Matrix methods – Regular Markov chains

MATH 2740 – Mathematics of Data Science – Lecture 13

Julien Arino

julien.arino@umanitoba.ca

Department of Mathematics @ University of Manitoba

Fall 202X

The University of Manitoba campuses are located on original lands of Anishinaabeg, Ininew, Anisininew, Dakota and Dene peoples, and on the National Homeland of the Red River Métis. We respect the Treaties that were made on these territories, we acknowledge the harms and mistakes of the past, and we dedicate ourselves to move forward in partnership with Indigenous communities in a spirit of Reconciliation and collaboration.

Outline

- Markov chains
- Running example – Mendelian inheritance
- Repetition of the process
- Regular Markov chains

Markov chains

Running example – Mendelian inheritance

Repetition of the process

Regular Markov chains

Markov chain

A Markov chain is a *stochastic process* in which the evolution through time depends only on the current state of the system (we say the process is *memoryless*)

Markov chains are an interesting combination of matrix theory and graph theory

They form the theoretical foundation for Hidden Markov processes or Markov Chain Monte Carlo (MCMC) methods, are used in ML

Basic principle

Conduct an experiment with a set of n possible outcomes

$$S = \{S_1, \ldots, S_n\}$$

Experiment repeated t times (with t large, potentially infinite)

Think of t as *time*

System has *no memory*: the next state depends only on the present state

Probability of S_i occurring at time $t + 1$ given that S_j occurred at time t is

$$p_{ij} = \mathbb{P}(S_i \mid S_j)$$

Suppose that S_i is the current state, then one of $S_1, \ldots, S_n$ must be the next state, so

$$p_{1i} + p_{2i} + \cdots + p_{ni} = 1, \quad 1 \leq i \leq n$$

Some of the p_{ij} can be zero, all that is needed is that $\sum_{k=1}^{n} p_{ki} = 1$ for all $i = 1, \ldots, n$

Definition 90 (Markov chain)

An experiment with finite number of possible outcomes $S_1, \ldots, S_n$ is repeated. The sequence of outcomes is a **Markov chain** if there is a set of n^2 numbers $\{p_{ij}\}$ such that the conditional probability of outcome S_i on any experiment given outcome S_j on the previous experiment is p_{ij}, i.e., for $1 \leq i, j \leq n$, $t = 1, \ldots$,

$$p_{ij} = \mathbb{P}(S_i \text{ on experiment } t+1 \mid S_j \text{ on experiment } t)$$

Outcomes $S_1, \ldots, S_n$ are **states** and p_{ij} are **transition probabilities**. $P = [p_{ij}]$ the **transition matrix**

In the following, we often write

$$\mathbb{P}(S_i \text{ on experiment } t+1 \mid S_j \text{ on experiment } t) \text{ as } \mathbb{P}(S_i(t+1) \mid S_j(t))$$

The matrix

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix}$$

has

- entries that are probabilities, i.e., $0 \leq p_{ij} \leq 1$
- column sum 1, which we write

$$\sum_{i=1}^{n} p_{ij} = 1, \quad j = 1, \ldots, n$$

or, using the notation $\mathbb{1}^T = (1, \ldots, 1)$,

$$\mathbb{1}^T P = \mathbb{1}^T$$

Markov chains

Running example – Mendelian inheritance

Repetition of the process

Regular Markov chains

The “orange” gene

A cat’s coat color is determined by many genes. The "orange" trait comes from a specific gene called the **orange locus**

It has two *alleles* (versions):

- **O** $\Longrightarrow$ Produces *phaeomelanin* (red/orange pigment)
- **o** $\Longrightarrow$ Produces *eumelanin* (black/brown pigment)

This gene is *sex-linked*. It is located on the **X chromosome**

- This changes the rules of inheritance!

