Exercise session n°1 : learning with experts and concentration inequalities

In this session, we consider online learning with experts (see Lecture #1) with linear losses. The losses ℓ_{jt} are in $[0,1]$ when not precised otherwise.

Exercise 1 :

Show that no strategy satisfies for all sequence $(\ell_{1t},\dots,\ell_{Nt})_t \in ([0,1]^N)^{\mathbb{N}}$:

$$\sum_{t=1}^{T}\sum_{j=1}^{N} p_{jt}\ell_{jt} - \sum_{t=1}^{T}\min_{k\in[N]}\ell_{kt} = o(T).$$

Exercise 2 :

1) Give an example where both

(a) $Z_T \xrightarrow{\mathcal{L}} Z$,

(b) f is continuous,

but $\lim_{T\to\infty}\mathbb{E}[Z_T] \neq \mathbb{E}[Z]$.

Definition. *We say that* $(Y_T)_T$ *is* ***uniform asymptotic integrable*** *(uai) if*

$$\lim_{L\to\infty}\lim_{T\to\infty}\mathbb{E}[\|Y_T\|\mathbb{1}_{\|Y_T\|>L}] = 0.$$

2) Show that if $Z_T \xrightarrow{\mathcal{L}} Z$ and $(f(Z_T))_T$ is uai, then

(a) $f(Z_T)\in\mathbb{L}^1$ for T large enough ;

(b) $f(Z)\in\mathbb{L}^1$;

(c) $\mathbb{E}[f(Z_T)] \to_{T\to\infty} \mathbb{E}[f(Z)]$.

Hint: for b), use Skorokhod's theorem.

3) Show that if $(Y_T)_T$ is bounded in $\mathbb{L}^p$ for $p>1$, i.e. $\sup_{T\geq 1}\mathbb{E}[\|Y_T\|^p] = B < +\infty$, then $(Y_T)_T$ is uai.

Exercise 3 :

Assume in this exercise that $\ell_{jt}\in[m,M]$, with $m,M\in\mathbb{R}$ unknown. How can we tune η ?
We consider in the following EWA with adaptive rates $(\eta_t)_t$:

$$p_{jt} = \frac{e^{-\eta_t\sum_{s=1}^{t-1}\ell_{js}}}{\sum_{k=1}^{N} e^{-\eta_t\sum_{s=1}^{t-1}\ell_{ks}}}.$$

1) Show that if (η_t) are non-increasing, then the regret of EWA satisfies :

$$R_T \leq \frac{\ln N}{\eta_T} + \sum_{t=1}^{T}\delta_t,$$

where $\delta_t = \sum_{j=1}^{N} p_{jt}\ell_{jt} + \frac{1}{\eta_t}\ln\left(\sum_{j=1}^{N} p_{jt}e^{-\eta_t \ell_{jt}}\right)$.

2) Prove the following Bernstein's inequality. For a random variable $X \in [m, M]$:

$$\forall \eta > 0,\ \ln \mathbb{E}[e^{\eta X}] \leq \eta \mathbb{E}[X] + \frac{e^{\eta(M-m)} - 1 - \eta(M-m)}{(M-m)^2}\mathrm{Var}\,(X)\,.$$

Hint: consider the function $\varphi : x \mapsto \frac{e^x - x - 1}{x^2}$.

We now consider EWA with $\eta_t = \frac{\ln N}{\sum_{s=1}^{t-1}\delta_s}$, with the convention that $\frac{\ln N}{0} = +\infty$.

3) Let $v_t = \sum_{j=1}^{N}(\ell_{jt} - \sum_{k=1}^{N} p_{kt}\ell_{kt})^2 p_{jt}$.

(a) Show that $v_t \geq \dfrac{\eta_t(M-m)}{e^{\eta_t(M-m)} - \eta_t(M-m) - 1}(M-m)\delta_t$.

(b) Deduce that $v_t \geq \frac{2\delta_t}{\eta_t} - \frac{2}{3}(M-m)\delta_t$.

