Exercise sheet n°2

Exercise 1 :

In this exercise, we are going to compare the $\frac{1}{K_{\text{inf}}(\nu_k,\mathcal{D},\mu^\star)}$ lower bound, with the $\frac{8}{\Delta_k^2}$ upper bound of UCB on $\mathbb{E}[N_k(T)]$.

1) For $p,q \in [0,1]$, we denote $\text{kl}(p,q) = \text{KL}(\text{Ber}(p),\text{Ber}(q))$. Show that for any $p,q \in [0,1]$,

$$\text{kl}(p,q) \geq 2(p-q)^2.$$

2) Let $(\Omega,\mathcal{F})$ be a measurable space and $\mathbb{P},\mathbb{Q}$ be two probability distributions over $(\Omega,\mathcal{F})$. Show that

$$\sup_{\substack{Z,\ Z \text{ is } \mathcal{F} \text{ measurable} \\ \text{taking values in } [0,1]}} |\mathbb{E}_{\mathbb{P}}[Z] - \mathbb{E}_{\mathbb{Q}}[Z]| \leq \sqrt{\frac{1}{2}\text{KL}(\mathbb{P},\mathbb{Q})}.$$

3) Pinsker's inequality: Show that under the same conditions as 2), we have

$$\|\mathbb{P}-\mathbb{Q}\|_{\text{TV}} \coloneqq \sup_{A\in\mathcal{F}} |\mathbb{P}(A) - \mathbb{Q}(A)| \leq \sqrt{\frac{1}{2}\text{KL}(\mathbb{P},\mathbb{Q})}.$$

Using refined versions of UCB (and its analysis), we can even get the following asympotic upper bound for any $\mathcal{D} \subset \{\nu \mid \nu \text{ is } \sigma \text{ sub-Gaussian}\}$ and $\nu \in \mathcal{D}$:

$$\limsup_{T\to\infty} \frac{\mathbb{E}[N_k(T)]}{\ln(T)} \leq \frac{2\sigma^2}{\Delta_k^2}.$$

4) Assume in this question that $\mathcal{D} \subset \mathcal{P}([0,1])$

(a) What does the above upper bound becomes when $\mathcal{D} \subset \mathcal{P}([0,1])$?

(b) Exhibit a lower bound on $K_{\text{inf}}(\nu_k,\mathcal{D},\mu^\star)$ in that case and compare with the above upper bound.

(c) Can you give an example where the known lower bound and the above upper bound differ?

5) Show that if $\mathcal{D} = \{\mathcal{N}(\mu,1) \mid \mu \in \mathbb{R}\}$, then $K_{\text{inf}}(\nu_k,\mathcal{D},\mu^\star) = \frac{2}{\Delta_k^2}$ and comment.

Solution: **1)** Fix $q \in (0,1)$ and define $f(p) = \text{kl}(p,q)$. Computing the derivatives

$$f'(p) = \ln\left(\frac{p(1-q)}{q(1-p)}\right)$$
$$f''(p) = \frac{1}{p(1-p)} \geq 4.$$

So a second order Taylor expansion yields:

$$\begin{aligned} f(p) &\geq f(q) + (p-q)f'(q) + 2(p-q)^2 \\ &= 2(p-q)^2. \end{aligned}$$

2) This is a consequence of the data processing inequality with expectations:

$$\mathrm{KL}(\mathbb{P}, \mathbb{Q}) \geq \mathrm{kl}(\mathbb{E}_{\mathbb{P}}[Z], \mathbb{E}_{\mathbb{Q}}[Z]).$$

