COMPSCI 614: Randomized Algorithms with Applications to Data Science

Prof. Cameron Musco

University of Massachusetts Amherst. Spring 2024.

Lecture 20

Summary

Last Time: Markov Chain Fundamentals

- The gambler's ruin problem.
- Aperiodicity and stationary distribution of a Markov chain.
- The fundamental theorem of Markov chains.
- Example of a uniform stationary distribution for a symmetric Markov chain (shuffling).

Summary

Last Time: Markov Chain Fundamentals

- The gambler's ruin problem.
- Aperiodicity and stationary distribution of a Markov chain.
- The fundamental theorem of Markov chains.
- Example of a uniform stationary distribution for a symmetric Markov chain (shuffling).

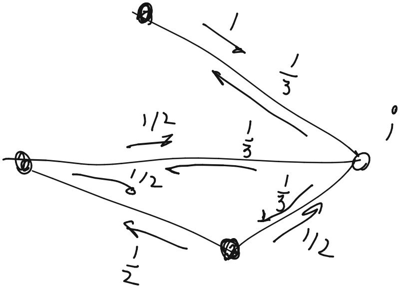
Today: Mixing Time Analysis

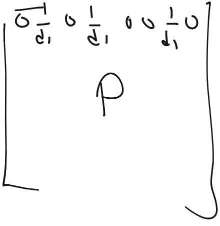
- How quickly does a Markov chain actually converge to its stationary distribution?
- Mixing time and its analysis via coupling.

Stationary Distribution Example 2

Random Walk on an Undirected Graph: Consider a random walk on an undirected graph. If it is at node i at step t, then it moves to any of i's neighbors at step $t+1$ with probability $\frac{1}{d_i}$.

- What is the state space of this chain? nodes of graph
- What is the transition probability $P_{i,j}$? $\frac{1}{d_i}$

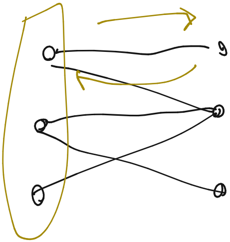




Stationary Distribution Example 2

Random Walk on an Undirected Graph: Consider a random walk on an undirected graph. If it is at node i at step t, then it moves to any of i's neighbors at step $t+1$ with probability $\frac{1}{d_i}$.

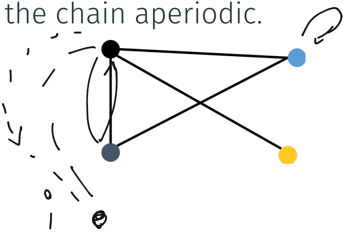
- What is the state space of this chain?
- What is the transition probability $P_{i,j}$?
- Is this chain aperiodic?



Stationary Distribution Example 2

Random Walk on an Undirected Graph: Consider a random walk on an undirected graph. If it is at node i at step t, then it moves to any of i's neighbors at step $t+1$ with probability $\frac{1}{d_i}$.

- What is the state space of this chain?
- What is the transition probability $P_{i,j}$?
- Is this chain aperiodic?
- If the graph is not bipartite, then there is at least one odd cycle, making the chain aperiodic.



Stationary Distribution Example 2

Random Walk on an Undirected Graph: Consider a random walk on an undirected graph. If it is at node i at step t, then it moves to any of i's neighbors at step $t+1$ with probability $\frac{1}{d_i}$.

Claim: When the graph is not bipartite, the unique stationary distribution of this Markov chain is given by $\pi(i) = \frac{d_i}{2|E|}$.

Stationary Distribution Example 2

Random Walk on an Undirected Graph: Consider a random walk on an undirected graph. If it is at node i at step t, then it moves to any of i's neighbors at step $t+1$ with probability $\frac{1}{d_i}$.

Claim: When the graph is not bipartite, the unique stationary distribution of this Markov chain is given by $\pi(i) = \frac{d_i}{2|E|}$.

$$\pi(i) = \pi P_{:,i} = \sum_j \pi(j) P_{j,i}$$

Stationary Distribution Example 2

Random Walk on an Undirected Graph: Consider a random walk on an undirected graph. If it is at node i at step t, then it moves to any of i's neighbors at step $t+1$ with probability $\frac{1}{d_i}$.

