Randomized Algorithms

Lecture 17: Path Coupling

Raul Garcia-Patron

School of Informatics
University of Edinburgh

Recap: TV and Coupling

Our goal: We want to sample from a MC with stationary distribution π in time poly(n) and $\log(1/\epsilon)$.

- TV distance: $\|D_1 - D_2\| \;=\; \frac{1}{2}\sum_{x\in\Omega} |D_1(x) - D_2(x)|$
- Lower bound mixing time:
  $|D_1(A) - D_2(A)| \leq \max_{A\subseteq\Omega} |D_1(A) - D_2(A)| = \|D_1 - D_2\|$
- Upper-bounds on mixing time (Coupling):
  $\|D_1 - D_2\| \leq \inf \Pr(X \neq Y)$ for a coupling (X, Y) of D_1 and D_2.
- We want to prove that $\|P^t(x,\cdot) - \pi\| \leq \epsilon$ fast enough
- Coupling lemma:
  $\Pr(X_T \neq Y_T | X_0 = x, Y_0 = y) \leq \epsilon \quad \Rightarrow \quad \tau(\epsilon) \leq T$

Mixing time

Definition (Definition 12.2)

Let M be a finite, irreducible and aperiodic Markov chain over the state space Ω and let π be its stationary distribution. We define $\Delta_x(t), \Delta(t)$ as

$$\Delta_x(t) \;=\; \|M^t[x,\cdot] - \pi\|, \quad \Delta(t) \;=\; \max_{x\in\Omega} \Delta_x(t).$$

We also define

$$\tau_x(\epsilon) \;=\; \min\{t : \Delta_x(t) \leq \epsilon\}, \quad \tau(\epsilon) \;=\; \max_{x\in\Omega} \tau_x(\epsilon).$$

1. $\tau(\epsilon)$ is called mixing time.
2. A chain is rapidly mixing if $\tau(\epsilon)$ is polynomial in $\log(1/\epsilon)$ and the size of the problem.

Coupling as upper-bound of TV distance

Definition (Definition 12.2)

A coupling of two probability distributions μ and ν is a pair of random variables (X, Y) defined on a single probability space, i.e., a joint probability distribution q on $\Omega \times \Omega$ such that

$$\sum_{y \in \Omega} q(x, y) = \mu(x) \textit{and} \sum_{x \in \Omega} q(x, y) = \nu(x) \quad (1)$$

Definition (Lemma 12.3)

Given distributions $\mu(x)$ and $\nu(x)$ on state space Ω. All couplings (X, Y) satisfy the condition

$$\inf \Pr(X \neq Y) \geq \|\mu - \nu\|. \quad (2)$$

- This will allow us to upper-bound distances between two Markov chains at step t, and also with respect to the stationary distribution, which leads to upper-bounds on mixing times.

Path coupling: the big picture (Book section 12.6)

Neighborhood: states $y \in \Omega$ reachable from x in a single step.

- Distance $d(X, Y)$: the amount of steps to reach y from x. Neighbors if $d(X, Y) = 1$. Many times $d(X, Y) \leq |V|$.
  - Distance at step t of MC: $d_t = d(X_t, Y_t)$
- $\Pr(X_T \neq Y_T | X_0 = x, Y_0 = y) \leq \max_{x,y} \Pr(d_T \geq 1) \leq \max_{x,y} E[d_T]$
- Our goal is to bound $\max_{x,y} \mathbf{E}[d_T] \leq \epsilon$.
- After some work... (see next slides) $\mathbf{E}[d_{t+1}] =\leq \beta \mathbf{E}[d_t]$, with $\beta < 1$ (Contraction of expect. distance)
- Iterate $\mathbf{E}[d_T] =\leq \beta^T d_0 \leq \beta^T |V|$

Therefore the chain is guaranteed to have mixed for all times, such that $\beta^T |V| \leq \epsilon$, leading to

$$\tau(\epsilon) = \frac{1}{\log(1/\beta)} \left(\log |V| + \log(1/\epsilon)\right).$$

Many times we can write $\beta = e^{-\alpha/|V|}$, leading to $\tau(\epsilon) = \frac{|V|}{\alpha} \left(\log |V| + \log(1/\epsilon)\right)$, where α can itself depend on parameters of the problem.

Path Coupling in a nutshell II

- We define a distance $d(X, Y)$. Neighbors if $d(X, Y) = 1$.
- Our goal is to prove concentration of expect. of distance: $\mathbf{E}[d_{t+1}] \leq \beta \mathbf{E}[d_t]$, with $\beta < 1$.
- Path: $X_t = Z_0, Z_1, ..., Z_{d_t} = Y_t$ where $d(Z_{i+1}, Z_i) = 1$
  - $d_t = \sum_{i=1}^{d_t} d(Z_{i+1}, Z_i)$ (by construction)
- Updated path: $X_{t+1} = Z'_0, Z'_1, ..., Z'_{d_t} = Y_{t+1}$.
  - $d_{t+1} \leq \sum_{i=1}^{d_t} d(Z'_{i+1}, Z'_i)$ (by triangle inequality)

1. For our problem of interest prove $\mathbf{E}[d(Z'_{i+1}, Z'_i)] \leq \beta d(Z_{i+1}, Z_i) = \beta$.
2. Leads to $\mathbf{E}[d_{t+1} | d_t] \leq \sum_{i=1}^{d_t} \mathbf{E}[d(Z'_{i+1}, Z'_i)] \leq \beta d_t$.
3. Then $\mathbf{E}[d_{t+1}] \leq \mathbf{E}[\mathbf{E}[d_{t+1} | d_t]] \leq \beta \mathbf{E}[d_t]$.

