The coupling method - Simons Counting Complexity Bootcamp, 2016

Nayantara Bhatnagar (University of Delaware)
Ivona Bezáková (Rochester Institute of Technology)

January 26, 2016

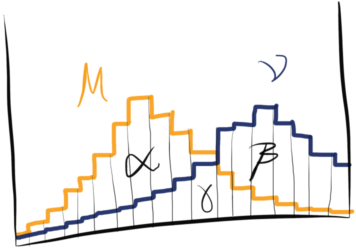
Techniques for bounding the mixing time

- Probabilistic techniques - coupling, martingales, strong stationary times, coupling from the past.
- Eigenvalues and eigenfunctions.
- Functional, isoperimteric and geometric inequalities - Cheeger's inequality, conductance, Poincaré and Nash inequalities, discrete curvature.
- (Levin-Peres-Wilmer 2009, Aldous-Fill 1999/2014) have comprehensive accounts.

In this part of the tutorial

- Coupling distributions
- Coupling Markov chains
- Path Coupling
- Exact sampling - coupling from the past

Definition - Total variation distance



$1 - \gamma = \alpha = \beta = \|\mu - \nu\|_{tv}$

Let μ and ν be probability measures on the same measurable space $(\Omega, \mathcal{F})$. The **total variation distance** *between μ and ν is given by*

$$\|\mu - \nu\|_{tv} = \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$$

Here Ω is finite and $\mathcal{F} = 2^{\Omega}$

$$\|\mu - \nu\|_{tv} = \max_{A \subset \Omega} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|$$

Definition - Coupling of distributions

(Doeblin 1938)

Let μ *and* ν *be probability measures on the same measurable space* $(\Omega, \mathcal{F})$*. A* **coupling** *of* μ *and* ν *is a pair of random variables* (X, Y) *on the probability space* $(\Omega \times \Omega, \mathcal{F} \times \mathcal{F}, \mathbb{P})$ *such that the marginals coincide*

$$\mathbb{P}(X \in A) = \mu(A), \quad \mathbb{P}(Y \in A) = \nu(A), \quad \forall A \in \mathcal{F}.$$

Example - Biased coins

μ: *Bernoulli*(p), probability p of "heads" $(= 1)$.
ν: *Bernoulli*(q), probability $q > p$ of heads.

$H_\mu(n) =$ no. of heads in n tosses.
$H_\nu(n) =$ no. of heads in n tosses.

Proposition 1. $\mathbb{P}(H_\mu(n) > k) \leq \mathbb{P}(H_\nu(n) > k)$

Example - Biased coins

μ: *Bernoulli*(p), probability p of "heads" $(= 1)$.
ν: *Bernoulli*(q), probability $q > p$ of heads.

$H_\mu(n) =$ no. of heads in n tosses.
$H_\nu(n) =$ no. of heads in n tosses.

Proposition 1. $\mathbb{P}(H_\mu(n) > k) \leq \mathbb{P}(H_\nu(n) > k)$

Proof.
For $1 \leq i \leq n$, (X_i, Y_i) a coupling of μ and ν. Indep. (X_i, Y_i).

- Let $X_i \sim \mu$.
- If $X_i = 1$, set $Y_i = 1$.
- If $X_i = 0$, set $Y_i = 1$ w.p. $\frac{q-p}{1-p}$ and 0 otherwise.

$$\mathbb{P}(X_1 + \ldots + X_n > k) \leq \mathbb{P}(Y_1 + \ldots + Y_n > k)$$

$\square$

Coupling and Total Variation Distance

Lemma 2. Let μ and ν be distributions on Ω. Then

$$\|\mu - \nu\|_{tv} \leq \inf_{\text{couplings } (X,Y)} \{\mathbb{P}(X \neq Y)\}$$

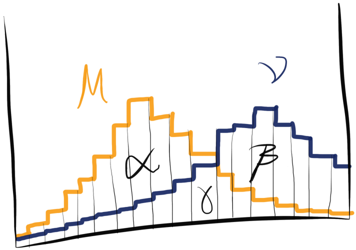
Proof.
For any coupling (X, Y) and $A \subset \Omega$

$$\begin{aligned}
\mu(A) - \nu(A) &= \mathbb{P}(X \in A) - \mathbb{P}(Y \in A) \\
&\leq \mathbb{P}(X \in A, Y \notin A) \\
&\leq \mathbb{P}(X \neq Y).
\end{aligned}$$

Similarly, $\nu(A) - \mu(A) \leq \mathbb{P}(X \neq Y)$.

Therefore, $\|\mu - \nu\|_{tv} \leq \mathbb{P}(X \neq Y)$. $\square$

Maximal coupling



$1 - \gamma = \alpha = \beta = \|\mu - \nu\|_{tv}$

Lemma 3. Let μ and ν be distributions on Ω. Then

$$\|\mu - \nu\|_{tv} = \inf_{\text{couplings } (X,Y)} \{\mathbb{P}(X \neq Y)\}$$

Coupling Markov chains

(Pitman 1974, Griffeath 1974/5, Aldous 1983)

A **coupling of Markov chains** with transition matrix P is a process $(X_t, Y_t)_{t=0}^{\infty}$ such that both (X_t) and (Y_t) are Markov chains with transition matrix P.

