ASSIGNMENT 2 REPORT

Multi-Armed Bandits: CS 6046

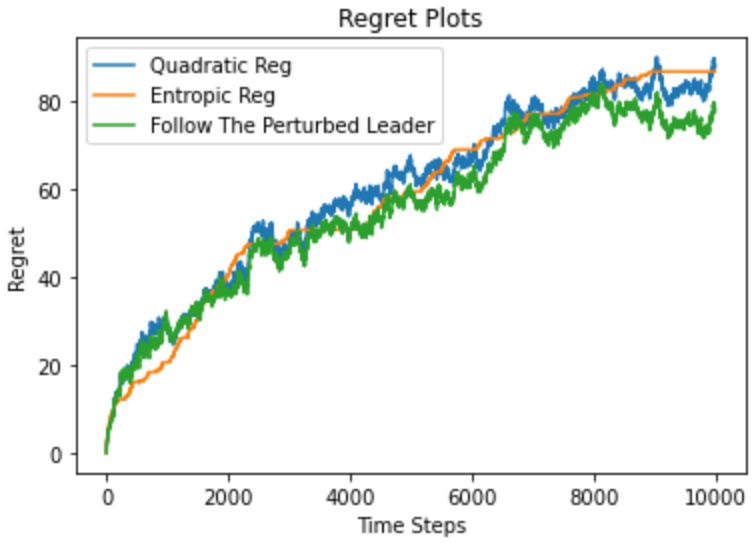
PREPARED BY

Pragalbh Vashishtha
MM19B012

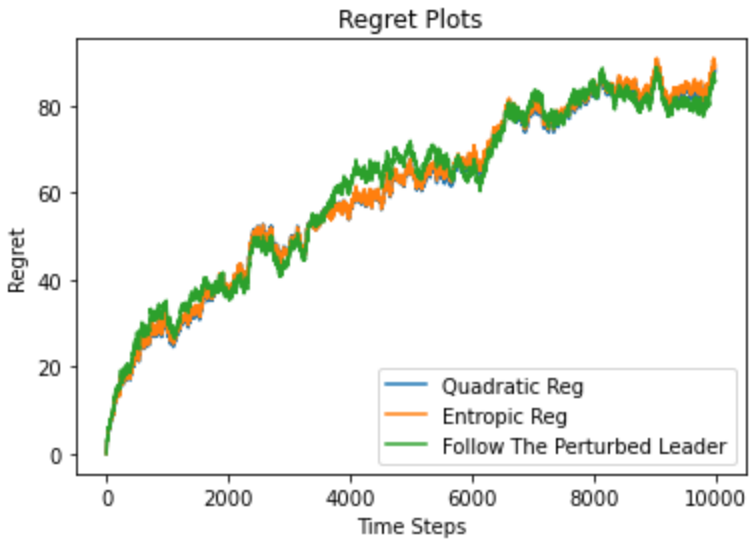
1. Question 1:-

Regret for the perturbed leader is similar to that of quadratic and entropic regularizer. It follows the leader with some noise added. Since this is a randomized algorithm, upon solving the regret bound for the algorithm, a regret bound of $O\left(\sqrt{Td}\right)$ is expected.

Note that FTPL performs marginally better than Entropic and Quadratic regularizers.



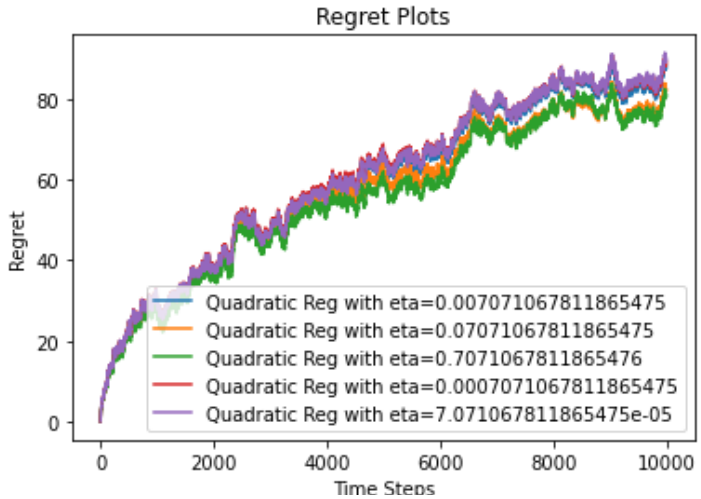
Regret plots with best values of η for each

