MAT 1302B – Mathematical Methods II

Alistair Savage

Mathematics and Statistics
University of Ottawa

Winter 2015 – Lecture 20

Announcements

Review class (next class):

- The last class is open for review.
- We can go over theory, do extra examples, or go through an old exam.
- Topics?

Exam period office hours:

- Monday, April 13, 2–3pm

All students may attend the DGDs on Monday, April 13.

Difference equations and Markov Chains

Suppose we want to model a system that changes over time.

Examples:

1. Populations of cities, countries, etc.
2. Populations of certain types of wildlife.
3. Customers of various competing companies.

We measure the state of the system at discrete time intervals.

Examples:

1. Census every few years.
2. Wildlife study (tag/release) conducted once a year.
3. Customer data tabulated every month.

Example: Market share

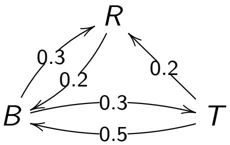
Suppose in a country (say, Canada), the cell phone market is dominated by three companies R, B and T (say, Rogers, Bell and Telus).

Assumption: Due to the limited options, we assume that these three companies control the entire market.

Instead of working with absolute numbers, we will work with market share (percentage of the market that each company has).

Example: Market share

Suppose each month the following migration occurs:



Suppose that currently

- R has $1/2$ of the market share, and
- B and T each have $1/4$ of the market share.

Question: What will be the market share

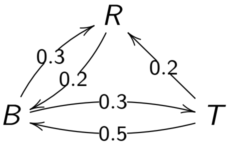
1. after one month?
2. after six months?

Example: Market shares

Let

- r_k denote R's market share k months in the future.
- b_k denote B's market share k months in the future.
- t_k denote T's market share k months in the future.

Then we have



$$\begin{aligned} r_{k+1} &= 0.8r_k + 0.3b_k + 0.2t_k \\ b_{k+1} &= 0.2r_k + 0.4b_k + 0.5t_k \\ t_{k+1} &= 0r_k + 0.3b_k + 0.3t_k \end{aligned}$$

Let's turn this into a matrix/vector equation. If

$$\vec{x}_k = \begin{bmatrix} r_k \\ b_k \\ t_k \end{bmatrix}, \quad \text{then} \quad \vec{x}_{k+1} = \begin{bmatrix} .8 & .3 & .2 \\ .2 & .4 & .5 \\ 0 & .3 & .3 \end{bmatrix} \vec{x}_k.$$

Example: Market share

$$\text{Let} \quad M = \begin{bmatrix} .8 & .3 & .2 \\ .2 & .4 & .5 \\ 0 & .3 & .3 \end{bmatrix}.$$

Note: The entries in each column add up to one (since every customer of a company must move/stay to/with some company).

The initial market share is given by $\vec{x}_0 = \begin{bmatrix} .5 \\ .25 \\ .25 \end{bmatrix}$. Therefore,

$$\vec{x}_1 = M\vec{x}_0 = \begin{bmatrix} .8 & .3 & .2 \\ .2 & .4 & .5 \\ 0 & .3 & .3 \end{bmatrix} \begin{bmatrix} .5 \\ .25 \\ .25 \end{bmatrix} = \begin{bmatrix} 0.525 \\ 0.325 \\ 0.15 \end{bmatrix}.$$

So after one month, R has 52.5%, B has 32.5% and T has 15% of the market share (note these add up to 100%!).

Example: Market share

To find the market shares after six months, we compute

$$\vec{x}_6 = M^6 \vec{x}_0 = \begin{bmatrix} .8 & .3 & .2 \\ .2 & .4 & .5 \\ 0 & .3 & .3 \end{bmatrix}^6 \begin{bmatrix} .5 \\ .25 \\ .25 \end{bmatrix} \approx \begin{bmatrix} 0.572 \\ 0.299 \\ 0.129 \end{bmatrix}.$$

Therefore

- R has 57.2% of the market share,
- B has 29.9% of the market share,
- T has 12.9% of the market share.

Note: The market shares add up to 100%.

Difference equations

We have a sequence of vectors

$$\vec{x}_0, \quad \vec{x}_1, \quad \vec{x}_2, \ldots$$

where the entries in $\vec{x}_k$ provide information about the state of some system at time k.

