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.
This exam is made of 3 parts. The first part contains varied questions on the course. Parts 2 and 3 are exercices on adversarial and stochastic online learning respectively.

Part 1. Appetizers

1. Let $(x_1, y_1), \dots, (x_T, y_T)$ be a sequence of i.i.d. random variables following a distribution ν in $[-X, X]^d \times [-Y, Y]$ for some $X, Y > 0$ and $d \geq 1$. We consider the decision set $\Theta = \{\theta \in \mathbb{R}^d : \|\theta\|_1 \leq B\}$ and the loss $\ell_t(\theta) = \big(\langle \theta, x_t \rangle - y_t\big)^2$.
   (a) Give the definition of the (adversarial) regret R_T of an algorithm that chooses $\theta_t \in \Theta$ at each round t.
   (b) Provide the pseudo-code of an algorithm that controls the regret.
   (c) Give the regret upper-bound associated to the above algorithm.
   (d) What are the hyper-parameters of the algorithm? How would you calibrate them?
   (e) Denoting by $\bar{\theta}_T = \frac{1}{T}\sum_{t=1}^T \theta_t$ the average iterate, show that
   $$\mathcal{R}(\bar{\theta}_t) - \inf_{\theta \in \Theta} \mathcal{R}(\theta) \leq \frac{\mathbb{E}[R_T]}{T} \quad \text{where} \quad \mathcal{R}(\theta) = \mathbb{E}\big[(\langle \theta, X \rangle - Y)^2\big], \quad (X, Y) \sim \nu\,.$$
   (f) What property (give the definition) of ℓ_t could be used to improve the rate?
2. 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 regret.
3. In stochastic bandits, what are the drawbacks of the Explore-Then-Commit algorithm compared to UCB?

Part 2. Internal regret

We consider the problem of prediction with expert advice. At each round $t = 1, \dots, T$, a learner chooses a weight vector $p_t \in \Delta_K = \{p \in \mathbb{R}_+^K : \sum_k p(k) = 1\}$, samples $k_t \sim p_t$, observes a loss vector $\ell_t \in [0,1]^K$ and suffers the loss $\ell_t(k_t)$. We would like to minimize the internal regret defined as:

$$R_T^{(\mathrm{int})} = \max_{1 \le i,j \le K} \mathbb{E}\left[\sum_{t=1}^T \big(\ell_t(k_t) - \ell_t(j)\big)\mathbb{1}_{\{k_t = i\}}\right].$$

Basically, a player has small internal regret if for all pairs of action $(i,j) \in [K]^2$, he does not regret of not having chosen action $j \in [K]$ when he selected $k_t = i$.

4. Denote by $R_T = \max_i \mathbb{E}\Big[\sum_{t=1}^T \ell_t(k_t) - \ell_t(i)\Big]$ the standard regret. We show here that internal regret is a stronger notion of regret.
   (a) Show that any algorithm with a sublinear internal regret $R_T^{(\mathrm{int})}$ has also a sublinear standard regret R_T.
   (b) Provide, for $K = 3$, a sequence of losses $\ell_1, \dots, \ell_T \in [0,1]^3$ and a sequence of decisions $k_1, \dots, k_T \in [K]$ to show that an algorithm can have $R_T^{(\mathrm{int})} = T/3$ although $R_T = 0$.

Now, we would like to design a low internal regret algorithm. For all $1 \le i \ne j \le K$, denote by $p_t^{i \to j} \in \Delta_K$ the vector obtained from p_t by putting probability mass 0 on i and $p_t(i) + p_t(j)$ on j.

