Spectral Statistics of Random Matrices

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

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

Ali Taqi

May 2021

Approved for the Division
(Mathematics)

_____

Jonathan M. Wells

Table of Contents

Abstract

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

Dedication

For my mother.

Introduction

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

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

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

Chapter 1

Random Matrices

1.1 Introduction

As promised, here is what a random matrix is.

Definition 1.1.1 (Random Matrix)**.** *A (homogenous) random matrix is any matrix $M \in \mathbb{F}^{N\times N}$ is a matrix whose entries are i.i.d random variables. So, if a random matrix $M = (m_{ij})$ is $\mathcal{D}$-distributed, this is equivalent to saying $m_{ij} \sim \mathcal{D}$. In the scope of this thesis, we will only work with homogenous random matrices. From thereonafter, assume every random matrix to be homogenously distributed.*

Example 1.1.1. *Let $\mathcal{D} = \mathcal{N}(0, 1)$.*

1.2 Random Matrix Ensembles

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

1.2.1 Hermite Beta-Ensembles

Definition 1.2.2 (Hermite-Gaussian β-ensemble)**.** *The Hermite β-ensemble is the ensemble of random matrices whose eigenvalues have the joint probability density function:*

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

where the normalization constant c_H^β is given by:

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

They represent a matrix whose entries have β real number components.

Reference the person's work here and refer to the algorithm appendix on how to simualate general beta matrices.

1.3 Analytical Results

Chapter 2

Spectral Statistics

2.1 Introduction

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

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

2.1.1 The Quintessential Spectral Statistic: the Eigenvalue

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

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

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

2.1.2 Interlude: Ensembles

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

[Definition was here]

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

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

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

2.2 Eigenvalue Spectra

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. Consider the two following *ordering schema*:

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

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

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

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

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

2.2.2 Order Statistics

With eigenvalue ordering unambiguous and well-defined, we may proceed to start talking about their order statistics. In short, given a random sample of fixed size, order statistics are random variables defined as the value of an element conditioning on its rank within the sample. (See A.2.2)

For example, the maximum of a sample is an order statistic concerned with the highest ranked element. In our case, this could corrospond the largest eigenvalue of a spectrum. After all, a spectrum is a random sample of fixed size, so this statistic is well-defined. To obtain the distribution of the largest eigenvalue for some distribution $\mathcal{D}$, we would simulate an ensemble $\mathcal{E}$ and observe λ_1 for each of its matrices.

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.

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

Definition 2.2.1 (Ranking Difference)**.** *The ranking difference is a function* $\delta : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ *which takes the index of two eigenvalues (from an ordered spectrum) and returns their difference. In other words,* $\delta : (\lambda_i, \lambda_j) \mapsto (i - j)$.

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

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

Note that the sizes of each equivalence class are **never equal**. With this partition in mind, we can consider various statistics conditioning on the value of ρ.

Conditioning on ρ will be especially useful in the cases where we are considering matrices like the Hermite-β matrices; the eigenvalues of those matrices tend to *repel*, so to speak, and we can observe these patterns using ρ.

2.2.3 Symmetric and Hermitian Matrices

A very important class of matrices in Linear Algebra is that of Symmetric or Hermitian matrices (See A.1.2). Simply put, those are matrices which are equal to their conjugate transpose.

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

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

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

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

2.3 Eigenvalue Dispersions

In this section, we define the final spectral statistic studied in this chapter: eigenvalue dispersion. As the name suggests, these statistics concern the distirbution of the spacings between the eigenvalues. Funnily enough, 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 motivate a few definitions and formalisms in this section, then motivate the definition of a matrix's eigenvalue dispersion. To start off, we define an object useful for pairing our eigenvalues together, the pairing scheme.

2.3.1 Pairing Schema

Next up, we introduce a new notation for a pairing scheme denoted Π. What are pairing schemes and why do they matter? Recall that our goal is to study the spacings between eigenvalues. If we are studying spacing, then a priori, we are concerned with pairs of eigenvalues! Spacing, after all, is a binary relationship.

So, with the definitions of spectra well motiviated, a natural definition of pairing schema follows. Essentially, Π is just a subset of the Cartesian product of a spectrum with itself. In other words, if we denote $\mathbb{S} := \sigma(P)$, then we say that a pairing scheme is simply a subset $\Pi \subseteq \mathbb{S}^2$. Here are some pairing schema that we will consider:

1. Let $\Pi_>$ be the set of unique combinations of ordered eigenvalues. This will be the standard ordered pair scheme used in lieu of our dispersion metric argument orders (more later).

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

2. Let Π_1 be the largest pair of eigenvalues of a spectrum. Nice and simple.

$$\Pi_1 = \{(\lambda_2, \lambda_1)\}$$

3. Let Π_C be the consecutive pairs of eigenvalues in a spectrum. This pairing scheme gives us the minimal information needed to express important bounds and spacings in terms of its elements.

$$\Pi_C = \{\pi_{ij} = (\lambda_i, \lambda_j) \mid i = j + 1\}_{i=1}^{N-1}$$

2.3.2 Dispersion Metrics

When we define the dispersion of a matrix, we will see that there is a free argument d called the dispersion metric. This function d is a general function whose domain is always two eigenvalues. In set notation, this is the set $\mathbb{C} \times \mathbb{C}$ - a pair of two complex numbers. Its range will often be the positive reals $\mathbb{R}^+$; this is because the dispersion metric often will be substitutable with distance metric. Sometimes, the range will be $\mathbb{C}$. So, the dispersion metric will take the following form:

$$d : \mathbb{C} \times \mathbb{C} \to \{\mathbb{R}^+, \mathbb{C}\}$$

Consider the following dispersion metrics below. Out of those 4 dispersion metrics, only the first one has a range of $\mathbb{C}$. The rest have a range of $\mathbb{R}^+$. Additionally, the second and third metrics are *symmetric* operations while the rest are not. The β-norm is only a symmetric operation when β is even.

