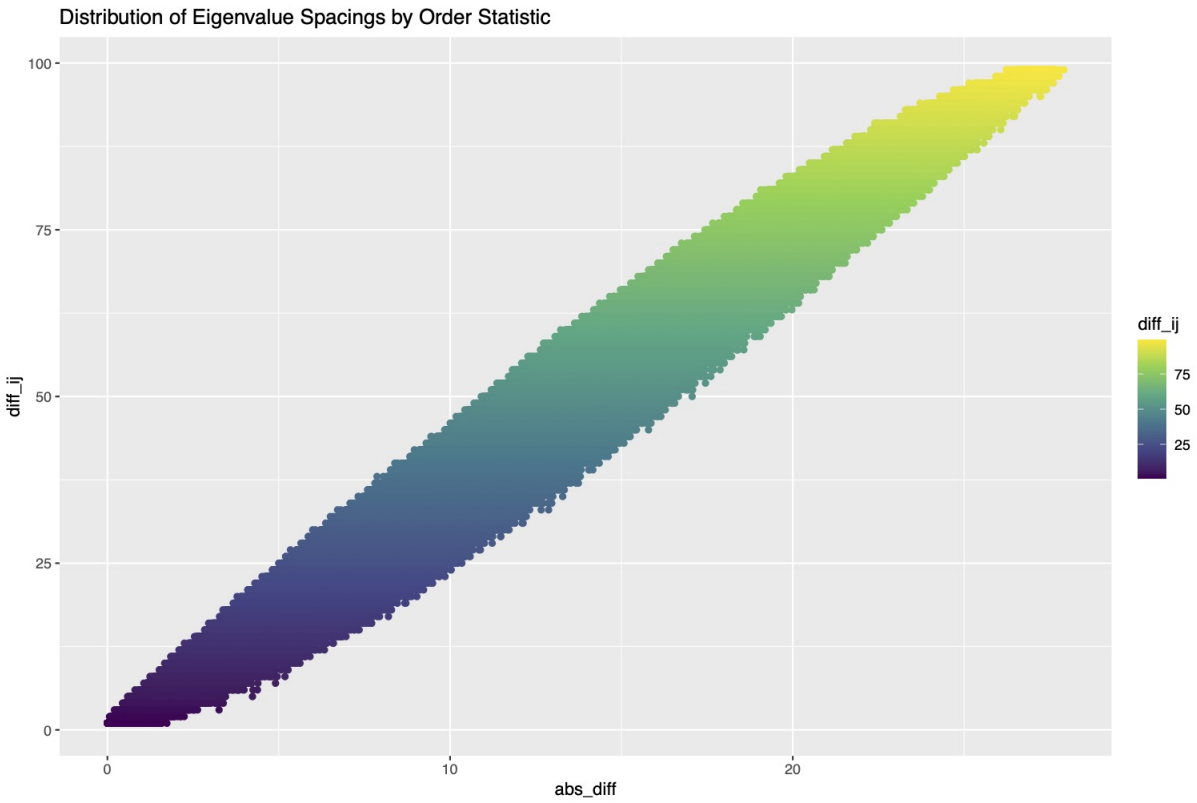


4.3 Dispersions

In this section, we will take a look at some dispersion statistics of the β-ensembles.

4.3.1 Order Statistics

Firstly, we will consider the dispersion $\Delta_{\text{abs}}(\mathcal{E} \mid \Pi_<)$ where $\mathcal{E} \sim \mathcal{H}(\beta = 4)$ is an ensemble of 100×100 matrices. Since we are using $\Pi_<$, we can say that we are interested in the difference of absolutes for an arbitrary pair of eigenvalues. Additionally, assume that we are using the norm-ordering scheme.



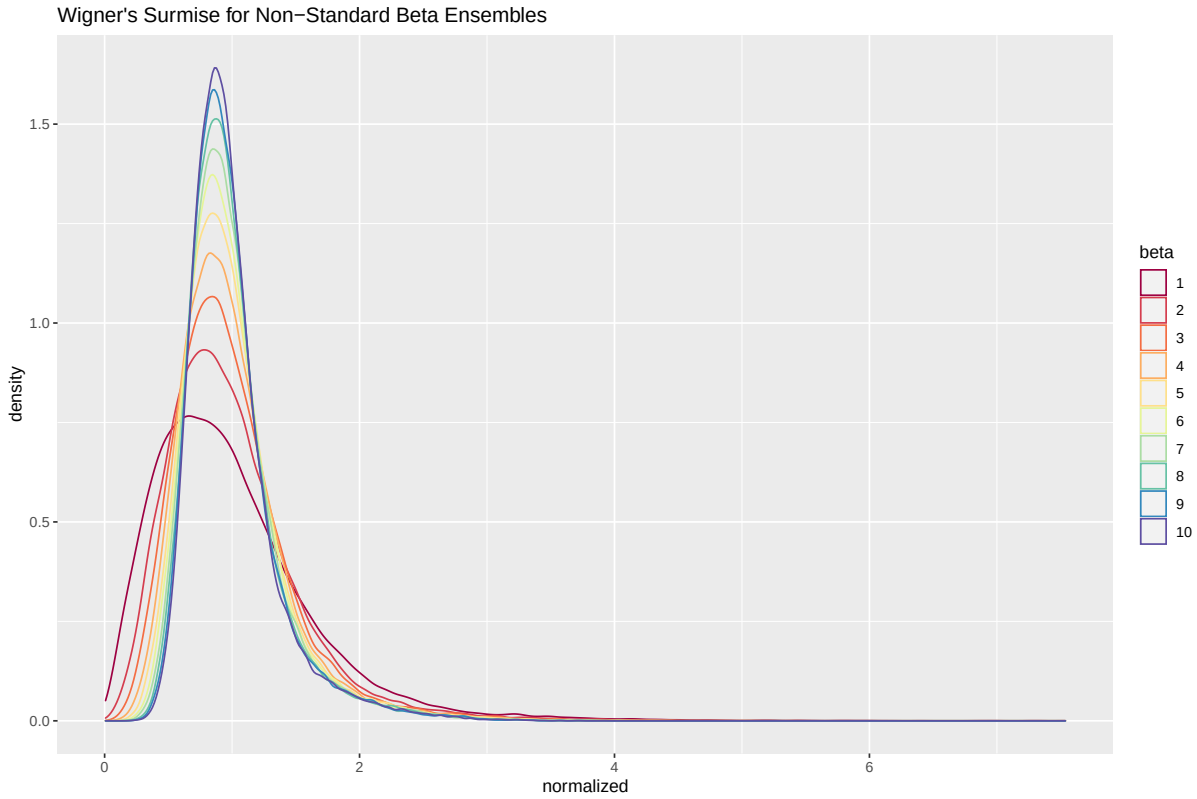
In the plot above, we are considering the absolute differences dispersion metric $\delta = \delta_{\text{abs}}$. Additionally, we are using the lower pairing scheme Π_C so we can interpret these pairs as the pairs we would obtain from randomly picking two distinct eigenvalues. Additionally, since we are using the norm-ordering scheme, this means that the ranking difference tells us the difference of sizes of two eigenvalues that are ranked a certain distance ρ apart.

As such, this explains why the bulk about $\rho = 50$ is the widest. This is because it represents the dispersion between pairs of eigenvalues in the origin and those near the boundary. Analogously, this explains why both $\rho = 0$ and $\rho = 100$ exhibit the narrowest dispersions. This is because they are comparing the largest eigenvalues with themselves (since the Semicircle distribution is symmetric). Lastly, note that this graph is symmetric about the identity line.

