Part 1: The Hoeffding–Azuma inequality

The Hoeffding-Azuma inequality

Theorem: Let $(\mathcal{F}_t)_{t \geq 0}$ be a filtration and let $(X_t)_{t \geq 1}$ be a sequence of adapted random variables (ie, $\forall t \geq 1$, X_t is $\mathcal{F}_t$-measurable), that are bounded: $\forall t$, $a_t \leq X_t \leq b_t$ a.s., where $a_t, b_t \in \mathbb{R}$.

Then (« probabilistic version »)

$$\forall \varepsilon > 0, \qquad \mathbb{P}\left\{ \sum_{t=1}^{T} X_t - \sum_{t=1}^{T} \mathbb{E}[X_t \mid \mathcal{F}_{t-1}] \geq \varepsilon \right\} \leq \exp\left(- \frac{2\varepsilon^2}{\sum_{t=1}^{T} (b_t - a_t)^2} \right)$$

or (« statistical version », totally equivalent)

$\forall \delta \in (0,1)$, with probability at least $1-\delta$,

$$\sum_{t=1}^{T} X_t - \sum_{t=1}^{T} \mathbb{E}[X_t \mid \mathcal{F}_{t-1}] \leq \sqrt{\frac{\sum_{t=1}^{T}(b_t - a_t)^2}{2} \ln \frac{1}{\delta}}$$

Note: Hoeffding's inequality is the special case when all X_t are independent and $\mathcal{F}_{t-1} = \sigma(X_1 \ldots X_{t-1})$, so that $\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] = \mathbb{E}[X_t]$.

Basic ingredient of the proof: extension of Hoeffding's lemma to conditional expectations

Lemma: X random variable s.t. $X \in [a,b]$ a.s.

Then, for all σ-algebras $\mathcal{G}$, for all $s \in \mathbb{R}$,

$$\ln \mathbb{E}\left[e^{s(X - \mathbb{E}[X|\mathcal{G}])} \,\middle|\, \mathcal{G} \right] = \ln\left(\mathbb{E}\left[e^{sX} \mid \mathcal{G} \right] \right) - s\,\mathbb{E}[X \mid \mathcal{G}] \leq \frac{s^2}{8}(b-a)^2$$

(We will discuss the proof later on... let's first prove the theorem based on this lemma.)

Proof (of the theorem):

Markov-Chernov bounding (= Markov's inequality after taking exponents):

We denote

$$\underbrace{S_T}_{\text{martingale}} = \underbrace{\sum_{t=1}^{T}}_{\text{sum of}} \underbrace{X_t - \mathbb{E}[X_t \mid \mathcal{F}_{t-1}]}_{\text{martingale increments or martingale differences}}$$

(martingale = sum of martingale increments)
or martingale differences

The « probabilistic version » is about upper bounding $\mathbb{P}\{S_T > \varepsilon\}$:

$$\mathbb{P}\{S_T > \varepsilon\} \underset{\forall s>0}{=} \mathbb{P}\{e^{sS_T} > e^{s\varepsilon}\} \underset{\text{Markov's inequality}}{\leqslant} e^{-s\varepsilon}\,\mathbb{E}\left[e^{sS_T}\right]$$

We show by induction that $\mathbb{E}\left[e^{sS_T}\right] \leqslant \exp\left(\frac{s^2}{8}\sum_{t=1}^{T}(b_t - a_t)^2\right)$

- For $T=1$, true by the conditional version of Hoeffding's lemma and the fact that $S_1 = X_1 - \mathbb{E}[X_1 \mid \mathcal{F}_0]$ with $X_1 \in [a_1, b_1]$ + taking expectations by tower rule: $\mathbb{E} = \mathbb{E}[\mathbb{E}[\;\mid \mathcal{F}_0]]$
- For $T-1 \to T$, where $T \geqslant 2$:

The extension of Hoeffding's lemma ensures that

$$\mathbb{E}\left[e^{s(X_T - \mathbb{E}[X_T \mid \mathcal{F}_{T-1}])} \,\middle|\, \mathcal{F}_{T-1}\right] \leqslant e^{s^2(b_T - a_T)^2/8}$$

so that

$$\begin{aligned}
\mathbb{E}\left[e^{sS_T}\right] &= \mathbb{E}\left[\mathbb{E}\left[e^{sS_T} \mid \mathcal{F}_{T-1}\right]\right] \\
&= \mathbb{E}\left[e^{sS_{T-1}}\,\mathbb{E}\left[e^{s(X_T - \mathbb{E}[X_T \mid \mathcal{F}_{T-1}])} \,\middle|\, \mathcal{F}_{T-1}\right]\right] \\
&\leqslant e^{s^2(b_T - a_T)^2/8} \times \mathbb{E}\left[e^{sS_{T-1}}\right] \\
&\underset{\text{by the induction hypothesis}}{\leqslant} \exp\left(s^2 \sum_{t \leqslant T} (b_t - a_t)^2/8\right)
\end{aligned}$$

