Stochastic Matrices

This is a special kind of difference equation in which the state change matrix encodes probabilities.

**Red Box Example:**

Pretend there are 3 RedBox kiosks in Durham, and that everyone who rents *Prognosis Negative* today will return it tomorrow. Suppose that someone from kiosk i will return to kiosk j with the following probabilities:

Renting

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Return</td>
<td>30%</td>
<td>40%</td>
<td>50%</td>
</tr>
<tr>
<td>Rent</td>
<td>30%</td>
<td>40%</td>
<td>30%</td>
</tr>
<tr>
<td>Refund</td>
<td>40%</td>
<td>20%</td>
<td>20%</td>
</tr>
</tbody>
</table>

If $v_k = (x_k, y_k, z_k) = \#\text{moves in kiosk } k$ on day $k$ then

$\begin{align*}
x_{k+1} &= 0.3x_k + 0.4y_k + 0.5z_k \\
y_{k+1} &= 0.3x_k + 0.4y_k + 0.3z_k \\
z_{k+1} &= 0.4x_k + 0.2y_k + 0.3z_k
\end{align*}$

Thus $v_{k+1} = \begin{pmatrix} 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.3 \end{pmatrix} v_k = A v_k$

Note the columns of $A$ sum to 1 because we're assuming every movie has a 100% chance of being returned somewhere.

→ this means the total # movies won't change.
**Def:** A square matrix is **stochastic** if its entries are nonnegative & the entries in each column sum to 1. A stochastic matrix is **positive** if all entries are positive (i.e., nonzero).

**Eg:**

- Positive stochastic matrix:
  \[
  A = \begin{pmatrix}
  0.3 & 0.4 & 0.5 \\
  0.3 & 0.4 & 0.5 \\
  0.4 & 0.2 & 0.2
  \end{pmatrix}
  \]

- Stochastic matrix:
  \[
  B = \begin{pmatrix}
  0.6 & 0.4 & 0.5 \\
  0.4 & 0.2 & 0.2 \\
  0.4 & 0.2 & 0.2
  \end{pmatrix}
  \]

- Not stochastic matrix:
  \[
  C = \begin{pmatrix}
  0.6 & 0.4 & 0.5 \\
  -0.1 & 0.4 & 0.3 \\
  -0.5 & 0.2 & 0.2
  \end{pmatrix}
  \]

**NB:** Columns sum to 1 means \(A^T(1) = (1)\):

\[
A = \begin{pmatrix}
0.3 & 0.4 & 0.5 \\
0.3 & 0.4 & 0.5 \\
0.4 & 0.2 & 0.2
\end{pmatrix} \quad A^T = \begin{pmatrix}
0.3 & 0.3 & 0.4 \\
0.4 & 0.4 & 0.2 \\
0.5 & 0.3 & 0.2
\end{pmatrix}
\]

\[
A^T(1) = \begin{pmatrix}
0.3 + 0.3 + 0.4 \\
0.4 + 0.4 + 0.2 \\
0.5 + 0.3 + 0.2
\end{pmatrix} = (1)
\]

**Fact:** If \(A\) is stochastic then 1 is an eigenvalue.

\[
\det(A - \lambda I) = \det((A - \lambda I)^T) = \det(A^T - \lambda I)
\]

(HW)

so \(A\) & \(A^T\) have the same eigenvalues, and \((1)\) is a 1-eigenvector of \(A^T\).
Fact: If $\lambda$ is an eigenvalue of a stochastic matrix then $|\lambda| \leq 1$.

Why? $\lambda$ is also an eigenvalue of $A^T$.

Let $v$ be an eigenvector: $A^T v = \lambda v$

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

$$v = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda x_3 \end{pmatrix} = A^T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_{11} x_1 + a_{12} x_2 + a_{13} x_3 \\ a_{21} x_1 + a_{22} x_2 + a_{23} x_3 \\ a_{31} x_1 + a_{32} x_2 + a_{33} x_3 \end{pmatrix}$$

Suppose $|x_1| \geq |x_2|$ and $|x_1| \geq |x_3|$

(choose the coordinate with largest abs. value)

1st coordinate:

$$\lambda x_1 = a_{11} x_1 + a_{12} x_2 + a_{13} x_3$$

$$\Rightarrow |x_1| |x_1| = |a_{11} x_1 + a_{12} x_2 + a_{13} x_3|$$

$$\leq |a_{11} x_1| + |a_{12} x_2| + |a_{13} x_3|$$

$$\leq (a_{11} + a_{12} + a_{13}) |x_1| = |x_1|$$

$$\Rightarrow |x_1| \leq 1 \quad \checkmark$$

Better Fact: If $\lambda \neq 1$ is an eigenvalue of a positive stochastic matrix then $|\lambda| < 1$.

