Homework: Sequential Learning

What I care about. I care about well-written proofs: with sufficient details, with calculations worked out and leading to pleasant and readable bounds. I favor quality of the writing over the quantity of questions answered. I give bonus points for elegant solutions.

Formats of your submission, deadline. Please send your final homework by email at `etienne.boursier@inria.fr`. I expect to receive PDF files, with answers either handwritten and neatly scanned or typed in LATEX. The homework can be written in **either French or English**, depending on your own preference. Deadline is Wednesday, November 13, at 6pm. This is a strict deadline: submitting after this deadline will negatively impact your grade, with the impact depending on the delay.

Beware: Typos. Most likely the statement comes with typos. This is part of the job. Try to correct them on your own!

Exercise 1: Learning with experts with sparse losses.

The aim of this exercise is to study what happens when both a non-negativity and a sparsity assumptions are made on the vectors of losses picked by the opponent. More formally, we consider the setting of expert learning with linear losses. We consider N experts, where at each round t, at most s components are positive while the other components are null. The parameter $s \in \{1, \dots, N\}$ is fixed throughout the game but is unknown to the agent (algorithm). The online protocol is the following. At each time step $t \in \mathbb{N}$:

1. The agents picks a convex combination $(p_{j,t})_{1\leq j\leq N}$, while the environment simultaneously picks a loss vector $(\ell_{j,t})_{1\leq j\leq N} \in [0,1]^N$, with at most s non-null components;
2. the choices are publicly revealed.

The agent aims to minimise the regret

$$R_T = \sum_{t=1}^{T}\sum_{j=1}^{N} p_{j,t}\ell_{j,t} - \min_{j\in[N]} \sum_{t=1}^{T} \ell_{j,t}.$$

The goal of this exercise is to determine the optimal order of magnitude of the regret under the non-negativity and sparsity assumptions.

Lower bound on the regret.

Consider the joint distribution over $\{0,1\}^N$ defined as the law of a random vector $\mathbf{L} = (L_1, \dots, L_N)$ drawn in two steps. First, we pick s components uniformly at random among $\{1, \dots, N\}$; we call

them $K_1, \dots, K_s$. Then, the components not picked ($k \neq K_j$ for all j) are associated with zero losses, $L_k = 0$. The losses L_k for picked components $K_1, \dots, K_s$ are then drawn according to a Bernoulli distribution with parameter $\frac{1}{2}$. The loss vector $L \in [0,1]^N$ thus generated is indeed s–sparse and non-negative. We fix an algorithm for the agent, consider an i.i.d. sequence $\mathbf{L}_1, \mathbf{L}_2, \dots$ of random vectors thus generated, and study the corresponding regret.

$$R_T = \sum_{t=1}^{T} \sum_{j=1}^{N} p_{j,t} L_{j,t} - \min_{j \in [N]} \sum_{t=1}^{T} L_{j,t}.$$

1. Show that the expectation of the regret can be written as

$$\mathbb{E}\left[\frac{R_T}{\sqrt{T}}\right] = \mathbb{E}\left[\max_{i \in [N]} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_t^{(i)}\right],$$

where the $(X_t^{(1)}, \dots, X_t^{(N)})$ are i.i.d. (in t) centered random vectors taking values in $[-1,1]^N$ **that do not depend on** p_t, with some covariance matrix denoted by Γ. Give a closed-form definition of the $X_t^{(i)}$ based on the $L_{i,t}$ and an explicit expression for Γ.

2. Explain why

$$\mathbb{E}\left[\max_{i \in [N]} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_t^{(i)}\right] \longrightarrow \mathbb{E}\left[\max_{i \in [N]} Z^{(i)}\right]$$

when $T \to \infty$, where $(Z_1, \dots, Z_N)$ follows the normal distribution $\mathcal{N}(0, \Gamma)$, i.e., the centered normal distribution with covariance matrix Γ.

3. Consider the Gaussian random vector $(W_1, \dots, W_N)$ with i.i.d. components W_i with distribution $\mathcal{N}(0, \mathrm{Var}(Z_1))$. Show that Slepian's lemma (stated below) is applicable and that it entails

$$\mathbb{E}\left[\max_{i \in [N]} Z^{(i)}\right] \geq \mathbb{E}\left[\max_{i \in [N]} W^{(i)}\right].$$

