COMPSCI 614: Randomized Algorithms with Applications to Data Science

Prof. Cameron Musco

University of Massachusetts Amherst. Spring 2024.

Lecture 21

Logistics

- I released Problem Set 5 yesterday, due 5/13 at 11:59pm.
- This problem set is optional – it can be used to replace your lowest grade on the first four problem sets.

Summary

Last Time: Markov Chain Mixing Times

- Total variation distance and its dual characterizations.
- Basic results on mixing time.
- Coupling as a technique for bounding mixing time.

  focused on TV distance

Summary

Last Time: Markov Chain Mixing Times

- Total variation distance and its dual characterizations.
- Basic results on mixing time.
- Coupling as a technique for bounding mixing time.

Today: Mixing Time Analysis

- Finish up coupling and example applications.
- Start on algorithmic applications – Markov Chain Monte Carlo (MCMC).

Total Variation Distance

Definition (Total Variation (TV) Distance)

For two distributions $p, q \in [0,1]^m$ over state space $[m]$, the total variation distance is given by:

$$\|p - q\|_{TV} = \frac{1}{2} \sum_{i \in [m]} |p(i) - q(i)| = \max_{A \subseteq [m]} |p(A) - q(A)|.$$

Kontorovich-Rubinstein duality: Let P, Q be possibly correlated random variables with marginal distributions p, q. Then

$$\|p - q\|_{TV} \leq \Pr[\mathsf{P} \neq \mathsf{Q}].$$

$$\|p - q\|_{TV} = \min_{P,Q} \Pr[P \neq Q]$$

This dual notion is the key idea behind mixing time analysis via coupling.

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:

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

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:

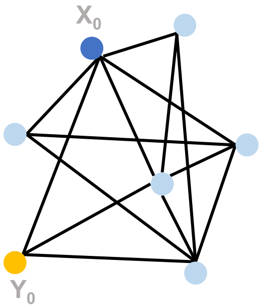
1. $X_0 = i$ and $Y_0 = j$ for some $i, j \in [m]$.
2. $\Pr[X_t = j | X_{t-1} = i] = \Pr[Y_t = j | Y_{t-1} = i] = P_{i,j}$
3. If $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.

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 $X_{t+1} = Y_{t+1}$.

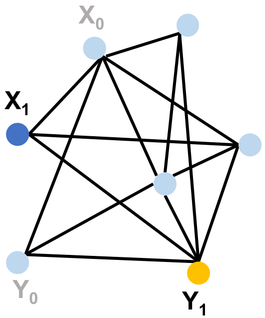


Formal Coupling Definition

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 $X_{t+1} = Y_{t+1}$.

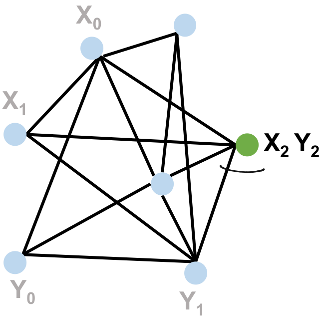


Formal Coupling Definition

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 $X_{t+1} = Y_{t+1}$.

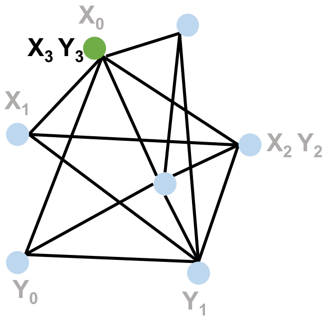


Formal Coupling Definition

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 $X_{t+1} = Y_{t+1}$.

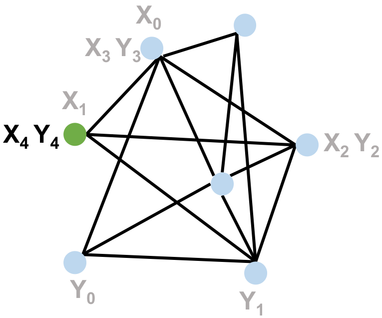


Formal Coupling Definition

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. 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* $T_{i,j} = \min\{t : X_t = Y_t | X_0 = i, Y_0 = j\}$,

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

$\Rightarrow \tau(\varepsilon) \leq t$

characterization in terms of spectrum?

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 $X_0, X_1, \ldots$ 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.

