Problem Set 7
Super Markov Brothers: Markov Chains and MDPs

AM121/ES121 — Fall 2016

Due 5 PM, Friday, November 18, 2016

Announcements

- The assignment is due by 5 PM, Friday, November 18, 2016.
- You may work with another student on this assignment and submit one writeup, but you must work together on every problem and state that you did this on your submission. It is ok to divide the writing up of the solutions, but not solving the problems.
- Readings: Lecture notes, section notes.
- **Extra credit** parts are truly optional and can only affect the final letter grade of students who are at grade boundaries.

Goals

This assignment will give you a feel for modeling, solving, and analyzing Markov Chains and Markov decision processes. It will also immerse you in the Mushroom Kingdom universe.

Contents

1 Let's gamble

Mario and Luigi are hanging out one afternoon and decide to put their coins where their game is. "The rules are simple," Mario says. "At each round, we play a Nintendo game. If I win, you give me a coin. If you win, I give you a coin. We play until I lose all my coins or you lose all your coins."

Mario has m coins at the start and Luigi has ℓ coins. They are equally matched in all Nintendo games, such that each wins in a game with probability 1/2.

This game can be described as follows: States i denote the amount Mario has won or lost at any given time. Since Mario can win no more than ℓ coins (all of Luigi's coins) and lose no more than m coins (all his coins), we have states i for $-m \leq i \leq \ell$. To model the absorbing states, the transition probability is:

$$p_{-m,-m} = 1 \text{ and } p_{\ell,\ell} = 1$$

For all states in between $-m$ and ℓ, there is equal probability of winning or losing a coin, and the transition probability $p_{i,j}$ is:

$$p_{i,j} = \begin{cases} 1/2, & \text{if } -m < i < \ell, j = i+1 \\ 1/2, & \text{if } -m < i < \ell, j = i-1 \\ 0, & \text{otherwise} \end{cases}$$

Task 1

Mario wants you to help him figure out his probability of winning. Answer the following questions.

1. Let $P(S_t = i)$ denote the probability that the chain is in state i at time t. What is $\lim_{t\to\infty} P(S_t = i)$ for any transient state i?
2. Which (if any) of the states are transient? Briefly justify your answer.
3. Which (if any) of the states are recurrent? Briefly justify your answer.
4. Let W_t be a random variable representing Mario's gains after t steps. Since the game is fair, the expectation $\mathbf{E}[W_t]$ is 0 for all t. Clearly then, $\lim_{t\to\infty} \mathbf{E}[W_t] = 0$ as well.

   Derive an expression for:

   $$\lim_{t\to\infty} \mathbf{E}[W_t]$$

   in terms of $q_i = \lim_{t\to\infty} P(S_t = i)$. (Note: remember some of the states may be recurrent states.)
5. Solve this expression for the probability that Mario wins.

End Task 1

2 Let's go clubbing

Mario and Luigi are running a new club called 'Peaches'. They expect the club to draw a whole lot of customers, and want to know how long the lines of people waiting to get in are likely to be.

By Mushroom Kingdom's restrictions, the length of the line can never exceed n people long. At any time step, exactly one of the following events occur:

- If the line is not empty, then with probability α the person in the front of the line gets into the club.
- If the line has fewer than n customers, with probability σ a new customer joins the line.

- The line remains unchanged with probability $1-\alpha-\sigma$ if the line is neither empty nor full, probability $1-\sigma$ if it is empty, and probability $1-\alpha$ if it is full

Assume $\alpha > 0, \sigma > 0$. This can be described as follows: Let S_t denote the number of customers in line at time t. We have $n+1$ states (from no customers to n customers) with the following transition probabilities:

$$p_{i,i+1} = \sigma \text{ if } i < n$$

$$p_{i,i-1} = \alpha \text{ if } i > 0$$

$$p_{i,i} = \begin{cases} 1-\sigma & \text{if } i = 0 \\ 1-\sigma-\alpha & \text{if } 1 \le i \le n-1 \\ 1-\alpha & \text{if } i = n \end{cases}$$

All other entries in the transition matrix are 0.

Task 2

1. Does there exist a stationary distribution (i.e., steady state probabilities that are independent of the start state)? Justify your answer.
2. Derive an expression for the stationary distribution π (for each state) as a function of α and σ. (Recall the balance equation $\pi = \pi \mathbf{P}$ for the steady state distribution.) [Hint: solve for the stationary distribution one equation at a time. By doing this, you should be able to form an expression for each state in terms of π_0.]
3. How does the stationary distribution look when $\alpha > \sigma$? When $\alpha < \sigma$? When $\alpha = \sigma$? You may wish to sketch a few graphs to help illustrate your description.
4. Assume that $n = 5$, and the line starts with 3 people when the club first opens. Under Mario's estimates for $\alpha = \frac{1}{4}$ and $\sigma = \frac{1}{10}$, what is the probability that the line has 2 people after 3 time steps? After 10 time steps? After 150 time steps? How do these probabilities compare to the stationary distribution? (Hint: you will want to use MATLAB or another mathematical software here.)
5. Present a 1-2 sentence argument for why looking at steady-state probabilities is useful in this setting. Then, present a 1-2 sentence argument for why steady-state probabilities are not useful for this setting (hint: what if the club is only open for 6 hours each night and each time step is 1 hour?).

