MVA 2024-25 P. Gaillard and R. Degenne

Sequential Learning

Final Examination

The duration of the exam is 2 hours. A single two-sided sheet of handwritten notes (with any content) is allowed. Answers can be written in French or English.

The exam is divided into **two separate parts**, each of which should be submitted on a **separate sheet**.

Part 1. Online Convex Optimization

1. Let $\Theta \subset \mathbb{R}^d$ be a convex compact set and $(\ell_t : \Theta \to \mathbb{R})_{1 \le t \le T}$ be a sequence of convex functions.

   (a) Provide the pseudocode of Online Gradient Descent (OGD).

   Solution: Intialize $\theta_1 \in \Theta$, (η_t) sequence of learning rate, and for $t \geq 1$ update

   $$\theta_{t+1} = \mathrm{Proj}_\Theta\Big(\theta_t - \eta_t \nabla \ell_t(\theta_t)\Big)$$

   (b) Prove that OGD with step size η_t is equivalent to the update:

   $$\theta_{t+1} = \operatorname*{arg\,min}_{\theta \in \Theta} \big\{ \langle \nabla \ell_t(\theta_t), \theta \rangle + \lambda_t \|\theta - \theta_t\|^2 \big\}$$

   for some regularization parameter λ_t. Give the expression of λ_t as a function of η_t.

   Solution:

   $$\begin{aligned}\mathrm{Proj}_\Theta\Big(\theta_t - \eta_t \nabla \ell_t(\theta_t)\Big) &= \operatorname*{arg\,min}_{\theta \in \Theta} \Big\|\theta - \theta_t + \eta_t \nabla \ell_t(\theta_t)\Big\|^2 \\ &= \operatorname*{arg\,min}_{\theta \in \Theta} \Big\{ \|\theta - \theta_t\|^2 + 2\eta_t \langle \nabla \ell_t(\theta_t), \theta - \theta_t \rangle + \eta_t^2 \|\nabla \ell_t(\theta_t)\|^2 \Big\} \\ &= \operatorname*{arg\,min}_{\theta \in \Theta} \Big\{ \langle \nabla \ell_t(\theta_t), \theta \rangle + \frac{1}{2\eta_t} \|\theta - \theta_t\|^2 \Big\}\end{aligned}$$

   (c) What is the connexion between OGD and Online Mirror Descent (OMD)? (Justify briefly)

   Solution: The agile version of OMD defined by $\theta_{t+1} \in \arg\min_\theta \{\eta \langle \nabla \ell_t(\theta_t), \theta \rangle + D_\psi(\theta, \theta_t)\}$ for some mirror map ψ is equivalent to OGD when $\psi = \frac{1}{2}\|\cdot\|^2$ as shown by previous question.

2. Assume that ℓ_t are i.i.d.. Let $L = \mathbb{E}[\ell_t(\,\cdot\,)]$, which we assume μ-strongly convex, and $\theta_* \in \arg\min_{\theta\in\Theta} L(\theta)$. Let $R_T(\theta_*) := \sum_{t=1}^T \ell_t(\theta_t) - \ell_t(\theta_*)$ be the regret of some online algorithm.
   (a) Design a point $\bar\theta_T$ that controls its excess risk $\mathbb{E}\big[L(\bar\theta_T) - L(\theta_*)\big]$ as a function of $\mathbb{E}\big[R_T(\theta_*)\big]$.

Solution: Defining $\bar\theta_T = \frac{1}{T}\sum_{t=1}^T \theta_t$ we have by convexity

$$\begin{aligned}\mathbb{E}\big[L(\bar\theta_T) - L(\theta_*)\big] &= \mathbb{E}\left[L\Big(\frac{1}{T}\sum_{t=1}^T \theta_t\Big) - L(\theta_*)\right]\\ &\leq \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^T L(\theta_t) - L(\theta_*)\right]\\ &= \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^T \ell_t(\theta_t) - \ell_t(\theta_*)\right] = \frac{\mathbb{E}[R_T(\theta_*)]}{T}\,.\end{aligned}$$

   (b) Prove an upper bound on $\mathbb{E}\big[\|\bar\theta_T - \theta_*\|^2\big]$.

