IE613:Online Learning
Assignment-3

Vinay Chourasiya (173190011)

April 23, 2018

Question 1

For ϵ-greedy Algorithm

- At round 1, start algorithm with probability vector uniform for all arms
- As we don't about the arm means at start, so I start mean vector (that contain all arms(K)) with zero (0) and according to the algorithm arm selection criteria Mean vector update further.

For UCB-V algorithm [1]

- According to research Paper [1] ,

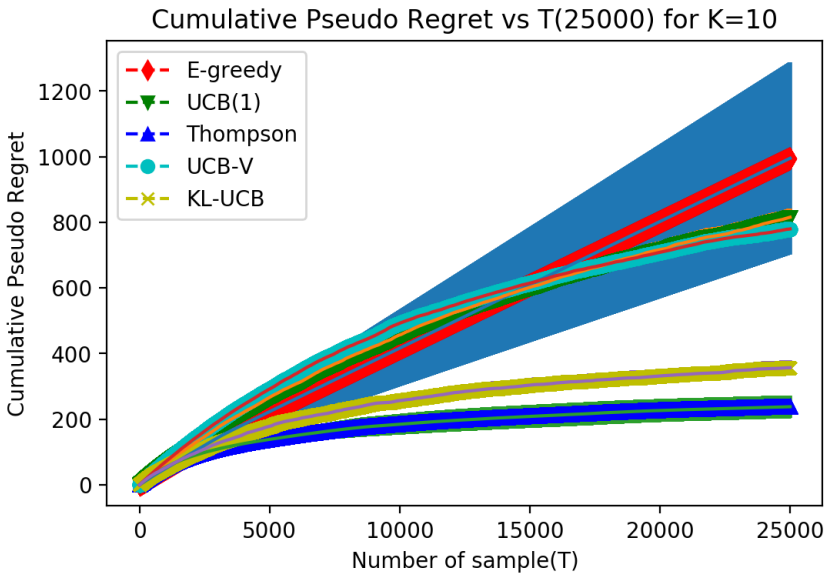
$$B_{i,N_{(i,t)}} = \hat{\mu}_{i,N_{(i,t)}} + \sqrt{\frac{2\zeta V_{i,N_{(i,t)}} \log t}{N_{(i,t)}}} + \frac{3\zeta \log t}{N_{(i,t)}}$$

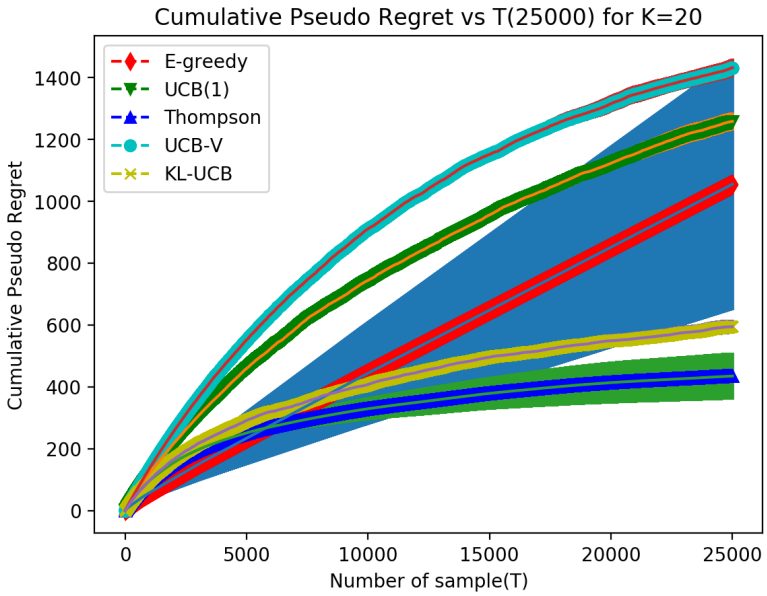
- where $V_{i,N_{(i,t)}}$ is Variance of arm i, and $\zeta = 1.2$
- In every round, $I_t = \arg\max_i(B_{i,N_{(i,t)}})$

