CS 263: Counting and Sampling

Nima Anari

Stanford University

slides for

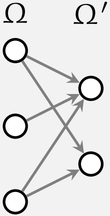
Dobrushin's Influence Matrix

Review

Markov kernel

$$N \in \mathbb{R}_{\geqslant 0}^{\Omega \times \Omega'}$$

$$\sum_y N(x, y) = 1$$

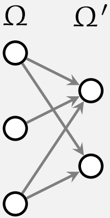


Review

Markov kernel

$$N \in \mathbb{R}_{\geqslant 0}^{\Omega \times \Omega'}$$

$$\sum_y N(x,y) = 1$$



Time-reversal:

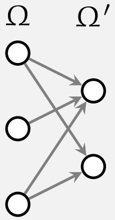
$$\mu(x)\underset{\substack{\uparrow \\ Q(x,y)}}{N}(x,y) = \mu^\circ(y)\underset{\substack{\uparrow \\ Q^\circ(y,x)}}{N^\circ}(y,x)$$

Review

Markov kernel

$$N \in \mathbb{R}_{\geqslant 0}^{\Omega \times \Omega'}$$

$$\sum_y N(x, y) = 1$$



Time-reversal:

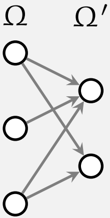
$$\underbrace{\mu(x) N(x, y)}_{Q(x,y)} = \underbrace{\mu^\circ(y) N^\circ(y, x)}_{Q^\circ(y,x)}$$

- Design technique: $N \mapsto N N^\circ$

Markov kernel

$$N \in \mathbb{R}_{\geqslant 0}^{\Omega\times\Omega'}$$

$$\textstyle\sum_y N(x,y) = 1$$



Time-reversal:

$$\mu(x)\underset{\substack{\uparrow\\ Q(x,y)}}{N(x,y)} = \mu^\circ(y)\underset{\substack{\uparrow\\ Q^\circ(y,x)}}{N^\circ(y,x)}$$

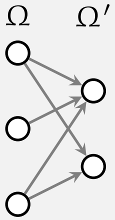
- Design technique: $N \mapsto NN^\circ$
- Examples: Glauber, etc.

Review

Markov kernel

$$N \in \mathbb{R}_{\geqslant 0}^{\Omega \times \Omega'}$$

$$\sum_y N(x,y) = 1$$



Time-reversal:

$$\underbrace{\mu(x) N(x,y)}_{Q(x,y)} = \underbrace{\mu^\circ(y) N^\circ(y,x)}_{Q^\circ(y,x)}$$

- Design technique: $N \mapsto NN^\circ$
- Examples: Glauber, etc.

- Wasserstein w.r.t. d is $\mathcal{W}(\mu, \nu) =$

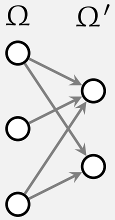
$$\min\{\mathbb{E}_{(X,Y)\sim\pi}[d(X,Y)] \mid \pi \text{ coupling}\}$$

Review

Markov kernel

$$N \in \mathbb{R}_{\geqslant 0}^{\Omega \times \Omega'}$$

$$\sum_y N(x,y) = 1$$



Time-reversal:

$$\underbrace{\mu(x)N(x,y)}_{Q(x,y)} = \underbrace{\mu^\circ(y)N^\circ(y,x)}_{Q^\circ(y,x)}$$

- Design technique: $N \mapsto NN^\circ$
- Examples: Glauber, etc.

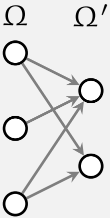
- Wasserstein w.r.t. d is $\mathcal{W}(\mu, \nu) = \min\{\mathbb{E}_{(X,Y)\sim\pi}[d(X,Y)] \mid \pi \text{ coupling}\}$
- For Metropolis chain on colorings:
  $\mathcal{W}(\nu P, \nu' P) \leqslant \left(1 - \frac{q-4\Delta}{qn}\right) \mathcal{W}(\nu, \nu')$

Review

Markov kernel

$$N \in \mathbb{R}_{\geqslant 0}^{\Omega \times \Omega'}$$

$$\textstyle\sum_y N(x,y) = 1$$



Time-reversal:

$$\mu(x)\underset{\overset{\uparrow}{Q(x,y)}}{N(x,y)} = \mu^\circ(y)\underset{\overset{\uparrow}{Q^\circ(y,x)}}{N^\circ(y,x)}$$

- Design technique: $N \mapsto NN^\circ$
- Examples: Glauber, etc.

- Wasserstein w.r.t. d is $\mathcal{W}(\mu,\nu) =$

  $\min\{\mathbb{E}_{(X,Y)\sim\pi}[d(X,Y)] \mid \pi \text{ coupling}\}$

- For Metropolis chain on colorings:

  $\mathcal{W}(\nu P, \nu' P) \leqslant \left(1 - \frac{q-4\Delta}{qn}\right) \mathcal{W}(\nu,\nu')$

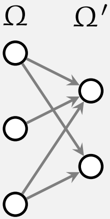
- Mixing when $q \geqslant 4\Delta + 1$ 🙂

Review

Markov kernel

$$N \in \mathbb{R}_{\geqslant 0}^{\Omega\times\Omega'}$$

$$\sum_y N(x,y) = 1$$



Time-reversal:

$$\mu(x)\underset{\substack{\uparrow\\ Q(x,y)}}{N(x,y)}=\mu^\circ(y)\underset{\substack{\uparrow\\ Q^\circ(y,x)}}{N^\circ(y,x)}$$

- Design technique: $N \mapsto NN^\circ$
- Examples: Glauber, etc.

- Wasserstein w.r.t. d is $\mathcal{W}(\mu,\nu) =$

  $$\min\{\mathbb{E}_{(X,Y)\sim\pi}[d(X,Y)] \mid \pi \text{ coupling}\}$$

- For Metropolis chain on colorings:

  $$\mathcal{W}(\nu P, \nu' P) \leqslant \left(1 - \frac{q-4\Delta}{qn}\right)\mathcal{W}(\nu,\nu')$$

- Mixing when $q \geqslant 4\Delta + 1$

Path coupling lemma

Suppose for all adjacent $X_0 \sim X_0'$ we can couple X_1, X_1' s.t.

$$\mathbb{E}[d(X_1,X_1')] \leqslant (1-c)d(X_0,X_0').$$

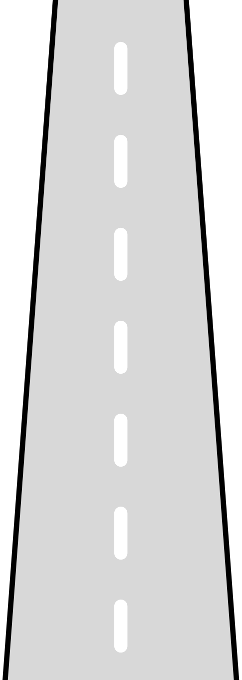
Then $\mathcal{W}(\nu P, \nu' P) \leqslant (1-c)\,\mathcal{W}(\nu,\nu')$.

Mixing via Transport

- Path coupling
- Dobrushin's condition
- Hardcore model
- Ising model
- Dobrushin++

Intro to Functional Analysis

- Divergences

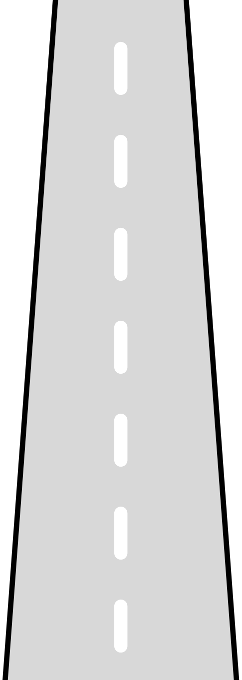


Mixing via Transport

- Path coupling
- Dobrushin's condition
- Hardcore model
- Ising model
- Dobrushin++

Intro to Functional Analysis

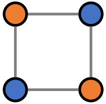
- Divergences



Path coupling for colorings [Jerrum]

Take the Metropolis chain for colorings:

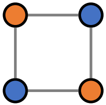
- Pick u.r. vertex v
- Pick u.r. color c
- Color v with c if valid



Path coupling for colorings [Jerrum]

Take the Metropolis chain for colorings:

- Pick u.r. vertex v
- Pick u.r. color c
- Color v with c if valid

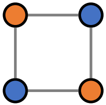


Coupling:

Path coupling for colorings [Jerrum]

Take the Metropolis chain for colorings:

- Pick u.r. vertex v
- Pick u.r. color c
- Color v with c if valid



Coupling:

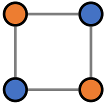
- Take X_0, X_0' differing on vertex w.

there is exactly one

Path coupling for colorings [Jerrum]

Take the Metropolis chain for colorings:

- Pick u.r. vertex v
- Pick u.r. color c
- Color v with c if valid



Coupling:

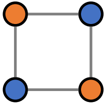
- Take X_0, X_0' differing on vertex w.
- Pick same v.

there is exactly one

Path coupling for colorings [Jerrum]

Take the Metropolis chain for colorings:

- Pick u.r. vertex v
- Pick u.r. color c
- Color v with c if valid



Coupling:

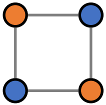
- Take X_0, X_0' differing on vertex w.
- Pick same v.

there is exactly one

- If $v = w$ or $v \not\sim w$, pick same c.

