Stochastic Matrices

Stochastic Matrices (1)

Definition: A **difference equation** is an equation of the form

$$v_{t+1} = Av_t$$

where A is an $n \times n$ matrix and $v_0, v_1, v_2, \ldots$ are vectors in $\mathbb{R}^n$

- v_t is the state at time t
- v_{t+1} is the state at time $t+1$
- $v_{t+1} = Av_t$ means that A is the *change of state matrix*.

$$v_t = Av_{t-1} = A^2 v_{t-2} = \cdots = A^t v_0$$

We are interested in the long-term behavior of this difference equation which can be analyzed through the dynamics of the change of state matrix and its consequences.

The problem

There are three video rental kiosks in a small town. Every customer returns their movie the next day. Let v_t be the vector whose entries x_t, y_t, z_t are the numbers of copies of Scary Movie at kiosks 1,2 and 3, respectively. Let A be the matrix whose i, j-entry is the probability that a customer renting Scary Movie from kiosk j returns it to kiosk i.

Stochastic Matrices (2)

Here's an example of a change of state matrix for this problem

$$A = \begin{array}{c} \\ \\ \\ \\ \end{array}\begin{array}{cc} & \begin{array}{ccc} k1 & k2 & k3 \end{array} \\ & \begin{bmatrix} 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.2 \end{bmatrix} \begin{array}{c} k1 \\ k2 \\ k3 \end{array} \end{array}$$

There is a 50% chance that a person renting from k3 returns it to k1.

For example, the number of movies returned to kiosk 2 will be:

30% of the movies from kiosk 1
40% of the movies from kiosk 2
30% of the movies from kiosk 3

When applied to all three rows we get $A\underbrace{\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix}}_{v_t} = \begin{bmatrix} .3x_t + .4y_t + .5z_t \\ .3x_t + .4y_t + .3z_t \\ .4x_t + .2y_t + .2z_t \end{bmatrix} = \underbrace{\begin{bmatrix} x_{t+1} \\ y_{t+1} \\ z_{t+1} \end{bmatrix}}_{v_{t+1}}$

$Av_t = v_{t+1}$ so, Av_t represents the number of movies in each kiosk the next day. This system is modeled by a difference equation.

Again, we are interested in the long-term behavior of this system, like how many movies will there be in each kiosk after 10 days, 100 days, 1000000 days and further in the future.

Stochastic Matrices (3)

Systems that involve difference equations representing probabilities are called Markov chains. The **Perron-Frobenius theorem** describes the long-term behavior of Markov chains.

Definitions:

- A square matrix is **stochastic** if all of its entries are nonnegative, and the entries of each column add up to 1.
- A **positive stochastic** matrix is a stochastic matrix whose entries are all positive numbers. In particular, no entry is equal to zero.
- A **regular stochastic** matrix is a stochastic matrix A such that A^n is positive for some integer $n \geq 1$ (that is, it is eventually positive).

Back to the video kiosks problem

The change of state matrix $A = \begin{bmatrix} 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.2 \end{bmatrix}$ is a positive stochastic matrix. The fact that columns add up to 1 says that all rented movies from one kiosk must be returned to some other kiosk, possibly itself. For example, of the movies rented from kiosk 1, 30% will be returned to kiosk 1, 30 % will be returned to kiosk 2, 40 % will be returned to kiosk 3

Stochastic Matrices (4)

Note that the sum of the column entries is 100%, as all of the movies are returned to one of the three kiosks. So, what we have so far is

$$\underset{v_{t+1}}{\begin{bmatrix} x_{t+1} \\ y_{t+1} \\ z_{t+1} \end{bmatrix}} = A \underset{v_t}{\begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix}} = \underset{v_{t+1}}{\begin{bmatrix} .3x_t + .4y_t + .5z_t \\ .3x_t + .4y_t + .3z_t \\ .4x_t + .2y_t + .2z_t \end{bmatrix}} \begin{matrix} k1 \\ k2 \\ k3 \end{matrix}$$

Note that if we add up the entries of v_{t+1}, we get $\underline{(.3x_t + .4y_t + .5z_t)}$+$(.3x_t + .4y_t + .3z_t)$+ $(.4x_t + .2y_t + .2z_t)$=$(.3 + .3 + .4)x_t + (.4 + .4 + .2)y_t + (.5 + .3 + .2)z_t = \underline{x_t} + \underline{y_t} + \underline{z_t}$

This says that the total number of copies of Scary Movie in the three kiosks does not change from day to day.

