Exercise session n°4 : Kullback-Leibler divergence and lower bounds

Exercise 1 :

In this exercise, we are going to compare the $\frac{1}{K_{\inf}(\nu_k,\mathcal{D},\mu^\star)}$ lower bound, with the $\frac{8}{\Delta_k^2}$ upper bound of UCB on $\mathbb{E}[N_k(T)]$.

1) For $p, q \in [0,1]$, we denote $\mathrm{kl}(p,q) = \mathrm{KL}(\mathrm{Ber}(p), \mathrm{Ber}(q))$. Show that for any $p, q \in [0,1]$,

$$\mathrm{kl}(p,q) \geq 2(p-q)^2.$$

2) Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathbb{P}, \mathbb{Q}$ be two probability distributions over $(\Omega, \mathcal{F})$. Show that

$$\sup_{\substack{Z,\ Z \text{ is } \mathcal{F} \text{ measurable} \\ \text{taking values in } [0,1]}} |\mathbb{E}_{\mathbb{P}}[Z] - \mathbb{E}_{\mathbb{Q}}[Z]| \leq \sqrt{\frac{1}{2}\mathrm{KL}(\mathbb{P}, \mathbb{Q})}.$$

3) Pinsker's inequality: Show that under the same conditions as 2), we have

$$\|\mathbb{P} - \mathbb{Q}\|_{\mathrm{TV}} \coloneqq \sup_{A \in \mathcal{F}} |\mathbb{P}(A) - \mathbb{Q}(A)| \leq \sqrt{\frac{1}{2}\mathrm{KL}(\mathbb{P}, \mathbb{Q})}.$$

Using refined versions of UCB (and its analysis), we can even get the following asympotic upper bound for any $\mathcal{D} \subset \{\nu \mid \nu \text{ is } \sigma \text{ sub-Gaussian}\}$ and $\nu \in \mathcal{D}$:

$$\limsup_{T \to \infty} \frac{\mathbb{E}[N_k(T)]}{\ln(T)} \leq \frac{2\sigma^2}{\Delta_k^2}.$$

4) Assume in this question that $\mathcal{D} \subset \mathcal{P}([0,1])$

(a) What does the above upper bound becomes when $\mathcal{D} \subset \mathcal{P}([0,1])$?

(b) Exhibit a lower bound on $K_{\inf}(\nu_k, \mathcal{D}, \mu^\star)$ in that case and compare with the above upper bound.

(c) Can you give an example where the known lower bound and the above upper bound differ?

5) Show that if $\mathcal{D} = \{\mathcal{N}(\mu, 1) \mid \mu \in \mathbb{R}\}$, then $K_{\inf}(\nu_k, \mathcal{D}, \mu^\star) = \frac{2}{\Delta_k^2}$ and comment.

Exercise 2 :

This exercise aims at giving a lower bound on the number of pulls of a suboptimal arm for small time horizons. We use the same notations as in the previous exercise.

1)

(a) Establish the following local version of Pinsker's inequality:

$$\text{for any } 0 \le p < q \le 1, \quad \mathrm{kl}(p,q) \ge \frac{1}{2\max_{x\in[p,q]} x(1-x)}(p-q)^2.$$

Why is it stronger than Pinsker's inequality?

(b) Deduce that it yields

$$\text{for any } 0 \le p < q \le 1, \quad \mathrm{kl}(p,q) \ge \frac{1}{2q}(p-q)^2.$$

2) A strategy is said *non-naive* if for all bandit instances and k such that $\mu_k = \mu^\star$, $\mathbb{E}[N_k(T)] \ge \frac{T}{K}$. Show that for all non-naive strategies and for any instance ν:

$$\forall T \le \frac{1}{8\mathrm{KL}^\star}, \forall k \in [K], \quad \mathbb{E}[N_k(T)] \ge \frac{T}{2K},$$
$$\text{where} \quad \mathrm{KL}^\star := \max_{k, \Delta_k > 0} K_{\inf}(\nu_k, \mathcal{D}, \mu^\star).$$

Hint: Consider the same alternative bandits instance ν' as we did in the course, when proving the asymptotic lower bound.

Exercise 3 :

This exercise aims at showing a minimax lower bound of the regret of the form $R_T \ge c\sqrt{KT}$. We restrict ourselves to the bandit model $\mathcal{D} = \{\mathcal{N}(\mu, 1) \mid \mu \in \mathbb{R}\}$, but similar arguments can be used for more general models (e.g. Bernoulli bandits). Fix in the following $K \ge 2$ and $T \ge \frac{K-1}{2}$. The minimax regret is defined as

$$R_T^\star = \inf_{\text{strategy } \pi} \sup_{\text{instance } \nu} R_T(\pi, \nu).$$

Let $\varepsilon > 0$. We consider in the following $K+1$ bandit instances $(\nu^j)_{j\in[K+1]}$, where

$$\nu_k^j = \mathcal{N}(0,1) \quad \text{for any } k \in [K] \text{ such that } j \ne k$$
$$\nu_k^k = \mathcal{N}(\varepsilon,1) \quad \text{for any } k \in [K].$$

1) Justify that

$$R_T^\star \ge \inf_{\pi} \sup_{\varepsilon\in(0,1)} \max_{i\in[K]} \varepsilon(T - \mathbb{E}_{\nu^i}^\pi[N_i(T)]),$$

and that for any strategy π, there exists k_0 such that $\mathbb{E}_{\nu^0}[N_{k_0}(T)] \le \frac{T}{K}$.

2) Use the fundamental inequality and Pinsker's inequality to show that

$$\mathbb{E}_{\nu^0}[N_{k_0}(T)]\frac{\varepsilon^2}{2} \ge 2\left(\mathbb{E}_{\nu^0}[\frac{N_{k_0}(T)}{T}] - \mathbb{E}_{\nu^{k_0}}[\frac{N_{k_0}(T)}{T}]\right)^2.$$

3) Combine the above results to derive

$$R_T^\star \ge \sup_{\varepsilon\in(0,1)} \varepsilon T\left(1 - \frac{1}{K} - \varepsilon\sqrt{\frac{T}{2K}}\right)$$

and conclude that $R_T^\star \ge \frac{1}{8\sqrt{2}}\sqrt{KT}$.