Exercise 2: Distribution-free lower bound for K–armed bandits (can be solved after Course #5)

As indicated in class, one of the exercises of the present homework is devoted to proving that in the stochastic K–armed bandit setting, i.e., when K arms with respective distributions $\nu_1, \ldots, \nu_K$ over $[0,1]$ (with expectations denoted by $\mu_1, \ldots, \mu_K$) are available, no strategy $\mathcal{S}$ can have a sharper distribution-free regret bound than one of the order $\sqrt{KT}$.

More precisely, we denote by Y_t the reward obtained at each round, when picking arm I_t; we recall that Y_t is drawn at random according to ν_{I_t} conditionally to I_t. The regret is defined as

$$R_T = T \max_{k=1,\ldots,K} \mu_k - \mathbb{E}\left[\sum_{t=1}^{T} Y_t\right].$$

You will prove that for all $K \geqslant 2$ and all $T \geqslant K/5$,

$$R_T^\star = \inf_{\mathcal{S}} \sup_{\underline{\nu}} R_T \geqslant \frac{1}{20}\sqrt{KT}\,,$$

where the defining infimum of $R_T^\star$ is over all strategies $\mathcal{S}$ and the supremum is over all K–tuples of distributions $\underline{\nu} = (\nu_1, \ldots, \nu_K)$ over $[0,1]$.

As the proof will reveal, it actually suffices to consider Bernoulli distributions. Indeed, let $\varepsilon \in (0,1)$ and consider the K–tuples $\underline{\nu}^{(0)}, \underline{\nu}^{(1)}, \ldots, \underline{\nu}^{(K)}$ defined based on the Bernoulli distributions $\mathrm{B}_+ = \mathrm{Ber}(1/2+\varepsilon/2)$ and $\mathrm{B}_- = \mathrm{Ber}(1/2-\varepsilon/2)$ as follows:

- In Model 0, all arms are associated with B_-, that is, $\underline{\nu}^{(0)} = (\mathrm{B}_-, \ldots, \mathrm{B}_-)$.
- In Model $i \in \{1, \ldots, K\}$, all arms are associated with B_- except the i–th arm, which is associated with B_+.

We denote by $\mathbb{P}_i$ the probability induced by Model i, for $i \in \{0, 1, \ldots, K\}$, and by $\mathbb{E}_i$ the corresponding expectation. We denote by $N_k(T)$ the number of times arm k was pulled by the considered strategy till round T included.

1. Explain why

$$R_T^\star \geqslant \inf_{\mathcal{S}} \sup_{\varepsilon\in(0,1)} \max_{i\in\{1,\ldots,K\}} \varepsilon\Big(T - \mathbb{E}_i\big[N_i(T)\big]\Big)$$

and why there exists k_0 such that $\mathbb{E}_0\big[N_{k_0}(T)\big] \leqslant T/K$.

2. Use the fundamental inequality for proving lower bounds in stochastic bandit problems and Pinsker's inequality to get, for all strategies $\mathcal{S}$,

$$\mathbb{E}_0\big[N_{k_0}(T)\big]\ \mathrm{KL}(\mathrm{B}_-, \mathrm{B}_+) \geqslant 2\Big(\mathbb{E}_0\big[N_{k_0}(T)/T\big] - \mathbb{E}_{k_0}\big[N_{k_0}(T)/T\big]\Big)^2.$$

3. Combine the results above to derive

$$R_T^\star \geqslant \inf_{\mathcal{S}} \sup_{\varepsilon\in(0,1)} \varepsilon\, T\left(1 - \frac{1}{K} - \sqrt{\frac{T}{2K}\,\mathrm{KL}(\mathrm{B}_-, \mathrm{B}_+)}\right)$$

and conclude to the desired bound. You may use that

$$\varepsilon \in (0,\, 1/2) \longmapsto 2.5\,\varepsilon^2 - \varepsilon \ln\frac{1+\varepsilon}{1-\varepsilon}$$

takes positive values.