Substituting above:

$$\mathbb{P}\{S_T > \varepsilon\} \leqslant \inf_{s>0} \exp\left(-s\varepsilon + s^2 \sum_{t \leqslant T} \frac{(b_t - a_t)^2}{8}\right)$$

$$= \exp\left(-2\varepsilon^2 \Big/ \sum_{t \leqslant T} (b_t - a_t)^2\right).$$

strictly convex function to minimize in the exponent: minimum achieved at s^* such that $s^* \sum_{t \leqslant T} (b_t - a_t)^2/4 = \varepsilon$ (gradient vanishes)

i.e. $s^* = 4\varepsilon \Big/ \sum_{t \leqslant T} (b_t - a_t)^2$

$\to$ It only remains to prove the extension of Hoeffding's lemma to conditional expectations.

 Lecture 4 — Spring 2021 term Lecture 4 — Spring 2021 term

But first (reminder) ↓ unconditional version

Lemma (Hoeffding): X random variable s.t. $X \in [a,b]$ a.s.

Then $\forall s \in \mathbb{R}$,

$$\ln \mathbb{E}\left[e^{s(X-\mathbb{E}X)}\right] = \ln \mathbb{E}\left[e^{sX}\right] - s\,\mathbb{E}[X] \leqslant \frac{s^2}{8}(b-a)^2$$

Proof (most elegant one I know of):

$$\Psi(s) = \ln \mathbb{E}\left[e^{sX}\right] \quad \text{defined for all } s \in \mathbb{R}$$

Ψ is differentiable at each $s \in \mathbb{R}$: cf. X bounded, thus $\eta \mapsto X e^{\eta X}$ locally dominated around s by an integrable r.v. independent of η, thus $\eta \mapsto \mathbb{E}[e^{\eta X}]$ differentiable at s with derivative $\mathbb{E}[X e^{sX}]$

with

$$\Psi'(s) = \frac{\mathbb{E}[X e^{sX}]}{\mathbb{E}[e^{sX}]}$$

Similarly, Ψ is twice differentiable at each $s \in \mathbb{R}$, with:

$$\Psi''(s) = \frac{\mathbb{E}[X^2 e^{sX}]\,\mathbb{E}[e^{sX}] - \left(\mathbb{E}[X e^{sX}]\right)^2}{\left(\mathbb{E}[e^{sX}]\right)^2} = \mathrm{Var}_{\mathbb{Q}}(X)$$

under the probability $\mathbb{Q}$ defined by

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(\omega) = \frac{e^{sX(\omega)}}{\mathbb{E}[e^{sX}]}$$

$X \in [a,b]$:

$$\mathrm{Var}_{\mathbb{Q}}(X) = \inf_{\mu \in \mathbb{R}} \mathbb{E}_{\mathbb{Q}}\left[(X-\mu)^2\right] \leqslant \mathbb{E}_{\mathbb{Q}}\left[\left(X - \frac{a+b}{2}\right)^2\right] \leq \frac{(b-a)^2}{4}$$

Taylor: $\exists x$ s.t.

$$\Psi(s) = \underbrace{\Psi(0)}_{=0} + s\underbrace{\Psi'(0)}_{=\mathbb{E}[X]} + \frac{s^2}{2}\underbrace{\Psi''(x)}_{\leq (b-a)^2/4}$$

Cf. Ψ is actually C^2 smooth

i.e.,

$$\ln \mathbb{E}\left[e^{sX}\right] \leqslant s\,\mathbb{E}[X] + \frac{s^2}{8}(b-a)^2$$

Back to Hoeffding's lemma with conditional expectations:

Proof 1? Can we take the proof of Hoeffding's lemma we just saw and replace all $\mathbb{E}$ by $\mathbb{E}[\ |\mathcal{G}_j]$?

