Grade 11/12 Math Circles

March 29, 2023

Reinforcement Learning 2 - Solutions

Problem Set 2 Solutions

1. (Continuing Problem 2 from Problem Set 1) Consider the following experiment in evolution. We are interested in the evolution of a gene by the name mTOR. The gene occurs in two variants T and S and plays an important role in determining the height of an individual. Each individual has a pair of this gene, either TT(tall) or TS(medium height) or SS(short). In the case where the individual is of medium height, the order of gene pairing is irrelevant (TS or ST is same). In evolution, an offspring inherits a pair of gene from each of his/her biological parents with equal probability. Thus if one partner is tall(TT) and the other partner is medium(TS), the offspring has 1/2 probability of being tall or 1/2 probability of being medium. We shall start with an individual of an arbitrary but fixed trait(TT or TS or SS). The other partner is of medium height (ST or TS). The offspring of such partners again finds a medium height partner. This pattern repeats for a number of generations, wherein the resultant offspring always has a medium height partner.

   (a) Write out the states and transition probabilities of this Markov process. (Note: you may have already answered this part)

   (b) Suppose we start with a tall height individual (as the first partner). What are the probabilities that any offspring belonging to the second generation would be tall, medium or short in height?

   (c) Suppose we start with a medium height individual (as the first partner). What are the probabilities that any offspring belonging to the second or third generation would be tall, medium or short in height?

   (d) What would be the answer to part (c) for any n-generations into future?

Solution:

(a) $States = \{TT, SS, ST\}$,

$$\mathscr{P} = \begin{array}{c} \\ TT \\ SS \\ ST \end{array} \begin{array}{c} \begin{array}{ccc} TT & SS & ST \end{array} \\ \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \end{array}$$

(b) $P^2(TT|TT) = \frac{3}{8}, P^2(SS|TT) = \frac{1}{8}, P^2(ST|TT) = \frac{1}{2}$

(c) $P^2(TT|ST) = \frac{1}{4}, P^2(SS|ST) = \frac{1}{4}, P^2(ST|ST) = \frac{1}{2}$
$P^3(TT|ST) = \frac{1}{4}, P^3(SS|ST) = \frac{1}{4}, P^3(ST|ST) = \frac{1}{2}$

(d) It will remain the same.
$P^n(TT|ST) = \frac{1}{4}, P^n(SS|ST) = \frac{1}{4}, P^n(ST|ST) = \frac{1}{2}$

2. Consider the game depicted in the following figure. 1

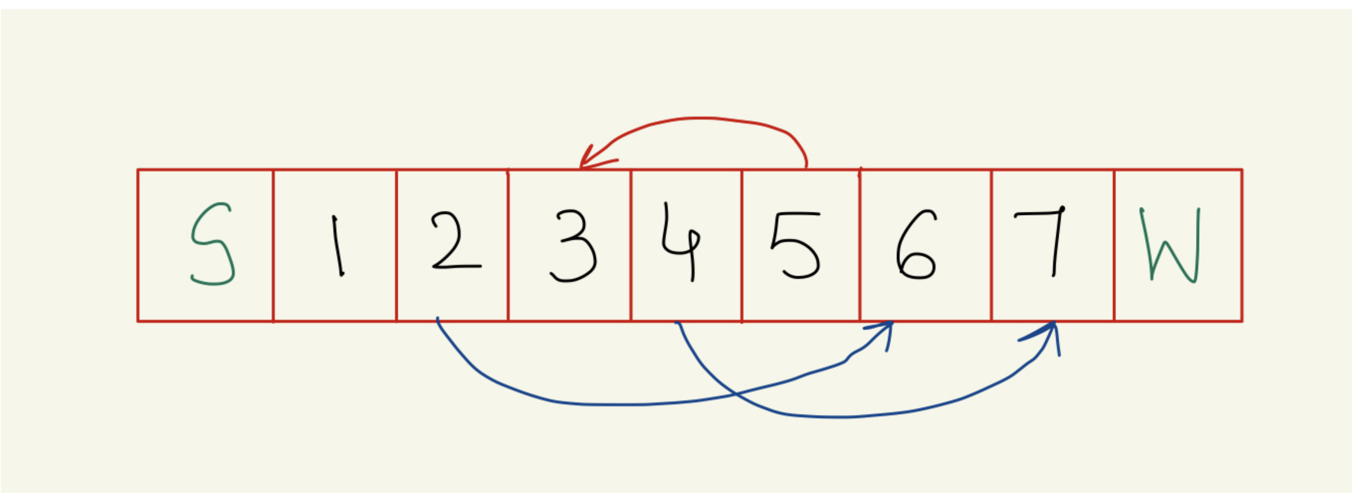

Figure 1: A snake and ladders equivalent game

- Initial state is S
- A fair four sided die is used to decide the next state at each time
- Player must land exactly on state W to win
- Die throws that take you further than state W leave the state unchanged
- If you land on 2 with a die throw, you climb the ladder to 6 directly. Similarly, if you land at 4 with a die throw, you climb the ladder to 7 directly.

- If you land at 5 with a die throw you are bit by a snake and have to go back to 3.

Based on this game:

(a) Identify the states, transition matrix of this Markov process.

(b) From the transition matrix, pick out absorbing state(s), if any.

(c) Construct a suitable reward function, discount factor and use the Bellman equation for the Markov reward process to compute how long does it take "on average" (the expected number of die throws) to reach the state W from any other state.

Solution:

(a) $States = \{S, 1, 3, 6, 7, W\}$,

$$\mathscr{P} = \begin{array}{c} \\ S \\ 1 \\ 3 \\ 6 \\ 7 \\ W \end{array} \begin{array}{c} \begin{array}{cccccc} S & 1 & 3 & 6 & 7 & W \end{array} \\ \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & \frac{2}{4} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{2}{4} & 0 \\ 0 & 0 & 0 & \frac{2}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

(b) State 'W' is an absorbing state.

(c) Reward: -1 for every state until state 'W' is achieved. Discount factor: $\gamma = 1$. Corresponding value function is:

$$\begin{bmatrix} V(S) \\ V(1) \\ V(3) \\ V(6) \\ V(7) \\ V(W) \end{bmatrix} = \begin{bmatrix} \frac{-23}{4} \\ \frac{-17}{3} \\ \frac{-16}{3} \\ -4 \\ -4 \\ 0 \end{bmatrix} = \begin{bmatrix} -5.7500 \\ -5.6667 \\ -5.3333 \\ -4.0000 \\ -4.0000 \end{bmatrix}.$$

Thus, the expected number of die throws to reach state W from states $S, 1, 3, 6, 7, W$ is $6, 6, 6, 4, 4$ respectively.