Sambuddha Chakrabarti, 220947

Problem 1. Using log sum inequality, show that $KL(P,Q) \geq 0$

Solution. The KL divergence between two probability distributions P and Q is defined as:

$$KL(P,Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)},$$

The log sum inequality states that for non-negative numbers $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$:

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i \right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i},$$

with equality if and only if $\frac{a_i}{b_i}$ is constant for all i.

Let $a_i = P(x_i)$ and $b_i = Q(x_i)$. Applying the log sum inequality to the KL divergence:

$$KL(P,Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)} \geq \left(\sum_x P(x) \right) \log \frac{\sum_x P(x)}{\sum_x Q(x)}.$$

Since P and Q are probability distributions, $\sum_x P(x) = 1$ and $\sum_x Q(x) = 1$. Substituting these values:

$$KL(P,Q) \geq 1 \cdot \log \frac{1}{1} = 0.$$

□

Problem 2. Consider an algorithm that always pulls arm 1 (throughout T rounds). Give an instance on which this algorithm achieves an expected regret of 0. Also give an instance on which it achieves an expected regret of T.

Solution. For the first instance, we consider a setup with two arms, with reward of 1 for arm 1 and 0 for arm 2. Thus, an algorithm that always pulls arm 1 will have 0 regret.

For the second instance, we consider a similar setup, with arm 1 having reward of 0 and arm 2 having reward of 1. Thus, an algorithm that always pulls arm 1 will have T regret. □

Problem 3. Which of the following statement(s) is/are correct? (Justify briefly - no marks for just writing the answer.)

(a) The expected regret of the Successive Elimination algorithm is $O(\sqrt{KT \log T})$ on all instances.

(b) The expected regret of the Successive Elimination algorithm can be $\leq (KT \log T)^{1/4}$ on some instances.

(c) The expected regret of the Successive Elimination algorithm cannot be $\leq (KT \log T)^{1/4}$ on all instances.

Solution.

Statement (a) This statement is **correct**. The Successive Elimination algorithm is designed to achieve an expected regret of $O(\sqrt{KT \log T})$ on all instances. This was proved in class in the analysis of Successive Elimination algorithms.

Statement (b) This statement is **correct**. On some instances (e.g., when the gaps are very large or the instance is trivial), the expected regret of the Successive Elimination algorithm can be much smaller than $O(\sqrt{KT \log T})$ as inferior arms get eliminated much quicker. In such cases, the regret can indeed be $\leq (KT \log T)^{1/4}$.

Statement (c) This statement is **correct**. The expected regret of the Successive Elimination algorithm cannot be $\leq (KT \log T)^{1/4}$ on all instances. There exist instances (e.g., when the suboptimality gaps are very small) where the regret is lower-bounded by $o(\sqrt{KT})$, which is larger than $(KT \log T)^{1/4}$ for sufficiently large T. □

Problem 4. Assuming *Good* (as defined in UCB algorithm), show that $\mu_a \leq UCB_a(t)$ for UCB for any t

Solution. The definition of *Good* is given as

$$\mu_a \leq \hat{\mu_a}(t) + \epsilon_a(t),$$

where $\epsilon_a(t) = \sqrt{\frac{5 \log T}{n_a(t)}}$ Hence, it follows trivially that $\mu_a \leq UCB_a(t)$ □

Problem 5. Assuming *Good* (as defined in Successive Elimination algorithm), prove that the optimal arm will never be eliminated in the Succesive Elimination Algorithm.

Solution. For arm a^* to be eliminated, $UCB_{a^*}(t) < LCB_a(t)$ for some a. This implies that,

$$\hat{\mu_{a^*}}(t) + \epsilon_{a^*}(t) < \hat{\mu_a}(t) - \epsilon_a(t) \implies \hat{\mu_a}(t) - \hat{\mu_{a^*}}(t) < 2\epsilon_a(t)$$

Assuming *Good* however,

$$\mu_{a^*} - \epsilon_{a^*}(t) < \hat{\mu_{a^*}}(t) < \mu_{a^*} + \epsilon_{a^*}(t)$$

$$\mu_a - \epsilon_a(t) < \hat{\mu_a}(t) < \mu_a + \epsilon_a(t)$$

, This implies,

$$\mu_{a^*} < \mu_a$$

, which is a contradiction. Hence, the best arm is never eliminated. □

Problem 6. In the class we show a lower bound of $\Omega(\sqrt{KT})$ on the expected regret of any algorithm. But we also see that the UCB can achieve an instance dependent upper bound of $O(logT)\Sigma_{a:\Delta_a>0}\frac{1}{\Delta_a}$. Explain why they do not contradict each other.

