Assignmnet 2

Rohan (241110057)

(1 marks each):

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

   Solution:

   The Kullback-Leibler (KL) divergence between two probability distributions P and Q over a discrete sample space X is defined as:

   $$KL(P, Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)}$$

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

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

   In the context of probability distributions $P(x)$ and $Q(x)$, the log-sum inequality becomes:

   $$\sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)} \geq \left(\sum_{x \in X} P(x) \right) \log \frac{\sum_{x \in X} P(x)}{\sum_{x \in X} Q(x)}$$

   Since $P(x)$ and $Q(x)$ are probability mass functions:

   $$\sum_{x \in X} P(x) = 1 \quad \text{and} \quad \sum_{x \in X} Q(x) = 1$$

   Thus, the right-hand side of the inequality simplifies to:

   $$1 \cdot \log \frac{1}{1} = \log 1 = 0$$

   Therefore, applying the log-sum inequality:

   $$KL(P, Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)} \geq 0$$

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:

   Regret is defined as the difference between the expected reward of the optimal arm and the expected reward of the chosen arm:

   $$R(T) = T\mu^* - \sum_{t=1}^{T} \mu_{a_t}$$

   where:

- μ^* is the expected reward of the best arm.
- μ_{a_t} is the expected reward of the arm chosen at time t.
- T is the total number of rounds.

Case 1: For regret to be 0, the algorithm must always select the optimal arm. This happens when arm 1 is the best arm.

Let the reward distributions be: $\mu_1 = 1, \quad \mu_2, \mu_3, \ldots, \mu_K \leq 1$.

The regret is:

$$R(T) = T(1) - \sum_{t=1}^{T} 1 = 0$$

Case 2: For regret to be T, the algorithm must always pull a suboptimal arm while another arm is strictly better.

Let the reward distributions be: $\mu_1 = 0, \quad \mu_2 = 1$.

Since the algorithm always pulls arm 1, and arm 1 has an expected reward of 0 while the optimal arm (arm 2) has an expected reward of 1, the regret is:

$$R(T) = T(1) - \sum_{t=1}^{T} 0 = T$$

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

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

Solution: True

The Successive Elimination algorithm adaptively eliminates suboptimal arms by maintaining confidence intervals and removing arms that are statistically unlikely to be optimal.

A well-known bound on the expected regret of SE is:

$$E[R(T)] = O\left(\sqrt{KT \log T}\right)$$

This bound holds for all instances because:

- Successive Elimination performs *optimistic exploration*, ensuring that all arms are pulled sufficiently before elimination.
- The regret accumulates over successive elimination phases, leading to a worst-case bound of $O(\sqrt{KT \log T})$.

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

Solution: False

The Successive Elimination (SE) algorithm has a worst-case expected regret bound of $O(\sqrt{KT \log T})$ which holds for all instances. While the regret can be smaller than this worst-case bound on "easy" instances (where the suboptimality gaps Δ_i are large), it does not achieve a regret as low as $O((KT \log T)^{1/4})$ on any instance.

1. **Worst-case regret**: The worst-case regret of SE is $O(\sqrt{KT \log T})$ and this bound is tight in the sense that there exist instances where the regret matches this scaling.

2. **Instance-dependent regret:** On easy instances where the suboptimality gaps Δ_i are large, the regret of SE can be much smaller than $O(\sqrt{KT \log T})$. However, the regret in such cases is typically expressed in terms of the gaps Δ_i, not in terms of K and T alone. Specifically, the regret on easy instances scales as:

$$E[R(T)] = O\left(\sum_{i \neq i^*} \frac{\log T}{\Delta_i}\right)$$

where i^* is the optimal arm. This bound can be much smaller than $O(\sqrt{KT \log T})$ when the gaps Δ_i are large, but it does not scale as $O((KT \log T)^{1/4})$.

3. **No $O((KT \log T)^{1/4})$ Regret Guarantee:** There is no theoretical justification or known result that guarantees the regret of SE can be as low as $O((KT \log T)^{1/4})$ on any instance. The $O((KT \log T)^{1/4})$ bound does not align with the known regret bounds for SE or other standard bandit algorithms.

(c) The expected regret of Successive Elimination algorithm can not be $\leq (KTlogT)^{1/4}$ on all instances.

Solution: True

This means that *there exist instances where the regret is at least* $O(\sqrt{KT \log T})$, which is indeed correct.