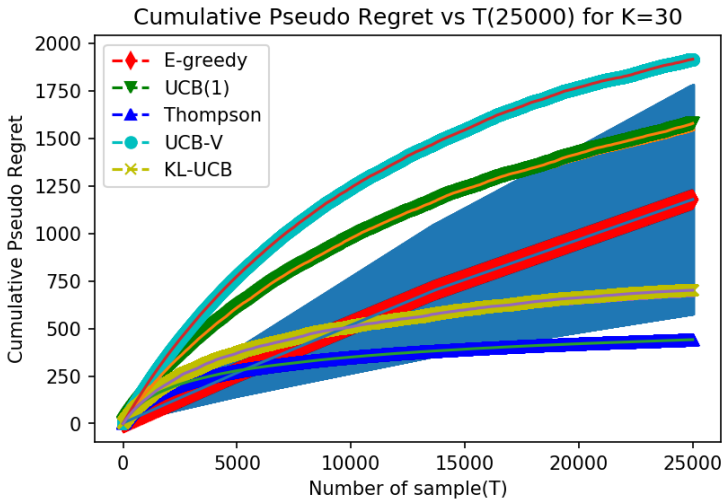
Programme Files

- For K=10 arms file $\rightarrow$ *Q1K10.py*
- For K=20 arms file $\rightarrow$ *Q1K20.py*
- For K=30 arms file $\rightarrow$ *Q1K30.py*

Plot for All algorithm (ϵ-greedy,UCB(α),UCB-V,KL-UCB,Thompson) as follow







Summary

1 UCB(α),UCB-V,KL-UCB,Thompson algorithm attain sub-linearity after some number of round , but ϵ-greedy algorithm gives linear regret for all Figure(1),Figure(2),Figure(2).

2 Thompson Algorithm perform better than other algorithm. it has smallest regret among all.

3 KL-UCB is also good in term of regret and sub-linearity (similar to Thompson)

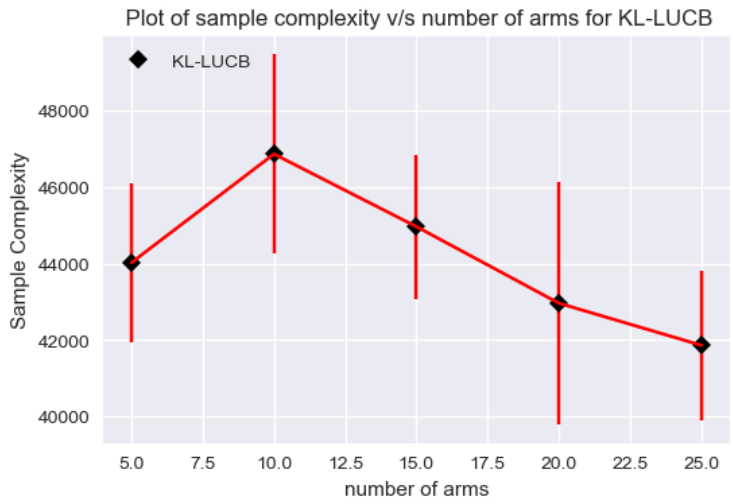
4 confidence interval for ϵ-greedy algorithm is wide, because of the randomization in choosing the arms,