If there is a matrix M such that

$$\vec{x}_{k+1} = M\vec{x}_k \tag{1}$$

(so $\vec{x}_1 = M\vec{x}_0$, $\vec{x}_2 = M\vec{x}_1$, etc.) then the equation (1) is called a linear difference equation or recurrence relation. The matrix M is called the migration matrix or transition matrix.

If we know the initial state of the system ($\vec{x}_0$), then we can use (1) to compute the state of the system at any later time ($\vec{x}_1$, $\vec{x}_2$, etc.).

Difference equations: the distant future

To find the state of the system in the distant future, we need to be able to find $\vec{x}_k$ for large value of k.

Since

$$\vec{x}_k = M^k \vec{x}_0,$$

this amounts to computing large powers of the matrix M.

Recall: We used diagonalization to compute high powers of matrices (provided the matrices were diagonalizable).

This is one of the reasons why diagonalization is useful.

Example: Probabilities

The exact same type of problem can be worded in terms of probabilities.

Example: Suppose the weather is either sunny or rainy.

- If it is sunny today, there is a 75% chance it will be sunny tomorrow.
- If it is rainy today, there is a 60% chance it will be rainy tomorrow.

If it is sunny today, what is the probability that it will be sunny two days from now?

Solution: Probabilities can be represented by numbers between 0 and 1. The migration matrix is

$$M = \begin{bmatrix} 0.75 & 0.4 \\ 0.25 & 0.6 \end{bmatrix}.$$

The current state is

$$\vec{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

(the first entry is the probability it is sunny and second is the probability it is rainy).

Example: Probabilities

$$M = \begin{bmatrix} 0.75 & 0.4 \\ 0.25 & 0.6 \end{bmatrix}, \quad \vec{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

We want the probabilities 2 days from now. So we want to compute $\vec{x}_2$.

We have

$$\vec{x}_1 = M\vec{x}_0 = \begin{bmatrix} 0.75 & 0.4 \\ 0.25 & 0.6 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.75 \\ 0.25 \end{bmatrix}.$$

Therefore,

$$\vec{x}_2 = M\vec{x}_1 = \begin{bmatrix} 0.75 & 0.4 \\ 0.25 & 0.6 \end{bmatrix} \begin{bmatrix} 0.75 \\ 0.25 \end{bmatrix} = \begin{bmatrix} 0.6625 \\ 0.3375 \end{bmatrix}.$$

(Note that the entries add up to 1.)

So there is a 66.25% chance that it will be sunny two days from now.

Example: A tale of two cities

Suppose two cities, Appolonia and Boondocks, have no contact with the outside world. Each year, 30% of the population of Appolonia moves to Boondocks and 20% of the population of Boondocks moves to Appolonia.

In the year 2013, the populations of A and B are both 500 000. What are the populations in 2014 and 2015?

Solution: We form the migration matrix

$$M = \begin{bmatrix} .7 & .2 \\ .3 & .8 \end{bmatrix}.$$

Note: The entries in each column add up to one (so M is a stochastic matrix).

If we work with fractions of the total population (instead of absolute numbers), the initial state is half the population in each city:

$$\vec{x}_0 = \begin{bmatrix} .5 \\ .5 \end{bmatrix} \qquad \text{(entries add up to one).}$$

Example: A tale of two cities

After one year, the populations are given by

$$\vec{x}_1 = M\vec{x}_0 = \begin{bmatrix} .7 & .2 \\ .3 & .8 \end{bmatrix} \begin{bmatrix} .5 \\ .5 \end{bmatrix} = \begin{bmatrix} (.7)(.5) + (.2)(.5) \\ (.3)(.5) + (.8)(.5) \end{bmatrix} = \begin{bmatrix} .45 \\ .55 \end{bmatrix}.$$

Since the total population of the two cities is 1 million, in 2014,

- the population of A is $(.45)(1\,000\,000) = 450\,000$, and
- the population of B is $(.55)(1\,000\,000) = 550\,000$.

