Lecture 4 - Variation Distance and Mixing Time

Monday, August 23

Now that we can construct good Markov chains a method is needed to measure how long it takes to "approach" a stationary distribution.

Definition 7.1. Given two distributions μ and π, the total variation (TV) distance is given by

$$\|\mu - \pi\|_{TV} = \sup_{A \subset V} \mu(A) - \pi(A).$$

Exercise 1. Show that $\|\mu - \pi\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \pi(x)|$.

Why is variation distance the appropriate notion to use? Recall the example of counting proper k-colorings in a graph G. In order to approximately count it was necessary to estimate

$$\rho_1 = \frac{|\chi(G : k)|}{|\chi(G - e_1 : k)|}$$

by generating uniform k-colorings of $G - e_1$ and checking the fraction that color G.

Now, suppose that samples are drawn by running a Markov chain for t steps. Then let $\hat{\rho_1}$ be the estimate of ρ_1 given by averaging over an infinite number of such samples.

If A is the collection of proper k-colorings of G then A is a subset of Ω, the proper k-colorings of $G - e_1$. Moreover, $\rho_1 = \pi(A)$ and $\hat{\rho_1} = \mathsf{p}^{(t)}(A)$. Therefore,

$$\pi(A) - \|\mathsf{p}^{(t)} - \pi\|_{TV} \leq \hat{\rho_1} = \pi(A) + (\mathsf{p}^{(t)}(A) - \pi(A)) \leq \pi(A) + \|\mathsf{p}^{(t)} - \pi\|_{TV}$$

and total variation distance bounds the accuracy of our estimate.

We are more interested in knowing relative error. In this case

$$1 - \frac{\|\mathsf{p}^{(t)} - \pi\|_{TV}}{\pi(A)} \leq \frac{\hat{\rho_1}}{\rho_1} \leq 1 + \frac{\|\mathsf{p}^{(t)} - \pi\|_{TV}}{\pi(A)}$$

and in order to estimate ρ_1 then we need $\|\mathsf{p}^{(t)} - \pi\|_{TV}$ to be small relative to $\pi(A)$.

Definition 7.2. The mixing time $\tau(\epsilon)$ is given by

$$\tau(\epsilon) = \max_{\sigma} \min\{t : \|\sigma \, \mathsf{P}^t - \pi\|_{TV} \leq \epsilon\}$$

where σ is an arbitrary initial distribution.

This says that no matter how bad the initial may be, the t-step distribution $\mathsf{p}^{(t)}$ is ϵ close to stationary in TV distance.

The following lemma makes it easier to compute $\tau(\epsilon)$.

Lemma 7.3. *The worst initial distribution is a point mass, i.e.* $\mathsf{p}^{(0)} = 1_{\{x\}}$ *for some* $x \in \Omega$.

Proof. The distribution $\mathsf{p}^{(t)}(y) = \sum_{x\in\Omega} \mathsf{p}^{(0)}(y)\, \mathsf{P}^t(x,y)$, and therefore

$$\begin{aligned}
\|\mathsf{p}^{(t)} - \pi\|_{TV} &= \frac{1}{2}\sum_{y\in\Omega} |\mathsf{p}^{(t)}(y) - \pi(y)| \\
&= \frac{1}{2}\sum_{y\in\Omega}\left|\sum_{x\in\Omega} \mathsf{p}^{(0)}(x)\ \left(\mathsf{P}^t(x,y) - \pi(y)\right)\right| \\
&\le \frac{1}{2}\sum_{y\in\Omega}\sum_{x\in\Omega} \mathsf{p}^{(0)}(x)\ \left|\mathsf{P}^t(x,y) - \pi(y)\right| \\
&= \sum_{x\in\Omega} \mathsf{p}^{(0)}(x)\,\frac{1}{2}\sum_{y\in\Omega}\left|\mathsf{P}^t(x,y) - \pi(y)\right| \\
&\le \max_{x\in\Omega}\frac{1}{2}\sum_{y\in\Omega}\left|\mathsf{P}^t(x,y) - \pi(y)\right|
\end{aligned}$$

□

In general it is nearly impossible to determine $\tau(\epsilon)$ exactly. However, there are a few cases where everything is known. Let us now work out a simple example.

Example 7.4. Determine $\tau(\epsilon)$ for the walk on the uniform two-point space with laziness γ, that is

$$\mathsf{P} = \begin{pmatrix} \gamma & 1-\gamma \\ 1-\gamma & \gamma \end{pmatrix} \qquad \textit{with} \qquad \pi = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

To find the variation distance we need to determine $\mathsf{p}^{(t)}$, and in particular P^t. Follow the standard diagonalization routine.

The eigenvalues of P are easily found to be $\lambda_1 = 1$ and $\lambda_2 = 2\gamma - 1$. The left eigenvector for $\lambda_1 = 1$ is just $\vec{v_1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ while that for $\lambda_2 = 2\gamma - 1$ is $\vec{v_2} = \begin{pmatrix} 1 & -1 \end{pmatrix}$.

Then let

$$V = \begin{pmatrix} \frac{\vec{v_1}}{\|\vec{v_1}\|} & \frac{\vec{v_2}}{\|\vec{v_2}\|} \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

and it follows that

$$\mathsf{P}^t = V\,\Lambda^t\,V^T = \begin{pmatrix} \frac{1}{2} + \frac{1}{2}\,(2\gamma-1)^t, & \frac{1}{2} - \frac{1}{2}\,(2\gamma-1)^t \\ \frac{1}{2} + \frac{1}{2}\,(2\gamma-1)^t, & \frac{1}{2} + \frac{1}{2}\,(2\gamma-1)^t \end{pmatrix}.$$

where $\Lambda = diag(\lambda_1,\, \lambda_2)$.

By Lemma 7.3 it can be assumed that the initial distribution is a point mass. Our problem is completely symmetric, so we may further assume that $\mathsf{p}^{(0)} = \begin{pmatrix} 1 & 0 \end{pmatrix}$. The t-step transition matrix P^t was already found above, so

$$\mathsf{p}^{(t)} = \mathsf{p}^{(0)}\,\mathsf{P}^t = \begin{pmatrix} \frac{1}{2} + \frac{1}{2}\,(2\gamma-1)^t, & \frac{1}{2} - \frac{1}{2}\,(2\gamma-1)^t \end{pmatrix}.$$

Finally, the variation distance is

$$\|\mathsf{p}^{(t)} - \pi\|_{TV} = \frac{1}{2}\sum_{x\in\Omega} |\mathsf{p}^{(t)}(x) - \pi(x)| = \frac{1}{2}\ |2\gamma - 1|^t$$

Solve for when this is less than or equal to ϵ to find that

$$\tau(\epsilon) = \left\lceil \frac{\log(1/2\epsilon)}{\log(1/|2\gamma-1|)} \right\rceil.$$

Since we were able to determine variation distance and mixing time exactly in this problem it will serve as a useful test case as we learn methods for bounding mixing times.

Exercise 2. Generalize this to a walk on the complete graph K_n with

$$\mathsf{P}(x,y) = \begin{cases} \gamma & if\ x = y \\ \frac{1-\gamma}{n-1} & if\ x \neq y \end{cases}.$$

Hint: Make the complete graph into a weighted two-point space.