EL2805 Reinforcement Learning

Exercise Session 1

October 27, 2019

Department of Automatic Control
School of Electrical Engineering
KTH Royal Institute of Technology

1 Exercises

Some of these exercises have been inspired by, or taken from, [1–4]. If you want to solve more exercises, see any of those books.

1.1

A fair die is tossed repeatedly. The maximum of the first n outcomes is denoted by X_n. Let $X_1 = 1$. Show that $(X_n)_{n\geq 1}$ is a Markov chain. Calculate the transition probability matrix, specify the classes, and determine which ones are recurrent and which ones are transient.

1.2

On a road, three trucks in four are followed by a car and one car in five is followed by a truck.

(i) What is the proportion of cars on this road?

(ii) If I see a truck pass me by on the road, on average how many vehicles pass before I see another truck?

1.3

Assume that $(X_n)_{n\geq 1}$ is a time-homogeneous discrete time Markov chain. Show that the n-step transition probabilities $P_{ij}^{(n)}$ for all $r < n$ satisfy

$$P_{ij}^{(n)} = \sum_k P_{ik}^{(r)} P_{kj}^{(n-r)}.$$

These are referred to as Chapman-Kolmogorov equations.

1.4

Consider the two-state Markov chain with the following transition matrix:

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix},$$

where $p, q \in [0, 1]$ and $p + q \neq 0$. Calculate the n-step transition probability matrix.

1.5

Let $(X_n)_{n\geq 1}$ be a random walk with state space $\{0, 1, 2, \dots\}$ and transition probabilities given by, $P_{00} = 1 - p$, $P_{01} = p$, and for any $i > 1$,

$$P_{ij} = \begin{cases} p & \text{if } j = i + 1 \\ 1 - p & \text{if } j = i - 1 \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$. Determine the value of p such that for each $j \geq 0$, $\lim_{n\to\infty} P_{ij}^n$ exists and is independent of i. Moreover, calculate the limiting probabilities.

1.6

A transition matrix P is said to be doubly stochastic if the sum of elements of each column of P is 1. If such a chain is irreducible and aperiodic, show that the uniform distribution is the stationary distribution.

1.7

Let X_n be the sum of n independent rolls of a fair die. Show that for any $k \geq 2$,

$$\lim_{n\to\infty} \mathbb{P}(X_n \text{ is divisible by } k) = \frac{1}{k}.$$

1.8

Consider a Markov chain with state space $\{-3, -2, \dots, 3\}$. Let $X_0 = 0$. For $n \geq 1$, if $X_n \neq -3, 3$, then X_n is $X_{n-1} - 1$ or $X_{n-1} + 1$ with equal probability. Otherwise, the chain gets reflected off the endpoint, i.e., from 3 it always goes to 2 and from -3 it always goes to -2.

(i) Is $|X_0|, |X_1|, \dots$ also a Markov chain?

(ii) Let sgn be the sign function, i.e., $\text{sgn}(x) = 1$ if $x > 0$, $\text{sgn}(x) = -1$ if $x < 0$, and $\text{sgn}(0) = 0$. Is $\text{sgn}(\text{X}_0), \text{sgn}(\text{X}_1), \dots$ a Markov chain?

1.9

Consider a finite Markov chain with n states, stationary distribution π, and transition probabilities P_{ij}. Imagine starting the chain at time 0 and running it for m steps, obtaining the sequence of states $X_0, X_1, \dots, X_m$. Consider the states in reverse order, $X_m, X_{m-1}, \dots, X_0$.

(i) Argue that given X_{k+1}, the state X_k is independent of $X_{k+2}, X_{k+3}, \dots, X_m$. Thus the reverse sequence is Markovian.

(ii) Argue that for the reverse sequence, the transition probabilities Q_{ij} are given by

$$Q_{ij} = \frac{\pi_j P_{ji}}{\pi_i}.$$

(iii) Prove that if the original Markov chain is time reversible, so that $\pi_i P_{ij} = \pi_j P_{ji}$, then $Q_{ij} = P_{ij}$. That is, the states follow the same transition probabilities whether viewed in forward order or reverse order.