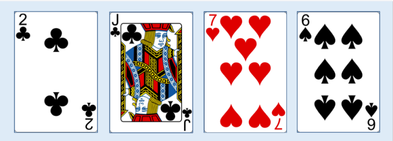
Coupling Example: Mixing Time of Shuffling

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?

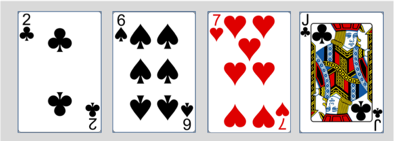
Coupling:

- 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.

$\mathbf{X_0}$



$\mathbf{Y_0}$

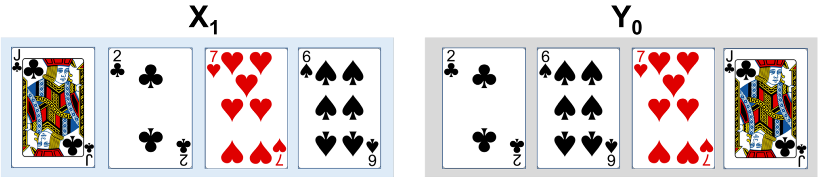


Coupling Example: Mixing Time of Shuffling

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.



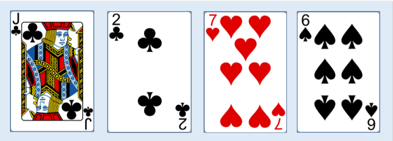
Coupling Example: Mixing Time of Shuffling

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?

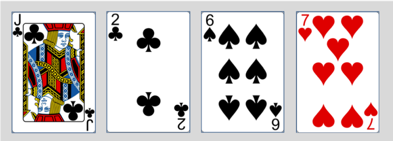
Coupling:

- Let $X_0, X_1, \ldots$ 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.

$\mathbf{X_1}$



$\mathbf{Y_1}$



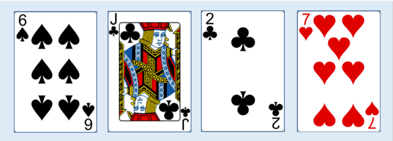
Coupling Example: Mixing Time of Shuffling

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?

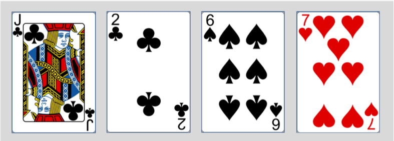
Coupling:

- 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 $\mathbf{X}$ chain, swap S to the top in the $\mathbf{Y}$ chain as well.

$\mathbf{X_2}$



$\mathbf{Y_1}$



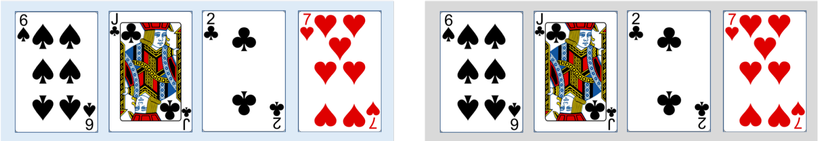
Coupling Example: Mixing Time of Shuffling

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 $\mathbf{X}$ chain, swap S to the top in the $\mathbf{Y}$ chain as well.

$\mathbf{X_2}$ $\mathbf{Y_2}$



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?

Coupling Example 1: Mixing Time of Shuffling

c cards in the deck

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

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

Coupling Example 1: Mixing Time of Shuffling

$$\max_{i \in [m]} \|q_{i,t} - \pi\|_{TV} \leq \max_{i,j \in [m]} \Pr[\mathrm{T}_{i,j} > t]$$
$$\leq \Pr[< c \text{ unique cards are swapped in } t \text{ swaps}]$$

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.

Coupling Example 1: Mixing Time of Shuffling

$$\max_{i\in[m]} \|q_{i,t} - \pi\|_{TV} \leq \max_{i,j\in[m]} \Pr[T_{i,j} > t]$$
$$\leq \Pr[< c \text{ unique cards are swapped in } t \text{ swaps}]$$

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.

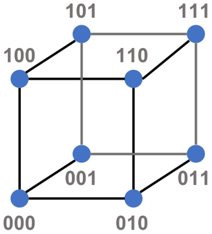
Thus, for $t \geq c\ln(c/\epsilon)$,
$\max_{i\in[m]} \|q_{i,t} - \pi\|_{TV} \leq \max_{i,j\in[m]} \|q_{i,t} - q_{j,t}\|_{TV} \leq \epsilon$.