Solution: Since L is μ-strongly convex, we have

$$\|\bar\theta_T - \theta_*\|^2 \leq \frac{2}{\mu}\Big(L(\bar\theta_T) - L(\theta_*) + \underbrace{\langle \nabla L(\theta_*), \theta_* - \bar\theta_T\rangle}_{\leq 0}\Big)$$

Taking the expectation we conclude $\mathbb{E}\big[\|\bar\theta_T - \theta_*\|^2\big] \leq \frac{2}{\mu T}\mathbb{E}\big[R_T(\theta_*)\big]$.

Problem: Optimistic Follow The Regularized Leader

Let $\Theta \subseteq \mathbb{R}^d$ such that $\theta_1 := 0 \in \Theta$. We assume that at each $t \geq 1$, the learner tries to guess the next gradient with some $\widehat{g}_{t+1} \in \mathbb{R}^d$ and updates

$$\theta_{t+1} \in \operatorname*{arg\,min}_{\theta\in\Theta} \Phi_t(\theta), \qquad \text{with} \quad \Phi_t(\theta) := \big\langle \theta, \textstyle\sum_{s=1}^t g_s + \widehat{g}_{t+1}\big\rangle + \frac{\lambda}{2}\|\theta\|^2$$

where $g_s = \nabla \ell_s(\theta_s)$ and $\lambda > 0$ is a regularization parameter. We assume $\widehat{g}_1 = 0$ and $\Phi_0(\theta) = \frac{\lambda}{2}\|\theta\|^2$.

3. Show that for all $t \geq 1$ and $\theta_* \in \Theta$, $\Phi_{t-1}(\theta_t) - \Phi_{t-1}(\theta_*) \leq -\frac{\lambda}{2}\|\theta_t - \theta_*\|^2$.

Solution: Let $\theta \in \Theta$. Since $\theta_t \in \arg\min_\theta \Phi_{t-1}(\theta)$, we have $\langle \nabla\Phi_{t-1}(\theta_t), \theta_t - \theta\rangle \leq 0$, which entails by λ-strong convexity of $\frac{\lambda}{2}\|\cdot\|^2$ and thus Φ_{t-1},

$$\Phi_{t-1}(\theta_t) - \Phi_t(\theta) \leq \langle \nabla\Phi_{t-1}(\theta_t), \theta_t - \theta\rangle - \frac{\lambda}{2}\|\theta_t - \theta\|^2 \leq -\frac{\lambda}{2}\|\theta_t - \theta_*\|^2\,.$$

4. Define $\tilde\theta_{t+1} \in \arg\min_{\theta\in\Theta}\big\{\langle \theta, \sum_{s=1}^t g_s\rangle + \frac{\lambda}{2}\|\theta\|^2\big\}$. Show that for all $t \geq 1$ and $\theta_* \in \Theta$

$$\Phi_{t-1}(\theta_t) \leq \langle \theta_*, \textstyle\sum_{s=1}^t g_s\rangle + \langle \tilde\theta_{t+1}, \widehat{g}_t - g_t\rangle - \frac{\lambda}{2}\|\tilde\theta_{t+1} - \theta_t\|^2 + \frac{\lambda}{2}\|\theta_*\|^2$$

Solution:

$$\begin{aligned}
\Phi_{t-1}(\theta_t) &\le \Phi_{t-1}(\tilde\theta_{t+1}) - \frac{\lambda}{2}\|\tilde\theta_{t+1} - \theta_t\|^2 \\
&= \big\langle \theta, \textstyle\sum_{s=1}^{t-1} g_s + \widehat{g}_t \big\rangle + \tfrac{\lambda}{2}\|\tilde\theta_{t+1}\|^2 - \tfrac{\lambda}{2}\|\tilde\theta_{t+1} - \theta_t\|^2 \qquad \leftarrow \text{by Q.4} \\
&= \big\langle \tilde\theta_{t+1}, \textstyle\sum_{s=1}^{t} g_s \big\rangle + \tfrac{\lambda}{2}\|\tilde\theta_{t+1}\|^2 + \langle \tilde\theta_{t+1}, \widehat{g}_t - g_t\rangle - \tfrac{\lambda}{2}\|\tilde\theta_{t+1} - \theta_t\|^2 \\
&\le \big\langle \theta_*, \textstyle\sum_{s=1}^{t} g_s \big\rangle + \tfrac{\lambda}{2}\|\theta_*\|^2 + \langle \tilde\theta_{t+1}, \widehat{g}_t - g_t\rangle - \tfrac{\lambda}{2}\|\tilde\theta_{t+1} - \theta_t\|^2 \qquad \leftarrow \text{by def of } \tilde\theta_{t+1}
\end{aligned}$$