4.3.2 Wigner's Surmise

Extending the β-Ensembles

Below, we extend the simulations of the Wigner dispersions as discussed in **Section 3.3** beyond the standard values of $\beta = 1, 2, 4$ to any $\beta \in \mathbb{N}$.



Takeaways

As mentioned in **Section 3.3.2**, we notice that there is a general pattern of decreasing variance, indicating that the pattern is consistent for other values of β! In other words, this seems to provide evidence that as $\beta \to \infty$, the variance of the consecutive dispersions shrinks, $\text{Var}(\delta(\tilde{\pi})) \to 0$. Recall from the physical interpretation of the model in **Section 4.1.3**, we discuss the limiting behaviour of the ensemble as $\beta \to \infty$. Specifically, it was mentioned that as β grows, the ensemble represents a deterministic system of particles of low temperature and maximum interaction which tries to align itself equidistantly. This would confirm our prediction, because an equidistant system of particles is precisely a system of point charges where there is no variance in the distance between consecutive particles, or in other words — where $\text{Var}(\delta(\tilde{\pi})) = 0$!

Conclusion

In this thesis, the concepts of random matrix distribution, ordered spectra, and eigenvalue dispersions were rigorously formalized. While those formulations were inspired in part by the pragmatic demand of making rigourous the mathematical methods of the RMAT package, they by no means are restricted to the domain of the package. Namely, the concepts of $\mathcal{D}$-distributions and eigenvalue ordering are timelessly critical concepts to understand in order to think about random matrices. Furthermore, carefully choosing dispersion metrics and selecting subsets of eigenvalue pairs in a systematic fashion is a necessary hurdle one must cross before even starting to study dispersions.

All in all, this toolkit of definitions and formalizations is a toolkit that will hopefully make any sort of communication involving random matrices unambiguous. While these definitions appear polished in this thesis, they were always seeds of something smaller. So, do not expect to understand everything immediately. The document does try to accomodate this by sprinkling thoughtful examples, demonstrations, and remarks, so be sure to leave no stone unturned. So for having reached the end, the reader should be proud. You are certainly more ready to delve in the deep waters of random matrix theory.

Other than the formalizations, in this thesis we survey several results through analyzing and visualizing simulations. In Chapter 2, we provide a heuristic demonstration of the Perron-Frobenius Theorem for Markov Chains by using computational evidence. Additionally, we find that the radius of the complex disk of p-Erdos matrix ensemble is inversely proportional to p; this demonstrates the relationship between mixing time (longer for sparser graphs) and the eigenvalues of stochastic matrices. Also in Chapter 2, we survey symmetric/hermitian matrices and empirically describe their spectra. Then, we take in a case study of what turns out to be the GOE ($\beta = 1$).

In Chapter 3, after we formalize dispersions, we simulate Wigner's surmise - a statement about the distribution of normalized consecutive eigenvalues of symmetric matrices. We simulate the distributions for the three standard β-ensembles before we launch into the next chapter. In Chapter 4, we bring everything together and make a case study out of the β-ensembles. After discussing the Wigner's surmise result at the end of Chapter 3, we formally define and motivate the β-ensemble by its joint eigenvalue p.d.f. Then, we discuss more abstract characterizations of the ensembles such as the invariance criterion and the physical interpretation of the model. The chapter ends by surveying some spectral statistics and extending the Wigner's surmise model for any $\beta \in \mathbb{N}$.

Open Questions

Spectral Statistics. What are the spectral statistics of the following $\mathcal{D}$-distributions?

1. The following e.h. distributions
   (a) Poisson
   (b) Beta
   (c) Gamma
2. Band matrices with various $\mathcal{D}$-distributions.
   (a) General k-band matrices
   (b) Tridiagonals ($k = 1$)

Dispersions. For the generalized β-ensembles, what are the moments of the Wigner Dispersion distributions?

Other. How can Wishart matrices (non-square Normal matrices) be studied? Generally, the approach to doing so is by using singular values. However, the question remains, what are their spectral statistics?

Appendix A

Math Review

In the first section of this appendix, we list a few items in linear algebra worth reviewing to make sure there are no gaps in knowledge necessary to understand concepts in this thesis. For a helpful reference, consult [Horn (2012)]. Additionally, there is another subsection including a proof that was written in the research process of this thesis. Albeit this proof has no direct relation to the primary topic, spectral statistics, it is loosely relevant and displays the tools and techniques relevant to the material.

A.1 Linear Algebra

A.1.1 Matrices

Definition A.1.1 (Eigenvalue)**.** *Suppose $P \in \mathbb{F}^{n\times n}$ is a square matrix. Then, the eigenvalues of the matrix P are precisely the roots of the characteristic polynomial of P, given by $char_P(\lambda) = \det(P - \lambda I)$. The polynomial $char_P(\lambda)$ has degree n. So by the Fundamental Theorem of Algebra, P has a multiset of n eigenvalues.*

Definition A.1.2 (Inverse Matrix)**.** *Suppose there is a matrix $P \in \mathbb{F}^{m\times n}$. Then, P^{-1} is its inverse matrix iff multiplying it by P returns the identity matrix. That is, the inverse matrix must satisfy:*

$$P^{-1} \text{ is the inverse of } P \iff P^{-1}P = I$$

Definition A.1.3 (Transpose Matrix)**.** *Suppose there is a matrix $P = (p_{ij}) \in \mathbb{F}^{m\times n}$. Then, its transpose matrix, $P^T = (t_{ij}) = (p_{ji}) \in \mathbb{F}^{n\times m}$ matrix whose columns are the rows of the original matrix.*

Definition A.1.4 (Conjugate Transpose Matrix)**.** *Suppose there is a matrix $P = (p_{ij}) \in \mathbb{C}^{m\times n}$. Then, its conjugate transpose matrix, $P^\dagger = (t_{ij}) = (\overline{p_{ji}}) \in \mathbb{F}^{n\times m}$ is a matrix whose columns are the rows of the original matrix.*

Definition A.1.5 (Symmetric Matrix)**.** *A matrix P is symmetric iff it is equal to its transpose:*

$$P \text{ is Symmetric} \iff P = P^T$$

Definition A.1.6 (Hermitian Matrix)**.** *A matrix P is Hermitian iff it is equal to its conjugate transpose:*

$$P \text{ is Hermitian} \iff P = P^\dagger$$

Definition A.1.7 (Orthogonal Matrix)**.** *A matrix P is called orthogonal iff its transpose is its inverse:*

$$P \text{ is Orthogonal} \iff P^T = P^{-1}$$

Definition A.1.8 (Unitary Matrix)**.** *A matrix P is called unitary iff its conjugate transpose is its inverse:*

$$P \text{ is Unitary} \iff P^\dagger = P^{-1}.$$

A.1.2 Other

For β-ensembles, we mention the invariance criterion, which is defined in lieu of the following groups.

Theorem A.1.1 (Orthogonal Group)**.** *The set of* all *orthogonal matrices in $\mathbb{F}^{n\times n}$ is a matrix group. It is called the orthogonal group.*

Theorem A.1.2 (Unitary Group)**.** *The set of* all *unitary matrices in $\mathbb{C}^{n\times n}$ is a matrix group. It is called the unitary group.*

As mentioned in **Section 2.3**, here is the statement of the Perron-Frobenius Theorem. Specifically, we consider only the application to ergodic Markov Chains by means of limiting scope to stochastic/transition matrices. For a reference, see [Levin (2008)].

