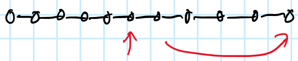
5. Markov chains convergence

Friday, September 17, 2021 3:00 PM

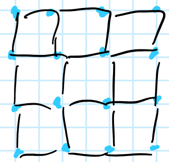
Last time: Markov chains & MCMC

Today: Convergence rates
Maybe applications of MCMC?

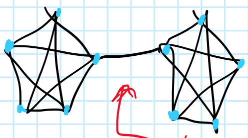
Convergence of random walks on undirected graphs



slow convergence



faster convergence



slow convergence because of bottleneck

Given an edge-weighted undirected graph with weight w_{xy} ($= w_{yx}$) b/t vertices x & y.

Let $w_x = \sum_y w_{xy}$ (degree), and let $p_{xy} = \frac{w_{xy}}{w_x}$ be the Markovian trans. prob.

(Note $p_{xy} \neq p_{yx}$)

(If adjacency matrix $A = \{w_{xy}\}$, let $D = \mathrm{diag}(\underbrace{A\vec{1}}_{\begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}})$, so $P = \underbrace{D^{-1}}_{\begin{bmatrix} w_1^{-1} & & 0 \\ & \ddots & \\ 0 & & w_n^{-1} \end{bmatrix}} A$)

Then the stationary dist. $\vec{\pi}$ has $\pi_x = \frac{w_x}{w_{total}}$, $w_{total} = \sum w_x$ because

$$\pi_x p_{xy} = \frac{w_x}{w_{total}} \cdot \frac{w_{xy}}{w_x} = \frac{w_y}{w_{total}} \cdot \frac{w_{yx}}{w_y} = \pi_y \cdot p_{yx}.$$

Def. Fix $\varepsilon > 0$. The ε-mixing time of a Markov chain is

$$\min_{\vec{p}(0) \text{ prob. dist.}} \left(t \;\middle|\; |\vec{a}(t) - \vec{\pi}|_1 < \varepsilon \right), \text{ where } \vec{a}(t) = \frac{1}{t}\left(\vec{p}(0) + \vec{p}(1) + \dots + \vec{p}(t-1)\right)$$

Def. For a subset $S \subseteq V$ vertices, let $\pi(S) = \sum_{x \in S} \pi_x$. The normalized conductance

$$\Phi(S) = \frac{\sum_{(x,y) \in (S, \bar{S})} \pi_x p_{xy}}{\min(\pi(S), \pi(\bar{S}))}$$

$(\bar{S} = V \setminus S)$

$$\Phi(S) = \frac{\sum_{(x,y) \in (S,\bar{S})} \ldots}{\min(\pi(S), \pi(\bar{S}))}$$

Interpretation: WLOG say $\pi(S) \le \pi(\bar{S})$. Then $\Phi = \sum_{x \in S} \underbrace{\frac{\pi_x}{\pi(S)}}_{\text{prob of being at } x \text{ if in stationarity and in } S} \underbrace{\sum_{y \in \bar{S}} p_{xy}}_{\text{prob of leaving } S \text{ in 1 step.}}$

Note: $\mathbb{E}[\#\text{ of steps to leave } S] = \underbrace{\Phi(S)}_{\text{1 step}} + \underbrace{2(1-\Phi(S))\Phi(S)}_{\text{2 steps}} + \underbrace{3(1-\Phi(S))^2\Phi(S)}_{\text{3 steps}} + \cdots$

Recall $\frac{1}{1-x} = \sum_{i=1}^{\infty} x^i$

$\frac{1}{(1-x)^2} = \sum_{i=1}^{\infty} i x^{i-1}$

$$= \Phi(S)\left[1 + 2(1-\Phi(S)) + 3(1-\Phi(S))^2 + \cdots\right]$$

$$= \Phi(S) \cdot \frac{1}{[1-(1-\Phi(S))]^2} = \frac{1}{\Phi(S)}.$$

Def. The normalized conductance of the Markov chain, denoted Φ, is

$$\Phi = \min_{S \subseteq V,\ S \ne \emptyset} \Phi(S)$$

Note: $\Omega\left(\frac{1}{\Phi}\right)$ is a lower bound on mixing time.