5. Deduce by induction on t that for all $t \ge 1$ and $\theta_* \in \Theta$:

$$\sum_{s=1}^{t} \langle \theta_s - \tilde\theta_{s+1}, \widehat{g}_s\rangle + \langle \tilde\theta_{s+1}, g_s\rangle \le \langle \theta_*, \textstyle\sum_{s=1}^{t} g_t\rangle - \tfrac{\lambda}{2}\textstyle\sum_{s=1}^{t} \|\theta_t - \tilde\theta_{t+1}\|^2 + \tfrac{\lambda}{2}\|\theta_*\|^2 \,.$$

Solution: The base case $t = 1$ is immediate since $\widehat{g}_1 = 0$. Assume the above inequality holds at $t - 1$, then applying it with $\theta_* = \theta_t$ yields

$$\begin{aligned}
&\sum_{s=1}^{t} \langle \theta_s - \tilde\theta_{s+1}, \widehat{g}_s\rangle + \langle \tilde\theta_{s+1}, g_s\rangle \\
&\quad \overset{\text{induction}}{\le} \quad \langle \theta_t, \textstyle\sum_{s=1}^{t-1} g_s\rangle + \tfrac{\lambda}{2}\|\theta_t\|^2 - \tfrac{\lambda}{2}\textstyle\sum_{s=1}^{t-1}\|\tilde\theta_{s+1} - \theta_s\|^2 + \langle \theta_t - \tilde\theta_{t+1}, \widehat{g}_t\rangle + \langle \tilde\theta_{t+1}, g_t\rangle \\
&\quad = \quad \Phi_{t-1}(\theta_t) - \langle \tilde\theta_{t+1}, \widehat{g}_t - g_t\rangle - \frac{\lambda}{2}\sum_{s=1}^{t-1}\|\tilde\theta_{s+1} - \theta_s\|^2 \\
&\quad = \quad \langle \theta_*, \textstyle\sum_{s=1}^{t} g_s\rangle - \tfrac{\lambda}{2}\textstyle\sum_{s=1}^{t}\|\tilde\theta_{s+1} - \theta_s\|^2 + \tfrac{\lambda}{2}\|\theta_*\|^2
\end{aligned}$$

6. Deduce that for any $\theta_* \in \Theta$

$$\sum_{t=1}^{T} \langle \theta_t - \theta_*, g_t\rangle \le \sum_{t=1}^{T} \langle \theta_t - \tilde\theta_{t+1}, g_t - \widehat{g}_t\rangle - \frac{\lambda}{2}\|\tilde\theta_{t+1} - \theta_t\|^2 + \frac{\lambda}{2}\|\theta_*\|^2 \,.$$

Solution:

$$\begin{aligned}
\sum_{t=1}^{T} \langle \theta_t - \theta_*, g_t\rangle &= \sum_{t=1}^{T} \langle \theta_t - \tilde\theta_{t+1}, g_t - \widehat{g}_t\rangle + \langle \theta_t - \tilde\theta_{t+1}, \widehat{g}_t\rangle + \langle \tilde\theta_{t+1}, g_t\rangle - \langle \theta_*, g_t\rangle \\
&\le \sum_{t=1}^{T} \langle \theta_t - \tilde\theta_{t+1}, g_t - \widehat{g}_t\rangle - \frac{\lambda}{2}\|\tilde\theta_{t+1} - \theta_t\|^2 + \frac{\lambda}{2}\|\theta_*\|^2
\end{aligned}$$

where the inequality is by previous question.

7. Conclude by proving a regret upper bound that depends on $V_T := \sum_{t=1}^T \|g_t - \widehat{g}_t\|^2$ for a well-optimized λ to be explicitly indicated.