After two years,

$$\vec{x}_2 = M^2\vec{x}_0 = M\vec{x}_1 = \begin{bmatrix} .7 & .2 \\ .3 & .8 \end{bmatrix} \begin{bmatrix} .45 \\ .55 \end{bmatrix} = \begin{bmatrix} .425 \\ .575 \end{bmatrix}.$$

Therefore, in 2015,

- the population of A is $(.425)(1\,000\,000) = 425\,000$, and
- the population of B is $(.575)(1\,000\,000) = 575\,000$.

Terminology

Probability vector: a vector with nonnegative entries that add up to one.

Stochastic matrix: a square matrix whose columns are probability vectors. In other words, a square matrix with nonnegative entries where the entries in each column add up to one.

Definition (Markov chain)

A Markov chain is a sequence of probability vectors

$$\vec{x}_0,\ \vec{x}_1,\ \vec{x}_2,\ \dots$$

together with a stochastic matrix P such that

$$\vec{x}_1 = P\vec{x}_0, \quad \vec{x}_2 = P\vec{x}_1, \quad \vec{x}_3 = P\vec{x}_2, \ \dots.$$

The Markov chain is described by the difference equation

$$\vec{x}_{k+1} = P\vec{x}_k \quad \text{for } k = 0, 1, 2, \dots.$$

The vectors $\vec{x}_k$ are called state vectors.

Example: A tale of two cities (cont.)

We had

$$M = \begin{bmatrix} .7 & .2 \\ .3 & .8 \end{bmatrix}, \ \vec{x}_0 = \begin{bmatrix} .5 \\ .5 \end{bmatrix}, \ \vec{x}_1 = M\vec{x}_0 = \begin{bmatrix} .45 \\ .55 \end{bmatrix}, \ \vec{x}_2 = M^2\vec{x}_0 = \begin{bmatrix} .425 \\ .575 \end{bmatrix}.$$

Question: What are the populations in the long term?

$$\vec{x}_3 = M\vec{x}_2 = \begin{bmatrix} 0.4125 \\ 0.5875 \end{bmatrix}, \ \vec{x}_4 = M\vec{x}_3 = \begin{bmatrix} 0.40625 \\ 0.59375 \end{bmatrix}, \ \vec{x}_5 = M\vec{x}_4 = \begin{bmatrix} 0.403125 \\ 0.596875 \end{bmatrix}$$

It seems like the $\vec{x}_k$ are approaching $\begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}$. Note that

$$M \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix} = \begin{bmatrix} .7 & .2 \\ .3 & .8 \end{bmatrix} \begin{bmatrix} .4 \\ .6 \end{bmatrix} = \begin{bmatrix} .4 \\ .6 \end{bmatrix}$$

So $\vec{q} = \begin{bmatrix} .4 \\ .6 \end{bmatrix}$ is an eigenvector with eigenvalue 1.

Some more terminology

If P is a stochastic matrix, then an eigenvector of P with eigenvalue 1 is called a steady-state vector or equilibrium vector.

A stochastic matrix P is regular if some power P^k $(k \geq 1)$ contains only strictly positive (that is, > 0) entries.

Example:

$$P = \begin{bmatrix} .2 & .3 & .6 \\ .2 & .4 & 0 \\ .6 & .3 & .4 \end{bmatrix}, \qquad P^2 = \begin{bmatrix} .46 & .36 & .36 \\ .12 & .22 & .12 \\ .42 & .42 & .52 \end{bmatrix}$$

Since P is stochastic (the entries are nonnegative and the entries in each column add up to one) and the entries of P^2 are strictly positive, P is a regular stochastic matrix.

Steady-state vectors

Theorem

If P is a regular stochastic matrix, then P has a unique steady-state vector $\vec{q}$.

If $\vec{x}_0$ is any initial state and

$$\vec{x}_{k+1} = P\vec{x}_k \qquad \text{for } k = 0, 1, 2, \ldots,$$

then the Markov chain $\vec{x}_0, \vec{x}_1, \ldots$ converges to $\vec{q}$ as $k \to \infty$. That is, the entries of $\vec{x}_k$ approach the entries of $\vec{q}$ as $k \to \infty$.

So for a Markov chain whose migration matrix is regular stochastic, the long term behavior is given by the eigenvector of eigenvalue 1.

