STA4502: Topics in Stochastic Processes	**Winter 2018**

Lecture 3: Eigenvalue Connection — March 14, 2018

Lecturer: Jeffrey Rosenthal *Scriber: Yuenan Joseph Cai*

Challenge 26 Solution

Starting from any state, there is at least $\frac{1}{3^{N-1}}$ probability to get to any other state in $N-1$ steps. This suggests using Minorization technique with $\epsilon\rho(\cdot) = \frac{1}{3^{N-1}}$ where $\rho(\cdot)$ is a uniform distribution on $\mathcal{X}$. Therefore, $\epsilon = \frac{N}{3^{N-1}}$. For $\delta > 0$, find a bound k^* s.t. $||\mu_k - \pi||_{TV} \leq (1-\epsilon)^{\lfloor k/(N-1) \rfloor} \leq \delta$.

$$\Rightarrow k^* \geq (N-1)\left(\frac{\ln(\delta)}{\ln(1-\frac{N}{3^{N-1}})} + 1\right).$$

When $N = 5$ and $\delta = 0.01$, $k^* = 294$. Can this be improved?

A Note on Coupling under Minorization Condition

The coupling method does not modify the marginal transition probabilities of $\{X_n\}$ and $\{Y_n\}$. Recall (X_n, Y_n) is jointly updated in the following manner:

- If $X_{n-1} = x \neq y = Y_{n-1}$, $\begin{cases} \text{w.p. } \epsilon \text{, choose } X_n = Y_n \sim \rho(\cdot) \\ \text{w.p. } 1-\epsilon \text{, choose } X_n \sim \frac{1}{1-\epsilon}(P(x,\cdot) - \epsilon\rho(.)) \text{ and } Y_n \sim \frac{1}{1-\epsilon}(P(y,\cdot) - \epsilon\rho(\cdot)) \end{cases}$;

- Otherwise, let $X_n = Y_n \sim P(x, \cdot)$.

In the nontrivial case $X_{n-1} = x \neq y = Y_{n-1}$,

$$P(X_n \in \cdot | X_{n-1} = x) = \epsilon\rho(\cdot) + (1-\epsilon)\Big(\frac{1}{1-\epsilon}(P(x,\cdot) - \epsilon\rho(\cdot))\Big) = P(x, \cdot).$$

Similarly, $P(Y_n \in \cdot | Y_{n-1} = y) = P(y, \cdot)$.

Example 29. (Coupling - Card Shuffling) Suppose a deck of cards is to be shuffled by taking the top card and place it randomly back into the deck. How long will it take to well scrambled the deck (i.e. "almost" equally likely to obtain any card arrangement)?

To make our analysis easier, consider an alternative shuffling method that takes a card at random from the deck and place it on top. It can be shown that the "random-to-top" shuffling method is equivalent to "top-to-random" method [1].

How to apply coupling? Consider using two desks of cards, a well mixed deck on the left and the one to be shuffled on the right. We can choose a card randomly from a deck, find the same card in the other one, and place both cards on top of the respective deck. Despite having the same card drawn from both decks, it is still a random draw from each in terms of marginals. Once all the cards have been selected once, both decks must be in the same order. Now the question can be viewed as a coupon collector's problem in finding the probability that more than k shuffles are needed to touch all n cards. Formally, the left deck starts in

[1] They are random walks on a group. One way to relate the two methods is that they are "time reversal" version of each other ($\widetilde{p}(x,y) = p(y,x)\frac{\pi(y)}{\pi(x)}$, where $\widetilde{p}$ is the new description under "random-to-top" method). In terms of group operations, each "top-to-random" draw t can be matched by a "random-to-top" draw and $\widetilde{p}(x, xt) = p(x, xt^{-1})$. We have $\widetilde{\mu}_k(x) = \mu_k(x^{-1})$ where x and x^{-1} are a permutation operation and its inverse.

stationary distribution, and after k steps,

$$
\begin{aligned}
&||\mu_k - \pi||_{TV} \\
\leq &P(T > k, \text{ where } T \text{ is the time taken to select all cards at least once}) \\
= &P(\text{Have not touch all cards by time } k) \\
= &P(\cup_{i=1}^{n} \text{Have not touch card } i \text{ by time } k) \\
\leq &\sum_{i=1}^{n} P(\text{Have not touch card } i \text{ by time } k) \\
= &n(1 - \frac{1}{n})^k \\
\leq &ne^{-\frac{k}{n}} \\
= &e^{-(\frac{k}{n} - \ln(n))}
\end{aligned}
$$

If $k = cn\ln(n)$, then $||\mu_k - \pi||_{TV} \leq n^{1-c}$. The bound is small when $c > 1$ and n is large. For a deck with 52 suit cards, $||\mu_k - \pi||_{TV} \leq 0.01$ when $c \approx 2.2$, or $k_* \approx 452$.

