18.175: Lecture 31

More Markov chains

Scott Sheffield

MIT

Outline

Recollections

General setup and basic properties

Recurrence and transience

Outline

Recollections

General setup and basic properties

Recurrence and transience

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 $\{0, 1, \ldots, M\}$.

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.

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.

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,
  $P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} = P_{ij}$.

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).

Matrix representation

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

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:

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

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.

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

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

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.

Ergodic Markov chains

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

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.

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., $\pi A = \pi$.

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.

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.

Examples

- Random walks on $\mathbb{R}^d$.

Examples

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

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 (deterministic unit decreases, random jump when zero hit).

Examples

- Random walks on $\mathbb{R}^d$.
- Branching processes: $p(i,j) = P\big(\sum_{m=1}^{i} \xi_m = j\big)$ where ξ_i are i.i.d. non-negative integer-valued random variables.
- Renewal chain (deterministic unit decreases, random jump when zero hit).
- Card shuffling.

Examples

- Random walks on $\mathbb{R}^d$.
- Branching processes: $p(i,j) = P\big(\sum_{m=1}^{i} \xi_m = j\big)$ where ξ_i are i.i.d. non-negative integer-valued random variables.
- Renewal chain (deterministic unit decreases, random jump when zero hit).
- Card shuffling.
- Ehrenfest chain (n balls in two chambers, randomly pick ball to swap).

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 (deterministic unit decreases, random jump when zero hit).
- Card shuffling.
- Ehrenfest chain (n balls in two chambers, randomly pick ball to swap).
- Birth and death chains (changes by ± 1). Stationarity distribution?

Examples

- Random walks on $\mathbb{R}^d$.
- Branching processes: $p(i,j) = P\big(\sum_{m=1}^{i} \xi_m = j\big)$ where ξ_i are i.i.d. non-negative integer-valued random variables.
- Renewal chain (deterministic unit decreases, random jump when zero hit).
- Card shuffling.
- Ehrenfest chain (n balls in two chambers, randomly pick ball to swap).
- Birth and death chains (changes by ± 1). Stationarity distribution?
- M/G/1 queues.

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 (deterministic unit decreases, random jump when zero hit).
- Card shuffling.
- Ehrenfest chain (n balls in two chambers, randomly pick ball to swap).
- Birth and death chains (changes by ± 1). Stationarity distribution?
- M/G/1 queues.
- Random walk on a graph. Stationary distribution?

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 (deterministic unit decreases, random jump when zero hit).
- Card shuffling.
- Ehrenfest chain (n balls in two chambers, randomly pick ball to swap).
- Birth and death chains (changes by ± 1). Stationarity distribution?
- M/G/1 queues.
- Random walk on a graph. Stationary distribution?
- Random walk on directed graph (e.g., single directed chain).

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 (deterministic unit decreases, random jump when zero hit).
- Card shuffling.
- Ehrenfest chain (n balls in two chambers, randomly pick ball to swap).
- Birth and death chains (changes by ± 1). Stationarity distribution?
- M/G/1 queues.
- Random walk on a graph. Stationary distribution?
- Random walk on directed graph (e.g., single directed chain).
- Snakes and ladders.

Outline

Recollections

General setup and basic properties

Recurrence and transience

Outline

Recollections

General setup and basic properties

Recurrence and transience

Markov chains: general definition

- Consider a measurable space $(S, \mathcal{S})$.

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

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 $S, \mathcal{S})$.

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.

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 $P(X_{n+1} \in B | \mathcal{F}_n) = p(X_n, B)$.

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 $P(X_{n+1} \in B | \mathcal{F}_n) = p(X_n, B)$.

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

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

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

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 $(S^{0,1,\ldots}, \mathcal{S}^{0,1,\ldots})$.

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.

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.

