COMPSCI 690RA: Randomized Algorithms and Probabilistic Data Analysis

---

Prof. Cameron Musco
University of Massachusetts Amherst. Spring 2022.
Lecture 11

Logistics

- Problem Set 4 is due next Tuesday 5/~~3~~2 at 8pm.
- The final exam is next Friday 5/5 at 10:30am for those that are taking it.
- I will hold extended office hours Wed. 5/3 from 2-4pm and Thurs. 5/4 from 4-6pm.
- I will accept final project submissions up until Sunday 5/8 at 11:59pm.
- SRTI’s are open for this course. It would be very helpful to me if you can fill them out!
- This was my first time teaching this course, so feedback on what worked and what didn’t is really useful to me.

Summary

Last Week: More Advanced Markov Chains.

- The gambler's ruin problem.
- Start on Markov chains and their analysis.
- Aperiodicity and stationary distribution of a Markov chain.
- Start on mixing time.

Summary

Last Week: More Advanced Markov Chains.

- The gambler's ruin problem.
- Start on Markov chains and their analysis.
- Aperiodicity and stationary distribution of a Markov chain.
- Start on mixing time.

Today: Finish up Markov Chains

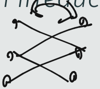
- Mixing time analysis via coupling.
- Markov Chain Monte Carlo (MCMC) methods.
- Sampling to counting · reductions.

Probabilistic Method

Fundamental Theorem of Markov Chains

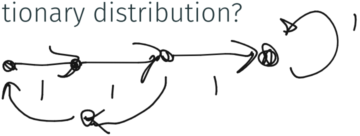
Theorem (The Fundamental Theorem of Markov Chains)

Let $X_0, X_1, \ldots$ *be a Markov chain with a finite state space and transition matrix* $P \in \underline{[0,1]^{m \times m}}$. *If the chain is both irreducible and aperiodic,*



1. *There exists a* *unique* *stationary distribution* $\pi \in [0,1]^m$ *with* $\underline{\pi = \pi P}$.
2. *For any states* i, j, $\lim_{t \to \infty} \Pr[X_t = i | X_0 = j] = \pi(i)$. *I.e., for any initial distribution* q_0, $\lim_{t \to \infty} q_t = \lim_{t \to \infty} q_0 P^t = \pi$.

Question for today: How long does it take us to converge close to this stationary distribution?



period
reducible
has unique π.

Mixing Time

Definition (Mixing Time)

Consider a Markov chain $X_0, X_1, \dots$ with unique stationary distribution π. Let $\underline{q_{i,t}}$ be the distribution over states at time t assuming $X_0 = i$. The mixing time is defined as:

$$\tau(\epsilon) = \min \left\{ t : \max_{\underline{i \in [m]}} \|\underline{q_{i,t}} - \underline{\pi}\|_{TV} \leq \epsilon \right\}.$$

Mixing Time

Definition (Mixing Time)

Consider a Markov chain $X_0, X_1, \ldots$ with unique stationary distribution π. Let $q_{i,t}$ be the distribution over states at time t assuming $X_0 = i$. The mixing time is defined as:

$$\tau(\epsilon) = \underline{\min} \left\{ t : \max_{i \in [m]} \|\underline{q_{i,t} - \pi}\|_{TV} \leq \epsilon \right\}.$$

Note: If $\|q_{i,t} - \pi\|_{TV} \leq \epsilon$ then for any $t' \geq t$, $\|q_{i,t'} - \pi\|_{TV} \leq \epsilon$.

$$q_{i,t} = \pi + e \qquad \|e\|_1 < \varepsilon$$

$$q_{i,t+1} = q_{i,t} P = \pi P + eP = \pi + \underset{\sim}{eP}$$

$$\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & \\ & & \end{bmatrix}$$

$$\|eP\|_1 < \|e\|_1$$

$$\leq \|e\|_1 \cdot \max \underbrace{\|P_i\|_1}_{1}$$

$$\leq \|e\|_1$$

Mixing Time

Definition (Mixing Time)

Consider a Markov chain $X_0, X_1, \ldots$ with unique stationary distribution π. Let $q_{i,t}$ be the distribution over states at time t assuming $X_0 = i$. The mixing time is defined as:

$$\tau(\epsilon) = \min \left\{ t : \max_{i \in [m]} \|q_{i,t} - \pi\|_{TV} \leq \epsilon \right\}.$$

Coupling Motivation: Last time we showed that

$$\underline{\max_{i \in [m]} \|q_{i,t} - \pi\|_{TV}} \leq \underline{\max_{i,j \in [m]} \|q_{i,t} - q_{j,t}\|_{TV}}.$$

By Kontorovich-Rubinstein duality, for $\underline{X_t}$, Y_t distributed by evolving the chain for t steps starting from state $\underset{X_0}{i}$ or $\underset{Y_0}{j}$ respectively, we have:

$$\underline{\max_{i,j \in [m]} \|q_{i,t} - q_{j,t}\|_{TV}} \leq \underline{\max_{i,j \in [m]} \Pr[X_t \neq Y_t]}.$$

Formal Coupling Definition

Definition (Coupling)

For a finite Markov chain $X_0, X_1, \ldots$ with transition matrix $P \in \mathbb{R}^{m \times m}$, a coupling is a joint process $(X_0, Y_0), (X_1, Y_1), \ldots$ such that:

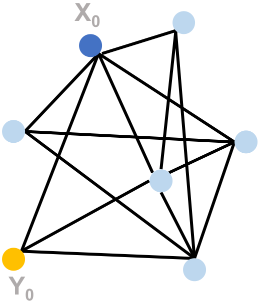
1. $X_0 = i$ and $Y_0 = j$ for some $i, j \in [m]$.
2. $\Pr[X_t = j | X_{t-1} = i] = \Pr[Y_t = j | Y_{t-1} = i] = P_{i,j}$
3. If $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.

Formal Coupling Definition

Definition (Coupling)

For a finite Markov chain $X_0, X_1, \ldots$ with transition matrix $P \in \mathbb{R}^{m \times m}$, a coupling is a joint process $(X_0, Y_0), (X_1, Y_1), \ldots$ such that:

1. $X_0 = i$ and $Y_0 = j$ for some $i, j \in [m]$.
2. $\Pr[X_t = j | X_{t-1} = i] = \Pr[Y_t = j | Y_{t-1} = i] = P_{i,j}$
3. If $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.

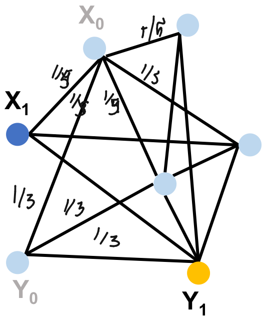


Formal Coupling Definition

Definition (Coupling)

For a finite Markov chain $X_0, X_1, \ldots$ with transition matrix $P \in \mathbb{R}^{m \times m}$, a coupling is a joint process $(X_0, Y_0), (X_1, Y_1), \ldots$ such that:

1. $X_0 = i$ and $Y_0 = j$ for some $i, j \in [m]$.
2. $\Pr[X_t = j | X_{t-1} = i] = \Pr[Y_t = j | Y_{t-1} = i] = P_{i,j}$
3. If $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.

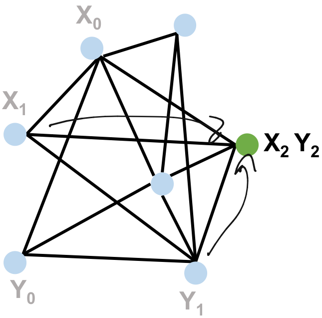


Formal Coupling Definition

Definition (Coupling)

For a finite Markov chain $X_0, X_1, \dots$ with transition matrix $P \in \mathbb{R}^{m \times m}$, a coupling is a joint process $(X_0, Y_0), (X_1, Y_1), \dots$ such that:

1. $X_0 = i$ and $Y_0 = j$ for some $i, j \in [m]$.
2. $\Pr[X_t = j | X_{t-1} = i] = \Pr[Y_t = j | Y_{t-1} = i] = P_{i,j}$
3. If $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.

