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, and $p(x) > 0, q(x) > 0$ for any $x \in X$. 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**) $KL(p, q) \geq 0$. Moreover, $KL(p, q) = 0$ if and only if $p = q$.

   Proof. Observe that $\log$ is concave, i.e. $\alpha \log(x) + (1-\alpha)\log(y) \leq \log(\alpha x + (1-\alpha)y), \forall \alpha \in [0, 1]$. We plug into the definition of KL-divergence and interpret the $p(x)$ terms as the weights of the convex combination:

$$KL(p,q) = \sum_{x \in X} p(x) \log\left(\frac{p(x)}{q(x)}\right) = -\sum_{x \in X} p(x) \log\left(\frac{q(x)}{p(x)}\right)$$
$$\geq -\log\left(\sum_{x \in X} p(x) \frac{q(x)}{p(x)}\right) = -\log 1 = 0$$

Note that because $\log(x)$ is *strictly* convex, Jensen's inequality holds with equality if and only if all inputs are equal. Thus, $\forall x \in X, \frac{p(x)}{q(x)} = k$ and k must be 1 since p and q are both probability distributions. So, $KL(p,q) = 0 \iff p = q$. $\square$

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 $KL(p,q) = \sum_{i=1}^n KL(p_i, q_i)$.

Proof. Using properties of logarithms and plugging in the definition of p and q.

$$\begin{aligned}
KL(p,q) &= \sum_{x \in X} p(x) \log\left(\frac{p(x)}{q(x)}\right) \\
&= \sum_{x \in X} p(x) \log\left(\frac{\prod_{i=1}^n p_i(x_i)}{\prod_{i=1}^n q_i(x_i)}\right) \\
&= \sum_{x \in X} p(x) \sum_{i=1}^n \log\left(\frac{p_i(x_i)}{q_i(x_i)}\right) \\
&= \sum_{i=1}^n \sum_{x \in X} p(x) \log\left(\frac{p_i(x_i)}{q_i(x_i)}\right) \\
&= \sum_{i=1}^n \sum_{x \in X} \prod_{j=1}^n p_j(x_j) \log\left(\frac{p(x_i)}{q(x_i)}\right)
\end{aligned}$$

Now, we separate the $j = i$ term in the first $\Pi_{j=1}^n p_j(x_j)$ term:

$$\begin{aligned}
\sum_{i=1}^n \sum_{x \in X} \prod_{j=1}^n p_j(x_j) \log\left(\frac{p(x_i)}{q(x_i)}\right) &= \sum_{i=1}^n \sum_{x \in X} p_i(x_i) \log\left(\frac{p(x_i)}{q(x_i)}\right) \prod_{i \neq j} p_j(x_j) \\
&= \sum_{i=1}^n \sum_{x_i \in X_i} p_i(x_i) \log\left(\frac{p(x_i)}{q(x_i)}\right) \left(\prod_{i \neq j} \sum_{x_j \in X_j} p_j(x_j)\right) \\
&= \sum_{i=1}^n \sum_{x_i \in X_i} p_i(x_i) \log\left(\frac{p(x_i)}{q(x_i)}\right) \left(\prod_{i \neq j} 1\right) \\
&= \sum_{i=1}^n \sum_{x_i \in X_i} p(x_i) \log\left(\frac{p(x_i)}{q(x_i)}\right) \\
&= \sum_{i=1}^n KL(p_i, q_i)
\end{aligned}$$

$\square$

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 $1 > p_1, q_1 > 0$ 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 $(p_1 + p_2) \ln(\frac{p_1+p_2}{q_1+q_2})$?

Proof. We work with hint (b). Specifically, we show that $p_1 \ln(\frac{p_1}{q_1}) + p_2 \ln(\frac{p_2}{q_2}) \geq (p_1 + p_2) \ln(\frac{p_1+p_2}{q_1+q_2})$ (call this inequality (b)). We defer the proof of this to the end.

First, assume that the inequality holds. Then we can condition on any event A to say that:[1]

$$\begin{aligned} KL(p, q) &= \sum_{x \in A} p(x) \ln\left(\frac{p(x)}{q(x)}\right) + \sum_{x \in \overline{A}} p(x) \ln\left(\frac{p(x)}{q(x)}\right) \\ &\geq p(A) \ln\left(\frac{p(A)}{q(A)}\right) + p(\overline{A}) \ln\left(\frac{p(\overline{A})}{q(\overline{A})}\right) \end{aligned}$$