6 Eigenvectors and Eigenvalues

In this section we will study the connections between Markov chains and eigenvalues. Assume a Markov Chain on finite state space $\mathcal{X}$ of size d, the transition probability P is diagonalizable [2] with elements in $\mathbb{C}$. Recall that the eigenpairs (λ_m, v_m) of P satisfies $v_m P = \lambda_m v_m$ for $m = 0, 1, ..., d-1$. Since $\pi P = \pi$ if π is a stationary distribution, we can always assign $(1, \pi)$ to (λ_0, v_0).

For an initial distribution written in terms basis of (left) eigenvectors $\mu_0 = a_0 v_0 + ... + a_{d-1} v_{d-1}$, the k-step distribution is $\mu_k = \mu_0 P^k = a_0 \lambda_0^k v_0 + ... + a_{d-1} \lambda_{d-1}^k v_{d-1}$. Let $\lambda_* = \max_{i \geq 1} |\lambda_i|$, then

$$
\begin{aligned}
&|\mu_k(x) - \pi(x)| \\
= &|a_1 \lambda_1^k v_1(x) + ... a_{d-1} \lambda_{d-1}^k v_{d-1}(x)| \\
\leq &|\lambda_1|^k |a_1 v_1(x)| + ... + |\lambda_{d-1}|^k |a_{d-1} v_{d-1}(x)| \\
\leq &|\lambda_*|^k \big(|a_1 v_1(x)| + ... + |a_{d-1} v_{d-1}(x)|\big) \\
= &C_{\mu_0, x} |\lambda_*|^k
\end{aligned}
$$

Remark 30. If $|\lambda_1|, ..., |\lambda_{d-1}| < 1$, then as $k \to \infty$, $\mu_k \to a_0 \pi$ where $a_0 = 1$. For the other direction, if the Markov Chain with stationary distribution π is irreducible and aperiodic, then $|\lambda_i| < 1$ $\forall i \geq 1$. A few comments:

1. We may have other $|\lambda_i| = 1$ for some $i \in \{1, ..., d-1\}$ when the chain is periodic. Consider an example on $\{1, 2\}$ of period 2:

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

   The eigenpairs are $\{\big(1, (1,1)\big), \big(-1, (-1, 1)\big)\}$. In this case the k-step distribution does not converge to the stationary distribution $(\frac{1}{2}, \frac{1}{2})$ for general μ_0. Instead we require periodic version of the Markov Chain Convergence Theorem.

[2] If P is not diagonalizable (when some eigenvalues have multiplicity ≥ 2), then we obtain the Jordan canonical form.

2. If the chain is not irreducible, we may still be able to find irreducible sub-chains where the statement still holds.

Next we will review a few results from Linear Algebra

Definition 31. ($L^2(\pi)$ norm) The $L^2(\pi)$ norm between v and w is $< v, w >_{L^2(\pi)} := \sum_{x \in \mathcal{X}} v(x)w(x)\pi(x)$.

If the eigenvectors $\{v_i\}$ are orthonormal in $L^2(\pi)$, then $< v_i, v_j >_{L^2(\pi)} = \delta_{ij}$. We seek for conditions in which P is diagonalizable w.r.t. some orthonormal vectors.

Definition 32.
1. The adjoint operator P^* of P satisfies $< v, wP > = < vP^*, w >$.
2. P is normal if $PP^* = P^*P$.
3. P is self-adjoint if $P = P^*$.

Fact 33.
1. If P is self-adjoint, then it is also normal, with real eigenvalues.
2. If P is either normal or self-adjoint, then there exists an orthonormal basis $\{v_i\}$.
3. If π is a uniform distribution, then $P^* = P^\dagger$ where $P^\dagger$ is the conjugate transpose of P. Furthermore P is self-adjoint iff P is symmetric.

Fact 34. P is self-adjoint iff the chain is reversible w.r.t P.