$\psi(s) = \ln \mathbb{E}[e^{sX} | \mathcal{G}_j]$ → The theorem of differentiation under $\mathbb{E}[\ ]$ only requires dominated convergence, which holds true for $\mathbb{E}[\ |\mathcal{G}_j]$ as well. Thus we also have a theorem of differentiation under $\mathbb{E}[\ |\mathcal{G}_j]$:

a.s., $\psi''(s)$ exists and equals

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

= some conditional variance under a different probability measure?

Yes, see details in some pages.

However, there are two other proofs that I find more elementary:

Proof 2 Too bad for elegance, let's get back to the original proof of Hoeffding's (unconditional) lemma, which only relies on calculus:

$Y = X - \mathbb{E}[X|\mathcal{G}_j] \in [A, B]$ where $A = a - \mathbb{E}[X|\mathcal{G}_j] \leqslant 0$, $B = b - \mathbb{E}[X|\mathcal{G}_j] \geqslant 0$ are both $\mathcal{G}_j$-measurable and $B-A = b-a > 0$

$$Y = \frac{B-Y}{B-A} A + \frac{Y-A}{B-A} B$$

convex weights

Since $y \mapsto e^{sy}$ is convex:

$$e^{sY} \leqslant \frac{B-Y}{B-A} e^{sA} + \frac{Y-A}{B-A} e^{sB}$$

Taking $\mathbb{E}[\ |\mathcal{G}_j]$: using $\mathbb{E}[Y|\mathcal{G}_j] = 0$ and A, B $\mathcal{G}_j$-measurable:

$$\mathbb{E}[e^{sY} | \mathcal{G}_j] \leqslant \frac{B}{B-A} e^{sA} - \frac{A}{B-A} e^{sB}$$

note that $B/B-A$ and $-A/B-A$ are convex weights

 Lecture 4 — Spring 2021 term Lecture 4 — Spring 2021 term

Now, by a function study (the very same as the one we performed in the proof of the unconditional version of Hoeffding's lemma) — or even by the latter lemma itself:

$\forall u, v \in \mathbb{R}, \ \forall p \in [0,1], \ \forall s \in \mathbb{R},$

$$\ln\left(p e^{su} + (1-p) e^{sv}\right) \leqslant s\big(pu + (1-p)v\big) + \frac{s^2}{8}(v-u)^2$$

← ln of expected value of e^{sZ} where $Z = \begin{cases} u & \text{w.p. } p \\ v & 1-p \end{cases}$

expected value of Z

range is $[u,v]$

In particular,

$$\frac{B}{B-A} e^{sA} - \frac{A}{B-A} e^{sB} \leqslant \exp\left(s\left(\frac{BA}{B-A} - \frac{AB}{B-A}\right) + \frac{s^2}{8}(B-A)^2 \right)$$
$$= \exp\left(\frac{s^2}{8}(b-a)^2\right)$$

recall that a.s., $B-A = b-a$

Summarizing:

$$\mathbb{E}\left[e^{sY} \mid \mathcal{G}\right] \leqslant \exp\left(\frac{s^2}{8}(b-a)^2\right)$$
$$= \mathbb{E}\left[e^{sX} \mid \mathcal{G}\right] \times \exp\left(-s\, \mathbb{E}[X \mid \mathcal{G}]\right)$$

Proof 3 My preferred (not only because I found it by myself): Hoeffding's lemma in its unconditional version ENTAILS the conditional version! This is because Hoeffding's lemma holds for all probability distributions — we should play with this fact.

For all $A \in \mathcal{G}$ s.t. $\mathbb{P}(A) > 0$, let $\mathbb{P}_A = \mathbb{P}(\cdot \mid A)$, the conditional distribution given the event A.

The unconditional version of Hoeffding's lemma ensures that

$$\forall A \in \mathcal{G} \text{ s.t. } \mathbb{P}(A) > 0, \quad \forall s \in \mathbb{R}, \quad \ln \mathbb{E}_A\left[e^{sX}\right] \leqslant s\, \mathbb{E}_A[X] + \frac{s^2}{8}(b-a)^2$$

Why do we consider the $\mathbb{E}_A$? Because $\mathbb{E}[X \mid \mathcal{G}]$ is the unique $\mathcal{G}$-measurable random variable such that $\forall A \in \mathcal{G}, \ \mathbb{E}[X \mathbb{1}_A] = \mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}]\, \mathbb{1}_A\big]$

or, equivalently, $\forall A \in \mathcal{G}$ s.t. $\mathbb{P}(A) > 0$, $\mathbb{E}_A[X] = \mathbb{E}_A\big[\mathbb{E}[X \mid \mathcal{G}]\big]$.