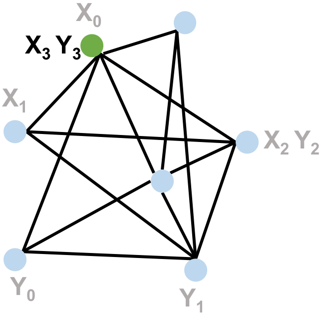


Formal Coupling Definition

Definition (Coupling)

For a finite Markov chain $X_0, X_1, \ldots$ with transition matrix $P \in \mathbb{R}^{m \times m}$, a coupling is a joint process $(X_0, Y_0), (X_1, Y_1), \ldots$ such that:

1. $X_0 = i$ and $Y_0 = j$ for some $i, j \in [m]$.
2. $\Pr[X_t = j | X_{t-1} = i] = \Pr[Y_t = j | Y_{t-1} = i] = P_{i,j}$
3. If $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.

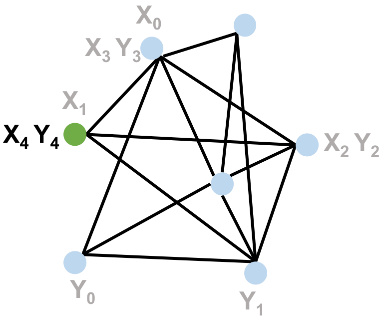


Formal Coupling Definition

Definition (Coupling)

For a finite Markov chain $X_0, X_1, \ldots$ with transition matrix $P \in \mathbb{R}^{m \times m}$, a coupling is a joint process $(X_0, Y_0), (X_1, Y_1), \ldots$ such that:

1. $X_0 = i$ and $Y_0 = j$ for some $i, j \in [m]$.
2. $\Pr[X_t = j | X_{t-1} = i] = \Pr[Y_t = j | Y_{t-1} = i] = P_{i,j}$
3. If $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.

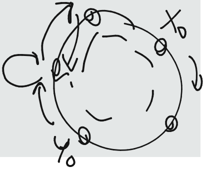


Formal Coupling Definition

Definition (Coupling)

For a finite Markov chain $X_0, X_1, \ldots$ with transition matrix $P \in \mathbb{R}^{m \times m}$, a coupling is a joint process $(X_0, Y_0), (X_1, Y_1), \ldots$ such that:

1. $X_0 = i$ and $Y_0 = j$ for some $i, j \in [m]$.
2. $\Pr[X_t = j | X_{t-1} = i] = \Pr[Y_t = j | Y_{t-1} = i] = P_{i,j}$
3. $\underline{\text{If } X_t = Y_t}$, then $X_{t+1} = Y_{t+1}$.



Theorem (Mixing Time Bound via Coupling)

For a finite, irreducible, and aperiodic Markov chain $X_0, X_1, \ldots$ *and any valid coupling* $(X_0, Y_0), (X_1, Y_1), \ldots$ *letting*
$\underline{T_{i,j}} = \min\{t : X_t = Y_t | X_0 = i, Y_0 = j\}$,

$$\max_{i \in [m]} \|q_{i,t} - \pi\|_{TV} \overset{\leq \epsilon}{\leq} \underline{\max_{i,j \in [m]} \|q_{i,t} - q_{j,t}\|_{TV}} \leq \underline{\max_{i,j \in [m]} \Pr[T_{i,j} > t]}.$$

Coupling Example: Mixing Time of Shuffling

How many times do we need to swap a random card to the top of the deck so that the distribution of orderings on our cards is ϵ-close in TV distance to the uniform distribution over all permutations?

Coupling Example: Mixing Time of Shuffling

How many times do we need to swap a random card to the top of the deck so that the distribution of orderings on our cards is ϵ-close in TV distance to the uniform distribution over all permutations?

Coupling:

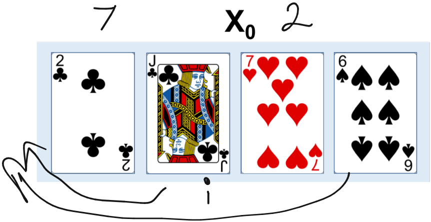
- Let $\mathsf{X}_0, \mathsf{X}_1, \ldots$ be the Markov chain where a random card is moved to the top in each step.
- Let $\mathsf{Y}_0, \mathsf{Y}_1$ be a correlated Markov chain. When card S is swapped to the top in the X chain, swap S to the top in the Y chain as well.

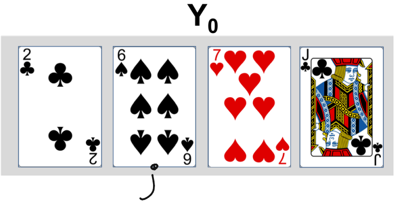
Coupling Example: Mixing Time of Shuffling

How many times do we need to swap a random card to the top of the deck so that the distribution of orderings on our cards is ϵ-close in TV distance to the uniform distribution over all permutations?

Coupling:

- Let $\mathbf{X}_0, \mathbf{X}_1, \ldots$ be the Markov chain where a random card is moved to the top in each step.
- Let $\mathbf{Y}_0, \mathbf{Y}_1$ be a correlated Markov chain. When card S is swapped to the top in the $\mathbf{X}$ chain, swap S to the top in the $\mathbf{Y}$ chain as well.



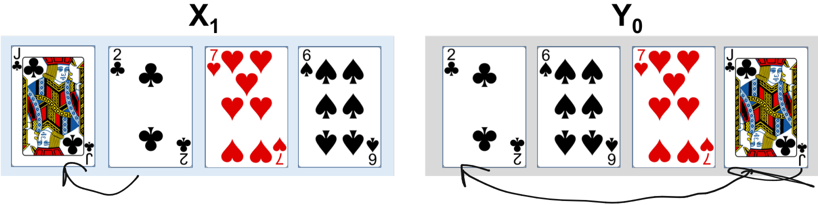


Coupling Example: Mixing Time of Shuffling

How many times do we need to swap a random card to the top of the deck so that the distribution of orderings on our cards is ϵ-close in TV distance to the uniform distribution over all permutations?

Coupling:

- Let $X_0, X_1, \ldots$ be the Markov chain where a random card is moved to the top in each step.
- Let Y_0, Y_1 be a correlated Markov chain. When card S is swapped to the top in the $\mathbf{X}$ chain, swap S to the top in the $\mathbf{Y}$ chain as well.

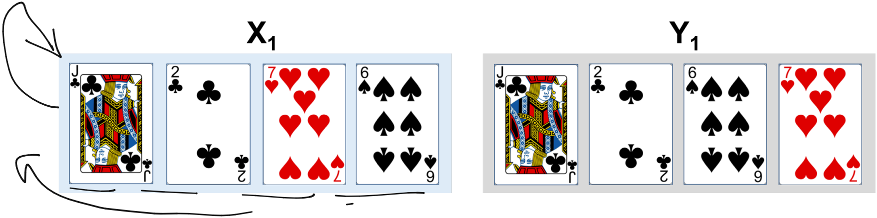


Coupling Example: Mixing Time of Shuffling

How many times do we need to swap a random card to the top of the deck so that the distribution of orderings on our cards is ϵ-close in TV distance to the uniform distribution over all permutations?

Coupling:

- Let $\mathbf{X}_0, \mathbf{X}_1, \ldots$ be the Markov chain where a random card is moved to the top in each step.
- Let $\mathbf{Y}_0, \mathbf{Y}_1$ be a correlated Markov chain. When card S is swapped to the top in the $\mathbf{X}$ chain, swap S to the top in the $\mathbf{Y}$ chain as well.