This quantity is larger tha $2(\mathbb{E}_{\mathbb{P}}[Z] - \mathbb{E}_{\mathbb{Q}}[Z])^2$, thanks to the last question. And we can take the sup over all such Z.
n **3)** This is taking $Z = \mathbb{1}_A$.
4) a) Replace σ^2 by $\frac{1}{4}$.
b) $K_{\inf}(\nu_k, \mathcal{D}, \mu^\star) \geq 2\Delta_k^2$. So the lower bound is smaller than the upper bound (logic!).
c) Taking $\mathcal{D}$ containing only Bernoulli variables does the trick.
5) Let p (resp. q) be the probability density of a Gaussian of mean μ_1 (resp. μ_2) and variance 1. Since $\frac{p(x)}{q(x)} = e^{\frac{\mu_2^2-\mu_1^2}{2}+(\mu_1-\mu_2)x}$, as simple computation leads to the answer.

$$\begin{aligned} \mathrm{KL}(p,q) &= \int_{\mathbb{R}} \ln(\frac{p}{q})p(x)\mathrm{d}x \\ &= \int (\frac{\mu_2^2 - \mu_1^2}{2} + (\mu_1 - \mu_2)x)p(x)\mathrm{d}x \\ &= \frac{\mu_2^2 - \mu_1^2}{2} + (\mu_1 - \mu_2)\mathbb{E}(p) \\ &= \frac{\mu_2^2 - \mu_1^2}{2} + (\mu_1 - \mu_2)\mu_1 \\ &= \frac{(\mu_2 - \mu_1)^2}{2}. \end{aligned}$$

Exercise 2 :

This exercise aims at giving a lower bound on the number of pulls of a suboptimal arm for small time horizons. We use the same notations as in the previous exercise.

1)

(a) Establish the following local version of Pinsker's inequality:

$$\text{for any } 0 \leq p < q \leq 1, \quad \mathrm{kl}(p,q) \geq \frac{1}{2\max_{x\in[p,q]} x(1-x)}(p-q)^2.$$

Why is it stronger than Pinsker's inequality?

(b) Deduce that it yields

$$\text{for any } 0 \leq p < q \leq 1, \quad \mathrm{kl}(p,q) \geq \frac{1}{2q}(p-q)^2.$$

2) A strategy is said *non-naive* if for all bandit instances and k such that $\mu_k = \mu^\star$, $\mathbb{E}[N_k(T)] \geq \frac{T}{K}$. Show that for all non-naive strategies and for any instance ν:

$$\forall T \leq \frac{1}{8\mathrm{KL}^\star}, \forall k \in [K], \quad \mathbb{E}[N_k(T)] \geq \frac{T}{2K},$$
$$\text{where} \quad \mathrm{KL}^\star := \max_{k,\Delta_k>0} K_{\inf}(\nu_k, \mathcal{D}, \mu^\star).$$

Hint: Consider the same alternative bandits instance ν' as we did in the course, when proving the asymptotic lower bound.

Solution: 1)a) We extract from the question 1) in Exercise 1:

$$\exists r \in [p,q] \text{ s.t. } \mathrm{kl}(p,q) = \frac{1}{2r(1-r)}(p-q)^2.$$

It is a tighter as soon as $\frac{1}{2} \notin [p,q]]$. b) This a direct consequence.
2) We can again assume, without loss of generality, that $KL^\star < +\infty$. Then for any suboptimal k (otherwise it is automatic from definition of non-naive algorithm), we can consider ν' as

$$\begin{cases} \nu'_j = \nu_j if j \neq k \\ \nu'_k \in \mathcal{D} s.t. \mathbb{E}(\nu'_k) > \mu^\star. \end{cases}$$

Again, we have

$$\mathbb{E}_\nu[N_k(T)]\mathrm{KL}(\nu_k,\nu'_k) \geq \mathrm{kl}(\mathbb{E}_\nu[\frac{N_k(T)}{T}], \mathbb{E}_{\nu'}[\frac{N_k(T)}{T}]).$$