Thm 4.5 The ε-mixing time of a random walk on an undirected graph is $O\left(\frac{\ln\left(\frac{1}{\pi_{min}}\right)}{\Phi^2 \varepsilon^3}\right)$, where $\pi_{min} = \min_x \pi_x$.

proof. Let $t = \frac{c \ln\left(\frac{1}{\pi_{min}}\right)}{\Phi^2 \varepsilon^3}$ for some suitable constant c.

Let $\vec{a} = \vec{a}(t) = \frac{1}{t}\left(\vec{p}(0) + \cdots + \vec{p}(t-1)\right)$.

Need to show $|\vec{a} - \vec{\pi}|_1 < \varepsilon$.

Let $v_i = \frac{a_i}{\pi_i}$. If $v_i > 1$, then node i is "heavy" because more mass is on i at time t than ∞.

WLOG, re-index so $v_1 \ge v_2 \ge v_3 \ge \cdots \ge \underbrace{v_{i_0}}_{\text{last heavy state}} > 1 \ge v_{i_0+1} \ge \cdots \ge v_n$

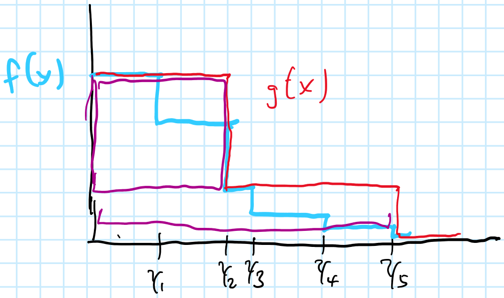
Then $|\vec{a} - \vec{\pi}|_1 = 2\sum_{i=1}^{i_0}(a_i - \pi_i) = 2\sum_{i=i_0+1}^{n}(\pi_i - a_i)$

Then $|\vec{a} - \vec{\pi}|_1 = 2\sum_{i=1}^{i_0}(a_i - \pi_i) = 2\sum_{i \geq i_0+1}(\pi_i - a_i)$

$$= 2\sum_{i=1}^{i_0}(v_i - 1)\pi_i = 2\sum_{i\geq i_0+1}(1-v_i)\pi_i.$$

Let $\gamma_i = \pi_1 + \dots + \pi_i$

Define $f : [0, \gamma_{i_0}] \to \mathbb{R}$ by $f(x) = v_i - 1$ for $x \in [\gamma_{i-1}, \gamma_i]$



Then $\sum_{i=1}^{i_0}(v_i - 1)\pi_i = \int_0^{\gamma_{i_0}} f(x)\,dx$

Also, divide $\{1, \dots, i_0\}$ into contiguous subsets $G_1, G_2, \dots, G_r$ to be specified later

Ex $G_1 = \{1\}$ $\quad G_2 = \{2,3,4\}$ $\quad G_3 = \{5,6\}$ ...

Define $g : [0, \gamma_{i_0}] \to \mathbb{R}$ by $g(x) = u_s - 1$ for $x \in \bigcup_{i \in G_s}[\gamma_{i-1}, \gamma_i]$

$u_s = \max_{i \in G_s} v_i$, the max value in each group.

Since $g(x) \geq f(x)$, $\int_0^{\gamma_{i_0}} f(x)\,dx \leq \int_0^{\gamma_{i_0}} g(x)\,dx$.

$$\int_0^{\gamma_{i_0}} g(x)\,dx = \sum_{t=1}^{r} \pi(G_1 \cup \dots \cup G_t)(u_t - u_{t+1})$$ (where $u_{t+1} = 1$)

Note: If $2\sum_{i \geq i_0+1}(1 - v_i)\pi_i \leq \varepsilon$, then we're already done.