- In **hard instances**, where the gap Δ between the best and suboptimal arms is very small, it takes longer to eliminate suboptimal arms.
- In **worst-case instances**, Successive Elimination must explore all arms extensively before distinguishing the optimal arm, leading to regret of $O(\sqrt{KT \log T})$.

Thus, it is **not possible** to achieve $O((KT \log T)^{1/4})$ regret on all instances.

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

Solution:

UCB Algorithm:

$$\text{UCB}_a(t) = \hat{\mu}_a(t) + \sqrt{\frac{2 \log t}{N_a(t)}}$$

"Good" Event:

$$|\hat{\mu}_a(t) - \mu_a| \leq \sqrt{\frac{2 \log t}{N_a(t)}}$$

Rearranging the inequality from the "Good" event:

$$\mu_a \leq \hat{\mu}_a(t) + \sqrt{\frac{2 \log t}{N_a(t)}}$$

Since the right-hand side is exactly the definition of $\text{UCB}_a(t)$, we obtain:

$$\mu_a \leq \text{UCB}_a(t)$$

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

Solution:

Let:

- μ^* be the expected reward of the optimal arm.
- $\hat{\mu}_a(t)$ be the empirical mean of arm a at time t.
- $N_a(t)$ be the number of times arm a has been played up to time t.

The confidence interval around the empirical mean of an arm a is given by:

$$\mathrm{CI}_a(t) = \sqrt{\frac{2\log T}{N_a(t)}}$$

Thus, the confidence bounds for arm a are:

- Upper Confidence Bound (UCB): $\mathrm{UCB}_a(t) = \hat{\mu}_a(t) + \mathrm{CI}_a(t)$
- Lower Confidence Bound (LCB): $\mathrm{LCB}_a(t) = \hat{\mu}_a(t) - \mathrm{CI}_a(t)$

The SE algorithm eliminates an arm a if there exists another arm b such that:

$$\mathrm{UCB}_a(t) < \mathrm{LCB}_b(t)$$

The "Good" Event Assumption:

The "Good" event ensures that for all arms a and all times t:

$$|\hat{\mu}_a(t) - \mu_a| \leq \mathrm{CI}_a(t)$$

This implies:

$$\mu_a - \mathrm{CI}_a(t) \leq \hat{\mu}_a(t) \leq \mu_a + \mathrm{CI}_a(t)$$

So, the **true mean** of any arm always lies within its confidence bounds:

$$\mathrm{LCB}_a(t) \leq \mu_a \leq \mathrm{UCB}_a(t)$$

Applying this to the optimal arm a^* with mean μ^*:

$$\mathrm{LCB}_{a^*}(t) \leq \mu^* \leq \mathrm{UCB}_{a^*}(t)$$

Since the optimal arm has the highest true mean, for any suboptimal arm a:

$$\mu_a < \mu^*$$

For the optimal arm a^* to be **eliminated**, there must exist some suboptimal arm a such that:

$$\mathrm{UCB}_{a^*}(t) < \mathrm{LCB}_a(t)$$

However, using the bounds derived from the "Good" event:

$$\mu^* \leq \mathrm{UCB}_{a^*}(t), \quad \text{and} \quad \mathrm{LCB}_a(t) \leq \mu_a$$

Since $\mu_a < \mu^*$:

$$\mathrm{LCB}_a(t) \leq \mu_a < \mu^* \leq \mathrm{UCB}_{a^*}(t)$$

Thus, it is **impossible** for $\mathrm{UCB}_{a^*}(t) < \mathrm{LCB}_a(t)$ to ever hold, meaning that the optimal arm is **never eliminated**.

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 $O(logT)\sum_{a:\Delta_a>0}\frac{1}{\Delta_a}$. Explain why they do not contradict each other.**

Solution:

(a) **The Lower Bound is a Worst-Case Guarantee:**
- The $\Omega(\sqrt{KT})$ bound does not state that every instance has this regret.
- It only ensures that **some hard instances** exist where any algorithm must suffer at least this much regret.

(b) **The Upper Bound is Instance-Dependent:**
- The bound $O\left(\sum_{a:\Delta_a>0} \frac{\log T}{\Delta_a}\right)$ holds only for certain problem instances.
- In easy instances, where Δ_a values are large, UCB eliminates suboptimal arms quickly and achieves much lower regret.

