STAT 31512 Spring 2024 — Scribe: Zhao Lyu
April 10, 2024 — Lecturer: Frederic Koehler
These notes have not received the scrutiny of publication. They could be missing important references, etc.

Mixing and distance measure

1 Warm-up

We first warm up by computing the spectral gap in an important example. This example behaves quite differently from the Gibbs sampler for product measures which we saw last class.

Question: Compute the spectral gap of simple random walk on a cycle on n vertices.

Let's construct the graph of the cycle on n vertices. The cycle is a 2-regular graph, as shown in fig. 1.

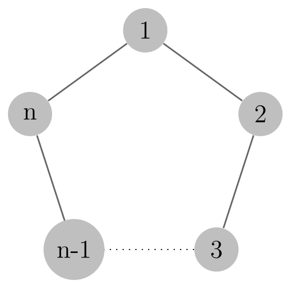

Figure 1: A 2-regular graph with n vertices.

The spectral gap is defined as $\lambda_1 - \lambda_2$, where λ_2 is the second largest eigenvalue of the transition matrix P of the simple random walk. Based on fig. 1, the transition matrix P of the simple random walk on the cycle follows

$$P = \frac{1}{2}A = \frac{1}{2}\begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 1 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & 1 & 0 & \dots \\ & \ddots & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & 1 & 0 & 1 \\ 1 & 0 & \dots & 0 & 1 & 0 \end{bmatrix},$$

and the Laplacian matrix of the cycle is $L = I - P$. We want compute $1 - \lambda_2(P)$.

Answer: $1 - \lambda_2(P) = \frac{c}{n^2}$ for some constant c.

Proof. Let $\varphi_w(x) = e^{2\pi i w x}$ for $w = 0, 1, \dots, n-1$. We have

$$\begin{aligned}(P\varphi_w)(x) &= \frac{1}{2}\left(e^{2\pi i(w+1)x/n} + e^{2\pi i(w-1)x/n}\right) \\ &= e^{2\pi i w x/n}\frac{e^{2\pi i w/n} + e^{-2\pi i w/n}}{2} \\ &= \varphi_w(x)\cos(2\pi w/n).\end{aligned}$$

Thus φ_w is an eigenvector of P with eigenvalue $\cos(2\pi w/n)$. By Taylor expansion, we have

$$\lambda_2 = \cos(2\pi/n) = 1 - \frac{2\pi^2}{n^2} + O(n^{-4})$$
$$1 - \lambda_2 = \frac{2\pi^2}{n^2} + O(n^{-4}).$$

□

Implication: By Poincaré inequality,

$$\begin{aligned} \mathrm{Var}_{X\sim \mathrm{Unif}([n])}(f) &\leq \frac{1}{1-\lambda_2}\langle f, Lf\rangle \\ &\approx \frac{n^2}{c}\langle f, Lf\rangle \end{aligned}$$

2 Distance measure

Definition 1. Let f be a convex function, and let τ, π be probability measures. Define

$$d_f(\tau,\pi) = \mathbb{E}_\pi\left[f\left(\frac{d\tau}{d\pi}\right)\right] - \underbrace{f\left(\mathbb{E}_\pi\left[\frac{d\tau}{d\pi}\right]\right)}_{f(1)}. \tag{1}$$

By Jensen's inequality, $d_f(\tau,\pi) \geq 0$.