Let A be a stochastic matrix, let v_t be a vector and let $v_{t+1} = Av_t$. Then the sum of entries of v_t equals the sum of the entries of v_{t+1}

Fact: Let A be a stochastic matrix. Then

- 1 is an eigenvalue of A
- If λ is an eigenvalue of A and $\lambda \neq 1$ then $|\lambda| < 1$ (λ could be real or nonreal complex).

Stochastic Matrices (5)

Fact: Let A be a stochastic matrix. Then

- ➢ 1 is an eigenvalue of A
- ➢ If λ is an eigenvalue of A and $\lambda \neq 1$ then $|\lambda| < 1$ (λ could be real or nonreal complex).

Proof of Part 1 above:

Recall these properties of transpose: $A^T + B^T = (A+B)^T$; if D is diagonal then $D^T = D$, and finally, $\det(A) = \det(A^T)$

If A is stochastic, consider the matrix A^T. The rows of A^T add up to 1 and multiplying a matrix by the vector $\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ simply sums up the rows. For example, using the matrix A from our kiosk

$A^T \cdot v = 1 \cdot v$

problem we get $A^T \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} .3 & .3 & 4 \\ .4 & .4 & .2 \\ .5 & .3 & .2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} .3+.3+.4 \\ .4+.4+.2 \\ .5+.3+.2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. This shows that $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is an eigenvector for A^T with eigenvalue 1.

Stochastic Matrices (6)

Fact: Let A be a stochastic matrix. Then

- 1 is an eigenvalue of A
- For a positive stochastic matrix A, if λ is an eigenvalue of A and $\lambda \neq 1$ then $|\lambda| < 1$ (λ could be real or nonreal complex).

Proof of Part 1 above (continued):

$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is an eigenvector for A^T with eigenvalue 1. A and A^T have the same characteristic polynomial: $\det(A^T - \lambda I) = \det(A^T - (\lambda I)^T) = \det(A - \lambda I)^T = \det(A - \lambda I)$. It follows that 1 must be an eigenvalue of A.

We will not prove Part 2 but its usefulness will become clear.

Definition: A **steady state vector** of a stochastic matrix A is an eigenvector with eigenvalue 1 such the entries are positive and add up to 1.

Stochastic Matrices (7)

Perron-Frobenius Theorem: Let A be a positive stochastic matrix. Then A admits a unique steady state vector w, which spans the 1-eigenspace. Moreover, for any vector v_0 with entries summing to some number c, the iterates

$$v_1 = Av_0, \qquad v_2 = Av_1, \qquad \dots, \qquad v_t = Av_{t-1}$$

approach cw as t gets large.

Translation:

- The 1-eigenspace of a positive stochastic matrix is a line
- The 1-eigenspace contains a vector with positive entries
- All vectors approach the 1-eigenspace upon repeated multiplication by A.

We can think of the steady state vector w as a vector of fractions or percentages. Using the movie kiosks example, if the movies are distributed according to the percentages of the steady state vector today, they will have exactly the same distribution tomorrow, since for w, $Aw = w$. No matter what the initial distribution is, the long-term distribution, in terms of percentages, will always be the steady state vector.

The sum c of the entries of v_0 is the total number of things being modeled and does not change. So, the long-term state of the system must approach cw: it is a multiple of w because it's in the 1-eigenspace, and the entries of cw add up to c.

Stochastic Matrices (8)

Recipe 1: Compute the steady state vector. Let A be a positive stochastic matrix. Here is how to compute the steady-state vector of A.

1. Find any eigenvector v of A with eigenvalue 1 by solving $(A-I_n)v = 0$.
2. Divide v by the sum of the entries of v to obtain a vector w whose entries sum to 1.
3. This vector automatically has positive entries. It is the unique steady-state vector.

Recipe 2: Approximate the steady state vector by computer. Let A be a positive stochastic matrix. Here is how to approximate the steady-state vector of A with a computer.

1. Choose any vector v_0 whose entries sum to 1 (e.g., a standard coordinate vector).
2. Compute $v_1 = Av_0$, $v_2 = Av_1$, $v_3 = Av_2$, etc.
3. These converge to the steady state vector w.