Theorem A.1.3 (Perron-Frobenius Theorem)**.** *The Perron–Frobenius theorem asserts that a real square matrix with positive entries has a unique largest real eigenvalue and that the corresponding eigenvector can be chosen to have strictly positive components.*

A.1.3 Proof: Real Symmetric Matrices have Real Eigenvectors

How does this relate to the rest of the thesis?

Eigenvectors have a wide range of applications in various fields. As such, a result **guaranteeing** the existence of real-valued eigenvectors is a strong, non-trivial one. For instance, the stationary distribution of a Markov chain is an eigenvector of its transition matrix of eigenvalue 1. In other scenarios, right-eigenvectors represent the conditional expectation of a transition matrix, and so forth. Altogether, this proof showcases techniques and tools used in this related domains of study.

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

The Proof

In this section, we will prove that for any $M \times M$ real symmetric matrix, $S_M \in \mathbb{R}^{M \times M}$, there exists for some eigenvalue λ, a corresponding **real** eigenvector $\vec{v} \in \mathbb{R}^M$. Prior to starting the main proof, we begin by proving a auxiliary lemma.

Lemma. Suppose we have a $M \times M$ real symmetric matrix with some eigenvalue λ. If there we have a corresponding 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:

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

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

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

We obtain (1) by expanding the equality $A\vec{v} = \lambda\vec{v}$ 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 $d_i v_i$. Next, we collect the terms to solve for v_i:

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

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

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

Finally, divide both sides by $(\lambda - d_i)$. On the right hand side, the coefficients of v_j become $\frac{a_j}{(\lambda - d_i)}$. Denote this coefficient $a'_j = \frac{a_j}{(\lambda - d_i)}$. Since S_M is a real symmetric matrix, we know its eigenvalues are real so $\lambda \in \mathbb{R}$ and that its entries are real so

$a_i, d_i \in \mathbb{R}$. Since a'_j is an arithmetic expression involving real numbers, then it follows that for every j, $a'_j \in \mathbb{R}$. As such, we can rewrite v_j as the following sum.

$$\forall i \in \mathbb{N}_M : v_i = \sum_{j \neq i} c_j v_j \quad (\forall j : a'_j \in \mathbb{R})$$

Thus, for any $M \in \mathbb{N}$, a real symmetric matrix $S_M \in \mathbb{R}^{M \times M}$ with eigenvalue λ must have a corresponding eigenvector v such that each of its entries is expressible as a real linear combination of the other entries. So, the proof is complete. $\square$

Now, leveraging the result of this lemma, we will prove the main theorem.

Theorem A.1.4 (Taqi)**.** *Suppose we have a $M \times M$ real symmetric matrix, S_M. Then, we will show that there exists for some eigenvalue λ, a corresponding **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:

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

Base Case. Take the base case $M = 2$. This proof is left to the reader as an exercise. Begin by taking a 2×2 symmetric matrix, and show that there exist real coefficients for the eigenvector corresponding to λ by using Gaussian Elimination.

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 $v \in \mathbb{R}^M$ corresponding to λ.

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 lemma to obtain (1) and our assumption of $f(M)$ to obtain (2), listed below:

$$(1) : \forall i \in \mathbb{N}_{M+1} : v_i = \sum_{j \neq i} c_j v_j \quad (c_j \in \mathbb{R})$$

$$(2) : \forall i \in \mathbb{N}_M : v_i \in \mathbb{R}$$

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 the $(m+1)^{th}$ row, $v_{m+1} = \sum_{j \neq m+1} c_j v_j$ for real coefficients $c_j \in \mathbb{R}$. By (2), this is a linear combination of real entries v_i. Since $v_{m+1} \in \mathbb{R}$, this means we have shown that:

$$\forall i \in \mathbb{N}_{M+1} : v_i \in \mathbb{R}$$

In other words, we have established that $f(M) \Rightarrow f(M+1)$. By induction, the proof of the theorem is complete. $\square$.

A.2 Probability Theory

Please refer to [Hwang & Blitzstein (2019)] as a resource; the definitions were sourced from there. For the definitions and theorems covered in Section A.1, consider checking out Chapter 3: Random Variables and their distributions. For Order Statistics, consider Chapter 8, Section 6: Order Statistics.

A.2.1 Random Variables

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 corresponding to the random variable, p_X exhausts over all the events in Ω. In other words, we expect $\int_\Omega p_X(s) = 1$.*

PDFs

Definition A.2.2 (Probability Density Function)**.** *For a continuous r.v. X with CDF F, the probability density function (PDF) of X is the derivative f of the CDF, given by $f(x) = F'(x)$. The support of X, and of its distribution, is the set of all x where $f(x) > 0$.*

Theorem A.2.1 (Characterizing the PDF)**.** *A probability density function is characterized by a few properties that are necessary for it to be valid. They are as follows:*

1. *Non-negativity: the PDF must be a non-negative valued function everywhere.*

$$f(x) \geq 0$$

2. *Integrates to 1: the PDF must integrate to 1 when integrated over its entire support.*

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Example (Normal PDF)**.** *For example, consider the probability density function for a r.v. $X \sim \mathcal{N}(\mu, \sigma)$.*

$$\mathbb{P}(X = x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

CDFs

Definition A.2.3 (Cumulative Distribution Function)**.** *The cumulative distribution function (CDF) of an r.v. X is the function F_X given by $F_X(x) = \mathbb{P}(X \leq x)$. When there is no risk of ambiguity, we sometimes drop the subscript and just write F (or some other letter) for a CDF.*

Theorem A.2.2 (Characterizing the CDF)**.** *A cumulative distribution function is characterized by a few properties that are necessary for it to be valid. They are as follows:*

1. *Monotonic function: the CDF must always be a monotonically increasing function. That is:*
$$x_0 \leq x_1 \implies F(x_0) \leq F(x_1)$$
2. *Right-continuous: The CDF is continuous except possibly for having some jumps. Wherever there is a jump, the CDF is continuous from the right. That is, for any a, we have:*
$$F(a) = \lim_{x \to a^+} F(x)$$
3. *Converges to 0 and 1 in its limits: since the CDF represents a cumulative probability, its limits must reflect that aspect of probability spaces. So, the CDF must satisfy:*
$$\lim_{x \to -\infty} F(x) = 0 \text{ and } \lim_{x \to \infty} F(x) = 1$$

Independence

Definition A.2.4 (Independence)**.** *Random variables X and Y are said to be independent if $\forall x, y \in \mathbb{R}$:*
$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \cdot \mathbb{P}(Y \leq y)$$
If the variables are discrete, then this is equivalent to the condition:
$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$$

Definition A.2.5 (i.i.d)**.** *A vector of random variables $\vec{X} = (X_i)_{i=1}^N$ is said to be i.i.d (independent and identically distributed) is each of its entries X_i are exactly so.*

A.2.2 Statistics

Definition A.2.6 (Statistic)**.** *A statistic is formally defined as a function of a vector of random variables. So, for instance f is a statistic of a vector of random variables $\bar{X}$. Its observed value is given by $f(\vec{X})$.*

Example (Mean Statistic)**.** *For instance, the mean of a random sample $\vec{X}$ is a statistic. Formally, we would define the statistic $f : \vec{X} \to \frac{\sum_i x_i}{N}$. So the mean of the random sample is defined as $\bar{x} = f(\vec{X})$.*

Order Statistics

Definition A.2.7 (Order Statistic)**.** *Suppose $\vec{X} = \{X_i\}_{i=1}^N$ is an ordered vector of random variables. Then, the i^{th} order statistic of X is given by X_i.*

Example (Order Statistic)**.** *Suppose $X = (20, 7, 2, 1)$. The smallest (fourth) order statistic is 1. The second order statistic is 7. The largest (first) order statistic is 20.*

Theorem A.2.3 (Order Statistics of Uniform i.i.d Variables)**.** *Let $U_0, U_1, \ldots, U_n$ be i.i.d. Unif$(0, 1)$. Then, the j^{th} order statistic $U_{(j)}$, is distributed as $U_{(j)} \sim Beta(j, n - j + 1)$.*

A.3 Markov Chains

In this section, we will define and discuss Markov chains, which are a very useful construct with countless applications. In short, a Markov chain is a stochastic model that describes a sequence of possible events in which the probability of each event is independent of every other state **except** the state directly preceding it. Markov chains are embedded in this thesis through our study of stochastic/transition matrices. Additionally, they also show up in **Appendix D** when we talk about the mixing time of Markov chains. There are many types of Markov chains, but we will stick to a version that is ubiquitous and canonical.