4. Conclude to an asymptotic lower bound of the order of $\sqrt{\frac{Ts\ln(N)}{N}}$; state it carefully and rigorously.

Lemma (Slepian's lemma, 1962)**.** *Let $(Z_1, \dots, Z_N)$ and $(W_1, \dots, W_N)$ be two centered Gaussian random vectors in $\mathbb{R}^N$. If for all $i \in [N]$, $\mathbb{E}[Z_i^2] = \mathbb{E}[W_i^2]$ and*

$$\forall i, j \in [N], \mathbb{E}[Z_i Z_j] \leq \mathbb{E}[W_i W_j],$$

then for all $t \in \mathbb{R}$:

$$\mathbb{P}\left[\max_{i \in [N]} Z_i \geq t\right] \geq \mathbb{P}\left[\max_{i \in [N]} W_i \geq t\right].$$

Upper bound on the regret

5. Consider in this question the non-sparse setting considered in the course (with losses in $[0,1]$).

(a) First show that

$$\sum_{t=1}^{T}\sum_{j=1}^{N} p_{j,t}(\ell_{j,t} - \sum_{k=1}^{N} p_{k,t}\ell_{k,t})^2 \leq \sum_{t=1}^{T}\sum_{j=1}^{N} p_{j,t}\ell_{j,t}$$

(b) We admit for the tuned EWA (Exercise session 1, Exercise 3) that

$$\sum_{t=1}^{T}\sum_{j=1}^{N} p_{j,t}\ell_{j,t} - \min_{k\in[N]}\sum_{t=1}^{T}\ell_{k,t} \leq 2\sqrt{\ln(N)\sum_{t=1}^{T}\sum_{j=1}^{N} p_{j,t}\ell_{j,t}} + \left(2 + \frac{4}{3}\ln(N)\right).$$

Show that it here leads to a regret bound of order $\ln(N) + \sqrt{\ln(N)\min_{k\in[N]}\sum_{t=1}^{T}\ell_{k,t}}$.

6. Show that tuned EWA leads in the sparse setting to a regret bound scaling as

$$R_T \leq \mathcal{O}\left(\ln(N) + \sqrt{\ln(N)\frac{Ts}{N}}\right).$$

Comment on the optimality of the bound and compare it with the non-sparse case.

Exercise 2: The (α, ψ)–UCB algorithm

Let $\psi : \mathbb{R} \to \mathbb{R}$ be a convex function such that $\psi(x) = \psi(-x)$ for all $x \in \mathbb{R}$. Consider a bandit model $\mathcal{D}$ such that for all $\nu \in \mathcal{D}$, if X denotes a random variable with distribution ν, then

$$\forall \lambda \geq 0, \max\left\{\ln \mathbb{E}_\nu\left[e^{\lambda(X-\mathbb{E}[X])}\right], \ln \mathbb{E}_\nu\left[e^{-\lambda(X-\mathbb{E}[X])}\right]\right\} \leq \psi(\lambda). \tag{1}$$

For all $x \geq 0$, we define the convex conjugate of ψ,

$$\psi^\star(x) = \sup\left\{\lambda x - \psi(\lambda) \mid \lambda \geq 0\right\},$$

and assume that $\psi^\star$ is invertible, with inverse denoted by $(\psi^\star)^{-1}$.

1. Provide such a function ψ for the model $\mathcal{D} = \mathcal{P}([0,1])$ of all probability distributions over $[0,1]$. Compute $\psi^\star$ and its inverse.

We generalize the UCB algorithm for stochastic bandits in the following way. We consider the same setting and use the same notation as the ones used in class, with the exception that the reward distributions of the arms $\nu_k \in \mathcal{D}$ correspond to random variables $X_k(t) \in \mathbb{R}$, which satisfy Equation (1).

Algorithm: (α, ψ)-UCB
Input: $\alpha > 0$ and $\psi : \mathbb{R} \to \mathbb{R}$ with $\psi(x) = \psi(-x)$ for all $x \geq 0$

Play each arm once
for $t \geq K + 1$ **do**
  Pick an arm (ties broken arbitrarily)

$$a_t \in \operatorname*{argmax}_{k \in [K]} \hat{\mu}_k(t-1) + (\psi^\star)^{-1}\left(\frac{\alpha \ln t}{N_k(t-1)}\right)$$

  Observe reward $X_{a_t}(t)$ and update $\hat{\mu}_{a_t}$, N_{a_t} in consequence

We want to upper bound the regret of the (α, ψ)-UCB algorithm as follows: for $\alpha > 2$,

$$\mathbb{E}[R_T] \leq \sum_{k, \Delta_k > 0} \Delta_k \left(\frac{\alpha}{\psi^\star(\Delta_k/2)} \ln T + \frac{2\alpha}{\alpha - 2}\right). \tag{2}$$

To that end, we first show that for each arm k and $t \geq K + 1$, an upper confidence bound on μ_k is given by

$$\hat{\mu}_k(t-1) + (\psi^\star)^{-1}\left(\frac{\alpha \ln t}{N_k(t-1)}\right).$$

2. (a) Prove that for all $t \geq 1$ and all $\lambda \geq 0$,

$$\mathbb{E}\left[\exp\left(-\lambda(X_k(t) - \mu_k)\mathbb{1}_{a_t = k}\right) \mid \mathcal{F}_{t-1}\right] \leq \exp(\psi(\lambda)\mathbb{1}_{a_t = k}),$$

for a filtration $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$ to specify explicitly.

