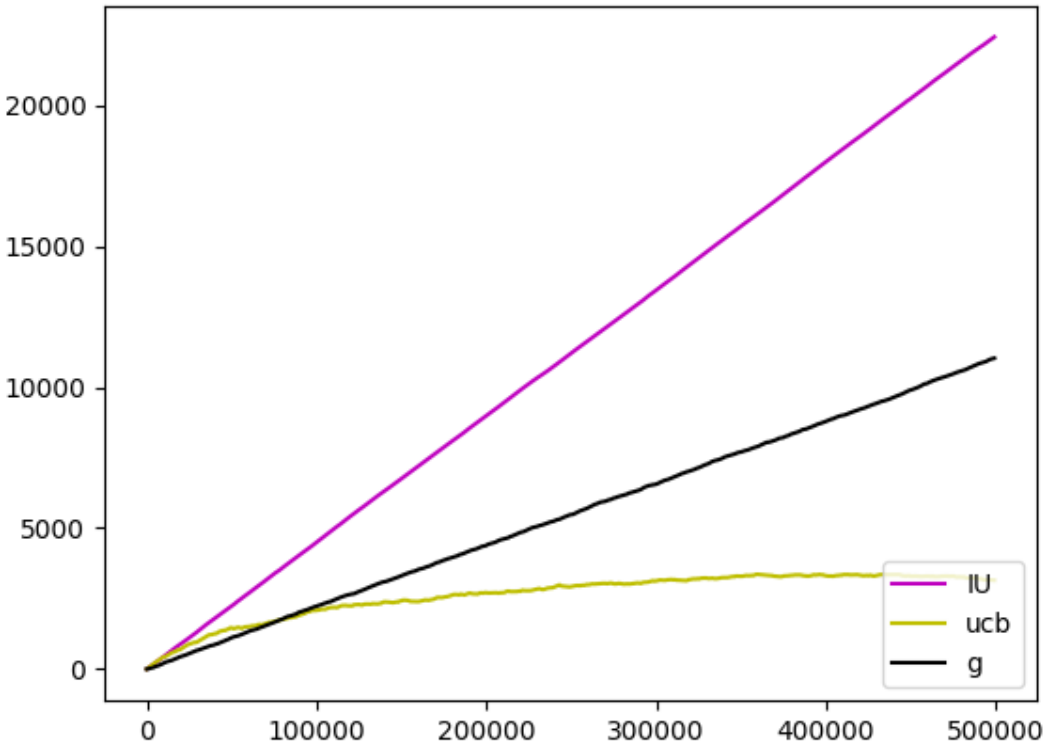
Assignment #5

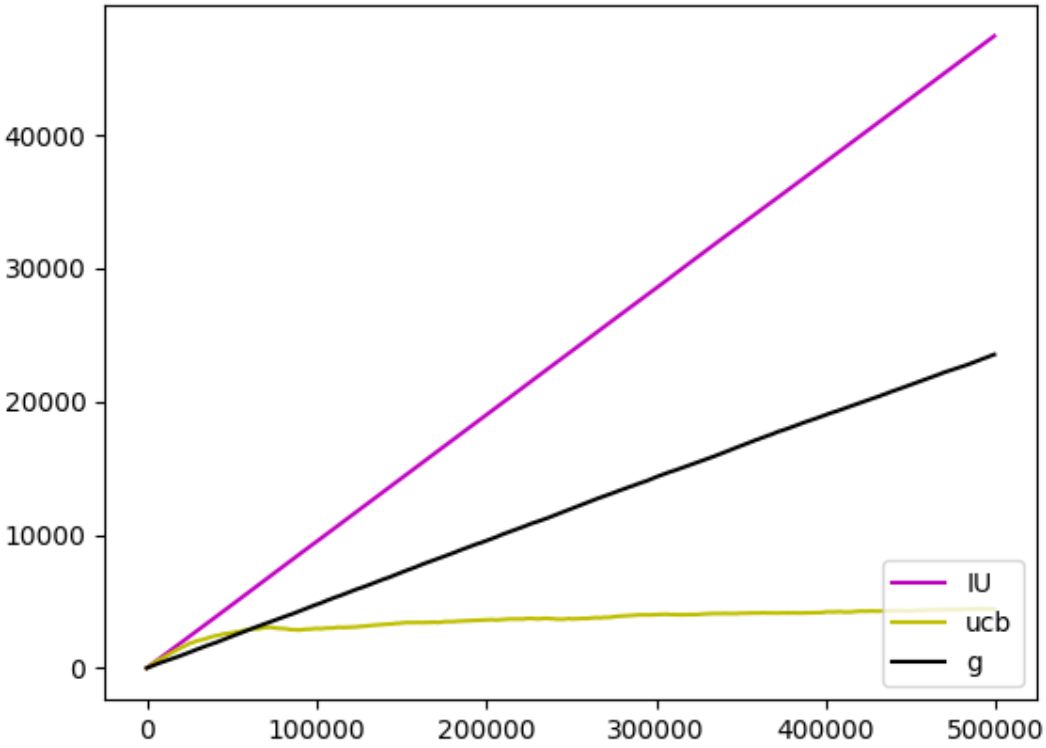
Sudhanshu Pathak