Now, consider the random variable $H = e^{s\mathbb{E}[X \mid \mathcal{G}]} e^{s^2(b-a)^2/8} - \mathbb{E}\left[e^{sX} \mid \mathcal{G}\right]$

We want to prove that $H \geq 0$ a.s.

H is $\mathcal{G}_j$-measurable & thus suffices to show that for all $A \in \mathcal{G}_j$ with $\mathbb{P}(A) > 0$,

$$\mathbb{E}_A[H] \geq 0 \qquad \text{that is,} \qquad \mathbb{E}[H \mathbb{1}_A] \geq 0$$

Indeed,
$$\mathbb{E}_A[H] = e^{s^2(b-a)^2/8}\, \mathbb{E}_A\big[e^{s\mathbb{E}[X|\mathcal{G}_j]}\big] - \mathbb{E}_A\big[\mathbb{E}[e^{sX}|\mathcal{G}_j]\big]$$
$$= e^{s^2(b-a)^2/8}\, \mathbb{E}_A\big[e^{s\mathbb{E}[X|\mathcal{G}_j]}\big] - \mathbb{E}_A\big[e^{sX}\big]$$
$$\underset{\text{(Jensen)}}{\geq} e^{s^2(b-a)^2/8}\, e^{s\mathbb{E}_A[X]} - \mathbb{E}_A\big[e^{sX}\big] \geq 0$$

(unconditional version of Hoeffding's lemma)

<u>Proof 1</u> Let's get back to it. In what follows all expectations relative to the original probability distribution $\mathbb{P}$ will be denoted by $\mathbb{E}$, and expectations under alternative distributions $\mathbb{Q}$ will be denoted by $\mathbb{E}_{\mathbb{Q}}$.

(1) Consider the random variable $L_s = \dfrac{e^{sX}}{\mathbb{E}[e^{sX}|\mathcal{G}_j]} \geq 0$

Since $\mathbb{E}[L_s] = \mathbb{E}\big[\mathbb{E}[L_s|\mathcal{G}_j]\big] = 1$, L_s is a density

We define the probability $\mathbb{Q}_s$ as : $\dfrac{d\mathbb{Q}_s}{d\mathbb{P}} = L_s$

(2) We show that $\varphi'(s) = \dfrac{\mathbb{E}[X e^{sX}|\mathcal{G}_j]}{\mathbb{E}[e^{sX}|\mathcal{G}_j]}$ also equals $\mathbb{E}_{\mathbb{Q}_s}[X|\mathcal{G}_j]$

$\varphi''(s) =$ [see expression some pages ago] $\mathrm{Var}_{\mathbb{Q}_s}(X|\mathcal{G}_j)$

To do so, it suffices to prove that for all bounded random variables Z (we'll pick $Z = X$ and $Z = X^2$), we have :

$$\mathbb{E}[Z L_s | \mathcal{G}_j] = \mathbb{E}_{\mathbb{Q}_s}[Z|\mathcal{G}_j]$$

or equivalently, that $\forall A \in \mathcal{G}_j$, (since both sides are $\mathcal{G}_j$-measurable)
$$\mathbb{E}\big[\mathbb{E}[Z L_s|\mathcal{G}_j]\,\mathbb{1}_A\big] = \mathbb{E}\big[\mathbb{E}_{\mathbb{Q}_s}[Z|\mathcal{G}_j]\,\mathbb{1}_A\big]$$

But :
$$\mathbb{E}\big[\mathbb{E}[Z L_s|\mathcal{G}_j]\,\mathbb{1}_A\big] \underset{\substack{\text{one characterization}\\ \text{of } \mathbb{E}[\ |\mathcal{G}_j]}}{=} \mathbb{E}[Z L_s \mathbb{1}_A] \underset{L_s \text{ is } \frac{d\mathbb{Q}_s}{d\mathbb{P}}}{=} \mathbb{E}_{\mathbb{Q}_s}[Z \mathbb{1}_A]$$