Solution:

$$\begin{aligned}
\sum_{t=1}^T \ell_t(\theta_t) - \ell_t(\theta_*) &\le \sum_{t=1}^T \langle \theta_t - \theta_*, g_t \rangle \qquad \leftarrow \text{by convexity} \\
&\le \sum_{t=1}^T \langle \theta_t - \tilde{\theta}_{t+1}, g_t - \widehat{g}_t \rangle - \frac{\lambda}{2}\|\tilde{\theta}_{t+1} - \theta_t\|^2 + \frac{\lambda}{2}\|\theta_*\|^2 \\
&\le \sum_{t=1}^T \frac{\lambda}{2}\cancel{\|\theta_t - \tilde{\theta}_{t+1}\|^2} + \frac{1}{2\lambda}\|g_t - \widehat{g}_t\|^2 - \frac{\lambda}{2}\cancel{\|\tilde{\theta}_{t+1} - \theta_t\|^2} + \frac{\lambda}{2}\|\theta_*\|^2 \\
&= \|\theta_*\| \sqrt{\textstyle\sum_{t=1}^T \|g_t - \widehat{g}_t\|^2}
\end{aligned}$$

for $\lambda = \|\theta_*\|^{-1}\sqrt{\sum_{t=1}^T \|g_t - \widehat{g}_t\|^2}$.

Part 2. Stochastic bandits

8. Give an example of a stochastic bandit problem on which the Follow The Leader algorithm has linear expected regret. Prove that linear lower bound on the expected regret.

Solution: Bernoulli bandit with two arms with means $\mu_1 > \mu_2$. With probability $(1 - \mu_1)\mu_2$ the first rewards seen are 0 for arm 1 and 1 for arm 2. Then the empirical mean of arm 2 is positive while the empirical mean of arm 1 is 0: FTL will pull arm 2, and the empirical mean of arm 2 remains positive while the empirical mean of arm 1 is still 0. FTL pulls arm 2 for $T - 1$ rounds and gets regret $(\mu_1 - \mu_2)(T - 1)$. The expected regret is bounded from below by $(1 - \mu_1)\mu_2(\mu_1 - \mu_2)(T - 1)$.

9. In fixed confidence best arm identification (BAI), an algorithm is δ-correct if it returns the best arm with probability at least $1 - \delta$.

   (a) If an algorithm is δ-correct on all bandits with Gaussian rewards, it satisfies a lower bound on its expected stopping time $\mathbb{E}[\tau_\delta]$. How does that lower bound depend on δ, for small δ?

   Solution: $\log(1/\delta)$ (or $kl(\delta, 1 - \delta)$ but that's the same as $\delta \to 0$)

   (b) Give an example of a δ-correct algorithm for BAI with Gaussian rewards with variance 1. Note that we are not asking for an algorithm with small sample complexity: any δ-correct algorithm suffices.

Solution: Sample arms uniformly. Maintain confidence intervals on each arm (based on Hoeffding's inequality for example). Stop when the interval of the best arm (empirically) does not intersect any other interval. When the algorithm stops, the arm is the best arm unless one of the intervals did not contain the true mean of its arm. The intervals are tuned such that this happens with probability less than δ.

Problem: ε-greedy algorithm for stochastic bandits

We consider the stochastic bandit setting: an algorithm sequentially interacts with $K \in \mathbb{N}$ arms with $K > 1$, where each arm $k \in \{1, \dots, K\}$ is described by a distribution ν_k supported on $[0, 1]$ with mean μ_k. We suppose that $\mu_1 > \mu_k$ for all $k \in \{2, \dots, K\}$. We call $\Delta_k = \mu_1 - \mu_k$ the gap of arm k. When the algorithm pulls arm k_t at time t, it observes a reward X_{t,k_t} sampled from ν_{k_t}.

For $i \in \mathbb{N}$ with $i \geq 1$, we write $[i] = \{1, \dots, i\}$. For two propositions p_1 and p_2, the expression $p_1 \wedge p_2$ means p_1 and p_2.

The ε-greedy algorithm depends on a sequence of parameters $\varepsilon_1, \varepsilon_2, \dots$ in $[0, 1]$. First, the algorithm pulls each arm once. Then at time $t > K$, let $N^k_{t-1} = \sum_{s=1}^{t-1} \mathbb{1}_{\{k_s = k\}}$ be the number of times arm k was chosen up to time $t-1$ and let $\widehat{\mu}^k_{t-1} = \frac{1}{N^k_{t-1}} \sum_{s=1}^{t-1} X_{s,k_s} \mathbb{1}_{\{k_s = k\}}$. With probability $1 - \varepsilon_t$, the ε-greedy algorithm pulls the arm $k_t = \arg\max_k \widehat{\mu}^k_{t-1}$; with probability ε_t, it pulls an arm uniformly at random. Let Z_t be the Bernoulli random variable with expectation ε_t with value 1 if the arm is chosen uniformly and 0 otherwise.

