Math 2000 Project 5: Markov Chains, the Perron-Frobenius theorem and Google's PageRank Algorithm*

Subhadip Chowdhury

- **Purpose:** To analyze Markov chains and investigate steady state vectors.
- **Prerequisite:** Eigenvalues and eigenvectors.
- **Resources:** Use Mathematica as needed. You might also want to take a look at http://setosa.io/ev/markov-chains/ and http://setosa.io/markov/index.html

Web Surfing

Definition 1. A **Stochastic matrix** (aka Markov Matrix) is a square matrix, all of whose entries are between 0 and 1 (inclusive), and such that the entries in each column add up to 1.

We can think of the matrix entries as probabilities of different events happening. For example, consider the matrix

$$A = \begin{bmatrix} 0.7 & 0.1 & 0.2 \\ 0.2 & 0.4 & 0.2 \\ 0.1 & 0.5 & 0.6 \end{bmatrix}.$$

(Note that the entries in each column add up to 1.) We can use this matrix to make a *very very* simple model of the internet, as follows.

Pretend there are only 3 domains on the internet: reddit.com, google.com, and instagram.com (Hereafter referred to as Domains 1, 2, and 3, respectively.) In every five-minute period, say, the people surfing this web have some probability of switching to another domain. For example, let's say that, in a given five-minute period, out of all the people clicking around on Domain 1, 70% will remain on Domain 1, 20% will end up clicking on a link to Domain 2, and 10% will end up on Domain 3. Notice that these are precisely the numbers in the first column of A. I.e., the first column encodes what happens to the people surfing Domain 1. Similarly, the second and third columns tells the probabilities of what happens to the people on Domain 2 and Domain 3, respectively.

Another way to say the same thing: if we identify matrix entries in the standard way, where a_{ij} represents the entry in row i and column j, then a_{ij} here is the probability that someone surfing Domain j will end up on Domain i five minutes from now.

*Most of this project is made using or copied from Lay's Linear Algebra book and Interactive Linear Algebra by Dan Margalit, Joseph Rabinoff.

Definition 2. The matrix A is called the *Transition Matrix* of this system and the columns of A are called the *Transition Probability Vectors*.

By our definition, the transition matrix is a stochastic matrix. We will see more examples of stochastic matrices later.

Exercise 1

Draw a graph illustrating this situation: make a vertex(node) for each of Domain 1, 2, and 3, and draw arrows between the nodes, each labelled with the probability of moving from one bubble to the next. You might want to check out the links above for some pretty neat animations.

Suppose initially, at time $t = 0$, **50%** of the surfers are on Domain **1**, **30%** are on Domain **2**, and **20%** are on Domain **3**. Encode this by the vector

$$\vec{\mathbf{x}}_0 = \begin{bmatrix} 0.5 \\ 0.3 \\ 0.2 \end{bmatrix}.$$

Then $\vec{\mathbf{x}}_1 = A\vec{\mathbf{x}}_0$ tells us what proportion of the surfers are on each domain after one time increment. And $\vec{\mathbf{x}}_2 = A\vec{\mathbf{x}}_1$ tells us where the surfers are after two time increments. And so on.

Exercise 2

Compute $\vec{\mathbf{x}}_1$ and $\vec{\mathbf{x}}_2$.

Note that for each of $\vec{\mathbf{x}}_0$, $\vec{\mathbf{x}}_1$, and $\vec{\mathbf{x}}_2$, the entries add up to 1. This makes sense for $\vec{\mathbf{x}}_0$, since its entries are probabilities covering all the cases. But it's not so obvious for $\vec{\mathbf{x}}_1$ and $\vec{\mathbf{x}}_2$.

Exercise 3

Show in general that, given a Markov matrix M and a vector $\vec{\mathbf{v}}$ whose entries add up to 1, the entries of $M\vec{\mathbf{v}}$ also add up to 1.

Definition 3. The sequence of vectors $\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2, \vec{\mathbf{x}}_3, \ldots, \vec{\mathbf{x}}_n, \ldots$ is called a **Markov chain**.

It lets us track the evolution of this system, seeing where the populations end up over time. However, in order to find something like $\vec{\mathbf{x}}_{100}$, we would need to compute A^{100}. In other words, we need eigenstuff and diagonalization!

