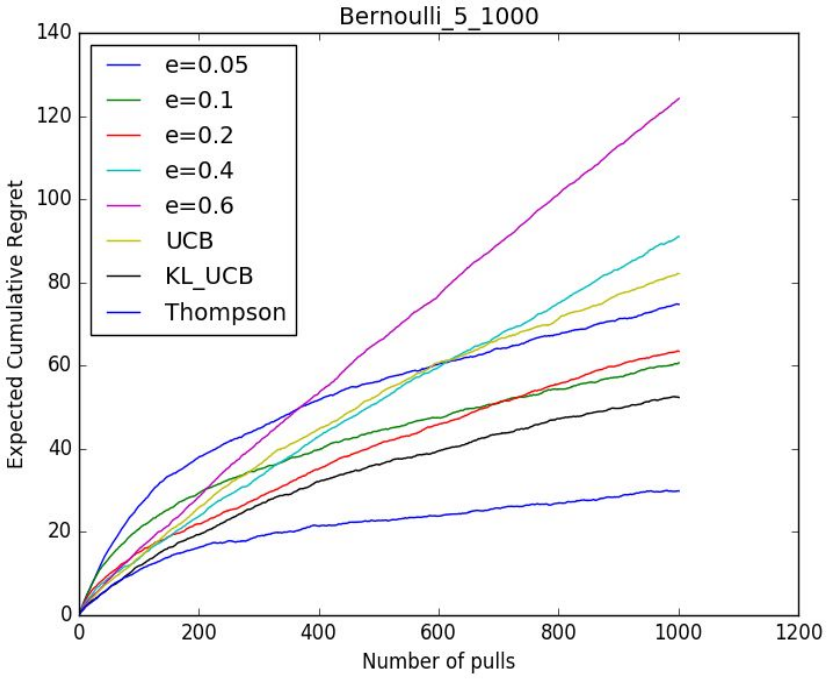
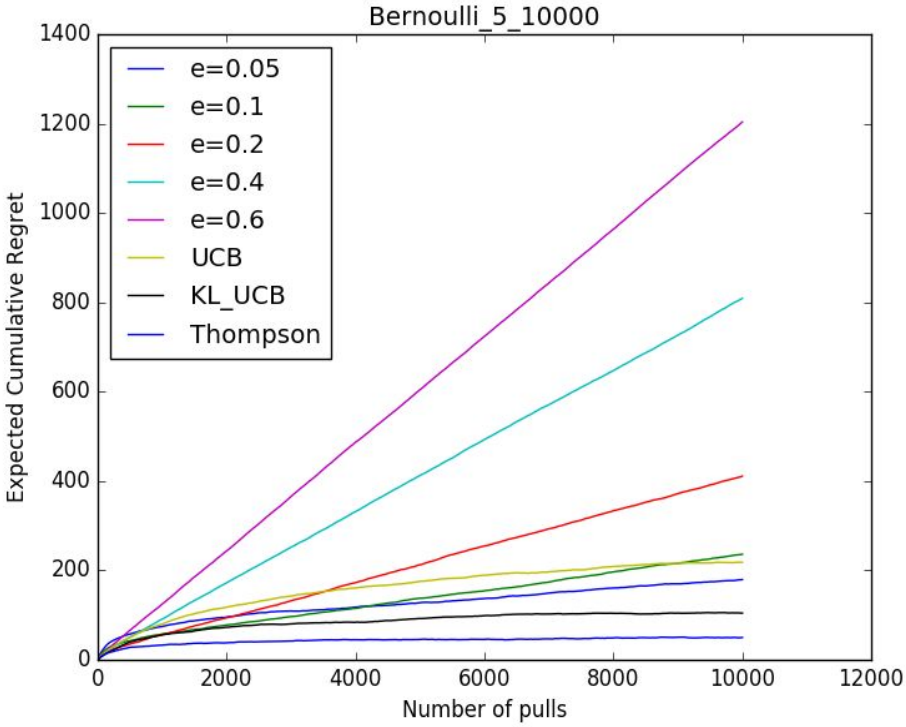
CS 747 Assignment 1

