CS747-Assignment 1

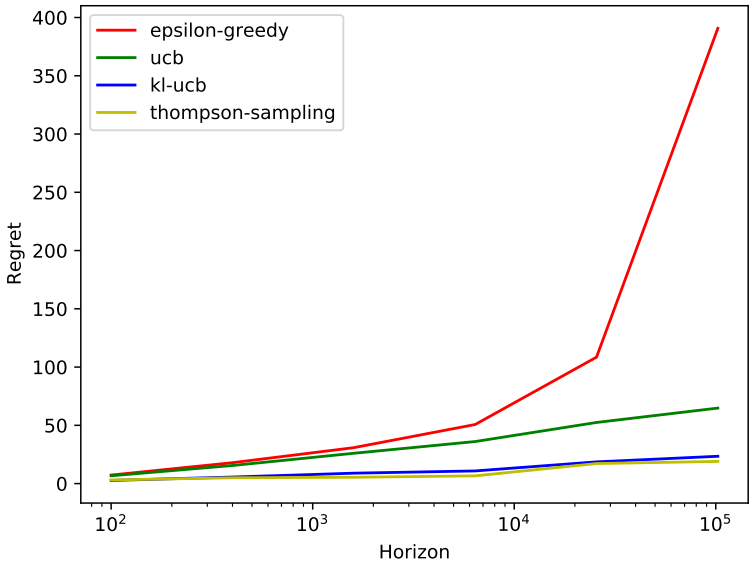
Yash Khemchandani - 170050025

September 24, 2020

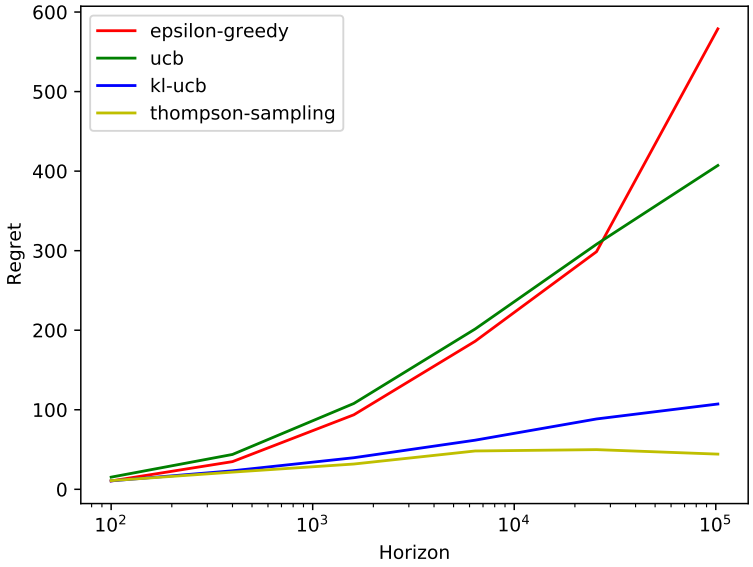
1 Implementation Details

- **Epsilon Greedy**: For the first N pulls, where N is the number of arms in the bandit, each arm is pulled once so that the empirical mean for all the arms is well defined in the future. If the algorithm decides to exploit (with probability 1-ϵ) and more than one arm has the maximum empirical mean, the first arm in the sequence is pulled (due to the nature of `numpy.argmax`)
- **UCB**: For the first N pulls, each arm is pulled once. If more than one arm has the maximum ucb_a^t, the first arm is pulled.
- **KL-UCB**: For the first N pulls, each arm is pulled once. It was empirically found that taking the value of c to be 0 instead of 3 in the term "$\ln(t) + c\ln(\ln(t))$" results in less regret, which is why the experiments have been performed using $c = 0$. The precision upto which binary search is performed to get the ucb-kl_a^t is taken to be 1e-3. If more than one arm has the maximum ucb-kl_a^t, the first arm is pulled.
- **Thompson Sampling**: No round-robin pulling for the first N pulls is performed. If there are more than one maximum samples, the first arm with the maximum sample is pulled.
- **Thompson Sampling with hint**: No round-robin pulling for the first N pulls is performed. The algorithm implemented is as follows:
  1. Given the sorted array of true means of the arms, the Beta PDF for each of the true mean is calculated for each arm according to the parameters $(s_a^t + 1, f_a^t + 1)$ where s_a^t and f_a^t are the number of successes/1's and failures/0's for arm a at time t respectively.
  2. For each arm a the probabilities of the true means are normalized so that their sum is equal to 1.
  3. Let X be the maximum of the true means. At any given time, the arm with the maximum probability for X is pulled. In case of any ties, the first arm is pulled.