How sex-linked genes work

Because the gene is on the X chromosome, males and females inherit it differently

Females have two X chromosomes (XX)

- They get two alleles for this gene (one from each parent)
- Possible genotypes: $X^O X^O$, $X^o X^o$, or $X^O X^o$

Males have one X and one Y chromosome (XY)

- They get *only one* allele for this gene (always from the mother)
- Possible genotypes: $X^O Y$ or $X^o Y$

Genotype vs. phenotype

Males (simple):

- $X^O Y \implies$ **orange cat**
- $X^o Y \implies$ **non-orange cat** (e.g., black)

Females (the special case):

- $X^O X^O \implies$ **orange cat**
- $X^o X^o \implies$ **non-orange cat** (e.g., black)
- $X^O X^o \implies$ **tortoiseshell cat**

A “tortie” isn’t a simple hybrid. Both alleles (O and o) are active in different patches of skin, creating the orange and black mottled pattern

Example 1: Orange dad + black Mom

Let's cross an **orange male ($X^O Y$)** with a **black female ($X^o X^o$)**

		Father	
		X^O	Y
Mother	X^o	$X^O X^o$	$X^o Y$
	X^o	$X^O X^o$	$X^o Y$

Results for their offspring:

- **All females** ($X^O X^o$) will be **tortoiseshell**
- **All males** ($X^o Y$) will be **black** (non-orange)

Example 2: black dad + tortoiseshell mom

Let's cross a **black male (X^oY)** with a **tortoiseshell female (X^OX^o)**

		Father	
		X^o	Y
Mother	X^O	X^OX^o	X^OY
	X^o	X^oX^o	X^oY

Results for their offspring (1/4 chance for each):

- $X^OX^o \implies$ **Tortoiseshell Female**
- $X^oX^o \implies$ **Black Female**
- $X^OY \implies$ **Orange Male**
- $X^oY \implies$ **Black Male**

Fun fact: what about male tortoiseshells?

- As we saw, a male is XY. He can only get X^O or X^o from his mother, not both
- A male tortoiseshell is possible, but *extremely rare*
- It's a genetic anomaly where the cat has an extra X chromosome: **XXY**
- This genotype (e.g., $X^O X^o Y$) allows the cat to be male (Y) but also express both orange and non-orange alleles ($X^O X^o$), just like a female

Markov chains

Running example – Mendelian inheritance

Repetition of the process

Regular Markov chains

Deriving the update equation

For $1 \leq i \leq n$, let $p_i(t)$ be probability that state S_i occurs at time t, which we also write $p_i(t) = \mathbb{P}(S_i(t))$

Since one of the states S_i must occur on the t^{th} repetition, we must have

$$p_1(t) + p_2(t) + \cdots + p_n(t) = 1$$

We want to use this information to derive $p_i(t+1)$, the probability that state S_i occurs at time $t + 1$

How do we get to S_i?

List all possible (n) ways to be in state S_i at time $t+1$:

1. We were in S_1, which happened with probability $p_1(t)$, then moved from S_1 to S_i, which has probability p_{i1}. Thus,

$$\mathbb{P}(S_i(t+1) \mid S_1(t)) = \mathbb{P}(S_i(t+1) \mid S_1(t))\, \mathbb{P}(S_1(t)) = p_{i1} p_1(t)$$

2. Likewise, if we were in S_2 at time t, then

$$\mathbb{P}(S_i(t+1) \mid S_2(t)) = \mathbb{P}(S_i(t+1) \mid S_2(t))\, \mathbb{P}(S_2(t)) = p_{i2} p_2(t)$$

..

n. Finally, if we were in S_n,

$$\mathbb{P}(S_i(t+1) \mid S_n(t)) = \mathbb{P}(S_i(t+1) \mid S_n(t))\, \mathbb{P}(S_n(t)) = p_{in} p_n(t)$$

Sum things up

So, for a given state $i = 1, \ldots, n$,

$$\begin{aligned} p_i(t+1) &= P(S_i(t+1) \mid S_1(t)) + \cdots + P(S_i(t+1) \mid S_n(t)) \\ &= p_{i1} p_1(t) + \cdots + p_{in} p_n(t) \end{aligned}$$

Therefore, since this must be true for all states $i = 1, \ldots, n$

$$\begin{aligned} p_1(t+1) &= p_{11} p_1(t) + p_{12} p_2(t) + \cdots + p_{1n} p_n(t) \\ &\vdots \\ p_n(t+1) &= p_{n1} p_1(t) + p_{n2} p_2(t) + \cdots + p_{nn} p_n(t) \end{aligned}$$

In matrix form

$$p(t+1) = Pp(t), \quad n = 1, 2, 3, \ldots$$

where $p(t) = (p_1(t), p_2(t), \ldots, p_n(t))^T$ is a probability vector and $P = (p_{ij})$ is an $n \times n$ *transition matrix*,

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ & & & \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix}$$