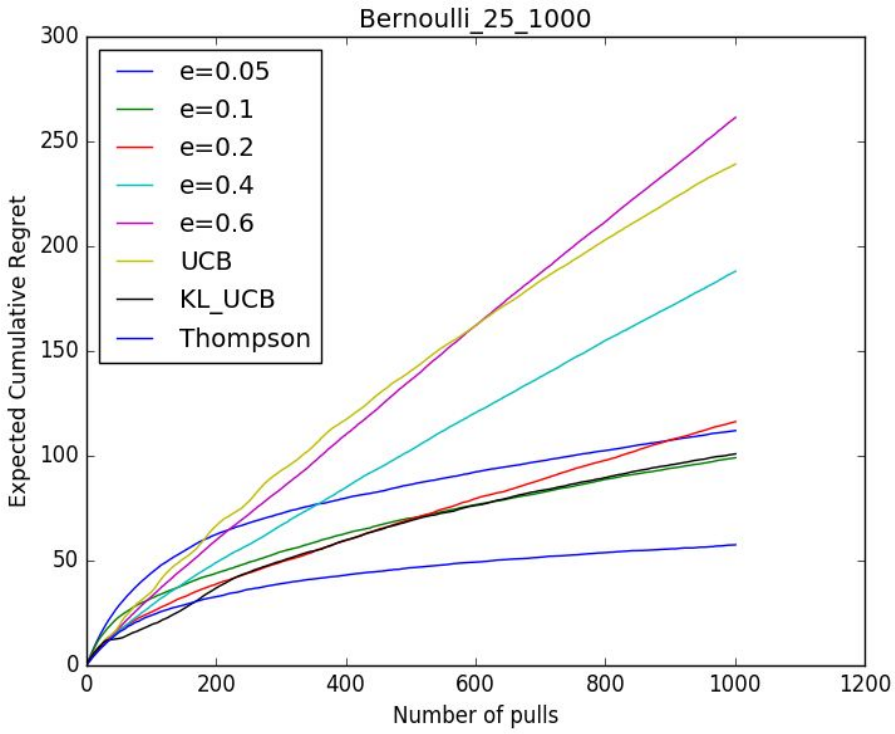
Bhavya Bahl

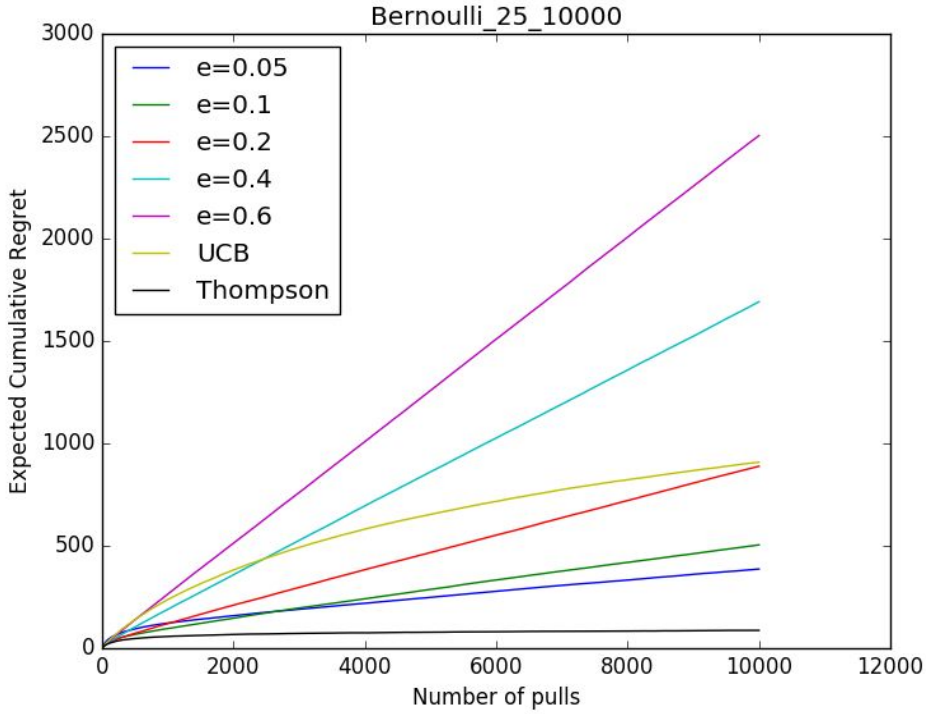
150050110

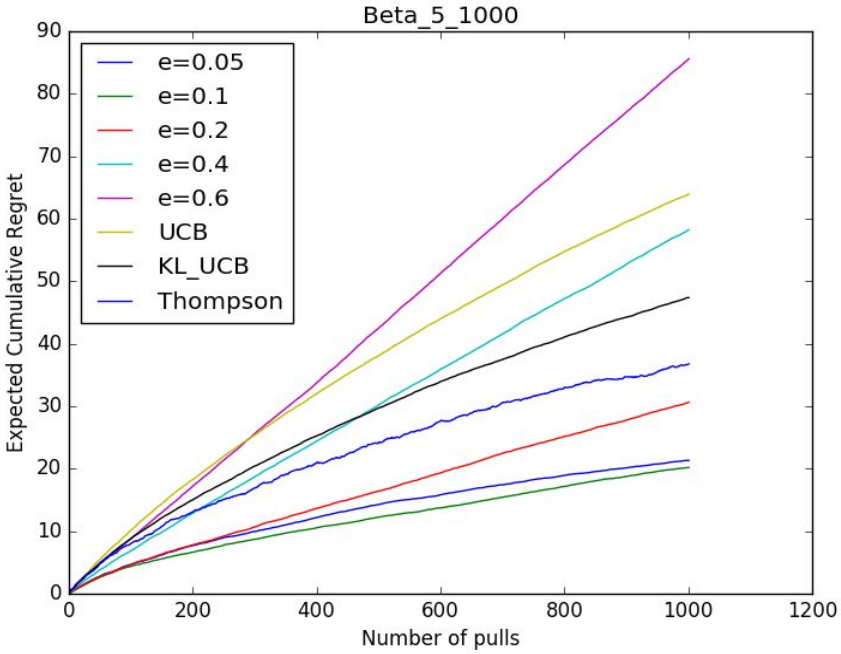
OBSERVATIONS

