Lecture 28 : The Spectral Gap

Stat 205B *Lecturer: Ben Hough* Scribe: *Ben Hough* <jbhough@math.berkeley.edu>

28.1 Introduction

In this lecture we will attempt to give a quantitative answer to the question "How long does it take for an irreducible finite state Markov chain to converge to equilibrium?" This is a question of significant practical importance, as Markov Chain Monte Carlo simulations are used widely in the scientific community to simulate Gibb's measures and to derive approximate solutions to difficult combinatorial questions. For example MCMC algorithms may be used to study the statistical properties of gases, model the flow of information in a network or count the number of dimer coverings of a finite graph.

Of course, the first question that one must answer before using an MCMC algorithm is "how long does the program need to run?" If the program doesn't run long enough, the distribution generated by the simulation will differ significantly from the equilibrium distribution leading to invalid results. On the other hand, running the simulation for too long can be very costly in terms of processing time. Thus, it is of immense practical importance to derive sharp quantitative bounds on the rates of convergence of Markov chains.

Unfortunately, this is a very difficult problem to solve in general, but significant progress has been made using analytic methods. In what follows, we shall shall introduce these techniques and illustrate their applications. For simplicity, we shall deal only with continuous time Markov Chains, although with some work many of these results may be extended to discrete time. Our discussion is closely based on lecture notes by Laurent Saloff-Coste[1].

28.2 Basic Definitions

We shall work on a finite state space $\mathcal{X}$. Recall that a Markov operator K with kernel $K(x,y)$ satisfies

$$Kf(x) = \sum_{y\in\mathcal{X}} K(x,y)f(y) \tag{28.1}$$

where $K(x,y)$ is a stochastic matrix. The continuous time semigroup associated to K is defined by

$$H_t f(x) = e^{-t(I-K)} = e^{-t}\sum_{i=0}^{\infty}\frac{t^i K^i f}{i!} \tag{28.2}$$

with kernel

$$H_t(x,y) = e^{-t}\sum_{i=0}^{\infty}\frac{t^i K^i(x,y)}{i!}. \tag{28.3}$$

[1]Saloff-Coste, et al., Lectures on Probability and Statistics, Springer.

We define $H_t^x(y) = H_t(x,y)$, which is a probability measure on $\mathcal{X}$ specifying the distribution at time t of a continuous time Markov chain $(X_t)_{t>0}$ with transition matrix K starting from x. Conceptually, the process (X_t) may be described as follows. We have a Poisson clock which rings according to a Poisson(1) distribution. Each time the clock rings, the process (X_t) jumps according to the transition matrix $K(x,y)$.

It is a familiar fact that if the matrix $K(x,y)$ is irreducible (which we shall henceforth assume without mention) then there is a stationary distribution $\pi(\cdot)$ so that $||H_t^x(\cdot) - \pi(\cdot)|| \to 0$. For convenience, we shall also consider the densities of the probability measures K_x^l and H_t^x with respect to π:

$$\begin{aligned} k_x^l(y) &= k^l(x,y) = \frac{K^l(x,y)}{\pi(y)} \\ h_t^x(y) &= h_t(x,y) = \frac{H_t^x(y)}{\pi(y)}. \end{aligned}$$

The adjoint K^* of K on $l^2(\pi)$ has kernel $K^*(x,y) = \pi(y)K(y,x)/\pi(x)$. Using the fact that π is the stationary distribution for K one readily checks that K^* is a Markov operator. It's semigroup, which represents the time reversal of H_t, is $H_t^* = e^{-t(I-K^*)}$ with kernel $H_t^*(x,y) = \pi(y)H_t(y,x)/\pi(x)$ and density $h_t^*(x,y) = h_t(y,x)$.