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

Coupling Example 2: 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.

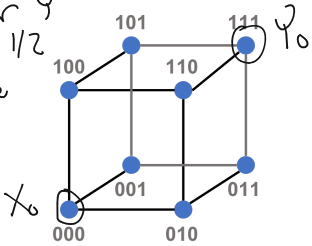
aperiodic?
irreducible? ✓



Coupling Example 2: 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.

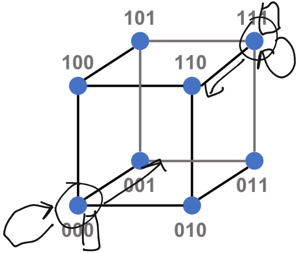
1) have either X or Y stay put w/ prob 1/2
- have other one move closer



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?

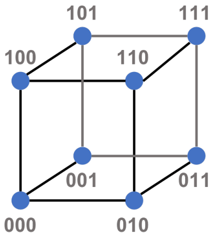
Coupling Example 2: Random Walk on a Hypercube

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.



Coupling Example 2: Random Walk on a Hypercube

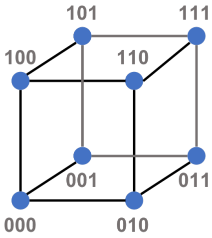
In each step, pick a single random position $i \in [n]$ and let $X_t(i) = Y_t(i) = 0$ with probability 1/2 and $X_t(i) = Y_t(i) = 1$ with probability 1/2.



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

Coupling Example 2: 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 $\tau(\epsilon) = O(n \log(n/\epsilon))$.

Coupling Example 3: Geometric Convergence of TV Distance

Claim: If $X_0, X_1, \ldots$ is finite, irreducible, and aperiodic, then for any $c < 1/2$ and any $\epsilon > 0$, $\tau(\epsilon) \leq \tau(c) \cdot O(\log(1/\epsilon))$.

I.e., it suffices to bound the mixing time for any small constant c and then can boost this result to any $\epsilon > 0$.

Coupling Example 3: Geometric Convergence of TV Distance

Claim: If $X_0, X_1, \ldots$ is finite, irreducible, and aperiodic, then for any $c < 1/2$ and any $\epsilon > 0$, $\tau(\epsilon) \leq \tau(c) \cdot O(\log(1/\epsilon))$.

I.e., it suffices to bound the mixing time for any small constant c and then can boost this result to any $\epsilon > 0$.

$\tau(c) \Rightarrow$ coupling $\Rightarrow \tau(\epsilon)$

Proof:

- After $t = \tau(c)$ steps, for any i we have $\|q_{i,t} - \pi\|_{TV} \leq c$. So, for any i, j we have $\|q_{i,t} - q_{j,t}\|_{TV} \leq 2c < 1$.

Coupling Example 3: Geometric Convergence of TV Distance

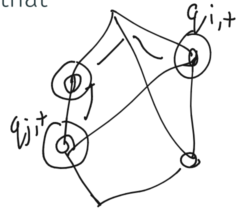
Claim: If $X_0, X_1, \ldots$ is finite, irreducible, and aperiodic, then for any $c < 1/2$ and any $\epsilon > 0$, $\tau(\epsilon) \leq \tau(c) \cdot O(\log(1/\epsilon))$.

I.e., it suffices to bound the mixing time for any small constant c and then can boost this result to any $\epsilon > 0$.

Proof:

- After $t = \tau(c)$ steps, for any i we have $\|q_{i,t} - \pi\|_{TV} \leq c$. So, for any i, j we have $\|q_{i,t} - q_{j,t}\|_{TV} \leq 2c < 1$.
- This implies a coupling between two chains $X_0, X_1, \ldots$ and $Y_0, Y_1, \ldots$ starting in any initial states such that $\Pr[X_t \neq Y_t] \leq 2c < 1$.