Applied to bounding the rate of convergence to stationarity of MC's for sampling tilings of lattice regions, particle processes, card shuffling, random walks on lattices and other natural graphs, Ising and Potts models, colorings, independent sets...

(Aldous-Diaconis 1986, Bayer-Diaconis 1992, Bubley-Dyer 1997, Luby-Vigoda 1999, Luby-Randall-Sinclair 2001, Vigoda 2001, Wilson 2004, ...)

Simple random walk on $\{0, 1 \ldots, n\}$

SRW:

- Move up or down with probability $\frac{1}{2}$ if possible.
- Do nothing if attempt to move outside interval.

Claim 4. If $0 \leq x \leq y \leq n$, $P^t(y, 0) \leq P^t(x, 0)$.

Simple random walk on $\{0, 1 \ldots, n\}$

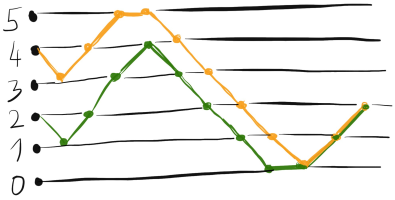
SRW:

- Move up or down with probability $\frac{1}{2}$ if possible.
- Do nothing if attempt to move outside interval.

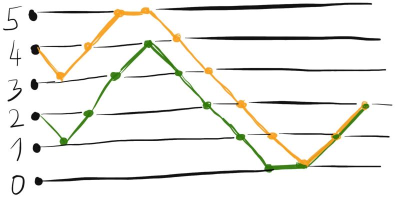
Claim 4. If $0 \leq x \leq y \leq n$, $P^t(y, 0) \leq P^t(x, 0)$.

A coupling (X_t, Y_t) of $P^t(x, \cdot)$ and $P^t(y, \cdot)$:

- $X_0 = x$, $Y_0 = y$.
- Let $b_1, b_2 \ldots$ be i.i.d. $\{\pm 1\}$-valued *Bernoulli*$(1/2)$.
- At the ith step, attempt to add b_i to both X_{i-1} and Y_{i-1}.



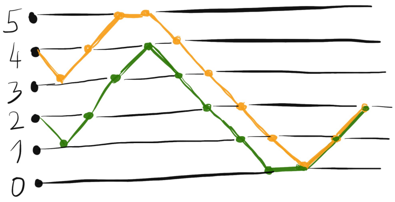
Simple random walk on $\{0, 1 \ldots, n\}$



For all t, $X_t \leq Y_t$. Therefore,

$$P^t(y, 0) = \mathbb{P}(Y_t = 0) \leq \mathbb{P}(X_t = 0) = P^t(x, 0).$$

Simple random walk on $\{0, 1 \ldots, n\}$



For all t, $X_t \leq Y_t$. Therefore,

$$P^t(y, 0) = \mathbb{P}(Y_t = 0) \leq \mathbb{P}(X_t = 0) = P^t(x, 0).$$

Note: Can modify any coupling so that the chains stay together after the first time they meet.

We'll assume this to be the case.

Distance to stationarity and the mixing time

Define

$$d(t) := \max_{x \in \Omega} \|P^t(x,\cdot) - \pi\|_{tv}, \qquad \bar{d}(t) := \max_{x,y \in \Omega} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{tv}$$

By the triangle inequality, $d(t) \leq \bar{d}(t) \leq 2d(t)$.

The mixing time is

$$t_{mix}(\varepsilon) := \min\{t : d(t) \leq \varepsilon\}$$

It's standard to work with $t_{mix} := t_{mix}(1/4)$ (any constant $\varepsilon < 1/2$) because

$$t_{mix}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil t_{mix}.$$

Bound on the mixing time

Theorem 5. Let (X_t, Y_t) be a coupling with $X_0 = x$ and $Y_0 = y$. Let τ^* be the first time they meet (and thereafter, coincide). Then

$$\|P^t(x,\cdot) - P^t(y,\cdot)\|_{tv} \le \mathbb{P}_{x,y}(\tau^* > t)$$

Proof.
By Lemma 2,

$$\|P^t(x,\cdot) - P^t(y,\cdot)\|_{tv} \le \mathbb{P}(X_t \neq Y_t) = \mathbb{P}_{x,y}(\tau^* > t) \qquad \square$$

Bound on the mixing time

Theorem 5. Let (X_t, Y_t) be a coupling with $X_0 = x$ and $Y_0 = y$. Let τ^* be the first time they meet (and thereafter, coincide). Then

$$\|P^t(x,\cdot) - P^t(y,\cdot)\|_{tv} \leq \mathbb{P}_{x,y}(\tau^* > t)$$

Proof.
By Lemma 2,

$$\|P^t(x,\cdot) - P^t(y,\cdot)\|_{tv} \leq \mathbb{P}(X_t \neq Y_t) = \mathbb{P}_{x,y}(\tau^* > t) \qquad \square$$

Corollary 6. $t_{mix} \leq 4 \max_{x,y} \mathbb{E}_{x,y}(\tau^*)$.