5. Show that
$$R_T^{(\mathrm{int})} \le \mathbb{E}\left[\sum_{t=1}^T \langle p_t, \ell_t\rangle - \min_{i \ne j} \sum_{t=1}^T \langle p_t^{i\to j}, \ell_t\rangle\right].$$

6. Denoting by $W_t(i,j) = \exp\Big(-\eta \sum_{s=1}^t \langle p_s^{i\to j}, \ell_s\rangle\Big)$ and $W_t = \sum_{i\ne j} W_t(i,j)$.
   (a) Show that
$$W_t \le W_{t-1} \exp\left(\eta^2 - \eta \sum_{i\ne j} q_t(i,j)\langle p_t^{i\to j}, \ell_t\rangle\right).$$
   (b) Show that
$$W_T \ge \exp\left(-\eta \min_{i\ne j} \sum_{t=1}^T \langle p_t^{i\to j}, \ell_t\rangle\right).$$

7. Deduce that
$$R_T^{(\mathrm{int})} \le \eta T + \frac{2\log K}{\eta},$$
   and optimize in η.

8. Assume that instead of observing $\ell_t \in [0,1]^K$ the learner would only observe the bandit feedback $\ell_t(k_t) \in [0,1]$.
   (a) How would you modify the above algorithm?

Input: learning rate $\eta > 0$
Init: $q_t \in \Delta_E$ uniform distribution on $E := \{(i,j) \in [K]^2 : i \neq j\}$
For $t = 1$ to T do

- Define $p_t \in \Delta_K$ by solving the fixed-point equation (we accept that this can be solved)

$$p_t = \sum_{i \neq j} q_t(i,j) p_t^{i \to j} .$$

- Play $k_t \sim p_t$ and observe $\ell_t \in [0,1]^K$.
- For $(i,j) \in E$ update

$$q_t(i,j) = \frac{\exp\Big(- \eta \sum_{s=1}^{t-1} \langle p_s^{i \to j}, \ell_s \rangle \Big)}{\sum_{k \neq l} \exp\Big(- \eta \sum_{s=1}^{t-1} \langle p_s^{k \to l}, \ell_s \rangle \Big)} .$$

End for

Algorithm 1: Exponentially Weighted Average Forecaster for Internal Regret

(b) What internal regret do you expect in terms of K and T (short justification)?

Part 3. Stochastic bandits

We consider the stochastic bandit setting: an algorithm sequentially interacts with $K \in \mathbb{N}$ arms with $K > 1$, where each arm $k \in \{1, \ldots, K\}$ is described by a distribution ν_k supported on $[0,1]$ with mean μ_k. We suppose that $\mu_1 \geq \mu_k$ for all $k \in \{1, \ldots, K\}$. Let $T \in \mathbb{N}$ be such that $T \geq K + 1$.

For $i, j \in \mathbb{N}$ with $i \geq 1$ and $i \leq j$, we write $[i] = \{1, \ldots, i\}$ and $[i : j] = \{i, i+1, \ldots, j\}$.

We study a variant of the UCB algorithm, denoted by UCB(δ) and described in Algorithm 2.

Input: Confidence level $\delta \in (0,1)$
For $t = 1$ to K do

- Pull arm $k_t = t$
- Observe $X_{t,k_t} \sim \nu_{k_t}$

Set $N_{K,k} = 1$ and $\widehat{\mu}_{K,k} = X_{k,k}$ for all $k \in [K]$.
For $t = K+1$ to T do

- Pull arm $k_t = \arg\max_{k \in [K]} \widehat{\mu}_{t-1,k} + \sqrt{\frac{1}{2} \frac{\log(2KT^2/\delta)}{N_{t-1,k}}}$
- Observe $X_{t,k_t} \sim \nu_{k_t}$
- Update $N_{t,k_t} = N_{t-1,k_t} + 1$ and $\widehat{\mu}_{t,k_t} = \widehat{\mu}_{t-1,k_t} + \frac{1}{N_{t,k_t}}(X_{t,k_t} - \widehat{\mu}_{t-1,k_t})$. For $k \neq k_t$, set $N_{t,k} = N_{t-1,k}$ and $\widehat{\mu}_{t,k} = \widehat{\mu}_{t-1,k}$.