(since 1 is the dominant eigenvalue)
The Red Box matrix has characteristic polynomial
\[ p(\lambda) = -\lambda^3 + 0.9\lambda + 0.12\lambda - 0.02 \]
\[ = -(\lambda-1)(\lambda+0.2)(\lambda-0.1) \]

Eigenvalues are \(1, -0.2, 0.1\)
and \( |-0.2| < 1, \ 10.1 < 1\)

In this case, there are 3 (different) eigenvalues, so the matrix is diagonalizable. In fact, the eigenvectors are

1: \( w_1 = \begin{pmatrix} \frac{7}{5} \\ -\frac{1}{5} \end{pmatrix} \)
2: \( w_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \)
3: \( w_3 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \)

Suppose you start with \( v_0 = \begin{pmatrix} 48 \\ 36 \\ 42 \end{pmatrix} \) movies.

Expand in the eigenbasis:

\[ v_0 = x_1 w_1 + x_2 w_2 + x_3 w_3 \Rightarrow x_1 = 7, \ x_2 = 3, \ x_3 = 2 \]
\[ v_0 = 7w_1 + 3w_2 + 2w_3 \]

Solve the difference equation:

\[ v_k = A^k v_0 = (1)^k 7w_1 + (-0.2)^k 3w_2 + (0.1)^k 2w_3 \]
\[ \Rightarrow 7w_1 = \begin{pmatrix} 49 \\ 42 \\ 35 \end{pmatrix} \]
Observation 1:

If \( V_0 = x_1 w_1 + x_2 w_2 + x_3 w_3 \),

then \( V_k = x_1 w_1 + (-0.2)^k x_2 w_2 + (0.1)^k x_3 w_3 \)

\[ \xrightarrow{k \to \infty} x_1 w_1 \] (if \( x_i \neq 0 \))

So \( V_k \) converges to a 1-eigenvector \[ \text{[demo]} \]

Observation 2:

Since the total #movies doesn't change, we even knew which eigenvector: it's the multiple of \( w_1 \) whose entries have the same sum as \( V_0 \).

In our case, we started with

\[ V_0 = \left( \frac{48}{36} \right) \to \text{126 total movies} \]

The sum of the entries of \( w_1 = \left( \frac{7}{5} \right) \) is 18, so

the sum of the entries of \( \frac{126}{18} \cdot w_1 = 7 \cdot w_1 \) is 126,

So \( V_k \xrightarrow{k \to \infty} 7 \cdot w_1 = \left( \frac{49}{35} \right) \) \[ \checkmark \]

This would've been easier if we'd replaced \( w_1 \) by \( \frac{1}{18} w_1 \) to assume the entries of \( w_1 \) sum to 1.
Observation 3:
The coordinates of $w_i = \begin{pmatrix} \frac{7}{5} \\ \frac{3}{5} \end{pmatrix}$ are positive numbers.

It's good they're not negative — that would mean negative movies in some kiosk!

These observations turn out to hold for any positive stochastic matrix, even if it's not diagonalizable.

**Perron–Frobenius Theorem:** If $A$ is a positive stochastic matrix, then there is a unique 1-eigenvector $w$ with positive coordinates summing to 1.

If $v_0$ is a vector with coordinates summing to $c$, then $v_k = A^k v_0 \xrightarrow{k \to \infty} c \cdot w$.

**Def:** The 1-eigenvector of a positive stochastic matrix whose coordinates sum to 1 is the steady state of that matrix.

This is easy to compute!

\[ \rightarrow \text{Find a 1-eigenvector } v \in \text{Null}(A - I_n) \]

\[ \rightarrow w = \frac{v}{\text{sum of coords of } v} \]
So the steady state of the Red Box matrix is
\[ w = \frac{1}{18} \left( \frac{7}{5} \right). \]

Positive Stochastic Matrices: Summary

If \( A \) is positive stochastic, then:
- The 1-eigenspace of \( A \) is a line.
- There is a 1-eigenvector with positive coordinates.
  Divide by the sum of the coordinates.
- There is a unique 1-eigenvector \( w \) with
  positive coordinates summing to 1.
- \( |\lambda| < 1 \) for all other eigenvalues, so
  1 is the dominant eigenvalue.
- If \( v_0 \) is any vector then
  \[ V_k = A^k v_0 \xrightarrow{k \to \infty} c \cdot w \]
- The scalar multiple \( c \) is the
  sum of the coordinates of \( v_0 \)
  (the total #movies doesn't change.)
Google's PageRank

or, how Larry Page & Sergei Brin used linear algebra to make the internet searchable.

