150070036
Foundations of intelligent learning agents
Assignment -2

Gamblers problem

Formulation:

The MDP formulation has been mathematically defined as follows , for any MDP defining the transition probabilities and the reward function is sufficient to solve an MDP.

Transitions[s,a,s']= p if s'=s+a+1 and a in min(s,100-s) and s not in {0,100}

Transitions[s,a,s']=1-p if s'=s-a-1 and a in min(s,100-s) and s not in {0,100}

Transitions[s,a,s]=1 for all a and s in {0,100}

Reward[s,a,s']=1 s'=s+a+1 and s'=100 and a in min(s,100-s) and s not in {0,100}

Reward[s,a,s']=0 else

But,

As our MDP has states which are terminal states, while solving we need to make the value function of the terminal states as zero.This is because we do not have any equation for terminal state providing any information if we were to solve using linear programming and also if we were to find V π using the iterative update the value function of terminal states remains the value randomly initialized and all states which lead to terminal state might not converge to their value function and might simply explode and this was an observation while performing the experiments.

Plots and observation:

The plots are of value function of states s∈ S and the plots of policy is of all possible policies. Policies are considered alternative optimal policy if their corresponding action value function is within epsilon of each other.

I.e.

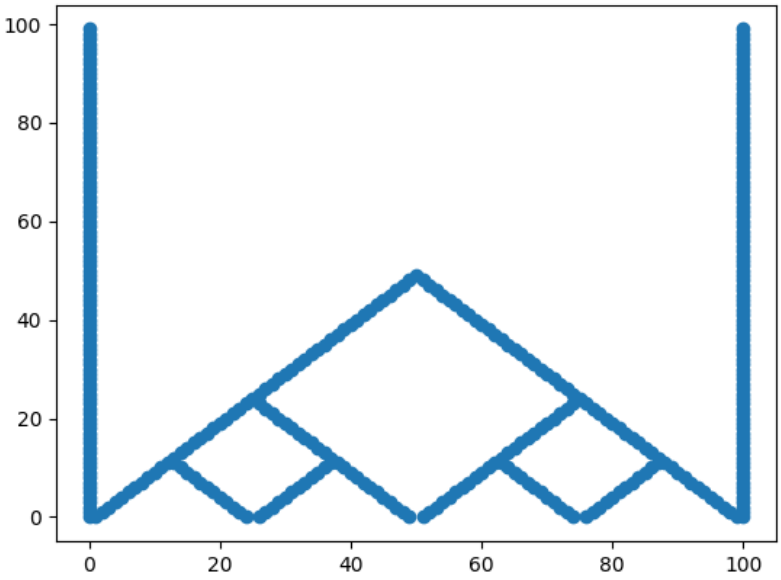
$$ABS\left(Q^{\pi}(a,s) - Q^{*}(a,s)\right) < \varepsilon \; for \; all \; s$$

THEN

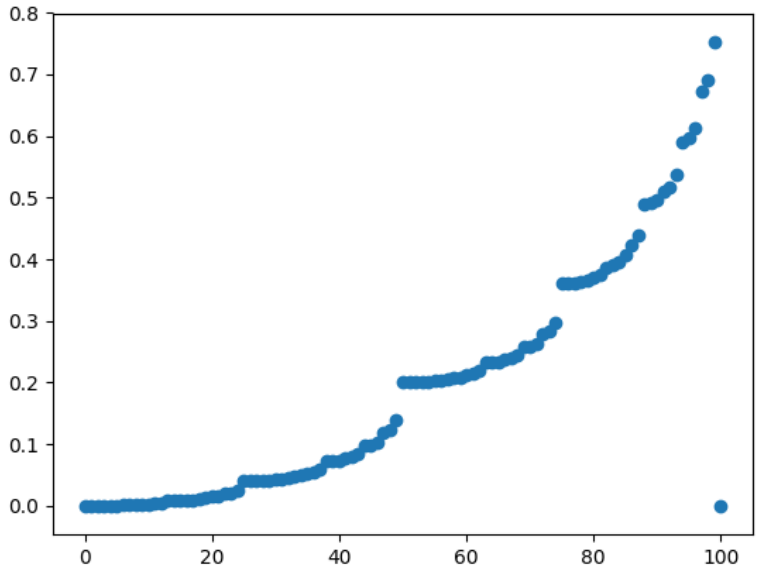
$$Q^{\pi}(a,s) \; is \; another \; optimal \; policy$$

P=0.2

Possible policies



value function



As states 0 and 100 are terminal states the the action value function is 0 for all actions hence all possible are optimal but are redundant at the terminal state.

