ECS 256 - Professor N. Matloff
Fall 2022

Group: Yun Qin, Jane Park
20 October 2022

Problem Set 1

1. *Consider a finite-state space, aperiodic Markov chain with transition matrix P. Our textbook shows a way to find the stationary probability vector π that is alternative to the usual method that solves a system of linear equations. Say the matrix is of size $r \times r$. The thrust of the argument is as follows:*

- *The power P^k is the k-step transition matrix.*
- *So, row 1 of the matrix is $\mathbb{P}[X_k = j|X_0 = 1]$, $j = 1, \cdots, r$. Row 2 is $\mathbb{P}[X_k = j|X_0 = 2]$, $j = 1, \cdots, r$, and so on.*
- *We know that for an aperiodic chain, $\lim_{k\to\infty} \mathbb{P}[X_k = j|X_0 = i] \to \pi_j$.*
- *Thus for large k, each row of P^k will be approximately the vector π'.*
- *So, to find π, we could raise P to a large power, then use row 1 as π.*
- *But even better, we could average all the rows, which presumably will give us a better estimate.*

Investigate the accuracy of this approach, as follows:

- *Do your experiments for one small P and one large one, of your choice.*
- *For values of $k = 1, 2, 3, \cdots, m$ (your choice of m): Find the estimates of π, both using row 1 only and averaging all the rows. Take as your accuracy criterion the ℓ^1 distance (sum of absolute differences) of the estimated π to the actual value.*
- *Show your results graphically, with commentary in a* `.tex` *file.*

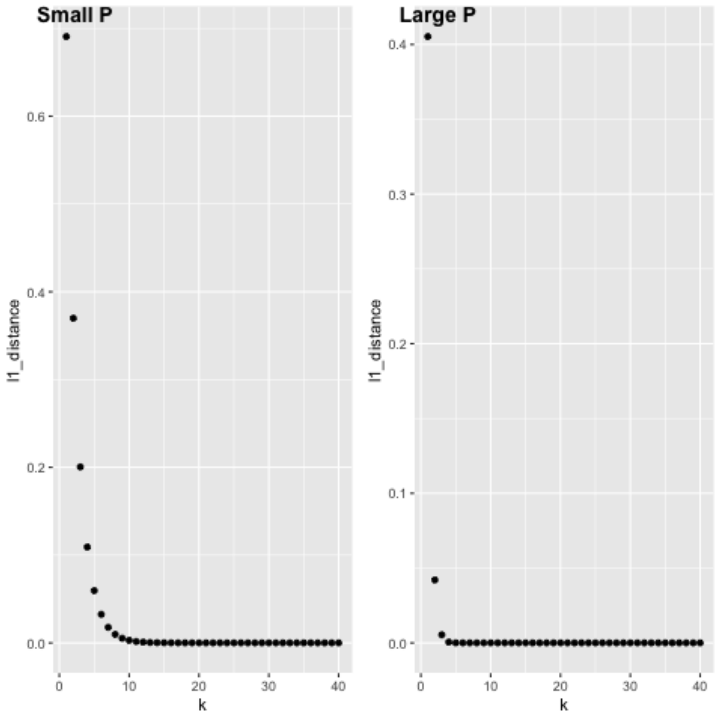

Figure 1: Convergent trend of the ℓ^1-norm difference over powers $k = 1, \cdots, m$ of $(P^k)[1,]$ and our π_k (defined below). We set `m=40`. (Left to right) `ggplots` of the desired compared quantities for 3×3 transition matrix example; and of our 20×20 transition matrix example, respectively.

Notice that the graphs, for both small and large dimensional P, we have a quick deflation of the difference between the first row of P^k, and π_k where $\pi_k[j] := \frac{1}{r}\sum_{j=1}^{r} P^k[,j]$, $\lim_{k\to\infty} \|P^k[1,] - \pi_k\|_{\ell^1} = \lim_{k\to\infty} \sum_{j=1}^{r} \left|P^k[1,j] - \pi_k[j]\right|$. This implies that we have convergence of $P^k[1,]$ to π_k, as $k \to \infty$. I.e.:

$$\lim_{k\to\infty} P^k[1,] = \lim_{k\to\infty} \pi_k =: \pi.$$

Conceptually, we know that the transition matrix always has 1 as an eigenvalue, and that te corresponding left eigenvector is the stationary distribution in vector form, given by π. The transition matrix has its full set of eigenvectors comprise of its eigenbasis, and all eigenvalues aside from 1 are values in $[0,1) \subset \mathbb{R}$. This implies that as we take higher powers of P, all eigenvalues not 1 vanish. This means that we have a Rank 1 matrix, entailing that all rows are multiples of each other. Thus, taking the average of all column entry for higher k powers will only make entries of π_k converge to the desired stationary probability distribution, put in vector form.

