Assignment 3

Kartik Bharadwaj CS20S020

June 6, 2021

1 Question 1

Part 1

Check method *setT0()* in **q1.py** file.

Part 2

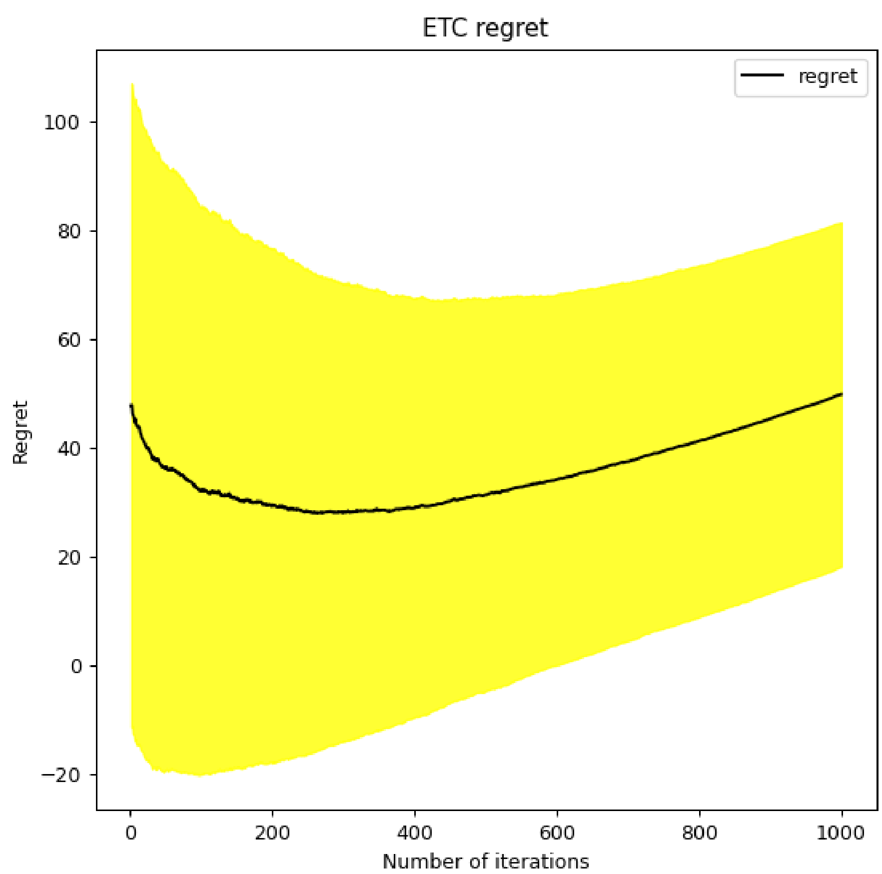

Figure 1: ETC Regret

Part 3

Our theoretical value $T_0 = 366$ for $T = 1000$. From the above graph, we can see that pseudo-regret of the algorithm decreases and reaches the minimum value at T_0, and increases from thereafter. From this observation, we conclude that if we explore too little or too much (i.e., ¿ or ¡ T_0), we accumulate a lot of regret. Also, the variance of the algorithm decreases as we get closer to T_0 and increases again as we move away from T_0. This reaffirms our conclusion. Moreover, to reduce variance one must increase T, and hence T_0.

Part 4

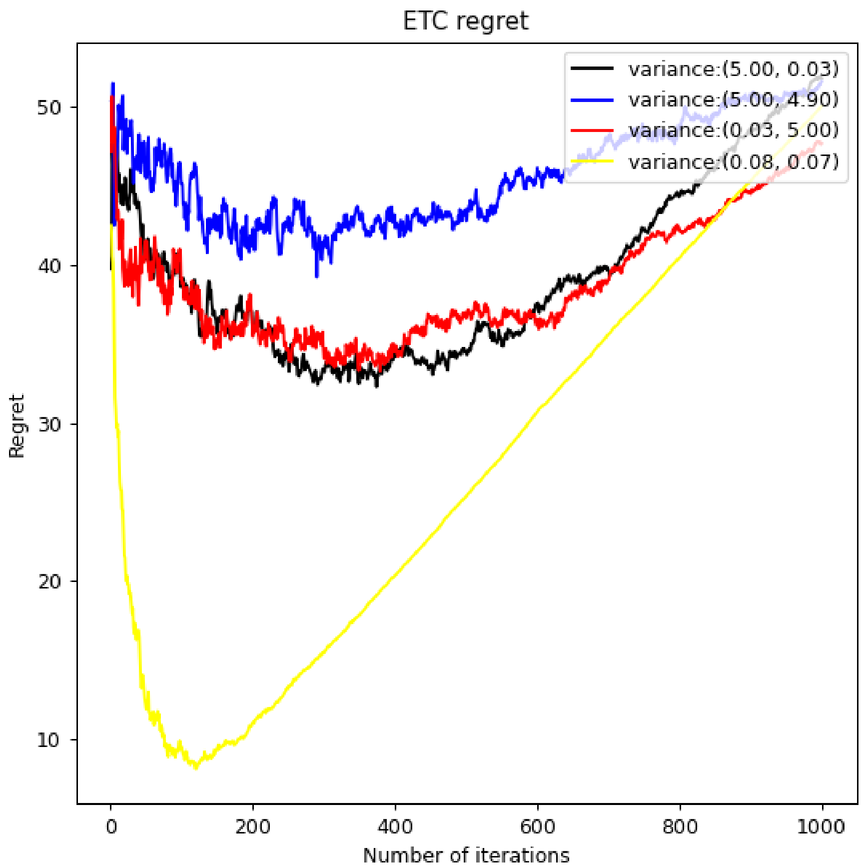

Figure 2: ETC Regret for different variances

From the above figure, we can say larger variance means larger expected squared difference. For smaller variances, like [0.08, 0.07], we reach minimal regret at point less than $T0$, while for larger variances it takes more iterations than $T0$. It is easy to see that the asymptotic behavior of an algorithm is a poor predictor of its finite time behavior. Again, this reaffirms our conclusion that for larger variances, we need to increase T, and hence, T_0 to reduce variance of regret.

