Lecture 16 - Evolving sets and distances

Wednesday, September 22

The past five weeks have been spent developing a basic theory for mixing times based on analytic / functional arguments. Today we return to methods discussed in Lecture 10 and the Seminar, on isoperimetric (geometric) arguments for studying mixing times.

Let me refresh your memory on a bit of notation. Recall that a Markov chain can be represented by a directed weighted graph, with vertices given by elements of the state space and edge weights $q(x, y) = \pi(x)\mathsf{P}(x, y)$. A reversible Markov chain is then represented by an undirected graph, as $q(x, y) = \pi(x)\mathsf{P}(x, y) = \pi(y)\mathsf{P}(y, x) = q(y, x)$. If $A \subset \Omega$ then let the *ergodic flow* from a set A to a set B be given by $\mathsf{Q}(A, B) = \sum_{x\in A,\, y\in A^c} \pi(x)\mathsf{P}(x, y)$. Then $\mathsf{Q}(A, A^c)$ measures the weight of the edges from a set A to its complement. Lawler and Sokal, and Jerrum and Sinclair showed a discrete Markov version of Cheeger's Inequality, that $2\Phi \geq \lambda \geq \Phi^2/2$, where $\Phi = \min_{A\subset\Omega} \frac{\mathsf{Q}(A,A^c)}{\min\{\pi(A),\pi(A^c)\}}$. This says that a Markov chain mixes fast if and only if there are always a lot of edges out of every set A, relative to the set size. That is, if there are no edge bottlenecks.

We will prove various mixing time bounds, including variations on Cheeger's inequality. Our main tool will be the *Evolving Set Process* of Morris and Peres [2]. This is a random walk on subsets of the state space Ω which exactly mimics the behaviour of the original Markov chain, and thus makes it easy to translate a Markov problem into a set problem.

Definition 16.1. The *evolving set process* is defined inductively. Let $S_0 = A$ be some initial set. If the current state is the set $A \subseteq \Omega$, a step is made by choosing a value of $u \in [0, 1]$ uniformly at random, and then moving to the set $A_u = \{y \in \Omega : \mathsf{Q}(A, y) \geq u\,\pi(y)\}$. The walk is denoted by S_0, S_1, S_2, …, S_n, with $\mathsf{P}_A(y \in S_n)$ indicating the probability that if $S_0 = A$ then $y \in S_n$. Also, let $\mathsf{K}^n(A, S)$ denote the probability of the set process evolving from A to S after n steps, and $\mathsf{E}_A f(\pi(S_n))$ be an expectation given that $S_0 = A$.

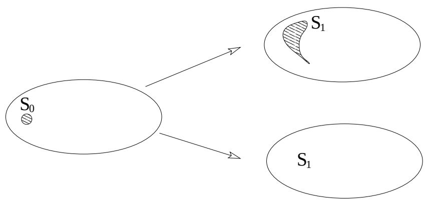

Figure 1: When u is small then the set grows, but when u is big then it shrinks.

Remarkably, the n-step probability distribution $\mathsf{P}^n(x, y)$ of the original Markov chain can be written *exactly* in terms of this random walk on sets.

Lemma 16.2. *Given an irreducible Markov chain, for all $n \geq 0$ and $x, y \in \Omega$ we have*

$$\mathsf{P}^n(x, y) = \frac{\pi(y)}{\pi(x)} \mathsf{P}_{\{x\}}(y \in S_n)$$

Proof. Our proof is taken directly from Morris and Peres [2]. We prove this by induction on n. The case $n = 0$ is trivial. Fix $n > 0$ and suppose that the result holds for $n-1$. Let U be the uniform random variable used to generate S_n from S_{n-1}. Then

$$\begin{aligned}
\mathsf{P}^n(x,y) &= \sum_{z\in\Omega} p^{n-1}(x,z)p(z,y) \\
&= \sum_{z\in\Omega} \mathsf{P}_{\{x\}}(z\in S_{n-1})\frac{\pi(z)}{\pi(x)}p(z,y) \\
&= \frac{\pi(y)}{\pi(x)}\mathsf{E}_{\{x\}}\left(\frac{1}{\pi(y)}Q(S_{n-1},y)\right) = \frac{\pi(y)}{\pi(x)}\mathsf{P}_{\{x\}}(y\in S_n).
\end{aligned}$$

□

Another curious property is the following.