1.10

Consider a pool of N interconnected web pages. Let r_i be the rank of page i. The PageRank algorithm used in the early days of Google, determines the page ranks by the following linear equations:

$$r_i = 1 - d + \sum_{j:i \text{ is linked to } j} \frac{dr_j}{C_j}, \quad i = 1, \ldots, N,$$

where $0 < d < 1$ is a given parameter and C_j is the total number of links contained in page j.

Model the network of web pages as a Markov chain, in which the page ranks are proportional to the corresponding stationary probabilities. The implies that if you wander around the web randomly according to this Markov chain, in the long run, the probability of visiting any page converges, and is independent of your start point.

1.11

Suppose a virus can exist in N different strains and in each generation either stays the same, or with probability μ mutates to another strain, which is choosen (uniformly) at random. What is the probability that the strain in the nth generation is the same as the initial strain (0th generation)?

Hint: Use the result of Problem 1.4.

1.12

Consider a sequence of zero-mean, finite variance, i.i.d. (independent and identically distributed) random variables $X_1, X_2, \ldots, X_N$. How does the variance of their sum scale with N?

2 Solutions

Solution to Problem 1.1

Let Z_n be the outcome of the nth toss. Then $X_{n+1} = \max(X_n, Z_n)$, and hence $(X_n)_{n\geq 1}$ is a Markov chain with state space $\{1, 2, \ldots, 6\}$ and transition probability matrix given by:

$$P = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 2/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 3/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 4/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 0 & 5/6 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

No two states communicate with each other, and hence we have six communication classes $\{1\}, \{2\}, \ldots, \{6\}$. The class $\{6\}$ is recurrent and the rest are transient.

Solution to Problem 1.2

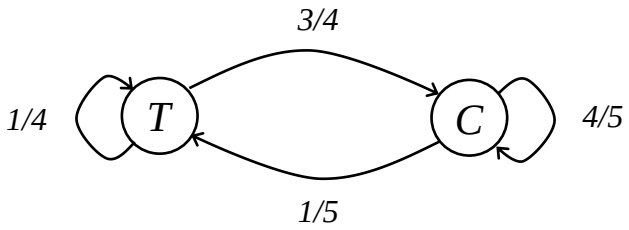

Figure 1: Markov chain for Problem 1.2

(Part (i).) The sequence of vehicles can be described by a Markov chain with state space $\{T, C\}$, with the transition graph shown in Figure 2. The proportion of cars is simply given by $\pi(C)$, where $\pi = [\pi(T),\ \pi(C)]$ is the stationary distribution of the Markov chain satisfying

$$\pi = \pi \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \\ \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix}, \quad \pi\mathbf{1} = 1, \quad \pi \geq 0.$$

Simple calculation gives $\pi(C) = \frac{15}{19}$.

(Part (ii).) If a truck is seen, then the average number of vehicles that pass by before another truck is seen corresponds to the mean recurrence time to state T, given that the chain is currently in state T. The mean recurrence time to state T is $h_{T,T} = \frac{1}{\pi(T)} = \frac{19}{4}$.

Solution to Problem 1.3

Fix $r < n$. By the law of total probability and the Markov property, it then follows that

$$\begin{aligned}
P_{ij}^{(n)} &= \mathbb{P}(X_n = j | X_0 = i) \\
&= \sum_k \mathbb{P}(X_n = j,\ X_r = k | X_0 = i) \\
&= \sum_k \frac{\mathbb{P}(X_n = j,\ X_r = k,\ X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_k \frac{\mathbb{P}(X_n = j | X_r = k,\ X_0 = i)\mathbb{P}(X_r = k,\ X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_k P_{kj}^{(n-r)}\ \frac{\mathbb{P}(X_r = k,\ X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_k P_{kj}^{(n-r)} P_{ik}^{(r)}.
\end{aligned}$$

Note that we used the Markov property in the second last equality to conclude that $\mathbb{P}(X_n = j | X_r = k,\ X_0 = i) = \mathbb{P}(X_n = j | X_r = k)$.