Stochastic Matrices (9)

Example 1

Find the eigenvalues and corresponding eigenvectors of the given matrix. Then find the steady state vector. Verify by iterating A applied to the vector $u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$$A = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

$$f(\lambda) = \lambda^2 - Tr(A)\lambda + det(A) \quad Tr(A) = \frac{3}{2} \quad det(A) = \frac{9}{16} - \frac{1}{16} = \frac{1}{2}$$

$$f(\lambda) = \lambda^2 - \frac{3}{2}\lambda + \frac{1}{2} = 0 \Rightarrow 2\lambda^2 - 3\lambda + 1 = 0 \Rightarrow (2\lambda - 1)(\lambda - 1) = 0$$

$$\lambda_1 = 1 \quad \lambda_2 = \frac{1}{2}$$

$$\lambda_1 = 1 \quad A - I = \begin{bmatrix} \frac{3}{4} - 1 & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} - 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} x \\ y \end{bmatrix} = y\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\lambda_2 = \frac{1}{2} \quad A - \frac{1}{2}I = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} x \\ y \end{bmatrix} = y\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad v_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$w = \frac{1}{1+1}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

Verify by iterating A applied to the vector $u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. See next slide.

Stochastic Matrices (10)

Example 1 (continued)

Verify by iterating A applied to the vector $u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$$A = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix} \qquad \lambda_1 = 1 \qquad v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad W = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

$$\lambda_2 = \frac{1}{2} \qquad v_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

As the Perron-Frobenius Theorem asserts, the vectors u_t approach a vector whose entries are the same with 50% of total in first entry and 50% of total in the second. This can also be seen using algebra and taking the appropriate limit: the eigenvectors v_1 and v_2 span $\mathbb{R}^2$ and for any $x = a_1 v_1 + a_2 v_2 \implies$

$$Ax = A(a_1 v_1 + a_2 v_2) = a_1 \underbrace{Av_1}_{v_1} + a_2 \underbrace{Av_2}_{\frac{1}{2}v_2} = a_1 v_1 + \frac{1}{2} a_2 v_2$$

Iterating multiplication by A results in

$$A^t x = \underline{a_1 v_1} + \underbrace{\frac{a_2}{2^t}}_{0} v_2 \xrightarrow[as\ t \to \infty]{} a_1 v_1 = \begin{bmatrix} a_1 \\ a_1 \end{bmatrix}$$

Let $u_t = A^t u_0$

u_0	$(1,0)$
u_1	$\left(\frac{3}{4}, \frac{1}{4}\right)$
u_2	$\left(\frac{5}{8}, \frac{3}{8}\right)$
u_3	$\left(\frac{9}{16}, \frac{7}{16}\right)$
u_4	$\left(\frac{17}{32}, \frac{15}{32}\right)$
u_5	$\left(\frac{33}{64}, \frac{31}{64}\right)$
u_{10}	$\left(\frac{1025}{2048}, \frac{1023}{2048}\right)$
u_{15}	$(0.500015, 0.499985)$

Stochastic Matrices (11)

The dynamics of $A = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$

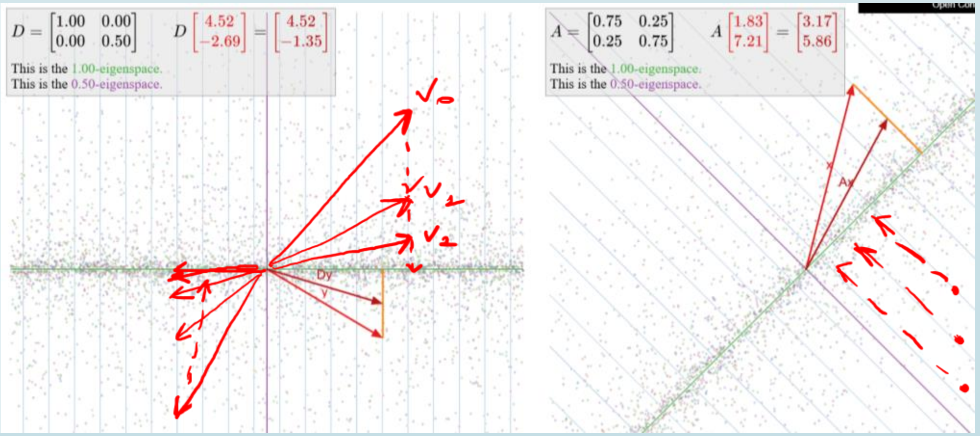
$\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ \frac{1}{2}y \end{bmatrix}$