Let the regret of the algorithm at time T be $R_T = T\mu_1 - \sum_{t=1}^T \mu_{k_t}$.

10. Prove that $\mathbb{E}[R_T] = \sum_{k=2}^K \Delta_k \mathbb{E}[N^k_T]$.

Solution: See lecture notes.

11. What is the expectation m_T of $\sum_{t=1}^T Z_t$? Give an upper bound on $\mathbb{P}\left\{\sum_{t=1}^T Z_t - m_T \geq Tx\right\}$ that is exponentially decreasing in T.

Solution: $m_T = \sum_{t=1}^T \varepsilon_t$.

Since $Z_t - \varepsilon_t$ is bounded in an interval of length 1, it is (1/4)-sub-Gaussian. By Hoeffding's inequality,

$$\mathbb{P}\left\{\sum_{t=1}^T Z_t - \sum_{t=1}^T \varepsilon_t \geq Tx\right\} \leq \exp(-2Tx^2) \,.$$

In the next questions, we suppose that $\varepsilon_t = \varepsilon \in [0, 1]$ for all $t \in \mathbb{N}$.

We introduce the notation $N^k_{Z,t} = \sum_{s=1}^t \mathbb{1}_{\{Z_s = 1 \,\wedge\, k_s = k\}}$, which corresponds to the number of pulls of arm k due to the uniform exploration. Let $\mathcal{E}_T$ be the event that for all $k \in [K]$ and $t \in [T]$, $\left|N^k_{Z,t} - t\frac{\varepsilon}{K}\right| \leq \sqrt{\frac{t}{2}\log(2KT^2)}$.

12. Prove that

$$\mathbb{E}[R_T] \le T\mathbb{P}(\mathcal{E}_T^c)\max_k \Delta_k + \sum_{k=1}^K \Delta_k \sum_{t=1}^T \mathbb{P}\{\mathcal{E}_T \ \wedge\ Z_t = 0 \ \wedge\ k_t = k\} + \sum_{k=1}^K \Delta_k \mathbb{E}[N_{Z,T}^k \mathbb{1}_{\{\mathcal{E}_T\}}] \,. \quad (1)$$

Solution:

$$\begin{aligned}\mathbb{E}[R_T] &= \mathbb{E}[\sum_{t=1}^T \Delta_{k_t}] \\ &= \mathbb{E}[\sum_{t=1}^T \Delta_{k_t}\mathbb{1}_{\{\mathcal{E}_T^c\}}] + \mathbb{E}[\sum_{t=1}^T \Delta_{k_t}\mathbb{1}_{\{\mathcal{E}_T\}}]\end{aligned}$$

The first term is less than $T\Delta_{\max}\mathbb{P}(\mathcal{E}_T^c)$, which is the first term in the bound we want to prove.

$$\begin{aligned}\mathbb{E}[\sum_{t=1}^T \Delta_{k_t}\mathbb{1}_{\{\mathcal{E}_T\}}] &= \sum_{k=1}^K \Delta_k \mathbb{E}[\sum_{t=1}^T \mathbb{1}_{\{\mathcal{E}_T \wedge k_t = k\}}] \\ &= \sum_{k=1}^K \Delta_k \mathbb{E}[\sum_{t=1}^T \mathbb{1}_{\{\mathcal{E}_T \wedge Z_t=0 \wedge k_t = k\}}] + \sum_{k=1}^K \Delta_k \mathbb{E}[\sum_{t=1}^T \mathbb{1}_{\{\mathcal{E}_T \wedge Z_t=1 \wedge k_t = k\}}] \\ &= \sum_{k=1}^K \Delta_k \sum_{t=1}^T \mathbb{P}\{\mathcal{E}_T \ \wedge\ Z_t = 0 \ \wedge\ k_t = k\} + \sum_{k=1}^K \Delta_k \mathbb{E}[\mathbb{1}_{\{\mathcal{E}_T\}} \sum_{t=1}^T \mathbb{1}_{\{Z_t=1 \wedge k_t=k\}}] \,.\end{aligned}$$

13. (a) For $t \ge 1$, $k \in [K]$, what is the law of the random variable with value 1 if both $Z_t = 1$ and $k_t = k$, and value 0 otherwise?