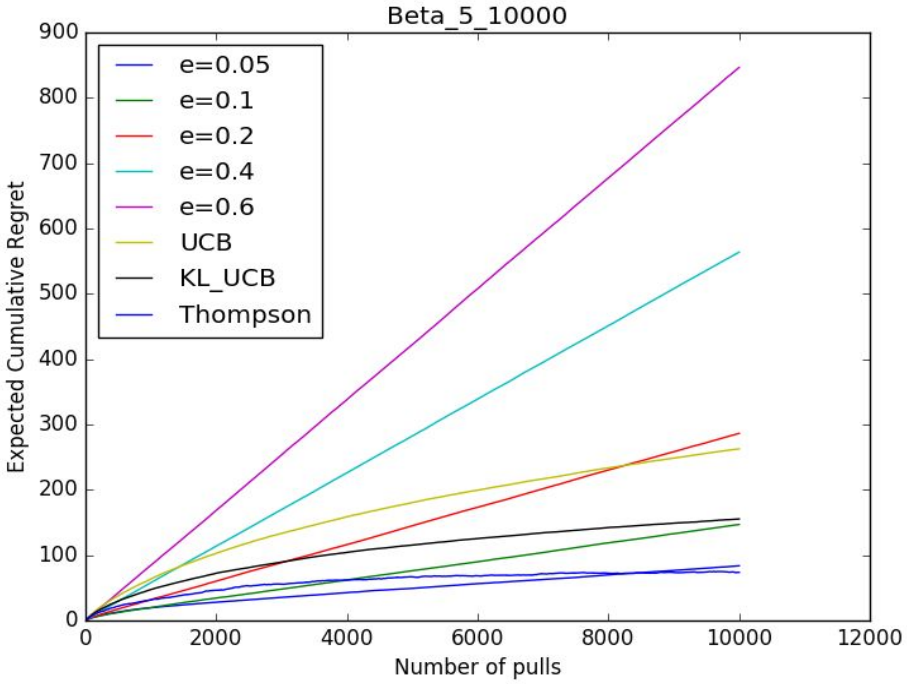


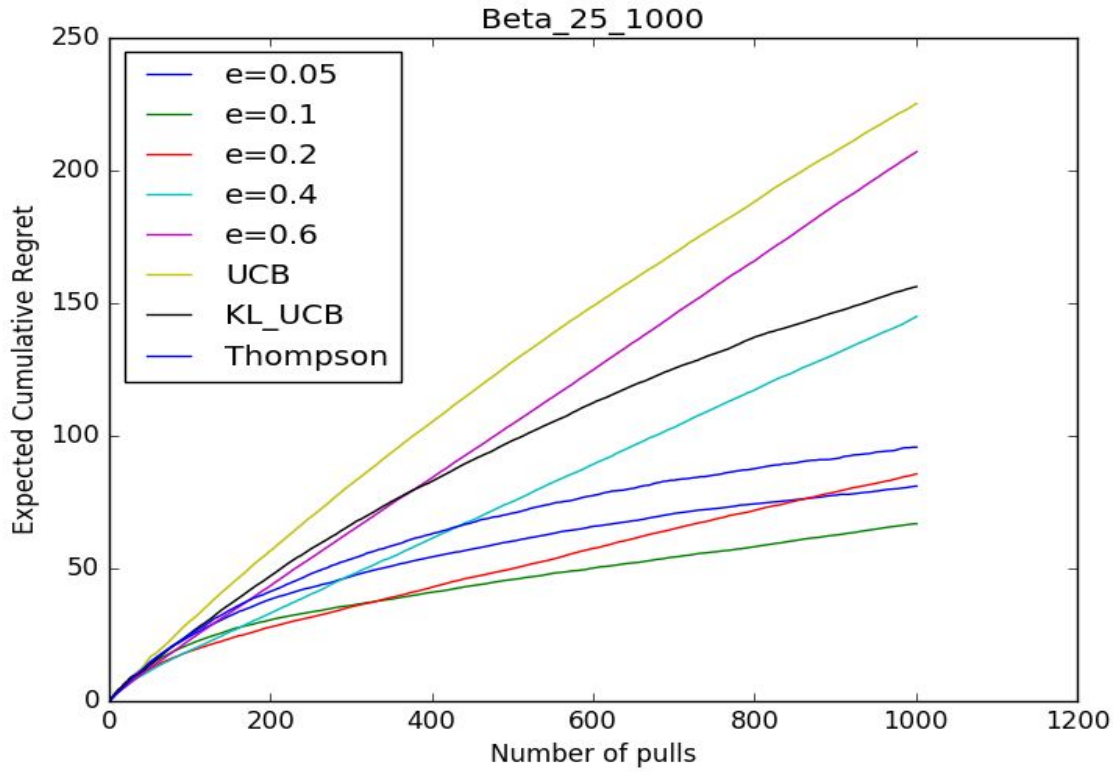


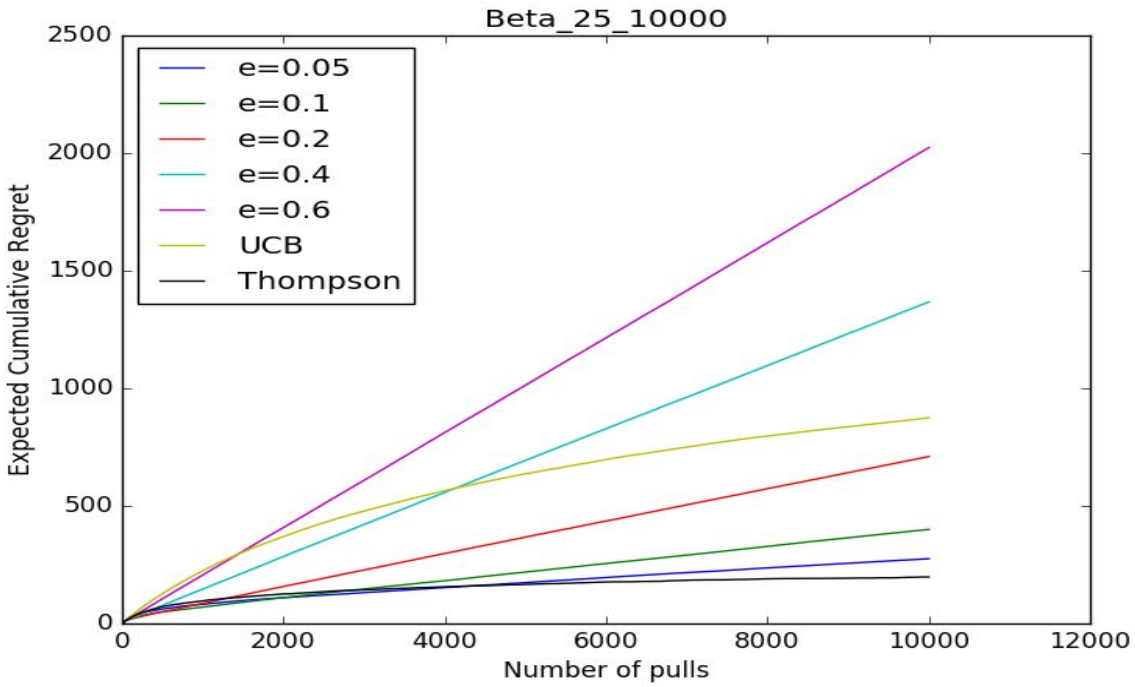

Bernoulli_25_1000
e=0.05
e=0.1
e=0.2
e=0.4