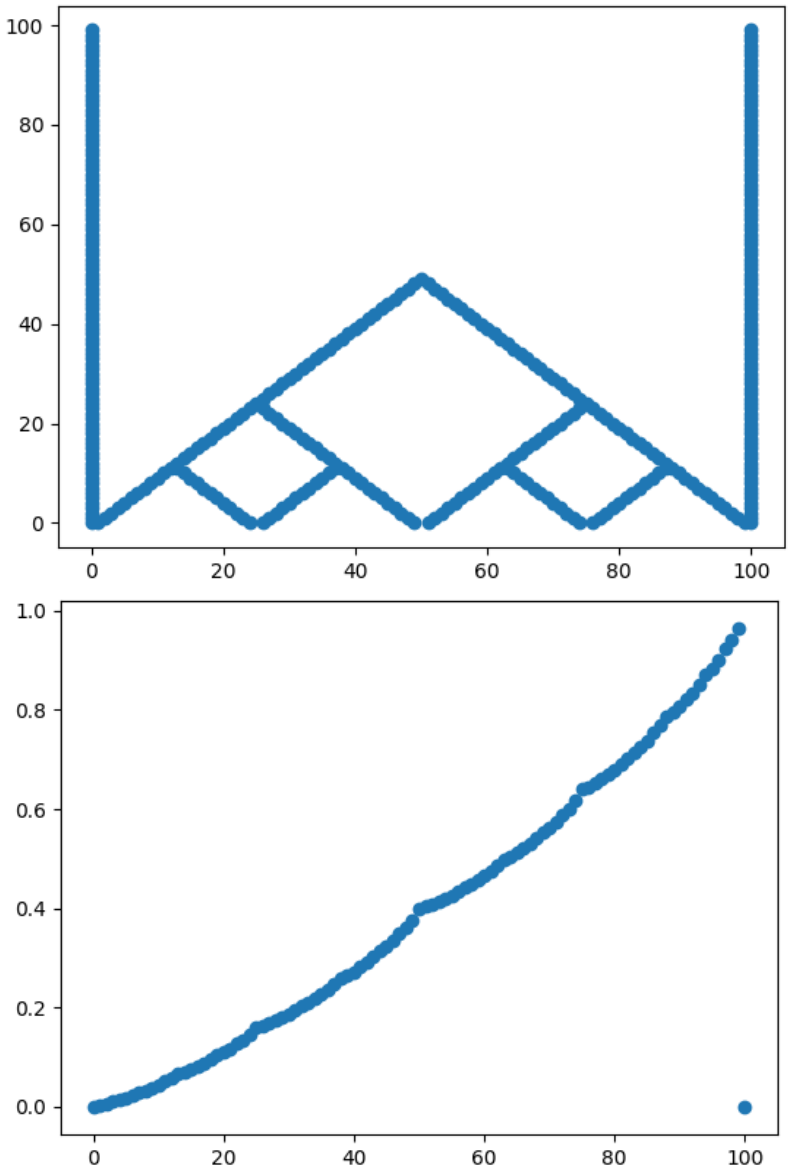
There is a sequence of triangles which suggests that the family of optimal policies include the policies that involve the gambler to set a target goal state less than s=100 and once he reaches that tries to bet everything to reach the higher state i.e. $\pi(a, s) = s - 1$ when s is one of the local goals.

There are some possible smaller sub goals which are not quite observable due to epsilon being larger than the max-norm of the corresponding policy and its action value function.

As for the value function there are some discrete jumps in value function at states which could act as a sub goal for the MDP.

P=0.4

Possible policies



As states 0 and 100 are terminal states the the action value function is 0 for all actions hence all possible are optimal but are redundant at the terminal state.