(b) Construct an $\mathcal{F}$-adapted supermartingale $(M_t)_{t \geq 0}$ based on this inequality.

3. Prove that for all $t \geq K + 1$, all $\ell \geq 1$, and all $\varepsilon > 0$,

$$\mathbb{P}\Big(\hat{\mu}_k(t-1) + \varepsilon \leq \mu_k \text{ and } N_k(t-1) = \ell\Big) \leq \exp(-\ell\psi^\star(\varepsilon)).$$

4. Provide a bound, for $t \geq K + 1$, on

$$\mathbb{P}\left(\hat{\mu}_k(t-1) + (\psi^\star)^{-1}\left(\frac{\alpha \ln t}{N_k(t-1)}\right) \leq \mu_k\right).$$

5. Briefly indicate how to bound, for $t \geq K + 1$,

$$\mathbb{P}\left(\hat{\mu}_k(t-1) - (\psi^\star)^{-1}\left(\frac{\alpha \ln t}{N_k(t-1)}\right) > \mu_k\right).$$

To establish the regret bound, we first fix a suboptimal arm j and an optimal arm $k^\star$.

6. Explain why $a_t = j$ for $t \geq K+1$ entails one of the following events:

$$\hat{\mu}_{k^\star}(t-1) + (\psi^\star)^{-1}\left(\frac{\alpha \ln t}{N_{k^\star}(t-1)}\right) \leq \mu^\star$$

$$\text{or} \quad \hat{\mu}_j(t-1) - (\psi^\star)^{-1}\left(\frac{\alpha \ln t}{N_j(t-1)}\right) > \mu_j,$$

$$\text{or} \quad N_j(t-1) < \frac{\alpha \ln t}{\psi^\star(\Delta_j/2)}.$$

7. Establish the regret bound given by Equation (2).

We conclude this exercise with a discussion of the bound for the model $\mathcal{D} = \mathcal{P}([0,1])$.

8. Provide also a distribution-free bound for (α, ψ)-UCB on this model, i.e., a bound over all distributions satisfying Equation (1). You need first to think of a suitable value for α.

Exercise 3: Phased Exploration

In this exercise, we consider the phased algorithm described below for some parameter $a \in \mathbb{N}^*$.

Algorithm: Phased Exploration
input: T, $a \in \mathbb{N}^*$

$\mathcal{K} \leftarrow [K]$
$\ell \leftarrow 0$
while $\mathrm{Card}(\mathcal{K}) > 1$ **do**
  For each arm in $\mathcal{K}$, pull it a^ℓ times
  for *all* $k \in \mathcal{K}$ *such that* $\hat{\mu}_k + \sqrt{\frac{2\ln T}{N_k(T)}} \leq \max_{i\in\mathcal{K}} \hat{\mu}_i - \sqrt{\frac{2\ln T}{N_i(T)}}$ **do** $\mathcal{K} \leftarrow \mathcal{K} \setminus \{k\}$
  $\ell \leftarrow \ell + 1$
repeat pull only arm in $\mathcal{K}$ **until** $t = T$

We consider the classical bandits setting of the course, with stochastic rewards $X_k(t) \in [0,1]$. Let $k^\star \in \mathrm{argmax}_k\, \mu_k$ and define the clean event

$$\mathcal{E} = \left\{ \forall t \in [T], \forall k \neq k^\star, \hat{\mu}_k(t) - \mu_k \leq \sqrt{\frac{2\ln T}{N_k(t)}} \quad \text{and} \quad \hat{\mu}_{k^\star}(t) - \mu_{k^\star} \leq \sqrt{\frac{2\ln T}{N_{k^\star}(t)}} \right\} \tag{3}$$

1. For $a = 1$, the above algorithm corresponds to an algorithm of the course. Which one? Also, show that $\mathcal{P}(\mathcal{E}) \geq 1 - \frac{K}{T^2}$.

2. Now consider $a \geq 2$ and prove that for any k such that $\Delta_k > 0$,

$$\mathbb{E}[N_k(T)\mathbb{1}_{\mathcal{E}}] \leq \left\lceil \frac{32a \ln T}{\Delta_k^2} \right\rceil.$$

3. From there, prove an upper bound (to precise) for the expected regret of this algorithm.

4. What is the role played by a? What are the pros and cons for taking a large a?