Claim: When the graph is not bipartite, the unique stationary distribution of this Markov chain is given by $\pi(i) = \frac{d_i}{2|E|}$.

$$\pi P_{:,i} = \sum_{j=1}^{n} \pi(j) P_{j,i} = \sum_{j \in N(i)} \frac{d_j}{2|E|} \cdot \frac{1}{d_j}$$

Stationary Distribution Example 2

Random Walk on an Undirected Graph: Consider a random walk on an undirected graph. If it is at node i at step t, then it moves to any of i's neighbors at step $t+1$ with probability $\frac{1}{d_i}$.

Claim: When the graph is not bipartite, the unique stationary distribution of this Markov chain is given by $\pi(i) = \frac{d_i}{2|E|}$.

$$\pi P_{:,i} = \sum_{j=1}^{n} \pi(j) P_{j,i} = \sum_{j \in N(i)} \frac{d_j}{2|E|} \cdot \frac{1}{d_j} = \sum_{j \in N(i)} \frac{1}{2|E|}$$

Stationary Distribution Example 2

Random Walk on an Undirected Graph: Consider a random walk on an undirected graph. If it is at node i at step t, then it moves to any of i's neighbors at step $t+1$ with probability $\frac{1}{d_i}$.

Claim: When the graph is not bipartite, the unique stationary distribution of this Markov chain is given by $\pi(i) = \frac{d_i}{2|E|}$.

$$\pi P_{:,i} = \sum_j \pi(j) P_{j,i} = \sum_j \frac{d_j}{2|E|} \cdot \frac{1}{d_j} = \sum_j \frac{1}{2|E|} = \frac{d_i}{2|E|} = \pi(i).$$

$(\sum_{i=1}^{n} d_i$

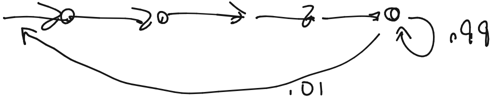
Stationary Distribution Example 2

Random Walk on an Undirected Graph: Consider a random walk on an undirected graph. If it is at node i at step t, then it moves to any of i's neighbors at step $t+1$ with probability $\frac{1}{d_i}$.

Claim: When the graph is not bipartite, the unique stationary distribution of this Markov chain is given by $\pi(i) = \frac{d_i}{2|E|}$.

$$\pi P_{:,i} = \sum_j \pi(j) P_{j,i} = \sum_j \frac{d_j}{2|E|} \cdot \frac{1}{d_j} = \sum_j \frac{1}{2|E|} = \frac{d_i}{2|E|} = \pi(i).$$

I.e., the probability of being at a given node i is dependent only on the node's degree, not on the structure of the graph in any other way.



Random Walk on an Undirected Graph: Consider a random walk on an undirected graph. If it is at node i at step t, then it moves to any of i's neighbors at step $t+1$ with probability $\frac{1}{d_i}$.

Claim: When the graph is not bipartite, the unique stationary distribution of this Markov chain is given by $\pi(i) = \frac{d_i}{2|E|}$.

$$\pi P_{:,i} = \sum_j \pi(j) P_{j,i} = \sum_j \frac{d_j}{2|E|} \cdot \frac{1}{d_j} = \sum_j \frac{1}{2|E|} = \frac{d_i}{2|E|} = \pi(i).$$

I.e., the probability of being at a given node i is dependent only on the node's degree, not on the structure of the graph in any other way.

What is the stationary distribution over the edges? $\frac{1}{|E|}$

Mixing Times

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:

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

$\frac{1}{2} \|p - q\|_1$

$$p = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix} \quad q = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$

$\|p - q\|_{TV} = \frac{1}{2}\left(\frac{1}{6} + \frac{1}{6} + \frac{1}{3}\right) = 1/3$

A = event of being in state 3

A^c = event . . . state 1 or 2

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

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

$p = q$

P? Q?

P = Q

$= \min_{P,Q} Pr(P \neq Q)$

$p: \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$ $q: \begin{bmatrix} 3/4 \\ 1/4 \end{bmatrix}$

P? Q?

$Pr(P \neq Q) = 1/4 = \|p-q\|_{TV}$

If P = H then Q = H

P = T then Q = H w.p. 1/2, T w.p. 1/2

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:

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

I.e., what is the maximum time it takes the Markov chain to converge to within ϵ in TV distance of the stationary distribution?