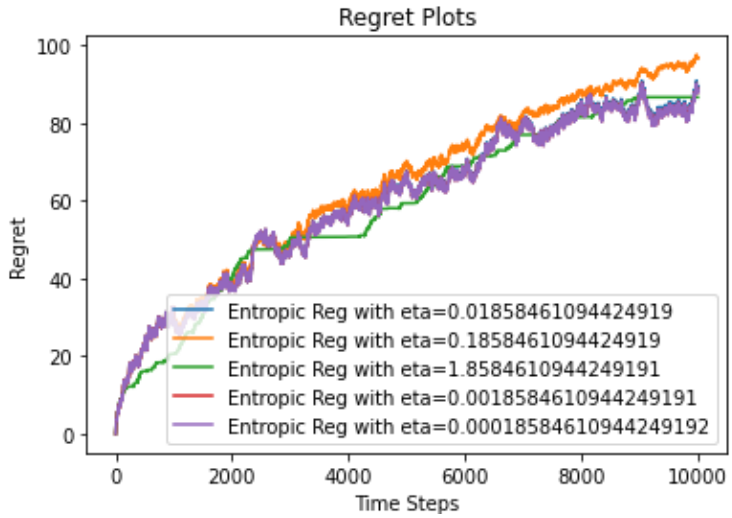


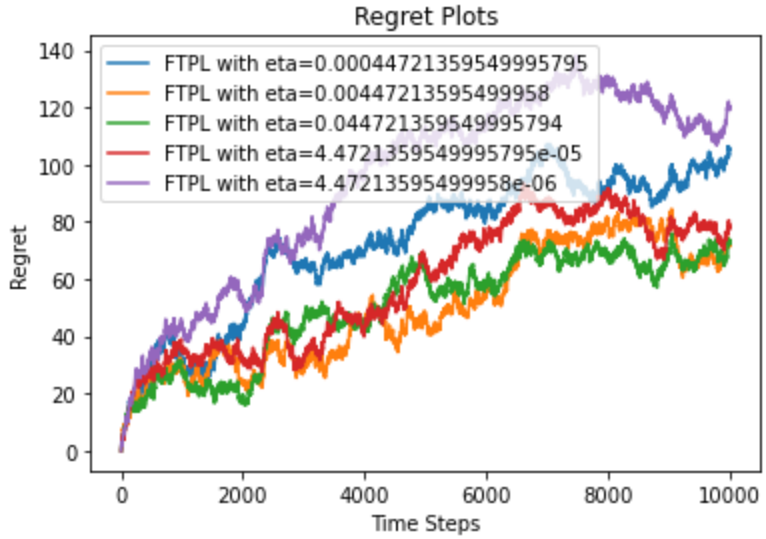
Regret plots with the theoretic values of η

Regret plot for Quadratic ,entropic and

perturbed leader with different η values







For finding out the values of the η which worked the best, the theoretical value was taken and compared to 10x, 100x, 0.1x, 0.01x.