Path coupling for colorings [Jerrum]

Take the Metropolis chain for colorings:



- Pick u.r. vertex v
- Pick u.r. color c
- Color v with c if valid

Coupling:

- Take X_0, X_0' differing on vertex w.
- Pick same v. there is exactly one
- If $v = w$ or $v \not\sim w$, pick same c.
- Let $a = X_0(w), b = X_0'(w)$. If $v \sim w$, use this coupling for color:

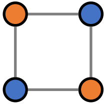
$$c: \quad 1 \quad 2 \cdots a \cdots b \cdots q$$

$$c': \quad 1 \quad 2 \cdots a \cdots b \cdots q$$

Path coupling for colorings [Jerrum]

Take the Metropolis chain for colorings:

- Pick u.r. vertex v
- Pick u.r. color c
- Color v with c if valid



Coupling:

- Take X_0, X'_0 differing on vertex w.
- Pick same v.

there is exactly one

- If $v = w$ or $v \not\sim w$, pick same c.
- Let $a = X_0(w), b = X'_0(w)$. If $v \sim w$, use this coupling for color:

$$\begin{array}{lcccccccc} c: & 1 & 2 & \cdots & a & \cdots & b & \cdots & q \\ & | & | & & & \times & & & | \\ c': & 1 & 2 & \cdots & a & \cdots & b & \cdots & q \end{array}$$

- We have

$$\mathbb{E}[d(X_1, X'_1)] \leqslant 1 - \frac{1}{n} \cdot \frac{q - \Delta}{q} + \frac{\Delta}{n} \cdot \frac{1}{q}$$

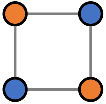
which simplifies to

$$1 - \frac{q - 2\Delta}{qn}.$$

Path coupling for colorings [Jerrum]

Take the Metropolis chain for colorings:

- Pick u.r. vertex v
- Pick u.r. color c
- Color v with c if valid



Coupling:

- Take X_0, X_0' differing on vertex w.
- Pick same v. there is exactly one
- If $v = w$ or $v \not\sim w$, pick same c.
- Let $a = X_0(w)$, $b = X_0'(w)$. If $v \sim w$, use this coupling for color:

$$\begin{array}{cccccccccc} c: & 1 & 2 & \cdots & a & \cdots & b & \cdots & q \\ & | & | & & & \times & & & | \\ c': & 1 & 2 & \cdots & a & \cdots & b & \cdots & q \end{array}$$

- We have

$$\mathbb{E}[d(X_1, X_1')] \leqslant 1 - \frac{1}{n} \cdot \frac{q - \Delta}{q} + \frac{\Delta}{n} \cdot \frac{1}{q}$$

which simplifies to

$$1 - \frac{q - 2\Delta}{qn}.$$

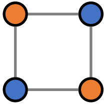
- As long as $q \geqslant 2\Delta + 1$, we get contraction! ☺

$$t_{\mathrm{mix}}(\epsilon) = O\left(\frac{q}{q - 2\Delta} \cdot n \log(n/\epsilon)\right)$$

Path coupling for colorings [Jerrum]

Take the Metropolis chain for colorings:

- Pick u.r. vertex v
- Pick u.r. color c
- Color v with c if valid



Coupling:

- Take X_0, X_0' differing on vertex w.
- Pick same v. there is exactly one
- If $v = w$ or $v \not\sim w$, pick same c.
- Let $a = X_0(w), b = X_0'(w)$. If $v \sim w$, use this coupling for color:

$$\begin{array}{lcccccccc} c: & 1 & 2 & \cdots & a & \cdots & b & \cdots & q \\ c': & 1 & 2 & \cdots & a & \cdots & b & \cdots & q \end{array}$$

- We have

$$\mathbb{E}[d(X_1, X_1')] \leqslant 1 - \frac{1}{n} \cdot \frac{q-\Delta}{q} + \frac{\Delta}{n} \cdot \frac{1}{q}$$

which simplifies to

$$1 - \frac{q-2\Delta}{qn}.$$

- As long as $q \geqslant 2\Delta + 1$, we get contraction! ☺

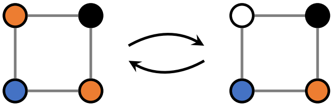
$$t_{\mathrm{mix}}(\epsilon) = O\left(\frac{q}{q-2\Delta} \cdot n \log(n/\epsilon)\right)$$

- What about Glauber?

Dobrushin influence matrix

Glauber for μ on $\Omega_1 \times \cdots \times \Omega_n$:

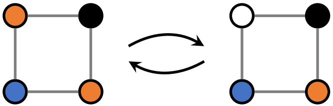
- Pick u.r. coord $i \in [n]$
- Replace coord i's value w.p. $\propto \mu(\text{result})$



Dobrushin influence matrix

Glauber for μ on $\Omega_1 \times \cdots \times \Omega_n$:

- Pick u.r. coord $i \in [n]$
- Replace coord i's value w.p. $\propto \mu(\text{result})$

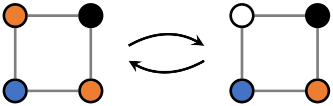


- We pick $\omega \in \Omega_i$ from $\mathsf{dist}(X_i \mid X_{-i})$.

Dobrushin influence matrix

Glauber for μ on $\Omega_1 \times \cdots \times \Omega_n$:

- Pick u.r. coord $i \in [n]$
- Replace coord i's value w.p. $\propto \mu(\text{result})$



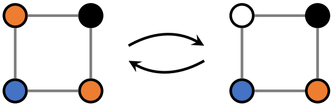
- We pick $\omega \in \Omega_i$ from $\mathsf{dist}(X_i \mid X_{-i})$.
- Influence: take worst case X, X' that differ in coord j:

$$d_{TV}\left(\mathsf{dist}(X_i \mid X_{-i}), \mathsf{dist}(X'_i \mid X'_{-i})\right)$$

Dobrushin influence matrix

Glauber for μ on $\Omega_1 \times \cdots \times \Omega_n$:

- Pick u.r. coord $i \in [n]$
- Replace coord i's value w.p. $\propto \mu(\text{result})$



- We pick $\omega \in \Omega_i$ from $\mathsf{dist}(X_i \mid X_{-i})$.
- Influence: take worst case X, X' that differ in coord j:

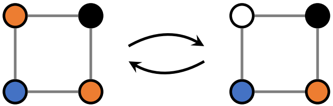
  $$d_{TV}\left(\mathsf{dist}(X_i \mid X_{-i}), \mathsf{dist}(X'_i \mid X'_{-i})\right)$$

- Call maximum value $\mathcal{I}[j \to i]$.

Dobrushin influence matrix

Glauber for μ on $\Omega_1 \times \cdots \times \Omega_n$:

- Pick u.r. coord $i \in [n]$
- Replace coord i's value w.p. $\propto \mu(\text{result})$



- We pick $\omega \in \Omega_i$ from $\mathsf{dist}(X_i \mid X_{-i})$.
- Influence: take worst case X, X' that differ in coord j:

$$d_{\mathsf{TV}}\left(\mathsf{dist}(X_i \mid X_{-i}), \mathsf{dist}(X_i' \mid X_{-i}')\right)$$

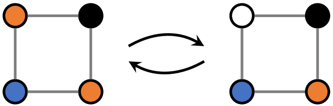
- Call maximum value $\mathcal{I}[j \to i]$.

