STATS 217: Introduction to Stochastic Processes I

Lecture 21

today + next time: mixing times

next week: CTMC

last week: CTMC + martingales

advertisement: next quarter
Stats 318
Modern Markov Chains

(Diaconis)

Stationary times

- Last time, we used the following description of the lazy random walk on the hypercube: suppose the current state is x and the current time is t. We choose a coordinate $i_{t+1} \in \{1, \ldots, n\}$ uniformly at random and an unbiased bit $b_{t+1} \in \{0, 1\}$, also uniformly at random, and independently of the coordinate i_{t+1}. Then, we set the value of coordinate i_{t+1} to b_{t+1} and keep all other coordinates unchanged.

Stationary times

- Last time, we used the following description of the lazy random walk on the hypercube: suppose the current state is x and the current time is t. We choose a coordinate $i_{t+1} \in \{1, \ldots, n\}$ uniformly at random and an unbiased bit $b_{t+1} \in \{0, 1\}$, also uniformly at random, and independently of the coordinate i_{t+1}. Then, we set the value of coordinate i_{t+1} to b_{t+1} and keep all other coordinates unchanged.
- Let τ_{refresh} be the first time that all coordinates have been chosen to be updated.

Stationary times

- Last time, we used the following description of the lazy random walk on the hypercube: suppose the current state is x and the current time is t. We choose a coordinate $i_{t+1} \in \{1, \ldots, n\}$ uniformly at random and an unbiased bit $b_{t+1} \in \{0, 1\}$, also uniformly at random, and independently of the coordinate i_{t+1}. Then, we set the value of coordinate i_{t+1} to b_{t+1} and keep all other coordinates unchanged.
- Let τ_{refresh} be the first time that all coordinates have been chosen to be updated.
- Note that τ_{refresh} is a stopping time with respect to the collection of random tuples $\{(i_k, b_k)\}_{k \geq 1}$. (Why?)

Recall this means that the event

$\tau_{\text{refresh}} = n$ is determined by $i_1, b_1, \ldots, i_n, b_n$.

this is not a

stopping time wrt X_t

Stationary times

$n=8$ 0 0 0

$i_1 = 1$ $b_1 = 0$

$i_2 = 2$ $b_2 = 0$

$i_3 = 3$ $b_3 = 0$

- Last time, we used the following description of the lazy random walk on the hypercube: suppose the current state is x and the current time is t. We choose a coordinate $i_{t+1} \in \{1, \ldots, n\}$ uniformly at random and an unbiased bit $b_{t+1} \in \{0, 1\}$, also uniformly at random, and independently of the coordinate i_{t+1}. Then, we set the value of coordinate i_{t+1} to b_{t+1} and keep all other coordinates unchanged.
- Let τ_{refresh} be the first time that all coordinates have been chosen to be updated.
- Note that τ_{refresh} is a stopping time with respect to the collection of random tuples $\{(i_k, b_k)\}_{k \geq 1}$. (Why?)
- Moreover, note that $X_{\tau_{\text{refresh}}}$ is distributed uniformly on $\{0, 1\}^n$. (Why?)

τ_{refresh} is random

$X_{\tau_{\text{refresh}}}$

* τ_{refresh} depends only on $i_1, i_2, i_3 \ldots$

$(X_{\tau_{\text{refresh}}})_j = \{ b_\ell$ where ℓ is max i

Strong stationary times

before $\tau_{refresh}$ s.t. coord chosen at $i = j$ }

- τ_{refresh} is an example of a **stationary time**.
- By this, we mean a stopping time τ (with respect to the initial state x of the chain as well as any auxiliary randomness) which satisfies

$$\mathbb{P}[X_\tau = y \mid X_0 = x] = \pi(y) \quad \forall y \in S,$$

where π is the stationary distribution of the chain.

Strong stationary times

- τ_{refresh} is an example of a **stationary time**.
- By this, we mean a stopping time τ (with respect to the initial state x of the chain as well as any auxiliary randomness) which satisfies

$$\mathbb{P}[X_\tau = y \mid X_0 = x] = \pi(y) \quad \forall y \in S,$$

  where π is the stationary distribution of the chain.
- Let $x \in S$ and consider the chain with initial state $X_0 = x$. A stationary time τ for which τ and X_τ are independent i.e. for all integers t and $y \in S$

$$\mathbb{P}[\tau = t, X_\tau = y \mid X_0 = x] = \mathbb{P}[\tau = t \mid X_0 = x] \cdot \pi(y)$$

  is called a **strong stationary time** for the starting state x.