(c) **There Exists a Regime Where the Two Bounds Match:**
- If the gaps Δ_a are very small, then $\frac{\log T}{\Delta_a}$ becomes large.
- In such cases, the instance-dependent bound of UCB approaches $O(\sqrt{KT})$, matching the worst-case bound.

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

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

Solution:

The Total Variation (TV) distance between two probability distributions P and Q over a sample space $\mathcal{X}$ is defined as:

$$d_{tv}(P, Q) = \frac{1}{2} \sum_{x \in \mathcal{X}} |P(x) - Q(x)|$$

For the given distributions $P = (p, 1-p)$ and $Q = (q, 1-q)$:

$$d_{tv}(P, Q) = \frac{1}{2} \left(|p - q| + |(1-p) - (1-q)|\right) = |p - q|$$

The probability mass functions of the product distributions $P^{\otimes 2}$ and $Q^{\otimes 2}$ are given by:

$$P^{\otimes 2}(x_1, x_2) = P(x_1)P(x_2), \quad Q^{\otimes 2}(x_1, x_2) = Q(x_1)Q(x_2)$$

Explicitly, the probabilities for each outcome $(x_1, x_2) \in \{0,1\}^2$ under $P^{\otimes 2}$ and $Q^{\otimes 2}$ are:

$$P^{\otimes 2}(0,0) = p^2, \quad P^{\otimes 2}(0,1) = p(1-p), \quad P^{\otimes 2}(1,0) = (1-p)p, \quad P^{\otimes 2}(1,1) = (1-p)^2$$

$$Q^{\otimes 2}(0,0) = q^2, \quad Q^{\otimes 2}(0,1) = q(1-q), \quad Q^{\otimes 2}(1,0) = (1-q)q, \quad Q^{\otimes 2}(1,1) = (1-q)^2$$

Computing $d_{tv}(P^{\otimes 2}, Q^{\otimes 2})$:
By definition of TV distance:

$$d_{tv}(P^{\otimes 2}, Q^{\otimes 2}) = \frac{1}{2} \sum_{(x_1,x_2) \in \{0,1\}^2} \left|P^{\otimes 2}(x_1, x_2) - Q^{\otimes 2}(x_1, x_2)\right|$$

Expanding:

$$d_{tv}(P^{\otimes 2}, Q^{\otimes 2}) = \frac{1}{2}\Big(|p^2 - q^2| + |p(1-p) - q(1-q)| + |(1-p)p - (1-q)q| + |(1-p)^2 - (1-q)^2|\Big)$$

Using the identity $a^2 - b^2 = (a-b)(a+b)$:

$$|p^2 - q^2| = |p - q|(p + q), \quad |(1-p)^2 - (1-q)^2| = |p - q|(2 - p - q)$$

For the middle terms:

$$|p(1-p) - q(1-q)| = |p-q|(1-(p+q-pq)), \quad |(1-p)p - (1-q)q| = |p-q|(1-(p+q-pq))$$

Summing up:

$$d_{tv}(P^{\otimes 2}, Q^{\otimes 2}) = \frac{1}{2}|p-q|\Big((p+q) + (2-p-q) + 2(1-(p+q-pq))\Big)$$

Simplifying:

$$d_{tv}(P^{\otimes 2}, Q^{\otimes 2}) = |p-q| \cdot \frac{4-2(p+q-pq)}{2} = |p-q|(2-(p+q-pq))$$

Since $2-(p+q-pq) \leq 2$:

$$d_{tv}(P^{\otimes 2}, Q^{\otimes 2}) \leq 2|p-q|$$

Since $d_{tv}(P,Q) = |p-q|$:

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

8. **Show for any n,**

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

Solution:

For the product distributions $P^{\otimes n}$ and $Q^{\otimes n}$, the KL divergence is defined as:

$$KL(P^{\otimes n}, Q^{\otimes n}) = \sum_{x_1,x_2,\ldots,x_n} P^{\otimes n}(x_1,x_2,\ldots,x_n) \log \frac{P^{\otimes n}(x_1,x_2,\ldots,x_n)}{Q^{\otimes n}(x_1,x_2,\ldots,x_n)}$$

Since $P^{\otimes n}$ and $Q^{\otimes n}$ are independent product distributions:

$$P^{\otimes n}(x_1,x_2,\ldots,x_n) = P(x_1)P(x_2)\cdots P(x_n)$$

$$Q^{\otimes n}(x_1,x_2,\ldots,x_n) = Q(x_1)Q(x_2)\cdots Q(x_n)$$

So, the KL divergence simplifies to:

$$KL(P^{\otimes n}, Q^{\otimes n}) = \sum_{x_1,x_2,\ldots,x_n} P(x_1)P(x_2)\cdots P(x_n) \log \frac{P(x_1)P(x_2)\cdots P(x_n)}{Q(x_1)Q(x_2)\cdots Q(x_n)}$$