Left to do: prove $\mathbf{E}[d(Z'_{i+1}, Z'_i)] \leq \beta$ for our problem of interest.

Glauber dynamics for independent sets

Consider Glauber dynamics to sample from the Gibbs distribution of independent sets of a graph $G = (V, E)$.

- Let $\pi(x)$ the Gibbs distribution on independent sets:

$$\pi(x) = \begin{cases} \frac{\lambda^{|x|}}{Z(\lambda)} & \text{if } x(v)x(w) = 0 \quad \forall \{v, w\} \in E \\ 0 & \text{Otherwise.} \end{cases} \tag{3}$$

where $|x| = \sum_{v \in V} x(v)$ and $Z(\lambda) = \sum_{x \in \chi} \lambda^{|x|}$ normalizes π.

Algorithm GLAUBERIS$(G = (V, E))$

1. Start with an arbitrary IS X_0
2. **for** $i \leftarrow 0$ **to** "whenever"
3. Choose v uniformly at random from V.
4. Set $X_{t+1}(w) = X_t(w) \forall w \neq v$.
5. **if** $\exists w' \in N(v)$, s.t. $X_t(w') = 1$ **set** $X_t(v) = 0$ (M1)
6. **else** $X_{t+1}(v) = 1$ with probability $\lambda/(1 + \lambda)$ (M2)
7. or $X_{t+1}(v) = 0$ with probability $1/(1 + \lambda)$

Proof Strategy

1. Construct a coupling (X_t, Y_t): chose same v for both chains.
2. We define a distance $d_t = d(X_t, Y_t) = |Y_t \setminus X_t| + |X_t \setminus Y_t|$
3. We will construct a path coupling by having a path $X_t = Z_0, Z_1, ..., Z_{d_t} = Y_t$ where Z_{i+1} and Z_i are neighbors.
4. We will prove $\mathbf{E}[d_{t+1} | d_t = 1] \leq 1 - c(\lambda)/n \leq e^{-c(\lambda)/n}$
5. This will lead to $\mathbf{E}[d_{t+1}] \leq e^{-c(\lambda)/n}\mathbf{E}[d_t]$

Therefore the chain is guaranteed to have mixed for all times

$$\tau(\epsilon) = \frac{n}{c(\lambda)} \left(\log n + \log(1/\epsilon)\right).$$

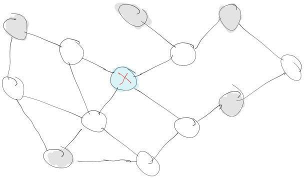
where $c(\lambda) = 1 - \lambda(\Delta - 1)/(1 + \lambda)$, which lead to rapid mixing sufficient condition $\lambda < (\Delta - 1)^{-1}$.

Key aspects of the proof

We will now prove $\mathbf{E}[d_{t+1}|d_t = 1] \leq 1 - \frac{c(\lambda)}{n}$.

- Without loss of generality, let $X_t = I$ and $Y_t = I \cup \{x\}$
- We do not care about how X_t or Y_t change but on when they have different updates that lead to $d_{t+1} \neq d_t$.
- The discussion needs to consider three scenarios:
    1. Case I: when $v = x$ ($d_{t+1} d_t = d_t - 1$).
    2. Case II: when $v \notin N(x) \cup \{x\}$ ($d_{t+1} = d_t + 1$).
    3. Case III: when $v \in N(x)$ ($d_{t+1} = d_t + 1$).

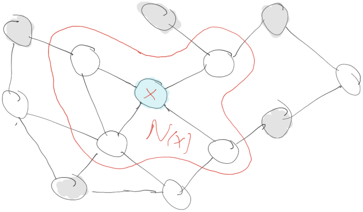
Case I: when $v = x$ $(d_{t+1} - d_t = 0)$



- If we select the vertex v where X_t and Y_t differ, all neighbors are the same in both chains, which implies the chain applies same move to both.
- Because either $X_t(v) = 1(Y_t(v) = 0)$ or $Y_t(v) = 1(X_t(v) = 0)$, all its neighbors have to be 0.
- Therefore the MC will chose move M2.
- Whatever is the update, now we have $X_t(v) = Y_t(v)$ where before we had $X_t(v) \neq Y_t(v)$, which means that $(X_{t+1}, Y_{t+1}) = 0$.