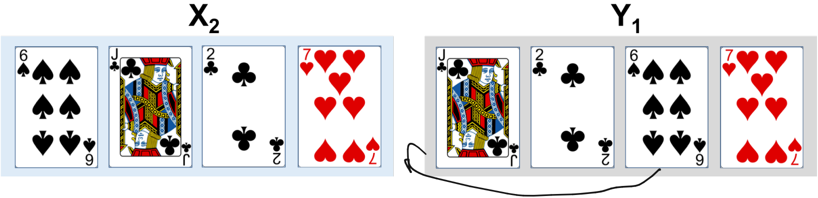


Coupling Example: Mixing Time of Shuffling

How many times do we need to swap a random card to the top of the deck so that the distribution of orderings on our cards is ϵ-close in TV distance to the uniform distribution over all permutations?

Coupling:

- Let $\mathbf{X}_0, \mathbf{X}_1, \ldots$ be the Markov chain where a random card is moved to the top in each step.
- Let $\mathbf{Y}_0, \mathbf{Y}_1$ be a correlated Markov chain. When card S is swapped to the top in the $\mathbf{X}$ chain, swap S to the top in the $\mathbf{Y}$ chain as well.

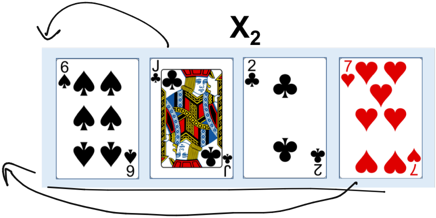


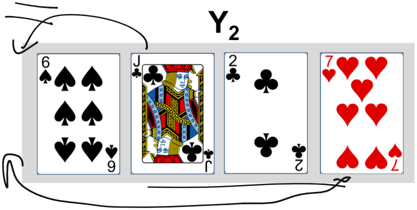
Coupling Example: Mixing Time of Shuffling

How many times do we need to swap a random card to the top of the deck so that the distribution of orderings on our cards is ϵ-close in TV distance to the uniform distribution over all permutations?

Coupling: J672 J627 6J72 6J27

- Let $X_0, X_1, \ldots$ be the Markov chain where a random card is moved to the top in each step.
- Let Y_0, Y_1 be a correlated Markov chain. When card S is swapped to the top in the $\mathbf{X}$ chain, swap S to the top in the $\mathbf{Y}$ chain as well.





Coupling Example: Mixing Time of Shuffling

How many times do we need to swap a random card to the top of the deck so that the distribution of orderings on our cards is ϵ-close in TV distance to the uniform distribution over all permutations?

Coupling:

- Let $X_0, X_1, \ldots$ be the Markov chain where a random card is moved to the top in each step.
- Let Y_0, Y_1 be a correlated Markov chain. When card S is swapped to the top in the X chain, swap S to the top in the Y chain as well.
- Can check that this is a valid coupling since X_t, Y_t have the correct marginal distributions, and since $X_t = Y_t \implies X_{t+1} = Y_{t+1}$

Coupling Example: Mixing Time of Shuffling

How many times do we need to swap a random card to the top of the deck so that the distribution of orderings on our cards is ϵ-close in TV distance to the uniform distribution over all permutations?

Coupling:

- Let $X_0, X_1, \ldots$ be the Markov chain where a random card is moved to the top in each step.
- Let Y_0, Y_1 be a correlated Markov chain. When card S is swapped to the top in the X chain, swap S to the top in the Y chain as well.
- Can check that this is a valid coupling since X_t, Y_t have the correct marginal distributions, and since $X_t = Y_t \implies X_{t+1} = Y_{t+1}$
- Observe that $X_t = Y_t$ as soon as all c unique cards have been swapped at least once. How many swaps does this take?

$C \log c$ steps in expectation

Coupling Example: Mixing Time of Shuffling

$$\begin{aligned}\max_{i\in[m]} \|q_{i,t} - \pi\|_{TV} &\le \max_{i,j\in[m]} \Pr[\underline{\mathsf{T}_{i,j} > t}] \\ &\le \Pr[<\underline{c \text{ unique}} \text{ cards are swapped in } t \text{ swaps}]\end{aligned}$$

Coupling Example: Mixing Time of Shuffling

$\underline{c \ln c}$

$$\max_{i\in[m]} \|q_{i,t} - \pi\|_{TV} \leq \max_{i,j\in[m]} \Pr[\mathsf{T}_{i,j} > t]$$

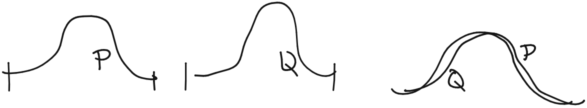
$$\leq \Pr[< c \text{ unique cards are swapped in } t \text{ swaps}]$$

By coupon collector analysis for $\underline{t \geq c\ln(\underline{c/\epsilon})}$, this probability is bounded by ϵ. In particular, by the fact that $\underbrace{\left(1-\frac{1}{c}\right)^{c\ln c/\epsilon}}_{\leq e^{-\ln(c/\epsilon)}} \leq \frac{\epsilon}{c}$ plus a union bound over c cards.

Coupling Example: Mixing Time of Shuffling

$Q = [\frac{1}{6}\ \frac{1}{6}\ \cdots\ \frac{1}{6}]$

$P = [\frac{1}{5}\ \frac{1}{5}\ , \cdots\ 0]$



$$\max_{i \in [m]} \|q_{i,t} - \pi\|_{TV} \leq \max_{i,j \in [m]} \Pr[\mathsf{T}_{i,j} > t]$$

$$\leq \Pr[< c \text{ unique cards are swapped in } t \text{ swaps}]$$

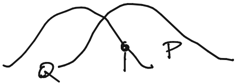
By coupon collector analysis for $t \geq c \ln(c/\epsilon)$, this probability is bounded by ϵ. In particular, by the fact that $\left(1 - \frac{1}{c}\right)^{c \ln c/\epsilon} \leq \frac{\epsilon}{c}$ plus a union bound over c cards.

$1 - \frac{5}{6}$

Thus, for $t \geq c \ln(c/\epsilon)$,
$\max_{i \in [m]} \|q_{i,t} - \pi\|_{TV} \leq \max_{i,j \in [m]} \|q_{i,t} - q_{j,t}\|_{TV} \leq \epsilon$.

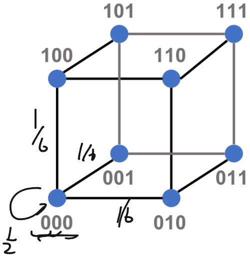
I.e., $\tau(\epsilon) \leq c \ln(c/\epsilon)$.

$\|q_{i,t} - q_{j,t}\|_{TV} = \min_{\substack{X_t \sim q_{i,t} \\ Y_t \sim q_{j,t}}} \Pr(X_t \neq Y_t)$



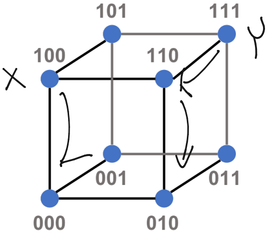
Coupling Example: Random Walk on a Hypercube

Let X_0, X_1 be a Markov chain over state space $\{0,1\}^n$. In each step, pick a random position $i \in [n]$ and set $X_t(i) = 0$ with probability 1/2 and $X_t(i) = 1$ with probability 1/2.



Coupling Example: Random Walk on a Hypercube

Let X_0, X_1 be a Markov chain over state space $\{0,1\}^n$. In each step, pick a random position $i \in [n]$ and set $X_t(i) = 0$ with probability 1/2 and $X_t(i) = 1$ with probability 1/2.