Using the logarithm property $\log(ab) = \log a + \log b$, we expand the log term:

$$\log \frac{P(x_1)P(x_2)\cdots P(x_n)}{Q(x_1)Q(x_2)\cdots Q(x_n)} = \sum_{i=1}^{n} \log \frac{P(x_i)}{Q(x_i)}$$

Now the KL divergence becomes:

$$KL(P^{\otimes n}, Q^{\otimes n}) = \sum_{x_1,x_2,\ldots,x_n} P(x_1)P(x_2)\cdots P(x_n) \sum_{i=1}^{n} \log \frac{P(x_i)}{Q(x_i)}$$

Since summation over product distributions can be separated, so:

$$KL(P^{\otimes n}, Q^{\otimes n}) = \sum_{i=1}^{n} \sum_{x_1,x_2,\ldots,x_n} P(x_1)P(x_2)\cdots P(x_n) \log \frac{P(x_i)}{Q(x_i)}$$

Since each x_i is independent and follows distribution $P(x_i)$, summing over all other x_j for $j \neq i$ results in:

$$KL(P^{\otimes n}, Q^{\otimes n}) = \sum_{i=1}^{n} \sum_{x_i} P(x_i) \log \frac{P(x_i)}{Q(x_i)}$$

Recognizing that the inner sum is just the definition of KL divergence for a single sample:

$$KL(P^{\otimes n}, Q^{\otimes n}) = \sum_{i=1}^{n} KL(P, Q)$$

Since there are n identical terms:

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

(2-mark)

1. **Prove for distributions P and Q defined on $[n]$,**

$$d_{tv}(P, Q) = \underset{S \subseteq [n]}{max} |P(S) - Q(S)|$$

Solution:

By the definition of TV distance:

$$d_{tv}(P, Q) = \frac{1}{2} \sum_{x \in [n]} |P(x) - Q(x)|$$

Let the subset S^+ as the set of elements where $P(x) \geq Q(x)$:

$$S^+ = \{x \in [n] \mid P(x) \geq Q(x)\}$$

Summing over S^+:

$$\sum_{x \in S^+} (P(x) - Q(x)) = \frac{1}{2} \sum_{x \in [n]} |P(x) - Q(x)|$$

Similarly, summing over the complementary set $S^- = [n] \setminus S^+$:

$$\sum_{x \in S^-} (Q(x) - P(x)) = \frac{1}{2} \sum_{x \in [n]} |P(x) - Q(x)|$$

Taking the maximum over all subsets S:

$$d_{tv}(P, Q) \leq \max_{S \subseteq [n]} |P(S) - Q(S)|$$

The above equation is eq-1.

Now, consider any subset $S \subseteq [n]$:

$$|P(S) - Q(S)| = \left| \sum_{x \in S} P(x) - Q(x) \right|$$

Using the triangle inequality:

$$|P(S) - Q(S)| \leq \sum_{x \in S} |P(x) - Q(x)|$$

Taking the maximum over all subsets S:

$$\max_{S\subseteq[n]} |P(S) - Q(S)| \leq \sum_{x\in[n]} |P(x) - Q(x)|$$

Since the TV distance is defined as half of this sum:

$$\max_{S\subseteq[n]} |P(S) - Q(S)| \leq 2d_{tv}(P, Q)$$

Dividing by 2:

$$\max_{S\subseteq[n]} |P(S) - Q(S)| \leq d_{tv}(P, Q)$$

The above equation is eq-2.

From eq-1 and eq-2:

$$d_{tv}(P, Q) = \max_{S\subseteq[n]} |P(S) - Q(S)|$$

(1-mark)

1. **In the first assignment, you showed that Testing Coin**$(\epsilon, 1/5)$ **problem can solved in** $O(1/\epsilon^2)$ **coin tosses. Show that Testing Coin** (ϵ, δ) **can be solved in** $O(\frac{log1/\delta}{\epsilon^2})$ **coin tosses for any** $\delta > 0$**.**

   Solution:

   We will generalize the result from first assignment to an arbitrary **failure probability** δ.