- Dobrushin influence matrix: matrix with entries $\mathcal{I}[j \to i]$.

Dobrushin influence matrix

Glauber for μ on $\Omega_1 \times \cdots \times \Omega_n$:

- Pick u.r. coord $i \in [n]$
- Replace coord i's value w.p. $\propto \mu(\text{result})$



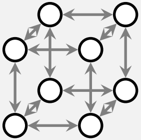
- We pick $\omega \in \Omega_i$ from $\mathsf{dist}(X_i \mid X_{-i})$.
- Influence: take worst case X, X' that differ in coord j:

  $$d_{\mathrm{TV}}\left(\mathsf{dist}(X_i \mid X_{-i}), \mathsf{dist}(X'_i \mid X'_{-i})\right)$$
- Call maximum value $\mathcal{I}[j \to i]$.

- Dobrushin influence matrix: matrix with entries $\mathcal{I}[j \to i]$.

Example: hypercube

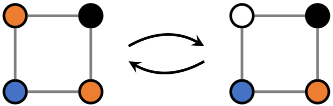
- $\Omega = \{0, 1\}^n$
- $\mu(x) = \text{uniform}$



Dobrushin influence matrix

Glauber for μ on $\Omega_1 \times \cdots \times \Omega_n$:

- Pick u.r. coord $i \in [n]$
- Replace coord i's value w.p. $\propto \mu(\text{result})$



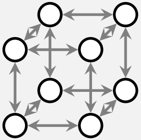
- We pick $\omega \in \Omega_i$ from $\mathsf{dist}(X_i \mid X_{-i})$.
- Influence: take worst case X, X' that differ in coord j:

  $$d_{\mathsf{TV}}\big(\mathsf{dist}(X_i \mid X_{-i}), \mathsf{dist}(X'_i \mid X'_{-i})\big)$$
- Call maximum value $\mathcal{I}[j \to i]$.

- Dobrushin influence matrix: matrix with entries $\mathcal{I}[j \to i]$.

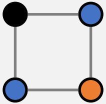
Example: hypercube

- $\Omega = \{0, 1\}^n$
- $\mu(x) = \text{uniform}$
- $\mathcal{I}[j \to i] = 0$



Example: coloring

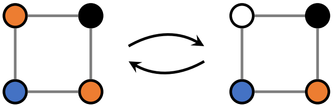
- $\Omega = [q]^n$
- $\mu = \text{proper coloring}$



Dobrushin influence matrix

Glauber for μ on $\Omega_1 \times \cdots \times \Omega_n$:

- Pick u.r. coord $i \in [n]$
- Replace coord i's value w.p. $\propto \mu(\text{result})$



- We pick $\omega \in \Omega_i$ from $\mathsf{dist}(X_i \mid X_{-i})$.
- Influence: take worst case X, X' that differ in coord j:

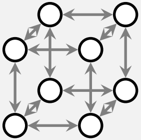
$$d_{\mathsf{TV}}\left(\mathsf{dist}(X_i \mid X_{-i}), \mathsf{dist}(X'_i \mid X'_{-i})\right)$$

- Call maximum value $\mathcal{I}[j \to i]$.

- Dobrushin influence matrix: matrix with entries $\mathcal{I}[j \to i]$.

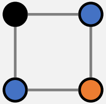
Example: hypercube

- $\Omega = \{0, 1\}^n$
- $\mu(x) = \text{uniform}$
- $\mathcal{I}[j \to i] = 0$



Example: coloring

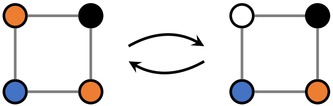
- $\Omega = [q]^n$
- $\mu = \text{proper coloring}$
- $\mathcal{I}[j \to i] = 0$ when $j \not\sim i$



Dobrushin influence matrix

Glauber for μ on $\Omega_1 \times \cdots \times \Omega_n$:

- Pick u.r. coord $i \in [n]$
- Replace coord i's value w.p. $\propto \mu(\text{result})$



- We pick $\omega \in \Omega_i$ from $\mathsf{dist}(X_i \mid X_{-i})$.
- Influence: take worst case X, X' that differ in coord j:

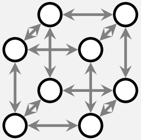
$$d_{TV}\big(\mathsf{dist}(X_i \mid X_{-i}), \mathsf{dist}(X'_i \mid X'_{-i})\big)$$

- Call maximum value $\mathcal{I}[j \to i]$.

- Dobrushin influence matrix: matrix with entries $\mathcal{I}[j \to i]$.

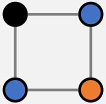
Example: hypercube

- $\Omega = \{0, 1\}^n$
- $\mu(x) = \text{uniform}$
- $\mathcal{I}[j \to i] = 0$



Example: coloring

- $\Omega = [q]^n$
- $\mu = \text{proper coloring}$
- $\mathcal{I}[j \to i] = 0$ when $j \not\sim i$
- $\mathcal{I}[j \to i] \leqslant 1/(q - \Delta)$



Dobrushin's condition

If columns of $\mathcal{I}$ sum to $\leqslant 1-\delta$, then

$$\mathcal{W}(\nu P, \nu' P) \leqslant (1-\delta/n)\, \mathcal{W}(\nu, \nu')$$

Dobrushin's condition

If columns of $\mathcal{J}$ sum to $\leqslant 1-\delta$, then

$$\mathcal{W}(\nu P, \nu' P) \leqslant (1-\delta/n)\, \mathcal{W}(\nu, \nu')$$

- This implies fast mixing:

$$t_{\mathsf{mix}}(\epsilon) = O\left(\tfrac{1}{\delta} n \log(n/\epsilon)\right)$$

Dobrushin's condition

If columns of $\mathcal{J}$ sum to $\leqslant 1-\delta$, then

$$\mathcal{W}(\nu P, \nu' P) \leqslant (1-\delta/n)\, \mathcal{W}(\nu, \nu')$$

- This implies fast mixing:

$$t_{\mathsf{mix}}(\epsilon) = O\big(\tfrac{1}{\delta} n \log(n/\epsilon)\big)$$

- For coloring, $\mathcal{J} \leqslant \mathrm{adj}/(q-\Delta)$, so column sums are $\Delta/(q-\Delta)$:

$$\frac{\Delta}{q-\Delta} < 1 \;\leftrightarrow\; q \geqslant 2\Delta + 1$$

same cond as Metropolis

Dobrushin's condition

If columns of $\mathcal{J}$ sum to $\leqslant 1-\delta$, then

$$\mathcal{W}(\nu P, \nu' P) \leqslant (1-\delta/n)\, \mathcal{W}(\nu, \nu')$$

- This implies fast mixing:

$$t_{\mathrm{mix}}(\epsilon) = O\left(\tfrac{1}{\delta} n \log(n/\epsilon)\right)$$

- For coloring, $\mathcal{J} \leqslant \mathrm{adj}/(q-\Delta)$, so column sums are $\Delta/(q-\Delta)$:

$$\frac{\Delta}{q-\Delta} < 1 \;\leftrightarrow\; q \geqslant 2\Delta + 1$$

same cond as Metropolis

- Note: $\mathcal{J}[i \to i] = 0$ always.

Dobrushin's condition

If columns of $\mathcal{J}$ sum to $\leqslant 1-\delta$, then

$$\mathcal{W}(\nu P, \nu' P) \leqslant (1-\delta/n)\, \mathcal{W}(\nu, \nu')$$

Proof:

- This implies fast mixing:
$$t_{\text{mix}}(\epsilon) = O\big(\tfrac{1}{\delta} n \log(n/\epsilon)\big)$$
- For coloring, $\mathcal{J} \leqslant \text{adj}/(q-\Delta)$, so column sums are $\Delta/(q-\Delta)$:
$$\tfrac{\Delta}{q-\Delta} < 1 \;\leftrightarrow\; q \geqslant 2\Delta + 1$$
same cond as Metropolis
- Note: $\mathcal{J}[i \to i] = 0$ always.

Dobrushin's condition

If columns of $\mathcal{J}$ sum to $\leqslant 1-\delta$, then

$$\mathcal{W}(\nu P, \nu' P) \leqslant (1-\delta/n)\, \mathcal{W}(\nu, \nu')$$

- This implies fast mixing:

$$t_{\text{mix}}(\epsilon) = O\big(\tfrac{1}{\delta} n \log(n/\epsilon)\big)$$

- For coloring, $\mathcal{J} \leqslant \text{adj}/(q-\Delta)$, so column sums are $\Delta/(q-\Delta)$:

$$\frac{\Delta}{q-\Delta} < 1 \;\leftrightarrow\; q \geqslant 2\Delta + 1$$

same cond as Metropolis

- Note: $\mathcal{J}[i \to i] = 0$ always.