Solution. The two bounds do not contradict each other, as $\Omega(\sqrt{KT})$ applies to any instance and holds for the hardest possible problem instance. Thus it ensures the performance regardless of the gaps Δ_a. The instance dependent upper bound of $O(logT)\Sigma_{a:\Delta_a>0}\frac{1}{\Delta_a}$ applies to only a fixed instance with gaps Δ_a. If the gaps are small, the regret might be large, potentially approaching the bound of $\Omega(\sqrt{KT})$, i.e. the worst case instance. □

Problem 7. Let $P = (p, 1-p)$ and $Q = (q, 1-q)$ be two distributions on two elements. Show

$$d_{\mathrm{tv}}(P^{\otimes 2}, Q^{\otimes 2}) \leq 2\, d_{\mathrm{tv}}(P, Q)$$

Solution. For any two distributions P and Q,

$$d_{tv}(P, Q) = \frac{1}{2}\Sigma_{\omega\in\Omega}|P(\omega) - Q(\omega)|$$

Also, $P^{\otimes 2}(\omega) = P(\omega)P(\omega)$ and same for Q. Thus,

$$d_{tv}(P^{\otimes 2}, Q^{\otimes 2}) = \frac{1}{2}\Sigma_{\omega\in\Omega}|P(\omega)^2 - Q(\omega)^2| = \frac{1}{2}\Sigma_{\omega\in\Omega}|P(\omega) - Q(\omega)||P(\omega) + Q(\omega)|$$

and we observe $|P(\omega) + Q(\omega)| \leq 2$, thus,

$$d_{tv}(P^{\otimes 2}, Q^{\otimes 2}) = \frac{1}{2}\Sigma_{\omega\in\Omega}|P(\omega) - Q(\omega)||P(\omega) + Q(\omega)| \leq \Sigma_{\omega\in\Omega}|P(\omega) - Q(\omega)| = 2\, d_{\mathrm{tv}}(P, Q)$$

□

Problem 8. Show that for any n,

$$\mathrm{KL}(P^{\otimes n}, Q^{\otimes n}) = n \cdot \mathrm{KL}(P, Q).$$

Solution. Using the property of logarithms, $\log\left(\prod_{i=1}^{n}\frac{P(\omega)}{Q(\omega)}\right) = \sum_{i=1}^{n}\log\frac{P(\omega)}{Q(\omega)}$, we can rewrite the KL divergence as:

$$\mathrm{KL}(P^{\otimes n}, Q^{\otimes n}) = \sum_{\omega\in\Omega}\left(\prod_{i=1}^{n} P(\omega)\right)\left(\sum_{i=1}^{n}\log\frac{P(\omega)}{Q(\omega)}\right).$$

By linearity of expectation and the independence of the ω, this simplifies to:

$$\mathrm{KL}(P^{\otimes n}, Q^{\otimes n}) = \sum_{i=1}^{n}\sum_{\omega\in\Omega} P(\omega)\log\frac{P(\omega)}{Q(\omega)}.$$

Since each term in the sum is identical and equal to $\mathrm{KL}(P, Q)$, we have:

$$\mathrm{KL}(P^{\otimes n}, Q^{\otimes n}) = n \cdot \mathrm{KL}(P, Q).$$

□

Problem 9. In the first assignment, you showed that Testing Coin$(\epsilon, 1/5)$ problem can solved in $O(\frac{1}{\epsilon^2})$ coin tosses. Show that Testing Coin(ϵ, δ) can be solved in $O(\frac{\log\frac{1}{\delta}}{\epsilon^2})$ coin tosses for any $\delta > 0$.