   Using Hoeffding's inequality, for independent Bernoulli trials $X_1, X_2, \ldots, X_n$, the empirical mean:

$$\hat{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

   satisfies the concentration bound:

$$P(|\hat{X} - E[X]| \geq \epsilon) \leq 2\exp\left(-2n\epsilon^2\right)$$

   Setting the right-hand side to be at most δ:

$$2\exp\left(-2n\epsilon^2\right) \leq \delta$$

   Taking the natural logarithm:

$$-2n\epsilon^2 \leq \log\frac{\delta}{2}$$

   Rearranging for n:

$$n \geq \frac{\log(2/\delta)}{2\epsilon^2}$$

   Since we are interested in asymptotic complexity:

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

(3-mark)

1. **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 $O(\frac{log 1/\delta}{\epsilon^2})$. Clearly, the upper and lower bounds are tight in terms of ϵ but not on δ. Show that the lower bound is also $\Omega(\frac{log 1/\delta}{\epsilon^2})$ for deterministic algorithms. You can assume $\epsilon \le 1/3$ (in fact, if you want, you can assume both $\epsilon, \delta \le c$ for any constant $c < 1$ of your choice). (Hint: Instead of Pinsker Inequality, use the following stronger bound**

$$d_{tv}(P, Q) \le \sqrt{1 - \epsilon^{-KL(P,Q)}}$$

)

Solution:

For the Testing Coin(ϵ, δ) problem:

- P corresponds to the distribution of coin tosses from a fair coin ($p = \frac{1}{2}$)
- Q corresponds to the distribution of coin tosses from a biased coin ($p = \frac{1}{2} + \epsilon$)

The KL divergence between these two distributions over n independent tosses is:

$$KL(P^{\otimes n}, Q^{\otimes n}) = nKL(P, Q)$$

Using the standard KL divergence formula for Bernoulli distributions:

$$KL(P, Q) = \left(\frac{1}{2}\right) \log \frac{\frac{1}{2}}{\frac{1}{2} + \epsilon} + \left(\frac{1}{2}\right) \log \frac{\frac{1}{2}}{\frac{1}{2} - \epsilon}$$

For small ϵ, using Taylor expansion:

$$KL(P, Q) \approx 2\epsilon^2$$

Thus, for n tosses:

$$KL(P^{\otimes n}, Q^{\otimes n}) \approx 2n\epsilon^2$$

Applying the stronger bound on TV distance:

$$d_{tv}(P^{\otimes n}, Q^{\otimes n}) \le \sqrt{1 - e^{-2n\epsilon^2}}$$

Setting the TV Distance Condition:

For a deterministic algorithm to succeed with probability at least $1 - \delta$:

$$d_{tv}(P^{\otimes n}, Q^{\otimes n}) \ge 1 - 2\delta$$

Using our bound:

$$\sqrt{1 - e^{-2n\epsilon^2}} \ge 1 - 2\delta$$

Squaring both sides:

$$1 - e^{-2n\epsilon^2} \ge (1 - 2\delta)^2$$

Rearranging:

$$e^{-2n\epsilon^2} \le 1 - (1 - 4\delta + 4\delta^2)$$

Approximating for small δ, we take $1 - 4\delta$ as the dominant term:

$$e^{-2n\epsilon^2} \le 4\delta$$

Taking the natural logarithm:

$$-2n\epsilon^2 \le \log 4\delta$$

Rearranging for n:

$$n \geq \frac{\log(1/4\delta)}{2\epsilon^2}$$

Since we only care about asymptotics, we simplify to:

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

(3+3 = 6 marks)

1. **Consider bandit instances in which the them mean reward of all arms lie in** $[1/2, 1/2+\gamma]$ **for some** $\gamma > 0$**. Modify Successive Elimination and UCB algorithm to achieve better regret bounds (both instance independent and instance dependent bounds) for these instances. Just to clarify, the algorithm knows the value of** γ**. Your answer should have both description of the modified algorithms as well as their regret analysis.**

   Solution:

   Modified Successive Elimination Algorithm

   The Successive Elimination algorithm proceeds in rounds, maintaining a set of active arms. In each round:

   (a) Pull each active arm a sufficient number of times.
   (b) Compute the empirical mean reward $\hat{\mu}_a(t)$ for each active arm.
   (c) Eliminate any arm whose upper confidence bound (UCB) is lower than the lower confidence bound (LCB) of another arm.
   (d) Repeat until only one arm remains.