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:

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

I.e., what is the maximum time it takes the Markov chain to converge to within ϵ in TV distance of the stationary distribution?

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

$q_{i,t} = \pi + e_t$

$\|e_t\| = \|q_{i,t} - \pi\|_1 \cdot 2 \cdot TV$

$q_{i,t+1} = q_{i,t} P = \pi P + e_t P$

$= \pi + e_t P$

$e_{t+1} = e_t P$

$\|e_{t+1}\|_1 \leq \|e_t\|_1$

(by Δ inequality)

Mixing Time Convergence

Typically, it suffices to focus on the mixing time for $\epsilon = 1/2$. We have:

Claim: If $X_0, X_1, \ldots$ is finite, irreducible, and aperiodic, then $\tau(\epsilon) \leq \tau(1/2) \cdot c \log(1/\epsilon)$ for large enough constant c.

$q_{i,t} = \pi + e_t$

$q_{i,2t} = \pi + e_t P^t$

↳ $\pi + e_{t+1}$

I'll do this later

Coupling Motivation

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

Coupling Motivation

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

$$\|q_{i,t} - \pi\|_{TV} = \|q_{i,t} - \pi P^t\|_{TV}$$

Coupling Motivation

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

$$\begin{aligned}\|q_{i,t} - \pi\|_{TV} &= \|q_{i,t} - \pi P^t\|_{TV} \\ &= \|q_{i,t} - \sum_j \pi(j) e_j P^t\|_{TV}\end{aligned}$$

$e_j P^t = q_{j,t}$

$$[\pi(1) \ \ldots \ \pi(m)]$$

$$= \pi(1) \cdot [1\ 0\ 0\ 0\ 0] + \pi(2)\, [0\ 1\ 0\ 0] + \ldots$$

Coupling Motivation

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

$$\begin{aligned}
\|q_{i,t} - \pi\|_{TV} &= \|q_{i,t} - \pi P^t\|_{TV} \\
&= \|q_{i,t} - \sum_j \pi(j) e_j P^t\|_{TV} \\
&= \|q_{i,t} - \sum_j \pi(j) q_{j,t}\|_{TV}
\end{aligned}$$

Coupling Motivation

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

$$
\begin{aligned}
\|q_{i,t} - \pi\|_{TV} &= \|q_{i,t} - \pi P^t\|_{TV} \\
&= \|q_{i,t} - \sum_j \pi(j) e_j P^t\|_{TV} \\
&= \|q_{i,t} - \sum_j \pi(j) q_{j,t}\|_{TV} \\
&\leq \sum_j \|\pi(j) q_{i,t} - \pi(j) q_{j,t}\|_{TV}
\end{aligned}
$$

Coupling Motivation

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

$$\begin{aligned}
\|q_{i,t} - \pi\|_{TV} &= \|q_{i,t} - \pi P^t\|_{TV} \\
&= \|q_{i,t} - \sum_j \pi(j) e_j P^t\|_{TV} \\
&= \|q_{i,t} - \sum_j \pi(j) q_{j,t}\|_{TV} \\
&\leq \sum_j \|\pi(j) q_{i,t} - \pi(j) q_{j,t}\|_{TV} \\
&\leq \sum_j \pi(j) \cdot \|q_{i,t} - q_{j,t}\|_{TV}
\end{aligned}$$

Coupling Motivation

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

$$
\begin{aligned}
\|q_{i,t} - \pi\|_{TV} &= \|q_{i,t} - \pi P^t\|_{TV} \\
&= \|q_{i,t} - \sum_j \pi(j) e_j P^t\|_{TV} \\
&= \|q_{i,t} - \sum_j \pi(j) q_{j,t}\|_{TV} \\
&\le \sum_j \|\pi(j) q_{i,t} - \pi(j) q_{j,t}\|_{TV} \\
&\le \sum_j \pi(j) \cdot \|q_{i,t} - q_{j,t}\|_{TV} \\
&\le \max_{j\in[m]} \|q_{i,t} - q_{j,t}\|_{TV}.
\end{aligned}
$$