e=0.6
UCB
KL_UCB
Thompson
Expected Cumulative Regret
Number of pulls
0
50
100
150
200
250
300
0
200
400
600
800
1000
1200

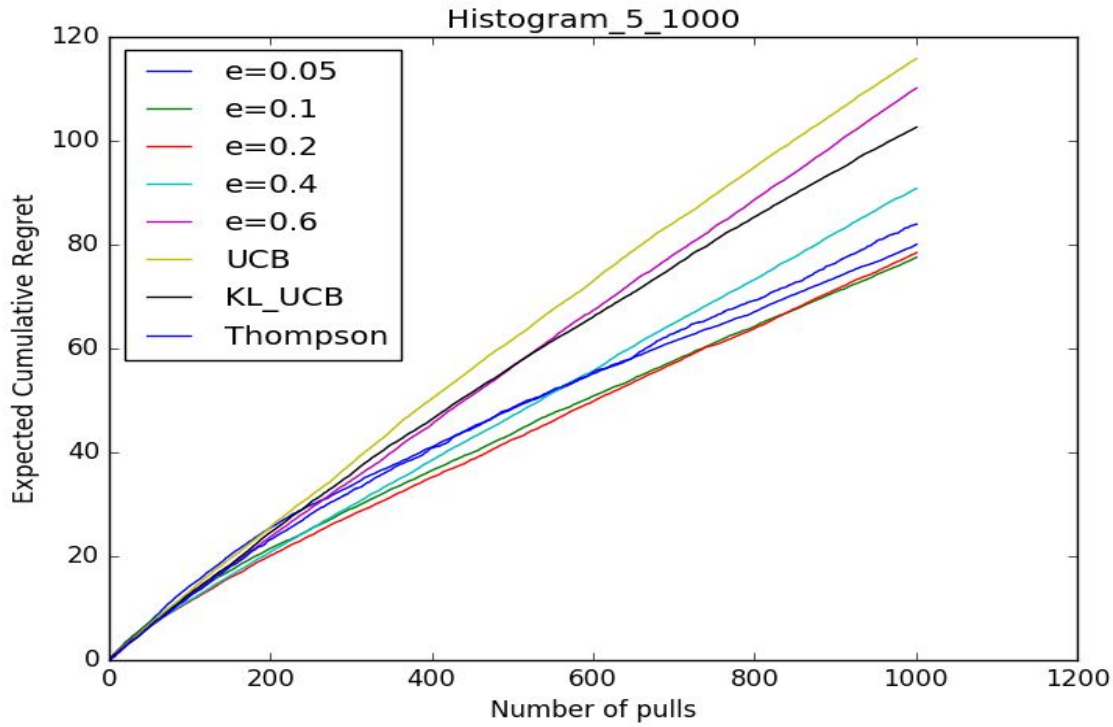

Bernoulli_25_10000
e=0.05
e=0.1
e=0.2
e=0.4
e=0.6
UCB
Thompson
Expected Cumulative Regret
Number of pulls
0
500
1000
1500
2000
2500
3000
0
2000
4000
6000
8000
10000
12000

