Stats 202C: Homework #1

Professor S.C. Zhu
Assignment: 1-4

Eric Chuu
UID: 604406828

April 28, 2017

Problem 1

(1) The period of a state i is given by

$$d_i = \gcd\{n \geq 1 : K_{ii}^n > 0\} \tag{1}$$

We also know that a state i is aperiodic if $d_i = 1$.

For K_0, we draw the transition graph diagram and see that all states can be reached in finite time with nonzero probability, so it is irreducible. Since the Markov chain is irreducible, then all states have the same period. Thus, it suffices to find the period of any state. If we begin in state 1, we can return to state 1 in two steps by going to state 2 and then back to state 1. We can also return to state 1 in three steps by going to state 2 then state 3, then returning back to state 1 directly. Then, $d_1 = 1$, and state 1 is aperiodic, and we conclude that K_0 is aperiodic.

For K_1, we can see from the transition graph diagram, $d_i = 1$ for $i = 1, \ldots, 5$, so K_1 is aperiodic. We can see that K_1 is reducible by considering states 5 and 4. If we are in states 4 or 5, the probability of going into states 1, 2, or 3 is 0.

For K_2, we draw the transition graph diagram and see that all states can be reached in finite time with nonzero probability, so it is irreducible. For each state, we can see that $d_i = 2$ for $i = 1, \ldots, 5$, so K_2 is periodic.

(2) The eigenvalues and left eigen-vectors for each matrix are given below.

K_0 left-eigenvectors (stored as column vectors):

$$
\begin{bmatrix}
-0.167 & -0.07 - 0.118i & -0.07 + 0.118i & 0.402 & 0.503 \\
-0.517 & 0.148 + 0.379i & 0.148 - 0.379i & 0.491 & -0.159 \\
-0.632 & 0.3 - 0.241i & 0.3 + 0.241i & 0.139 & -0.638 \\
-0.492 & -0.735 & -0.735 & -0.667 & -0.221 \\
-0.251 & 0.357 - 0.019i & 0.357 + 0.019i & -0.365 & 0.516
\end{bmatrix}, \quad
\Lambda_1 =
\begin{bmatrix}
1 \\
-0.544 + 0.182i \\
-0.544 - 0.182i \\
0.279 \\
-0.19
\end{bmatrix}
$$

K_1 left-eigenvectors (stored as column vectors):

$$\begin{bmatrix} 0 & -0.284 & 0 & 0.78 & -0.697 \\ 0 & -0.42 & 0 & -0.528 & 0 \\ 0 & -0.386 & 0 & -0.201 & 0.697 \\ -0.707 & 0.541 & -0.707 & -0.213 & 0.116 \\ -0.707 & 0.549 & 0.707 & 0.162 & -0.116 \end{bmatrix}, \quad \Lambda_2 = \begin{bmatrix} 1 \\ 0.939 \\ -0.4 \\ -0.139 \\ 0.2 \end{bmatrix}$$

K_2 left-eigenvectors (stored as column vectors):

$$\begin{bmatrix} -0.149 & 0.149 & 0.772 & -0.242 & -0.242 \\ -0.45 & 0.45 & -0.617 & 0.645 & 0.645 \\ -0.46 & 0.46 & -0.154 & -0.403 & -0.403 \\ 0.563 & 0.563 & 0 & 0.426i & 0.426i \\ 0.497 & 0.497 & 0 & 0.426i & 0.426i \end{bmatrix}, \quad \Lambda_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0.529i \\ 0.529i \end{bmatrix}$$

(3) For K_0, there is one invariant probability:

$$\begin{bmatrix} 0.08094 & 0.2512 & 0.3071 & 0.2389 & 0.1219 \end{bmatrix}$$

with eigenvalue 1.
For K_1, there is one invariant probability:

$$\begin{bmatrix} 0 & 0 & 0 & 0.5000 & 0.5000 \end{bmatrix}$$

with eigenvalue 1.
For K_2, there is one invariant probability:

$$\begin{bmatrix} 0.0703 & 0.2125 & 0.2172 & -0.2656 - 0.2344 \end{bmatrix}$$

with eigenvalue 1.
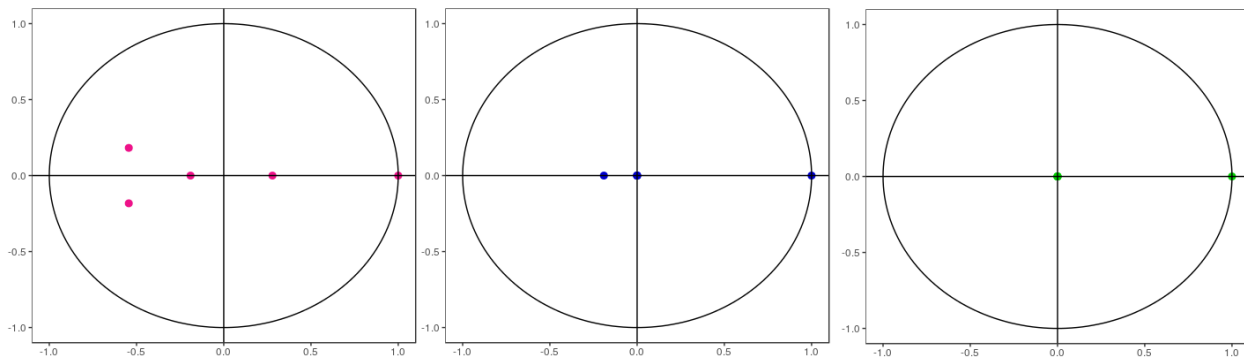**(4)** The eigen-values for K_0, K_0^{50}, K_0^{200} are plotted below.

Figure 1: Left: Eigenvalues for K_0. Center: Eigenvalues for K_0^{50}. Right: Eigenvalues for K_0^{200}

(5) We compute the matrix K_0^{200}, and we see that it becomes the "ideal" matrix, where each row is π.

$$K_0^{200} = \begin{bmatrix} 0.809 & 0.2512 & 0.3071 & 0.2389 & 0.1219 \\ 0.809 & 0.2512 & 0.3071 & 0.2389 & 0.1219 \\ 0.809 & 0.2512 & 0.3071 & 0.2389 & 0.1219 \\ 0.809 & 0.2512 & 0.3071 & 0.2389 & 0.1219 \\ 0.809 & 0.2512 & 0.3071 & 0.2389 & 0.1219 \end{bmatrix}$$

Problem 2

(1) The plot of $d_{\mathrm{TV}}(n), d_{\mathrm{KL}}(n)$ for the first 200 steps is shown below.

(2) We calculate the contraction coefficient K_0 using

$$C(K_0) = \max_{x,y} \|K_0(x,\cdot) - K_0(y,\cdot)\|_{\mathrm{TV}} \tag{2}$$

For two initial probbailities ν_1, ν_2,

$$\|\nu_1 \cdot K_0 - \nu_2 \cdot K_0\|_{\mathrm{TV}} \leq C(K_0)\|\nu_1 - \nu_2\|_{\mathrm{TV}} \tag{3}$$

As $\|\nu_1 - \nu_2\|_{\mathrm{TV}} \leq 1$, if $C(K) < 1$, then the convergence rate can be upper bounded by

$$A(n) = \|\nu_1 \cdot K_0^n - \nu_2 \cdot K_0^n\|_{\mathrm{TV}} \leq C^n(K_0)\|\nu_1 - \nu_2\|_{\mathrm{TV}} \leq C^n(K_0), \quad \forall \nu_1, \nu_2 \tag{4}$$

The plot for the bound $C^n(K_0)$ over $n = 1, \ldots, 100$ is shown below.

(3) The Diaconis-Hanlon bound is given by

$$B(n) = \|\pi - \nu K_0^n\|_{\mathrm{TV}} \leq \sqrt{\frac{1 - \pi(x_0)}{4\pi(x_0)}} \lambda_{\mathrm{slem}}^n \tag{5}$$

where $x_0 = 1$ is the initial state and $\pi(x_0)$ is the target probability at $x_0 = 1$ and $\lambda_{\mathrm{slem}}^n$ is the second largest eigen-value modulus. The plot of the real convergence d_{TV} in comparison with $A(n), B(n)$ is given below.