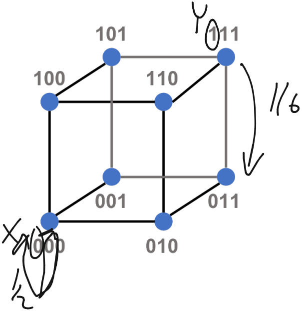


What is a coupling $(X_0, Y_0), (X_1, Y_1), \ldots$ on this chain that we can use to bound the mixing time of this walk?

Coupling Example: Random Walk on a Hypercube

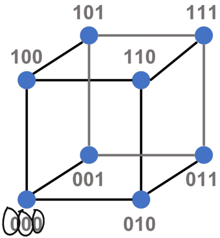
In each step, pick a single random position $i \in [n]$ and let $\underline{X_t(i) = Y_t(i) = 0}$ with probability 1/2 and $X_t(i) = Y_t(i) = \underline{1}$ with probability 1/2.

$\frac{1}{3} \cdot \frac{1}{2}$



Coupling Example: Random Walk on a Hypercube

In each step, pick a single random position $i \in [n]$ and let $X_t(i) = Y_t(i) = 0$ with probability 1/2 and $X_t(i) = Y_t(i) = 1$ with probability 1/2.



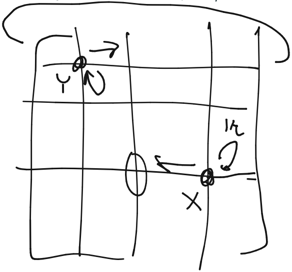
$\tau(\epsilon)$

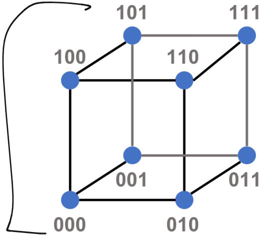
How large must we set t so that $\Pr[X_t \neq Y_t] \leq \epsilon$?

(coupon collection

Coupling Example: Random Walk on a Hypercube

In each step, pick a single random position $i \in [n]$ and let $X_t(i) = Y_t(i) = 0$ with probability 1/2 and $X_t(i) = Y_t(i) = 1$ with probability 1/2.





How large must we set t so that $\Pr[X_t \neq Y_t] \leq \epsilon$?

Upshot: The mixing time of the n-dimensional hypercube is $\tau(\epsilon) = \underline{O(n \log(n/\epsilon))}$.

Markov Chain Monte Carlo

Markov Chain Monte Carlo

Many applications in computational biology, machine learning, theoretical computer science, etc. require sampling from complex distributions, which are difficult to write down in closed form, and difficult to directly sample from.

A very common approach is to design a Markov chain whose stationary distribution π is equal to the distribution of interest.

By running this Markov chain for at least $\tau(\epsilon)$ steps (burn-in time), one can draw a sample which is nearly from the distribution of interest.

Markov Chain Monte Carlo

Many applications in computational biology, machine learning, theoretical computer science, etc. require sampling from complex distributions, which are difficult to write down in closed form, and difficult to directly sample from.

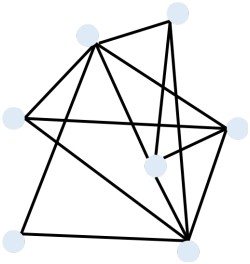
A very common approach is to design a Markov chain whose stationary distribution π is equal to the distribution of interest.

By running this Markov chain for at least $\tau(\epsilon)$ steps (burn-in time), one can draw a sample which is nearly from the distribution of interest.

Note: A major focus is on designing and analyzing Markov chains where $\tau(\epsilon)$ is small. For today, we'll just focus on getting the stationary distribution right, and mostly ignore runtime.

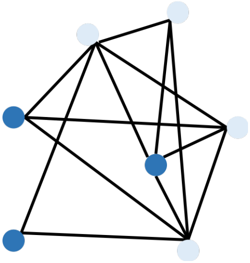
Sampling Independent Sets

Suppose we would like to sample a uniformly random independent set from a graph G.



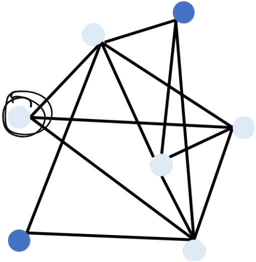
Sampling Independent Sets

Suppose we would like to sample a uniformly random independent set from a graph G.



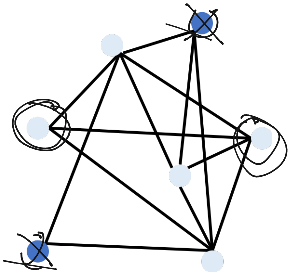
Sampling Independent Sets

Suppose we would like to sample a uniformly random independent set from a graph G.



Sampling Independent Sets

Suppose we would like to sample a uniformly random independent set from a graph G.



Very non-obvious how to sample from this distribution. Exactly counting the number of independent sets, which is closely related to sampling, is #~~P-hard~~.

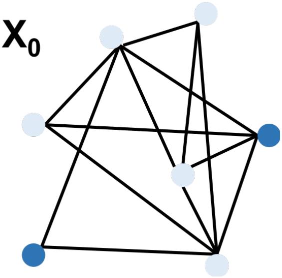
NP-Hard

Markov Chain on Independent Sets

Design a Markov chain $X_0, X_1, \ldots$ whose states are exactly the independent sets. E.g., let X_{t+1} be chosen uniformly at random from $\mathcal{N}(X_t) = \{Y : \text{independent set formed by adding/removing a node from } X_t\}$.

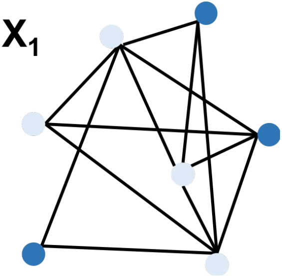
Markov Chain on Independent Sets

Design a Markov chain $X_0, X_1, \ldots$ whose states are exactly the independent sets. E.g., let X_{t+1} be chosen uniformly at random from $\mathcal{N}(X_t) = \{Y : \text{independent set formed by adding/removing a node from } X_t\}$.



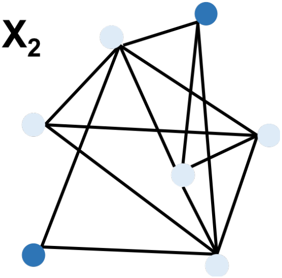
Markov Chain on Independent Sets

Design a Markov chain $X_0, X_1, \ldots$ whose states are exactly the independent sets. E.g., let X_{t+1} be chosen uniformly at random from $\mathcal{N}(X_t) = \{Y : \text{independent set formed by adding/removing a node from } X_t\}$.



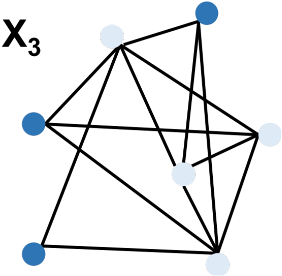
Markov Chain on Independent Sets

Design a Markov chain $X_0, X_1, \ldots$ whose states are exactly the independent sets. E.g., let X_{t+1} be chosen uniformly at random from $\mathcal{N}(X_t) = \{Y : \text{independent set formed by adding/removing a node from } X_t\}$.



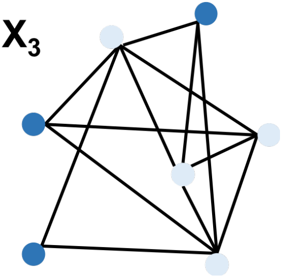
Markov Chain on Independent Sets

Design a Markov chain $X_0, X_1, \ldots$ whose states are exactly the independent sets. E.g., let X_{t+1} be chosen uniformly at random from $\mathcal{N}(X_t) = \{Y : \text{independent set formed by adding/removing a node from } X_t\}$.



