CS 747: Assignment 2 Report

Karthikeya

Autumn 2022

Contents

1 Task 1

In value iteration the condition I have used to end the algorithm is the maximum absolute difference between current and previous value functions should be less than 10^{-12}. I chose this value to be of that order because we need the accurate value upto 6 decimal places. To break tie-breaks in Howard policy(and other policies) I chose the action with the least serial number.

2 Task 2

States in my MDP are of the form (bb,rr,strike) where strike(0 or 1) represents the batsmen on strike. And also 2 extra states to encode Win and Loss. So in total number of states are $2 * |S| + 2$ where S is the set of states given in states.txt file. Now computing Transition and Reward functions is trivial. Also I have assumed that Batsmen A is initially on strike. I have used paramters in sample-p1.txt to get all the plot data.

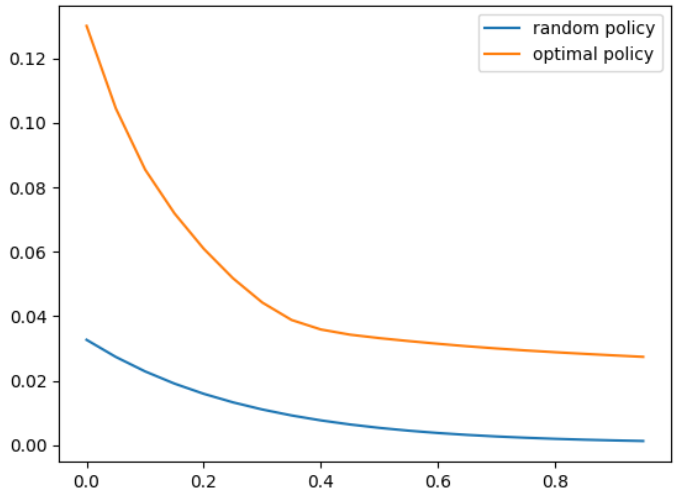

Figure 1: Win probability vs B's degree of weakness

As expected win probability of the optimal policy decreases with increase in B's weakness, also optimal policy performs much better than random policy.

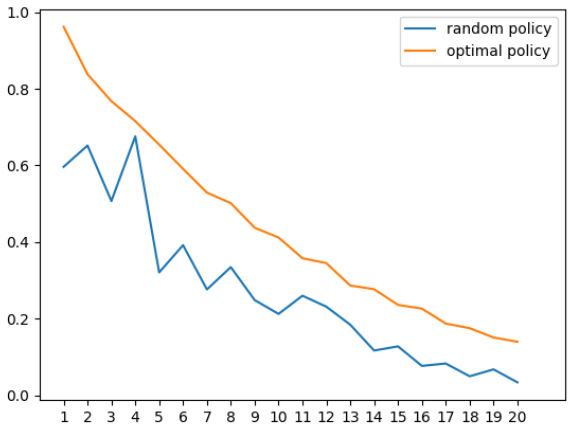

Figure 2: Win probability vs Runs required(with 10 remaining balls)

As expected win probability of the optimal policy decreases with increase in runs required, also optimal policy performs much better than random policy.

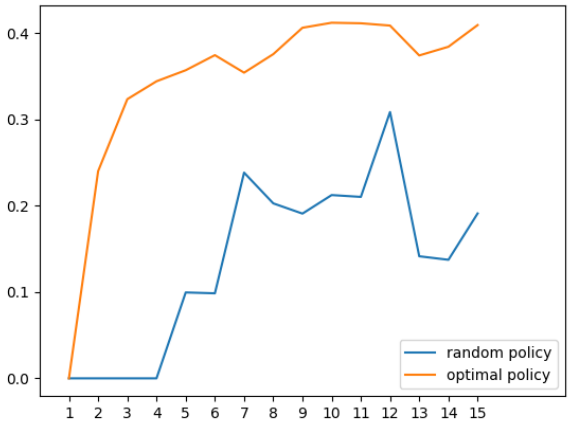

Figure 3: Win probability vs Balls remaining(with 10 runs required)

Here also optimal policy outperforms the random policy. But here we can notice that the optimal policy has some local minimas. We can see that this is happening only when x = 7 13. This is because after this ball the strike would most probably

change(because of the parameters I chose) and it most probably be difficult for B to change strike or score any runs in the next over.
Note that the random policy trend is completely random for the last 2 plots as one would expect.