AI4010: Online Learning
Second Midterm Exam
Oct 2025

Instructions:

- *The total number of marks is 20.*
- *The total duration of the exam is 90 minutes. No electronic aids are allowed. You can keep a maximum of one sheet of paper with formulas/notes.*
- *All questions are mandatory. A yes/no answer without proper proof or justification will be given zero marks even if it is correct.*
- *Any plagiarism case, if detected, will attract F grade in the course irrespective of overall performance.*
- *Use* $0\log(0) = 0$ *wherever required.*

Problem 0.1 (2 marks). *Let X and Y be independent σ_1 and σ_2 subgaussian random variables with zero mean respectively. Show that $X+Y$ is $(\sigma_1+\sigma_2)$-subgaussian.*

Problem 0.2 (2 marks). *We call a random vector $Z = (Z_1, Z_2, \cdots Z_n)^T$ σ-subgaussian if for any $u \in \Delta_n$, $u^T Z$ is σ-subgaussian. Prove that if Z_i's are mutually independent and each Z_i is σ-subgaussian then the random vector Z is σ-subgaussian.*

Problem 0.3 (3 marks). *In this question, we will see that the Hoeffding's bound may not always be tight. Let $X_1, X_2, \cdots X_n$ be iid Bernoulli random variables with mean μ. Chernoff's multiplicative bound is given by*

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \leq (1-\varepsilon)\mu\right) \leq exp\left(\frac{-n\mu\varepsilon^2}{2}\right) \forall \varepsilon \in [0,1].$$

Show that for certain values of μ, Chernoff's bound above gives a better tail bound than Hoeffding's inequality. Show your work.

Problem 0.4 (5 marks). *What is the probability (upper bound) that sequential halving algorithm with fixed budget $T >> N$ returns a sub-optimal arm. Write detailed proof.*

Hint: Use Hoeffding's lemma and Union Bound.

Problem 0.5 (3+3+2 = 8 marks). *(Two arm Bernoulli bandit with side information) Consider a stochastic 2-armed bandit problem where each arm i's reward sequence is generated independently from a Bernoulli distribution with parameter $\mu_i, i = 1, 2$. Further, it is known that $\mu_1 \neq \mu_2$ and*

$\mu_1, \mu_2 \in \{a, b\}$ *where* $0 < a < b < 1$ *are known constants; the only uncertainty is in the order. Denote* $\Delta := b - a$.

Consider the following simple algorithm. In the beginning, play each arm once, i.e., $i_1 = 1, i_2 = 2$. *At every subsequent time* $t \geq 3$, *if there exists an arm whose observed empirical mean so far exceeds* $(a+b)/2$, *then play the arm with the highest empirical mean. Else, play both arms one after another, i.e.,* $i_t = 1$ *followed by* $i_{t+1} = 2$.[1]

Without loss of generality, let arm 1 be the optimal arm when running the algorithm. Split the set of times when arm 2 is played according to whether its observed empirical mean so far is (i) greater than or (ii) at most $(a + b)/2$.

1. *Bound (from above) the sum of probabilities of playing arm 2 at all times when event (i) occurs. (3 marks)*

2. *Bound (from above) the sum of probabilities of playing arm 2 at all times when event (ii) occurs by using the definition of the algorithm and relating event (ii) to an event involving the empirical mean of arm 1. (3 marks)*

3. *Put together the conclusions of the previous parts to derive a regret bound independent of* T *and depending only on* Δ. *(2 marks)*

[1] If at time t the empirical rewards of both the arms are at-most $(a+b)/2$, skip the empirical rewards camparison at $t+1$.