The first claim is that A (or any Markov matrix) always has an eigenvalue of **1**. It's not trivial to find a $\vec{\mathbf{v}}$ such that $A\vec{\mathbf{v}} = \vec{\mathbf{v}}$ in general, but...

Exercise 4

Find an easy nonzero $\vec{\mathbf{v}}$ such that $A^T\vec{\mathbf{v}} = \vec{\mathbf{v}}$. This shows that **1** is an eigenvalue of A^T.
Hint: use the fact that, since we've transposed, the entries in each *row* of A^T add up to 1.

Recall that A and A^T have the same eigenvalues. Thus 1 is an eigenvalue of A as well.

Exercise 5

(a) Find the characteristic polynomial of A, and use the fact that we already know $(\lambda - 1)$ will appear in the factorization to find the other eigenvalue(s) of A.

(b) Find the eigenspace for each eigenvalue of A, and write down a nice eigenbasis for $\mathbb{R}^3$.

(c) Write down the diagonalization $A = BDB^{-1}$.

(d) Compute A^{100} and $\vec{x}_{100}$.

(e) Find $\lim_{n\to\infty} \vec{x}_n$.

(f) In the long-term, what percentage of surfers end up on reddit.com, what percentage end up on google.com, and what percentage end up on instagram.com?

(g) Did it matter here what our particular initial distribution $\vec{x}_0$ was? If 100% started at reddit.com, would we still end up with the same percentages on each domain over the long term?

Steady State and the Perron-Frobenius Theorem

The eigenvalues of stochastic matrices have very special properties.

Proposition 1. *Let A be a stochastic matrix. Then:*

(a) 1 is an eigenvalue of A.

(b) If λ is a (real or complex) eigenvalue of A, then $|\lambda| \leq 1$.

As we observed in the last section in exercise 4, we can prove that 1 is always an eigenvalue of a stochastic matrix. Let's prove the second part. We will restrict to the case of real eigenvalues for the sake of this project.

Exercise 6

1. Let λ be any real eigenvalue of A. Explain why we can always find a vector $\vec{\mathbf{x}}$ such that $A^T\vec{\mathbf{x}} = \lambda\vec{\mathbf{x}}$.

2. Let $\vec{\mathbf{x}} = [x_1 \quad x_2 \quad \dots \quad x_n]^T$. Choose x_j with the largest absolute value, so that $|x_i| \le |x_j|$ for all i. Explain the steps in the following chain of inequality

$$|\lambda| \cdot |x_j| = \left|\sum_{i=1}^{n} a_{ij}x_i\right| \le \sum_{i=1}^{n} (a_{ij} \cdot |x_i|) \le \left(\sum_{i=1}^{n} a_{ij}\right) \cdot |x_j| = 1 \cdot |x_j|$$

Hence we can conclude that $|\lambda| \le 1$.

Definition 4. We say that a matrix A is *positive* if all of its entries are positive numbers.

For a positive stochastic matrix A, one can show that if $\lambda \neq 1$ is a (real or complex) eigenvalue of A, then $|\lambda| < 1$. The $1-$eigenspace E_1 of a stochastic matrix is very important.

Definition 5. If A is a stochastic matrix, then a *steady-state vector* (or equilibrium vector) for A is a probability vector $\vec{\mathbf{q}}$ such that

$$A\vec{\mathbf{q}} = \vec{\mathbf{q}}$$

In other words, it is an eigenvector $\vec{\mathbf{q}}$ of A with eigenvalue 1, such that the entries are positive and sum to 1.

The Perron-Frobenius theorem describes the long-term behavior of such a process represented by a stochastic matrix. Its proof is complicated and is beyond the scope of this project.

Theorem 2 (Perron-Frobenius Theorem)**.** *Let A be a positive stochastic matrix. Then A admits a unique steady state vector $\vec{\mathbf{q}}$, which spans the $1-$eigenspace E_1. Further, if $\vec{\mathbf{x}}_0$ is any initial state and $\vec{\mathbf{x}}_{k+1} = A\vec{\mathbf{x}}_k$ then the Markov chain $\{\vec{\mathbf{x}}_k\}$ converges[1] to $\vec{\mathbf{q}}$ as $k \to \infty$.*

Why is this nontrivial? For two reasons:

- Apriori, we did not know whether all the entries of the eigenvector corresponding to the eigenvalue 1 are positive. We also did not know about the geometric multiplicity of the eigenvalue 1. P-F theorem tells us that in fact, $\dim(E_1) = 1$ and we can find a vector $\vec{\mathbf{q}} \in E_1$ such that all entries of $\vec{\mathbf{q}}$ are positive and sum to 1!