The *Dirichlet form* associated with $H_t = e^{-t(I-K)}$ is defined by:

$$\mathcal{E}(f,g) \overset{def}{=} \mathcal{R}(\langle (I-K)f, g\rangle). \tag{28.4}$$

It has the following properties:

Lemma 28.1 *The Dirichlet form $\mathcal{E}$ satisfies*

i. $\mathcal{E}(f,f) = \left\langle (I - \frac{1}{2}(K+K^*))f, f \right\rangle$

ii. $\mathcal{E}(f,f) = \frac{1}{2}\sum_{x,y} |f(x) - f(y)|^2 K(x,y)\pi(x)$

iii. $\frac{\partial}{\partial t}||H_t f||_2^2 = -2\mathcal{E}(H_t f, H_t f)$.

Proof:

i. Note that $\langle Kf, f\rangle = \langle f, K^* f\rangle = \overline{\langle K^* f, f\rangle}$.

ii. Observe that $\mathcal{E}(f,f) = ||f||_2^2 - \mathcal{R}(\langle Kf, f\rangle)$ and compute:

$$\begin{aligned} \frac{1}{2}\sum_{x,y} |f(x) - f(y)|^2 K(x,y)\pi(x) &= \frac{1}{2}\sum_{x,y} (|f(x)|^2 + |f(y)|^2 - 2\mathcal{R}(\overline{f(x)}f(y)))K(x,y)\pi(x) \\ &= ||f||_2^2 - \mathcal{R}(\langle Kf, f\rangle). \end{aligned} \tag{28.5}$$

iii. Calculus.

$$\begin{aligned} \frac{\partial}{\partial t}||H_t f||_2^2 &= \frac{\partial}{\partial t}\langle H_t f, H_t f\rangle \\ &= \left\langle \frac{\partial}{\partial t} H_t f, H_t f \right\rangle + \left\langle H_t f, \frac{\partial}{\partial t} H_t f \right\rangle \\ &= \langle -(I-K)H_t f, H_t f\rangle + \langle H_t f, -(I-K)H_t f\rangle \\ &= -2\mathcal{E}(H_t f, H_t f) \end{aligned} \tag{28.6}$$

■

Definition 28.2 *Let K be a Markov kernel with Dirichlet form $\mathcal{E}$. The* **spectral gap** $\lambda = \lambda(K)$ *is defined by:*

$$\lambda = \min\left\{\frac{\mathcal{E}(f,f)}{Var_\pi(f)}; Var_\pi(f) \neq 0\right\} \tag{28.7}$$

In general, $\lambda(K)$ is the smallest nonzero eigenvalue of $I - \frac{1}{2}(K + K^*)$. If (K, π) is reversible, then K is self adjoint so the Dirichlet form satisfies $\mathcal{E}(f,f) = \langle (I-K)f, f\rangle$ and λ is the smallest nonzero eigenvalue of $(I - K)$.

28.3 A Few Results

The most elementary result concerning the convergence of Markov chains is the Perron-Frobenius theorem for discrete time chains which asserts that an irreducible finite state Markov chain decays to equilibrium exponentially fast. However, the constants supplied by the theorem are generally very difficult to calculate and too conservative to be of any practical use. Fortunately, we can do much better with the spectral gap. To see why we should be interested in the spectral gap, we note the following exact result.

Theorem 28.3 *Suppose (K, π) is reversible and let $(\psi_i)_0^{n-1}$ be an orthonormal basis of $l^2(\pi)$ consisting of real eigenfunctions of $I - K$ with associated eigenvalues $(\lambda_i)_0^{n-1}$ listed in nondecreasing order. In particular, let $\psi \equiv 1$. Then*

$$||h_t^x - 1||_2^2 = \sum_{i=1}^{n-1} e^{-2t\lambda_i} |\psi_i(x)|^2. \tag{28.8}$$