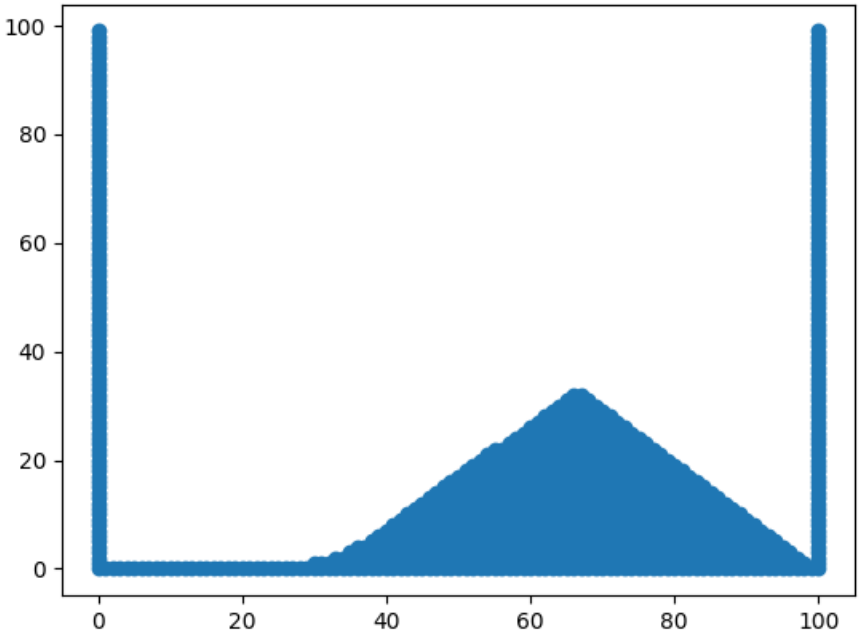
There is a sequence of triangles which suggests that the family of optimal policies include the policies that involve the gambler to set a target goal state less than s=100 and once he reaches that tries to bet everything to reach the higher state i.e. $\pi(a, s) = s - 1$ when s is one of the local goals.

There are some possible smaller sub goals which are not quite observable due to epsilon being larger than the max-norm of the corresponding policy and its action value function.

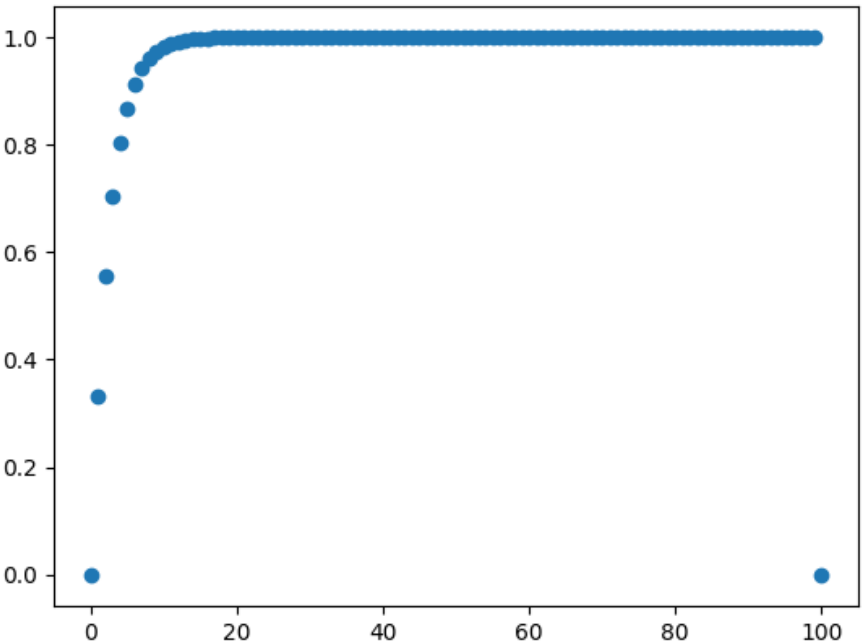
As for the value function there are some discrete jumps in value function at states which could act as a sub goal for the MDP but this time the jumps are much smaller.