Proof:

- Idea: use path coupling!

Dobrushin's condition

If columns of $\mathcal{J}$ sum to $\leqslant 1-\delta$, then

$$\mathcal{W}(\nu P, \nu' P) \leqslant (1-\delta/n)\, \mathcal{W}(\nu, \nu')$$

- This implies fast mixing:
$$t_{\text{mix}}(\epsilon) = O\left(\tfrac{1}{\delta} n \log(n/\epsilon)\right)$$
- For coloring, $\mathcal{J} \leqslant \text{adj}/(q-\Delta)$, so column sums are $\Delta/(q-\Delta)$:
$$\frac{\Delta}{q-\Delta} < 1 \;\leftrightarrow\; q \geqslant 2\Delta + 1$$
same cond as Metropolis
- Note: $\mathcal{J}[i \to i] = 0$ always.

Proof:

- Idea: use path coupling!
- Take X_0, X_0' differing in j.

Dobrushin's condition

If columns of $\mathcal{J}$ sum to $\leqslant 1-\delta$, then

$$\mathcal{W}(\nu P, \nu' P) \leqslant (1-\delta/n)\, \mathcal{W}(\nu, \nu')$$

- This implies fast mixing:
$$t_{\mathrm{mix}}(\epsilon) = O\left(\tfrac{1}{\delta} n \log(n/\epsilon)\right)$$
- For coloring, $\mathcal{J} \leqslant \mathrm{adj}/(q-\Delta)$, so column sums are $\Delta/(q-\Delta)$:
$$\frac{\Delta}{q-\Delta} < 1 \;\leftrightarrow\; q \geqslant 2\Delta + 1$$
same cond as Metropolis
- Note: $\mathcal{J}[i \to i] = 0$ always.

Proof:

- Idea: use path coupling!
- Take X_0, X_0' differing in j.
- Will produce coupling of X_1, X_1'.

Dobrushin's condition

If columns of $\mathcal{J}$ sum to $\leqslant 1-\delta$, then

$$\mathcal{W}(\nu P, \nu' P) \leqslant (1-\delta/n)\, \mathcal{W}(\nu, \nu')$$

- This implies fast mixing:

$$t_{\mathrm{mix}}(\epsilon) = O\left(\tfrac{1}{\delta} n \log(n/\epsilon)\right)$$

- For coloring, $\mathcal{J} \leqslant \mathrm{adj}/(q-\Delta)$, so column sums are $\Delta/(q-\Delta)$:

$$\frac{\Delta}{q-\Delta} < 1 \;\leftrightarrow\; q \geqslant 2\Delta + 1$$

same cond as Metropolis

- Note: $\mathcal{J}[i \to i] = 0$ always.

Proof:

- Idea: use path coupling!
- Take X_0, X_0' differing in j.
- Will produce coupling of X_1, X_1'.
- Pick same coord i.

Dobrushin's condition

If columns of $\mathcal{J}$ sum to $\leqslant 1-\delta$, then

$$\mathcal{W}(\nu P, \nu' P) \leqslant (1-\delta/n)\, \mathcal{W}(\nu, \nu')$$

- This implies fast mixing:

$$t_{\mathrm{mix}}(\epsilon) = O\left(\tfrac{1}{\delta} n \log(n/\epsilon)\right)$$

- For coloring, $\mathcal{J} \leqslant \mathrm{adj}/(q-\Delta)$, so column sums are $\Delta/(q-\Delta)$:

$$\frac{\Delta}{q-\Delta} < 1 \;\leftrightarrow\; q \geqslant 2\Delta + 1$$

same cond as Metropolis

- Note: $\mathcal{J}[i \to i] = 0$ always.

Proof:

- Idea: use path coupling!
- Take X_0, X_0' differing in j.
- Will produce coupling of X_1, X_1'.
- Pick same coord i.
- Maximally couple replacements.

the one defining d_{TV}

Dobrushin's condition

If columns of $\mathcal{J}$ sum to $\leqslant 1-\delta$, then

$$\mathcal{W}(\nu P, \nu' P) \leqslant (1-\delta/n)\, \mathcal{W}(\nu, \nu')$$

- This implies fast mixing:

$$t_{\text{mix}}(\epsilon) = O\left(\tfrac{1}{\delta} n \log(n/\epsilon)\right)$$

- For coloring, $\mathcal{J} \leqslant \text{adj}/(q-\Delta)$, so column sums are $\Delta/(q-\Delta)$:

$$\frac{\Delta}{q-\Delta} < 1 \;\leftrightarrow\; q \geqslant 2\Delta + 1$$

same cond as Metropolis

- Note: $\mathcal{J}[i \to i] = 0$ always.

Proof:

- Idea: use path coupling!
- Take X_0, X_0' differing in j.
- Will produce coupling of X_1, X_1'.
- Pick same coord i.
- Maximally couple replacements.

the one defining d_{TV}

- Possibilities for $d(X_1, X_1')$:
  - 0 (picked $i = j$)
  - 1 (equal replacements)
  - 2 (different replacements)

Dobrushin's condition

If columns of $\mathcal{J}$ sum to $\leqslant 1-\delta$, then

$$\mathcal{W}(\nu P, \nu' P) \leqslant (1-\delta/n)\, \mathcal{W}(\nu, \nu')$$

- This implies fast mixing:
$$t_{\mathrm{mix}}(\epsilon) = O\left(\tfrac{1}{\delta} n \log(n/\epsilon)\right)$$
- For coloring, $\mathcal{J} \leqslant \mathrm{adj}/(q-\Delta)$, so column sums are $\Delta/(q-\Delta)$:
$$\frac{\Delta}{q-\Delta} < 1 \;\leftrightarrow\; q \geqslant 2\Delta + 1$$
same cond as Metropolis
- Note: $\mathcal{J}[i \to i] = 0$ always.

Proof:

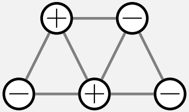
- Idea: use path coupling!
- Take X_0, X_0' differing in j.
- Will produce coupling of X_1, X_1'.
- Pick same coord i.
- Maximally couple replacements.

  the one defining d_{TV}
- Possibilities for $d(X_1, X_1')$:
  - 0 (picked $i = j$)
  - 1 (equal replacements)
  - 2 (different replacements)
- We get $\mathbb{E}[d(X_1, X_1')] \leqslant \frac{1}{n} \cdot 0 + \frac{1}{n} \sum_{i \neq j} (1 + \mathcal{J}[j \to i]) \leqslant 1 - \delta/n$

- Useful for spin systems.

▷ Useful for spin systems.

Spin systems



graph $G = (V, E)$

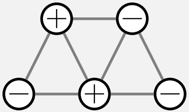
$$\Omega = \Omega_1 \times \cdots \times \Omega_n$$

local interaction

$$\mu(x) = \prod_v \phi_v(x_v) \cdot \prod_{u \sim v} \phi_{uv}(x_u, x_v)$$

- Useful for spin systems.

Spin systems



graph $G = (V, E)$

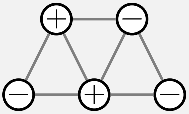
$$\Omega = \Omega_1 \times \cdots \times \Omega_n$$

$$\mu(x) = \prod_v \phi_v(x_v) \cdot \prod_{u \sim v} \underbrace{\phi_{uv}}_{\text{local interaction}}(x_u, x_v)$$

- $\mathcal{I}[j \to i] = 0$ when $i \not\sim j$

- Useful for spin systems.

Spin systems



graph $G = (V, E)$

$\Omega = \Omega_1 \times \cdots \times \Omega_n$

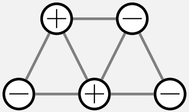
local interaction

$$\mu(x) = \prod_v \phi_v(x_v) \cdot \prod_{u\sim v} \phi_{uv}(x_u, x_v)$$

- $\mathcal{J}[j \to i] = 0$ when $i \not\sim j$
- $\mathcal{J}$ is weighted adjacency matrix

- Useful for spin systems.

