

Solution for Exercise on Lipschitz bandits.

[written in somewhat a rush:
let me know if there are
types!]

1) Fixed $K \geq 2$

→ Consider the bins $[(j-1)/K, j/K]$ for $j=1, \dots, K$

→ Master strategy

- * whenever the auxiliary strategy recommends $J_t \in \{1, \dots, K\}$,
pick $I_t \in [0, 1]$ uniformly at random in $[(J_t-1)/K, J_t/K]$
- * get a reward $\tilde{y}_t$ sampled according to $\tilde{\nu}_{I_t}$
- * send this reward to the auxiliary strategy

→ Auxiliary strategy: UCB

- * pick arms $J_t \in \{1, \dots, K\}$ according to the UCB strategy
- * get the associated rewards from the master strategy

The auxiliary strategy thus performs UCB on the bandit model $(\tilde{\nu}_j)_{j=1, \dots, K}$ where $\tilde{\nu}_j$ is the distribution of $\tilde{y}$ obtained from the following two-step randomization:

- draw X uniformly at random in $[(j-1)/K, j/K]$
- draw Y at random according to ν_X (given X).

In particular,

$$\tilde{\mu}_j = E(\tilde{\nu}_j) = K \int_{(j-1)/K}^{j/K} f(t) dt \quad \text{where } f(t) = E(\nu_t)$$

Performance of the (auxiliary) strategy as indicated by the distribution-free bound on UCB we exhibited in our earlier exercise:

$$T \max_{j=1, \dots, K} \tilde{\mu}_j - E\left[\sum_{t=1}^T \tilde{y}_t\right] \leq \sqrt{KT(8\ln T + 2)}$$

To get the performance of the (master) strategy, we only need to control the

approximation error

$$\max_{x \in [q_1]} f(x) - \max_j \tilde{f}_j$$

But $\forall x \in [j-1/k, j/k]$, $|\tilde{f}_j - f(x)| \leq K \int_{j-1/k}^{j/k} |f(t) - f(x)| dt$

$$\leq L \times K \int_{j-1/k}^{j/k} |t-x| dt$$

$$\leq L \times K \int_0^{1/k} t dt = \frac{L}{2K}$$

in particular,

$$|\max_j \tilde{f}_j - \max_{x \in [q_1]} f(x)| \leq T \frac{L}{2K}$$

worst-case (largest) value is when $x = j/k$ or $x = (j-1)/k$

The (total) regret is therefore:

$$T \max_{x \in [q_1]} f(x) - \mathbb{E} \left[\sum_{t=1}^T y_t \right] \leq \frac{LT}{2K} + \sqrt{KT(8\ln T + 2)}$$

2) How should we pick K?

→ If T is known, we can set K s.t. T/K is of the same order of magnitude as $\sqrt{KT}$ (the bound needs to hold $\forall L$, so we cannot have K depend on L): e.g., $K = \lceil T^{1/3} \rceil \leq 1 + T^{1/3}$, in which case the regret bound is $\leq \left(\frac{L}{2} + \sqrt{8\ln T + 2} \right) (T^{2/3} + \sqrt{T})$.

→ Otherwise, we resort to a (dirty) "doubling trick," by restarting the strategy of question (1) after times $t = 2^r$, with $r = 0, 1, 2, \dots$, for 2^r rounds and with $K = \lceil (2^r)^{1/3} \rceil$.

The total regret is equal to the sum of the regrets over these regimes:

$$\bar{R}_T \leq 2 + \sum_{r=1}^{r_T} \left(\frac{L}{2} + \sqrt{8\ln 2^r + 2} \right) \left((2^r)^{2/3} + \sqrt{2^r} \right) \leq \sum_{r=1}^{r_T} (2^r)^{2/3}$$

with r_T s.t. $2^{r_T+1} \leq T \leq 2^{r_T+1}$

$$\bar{R}_T \leq 2 + \left(\frac{L}{2} + \sqrt{8\ln T + 2} \right) \times \left(\sum_{r=0}^{r_T-1} (2^{2/3})^r \right) \times 2 \times 2^{2/3} \leq \frac{T^{2/3}}{2^{2/3-1}} \leq T^{2/3} / 2^{2/3-1}$$

regime r_T might not be incomplete for $t \leq T$ also hold

That is,

$$\bar{R}_T \leq 2 + \left(\frac{L}{2} + \sqrt{8 \ln T + 2} \right) \times \underbrace{\frac{2 \times 2^{2/3}}{2^{2/3} - 1}}_{\leq 6} \times T^{2/3}$$

Final clean bound:

$$\bar{R}_T \leq (3L + 6\sqrt{8 \ln T + 2}) T^{2/3} + 2$$

Note

it can be shown that the $T^{2/3}$ order of magnitude is optimal; the $\sqrt{\ln T}$ term can be dropped by resorting to more efficient auxiliary strategies than UCB.

Correction for the exercise providing a useful rewriting of KL:

- Given that when $P \ll Q$, we have

$$\begin{aligned} \text{KL}(P, Q) &= \int_{\Omega} \left(\frac{dP}{dQ} \ln \frac{dP}{dQ} \right) dQ && \text{by definition of KL} \\ &= \int_{\Omega} \left(\ln \frac{dP}{dQ} \right) dP && \text{by definition of } \frac{dP}{dQ} \end{aligned}$$

Also, by definition of the density functions: $dQ = g d\mathbb{P}$ and $dP = f d\mathbb{P}$

$$\text{Thus, to get } \text{KL}(P, Q) = \int_{\Omega} \left(\frac{f}{g} \ln \frac{f}{g} \right) g d\mathbb{P} = \int_{\Omega} \left(\ln \frac{f}{g} \right) f d\mathbb{P}$$

We only need to prove that $\frac{f}{g}$ is a density of P wrt Q .

- To that end, we need to be careful with the event $E = \{g=0\}$

$$\begin{aligned} \text{We have } Q(E) &= \int \mathbb{1}_E dQ = \int \mathbb{1}_{\{g=0\}} g d\mathbb{P} = 0, \\ \text{thus, by } P \ll Q: & \quad P(E) = 0 \text{ as well.} \end{aligned}$$

Therefore, for all $A \in \mathcal{F}$,

$$\begin{aligned} P(A) &= P(A \cap E^c) = \int \mathbb{1}_A \mathbb{1}_{\{g>0\}} f d\mathbb{P} \\ &= \int \mathbb{1}_A \mathbb{1}_{\{g>0\}} \frac{f}{g} g d\mathbb{P} \\ &= \int \mathbb{1}_A \frac{f}{g} \underbrace{(\mathbb{1}_{\{g>0\}} g)}_{= g d\mathbb{P} = dQ} d\mathbb{P} \end{aligned}$$

here, we heavily used the conventions
 $0/0 = 0$
 and
 $0 \times +\infty = 0$

$$\text{Thus, } P(A) = \int_{\Omega} \mathbb{1}_A \frac{f}{g} dQ.$$