One thing we noticed is that for the larger P, which rows of twenty entries are taken from a Unif$[1,20]$ distribution, and which sum we normalized (so that the sum of any row is 1 to fulfill that requirement for the definition of transition matrices), the more spread out the values are in that matrix, and the (possibly significantly) smaller each row entry will be in subsequent powers of that matrix. This is why, in our case, we have many data points far above the $y = 0$ line in our 3×3 case whereas in the 20×20 case, we only had three points for $k = 1, 2, 3$ where the distance between $P^k[1,]$ and π_k is noticeably greater than 0.

2. *Consider the states of a Markov chain over time, $X_0, X_1, X_2, \cdots$, with the state space being some subset of the set of all integers, which we will take to be $1, \cdots, m$. Suppose the chain has a stationary distribution π, and that X_0 has this distribution.*

There is a term from time series analysis, autocorrelation: $\rho(k)$, defined to be the correlation between X_i and X_{i+k}. Note that this quantity depends only on k, not i.

(a) *Explain why there is no dependence on i (`.tex` file).* Let $P \in [0,1]^{m\times m}$ denote the transition matrix from any state i to state j for $i, j \in \{1, \cdots, m\}$.

If π is the stationary distribution, then as a vector $[\pi] \in [0,1] \in \mathbb{R}^m$, we have that $[\pi]' \cdot P = [\pi]' = 1 \cdot [\pi]'$ (meaning that π is a left eigenvector to P). Since $X_0 \sim \pi$, so that $\mathbb{P}[X_0 = j] := \pi(j)$. We have that, since for all n, X_n has the same distribution as is suggested by the vector $[\pi]' \cdot P^n = 1^n \cdot [\pi]' = [\pi]$, so that $X_n \sim \pi$ for all $n \in \mathbb{Z}_{\geq 0}$. This implies that $\mu_{X_n} = \mu_{X_0} =: \mu$, and $\sigma_{X_n} = \sigma_{X_0} =: \sigma$ for all $n \in \mathbb{Z}_{\geq 0}$.

Since these distributions X_n form a Markov chain, we know that the probability of X_{i+k} achieving state t given that X_i is at state i, for $i, i+k \in \{1, \cdots, m\}$ is

$$\mathbb{P}[X_{i+k} = t | X_i = s] = \mathbb{P}[X_k = t | X_0 = s] = (P^k)_{st}. \tag{1}$$

Then the autocorrelation of X_i and X_{i+k} is:

$$\begin{aligned}
\rho(X_i, X_{i+k}) &:= \frac{\mathbb{E}[X_i X_{i+k}] - \mu_{X_i}\mu_{X_{i+k}}}{\sigma_{X_i}\sigma_{X_{i+k}}} \\
&=: \frac{\mathbb{E}[X_i X_{i+k}] - \mu^2}{\sigma^2}, && \text{as all } X_n \text{ are identically distributed;} \\
&= \frac{\sum_{s=1}^{m} \mathbb{P}[X_i = s]\mathbb{E}[X_i X_{i+k} | X_i = s] - \mu^2}{\sigma^2}, && \text{by the } \textbf{Law of Total Expectation}; \\
&= \frac{\sum_{s=1}^{m} \pi[s]\mathbb{E}[sX_{i+k} | X_i = s] - \mu^2}{\sigma^2}, && \text{as } \mathbb{P}[X_i = s] = \mathbb{P}[X_0 = s] = \pi[s]; \\
&= \frac{\sum_{s=1}^{m} s\ \pi[s]\mathbb{E}[X_{i+k} | X_i = s] - \mu^2}{\sigma^2}, && \text{by linearity of expectation;} \\
&= \frac{\sum_{s=1}^{m} s\pi[s] \sum_{t=1}^{m} t\mathbb{P}[X_{i+k} = t | X_i = s] - \mu^2}{\sigma^2} \\
&= \frac{\sum_{t=1}^{m}\sum_{s=1}^{m} st\ \pi[s]\mathbb{P}[X_{i+k} = t | X_i = s] - \mu^2}{\sigma^2}, && \text{as finite sums are reorderable;} \\
&= \frac{\sum_{t=1}^{m}\sum_{s=1}^{m} st\ \pi[s](P^k)_{st} - \mu^2}{\sigma^2}, && \text{by (1) above.}
\end{aligned}$$

