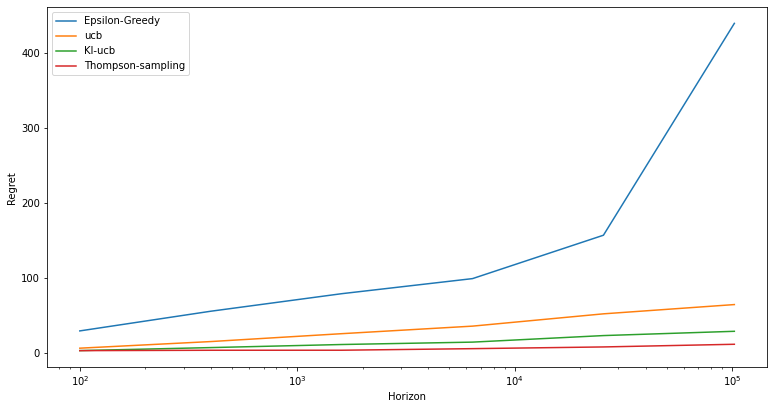
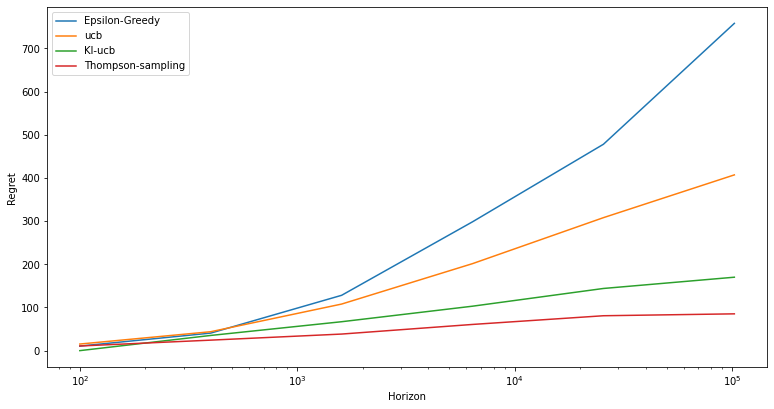
**Assignment 1**

**Task 1:**

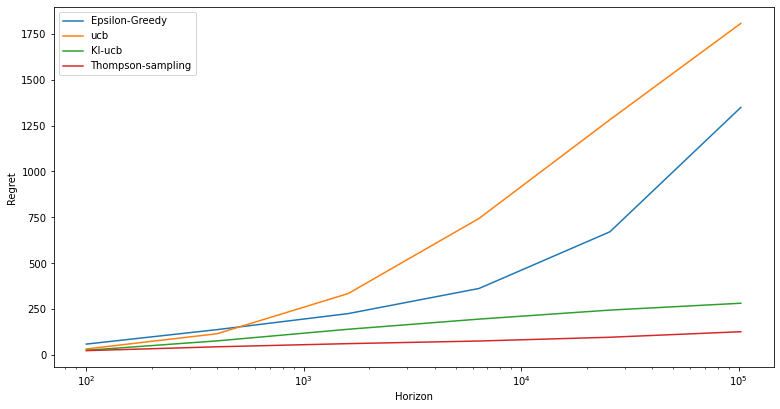
Instance 1



Instance 2

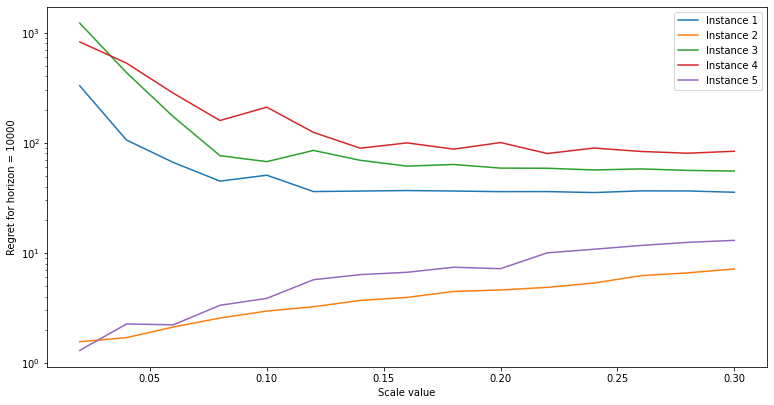


Instance 3



Here, the regret on the y axis is averaged over all the random seeds(0-49)

**TASK 2:**

****

Here, the regret on the y axis is averaged over all the random seeds(0-49)

Conclusions from graph:

* For instances 4,3,1 in task 2, the value of scale for which the regret is least is 0.3
* For instances 2 and 5 the value of scale for which the regret is least is 0.02