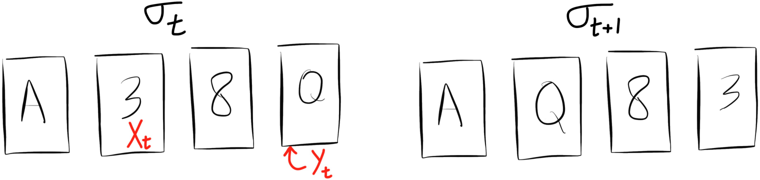
Proof.

$$d(t) \leq \overline{d}(t) \leq \max_{x,y} \mathbb{P}_{x,y}(\tau^* > t) \leq \max_{x,y} \frac{\mathbb{E}_{x,y}(\tau^*)}{t}$$

Example - Card shuffling by random transpositions

Shuffling MC on $\Omega = S_n$:

- Choose card X_t and an independent position Y_t uniformly.
- Exchange X_t with $\sigma_t(Y_t)$ (the card at Y_t).



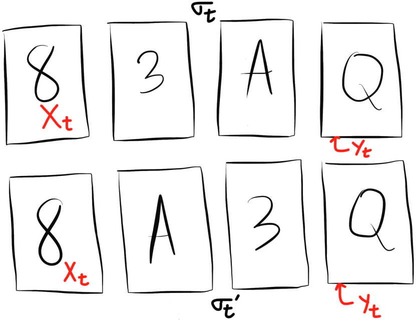
Stationary distribution is uniform over permutations S_n.

Example - Card shuffling by random transpositions

Coupling of σ_t, σ'_t:

- Choose card X_t and independent position Y_t uniformly.
- Use X_t and Y_t to update both σ_t and σ'_t

Let M_t = number of cards at the same position in σ and σ'.

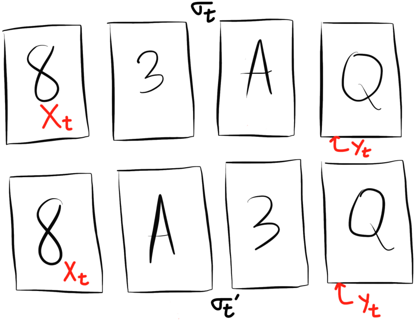


Example - Card shuffling by random transpositions

Coupling of σ_t, σ'_t:

- Choose card X_t and independent position Y_t uniformly.
- Use X_t and Y_t to update both σ_t and σ'_t

Let M_t = number of cards at the same position in σ and σ'.



Case 1:

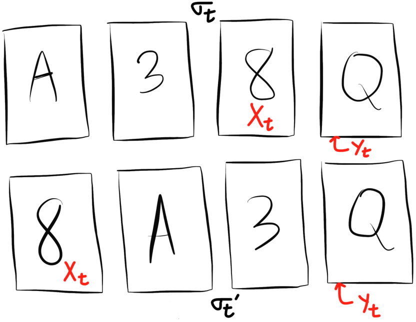
- X_t in same position
- $M_{t+1} = M_t$.

Example - Card shuffling by random transpositions

Coupling of σ_t, σ'_t:

- Choose card X_t and independent position Y_t uniformly.
- Use X_t and Y_t to update both σ_t and σ'_t

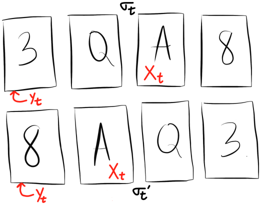
Let M_t = number of cards at the same position in σ and σ'.



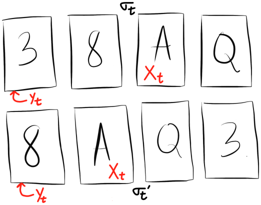
Case 2:

- X_t in different pos.
- $\sigma(Y_t) = \sigma'(Y_t)$.
- $M_{t+1} = M_t$.

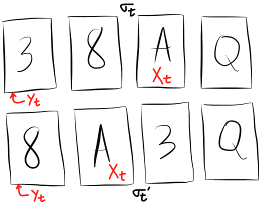
Example - Card shuffling by random transpositions



$M_{t+1} = M_t + 1$



$M_{t+1} = M_t + 2$



$M_{t+1} = M_t + 3$

Case 3:

- X_t in different pos.
- $\sigma(Y_t) \neq \sigma'(Y_t)$.
- $M_{t+1} > M_t$.