Strong stationary times

- τ_{refresh} is an example of a **stationary time**.
- By this, we mean a stopping time τ (with respect to the initial state x of the chain as well as any auxiliary randomness) which satisfies

$$\mathbb{P}[X_\tau = y \mid X_0 = x] = \pi(y) \quad \forall y \in S,$$

  where π is the stationary distribution of the chain.
- Let $x \in S$ and consider the chain with initial state $X_0 = x$. A stationary time τ for which τ and X_τ are independent i.e. for all integers t and $y \in S$

$$\mathbb{P}[\tau = t, X_\tau = y \mid X_0 = x] = \mathbb{P}[\tau = t \mid X_0 = x] \cdot \pi(y)$$

  is called a **strong stationary time** for the starting state x.
- τ_{refresh} is an example of a strong stationary time for any starting state x. (Why?)

Strong stationary times

- A strong stationary time for the starting state x is a stopping time (with respect to the chain and auxiliary randomness) satisfying

$$\mathbb{P}[\tau = t, X_\tau = y \mid X_0 = x] = \mathbb{P}[\tau = t \mid X_0 = x] \cdot \pi(y).$$

Strong stationary times

- A strong stationary time for the starting state x is a stopping time (with respect to the chain and auxiliary randomness) satisfying

$$\mathbb{P}[\tau = t, X_\tau = y \mid X_0 = x] = \mathbb{P}[\tau = t \mid X_0 = x] \cdot \pi(y).$$

$\Downarrow$ + law of total probability

- The property that will be useful to us is

$$\left[\ \mathbb{P}[\tau \leq t, X_t = y \mid X_0 = x] = \mathbb{P}[\tau \leq t \mid X_0 = x] \cdot \pi(y), \right]$$

which can be proved using the law of total probability.

Strong stationary times

Indeed,

$$\mathbb{P}[\tau \leq t, X_t = y \mid X_0 = x] = \sum_{s \leq t} \sum_{z \in S} \mathbb{P}[\tau = s, X_t = y, X_s = z \mid X_0 = x]$$

Strong stationary times

Indeed,

$$\begin{aligned}\mathbb{P}[\tau \leq t, X_t = y \mid X_0 = x] &= \sum_{s \leq t} \sum_{z \in S} \mathbb{P}[\tau = s, X_t = y, X_s = z \mid X_0 = x] \\ &= \sum_{s \leq t} \sum_{z \in S} \mathbb{P}[\tau = s, X_s = z \mid X_0 = x] \cdot P_{z,y}\end{aligned}$$

Strong stationary times

Indeed,

$$\begin{aligned}\mathbb{P}[\tau \leq t, X_t = y \mid X_0 = x] &= \sum_{s \leq t} \sum_{z \in S} \mathbb{P}[\tau = s, X_t = y, X_s = z \mid X_0 = x] \\ &= \sum_{s \leq t} \sum_{z \in S} \mathbb{P}[\tau = s, X_s = z \mid X_0 = x] \cdot P_{z,y} \\ &= \sum_{s \leq t} \sum_{z \in S} \left(\mathbb{P}[\tau = s \mid X_0 = x] \cdot \pi(z)\right) P_{z,y}\end{aligned}$$

Strong stationary times

Indeed,

$$\begin{aligned}
\mathbb{P}[\tau \leq t, X_t = y \mid X_0 = x] &= \sum_{s \leq t} \sum_{z \in S} \mathbb{P}[\tau = s, X_t = y, X_s = z \mid X_0 = x] \\
&= \sum_{s \leq t} \sum_{z \in S} \mathbb{P}[\tau = s, X_s = z \mid X_0 = x] \cdot P_{z,y} \\
&= \sum_{s \leq t} \sum_{z \in S} \left(\mathbb{P}[\tau = s \mid X_0 = x] \cdot \pi(z)\right) P_{z,y} \\
&= \sum_{s \leq t} \mathbb{P}[\tau = s \mid X_0 = x] \cdot \sum_{z \in S} \pi(z) P_{z,y}
\end{aligned}$$