Idea: each web page has an "importance", or rank. This is a positive number. If page P links to n other pages Q_1, ..., Q_n, then each Q_i inherits $\frac{1}{n}$ of P's importance.

→ so if an important page links to your page, then your page is important too.

→ or, if a million unimportant pages link to your pages, then your page is important.

→ but if only one crappy page links to you, then your page is not important.

Random surfer interpretation:

The random surfer sits at his computer all day clicking links at random. The pages he visits most often are the most important in the above sense, as it turns out.
Eg: Here's an internet with 4 pages. Links are indicated by arrows.

- Page A has 3 links
  - passes $\frac{1}{2}$ of its importance to B C D
- Page B has 2 links
  - passes $\frac{1}{2}$ of its importance to C D
- Page C has 1 link
  - passes all of its importance to A
- Page D has 2 links
  - passes $\frac{1}{2}$ of its importance to A C

So, if the pages have importance $a$ b c d then

$$a = c + \frac{1}{2}d$$
$$b = \frac{1}{2}a$$
$$c = \frac{1}{3}a + \frac{1}{2}b + \frac{1}{2}d$$
$$d = \frac{1}{3}a + \frac{1}{2}b$$

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$
Observation:

→ The importance matrix is stochastic
   (columns sum to 1: e.g. A has 3 links, each with importance \( \frac{1}{3} \)) (unless there’s a page with no links…)

→ The rank vector is an eigenvector with eigenvalue 1
   (the $25$ billion eigenvector)

In this case, the 1-eigenspace is spanned by

\[
\omega = \frac{1}{31} \begin{pmatrix} 12 \\ 9 \\ 6 \end{pmatrix} \quad \text{w.r.t.} \quad a = \frac{12}{31} \quad c = \frac{4}{31} \\
12 = \frac{4}{31} \quad d = \frac{6}{31}
\]

(normalize so they sum to 1).

→ A is most important!

Random Surfer Interpretation:

If the random surfer has probabilities \((a, b, c, d)\) of being on pages A B C D, then after the next click he has probabilities

\[
\begin{pmatrix}
\frac{1}{3}a \\
\frac{1}{3}a + \frac{1}{2}b \\
\frac{2}{3}a + \frac{1}{2}b
\end{pmatrix} =
\begin{pmatrix}
0 & 0 & 1 & \frac{12}{31} \\
\frac{1}{3} & 0 & 0 & 0 \\
\frac{1}{3} & \frac{1}{2} & 0 & 0
\end{pmatrix}
\begin{pmatrix}
a \\
b \\
c \\
d
\end{pmatrix}
\]

of being on each page.

So the rank vector is the steady state for the random surfer → spends more time on important pages.
Observation: this importance matrix is usually stochastic but not positive stochastic, so we can’t apply Perron-Frobenius. Does this cause problems? Yes!

Eg (Disconnected Internet): Consider this Internet:

\[
\begin{pmatrix}
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/2 & 1/2 & 0 \\
0 & 0 & 1/2 & 1/2 & 0 \\
0 & 0 & 0 & 1/2 & 1/2 \\
\end{pmatrix}
\]

But both (1,1,0,0,0) and (0,0,1,1,1) are 1-eigen vectors! rank vector is not unique!

The Google Matrix

Page & Brin’s solution is as follows.

For a damping factor \( p \in (0,1) \) (e.g., \( p = 0.15 \)), let \( A \) be the importance matrix and let

\[
B = \frac{1}{N} \left( \begin{array}{c}
1 \\
1 \\
\vdots \\
1
\end{array} \right) \quad N = \# \text{pages} \quad (N \times N)
\]
The Google Matrix is
\[ G = (1-p)A + pB \]

For example, in a disconnected internet example:
\[ G = (1-p) \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} + p \begin{pmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix} \]

Fact: the Google matrix is positive stochastic.
- Stochastic: the columns of \((1-p)A\) sum to \(1-p\)
  the columns of \(pB\) sum to \(p\)
  columns of \(G\) sum to 1
- Positive: because \(pB\) has positive entries.

Random Surfer Interpretation:
With probability \(p\), the random surfer navigates to a random page anywhere on the Internet; he clicks on a random link otherwise.

Larry Page

Def: The PageRank vector is the steady state of the Google matrix.
So the importance of a page is the value of its coordinate.