CS 747 (Spring 2025) Week 3 Test (Batch 1)

5.35 p.m. – 6.00 p.m., January 30, 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. You are interacting with a 2-armed bandit instance, in which each arm gives Bernoulli rewards. Suppose the arms are 1 and 2. It is known that the arm have *different* means, but the means themselves are unknown. Your aim is to determine the superior arm (that is, the arm with the higher mean) by pulling the arms and observing the rewards. The requirement is that you stop and correctly identify the superior arm with probability at least 1 - δ. You can either propose an independent solution, or use as a subroutine the procedure L described below .

> L is a blackbox routine that works as follows. Initially, a mistake probability $\delta_L \in (0, 1)$ and a "threshold" $c_L \in (0, 1)$ are given to L. Thereafter, L is fed a sequence of 0's and 1's. L provides the following guarantee: if the sequence of input bits are generated i.i.d. from a Bernoulli variable with mean $p \neq c_L$, then L will stop and return the sign of $(p - c_L)$ ("positive" if $p > c_L$, "negative" if $p < c_L$) correctly with probability at least 1 - δ_L.
>
> For example, L may be initialised with $\delta_L = 0.1$ and $c_L = 0.26$. While being fed the sequence $1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, \ldots$, it may stop after the fist ten entries and return "positive". If the input bit sequence is indeed generated by repeatedly sampling a Bernoulli distribution with mean 0.56, then the probability that L returns "positive" is at least $1 - 0.1 = 0.9$. You can assume that L is deterministic, although this assumption is not required for your task. You are also permitted to run L multiple times, in sequence or in parallel.

Describe your procedure (with or without the use of L) and explain why it provides the required probabilistic guarantee. [3 marks]

Answer 1.

We provide two solutions; in principle many other combinations are possible.

Without using L. We pull both arms at each time step $t \geq 1$. For each arm $a \in \{1, 2\}$, we maintain the empirical average $\hat{p}_a^t$ at each step t. We also construct

$$\text{lcb}_a^t = \hat{p}_a^t - \sqrt{\frac{1}{2t} \ln \frac{kt^\alpha}{\delta}};\ \text{ucb}_a^t = \hat{p}_a^t + \sqrt{\frac{1}{2t} \ln \frac{kt^\alpha}{\delta}},$$

where the constants k and α will be specified shortly. Suppose on some time step t, we observe that the LCB of an arm exceeds the UCB of the other arm, we declare the empirically superior arm as the winner.

The correctness of this algorithm is based on the classic recipe with confidence bounds. Observe that for the algorithm to make a mistake, there must exist an arm $a \in \{1, 2\}$ and a time step $t \geq 1$ for which either the LCB of the arm gos above its true mean, or the UCB of the arm falls below its true mean. For each arm $a \in \{1, 2\}$ and each $t = 1, 2, 3, \dots,$, by Hoeffding's inequality, (1) the probability that $\text{lcb}_a^t > p_a^t$ is at most $\frac{\delta}{kt^\alpha}$, and (2) the probability that $\text{ucb}_a^t < p_a^t$ is at most $\frac{\delta}{kt^\alpha}$. Hence, the probability that there exists an arm a and a time step t on which either of these bad events occurs is at most

$$\sum_{a \in \{1,2\}} \sum_{t=1}^{\infty} \left(\frac{\delta}{kt^\alpha} + \frac{\delta}{kt^\alpha} \right) = \frac{4\delta}{k} \sum_{t=1}^{\infty} \frac{1}{t^\alpha}.$$

If we set, for example, $k = 7, \alpha = 4$, this probability is smaller than δ.

We also need to show that the procedure will terminate with probability 1: that is, the LCB of of one arm will eventually exceed the UCB of the other. The intuition for this is that the square root terms in each confidence bound will become arbitrarily small, and since the empirical means, too, will converge to the true means, the separation must happen.

Using L. At each time step $t = 1, 2, 3, \dots$, we pull both arms. Suppose that their rewards are x_1^t and x_2^t. Create a new Bernoulli random variable y^t as follows.

- If $x_1^t = x_2^t$, then y^t is set to 0 with probability $\frac{1}{2}$ and to 1 with probability $\frac{1}{2}$.
- If $x_1^t > x_2^t$, then y^t is set to 1.
- If $x_1^t < x_2^t$, then y^t is set to 0.