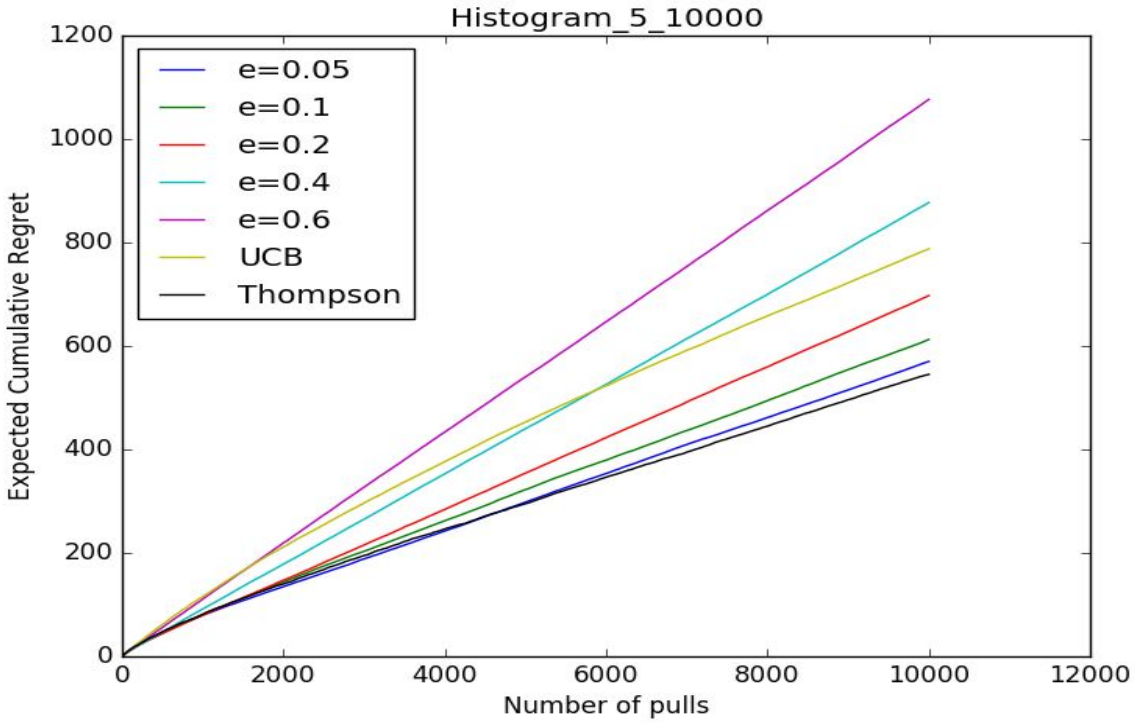

Beta_5_1000
e=0.05
e=0.1
e=0.2
e=0.4
e=0.6
UCB
KL_UCB
Thompson
Expected Cumulative Regret
Number of pulls
0
10
20
30
40
50
60
70
80
90
0
200
400
600
800
1000
1200