Examples:

1. Total variance $f(x) = \frac{1}{2}|x-1|$

$$\begin{aligned} d_{\mathrm{TV}}(\tau,\pi) &= \frac{1}{2}\mathbb{E}\left[\left|\frac{d\tau}{d\pi} - 1\right|\right] \\ &= \frac{1}{2}\sum_x \left|\frac{\tau(x)}{\pi(x)} - 1\right| \pi(x) \quad \text{discrete case} \\ &= \frac{1}{2}\sum_x |\tau(x) - \pi(x)| \\ &= \frac{1}{2}|\tau - \pi|_{l_1} \end{aligned}$$

2. KL divergence $f(x) = x\log(x)$ f is convex becuase $f'(x) = \log(x) + 1$ is an increasing function.

$$\begin{aligned} d_{\mathrm{KL}}(\pi,\tau) &= \mathbb{E}_\pi\left[\frac{d\tau}{d\pi}\log\frac{d\tau}{d\pi}\right] \\ &= \sum_x \tau(x)\log\left(\frac{\tau(x)}{\pi(x)}\right) \\ &= \mathbb{E}_\tau\left[\log\left(\frac{d\tau}{d\pi}\right)\right] \end{aligned}$$

3. χ^2 divergence $f(x) = (x-1)^2$

$$\begin{aligned} d_{\chi^2}(\pi, \tau) &= \mathbb{E}_\pi \left[\left(\frac{d\tau}{d\pi} - 1 \right)^2 \right] \\ &= \mathbb{E}_\pi \left[\left(\frac{d\tau}{d\pi} \right)^2 \right] - 1 \\ &= \mathrm{Var}_\pi \left(\frac{d\tau}{d\pi} \right) \end{aligned}$$

3 Mixing time

Definition 2. ϵ-mixing time of a Markov chain P with stationary distribution π is

$$t_{\text{mix}}(\epsilon) = \inf\{t \geq 1 : \forall \tau \quad d_{\text{TV}}(\tau P^t, \pi) \leq \epsilon\}.$$

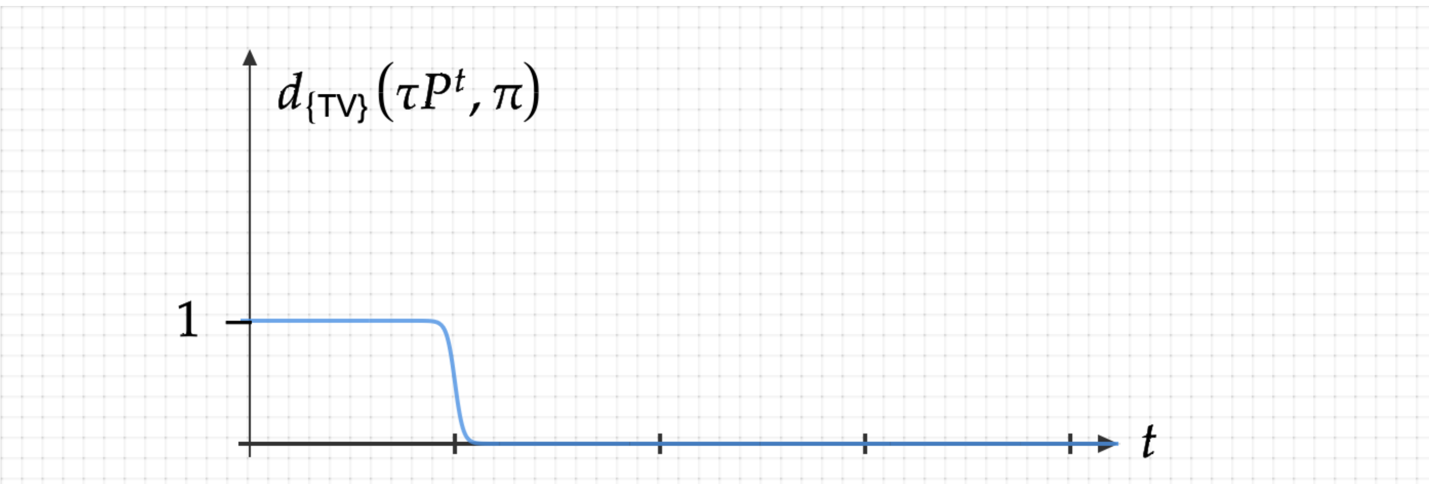

Figure 2: Cut-off phenomenon of mixing time. In many (but not all cases), mixing has a very sharp phase transition between TV distance 1 and 0 so the particular value of ϵ is not so important.