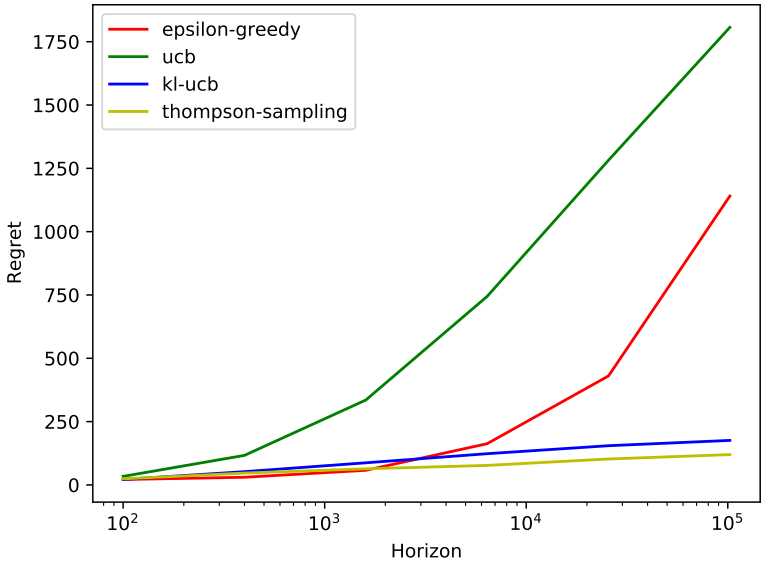
2 Plots for T1



(a) Instance 1



(b) Instance 2



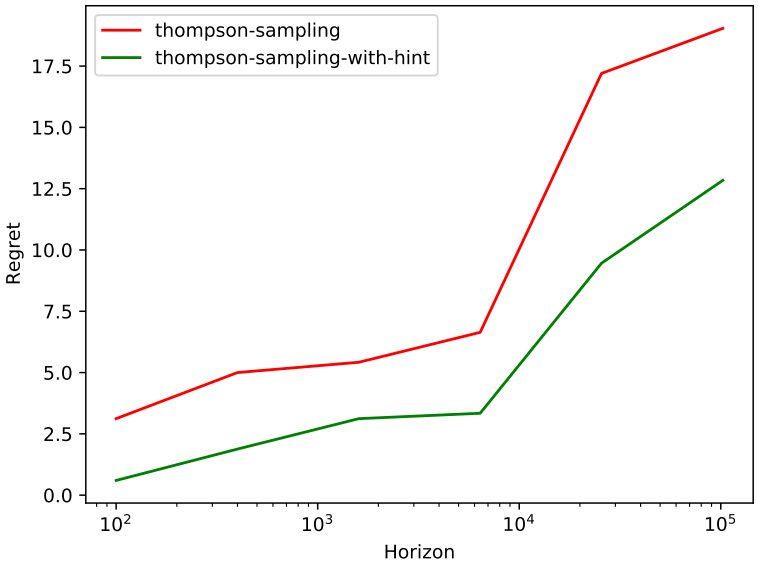
(c) Instance 3

Figure 1: Regret vs Horizon plots for T1

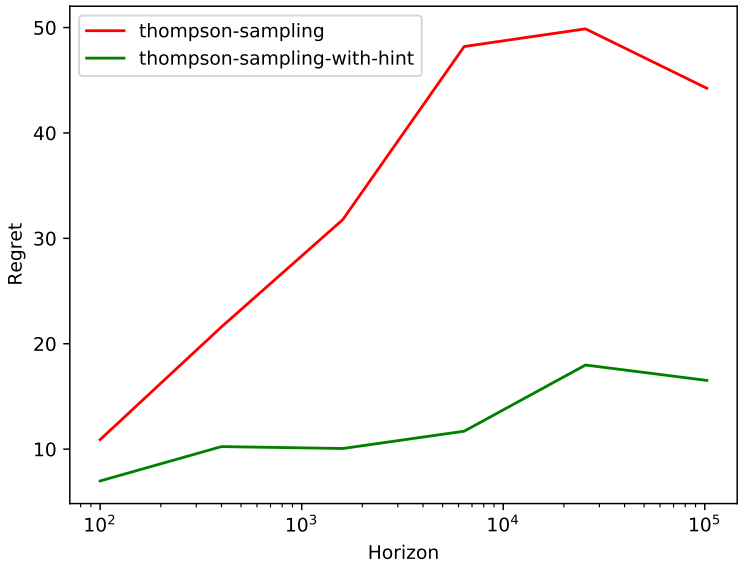
Interpretations:

- For all the instances, Thompson Sampling and KL-UCB algorithm achieve lowest regrets, which is justified since these algorithms have been proved to achieve the optimal regret.
- Moreover, we observe that Thompson Sampling algorithm performs better than KL-UCB algorithm which solidifies the argument that this algorithm is excellent in practice.
- UCB performs worse than KL-UCB since it doesn't achieve the Lai and Robbin's lower bound.
- For instance 3, Epsilon-Greedy seems to perform better than UCB. This might seem like a contradiction but theoretically it is quite possible since UCB achieves sub-linear regret asymptotically and 102400 may not be a