Where we apply inequality (b) on both of the summation terms individually. Now, we want to show that:

$$\begin{aligned} KL(p, q) \geq p(A) \ln\left(\frac{p(A)}{q(A)}\right) + p(\overline{A}) \ln\left(\frac{p(\overline{A})}{q(\overline{A})}\right) &\geq 2[p(A) - q(A)]^2 \\ \implies p(A) \ln\left(\frac{p(A)}{q(A)}\right) + p(\overline{A}) \ln\left(\frac{p(\overline{A})}{q(\overline{A})}\right) - 2[p(A) - q(A)]^2 &\geq 0 \\ \implies p(A) \ln\left(\frac{p(A)}{q(A)}\right) + (1 - p(A)) \ln\left(\frac{1 - p(A)}{1 - q(A)}\right) - 2[p(A) - q(A)]^2 &\geq 0 \qquad (1) \end{aligned}$$

Now, we want to find the $q(A)$ to minimize the difference, so we take the derivative with respect to $q(A)$:

$$\begin{aligned} &\frac{\partial}{\partial q(A)} \left(p(A) \ln\left(\frac{p(A)}{q(A)}\right) + (1 - p(A)) \ln\left(\frac{1 - p(A)}{1 - q(A)}\right) - 2[p(A) - q(A)]^2 \right) \\ &\qquad = -\frac{p(A)}{q(A)} + \frac{1 - p(A)}{1 - q(A)} + 4[p(A) - q(A)] \\ &\qquad = \frac{q(A) - p(A)}{q(A)(1 - q(A))} + 4[p(A) - q(A)] \\ &\qquad = (p(A) - q(A)) \left(4 - \frac{1}{q(A)(1 - q(A))}\right) \end{aligned}$$

To find that inequality 1 is always non-negative (observe that its derivative is minimized at 0 when $q(A) = p(A)$ and the derivative is an upside down parabola w.r.t. to any fixed value of $p(A)$ because when $q(A) \leq p(A)$, $\partial f / \partial q \leq 0$ and when $q(A) \geq p(A)$, $\partial f / \partial q \geq 0$).

[1] $\overline{A}$ denotes the complement of A.

Since the difference is always non-negative, we have that:

$$\begin{aligned} p(A)\ln\left(\frac{p(A)}{q(A)}\right) + (1-p(A))\ln\left(\frac{1-p(A)}{1-q(A)}\right) - 2[p(A)-q(A)]^2 \geq 0 \\ \implies p(A)\ln\left(\frac{p(A)}{q(A)}\right) + p(\overline{A})\ln\left(\frac{p(\overline{A})}{q(\overline{A})}\right) \geq 2[p(A)-q(A)]^2 \\ \implies KL(p,q) \geq 2[p(A)-q(A)]^2 \end{aligned}$$

As desired. To finish, we prove inequality (b) in the hint, namely:

$$\begin{aligned} & p_1 \ln\left(\frac{p_1}{q_1}\right) + p_2 \ln\left(\frac{p_2}{q_2}\right) \\ &= (p_1+p_2)\left(\frac{p_1}{p_1+p_2}\ln\left(\frac{(p_1+p_2)\frac{p_1}{(p_1+p_2)}}{(q_1+q_2)\frac{q_1}{(q_1+q_2)}}\right) + \frac{p_2}{p_1+p_2}\ln\left(\frac{(p_1+p_2)\frac{p_2}{(p_1+p_2)}}{(q_1+q_2)\frac{q_2}{(q_1+q_2)}}\right)\right) \\ &= (p_1+p_2)\ln\left(\frac{p_1+p_2}{q_1+q_2}\right)\left(1 + \frac{p_1}{p_1+p_2}\ln\left(\frac{p_1/(p_1+p_2)}{q_1/(q_1+q_2)}\right) + \frac{p_2}{p_1+p_2}\ln\left(\frac{p_2/(p_1+p_2)}{q_2/(q_1+q_2)}\right)\right) \\ &= (p_1+p_2)\ln\left(\frac{p_1+p_2}{q_1+q_2}\right) + (p_1+p_2)\underbrace{\left(\frac{p_1}{p_1+p_2}\ln\left(\frac{p_1/(p_1+p_2)}{q_1/(q_1+q_2)}\right) + \frac{p_2}{p_1+p_2}\ln\left(\frac{p_2/(p_1+p_2)}{q_2/(q_1+q_2)}\right)\right)}_{\text{KL divergence between the distributions given that we are either in } p_1 \text{ or } p_2 \text{, so } \geq 0.} \\ &\geq (p_1+p_2)\ln\left(\frac{p_1+p_2}{q_1+q_2}\right) \end{aligned}$$

So inequality (b) holds. □

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.

