Markov chains

Examples

Ergodicity and stationarity
Outline

Markov chains

Examples

Ergodicity and stationarity
Consider a sequence of random variables $X_0, X_1, X_2, \ldots$ each taking values in the same state space, which for now we take to be a finite set that we label by $\{0, 1, \ldots, M\}$.
Consider a sequence of random variables $X_0, X_1, X_2, \ldots$ each taking values in the same state space, which for now we take to be a finite set that we label by $\{0, 1, \ldots, M\}$.

Interpret $X_n$ as state of the system at time $n$. 

This kind of an “almost memoryless” property. Probability distribution for next state depends only on the current state (and not on the rest of the state history).
Consider a sequence of random variables $X_0, X_1, X_2, \ldots$ each taking values in the same state space, which for now we take to be a finite set that we label by $\{0, 1, \ldots, M\}$.

Interpret $X_n$ as state of the system at time $n$.

Sequence is called a **Markov chain** if we have a fixed collection of numbers $P_{ij}$ (one for each pair $i, j \in \{0, 1, \ldots, M\}$) such that whenever the system is in state $i$, there is probability $P_{ij}$ that system will next be in state $j$. 

Kind of an “almost memoryless” property. Probability distribution for next state depends only on the current state (and not on the rest of the state history).
Markov chains

Consider a sequence of random variables $X_0, X_1, X_2, \ldots$ each taking values in the same state space, which for now we take to be a finite set that we label by $\{0, 1, \ldots, M\}$.

Interpret $X_n$ as state of the system at time $n$.

Sequence is called a **Markov chain** if we have a fixed collection of numbers $P_{ij}$ (one for each pair $i, j \in \{0, 1, \ldots, M\}$) such that whenever the system is in state $i$, there is probability $P_{ij}$ that system will next be in state $j$.

Precisely, 

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0\} = P_{ij}.$$
Consider a sequence of random variables $X_0, X_1, X_2, \ldots$ each taking values in the same state space, which for now we take to be a finite set that we label by \{0, 1, \ldots, M\}.

Interpret $X_n$ as state of the system at time $n$.

Sequence is called a **Markov chain** if we have a fixed collection of numbers $P_{ij}$ (one for each pair $i, j \in \{0, 1, \ldots, M\}$) such that whenever the system is in state $i$, there is probability $P_{ij}$ that system will next be in state $j$.

Precisely, $$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0\} = P_{ij}.$$ Kind of an “almost memoryless” property. Probability distribution for next state depends only on the current state (and not on the rest of the state history).
For example, imagine a simple weather model with two states: rainy and sunny.
Simple example

- For example, imagine a simple weather model with two states: rainy and sunny.
- If it’s rainy one day, there’s a .5 chance it will be rainy the next day, a .5 chance it will be sunny.
For example, imagine a simple weather model with two states: rainy and sunny.

If it’s rainy one day, there’s a .5 chance it will be rainy the next day, a .5 chance it will be sunny.

If it’s sunny one day, there’s a .8 chance it will be sunny the next day, a .2 chance it will be rainy.
For example, imagine a simple weather model with two states: rainy and sunny.

If it’s rainy one day, there’s a .5 chance it will be rainy the next day, a .5 chance it will be sunny.

If it’s sunny one day, there’s a .8 chance it will be sunny the next day, a .2 chance it will be rainy.

In this climate, sun tends to last longer than rain.
For example, imagine a simple weather model with two states: rainy and sunny.

If it’s rainy one day, there’s a .5 chance it will be rainy the next day, a .5 chance it will be sunny.

If it’s sunny one day, there’s a .8 chance it will be sunny the next day, a .2 chance it will be rainy.

In this climate, sun tends to last longer than rain.