Therefore: $\mathbf{E}[d_{t+1} - d_t | d_t = 1] = -1/n$

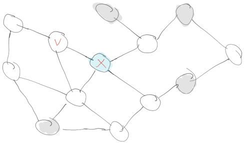
Case II: when $v \notin N(x) \cup \{x\}$ ($d_{t+1} = d_t + 1$)



- All neighbors of $w \notin N(v) \cup \{v\}$ are equal on both chains
- MC implements the same update (M1 or M2 with same output) on both chains.
- Because initially $X_t(w) = Y_t(w)$ and also $X_{t+1}(w) = Y_{t+1}(w)$ we have $d_{t+1} = d_t$.

We obtain: $\mathbf{E}[d_{t+1} - d_t | d_t = 1] = 0$

Case III: when $v \in N(x)$ ($d_{t+1} = d_t + 1$)



- If $v \in N(x)$ and assume without lost of generality that $X_t(x) = 1 (Y_t(x) = 0)$.
- What may happen is that the MC implements move M1 to X_t and M2 to Y_t potentially leading to $Y_{t+1}(v) = 1 \neq X_{t+1}(v)$ and $d_{t+1} = d_t + 1$.
- This can only happen if all neighbors of v in Y_t are 0 and the move M2 select output 1 for v, which has probability $\leq \lambda/(1+\lambda)$.

We obtain:

$$\mathbf{E}[d_{t+1} - d_t | d_t = 1] \leq \frac{\Delta}{n}\left(\frac{\lambda}{1+\lambda}\right). \tag{4}$$

An equivalent argument works for the case $X_t(v) = 0 (Y_t(v) = 1)$.

Final step

We will now prove $\mathbf{E}[d_{t+1}|d_t = 1] \leq 1 - \frac{c(\lambda)}{n}$.

- Without loss of generality, let $X_t = I$ and $Y_t = I \cup \{x\}$
- We do not care about how X_t or Y_t change but on when they have different updates that lead to $d_{t+1} \neq d_t$.
- The discussion depends on what is the neighborhood of y. Because $\Delta = 4$, there exist three cases:
  1. Case I: (when $v = x$) $\mathbf{E}[d_{t+1} - d_t|d_t = 1] = -1/n$.
  2. Case II ($v \notin N(x) \cup \{x\}$: $\mathbf{E}[d_{t+1} - d_t|d_t = 1] = 0$.
  3. Case III ($v \in N(x)$): $\mathbf{E}[d_{t+1} - d_t|d_t = 1] \leq \frac{\Delta}{n}\left(\frac{\lambda}{1+\lambda}\right)$.

We obtain:

$$\mathbf{E}[d_{t+1}|d_t = 1] \leq 1 - \frac{1}{n} + \frac{\Delta}{n}\left(\frac{\lambda}{1+\lambda}\right) = 1 - \frac{1}{n}\left(\frac{1-\lambda(\Delta-1)}{1+\lambda}\right). \quad (5)$$

If $\lambda < (\Delta - 1)^{-1}$ we obtain: $\mathbf{E}[d_{t+1}|d_t = 1] \leq 1 - \frac{c(\lambda)}{n} \leq e^{-c(\lambda)/n}$ with $c(\lambda) = 1 - \lambda(\Delta-1)/(1+\lambda) \geq 0$.

Recap of proof strategy of this example

1. We designed a coupling (X_t, Y_t): chose same v for both chains.
2. We defined a distance $d_t = d(X_t, Y_t) = |Y_t \setminus X_t| + |X_t \setminus Y_t|$
3. We construct a path coupling by having a path $X_t = Z_0, Z_1, ..., Z_{d_t} = Y_t$ where Z_{i+1} and Z_i are neighbors.
4. We proved $\mathbf{E}[d_{t+1}|d_t = 1] \leq (1 - c(\lambda)/n \leq e^{-c(\lambda)/n}$
5. This will lead to $\mathbf{E}[d_{t+1}] \leq e^{-c(\lambda)/n}\mathbf{E}[d_t]$

Therefore the chain is guaranteed to have mixed for all times

$$\tau(\epsilon) = \frac{n}{c(\lambda)} \left(\log n + \log(1/\epsilon)\right).$$

where $c(\lambda) = 1 - \lambda(\Delta - 1)/(1 + \lambda)$, which lead to rapid mixing sufficient condition $\lambda < (\Delta - 1)^{-1}$.

What if $\mathbf{E}[d(Z'_{i+1}, Z'_i)] = 1$?

Lemma (Bubley and Dyer 1997 (cont'd))

If $\mathbf{E}[d_{t+1}|d_t] \leq d_t$ *an there is some* $\alpha > 0$ *such that* $\Pr[d(X', Y') \neq d(X, Y)] \geq \alpha$ *for all* $(X, Y) \in \Omega \times \Omega$*, then*

$$\tau(\epsilon) \leq \left\lceil \frac{e|V|^2}{\alpha} \right\rceil \lceil \ln(\epsilon^{-1}) \rceil.$$