Strong stationary times

Indeed,

$$\begin{aligned}
\mathbb{P}[\tau \le t, X_t = y \mid X_0 = x] &= \sum_{s \le t} \sum_{z \in S} \mathbb{P}[\tau = s, X_t = y, X_s = z \mid X_0 = x] \\
&= \sum_{s \le t} \sum_{z \in S} \mathbb{P}[\tau = s, X_s = z \mid X_0 = x] \cdot P_{z,y} \\
&= \sum_{s \le t} \sum_{z \in S} \left(\mathbb{P}[\tau = s \mid X_0 = x] \cdot \pi(z)\right) P_{z,y} \\
&= \sum_{s \le t} \mathbb{P}[\tau = s \mid X_0 = x] \cdot \sum_{z \in S} \pi(z) P_{z,y} \\
&= \sum_{s \le t} \mathbb{P}[\tau = s \mid X_0 = x] \cdot \pi(y) \\
&= \mathbb{P}[\tau \le t \mid X_0 = x] \cdot \pi(y).
\end{aligned}$$

Bounding the mixing time using strong stationary times

- Strong stationary times are also a useful tool for bounding the mixing time. This is captured by the following:
- Suppose that τ is a strong stationary time for the starting state x. Then,

$$\mathrm{TV}(X_t \mid X_0 = x, \pi) \leq \mathbb{P}[\tau > t \mid X_0 = x].$$

* not as tautological.

Recall that last time:

$$\left\{ \begin{aligned} D_{xy}(t) &= \mathrm{TV}(X_t \mid X_0 = x,\ X_t \mid X_0 = y) \\ &\leq \mathbb{P}[\tau_{couple} > t] \end{aligned} \right\}$$

τ_{couple} = coalescence time for $(\hat{X}_t, \hat{Y}_t)$.

Bounding the mixing time using strong stationary times

- Strong stationary times are also a useful tool for bounding the mixing time. This is captured by the following:
- Suppose that τ is a strong stationary time for the starting state x. Then,

$$\mathsf{TV}(X_t \mid X_0 = x, \pi) \leq \mathbb{P}[\tau > t \mid X_0 = x].$$

- Note that

$$\frac{1}{2} \sum_{y} | \pi(y) - P^t_{x,y} |$$

$$\mathsf{TV}(\underbrace{X_t \mid X_0}_{P^t(x, \cdot)} = x, \pi) = \sum_{y:\pi(y) > P^t_{x,y}} [\pi(y) - P^t_{x,y}]$$

$$\mathsf{TV}(\mathrm{Ber}(p), \mathrm{Ber}(q)) = \frac{1}{2} |p-q| + \frac{1}{2} |1-p-1+q| = |p-q|$$

Bounding the mixing time using strong stationary times

- Strong stationary times are also a useful tool for bounding the mixing time. This is captured by the following:
- Suppose that τ is a strong stationary time for the starting state x. Then,

last time:

$$\mathrm{TV}(X_t \mid X_0 = x, \pi) \leq \mathbb{P}[\tau > t \mid X_0 = x].$$

e.g. $\{0,1\}^n$

$\tau_{mix}(\varepsilon) \leq n\log n + n\log(\varepsilon^{-1})$.

$\tau = \tau_{refresh}$.

- Note that

$$\begin{aligned}\mathrm{TV}(X_t \mid X_0 = x, \pi) &= \sum_{y:\pi(y) > P^t_{x,y}} [\pi(y) - P^t_{x,y}] \\ &= \sum_{y:\pi(y) > P^t_{x,y}} \pi(y) \underbrace{\left(1 - \frac{P^t_{x,y}}{\pi(y)}\right)}_{\text{nonnegative}}\end{aligned}.$$

