CS 747 (Spring 2025) Week 1 Test (Batch 1)

5.35 p.m. – 6.00 p.m., January 16, 2025, LA 001

Name: ______________________________ Roll number: ________________

Note. There are two questions, one on each page. Provide your answer to each question in the space given below it. Draw a line (either vertical or horizontal) and do all your rough work on one side of it.

Question 1. Let $\mathcal{I}$ be the set of all 2-armed bandit instances in which both arms yield Bernoulli rewards, and the means of the arms are different. Let $\mathcal{L}$ be the set of all deterministic algorithms that operate on instances in $\mathcal{I}$. State whether the following statement is true or false, and justify your answer. [1 mark]

> *There exists an algorithm $L \in \mathcal{L}$ whose (expected cumulative) regret on every bandit instance $I \in \mathcal{I}$ grows linearly with the horizon (that is, with the total number of pulls).*

Answer 1.

The statement is true. Suppose that the arms' means are p_1 and p_2, such that $p_1 > p_2$. Choose L to be the algorithm that pulls the arms in round-robin fashion, starting with a fixed arm. The arm is fixed by its name/index, not its mean (which is unknown to the algorithm).

Notice that after T pulls, the inferior arm has been pulled either $\lceil T/2 \rceil$ or $\lfloor T/2 \rfloor$ times. Hence the expected cumulative regret of L after T pulls is either $(p_1 - p_2)\lceil T/2 \rceil$ or $(p_1 - p_2)\lfloor T/2 \rfloor$—which in both cases is lower-bounded $(p_1 - p_2)(T/2 - 1)$—a linear function of T.

Question 2. In a particular instance $I \in \mathcal{I}$ from Question 1, one arm's mean reward is $p = 0.6$. If this arm is pulled 5 times, what is the probability that its empirical mean lies in the interval $(0.3, 0.7]$? Recall that the arm's empirical mean is the ratio of the sum of its rewards to its number of pulls. [2 marks]

Answer 2.

After 5 pulls, the empirical mean necessarily has to be from the set $\{0, 1/5, 2/5, 3/5, 4/5, 1\}$. The only values from this set that lie in $(0.3, 0.7]$ are $2/5$ and $3/5$. The probability that the empirical mean is $2/5$ is $\binom{5}{2}p^2(1-p)^3$, while the probability that the empirical mean is $3/5$ is $\binom{5}{3}p^3(1-p)^2$. The required answer is the sum of these probabilities, which works out to

$$10p^2(1-p)^2 = 0.576.$$

CS 747 (Spring 2025) Week 1 Test (Batch 2)

6.15 p.m. – 6.40 p.m., January 16, 2025, LA 001

Name: ______________________________ Roll number: ______________

Note. There is one question in this test. You can use the space on both pages for your answer. Draw a line (either vertical or horizontal) and do all your rough work on one side of it.

Question 1. Suppose the following algorithm (which resembles the ϵG1 algorithm presented in class) is executed on a 2-armed bandit instance in which both arms yield Bernoulli rewards. Out of the horizon of T, the first t pulls are given to the first arm, and the next t pulls to the second arm. An arm a^{best} is now selected among the two arms such that its empirical mean is the highest; if a tie occurs it is broken uniformly at random. Now, a^{best} is pulled for the remaining $T - 2t$ pulls (this decision is not changed under any circumstances). Assume that $t \geq 1$ and $T \geq 2t + 1$.

Assume that the (true) means of the arms are p_1 and p_2. What is the probability that the very last reward obtained (that is, the outcome of the T-th pull) is 0? Your answer does not have to be a closed-form expression. Explain the steps used to obtain your answer. [3 marks]

Answer 1.

Suppose arm 1 has mean p_1 and arm 2 has mean p_2. If a^{best} is arm 1, then the probability of the last pull giving a 0 is $(1 - p_1)$. On the other hand, if a^{best} is arm 2, then the probability of the last pull giving a 0 is $(1 - p_2)$. Our main task is to work out the probability of each arm being selected as a^{best}.

Arm 1 is selected as a^{best} if and only if its empirical mean exceeds that of arm 2's after t pulls, or the empirical means are equal and arm 1 wins the tiebreak. In turn, the relative order of the empirical means depends on the number of 1-rewards each arm has gathered in its first t pulls. We have

$$\mathbb{P}\{\text{Arm 1 is } a^{\text{best}}\} = \sum_{x=1}^{t} \sum_{y=0}^{x-1} \binom{t}{x} (p_1)^x (1 - p_1)^{t-x} \binom{t}{y} (p_2)^y (1 - p_2)^{t-y} +$$
$$\frac{1}{2} \sum_{x=0}^{t} \binom{t}{x} (p_1)^x (1 - p_1)^{t-x} \binom{t}{x} (p_2)^x (1 - p_2)^{t-x}.$$

The required probability is

$$\mathbb{P}\{\text{Arm 1 is } a^{\text{best}}\}(1 - p_1) + (1 - \mathbb{P}\{\text{Arm 1 is } a^{\text{best}}\})(1 - p_2).$$