FILA: Assignment-1 Report

Ashish Aggarwal
Roll no. : 203050015

Thursday 24th September, 2020

1 T1 & T2: Coding

1.1 epsilon-greedy

1.1.1 Implementation

I have implemented epsilon-greedy 3 Algorithm. I have pulled all the arms once to initialize the empirical mean values and ties are broken uniformly.

1.1.2 Observations

- It is performing Linearly as expected in all instances.
- In instance-3 though the regret is less than the UCB even for horizon 102400 but if you will see graph slope for epsilon will cross UCB soon. To verify this I checked for horizon value 200000 and compared the average over 50 seeds:
  - $Horizon : 200000, epsilon - greedy, epsilon : 0.02, Regret(over 0 - 49 seed) = 2237.98$
  - $Horizon : 200000, UCB Regret(over 0 - 49 seed) = 2077.14$

1.2 UCB

1.2.1 Implementation

Each arm is pulled once initially to set values of count pull and empirical mean of each arm. Pull the arm with highest UCB values and ties are broken uniformly if more than one arm has maximum UCB value.

$$UCB_a^t = p_a^t + \sqrt{2\ln(t)/u_a^t}$$

1.2.2 Observations

Performing sub-linearly w.r.t Horizon better than Epsilon greedy except in instance-3 where it is taking more time to overtake epsilon-greedy.Please see comparison given in epsilon-greedy section.

1.3 KL-UCB

1.3.1 Implementation

- Each arm is pulled once initially to set values of count pull and empirical mean of each arm.
- Pull the arm with highest Q values which satisfies following equation and ties are broken uniformly if more than one arm has maximum Q value. Precision used for Q values: 1e-5:
$$u_a^t * KL(p_a^t, Q) \leq ln(t) + 3ln(ln(t))$$
- Used binary search to get Max value of Q value which satisfy the equation and precision used for zero and 1 value passed to KL divergence is 1e-10 i.e. if number has more than 10 zeros after decimal it will be considered zero likewise if number has more than 10 nine after decimal it will be considered one.
- I have used the vectorized implementation of binary search to reduce running time of KL-UCB. It is running in <2min for horizon of 102400

1.3.2 Observations

Performing sub-linearly w.r.t Horizon better than Epsilon-greedy and UCB in all instances.

1.4 Thompson-Sampling

1.4.1 Implementation

Pulled each arm based on max values we get from Beta distribution at each time-step. Starting parameter to beta distribution are B(1,1) i.e. when no failure and success is recorded, at that time all arm get a sample from uniform distribution.

1.4.2 Observations

By far the Best performing Algorithm in all instances and regret is sub-linear.

Plots T1:

