10 FEB: MCMC READING GROUP
PATH - COUPLING

Last Time:

$$\tau_{mix}(\varepsilon) \leq \left\lceil \max_{x,y \in \Omega} \mathbb{E}[\tau_{couple}]\, e \ln\left(\frac{1}{\varepsilon}\right) \right\rceil$$

Usual steps:

1. Define a Markovian Coupling on the same Markov Chain starting at arbitrary states.
2. Bound the Expected time steps for the states to coalesce.
3. Plug in above.

Difficulty: How to bound $\mathbb{E}[\tau_{couple}]$ for arbitrary states?

One way: Let d be a metric on the state space Ω.

$\bar{d}(t) \leq \Pr(X_t \neq Y_t) = \Pr(d(X_t, Y_t) \geq 1)$

By Markov Inequality,

$$\leq \mathbb{E}[d(X_t, Y_t)]$$

If at each time step this expected distance decreases, then we can bound τ_{mix}.

Path couplings helps us create couplings where this happens.

Path Coupling Theorem

Consider a weighted graph $G = (\Omega, E_0)$ with edge weights l, and $l(x,y) \geq 1$.

<u>Note</u>: This need not be the graph of Markov chain transitions.

The <u>length</u> of a path in G, say $x_0, x_1, \ldots, x_r$ is given by -

$$\sum_{i=1}^{r} l(x_{i-1}, x_i).$$

Define a <u>distance metric</u> $d(X,Y)$ to be the shortest length path in the graph

Let $D = \text{diameter}(\Omega) = \max_{x,y} d(x,y)$

<u>Exc</u>: Verify this is a metric.

$d(X,X) = 0$ $\qquad$ $d(X,Y) + d(Y,Z) \geq d(X,Z)$

$d(X,Y) = d(Y,X)$

Theorem

Suppose that for each $(x, y) \in E$ there is a coupling (X_1, Y_1) of $P_r(x, \cdot)$ and $P_r(y, \cdot)$ such that –

$$\mathbb{E}[d(X_1, Y_1)] \leq \beta \, d(x, y)$$

$(\beta \in (0,1))$

Then

$$d(t) \leq D \beta^t$$

Corollary

For $D\beta^t \leq \varepsilon$

$$\Rightarrow \left(\frac{1}{\beta}\right)^t \geq \frac{D}{\varepsilon}$$

$$\Rightarrow t \log\left(\frac{1}{\beta}\right) \geq \log\left(\frac{D}{\varepsilon}\right)$$

$$\Rightarrow t \geq \frac{\log(D) - \log(\varepsilon)}{\log(\beta)^{-1}}$$

$$\boxed{\Rightarrow \tau_{mix}(\varepsilon) \leq \left\lceil \frac{\log(D) - \log(\varepsilon)}{\log(\beta)^{-1}} \right\rceil}$$

Example Application

Boolean Hypercube MC

1. Choose $i \in_R [n]$, $r \in_R \{0,1\}$
2. set $X_{t+1}[i] = r$

   $X_{t+1}[j] = X_t[j]$, $\forall j \neq i$

Here, $\Omega = \{0,1\}^n$

Let $E = \{(x,y) \mid x$ and y differ at exactly 1 position$\}$

Let $d(x,y) = \#$ positions at which x and y differ

(Hamming Metric)

Here $D = n$, since two states can differ at atmost n positions.

Now for any $(x,y) \in E$, WLOG assume that they differ at i'.

Define the coupling -

i) Choose $i \in_R [n]$, $r \in_R \{0,1\}$

ii) set $X_{t+1}[i] = Y_{t+1}[i] = r$

$$X_{t+1}[j] = X_t[j] \quad \forall j \neq i$$

$$Y_{t+1}[j] = Y_t[j] \quad \forall j \neq i$$

Now $d(x, y) = 1$

In order to bound $\mathbb{E}[d(X_1, Y_1)]$,

Case 1: $i \neq i'$

$\Rightarrow d(X_1, Y_1)$ remains same.

Case 2: $i = i'$

$\Rightarrow d(X_1, Y_1)$ reduces by 1.

$$\therefore \mathbb{E}[d(X_1, Y_1)] = 1 - \frac{1}{n}$$

$$\leq \exp(-1/n)$$

$$\Rightarrow \beta = \exp(-1/n)$$

$\therefore$ By path coupling theorem

$$\tau_{mix}(\varepsilon) \leq \frac{\log(n) - \log(\varepsilon)}{\log\left(\exp\left(\frac{-1}{n}\right)^{-1}\right)}.$$

$$\leq n\left(\log(n) - \log(\varepsilon)\right)$$

$$= O(n \log(n))$$

Note: The same MC analysed using coupling needed some intelligently defined Geometric r.v.s. This is much simpler!