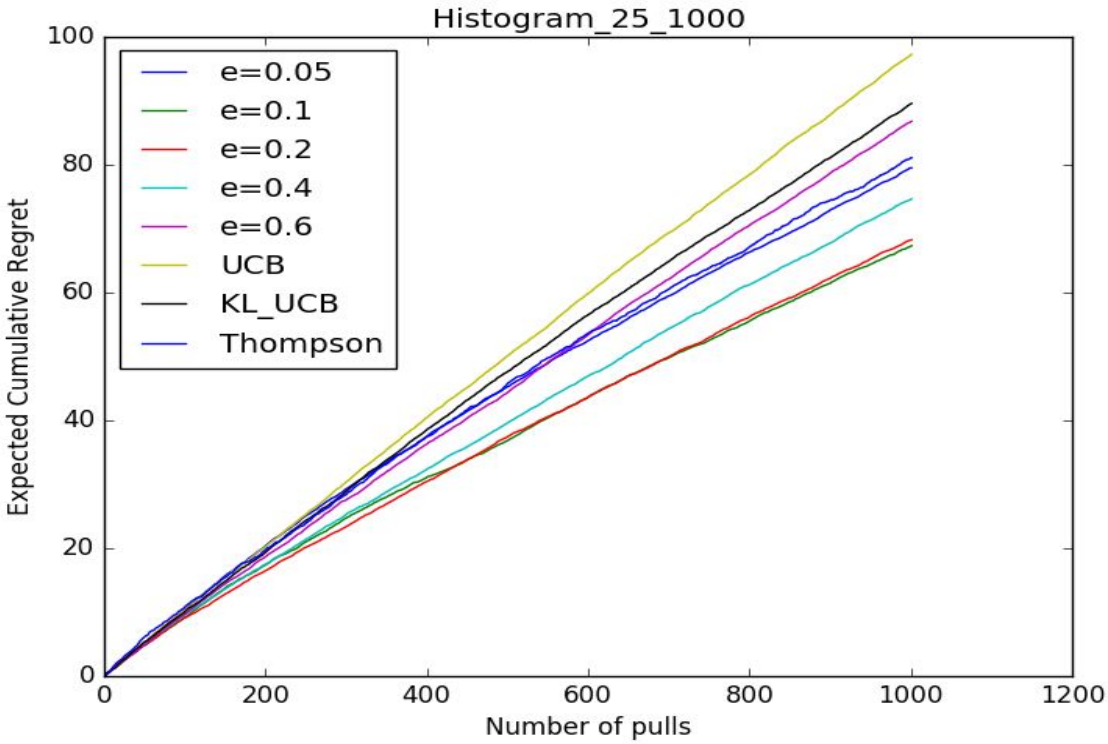

Beta_5_10000
e=0.05
e=0.1
e=0.2
e=0.4
e=0.6
UCB
KL_UCB
Thompson
Expected Cumulative Regret
Number of pulls
0
100
200
300
400
500
600
700
800
900
0
2000
4000
6000
8000
10000
12000

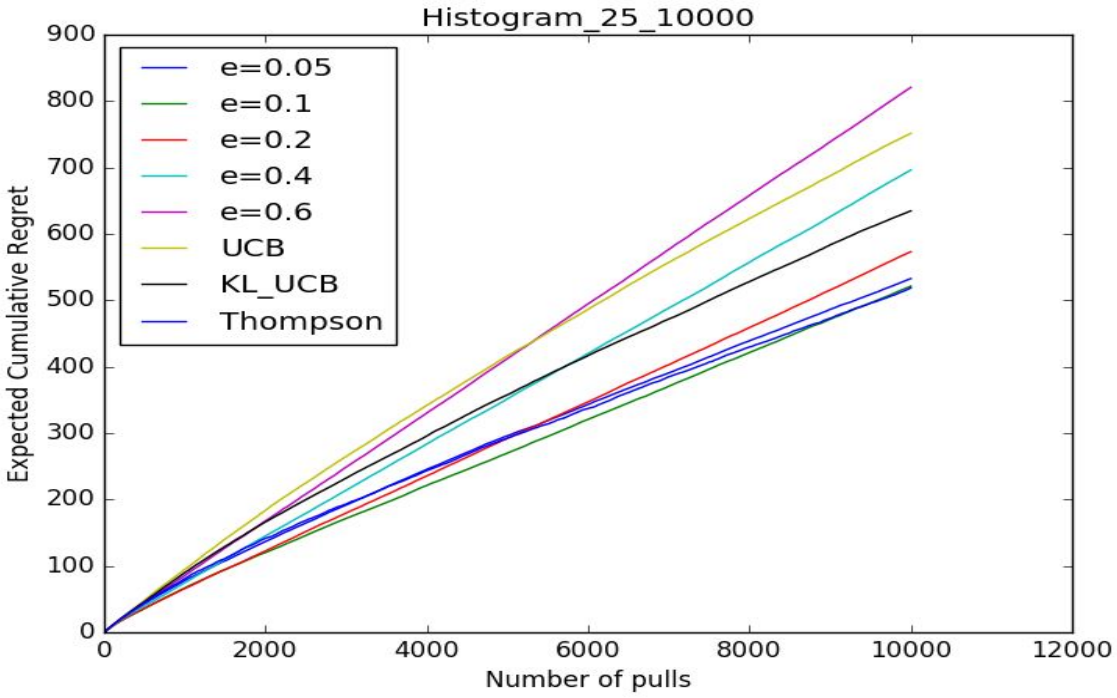

Beta_25_1000
e=0.05
e=0.1
e=0.2
e=0.4
e=0.6
UCB
KL_UCB
Thompson
Expected Cumulative Regret
Number of pulls
0
50
100
150
200
250
0
200
400
600
800
1000
1200