Given that it is rainy today, how many days do I expect to have to wait to see a sunny day?
Simple example

- For example, imagine a simple weather model with two states: rainy and sunny.
- If it’s rainy one day, there’s a .5 chance it will be rainy the next day, a .5 chance it will be sunny.
- If it’s sunny one day, there’s a .8 chance it will be sunny the next day, a .2 chance it will be rainy.
- In this climate, sun tends to last longer than rain.
- Given that it is rainy today, how many days do I expect to have to wait to see a sunny day?
- Given that it is sunny today, how many days do I expect to have to wait to see a rainy day?
For example, imagine a simple weather model with two states: rainy and sunny.

If it’s rainy one day, there’s a .5 chance it will be rainy the next day, a .5 chance it will be sunny.

If it’s sunny one day, there’s a .8 chance it will be sunny the next day, a .2 chance it will be rainy.

In this climate, sun tends to last longer than rain.

Given that it is rainy today, how many days do I expect to have to wait to see a sunny day?

Given that it is sunny today, how many days do I expect to have to wait to see a rainy day?

Over the long haul, what fraction of days are sunny?
Matrix representation

- To describe a Markov chain, we need to define $P_{ij}$ for any $i, j \in \{0, 1, \ldots, M\}$. 

\[
A = \begin{pmatrix}
P_{00} & P_{01} & \cdots & P_{0M} \\
P_{10} & P_{11} & \cdots & P_{1M} \\
\vdots & \vdots & \ddots & \vdots \\
P_{M0} & P_{M1} & \cdots & P_{MM}
\end{pmatrix}
\]

- For this to make sense, we require $P_{ij} \geq 0$ for all $i, j$ and $\sum_{j=0}^{M} P_{ij} = 1$ for each $i$. That is, the rows sum to one.
To describe a Markov chain, we need to define $P_{ij}$ for any $i, j \in \{0, 1, \ldots, M\}$.

It is convenient to represent the collection of transition probabilities $P_{ij}$ as a matrix:

$$A = \begin{pmatrix}
P_{00} & P_{01} & \cdots & P_{0M} \\
P_{10} & P_{11} & \cdots & P_{1M} \\
\vdots & \vdots & \ddots & \vdots \\
P_{M0} & P_{M1} & \cdots & P_{MM}
\end{pmatrix}$$
Matrix representation

To describe a Markov chain, we need to define $P_{ij}$ for any $i, j \in \{0, 1, \ldots, M\}$.

It is convenient to represent the collection of transition probabilities $P_{ij}$ as a matrix:

$$A = \begin{pmatrix}
    P_{00} & P_{01} & \ldots & P_{0M} \\
    P_{10} & P_{11} & \ldots & P_{1M} \\
    \vdots & \vdots & \ddots & \vdots \\
    P_{M0} & P_{M1} & \ldots & P_{MM}
\end{pmatrix}$$

For this to make sense, we require $P_{ij} \geq 0$ for all $i, j$ and $\sum_{j=0}^{M} P_{ij} = 1$ for each $i$. That is, the rows sum to one.
Suppose that $p_i$ is the probability that system is in state $i$ at time zero.

What does the following product represent?

$$
\begin{pmatrix}
p_0 & p_1 & \ldots & p_M \\
p_0 & p_1 & \ldots & p_M \\
\vdots & \vdots & \ddots & \vdots \\
p_0 & p_1 & \ldots & p_M
\end{pmatrix}
$$

Answer: the probability distribution at time one.

How about the following product?

$$
(p_0 p_1 \ldots p_M) A^n
$$

Answer: the probability distribution at time $n$. 

18.600 Lecture 33
Suppose that $p_i$ is the probability that system is in state $i$ at time zero.

What does the following product represent?

\[
\begin{pmatrix}
p_0 & p_1 & \ldots & p_M \\
\end{pmatrix}
\begin{pmatrix}
P_{00} & P_{01} & \ldots & P_{0M} \\
P_{10} & P_{11} & \ldots & P_{1M} \\
\cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot \\
P_{M0} & P_{M1} & \ldots & P_{MM} \\
\end{pmatrix}
\]

Answer: the probability distribution at time one.