Example - Card shuffling by random transpositions

Proposition 7 (Broder). Let τ^* be the first time $M_t = n$. For any x, y,

$$\mathbb{E}_{x,y}(\tau^*) < \frac{\pi^2}{6} n^2, \quad t_{mix} = O(n^2).$$

Proof.
Let $\tau_i =$ steps to increase M_t from $i-1$ to i so

$$\tau^* = \tau_1 + \tau_2 + \cdots + \tau_n.$$

$$\mathbb{P}(M_{t+1} > M_t \mid M_t = i) = \frac{(n-i)^2}{n^2} \Rightarrow \mathbb{E}(\tau_{i+1} | M_t = i) = \frac{n^2}{(n-i)^2}.$$

Therefore, for any x, y

$$\mathbb{E}_{x,y}(\tau^*) \leq n^2 \sum_{i=0}^{n-1} \frac{1}{(n-i)^2} < \frac{\pi^2}{6} n^2.$$

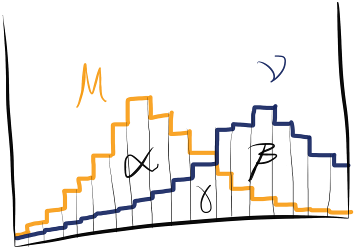
□

In this part of the tutorial

- Coupling distributions
- Coupling Markov chains
- Coloring MC
- Path Coupling
- Exact sampling - coupling from the past

A few remarks on yesterday's talk

- Shuffling by random transpositions has $t_{mix} \leq O(n \log n)$. Strong stationary times - random times at which the MC is guaranteed to be at stationarity.
- Maximal/optimal coupling

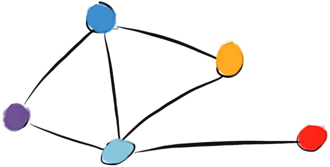


$1 - \gamma = \alpha = \beta = \|\mu - \nu\|_{tv}$

Lemma 3. Let μ and ν be distributions on Ω. Then

$$\|\mu - \nu\|_{tv} = \inf_{\text{couplings } (X,Y)} \{\mathbb{P}(X \neq Y)\}$$

Sampling Colorings



$[q] = \{$ ● , ● , ● , ● , ● $\}$

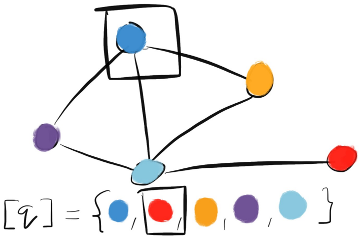
Graph $G = (V, E)$. Ω set of proper colorings of G.

Metropolis MC:

- Select $v \in V$ and $k \in [q]$ uniformly.
- If k is allowed at v, update.

Stationary distribution: uniform over colorings Ω $(q > \Delta + 1)$.

Sampling Colorings



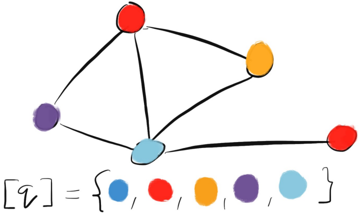
Graph $G = (V, E)$. Ω set of proper colorings of G.

Metropolis MC:

- Select $v \in V$ and $k \in [q]$ uniformly.
- If k is allowed at v, update.

Stationary distribution: uniform over colorings Ω $(q > \Delta + 1)$.

Sampling Colorings



Graph $G = (V, E)$. Ω set of proper colorings of G.

Metropolis MC:

- Select $v \in V$ and $k \in [q]$ uniformly.
- If k is allowed at v, update.

Stationary distribution: uniform over colorings Ω ($q > \Delta + 1$).

A coupling for colorings

Recall,

Theorem 8. Let (X_t, Y_t) be a coupling with $X_0 = x$ and $Y_0 = y$. Let τ^* be the first time they meet (and thereafter, coincide). Then

$$d(t) \leq \max_{x,y} \mathbb{P}_{x,y}(\tau^* > t).$$

Theorem 9. Let G have max. degree Δ. Let $q > 4\Delta$. Then,

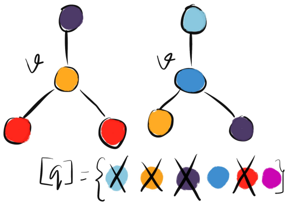
$$t_{mix}(\varepsilon) \leq \left\lceil \frac{1}{1 - 4\Delta/q} n(\log n + \log(\varepsilon^{-1})) \right\rceil.$$

Coupling:

- Generate a vertex-color pair (v, k) uniformly.
- Update the colorings X_t and Y_t by recoloring v with k if it is allowed.

Case analysis

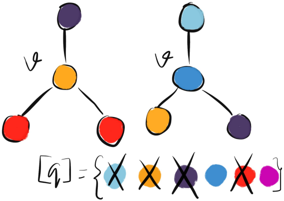
For colorings X_t, Y_t, let $D_t := |\{v \mid X_t(v) \neq Y_t(v)\}|$.