Solution: It's a Bernoulli(ε/K).

(b) Let $\delta \in (0,1)$, $t \in [T]$ and $k \in [K]$. By showing two concentration inequalities and doing an union bound, prove that with probability $1-\delta$,

$$\left|N_{Z,t}^k - t\frac{\varepsilon}{K}\right| \le \sqrt{\frac{t}{2}\log\frac{2}{\delta}} \,. \quad (2)$$

Deduce that with probability $1-\frac{1}{T}$, for all $k \in [K]$ and $t \in [T]$, $\left|N_{Z,t}^k - t\frac{\varepsilon}{K}\right| \le \sqrt{\frac{t}{2}\log(2KT^2)}$. That is, $\mathbb{P}(\mathcal{E}_T) \ge 1 - \frac{1}{T}$.

Solution: $N_{Z,t}^k$ is the sum of t Bernoulli random variables with expectation ε/K (which correspond to the uniform exploration pulls of k). The first inequality is the result of Hoeffding's inequality twice, once for each tail, and an union bound over the two events. The second inequality is the result of union bounds over $k \in [K]$ and $t \in [T]$ and the choice $\delta = 1/T$.

14. (a) Suppose that $T \geq \frac{2K^2}{\varepsilon^2}\log(2KT^2)$. For $t \in [T]$ such that $t \geq \frac{2K^2}{\varepsilon^2}\log(2KT^2)$ and $k \neq 1$, show an upper bound on $\mathbb{P}\{\mathcal{E}_T \wedge \widehat{\mu}_t^k > \widehat{\mu}_t^1\}$ of the form $C_1 t \exp(-tC_2)$ where C_1 and C_2 may depend on the parameters of the problem but not on t.

Solution: Using the fact that $N_t^k \geq N_{Z,t}^k$, the definition of $\mathcal{E}_T$ and finally the lower bound on t, we get

$$\begin{aligned}\mathbb{P}\{\mathcal{E}_T \wedge \widehat{\mu}_t^k > \widehat{\mu}_t^1\} &\leq \mathbb{P}\{\min\{N_t^k, N_t^1\} \geq t\frac{\varepsilon}{K} - \sqrt{\frac{t}{2}\log(2KT^2)} \wedge \widehat{\mu}_t^k > \widehat{\mu}_t^1\} \\ &\leq \mathbb{P}\{\min\{N_t^k, N_t^1\} \geq t\frac{\varepsilon}{2K} \wedge \widehat{\mu}_t^k > \widehat{\mu}_t^1\}\,.\end{aligned}$$

If $\widehat{\mu}_t^k > \widehat{\mu}_t^1$ then either $\widehat{\mu}_t^k \geq \mu_k + \Delta_k/2$ or $\widehat{\mu}_t^1 \leq \mu_1 - \Delta_k/2$. By a union bound,

$$\begin{aligned}&\mathbb{P}\{\min\{N_t^k, N_t^1\} \geq t\frac{\varepsilon}{2K} \wedge \widehat{\mu}_t^k > \widehat{\mu}_t^1\} \\ &\leq \mathbb{P}\{N_t^k \geq t\frac{\varepsilon}{2K} \wedge \widehat{\mu}_t^k \geq \mu_k + \Delta_k/2\} + \mathbb{P}\{N_t^1 \geq t\frac{\varepsilon}{2K} \wedge \widehat{\mu}_t^1 \leq \mu_1 - \Delta_k/2\}\,.\end{aligned}$$

We bound each of the two parts in a similar way. We first do an union bound over the possible values of N_t^k, then use Hoeffding's inequality.

$$\begin{aligned}\mathbb{P}\{N_t^k \geq t\frac{\varepsilon}{2K} \wedge \widehat{\mu}_t^k \geq \mu_k + \Delta_k/2\} &\leq \sum_{s=\lfloor t\frac{\varepsilon}{2K}\rfloor}^{t} \mathbb{P}\{N_t^k = s \wedge \widehat{\mu}_t^k \geq \mu_k + \Delta_k/2\} \\ &\leq \sum_{s=\lfloor t\frac{\varepsilon}{2K}\rfloor}^{t} \exp(-s\Delta_k^2/2) \\ &\leq t\exp(-t\frac{\varepsilon\Delta_k^2}{4K})\end{aligned}$$

(b) Deduce an upper bound on $\sum_{t=1}^T \mathbb{P}\{\mathcal{E}_T \wedge Z_t = 0 \wedge k_t = k\}$ for $k \neq 1$. Your bound can be expressed as a function of the quantity $C_{\exp}(a) := \sum_{t=1}^\infty te^{-ta}$.