Graphs plotted for N = 5000000
For SimpleRegret, N = 1000, trials = 100

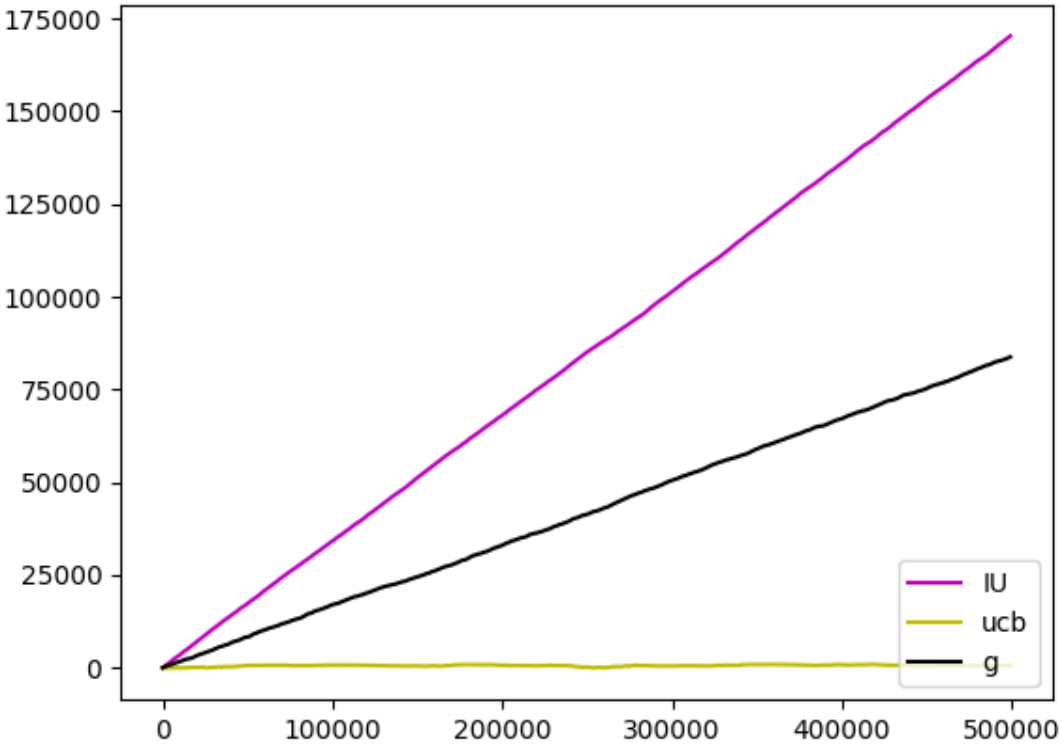
BanditA CumulativeRegret:



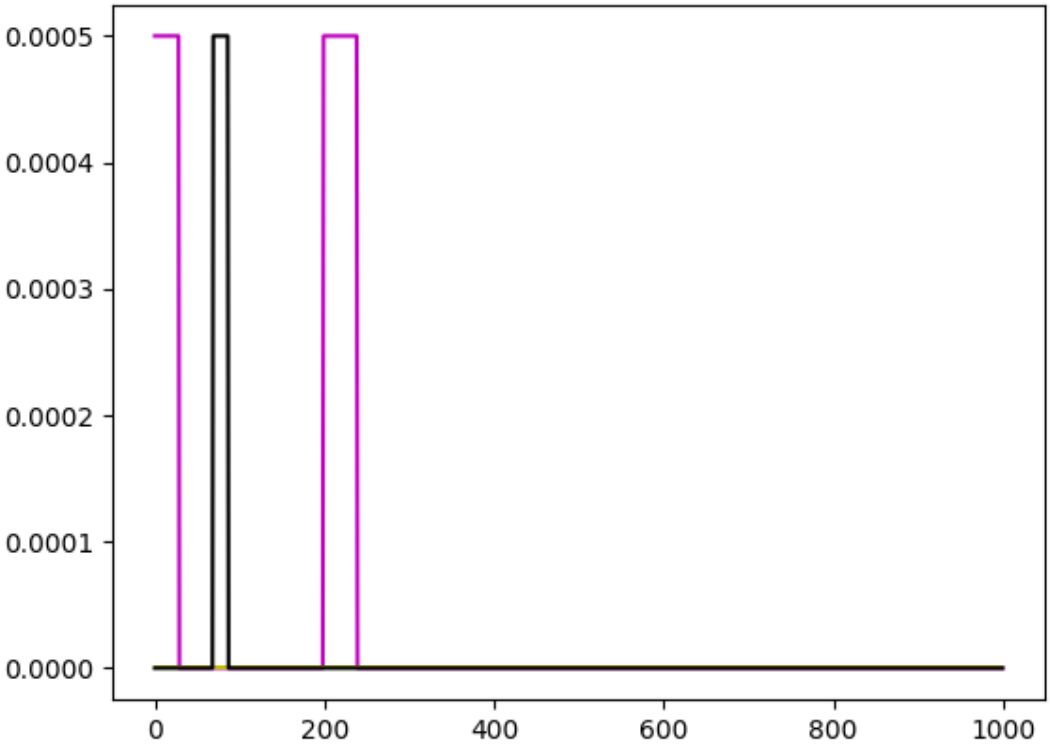
BanditB CumulativeRegret:



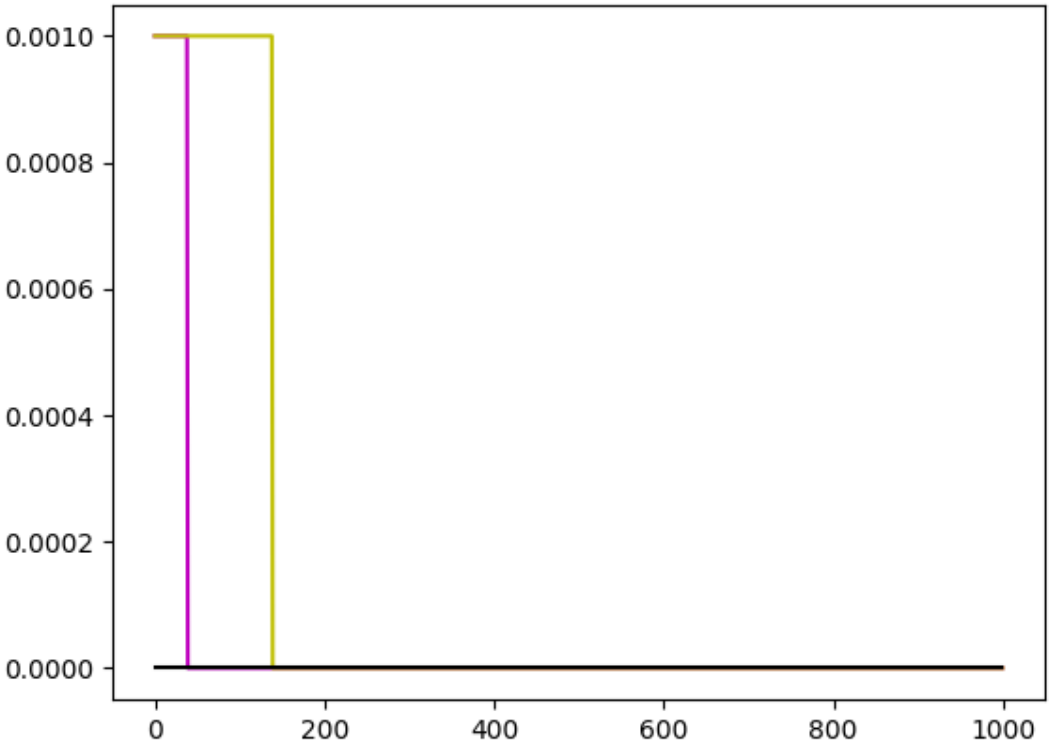
BanditC CumulativeRegret:



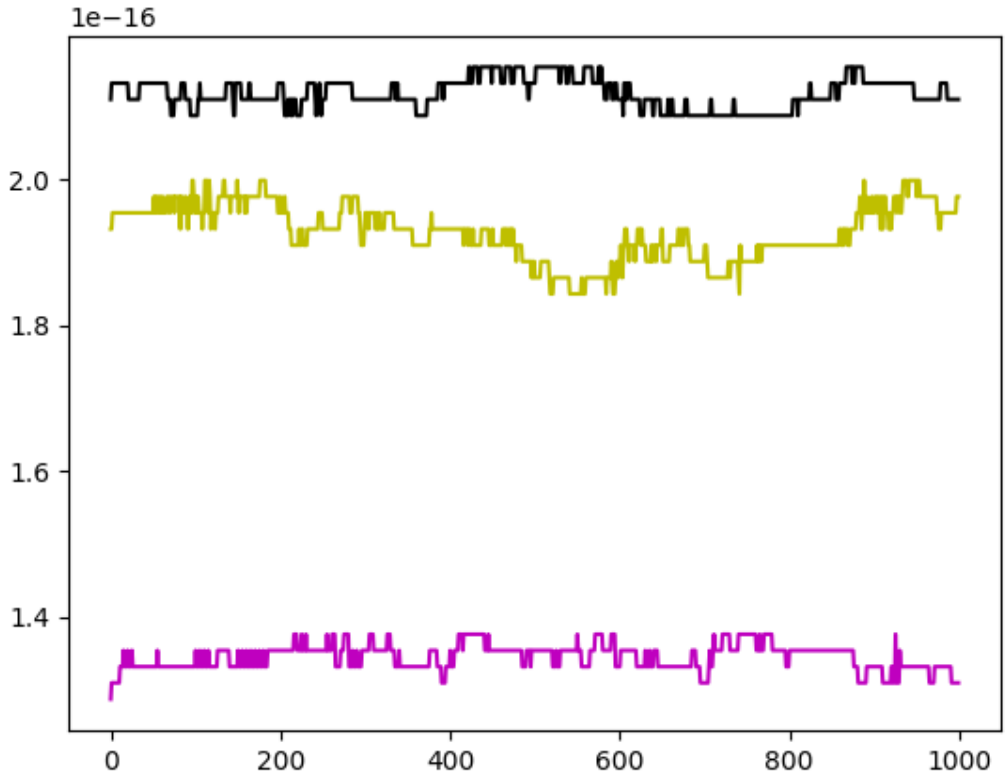
Bandit A Simple Regret:



Bandit B Simple Regret:



BanditC SimpleRegret:



Bandit3 Definition:
Consider bandit with six arms having following probability and reward distribution.

(Reward, Probability) = (6,0.2), (4,0.3), (2,0.6), (1,0.7, (0,0.8)

Observation:

1) Normally UCB algorithm converges faster than other two algorithms in the following order: This can also be seen from the graph.

   UCB
   Epsilon Greedy
   Increment Uniform

2) For simple regret for larger iterations regret is almost close to zero.