The strategy is non-naive, so $\mathbb{E}_{\nu'}[\frac{N_k(T)}{T}] \geq \frac{1}{K}$. If $\mathbb{E}_\nu[\frac{N_k(T)}{T}] \geq \frac{1}{K}$, then the lower bound is true. Otherwise, the local version of Pinsker's inequality yields (+using monotonicity)

$$\begin{aligned} \mathbb{E}_\nu[N_k(T)]\mathrm{KL}(\nu_k,\nu'_k) &\geq \mathrm{kl}(\mathbb{E}_\nu[\frac{N_k(T)}{T}], \frac{1}{K}) \\ \frac{T}{K}\mathrm{KL}(\nu_k,\nu'_k) \geq &\geq \frac{K}{2}(\mathbb{E}_\nu[\frac{N_k(T)}{T}] - \frac{1}{K})^2. \end{aligned}$$

Going to the infimum of such ν'_k yields $\frac{2T}{K^2}K_{\inf}(\nu_k,\mathcal{D},\mu^\star) \geq (\mathbb{E}_\nu[\frac{N_k(T)}{T}] - \frac{1}{K})^2$. We can then conclude when T is is in the considered range.

Exercise 3 :
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.

Solution: 1) We have for any $n \in \mathbb{N}$, following the same arguments as in the proof of MOSS:

$$\mathbb{P}\left(\mu_k - \hat{\mu}_k(t) \geq \sqrt{\frac{1}{N_k(t)} \ln_+\left(\frac{t}{N_k(t)}\right)} + \varepsilon \text{ and } 2n \geq N_k(t) \geq n\right) \leq e^{-2n\varepsilon^2}\frac{2n}{t}.$$

As a consequence, we can do the peeling:

$$\mathbb{P}\left(\mu_k - \hat{\mu}_k(t) \geq \sqrt{\frac{1}{N_k(t)} \ln_+\left(\frac{t}{N_k(t)}\right) + \varepsilon}\right) = \sum_{\ell=0}^{\infty} \mathbb{P}\left(\mu_k - \hat{\mu}_k(t) \geq \sqrt{\frac{1}{N_k(t)} \ln_+\left(\frac{t}{N_k(t)}\right)} + \varepsilon \text{ and } 2^{\ell+1}\right.$$

$$\frac{1}{t}\sum_{\ell=0}^{\infty} 2^{\ell+1} \exp\left(-2^{\ell+1}\varepsilon^2\right).$$

Note that $f : x \mapsto 2^{x+1}\exp(-2^{x+1}\varepsilon^2)$ is increasing and then decreasing on $\mathbb{R}_+$. As a consequence, we have the comparison $\sum_{\ell=0}^{\infty} f(l) \leq \max_x f(x) + \int_0^\infty f(x)\mathrm{d}x$. So

$$\begin{aligned}
\sum_{\ell=0}^{\infty} 2^{\ell+1}\exp\left(-2^{\ell+1}\varepsilon^2\right) &\leq \max_{x\in\mathbb{R}_+} 2^{x+1}\exp(-2^{x+1}\varepsilon^2) + \int_0^\infty 2^{x+1}\exp(-2^{x+1}\varepsilon^2)\mathrm{d}x \\
&\leq \max_{u\in[2,\infty)} u\exp -u\varepsilon^2 + \frac{1}{\ln(2)}\int_2^\infty \exp(-u\varepsilon^2)\mathrm{d}u \\
&\leq \frac{c}{\varepsilon^2}.
\end{aligned}$$

2) We define the clean event for the suboptimal arm k at time t as

$$\mathcal{E}_{k,t} = \left\{\hat{\mu}_k(t) \leq \mu_k + \sqrt{\frac{\ln(t/N_k(t))}{N_k(t)}} + \frac{\Delta_k}{3} \text{ and } \hat{\mu}_{k^*}(t) \geq \mu_k - \sqrt{\frac{\ln(t/N_k(t))}{N_k(t)}} - \frac{\Delta_k}{3}\right\}.$$

We have $\mathbb{P}(\neg\mathcal{E}_{k,t}) \leq \frac{18c}{t\Delta_k^2}$. Moreover, we can show that

$$\mathcal{E}_{k,t} \text{ and } a_{t+1} = k \implies N_k(t) \leq \frac{36}{\Delta_k^2}\ln(t/N_k(t)),$$

which can be rewritten for some constant c_1 as

$$\mathcal{E}_{k,t} \text{ and } a_{t+1} = k \implies N_k(t) \leq \frac{36}{\Delta_k^2}\left(\ln_+(t\Delta_k^2) + c_1\right).$$

We can then conclude using classical arguments.