How about the following product?

\[
(p_0 p_1 \ldots p_M) A^n
\]

Answer: the probability distribution at time $n$. 
Transitions via matrices

- Suppose that $p_i$ is the probability that system is in state $i$ at time zero.
- What does the following product represent?

\[
\begin{pmatrix}
P_{00} & P_{01} & \ldots & P_{0M} \\
P_{10} & P_{11} & \ldots & P_{1M} \\
\vdots & \vdots & \ddots & \vdots \\
P_{M0} & P_{M1} & \ldots & P_{MM}
\end{pmatrix}
\begin{pmatrix}
p_0 \\
p_1 \\
\vdots \\
p_M
\end{pmatrix}
\]

- Answer: the probability distribution at time one.

- How about the following product?

\[
(p_0 \ p_1 \ \ldots \ p_M)^T A^n
\]

- Answer: the probability distribution at time $n$. 
Transitions via matrices

▶ Suppose that $p_i$ is the probability that system is in state $i$ at time zero.

▶ What does the following product represent?

\[
\left( \begin{array}{cccc}
  p_0 & p_1 & \cdots & p_M \\
  P_{00} & P_{01} & \cdots & P_{0M} \\
  P_{10} & P_{11} & \cdots & P_{1M} \\
  \vdots & \vdots & \ddots & \vdots \\
  P_{M0} & P_{M1} & \cdots & P_{MM}
\end{array} \right)
\]

▶ Answer: the probability distribution at time one.

▶ How about the following product?

\[
\left( \begin{array}{cccc}
  p_0 & p_1 & \cdots & p_M \\
  P_{00} & P_{01} & \cdots & P_{0M} \\
  P_{10} & P_{11} & \cdots & P_{1M} \\
  \vdots & \vdots & \ddots & \vdots \\
  P_{M0} & P_{M1} & \cdots & P_{MM}
\end{array} \right) A^n
\]
Transitions via matrices

Suppose that $p_i$ is the probability that system is in state $i$ at time zero.

What does the following product represent?

$$( p_0 \ p_1 \ \ldots \ p_M ) \begin{pmatrix} P_{00} & P_{01} & \ldots & P_{0M} \\ P_{10} & P_{11} & \ldots & P_{1M} \\ \vdots & \vdots & \ddots & \vdots \\ P_{M0} & P_{M1} & \ldots & P_{MM} \end{pmatrix}$$

Answer: the probability distribution at time one.

How about the following product?

$$( p_0 \ p_1 \ \ldots \ p_M ) A^n$$

Answer: the probability distribution at time $n$. 
Powers of transition matrix

- We write $P_{ij}^{(n)}$ for the probability to go from state $i$ to state $j$ over $n$ steps.
Powers of transition matrix

- We write $P_{ij}^{(n)}$ for the probability to go from state $i$ to state $j$ over $n$ steps.

- From the matrix point of view

\[
\begin{pmatrix}
P_{00}^{(n)} & P_{01}^{(n)} & \cdots & P_{0M}^{(n)} \\
P_{10}^{(n)} & P_{11}^{(n)} & \cdots & P_{1M}^{(n)} \\
\vdots & \vdots & \ddots & \vdots \\
P_{M0}^{(n)} & P_{M1}^{(n)} & \cdots & P_{MM}^{(n)}
\end{pmatrix}
= 
\begin{pmatrix}
P_{00} & P_{01} & \cdots & P_{0M} \\
P_{10} & P_{11} & \cdots & P_{1M} \\
\vdots & \vdots & \ddots & \vdots \\
P_{M0} & P_{M1} & \cdots & P_{MM}
\end{pmatrix}^n
\]
We write $P_{ij}^{(n)}$ for the probability to go from state $i$ to state $j$ over $n$ steps.