So within the scope of this thesis, we only care about discrete, time-homogeneous Markov chains. Specifically, we will be considering Markov chains that represent a random walk on a fixed graph. This is a simple Markov chain with two characterizing sets of parameters: the number of vertices, and the weights between each vertex. In this representation, each vertex represents a state and each edge represents a transition probability. This way, we can obtain an $N \times N$ transition matrix for any graph with N given that we know every transition probability. This will be formalized further, but the idea is that each random stochastic matrix will represent a graph with random weights.

Please refer to [Hwang & Blitzstein (2019)], Chapter 11, for more information on 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 satisfied:*

(Closed Support) $\forall X_i : X_i$ has support and range $S_M = \{1, 2, ..., M\}$.

(Markov Property) The transition probability from state $i \to j$, given by $\mathbb{P}(X_{n+1} = j \mid X_n = i)$ 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. So, $\forall i, j \in S_M$, we observe:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i', \ldots, X_0 = i^{(n-1)})$$

Transition Matrices

Definition A.3.2 (Transition Matrix)**.** *Take a Markov Chain with states* $\{1, \ldots, M\}$*. Letting* $q_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$ *be the transition probability from* $i \to j$*, then the matrix* $Q = (q_{ij})$ *is the transition matrix of the chain. That being said, transition matrices must satisfy a few conditions such as those mentioned below.*

(Nonnegativity) Q is a non-negative matrix. So every entry $q_{ij} \in \mathbb{R}^+$ is non-negative. This follows because probabilities are necessarily non-negative values.

(Stochasticity) For this transition matrix to be valid, its rows have to be stochastic, meaning their entries sum to 1. This may be understood as applying the law of total probability to the event of transitioning from any given state $i \in S_M$. In other words, the chain has to go somewhere with probability 1.

$$\forall i \in \mathbb{N}_M : \sum_{j \in \mathbb{N}_M} q_{ij} = 1$$

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.3.3 (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* $q_{ij}^{(n)}$ *:*

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

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.3.4 (Marginal Distribution of X_n)**.** *Let* $\vec{\pi} = (\pi_1, \pi_2, \ldots, \pi_M)$ *represent a vector of probabilities such that for every state* i*,* $\pi_i = \mathbb{P}(X_0 = i)$ *represents the initial probability of being at state* i*. Then, the marginal distribution of* X_n *(the distribution of the chain at time* n*) is a vector where the* j^{th} *component is* $\mathbb{P}(X_n = j)$ *for* $j \in S_M$*. The marginal distribution of* X_n *is precisely given the vector* $\vec{\pi}Q^n \in [0, 1]^M$*. So, we call* $\vec{\pi}$ *an initial state distribution, and* $\vec{\pi}Q^n$ *the state of the distribution at time* n*.*

Classification of states

Having defined Markov chains and their relationship to transition matrices, we now discuss a few important properties of Markov chains through the process of classifying the behavior of some of its states.

(Recurrent State) 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.

That being said, here is a caveat to take into consideration. As we let $n \to \infty$, our Markov chain will guarantee 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 ε, then realizing that by the support of $\text{Geom}(\varepsilon)$ is always some finite value. Then, by the equivalence between the *Markov property* and independent Geometric trials, we are guaranteed the existence of some finite value such that there is a success of never returning to i.

Definition A.3.5 (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, we can equivalently say that a chain is *irreducible* if there is an integer $n \in \mathbb{N}$ such that the (i, j) entry of Q^n is positive for any i, j. That being said, we define a Markov chain to be **reducible** if it is not **irreducible**. Using our quantifier formulation, it means that it suffices to find *transient* states so that:

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

Lastly, another state feature that is important to our discussion is periodicity. In simple words, the period of a state is defined as an upper bound on the time needed to return to a state. First, we define the period of a given state.

Definition A.3.6 (Period)**.** *A state i has period k if any return to state i must occur in k steps. Formally, the period of state i, denoted ϕ_i is defined as follows:*

$$\phi(i) = \{\gcd n \mid p_{ii}^n > 0\}$$

That being said, if every state of the chain has a period of 1, then we call it **aperiodic**. Having defined irreducibility and aperiodicity, we are finally able to define our primary objects of interest: ergodic Markov chains.

Ergodicity

Lastly, we come to our discussion of ergodicity. Ergodicity is a term used in statistical physics to describe stochastic systems which are thoroughly "mixed". In an ergodic system, every state is expected to be visited uniformly and randomly. Roughly speaking, we could also say ergodicity is akin to a state of equilibrium. In fact, the origin of the word comes from "ergodic theory" developed by Boltzmann, who was interested in statistical mechanics. That being said, we define an ergodic Markov chain as follows.

Definition A.3.7 (Ergodic Markov Chain)**.** *An ergodic Markov chain is a Markov chain whose states are irreducible and aperiodic. Ergodic Markov chains guarantee the existence of a unique steady-state vector $\vec{\pi}$ called the stationary distribution, which represent a left eigenvector of the corresponding transition matrix. Additionally, if we denote π_i as the steady-state probability of state i and $\eta(i,t)$ as the number of visits to state i in the time t, then:*

$$\lim_{t\to\infty} \frac{\eta(i,t)}{t} = \pi(i)$$

In other words, $\pi(i)$ represents the average time the Markov chain spends at state i in the long term, which converges as a result of erodicity or "thorough mixing".

Informally, irreducibility ensures that there is a sequence of transitions of non-zero probability from any state to any other. Aperiodicity, on the other hand, ensures that the states are not partitioned into sets such that all state transitions occur cyclically from one set to another. That being said, this wraps up our discussion of Markov chains for this section. This subject is revisited in the mixing time simulation chapter in **Appendix D**, where we discuss the time it takes a matrix to send an arbitrary probability vector to the stationary distribution.

Appendix B

Algorithm Appendix

In this appendix, we will enumerate the various algorithms used in this thesis for the purpose of transparency and reproducibility. Every one of these algorithms have their code counterpart in the **RMAT** package. For the most part, the implicit $\mathcal{D}$-matrices are the non-trivial algorithms to peruse. Changing methodology could impact results, and as such, all results are tied to the algorithms in this appendix.