Markov properties

- **Markov property:** Take $(\Omega_0, \mathcal{F}) = \left(S^{\{0,1,\ldots\}}, \mathcal{S}^{\{0,1,\ldots\}}\right)$, and let P_μ be Markov chain measure and θ_n the shift operator on Ω_0 (shifts sequence n units to left, discarding elements shifted off the edge). If $Y : \Omega_0 \to \mathbb{R}$ is bounded and measurable then

$$E_\mu(Y \circ \theta_n | \mathcal{F}_n) = E_{X_n} Y.$$

Markov properties

- **Markov property:** Take $(\Omega_0, \mathcal{F}) = \left(S^{\{0,1,\ldots\}}, \mathcal{S}^{\{0,1,\ldots\}}\right)$, and let P_μ be Markov chain measure and θ_n the shift operator on Ω_0 (shifts sequence n units to left, discarding elements shifted off the edge). If $Y : \Omega_0 \to \mathbb{R}$ is bounded and measurable then

$$E_\mu(Y \circ \theta_n | \mathcal{F}_n) = E_{X_n} Y.$$

- **Strong Markov property:** Can replace n with a.s. finite stopping time N and function Y can vary with time. Suppose that for each n, $Y_n : \Omega_n \to \mathbb{R}$ is measurable and $|Y_n| \leq M$ for all n. Then

$$E_\mu(Y_N \circ \theta_N | \mathcal{F}_N) = E_{X_N} Y_N,$$

where RHS means $E_x Y_n$ evaluated at $x = X_n, n = N$.

Properties

- **Property of infinite opportunities:** Suppose X_n is Markov chain and

$$P(\cup_{m=n+1}^{\infty}\{X_m \in B_m\}|X_n) \geq \delta > 0$$

on $\{X_n \in A_n\}$. Then $P(\{X_n \in A_n \, i.o.\} - \{X_n \in B_n \, i.o.\}) = 0$.

Properties

- **Property of infinite opportunities:** Suppose X_n is Markov chain and

$$P(\cup_{m=n+1}^{\infty}\{X_m \in B_m\}|X_n) \geq \delta > 0$$

on $\{X_n \in A_n\}$. Then $P(\{X_n \in A_n \, i.o.\} - \{X_n \in B_n \, i.o.\}) = 0$.

- **Reflection principle:** Symmetric random walks on $\mathbb{R}$. Have $P(\sup_{m\geq n} S_m > a) \leq 2P(S_n > a)$.

Properties

- **Property of infinite opportunities:** Suppose X_n is Markov chain and

$$P(\cup_{m=n+1}^{\infty}\{X_m \in B_m\}|X_n) \geq \delta > 0$$

on $\{X_n \in A_n\}$. Then $P(\{X_n \in A_n \, i.o.\} - \{X_n \in B_n \, i.o.\}) = 0$.

- **Reflection principle:** Symmetric random walks on $\mathbb{R}$. Have $P(\sup_{m\geq n} S_m > a) \leq 2P(S_n > a)$.
- **Proof idea:** Reflection picture.

Outline

Recollections

General setup and basic properties

Recurrence and transience

Outline

Recollections

General setup and basic properties

Recurrence and transience

Query

- **Interesting question:** If A is an infinite probability transition matrix on a countable state space, what does the (infinite) matrix $I + A + A^2 + A^3 + \ldots = (I - A)^{-1}$ represent (if the sum converges)?

Query

- **Interesting question:** If A is an infinite probability transition matrix on a countable state space, what does the (infinite) matrix $I + A + A^2 + A^3 + \ldots = (I - A)^{-1}$ represent (if the sum converges)?
- **Question:** Does it describe the expected number of y hits when starting at x? Is there a similar interpretation for other power series?

Query

- **Interesting question:** If A is an infinite probability transition matrix on a countable state space, what does the (infinite) matrix $I + A + A^2 + A^3 + \ldots = (I - A)^{-1}$ represent (if the sum converges)?
- **Question:** Does it describe the expected number of y hits when starting at x? Is there a similar interpretation for other power series?
- How about e^A or $e^{\lambda A}$?

Query

- **Interesting question:** If A is an infinite probability transition matrix on a countable state space, what does the (infinite) matrix $I + A + A^2 + A^3 + \ldots = (I - A)^{-1}$ represent (if the sum converges)?
- **Question:** Does it describe the expected number of y hits when starting at x? Is there a similar interpretation for other power series?
- How about e^A or $e^{\lambda A}$?
- Related to distribution after a Poisson random number of steps?

Recurrence

- Consider probability walk from y ever returns to y.

Recurrence

- Consider probability walk from y ever returns to y.
- If it's 1, return to y infinitely often, else don't. Call y a **recurrent state** if we return to y infinitely often.