Solution to Problem 1.4

One way to solve this problem is to use the result of previous problem. In order to get more insights, we use another approach. In order to compute P^n, namely the n-th power of P, we first derive the spectral decomposition of matrix P, i.e., factorize P as

$$P = Q\Lambda Q^{-1},$$

where Λ is a diagonal matrix, whose diagonal elements are the eigenvalues of P, and where columns of Q are the corresponding eigenvectors. Let λ be an eigenvalues of P. Then,

$$P\lambda = q\lambda, \quad \lambda \neq \mathbf{0}. \tag{1}$$

Substituting the given P into (1), it follows that eigenvalues of P are the roots of following quadratic equation:

$$\lambda^2 - (2 - p - q)\lambda + 1 - p - q = 0,$$

which is satisfied by $\lambda = 1$ and $\lambda = 1 - p - q$. Using (1), it is easy to verify, through elementary calculations, that the eigenvector corresponding to $\lambda = 1$ is $[1,\ 1]^\top$, and that corresponding to $\lambda = 1 - p - q$ is $[p,\ -q]^\top$. Hence,

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 1 - p - q \end{bmatrix}, \qquad Q = \begin{bmatrix} 1 & p \\ 1 & -q \end{bmatrix}.$$

Now,

$$\begin{aligned}
P^n &= Q\Lambda^n Q^{-1} \\
&= \begin{bmatrix} 1 & p \\ 1 & -q \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 - p - q \end{bmatrix}^n \begin{bmatrix} 1 & p \\ 1 & -q \end{bmatrix}^{-1} \\
&= \frac{1}{p+q} \begin{bmatrix} 1 & p \\ 1 & -q \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (1 - p - q)^n \end{bmatrix} \begin{bmatrix} q & p \\ 1 & -1 \end{bmatrix} \\
&= \frac{1}{p+q} \begin{bmatrix} q + p\alpha^n & p(1 - \alpha^n) \\ q(1 - \alpha^n) & p + q\alpha^n \end{bmatrix},
\end{aligned}$$

where $\alpha = 1 - p - q$. The magnitude of α determines how rapidly the Markov chain forgets its initial condition.

Solution to Problem 1.5

The limit $\lim_{n\to\infty} P_{ij}^n$ exists for each j, and is independent of i, if there is a unique stationary distribution, namely the following set of equations has a unique solution:

$$\pi = \pi P, \quad \pi \mathbf{1} = 1, \quad \pi \geq 0.$$

By expanding $\pi = \pi P$, we observe that

$$\begin{aligned}
\pi_0 &= (1-p)\pi_0 + (1-p)\pi_1 \quad \Rightarrow \quad p\pi_0 = (1-p)\pi_1 \\
\pi_1 &= p\pi_0 + (1-p)\pi_2 \quad \Rightarrow \quad p\pi_1 = (1-p)\pi_2 \\
&\vdots \\
\pi_i &= p\pi_{i-1} + (1-p)\pi_{i+1} \quad \Rightarrow \quad p\pi_i = (1-p)\pi_{i+1}.
\end{aligned}$$

Hence, we deduce that, for all i,

$$\pi_i = \left(\frac{p}{1-p}\right)\pi_{i-1} = \left(\frac{p}{1-p}\right)^2 \pi_{i-2} = \ldots = \left(\frac{p}{1-p}\right)^i \pi_0.$$

In order to satisfy $\pi\mathbf{1} = 1$, we require that the geometric series $\sum_{i=1}^{\infty}\left(\frac{p}{1-p}\right)^i$ converges. Hence, we must have $p/(1-p) < 1$ or equivalently, $p < 1/2$. Thus, for $p < 1/2$, this irreducible, aperiodic Markov chain which is positively recurrent has limiting probabilities.