So

$$\begin{pmatrix} p_1(t+1) \\ \cdots \\ p_n(t+1) \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix} \begin{pmatrix} p_1(t) \\ \cdots \\ p_n(t) \end{pmatrix}$$

Easy to check that this gives the same expression as before

Stochastic matrices

Definition 91 (Stochastic matrix)

The nonnegative $n \times n$ matrix M is **row-stochastic** (resp. **column-stochastic**) if $\sum_{j=1}^{n} a_{ij} = 1$ for all $i = 1, \ldots, n$ (resp. $\sum_{i=1}^{n} a_{ij} = 1$ for all $j = 1, \ldots, n$)

We often say **stochastic** and let the context determine whether we mean row- or column-stochastic

If it is both row- and column-stochastic, the matrix is **doubly stochastic**

Theorem 92

Let $M \in \mathcal{M}_n$ be a stochastic matrix. Then all eigenvalues λ of M are such that $|\lambda| \leq 1$.

Theorem 93

Let $M \in \mathcal{M}_n$ be a stochastic matrix. $\lambda = 1$ is an eigenvalue of M. If M is row-stochastic, the eigenvalue 1 is associated to the column vector of ones (a right eigenvector of M); if M is column-stochastic, the eigenvalue 1 is associated to the row vector of ones (a left eigenvector of M)

Proof of Theorem 93

Suppose $M \in \mathcal{M}_n$ is row-stochastic. One way to write the requirement that each row sum equals 1 is

$$M\mathbf{1} = \mathbf{1} \tag{1}$$

where $\mathbf{1} = (1, \dots, 1) \in \mathbb{C}^n$ is a column vector

If $M \in \mathcal{M}_n$, then the eigenpair equation takes the form

$$M\boldsymbol{v} = \lambda\boldsymbol{v}, \quad \boldsymbol{v} \neq \mathbf{0}$$

So, in (1), $\boldsymbol{v} = \mathbf{1}$ and $\lambda = 1$

This works the same way for a column-stochastic matrix, except that here the relation is $\mathbf{1}M = \mathbf{1}$ with $\mathbf{1}$ a row vector and the (left)eigenpair relation is $\boldsymbol{v}^T M = \lambda \boldsymbol{v}^T$ with $\boldsymbol{v}^T$ a row vector

Long time behaviour

Let $p(0)$ be the initial distribution vector. Then

$$\begin{aligned} p(1) &= Pp(0) \\ p(2) &= Pp(1) \\ &= P\left(Pp(0)\right) \\ &= P^2 p(0) \end{aligned}$$

Continuing, we get, for any t,

$$p(t) = P^t p(0)$$

Therefore,

$$\lim_{t \to +\infty} p(t) = \lim_{t \to +\infty} P^t p(0) = \left(\lim_{t \to +\infty} P^t \right) p(0)$$

if this limit exists

The matrix P^t

Theorem 94

If M, N are nonsingular stochastic matrices, then MN is a stochastic matrix

Corollary 95

If M is a nonsingular stochastic matrix, then for any $k \in \mathbb{N}$, M^k is a stochastic matrix

So P^t is stochastic

Markov chains

Running example – Mendelian inheritance

Repetition of the process

Regular Markov chains

Regular Markov chains

Definition 96 (Regular Markov chain)

A **regular** Markov chain has P^k (entry-wise) positive for some integer $k > 0$, i.e., P^k has only positive entries

Definition 97 (Primitive matrix)

A nonnegative matrix M is **primitive** if, and only if, there is an integer $k > 0$ such that M^k is positive.

Theorem 98

Markov chain regular $\iff$ *transition matrix* P *primitive*

Matrices and graphs

Here and with absorbing chains, there is a lot to gain from using a bit of graph theory

Matrices and graphs are intimitely linked

Some matrix problems are easier considered with graphs, some graph problems are easier with matrices

Note that I say *graph*, but in other contexts, people speak of *networks*

What is a directed graph?