- If a Markov process has a positive transition matrix, the process will converge to *the* steady state $\vec{\mathbf{q}}$ *regardless of the initial state.*

[1] We say that a sequence of vectors $\{\vec{\mathbf{x}}_k\}$ converges to a vector $\vec{\mathbf{q}}$ as $k \to \infty$ if the entries in $\vec{\mathbf{x}}_k$ can be made as close as desired to the corresponding entries in $\vec{\mathbf{q}}$ by taking k sufficiently large.

Exercise 7

Let $A = \begin{pmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{pmatrix}$ be a stochastic matrix.

(a) Find the eigenvalues and corresponding eigenvectors of A.

(b) Using the eigenvector corresponding to the eigenvalue 1, find the steady-state vector $\vec{\mathbf{q}}$ of A.

Let's try to give a visual interpretation of the linear transformation defined by the matrix above. This matrix A is diagonalizable; we have $A = CDC^{-1}$ for

$$C = \begin{pmatrix} 7 & -1 & 1 \\ 6 & 0 & -3 \\ 5 & 1 & 2 \end{pmatrix} \quad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -.2 & 0 \\ 0 & 0 & .1 \end{pmatrix}$$

The matrix D leaves the $x-$coordinate unchanged, scales the $y-$ coordinate by $-1/5$, and scales the $z-$coordinate by $1/10$. Repeated multiplication by D makes the $y-$ and $z-$coordinates very small, so it *sucks all vectors into the* $x-$*axis.*

The matrix A does the same thing as D, but with respect to the coordinate system defined by the columns $\vec{\mathbf{u}}_1, \vec{\mathbf{u}}_2, \vec{\mathbf{u}}_3$ of C. This means that A *sucks all vectors into the* $1-$*eigenspace*, without changing the sum of the entries of the vectors.

Google's PageRank Algorithm

In 1996, Larry Page and Sergey Brin invented a way to rank pages by importance. They founded Google based on their algorithm. Here is how it works (Roughly). Each web page has an associated importance, or *rank*. This is a positive number. If a page P links to n other pages $Q_1, Q_2, \ldots, Q_n$, then each page Q_i inherits $\frac{1}{n}$ of P's importance.

Definition 6. Consider an Internet with n pages. The Rank matrix is the $n \times n$ matrix A whose i, j-entry is the importance that page j passes to page i.

Observe that the rank matrix is a stochastic matrix, assuming every page contains a link: if page i has m links, $m \leq n$, then the ith column contains the number $\frac{1}{m}$, a total of m times, and the number zero in the other entries.

The goal is to find the steady-state rank vector of this Rank matrix. We would like to use the Perron-Frobenius theorem to find the rank vector. Unfortunately, the Rank Matrix is not always a positive stochastic matrix.

Here is Page and Brin's solution. First we fix the rank matrix by replacing each zero column with a column of $\frac{1}{n}$s, where n is the number of pages.

So for example,

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} \quad \text{becomes} \quad A' = \begin{pmatrix} 0 & 0 & 1/3 \\ 0 & 0 & 1/3 \\ 1 & 1 & 1/3 \end{pmatrix}$$

The modified Rank Matrix A' is always stochastic.

Now we choose a number p in $(0, 1)$, called the damping factor. (A typical value is $p = 0.15$.)

Definition 7 (The Google Matrix)**.** Let A be the Rank Matrix for an Internet with n pages, and let A' be the modified Rank Matrix. The *Google Matrix* is the matrix

$$G = (1-p)\cdot A' + p\cdot B \quad \text{where} \quad B = \frac{1}{n}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

Exercise 8

Show that G is a positive stochastic matrix.

If we declare that the ranks of all of the pages must sum to one, then we find:

Definition 8 (The 25 Billion Dollar Eigenvector)**.** The PageRank vector is the steady state of the Google Matrix.

This exists and has positive entries by the Perron-Frobenius theorem. The hard part is calculating it: in real life, the Google Matrix has zillions of rows.