

## Homework #1

# Finding the Optimal State-Value Function

## Problem

### Description

The game DieN is played in the following way:

1. You will be given a die with  $N$  sides. You will know the size of  $N$ , and can assume that  $N$  is a value greater than 1 and less than or equal to 30.
2. You will be given a boolean mask vector *isBadSide* where the value of 1 indicates the sides of the die that will make you lose. The vector will be of size  $N$ , and 1 indexed. (there is no 0 side)
3. You start with 0 dollars.
4. At any time you have the option to roll the die or to quit the game
  - a. If you decide to roll:
    - i. And you roll a number not in *isBadSide*, you receive that many dollars. (eg. if you roll the number 2 and 2 is not a bad side -- meaning the second element of the vector is 0 -- in *isBadSide*, then you receive 2 dollars)  
  
Repeat step 4.
    - ii. And you roll a number in *isBadSide*, then you lose all money obtained in previous rolls and the game ends.
  - b. If you decide to quit:
    - i. You keep all the money gained from previous rolls and the game ends.

### Procedure

- For this problem, determine an optimal policy for playing the game DieN for  $N$  sides. You will be given  $N$  and the array *isBadSide* which indicates which sides are bad. As you will see, the optimal policy for this game will depend on *your current bankroll*.
- You can try solving this problem either by creating an MDP of the game (state, action, transition, reward function, and assume a gamma of 1) and then calculating the optimal

---

state-value function or you can plug-in values and solve directly using the Bellman Equations.

- What is the expected amount of dollars for this game if you follow an optimal policy? That is, what is the optimal state-value function for the initial state of the game? Provide answers for the problems you are given on Canvas. Your answer must be correct to 3 decimal places.

## Examples

The following examples can be used to verify your calculation is correct.

- Input:  $N = 21$ ,  $\text{isBadSide} = [1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0]$ , Output: 7.379
- Input:  $N = 22$ ,  $\text{isBadSide} = [1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0]$ , Output: 6.314
- Input:  $N = 6$ ,  $\text{isBadSide} = [1, 1, 1, 0, 0, 0]$ , Output: 2.583

## Resources

The concepts explored in this homework are covered by:

- Lectures
  - Lesson 1: Smoov & Curly's Bogus Journey
- Readings
  - Littman (1996) (chapters 1-2)
- Libraries
  - <https://pymdptoolbox.readthedocs.io/en/latest/api/mdptoolbox.html>

## Submission Details

**The due date is indicated on the Canvas page for this assignment.**

Make sure you have set your timezone in Canvas to ensure the deadline is accurate.

To complete the assignment calculate answers to the specific problems given and submit the results on Canvas. You have a maximum of 10 attempts.