$\Sigma \pi(y) \leq 1$

Bounding the mixing time using strong stationary times

- Strong stationary times are also a useful tool for bounding the mixing time. This is captured by the following:
- Suppose that τ is a strong stationary time for the starting state x. Then,

$$\mathrm{TV}(X_t \mid X_0 = x, \pi) \leq \mathbb{P}[\tau > t \mid X_0 = x].$$

- Note that

$$\begin{aligned}
\mathrm{TV}(X_t \mid X_0 = x, \pi) &= \sum_{y:\pi(y) > P^t_{x,y}} [\pi(y) - P^t_{x,y}] \\
&= \sum_{y:\pi(y) > P^t_{x,y}} \pi(y) \left(1 - \frac{P^t_{x,y}}{\pi(y)}\right) \\
&\leq \max_{y \in S} \underbrace{\left(1 - \frac{P^t_{x,y}}{\pi(y)}\right)}.
\end{aligned}$$

separation distance

Bounding the mixing time using strong stationary times

- Hence, it will suffice to show that for all $y \in S$,

$$1 - \frac{P^t_{x,y}}{\pi(y)} \leq \mathbb{P}[\tau > t \mid X_0 = x].$$

$$\Rightarrow \quad TV(P^t(x,\cdot), \pi) \leq \max_{y \in S} \left(1 - \frac{P^t_{x,y}}{\pi(y)}\right)$$

$$\leq \mathbb{P}[\tau > t \mid X_0 = x].$$

Bounding the mixing time using strong stationary times

- Hence, it will suffice to show that for all $y \in S$,

$$1 - \frac{P^t_{x,y}}{\pi(y)} \leq \mathbb{P}[\tau > t \mid X_0 = x].$$

- We have

$$\begin{aligned}
1 - \frac{P^t_{x,y}}{\pi(y)} &= 1 - \frac{\mathbb{P}[X_t = y \mid X_0 = x]}{\pi(y)} \\
&\leq 1 - \frac{\mathbb{P}[X_t = y, \tau \leq t \mid X_0 = x]}{\pi(y)} \\
&= 1 - \frac{\cancel{\pi(y)}\,\mathbb{P}[\tau \leq t \mid X_0 = x]}{\cancel{\pi(y)}} \\
&= \mathbb{P}[\tau > t \mid X_0 = x].
\end{aligned}$$

Bounding the mixing time using strong stationary times

- Hence, it will suffice to show that for all $y \in S$,

$$1 - \frac{P^t_{x,y}}{\pi(y)} \le \mathbb{P}[\tau > t \mid X_0 = x].$$

- We have

$$\begin{aligned} 1 - \frac{P^t_{x,y}}{\pi(y)} &= 1 - \frac{\mathbb{P}[X_t = y \mid X_0 = x]}{\pi(y)} \\ &\le 1 - \frac{\mathbb{P}[X_t = y, \tau \le t \mid X_0 = x]}{\pi(y)} \\ &= 1 - \frac{\pi(y)\mathbb{P}[\tau \le t \mid X_0 = x]}{\pi(y)} \end{aligned}$$

Bounding the mixing time using strong stationary times

- Hence, it will suffice to show that for all $y \in S$,

$$1 - \frac{P^t_{x,y}}{\pi(y)} \leq \mathbb{P}[\tau > t \mid X_0 = x].$$

$*$ strong stationary times

$*$ $TV(P^t(x,\cdot), \pi) \leq \mathbb{P}[\tau > t \mid X_0 = x]$

- We have

$$\begin{aligned}
1 - \frac{P^t_{x,y}}{\pi(y)} &= 1 - \frac{\mathbb{P}[X_t = y \mid X_0 = x]}{\pi(y)} \\
&\leq 1 - \frac{\mathbb{P}[X_t = y, \tau \leq t \mid X_0 = x]}{\pi(y)} \\
&= 1 - \frac{\pi(y)\mathbb{P}[\tau \leq t \mid X_0 = x]}{\pi(y)} \\
&= \mathbb{P}[\tau > t \mid X_0 = x].
\end{aligned}$$

$*$ in examples: we want to find a not wasteful τ.

Example: top-to-random shuffle

- Consider a deck of n cards.
- At each step of the top-to-random shuffle, we take the top card, and place it in any of the n available positions.