5 As number of arms increase the regret value also increase for all algorithm.

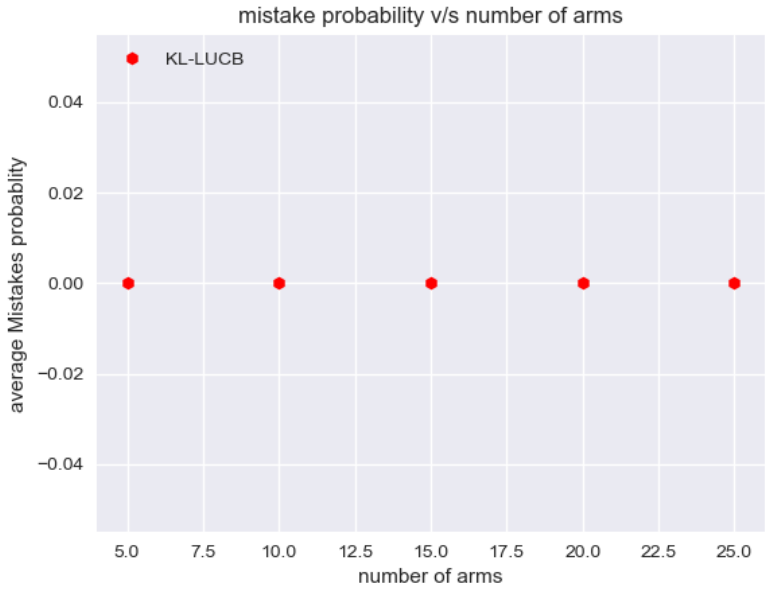
Question:2

KL-LUCB algorithm

- Program file for KL-LUCB algorithm → *Q2KLLUCB.py*
- Plot of average sample complexity v/s number of arms

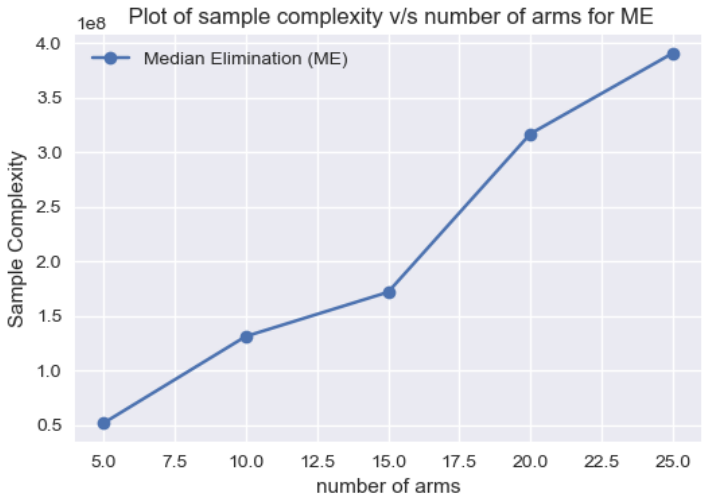


- plot of average mistake probability v/s number of arms

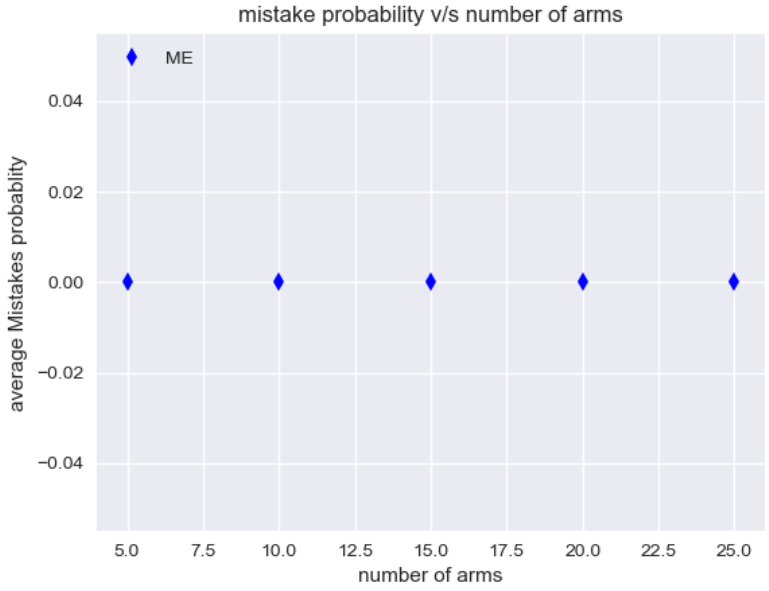


Median Elimination algorithm

- Program file for Median Elimination algorithm → *Q2ME.py*
- Plot of average sample complexity v/s number of arms



- plot of average mistake probability v/s number of arms



References

[1] Audibert, Jean-Yves, Rémi Munos, and Csaba Szepesvári. *Exploration–exploitation trade-off using variance estimates in multi-armed bandits.* Theoretical Computer Science 410.19 (2009): 1876-1902.