Lecture 17 - Mixing times and conductance

Friday, September 24

We continue where we left off last class. All of our results today come from a paper of myself [1], with some inspiration from the work of Morris and Peres [2]. *Note that almost all the results on evolving sets that we discuss for the next few weeks are in papers of myself that are currently being written up. Any suggestions regarding notation, clarity, examples, or whatever you like would be appreciated.*

Our goal is to say something about mixing times. Let $\tau(\epsilon)$ denote normal total variation mixing time, $\mathsf{D}(\epsilon)$ denote time to reach distance ϵ in informational divergences, and $L^2(\epsilon)$ denote the same for L^2-distance.

Corollary 17.1. *If $\mathcal{M}$ is a Markov chain with stationary distribution π then for any starting distribution $\mathsf{p}^{(0)}$ the n-step distribution $\mathsf{p}^{(n)} = \mathsf{p}^{(0)}\,\mathsf{P}^n$ satisfies*

$$\begin{aligned}
\|\mathsf{p}^{(n)} - \pi\|_{TV} &\leq \mathcal{C}_{TV}^n\,(1-\pi_*) \quad , \quad & \tau(\epsilon) &\leq \frac{\log \epsilon^{-1}}{1-\mathcal{C}_{TV}}, \\
\mathsf{D}(\mathsf{p}^{(n)}\|\pi) &\leq \mathcal{C}_D^n\,\log \pi_*^{-1} \quad , \quad & \mathsf{D}(\epsilon) &\leq \frac{\log\log \pi_*^{-1} + \log \epsilon^{-1}}{1-\mathcal{C}_D}, \\
\|\mathsf{p}^{(n)} - \pi\|_{2,\pi} &\leq \mathcal{C}_{L^2}^n \sqrt{\frac{1-\pi_*}{\pi_*}} \quad , \quad & L^2(\epsilon) &\leq \frac{\frac{1}{2}\log \pi_*^{-1} + \log \epsilon^{-1}}{1-\mathcal{C}_{L^2}},
\end{aligned}$$

where $\pi_* = \min_{y\in V}\pi(y)$ and for $A \subset \Omega$ let

$$\begin{aligned}
\mathcal{C}_{TV}(A) &= \int_0^1 \frac{\pi(A_u)(1-\pi(A_u))}{\pi(A)(1-\pi(A))}\,du, \\
\mathcal{C}_D(A) &= \int_0^1 \frac{\pi(A_u)\log\frac{1}{\pi(A_u)}}{\pi(A)\log\frac{1}{\pi(A)}}\,du, \\
\mathcal{C}_{L^2}(A) &= \int_0^1 \sqrt{\frac{\pi(A_u)(1-\pi(A_u))}{\pi(A)(1-\pi(A))}}\,du,
\end{aligned}$$

while $\mathcal{C}_{TV} = \max_{\pi(A)\leq 1/2}\mathcal{C}_{TV}(A)$, $\mathcal{C}_D = \max_{A\subset\Omega}\mathcal{C}_D(A)$ and $\mathcal{C}_{L^2} = \max_{\pi(A)\leq 1/2}\mathcal{C}_{L^2}(A)$.

Proof. We work out only the total variation case. The other cases are similar.

Recall from Lemma 7.3 that the worst initial distribution is a point, so it suffices to consider the case where the initial distribution is $\delta_{\{x\}}$ for some $x \in \Omega$. Let $\mathsf{E}_A f(\pi(S_n))$ denote the expectation when $S_0 = A$.

$$\begin{aligned}
\frac{1}{\pi(x)}\,\mathsf{E}_{\{x\}}\pi(S_n)(1-\pi(S_n)) &= \sum_{S\subset\Omega}\left(\int_0^1 \pi(S_u)(1-\pi(S_u))\,du\right)\frac{\mathsf{K}^{n-1}(\{x\},S)}{\pi(x)} \\
&= \sum_{S_{n-1}\subset\Omega}\mathcal{C}_{TV}(S_{n-1})\,\pi(S_{n-1})(1-\pi(S_{n-1}))\,\frac{\mathsf{K}^{n-1}(\{x\},S_{n-1})}{\pi(x)} \\
&= \left(\mathsf{E}_{\{x\}}\prod_{i=0}^{n-1}\mathcal{C}_{TV}(S_i)\right)\frac{\pi(x)(1-\pi(x))}{\pi(x)} \\
&\leq (1-\pi(x))\,\mathcal{C}_{TV}^n
\end{aligned}$$

The final equality followed by induction, the inequality is because $\mathcal{C}_{TV}$ is the worst case.