Glauber Dynamics

1. Choose a vertex $u \in_R V$, $c \in_R C$
2. If $c \notin \{\sigma_t(w) \mid w \in N(u)\}$ then

$$\sigma_{t+1}(u) = c$$

$$\sigma_{t+1}(w) = \sigma_t(w) \quad \forall\, w \neq u$$

3. Else $\sigma_{t+1} = \sigma_t$

t

Here $\Omega =$ Set of all valid k-colorings

Define $d(\sigma, \tau) = |\{v \mid \sigma(v) \neq \tau(v)\}|$

Diameter $D = n. = |V|$

Let $E = \{(\sigma, \tau) \mid d(\sigma, \tau) = 1\}$

Let $(\sigma_0, \tau_0) \in E$ differ at vertex v.

consider the coupling -

1. Choose $u \in_R V$, $c \in_R C$
2. If $c \notin \{\sigma_t(w) \mid w \in N(u)\}$ then

$$\sigma_{t+1}(u) = c$$

$$\sigma_{t+1}(w) = \sigma_t(w) \quad \forall\ w \neq u$$

3. Else $\sigma_{t+1} = \sigma_t$

4. If $c \notin \{\tau_t(w) \mid w \in N(u)\}$ then

$$\tau_{t+1}(u) = c$$

$$\tau_{t+1}(w) = \tau_t(w) \quad \forall\ w \neq u$$

5. Else $\tau_{t+1} = \tau_t$.

In order to bound $\mathbb{E}[d(\sigma_1, \tau_1)]$,

Case 1: If $u \neq v$ and $u \notin N(v)$

Here distance remains same since the colors of the neighborhood are identical in σ_0 and τ_0.

Case 2: If $u = v$

Distance decreases by 1 if the update goes through in both chains. This occurs when none of the colors in $N(v)$ are chosen.

i.e. $\Pr(u=v,\ c \notin \{\sigma_0(w) \mid w \in N(v)\})$

$\geq \frac{1}{n} \frac{k-\Delta}{k}$ worst case v has Δ neighbors each having a distinct color. (least possible available colors)

Case 3: If $u \in N(v)$

If $c = \sigma(v)$ or $\tau(v)$, the update will go through in only one chain. Hence the distance increases by 1.

This occurs with prob $\leq \frac{2}{kn} \cdot \Delta$

$$\Rightarrow \mathbb{E}[d(\sigma_1, \tau_1)] \leq 1 + \left(-1 \cdot \frac{k-\Delta}{kn} + 1 \cdot \frac{2\Delta}{kn}\right)$$

$$\leq 1 + \frac{1}{kn}(3\Delta - k)$$

For $k = 3\Delta + 1$,

$$\leq 1 - \frac{1}{kn}$$

$$\leq \exp\left(-\frac{1}{kn}\right)$$

By Path Coupling Theorem,

$$\tau_{mix}(\varepsilon) = \frac{\log(n) - \log(\varepsilon)}{1/kn}$$

$$= O(nk \log n)$$

Proof of the Path Coupling Theorem

Suppose the coupling was global. i.e. for all $x, y \in \Omega$ there is a coupling (X_1, Y_1) of $P(x, \cdot)$, $P(\cdot, y)$ such that

$$\mathbb{E}[d(X_1, Y_1)] \leq \beta\, d(X_0, Y_0)$$

Then,

$$\mathbb{E}[d(X^t, Y^t)] \leq \beta^t d(X_0, Y_0)$$

$$\leq \beta^t D$$

Recall,

$$\bar{d}(t) \leq \Pr(X_t \neq Y_t) = \Pr(d(X_t, Y_t) \geq 1)$$

$$\begin{pmatrix}\text{Markov} \\ \text{Inea}\end{pmatrix} \leq \mathbb{E}[d(X_t, Y_t)] \leq \beta^t D$$

Thus if we show that a coupling for every edge implies a global coupling we are done.

<u>Proof</u>: Consider any two ~~states~~ $x', y' \in \Omega$

Let $(x_0, x_1, \ldots, x_r)$ be the min length path.

Let X, Y be a coupling for $P(x', \cdot)$ and $P(y', \cdot)$

Let $A^{(i)}$, $B^{(i)}$ be a coupling for $P(x_i, \cdot)$, $P(x_{i+1}, \cdot)$ such that the contraction property holds.

Now,

$$\mathbb{E}[d(X_1, Y_1)] \leq \mathbb{E}\left[\sum_{i=0}^{r-1} d(A_1^{(i)}, B_1^{(i)})\right]$$

(exc: why is this true)

$$\leq \sum_{i=0}^{r-1} \mathbb{E}\left[d(A_1^{(i)}, B_1^{(i)})\right]$$

(by contraction property)

$$\leq \beta \sum_{i=0}^{r-1} d\left(A_0^{(i)}, B_0^{(i)}\right)$$

$$\leq \beta \cdot d(X_0, Y_0)$$