Homework #1
DATA 37200: Learning, Decisions, and Limits (Winter 2025)

Due Friday 01/31, 9:00 pm

General Instructions The assignment is meant to practice your understanding of course materials, and some of them are challenging. You are allowed to discuss with fellow students, however please **write up your solutions independently** (e.g., start writing solutions after a few hours of any discussion) and, equally importantly, acknowledge everyone you discussed the homework with on your writeup. All course materials are available here https://frkoehle.github.io/data37200-w2025/. Unless particularly stated, any results/theorems covered in our class can be used without the need of a proof.

Notably, attempt to consult outside sources, on the Internet or otherwise, for solutions to any of these homework problems is *not* allowed. **Needless to say, you are not allowed to query any Large Language Models (LLMs) for solving the HW problems.**

Whenever a question asks you to "show" or "prove" a claim, please provide a formal mathematical proof. These problems have been labeled based on their difficulties. `Short` problems are intended to take you 5-15 minutes each and `medium` problems are intended to take 15-30 minutes each. `Long` problems may take anywhere between 30 minutes to several hours depending on whether inspiration strikes. Note that, the total score is meant to *not* be normalized to 100 (for instance, this HW has 30 in total for regular students and additional 15 points for those who take it as elective).

Finally, please write your solutions in latex — **hand written solutions will not be accepted.** Hope you enjoy the homework!

Problem 1: Properties of KL-Divergences

Let p, q be two distributions with finite support X. Prove the following three properties about their KL-divergence.

Note: while for most HW problems, you can directly cite any results we covered in class as granted, for this problem you will have to (and should be able to) derive proofs from scratch using basic algebra.

1. (**10 points**) $KL(q, q) \geq 0$. Moreover, $KL(p, q) = 0$ if and only if $p = q$.
2. (**10 points**) For $i = 1, \cdots, n$, let p_i, q_i be two distributions supported on finite set X_i and let $p = \Pi_{i=1}^n p_i, q = \Pi_{i=1}^n q_i$ be their product distributions, respectively. Show that $KL(p, q) = \sum_{i=1}^n KL(p_i, q_i)$.
3. (**20 points**) For any event $A \subseteq X$, show that

$$2[p(A) - q(A)]^2 \leq KL(p, q)$$

where $p(A) = \sum_{x \in A} p(x)$.

Hint: To prove the third property, there are multiple approaches. For some approach, it may be helpful to check out the following basic algebraic conclusions or relations: (a) what's the relation between $|p(A) - q(A)|^2$ for any A and the l_1 distance between vectors p, q; (b) given any $1 > p_1 \geq q_1 > 0$ and $1 > p_2 > q_2 > 0$, which of the following two terms is larger: $p_1 \ln(\frac{p_1}{q_1}) + p_2 \ln(\frac{p_2}{q_2})$ and $(p_1 + p_2) \ln(\frac{p_1+p_2}{q_1+q_2})$?

Problem 2: KL-Divergences for Example Distribution

Recall from class that RC_ϵ is a Bernoulli distribution with mean $(1 + \epsilon)/2$.

Prove the following conclusions.

1. (**10 points**) $KL(RC_\epsilon, RC_0) = \Theta(\epsilon^2)$ for any $\epsilon \in (0, 1/4)$ where Θ comes from big O and big Theta notation.
2. (**10 points**) $KL(RC_0, RC_\epsilon) = \Theta(\epsilon^2)$ for any $\epsilon \in (0, 1/4)$.

Problem 3: Characterizing "neglected arms" under any MAB algorithm

(**20 points**) Consider any multi-armed bandit instance with k arms where each arm follows a Bernoulli distribution with realized rewards in $\{0, 1\}$. Prove that, for any *deterministic* bandit algorithm,[1] there exits a subset of arms $J \subseteq [k] = \{1, \cdots, k\}$ such that

1. $|J| \geq k/3$
2. for any $j \in J$, $\mathbb{E}(N_j^T) \leq \frac{3T}{k}$ where N_j^T is the total number of times arm j is pulled until time T in this given MAB instance.
3. for any $j \in J$, $\mathbf{Pr}(I^T = j) \leq \frac{3}{k}$ where I^T) denotes the (random) arm that this algorithm pulls at round T.

Problem 4: Improving UCB Gap-Independent Regret Analysis

(**20 points**) In the class, we showed an $O(k\sqrt{T \log T})$ gap-independent regret bound for UCB (specifically, see page 45, 46 of Lecture 2 slides in the above course link). In this question, you are tasked to improve this bound to $O(\sqrt{kT \log T})$ by refining the analysis in class, assuming gap Δ_i's, σ are all upper bounded by constants and $k < T$.

Formally, recall the following two lemma proved in class:

1. The regret decomposition lemma $Regret = \mathbb{E}\big[\sum_{i=1}^k \Delta_i N_i(T)\big]$ where random variable $N_i(T)$ denotes the number of times arm i is pulled until round T;

[1] Recall, deterministic bandit algorithm means it maps any realized past rewords to a deterministic choice of an arm at the next round.

2. For UCB, with probability at least $1 - 2/T$, $N_i \leq 8\sigma^2 \frac{\log(T)}{(\Delta_i)^2} + 1$ holds true with probability 1 simultaneously for every arm $i \in [k]$.

Use the two lemmas above to prove the following regret upper bound for UCB

$$\text{Regret}_T = O(\sqrt{kT \log T}) \tag{1}$$

Hint: the reason that we can strengthen the analysis in class is because the upper bound T used to upper bound $N_i(T)$ on slide 45 is too loose – to see this, we have $\sum_i N_i(T) = T$ with probability 1, whereas $\sum_i T = kT$. Think about how to leverage the equation $\sum_i N_i(T) = T$ to tighten the regret analysis.