For the bonus part, the trick is to bound the probability of the clean events, starting from $t = \lceil \frac{1}{\Delta_k^2} \rceil$.

3) a)

$$\begin{aligned}
\Delta_k\mathbb{E}[N_k(T)] &\leq c' \min\left(\Delta_k\mathbb{E}[N_k(T)], \frac{\ln_+(T\Delta_k^2)}{\Delta_k}\right) + c'\Delta_k \\
&\leq c' \sup_{\Delta>0}\min\left(\Delta T, \frac{\ln(T\Delta^2)}{\Delta}\right) + c' \\
&\leq c' \sup_{u>0}\min\left(u\frac{\mathbb{E}[N_k(T)]}{\sqrt{T}}, \frac{\sqrt{T}\ln_+(u^2)}{u}\right) + c' = c'\sqrt{T}\sup_{u>0}\min\left(\alpha_k u, \frac{\ln_+(u^2)}{u}\right) + c',
\end{aligned}$$

where $\alpha_k = \frac{\mathbb{E}[N_k(T)]}{T} \in [0,1]$. We can then use the admitted result to get

$$\Delta_k \mathbb{E}[N_k(T)] \leq c'' \sqrt{T} \max\left(\alpha_k, \sqrt{\alpha_k \ln(1/\alpha_k)}\right) + c'.$$

b) We have

$$\begin{aligned}
R_T &= \sum_k \Delta_k \mathbb{E}[N_k(T)] \\
&\leq c'\sqrt{T} \sum_k \left(\alpha_k + \sqrt{\alpha_k \ln(1/\alpha_k)}\right) + c'K \\
&\leq c'\sqrt{T} \sum_k \sqrt{\alpha_k \ln(1/\alpha_k)} + c'(K + \sqrt{T}) && \sum_k \alpha_k = 1 \\
&\leq c'\sqrt{KT} \sqrt{\sum_k \alpha_k} \sqrt{-\sum_k \frac{1}{K} \ln(\alpha_k)} + c'(K + \sqrt{T}) && \text{Cauchy-Schwarz} \\
&\leq c'\sqrt{KT} \sqrt{-\ln\left(\frac{1}{K}\sum_k \alpha_k\right)} + c'(K + \sqrt{T}) && -\ln \text{ is concave} \\
&\leq c'\sqrt{KT}\sqrt{\ln K} + c'(K + \sqrt{T}).
\end{aligned}$$

Exercise 4 :

Consider th K-armed stochastic contextual setting (setting 1 in lecture 7) and assume that $\mathcal{C} = [0,1]$ and the reward function is (L,α)-Hölder for $\alpha \in (0,1]$:

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

Build an algorithm with a regret bound (to prove) of order

$$R_T = \mathcal{O}\left(L^{\frac{1}{2\alpha+1}} K^{\frac{\alpha}{2\alpha+1}} T^{\frac{\alpha+1}{2\alpha+1}}\right).$$

Solution: The idea is to discretize $\mathcal{C}$ into M bins of size $1/M$ and run MOSS independently for each context bin.

The regret then scales as

$$\frac{TL}{M^\alpha} + \sum_{i=1}^{M} \sqrt{KT_i} \leq \frac{TL}{M^\alpha} + \sqrt{MKT}.$$

Taking $M = \left(L^2 \frac{T}{K}\right)^{\frac{1}{2\alpha+1}}$ leads to the result.

Exercise 5 :

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

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

1) Show that

$$\alpha^*\nu = \operatorname*{argmax}_{\alpha \in \mathcal{P}_K} \Phi(\nu, \alpha)$$

$$\text{where} \quad \Phi(\nu, \alpha) = \frac{1}{2} \min_{k \neq k^*} \frac{\alpha_{k^*} \alpha_k}{\alpha_{k^*} + \alpha_k} \Delta_k^2.$$

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.