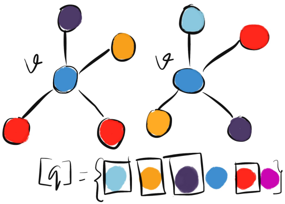


$$D_{t+1} = D_t - 1 \text{ w.p. } \geq \frac{D_t(q-2\Delta)}{qn}$$

Case analysis

For colorings X_t, Y_t, let $D_t := |\{v \mid X_t(v) \neq Y_t(v)\}|$.



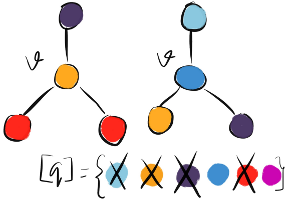


$D_{t+1} = D_t - 1$ w.p. $\geq \frac{D_t(q-2\Delta)}{qn}$

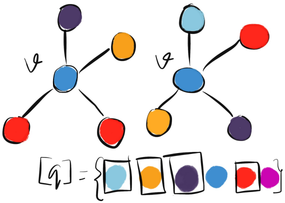
$D_{t+1} = D_t + 1$ w.p. $\leq \frac{D_t(2\Delta)}{qn}$

Case analysis

For colorings X_t, Y_t, let $D_t := |\{v \mid X_t(v) \neq Y_t(v)\}|$.



$$D_{t+1} = D_t - 1 \text{ w.p. } \geq \frac{D_t(q-2\Delta)}{qn}$$

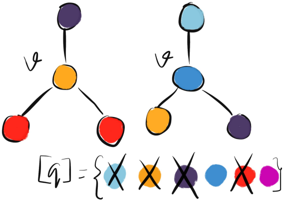


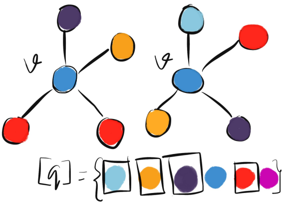
$$D_{t+1} = D_t + 1 \text{ w.p. } \leq \frac{D_t(2\Delta)}{qn}$$

$$\mathbb{E}(D_{t+1}|X_t, Y_t) \leq D_t - \frac{D_t(q-2\Delta)}{qn} + \frac{2D_t\Delta}{qn} = D_t\left(1 - \frac{q-4\Delta}{qn}\right)$$

Case analysis

For colorings X_t, Y_t, let $D_t := |\{v \mid X_t(v) \neq Y_t(v)\}|$.





$D_{t+1} = D_t - 1$ w.p. $\geq \frac{D_t(q-2\Delta)}{qn}$ $\qquad$ $D_{t+1} = D_t + 1$ w.p. $\leq \frac{D_t(2\Delta)}{qn}$

$$\mathbb{E}(D_{t+1}|X_t, Y_t) \leq D_t - \frac{D_t(q-2\Delta)}{qn} + \frac{2D_t\Delta}{qn} = D_t\left(1 - \frac{q-4\Delta}{qn}\right)$$

Iterating, $\mathbb{E}(D_t|X_0, Y_0) \leq D_0\left(1 - \frac{q-4\Delta}{qn}\right)^t \leq n\left(1 - \frac{q-4\Delta}{qn}\right)^t$

Mixing time

We've shown

$$\mathbb{E}(D_t | X_0 = x, Y_0 = y) \le n \left(1 - \frac{q - 4\Delta}{qn}\right)^t \le n \exp\left(-\frac{q - 4\Delta}{q} \frac{t}{n}\right).$$

By Theorem 8,

$$\begin{aligned} d(t) \le \max_{x,y} \mathbb{P}_{x,y}(\tau^* > t) &= \max_{x,y} \mathbb{P}(D_t > 1 | X_0 = x, Y_0 = y) \\ &\le \max_{x,y} \mathbb{E}(D_t | X_0 = x, Y_0 = y) \end{aligned}$$

Therefore,

$$t_{mix}(\varepsilon) \le \left\lceil \frac{1}{1 - 4\Delta/q} n(\log n + \log(\varepsilon^{-1})) \right\rceil.$$

Fast mixing for $q > 2\Delta$

(Jerrum 1995, Salas-Sokal 1997)

Theorem 10. Let G have max. degree Δ. If $q > 2\Delta$, the mixing time of the Metropolis chain on colorings is

$$t_{mix}(\varepsilon) \leq \left\lceil \left(\frac{q}{q - 2\Delta} \right) n(\log n + \log(\varepsilon^{-1})) \right\rceil$$

- Use path metrics on Ω to couple only colorings with a single difference and simplify the proof.
- Coupling colorings with just one difference in a smarter way.