Solution. In the first assignment, it was shown that the Testing Coin$(\epsilon, 1/5)$ problem can be solved in $O(1/\epsilon^2)$ coin tosses. We now extend this result to show that Testing Coin(ϵ, δ) can be solved in $O\left(\frac{\log 1/\delta}{\epsilon^2}\right)$ coin tosses for any $\delta > 0$.

The key idea is to use the Chernoff bound to control the probability of error. Given a fair coin, we perform N independent coin tosses and estimate the empirical bias. The number of required samples N is determined by ensuring that the probability of deviation from the true bias is at most δ.

Using standard Chernoff bounds, the number of samples required to distinguish between two distributions within error ϵ with probability at least $1-\delta$ satisfies:

$$N = O\left(\frac{\log 1/\delta}{\epsilon^2}\right)$$

Thus, we conclude that Testing Coin(ϵ, δ) can indeed be solved in $O\left(\frac{\log 1/\delta}{\epsilon^2}\right)$ coin tosses.
□

Problem 10. In the class, we saw a lower bound of $\frac{(1-2\delta)^2}{\epsilon^2}$ on coin tosses for deterministic algorithms for TestingCoin(ϵ, δ) problem. The above question shows, the upper bound is $O\left(\frac{\log 1/\delta}{\epsilon^2}\right)$. Clearly, the upper and lower bounds are tight in terms of ϵ but not on δ. Show that the lower bound is also $O\left(\frac{\log 1/\delta}{\epsilon^2}\right)$ for deterministic algorithms. You can assume $\epsilon \leq 1/3$ (in fact, if you want, you can assume both $\epsilon, \delta \leq c$ for any constant $c < 1$ of your choice).

Solution. We now show that the lower bound is also $\Omega\left(\frac{\log 1/\delta}{\epsilon^2}\right)$ for deterministic algorithms. We assume $\epsilon \leq 1/3$.

Using the inequality:

$$d_{\text{tv}}(P, Q) \leq \sqrt{1 - e^{-KL(P,Q)}}$$

,we relate total variation distance to the Kullback-Leibler divergence. The KL divergence for two Bernoulli distributions P and Q satisfies:

$$KL(P, Q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}. \tag{1}$$

For small ϵ, we approximate:

$$KL(P, Q) \approx \frac{(p-q)^2}{q(1-q)} \approx \frac{\epsilon^2}{p(1-p)}. \tag{2}$$

By substituting into the bound on total variation distance, we obtain:

$$d_{\text{tv}}(P, Q) \leq \sqrt{1 - e^{-\frac{\epsilon^2}{p(1-p)}}} \approx 2\epsilon^2. \tag{3}$$

To achieve error probability at most δ, we require a number of samples N satisfying:

$$N \cdot d_{\text{tv}}(P, Q) \geq \Omega(\log 1/\delta). \tag{4}$$

Since $d_{\text{tv}}(P, Q) \approx 2\epsilon^2$, rearranging gives:

$$N \geq \Omega\left(\frac{\log 1/\delta}{\epsilon^2}\right). \tag{5}$$

Thus, we establish the required lower bound, completing the proof. □

Problem 11. We consider a multi-armed bandit problem with K arms, where the mean reward of each arm lies in the range $[1/2, 1/2+\gamma]$ for some known $\gamma > 0$. Our goal is to modify the Successive Elimination and UCB algorithms to achieve improved regret bounds by leveraging the knowledge of γ.

Solution. For Successive Elimination, we assume that $\epsilon(t) = \gamma\sqrt{\frac{c \log T}{n_a(t)}}$, as we can get away with keeping a narrow margin for this case as the means are all in the range $[1/2, 1/2+\gamma]$. The algorithm is given as,

1. **Initialization:** Pull each arm once.
2. **Iteration:** At round t, for each active arm a:
   - Compute its empirical mean:
   $$\hat{\mu}_a = \frac{1}{n_a(t)} \sum_{i=1}^{n_a(t)} X_i,$$
   - Compute $UCB_a(t)$:
   $$UCB_a(t) = \hat{\mu}_i + \epsilon_a(t)$$
   - Compute $LCB_a(t)$:
   $$UCB_a(t) = \hat{\mu}_i - \epsilon_a(t)$$
   - Eliminate arm i if there exists an arm j such that:
   $$UCB_i(t) < LCB_j(t)$$
3. Continue playing the remaining arms until T runs out.

