Solution for Exercise on Lipschitz bandits.

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

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 Y_t sampled according to ν_{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 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,

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

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

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

To get the performance of the (master) strategy, we only need to control the approximation error $\max_{x\in[0,1]} f(x) - \max_j \tilde{\mu}_j$

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

$$\leqslant L \times K \int_{(j-1)/K}^{j/K} |t-x|\,dt$$

worst-case (largest) value is when $x = j/K$ or $x = \frac{j-1}{K}$

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

in particular,

$$\left|\max_j \tilde{\mu}_j - \max_{x\in[0,1]} f(x)\right| \leqslant T\,\frac{L}{2K}$$

The (total) regret is therefore:

$$T\max_{x\in[0,1]} f(x) - \mathbb{E}\left[\sum_{t=1}^{T} Y_t\right] \leqslant \frac{LT}{2K} + \sqrt{KT(8\ln T+2)}$$

2) <u>How should we pick K?</u>

→ 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., $K = \lceil T^{1/3}\rceil \leqslant 1+T^{1/3}$, in which case the regret bound is $\leqslant \left(\frac{L}{2}+\sqrt{8\ln T+2}\right)\left(T^{2/3}+\sqrt{T}\right)$.

→ Otherwise, we resort to a (dirty) "doubling trick", by restarting the strategy of question (1) after times $t = 2^{r+1}$, 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:

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

regime r_T might be incomplete but the bound also holds for $t \leqslant T$

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

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

That is, $\overline{R}_T \leqslant 2 + \left(\frac{L}{2} + \sqrt{8\ln T + 2}\right) \times \underbrace{\frac{2 \times 2^{2/3}}{2^{2/3} - 1}}_{\leqslant 6} \times T^{2/3}$

Final clean bound: $\overline{R}_T \leqslant \left(3L + 6\sqrt{8\ln T + 2}\right) T^{2/3} + 2$

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 $\mathbb{P} \ll \mathbb{Q}$, we have

$$\mathrm{KL}(\mathbb{P},\mathbb{Q}) = \int_\Omega \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \ln \frac{d\mathbb{P}}{d\mathbb{Q}}\right) d\mathbb{Q} \qquad \text{by definition of KL}$$

$$= \int_\Omega \left(\ln \frac{d\mathbb{P}}{d\mathbb{Q}}\right) d\mathbb{P} \qquad \text{by definition of } \frac{d\mathbb{P}}{d\mathbb{Q}}$$

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

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

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

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

We have $\mathbb{Q}(E) = \int \mathbb{1}_E \, d\mathbb{Q} = \int \mathbb{1}_{\{g=0\}} \, g \, d\nu = 0,$

thus, by $\mathbb{P} \ll \mathbb{Q}$:

$\mathbb{P}(E) = 0$ as well.

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

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

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

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