$\min \Pr(X_t \neq Y_t) = \|q_{i,t} - q_{j,t}\|_{TV} \leq 2c < 1$



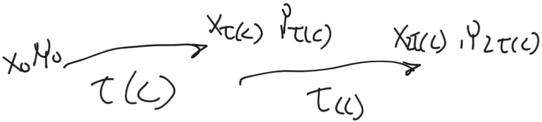
Coupling Example 3: Geometric Convergence of TV Distance

Claim: If $X_0, X_1, \ldots$ is finite, irreducible, and aperiodic, then for any $c < 1/2$ and any $\epsilon > 0$, $\tau(\epsilon) \leq \tau(c) \cdot O(\log(1/\epsilon))$.

I.e., it suffices to bound the mixing time for any small constant c and then can boost this result to any $\epsilon > 0$.

Proof:

- After $t = \tau(c)$ steps, for any i we have $\|q_{i,t} - \pi\|_{TV} \leq c$. So, for any i, j we have $\|q_{i,t} - q_{j,t}\|_{TV} \leq 2c < 1$.
- This implies a coupling between two chains $X_0, X_1, \ldots$ and $Y_0, Y_1, \ldots$ starting in any initial states such that $\Pr[X_t \neq Y_t] \leq 2c < 1$.
- So after $\tau(c) \cdot O(\log(1/\epsilon))$ steps, $\Pr[X_t \neq Y_t] \leq (2c)^{O(\log 1/\epsilon)} \leq \epsilon$



Coupling Example 3: Geometric Convergence of TV Distance

Claim: If $X_0, X_1, \ldots$ is finite, irreducible, and aperiodic, then for any $c < 1/2$ and any $\epsilon > 0$, $\tau(\epsilon) \leq \tau(c) \cdot O(\log(1/\epsilon))$.

I.e., it suffices to bound the mixing time for any small constant c and then can boost this result to any $\epsilon > 0$.

Proof:

- After $t = \tau(c)$ steps, for any i we have $\|q_{i,t} - \pi\|_{TV} \leq c$. So, for any i, j we have $\|q_{i,t} - q_{j,t}\|_{TV} \leq 2c < 1$.
- This implies a coupling between two chains $X_0, X_1, \ldots$ and $Y_0, Y_1, \ldots$ starting in any initial states such that $\Pr[X_t \neq Y_t] \leq 2c < 1$.
- So after $\tau(c) \cdot O(\log(1/\epsilon))$ steps, $\Pr[X_t \neq Y_t] \leq (2c)^{O(\log 1/\epsilon)} \leq \epsilon$
- This establishes that $\tau(\epsilon) \leq \tau(c) \cdot O(\log(1/\epsilon))$.

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.

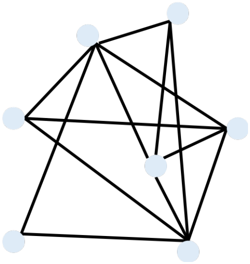
A very common approach is to design a Markov chain whose stationary distribution π is equal to the distribution of interest.

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.

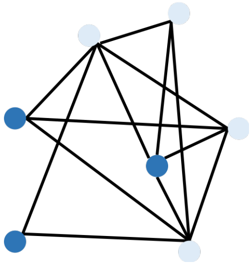
Sampling Independent Sets

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



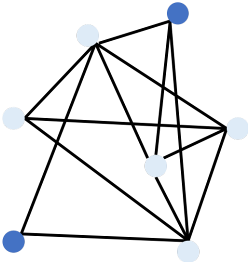
Sampling Independent Sets

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.

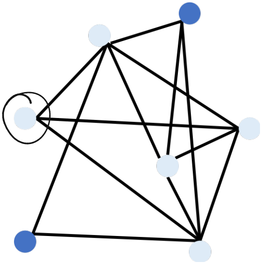


Sampling Independent Sets

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

X_t

$N(X_t)$



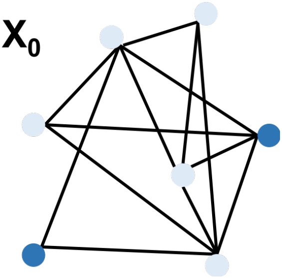
Very non-obvious how to sample from this distribution. Exactly counting the number of independent sets, which is closely related to sampling, is #P-hard.

Markov Chain on Independent Sets

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

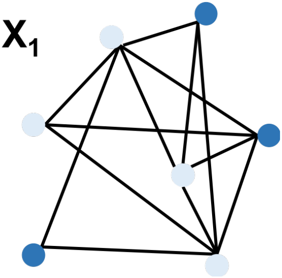
Markov Chain on Independent Sets

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 $\mathcal{N}(X_t) = \{Y : \text{independent set formed by adding/removing a node from } X_t\}$.