For the restriction to $\pi(A) \leq 1/2$ observe that $(A^c)_u = (A_{1-u})^c$. But then $\mathcal{C}_{TV}(A^c) = \mathcal{C}_{TV}(A)$.

Bounds on the various mixing times follow immediately from the bounds on the distances. □

The quantity $\mathcal{C}_f$ is a weighted measure of congestion. If we want to understand the mixing in various distances then this is only a matter of choosing the correct function f to weight $\pi(S_n)$ by.

Just how strong Corollary 17.1 is can be seen in the following example.

Example 17.2. Given $\alpha \in [0,1]$ consider the random walk on K_n with $\mathsf{P}(x,y) = (1-\alpha)/n$ for all $y \neq x$ and $\mathsf{P}(x,x) = \alpha + (1-\alpha)/n$, that is, choose a point uniformly at random and move there with probability $1-\alpha$, otherwise do nothing. Then

$$\pi(A_u) = \begin{cases} 0 & if\ u > \alpha + (1-\alpha)\pi(A), \\ \pi(A) & if\ u > (1-\alpha)\pi(A), \\ 1 & otherwise\,. \end{cases}$$

A quick calculation shows that $\mathcal{C}_{TV} = \mathcal{C}_D = \mathcal{C}_{L^2} = \alpha$, and so Corollary 17.1 implies $\|\mathsf{p}^{(t)} - \pi\|_{TV} \leq \alpha^t\,(1-1/n)$, $\mathsf{D}(\mathsf{p}^{(t)}\|\pi) \leq \alpha^t\,\log n$ and $\|\mathsf{p}^{(t)} - \pi\|_{L^2} \leq \alpha^t\,\sqrt{n-1}$.

When $\alpha \in \left[\frac{-1}{n-1},\,0\right)$ then

$$\pi(A_u) = \begin{cases} 0 & if\ u > (1-\alpha)\pi(A), \\ \pi(A^c) & if\ u > \alpha + (1-\alpha)\pi(A), \\ 1 & otherwise \end{cases}$$

Once again, an easy calculation shows that $\mathcal{C}_{TV} = \mathcal{C}_{L^2} = -\alpha$ and the corollary implies $\|\mathsf{p}^{(t)} - \pi\|_{TV} \leq (-\alpha)^t\,(1-1/n)$ and $\|\mathsf{p}^{(t)} - \pi\|_{L^2} \leq (-\alpha)^t\,\sqrt{n-1}$.

The t step distribution is $\mathsf{p}^{(t)}(x,x) = \frac{1}{n} + \alpha^t\left(1-\frac{1}{n}\right)$ and $\mathsf{p}^{(t)}(x,y) = \frac{1}{n} - \frac{\alpha^t}{n}$ for all $y \neq x$. It follows that when $\alpha \geq 0$ then $\mathsf{D}(\mathsf{p}^{(t)}\|\pi) = (1+o_n(1))\alpha^t \log n$ as $n \to \infty$ and our bound is asymptotically correct, while for general α we have $\|\mathsf{p}^{(t)} - \pi\|_{TV} = |\alpha|^t\,(1-1/n)$ and $\|\mathsf{p}^{(t)} - \pi\|_{L^2} = |\alpha|^t\,\sqrt{n-1}$, and our total variation and L^2 bounds are exact at all times t, even for non-lazy chains!

We are now ready to state our main result, upper and lower bounds on the various $1-\mathcal{C}$ quantities in terms of $\Psi(A)$. Even for lazy Markov chains these can lead to better mixing time bounds than have been shown by other authors (a few examples are given in the next lecture).

Theorem 17.3. *If $\mathcal{M}$ is an irreducible Markov chain and $A \subset \Omega$, then*

$$\begin{aligned} \tilde{\phi}(A) &\geq 1 - \mathcal{C}_{L^2}(A) &&\geq 1 - \sqrt{1-\tilde{\phi}(A)^2} \geq \tilde{\phi}(A)^2/2 \\ \tilde{\phi}(A) &\geq 1 - \mathcal{C}_D(A) &&\geq \frac{2\phi(A)^2}{\log(1/\pi(A))} \\ \tilde{\phi}(A) &\geq 1 - \mathcal{C}_{TV}(A) &&\geq 4\tilde{\phi}(A)^2\pi(A)(1-\pi(A)) \end{aligned}$$

where $\phi(A) = \frac{\Psi(A)}{\pi(A)}$ *and* $\tilde{\phi}(A) = \frac{\Psi(A)}{\pi(A)\,\pi(A^c)}$.