B.1 Implicit $\mathcal{D}$-Matrices

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

1. *To sample a row r of size N, fix $N \in \mathbb{N}$.*
2. *Sample a vector $\vec{X}$ with N i.i.d entries between $[0, 1]$. So, sample $\vec{X} \sim Unif(0, 1)$.*
3. *Assign $r \leftarrow \vec{X}$, and then normalize the row by dividing each entry by the row sum; so assign $(r) \leftarrow (r)/\sum_{j=1}^{N} r_j$.*
4. *Return the stochastic row r.*

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

1. *To generate a stochastic square matrix P of size $N \times N$, fix $N \in \mathbb{N}$.*
2. *Then, for every row of P, randomly sample a stochastic row of size N and assign it.*
3. *Return the stochastic matrix P.*

Note that for Erdos-Renyi transition matrices, we cannot set the degree of the vertex to be 0. Otherwise, the matrix cannot be stochastic. As such, we sever $N-1$ vertices **at most**. This means that these matrices do not formally represent walks on Erdos-Renyi graphs, but rather an extremely close approximation of them.

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

1. *Fix $N \in \mathbb{N}$ and $p \in [0,1]$.*
2. *Generate a matrix $Q \sim Unif(0,1)$, i.e. with randomly uniform entries on [0,1].*
3. *For each row r in $\{1,\ldots,N\}$, generate $deg(r) \sim Bin(N-1,p)$.*
4. *Randomly choose $N - deg(r)$ vertices, then set the entries r_j in the j columns to 0 to sever them.*
5. *Renormalize the matrix by dividing each row by its sum; let $(r) \leftarrow (r)/\sum_{j=1}^{N}(r_j)$.*

B.2 Explicit $\mathcal{D}$-Matrices

Algorithm B.2.1 (Homogeneous Explicit $\mathcal{D}$-Matrix)**.**

1. *To simulate a $\mathcal{D}$-distributed square matrix P of size N, fix $N \in \mathbb{N}$.*
2. *Sample a vector $\vec{X}$ with N i.i.d entries from $\mathcal{D}$.*
3. *Assign the vector $\vec{X}$ as a row of the matrix P. Repeat for every other row.*
4. *Return the $\mathcal{D}$-distributed matrix P.*

Algorithm B.2.2 (Dumitriu's Beta Matrix)**.**

1. *To simulate an $N \times N$ beta matrix, 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 $\chi(df = c_j)$ where $c_j = \beta \cdot j$ for columns spanning $j = 1,\ldots,N-1$.*
4. *Normalize the entries by dividing by $\sqrt{2}$.*

Appendix C

Code Appendix

The RMAT Package.
The RMAT package is composed of two primary modules, the random matrix module (matrices.R) and the spectral statistics module (spectrum.R) and (dispersion.R). They can be further subdivided into smaller modules as follows.
Note. The source code was functional as of R 4.0.5.

1. **Random Matrix Module**
   (a) Explicitly Distributed Matrices
   (b) Implicitly Distributed Matrices
   (c) Ensemble Extensions
2. **Spectral Statistics Module**
   (a) Spectrum
   (b) Dispersions
   (c) Parallel Extensions

C.1 Matrix Module

Random matrices can either be explicitly or implicitly distributed. If they are explicitly distributed, their entries have a specific distribution. Otherwise, the entries have an implicit distribution imposed by the generative algorithm the matrix uses.

C.1.1 Explicitly Distributed Matrices

For (homogeneous) explicitly distributed matrices, we can use a "function factory" method to be concise. The actual implementation is more verbose for the purposes of argument documentation, but the following code is minimal and fully functional.

Homogeneously Explicit Distributions

```
# ... represents all the arguments taken in by the rdist function
RM_explicit <- function(rdist){
  function(N, ..., symm = FALSE){
    # Create an [N x N] matrix sampling the rows from rdist, passing ... to rdist
    P <- matrix(rdist(N^2, ...), nrow = N)
    # Make symmetric if prompted
    if(symm){P <- .makeHermitian(P)}
    # Return P
    P
  }
}

# A version where we add an imaginary component
RM_explicit_cplx <- function(rdist){
  RM_dist <- function(N, ..., symm = FALSE, cplx = FALSE, herm = FALSE){
    # Create an [N x N] matrix sampling the rows from rdist, passing ... to rdist
    P <- matrix(rdist(N^2, ...), nrow = N)
    # Make symmetric/hermitian if prompted
    if(symm || herm){P <- .makeHermitian(P)}
    # Returns a matrix with complex (and hermitian) entries if prompted
    if(cplx){
      # Recursively add imaginary components as 1i * instance of real-valued matrix.
      Im_P <- (1i * RM_dist(N, ...))
      # Make imaginary part Hermitian if prompted
      if(herm){P <- P + .makeHermitian(Im_P)}
      else{P <- P + Im_P}
    }
    P # Return the matrix
  }
}
```

With our function factories set up, we can quickly generate all the random matrix functions for all the distributions our hearts could desire.

```
RM_unif <- RM_explicit_cplx(runif)
RM_norm <- RM_explicit_cplx(rnorm)
```

Beta Matrices

For the β-ensemble matrices, we simply use the algorithm provided in Dimitriu's paper. Doing so, we get the function:

```
# Generate a Hermite beta matrix using Dumitriu's Matrix Model
RM_beta <- function(N, beta){
  # Set the diagonal ~ N(0,2)
  P <- diag(rnorm(n = N, mean = 0, sd = sqrt(2)))
  # Get degrees of freedom sequence for offdigonal
  df_seq <- beta * (N - seq(1, N-1))
  # Set the off-1 diagonals as chi squared variables with df(beta_i)
  P[row(P) - col(P) == 1] <- P[row(P) - col(P) == -1] <- sqrt(rchisq(N-1, df_seq))
  # Rescale the entries by 1/sqrt(2)
  P <- P/sqrt(2)
  # Return the beta matrix
  P
}
```

Helper Functions

```
# Returns a Hermitian version of a matrix by manual assignment
.makeHermitian <- function(P){
  for(i in 1:nrow(P)){
    for(j in 1:ncol(P)){
      # Select the entries in the upper triangle (i < j)
      if(i < j){
        # Make the upper triangle equal to the conjugate transpose of the lower triangle
        P[i,j] <- Conj(P[j,i])
      }
    }
  }
  # Return the Hermitian Matrix
  P
}

# Return the off-diagonal entries of row i
.offdiagonalEntries <- function(row, row_index){row[which(1:length(row) != row_index)]}
```

C.1.2 Implicitly Distributed Matrices

Stochastic Matrices. For stochastic matrices, we require slightly more setup. First, we setup the row functions to sample probability vectors:

```
# Generates stochastic rows of length N
.stoch_row <- function(N){
  # Sample a vector of probabilities
  row <- runif(n = N, min = 0, max = 1)
  # Return the normalized row (sums to one)
  row / sum(row)
}
```

For randomly introduced sparsity, we define the following row function.

```
# Generates same rows as in .stoch_row(N), but with randomly introduced sparsity
.stoch_row_zeros <- function(N){
  # Sample a vector of probabilities
  row <- runif(n = N, min = 0, max = 1)
  # Sample a vertex degree of at least one (as to ensure row is stochastic)
  degree_vertex <- sample(x = 1:(N-1), size = 1)
  # Sever a random selection of edges to set the vertex degree
  row[sample(1:N, size = N - degree_vertex)] <- 0
  # Return normalized row
  row / sum(row)
}
```

Once this is done, we can use this function iteratively. With some magic, we can incorporate an option to make the matrix symmetric, and we get the following function.

```
RM_stoch <- function(N, symm = F, sparsity = F){
  # Choose row function depending on sparsity argument
  if(sparsity){row_fxn <- .stoch_row_zeros} else {row_fxn <- .stoch_row}
  # Generate the [N x N] stochastic matrix stacking N stochastic rows
  P <- do.call("rbind", lapply(X = rep(N, N), FUN = row_fxn))
  # Make symmetric (if prompted)
  if(symm){ P <- .makeStochSymm(P) }
  # Return the matrix
  P
}
```