Contraction in D_t

Recall, we showed contraction in one step: for some $\alpha > 0$

$$\mathbb{E}(D_{t+1} \mid X_t, Y_t) \leq D_t e^{-\alpha} \quad \Rightarrow \quad d(t) \leq n e^{-\alpha t}$$

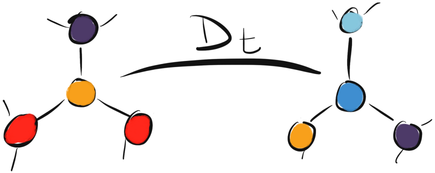
$$\Rightarrow \quad t_{mix}(\varepsilon) \leq \left\lceil \frac{\log n + \log \varepsilon^{-1}}{\alpha} \right\rceil$$

Contraction in D_t

Recall, we showed contraction in one step: for some $\alpha > 0$

$$\mathbb{E}(D_{t+1} \mid X_t, Y_t) \leq D_t e^{-\alpha} \quad \Rightarrow \quad d(t) \leq n e^{-\alpha t}$$

$$\Rightarrow \quad t_{mix}(\varepsilon) \leq \left\lceil \frac{\log n + \log \varepsilon^{-1}}{\alpha} \right\rceil$$

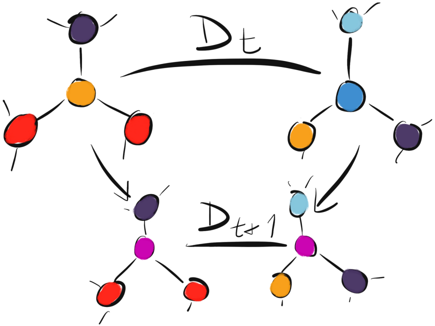


Contraction in D_t

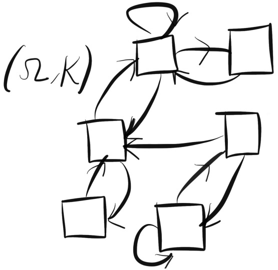
Recall, we showed contraction in one step: for some $\alpha > 0$

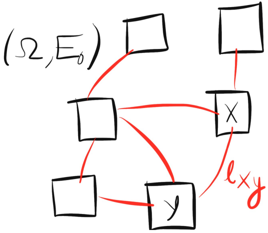
$$\mathbb{E}(D_{t+1} \mid X_t, Y_t) \leq D_t e^{-\alpha} \quad \Rightarrow \quad d(t) \leq ne^{-\alpha t}$$

$$\Rightarrow \quad t_{mix}(\varepsilon) \leq \left\lceil \frac{\log n + \log \varepsilon^{-1}}{\alpha} \right\rceil$$



Path Coupling (Bubley-Dyer 1997)





Connected graph (Ω, E_0) on Ω.

Length function $\ell : E_0 \to [1, \infty)$.

A **path** from x_0 to x_r is $\xi = (x_0, x_1, \ldots, x_r)$, and $(x_{i-1}, x_i) \in E_0$.

Length of path ξ is $\ell(\xi) := \sum_{i=1}^{r} \ell((x_{i-1}, x_i))$.

Path metric on Ω is

$$\rho(x, y) = \min\{\ell(\xi) \mid \xi \text{ is a path between } x, y\}$$

Path Coupling (Bubley-Dyer 1997)

Theorem 11 (Bubley-Dyer '97). If for each $(x, y) \in E_0$ there is a coupling (X_1, Y_1) of $P(x, \cdot)$ and $P(y, \cdot)$ so for some $\alpha > 0$,

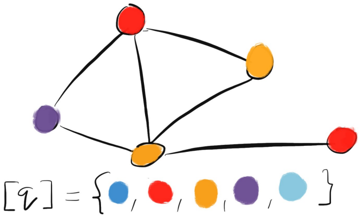
$$\mathbb{E}_{x,y}(\rho(X_1, Y_1)) \leq e^{-\alpha}\rho(x, y),$$

then

$$t_{mix}(\varepsilon) \leq \left\lceil \frac{\log(diam(\Omega) + \log(\varepsilon))}{\alpha} \right\rceil$$

where $diam(\Omega) = \max_{x,y} \rho(x, y)$.

Metropolis chain on extended space



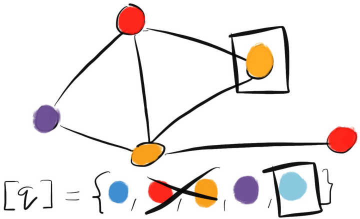
Graph $G = (V, E)$. $\tilde{\Omega}$ set of **all** colorings of G (including improper).

Metropolis MC:

- Select $v \in V$ and $k \in [q]$ uniformly.
- If k is allowed at v, update.

Stationary distribution: uniform on Ω, proper colorings.

Metropolis chain on extended space



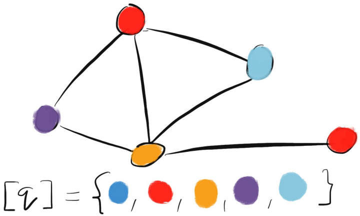
Graph $G = (V, E)$. $\tilde{\Omega}$ set of **all** colorings of G (including improper).