Proof: Note that the functions ψ_i are eigenfunctions of H_t with eigenvalues $e^{-\lambda_i t}$. Also $H_t(x,y) = H_t\delta_y(x)$ where $\delta_y(z) = 1$ if $z = y$ and is 0 otherwise. Now

$$\delta_y = \sum_i \langle \delta_y, \psi_i \rangle\, \psi_i \tag{28.9}$$

and $\langle \delta_y, \psi_i\rangle = \overline{\psi_i(y)}\pi(y)$. Thus we obtain

$$\begin{aligned}
H_t(x,y) &= H_t\left(\sum_i \overline{\psi_i(y)}\pi(y)\psi_i\right)(x) \\
h_t^x(y) &= \sum_{i=0}^{n-1} e^{-\lambda_i t}\overline{\psi_i(y)}\psi_i(x) \\
||h_t^x - 1||_2^2 &= \left\langle \sum_{i=0}^{n-1} e^{-\lambda_i t}\overline{\psi_i}\psi_i(x) - 1, \sum_{i=0}^{n-1} e^{-\lambda_i t}\overline{\psi_i}\psi_i(x) - 1 \right\rangle \\
&= \sum_{i=1}^{n-1} e^{-2t\lambda_i}|\psi_i(x)|^2
\end{aligned} \tag{28.10}$$

where we have used the fact that the ψ_i's are orthonormal and that $\psi_0 \equiv 1$ to deduce the last line.
■

The expression above exhibits the fact that for reversible Markov chains the spectral gap controls the rate of exponential decay to the stationary distribution.

Lemma 28.4 *Let K be a Markov kernel with spectral gap $\lambda = \lambda(K)$. Then the semigroup $H_t = e^{-t(I-K)}$ satisfies*

$$\forall f \in l^2(\pi),\ ||H_t f - \pi(f)||_2^2 \leq e^{-2\lambda t} Var_\pi(f). \tag{28.11}$$

Proof: Define $u(t) = Var_\pi(H_t f) = ||H_t(f - \pi(f))||_2^2 = ||H_t f - \pi(f)||_2^2$. Using the lemma above, we obtain:

$$u'(t) = -2\mathcal{E}(H_t(f - \pi(f)), H_t(f - \pi(f))) \leq -2\lambda u(t).$$

It follows that $u(t) \leq e^{-2\lambda t} u(0)$. Since $u(0) = Var_\pi(f)$ the proof is complete. ■

From the above lemma we obtain a simple yet extremely useful quantitative bound on the rate of convergence of a finite state Markov chain:

Theorem 28.5 *Let K be a Markov kernel with spectral gap $\lambda = \lambda(K)$. Then the density $h_t^x(\cdot) = H_t^x(\cdot)/\pi(\cdot)$ satisfies*

$$||h_t^x - 1||_2 \leq \sqrt{1/\pi(x)}\, e^{-\lambda t}. \tag{28.12}$$

It follows that

$$|H_t(x,y) - \pi(y)| \leq \sqrt{\pi(y)/\pi(x)}\, e^{-\lambda t}. \tag{28.13}$$

Proof: Let H_t^* be the adjoint of H_t on $l^2(\pi)$. This is a Markov semigroup with spectral gap $\lambda(K^*) = \lambda(K)$. Set $\delta_x(y) = 1/\pi(x)$ if $y = x$ and $\delta_x(y) = 0$ otherwise. Then

$$h_t^x(y) = \frac{H_t^x(y)}{\pi(y)} = H_t^* \delta_x(y) \tag{28.14}$$

and applying the above lemma to K^*:

$$||H_t^* \delta_x - 1||_2^2 \leq e^{-2\lambda t} Var_\pi(\delta_x). \tag{28.15}$$

Combining the two equations above we obtain

$$||h_t^x - 1||_2 \leq \sqrt{(1 - \pi(x))/\pi(x)}\, e^{-\lambda t} \leq 1/\sqrt{\pi(x)}\, e^{-\lambda t}. \tag{28.16}$$

Since the same result clearly holds for h_t^* we readily obtain

$$\begin{aligned} |h_t(x,y) - 1| &= \left| \sum_z (h_{t/2}(x,z) - 1)(h_{t/2}(z,y) - 1)\pi(z) \right| \\ &\leq ||h_{t/2}^x||_2 ||h_{t/2}^{*y} - 1||_2 \\ &\leq \frac{1}{\sqrt{\pi(x)\pi(y)}} e^{-\lambda t}. \end{aligned} \tag{28.17}$$

Now multiply each side of the above inequality by $\pi(y)$ to finish the job. ■

Definition 28.6 *The p-mixing time for a Markov chain with kernel $K(x,y)$ is defined to be the constant $T_p = \min\{t > 0 : \max_x ||h_t^x - 1||_p \leq 1/e\}$.*

Definition 28.7 *Let* $\omega = \omega(K) = \min\{\mathcal{R}(\psi) : \psi \neq 0$ *is an eigenvalue of* $I - K\}$.