Erdos-Renyi Stochastic Matrices. For the Erdos-Renyi walks, we do something similar by defining a parameterized row function.

```
# Generates a stochastic row with parameterized sparsity of p
.stoch_row_erdos <- function(N, p){
  # Sample a vector of probabilities
  row <- runif(n = N, min = 0, max = 1)
  # Sample the vertex degree so that it is ~ Bin(n,p)
  degree_vertex <- rbinom(n = 1, size = N, prob = 1 - p)
  # Sever a random selection of edges to set the vertex degree
  row[sample(1:N, degree_vertex)] <- 0
  # Return normalized row only if non-zero (cannot divide by 0)
  if(sum(row) != 0){
    row / sum(row)
  } else{
    .stoch_row_erdos(N, p) # Otherwise, try again
  }
}
```

And we again use the row function iteratively to get the following function.

```
RM_erdos <- function(N, p, stoch = T){
  # Generate an [N x N] Erdos-Renyi stochastic matrix by stacking N p-stochastic rows
  P <- do.call("rbind", lapply(X = rep(N, N), FUN = .stoch_row_erdos, p = p))
  # Return the Erdos-Renyi transition matrix
  P
}
```

And as such, we have minimal, functional implementations of functions that sample random matrices!

C.1.3 Ensemble Extensions

Lastly, we have the ensemble extensions. These functions are quite simple to implement user a "function factory". Again, the actual implementations are more verbose due to the argument descriptions, but otherwise, are exactly the same.

```
# Extends a RM_dist function to its RME_dist ensemble counterpart
RME_extender <- function(RM_dist){
  # Function returns a list of replicates of the RM_dist function with '...' as arguments
  function(N, ..., size){
    lapply(X = rep(N, size), FUN = RM_dist, ...)
  }
}
```

Now, we extend the functions as follows, and we are done with the matrix module!

```
RME_unif <- RME_extender(RM_unif)
RME_norm <- RME_extender(RM_norm)
RME_beta <- RME_extender(RM_beta)
RME_stoch <- RME_extender(RM_stoch)
RME_erdos <- RME_extender(RM_erdos)
```

C.2 Spectral Statistics Module

C.2.1 Spectrum

```
spectrum <- function(array, components = T, norm_order = T, singular = F, order = NA){
  # Digits to round values to
  digits <- 4
  # Get the type of array
  array_class <- .arrayClass(array)
  # For ensembles, iteratively rbind() each matrix's spectrum
  if(array_class == "ensemble"){
    map_dfr(array, .spectrum_matrix, components, norm_order, singular, order, digits)
  }
  # From matrices, call the function returning the ordered spectrum for a singleton matrix
  else if(array_class == "matrix"){
    .spectrum_matrix(array, components, norm_order, singular, order, digits)
  }
}
```

```
# Helper function returning tidied eigenvalue array for a matrix
.spectrum_matrix <- function(P, components, norm_order, singular, order, digits = 4){
  # For singular values, take P as product of the itself and its tranpose
  if(singular){P <- P %*% t(P)}
  # Get the eigenvalues of P
  eigenvalues <- eigen(P, only.values = TRUE)$values
  # Take the square root of the eigenvalues to obtain singular values
  if(singular){eigenvalues <- sqrt(eigenvalues)}
  # Sort the eigenvalues to make it an ordered spectrum
  eigenvalues <- .sortValues(eigenvalues, norm_order)
  # If uninitialized, select all orders; otherwise, use c() so singletons => vectors
  if(class(order) == "logical"){order <- 1:nrow(P)} else{order <- c(order)}
  # Return the spectrum of the matrix
  purrr::map_dfr(order, .resolve_eigenvalue, eigenvalues, components, digits)
}
```

```
# Read and parse an eigenvalue from a sorted eigenvalue array
.resolve_eigenvalue <- function(order, eigenvalues, components, digits){
  # Read from a sorted eigenvalue array at that order
  eigenvalue <- eigenvalues[order]
  # Get norm and order columns
  features <- data.frame(Norm = abs(eigenvalue), Order = order)
  if(components){
    # If components are sought, resolve the eigenvalue into seperate columns first
    res <- cbind(data.frame(Re = Re(eigenvalue), Im = Im(eigenvalue)), features)
  } else{
    # Otherwise, don't resolve the eigenvalue components
    res <- cbind(data.frame(Eigenvalue = eigenvalue), features)
  }
  # Round entries and return the resolved eigenvalue
  res <- round(res, digits)
  return(res)
}
```

Helper Functions

```
# Parses an array to see classify it as a matrix or an ensemble of matrices.
.arrayClass <- function(array){
  # Sample an element from the array and get its class
  elem <- array[[1]]
  types <- class(elem)
  # Classify it by analyzing the element class
  if("numeric" %in% types || "complex" %in% types){
    return("matrix")
  }
  else if("matrix" %in% types){
    return("ensemble")
  }
}

# Sort an array of numbers by their norm (written for eigenvalue sorting)
.sortValues <- function(vals, norm_order){
  values <- data.frame(value = vals)
  # If asked to sort by norms, arrange by norm and return
  if(norm_order){
    values$norm <- abs(values$value)
    values <- values %>% arrange(desc(norm))
    # Return the norm-sorted values
    values$value
  }
  # Otherwise, sort by sign and return
  else{ sort(vals, decreasing = TRUE) }
}
```

C.2.2 Dispersions

```
# Compute the dispersion of a matrix or matrix ensemble
dispersion <- function(array, pairs = NA, norm_order = T, singular = F, pow_norm = 1){
  # Digits to round values to
  digits <- 4
  # Get the type of array
  array_class <- .arrayClass(array)
  # Parse input and generate pair scheme (default NA), passing on array for dimension
  pairs <- .parsePairs(pairs, array, array_class)
  # For ensembles; iteratively rbind() each matrix's dispersion
  if(array_class == "ensemble"){
    map_dfr(array, .dispersion_matrix, pairs, norm_order, singular, pow_norm, digits)
  }
  # Array is a matrix; call function returning dispersion for singleton matrix
  else if(array_class == "matrix"){
    .dispersion_matrix(array, pairs, norm_order, singular, pow_norm, digits)
  }
}
```

```
# Find the eigenvalue dispersions for a given matrix
.dispersion_matrix <- function(P, pairs, norm_order, singular, pow_norm, digits = 4){
  # Get the ordered spectrum of the matrix
  eigenvalues <- spectrum(P, norm_order = norm_order, singular = singular)
  # Generate norm function to pass along as argument (Euclidean or Beta norm)
  norm_fn <- function(x){ (abs(x))^pow_norm }
  # Compute and return the dispersion
  map2_dfr(pairs[["i"]], pairs[["j"]], .resolve_dispersion, eigenvalues, norm_fn, digits)
}
```

```
# Read and parse a dispersion observation between eigenvalue i and j.
.resolve_dispersion <- function(i, j, eigenvalues, norm_fn, digits){
  # Initialize dispersion dataframe by adding order of eigenvalues compared
  disp <- data.frame(i = i, j = j)
  # Add the eigenvalues
  disp$eig_i <- .read_eigenvalue(i, eigenvalues)
  disp$eig_j <- .read_eigenvalue(j, eigenvalues)
  # Get the identity difference
  disp$id_diff <- disp$eig_j - disp$eig_i
  # Compute norm of the identity difference (standard norm metric)
  disp$id_diff_norm <- norm_fn(disp$id_diff)
  # Compute the difference of absolutes
  disp$abs_diff <- norm_fn(disp$eig_j) - norm_fn(disp$eig_i)
  # Round digits
  disp <- round(disp, digits)
  # Get the ranking difference
  disp$diff_ij <- disp$i - disp$j
  # Return the resolved dispersion observation
  disp
}
```

