Exercise session n°6 : Contextual bandits and Best Arm Identification

Exercise 1 :

Consider th K-armed stochastic contextual setting (setting 1 in lecture 8) and assume that $\mathcal{C} = [0,1]$ and the reward function is (L, α)-Hölder for $\alpha \in (0,1]$:

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

Using ideas similar to Exercise 1 of the Exercise Session #5, 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 2 :

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 $\nu' \in \mathcal{D}^K$:

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

1) Show that

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

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.

Exercise 3 :

We aim at proving the regret bound of Track-And-Stop algorithm in this exercise. Assume that ν has a unique optimal arm (recall that $\nu_k = \mathcal{N}(\mu_k, 1)$). Make $\mathcal{D}^K$ a metric space via the metric $d(\nu, \nu') = \max_{k \in [K]} |\mathbb{E}(\nu_k) - \mathbb{E}(\nu'_k)|$.

We also define in the following $\hat{\nu}_k^t = \mathcal{N}(\hat{\mu}_k(t), 1)$ and use the same notations as in the previous exercise (we are going to use the results of the previous exercise).

Let $\varepsilon > 0$ be a small constant and define the random times

$$\tau_\nu(\varepsilon) = 1 + \max\{t \mid d(\hat{\nu}^t, \nu) \geq \varepsilon\}$$
$$\tau_\alpha(\varepsilon) = 1 + \max\{t \mid \|\alpha^*(\nu) - \alpha^*(\hat{\nu}^t)|_\infty \geq \varepsilon\}$$
$$\tau_T(\varepsilon) = 1 + \max\{t \mid \|\alpha^*(\nu) - \frac{N(t)}{t}\|_\infty \geq \varepsilon\}.$$

Note these are not stopping times.

1) We are gonna use the first concentration inequality admitted in the proof of the Lemma that guarantees soudness of Track-and-Stop.

(a) Define the random variable:

$$\Lambda = \min\{\lambda \geq 1 \mid d(\hat{\nu}^t, \nu) \leq \sqrt{\frac{2\ln(\lambda K t(t+1))}{\min_k N_k(t)}} \text{ for all t}\}.$$

Show that $\mathbb{E}[\ln(\Lambda)^2] < \infty$.

(b) Prove that $\mathbb{E}[\tau_\nu(\varepsilon)] < \infty$ for all $\varepsilon > 0$.

2) Prove that $\mathbb{E}[\tau_\alpha(\varepsilon)] < \infty$ for all $\varepsilon > 0$.

3) Prove that $\mathbb{E}[\tau_T(\varepsilon)] < \infty$ for all $\varepsilon > 0$.

4)

(a) Define for any $\varepsilon > 0$

$$\tau_\beta(\varepsilon, \delta) = 1 + \max\{t \mid t\Phi(\nu, \alpha^*(\nu)) < \beta_t(\delta) + \varepsilon t\}$$
$$\text{and} \quad u(\varepsilon) = \sup_{\nu', \alpha}\{\Phi(\nu, \alpha^*(\nu) - \Phi(\nu', \alpha)) \mid d(\nu', \nu) \leq \varepsilon, \|\alpha - \alpha^*(\nu)\|_\infty \leq \varepsilon\}.$$

Show that $\mathbb{E}[\tau] \leq \mathbb{E}[\tau_\nu(\varepsilon)] + \mathbb{E}[\tau_T(\varepsilon)] + \mathbb{E}[\tau_\beta(u(\varepsilon), \delta)]$.

(b) Conclude that $\lim_{\delta \to 0^+} \frac{\mathbb{E}[\tau]}{\ln(1/\delta)} \leq c^*(\nu)$.