   Since the **variance** of each arm is now at most $\gamma(1-\gamma)$, we replace the usual confidence width:

$$\sqrt{\frac{2\log T}{N_a(t)}}$$

   with a tighter bound:

$$\sqrt{\frac{\gamma \log T}{N_a(t)}}$$

   Thus, the new confidence bounds are:

$$\text{UCB}_a(t) = \hat{\mu}_a(t) + \sqrt{\frac{\gamma \log T}{N_a(t)}}$$

$$\text{LCB}_a(t) = \hat{\mu}_a(t) - \sqrt{\frac{\gamma \log T}{N_a(t)}}$$

   Regret Analysis:

   Using these modified confidence intervals:

   - The instance independent regret improves to: $O\left(\sqrt{K\gamma T \log T}\right)$.
   - The instance dependent regret improves to: $O\left(\sum_{a:\Delta_a>0} \frac{\log T}{\gamma \Delta_a}\right)$.

Modified UCB Algorithm

The UCB algorithm selects arms based on:

$$\text{UCB}_a(t) = \hat{\mu}_a(t) + \sqrt{\frac{2\log t}{N_a(t)}}$$

Modification for $\left[\frac{1}{2}, \frac{1}{2} + \gamma\right]$ Rewards

Since the variance is at most $\gamma(1-\gamma)$, we modify the confidence term to: $\sqrt{\frac{\gamma \log t}{N_a(t)}}$

Thus, the new UCB selection rule is: $\text{UCB}_a(t) = \hat{\mu}_a(t) + \sqrt{\frac{\gamma \log t}{N_a(t)}}$

Regret Analysis

With this modification:

- The instance independent regret improves to: $O\left(\sqrt{K\gamma T \log T}\right)$.
- The instance dependent regret improves to: $O\left(\sum_{a:\Delta_a>0} \frac{\log T}{\gamma \Delta_a}\right)$.

(3-marks)

1. **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. That is, if the input coins satisfy the first item, algorithm should return 'Fair' with at least 4/5 probability. On the other hand, if the input coins satisfy the second item, algorithm should return 'Biased' with at least 4/5 probability.
   Modify the lower bound proof for Biased Coin Identification problem to show the same lower bound $(of \Omega(K/\epsilon^2))$ for the above problem.

   Solution:

   - Under H_0, all coins are fair. The probability of observing heads for any coin is $P_0(H) = 1/2$.
   - Under H_1, exactly one coin is biased, with $P_1(H) = 1/2 + \epsilon$, and the rest are fair.

   For a single coin flip, the KL divergence between H_1 (biased coin) and H_0 (fair coin) is:

   $$KL(P_1, P_0) = \left(\frac{1}{2} + \epsilon\right) \log\left(\frac{\frac{1}{2} + \epsilon}{\frac{1}{2}}\right) + \left(\frac{1}{2} - \epsilon\right) \log\left(\frac{\frac{1}{2} - \epsilon}{\frac{1}{2}}\right)$$

   For small ϵ, using a Taylor expansion, this simplifies to:

   $$KL(P_1, P_0) \approx 2\epsilon^2$$

   Suppose we flip each coin n times. The total KL divergence contributed by the biased coin is:

   $$n \cdot KL(P_1, P_0) \approx 2n\epsilon^2$$

   Since the biased coin is unknown, the average KL divergence over all K coins is:

   $$\frac{2n\epsilon^2}{K}$$

To distinguish between H_0 and H_1 with error probability at most 1/5, the change of measure method requires:

$$KL(P_1, P_0) = \Omega(1)$$

Substituting the average KL divergence:

$$\frac{2n\epsilon^2}{K} = \Omega(1)$$

Solving for n, we obtain:

$$n = \Omega\left(\frac{K}{\epsilon^2}\right)$$

Prove the Pinsker Inequality:

- **(2-mark)**

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

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

  (Hint: use elementary calculus)

  Solution:

  – For binary distributions, the total variation distance is:

  $$d_{tv}(P, Q) = \frac{1}{2}\left(|p-q| + |(1-p)-(1-q)|\right) = |p-q|$$

  – The KL divergence between P and Q is:

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

  The inequality $\log x \leq x - 1$ holds for all $x > 0$, with equality if and only if $x = 1$.
  Using this, we can bound the KL divergence.