Example: top-to-random shuffle

- Consider a deck of n cards.
- At each step of the top-to-random shuffle, we take the top card, and place it in any of the n available positions.
- On the homework, you showed that this is an aperiodic and irreducible chain with the unique stationary distribution π given by the uniform distribution on all possible permutations of the n cards.

\# of permutation of $\{1, \ldots, n\}$

$$= n!$$

$$\sim \ldots \left(\frac{n}{e}\right)^{n}$$

Example: top-to-random shuffle

- Consider a deck of n cards.
- At each step of the top-to-random shuffle, we take the top card, and place it in any of the n available positions.
- On the homework, you showed that this is an aperiodic and irreducible chain with the unique stationary distribution π given by the uniform distribution on all possible permutations of the n cards.
- Now, we will bound the mixing time by constructing a suitable strong stationary time.

Example: top-to-random shuffle

- Let τ_{top} be first time when the original bottom card has moved to the top.

A
B
C

- is τ_{top} a stationary time?

$$\mathbb{P}[X_{\tau_{top}} = y \mid X_0 = \begin{smallmatrix} A \\ B \\ C \end{smallmatrix}] = \pi(y).$$

Example: top-to-random shuffle

- Let τ_{top} be first time when the original bottom card has moved to the top.
- Then, $\tau = \tau_{\text{top}} + 1$ is a strong stationary time.

$$X_0 = \begin{matrix} A \\ B \\ C \\ D \end{matrix} \quad \cdots\cdots \quad X_{\tau_{top}} \quad \begin{matrix} D \\ B \\ A \\ C \end{matrix}$$

now select D

$$\begin{matrix} \times \\ B \\ \times \\ A \\ \times \\ C \\ \times \end{matrix}$$

Example: top-to-random shuffle

- Let τ_{top} be first time when the original bottom card has moved to the top.
- Then, $\tau = \tau_{\text{top}} + 1$ is a strong stationary time.
- Clearly, this is a stopping time.

Example: top-to-random shuffle

- Let τ_{top} be first time when the original bottom card has moved to the top.
- Then, $\tau = \tau_{\text{top}} + 1$ is a strong stationary time.
- Clearly, this is a stopping time.
- The key to showing that it is a strong stationary time is the following observation, which can be proved by induction: let τ_k denote the first time when there are k cards under the original bottom card. Then, at X_{τ_k}, all $k!$ permutations of these cards are equally likely. (Why?)

e.g.

$X_0 =$ A
B
C
D

e.g.: X_1 B
C
D
A

e.g. C
D
x
A
x

we know that B goes below D

Example: top-to-random shuffle

D

1 —
2 —
3 —
4 —

- Let τ_{top} be first time when the original bottom card has moved to the top.
- Then, $\tau = \tau_{\text{top}} + 1$ is a strong stationary time.
- Clearly, this is a stopping time.
- The key to showing that it is a strong stationary time is the following observation, which can be proved by induction: let τ_k denote the first time when there are k cards under the original bottom card. Then, at X_{τ_k}, all $k!$ permutations of these cards are equally likely. (Why?)
- Therefore, by the relationship between strong stationary times and total variation distance,

$$\mathrm{TV}(X_t \mid X_0 = x, \pi) \leq \mathbb{P}[\tau > t \mid X_0 = x].$$

A
B
C
D

Prob [A goes below D] = 1/4

D
A

Example: top-to-random shuffle

$\tau = \tau_{top} + 1$

first time when all cards are below original bottom card.

D —
1
2
3

- Note that, when there are k cards under the bottom card, the probability that the top card goes under it is $(k+1)/n$.
- Therefore, we see that

$$\tau \sim \mathrm{Geom}(1/n) + \mathrm{Geom}(2/n) + \cdots + \mathrm{Geom}(1),$$

where the geometric random variables on the right hand side are independent. This is the same as the distribution of the coupon collector's time.

- Therefore, as for the lazy random walk on the hypercube, we get that

$$\tau_{\mathrm{mix}}(\varepsilon) \le \underbrace{n \log n + n \log(\varepsilon^{-1})}_{\sim 193.2}.$$

$n = 52$

$\varepsilon = 0.01$

next time!
Riffle shuffle
$\tau_{mix}(1/4) \sim (\log) n$