Spin systems



graph $G = (V, E)$

$\Omega = \Omega_1 \times \cdots \times \Omega_n$

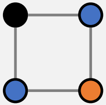
$$\mu(x) = \prod_v \phi_v(x_v) \cdot \prod_{u \sim v} \underbrace{\phi_{uv}(x_u, x_v)}_{}$$

local interaction

- $\mathcal{J}[j \to i] = 0$ when $i \not\sim j$
- $\mathcal{J}$ is weighted adjacency matrix

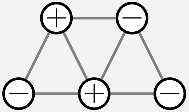
Example: coloring

- $\Omega_i = [q]$
- $\phi_{uv} = \mathbb{1}[x_u \neq x_v]$
- $\mathcal{J} \leqslant \mathrm{adj}/(q - \Delta)$



- Useful for spin systems.

Spin systems



graph $G = (V, E)$

$\Omega = \Omega_1 \times \cdots \times \Omega_n$

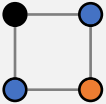
local interaction

$$\mu(x) = \prod_v \phi_v(x_v) \cdot \prod_{u\sim v} \phi_{uv}(x_u, x_v)$$

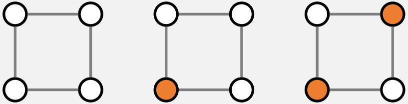
- $\mathcal{J}[j \to i] = 0$ when $i \not\sim j$
- $\mathcal{J}$ is weighted adjacency matrix

Example: coloring

- $\Omega_i = [q]$
- $\phi_{uv} = \mathbb{1}[x_u \neq x_v]$
- $\mathcal{J} \leqslant \mathrm{adj}/(q - \Delta)$



Example: hardcore



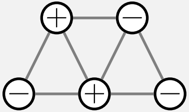
$\mu \propto 1$ $\quad$ $\mu \propto \lambda$ $\quad$ $\mu \propto \lambda^2$

$\Omega_i = \{0, 1\}$ $\quad$ $\phi_v = \lambda^{x_v}$ $\quad$ $\phi_{uv} = 1 - x_u x_v$

fugacity

- Useful for spin systems.

Spin systems



graph $G = (V, E)$

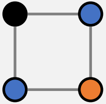
$\Omega = \Omega_1 \times \cdots \times \Omega_n$

$$\mu(x) = \prod_v \phi_v(x_v) \cdot \prod_{u\sim v} \underbrace{\phi_{uv}}_{\text{local interaction}}(x_u, x_v)$$

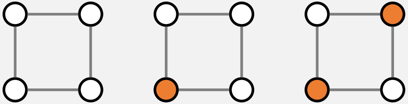
- $\mathcal{J}[j \to i] = 0$ when $i \not\sim j$
- $\mathcal{J}$ is weighted adjacency matrix

Example: coloring

- $\Omega_i = [q]$
- $\phi_{uv} = \mathbb{1}[x_u \neq x_v]$
- $\mathcal{J} \leqslant \mathrm{adj}/(q - \Delta)$



Example: hardcore

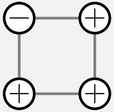


$\mu \propto 1$ $\quad$ $\mu \propto \lambda$ $\quad$ $\mu \propto \lambda^2$

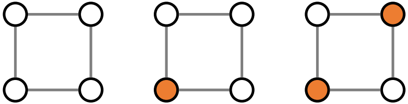
$\Omega_i = \{0, 1\}$ $\quad$ $\phi_v = \underbrace{\lambda}_{\text{fugacity}}{}^{x_v}$ $\quad$ $\phi_{uv} = 1 - x_u x_v$

Example: Ising

- $\Omega_i = \{\pm 1\}$
- $\phi_v = \exp(\underbrace{h_v}_{\text{external field}} x_v)$
- $\phi_{uv} = \exp(\underbrace{\beta_{uv}}_{\text{ferro/anti-ferromagnetic interaction}} x_u x_v)$

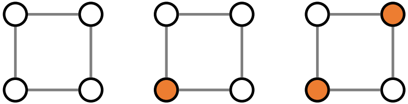


Hardcore model



$\mu(\text{ind set } S) \propto \lambda^{|S|}$

Hardcore model

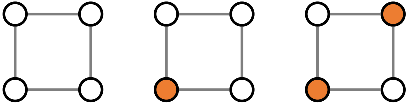


$$\mu(\text{ind set } S) \propto \lambda^{|S|}$$

- Large λ is hard. ☹

  max ind set is **NP**-hard

Hardcore model



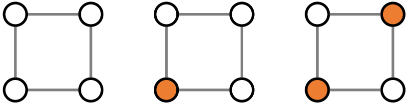
$$\mu(\text{ind set } S) \propto \lambda^{|S|}$$

- Large λ is hard. ☹

  max ind set is **NP**-hard

- For what λ is it easy to sample?

Hardcore model



$$\mu(\text{ind set } S) \propto \lambda^{|S|}$$

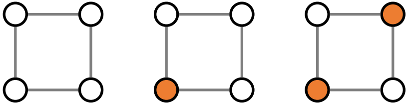
- Large λ is hard. ☹

  max ind set is **NP**-hard

- For what λ is it easy to sample?
- Dobrushin: $\mathcal{I} \leqslant c \cdot \text{adj}$ where $c =$

  $$d_{\text{TV}}(\text{Ber}(0), \text{Ber}(\lambda/(1+\lambda)))$$

Hardcore model



$$\mu(\text{ind set } S) \propto \lambda^{|S|}$$

- Large λ is hard. ☹

  max ind set is NP-hard

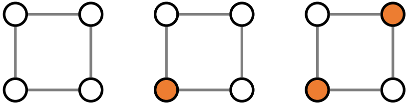
- For what λ is it easy to sample?
- Dobrushin: $\mathcal{I} \leqslant c \cdot \text{adj}$ where $c =$

  $$d_{\mathsf{TV}}(\mathsf{Ber}(0), \mathsf{Ber}(\lambda/(1+\lambda)))$$

- When $\lambda \leqslant (1-\delta)/\Delta$, col sums are

  $$\leqslant \frac{\lambda\Delta}{1+\lambda} \leqslant \lambda\Delta \leqslant 1-\delta$$

Hardcore model



$$\mu(\text{ind set } S) \propto \lambda^{|S|}$$

- Large λ is hard. 🙁

  max ind set is NP-hard

- For what λ is it easy to sample?
- Dobrushin: $\mathcal{I} \leqslant c \cdot \text{adj}$ where $c =$

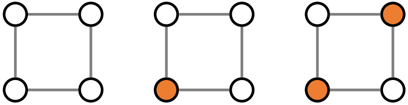
$$d_{\mathsf{TV}}(\mathsf{Ber}(0), \mathsf{Ber}(\lambda/(1+\lambda)))$$

- When $\lambda \leqslant (1-\delta)/\Delta$, col sums are

$$\leqslant \frac{\lambda\Delta}{1+\lambda} \leqslant \lambda\Delta \leqslant 1-\delta$$

- Mixing in $O(n \log n)$ steps. 🙂

Hardcore model



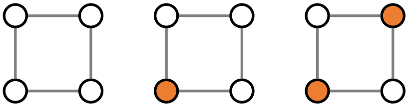
$$\mu(\text{ind set } S) \propto \lambda^{|S|}$$

- Large λ is hard. ☹

  max ind set is NP-hard

- For what λ is it easy to sample?
- Dobrushin: $\mathcal{I} \leqslant c \cdot \text{adj}$ where $c =$
  $$d_{TV}(\text{Ber}(0), \text{Ber}(\lambda/(1+\lambda)))$$
- When $\lambda \leqslant (1-\delta)/\Delta$, col sums are
  $$\leqslant \frac{\lambda\Delta}{1+\lambda} \leqslant \lambda\Delta \leqslant 1-\delta$$
- Mixing in $O(n \log n)$ steps. ☺
- This is NOT the tightest result. Dobrushin is suboptimal.