4)

(a) Show that $\left(\sum_{t=1}^{T}\delta_t\right)^2 \leq \sum_{t=1}^{T} v_t \ln N + (M-m)(1 + \frac{2}{3}\ln N)\sum_{t=1}^{T}\delta_t$.

(b) Finally, show that $R_T \leq (M-m)\sqrt{T\ln N} + (M-m)(2 + \frac{4}{3}\ln N)$.

Exercise 4 :

In this exercise, we are trying to prove the conditional Hoeffding's lemma with a similar proof technique we used for the Hoeffding lemma (without conditioning). Consider a random variable X such that $X \in [a,b]$ almost surely and a σ-algebra $\mathcal{G}$. Define the function $\psi : s \mapsto \ln(\mathbb{E}[e^{sX}] \mid \mathcal{G})$.

1) Justify that ψ is twice continuously differentiable on $\mathbb{R}$ and that for any $s \in \mathbb{R}$:

$$\psi'(s) = \frac{\mathbb{E}[Xe^{sX} \mid \mathcal{G}]}{\mathbb{E}[e^{sX} \mid \mathcal{G}]}$$

$$\psi''(s) = \frac{\mathbb{E}[X^2 e^{sX} \mid \mathcal{G}]\mathbb{E}[e^{sX} \mid \mathcal{G}] - \left(\mathbb{E}[Xe^{sX} \mid \mathcal{G}]\right)^2}{\left(\mathbb{E}[e^{sX} \mid \mathcal{G}]\right)^2}.$$

2) Show that we can define the probability distribution $\mathbb{Q}_s$ as

$$\frac{\mathrm{d}\mathbb{Q}_s}{\mathrm{d}\mathbb{P}} = \frac{e^{sX}}{\mathbb{E}[e^{sX} \mid \mathcal{G}]}$$

where $X \sim \mathbb{P}$.

3) Show that for any random variable Z, we have

$$\mathbb{E}_{\mathbb{P}}[Z\frac{e^{sX}}{\mathbb{E}[e^{sX} \mid \mathcal{G}]} \mid \mathcal{G}] = \mathbb{E}_{\mathbb{Q}_s}[Z \mid \mathcal{G}].$$

4) Deduce that

$$\ln(\mathbb{E}[e^{s(X - \mathbb{E}[X]})]) \leq \frac{s^2}{8}(b-a)^2.$$

Exercise 5 :

In this exercise, we aim at showing a version of Hoeffding-Azuma inequality for unbounded sub-Gaussian variables.

1) Show that if a random variable X is σ sub-Gaussian, then for any $\varepsilon > 0$:

$$\max(\mathbb{P}(X \geq \varepsilon), \mathbb{P}(X \leq -\varepsilon)) \leq \exp(-\frac{\varepsilon^2}{2\sigma^2}).$$

2) Conversely, show that if there exist $b \geq 1, c > 0$ such that

$$\max(\mathbb{P}(X \geq \varepsilon), \mathbb{P}(X \leq -\varepsilon)) \leq b\exp(-c\varepsilon^2),$$

then X is $\sqrt{\frac{14b}{c}}$ sub-Gaussian.

Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration and let $(X_t)_{t\geq 1}$ be a sequence of adapted random variables and suppose there are constants σ_t such that for any $t \in \mathbb{N}$ and $\varepsilon > 0$,

$$\max(\mathbb{P}(X_t - \mathbb{E}[X_t] > \varepsilon \mid \mathcal{F}_{t-1}), \mathbb{P}(X_t - \mathbb{E}[X_t] < -\varepsilon \mid \mathcal{F}_{t-1})) \leq b\exp(-\frac{\varepsilon^2}{2\sigma_t^2}).$$

3) Show that for any $T \in \mathbb{N}$ and $\varepsilon > 0$:

$$\mathbb{P}\left(\sum_{t=1}^{T} X_t - \mathbb{E}[X_t \mid \mathcal{F}_{t-1} \geq \varepsilon]\right) \leq \exp\left(-\frac{\varepsilon^2}{56b\sum_{t=1}^{T}\sigma_t^2}\right).$$