Metropolis MC:

- Select $v \in V$ and $k \in [q]$ uniformly.
- If k is allowed at v, update.

Stationary distribution: uniform on Ω, proper colorings.

Metropolis chain on extended space



Graph $G = (V, E)$. $\tilde{\Omega}$ set of **all** colorings of G (including improper).

Metropolis MC:

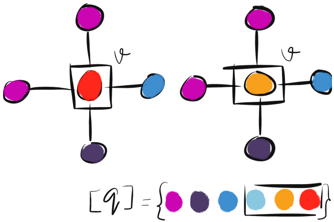
- Select $v \in V$ and $k \in [q]$ uniformly.
- If k is allowed at v, update.

Stationary distribution: uniform on Ω, proper colorings.

Path coupling for colorings

Let $x \sim y$ in $(\tilde{\Omega}, E_0)$ if their color differs at 1 vertex.

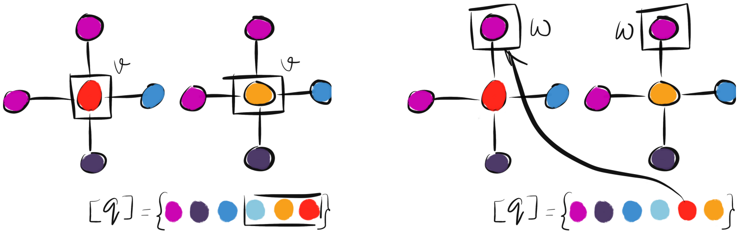
For $(x, y) \in E_0$, let $\ell(x, y) = 1$.



Path coupling for colorings

Let $x \sim y$ in $(\tilde{\Omega}, E_0)$ if their color differs at 1 vertex.

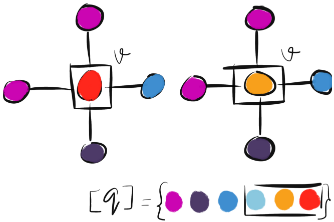
For $(x, y) \in E_0$, let $\ell(x, y) = 1$.

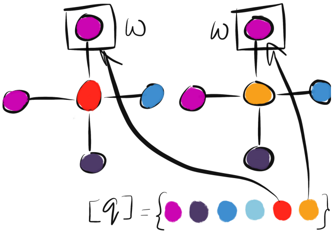


Path coupling for colorings

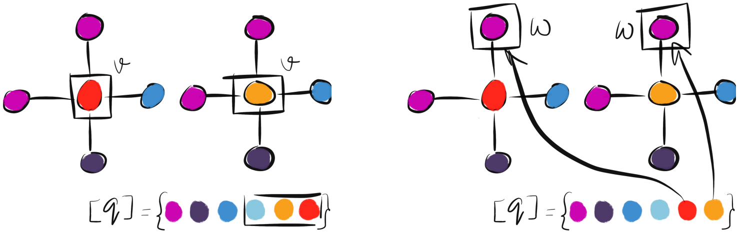
Let $x \sim y$ in $(\tilde{\Omega}, E_0)$ if their color differs at 1 vertex.

For $(x, y) \in E_0$, let $\ell(x, y) = 1$.





Coupling



Let v be the disagreeing vertex in colorings x and y.

- Pick a vertex u and color k uniformly at random.
- If $u \notin N(v)$ attempt to update u with k.
- If $u \in N(v)$
  - If $k \notin \{x(v), y(v)\}$ attempt to update u with k.
  - Otherwise, attempt to update u in x with k and in y with the color in $\{x(v), y(v)\} \setminus \{k\}$.

$$\mathbb{E}_{x,y}(\rho(X_1, Y_1)) - \rho(x, y) \leq -\frac{q - \Delta}{qn} + \frac{\Delta}{qn} = -\frac{q - 2\Delta}{qn}$$

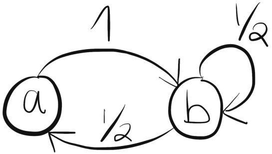
Sampling colorings

- Conjecture: $O(n \log n)$ mixing for $q \geq \Delta + 2$.
- (Vigoda 1999) $O(n^2 \log n)$ mixing for $q \geq \frac{11}{6}\Delta$.
- (Hayes-Sinclair 2005) $\Omega(n \log n)$ lower bound.
- Restricted cases - triangle free graphs, large max. degree. Non-Markovian couplings. (Dyer, Flaxman, Frieze, Hayes, Molloy, Vigoda)

Exact Sampling - Coupling from the past

(Propp-Wilson 1996)

Two copies of the chain are run until they meet. When the chains meet, are they at stationarity?



No: $\pi(a) = \frac{1}{3}, \pi(b) = \frac{2}{3}$, but the chains never meet at a.

Coupling from the past (CFTP): If we "run the chain backwards/ from the past" from a time so that all trajectories meet, guaranteed to be at stationarity.