Helper Functions

```
# Parse a string argument for which pairing scheme to utilize
.parsePairs <- function(pairs, array, array_class){
  # Valid schemes for printing if user is unaware of options
  valid_schemes <- c("largest", "lower", "upper", "consecutive", "all")
  # Set default to be the consecutive pair scheme
  if(class(pairs) == "logical"){pairs <- "consecutive"}
  # Stop function call if the argument is invalid
  if(!(pairs %in% valid_schemes)){
    scheme_list <- paste(valid_schemes, collapse = ", ")
    stop(paste("Invalid pair scheme. Try one of the following: ", scheme_list, ".", ""))
  }
  # // Once we verify that we have a valid pair scheme string, try to parse it.
  # First, obtain a matrix by inferring array type; if ensemble take first matrix
  if(array_class == "ensemble") { P <- array[[1]] }
  else if(array_class == "matrix") { P <- array }
  # Obtain the dimension of the matrix
  N <- nrow(P)
  # Parse the pair string and evaluate the pair scheme
  if(pairs == "largest"){pair_scheme <- data.frame(i = 2, j = 1)}
  else if(pairs == "consecutive"){pair_scheme <- .consecutive_pairs(N)}
  else if(pairs == "lower"){pair_scheme <- .unique_pairs_lower(N)}
  else if(pairs == "upper"){pair_scheme <- .unique_pairs_upper(N)}
  else if(pairs == "all"){pair_scheme <- .all_pairs(N)}
  # Return pair scheme
  return(pair_scheme)
}
```

Appendix D

Mixing Time Simulations

D.1 Introduction

This section goes hand-in-hand with **Appendix A.3** on Markov Chains, so please read that section first. In this supplementary chapter, we will be overviewing the topic of simulating mixing time distributions for Markov chains. While this topic is not directly related to spectral statistics, mixing time simulations involve numerically approximating how many powers of a stochastic matrix are needed to obtain a stationary distribution. As such, we are simulating the time it takes an arbitrary initial probability distribution to "mix" or approach an eigenvector of the transition matrix.

That being said, we will start to define the mixing time of a Markov chain by using the framework of transition matrices. However, before we do so, we must formalize probability vectors, which are the vectors with which we will iterate our transition matrices.

Definition D.1.1 (Probability Vector)**.** *An N-dimensional probability vector, denoted $\vec{\pi} \in [0,1]^N$ is a vector of probabilities such that the entries of $\vec{\pi}$ sum to one. That is,*

$$\vec{\pi} = (\pi_j)_{j=1}^N \text{ where } \sum_{j=1}^N \pi_j = 1$$

*In other words, a probability vector is a stochastic row, so we use the same algorithm to randomly generate one as documented in **Algorithm B.1.1**. To notate a randomly sampled probability vector of size N, write $\vec{\pi} \sim \Phi(N)$.*

Now, given a probability vector $\vec{\pi}$, we are interested in observing the expected time spent at each vertex in the graph after n steps. See the definition of the n-step probability in **Appendix A.3**. From that section, we know that this is equivalent to taking powers of the matrix. In other words, we are interested in observing the advancing probability vector $\vec{\pi}^{(n)} = \vec{\pi}Q^n$.

In other words, to observe how the Markov chain advances an arbitrary probability vector $\vec{\pi}$, we would seek the following sequence of N-dimensional probability vectors:

$$\vec{\pi} \to \vec{\pi}^{(1)} \to \vec{\pi}^{(2)} \to \ldots \to \vec{\pi}^{(T)}$$

As such, it is worthwhile to define the power sequence of a probability vector w.r.t. a given transition matrix, which we do below.

Definition D.1.2 (Power Sequences)**.** *Suppose that $\vec{\pi} \in [0,1]^N$ is a probability vector and Q is an $N \times N$ transition matrix. The power sequence of $\vec{\pi}$ w.r.t. Q is defined as the sequence $\mathcal{P}(\vec{\pi} \mid Q)$ where:*

$$\mathcal{P}(\vec{\pi} \mid Q) = \langle \vec{\pi}^{(n)} \rangle \text{ where } \pi_n = \vec{\pi}Q^n \text{ for } n \in \mathbb{N}$$

In practice, we usually take finite sequences since infinite sequences are computationally intractable. To notate a sequence that goes from $n = 0, \ldots, T$, we denote this $\mathcal{P}^{(T)}(\vec{\pi} \mid Q)$.

Now, what we are interested in is observing when a point becomes a stationary distribution, or in other words, an eigenvector where $\lambda = 1$. In other words, we seek to find the power τ where $\vec{\pi}^{(\tau)}$ satisfies:

$$\vec{\pi}^{(\tau)}Q = \vec{\pi}^{(\tau)}$$

As we said, this is akin to solving the eigenvector equation for a particular eigenvalue. So, what we seek is a vector $\vec{\pi}'$:

$$\vec{\pi}'Q = \lambda\vec{\pi}'$$

Power Sequence and Ratio Approximations

Suppose we are generally trying to solve the eigenvector equation:

$$\vec{\pi}Q = \lambda\vec{\pi}$$

If we multiply both sides by Q^{n-1}, we get the equation:

$$\vec{\pi}Q^n = \lambda\vec{\pi}Q^{n-1}$$

Notice, we could use our power sequence notation to simplify this and get:

$$\vec{\pi}^{(n)} = \lambda\vec{\pi}^{(n-1)}$$

If we represent lambda as a vector of constants by saying $\vec{\lambda} = (\lambda, \lambda, \ldots, \lambda)$, we get:

$$\vec{\pi}^{(n)} = (\lambda, \lambda, \ldots, \lambda)\vec{\pi}^{(n-1)}$$

This means that we could express $\vec{\pi}^{(n)}$ as the vector $\vec{\pi}^{(n-1)}$ where every entry is scaled by λ. This means that for every entry j,

$$\vec{\pi}_j^{(n)} = \lambda\vec{\pi}_j^{(n-1)}$$

If we interpret division on vectors to mean component-wise division of the vector entries, we could rewrite:

$$\frac{\vec{\pi}^{(n)}}{\vec{\pi}^{(n-1)}} = \vec{\lambda}$$

There isn't a precise notion for this. As such, the quantity $\frac{\vec{\pi}^{(n)}}{\vec{\pi}^{(n-1)}}$ is a special quantity which we will denote $\rho^{(n)}$. It is a vector of ratios, and for this reason, we will call the following the consecutive ratio sequence.

