## Assignment 6: DTMC and Exponentials: Some Numerical Problems
- Due on October 2

Problem 4.76 in Chapter 4.

Hint 1: A knight can move to a square that is 2 squares horizontally and one square vertically, or two squares vertically and one square horizontally. In the figure, the black knight may move to any of eight squares (black dots); the white knight in the bottom-right corner (in our case) is limited to two squares (white dots).

Hint 2: Try to use RWoWG.
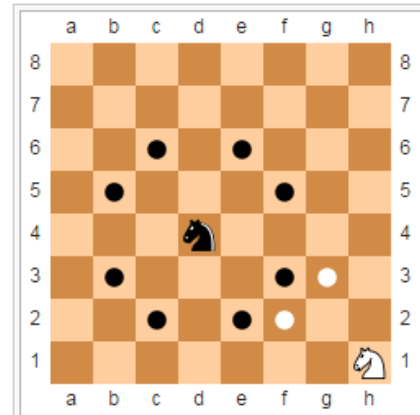
Problems 5.3 and 5.9 in Chapter 5.

Numerical problems:

1. **Pooh Bear.** Consider the Pooh bear example (ICPS-4, Prob.1). Let $\mathbf{P} \in \mathbb{R}^{3 \times 3}$ be the TPM.

    (a) Use Matlab (or any other language) to numerically compute (NOT by hand) $\boldsymbol{\pi} \equiv (\pi_A, \pi_B, \pi_C)$ following the approach described in the Appendix here.

    (b) Compute $\mathbf{P}^{10}$ and $\mathbf{P}^{50}$. What do you observe?

    (c) Plot $P_{A,A}^{(n)}$, $P_{A,B}^{(n)}$ and $P_{A,C}^{(n)}$ as functions of time n, for $n = 0, 1, 2, \ldots, 50$.
    Remark: This will visualize the convergence to steady state as claimed in Theorem 1(b).

2. **Markov Mouse in a Closed Maze.** Consider the Markov mouse in a closed maze (ICPS-4, Problem 2). Let $\mathbf{P} \in \mathbb{R}^{9 \times 9}$ be the TPM.

    (a) Use MatLab (or any other language) to numerically compute (NOT by hand) $\boldsymbol{\pi} \equiv (\pi_1, \ldots, \pi_9)$ using the approach described in the Appendix here.

    (b) Compute $\mathbf{P}^8$, $\mathbf{P}^9$, $\mathbf{P}^{18}$ and $\mathbf{P}^{19}$. What do you observe?

    (c) Plot $P_{1,1}^{(n)}$, $P_{1,7}^{(n)}$, $P_{1,2}^{(n)}$ and $P_{1,6}^{(n)}$ as functions of time n, for $n = 0, 1, 2, \ldots, 50$.
    Remark: This will visualize the convergence of the subsequence (and divergence of the whole sequence) as claimed in Theorem 3(b).

3. **Markov Mouse in an Open Maze**
    Consider the Markov mouse in an open maze (as in ICPS-6, Problem 2).

    (a) Let state 0 be the "outside world". Write the TMP $\mathbf{P} \in \mathbb{R}^{10 \times 10}$ in its canonical form. Compute $\mathbf{P}^{10}$ and $\mathbf{P}^{50}$.

    (b) Suppose we now want to know from which room the mouse escapes the maze. Let states 10, 11 and 12 denote the three exists. Write the TMP $\mathbf{P} \in \mathbb{R}^{12 \times 12}$ in its canonical form. Compute $\mathbf{P}^{100}$.

(c) Compute the matrix $\mathbf{B} = \mathbf{SR}$ and compare with $\mathbf{P}^{100}$. What do you see?

Remark: This helps visualize the asymptotic behavior of a reducible DTMC with one or more absorbing states.

4. **General Canonical Form of A Reducible DTMC**
   Consider the canonical form $\mathbf{P}$ in Other Problem 4 of HW5. Compute $\mathbf{P}^{100}$. What do you observe? How to interpret each number in $\mathbf{P}^{100}$?
   Remark: This helps visualize the asymptotic behavior of a reducible DTMC with one or more absorbing classes, and the asymptotic behavior of each recurrent/absorbing class.

5. **Inter-finish time distribution of the Skinny Turkey Half Marathon**.

   Every year on the morning of Thanksgivings, runners in Raleigh wait to start running the annul Skinny Turkey half marathon (2200 Wakefield Pines Drive, Raleigh). After the race, people can enjoy free turkeys (provided by the race organizer). I have run this half marathon twice in the past (in 2012 and 2013) and I enjoyed it very much.