  $$p\log\frac{p}{q} \leq p\left(\frac{p}{q} - 1\right) = \frac{p(p-q)}{q}$$

  $$(1-p)\log\frac{1-p}{1-q} \leq (1-p)\left(\frac{1-p}{1-q} - 1\right) = \frac{(1-p)(q-p)}{1-q}$$

  Summing these two terms:

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

  Simplifying:

  $$KL(P, Q) \leq \frac{(p-q)^2}{q(1-q)}$$

  To relate the KL divergence to the total variation distance, we use the fact that:

  $$d_{tv}(P, Q) = |p-q|$$

  From the bound on the KL divergence:

  $$KL(P, Q) \leq \frac{(p-q)^2}{q(1-q)}$$

Since $q(1-q) \leq \frac{1}{4}$ for $q \in [0,1]$:

$$KL(P,Q) \leq \frac{(p-q)^2}{\frac{1}{4}} = 4(p-q)^2$$

Taking square roots:

$$|p-q| \leq \sqrt{\frac{1}{2} KL(P,Q)}$$

- **(1-mark)**

  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 $P(S) = \sum_{i \in S} p_i$ and $Q(S) = \sum_{i \in S} q_i$. Show

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

  (Hint: use log sum inequality)

  Solution:

  For distributions P and Q the KL divergence is:

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

  For the binary distributions P' and Q' the KL divergence is:

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

  Apply the Log Sum Inequality:

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

  Partition the sum in $KL(P,Q)$ into two parts: one over S and one over $S^c = [n] \setminus S$:

  $$KL(P,Q) = \sum_{i \in S} p_i \log \frac{p_i}{q_i} + \sum_{i \in S^c} p_i \log \frac{p_i}{q_i}$$

  Apply the Log Sum Inequality to Each Partition

  - For the subset S: $\sum_{i \in S} p_i \log \frac{p_i}{q_i} \geq P(S) \log \frac{P(S)}{Q(S)}$
  - For the subset S^c: $\sum_{i \in S^c} p_i \log \frac{p_i}{q_i} \geq (1 - P(S)) \log \frac{1-P(S)}{1-Q(S)}$

  Adding the two inequalities:

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

  The right-hand side is exactly $KL(P', Q')$

  Thus:

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

- **(2-mark)**

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

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

  (Proof of $E[\delta] \le O(\sqrt{K/T})$ for the MOSS algorithm)

  Solution:

  - **Total Variation Distance**:

$$d_{tv}(P, Q) = \frac{1}{2} \sum_{i=1}^{n} |p_i - q_i|$$

  - **Kullback-Leibler Divergence**:

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

  The inequality $\log x \le x - 1$ holds for all $x > 0$, with equality if and only if $x = 1$. Using this, we can bound the KL divergence.

  Using the inequality $\log x \le x - 1$:

$$p_i \log \frac{p_i}{q_i} \le p_i \left(\frac{p_i}{q_i} - 1 \right) = \frac{p_i(p_i - q_i)}{q_i}$$

  Summing over all i:

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

  To relate the KL divergence to the total variation distance, we use the fact that:

$$d_{tv}(P, Q) = \frac{1}{2} \sum_{i=1}^{n} |p_i - q_i|$$

  From the bound on the KL divergence:

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

  Since $q_i \le 1$ for all i:

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

  Taking square roots:

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

Let $r_1, ..., r_T$ be T independent samples from a distribution D with mean μ (D is supported on $[0,1]$ so all r_i and μ is in $[0,1]$). For any $1 \leq x \leq T$, let $\hat{\mu}_x = \frac{\sum_{i=1}^{x} r_i}{x}$ and

$$I_x = \hat{\mu}_x + \sqrt{\frac{log^+(\frac{T}{Kx})}{x}}$$

where $log^+(z) = max(z, 0)$. Let $\delta = max(0, \mu - min_{1\leq x\leq T} I_x)$

- **(3-marks)**

For any y, prove that

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

(Hint: use Hoffding's maximal inequality - Suppose $X_1, X_2, ...$ are iid such that each X_i is in $[0,1]$ and $E[X_i] = \mu$ then for any $t > 0$ and $m \geq 1$, we have