Now, for $p < 1/2$, the condition

$$\pi_0 \sum_{i=1}^{\infty}\left(\frac{p}{1-p}\right)^i = 1$$

yields

$$\pi_0 = \frac{1-2p}{1-p}.$$

Hence, the limiting probabilities are given by

$$\pi_i = \left(\frac{p}{1-p}\right)\frac{1-2p}{1-p}, \quad \forall i \geq 0.$$

Solution to Problem 1.6

Assume that there are M states in the Markov chain indexed by $i = 1, \ldots, M$. Irreducibility implies that all states are communicating with other, and aperiodic means that all states are aperiodic. These two assumptions ensure that the Markov chain has a unique stationary distribution, namely the following set of equations has a solution, and that it is unique:

$$\pi = \pi P, \quad \pi \mathbf{1} = 1, \quad \pi \geq 0.$$

It is easy to verify that if P is doubly stochastic, then $\pi = [\frac{1}{M}, \frac{1}{M}, \ldots, \frac{1}{M}]$ satisfies the above equation. To do so, writing for state j, we have that:

$$\pi_j = \sum_{i=1}^{M} \pi_i P_{ij}.$$

Now substituting $\pi_i = \frac{1}{M}$ into the above equation verifies the claim:

$$\sum_{i=1}^{M} \frac{1}{M} P_{ij} = \frac{1}{M}\sum_{i=1}^{M} P_{ij} = \frac{1}{M}.$$

Solution to Problem 1.7

Fix $k \geq 2$ and let $Y_t = \text{mod}(X_t, k)$. Then $Y_0, Y_1, \ldots$ is a Markov chain with k states and $Y_0 = 0$. The transition probabilities of this Markov chain are given by

$$P_{ij} = \sum_{\ell=1}^{6} I\{(i+\ell) \bmod k = j\} \times \frac{1}{6}.$$

X_t is divisible by k if and only if $Y_n = 0$. Furthermore, $\mathbb{P}(Y_n = 0) = P_{00}^n$. This is a finite state ergodic Markov chain and we show that the corresponding transition matrix P is doubly stochastic, namely each column of P and each row of P add up to one. For every j we have

$$\begin{aligned}
\sum_{i=0}^{k-1} P_{ij} &= \sum_{i=0}^{k-1} \sum_{\ell=1}^{6} I\{(i+\ell) \bmod k = j\} \times \frac{1}{6} \\
&= \frac{1}{6} \sum_{\ell=1}^{6} \sum_{i=0}^{k-1} I\{(i+\ell) \bmod k = j\} \\
&= \frac{1}{6} \sum_{\ell=1}^{6} 1 = 1.
\end{aligned}$$

Since P is doubly stochastic, the corresponding stationary distribution is uniform distribution (see, e.g., the previous question). Hence,

$$\lim_{n\to\infty} \mathbb{P}(X_n \text{ is divisible by } k) = \frac{1}{k}.$$

Solution to Problem 1.8

(Part (i).) The sequence $|X_0|, |X_1|, \ldots$ is also a Markov Chain. It can be viewed as the chain on state space $\{0, 1, 2, 3\}$ that moves left or right with equal probability, except that at 0 it bounces back to 1 and at 3 it bounces back to 2. Given that $|X_n| = k$, we know that $X_n = k$ or $X_n = -k$, and the information about $X_{n-1}, X_{n-2}, \ldots$ does not affect the conditional distribution of $|X_{n+1}|$.

(Part (ii).) No, this is not a Markov chain because knowing that the chain was at 0 recently affects how far the chain can be from the origin. For example,

$$\mathbb{P}(\text{sgn}(X_2) = 1 | \text{sgn}(X_1) = 1) > \mathbb{P}(\text{sgn}(X_2) = 1 | \text{sgn}(X_1) = 1,\ \text{sgn}(X_0) = 0),$$

since the conditioning information on the right-hand side implies $X_1 = 1$, whereas the conditioning information on the left-hand side says exactly that X_1 is 1, 2, or 3.