Focusing on the first summand in the numerator, we observe that this expression linearly standardizes $\mathbb{E}[X_i X_{i+k}] = \mathbb{E}[X_0 X_k]$, for all $k \in \mathbb{Z}^+$.

The expression $\mathbb{E}[X_0 X_k] = \sum_{t=1}^{m}\sum_{s=1}^{m} s\pi[s] \cdot t\ (P^k)_{st}$, which is akin to the *Hadamaard product* of $[s]_{s=1}^m$ and $[\pi]$, denoted $[s]_{s=1}^m \circ [\pi] := [s \cdot \pi[s]]_{s=1}^m$, dotted with $[t(P^k)_{st}]_{s=1}^m$, which is t-times the t^{th} column of P^k, summed over $t = 1, \cdots, m$, which is a multivariate expectation of the product random variable $X_0 X_k$, where X_k depends on X_0.

(b) *Write an* `R` *function with call form* `MCautocor(k,P)` *(Note that an argument* `m` *is not needed. Explain why.)*

The argument m is unnecessary because the conception of m is built into the dimension of the matrices side (equal by design of the transition matrix). In other words, the transition matrix already tells us the number of states for our random variable—which m denotes.

3. *Consider the ALOHA model with 3 stations. Derive the long- run average time between collisions. Expression your answer in terms of p, q and π, and evaluate for the case $p = 0.4$, $q = 0.3$. Show your reasoning in a* `.tex` *file.*

The long-run average time of collisions would be, for any given trial in this scenario, the quantity time units divided by the number of collisions that occur in that trial.

The long running average of the number of collisions is determined by $\mathbb{P}[C]$ per epoch [time unit], where $\mathbb{P}[C]$ is the $\mathbb{P}$robability of Collisions. By **the law of total probability**, $\mathbb{P}[C] = \sum_{i=0}^{3} \mathbb{P}[C|S = i]\mathbb{P}[S = i]$, where S is denotes the current state of the system. If you are given that you are at State i, we need only care about the summands in each P_{ij} which contains a p^2 or p^3 in the product, where P_{ij} is the entry in the (i, j) position of P, for each fixed i and $j = 0, 1, 2, 3$. This is because

those terms with p^2 or p^3 denote times when there are 2 or 3 active nodes attempting to send a message at the same time, respectively. If more than one node attempts to send a message, we then have a collision.

Thus:

$$\begin{aligned}
\mathbb{P}[C|S=0] &= \binom{3}{2}q^2(1-q)p^2 + q^3p^3 + \binom{3}{2}q^3(1-p)p^2 \\
\mathbb{P}[C|S=1] &= \binom{2}{1}q(1-q)p^2 + q^2p^3 + \binom{3}{2}q^2p^2(1-p) \\
\mathbb{P}[C|S=2] &= (1-q)p^2 + qp^3 + \binom{3}{2}q(1-p)p^2 \\
\mathbb{P}[C|S=3] &= p^3 + \binom{3}{2}p^2(1-p).
\end{aligned}$$

And $\mathbb{P}[S=i] = \pi[i]$, where π is the eigenvector of eigenvalue 1 for $I_4 - P'$, which are the stationary state probabilities of the ALOHA model.

So for $p = 0.4$ and $q = 0.3$ this sum $\mathbb{P}[C] = \sum_{i=0}^{3} \mathbb{P}[C|S=i]\mathbb{P}[S=i] = 0.1791231$, according to our `Problem3.R`.

By the Markov property in Equation (1) in **1.** above, we have that we have, on average, $\mathbb{P}[C]$ collisions happen with ever next epoch. Thus, if we run over some sufficiently large N epochs [time unit], we will have around $N\mathbb{P}[c]$ collisions. The the average time between collisions over these N epochs is

$$\lim_{N\to\infty} \frac{N}{N\mathbb{P}[C]} = \lim_{N\to\infty} \frac{1}{\mathbb{P}[C]} = \frac{1}{0.1791231} = 5.582754.$$

Thus, we have on average 5.58 epochs between each collision for our ALOHA net with 3 stations.