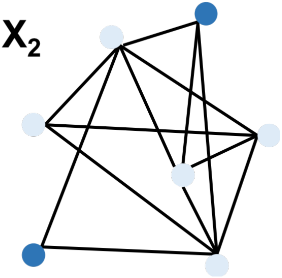
Markov Chain on Independent Sets

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 $\mathcal{N}(X_t) = \{Y : \text{independent set formed by adding/removing a node from } X_t\}$.

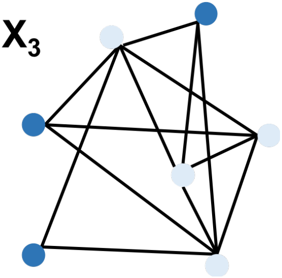
$\cup \; X_t \}$

aperiodic?

irreducible?

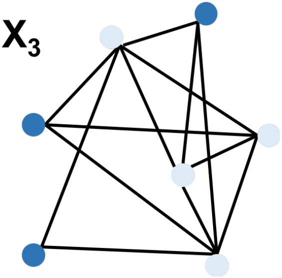
$X \rightleftarrows Y$

$P_{xy} \neq P_{yx}$



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 $\mathcal{N}(X_t) = \{Y : \text{independent set formed by adding/removing a node from } X_t\}$.

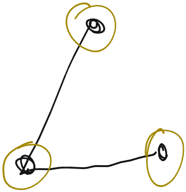


Unfortunately, the stationary distribution of this chain may not be uniform. It places higher probability on independent sets with lots of neighboring independent sets.

Achieving a Uniform Stationary Distribution

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

- Pick 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 $X_{t+1} = X_t$.



C

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?

↳ as long there is at least 1 edge in graph

C

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.

C

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 $P_{i,j}$?

$j \notin N(i)$ $\quad P_{ij} = 0$

$j \in N(i)$ $\quad P_{ij} = \frac{1}{|V|}$

c

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.c

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.c

Thus, the Markov chain is symmetric, so by our claim from two classes ago, 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:

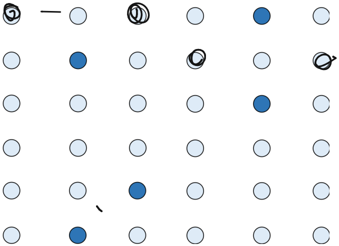
$$\pi(X) = \frac{\lambda^{|X|}}{\sum_{Y \text{ independent}} \lambda^{|Y|}},$$

$e^{\ln \lambda |X|}$

log-linear models
softmax
ising model

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

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



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

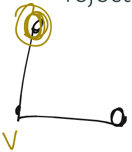
A very generic way of designing a Markov chain over state space $[m]$ with stationary distribution $\pi \in [0,1]^m$.

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

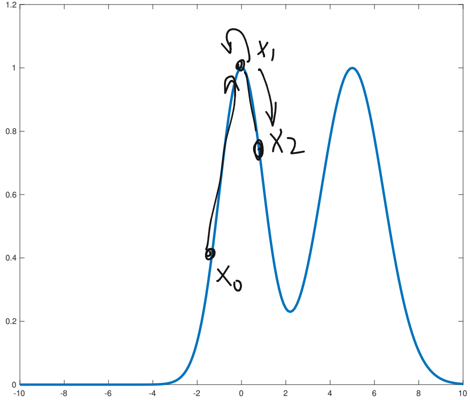
Metropolis-Hastings Algorithm

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



Metropolis-Hastings Analysis

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

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.

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.

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

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.

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

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.

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

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

$$= p(i) \cdot \sum_j Q_{i,j} = p(i).$$

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 $\mathsf{X}_{t+1} = \mathsf{X}_t \setminus \{v\}$
  - 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 $\mathsf{X}_{t+1} = \mathsf{X}_t \setminus \{v\}$
    - 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 $\mathsf{X}_{t+1} = \mathsf{X}_t$
- Need to accept the transition with probability $\min\left(1, \frac{p(\mathsf{X}_{t+1})}{p(\mathsf{X}_t)}\right)$.

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 $\min\left(1, \frac{p(\mathsf{X}_{t+1})}{p(\mathsf{X}_t)}\right)$.

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.