   Proof. To establish a lower bound on the KL-divergence, we will use Pinsker's inequality. We will look at the event that our coin comes up heads (denoted H). Then, by Pinsker's inequality:

$$\begin{aligned} 2[RC_\epsilon(H) - RC_0(H)]^2 &\leq KL(RC_\epsilon, RC_0) \\ 2\left[\frac{1+\epsilon}{2} - \frac{1}{2}\right]^2 &\leq KL(RC_\epsilon, RC_0) \\ \frac{\epsilon^2}{2} &\leq KL(RC_\epsilon, RC_0) \end{aligned}$$

Thus, $KL(RC_\epsilon, RC_0) = \Omega(\epsilon^2)$. To establish an upper bound, we plug into the definition of KL divergence:

$$\begin{aligned} KL(RC_\epsilon, RC_0) &= \frac{1+\epsilon}{2}\log\left(\frac{\frac{1+\epsilon}{2}}{1/2}\right) + \frac{1-\epsilon}{2}\log\left(\frac{\frac{1-\epsilon}{2}}{1/2}\right) \\ &= \frac{1+\epsilon}{2}\log\left(1+\epsilon\right) + \frac{1-\epsilon}{2}\log\left(1-\epsilon\right) \end{aligned}$$

We will use the taylor expansion: when ϵ is close to 0 ($\epsilon \leq 1$), $\log(1-\epsilon) = -\epsilon - \epsilon^2/2 - O(\epsilon^3)$ and $\log(1+\epsilon) = \epsilon - \epsilon^2/2 + O(\epsilon^3)$. Plugging back in:

$$\begin{aligned} KL(RC_\epsilon, RC_0) &= \frac{1+\epsilon}{2}\log\left(1+\epsilon\right) + \frac{1-\epsilon}{2}\log\left(1-\epsilon\right) \\ &= \frac{1+\epsilon}{2}\left(\epsilon^2 - \frac{\epsilon^2}{2} + O(\epsilon^3)\right) + \frac{1-\epsilon}{2}\left(-\epsilon^2 - \frac{\epsilon^2}{2} - O(\epsilon^3)\right) \\ &= \frac{1}{2}(\epsilon^2 + O(\epsilon^4)) = O(\epsilon^2) \end{aligned}$$

Since $KL(RC_\epsilon, RC_0) = \Omega(\epsilon^2)$ and $KL(RC_\epsilon, RC_0) = O(\epsilon^2)$, $KL(RC_\epsilon, RC_0) = \Theta(\epsilon^2)$. □

2. (**10 points**) $KL(RC_0, RC_\epsilon) = \Theta(\epsilon^2)$ for any $\epsilon \in (0, 1/4)$.

Proof. To get a lower bound, we use Pinsker's inequality (for the event H of getting heads):

$$\begin{aligned} 2[RC_0(H) - RC_\epsilon(H)]^2 &\leq KL(RC_\epsilon, RC_0) \\ 2\left[\frac{1}{2} - \frac{1+\epsilon}{2}\right]^2 &\leq KL(RC_\epsilon, RC_0) \\ \frac{\epsilon^2}{2} &\leq KL(RC_\epsilon, RC_0) \end{aligned}$$

For our upper bound, by plugging into the definition of KL divergence:

$$\begin{aligned} KL(RC_0, RC_\epsilon) &= \frac{1}{2}\log\left(\frac{1/2}{\frac{1+\epsilon}{2}}\right) + \frac{1}{2}\log\left(\frac{1/2}{\frac{1-\epsilon}{2}}\right) \\ &= \frac{1}{2}\left[\log\left(\frac{1}{1+\epsilon}\right) + \log\left(\frac{1}{1-\epsilon}\right)\right] \\ &= -\frac{1}{2}\left[\log\left(1-\epsilon^2\right)\right] \end{aligned}$$

We will prove that, when $\epsilon \in (0, 1/4)$, $-1/2\ln(1-\epsilon^2) \leq \epsilon^2$. To prove this, we will show that the difference $\epsilon^2 - (-1/2\ln(1-\epsilon^2)) \geq 0, \forall \epsilon \in [0, 1/4)$[2]. First, observe that when $\epsilon = 0, \epsilon^2 + 1/2\ln(1-\epsilon^2) = 0$. Next we show that for the interval $\epsilon \in (0, 1/4)$, the derivative of the difference is always positive:

$$\begin{aligned} \frac{d}{d\epsilon}\left(\epsilon^2 + \frac{1}{2}\ln(1-\epsilon^2)\right) &= 2\epsilon + \frac{1}{2}\left(\frac{1}{1-\epsilon^2}\right)(-2\epsilon) \\ &= 2\epsilon - \frac{\epsilon}{1-\epsilon^2} = \epsilon\left(2 - \frac{1}{1-\epsilon^2}\right) \end{aligned}$$

[2] A taylor expansion (as used in part (a)) could be used, but this is just to highlight another technique that suffices.