Definition 99 (Digraph)

A **directed graph** (or **digraph**) G is a pair (V, A) where:

- V is a finite set of elements called **vertices** or **nodes**
- $A \subseteq V \times V$ is a set of ordered pairs of vertices called **arcs** or **directed edges**

Definition 100 (Arc)

An **arc** $a = (u, v) \in A$ represents a connection **from** vertex u **to** vertex v

- u is the **tail** of the arc
- v is the **head** of the arc

In the context of Markov chains

- Vertices (nodes) represent the **states** of the system
- Arcs represent possible **transitions** between states
- The weights on the arcs represent the probability to make a given transition

Matrix $\leftrightarrow$ Graph

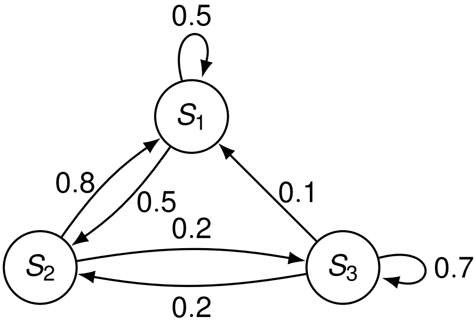
Given a transition matrix $P = [p_{ij}]$, define an induced digraph $\mathcal{G} = (V, A)$

- Vertices V correspond to the states
- An arc (j, i) exists in A if and only if $p_{ij} > 0$

Transition matrix

$$P = \begin{pmatrix} 0.5 & 0.8 & 0.1 \\ 0.5 & 0 & 0.2 \\ 0 & 0.2 & 0.7 \end{pmatrix}$$

Transition graph



Definition 101 (Reducible/irrecible matrix)

A matrix $M \in \mathcal{M}_n$ is **reducible** if there exists a permutation matrix P such that

$$P^T M P = \begin{pmatrix} P & Q \\ \mathbf{0} & R \end{pmatrix},$$

i.e., M is similar to a block upper triangular matrix. The matrix M is **irreducible** if no such matrix exists

Definition 102 (Strongly connected digraph)

A digraph $\mathcal{G} = (V, A)$ is **strongly connected** if for any pair of vertices $u, v \in V$, there is a directed path from u to v

Theorem 103

$P \in \mathcal{M}_n$ irreducible $\iff$ $\mathcal{G}(P)$ strongly connected

A sufficient condition for primitivity

Theorem 104

Let $M \in \mathcal{M}_n$ be a nonnegative matrix. If $\mathcal{G}(M)$ is strongly connected and at least one of the diagonal entries m_{ii} of M is positive, then M is primitive

Behaviour of a regular MC

Theorem 105

If P is the transition matrix of a regular Markov chain, then

1. *the powers P^t approach a stochastic matrix W*
2. *each column of W is the same (column) vector $w = (w_1, \ldots, w_n)^T$*
3. *the components of w are positive*

So if the Markov chain is regular

$$\lim_{t \to +\infty} p(t) = \lim_{t \to +\infty} P^t p(0) = Wp(0)$$

Computing W

Recall that since P is a stochastic matrix, 1 is an eigenvalue of P. As P is column stochastic, 1 is associated to the left (row) eigenvector $\mathbb{1}$

Now, if $\boldsymbol{p}(t)$ converges, then $\boldsymbol{p}(t+1) = P\boldsymbol{p}(t)$ at the limit, so $\boldsymbol{w} = \lim_{t\to\infty} \boldsymbol{p}(t)$ is a **fixed point** of the system. Replacing $\boldsymbol{p}$ with its limit, we have

$$\boldsymbol{w} = P\boldsymbol{w}$$

Solving for $\boldsymbol{w}$ thus amounts to finding $\boldsymbol{w}$ as a (right) eigenvector corresponding to the eigenvalue 1

Remember to normalise

$\boldsymbol{w}$ might have to be normalized since you want a probability vector

Check that the norm $\|\boldsymbol{w}\|_1$ defined by

$$\|\boldsymbol{w}\|_1 = |w_1| + \cdots + |w_n| = w_1 + \cdots + w_n$$

(since $\boldsymbol{w} \geq \boldsymbol{0}$) is equal to one

If not, use

$$\tilde{\boldsymbol{w}} = \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_1}$$

Back to orange cats

Create a chain by tracking the 3 female genotypes:

- S_1: $X^O X^O$ (orange)
- S_2: $X^o X^o$ (black)
- S_3: $X^O X^o$ (tortoiseshell)

To make the chain regular, we mate our female with a male chosen randomly from a **fixed population** that is:

- 50% orange males ($X^O Y$)
- 50% black males ($X^o Y$)

State 1: orange female ($X^O X^O$)

The mother is $X^O X^O$. We pick a father with 50/50 probability.