This will result in $E[R(T)|Good] = \sqrt{\gamma KT \log T}$. Since we know the value of γ, we can choose c such that the $\Pr(Bad_a)$ is small, i.e.

$$\Pr(Bad_a) \leq \frac{2}{e^{2c\gamma^2 \log T}} = O(\frac{1}{T^{2c\gamma^2}}) \approx O(\frac{1}{T^{10}}),$$

for $c = \frac{5}{\gamma^2}$ Thus, $E[R(T)] = O(\frac{1}{T^8}) + E[R(T)|Good] = \sqrt{\gamma KT \log T}$ Similarly, for the instance dependent regret being $R(T) = O(\log T)\Sigma_a \frac{\gamma}{\Delta_a}$. The regret gets scaled down by γ. The UCB algorithm is also defined similarly, with $\epsilon(t) = \gamma\sqrt{\frac{c \log T}{n_a(t)}}$

1. **Initialization:** Pull each arm once.

2. **At round** t:

   - Compute:
   $$UCB_a(t) = \hat{\mu}_a(t) + \epsilon_a(t).$$
   - Play the arm with the highest $UCB_a(t)$.

Therefore, we choose $c = \frac{5}{\gamma^2}$ Thus, $E[R(T)] = O(\frac{1}{T^8}) + E[R(T)|Good] = \sqrt{\gamma KT \log T}$ Similarly, for the instance dependent regret being $R(T) = O(\log T)\Sigma_a \frac{\gamma}{\Delta_a}$. The regret gets scaled down by γ. $\square$

Problem 12. Consider a following problem : input consists of K coins and $\epsilon > 0$. It is promised that either
• all coins are fair or
• there is exactly one coin which is biased and have $\Pr(H) = 1/2 + \epsilon$ and rest other coins are fair.

The goal is to correctly determine the one of the above possibility with at least 4/5 probability. Modify the lower bound proof for Biased Coin Identification problem to show the same lower bound (of $\Omega(\frac{K}{\epsilon^2})$) for the above problem.

Solution. From earlier instances, we know to identify a single coin with the required probability requires $\Omega(\frac{1}{\epsilon^2})$. Thus, from information theoretic limits, since we need to repeat the same experiment K times in the worst case (the biased coin is the last tested), we require a worst case lower bound of $\Omega(\frac{K}{\epsilon^2})$. $\square$

Problem 13. Let $P = (p, 1-p)$ and $Q = (q, 1-q)$ be two binary distributions (distributions on two elements). Show

$$d_{\text{tv}}(P, Q) \leq \sqrt{\frac{1}{2} KL(P, Q)}$$

Solution. The total variation distance between P and Q is defined as:

$$d_{\text{tv}}(P, Q) = \frac{1}{2} \sum_x |P(x) - Q(x)| = |p - q|$$

The KL divergence is given as

$$KL(P, Q) = p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q}$$

We will try to show that

$$2(p - q)^2 \leq p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q}$$

In this case, we introduce the function $f : (0, 1) \to R$ defined by $f(x) = p \log x + (1 - p) \log(1 - x)$. The RHS is exactly $f(p) - f(q)$. Thus,

$$f(p) - f(q) = \int_q^p f'(x)dx = \int_q^p \frac{p - x}{x(1 - x)}dx \geq 4 \int_q^p (p - x)dx = 2(p - q)^2$$

which is the required result. $\square$

Problem 14. Consider two distributions $P = (p_1, ..., p_n)$ and $Q = (q_1, ..., q_n)$ on $[n]$. Consider any set $S \subseteq [n]$. Let $P' = (P(S), 1-P(S))$ and $Q' = (Q(S), 1-Q(S))$, be binary distributions where recall that $P(S) = \Sigma_{i \in S} p_i$ and $Q(S) = \Sigma_{i \in S} q_i$. Show $KL(P', Q') \leq KL(P, Q)$