Fact 35. Using Cauchy–Schwarz inequality, one can show that

$$\begin{aligned}
&||\mu_k - \pi||_{TV} \\
=&\frac{1}{2}\sum_x |\mu(x) - \pi(x)| \\
=&\frac{1}{2}\sum_x |\mu(x) - \pi(x)|\sqrt{\pi(x)})\frac{1}{\sqrt{\pi(x)}} \\
\leq&\frac{1}{2}\sqrt{\sum_x (\mu_k(x) - \pi(x))^2\pi(x)\sum_x \frac{1}{\pi(x)}} \\
\leq&\frac{1}{2}\sqrt{\sum_x (\mu_k(x) - \pi(x))^2\pi(x)\frac{n}{\min_x \pi(x)}} \\
=&\frac{1}{2}\sqrt{\frac{n}{\min_x \pi(x)}}||\mu_k - \pi||_{L^2(\pi)}
\end{aligned}$$

Fact 36. The result below will allow us to quantify the convergence rate in terms of eigenvalues.

$$\begin{aligned}
&||\mu_k - \pi||^2_{L^2(\pi)} \\
&= < \mu_k - \pi, \mu_k - \pi >_{L^2(\pi)} \\
&= < \sum_{i=1}^{n-1} a_i \lambda_i^k v_i, \sum_{i=1}^{n-1} a_i \lambda_i^k v_i >_{L^2(\pi)} \\
&= \sum_{i,j=1}^{n-1} a_i a_j \lambda_i^k \lambda_j^k < v_i, v_j >_{L^2(\pi)} \\
&= \sum_{i=1}^{n-1} a_i^2 \lambda_i^{2k} \\
&\leq \lambda_*^{2k} \sum_{i=1}^{n-1} a_i^2 \\
&\leq \lambda_*^{k} \sum_{i=1}^{n-1} a_i^2
\end{aligned}$$

Remark 37. In continuous case, we are interested in the operator norm $||P_0||_{L^2(\pi)\to L^2(\pi)}$ [3]. The notation $P_0 := P|_{\pi^\perp}$ stands for P restricted to signed measure of total mass 0 (analog of $\{v_1, ..., v_{d-1}\}$, which are orthogonal elements of $v_0 = \pi$). In general, $||P||_{L^2(\pi)\to L^2(\pi)} = 1$. If $||P_0|| < 1$, the chain is geometric ergodic.

6.1 A Few Motivating Examples

Example 38. (Frog Walk) Suppose there are n lily pads arranged in a circle. A frog starts at pad 0 and at each step, with equal probability, either moves clockwise, counter-clockwise or remains at where it was. The transition matrix associated with the Markov Chain on $\mathcal{X} = \{0, 1, ..., n-1\} = \mathbb{Z}/(n)$ is:

$$P = \begin{bmatrix}
1/3 & 1/3 & 0 & ... & & 0 & 1/3 \\
1/3 & 1/3 & 1/3 & 0 & ... & & 0 \\
0 & 1/3 & 1/3 & 1/3 & 0 & ... & 0 \\
... & & & & & & ... \\
0 & ... & 0 & 1/3 & 1/3 & 1/3 & 0 \\
0 & ... & & 0 & 1/3 & 1/3 & 1/3 \\
1/3 & 0 & ... & & 0 & 1/3 & 1/3
\end{bmatrix}$$

This chain is reversible w.r.t the uniform distribution on $\mathcal{X}$ and therefore, $\pi(x) = \frac{1}{N}$ is a stationary distribution. For fixed ϵ, find k^* such that $\forall k \geq k^*$, $||\mu_{k^*} - \pi||_{TV} \leq \epsilon$; in particular, solve for $n = 1000$ and $\epsilon = 0.01$.

This is a random walk on an abelian group which will be covered in the following lecture. One may simplify the notations in terms of step distributions for random walk on abelian groups.

$$Q(-1) = Q(0) = Q(1) = \frac{1}{3}.$$

[3] Norm of operator A is defined as $||A|| = \sup_{u\neq 0} \frac{||Au||}{||u||}$

Example 39. (Bit Flipping) Suppose the state space $\mathcal{X} = (\mathbb{Z}/(2))^d$ represents the set of bits of length d (for example $d = 8$, $x = (0, 0, 1, 1, 1, 0, 1, 0) \in \mathcal{X}$). Define a Markov Chain with the following transition probabilities:

$$\begin{cases} \text{w.p. } \frac{1}{d+1}, \text{ do nothing to the list} \\ \text{w.p. } \frac{d}{d+1}, \text{ change a random bit} \end{cases}.$$

Equivalently,

$$\begin{cases} \text{w.p. } \frac{1}{d+1}, \text{ set } X_n = X_{n-1} \\ \text{w.p. } \frac{d}{d+1}, \text{ set } \begin{cases} X_{n,i} = X_{n-1,i}, \ i \neq j \\ X_{n,i} = 1 - X_{n-1,i}, \ i = j \end{cases} \quad \text{where } j \text{ is chosen uniformly from } \{1, ..., d\} \end{cases}.$$

This chain is also reversible w.r.t the uniform distribution on $\mathcal{X}$ and therefore, $\pi(x) = \frac{1}{2^d}$ is a stationary distribution. How fast does the k-th step distribution converge to π?