$\pi = \sum \pi(j) e_j$

$[1 \quad 2 \quad 4] = 1\cdot[1,0,0] + 2\cdot[0,1,0] + 4\cdot[0,0,1]$

$= \|\sum_j \pi(j) q_{i,t} - \sum_j \pi(j) q_{j,t}\|$

use that $\sum \pi(j) = 1$

~~$\|q_{i,t} - \pi\| \le \|q_{i,t} - q_{j,t}\| + \|q_{j,t} - \pi\|$~~

Coupling Motivation

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

$$
\begin{aligned}
\|q_{i,t} - \pi\|_{TV} &= \|q_{i,t} - \pi P^t\|_{TV} \\
&= \|q_{i,t} - \sum_j \pi(j) e_j P^t\|_{TV} \\
&= \|q_{i,t} - \sum_j \pi(j) q_{j,t}\|_{TV} \\
&\leq \sum_j \|\pi(j) q_{i,t} - \pi(j) q_{j,t}\|_{TV} \\
&\leq \sum_j \pi(j) \cdot \|q_{i,t} - q_{j,t}\|_{TV} \\
&\leq \max_{j\in[m]} \|q_{i,t} - q_{j,t}\|_{TV}.
\end{aligned}
$$

Coupling: A common technique for bounding the mixing time by showing that $\max_{i,j\in[m]} \|q_{i,t} - q_{j,t}\|_{TV}$ is small.

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:

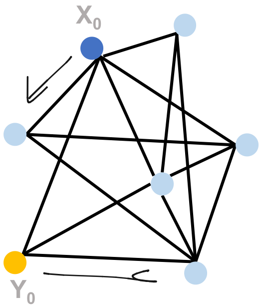
1. $X_0 = i$ and $Y_0 = j$ for some $i, j \in [m]$.
2. $\Pr[X_t = j | X_{t-1} = i] = \Pr[Y_t = j | Y_{t-1} = i] = P_{i,j}$
3. If $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.

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

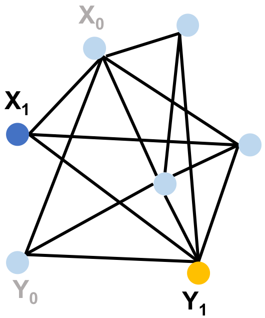


Formal Coupling Definition

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

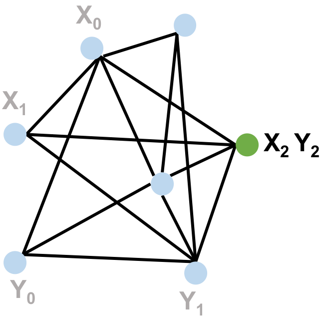


Formal Coupling Definition

Definition (Coupling)

For a finite Markov chain $X_0, X_1, \ldots$ 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)

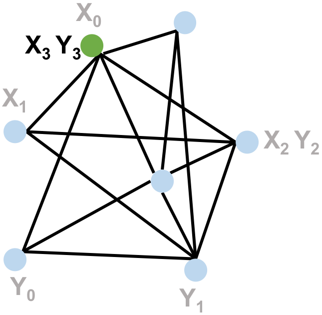
For a finite Markov chain $X_0, X_1, \ldots$ with transition matrix $P \in \mathbb{R}^{m \times m}$, a coupling is a joint process $(X_0, Y_0), (X_1, Y_1), \ldots$ such that:

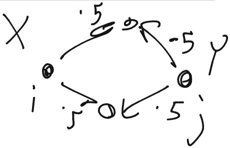
1. $X_0 = i$ and $Y_0 = j$ for some $i, j \in [m]$.
2. $\Pr[X_t = j | X_{t-1} = i] = \Pr[Y_t = j | Y_{t-1} = i] = P_{i,j}$
3. If $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.

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

$\Pr(X_{t+1} = j \mid X_t = i) = P_{ij}$

⋮



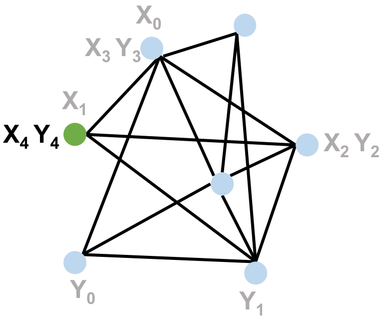


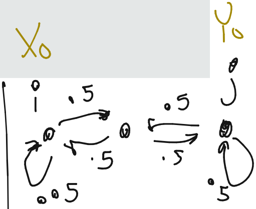
Formal Coupling Definition

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





$Pr(T_{ij}=1) = 1/2$

$Pr(T_{ij}=2) = 1/4$

⋮

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

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

Coupling Theorem Proof

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

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

Follows from Kontorovich-Rubinstein duality.