Markov Chain on Independent Sets

Design a Markov chain $X_0, X_1, \ldots$ whose states are exactly the independent sets. E.g., let X_{t+1} be chosen uniformly at random from $\mathcal{N}(X_t) = \{Y : \text{independent set formed by adding/removing a node from } X_t\}$.



Unfortunately, the stationary distribution of this chain may not be uniform. It places higher probability on independent sets with lots of neighboring independent sets.

Achieving a Uniform Stationary Distribution

Define a Markov chain $X_0, X_1, \ldots$ over independent sets with transition function:

- Pick $\underline{\text{a random vertex } v}$.
- If $\underline{v \in}\ X_t$, set $X_{t+1} = \underline{X_t \setminus \{v\}}$.
- If $\underline{v \notin X_t}$ and $X_t \cup \{v\}$ is independent, set $\underline{X_{t+1} = X_t} \cup \{v\}$.
- Else set $\underline{X_{t+1} = X_t}$.

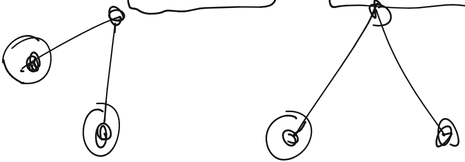
Achieving a Uniform Stationary Distribution

Define a Markov chain $X_0, X_1, \ldots$ over independent sets with transition function:

- Pick a random vertex v.
- If $v \in X_t$, set $X_{t+1} = X_t \setminus \{v\}$.
- If $v \notin X_t$ and $X_t \cup \{v\}$ is independent, set $X_{t+1} = X_t \cup \{v\}$.
- Else set $X_{t+1} = X_t$.

at least one self loop.

Is this chain irreducible and aperiodic?



Achieving a Uniform Stationary Distribution

Define a Markov chain $X_0, X_1, \ldots$ over independent sets with transition function:

- Pick a random vertex v.
- If $v \in X_t$, set $X_{t+1} = X_t \setminus \{v\}$.
- If $v \notin X_t$ and $X_t \cup \{v\}$ is independent, set $X_{t+1} = X_t \cup \{v\}$.
- Else set $X_{t+1} = X_t$.

Is this chain irreducible and aperiodic? Yes.

Achieving a Uniform Stationary Distribution

Define a Markov chain $X_0, X_1, \ldots$ over independent sets with transition function:

- Pick a random vertex v.
- If $v \in X_t$, set $X_{t+1} = X_t \setminus \{v\}$.
- If $v \notin X_t$ and $X_t \cup \{v\}$ is independent, set $X_{t+1} = X_t \cup \{v\}$.
- Else set $X_{t+1} = X_t$.

Is this chain irreducible and aperiodic? Yes. if ≥ 1 edge

For any two independent sets i, j, what is $P_{i,j}$?

$P_{ij} = 0$ if i, j are not neighbors

$P_{ij} = \frac{1}{|V|}$ if i, j differ by 1 vertex

Achieving a Uniform Stationary Distribution

Define a Markov chain $X_0, X_1, \ldots$ over independent sets with transition function:

- Pick a random vertex v.
- If $v \in X_t$, set $X_{t+1} = X_t \setminus \{v\}$.
- If $v \notin X_t$ and $X_t \cup \{v\}$ is independent, set $X_{t+1} = X_t \cup \{v\}$.
- Else set $X_{t+1} = X_t$.

Is this chain irreducible and aperiodic? Yes.

For any two independent sets i, j, what is $P_{i,j}$? $P_{i,j} = P_{j,i} = 1/|V|$ if i, j differ by one vertex, $P_{i,j} = P_{j,i} = 0$ otherwise.

symmetric

Achieving a Uniform Stationary Distribution

Define a Markov chain $X_0, X_1, \ldots$ over independent sets with transition function:

- Pick a random vertex v.
- If $v \in X_t$, set $X_{t+1} = X_t \setminus \{v\}$.
- If $v \notin X_t$ and $X_t \cup \{v\}$ is independent, set $X_{t+1} = X_t \cup \{v\}$.
- Else set $X_{t+1} = X_t$.

Is this chain irreducible and aperiodic? Yes.

For any two independent sets i, j, what is $P_{i,j}$? $P_{i,j} = P_{j,i} = 1/|V|$ if i, j differ by one vertex, $P_{i,j} = P_{j,i} = 0$ otherwise.

Thus, the Markov chain is symmetric, so by our claim from last class, the stationary distribution is uniform.

Achieving a Non-Uniform Stationary Distribution

Suppose we want to sample an independent set X from our graph with probability:

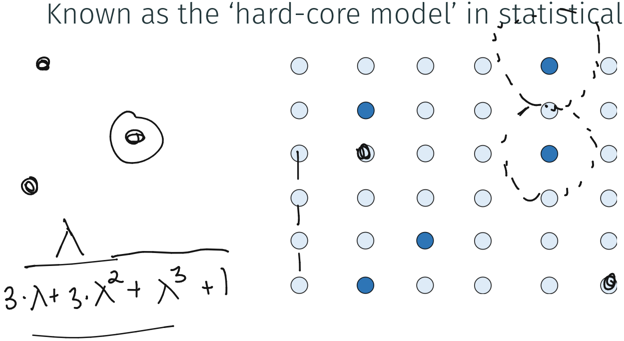
$P(X) = \lambda^{|X|}$

$$\pi(X) = \frac{\lambda^{|X|}}{\sum_{Y \text{ independent}} \lambda^{|Y|}},$$

$\pi(X) = \frac{P(X)}{\sum_{Y} P(Y)}$

for some 'fugacity' parameter $\lambda > 0$.

Known as the 'hard-core model' in statistical physics.



$\frac{\lambda}{3 \cdot \lambda + 3 \cdot \lambda^2 + \lambda^3 + 1}$

Metropolis-Hastings Algorithm

A very generic way of designing a Markov chain over state space $[m]$ with stationary distribution $\pi \in [0,1]^m$.

Metropolis-Hastings Algorithm

A very generic way of designing a Markov chain over state space $[m]$ with stationary distribution $\underline{\pi \in [0,1]}^m$. $p(x) = \lambda^{|x|}$

- Assume the ability to efficiently compute a density $\underline{p(X)} \propto \underline{\pi(X)}$.
- Assume access to some symmetric transition function with transition probability matrix $Q \in [0,1]^{m \times m}$.

Metropolis-Hastings Algorithm

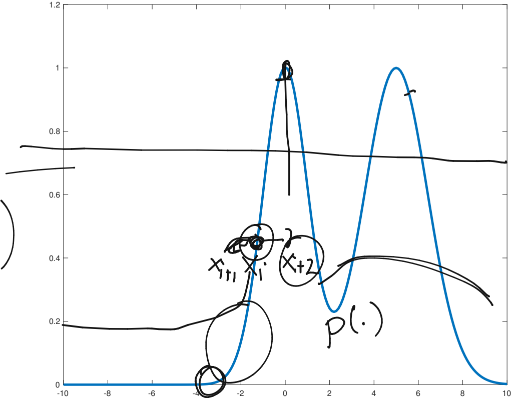
A very generic way of designing a Markov chain over state space $[m]$ with stationary distribution $\pi \in [0,1]^m$.

- Assume the ability to efficiently compute a density $p(X) \propto \pi(X)$.
- Assume access to some symmetric transition function with transition probability matrix $Q \in [0,1]^{m \times m}$.
- At step t, generate a 'candidate' state X_{t+1} from X_t according to Q.
- With probability $\min\left(1, \frac{p(\mathrm{X}_{t+1})}{p(\mathrm{X}_t)}\right)$, 'accept' the candidate. Else 'reject' the candidate, setting $\mathrm{X}_{t+1} = \mathrm{X}_t$.