Definition D.1.3 (Consecutive Ratio Sequences)**.** *Now, suppose we have a probability vector $\vec{\pi}$ and its corresponding (finite) power sequence $\mathcal{P}^{(T)}(\vec{\pi} \mid Q)$ w.r.t. some transition matrix Q. Accordingly, define the consecutive ratio sequence (CRS) of $\vec{\pi}$ as follows:*

$$\mathcal{R}(\vec{\pi} \mid Q) = \langle \rho^{(n)} \rangle_{n=2}^{T} \text{ where } \rho_j^{(n)} = \frac{\pi_j^{(n)}}{\pi_j^{(n-1)}} \text{ for } j = 1, \ldots, N$$

In other words, the consecutive ratio sequence of $\vec{\pi}$ can be obtained by performing ***component-wise division*** *on two consecutive elements of the power sequence of $\vec{\pi}$.*

Finally, we can formalize the notion of numerically obtaining eigenvectors using the power and consecutive ratio sequences by the following fact:

$$\vec{\pi}^{(n)} \text{ is an eigenvector } \iff \rho^{(n)} = (\lambda, \lambda, \ldots, \lambda)$$

As such, stationary distributions are reached when we obtain an eigenvector of eigenvalue one. That is,

$$\vec{\pi}^{(n)} \text{ is a stationary distribution } \iff \rho^{(n)} = (1, 1, \ldots, 1)$$

We can interpret this as saying that for a vector to be a stationary distribution, multiplying it by the transition matrix should not alter any of the entries.

D.2 Convergence

Because these ratios never actually converge to the eigenvectors for a finite power $n \in \mathbb{N}$, we must define a notion of "near convergence". So as a preliminary, we first must define ε-equivalence, where we consider two vectors to be equivalent when they are close enough to each other.

Definition D.2.1 (ε-Equivalence)**.** *Let $\mathbb{F}$ be a field, and fix $\varepsilon \in \mathbb{R}^+$. Suppose we have vectors $v, v' \in \mathbb{F}^N$. Then, v, v' are ε-equivalent if and only if each of their components have a difference of at most ε. In other words,*

$$v \sim_\varepsilon v' \iff \forall i \in \mathbb{N}_N : |v_i - v_i'| < \varepsilon$$

This makes defining "near convergence" quite simple.

Definition D.2.2 (ε-Convergence)**.** *Let $\varepsilon \in \mathbb{R}^+$, and suppose we have a* ***finite*** *evolution sequence $\mathcal{P}^{(T)}(\vec{\pi} \mid Q)$. Then, $\vec{\pi}$ ε-converges w.r.t Q within $\mathcal{P}^{(T)}(\vec{\pi} \mid Q)$ if and only if:*

$$\exists \tau \in \mathbb{N} \mid \tau \leq T \text{ such that } \forall k \geq \tau \mid \vec{\pi}^{(\tau)} \sim_\varepsilon \vec{\pi}^{(k)}$$

Finally, we are able to define mixing time. This definition is inspired in part by Mark Levin's book "Markov Chains and Mixing Times". Please see [Levin (2008)] as a resource.

Definition D.2.3 (Mixing Time)**.** *Suppose we have a Markov chain represented by the transition matrix Q, and suppose that $\vec{\pi}$ is an probability vector. Then, the mixing time of $\vec{\pi}$ with respect to Q is the positive integer τ such that $\vec{\pi}Q^\tau = \vec{\pi}^{(\tau)} \approx \vec{\pi}^{(\tau+1)}$. In other words, a vector $\vec{\pi}$ is mixed at τ when $\vec{\pi}^{(\tau)}$ becomes a stationary distribution or eigenvector of Q. To be concrete, we say that $\vec{\pi}$ is ε-mixed w.r.t. Q with a mixing time of τ given the following:*

$$\tau = \min\left\{N \in \mathbb{N} \mid \forall n \geq N : \vec{\pi}^{(n)} \sim_\varepsilon \vec{\pi}^{(N)}\right\}$$

D.3 Simulation

Finally, we are ready to simulate the mixing time distribution of a random matrix or random matrix ensemble. We begin by defining a random batch of points.

Definition D.3.1 (Monte-Carlo Stochastic Batch)**.** *A Monte-Carlo stochastic batch of dimension N and size K is a set of K random probability vectors of size B. That is,*

$$\mathcal{B} = \{\vec{\pi}_i \sim \Phi(N)\}_{i=1}^{B}$$

After which, we define the mixing time distribution to be the set of their mixing times for each matrix or matrices.

Definition D.3.2 (Mixing Time Distribution)**.** *Suppose we have an ensemble of stochastic matrices $\mathcal{E}$ and that we have a batch of probability vectors $\mathcal{B}$. Then, the mixing time distribution of $\mathcal{E}$ is defined as the following:*

$$\{\tau(P, \vec{\pi}) \mid P \in \mathcal{E}, \vec{\pi} \in \mathcal{B}\}$$

D.4 Generalizations

Generally, we could do this for any matrix, not just stochastic matrices! This means that we have a notion of mixing even for non-stochastic matrices.

Definition D.4.1 (Random Batch)**.** *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 λ. That is,*

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

The only difference between these batches and the previous ones is that for stochastic matrices, our points had to be probability vectors. Here, we are free to choose any random point over $\mathbb{F}^N$. That being said, note that if $\mathbb{F} = \mathbb{C}$, then take $\vec{x} \in \mathcal{B}_\lambda$ to mean $\vec{x} = a + bi$ where $a, b \sim \text{Unif}(-\lambda, \lambda)$.

A Generalization of Mixing

We could generalize the notion of mixing to any vector $\vec{v}$ and any matrix M. To do so, we rewrite the approximate eigenvector statement as follows:

$$\vec{v}^{(n)} \text{ is an eigenvector of } Q \iff \rho^{(n)} = (\lambda_1, \lambda_1, \ldots, \lambda_1)$$

Note that above, we specify that $\lambda = \lambda_1$. One fact that we omitted but implied was that we were using the largest eigenvalue of our stochastic matrices, $\lambda_1 = 1$. Indeed, from computation evidence, we find from the consecutive ratio sequence that regardless of the initial vector, the power sequence always tends towards an eigenvector of λ_1. **So, we find extensive computational evidence that for any matrix P and vector v, iterating P and multiplying it by v will eventually return an eigenvector of the largest eigenvalue.**

Open Questions

Generally speaking, the ratio sequence has a distribution of terms that is Cauchy-like (a ratio distribution). We occasionally get extreme values due to the instability of ratio distributions. That being said, here are some open questions:

1. How are the entries of the consecutive ratio sequence distributed?
2. Are there any cases when the power sequences do not converge to an eigenvector of λ_1?
3. How do we obtain a full eigenvector basis using these techniques?

References

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

Dummit, D., & Foote, R. (2003). *Abstract Algebra, 3rd Edition*. Wiley, 3 ed.

Horn, R. (2012). *Matrix Analysis*. Cambridge University Press, 2 ed.

Hwang, J., & Blitzstein, J. K. (2019). *Introduction to Probability*. CRC Press, Taylor and Francis Group.

Kublanovskaya, V. (1962). On some algorithms for the solution of the complete eigenvalue problem. *USSR Computational Mathematics and Mathematical Physics*, *1*(3), 637–657.

Levin, D. (2008). *Markov Chains and Mixing Times*. American Mathematical Society, 1 ed.

Mehta, M. L. (2004). *Random Matrices, 3rd Edition (Vol. 142)*. Academic Press.

Meyer, C. (2004). *Matrix analysis and applied linear algebra*.

Miller, S. J., & Takloo-Bighash, R. (2004). *An Invitation to Modern Number Theory*. Academic Press.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. `https://www.R-project.org/`

Steen, L. (1973). Highlights in the History of Spectral Theory. *The American Mathematical Monthly*, *80*(4), 359–381.

Tao, T. (2012). *Introduction to Random Matrix Theory*. American Mathematical Society.

Taqi, A. (2021). *RMAT: Random Matrix Analysis Toolkit*. R package version 0.2.0. `https://CRAN.R-project.org/package=RMAT`

Wigner, E. (1955). *Characteristic Vectors of Bordered Matrices with Infinite Dimensions*.