End Task 2

3 Let's play Monopoly

After a long night of partying at the club the previous day, Mario and Luigi decide to take it easy today by inviting Peach over to play a game of Monopoly. Bowser had tapped in on the phone call, and upon hearing this news plans to join in on the fun. Being competitive, Bowser asks you for some support on how to win at this game. He assures you that he is not intending to cause any harm, but just wants to trade in his reputation of being brawny for being brainy.

You agree to help him. After thinking about this a little bit, you realize that while luck is a large part of Monopoly, there is some optimization that can be done. In particular, you think that knowing how likely a player is to land in any given square on the board may be useful information for making informed decisions about which properties to go for. You analyze the game board as shown in Figure 1.[1]

[1] `http://www.worldofmonopoly.co.uk/history/images/bd-usa.jpg`

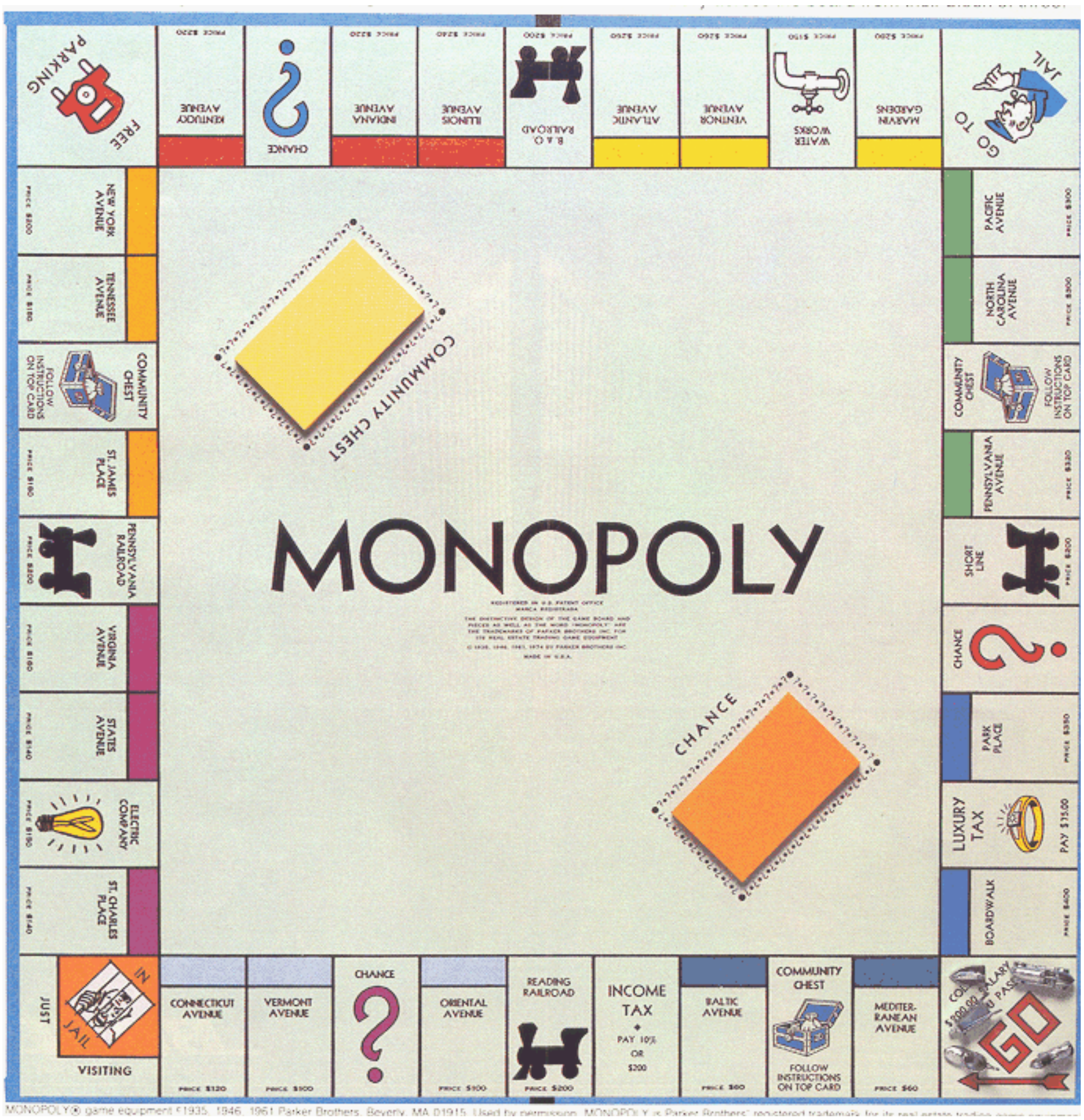

Figure 1: Monopoly game board

Task 3

You think you can model the problem as a Markov Chain. You can find the official rules of Monopoly here `https://en.wikipedia.org/wiki/Monopoly_(game)`, however, as a good applied mathematician you decide to start with the following simplifying assumptions on the rules:

- A player is sent to jail when she is landing directly on "Go to Jail." There is no need to model other cases that a player is sent to jail.
- A player in Jail stays there until either of the two following cases happens:
  1. she rolls a double (e.g., both dice show same number), at the same time, she will advance by the number represented by the roll.
  2. she has spent three turns in Jail. In this case, although the player is still in the same location, the state transforms from "In Jail" into "Just Visiting" and the player can participate as normal in the next turn.
- There is no need to model a "Get out of Jail" card.
- There is no need to model bankruptcy, winning the game, "Community Chest" or "Chance" cards. We're looking to simply model the basic navigation around the board (including the effect of jail.)
- There is no "doubles go twice" rule. If you roll a double, you do not get to go again.
- There is no need to model buying houses, collecting rent, or spending money.

You ask yourself the following questions.

1. Give a complete description of a Markov Chain that models a player moving about the board, where at every turn a player roll a pair of fair dice. Your description should be complete but concise, e.g., by making use of appropriate mathematical notation to avoid enumerating the transition probabilities from every state. Be sure to explain how your Markov Chain deals with the jail situation, e.g., "Just Visiting", "In Jail", and "Go to Jail".
2. Does the Markov Chain have a stationary distribution? Justify your answer.

   Suppose that you now compute this stationary distribution [**you don't actually need to to this unless you want to!**]. You want to make use of this information to come up with some advice for Bowser.
3. After you share your findings with Bowser, he buys "'St. James Place." Then to prove he is a good friend, he gives two "Get out of jail free" cards to his opponents. Explain why Bowser may not be the good friend he says he is. [Hint: for this you can reason about transitions to St. James Place but you don't need to think specifically about the stationary distribution.]
4. In thinking of possible strategies for Bowser, you would like to measure the cost-effectiveness of various monopolies (i.e. buying all of one block of properties). Making use of the stationary distribution, describe a simple method for measuring this. [Hint: you should take into account the gains from having property but also the cost of acquiring the property. You need not worry about second-order effects such as making another player bankrupt or needing to collect money from a specific other player.]
5. How might you use this cost-effectiveness measure to decide which properties to trade with other players? (A trade corresponds to property sold or gained with another player, and the cost of the trade.)

End Task 3

4 Let's help find that song

Enterprising plumbers that they are, Mario and Luigi have parlayed their heroic success into launching a new product line– selling portable music players online. To their (and our) surprise, the portable music player is selling extremely well, especially among Toads. The reviews for the product has been mostly positive, but there is one issue that keeps coming up. The playlist is sorted by song name, but there is no simple way to search and seek to a particular song. All the interface allows is 'next' and 'previous' in regular mode and 'next' in shuffle mode.

"I was listening to the Overworld theme," one blue Toad writes. "But I had this strong urge to hear the Underworld theme. I have 50 songs in the list between these two, and had to press next like 50 times. Maybe I should've gotten a music player with a scroll wheel." A yellow Toad replies: "There is a shuffle mode that goes to a random song. Maybe there is a smart way to find the songs we want?"

Mario lets you know about the situation and you think you can help. In particular, you think the following idea may make sense. First you can listen to the current song long enough to identify it, and if it is close in the playlist to the desired song, then you can navigate forward or back to find the song (the playlist is circular, and so either direction would eventually succeed). If the current song is far away, you can use the shuffle model until you identify a song that is close to the desired song, at which point you can switch to searching forward or back sequentially.

But how to do this in the best way? To begin answering these questions, you decide to model the problem as a Markov decision process (MDP).

Task 4

1. Model the state space of the MDP, assuming n songs. How many states are there?
2. Model the set of actions. Be precise about what the action does and how it relates to actual operations on the music player.
3. Model the reward function $R(s, a)$. Justify how you have assigned rewards to different actions in each states. [Note: there are different possible, reasonable answers here.]
4. Model the transition function $P(s, a, s')$. Recall that the playlist is circular.
5. State an objective criterion for solving the MDP.
6. (**Extra credit**) Implement a LP model for solving MDPs in AMPL. Encode your song-seeking MDP model in a data file (for a problem of a certain size) and solve. When should one seek sequentially and when should one shuffle? (note: you will probably want to generate the corresponding data via another tool and use AMPL's read command to load this into the data file.)
7. (**Extra Credit**) By experimenting with different size problems, when does one choose to seek instead of shuffle (as a function of the number of songs)?

End Task 4

5 Let's turn in the assignment

Final Task 5

You must turn in your assignment to the Canvas site by 5 PM, Friday, November 18, 2016. You may use 1 late day. (If you did the optimal AMPL exercise then turn this in as well.

End Task 5

Congratulations on completing your final AM121 assignment!