We now state a result bounding the p-mixing times in terms of the spectral gap.

Theorem 28.8 *Let* K *be an irreducible Markov kernel and let* $\pi_* = \min\{\pi(x) : x \in \mathcal{X}\}$. *Then, for* $1 \leq p \leq 2$,

$$\frac{1}{\omega} \leq T_p \leq \frac{1}{2\lambda}\left(2 + \log\frac{1}{\pi^*}\right) \tag{28.18}$$

and for $2 < p \leq \infty$,

$$\frac{1}{\omega} \leq T_p \leq \frac{1}{\lambda}\left(1 + \log\frac{1}{\pi^*}\right). \tag{28.19}$$

Proof: We prove only the upper bounds. By Holder's inequality we see that if $r < s$ then $||\cdot||_r \leq ||\cdot||_s$. Now use (28.12) for the case $p \leq 2$ and (28.17) for the case $2 \leq p \leq \infty$. ■

28.4 An Example

We now illustrate how these techniques may be used to bound the mixing time of a specific Markov chain. Let $\mathcal{X} = \{0,1\}^n$ and set $K(x,y) = 0$ unless $|x - y| = \sum_i |x_i - y_i| = 1$ in which case $K(x,y) = 1$. The functions

$$f_y : x \mapsto (-1)^{y \cdot x},\ y \in \{0,1\}^n \tag{28.20}$$

where $x \cdot y = \sum_i x_i y_i$ form an orthonormal basis of $l^2(\pi)$ and it is easy to see that $\pi(x) = 2^{-n}$ is the stationary distribution. Now observe that

$$\begin{aligned} Kf_y(x) &= \sum_z K(x,z) f_y(z) \\ &= \left(\frac{1}{n}\sum_i (-1)^{e_i \cdot y}\right) f_y(x) \\ &= \frac{n - 2|y|}{n} f_y(x) \end{aligned} \tag{28.21}$$

so f_y is an eigenfuction of $I - K$ with eigenvalue $2|y|/n$ where $|y|$ denotes the number of 1's in y. Thus, $\omega = \lambda = 2/n$ and $\pi_* = 2^{-n}$ so our theorem yields the bounds:

$$\frac{n}{2} \leq T_2 \leq n(2 + n). \tag{28.22}$$

Using the exact formula established above, a straightforward computation shows that T_2 is $\mathcal{O}(n \log n)$.

28.5 The log-Sobolev Constant

Superior bounds on the mixing time may be obtained from the log-Sobolev constant, α, which is defined in an analogous manner to the spectral gap.

Definition 28.9 *Let K be an irreducible Markov chain with stationary measure π. The log-Sobolev constant $\alpha = \alpha(K)$ is defined by*

$$\alpha = \min\left\{\frac{\mathcal{E}(f,f)}{\mathcal{L}(f)}; \mathcal{L}(f) \neq 0\right\} \tag{28.23}$$

where

$$\mathcal{L}(f) = \sum_{x\in\mathcal{X}} |f(x)|^2 \log\left(\frac{|f(x)|^2}{||f||_2^2}\right)\pi(x). \tag{28.24}$$

The log-Sobolev constant yields the following bounds on the mixing times.

Theorem 28.10 *Let (K,π) be a finite reversible Markov chain. For $1 \leq p \leq 2$,*

$$\frac{1}{\omega} \leq T_p \leq \frac{1}{4\alpha}\left(4 + \log_+ \log \frac{1}{\pi_*}\right) \tag{28.25}$$

and for $2 < p \leq \infty$,

$$\frac{1}{\omega} \leq T_p \leq \frac{1}{2\alpha}\left(3 + \log_+ \log \frac{1}{\pi_*}\right). \tag{28.26}$$