1. The identity difference: $d_{id}(z, z') = z' - z$

2. The standard norm: $d_n(z, z') = |z' - z|$

3. The β-norm: $d_\beta(z, z') = |z' - z|^\beta$

4. The difference of absolutes: $d_{ad}(z, z') = |z'| - |z|$

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

2.3.3 Matrix and Ensemble Dispersions

Definition 2.3.1 (Dispersion)**.** *The dispersion of a matrix $P \in \mathbb{F}^{N\times N}$ with respect to a dispersion metric $d : \mathbb{C}\times\mathbb{C} \to \mathbb{F}$ and pairing scheme Π, call it $\Delta_d(P, \Pi)$, is defined as follows. Suppose $\sigma(P) := \mathbb{S}$ is the ordered spectrum of P where $\sigma(P) = \{\lambda_1, \ldots, \lambda_N\}$. Then, let $\Pi = \{\pi_{ij} = (\lambda_i, \lambda_j)\} \subseteq \mathbb{S}^2$ be a subset of eigenvalue ordered pairs. Then, the dispersion of P with respect to d is simply the set $\Delta_d(P, \Pi) = \{\delta_{ij} = d(\pi_{ij}) \mid \pi_{ij} = (\lambda_i, \lambda_j) \in \Pi\}$.*

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

Definition 2.3.2 (Ensemble Dispersion)**.** *If we have an ensemble $\mathcal{E}$, then we can naturally extend the definition of $\Delta_d(\mathcal{E}, \Pi)$. To take the dispersion of an ensemble, simply take the union of the dispersions of each of its matrices. In other words, if $\mathcal{E} = \{P_i \sim \mathcal{D} \mid P_i \in \mathbb{F}^{N\times N}\}_{i=1}^{K}$, then $\Delta_d(\mathcal{E}, \Pi) = \bigcup_{i=1}^{K} \Delta_d(P_i, \Pi)$.*

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

Appendix A

Math Appendix

A.1 Linear Algebra

Symmetric Matrix.

A.1.1 Real Symmetric Matrices have Real Eigenvectors

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

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

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

So, we will show that:

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

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

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

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

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

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

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

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

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

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

Now, we will prove the main theorem.

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

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

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

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

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

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

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

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

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

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

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

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

A.2 Probability Theory

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

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

A.3 Markov Chains

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

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

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

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

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

- Q is a non-negative matrix. That is, note that $Q \in \mathcal{M}_{\mathbb{R}^+}[M \times M]$ so every $q_{ij} \in \mathbb{R}^+$. This follows because probabilities are necessarily non-negative values.
- The entries of every row i of Q must sum up to 1. This may be understood as applying the law of total probability to the event of transitioning from any given state $\forall i \in S_M$. In other words, the chain has to go somewhere with probability 1.

  $$\forall i \in S_M : \sum_{j \in S_M} q_{ij} = 1$$

- Note, it is NOT necessary that the converse holds. The columns of our transition matrix need not sum to 1 for it to be a valid transition matrix.

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

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

Realize:

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

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

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

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

A.3.1 Classification of states

- A state $i \in S_M$ is said to be **recurrent** if starting from i, the probability is 1 that the chain will *eventually* return to i. If the chain is not recurrent, it is **transient**, meaning that if it starts at i, there is a non-zero probability that it never returns to i.

- Caveat: As we let $n \to \infty$, our Markov chain will gurantee that all transient states will be left forever, no matter how small the probability is. This can be proven by letting the probability be some ε, then realizing that by the support of $\mathrm{Geom}(\varepsilon)$ is always some finite value, then the equivalence between the Markov property and independent Geometric trials gurantees the existence of some finite value such that there is a success of never returning to i.

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

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

- From our quantifier formulation of irreducible Markov chains, note that we can equivalently say that a chain is irreducible if there is integer $n \in \mathbb{N}$ such that the (i, j) entry of Q^n is positive for any i, j.

- A Markov chain is **reducible** if it is not **irreducible**. Using our quantifier formulation, it means that it suffices to find transient states so that:

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

Appendix B

Algorithm Appendix

B.1 Matrix Simulation

B.1.1 Stochastic Matrices

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

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

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

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

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

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

B.1.2 Normal Matrices

Algorithm B.1.4 (Hermite β-Matrix)**.**

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

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

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

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

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

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

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

Appendix C

Code Appendix

Appendix D

Mixing Time Simulations

D.1 Introduction

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

D.2 Mixing Time Simulations

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

D.3 Erdos-Renyi Ensemble Simulations

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

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

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

D.4 Questions

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

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

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

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

D.4.1 Cauchy Distributed Ratios

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

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