Metropolis-Hastings Intuition

Q

$\min\left(1, \frac{p(x'_{i+1})}{p(x'_i)}\right)$



Metropolis-Hastings Analysis

Need to check that for the Metropolis-Hastings algorithm, $\underline{\pi P} = \pi$.

Metropolis-Hastings Analysis

$p(x) = \lambda^{|x|}$

$\pi(x) = \frac{\lambda^{|x|}}{\sum \lambda^{|y|}}$

Need to check that for the Metropolis-Hastings algorithm, $\underline{\pi P = \pi}$.

Suffices to show that $pP = p$ where $p(i) \propto \pi(i)$ is our efficiently computable density.

Metropolis-Hastings Analysis

Need to check that for the Metropolis-Hastings algorithm, $\pi P = \pi$.

Suffices to show that $pP = p$ where $p(i) \propto \pi(i)$ is our efficiently computable density.

$$[pP](i) = \underbrace{\sum_j p(j) \cdot Q_{j,i} \cdot \min\left(1, \frac{p(i)}{p(j)}\right)}_{\textit{aceptances}} + \underbrace{p(i) \cdot \sum_j Q_{i,j}\left(1 - \min\left(1, \frac{p(j)}{p(i)}\right)\right)}_{\textit{rejections}}$$

$$[p]\begin{bmatrix} P \\ \\ \end{bmatrix} = [pP_i]$$

Metropolis-Hastings Analysis

Need to check that for the Metropolis-Hastings algorithm, $\pi P = \pi$.

Suffices to show that $pP = p$ where $p(i) \propto \pi(i)$ is our efficiently computable density.

$$[pP](i) = \underbrace{\sum_j p(j) \cdot Q_{j,i} \cdot \min\left(1, \frac{p(i)}{p(j)}\right)}_{\text{aceptances}} + \underbrace{p(i) \cdot \sum_j Q_{i,j}\left(1 - \min\left(1, \frac{p(j)}{p(i)}\right)\right)}_{\text{rejections}}$$

$$= \sum_j Q_{i,j} \cdot \min\left(p(j), p(i)\right) + p(i) \cdot \sum_j Q_{i,j} - \sum_j Q_{i,j} \cdot \min(p(i), p(j))$$

Metropolis-Hastings Analysis

Need to check that for the Metropolis-Hastings algorithm, $\pi P = \pi$.

Suffices to show that $pP = p$ where $p(i) \propto \pi(i)$ is our efficiently computable density.

$$\underline{[pP](i)} = \underbrace{\sum_j p(j) \cdot Q_{j,i} \cdot \min\left(1, \frac{p(i)}{p(j)}\right)}_{\textit{aceptances}} + \underbrace{p(i) \cdot \sum_j Q_{i,j} \left(1 - \min\left(1, \frac{p(j)}{p(i)}\right)\right)}_{\textit{rejections}}$$

$$= \sum_j Q_{i,j} \cdot \min\left(p(j), p(i)\right) + p(i) \cdot \sum_j Q_{i,j} - \sum_j Q_{i,j} \cdot \min(p(i), p(j))$$

$$= \underline{p(i)} \cdot \underbrace{\sum_j Q_{i,j}}_{1} = \underline{p(i)}$$

Metropolis-Hastings Analysis

Need to check that for the Metropolis-Hastings algorithm, $\pi P = \pi$.

Suffices to show that $pP = p$ where $p(i) \propto \pi(i)$ is our efficiently computable density.

$$\begin{aligned}
[pP](i) &= \underbrace{\sum_j p(j) \cdot Q_{j,i} \cdot \min\left(1, \frac{p(i)}{p(j)}\right)}_{\textit{aceptances}} + \underbrace{p(i) \cdot \sum_j Q_{i,j} \left(1 - \min\left(1, \frac{p(j)}{p(i)}\right)\right)}_{\textit{rejections}} \\
&= \sum_j Q_{i,j} \cdot \min\left(p(j), p(i)\right) + p(i) \cdot \sum_j Q_{i,j} - \sum_j Q_{i,j} \cdot \min(p(i), p(j)) \\
&= p(i) \cdot \sum_j Q_{i,j} = p(i).
\end{aligned}$$

Metropolis-Hastings for the Hard-Core Model

Want to sample an independent set X with probability $\underline{\pi(X)} = \frac{\underline{\lambda^{|X|}}}{\sum_{Y \text{ independent}} \lambda^{|Y|}}$.

- Let $\underline{p(X) = \lambda^{|X|}}$ and let the transition function Q be given by:
  - Pick a random vertex v.
  - If $v \in \mathsf{X}_t$, set $\mathsf{X}_{t+1} = \mathsf{X}_t \setminus \{v\}$
  - If $v \notin \mathsf{X}_t$ and $\mathsf{X}_t \cup \{v\}$ is independent, set $\mathsf{X}_{t+1} = \mathsf{X}_t \cup \{v\}$.
  - Else set $\mathsf{X}_{t+1} = \mathsf{X}_t$

Metropolis-Hastings for the Hard-Core Model

Want to sample an independent set X with probability $\pi(X) = \frac{\lambda^{|X|}}{\sum_{Y \text{ independent}} \lambda^{|Y|}}$.

- Let $p(X) = \lambda^{|X|}$ and let the transition function Q be given by:
    - Pick a random vertex v.
    - If $v \in \mathsf{X}_t$, set $\underline{\mathsf{X}_{t+1} = \mathsf{X}_t \setminus \{v\}}$
    - If $v \notin \mathsf{X}_t$ and $\underline{\mathsf{X}_t \cup \{v\}}$ is independent, set $\mathsf{X}_{t+1} = \mathsf{X}_t \cup \{v\}$.
    - Else set $\mathsf{X}_{t+1} = \mathsf{X}_t$
- Need to accept the transition with probability $\min\left(1, \underset{\sim}{\frac{p(\mathsf{X}_{t+1})}{p(\mathsf{X}_t)}}\right)$.

$\frac{1}{\lambda}$ if removing vertex

λ if adding vertex

Metropolis-Hastings for the Hard-Core Model

Want to sample an independent set X with probability $\pi(X) = \frac{\lambda^{|X|}}{\sum_{Y \text{ independent}} \lambda^{|Y|}}$.

- Let $p(X) = \lambda^{|X|}$ and let the transition function Q be given by:
  - Pick a random vertex v.
  - If $v \in \mathrm{X}_t$, set $\mathrm{X}_{t+1} = \mathrm{X}_t \setminus \{v\}$ with probability $\min(1, 1/\lambda)$.
  - If $v \notin \mathrm{X}_t$ and $\mathrm{X}_t \cup \{v\}$ is independent, set $\mathrm{X}_{t+1} = \mathrm{X}_t \cup \{v\}$.
  - Else set $\mathrm{X}_{t+1} = \mathrm{X}_t$ with probability $\min(1, \lambda)$.
- Need to accept the transition with probability $\min\left(1, \frac{p(\mathrm{X}_{t+1})}{p(\mathrm{X}_t)}\right)$.

Metropolis-Hastings for the Hard-Core Model

Want to sample an independent set X with probability $\pi(X) = \frac{\lambda^{|X|}}{\sum_{Y \text{ independent}} \lambda^{|Y|}}$.

- Let $p(X) = \lambda^{|X|}$ and let the transition function Q be given by:
    - Pick a random vertex v.
    - If $v \in \mathsf{X}_t$, set $\mathsf{X}_{t+1} = \mathsf{X}_t \setminus \{v\}$ with probability $\min(1, 1/\lambda)$.
    - If $v \notin \mathsf{X}_t$ and $\mathsf{X}_t \cup \{v\}$ is independent, set $\mathsf{X}_{t+1} = \mathsf{X}_t \cup \{v\}$.
    - Else set $\mathsf{X}_{t+1} = \mathsf{X}_t$ with probability $\min(1, \lambda)$.
- Need to accept the transition with probability $\min\left(1, \frac{p(\mathsf{X}_{t+1})}{p(\mathsf{X}_t)}\right)$.

