CHENNAI MATHEMATICAL INSTITUTE

Reinforcement learning

Assignment 4
Date: April 30, 2020. Due date: May 10, 2020.

Use the code available at https://github.com/johnmyleswhite/BanditsBook which gives implementations of ϵ-greedy,UCB1 and a version of γ-greedy with EXP3, γ-EXP3. Understand each code well. And write code for EXP3 as we have discussed it in class. Solve the following using the code:

(1) Compare the performance of ϵ-greedy, UCB1, EXP3 and γ-EXP3 and (with $\gamma = 0.05$) for 5 Bernoulli bandits for a horizon of 1000. And use for the bandits $[0.1, 0.1, 0.1, 0.1, 0.6]$ in one graph and use $[0.1, 0.2, 0.5, 0.8, 0.95]$ in the other. Plot the pseudo regret of the algorithms on the same graph. To plot the regret at time instant t simulate each bandit a 1000 times and take the average.

(2) For each of the algorithms plot the average number of times an arm was choosen upto time instant t (averaged over the 1000 simulations).

(3) Implement Problem 11.8, a,b,c from Lattimore and Szepesvari's book.