2 Question 2

Part 1

For Thompson sampling, we use normal distribution as conjugate prior for the posterior distribution.

Part 2 & Part 3

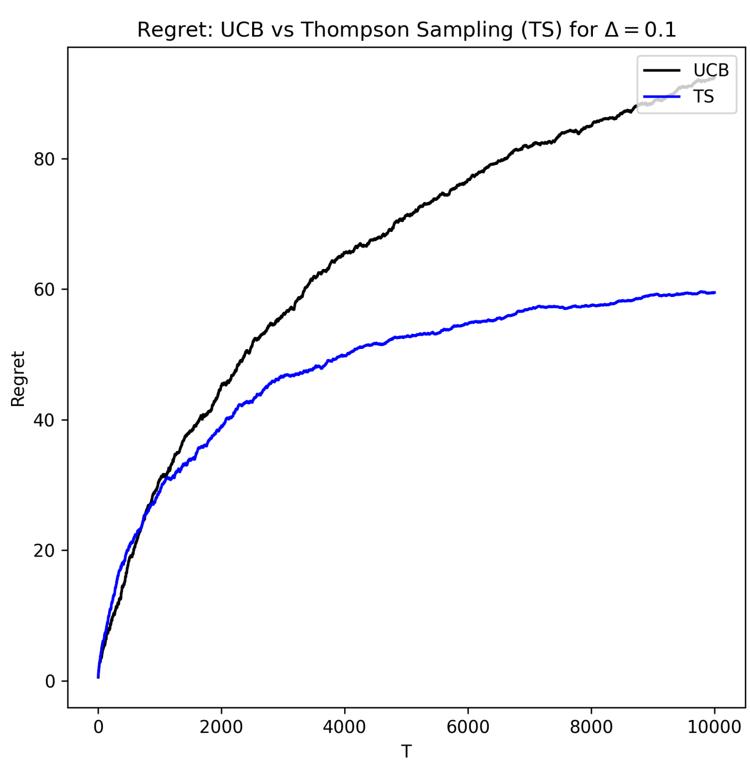

Figure 3: UCB vs Thompson Sampling

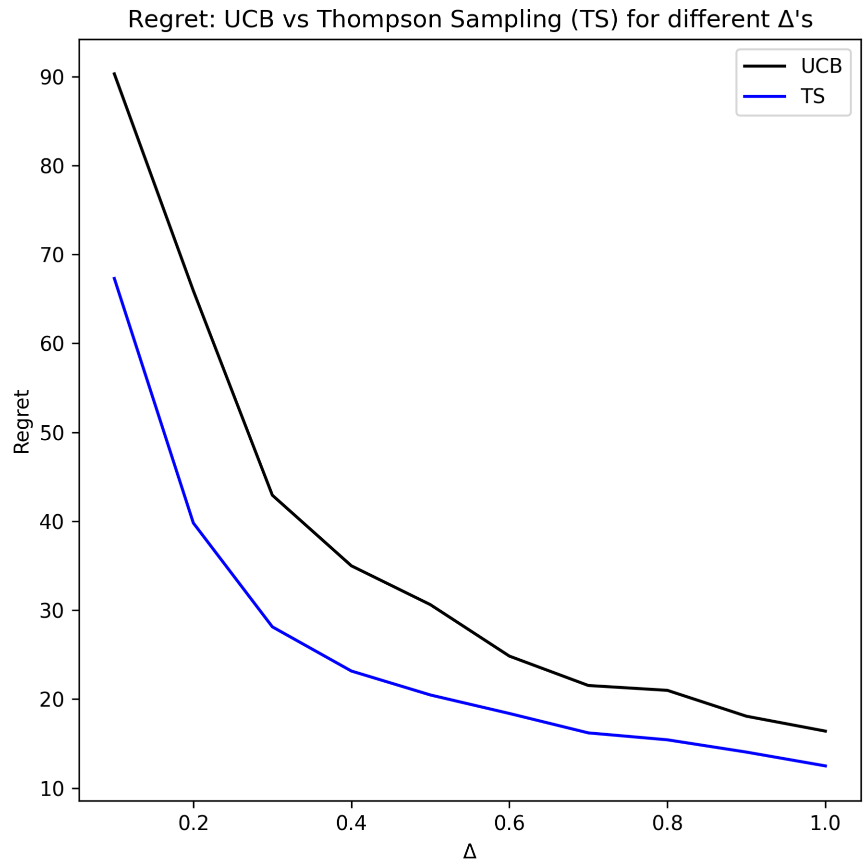

Figure 4: UCB vs Thompson Sampling for different Δs'

From the above figures, we can see that Thompson Sampling (TS) outperforms UCB. TS reduces the influence of delayed feedback by randomizing over actions, while UCB is deterministic and suffers a larger regret when picking a sub-optimal arm. Another reason is that UCB regret bounds depend on the specific choice of upper bound U_t used by the algorithm in question. With TS, U_t plays no role in the algorithm. This suggests that, while the regret of a UCB algorithm depends critically on the specific choice of upper-confidence bound, TS depends only on the best possible choice.