From the matrix point of view

\[
\begin{pmatrix}
P_{00}^{(n)} & P_{01}^{(n)} & \cdots & P_{0M}^{(n)} \\
P_{10}^{(n)} & P_{11}^{(n)} & \cdots & P_{1M}^{(n)} \\
\vdots & \vdots & \ddots & \vdots \\
P_{M0}^{(n)} & P_{M1}^{(n)} & \cdots & P_{MM}^{(n)}
\end{pmatrix}
= 
\left(\begin{array}{cccc}
P_{00} & P_{01} & \cdots & P_{0M} \\
P_{10} & P_{11} & \cdots & P_{1M} \\
\vdots & \vdots & \ddots & \vdots \\
P_{M0} & P_{M1} & \cdots & P_{MM}
\end{array}\right)^n
\]

If $A$ is the one-step transition matrix, then $A^n$ is the $n$-step transition matrix.
What does it mean if all of the rows are identical?
Questions

What does it mean if all of the rows are identical?

Answer: state sequence $X_i$ consists of i.i.d. random variables.
Questions

▶ What does it mean if all of the rows are identical?
▶ Answer: state sequence $X_i$ consists of i.i.d. random variables.
▶ What if matrix is the identity?
Questions

- What does it mean if all of the rows are identical?
  - Answer: state sequence $X_i$ consists of i.i.d. random variables.

- What if matrix is the identity?
  - Answer: states never change.
Questions

- What does it mean if all of the rows are identical?
  - Answer: state sequence $X_i$ consists of i.i.d. random variables.
- What if matrix is the identity?
  - Answer: states never change.
- What if each $P_{ij}$ is either one or zero?
Questions

- What does it mean if all of the rows are identical?
  - Answer: state sequence $X_i$ consists of i.i.d. random variables.

- What if matrix is the identity?
  - Answer: states never change.

- What if each $P_{ij}$ is either one or zero?
  - Answer: state evolution is deterministic.
Outline

Markov chains

Examples

Ergodicity and stationarity
Outline

Markov chains

Examples

Ergodicity and stationarity
Consider the simple weather example: If it’s rainy one day, there’s a .5 chance it will be rainy the next day, a .5 chance it will be sunny. If it’s sunny one day, there’s a .8 chance it will be sunny the next day, a .2 chance it will be rainy.
Consider the simple weather example: If it’s rainy one day, there’s a .5 chance it will be rainy the next day, a .5 chance it will be sunny. If it’s sunny one day, there’s a .8 chance it will be sunny the next day, a .2 chance it will be rainy.

Let rainy be state zero, sunny state one, and write the transition matrix by

\[
A = \begin{pmatrix} .5 & .5 \\ .2 & .8 \end{pmatrix}
\]
Consider the simple weather example: If it’s rainy one day, there’s a .5 chance it will be rainy the next day, a .5 chance it will be sunny. If it’s sunny one day, there’s a .8 chance it will be sunny the next day, a .2 chance it will be rainy.

Let rainy be state zero, sunny state one, and write the transition matrix by

\[
A = \begin{pmatrix}
.5 & .5 \\
.2 & .8 \\
\end{pmatrix}
\]

Note that

\[
A^2 = \begin{pmatrix}
.64 & .35 \\
.26 & .74 \\
\end{pmatrix}
\]
Consider the simple weather example: If it’s rainy one day, there’s a .5 chance it will be rainy the next day, a .5 chance it will be sunny. If it’s sunny one day, there’s a .8 chance it will be sunny the next day, a .2 chance it will be rainy.

Let rainy be state zero, sunny state one, and write the transition matrix by

\[
A = \begin{pmatrix}
0.5 & 0.5 \\
0.2 & 0.8
\end{pmatrix}
\]

Note that

\[
A^2 = \begin{pmatrix}
0.64 & 0.35 \\
0.26 & 0.74
\end{pmatrix}
\]

Can compute \( A^{10} \) =

\[
\begin{pmatrix}
0.285719 & 0.714281 \\
0.285713 & 0.714287
\end{pmatrix}
\]
Does relationship status have the Markov property?

In a relationship

Single

Married

It’s complicated

Engaged

Can we assign a probability to each arrow?

Markov model implies time spent in any state (e.g., a marriage) before leaving is a geometric random variable.