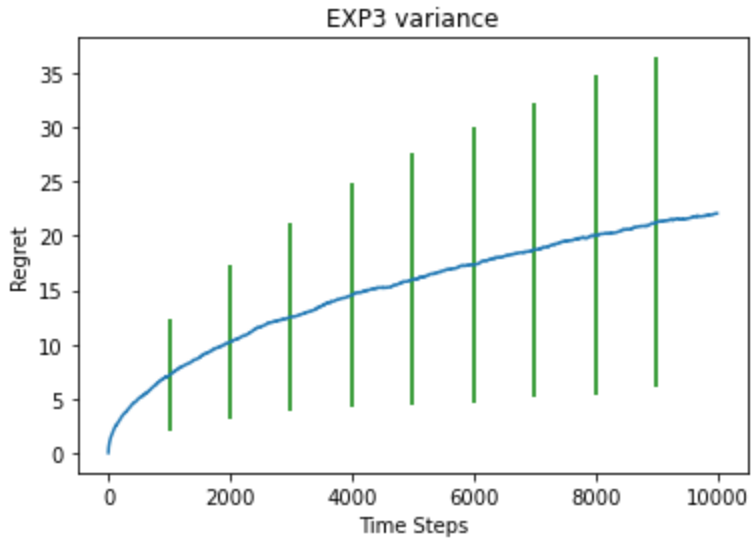
The following graphs were obtained:-

Note the values for the best regret for each case:-

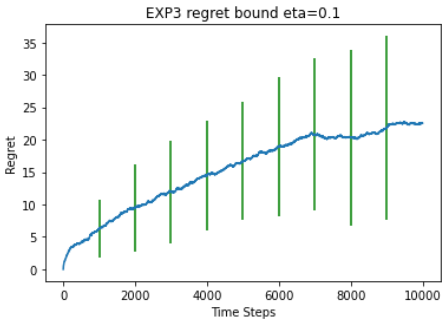
- Quadratic Regularization:- 0.070716 (10x theoretical)
- Entropic Regularization:- 1.8584 (100x theoretical)
- Follow the Regularized Leader:- 0.0044721 (10x theoretical)

2.Question 2:-

The following graph shows the error bars for 1000 iterations of EXP3.

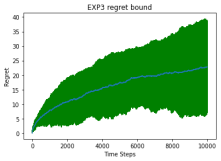


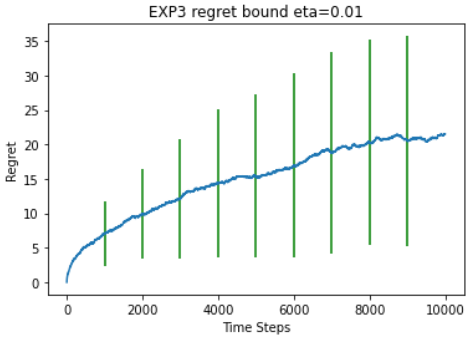
Below the exp3 variance is shown for various values of η. The variance is relatively the same, and very high. However on closer inspection, variance is least for 0.1 and increases with increasing η. With η less than 0.1 also it increases again

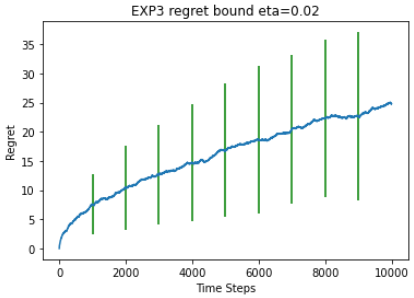


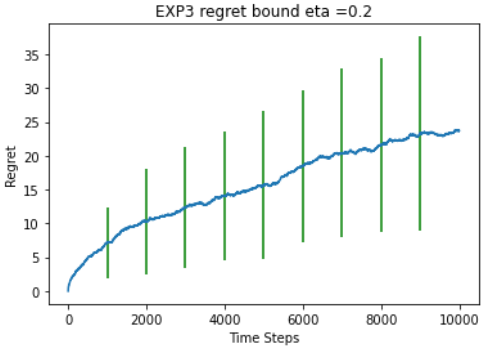
The variance is least when η =0. It is of order O(n). The order of variance can be as high as $\Omega(n^2)$ and is on average O(n).[1]

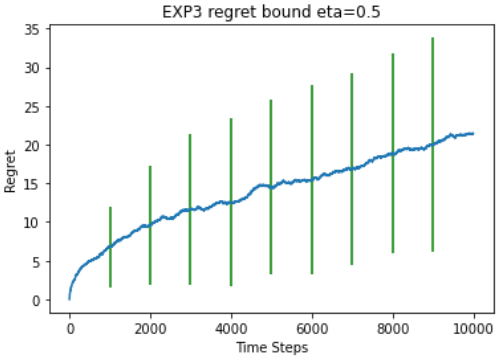
Note that variance is plotted only for sparse points. Else the graph shown below will originate