Lemma 16.3. *The sequence $\{\pi(S_n)\}_{n\geq 0}$ forms a martingale, that is $\mathsf{E}\pi(S_n) = \pi(S_0)$.*

Proof. Again, taken from [2]. We have

$$\begin{aligned}
E\left(\pi(S_{n+1})\,|S_n\,\right) &= \sum_{y\in\Omega}\pi(y)\mathsf{P}\left(y\in S_{n+1}\,|S_n\,\right) \\
&= \sum_{y\in\Omega} Q(S_n,y) = \pi(S_n).
\end{aligned}$$

□

A few more facts are not needed immediately, but will prove useful later. We leave the proofs to the reader as an exercise.

- If $\mathcal{M}$ is lazy then $\mathsf{Q}(A,A^c) = \int_0^{1/2}(\pi(A_u)-\pi(A))\,du = \int_{1/2}^1(\pi(A)-\pi(A_u))\,du.$
- If $\mathcal{M}$ is lazy then $\Psi(A) := \frac{1}{2}\int_0^1|\pi(A_u)-\pi(A)|\,du = \mathsf{Q}(A,A^c)$.
- In general, if $\wp = \max\{y:\,\pi(A_y)\geq\pi(A)\}$ then $\Psi(A) = \int_0^{\wp}(\pi(A_u)-\pi(A))\,du = \int_{\wp}^1(\pi(A)-\pi(A_u))\,du.$

Armed with only the two lemmas on evolving sets we can show a connection between convergence of the evolving set walk and convergence of the original Markov chain.

Theorem 16.4. *If $\mathcal{M}$ is an irreducible Markov chain with stationary distribution π then*

$$\begin{aligned}
\|\mathsf{P}^n(x,\cdot)-\pi\|_{TV} &\leq \frac{1}{\pi(x)}\,\mathsf{E}_{\{x\}}\,\pi(S_n)(1-\pi(S_n)), \\
\mathsf{D}(\mathsf{P}^n(x,\cdot)\|\pi) &\leq \frac{1}{\pi(x)}\,\mathsf{E}_{\{x\}}\,\pi(S_n)\log\frac{1}{\pi(S_n)}, \\
\|\mathsf{P}^n(x,\cdot)-\pi\|_{2,\pi} &\leq \frac{1}{\pi(x)}\,\mathsf{E}_{\{x\}}\sqrt{\pi(S_n)(1-\pi(S_n))},
\end{aligned}$$

for any $x\in\Omega$.

We require a simple fact from analysis. Recall that a *convex function* f is such that for every x and y in the domain, and every $\lambda\in[0,1]$ then $f(\lambda\,x+(1-\lambda)\,y)\leq\lambda\,f(x)+(1-\lambda)\,f(y)$. This just says that the function is bowl shaped, or more formally the average of the values at endpoints is at least as big as the value of the function evaluated at the average.

Jensen's Inequality says that if g is a function with domain $[a,b]$, μ is a probability measure on $[a,b]$, and if a function $f(x)$ is convex on the range of g, then

$$f\left(\int_a^b g(x)\,\mu(dx)\right) \le \int_a^b f\circ g(x)\,\mu(dx)\,.$$

Jensen's inequality just says that the average value of a convex function is at least as big as the value of that function evaluated at the average. This follows pretty much immediately from the definition of convexity.