Theorem 1.

$$\underbrace{d_{TV}(\tau, \pi) \leq \sqrt{2 d_{KL}(\tau, \pi)}}_{\textit{Pinsker's inequality}} \leq \sqrt{2 d_{\chi^2}(\tau, \pi)}. \tag{2}$$

Proof. The second inequality follows as below

$$\begin{aligned} d_{\text{KL}}(\tau, \pi) &= \mathbb{E}_\pi \left[\frac{d\tau}{d\pi} \log \left(1 + \frac{d\tau}{d\pi} - 1 \right) \right] \\ &\leq \mathbb{E}_\pi \left[\frac{d\tau}{d\pi} \left(\frac{d\tau}{d\pi} - 1 \right) \right] \\ &= \mathbb{E}_\pi \left[\left(\frac{d\tau}{d\pi} \right)^2 \right] - 1. \end{aligned}$$

□

Theorem 2. *Assume P reversible with respect to π. Then*

$$d_{\chi^2}(\tau P, \pi) \leq \max_{i \neq 1} \left[|\lambda_i|^2\right] d_{\chi^2}(\tau, \pi). \tag{3}$$

Corollary 1.

$$d_{\chi^2}(\tau P^t, \pi) \leq \max_{i \neq 1} \left[|\lambda_i|^2\right]^t d_{\chi^2}(\tau, \pi). \tag{4}$$

This trivially follows from the above theorem. By definition of ϵ-mixing time, we have (we use t in short)

$$t = \frac{\log(d_{\chi^2}(\tau, \pi)/\epsilon^2)}{\log(\max_{i \neq 1} |\lambda_i|)} \implies d_{\chi^2}(\tau P^t, \pi) \leq \epsilon^2.$$

Lemma 1.

$$\frac{d(\tau P)}{d\pi} = P \frac{d\tau}{d\pi} \tag{5}$$

Proof. This proof is for discrete case. Let $\Pi = \mathrm{diag}(\pi)$.

$$\frac{d\tau}{d\pi} = (\tau \Pi^{-1})^\top, \quad \frac{d(\tau P)}{d\pi} = ((\tau P)\Pi^{-1})^\top.$$

By reversibility, $\Pi P = P^\top \Pi$, so $P = \Pi^{-1} P^\top \Pi$. Thus

$$\begin{aligned}
(\tau P \Pi^{-1})^\top &= (\tau \Pi^{-1} P^\top)^\top \\
&= P(\tau \Pi^{-1})^\top \\
&= P \frac{d\tau}{d\pi}.
\end{aligned}$$

Similar proof follows for $\frac{d\tau P}{d\pi}$. □

Let's prove the theorem with above lemma. Let s be the number of state space.

Proof.

$$\begin{aligned}
d_{\chi^2}(\tau P, \pi) &= \mathbb{E}_\pi \left[\left(\frac{d(\tau P)}{d\pi} - 1 \right)^2 \right] \\
&= \mathbb{E}_\pi \left[\left(P \frac{d\tau}{d\pi} - 1 \right)^2 \right] \\
&= \sum_{i=2}^{s} \langle f_i, P \frac{d\tau}{d\pi} \rangle_\pi^2 + \underbrace{\langle f_1, P \frac{d\tau}{d\pi} - 1 \rangle_\pi^2}_{= \mathbb{E}_\pi \left[\frac{d\tau}{d\pi} - 1 \right] = 0} \\
&= \sum_{i=2}^{s} \langle P f_i, \frac{d\tau}{d\pi} \rangle_\pi^2 \\
&\leq \max_{i \neq 1} \left[|\lambda_i|^2\right] \underbrace{\sum_{i=2}^{s} \langle f_i, \frac{d\tau}{d\pi} \rangle_\pi^2}_{d_{\chi^2}(\tau, \pi)}.
\end{aligned}$$

□