Since $2 - \frac{1}{1-\epsilon^2} \geq 0$ for $\epsilon \in \left[0, \frac{1}{\sqrt{2}}\right]$, it is clear that the derivative of the difference is always positive in our range. Thus, the difference is always positive, and $-1/2\ln(1-\epsilon^2) \leq \epsilon^2$ for the range $\epsilon \in (0, 1/4)$. Because $\frac{\epsilon^2}{2} \leq KL(RC_0, RC_\epsilon) \leq \epsilon^2$ in the range $\epsilon \in (0, 1/4)$, $KL(RC_0, RC_\epsilon) = \Theta(\epsilon^2)$. □

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,[3] 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.

Proof. First, observe that it suffices to show that $\frac{2}{3}k$ arms satisfy (2) and $\frac{2}{3}k$ arms satisfy (3). Then, we can argue that if we select an arm uniformly at random and let the A denote the event that the arm drawn satisfies (2) and B denote the event that the arm satisfies (3), then:

$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &\geq 2/3 + 2/3 - P(A \cup B) \\ &= 4/3 - P(A \cup B) \\ &\geq 1/3 \end{aligned}$$

Since $P(A \cup B) \leq 1$. The probability that the arm drawn satisfies both (2) and (3) is $P(A \cup B) \geq 1/3$, which means more than $k/3$ arms satisfy both properties, so (1) is also satisfied.

Now, we show that at least $\frac{2}{3}k$ arms satisfy (2). Call the set of arms that have been pulled strictly more than $\frac{3T}{k}$ times by round T in expectation M. Suppose, for contradiction, that $|M| > \frac{k}{3}$. Then,

$$\mathbb{E}\left[\sum_{i \in M} N_i^T\right] = \sum_{i \in M} \mathbb{E}\left[N_i^T\right] > \frac{k}{3}\left(\frac{3T}{k}\right) = T$$

So we have pulled more arms than the number of rounds – contradiction.

Similarly, we show that at least $\frac{2}{3}k$ arms satisfy (3). Call the set of arms that have strictly more than $\frac{3}{k}$ probability to be pulled in round T in expectation H. Suppose, for contradiction, that $|H| > \frac{k}{3}$. Then,

$$\sum_{i \in H} \mathbf{Pr}(I^T = i) > \frac{k}{3}\left(\frac{3}{k}\right) = 1$$

violating the fact that the sum of the probabilities must sum to a value no greater than 1 – contradiction. □

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

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;
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

$$\text{Regret}_T = O(\sqrt{kT \log T}) \tag{2}$$

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.

Proof. First, for the bad case (where at least one arm does not concentrate around its true mean, which happens with probability 2δ for each of the T arms), we observe that every arm is pulled less than CT times, for some $C \in [0, 1]$. Thus in our bad case, our regret is:

$$CT \times 2T\delta = 2C$$

After plugging in for $\delta = 1/T^2$. This term will get dominated by the regret in the good case, in which the sample mean of every arm concentrates around its true mean. In the good case, we solve for Δ_i in the second lemma:

$$\begin{aligned} N_i &\leq 8\sigma^2 \frac{\log(T)}{(\Delta_i)^2} + 1 \\ \Delta_i^2(N_i - 1) &\leq 8\sigma^2 \log(T) \\ \Delta_i &\leq \sqrt{\frac{8\sigma^2 \log(T)}{N_i - 1}} \\ \Delta_i &\leq O\left(\sqrt{\frac{\log(T)}{N_i}}\right) \end{aligned}$$

Next, we plug our upper bound for Δ_i back into the regret term and use the fact that square root is a concave function (so $\sum_{i\in k}\sqrt{x_i} \le \sqrt{k\sum_{i\in k} x_i}$):

$$\begin{aligned}
\sum_{i\in k} \Delta_i N_i(T) &\le \sum_{i\in k} O\left(\sqrt{\frac{\log(T)}{N_i(T)}}\right) N_i(T) \\
&= O\left(\sqrt{\log(T)}\right) \sum_{i\in k} \sqrt{N_i(T)} \\
&\le O\left(\sqrt{\log(T)}\right) \sqrt{k\sum_{i\in k} N_i(T)} \\
&= O\left(\sqrt{\log(T)}\right) \sqrt{kT} \\
&= O\left(\sqrt{kT\log(T)}\right)
\end{aligned}$$

As desired. $\square$