Exercise sheet n°2

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 $K_{\inf}(\nu_k, \mathcal{D}, \mu^\star) = \frac{\Delta_k^2}{2}$ 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:

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

Why is it stronger than Pinsker's inequality?

(b) Deduce that it yields

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

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

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

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

Exercise 3 :

Consider an alternative version of MOSS algorithm, where $U_k(t)$ is replaced by the following value:

$$U_k(t) = \hat{\mu}_k(t) + \sqrt{\frac{1}{N_k(t)} \ln_+\left(\frac{t}{N_k(t)}\right)}.$$

1) Show that there is a universal constant $c > 0$, such that for any $\varepsilon > 0$ and any $t \in \mathbb{N}$,

$$\mathbb{P}\left(\mu_k - \hat{\mu}_k(t) \geq \sqrt{\frac{1}{N_k(t)} \ln_+\left(\frac{t}{N_k(t)}\right)} + \varepsilon\right) \leq \frac{c}{t\varepsilon^2}$$
$$\text{and } \mathbb{P}\left(\hat{\mu}_k(t) - \mu_k \geq \sqrt{\frac{1}{N_k(t)} \ln_+\left(\frac{t}{N_k(t)}\right)} + \varepsilon\right) \leq \frac{c}{t\varepsilon^2}.$$

Hint: Use a peeling argument as in the proof of MOSS.

2) Deduce that the regret of this algorithm can be bounded as

$$R_T \leq c' \left(\sum_{k, \Delta_k > 0} \frac{\ln(T)}{\Delta_k} + \Delta_k\right),$$

where c' is a universal constant.

Bonus: show that we can even have the tighter bound (for another constant c')

$$\mathbb{E}[N_k(T)] \leq c' \left(\frac{\ln_+(T\Delta_k^2)}{\Delta_k^2} + 1\right).$$

3) Admit for this question that for any $\alpha \in [0, 1]$,

$$\max_{u>0} \min\left(\alpha u, \frac{\ln_+(u^2)}{u}\right) \leq \max\left(e\alpha, \sqrt{\alpha \ln(1/\alpha)}\right).$$

(a) Using the previous bonus question, show that there is a universal constant c' such that for any $k \in [K]$,

$$\Delta_k \mathbb{E}[N_k(T)] \leq c' \max(\frac{\mathbb{E}[N_k(T)]}{\sqrt{T}}, \sqrt{\mathbb{E}[N_k(T)] \ln\left(\frac{T}{\mathbb{E}[N_k(T)]}\right)}) + c'.$$

(b) Show that the modified MOSS satisfies the following distribution free bound

$$R_T \leq c'(\sqrt{KT\ln(K)} + K),$$

where c' is a universal constant.

Exercise 4 :

Consider th K-armed stochastic contextual setting (setting 1 in lecture 7) and assume that $\mathcal{C} = [0,1]$ and the reward function is (L,α)-Hölder for $\alpha \in (0,1]$:

$$\forall k \in [K], \forall c, c' \in \mathcal{C}, |r(k,c) - r(k,c')| \leq L|c - c'|^\alpha.$$

Build an algorithm with a regret bound (to prove) of order

$$R_T = \mathcal{O}\left(L^{\frac{1}{2\alpha+1}} K^{\frac{\alpha}{2\alpha+1}} T^{\frac{\alpha+1}{2\alpha+1}}\right).$$

Exercise 5 :

Consider in this exercise a bandit instance $\nu \in \mathcal{D}^K$ such that

- $\mathcal{D} = \{\mathcal{N}(\mu, 1) \mid \mu \in \mathbb{R}\}$;
- ν has a unique optimal arm.

We define for any $\nu' \in \mathcal{D}^K$:

$$\alpha^*(\nu') = \operatorname*{argmax}_{\alpha \in \mathcal{P}_K} \inf_{\tilde{\nu}' \in \mathcal{D}_{\text{alt}(\nu')}} \sum_{k=1}^K \alpha_k \text{KL}(\nu'_k, \tilde{\nu}'_k).$$

1) Show that

$$\alpha^*\nu = \operatorname*{argmax}_{\alpha \in \mathcal{P}_K} \Phi(\nu, \alpha)$$
$$\text{where} \quad \Phi(\nu, \alpha) = \frac{1}{2} \min_{k \neq k^*} \frac{\alpha_{k^*}\alpha_k}{\alpha_{k^*} + \alpha_k} \Delta_k^2.$$

2) Justify that $\Phi(\nu, \alpha)$ is a concave function of α.

3) Show that $\alpha^*(\nu)$ is unique.

4) Show that α^* is continuous at ν.