Solution to Problem 1.9

(Part (i).) We have that:

$$\begin{aligned}
\mathbb{P}(X_k | X_{k+1}, \ldots, X_m) &= \frac{\mathbb{P}(X_k, X_{k+1}, \ldots, X_m)}{\mathbb{P}(X_{k+1}, \ldots, X_m)} \\
&= \frac{\mathbb{P}(X_k)\mathbb{P}(X_{k+1}|X_k)\mathbb{P}(X_{k+2}, \ldots, X_m | X_k, X_{k+1})}{\mathbb{P}(X_{k+1})\mathbb{P}(X_{k+2}, \ldots, X_m | X_k)} \\
&= \frac{\mathbb{P}(X_k)\mathbb{P}(X_{k+1}|X_k)\mathbb{P}(X_{k+2}, \ldots, X_m | X_k)}{\mathbb{P}(X_{k+1})\mathbb{P}(X_{k+2}, \ldots, X_m | X_k)} \\
&= \frac{\mathbb{P}(X_k)\mathbb{P}(X_{k+1}|X_k)}{\mathbb{P}(X_{k+1})}.
\end{aligned}$$

Since this is a function only of X_k and X_{k+1}, we have the desired Markovian dependency on only the previous state.

Part (ii). Using the result of Part (i), it follows that for any i and j:

$$Q_{ij} = \mathbb{P}(X_k = j | X_{k+1} = i) = \frac{\mathbb{P}(X_k = j)\mathbb{P}(X_{k+1} = j | X_k = i)}{\mathbb{P}(X_{k+1} = i)} = \frac{\pi_j P_{ji}}{\pi_i}.$$

Part (iii). This follows directly from Part (ii), where we obtain $\pi_i Q_{ij} = \pi_j P_{ji}$, which can be true only if $Q_{ij} = P_{ij}$.

Solution to Problem 1.10

We consider a Markov chain where every page is a state, and where there is an imaginary restart page, which we label as state 0. We define the transition probabilities satisfying the following:

(a) Every state (page) i, transits to state (page) 0 with probability $1 - d$. Hence $P_{i0} = 1 - d$ for all i.

(b) When at state 0 (restart page), you navigate to a page chosen uniformly at random, so that $P_{0i} = \frac{d}{N}$ for all $i \neq 0$ and $P_{00} = 1 - d$.

(c) For all $i \neq 0$ and $j \neq 0$, the probability of going from state j to state i is

$$P_{ji} = \frac{d}{C_j} I\{j \text{ links to } i\},$$

where $A \mapsto I(A)$ is an indicator function, namely $I(A)$ equals 1 if A happens, and is zero otherwise.

The stationary probability distribution of the above Markov chain, denoted by π, satisfies:

$$\begin{aligned}
\pi_0 &= \sum_{j=0}^{N} \pi_j P_{j0} = \sum_{j=0}^{N} \pi_j (1 - d) = 1 - d, \\
\pi_i &= \pi_0 P_{0i} + \sum_{j=1}^{N} \pi_j P_{ji} = (1 - d)\frac{d}{N} + \sum_{j=1}^{N} \pi_j \frac{d}{C_j} I\{j \text{ links to } i\} \\
&= (1 - d)\frac{d}{N} + \sum_{j:\ i \text{ linked to } j} \pi_j \frac{d}{C_j}.
\end{aligned}$$

Multiplying both sides by $\frac{N}{d}$, we get

$$\frac{N}{d}\pi_i = (1 - d) + \sum_{j:\ i \text{ linked to } j} \left(\frac{N}{d}\pi_j\right) \frac{d}{C_j}.$$

Finally defining the page rank $r_i = \frac{N}{d}\pi_j$, we derive the page rank equations.

Answer to Problem 1.11

$$\frac{1}{N} + \left(1 - \frac{1}{N}\right)\left(1 - \frac{\mu N}{N - 1}\right)^n$$

Solution to Problem 1.12

Let $Z = X_1 + \cdots + X_N$. Then $\mathbb{E}\{Z\} = 0$ and $\text{var}\{Z\} = \mathbb{E}\{Z^2\} = N\mathbb{E}\{X_i^2\} = N\text{var}\{X_i\}$.

References

[1] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge University Press, 2017.

[2] S. Ghahramani, *Fundamentals of probability*, vol. 2. Prentice Hall, 2000.

[3] S. M. Ross, *A first course in probability*. Pearson Education International, 2009.

[4] J. R. Norris, *Markov Chains*. Cambridge University Press, 1998.