Hardcore model



$$\mu(\text{ind set } S) \propto \lambda^{|S|}$$

- Large λ is hard. 🙁

  max ind set is NP-hard
- For what λ is it easy to sample?
- Dobrushin: $\mathcal{I} \leqslant c \cdot \text{adj}$ where $c =$

  $$d_{TV}(\text{Ber}(0), \text{Ber}(\lambda/(1+\lambda)))$$
- When $\lambda \leqslant (1-\delta)/\Delta$, col sums are

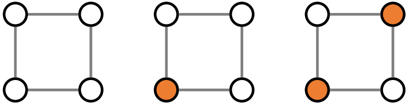
  $$\leqslant \frac{\lambda\Delta}{1+\lambda} \leqslant \lambda\Delta \leqslant 1-\delta$$
- Mixing in $O(n \log n)$ steps. 🙂
- This is NOT the tightest result. Dobrushin is suboptimal.
- We will see later

  $$\lambda \leqslant (1-\delta)\lambda_c(\Delta) \implies \text{fast mixing}$$

  for a specific critical threshold

  $$\lambda_c(\Delta) \simeq \frac{e}{\Delta}$$

Hardcore model



$$\mu(\text{ind set } S) \propto \lambda^{|S|}$$

- Large λ is hard. ☹

  max ind set is NP-hard

- For what λ is it easy to sample?
- Dobrushin: $\mathcal{I} \leqslant c \cdot \text{adj}$ where $c =$

  $$d_{TV}(\text{Ber}(0), \text{Ber}(\lambda/(1+\lambda)))$$

- When $\lambda \leqslant (1-\delta)/\Delta$, col sums are

  $$\leqslant \frac{\lambda\Delta}{1+\lambda} \leqslant \lambda\Delta \leqslant 1-\delta$$

- Mixing in $O(n \log n)$ steps. 🙂
- This is NOT the tightest result. Dobrushin is suboptimal.
- We will see later

  $$\lambda \leqslant (1-\delta)\lambda_c(\Delta) \implies \text{fast mixing}$$

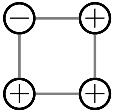
  for a specific critical threshold

  $$\lambda_c(\Delta) \simeq \frac{e}{\Delta}$$

- On the opposite side, [Sly'10] showed it is NP-hard to sample when

  $$\lambda \geqslant (1+\delta)\lambda_c(\Delta)$$

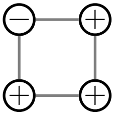
Ising model



$$\mu(x) \propto \exp\left(\tfrac{1}{2} \sum_{u,v} \beta_{uv} x_u x_v + \sum_v h_v x_v\right)$$

symmetric matrix

Ising model

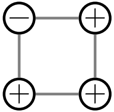


$$\mu(x) \propto \exp\left(\tfrac{1}{2}\sum_{u,v}\beta_{uv}x_u x_v + \sum_v h_v x_v\right)$$

symmetric matrix

- Ferromagnetic: all $\beta \geqslant 0$

Ising model

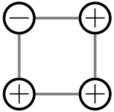


$$\mu(x) \propto \exp\left(\frac{1}{2}\sum_{u,v}\beta_{uv}x_u x_v + \sum_v h_v x_v\right)$$

symmetric matrix

- Ferromagnetic: all $\beta \geqslant 0$
- Anti-ferromagnetic: all $\beta \leqslant 0$

Ising model

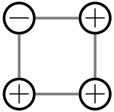


$$\mu(x) \propto \exp\left(\frac{1}{2}\sum_{u,v} \beta_{uv} x_u x_v + \sum_v h_v x_v\right)$$

symmetric matrix

- Ferromagnetic: all $\beta \geqslant 0$
- Anti-ferromagnetic: all $\beta \leqslant 0$
- Sherrington-Kirkpatrick: random β

Ising model

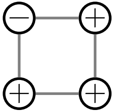


$$\mu(x) \propto \exp\left(\tfrac{1}{2}\sum_{u,v}\beta_{uv}x_u x_v + \sum_v h_v x_v\right)$$

symmetric matrix

- Ferromagnetic: all $\beta \geqslant 0$
- Anti-ferromagnetic: all $\beta \leqslant 0$
- Sherrington-Kirkpatrick: random β
- Note: every 2-spin system is a (limit of an) Ising model

Ising model



$$\mu(x) \propto \exp\left(\tfrac{1}{2}\sum_{u,v}\beta_{uv}x_u x_v + \sum_v h_v x_v\right)$$

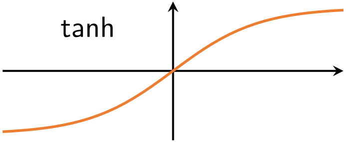
symmetric matrix

- Ferromagnetic: all $\beta \geqslant 0$
- Anti-ferromagnetic: all $\beta \leqslant 0$
- Sherrington-Kirkpatrick: random β
- Note: every 2-spin system is a (limit of an) Ising model

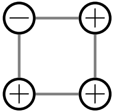
- We have

$$\mathbb{E}[x_v \mid x_{-v}] = \frac{e^c - e^{-c}}{e^c + e^{-c}} = \tanh(c)$$

where $c = h_v + \sum_u \beta_{uv} x_u$.



Ising model

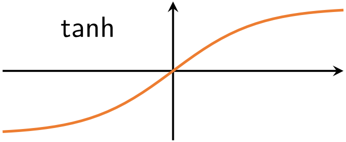


$\mu(x) \propto \exp(\frac{1}{2}\sum_{u,v}\beta_{uv}x_u x_v + \sum_v h_v x_v)$

symmetric matrix

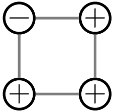
- Ferromagnetic: all $\beta \geqslant 0$
- Anti-ferromagnetic: all $\beta \leqslant 0$
- Sherrington-Kirkpatrick: random β
- Note: every 2-spin system is a (limit of an) Ising model

- We have
$$\mathbb{E}[x_v \mid x_{-v}] = \frac{e^c - e^{-c}}{e^c + e^{-c}} = \tanh(c)$$
where $c = h_v + \sum_u \beta_{uv} x_u$.



- So we get $\mathcal{I}[j \to i] \leqslant$
$$\frac{|\tanh(\beta_{ij}) - \tanh(-\beta_{ij})|}{2} \leqslant |\beta_{ij}|$$

Ising model

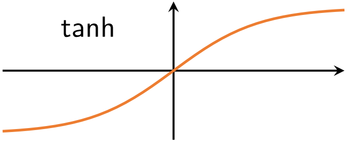


$$\mu(x) \propto \exp\left(\tfrac{1}{2}\sum_{u,v}\beta_{uv}x_u x_v + \sum_v h_v x_v\right)$$

symmetric matrix

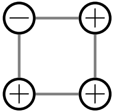
- Ferromagnetic: all $\beta \geqslant 0$
- Anti-ferromagnetic: all $\beta \leqslant 0$
- Sherrington-Kirkpatrick: random β
- Note: every 2-spin system is a (limit of an) Ising model

- We have
$$\mathbb{E}[x_v \mid x_{-v}] = \frac{e^c - e^{-c}}{e^c + e^{-c}} = \tanh(c)$$
where $c = h_v + \sum_u \beta_{uv} x_u$.



- So we get $\mathcal{I}[j \to i] \leqslant$
$$\frac{|\tanh(\beta_{ij}) - \tanh(-\beta_{ij})|}{2} \leqslant |\beta_{ij}|$$
- When β has ℓ_1 row/col norms $\leqslant 1 - \delta$, have fast mixing. 🙂

Ising model

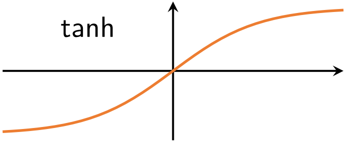


$$\mu(x) \propto \exp\Big(\tfrac{1}{2}\sum_{u,v} \underset{\uparrow}{\beta_{uv}} x_u x_v + \sum_v h_v x_v\Big)$$

symmetric matrix

- Ferromagnetic: all $\beta \geqslant 0$
- Anti-ferromagnetic: all $\beta \leqslant 0$
- Sherrington-Kirkpatrick: random β
- Note: every 2-spin system is a (limit of an) Ising model

- We have
  $$\mathbb{E}[x_v \mid x_{-v}] = \frac{e^c - e^{-c}}{e^c + e^{-c}} = \tanh(c)$$
  where $c = h_v + \sum_u \beta_{uv} x_u$.



- So we get $\mathcal{I}[j \to i] \leqslant$
  $$\frac{|\tanh(\beta_{ij}) - \tanh(-\beta_{ij})|}{2} \leqslant |\beta_{ij}|$$
- When β has ℓ_1 row/col norms $\leqslant 1 - \delta$, have fast mixing. 🙂
- Can be asymptotically tight for certain β (Curie-Weiss).

Dobrushin++

Why not d other than Hamming?

Dobrushin++

Why not d other than Hamming?

