Assignment 3: Investigating the policy

Liang Chen - 03730468

November 2019

Task 1:

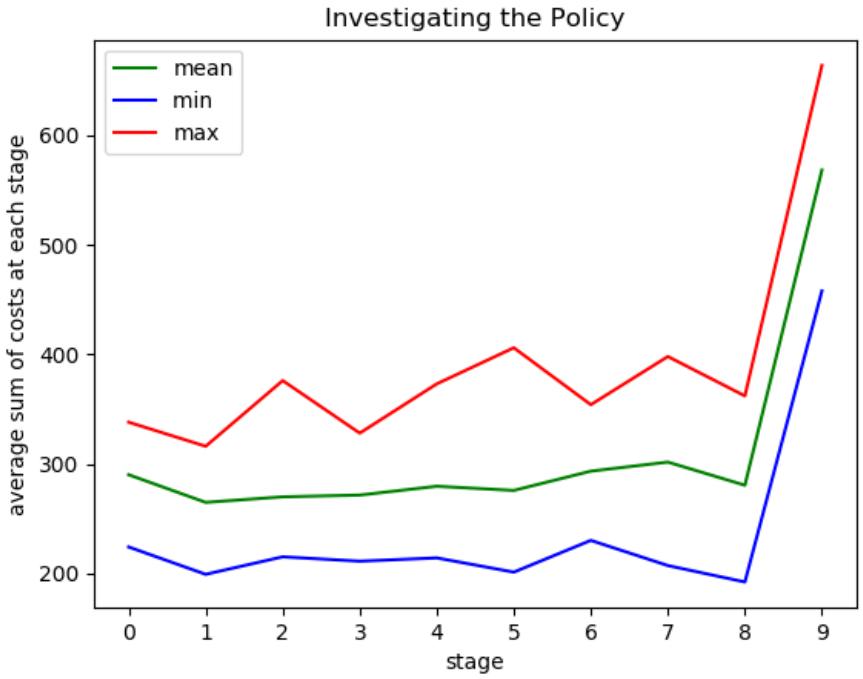
The following is a random generated probability matrix that denotes the chances of each worker to complete a task. The ij element of the matrix denotes the probability to complete the j-th type of job by the i-th worker.
Each time each worker draws a random number in [0,1]. The first job in the queue would be removed, if the random number drawn by the worker is larger than the corresponding entry in matrix,i.e. the chance to complete the second type of job by the first worker is 0.28(1-0.72). This means that the smaller entry in matrix, the higher the probability that the job would be removed from the queue.

$$\begin{bmatrix} 0.55 & 0.72 \\ 0.60 & 0.54 \\ 0.42 & 0.65 \end{bmatrix}$$

For verification of policy, a state with almost empty queues ((2,0,0),(0,0,0), (1,0,0),1) was selected. According to the results of the dynamic program, the action 3 was chosen because the third worker has a higher probability to complete the job in the next few round.
On the other hand, action 0 (being idle) was chosen in a state with almost full queues, which is defined as (((1,2,0),(1,2,1),(2,2,0),2)) in this case. Although there are two slots for this job, the scheduler decide to wait a round, to avoid the jam of queue.

Task 2:

The plot is shown in 1. From this plot, we could see that the trajectory, which was chosen by the policy, has a higher sum of cost-to-go from stage 8 to stage 9. This is because the policy at that time was not optimal. On the other hand, the curve is more stable because the policy was converging to an optimal policy, which causes the cost-to-go to be smaller.


Investigating the Policy
mean
min
max
average sum of costs at each stage
600
500
400
300
200
0
1
2
3
4
5
6
7
8
9
stage