Solution. The Kullback-Leibler (KL) divergence between P and Q is:

$$KL(P, Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}.$$

Similarly, the KL divergence between P' and Q' is:

$$KL(P', Q') = P(S) \log \frac{P(S)}{Q(S)} + (1 - P(S)) \log \frac{1 - P(S)}{1 - Q(S)}$$

By the log-sum inequality, which states that for positive numbers a_i and b_i:

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \log \frac{\sum_i a_i}{\sum_i b_i}$$

we apply this to the KL divergence sum:

$$\sum_{i \in S} p_i \log \frac{p_i}{q_i} + \sum_{i \notin S} p_i \log \frac{p_i}{q_i} \geq P(S) \log \frac{P(S)}{Q(S)} + (1 - P(S)) \log \frac{1 - P(S)}{1 - Q(S)}$$

Thus, we conclude:

$$KL(P', Q') \leq KL(P, Q)$$

This result follows from the data processing inequality (DPI), which states that applying a function (such as marginalization) cannot increase the KL divergence. This completes the proof. □

Problem 15. Consider two distributions $P = (p_1, ..., p_n)$ and $Q = (q_1, ..., q_n)$ on $[n]$. Show that

$$d_{\text{tv}}(P, Q) \leq \sqrt{\frac{1}{2} KL(P, Q)}$$

Solution. As proved earlier, $KL(P', Q') \leq KL(P, Q)$ and for binomial distributions P' and Q', we have $d_{\text{tv}}(P', Q') \leq \sqrt{\frac{1}{2} KL(P', Q')}$. Thus the given problem can be reduced to the binary case for Pinsker's as proved earlier in the previous two parts □

Problem 16. Proof of $E[\delta] \leq O(\sqrt{\frac{K}{T}}$ for the MOSS algorithm

Solution. We first try to prove

$$\Pr(\delta \geq y) \leq \frac{K}{T} O(\frac{1}{y^2})$$

Using Hoeffding's maximal inequality, for any $t > 0$,

$$\Pr\left(\exists 1 \leq r \leq m : \sum_{i=1}^{r}(\mu - X_i) \geq t\right) \leq \exp\left(-\frac{2t^2}{m}\right).$$

To apply this, we consider:

$$\Pr\left(\exists 1 \leq x \leq T : \mu - \hat{\mu}_x \geq \sqrt{\frac{\log^+(T/Kx)}{x}} + y\right) \leq \sum_j \Pr\left(\exists 2^j \leq x \leq 2^{j+1} : \sum_{i=1}^{x} \mu - \hat{\mu}_x \geq y\right)$$

This term reduces to

$$\sum_j \Pr\left(\exists 2^j \leq x \leq 2^{j+1} : \sum_{i=1}^{x} \mu - \hat{\mu}_x \geq y\right) \leq \sum_j 2^j \exp(-2^j y^2) = O\left(\frac{1}{y^2}\right)$$

For each of T rounds and all of K arms, this probability is written as

$$\Pr(\delta \geq y) \leq \frac{K}{T} O(\frac{1}{y^2})$$

Now, we define a discrete r.v. β as

$$\beta = 2^j \text{ for } 2^j \leq \delta < 2^{j+1}$$

Thus, the expectation of δ is bounded by:

$$\mathbb{E}[\delta] \leq \mathbb{E}[\beta] = \sum_j 2^j \Pr(\delta \geq 2^j).$$

Using the probability bound:

$$\Pr(\delta \geq 2^j) \leq \frac{K}{T} \cdot O\left(\frac{1}{(2^j)^2}\right).$$

Substituting this into the expectation sum:

$$\mathbb{E}[\beta] \leq \sum_j 2^j \cdot \frac{K}{T} \cdot O\left(\frac{1}{(2^j)^2}\right).$$

Simplifying:

$$\mathbb{E}[\beta] = O\left(\frac{K}{T} \sum_j \frac{1}{2^j}\right).$$

Since the summation $\sum_j \frac{1}{2^j}$ is a geometric series converging to $O(1)$, we obtain:

$$\mathbb{E}[\delta] \leq O\left(\frac{K}{T}\right).$$

(I am not sure of this answer and could only prove till this much) $\square$