- For $c \in \mathbb{R}^n_{\geqslant 0}$, define c-weighted Hamming as

$$d(x, y) = \sum \{c_i \mid x_i \neq y_i\}$$

Dobrushin++

Why not d other than Hamming?

- For $c \in \mathbb{R}^n_{\geqslant 0}$, define c-weighted Hamming as
$$d(x, y) = \sum\{c_i \mid x_i \neq y_i\}$$
- Take X_0, X_0' differing in j. Will produce coupling of X_1, X_1'.

Dobrushin++

Why not d other than Hamming?

- For $c \in \mathbb{R}^n_{\geqslant 0}$, define c-weighted Hamming as

$$d(x, y) = \sum\{c_i \mid x_i \neq y_i\}$$

- Take X_0, X_0' differing in j. Will produce coupling of X_1, X_1'.

Dobrushin++

Why not d other than Hamming?

- For $c \in \mathbb{R}^n_{\geqslant 0}$, define c-weighted Hamming as

$$d(x, y) = \sum\{c_i \mid x_i \neq y_i\}$$

- Take X_0, X'_0 differing in j. Will produce coupling of X_1, X'_1.
- Pick same coord i. Maximally couple replacements.

Dobrushin++

Why not d other than Hamming?

- For $c \in \mathbb{R}^n_{\geqslant 0}$, define c-weighted Hamming as
$$d(x, y) = \sum\{c_i \mid x_i \neq y_i\}$$
- Take X_0, X_0' differing in j. Will produce coupling of X_1, X_1'.
- Pick same coord i. Maximally couple replacements.

Dobrushin++

Why not d other than Hamming?

- For $c \in \mathbb{R}^n_{\geqslant 0}$, define c-weighted Hamming as
$$d(x, y) = \sum\{c_i \mid x_i \neq y_i\}$$
- Take X_0, X_0' differing in j. Will produce coupling of X_1, X_1'.
- Pick same coord i. Maximally couple replacements.
- We get $\mathbb{E}[d(X_1, X_1')] \leqslant$
$\frac{1}{n} \cdot 0 + \frac{1}{n} \sum_{i \neq j}(c_j + \mathcal{I}[j \to i]c_i)$

Dobrushin++

Why not d other than Hamming?

- For $c \in \mathbb{R}^n_{\geqslant 0}$, define c-weighted Hamming as
$$d(x, y) = \sum\{c_i \mid x_i \neq y_i\}$$
- Take X_0, X_0' differing in j. Will produce coupling of X_1, X_1'.
- Pick same coord i. Maximally couple replacements.
- We get $\mathbb{E}[d(X_1, X_1')] \leqslant$
$$\frac{1}{n} \cdot 0 + \frac{1}{n} \sum_{i \neq j} (c_j + \mathcal{I}[j \to i] c_i)$$
- Contraction: $c\,\mathcal{I} \leqslant (1 - \delta) c$

Dobrushin++

Why not d other than Hamming?

- For $c \in \mathbb{R}^n_{\geqslant 0}$, define c-weighted Hamming as
$$d(x, y) = \sum\{c_i \mid x_i \neq y_i\}$$
- Take X_0, X_0' differing in j. Will produce coupling of X_1, X_1'.
- Pick same coord i. Maximally couple replacements.
- We get $\mathbb{E}[d(X_1, X_1')] \leqslant$
$$\frac{1}{n} \cdot 0 + \frac{1}{n} \sum_{i \neq j} (c_j + \mathcal{I}[j \to i] c_i)$$
- Contraction: $c\,\mathcal{I} \leqslant (1 - \delta)c$
- When this happens,
$$\mathcal{W}(\nu P, \nu' P) \leqslant (1 - \delta/n)\, \mathcal{W}(\nu, \nu')$$

Dobrushin++

Why not d other than Hamming?

- For $c \in \mathbb{R}^n_{\geqslant 0}$, define c-weighted Hamming as
$$d(x, y) = \sum\{c_i \mid x_i \neq y_i\}$$
- Take X_0, X_0' differing in j. Will produce coupling of X_1, X_1'.
- Pick same coord i. Maximally couple replacements.
- We get $\mathbb{E}[d(X_1, X_1')] \leqslant$
$$\frac{1}{n} \cdot 0 + \frac{1}{n} \sum_{i \neq j} (c_j + \mathcal{J}[j \to i] c_i)$$
- Contraction: $c\,\mathcal{J} \leqslant (1 - \delta) c$
- When this happens,
$$\mathcal{W}(\nu P, \nu' P) \leqslant (1 - \delta/n)\, \mathcal{W}(\nu, \nu')$$

Implication for mixing

Given $c\,\mathcal{J} \leqslant (1 - \delta)c$, we have
$$t_{\mathsf{mix}}(\epsilon) = O\left(\frac{n}{\delta} \log\left(\frac{n \cdot c_{\mathsf{max}}}{\epsilon \cdot c_{\mathsf{min}}}\right)\right)$$

Dobrushin++

Why not d other than Hamming?

- For $c \in \mathbb{R}^n_{\geqslant 0}$, define c-weighted Hamming as
$$d(x, y) = \sum \{c_i \mid x_i \neq y_i\}$$
- Take X_0, X_0' differing in j. Will produce coupling of X_1, X_1'.
- Pick same coord i. Maximally couple replacements.
- We get $\mathbb{E}[d(X_1, X_1')] \leqslant$
$$\frac{1}{n} \cdot 0 + \frac{1}{n} \sum_{i \neq j} (c_j + \mathcal{I}[j \to i] c_i)$$
- Contraction: $c\,\mathcal{I} \leqslant (1 - \delta) c$
- When this happens,
$$\mathcal{W}(\nu P, \nu' P) \leqslant (1 - \delta/n)\, \mathcal{W}(\nu, \nu')$$

Implication for mixing

Given $c\,\mathcal{I} \leqslant (1 - \delta) c$, we have

$$t_{\mathsf{mix}}(\epsilon) = O\left(\frac{n}{\delta} \log\left(\frac{n \cdot c_{\max}}{\epsilon \cdot c_{\min}}\right)\right)$$

- Slightly careful about $c_{\max}/c_{\min}$.

Dobrushin++

Why not d other than Hamming?

- For $c \in \mathbb{R}^n_{\geqslant 0}$, define c-weighted Hamming as
$$d(x, y) = \sum\{c_i \mid x_i \neq y_i\}$$
- Take X_0, X'_0 differing in j. Will produce coupling of X_1, X'_1.
- Pick same coord i. Maximally couple replacements.
- We get $\mathbb{E}[d(X_1, X'_1)] \leqslant$
$$\frac{1}{n} \cdot 0 + \frac{1}{n} \sum_{i \neq j} (c_j + \mathcal{I}[j \to i] c_i)$$
- Contraction: $c\,\mathcal{I} \leqslant (1 - \delta) c$
- When this happens,
$$\mathcal{W}(\nu P, \nu' P) \leqslant (1 - \delta/n)\, \mathcal{W}(\nu, \nu')$$

Implication for mixing

Given $c\,\mathcal{I} \leqslant (1 - \delta) c$, we have
$$t_{\text{mix}}(\epsilon) = O\left(\frac{n}{\delta} \log\left(\frac{n \cdot c_{\max}}{\epsilon \cdot c_{\min}}\right)\right)$$

- Slightly careful about $c_{\max}/c_{\min}$.
- Influence matrix $\mathcal{I}$ is $\geqslant 0$. "Optimal" choice of c by [Perron-Frobenius] theory is the Perron eigenvector:
$$c\,\mathcal{I} = \lambda_{\max}(\mathcal{I}) c$$

Dobrushin++

Why not d other than Hamming?