The key challenge then becomes to analyze the mixing time.

For the related Glauber dynamics, Luby and Vigoda showed that for graphs with maximum degree Δ, when $\lambda < \frac{2}{\Delta - 2}$, the mixing time is $O(n \log n)$. But when $\lambda > \frac{c}{\Delta}$ for large enough constant c, it is NP-hard to approximately sample from the hard-core model.

Counting to Sampling Reductions

Often if one can efficiently sample from the distribution $\pi(X) = \frac{p(X)}{\sum_Y p(Y)}$, one can efficiently approximate the normalizing constant $Z = \sum_Y p(Y)$ (often called the partition function).

- If Z is hard to approximate, then this can give a proof that sampling is hard, and thus it is unlikely that any simple MCMC method for sampling from π mixes rapidly.
- This is e.g., how one can show that sampling from the hard-core model is hard when $\lambda = \Omega(1/\Delta)$.

Counting to Sampling Reductions

Often if one can efficiently sample from the distribution $\pi(X) = \frac{p(X)}{\underline{\sum_Y p(Y)}}$, one can efficiently approximate the normalizing constant $Z = \sum_Y p(Y)$ (often called the partition function).

- If Z is hard to approximate, then this can give a proof that sampling is hard, and thus it is unlikely that any simple MCMC method for sampling from π mixes rapidly.

$\frac{1}{Z}$

$Z = |S(G)|$

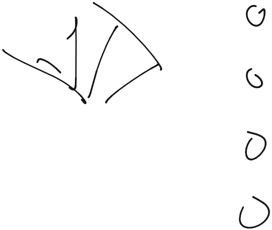
- This is e.g., how one can show that sampling from the hard-core model is hard when $\lambda = \Omega(1/\Delta)$.
- Let's consider the simple case of $\underline{\lambda = 1}$. I.e., we want to sample a uniformly random independent set.
- In this case, $Z = |S(G)|$, the number of independent sets in G. It is known that approximating $|S(G)|$ even up to a poly(n) factor is NP-Hard.

Counting Independent Sets

How can we count the number of independent sets $|S(G)|$ in a graph, given an oracle for sampling a uniform random independent set?

- take roughly $|S(G)|$ samples
- take roughly $\sqrt{|S(G)|}$ samples you'll see

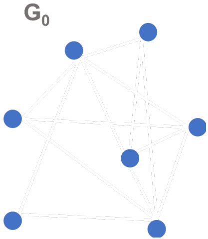
repeat.



Counting Independent Sets

How can we count the number of independent sets $|S(G)|$ in a graph, given an oracle for sampling a uniform random independent set?

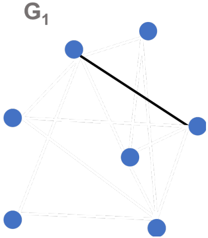
Let $G_0, G_1, \ldots, G_m$ be a sequence of graphs with $G_m = G$ and G_i obtained by removing an arbitrary edge from G_{i+1}.



Counting Independent Sets

How can we count the number of independent sets $|S(G)|$ in a graph, given an oracle for sampling a uniform random independent set?

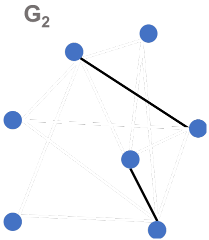
Let $G_0, G_1, \ldots, G_m$ be a sequence of graphs with $G_m = G$ and G_i obtained by removing an arbitrary edge from G_{i+1}.



Counting Independent Sets

How can we count the number of independent sets $|S(G)|$ in a graph, given an oracle for sampling a uniform random independent set?

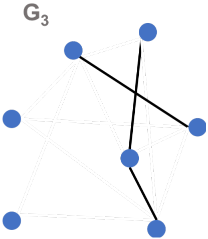
Let $G_0, G_1, \ldots, G_m$ be a sequence of graphs with $G_m = G$ and G_i obtained by removing an arbitrary edge from G_{i+1}.



Counting Independent Sets

How can we count the number of independent sets $|S(G)|$ in a graph, given an oracle for sampling a uniform random independent set?

Let $G_0, G_1, \ldots, G_m$ be a sequence of graphs with $G_m = G$ and G_i obtained by removing an arbitrary edge from G_{i+1}.

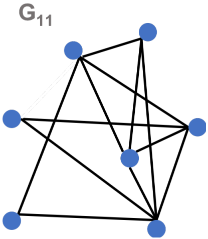


Counting Independent Sets

How can we count the number of independent sets $|S(G)|$ in a graph, given an oracle for sampling a uniform random independent set?

', G

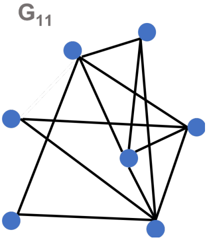
Let $G_0, G_1, \ldots, G_m$ be a sequence of graphs with $G_m = G$ and G_i obtained by removing an arbitrary edge from G_{i+1}.



Counting Independent Sets

How can we count the number of independent sets $|S(G)|$ in a graph, given an oracle for sampling a uniform random independent set?

Let $G_0, G_1, \ldots, G_m$ be a sequence of graphs with $G_m = G$ and G_i obtained by removing an arbitrary edge from G_{i+1}.



We can write:

$$\underline{|S(G)|} = \frac{\overset{S(G)}{\boxed{|S(G_m)|}}}{\cancel{|S(G_{m-1})|}} \cdot \frac{\cancel{|S(G_{m-1})|}}{\cancel{|S(G_{m-2})|}} \cdot \ldots \cdot \frac{\cancel{|S(G_1)|}}{\cancel{|S(G_0)|}} \cdot \cancel{|S(G_0)|}.$$

Counting Independent Sets

$$|S(G)| = \frac{|S(G_m)|}{|S(G_{m-1})|} \cdot \frac{|S(G_{m-1})|}{|S(G_{m-2})|} \cdot \ldots \cdot \frac{|S(G_1)|}{|S(G_0)|} \cdot \overset{2^n}{\boxed{|S(G_0)|}}$$

Counting Independent Sets

$$\underline{|S(G)|} = \frac{|S(G_m)|}{\underline{|S(G_{m-1})|}} \cdot \frac{|S(G_{m-1})|}{\underline{|S(G_{m-2})|}} \cdot \ldots \cdot \frac{|S(G_1)|}{\underline{|S(G_0)|}} \cdot 2^n$$

Counting Independent Sets

$$|S(G)| = \frac{|S(G_m)|}{|S\underbrace{(G_{m-1}}_{})|} \cdot \frac{|S(G_{m-1})|}{|S(\underbrace{G_{m-2}}_{})|} \cdot \ldots \cdot \frac{|S(G_1)|}{|S\underbrace{(G_0)}_{}|} \cdot 2^n = 2^n \cdot \Pi_{i=1}^{m} r_i,$$

where $r_i = \underbrace{\frac{|S(G_{i})|}{|S(G_{i-1})|}}_{}$.