Random function representation of a MC

Ergodic MC with transition matrix P. A distribution $\mathcal{F}$ over functions $f : \Omega \to \Omega$ is a **random function representation** (RFR) iff

$$\mathbb{P}_{\mathcal{F}}(f(x) = y) = P(x, y)$$

e.g. for the SRW on $\{0, 1 \ldots, n\}$ let

$$f(i) = \min\{i + 1, n\}, \qquad f'(i) = \max\{i - 1, 0\}.$$

and $\mathcal{F}$ uniform on f and f'.

Proposition 12. Every transition matrix has an RFR, not necessarily unique.

$\mathcal{F}$ defines a coupling on Ω via

$$(x, y) \xrightarrow{f \sim \mathcal{F}} (f(x), f(y))$$

Simulating forwards and backwards in time

Associate a random function $f_t \sim \mathcal{F}$ to each $t \in \{-\infty, \ldots, \infty\}$.

Forward simulation of the chain from x for t steps:

$$F_0^t(x) = f_{t-1} \circ \cdots \circ f_0(x), \qquad P^t(x, y) = \mathbb{P}(F_0^t(x) = y).$$

Backward simulation of the chain from x for t steps:

$$F_{-t}^0(x) = f_{-1} \circ \cdots \circ f_{-t}(x).$$

Let S be time so that $|F_0^S(\Omega)| = 1$. May not be stationary at S.

CFTP: Let S be such that $|F_{-S}^0(\Omega)| = 1$. The chain is stationary!

Why does CFTP work?

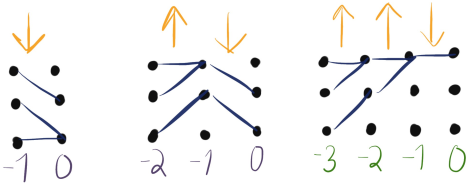
$$\lim_{t\to\infty} \mathbb{P}(F_0^t(x) = y) = \pi(y) = \lim_{t\to\infty} \mathbb{P}(F_{-t}^0(x) = y)$$

Let S be such that $|F_{-S}^0(\Omega)| = 1$ and $t > S$.

$$\begin{aligned} F_{-t}^0(x) &= f_{-1} \circ \cdots \circ f_{-t}(x) \\ &= F_{-S}^0 \circ f_{-S-1} \circ \cdots \circ f_{-t}(x) \\ &= F_{-S}^0(y) \\ &= F_{-S}^0(x). \end{aligned}$$

Since $t > S$ was arbitrary, F_{-t}^0 has the same distribution as $F_{-\infty}^0$.

Implementing CFTP



Issues/Benefits:

- Choosing $\mathcal{F}$ so $\mathbb{E}(S)$ is small.
- When Ω is large, detecting when $|F^0_{-S}(\Omega)| = 1$.
- Exact samples even in absence of bounds on mixing time.

Monotone CFTP:

- Partial order $\preceq$ on states respected by the coupling with a maximal state x_{max} and minimal state x_{min}.
- Run CFTP from x_{max} and x_{min}. All trajectories will have converged when these coalesce.
- Ising Model, dimer configs. of hexagonal grid, bipartite independent set.

Strong stationary times

A random variable τ is a **stopping time** for (X_t) if $\mathbf{1}_{\tau=t}$ is a function only of $(X_0, \ldots, X_t)$.

A **strong stationary time** (SST) for (X_t) with stationary dist π is a randomized stopping time such that

$$\mathbb{P}_x(\tau = t, X_\tau = y) = \mathbb{P}_x(\tau = t)\pi(y)$$

Theorem 13. If τ is an SST then

$$d(t) \leq \max_x \mathbb{P}_x(\tau > t).$$

Broder stopping time for random shuffling

MC on S_n:

- Choose L_t and R_t u.a.r. and transpose if different.

Stopping time: Mark R_t if both

- R_t is unmarked
- Either L_t is marked or $L_t = R_t$.

Time τ for all cards to be marked is an SST.

$$\tau = \tau_0 + \cdots + \tau_{n-1}$$

where τ_k = number of transpositions after kth card is marked and upto and including when $k+1$st card is marked .

$$\tau_k \sim Geom\left(\frac{(k+1)(n-k)}{n^2}\right)$$

Coupon collector estimate

$$\left(\frac{n^2}{(k+1)(n-k)}\right) = \frac{n^2}{n+1}\left(\frac{1}{k+1} + \frac{1}{n-k}\right)$$

$$\mathbb{E}(\tau) = \sum_{k=0}^{n-1} \mathbb{E}(\tau_k) = 2n(\log n + O(1))$$

Can also calculate

$$Var(\tau) = O(n^2).$$

Let $t_0 = \mathbb{E}(\tau) + 2\sqrt{Var(\tau)}$. By Chebyshev,

$$\mathbb{P}(\tau > t_0) \leq \frac{1}{4}.$$

$$t_{mix} \leq (2 + o(1)) n \log n.$$