Proof of Theorem. This comes from a paper of myself [1]. We begin with the first statement. Let $f(y) = (1-y)^+ = \max\{1-y,\,0\}$.

$$\begin{aligned}
\|\mathsf{P}^n(x,\cdot)-\pi\|_{TV} &= \sum_{y\in\Omega} f\left(\frac{\mathsf{P}^n(x,y)}{\pi(y)}\right)\pi(y) \\
&= \sum_{y\in\Omega} f\left(\frac{\mathsf{P}_{\{x\}}(y\in S_n)}{\pi(x)}\right)\pi(y) && \text{introduce evolving sets} \\
&= \sum_{y\in\Omega} f\left(\sum_{S_n\subset\Omega}\frac{1_{y\in S_n}}{\pi(S_n)}\,\frac{\pi(S_n)\,\mathsf{K}^n(\{x\},S_n)}{\pi(x)}\right)\pi(y) && \text{rewrite with indicator} \\
&\le \sum_{y\in\Omega}\sum_{S_n\subset\Omega} f\left(\frac{1_{y\in S_n}}{\pi(S_n)}\right)\frac{\pi(S_n)\,\mathsf{K}^n(\{x\},S_n)}{\pi(x)}\,\pi(y) && \text{pull out sum with Jensen} \\
&= \sum_{S_n\subset\Omega}\left[\sum_{y\in\Omega} f\left(\frac{1_{y\in S_n}}{\pi(S_n)}\right)\pi(y)\right]\frac{\pi(S_n)\,\mathsf{K}^n(\{x\},S_n)}{\pi(x)} && \text{change order of summation} \\
&= \sum_{S_n\subset\Omega}\left[\pi(S_n)\,f(\ge 1) + (1-\pi(S_n))\,f(0)\right]\frac{\pi(S_n)\,\mathsf{K}^n(\{x\},S_n)}{\pi(x)} && \text{evaluate sum} \\
&= \sum_{S_n\subset\Omega}(1-\pi(S_n))\,\frac{\pi(S_n)\,\mathsf{K}^n(\{x\},S_n)}{\pi(x)}
\end{aligned}$$

The same method holds for informational divergence with $f(x) = x\log x$.

For the χ^2 distance let $f(x) = (1-x)^2$, but use Minkowski's inequality

$$\forall p\ge 1\ :\ \left(\int\left|\int g(x,y)\,dx\right|^p dy\right)^{1/p} \le \int\left(\int |g(x,y)|^p\,dy\right)^{1/p} dx,$$

instead of Jensen's and the change in order of integration (Minkowski just says that the p-norm of a sum is dominated by the sum of the p-norms, basically a generalized triangle inequality). We write a few of the steps for clarity, the remaining steps are the same.

$$\begin{aligned}
\|\mathsf{P}^n(x,\cdot)-\pi\|_{2,\pi} &= \left(\sum_{v\in\Omega}\left(\sum_{S_n\subset\Omega}\left(\frac{1_{y\in S_n}}{\pi(S_n)}-1\right)\frac{\pi(S_n)\,\mathsf{K}^n(\{x\},S_n)}{\pi(x)}\right)^2\pi(v)\right)^{1/2} \\
&\le \sum_{S_n\subset\Omega}\left[\sum_{v\in\Omega}\left(\frac{1_{y\in S_n}}{\pi(S_n)}-1\right)^2\pi(v)\right]^{1/2}\frac{\pi(S_n)\,\mathsf{K}^n(\{x\},S_n)}{\pi(x)} \\
&= \sum_{S_n\subset\Omega}\sqrt{\frac{1-\pi(S_n)}{\pi(S_n)}}\,\frac{\pi(S_n)\,\mathsf{K}^n(\{x\},S_n)}{\pi(x)}
\end{aligned}$$

$\square$

Remark: The **chi-square distance** $\|\mu - \pi\|_{\chi^2(\pi)} = \left\|1 - \frac{\mu}{\pi}\right\|_{2,\pi}$ *is often used in the literature. Squaring the* L^2 *bound above gives a result for chi-square distance as well.*

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.