For lazy Markov chains $\Psi(A) = \mathsf{Q}(A,A^c)$ and so $\phi(A) = \Phi(A) = \frac{\mathsf{Q}(A,A^c)}{\pi(A)}$ and $\tilde{\phi}(A) = \tilde{\Phi}(A) = \frac{\mathsf{Q}(A,A^c)}{\pi(A)\pi(A^c)}$.

Proof. Jensen's inequality is the key to all the upper and lower bounds. We give only the lower bound on $\mathcal{C}_{L^2}(A)$ because it is the most interesting for our applications; the other lower bounds can be proven similarly. The upper bound require a more careful use of Jensen, however, rather than detail it here a more elegent argument for all of these cases will be given later next week.

$$\begin{aligned}
\mathcal{C}_{L^2}(A) &= \int_0^{\wp} \sqrt{\frac{\pi(A_u)(1-\pi(A_u))}{\pi(A)\pi(A^c)}}\,du + \int_{\wp}^1 \sqrt{\frac{\pi(A_u)(1-\pi(A_u))}{\pi(A)\pi(A^c)}}\,du \\
&\leq \wp \sqrt{\left(1+\frac{\Psi(A)}{\wp\,\pi(A)}\right)\left(1-\frac{\Psi(A)}{\wp\,\pi(A^c)}\right)} + (1-\wp)\sqrt{\left(1-\frac{\Psi(A)}{(1-\wp)\,\pi(A)}\right)\left(1+\frac{\Psi(A)}{(1-\wp)\,\pi(A^c)}\right)} \\
&= \sqrt{\left(\wp+\tilde{\phi}(A)\,\pi(A^c)\right)\left(\wp-\tilde{\phi}(A)\,\pi(A)\right)} + \sqrt{\left(1-\wp-\tilde{\phi}(A)\,\pi(A^c)\right)\left(1-\wp+\tilde{\phi}(A)\,\pi(A)\right)}
\end{aligned}$$

The inequality is by Jensen's inequality applied to $f(x)=\sqrt{x(1-x)}$, $g(u)=\pi(A_u)$ and probability measure $\frac{du}{\wp}$ on $[0,\wp]$ or $\frac{du}{1-\wp}$ on $[\wp,1]$.

Suppose that $\sqrt{X\,Y}+\sqrt{(1-X)(1-Y)}\leq\sqrt{1-(X-Y)^2}$ when $X,\,Y\in[0,1]$. Let $X=\wp+\tilde{\phi}(A)\,\pi(A^c)$ and $Y=\wp-\tilde{\phi}(A)\,\pi(A)$. It is easily checked that $X,Y\in[0,1]$, and the bound on $\mathcal{C}_{L^2}(A)$ follows immediately.

To prove the inequality let $g(X,Y)=\sqrt{X\,Y}+\sqrt{(1-X)(1-Y)}$. Then

$$g(X,Y)^2 = 1-(X+Y)+2\,X\,Y+\sqrt{[1-(X+Y)+2\,X\,Y]^2-[1-2(X+Y)+(X+Y)^2]}\,.$$

Now, $\sqrt{A^2-B}\leq A-B$ if $A^2\geq B$ and $A\leq\frac{1+B}{2}$ (square both sides to show this). These conditions are easily verified with $A=1-(X+Y)+2\,X\,Y$ and $B=1-2(X+Y)+(X+Y)^2$, and so

$$\begin{aligned}
g(X,Y)^2 &\leq 2\,[1-(X+Y)+2\,X\,Y]-\left[1-2(X+Y)+(X+Y)^2\right] \\
&= 1+2\,X\,Y-X^2-Y^2 = 1-(X-Y)^2
\end{aligned}$$

□

References

[1] R. Montenegro. Evolving sets and bounds on various mixing quantities. *Preprint*, 2004.

[2] B. Morris and Y. Peres. Evolving sets and mixing. *Proc. 35th Annual ACM Symposium on Theory of Computing*, pages 279–286, 2003.