Beta_25_10000
e=0.05
e=0.1
e=0.2
e=0.4
e=0.6
UCB
Thompson
Expected Cumulative Regret
Number of pulls
0
500
1000
1500
2000
2500
0
2000
4000
6000
8000
10000
12000


Histogram_5_1000
e=0.05
e=0.1
e=0.2
e=0.4
e=0.6
UCB
KL_UCB
Thompson
Expected Cumulative Regret
Number of pulls
0
20
40
60
80
100
120
0
200
400
600
800
1000
1200


Histogram_5_10000
e=0.05
e=0.1
e=0.2
e=0.4
e=0.6
UCB
Thompson
Expected Cumulative Regret
Number of pulls
0
200
400
600
800
1000
1200
0
2000
4000
6000
8000
10000
12000


Histogram_25_1000
e=0.05
e=0.1
e=0.2
e=0.4
e=0.6
UCB
KL_UCB
Thompson
Expected Cumulative Regret
Number of pulls
0
20
40
60
80
100
0
200
400
600
800
1000
1200


Histogram_25_10000
e=0.05
e=0.1
e=0.2
e=0.4
e=0.6
UCB
KL_UCB
Thompson
Expected Cumulative Regret
Number of pulls
0
100
200
300
400
500
600
700
800
900
0
2000
4000
6000
8000
10000
12000

1. For small horizons, larger values of epsilon perform better than smaller values. This is due to the following fact: If epsilon is small, then it does very little exploration and tends to pick optimal arm without exploring much. It chooses non-optimal arm in the beginning very often leading to rapid increase in the regret. As the number of pulls increase and more exploration happens, the empirical means get closer to the actual means and hence the optimal arm is selected more often. At each pull the regret for the epsilon greedy algorithm increases rapidly and is proportional to the value of epsilon

$$\sum_{t=0}^{T-1} p^* - E[r^t] = \sum_{t=0}^{T-1} p^* - \varepsilon * \bar{p} - (1-\varepsilon) * p^* = \sum_{t=0}^{T-1} \varepsilon * (p^* - \bar{p})$$

   That is why for small values of epsilon the regret increases slowly
2. In some cases epsilon-greedy algorithm with epsilon = 0.05 performs better than algorithms like UCB and KL-UCB. This is unexpected. But we know that the regret for UCB, KL-UCB and Thompson Sampling is of the order log(T). Due to large constants it might be the case that epsilon greedy performs better than these algorithms. Eventually if we increase the horizon, the curve for epsilon-greedy will cross the curve for UCB and KL-UCB as epsilon-greedy increases linearly with each pull. It can also be observed that Thompson Sampling performs the best in almost all the cases.
3. We can improve the performance for the epsilon greedy cases by incorporating a factor of time in it. Initially, a large value of epsilon can be chosen for more exploration and then as time progresses epsilon can be decreased leading to better exploitation.

4. To account for non-bernoulli rewards: The epsilon-greedy, UCB and KL-UCB algorithms remain the same because they depend only on the empirical reward and number of pulls of each arm. Although, KL-UCB could be adapted to new distributions by using the corresponding KL Divergence for those distributions but the paper mentions that the performance achieved by that is not much different from that obtained by using the same formula as Bernoulli distributions. For thompson sampling , I refered to this paper http://proceedings.mlr.press/v23/agrawal12/agrawal12.pdf

   In this paper, they sample the mean of each arm from beta distribution and pull the arm corresponding to the max mean and get the reward in the range [0,1]. Using the reward as the probability of success, another Bernoulli trial is performed and the parameters of the corresponding beta distribution are updated according to the result of this trial.
5. Standard way for implementing epsilon-greedy and UCB was used. For KL-UCB I uses bisection method to maximize for q starting with lower limit as p and higher limit as 1. 0log(0/0), 0log(x/0) is treated as 0. For thompson sampling too, the standard distribution was used.