Counting Independent Sets

$$|S(G)| = \frac{|S(G_m)|}{|S(G_{m-1})|} \cdot \frac{|S(G_{m-1})|}{|S(G_{m-2})|} \cdot \ldots \cdot \frac{|S(G_1)|}{|S(G_0)|} \cdot 2^n = 2^n \cdot \Pi_{i=1}^{m} r_i,$$

where $r_i = \frac{|S(G_m)|}{|S(G_{m-i})|}$. If we can estimate each r_i with $\tilde{r}_i$ satisfying

$$\left(1 - \frac{\epsilon}{2m}\right) \cdot r_i \le \underline{\tilde{r}_i} \le \left(1 + \frac{\epsilon}{2m}\right) \cdot r_i,$$

then:

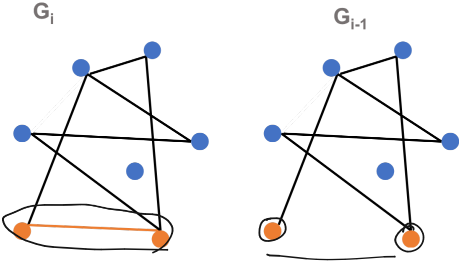
$$\underline{(1-\epsilon) \cdot |S(G)|} \le \underline{2^n \cdot \Pi_{i=1}^{m} \tilde{r}_i} \le \underline{(1+\epsilon) \cdot |S(G)|}$$

since $\underline{\left(1 + \frac{\epsilon}{2m}\right)}^m \le 1 + \epsilon$ and $\underline{\left(1 - \frac{\epsilon}{2m}\right)}^m \ge 1 - \epsilon$.

Independent Set Ratios

Consider the ratio $r_i = \frac{|S(G_i)|}{|S(G_{i-1})|}$. Observe that $r_i \leq 1$.

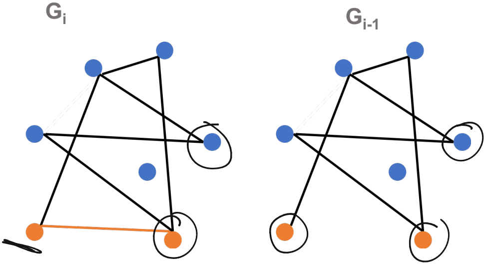
Further, $r_i \geq 1/2$. Let (u, v) be the edge removed from G_i to obtain G_{i-1}. Then each independent set in $S(G_{i-1}) \setminus S(G_i)$, must contain both u and v.



Independent Set Ratios

Consider the ratio $r_i = \frac{|S(G_i)|}{|S(G_{i-1})|}$. Observe that $r_i \leq 1$.

Further, $r_i \geq 1/2$. Let (u, v) be the edge removed from G_i to obtain G_{i-1}. Then each independent set in $S(G_{i-1}) \setminus S(G_i)$, must contain both u and v.



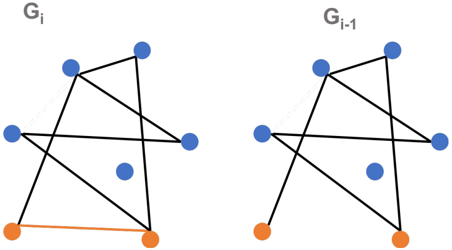
So, we can map each set in $S(G_{i-1}) \setminus S(G_i)$ to a unique set in $S(G_i)$ by simply removing v.

$$\underbrace{|S(G_{i-1}) \setminus S(G_i)|}_{\#\text{ new sets}} \leq \underbrace{|S(G_i)|}_{\#\text{ old sets}}$$

Independent Set Ratios

Consider the ratio $r_i = \frac{|S(G_i)|}{|S(G_{i-1})|}$. Observe that $r_i \leq 1$.

Further, $r_i \geq 1/2$. Let (u, v) be the edge removed from G_i to obtain G_{i-1}. Then each independent set in $S(G_{i-1}) \setminus S(G_i)$, must contain both u and v.



So, we can map each set in $S(G_{i-1}) \setminus S(G_i)$ to a unique set in $S(G_i)$ by simply removing v.

$$\underline{r_i} = \frac{\underline{|S(G_i)|}}{\underline{|S(G_{i-1})|}} = \frac{|S(G_i)|}{\underline{|S(G_i)|} + \underline{|S(G_{i-1}) \setminus S(G_i)|}} \geq \frac{1}{2}.$$

Independent Set Ratios

So Far: We have written $\underline{|S(G)|} = \underline{2^n \cdot \Pi_{i=1}^m r_i}$ where $r_i = \frac{|S(G_i)|}{|S(G_{i-1})|}$. Need to get a $\underline{1 \pm \epsilon/m}$ estimate to each r_i to get a $1 \pm \epsilon$ estimate to $|S(G)|$.

Independent Set Ratios

So Far: We have written $|S(G)| = 2^n \cdot \Pi_{i=1}^{m} r_i$ where $r_i = \frac{|S(G_i)|}{|S(G_{i-1})|}$. Need to get a $1 \pm \epsilon/m$ estimate to each r_i to get a $1 \pm \epsilon$ estimate to $|S(G)|$.

Let X be a random variable generated as follows: pick a random independent set from G_{i-1} and let $X = 1$ if the set is also independent in G_i. Otherwise let $X = 0$.

$$\left[\Pr(X=1) = r_i \right.$$

$$\Pr(X=0) = 1 - r_i$$

Independent Set Ratios

So Far: We have written $|S(G)| = 2^n \cdot \Pi_{i=1}^{m} r_i$ where $r_i = \frac{|S(G_i)|}{|S(G_{i-1})|}$. Need to get a $1 \pm \epsilon/m$ estimate to each r_i to get a $1 \pm \epsilon$ estimate to $|S(G)|$.

Let X be a random variable generated as follows: pick a random independent set from G_{i-1} and let $\mathsf{X} = 1$ if the set is also independent in G_i. Otherwise let $\mathsf{X} = 0$.

What is $\mathbb{E}[\mathsf{X}]$? $= r_i$

Independent Set Ratios

So Far: We have written $|S(G)| = 2^n \cdot \Pi_{i=1}^{m} r_i$ where $r_i = \frac{|S(G_i)|}{|S(G_{i-1})|}$. Need to get a $1 \pm \epsilon/m$ estimate to each r_i to get a $1 \pm \epsilon$ estimate to $|S(G)|$.

Let $\mathbf{X}$ be a random variable generated as follows: pick a random independent set from G_{i-1} and let $\mathbf{X} = 1$ if the set is also independent in G_i. Otherwise let $\mathbf{X} = 0$.

What is $\mathbb{E}[\mathbf{X}]$?

How many samples of $\mathbf{X}$ do we need to take to obtain a $1 \pm \epsilon/m$ approximation to r_i with high probability?

$\mathrm{Var}(x) \leq 1$

$\frac{1}{t} \cdot \sum_{i=1}^{t} X_i \approx \tilde{r}_i$

$\mathrm{Var}(\tilde{r}_i) = \frac{1}{t}$

$\frac{1}{\sqrt{t}} \approx \frac{\varepsilon}{m}$

$\boxed{t = \frac{m^2}{\varepsilon^2}}$

Counting Independent Sets

Upshot: For a graph G with m edges, making $\tilde{O}(m^2/\epsilon^2)$ calls to a uniform random independent set sampler on G or its subgraphs suffices to approximate the number of independent sets in G up to $1 \pm \epsilon$ relative error.

Counting Independent Sets

Upshot: For a graph G with m edges, making $\tilde{O}(m^2/\epsilon^2)$ calls to a uniform random independent set sampler on G or its subgraphs suffices to approximate the number of independent sets in G up to $1 \pm \epsilon$ relative error.

- So a polynomial time algorithm for uniform random independent set sampling, would lead to a polynomial time algorithm for counting independent sets, and hence the collapse of NP to P.

Counting Independent Sets

Upshot: For a graph G with m edges, making $\tilde{O}(m^2/\epsilon^2)$ calls to a uniform random independent set sampler on G or its subgraphs suffices to approximate the number of independent sets in G up to $1 \pm \epsilon$ relative error.

- So a polynomial time algorithm for uniform random independent set sampling, would lead to a polynomial time algorithm for counting independent sets, and hence the collapse of NP to P.
- Observe that near-uniform sampling (as would be obtained e.g., with an MCMC method) would also suffice.