Coupling Theorem Proof

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

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

Follows from Kontorovich-Rubinstein duality.

For X_t, Y_t distributed by evolving the chain for t steps starting from state i or j respectively, we have:

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

X Y

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:

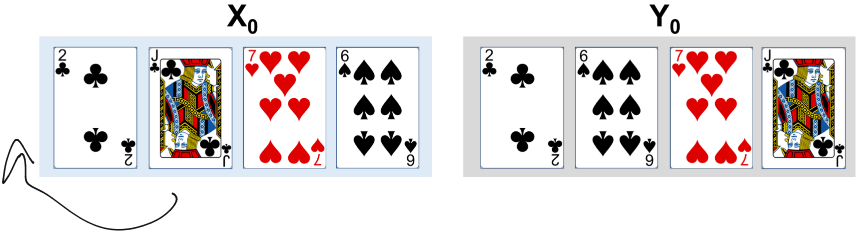
- Let $X_0, X_1, \ldots$ be the Markov chain where a random card is moved to the top in each step.
- Let Y_0, Y_1 be a correlated Markov chain. When card S is swapped to the top in the X chain, swap S to the top in the Y chain as well.

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.

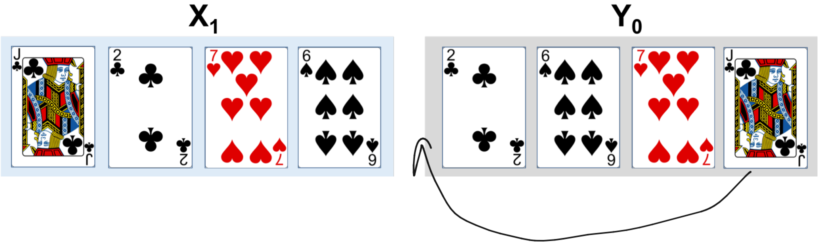


Coupling Example: Mixing Time of Shuffling

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.

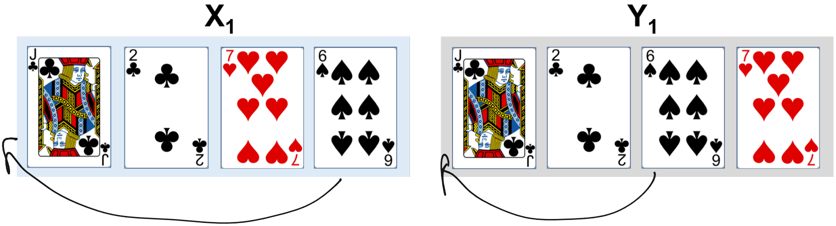


Coupling Example: Mixing Time of Shuffling

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.

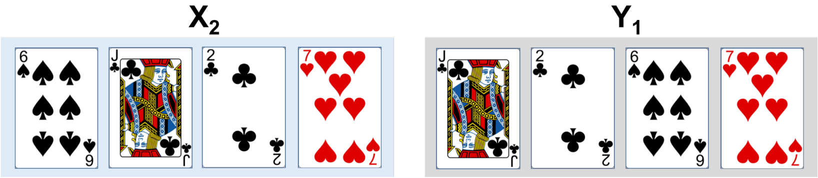


Coupling Example: Mixing Time of Shuffling

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.

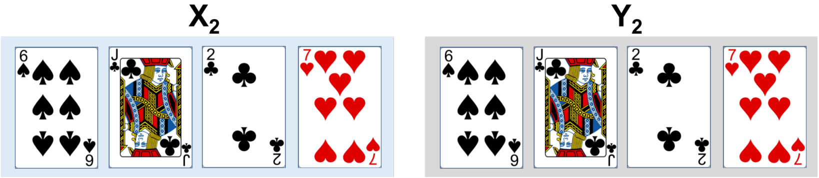


Coupling Example: Mixing Time of Shuffling

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.



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: Mixing Time of Shuffling

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

Coupling Example: Mixing Time of Shuffling

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

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.

Coupling Example: 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.

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