

Assignment 3: April 13

Instructions: Discussion among the class participants is highly encouraged. But please make sure that you understand the algorithms and write your own code.

Submit Solutions by 11:59PM, 21st April on Moodle. Late submission will not be evaluated.

Question 1 (15 points) You will implement and compare different algorithms for sampling the arms of a stochastic multi-armed bandit. Each arm provides i.i.d. rewards from a Bernoulli distribution with mean given in the table below. The objective is to minimise the regret. The algorithms you will implement are epsilon-greedy exploration, UCB, KL-UCB, Thompson Sampling.

Arm	Arm 1	Arm 2	Arm 3	Arm 4	Arm 5
Mean	0.4	0.3	0.5	0.2	0.1

Table 3.1: Table 1

For epsilon greedy, you can choose $\epsilon = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7$

Question 2 (15 Marks) Pure exploration, Best Arm Selection: Consider a K -armed bandit where each arm is a Bernoulli random variable. We would like to identify the best arm with probability atleast $1 - \delta$, i.e., $\Pr\{\hat{I} = I^*\} \geq 1 - \delta$, where \hat{I} is the estimated best arm and I^* is the best arm. Numerically compare the expected sample complexity and the mistake probability of the following algorithms for different values of K . Set (confidence parameter) $\delta = 0.1$.

- KL-LUCB [1] with parameters $\epsilon = 0, k_1 = 4e + 4, \beta(t, \delta) = \log(k_1 K t^\alpha / \delta) + \log \log(k_1 K t^\alpha / \delta)$
- lil'UCB [2] with parameters $\epsilon = 0.01, \beta = 1, \lambda = 9, \sigma = 0.5$.

Vary K between 10 to 50 in step size of 10. For each K , generate 1000 bandit instances with mean reward for each arm drawn uniformly at random from $[0, 1]$. Average the sample complexities over all the instances to obtain the average sample complexity and the mistake probability should be calculated as the fraction of times non best arm is returned after stop signals are received.

Submission Format and Evaluation: You should submit a report along with your code. Please zip all your files and upload via Moodle. The zipped folder should named as YourRegistrationNo.zip e.g. '154290002.zip'. The report should contain one figure with four plots corresponding to each algorithm in Q.1. Write a brief summary of your observations. We may also call you to a face-to-face session to explain your code.