Solution: 1) We are considering the optim problem

$$\sup_{\alpha} \inf_{\mu' \in \mathcal{M}_{\mathrm{alt}}(\mu)} \sum_k \alpha_k (\mu_k - \mu'_k)^2.$$

By continuity, we can extend $\inf_{\mu' \in \mathcal{M}_{\mathrm{alt}}(\mu)}$ to its closure. For a fixed α, the minimum over μ' is then reached for $\mu'_k = \mu_k$ except for $k = k^*$ and some suboptimal arm. I.e., for a fixed α, the infimum can be recast as

$$\inf_{\mu' \in \mathcal{M}_{\mathrm{alt}}(\mu)} \sum_k \alpha_k (\mu_k - \mu'_k)^2 = \min_{k \neq k^*} \inf_{x \in [0,1]} \alpha_{k^*} x^2 \Delta_k^2 + \alpha_k (1-x)^2 \Delta_k^2$$

$$\min_{k \neq k^*} \frac{\Delta_k^2}{\frac{1}{\alpha_{k^*}} + \frac{1}{\alpha_k}} \qquad \text{by noting that the minimal } x \text{ is } x_k = \frac{\alpha_k}{\alpha_k + \alpha_{k^*}}.$$

2) It is the minimum of concave functions.

3) The max over α is reached when all the $\frac{\Delta_k^2}{\frac{1}{\alpha_{k^*}} + \frac{1}{\alpha_k}}$ are equal, i.e. when for any $k, k' \neq k^*$

$$\frac{\Delta_k^2}{\frac{1}{\alpha_{k^*}} + \frac{1}{\alpha_k}} = \frac{\Delta_{k'}^2}{\frac{1}{\alpha_{k^*}} + \frac{1}{\alpha_{k'}}}.$$

Using the fact that $\sum_k \alpha_k = 1$, fixing the value of α_{k^*} then fixes the value of all α_k. From there for any $k \neq k^*$, noting $\Phi(\nu) = \max_{\alpha \in \mathcal{P}_K} \Phi(\nu, \alpha)$:

$$\alpha_k^* = \frac{2\alpha_{k^*}^* \Phi(\nu)}{\Delta_k^2 \alpha_{k^*}^* - 2\Phi(\nu)}.$$

Therefore,

$$\alpha_{k^*}^* + \sum_{k \neq k^*} \frac{2\alpha_{k^*}^* \Phi(\nu)}{\Delta_k^2 \alpha_{k^*}^* - 2\Phi(\nu)} = 1.$$

The solutions to this equation (in $\alpha_{k^*}^*$) are the roots of a polynomial, and are thus either finite or the polynomial is constant. The polynomial is obviously not constant here, so that there are a finite number of maximisers of $\max_{\alpha \in \mathcal{P}_K} \Phi(\nu, \alpha)$. The objective function is yet concave and thus either has a unique maximizer or an infinite number of maximizers. Hence, there is a unique maximizer $\alpha^*(\nu)$.

4) $\mathrm{argmax}_k\, \mathbb{E}(\nu_k)$ is constant in a neighborhood of ν. Hence by the previous part, Φ is continuous at (ν, α). Suppose that α^* is not continuous at ν. Then there exists a sequence (ν_n) converging to ν such that $\alpha^*(\nu_n) \not\to \alpha^*(\nu)$. By compactness, we can then extract a limit α_∞ of subsequence of $\alpha^*(\nu_n)$ such that $\alpha_\infty \neq \alpha^*(\nu)$. But then, we would have

$$\Phi(\alpha^*(\nu), \nu) = \lim_n \Phi(\alpha^*(\nu), \nu_n) \leq \lim_n \Phi(\alpha^*(\nu_{t_n}), \nu_{t_n}) = \Phi(\alpha_\infty, \nu).$$

By unicity of the maximizer, this then implies $\alpha_\infty = \alpha^*(\nu)$, so that α^* is continuous at ν.