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

Note that the above inequality is stronger than the one we used throughout the class. To apply the above inequality, t should not depend on r. However, you might come up with an expression where t may depend on r. In such case, use the following trick:

$$Pr(\exists 1 \leq r \leq m : \sum_{i=1}^{r}(\mu - X_i) \geq t(r)) \leq \sum_{j} Pr(\exists 2^j \leq r \leq 2^{j+1} : \sum_{i=1}^{r}(\mu - X_i) \geq f(j))$$

where $t(r) \geq f(j)$ for $r \in [2^j, 2^{j+1}]$. Now with some manipulation, you should be able to apply the Hoffding's maximal inequality.
Additionally, following inequality might be useful: $\sum_{j=1}^{\infty} 2^j exp(-2^j y^2) = O(\frac{1}{y^2})$

Solution:

The event $\delta \geq y$ implies:

$$\mu - \min_{1\leq x\leq T} I_x \geq y$$

This means there exists some x such that:

$$I_x \leq \mu - y$$

Substituting the definition of I_x:

$$\hat{\mu}_x + \sqrt{\frac{\log_+(T/Kx)}{x}} \leq \mu - y$$

Rearranging:

$$\hat{\mu}_x - \mu \leq -y - \sqrt{\frac{\log_+(T/Kx)}{x}}$$

Let $X_i = \mu - r_i$. Note that X_i are independent, $0 \leq X_i \leq 1$, and $E[X_i] = 0$.
Hoeffding's maximal inequality states that for any $t > 0$ and $m \geq 1$:

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

To handle the dependence of t on r, we partition the range $1 \leq x \leq T$ into intervals of the form $[2^j, 2^{j+1}]$.
For each j, let:

$$f(j) = y \cdot 2^j + \sqrt{\frac{\log_+(T/K \cdot 2^j)}{2^j}} \cdot 2^j$$

Using the summation trick:

$$\Pr\left(\exists 2^j \leq x \leq 2^{j+1} : \sum_{i=1}^{x} X_i \geq f(j)\right) \leq 2^j \exp\left(-2^j y^2\right)$$

Summing over all j:

$$\Pr(\delta \geq y) \leq \sum_{j=1}^{\infty} 2^j \exp\left(-2^j y^2\right)$$

Using the given inequality:

$$\sum_{j=1}^{\infty} 2^j \exp\left(-2^j y^2\right) = O\left(\frac{1}{y^2}\right)$$

The term $\log_+(T/Kx)$ ensures that the probability scales with K/T.
Thus:

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

- **(2-marks)**
  Prove that $E[\delta] = O(\sqrt{K/T})$.
  (Hint: note that δ is a continuous r.v.. Define a new r.v. β such that β is discrete and $\delta \leq \beta$ and so $E[\delta] \leq E[\beta]$. Obviously, you need to use the result of the first part, i.e., $Pr(\delta \geq y) \leq \frac{K}{T} \cdot O(\frac{1}{y})$ which should be hint on how to define β.)

  Solution:

  Since δ is a continuous random variable, we define a discrete random variable β such that $\delta \leq \beta$. This allows us to bound $E[\delta]$ as:

  $$E[\delta] \leq E[\beta]$$

  Let $y_j = 2^{-j}$ for $j = 1, 2, \ldots.$, So:

  $$\beta = \sum_{j=1}^{\infty} y_j \cdot \mathbf{1}(\delta \geq y_j)$$

  where $\mathbf{1}(\delta \geq y_j)$ is the indicator function for the event $\delta \geq y_j$.
  The expectation of β is:

  $$E[\beta] = \sum_{j=1}^{\infty} y_j \cdot \Pr(\delta \geq y_j)$$

  Using the result from the previous part:

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

  Substituting $y_j = 2^{-j}$:

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

  Substitute $y_j = 2^{-j}$ and the bound on $\Pr(\delta \geq y_j)$ into the expression for $E[\beta]$:

  $$E[\beta] = \sum_{j=1}^{\infty} 2^{-j} \cdot \frac{K}{T} \cdot O\left(2^{2j}\right)$$

Simplifying the expression:

$$E[\beta] = \frac{K}{T} \cdot O\left(\sum_{j=1}^{\infty} 2^j\right)$$

The sum $\sum_{j=1}^{\infty} 2^j$ diverges, but we can use the fact that $\sum_{j=1}^{\infty} 2^j \exp(-2^j y^2) = O(1/y^2)$ to bound the expectation:

$$E[\beta] = \frac{K}{T} \cdot O\left(\frac{1}{y^2}\right)$$

To bound $E[\delta]$, we take the square root of the bound on $E[\beta]$:

$$E[\delta] \leq E[\beta] = O\left(\sqrt{\frac{K}{T}}\right)$$