Using the results of Example 1 we know that $A = CDC^{-1}$ where $C = \begin{array}{cc} v_1 & v_2 \\ \hline \end{array}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and $D = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$

$\begin{matrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{matrix}$

D does not change the x-coordinate of a vector and scales the y-coordinate by a factor of $\frac{1}{2}$. Upon iteration the x-coordinate remains unchanged while the y-coordinate shrinks closer and closer to zero "sucking" the vector into the x-axis.

The matrix A does the same but with respect to the coordinate system defined by the eigenvector columns of C. This means that A "sucks" all vectors into the 1-eigenspace. In addition, it does so without changing the sum of the vector entries.



Stochastic Matrices (12)

Back to the video kiosks problem

$Av_2 = \lambda_2 v_2$

Recall that the change of state matrix was $A = \begin{bmatrix} 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.2 \end{bmatrix}$

Its characteristic polynomial is $f(\lambda) = -\lambda^3 + 0.12\lambda - 0.02 = -(\lambda - 1)(\lambda + 0.2)(\lambda - 0.1)$

$\lambda_1 = 1$ $\qquad \lambda_2 = -0.2$ $\qquad \lambda_3 = 0.1$

$v_1 = \begin{bmatrix} 7 \\ 6 \\ 5 \end{bmatrix}$ $\qquad v_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ $\qquad v_3 = \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}$

So, the steady state vector is $w = \frac{1}{18}\begin{bmatrix} 7 \\ 6 \\ 5 \end{bmatrix} = \begin{bmatrix} \frac{7}{18} \\ \frac{6}{18} \\ \frac{5}{18} \end{bmatrix}$

Stochastic Matrices (13)

$Av_2 = -0.2v_2 \rightsquigarrow 0$

$Av_3 = 0.1v_3 \rightsquigarrow 0$

The dynamics of the video kiosks problem

Eigenvectors v_1, v_2, v_3 form a basis for $\mathbb{R}^3$ and for any vector $x = a_1v_1 + a_2v_2 + a_3v_3$ we have

$$Ax = A(a_1v_1 + a_2v_2 + a_3v_3) = a_1\underbrace{Av_1}_{v_1} + a_2\underbrace{Av_2}_{-0.2v_2} + a_3\underbrace{Av_3}_{0.1v_3} = a_1v_1 - 0.2a_2v_2 + 0.1a_3v_3$$

$$A^tx = a_1v_1 + \underbrace{(-0.2)^t}_{0}a_2v_2 + \underbrace{(0.1)^t}_{0}\,a_3v_3 \xrightarrow[as\ t\to\infty]{} \underline{a_1v_1}$$

So, A^tx approaches a_1v_1, an eigenvector with eigenvalue 1

Let's say that there are a total of 180 movies initially distributed as $v_0 = (40,80,60)$

t	v_t
0	(40,80,60)
1	(72,62,44)
2	(69,60.2,50.8)
3	(70.18,60.02,49.8)
5	(70.0074,60.0002,49.9924)
10	(70,60,50)

We easily see that the long-term distribution is equal to $180w =$

$$180\begin{bmatrix}\frac{7}{18}\\ \frac{6}{18}\\ \frac{5}{18}\end{bmatrix} = \begin{bmatrix}70\\60\\50\end{bmatrix}$$ and this is independent

of the initial distribution.

Stochastic Matrices (14)

The dynamics of the video kiosks problem

$$A = CDC^{-1} \text{ where } C = \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{bmatrix} 7 & -1 & 1 \\ 6 & 0 & -3 \\ 5 & 1 & 2 \end{bmatrix} \end{array} \text{ and } D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.2 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}$$

The matrix D leaves the x-coordinate of a vector unchanged, scales the y-coordinate by $-\frac{1}{5}$ and shrinks the z-coordinate by $\frac{1}{10}$. Upon iteration, the y- and z-coordinates become very small while the x-coordinate remains the same "sucking" the vector into the x-axis.

A behaves similarly but with respect to the eigenspaces spanned by v_1, v_2 and v_3. So, under the action of A, all vectors get "sucked" into the 1-eigenspace, without changing the sum of the entries.