CS 747 - Programming Assignment 1

B.Nikhil 170050099

September 25, 2020

T1: Implementing the sampling algorithms - Assumptions

For `UCB` and `KL-UCB`, I assumed algorithm starts in a round robin fashion until each arm is pulled once.

T2: Thompson Sampling with hint

We know that in `Thompson Sampling` we assosiate a beta distribution for each arm. `Thompson Sampling with hint` will check if means associated with those beta distributions is close within $var = 0.01$ of Highest mean. If such mean for a beta distribution is found then pull that arm else pull the arm suggested by the `Thompson Sampling`

T3: Experiments on epsilon-greedy

for $\epsilon_1 = 0.0001, \epsilon_2 = 0.02, \epsilon_3 = 0.5$ given three bandit instances satisfy the given condition.(refer below figures)

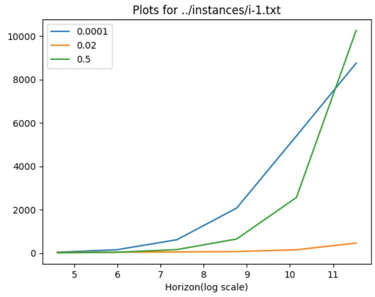

Figure 1: T3 for bandit instance-1.

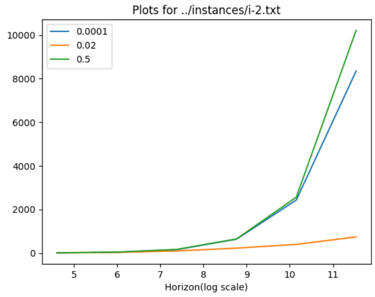

Figure 2: T3 for bandit instance-2.

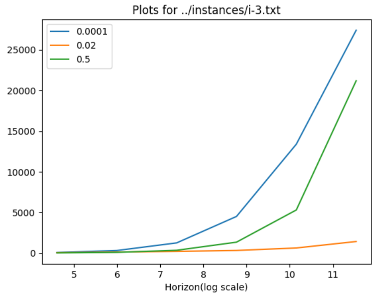

Figure 3: T3 for bandit instance-3.

T4: Output Data Interpretation and Plots

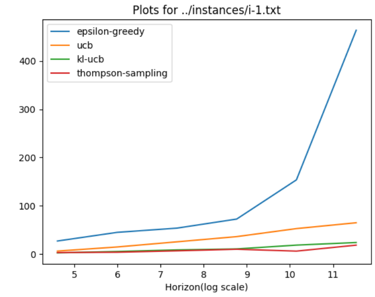

Figure 4: T1 for bandit instance-1.

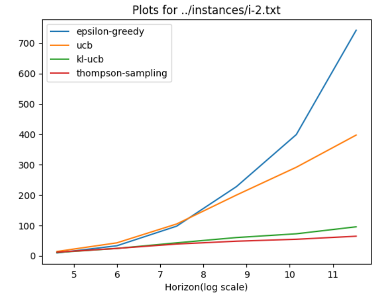

Figure 5: T1 for bandit instance-2

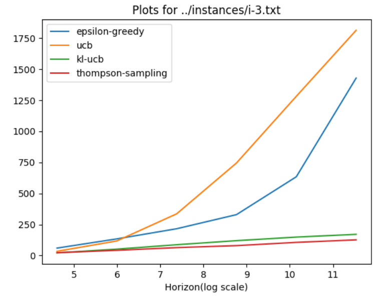

Figure 6: T1 for bandit instance-3

All the plots for T1 are as expected except `ucb` vs `epsilon-greedy` for bandit instance-3, as we might expect `epsilon-greedy` should have higher regret compared to `ucb`, to check if its correct, I ran `ucb` and `epsilon-greedy` for 204800 horizon(refer figure 7) and it proves `ucb` is suboptimal where as `epsilon-greedy` isn't.
Figures 8, 9 ,10 shows comparison between `Thompson Sampling` and `Thompson Sampling with hint` using the algorithm mentioned in section T2.

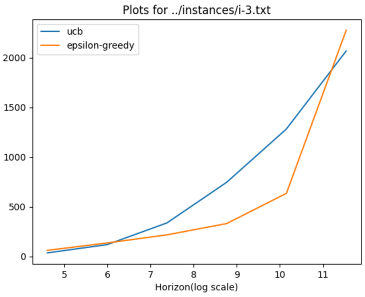

Figure 7: UCB vs epsilon-greedy for longer horizon of bandit instance-3

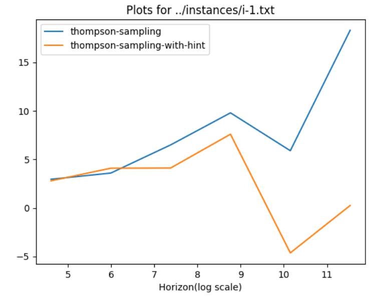

Figure 8: T2 for bandit instance-1

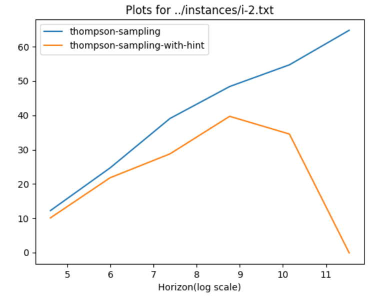

Figure 9: T2 for bandit instance-2

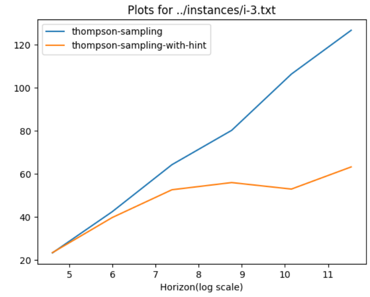

Figure 10: T2 for bandit instance-3