End for

Algorithm 2: UCB(δ)

In this part, we will call regret the quantity $R_T = T\mu_1 - \sum_{t=1}^{T} \mu_{k_t}$. We will bound the expected regret $\mathbb{E}[R_T]$ of UCB(δ).

9. Prove that for all $k \in [K]$ and $t \in [K+1:T]$, $\widehat{\mu}_{t,k} = \frac{1}{N_{t,k}} \sum_{s=1}^{t} X_{s,k_s} \mathbb{I}\{k_s = k\}$. Here $\mathbb{I}\{A\}$ is the indicator of event A, with value 1 if the event happens and 0 otherwise.

10. Let E_{bad} be the event $\{\exists t \in [K+1:T], \exists k \in [K], \ |\widehat{\mu}_{t,k} - \mu_k| > \sqrt{\frac{1}{2}\frac{\log(2KT^2/\delta)}{N_{t,k}}}\}$.
    (a) Prove that
    $$\mathbb{P}(E_{\text{bad}}) \le \sum_{t=K+1}^{T} \sum_{k=1}^{K} (\mathbb{P}(\widehat{\mu}_{t,k} - \mu_k > \sqrt{\frac{1}{2}\frac{\log(2KT^2/\delta)}{N_{t,k}}}) + \mathbb{P}(\widehat{\mu}_{t,k} - \mu_k < -\sqrt{\frac{1}{2}\frac{\log(2KT^2/\delta)}{N_{t,k}}})) \,.$$
    (b) Show that $\mathbb{P}(\widehat{\mu}_{t,k} - \mu_k > \sqrt{\frac{1}{2}\frac{\log(2KT^2/\delta)}{N_{t,k}}}) \le \frac{\delta}{2KT}$. Hint: $N_{t,k}$ is random, but takes values in $[1,t]$. We admit that the same bound is true for $\mathbb{P}(\widehat{\mu}_{t,k} - \mu_k < -\sqrt{\frac{1}{2}\frac{\log(2KT^2/\delta)}{N_{t,k}}})$.
    (c) Show that $\mathbb{P}(E_{\text{bad}}) \le \delta$.

11. Let E be the complement of E_{bad}, $E = \{\forall t \in [K+1:T], \forall k \in [K], |\widehat{\mu}_{t,k} - \mu_k| \le \sqrt{\frac{1}{2}\frac{\log(2KT^2/\delta)}{N_{t,k}}}\}$.
    (a) Write the regret as a sum over arms, using the suboptimality gaps $\Delta_k = \mu_1 - \mu_k$ for $k \in [K]$.
    (b) Show that for all $t \in [K+1:T]$, if event E holds then $\mu_{k_t} + 2\sqrt{\frac{1}{2}\frac{\log(2KT^2/\delta)}{N_{t-1,k_t}}} \ge \mu_1$.
    (c) Under event E, show that $N_{T,k} \le 1 + \frac{2\log(2KT^2/\delta)}{\Delta_k^2}$ for all $k \in [K]$ with $\Delta_k > 0$.
    (d) Use the questions above to give an upper bound on $\mathbb{E}[R_T \mathbb{I}\{E\}]$.

12. Give an upper bound on $\mathbb{E}[R_T]$, function of K, T, δ and the gaps $(\Delta_k)_{k\in[K]}$. Use that bound to show that for a well chosen $\delta \in (0,1)$, $\mathbb{E}[R_T] \le C_1 + C_2 \log T$ where C_1, C_2 can depend on parameters of the problem, but don't depend on T.

13. For a well chosen $\delta \in (0,1)$, prove an upper bound of the form $\mathbb{E}[R_T] \le C_1' + C_2'\sqrt{T\log T}$ where C_1', C_2' don't depend on T and C_2' does not depend on the gaps $(\Delta_k)_{k\in[K]}$ or the means $(\mu_k)_{k\in[K]}$.