and on the other end

$$
\begin{aligned}
&\mathbb{E}\big[\ \mathbb{E}_{Q_s}[Z|\mathcal{G}]\ \mathbb{1}_A\big] && \text{as } \mathbb{E}[L_s|\mathcal{G}]=1 \text{ as by definition of } L_s\\
&= \mathbb{E}\big[\ \mathbb{E}_{Q_s}[Z|\mathcal{G}]\ \mathbb{E}[L_s|\mathcal{G}]\ \mathbb{1}_A\big] && \mathbb{E}_{Q_s}[Z|\mathcal{G}] \text{ and } \mathbb{1}_A\ \mathcal{G}\text{-measurable and can go inside } \mathbb{E}[\ |\mathcal{G}]\\
&= \mathbb{E}\big[\ \mathbb{E}[\ L_s\ \mathbb{E}_{Q_s}[Z|\mathcal{G}]\ \mathbb{1}_A\ |\mathcal{G}]\big] && \text{“tower rule”}\\
&= \mathbb{E}\big[\ L_s\ \mathbb{E}_{Q_s}[Z|\mathcal{G}]\ \mathbb{1}_A\big] && L_s = \frac{dQ_s}{dP}\\
&= \mathbb{E}_{Q_s}\big[\ \mathbb{E}_{Q_s}[Z|\mathcal{G}]\ \mathbb{1}_A\big] && \\
&= \cancel{\mathbb{E}_{Q_s}\big[\mathbb{E}_{Q_s}[Z\mathbb{1}_A|\mathcal{G}]\big]} = \mathbb{E}_{Q_s}[Z\mathbb{1}_A] && \text{by a characterization of } \mathbb{E}[\ |\mathcal{G}]
\end{aligned}
$$

That's the real trick, but it's not a trick, it's what is called Bayes' formula for conditional expectations

which concludes the proof of (2).

(3) $\qquad \psi'(0) = \mathbb{E}[X|\mathcal{G}]$ $\qquad$ (clear)

and for all x, $\qquad \psi''(x) = \mathrm{Var}_{Q_x}(X|\mathcal{G}) \leqslant \dfrac{(b-a)^2}{4}$ (we prove that below)

so that we may conclude as in the case of the unconditional Hoeffding's lemma.

Indeed, $\forall c \in \mathbb{R}$,

$$
\begin{aligned}
\mathbb{E}_{Q_x}\big[(X-c)^2\,|\,\mathcal{G}\big] &= \mathbb{E}_{Q_x}\big[(X-\mathbb{E}_{Q_x}[X|\mathcal{G}]+\mathbb{E}_{Q_x}[X|\mathcal{G}]-c)^2\,|\,\mathcal{G}\big]\\
&= \underbrace{\mathbb{E}_{Q_x}\big[(X-\mathbb{E}_{Q_x}[X|\mathcal{G}])^2\,|\,\mathcal{G}\big]}_{\overset{\text{def.}}{=}\ \mathrm{Var}_{Q_x}(X|\mathcal{G})} + 2\times 0 + \underbrace{\cdots}_{\geqslant 0}
\end{aligned}
$$

and we take

$c = \dfrac{b+a}{2}$ and use $X \in [a,b]$

to get the a.s. bound $\qquad (X-c)^2 \leqslant \dfrac{(b-a)^2}{4}$

A final remark: Dealing with non-constant but predictable ranges in the Hoeffding–Azuma inequality

↳ Sometimes useful to get slightly better constants

Hoeffding's lemma: extension #1

Setting: X random variable s.t. there exists a bounded and $\mathcal{G}$-measurable random variable G, as well as $a, b \in \mathbb{R}$ with: $G + a \leq X \leq G + b$

Then:

$$\forall s \in \mathbb{R}, \qquad \ln \mathbb{E}\left[e^{sX} \mid \mathcal{G}\right] \leq s\,\mathbb{E}\left[X \mid \mathcal{G}\right] + \frac{s^2}{8}(b-a)^2$$

Rk: X bounded as well, we get this statement from the first statement by considering $a \leq X - G \leq b$

extension #2

Setting: What about when there exist U, V two $\mathcal{G}$-measurable random variables with $U \leq X \leq V$ and X bounded (for e^{sX} to be $\mathbb{L}^1$)?

