CS 188
Summer 2022

Regular Discussion 6A

1 MDPs: Micro-Blackjack

In micro-blackjack, you repeatedly draw a card (with replacement) that is equally likely to be a 2, 3, or 4. You can either Draw or Stop if the total score of the cards you have drawn is less than 6. If your total score is 6 or higher, the game ends, and you receive a utility of 0. When you Stop, your utility is equal to your total score (up to 5), and the game ends. When you Draw, you receive no utility. There is no discount ($\gamma = 1$). Let's formulate this problem as an MDP with the following states: $0, 2, 3, 4, 5$ and a *Done* state, for when the game ends.

1. What is the transition function and the reward function for this MDP?

2. Fill in the following table of value iteration values for the first 4 iterations.

States	0	2	3	4	5
V_0					
V_1					
V_2					
V_3					
V_4					

3. You should have noticed that value iteration converged above. What is the optimal policy for the MDP?

States	0	2	3	4	5
π^*					

4. Perform one iteration of policy iteration for one step of this MDP, starting from the fixed policy below:

States	0	2	3	4	5
π_i	Draw	Stop	Draw	Stop	Draw
V^{π_i}					
π_{i+1}					

5. Consider a variant of this problem where we use a discount factor of $\gamma = 0.5$. What are the new optimal values and policy for this formulation? (Hint: You shouldn't need to recalculate all values. Start by thinking about V^* of 3, 4, and 5.)

States	0	2	3	4	5
V^*					
π^*					

2 Something Fishy

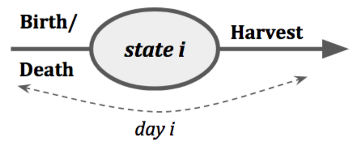
In this problem, we will consider the task of managing a fishery for an infinite number of days. (Fisheries farm fish, continually harvesting and selling them.) Imagine that our fishery has a very large, enclosed pool where we keep our fish.

Harvest (11pm): Before we go home each day at 11pm, we have the option to harvest some (possibly all) of the fish, thus removing those fish from the pool and earning us some profit, x dollars for x fish.

Birth/death (midnight): At midnight each day, some fish are born and some die, so the number of fish in the pool changes. An ecologist has analyzed the ecological dynamics of the fish population. They say that if at midnight there are x fish in the pool, then after midnight there will be exactly $f(x)$ fish in the pool, where f is a function they have provided to us. (We will pretend it is possible to have fractional fish.)

To ensure you properly maximize your profit while managing the fishery, you choose to model it using a Markov decision problem.

For this problem we will define States and Actions as follows:
State: the number of fish in the pool that day (before harvesting)
Action: the number of fish you harvest that day

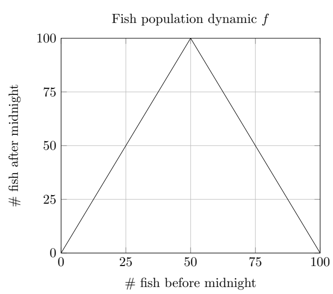


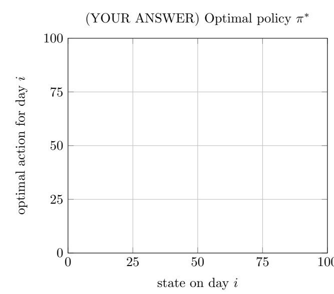
(a) How will you define the transition and reward functions?

$T(s, a, s') =$ ____

$R(s, a) =$ ____

(b) Suppose the discount rate is $\gamma = 0.99$ and f is as below. Graph the optimal policy π^*.





(c) Suppose the discount rate is $\gamma = 0.99$ and f is as below. Graph the optimal policy π^*.