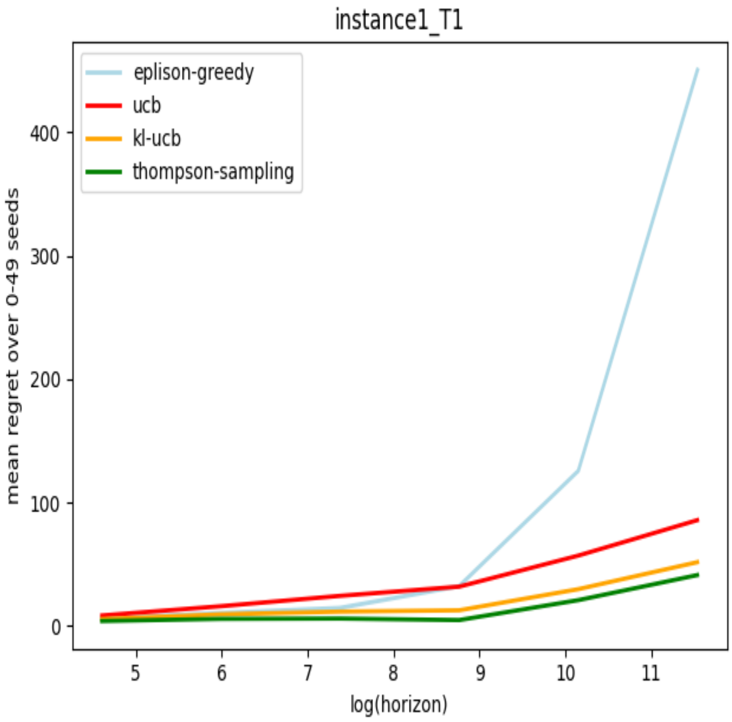

Figure 1: Instance-1 comparison for T1 experiment

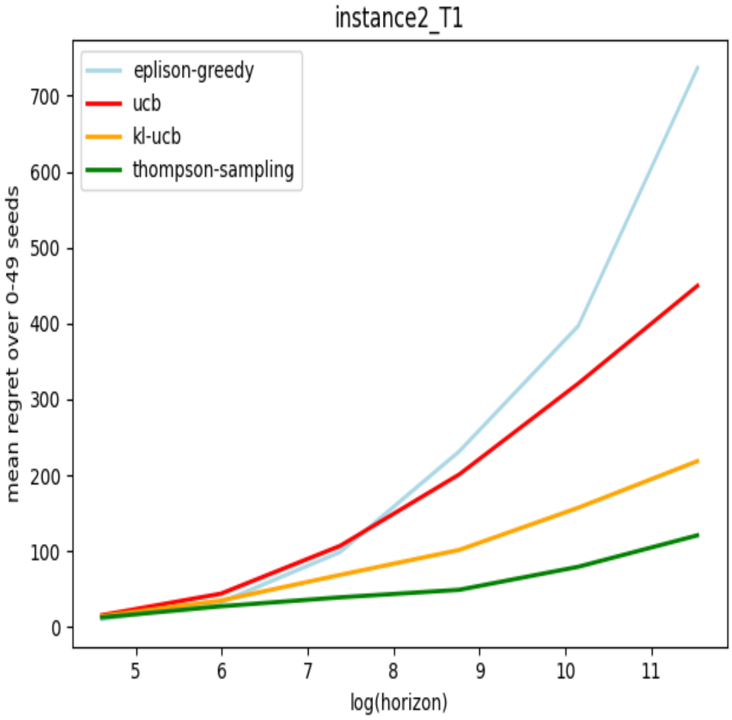

Figure 2: Instance-2 comparison for T1 experiment

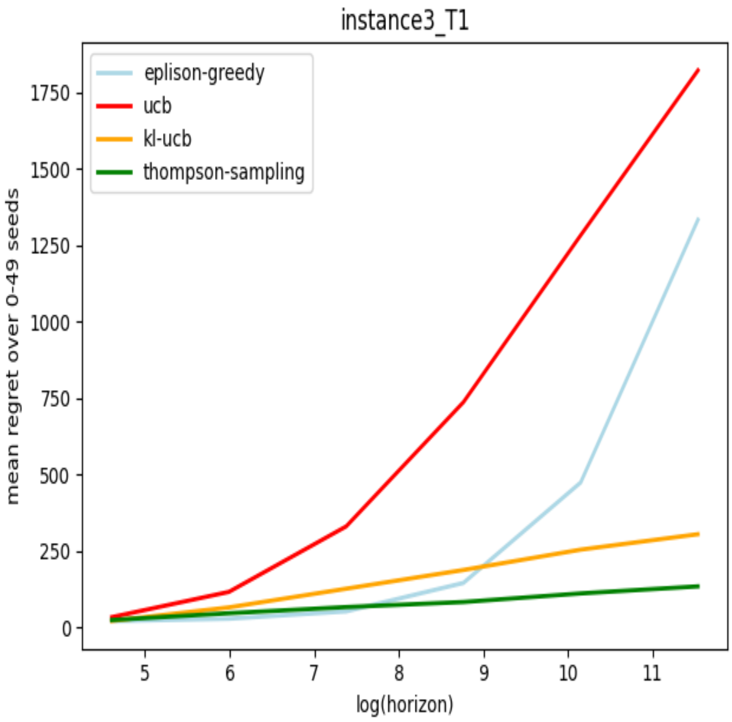

Figure 3: Instance-3 comparison for T1 experiment

1.5 Thompson-Sampling-with-hint

1.5.1 Implementation

- I tried calculating α_0 and β_0 parameter for beta distribution from the mean and variance of hint array and use that as prior belief while sampling from Beta Distribution i.e. sample from $B(\alpha_0 + success[arm], \beta_0 + failure[arm])$ but performance did not improved much. So dropped this idea.
- So Now, the idea behind is as we know the maximum true mean I'll try to find the arm which has empirical mean around this true mean and is better than others. Also, when Beta samples are much bigger than the true means I will sample that arm even if mean is not near true mean. It will help in initial convergence of beta distribution.
- This idea help eradicate pulling non-optimal arm which give better beta sample values sometimes than optimal arm.
- I have maintained the current empirical mean array of beta distribution calculated for each arm using $\frac{\alpha}{\alpha+\beta}$; where $\alpha = success[arm] + 1 \& \beta = failure[arm] + 1$ till that time-step.
- **Then took a sample from Beta distribution as we would do in thompson-sampling and checked if that sample is near the maximum true mean given in the hint and also the arm which gave this max sample is also has the highest empirical mean in the beta mean array maintained above. Then pull that arm.**
- If above condition does not match pull the arm with highest sample as we would do in thompson-sampling.

1.5.2 Observations

This improvement with hint has performed consistently well in comparison to thompson sampling.

Plots T2:

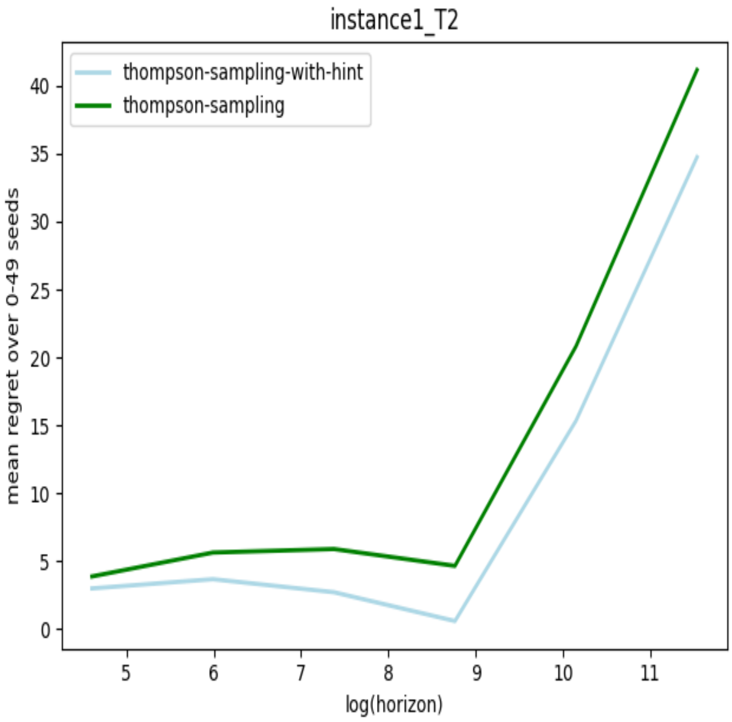

Figure 4: Instance-1 comparison for T2 experiment

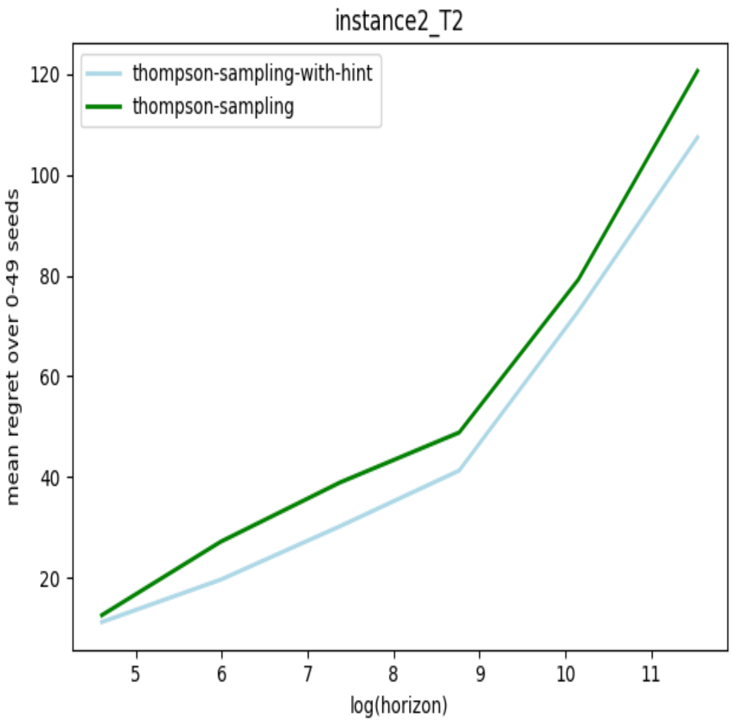

Figure 5: Instance-2 comparison for T2 experiment

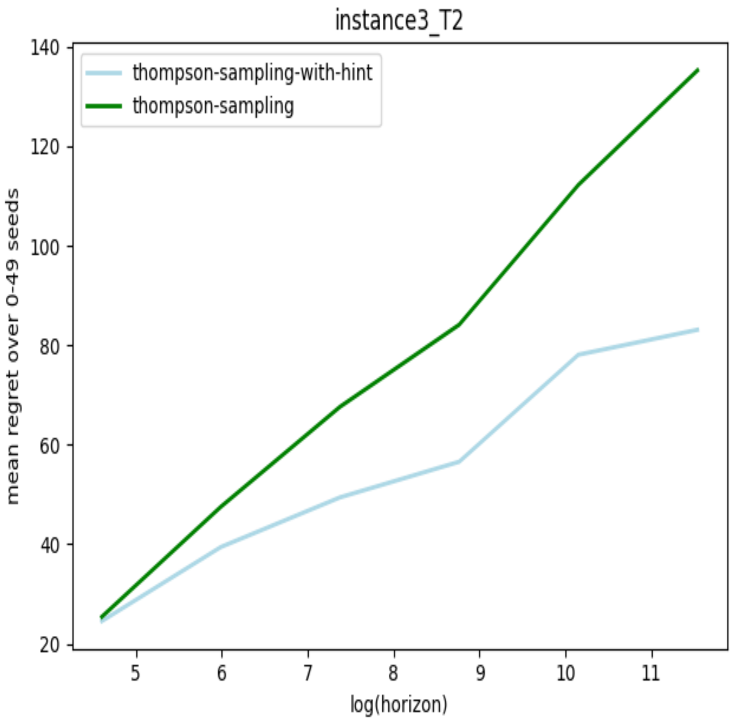

Figure 6: Instance-3 comparison for T2 experiment

2 T3: experiment with epsilon values

Following are the epsilon values obtained such that regret for $\epsilon 2$ is minimum compared to $\epsilon 1$ & $\epsilon 3$

	$\epsilon 1$	$\epsilon 2$	$\epsilon 3$
Instance1:	.001	.005	.01
Instance2:	.005	.01	.05
Instance3:	.001	.01	.05

3 T4: Output Data, Plots, Report

- All plots are given in the PDF already in respective sections.
- Please find the outputDataT1.txt and outputDataT2.txt in the submission folder.