large horizon. If we continue the experiment for more timesteps, UCB will eventually perform better than Epsilon-Greedy (which is also evident from the greater slope of Epsilon-Greedy plot at timestep 102400)

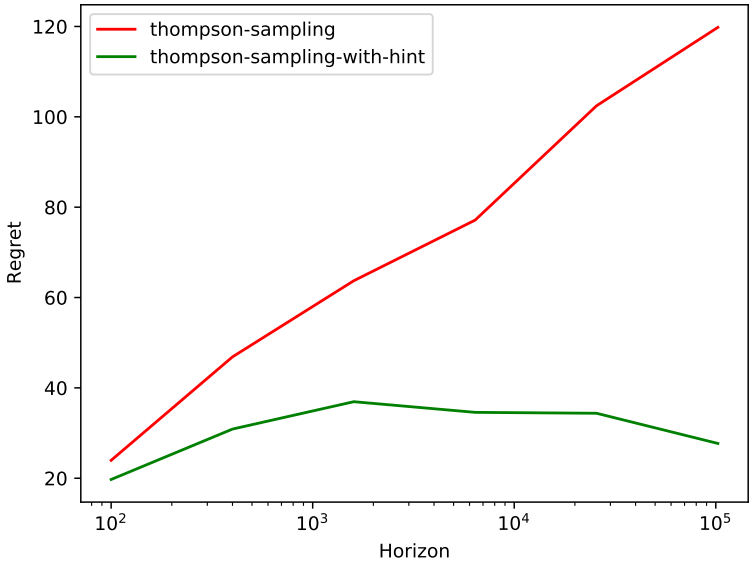
3 Plots for T2



(a) Instance 1



(b) Instance 2



(c) Instance 3

Figure 2: Regret vs Horizon plots for T2

Interpretations:
The Thompson-Sampling-with-Hint algorithm consistently performs better than the Thompson Sampling algorithm which is expected because of the extra knowledge of the true means used by the algorithm.

4 ϵ values for T3

By empirical analysis, the ϵ values obtained for epsilon-greedy algorithm are :

- ϵ_1 : 0.0002
- ϵ_2 : 0.02
- ϵ_3 : 0.9

The results for experiments are described below:

ϵ	Regret(Instance 1)	Regret(Instance 2)	Regret(Instance 3)
0.0002	730.5	4427.98	1968.88
0.02	390.72	578.88	1140.7
0.9	18447.96	18408.88	38090.48