So consider case $\sum_{i\geq i_0+1}(1-v_i)\pi_i > \frac{\varepsilon}{2} \implies \sum_{i \geq i_0+1}\pi_i > \frac{\varepsilon}{2}$

Then, $\forall$ subset A of heavy nodes

$$\min(\pi(A), \pi(\bar{A})) \geq \frac{\varepsilon}{2}\cdot\pi(A)$$

includes all light nodes

(because $\pi(\bar{A}) \geq \frac{\varepsilon}{2} \geq \frac{\varepsilon}{2}\cdot\pi(A)$)

(and $\frac{\varepsilon}{2} < 1$, so $\pi(A) \geq \frac{\varepsilon}{2}\cdot\pi(A)$)

Let $G_1 = \{1\}$. If $G_1, \dots, G_{s-1}$ have already been defined,

Let G_s start $k+1$, where k is the end of the nodes in G_{s-1}.

G_s end on an element ℓ, defined as the largest integer $i > k$ and $i \leq i_0$ s.t. $\sum_{j=k+1}^{\ell}\pi_j \leq \frac{\varepsilon\,\gamma_k}{4}$.

Correction

$\ln(1+x) \leq x$

$\ln(1+x) \geq \frac{x}{2}$ for $x \in [0,2]$

$$\sum_{j=k+1} \pi_j \le \frac{\quad}{4} .$$

$\ln(1+x) \ge \frac{x}{2}$ for $x \in [0,2]$

Then by def, $\quad \gamma_{\ell+1} = \gamma_k + \sum_{j=k+1}^{\ell+1} \pi_j > \gamma_k + \frac{\varepsilon\Phi}{4}\gamma_k = \left(1+\frac{\varepsilon\Phi}{4}\right)\gamma_k .$

Then each group $G_1, \dots, G_r$ scales up γ by $\left(1+\frac{\varepsilon\Phi}{4}\right)$, and $\gamma_1 = \pi_1$, so

$$r \le \log_{\left(1+\frac{\varepsilon\Phi}{4}\right)} \frac{1}{\pi_1} + 2 \le \frac{\ln\frac{1}{\pi_1}}{\ln\left(1+\frac{\varepsilon\Phi}{4}\right)} + 2 \le \frac{\ln\frac{1}{\pi_1}}{\frac{\varepsilon\Phi}{8}} + 2 = \frac{8}{\varepsilon\Phi}\ln\frac{1}{\pi_1} + 2$$

(**Lemma**: $\pi(G_1 \cup \dots \cup G_s)(u_s - u_{s+1}) \le \frac{8}{t\Phi\varepsilon}$ for all $s \le r$.)

Then $\displaystyle\sum_{s=1}^{r} \pi(G_1 \cup \dots \cup G_s)(u_s - u_{s+1}) \le r \cdot \frac{8}{t\Phi\varepsilon}$

$\left(\frac{4}{\varepsilon\Phi}\cdot\ln\frac{1}{\pi_1} + 2\right)$ $\qquad t = \frac{c\ln\left(\frac{1}{\pi_{min}}\right)}{\Phi^2\varepsilon^3}$

$$= \left(\frac{8}{\varepsilon\Phi}\cdot\ln\frac{1}{\pi_1} + 2\right)\cdot\frac{8}{\Phi\varepsilon}\cdot\left[\frac{\Phi^2\varepsilon^3}{c\ln\left(\frac{1}{\pi_{min}}\right)}\right]$$

$$= \frac{8}{\varepsilon\Phi}\cdot\frac{8}{\varepsilon\Phi}\cdot\frac{\Phi^2\varepsilon^3}{c}\cdot\underbrace{\frac{\ln\frac{1}{\pi_1}}{\ln\left(\frac{1}{\pi_{min}}\right)}}_{\le 1} + \frac{16}{\Phi\varepsilon}\cdot\frac{\Phi^2\varepsilon^3}{c\ln\left(\frac{1}{\pi_{min}}\right)}$$

$$\le \frac{64}{c}\cdot\varepsilon + \frac{16\Phi\varepsilon^2}{c\ln\left(\frac{1}{\pi_{min}}\right)} < \varepsilon \quad \text{for some } c \quad \left(\begin{array}{l}\Phi < 1 \\ \ln\frac{1}{\pi_{min}} > 1\end{array}\right)$$