Case 1: Father is $X^O Y$

		X^O	Y
Mother	X^O	$X^O X^O$ (S_1)	Male
	X^O	$X^O X^O$ (S_1)	Male

Daughters: 100% S_1

Case 2: Father is $X^o Y$

		X^o	Y
Mother	X^O	$X^O X^o$ (S_3)	Male
	X^O	$X^O X^o$ (S_3)	Male

Daughters: 100% S_3

Transitions from S_1:

- $\mathbb{P}(S_1 \to S_1) = 0.5 \times 1.0 = \mathbf{0.5}$
- $\mathbb{P}(S_1 \to S_2) = 0$
- $\mathbb{P}(S_1 \to S_3) = 0.5 \times 1.0 = \mathbf{0.5}$

State 2: black female (X^oX^o)

The mother is X^oX^o. We pick a father with 50/50 probability.

Case 1: Father is X^OY

		X^O	Y
Mother	X^o	X^OX^o (S_3)	Male
	X^o	X^OX^o (S_3)	Male

Daughters: 100% S_3

Case 2: Father is X^oY

		X^o	Y
Mother	X^o	X^oX^o (S_2)	Male
	X^o	X^oX^o (S_2)	Male

Daughters: 100% S_2

Transitions from S_2:

- $\mathbb{P}(S_2 \to S_1) = 0$
- $\mathbb{P}(S_2 \to S_2) = 0.5 \times 1.0 = \mathbf{0.5}$
- $\mathbb{P}(S_2 \to S_3) = 0.5 \times 1.0 = \mathbf{0.5}$

State 3: tortoiseshell female ($X^O X^o$)

The mother is $X^O X^o$. We pick a father with 50/50 probability.

Case 1: Father is $X^O Y$

Mother	X^O	Y
X^O	$X^O X^O$ (S_1)	Male
X^o	$X^O X^o$ (S_3)	Male

Daughters: 50% S_1, 50% S_3

Case 2: Father is $X^o Y$

Mother	X^o	Y
X^O	$X^O X^o$ (S_3)	Male
X^o	$X^o X^o$ (S_2)	Male

Daughters: 50% S_2, 50% S_3

Transitions from S_3:

- $\mathbb{P}(S_3 \to S_1) = 0.5 \times 0.5 = \mathbf{0.25}$
- $\mathbb{P}(S_3 \to S_2) = 0.5 \times 0.5 = \mathbf{0.25}$
- $\mathbb{P}(S_3 \to S_3) = (0.5 \times 0.5) + (0.5 \times 0.5) = \mathbf{0.5}$

Summary of the orange cat problem

The transition matrix P for states $\{S_1, S_2, S_3\}$ is:

$$P = \begin{pmatrix} 0.5 & 0 & 0.25 \\ 0 & 0.5 & 0.25 \\ 0.5 & 0.5 & 0.5 \end{pmatrix}$$

Is this chain regular?

- **Irreducible? Yes.** All states communicate.
  - $S_1 \to S_3 \to S_2$ (Path from S_1 to S_2)
  - $S_2 \to S_3 \to S_1$ (Path from S_2 to S_1)
  - All other paths are direct ($S_1 \to S_3$, $S_3 \to S_1$, etc.)
- **Aperiodic? Yes.** All states have self-loops ($p_{11}, p_{22}, p_{33} > 0$).

Since the chain is irreducible and aperiodic, it is **regular**.

We must solve the right eigenvector problem $w = Pw$, where $w = (w_1, w_2, w_3)^T$

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 0.5 & 0 & 0.25 \\ 0 & 0.5 & 0.25 \\ 0.5 & 0.5 & 0.5 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

This gives the system of (dependent) equations:

$$\begin{aligned} w_1 &= 0.5w_1 + 0.25w_3 \\ w_2 &= 0.5w_2 + 0.25w_3 \\ w_3 &= 0.5w_1 + 0.5w_2 + 0.5w_3 \end{aligned}$$

From the first two equations:

- $0.5w_1 = 0.25w_3 \implies w_3 = 2w_1$
- $0.5w_2 = 0.25w_3 \implies w_3 = 2w_2$

This means $w_1 = w_2$. Using the constraint $w_1 + w_2 + w_3 = 1$:

$$w_1 + (w_1) + (2w_1) = 1 \implies 4w_1 = 1 \implies w_1 = 0.25$$

The stationary distribution is $w = \begin{pmatrix} 0.25 & 0.25 & 0.5 \end{pmatrix}^T$