Dynamic Programming and Reinforcement Learning Assignment2 Report

Geoffrey van Driessel (2639310), Chih-Chieh Lin (2700266)

November 26, 2020

Introduction

This report describes our solutions and results for solving the problems in a slowly deteriorating system. Regarding the four questions mentioned in this assignment, the relevant results and solutions will be placed in the following sections.

1 Stationary Distribution and Long-run Average Cost

When it comes to the calculation of stationary distribution, we have to find a certain stable distribution that would hardly change by multiplying it with the probability matrix. We assume that the procedure start from state 0. We then multiply this initial distribution with probability matrix until the distribution converges. To be more specific, we compare the last distribution with its previous one. If the difference between these two distribution is extremely small, the stationary distribution is found. The stationary distribution is shown in Figure 1.

After obtaining the stationary distribution, we then use it to calculate the Average replacement cost. We derived the Average replacement cost by multiplying the stationary distribution with reward vector(a vector starting ranging from 0.1 until 1 with steps of 0.01). The Average replacement cost we got is: **0.1461**.

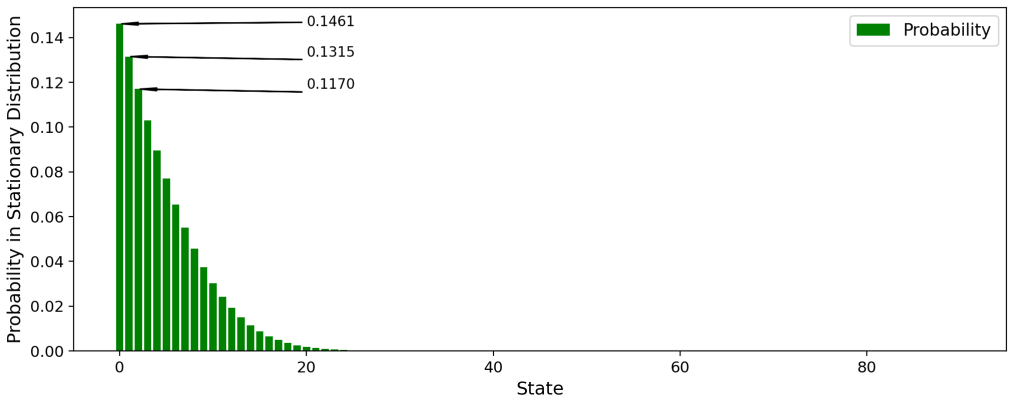

Figure 1: Stationary Distribution

2 Average Cost Solved by Poisson Equation

For the Poisson equation, we consider the following formula:

$$V + \phi = r + PV$$

Which we then rewrite in the following way to be able to solve the linear equation:

$$V(1 - P) + \phi = r$$

We then assume that V(0)=0 in order to solve the linear equations with enough conditions. Finally the average replacement cost calculated by Poisson equation is: **0.1461**.

3 Preventive replacement with policy iteration

We then introduce the possible decision to have a preventive replacement with cost 1/2. By doing so we transition from markov reward chains to markov decision chains. Our starting policy is simply always choosing to not do a preventive replacement. We then iterate from the last state to the first state to determine for each state the optimal policy. This is done by selecting the minimum cost. We then continue this process by going back one state recalculating the solution of the linear equation until we are at state 0.

We find a new long run average of **0.14492** which is slightly lower than before, as expected. We also find that from the state 0.23 (14th state) the optimal decision becomes to do a preventive replacement. Meaning that if the probability of a failure is 0.23 or higher, then the cost of paying for a preventive replacement is lower.

4 Preventive replacement with value iteration

Using value iteration we find the same long run average: **0.14492**. Also, we find that the optimal policy start to do preventive replacement after the 14th state, which has the same behavior as the one we solved in policy iteration.