$\square$

proof of lemma:

Consider a starting prob. $\vec{a}$. $\vec{a} - \vec{a}P$ is the net loss in prob. for each state after a step

$\left(\vec{a} = \frac{1}{t}\left(\vec{p}(0) + \dots + \vec{p}(t-1)\right)\right)$

Consider a group $G_s = \{ k+1, \dots, \ell \}$, $k < i_0$

Let $A = \{1, \dots, k\}$.

The net loss for A in one step at time t is

$$\sum_{i=1}^{k} \left(a_i - (\vec{a}P)_i\right) \le \frac{2}{t}$$

(by convergence proof, since $\vec{a} - \vec{a}P = \frac{1}{t}\left(\vec{p}(t) - \vec{p}(0)\right)$)

Consider another way to measure prob. loss. Take the difference of flows from A to $\bar{A}$ and $\bar{A}$ to A.

Consider another way to measure prob. loss. Take the difference of flows from A to $\bar{A}$ and $\bar{A}$ to A.

For any $i < j$

$$\text{net-flow}(i,j) = \text{flow}(i,j) - \text{flow}(j,i) = a_i p_{ij} - a_j p_{ji} = \underline{\pi_i p_{ij}} v_i - \underline{\pi_j p_{ji}} v_j$$

$$= \pi_j p_{ji} (v_i - v_j) \geq 0$$

For any two nodes, there is nonnegative flow from heavier to lighter

Since $\ell \geq k$, flow from A to $\{k+1, k+2, \ldots, \ell\}$ minus flow from $\{k+1, \ldots, \ell\}$ to A is nonnegative.

For $i \leq k$, $j \geq \ell$, $v_i \geq v_j$ and $v_j \leq v_{\ell+1}$, so the net loss from A is at least

$$\sum_{\substack{i \leq k \\ j > \ell}} \pi_j p_{ij} (v_i - v_j) \geq (v_k - v_{\ell+1}) \sum_{\substack{i \leq k \\ j \geq \ell}} \pi_j p_{ji}.$$

Thus, $$(v_k - v_{\ell+1}) \sum_{\substack{i \leq k \\ j > \ell}} \pi_j p_{ji} \leq \frac{2}{t}$$ (by other computation)

But $$\sum_{i=1}^{k} \sum_{j=k+1}^{\ell} \pi_j p_{ji} \leq \sum_{j=k+1}^{\ell} \pi_j \leq \frac{\varepsilon \Phi \pi(A)}{4}$$ by def of ℓ.

By def of Φ and using $\min(\pi(A), \pi(\bar{A})) \geq \frac{\varepsilon}{2} \pi(A)$,

$$\sum_{i \leq k < j} \pi_j p_{ji} \geq \Phi \min(\pi(A), \pi(\bar{A})) \geq \frac{\varepsilon \Phi \gamma_k}{2}$$

Then $$\sum_{\substack{i \leq k \\ j > \ell}} \pi_j p_{ji} = \sum_{i \leq k < j} \pi_j p_{ji} - \sum_{\substack{i \leq k \\ k < j \leq \ell}} \pi_j p_{ji} \geq \frac{\varepsilon \Phi \gamma_k}{2}$$

$$\Rightarrow v_k - v_{\ell+1} \leq \frac{8}{t \varepsilon \Phi \gamma_k}.$$

(similar but easier for $k = i_0$)