Not true... Can we make a better model with more states?
Does relationship status have the Markov property?

- Single
- In a relationship
- It’s complicated
- Engaged
- Married

Can we assign a probability to each arrow?
Does relationship status have the Markov property?

- Single
- In a relationship
- It's complicated
- Engaged
- Married

- Can we assign a probability to each arrow?
- Markov model implies time spent in any state (e.g., a marriage) before leaving is a geometric random variable.
Does relationship status have the Markov property?

- Can we assign a probability to each arrow?
- Markov model implies time spent in any state (e.g., a marriage) before leaving is a geometric random variable.
- Not true... Can we make a better model with more states?
Outline

Markov chains

Examples

Ergodicity and stationarity
Outline

Markov chains

Examples

Ergodicity and stationarity
Say Markov chain is **ergodic** if some power of the transition matrix has all non-zero entries.
Say Markov chain is **ergodic** if some power of the transition matrix has all non-zero entries.

Turns out that if chain has this property, then
\[ \pi_j := \lim_{n \to \infty} P_{ij}^{(n)} \]
exists and the \( \pi_j \) are the unique non-negative solutions of \( \pi_j = \sum_{k=0}^{M} \pi_k P_{kj} \) that sum to one.

This means that the row vector \( \pi = (\pi_0 \pi_1 \ldots \pi_M) \) is a left eigenvector of \( A \) with eigenvalue 1, i.e., \( \pi A = \pi \).

We call \( \pi \) the stationary distribution of the Markov chain.

One can solve the system of linear equations \( \pi_j = \sum_{k=0}^{M} \pi_k P_{kj} \) to compute the values \( \pi_j \). Equivalent to considering \( A \) fixed and solving \( \pi A = \pi \). Or solving \( (A - I) \pi = 0 \). This determines \( \pi \) up to a multiplicative constant, and fact that \( \sum \pi_j = 1 \) determines the constant.
Ergodic Markov chains

- Say Markov chain is **ergodic** if some power of the transition matrix has all non-zero entries.
- Turns out that if chain has this property, then \( \pi_j := \lim_{n \to \infty} P_{ij}^{(n)} \) exists and the \( \pi_j \) are the unique non-negative solutions of \( \pi_j = \sum_{k=0}^{M} \pi_k P_{kj} \) that sum to one.
- This means that the row vector

\[
\pi = \left( \begin{array}{c} \pi_0 \\ \pi_1 \\ \vdots \\ \pi_M \end{array} \right)
\]

is a left eigenvector of \( A \) with eigenvalue 1, i.e., \( \pi A = \pi \).
Say Markov chain is **ergodic** if some power of the transition matrix has all non-zero entries.

Turns out that if chain has this property, then $\pi_j := \lim_{n \to \infty} P_{ij}^{(n)}$ exists and the $\pi_j$ are the unique non-negative solutions of $\pi_j = \sum_{k=0}^{M} \pi_k P_{kj}$ that sum to one. This means that the row vector

$$\pi = \begin{pmatrix} \pi_0 & \pi_1 & \ldots & \pi_M \end{pmatrix}$$

is a left eigenvector of $A$ with eigenvalue 1, i.e., $\pi A = \pi$.

We call $\pi$ the **stationary distribution** of the Markov chain.
Ergodic Markov chains

- Say Markov chain is **ergodic** if some power of the transition matrix has all non-zero entries.

- Turns out that if chain has this property, then
  \[ \pi_j := \lim_{n \to \infty} P_{ij}^{(n)} \] exists and the \( \pi_j \) are the unique non-negative solutions of
  \[ \pi_j = \sum_{k=0}^{M} \pi_k P_{kj} \] that sum to one.

- This means that the row vector
  \[ \pi = \left( \begin{array}{c} \pi_0 \\ \pi_1 \\ \vdots \\ \pi_M \end{array} \right) \]
  is a left eigenvector of \( A \) with eigenvalue 1, i.e., \( \pi A = \pi \).