4. *Consider a Markov chain with finite state space.*

(a) *Write an* `R` *function with call form*
`eTij(P) # P is the transition matrix`
giving the values of ET_{ij}.
This is given in our `Problem4.R`.

(b) *Do the same for variance, writing a function*
`varTij(P) # P is the transition matrix`
that finds all Var(Tij), where Tij is the same it takes to go from state i to state j. Show your derivation in a `.tex` *file.*

The derivation is as follows: Given d states, fix $i, j \in \{1, \cdots, d\}$.

We already have the expression $\mathbb{E}[T_ij]$, as given by our function `ETij(P)` in our `Problem4.R`. That quantity squared will be part of our second summand in the formula,

$$\mathrm{Var}(T_{ij}) = \mathbb{E}[T_{ij}^2] - \mathbb{E}[T_{ij}]^2. \tag{2}$$

Then the problem lies with $\mathbb{E}[T_{ij}^2]$. This term is derived by definition of T_{ij} given $U = k \in \{1, \cdots, d\}$, the next step after "starting" at State i. We still have that if given $U = j$, then $T_{ij}^2 \equiv 1^2 = 1$. Now, if we're given $U = k \neq j$

$$T_{ij}^2 = (1 + T_{kj})^2 \quad \text{for } k \in \{1, \cdots, d\} \setminus \{j\}.$$

Thus, for all $i, j \in \{1, \cdots, d\}$:

$$\begin{aligned}
\mathbb{E}[T_{ij}^2] &= \sum_{k=1}^{d} p_{ik}\mathbb{E}[T_{ij}^2|U=k], && \text{by } \textbf{Total Law of Expectation};\\
&= p_{jj}\mathbb{E}[T_{jj}^2] + \sum_{k\neq j} p_{ik}\mathbb{E}[(1+T_{kj})^2], && \text{using } \textbf{PSB Equation (6.39)};\\
&= p_{jj} + \sum_{k\neq j} p_{ik}\mathbb{E}[1+2T_{kj}+T_{kj}^2], && \text{by above paragraph, } \mathbb{E}[T_{jj}^2]\equiv 1;\\
&= p_{jj} + \sum_{k\neq j} p_{ik}\left(1+2\mathbb{E}[T_{kj}]+\mathbb{E}[T_{kj}^2]\right), && \text{by linearity of expectation};\\
&= p_{jj} + \sum_{k\neq j} p_{ik} + 2\sum_{k\neq j} p_{ik}\mathbb{E}[T_{kj}] + \sum_{k\neq j}\mathbb{E}[T_{kj}^2]\\
&= 1 + 2\sum_{k\neq j} p_{ik}\mathbb{E}[T_{kj}] + \sum_{k\neq j}\mathbb{E}[T_{kj}^2]\\
&= 2\left(1+\sum_{k\neq j} p_{ik}\mathbb{E}[T_{kj}]\right) - 1 + \sum_{k\neq j}\mathbb{E}[T_{kj}^2], && \text{the ol' ``add zero'' trick};\\
&= 2\mathbb{E}[T_{ij}] - 1 + \sum_{k\neq j}\mathbb{E}[T_{kj}^2], && \text{by } \textbf{PSB Equation (6.41)}.
\end{aligned}$$

Here let β_i denote $E(T_{in}^2)$, and define $\beta = (\beta_1, \beta_2, ..., \beta_{n-1})'$, so we have

$$\beta_i = 2E[T_{in}] - 1 + \beta_k \tag{3}$$

Equation (3) is a system of linear equations, which we can write in matrix form. Then the code to solve the system is in our `Problem4.R` function `findbeta`.
Then, using Equation (2) above, we conclude that:

$$\text{Var}(T_{ij}) = \sum_{k\neq j}\mathbb{E}[T_{kj}^2] \quad - \mathbb{E}[T_{ij}]^2 + 2\mathbb{E}[T_{ij}] - 1.$$

The function `varTij(P)` in our `Problem4.R` uses this forulation to return the value of $\text{Var}(T_{ij})$, for any transition matrix `P`.