We see that both converge to 0 very quickly, with KL-Divergence converging to 0 slightly faster. Bound $A(n)$, calculated using the contraction coefficient is shown in the turquoise curve below, while Bound $B(n)$, the Diaconis-Hanlon bound, is shown in the purple curve below. We can see that $B(n)$ provides a tighter bound on the convergence than does $A(n)$.
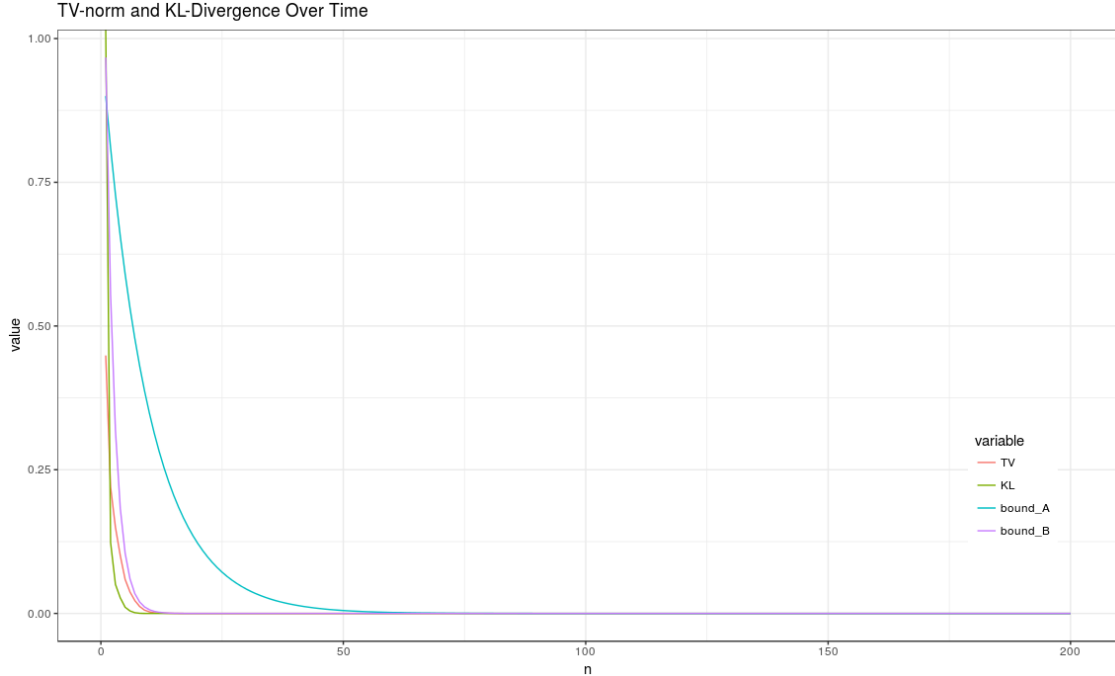

Figure 2: TV-norm and KL-Divergence for $n = 200$ steps. Bound A and Bound B refer to $A(n), B(n)$, respectively, both as calculated above.

Problem 3

In a finite state Ω, suppose at step t, a Markov chain MC has state X following probability ν. By applying the Markov kernel P once, its state in $t+1$ is Y which follows probability $\mu = \nu \cdot P$. $P(x,y)$ is the transition (conditional) probability from state x to y. Suppose we denote by $Q(x,y)$ the reverse transition probability of P at time t,

$$Q(y,x) = \frac{P(x,y)\nu(x)}{\mu(y)} \tag{6}$$

Show that the Kullback-Leibler divergence decreases monotonically,

$$KL\left(\pi||\nu\right) - KL\left(\pi||\mu\right) = \mathrm{E}\left[KL\left(P\left(y,x\right)||Q\left(y,x\right)\right)\right] \geq 0 \tag{7}$$

Solution
By definition of KL-divergence, we have

$$KL\left(\pi||\nu\right) = \sum_{y:\pi(y)>0} \pi(y)\cdot\log\frac{\pi(y)}{\mu(y)} = \sum_{y:\pi(y)>0} \pi(y)\log\frac{\pi}{\sum_x P(x,y)\nu(x)}$$