Suppose the true means of arm 1 and arm 2 are p_1 and p_2, respectively. By our process,

$$\mathbb{P}\{y^t = 1\} = p_1 p_2 \frac{1}{2} + (1 - p_1)(1 - p_2)\frac{1}{2} + p_1(1 - p_2) = \frac{1}{2} + \frac{p_1 - p_2}{2}.$$

Since $y^1, y^2, \dots$ are drawn i.i.d. from the same Bernoulli distribution, determining if this distribution has a mean exceeding $\frac{1}{2}$ is equivalent to determining if $p_1 > p_2$. L is a readymade device for this precise purpose. We set $c_L = \frac{1}{2}$ and $\delta_L = \delta$, and pass L the samples $y^1, y^2, \dots$ until L terminates and returns "positive" (in which case we declare "$p_1 > p_2$") or "negative" (in which case we declare "$p_1 < p_2$").

CS 747 (Spring 2025) Week 3 Test (Batch 2)

6.15 p.m. – 6.40 p.m., January 30, 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. In the proof done in class this week (to upper-bound the regret of UCB), in several places we used the following result: for any two events A and B,

$$\mathbb{P}\{A \text{ or } B\} \leq \mathbb{P}\{A\} + \mathbb{P}\{B\},$$

or equivalently, that

$$\mathbb{P}\{A\} + \mathbb{P}\{B\} - \mathbb{P}\{A \text{ or } B\} \geq 0.$$

Does the quantity $\mathbb{P}\{A\} + \mathbb{P}\{B\} - \mathbb{P}\{A \text{ or } B\}$ have a qualitative interpretation? Why is it guaranteed to be non-negative? [1 mark]

Answer 1.

$\mathbb{P}\{A\} + \mathbb{P}\{B\} - \mathbb{P}\{A \text{ or } B\}$ is nothing but $\mathbb{P}\{A \text{ and } B\}$. Since it is a probability, it is non-negative.

Question 2. A 3-armed bandit instance I has arms that yield Bernoulli rewards. The arms are 1, 2, and 3, with means p_1, p_2, and p_3, respectively, satisfying

$$p_1 > p_2 > p_3.$$

Fix the horizon of pulls $T \geq 3$. For any algorithm X that is executed on I, let u_1^X, u_2^X, and u_3^X denote the *expected* number of pulls performed by X over horizon T on arms 1, 2, and 3, respectively (thus $u_1^X + u_2^X + u_3^X = T$). Let R^X denote the (expected cimulative) regret of X on I after T pulls.

Y and Z are algorithms that can be executed on I. Both algorithms allot the same expected number of pulls to arm 1: that is,

$$u_1^Y = u_1^Z.$$

The expected number of pulls allotted to each arm by Y is consistent with the order of the means. That is:

$$u_1^Y > u_2^Y > u_3^Y.$$

However, the expected number of pulls under Z satisfy a different relation, which is:

$$u_1^Z > u_3^Z > u_2^Z.$$

Can we conclude that $R^Y < R^Z$. Can we conclude that $R^Y > R^Z$? Answer yes or no to both questions and justify your answers. [2 marks]

Answer 2. We can conclude that $R^Y < R^z$ (and not the other way round). Since both algorithms perform T pulls, and also pull arm 1 the same number of times, it means that their total number of pulls of arms 2 and 3 is the same. In turn, Y gives more of these pulls to arm 2, which has a higher mean than arm 3. Hence Y must incur lower regret than Z. Here is a mathematical working, after defining $\Delta_2 = p_1 - p_2$ and $\Delta_3 = p_1 - p_3$, and applying the observation that $u_3^Y < u_3^Z$.

$$\begin{aligned}
R^Y &= \Delta_2 u_2^Y + \Delta_3 u_3^Y \\
&= \Delta_2(u_2^Y + u_3^Y) + (\Delta_3 - \Delta_2)u_3^Y \\
&< \Delta_2(u_2^Y + u_3^Y) + (\Delta_3 - \Delta_2)u_3^Z \\
&= \Delta_2(u_2^Z + u_3^Z) + (\Delta_3 - \Delta_2)u_3^Z \\
&= \Delta_2 u_2^Z + \Delta_3 u_3^Z \\
&= R^Z.
\end{aligned}$$