Exercise session n°5 : MOSS algorithm and continuous bandits

Exercise 1 :

We consider the setting of stochastic with a continuum of arms indexed by $\mathcal{A} = [0, 1]$, with a mean-payoff function f that is α-Hölder with $\alpha \in (0, 1]$, i.e., there is $L > 0$ such that

$$\forall x, x' \in [0, 1], \quad |f(x) - f(x')| \leq L|x - x'|^\alpha.$$

We now consider the following algorithm that discretizes the action space into K bins:

- discretize $[0, 1]$ into K bins, with $B_i = [\frac{i-1}{K}, \frac{i}{K}]$ for any $i = 1, \ldots, K$,
- run MOSS algorithm over K arms, where picking the arm $I_t \in [K]$ corresponds to picking an action a_t (chosen arbitrarily) in the bin B_{I_t}.

1) Show that the regret of this "discretized" algorithm satisfies:

$$R_T \leq K + 45\sqrt{KT} + \frac{TL}{K^\alpha}.$$

2) Assume that T, α are known in advance. Show that for a good choice of K, the regret is of order $T^{\frac{\alpha+1}{2\alpha+1}}$.

Exercise 2 :

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.