18.175: Lecture 30

Markov chains

Scott Sheffield

MIT

Outline

Review what you know about finite state Markov chains

Finite state ergodicity and stationarity

More general setup

Outline

Review what you know about finite state Markov chains

Finite state ergodicity and stationarity

More general setup

Markov chains

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

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 to I expect to have to wait to see a sunny day?
- Given that it is sunny today, how many days to 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 $i, j \in \{0, 1, \ldots, M\}$.
- It is convenient to represent the collection of transition probabilities P_{ij} as a matrix:

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

- 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.

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?

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

- Answer: the probability distribution at time one.
- How about the following product?

$$\begin{pmatrix} p_0 & p_1 & \ldots & p_M \end{pmatrix} A^n$$

- 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.
- From the matrix point of view

$$\begin{pmatrix} P_{00}^{(n)} & P_{01}^{(n)} & \dots & P_{0M}^{(n)} \\ P_{10}^{(n)} & P_{11}^{(n)} & \dots & P_{1M}^{(n)} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ P_{M0}^{(n)} & P_{M1}^{(n)} & \dots & P_{MM}^{(n)} \end{pmatrix} = \begin{pmatrix} P_{00} & P_{01} & \dots & P_{0M} \\ P_{10} & P_{11} & \dots & P_{1M} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ P_{M0} & P_{M1} & \dots & P_{MM} \end{pmatrix}^n$$

- If A is the one-step transition matrix, then A^n is the n-step transition matrix.

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.

Simple example

- 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

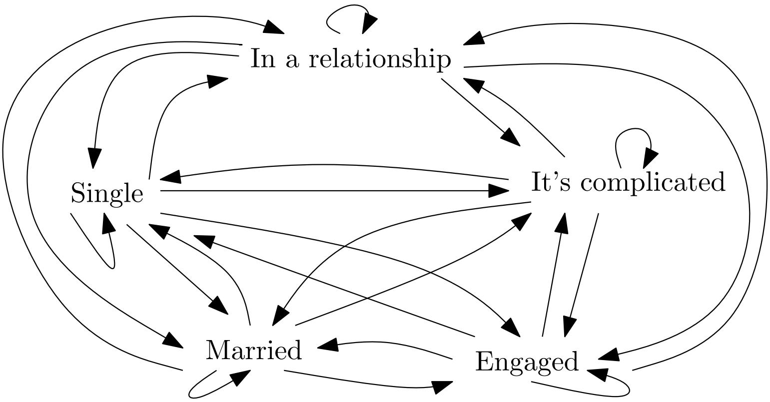
$$A = \begin{pmatrix} .5 & .5 \\ .2 & .8 \end{pmatrix}$$

- Note that

$$A^2 = \begin{pmatrix} .64 & .35 \\ .26 & .74 \end{pmatrix}$$

- Can compute $A^{10} = \begin{pmatrix} .285719 & .714281 \\ .285713 & .714287 \end{pmatrix}$

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

Review what you know about finite state Markov chains

Finite state ergodicity and stationarity

More general setup

Outline

Review what you know about finite state Markov chains

Finite state ergodicity and stationarity

More general setup

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 π_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

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

  is a left eigenvector of A with eigenvalue 1, i.e., $\pi A = \pi$.
- We call π 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 π_j. Equivalent to considering A fixed and solving $\pi A = \pi$. Or solving $(A - I)\pi = 0$. This determines π up to a multiplicative constant, and fact that $\sum \pi_j = 1$ determines the constant.

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 $\pi_1 + \pi_2 = 1$. 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
$A^{10} = \begin{pmatrix} .285719 & .714281 \\ .285713 & .714287 \end{pmatrix} \approx \begin{pmatrix} 2/7 & 5/7 \\ 2/7 & 5/7 \end{pmatrix} = \begin{pmatrix} \pi \\ \pi \end{pmatrix}$

Outline

Review what you know about finite state Markov chains

Finite state ergodicity and stationarity

More general setup

Outline

Review what you know about finite state Markov chains

Finite state ergodicity and stationarity

More general setup

Markov chains: general definition

- Consider a measurable space $(S, \mathcal{S})$.
- A function $p : S \times \mathcal{S} \to \mathbb{R}$ is a **transition probability** if
  - For each $x \in S$, $A \to p(x, A)$ is a probability measure on $S, \mathcal{S})$.
  - For each $A \in S$, the map $x \to p(x, A)$ is a measurable function.
- Say that X_n is a **Markov chain** w.r.t. $\mathcal{F}_n$ with transition probability p if $P(X_{n+1} \in B | \mathcal{F}_n) = p(X_n, B)$.
- How do we construct an infinite Markov chain? Choose p and initial distribution μ on $(S, \mathcal{S})$. For each $n < \infty$ write

$$P(X_j \in B_j, 0 \leq j \leq n) = \int_{B_0} \mu(dx_0) \int_{B_1} p(x_0, dx_1) \cdots$$

$$\int_{B_n} p(x_{n-1}, dx_n).$$

Extend to $n = \infty$ by Kolmogorov's extension theorem.

Markov chains

- **Definition, again:** Say X_n is a **Markov chain** w.r.t. $\mathcal{F}_n$ with transition probability p if $P(X_{n+1} \in B | \mathcal{F}_n) = p(X_n, B)$.
- **Construction, again:** Fix initial distribution μ on $(S, \mathcal{S})$. For each $n < \infty$ write

$$P(X_j \in B_j, 0 \leq j \leq n) = \int_{B_0} \mu(dx_0) \int_{B_1} p(x_0, dx_1) \cdots$$

$$\int_{B_n} p(x_{n-1}, dx_n).$$

  Extend to $n = \infty$ by Kolmogorov's extension theorem.
- **Notation:** Extension produces probability measure P_μ on sequence space $(S^{0,1,\ldots}, \mathcal{S}^{0,1,\ldots})$.
- **Theorem:** $(X_0, X_1, \ldots)$ chosen from P_μ is Markov chain.
- **Theorem:** If X_n is any Markov chain with initial distribution μ and transition p, then finite dim. probabilities are as above.

Examples

- Random walks on $\mathbb{R}^d$.
- Branching processes: $p(i,j) = P\left(\sum_{m=1}^{i} \xi_m = j\right)$ where ξ_i are i.i.d. non-negative integer-valued random variables.
- Renewal chain.
- Card shuffling.
- Ehrenfest chain.

18.175 Theory of Probability
Spring 2014