Note: There will be an entire 1-dimensional eigenspace of eigenvalue 1. However, we want an eigenvector that is a probability vector (nonnegative entries adding up to one). There is only one of these.

Example: Voting

Suppose we want to model the voting in Canadian federal elections.

$$\vec{x} = \begin{bmatrix} \text{\% voting Liberal (L)} \\ \text{\% voting Conservative (C)} \\ \text{\% voting NDP (N)} \end{bmatrix}$$

Assume that

- Of the people that vote L in one election, 30% vote C in the next election and 40% vote N.
- Of the people that vote C in one election, 40% vote L in the next election and 20% vote N.
- Of the people that vote N in one election, 50% vote L in the next election and 30% vote C.

So the migration matrix is

$$P = \begin{bmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{bmatrix}.$$

Example: Voting

Suppose the outcome of one election is given by

$$\vec{x}_0 = \begin{bmatrix} .3 \\ .5 \\ .2 \end{bmatrix}.$$

Determine the outcomes of the next two elections.

Solution:

$$\vec{x}_1 = P\vec{x}_0 = \begin{bmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{bmatrix} \begin{bmatrix} .3 \\ .5 \\ .2 \end{bmatrix} = \begin{bmatrix} .39 \\ .35 \\ .26 \end{bmatrix}$$

(Partial check: entries add up to one. $.39 + .35 + .26 = 1$)

Therefore, in the next election, 39% will vote L, 35% will vote C and 26% will vote N.

Example: Voting

$$\vec{x}_2 = P\vec{x}_1 = \begin{bmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{bmatrix} \begin{bmatrix} .39 \\ .35 \\ .26 \end{bmatrix} = \begin{bmatrix} .387 \\ .335 \\ .278 \end{bmatrix}$$

(Partial check: entries add up to one. $.387 + .335 + .278 = 1$)
Therefore, in the election after next, 38.7% will vote L, 33.5% will vote C and 27.8% will note N.

Question: What will be the voting behavior in the long term?

Note that in the migration matrix

$$P = \begin{bmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{bmatrix},$$

- all entries are strictly positive, and
- the sum of entries in each column is 1.

So P is regular stochastic.

Example: Voting

Since the migration is regular stochastic, the voting in the long term will approach the unique probability eigenvector of eigenvalue one.

We just need to find this eigenvector!

To find the eigenspace for the eigenvalue $\lambda = 1$, we must solve $(P - I)\vec{q} = \vec{0}$. So we row reduce

$$\left[\begin{array}{c|c} P - I & \vec{0} \end{array}\right] = \left[\begin{array}{ccc|c} -.7 & .4 & .5 & 0 \\ .3 & -.6 & .3 & 0 \\ .4 & .2 & -.8 & 0 \end{array}\right] \xrightarrow{\text{row reduce}} \left[\begin{array}{ccc|c} 1 & 0 & -7/5 & 0 \\ 0 & 1 & -6/5 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$\begin{array}{ll} q_1 & = \frac{7}{5} q_3 \\ q_2 & = \frac{6}{5} q_3 \\ q_3 & \text{free} \end{array} \implies \vec{q} = \begin{bmatrix} \frac{7}{5} q_3 \\ \frac{6}{5} q_3 \\ q_3 \end{bmatrix} = q_3 \begin{bmatrix} 7/5 \\ 6/5 \\ 1 \end{bmatrix}$$

Remember we want the entries of $\vec{q}$ to add up to one. So we want

$$1 = \frac{7}{5} q_3 + \frac{6}{5} q_3 + q_3 = \frac{18}{5} q_3 \implies q_3 = \frac{5}{18}.$$

Example: Voting

Thus, the long term voting is given by the steady-state vector

$$\vec{q} = \frac{5}{18}\begin{bmatrix} 7/5 \\ 6/5 \\ 1 \end{bmatrix} = \begin{bmatrix} 7/18 \\ 6/18 \\ 5/18 \end{bmatrix}.$$

So, in the long term

- 7/18 of the population will vote L,
- 6/18 of the population will vote C,
- 5/18 of the population will vote N.

Example: Voting

Check your answer!