Solution:

$$\begin{aligned}&\sum_{t=1}^T \mathbb{P}\{\mathcal{E}_T \wedge Z_t = 0 \wedge k_t = k\} \\ &\leq \frac{2K^2}{\varepsilon^2}\log(2KT^2) + \sum_{t \geq \frac{2K^2}{\varepsilon^2}\log(2KT^2)} 2t\exp\left(-t\frac{\varepsilon\Delta_k^2}{4K}\right) \\ &\leq \frac{2K^2}{\varepsilon^2}\log(2KT^2) + \sum_{t \geq \frac{2K^2}{\varepsilon^2}\log(2KT^2)} 2t\exp\left(-t\frac{\varepsilon\Delta_k^2}{4K}\right) \\ &\leq \frac{2K^2}{\varepsilon^2}\log(2KT^2) + 2C_{\exp}(\frac{\varepsilon\Delta_k^2}{4K})\,.\end{aligned}$$

(c) Prove that $\limsup_{T\to\infty} \frac{\mathbb{E}[R_T]}{T} \leq \frac{\varepsilon}{K}\sum_{k=2}^K \Delta_k$.

Solution: We proved in Question 12 that

$$\mathbb{E}[R_T] \le T\mathbb{P}(\mathcal{E}_T^c)\max_k \Delta_k + \sum_{k=1}^K \Delta_k \sum_{t=1}^T \mathbb{P}\{\mathcal{E}_T \ \wedge \ Z_t = 0 \ \wedge \ k_t = k\} + \sum_{k=1}^K \Delta_k \mathbb{E}[N_{Z,T}^k \mathbb{1}_{\{\mathcal{E}_T\}}] \,. \tag{3}$$

We also proved $\mathbb{P}(\mathcal{E}_T^c) \le 1/T$ in Question 13, such that the first term of the right hand side vanishes when divided by $1/T$. So we have

$$\begin{aligned}\limsup_{T\to\infty} \frac{\mathbb{E}[R_T]}{T} \le \limsup_{T\to\infty} \frac{1}{T} \sum_{k=1}^K \Delta_k \sum_{t=1}^T \mathbb{P}\{\mathcal{E}_T \ \wedge \ Z_t = 0 \ \wedge \ k_t = k\} \\ + \limsup_{T\to\infty} \frac{1}{T} \sum_{k=1}^K \Delta_k \mathbb{E}[N_{Z,T}^k \mathbb{1}_{\{\mathcal{E}_T\}}] \,.\end{aligned}$$

Question 14b shows that the first term vanishes.

The definition of $\mathcal{E}_T$ gives $N_{Z,T}^k \le T\frac{\varepsilon}{K} + o(T)$. The only non-vanishing part corresponds to that $T\frac{\varepsilon}{K}$ upper bounds and gives the result.

15. Prove that $\lim_{T\to\infty} \frac{\mathbb{E}[R_T]}{T} = \frac{\varepsilon}{K}\sum_{k=2}^K \Delta_k$.

Solution: One inequality is given by the previous question. We now prove the other one.

We have the lower bound on the regret $\mathbb{E}[R_T] \ge \sum_{k=1}^K \Delta_k \mathbb{E}[N_{Z,T}^k \mathbb{1}_{\{\mathcal{E}_T\}}]$.

Then use the definition of $\mathcal{E}_T$ to get a lower bound $T\frac{\varepsilon}{K} - o(T)$ for $N_{Z,T}^k$ and use that $\mathbb{P}(\mathcal{E}_T) \ge 1 - \frac{1}{T}$ to get

$$\begin{aligned}\mathbb{E}[R_T] &\ge \sum_{k=2}^K \Delta_k \mathbb{E}[N_{Z,T}^k \mathbb{1}_{\{\mathcal{E}_T\}}] \\ &\ge \sum_{k=2}^K \Delta_k T\frac{\varepsilon}{K} - o(T) \,.\end{aligned}$$