- For $c \in \mathbb{R}^n_{\geqslant 0}$, define c-weighted Hamming as
$$d(x,y) = \sum \{c_i \mid x_i \neq y_i\}$$
- Take X_0, X'_0 differing in j. Will produce coupling of X_1, X'_1.
- Pick same coord i. Maximally couple replacements.
- We get $\mathbb{E}[d(X_1, X'_1)] \leqslant$
$$\frac{1}{n} \cdot 0 + \frac{1}{n} \sum_{i \neq j} (c_j + \mathcal{I}[j \to i] c_i)$$
- Contraction: $c\,\mathcal{I} \leqslant (1-\delta)c$
- When this happens,
$$\mathcal{W}(\nu P, \nu' P) \leqslant (1 - \delta/n)\, \mathcal{W}(\nu, \nu')$$

Implication for mixing

Given $c\,\mathcal{I} \leqslant (1-\delta)c$, we have
$$t_{\mathrm{mix}}(\epsilon) = O\left(\frac{n}{\delta} \log\left(\frac{n \cdot c_{\max}}{\epsilon \cdot c_{\min}}\right)\right)$$

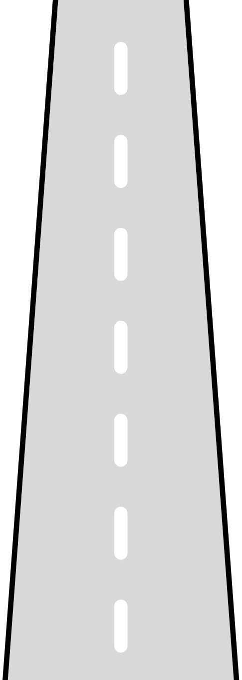
- Slightly careful about $c_{\max}/c_{\min}$.
- Influence matrix $\mathcal{I}$ is $\geqslant 0$. "Optimal" choice of c by [Perron-Frobenius] theory is the Perron eigenvector:
$$c\,\mathcal{I} = \lambda_{\max}(\mathcal{I}) c$$
- We can apply Dobrushin++ as long as $\lambda_{\max}(\mathcal{I}) < 1$

Mixing via Transport

- Path coupling
- Dobrushin's condition
- Hardcore model
- Ising model
- Dobrushin++

Intro to Functional Analysis

- Divergences

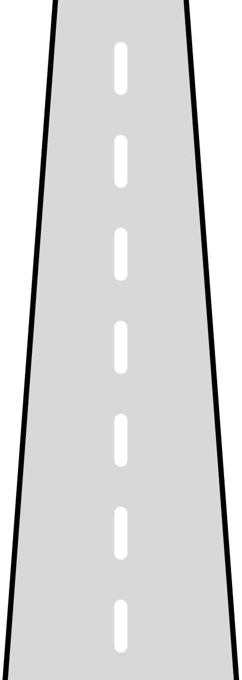


Mixing via Transport

- Path coupling
- Dobrushin's condition
- Hardcore model
- Ising model
- Dobrushin++

Intro to Functional Analysis

- Divergences

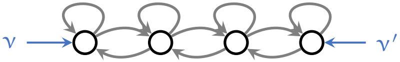


Divergences

- Prevalent strategy for analyzing mixing time: contraction

Divergences

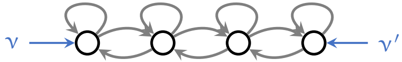
- Prevalent strategy for analyzing mixing time: contraction
- d_{TV} is too crude; doesn't contract every step



$$d_{TV}(\nu P, \nu' P) = d_{TV}(\nu, \nu')$$

Divergences

- Prevalent strategy for analyzing mixing time: contraction
- d_{TV} is too crude; doesn't contract every step



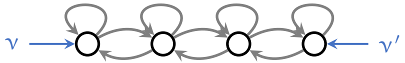
$$d_{TV}(\nu P, \nu' P) = d_{TV}(\nu, \nu')$$

- Fix: use a proxy for d_{TV}
  - Wasserstein distance
  - divergences, variance, entropy

functional analysis

Divergences

- Prevalent strategy for analyzing mixing time: contraction
- d_{TV} is too crude; doesn't contract every step



$$d_{TV}(\nu P, \nu' P) = d_{TV}(\nu, \nu')$$

- Fix: use a proxy for d_{TV}
  - Wasserstein distance
  - divergences, variance, entropy

functional analysis

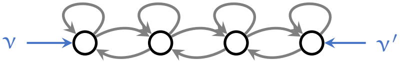
ϕ-entropy

For fn $\phi : \mathbb{R} \to \mathbb{R}$ and $f : \Omega \to \mathbb{R}$ define

$$\mathrm{Ent}_\mu^\phi[f] = \mathbb{E}_\mu[\phi \circ f] - \phi(\mathbb{E}_\mu[f]).$$

Divergences

- Prevalent strategy for analyzing mixing time: contraction
- d_{TV} is too crude; doesn't contract every step



$$d_{TV}(\nu P, \nu' P) = d_{TV}(\nu, \nu')$$

- Fix: use a proxy for d_{TV}
  - Wasserstein distance
  - divergences, variance, entropy

functional analysis

ϕ-entropy

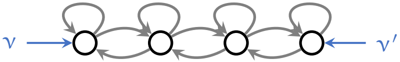
For fn $\phi : \mathbb{R} \to \mathbb{R}$ and $f : \Omega \to \mathbb{R}$ define

$$\mathrm{Ent}_\mu^\phi[f] = \mathbb{E}_\mu[\phi \circ f] - \phi(\mathbb{E}_\mu[f]).$$

- When ϕ is convex, ϕ-entropy is $\geqslant 0$ (Jensen's inequality).

Divergences

- Prevalent strategy for analyzing mixing time: contraction
- d_{TV} is too crude; doesn't contract every step



$$d_{TV}(\nu P, \nu' P) = d_{TV}(\nu, \nu')$$

- Fix: use a proxy for d_{TV}
  - Wasserstein distance
  - divergences, variance, entropy

functional analysis

ϕ-entropy

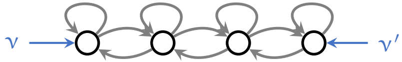
For fn $\phi : \mathbb{R} \to \mathbb{R}$ and $f : \Omega \to \mathbb{R}$ define

$$\mathrm{Ent}_{\mu}^{\phi}[f] = \mathbb{E}_{\mu}[\phi \circ f] - \phi(\mathbb{E}_{\mu}[f]).$$

- When ϕ is convex, ϕ-entropy is $\geqslant 0$ (Jensen's inequality).
- Equal to 0 when f is constant.

Divergences

- Prevalent strategy for analyzing mixing time: contraction
- d_{TV} is too crude; doesn't contract every step



$$d_{TV}(\nu P, \nu' P) = d_{TV}(\nu, \nu')$$

- Fix: use a proxy for d_{TV}
  - Wasserstein distance
  - divergences, variance, entropy

functional analysis

ϕ-entropy

For fn $\phi : \mathbb{R} \to \mathbb{R}$ and $f : \Omega \to \mathbb{R}$ define

$$\mathrm{Ent}_\mu^\phi[f] = \mathbb{E}_\mu[\phi \circ f] - \phi(\mathbb{E}_\mu[f]).$$

- When ϕ is convex, ϕ-entropy is $\geqslant 0$ (Jensen's inequality).
- Equal to 0 when f is constant.

usually f in the literature

ϕ-divergence

For measure ν and dist μ define

$$\mathcal{D}_\phi(\nu \parallel \mu) = \mathrm{Ent}_\mu^\phi\left[\frac{\nu}{\mu}\right]$$

Proxy for d_{TV}

Contraction: $\mathcal{D}_\phi(\nu P \parallel \mu) \leqslant (1-\delta)\, \mathcal{D}_\phi(\nu \parallel \mu)$ for stationary μ.

Proxy for d_{TV}

Contraction: $\mathcal{D}_\phi(\nu P \parallel \mu) \leqslant (1-\delta)\,\mathcal{D}_\phi(\nu \parallel \mu)$ for stationary μ.

Variance

$$\phi(x) := x^2$$

- $\mathrm{Ent}_\mu^\phi[f] = \mathrm{Var}_\mu[f]$
- $\mathcal{D}_\phi(\nu \parallel \mu) = \chi^2(\nu \parallel \mu)$
- It is a proxy by Cauchy-Schwarz:

$$d_{TV}(\nu, \mu) \leqslant O\left(\sqrt{\chi^2(\nu \parallel \mu)}\right)$$

- Contraction related to eigs of P.

called Poincaré inequality

Entropy

$$\phi(x) := x \log x$$

- $\mathrm{Ent}_\mu^\phi[f] = \mathrm{Ent}_\mu[f]$
- $\mathcal{D}_\phi(\nu \parallel \mu) = \mathcal{D}_{KL}(\nu \parallel \mu)$
- It is a proxy by Pinsker:

$$d_{TV}(\nu, \mu) \leqslant O\left(\sqrt{\mathcal{D}_{KL}(\nu \parallel \mu)}\right)$$

- Contraction: very hard!

called modified log-Sobolev inequality