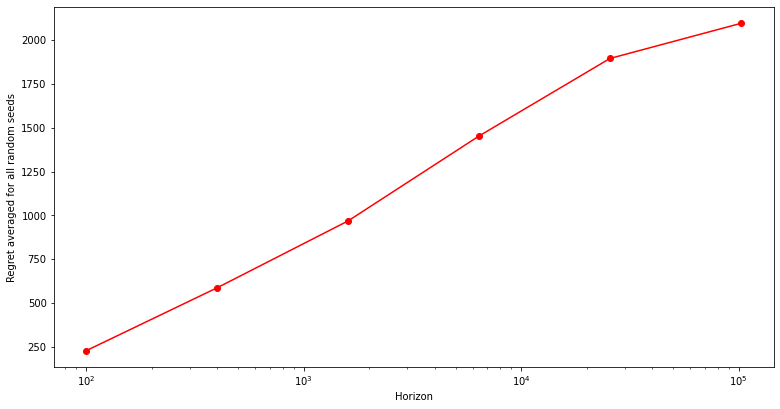
**TASK 3:**

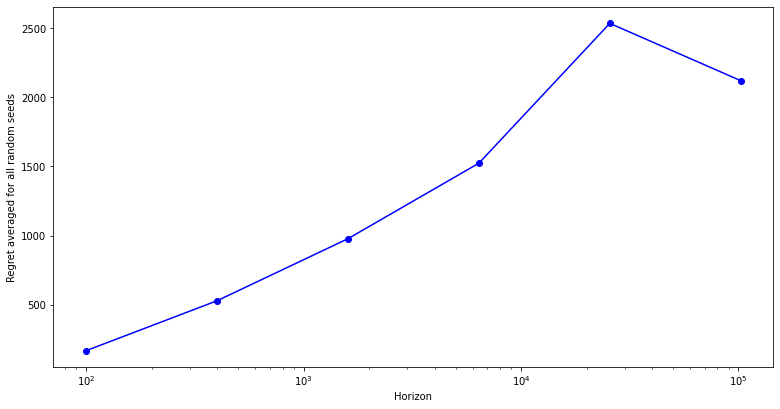
In this task, to maximize the reward, I have used Thompson’s sampling, in which the success at time step t for an arm is the sum of all the reward till that time from that arm and the failure is total no. of trials that arm is pulled minus the success.

After that I picked the arm which has max beta value where

Beta ~ beta(success+1, failure+1)

Instance 1



Instance 2 

**TASK 4:**

Here we are given multiple rewards ranging in [0,1] and their corresponding probability of occurrence .

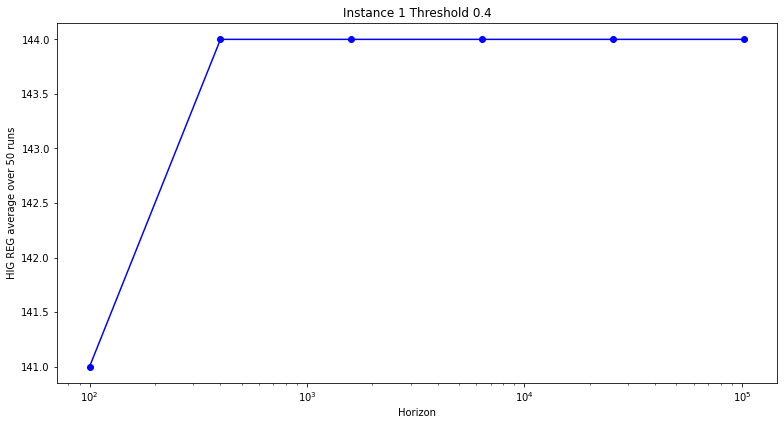
Now to calculate the number of HIGHS, we are given a threshold value.

Algorithm:

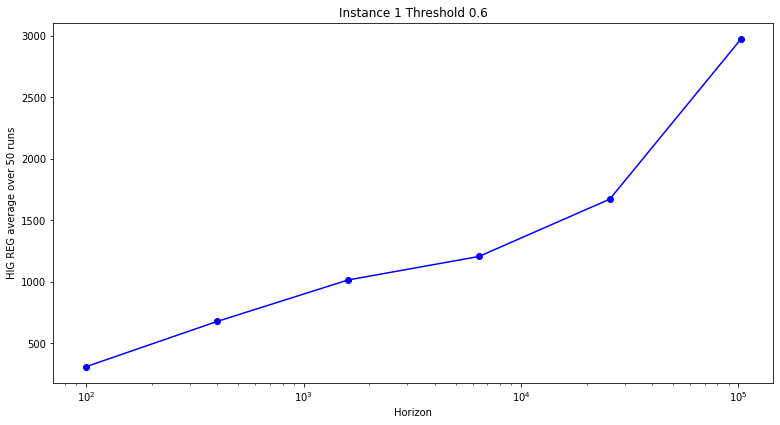
We will treat the process of pulling each arm from a given instance as a Bernoulli trial with reward 0 or 1, of which the probability of success is the sum of the probability of rewards greater than the threshold.

And now it is a normal multi-arm bandit problem with each pull having reward 0 or 1 of which we already know how to calculate the regret that is :

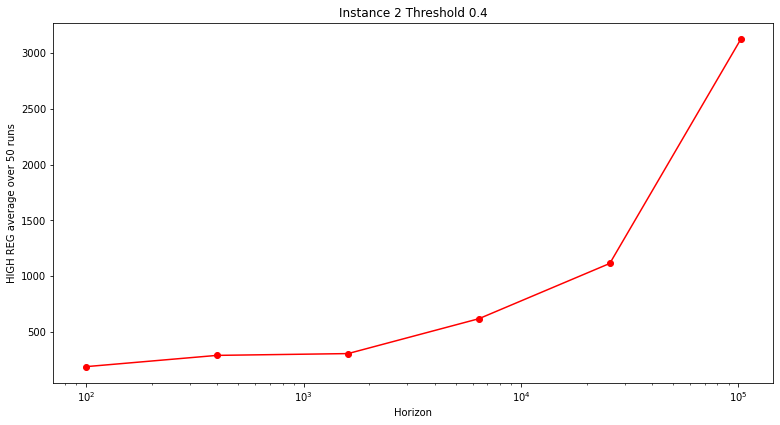
max(prob of success)\* horizon - Total reward

Instance 1 Threshold 0.4

Instance 1 Threshold 0.6



Instance 2 Threshold 0.4



Instance 2 Threshold 0.6