P=0.6

Possible policies



value function

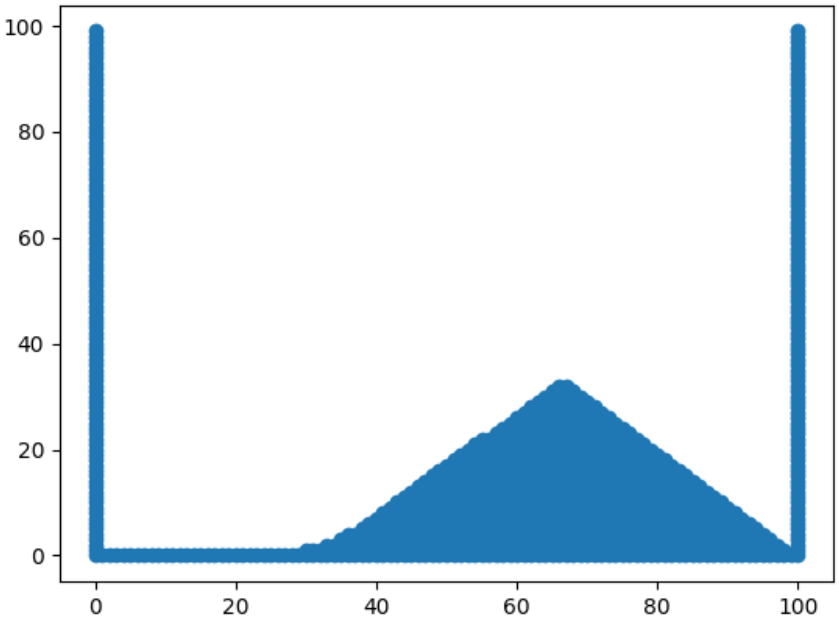


This is a peculiar case for p>0.5 . The family of optimal policies are such that as long as a certain capital is not established bet the minimum amount possible. Once we have reached the minimum capital we can bet any amount less than or equal to the current capital we have. As usual as states 0 and 100 are terminal states the the action value function is 0 for all actions hence all possible are optimal but are redundant at the terminal state.

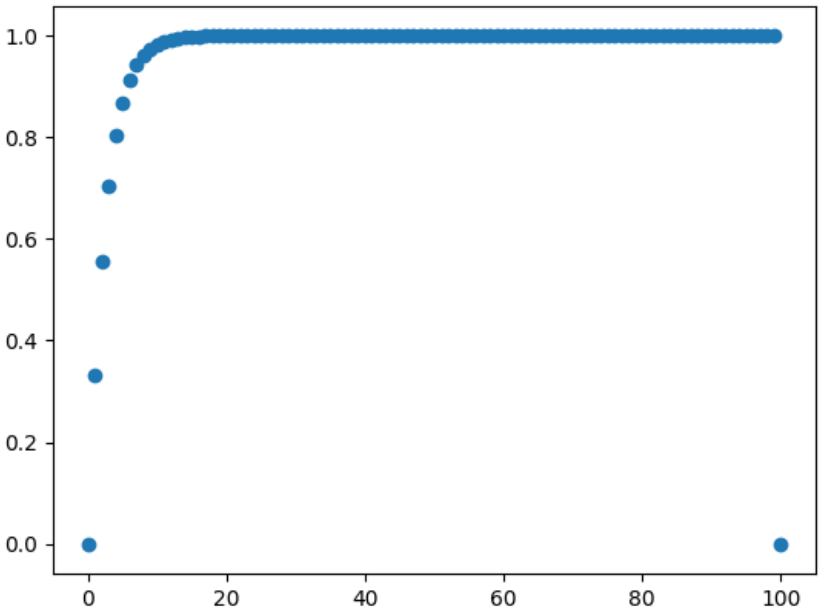
The value function has changed drastically as now it is almost constant above a certain threshold, and increases like a R-C charging curve. This might be due to the problem formulation and as p>0.5 as we have sufficient capital we may be able to bet any amount less than or equal to the current capital

P=0.8

Possible policies



value function



This is a peculiar case for p>0.5 . The family of optimal policies are such that as long as a certain capital is not established bet the minimum amount possible. Once we have reached the minimum capital we can bet any amount less than or equal to the current capital we have. As usual as states 0 and 100 are terminal states the the action value function is 0 for all actions hence all possible are optimal but are redundant at the terminal state.
The value function has changed drastically as now it is almost constant above a certain threshold, and increases like a R-C charging curve. This might be due to the problem formulation and as p>0.5 as we have sufficient capital we may be able to bet any amount less than or equal to the current capital. This plot is very similar for p=0.6