   The goal here is to analyze the inter-finish time distribution and see if it looks like exponential. To do so we need to download the data and plot the histogram. Follow the procedure below:

   (a) Go to http://www.racetecresults.com/results.aspx?CId=16247&RId=2057 and export data as a CSV file. (You can see that I finished the race and was placed overall #24 and age group #2) When you open the file, you can delete all other columns (gender, names, etc.) but keep only the *net time* (which is given in the descending order).
   (b) To calculate the inter-finish times, first calculate the differences of every two consecutive net times in the original format. Then convert the difference into seconds by multiplying 24*3600. (Change the "cell format" from "hh:mm:ss" to "number").
   (c) Use the histogram function in the analysis toolpack. You will need the analysis toolpox add-ins. https://support.microsoft.com/en-us/kb/214269
   (d) Set the bin size for the histogram. Say we want to count the number of data point with inter-arrival times in (0,5), (5,10), (10,15), ..., (115, 120).
   (e) It is unnecessary to include all the data point. That is, you may disregard some outliers to get a better plot (e.g., by setting a bound of the bins to 120, which means you do not consider inter-finish times that are bigger than 120 seconds).
   (f) Make your bins as x-axis and frequency as y-axis, which is given by histogram. Plot it in column.

   Base on the histogram, do the inter-finish times of Skinny Turkey half look like exponential distribution?

Appendix:
Numerical Solution of Stationary Distribution π

Consider a finite state space DTMC. When it is irreducible, we know that the DTMC is "good", since all states are positive-recurrent. Let the state space be $\mathcal{S} \equiv \{1, \ldots, n\}$. We know that the following *balance equation* (BE) has a unique solution.

$$\boldsymbol{\pi}\,\mathbf{P} = \boldsymbol{\pi},$$
$$\sum_{i=1}^{n} \pi_i = 1. \tag{1}$$

Solving the BE involves solving a system of linear equations, which is equivalent to inverting matrices. However, when the transition probability matrix (TPM) $\mathbf{P}$ is big, solving the BE can be a headache! Let's now discuss on how to numerically solve the BE by programming (such as using MatLab).

First, note that there are n unknowns $(\pi_1, \ldots, \pi_n)$ but $(n+1)$ equations in (1). However, we know from HW4 that one of the first n equations in (1) is redundant. Define the following vectors:

$$\begin{aligned} \text{a column vector} \quad \mathbf{e} &\equiv (1, \ldots, 1)^T \in \mathbb{R}^{n \times 1}, \\ \text{a row vector} \quad \mathbf{e}_n &\equiv (0, \ldots, 0, 1) \in \mathbb{R}^{1 \times n}. \end{aligned}$$

Note that (1) is equivalent to

$$\boldsymbol{\pi}\,(\mathbf{I} - \mathbf{P}) = \mathbf{0},$$
$$\boldsymbol{\pi}\,\mathbf{e} = 1,$$

or

$$\boldsymbol{\pi}\,[\mathbf{A}|\mathbf{e}] = \boldsymbol{\pi}\,[\mathbf{A}_1, \ldots, \mathbf{A}_n|\mathbf{e}] = (0, \ldots, 0, 1) \in \mathbb{R}^{1 \times (n+1)}, \quad \text{where } \mathbf{A} \equiv \mathbf{I} - \mathbf{P}, \tag{2}$$

and $\mathbf{A}_k$ is the kth column of $\mathbf{A}$, $k = 1, \ldots, n$. Since one of the first n equations in (1) is redundant, we can drop a column in $\mathbf{A}$. Without loss of generality, let's drop the last column $\mathbf{A}_n$, then (2) becomes

$$\boldsymbol{\pi}\,[\mathbf{A}_1, \ldots, \mathbf{A}_{n-1}|\mathbf{e}] = \mathbf{e}_n \in \mathbb{R}^{1 \times n}, \tag{3}$$

which suggests that

$$\boldsymbol{\pi} = \mathbf{e}_n\,\mathbf{B}^{-1}, \quad \text{where } \mathbf{B} \equiv [\mathbf{A}_1, \ldots, \mathbf{A}_{n-1}|\mathbf{e}] \in \mathbb{R}^{n \times n}. \tag{4}$$

Of course, we have to justify that the matrix $\mathbf{B}$ is indeed invertible. That will take us to an advanced matrix analysis course. In fact, the TPM for a "good" DTMC is a special case of what's called the M matrix, which has a lot of nice properties.

To summarize, we can numerically obtain π using MatLab (or any other programming software) according to Equation (4).