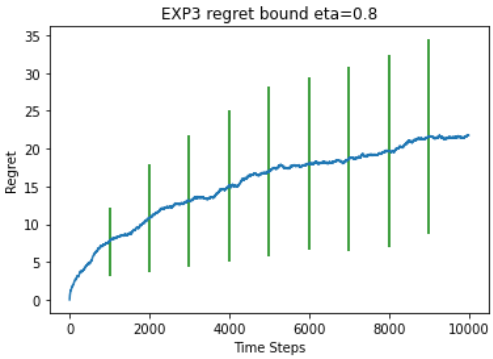


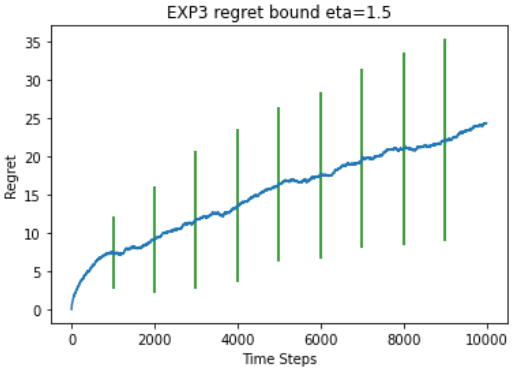












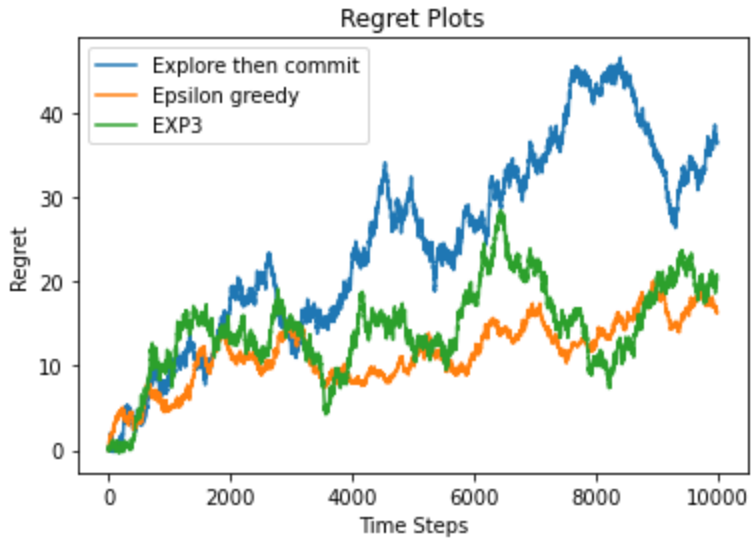
To combat this variance the following change is proposed:-
[original algorithm proposed by Peter Auer et al.]

$\tilde{P}_{ta} = (1-\gamma)P_{ta} + \gamma/k.$ is the update rule. The expected regret is at least Ω(nγ). So smaller the γ smaller the variance of the regret bound.

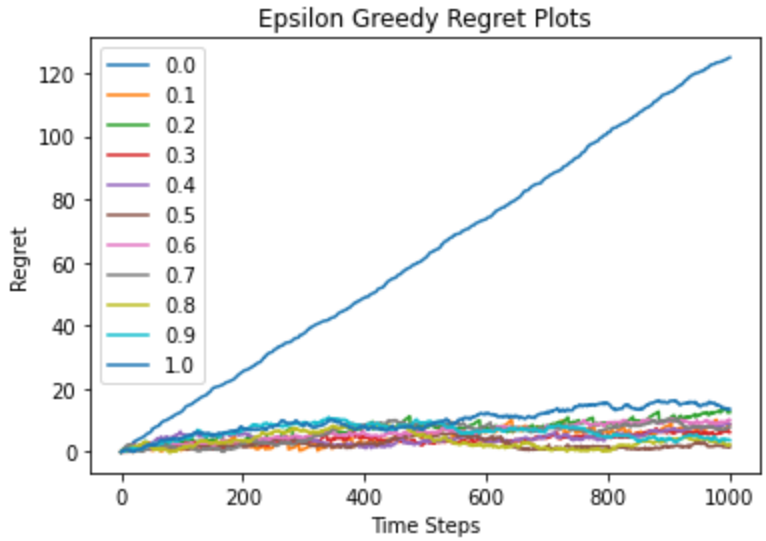
3. Question 3

The following image shows the performance of Epsilon Greedy, Round Robin (explore-then-commit) and EXP3:-

Explore then commit has the highest regret bound, while EXP3 and Epsilon greedy have relatively smaller regret bounds (about half)



The following is the effect of epsilon on the regret bounds



The regret is linear for epsilon =0 as it is pure exploration and no exploitation
Removing epsilon= 0, we note the following :-

Most regret is obtained for pure exploitation (epsilon=1). The best epsilon for regret is around 0.7 as seen in the graph