1. The entries in the steady-state vector should add to one:

$$\frac{7}{18} + \frac{6}{18} + \frac{5}{18} = \frac{18}{18} = 1 \qquad \checkmark$$

2. The steady-state vector (or any nonzero multiple of it) should be an eigenvector of the migration matrix with eigenvalue 1:

$$P(18\vec{q}) = \begin{bmatrix} .3 & .4 & .5 \\ .3 & .4 & .3 \\ .4 & .2 & .2 \end{bmatrix} \begin{bmatrix} 7 \\ 6 \\ 5 \end{bmatrix} = \begin{bmatrix} 7 \\ 6 \\ 5 \end{bmatrix} \qquad \checkmark$$

Note: we use $\begin{bmatrix} 7 \\ 6 \\ 5 \end{bmatrix} = 18\vec{q}$ instead of $\vec{q}$ itself to avoid the fractions.

Demonstration

Mathematica demonstration

http://demonstrations.wolfram.com/FiniteStateDiscreteTimeMarkovChains/

How to solve Markov chain problems

First steps:

1. Use the information given to write down the migration matrix M and initial state vector $\vec{x}_0$.

To find the state vector k units of time (unit depends on the problem) in the future:

2. Compute
$$\vec{x}_k = M^k \vec{x}_0$$

To find the state vector in the distant future (steady-state vector):

3. Check that the migration matrix M is regular stochastic:
    - stochastic: entries are nonnegative, sum of entries in each column is one
    - regular: some power of the matrix (maybe the matrix itself, i.e. the first power) has strictly positive (> 0) entries
4. Find the eigenspace corresponding to the eigenvalue 1 (this space should be one dimensional).
5. Find the value of the free variable that makes the entries of the eigenvector add up to one.

Warning!

Don't forget that when finding the long term behaviour, you need to justify that the steady-state vector (the eigenvector of eigenvalue one) really does describe the long term behaviour.

To do this, you need to check that the migration matrix is regular stochastic.

If you don't verify this, your solution is incomplete (and you'll lose marks on the final exam, for instance).

Example

If

$$M = \begin{bmatrix} .7 & .4 & .2 \\ .3 & .2 & .6 \\ 0 & .4 & .2 \end{bmatrix}, \quad \text{then} \quad M^2 = \begin{bmatrix} .61 & .41 & .42 \\ .27 & .4 & .3 \\ .12 & .16 & .28 \end{bmatrix}$$

Column sums of M are one, entries of M are nonnegative, and all entries of M^2 are strictly greater than zero, so M is a regular stochastic matrix. Therefore, the long term behavior is given by the steady-state vector.

Google PageRank

First internet search engines: Search results were based solely on the frequency of occurrence of the search term.

Google's idea: Frequency of the search term should be combined with the "importance of the page."

Question: How do we quantitatively measure the importance of page?

Basic idea: The more links from important pages pointing to a given page, the more important it is. This definition is recursive (importance of a page depends on the importance of the pages linking to it).

Same idea can be applied to many other situations (Facebook, academic journal rankings, etc.).

Google PageRank

Suppose a large number of web surfers start at random sites on the internet (even distribution).

Every unit of time, each surfer clicks on a random link on the page.

This is an example of a difference equation with huge matrices and vectors (billions of entries).

As we evolve the system (i.e. multiply by the migration matrix), the surfers will start to accumulate at the more important pages.

We look for a steady state: a state vector $\vec{q}$ such that

$$M\vec{q} = \vec{q} \qquad (M = \text{migration matrix}).$$

This is an eigenvector with eigenvalue one! The entries are the PageRank of the page (more or less).

Google PageRank

Finding the eigenspace of such a large matrix is impractical, so it is numerically approximated.

In the actual PageRank algorithm, the process is slightly more complicated:

- random jumps: there is some probability that a web surfer will jump to a random site on the web rather than following a link (this solves some problems like including sites not linked to any others)
- damping factor: eventually surfers stop clicking (this solves some other problems like all surfers eventually ending up at a relatively small number of sites)

However, the fundamental underlying idea is that of a Markov chain, which we now understand.

Next time

For next time:

- Look at final exams for previous years.
- Think of questions you want to ask during the review class.

Next class:

- Discussion of the final exam.
- Review.