- We call \( \pi \) the **stationary distribution** of the Markov chain.

- One can solve the system of linear equations
  \[ \pi_j = \sum_{k=0}^{M} \pi_k P_{kj} \] to compute the values \( \pi_j \). Equivalent to considering \( A \) fixed and solving \( \pi A = \pi \). Or solving
  \[ (A - I)\pi = 0. \] This determines \( \pi \) up to a multiplicative constant, and fact that \( \sum \pi_j = 1 \) determines the constant.
If \( A = \begin{pmatrix} .5 & .5 \\ .2 & .8 \end{pmatrix} \), then we know

\[
\pi A = \begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix} \begin{pmatrix} .5 & .5 \\ .2 & .8 \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix} = \pi.
\]
Simple example

If $A = \begin{pmatrix} .5 & .5 \\ .2 & .8 \end{pmatrix}$, then we know

$$\pi A = \begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix} \begin{pmatrix} .5 & .5 \\ .2 & .8 \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix} = \pi.$$ 

This means that $.5\pi_0 + .2\pi_1 = \pi_0$ and $.5\pi_0 + .8\pi_1 = \pi_1$ and we also know that $\pi_0 + \pi_1 = 1$. Solving these equations gives $\pi_0 = 2/7$ and $\pi_1 = 5/7$, so $\pi = \begin{pmatrix} 2/7 & 5/7 \end{pmatrix}$. 

Indeed, $\pi A = \begin{pmatrix} 2/7 & 5/7 \end{pmatrix} \begin{pmatrix} .5 & .5 \\ .2 & .8 \end{pmatrix} = \begin{pmatrix} 2/7 & 5/7 \end{pmatrix} = \pi.$
If $A = \begin{pmatrix} .5 & .5 \\ .2 & .8 \end{pmatrix}$, then we know

$$\pi A = \begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix} \begin{pmatrix} .5 & .5 \\ .2 & .8 \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix} = \pi.$$ 

This means that $.5\pi_0 + .2\pi_1 = \pi_0$ and $.5\pi_0 + .8\pi_1 = \pi_1$ and we also know that $\pi_0 + \pi_1 = 1$. Solving these equations gives $\pi_0 = 2/7$ and $\pi_1 = 5/7$, so $\pi = \begin{pmatrix} 2/7 & 5/7 \end{pmatrix}$.

Indeed,

$$\pi A = \begin{pmatrix} 2/7 & 5/7 \end{pmatrix} \begin{pmatrix} .5 & .5 \\ .2 & .8 \end{pmatrix} = \begin{pmatrix} 2/7 & 5/7 \end{pmatrix} = \pi.$$
Simple example

- If \( A = \begin{pmatrix} .5 & .5 \\ .2 & .8 \end{pmatrix} \), then we know

\[
\pi A = \begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix} \begin{pmatrix} .5 & .5 \\ .2 & .8 \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix} = \pi.
\]

- This means that \(.5\pi_0 + .2\pi_1 = \pi_0\) and \(.5\pi_0 + .8\pi_1 = \pi_1\) and we also know that \(\pi_0 + \pi_1 = 1\). Solving these equations gives \(\pi_0 = 2/7\) and \(\pi_1 = 5/7\), so \(\pi = \begin{pmatrix} 2/7 & 5/7 \end{pmatrix}\).

- Indeed,

\[
\pi A = \begin{pmatrix} 2/7 & 5/7 \end{pmatrix} \begin{pmatrix} .5 & .5 \\ .2 & .8 \end{pmatrix} = \begin{pmatrix} 2/7 & 5/7 \end{pmatrix} = \pi.
\]

- Recall that

\[
A^{10} = \begin{pmatrix} .285719 & .714281 \\ .285713 & .714287 \end{pmatrix} \approx \begin{pmatrix} 2/7 & 5/7 \\ 2/7 & 5/7 \end{pmatrix} = \begin{pmatrix} \pi \\ \pi \end{pmatrix}
\]