Autonomous Systems Lab 4

Eduard Alcobé, Aakash Maroti, Aaron Verdaguer

Provide the formalisation of the above problem as an infinite-horizon discounted reward MDP
M = ⟨S, A, (Pa)a∈A, r, s0, γ⟩

Since the dice is 6 sided, N here is 6.

Set S of states:

The state is represented by an array/list of length N+1. (N+1 is 7 for part A)
The ordering, while preserved in the list, is not essential to the representation of state.
Also, the length is N+1 (N+1 is 7 for part A) because of the pigeonhole principle, where the user will go bust at the N+1 th (N+1 is 7 for part A) number.

Values 1-N represent the value of the dice. -1 represents no roll.
N+1(N+1 is 7 for part A) represents the <Stop>.

Initial state s0,
[-1, -1, -1, -1, -1, -1, -1] for part A.

Set A of actions
Action 1: <Stop>
Action 2: Roll Dice

Transition probabilities Pa(s'|s) of going from state s to state s' when taking action a (for each action a and pair of states s, s'),
For Action 1: With probability 1 we add STOP to the state list, representing the end of the episode.

For Action 2: With probability 1/N (N is 6 for part A) we add the correspondent number from the dice to the state list. With probability 1/N times the number of rolls, you get reward 0 and the episode is ended if the rolled number is already in the state list.

Reward function r that encodes the reward r(a, s) obtained when taking action a in state s.
Accumulated reward for Action 1: Product of all numbers within the range (1,N) in the state list. So, the -1 are ignored.

Accumulated reward for Action 2: Product of all numbers within the range (1,N) in the state list but if there is a repeated number in the range (1,N) the final reward is 0.

To be faithful to the model seen in class, you also need to provide a discount factor γ
Γ = 0.9