Note that if $\mu(y) > 0$ when $\pi(y) > 0$, then by definition the KL divergence is infinite. Using this fact, we know that if $\sum_x P(x,y)\nu(x) = 0$ when $\pi(x) > 0$, this implies that $KL(\pi||\nu)$ is infinite and the inequality in (7) holds. If the sum is nonzero, then we have

$$
\begin{aligned}
\sum_{y:\pi(y)>0} \pi(y)\log\frac{\pi}{\sum_x P(x,y)\nu(x)} &\leq -\sum_{y:\pi(y)>0} \pi(y)\log\left[\sum_{x:\pi(x)>0} \frac{\nu(x)}{\pi(y)}\right] \\
&\leq -\sum_{y:\pi(y)>0} \pi(y)\log\left[\sum_{x:\pi(x)>0} Q(y,x)\cdot\frac{\nu(x)}{\pi(x)}\right] \\
&\leq -\sum_{y:\pi(y)>0} \pi(y)\sum_{x:\pi(x)>0} Q(y,x)\cdot\log\frac{\nu(x)}{\pi(x)} \quad \text{(Jensen's Inequality)} \\
&= \sum_x \pi(x)\log\frac{\pi(x)}{\nu(x)} \\
&= KL\left(\pi||\nu\right)
\end{aligned}
$$

where the second line follows from applying the equality in (6). Thus we conclude that

$$KL\left(\pi||\mu\right) \leq KL\left(\pi||\nu\right)$$
$$KL\left(\pi||\nu\right) - KL\left(\pi||\mu\right) \geq 0$$

and the inequality in (7) holds. $\qquad\square$

Problem 4

A Markov chain returning time $\tau_{\text{ret}(i)}$ is the minimum steps that a Markov chain needs to return to state i after leaving state i. Suppose we consider a random walk in the countable set of non-negative numbers $\Omega = \{1, 2, \ldots, \}$. At a step, the Markov chain state $x_t = n$, it has probability α to go up ($x_{t+1} = n + 1$) and probability $1 - \alpha$ to return to $x_{t+1} = 0$. Calculate the probability for returning to state 0 in finite step

$$\mathbf{Pr}\left(\tau_{\text{ret}}(0) < \infty\right) = \sum_{\tau(0)=1}^{\infty} \mathbf{Pr}\left(\tau(0)\right) \tag{8}$$

Calculate the expected return time

$$\mathrm{E}\left[\tau_{\text{ret}}(0)\right] \tag{9}$$

Solution
We consider the transition matrix,

$$K = \begin{bmatrix} 1-\alpha & \alpha & 0 & 0 & \cdots \\ 1-\alpha & 0 & \alpha & 0 & \cdots \\ 1-\alpha & 0 & 0 & \alpha & \cdots \\ 1-\alpha & 0 & 0 & 0 & \cdots \\ 1-\alpha & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Each step follows a bernoulli distribution with probability of failure α and probability of success $1 - \alpha$. Then the probability of returning to state 0 in finite step is equivalent to calculating the probability of success in finite step. Using the summation in (8) above, we can calculate

$$\begin{aligned} \mathbf{Pr}\left(\tau_{\text{ret}}(0) < \infty\right) &= (1-\alpha) + \alpha(1-\alpha) + \alpha^2(1-\alpha) + \cdots + \alpha^n(1-\alpha) + \cdots \\ &= (1-\alpha)\left[1 + \alpha + \alpha^2 + \alpha^3 + \cdots\right] \\ &= (1-\alpha) \cdot \frac{1}{1-\alpha} \end{aligned}$$

if $\alpha < 1$ (note we disregard absolute value since α is a probability). Thus, the probability of returning to state 0 in finite time is 1 if $\alpha < 1$, but 0 for $\alpha = 1$. The expected return time is then just the expectation of a geometric random variable, i.e., the expected number of times it takes for the first success, so

$$\mathrm{E}\left[\tau_{\text{ret}}(0)\right] \begin{cases} \frac{1}{1-\alpha}, & \text{if } 0 \leq \alpha < 1 \\ \infty, & \text{if } \alpha = 1 \end{cases}$$

We see that the expectation is finite only for $\alpha < 1$, and infinite for $\alpha = 1$. $\qquad\square$