An inspection of Proof 2 reveals that one can prove

$$\forall s \in \mathbb{R}, \qquad \ln \mathbb{E}\left[e^{sX} \mid \mathcal{G}\right] \leq s\,\mathbb{E}\left[X \mid \mathcal{G}\right] + \frac{s^2}{8}(V-U)^2$$

Note: To state our extension of the Hoeffding–Azuma inequality, we will need a constant bound on $V-U$:

$$V - U \leq \Delta \qquad \text{where} \quad \Delta \in \mathbb{R}^+$$

But actually $\left\{\begin{array}{l} U \leq X \leq V \\ V - U \leq \Delta \in \mathbb{R}^+ \end{array}\right.$ entail $\dfrac{U+V}{2} - \dfrac{\Delta}{2} \leq X \leq \dfrac{U+V}{2} + \dfrac{\Delta}{2}$

So we're back to extension #1 (in particular, $(U+V)/2$ is bounded)

Hoeffding–Azuma inequality

Let $(\mathcal{F}_t)$ be a filtration and (X_t) be a sequence of adapted random variables such that

(1) $\forall t$, $\left.\begin{array}{l} \exists\, G_t \ \ \mathcal{F}_{t-1}\text{-measurable and bounded} \\ \exists\, a_t, b_t \in \mathbb{R} \end{array}\right\}$ with $G_t + a_t \leq X_t \leq G_t + b_t$

possibly following from

(2) $\forall t$, $\left.\begin{array}{l} \exists\, U_t, V_t \ \ \mathcal{F}_{t-1}\text{-measurable} \\ X_t \ \text{bounded} \\ \exists\, \Delta_t \in \mathbb{R}^+ \end{array}\right\}$ with $\left\{\begin{array}{l} V_t - U_t \leq \Delta_t \\ U_t \leq X_t \leq V_t \end{array}\right.$

Then $\forall \delta \in (0,1)$

with probability at least $1-\delta$,

Case (1) $$\sum_{t=1}^{T} X_t - \sum_{t=1}^{T} \mathbb{E}[X_t \mid \mathcal{F}_{t-1}] \leqslant \sqrt{\frac{\sum_{t=1}^{T}(b_t - a_t)^2}{2} \ln \frac{1}{\delta}}$$

Case (2) $$\sum_{t=1}^{T} X_t - \sum_{t=1}^{T} \mathbb{E}[X_t \mid \mathcal{F}_{t-1}] \leqslant \sqrt{\frac{\sum_{t=1}^{T} \Delta_t^2}{2} \ln 1/\delta}$$

Part 2: Non-convex aggregation via randomization

What can we do when no convexity assumption holds?

↳ <u>Non-convex aggregation via randomization.</u>

<u>Example 1</u>: N-ary decisions in a game (4-ary if we have to pick paths in a graph: → ← ↑ ↓)

(binary if accept / reject actions)

1. Opponent picks state of the world y_t } simultaneously
2. Statistician picks action $j_t \in \{1, \dots N\}$ }
3. Loss $\ell(j_t, y_t)$ or reward $-\ell(j_t, y_t)$ is encountered, both y_t and j_t are made public

<u>Example 2</u>: Prediction with expert advice (the «meta-statistical» framework)

↳ when the prediction space is not convex:

1. Opponent picks observation $y_t \in \mathcal{Y}$
2. Simultaneously, experts provide forecasts $f_{j,t} \in \mathcal{Y}$, $j \in \{1 \dots N\}$ and Statistician picks forecast $\hat{y}_t \in \mathcal{Y}$
3. y_t and $\hat{y}_t$ are revealed, losses $\ell(\hat{y}_t, y_t)$ and $\ell(f_{j,t}, y_t)$ are suffered

No convexity: $\mathcal{Y}$ not convex [OR $\mathcal{Y}$ convex but $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ not convex in its first argument]
eg, $\mathcal{Y} = \{1 \dots M\}$ in M-ary classification

↳ $\hat{y}_t$ cannot be any convex / linear prediction of the $f_{j,t}$ we wish.

<u>Solution</u>: (at least, an easy solution, there might be others)

Draw $J_t \in \{1, \dots N\}$ at random

and pick $\begin{cases} \text{action } J_t & \text{(in Example 1)} \\ \text{forecast } \hat{y}_t = f_{J_t, t} & \text{(in Example 2)} \end{cases}$

⇣

<u>General setting</u>: Simultaneously $\begin{cases} \text{1. Opponent picks } \underline{\ell}_t = (\ell_{1t}, \dots \ell_{Nt}) \in \mathbb{R}^N \\ \text{2. Statistician draws } J_t \in \{1, \dots N\} \end{cases}$

3. J_t and $(\ell_{1t}, \dots \ell_{Nt})$ are revealed.

<u>Aim</u>: Minimize the regret $\displaystyle\sum_{t=1}^{T} \ell_{J_t, t} - \min_{k=1 \dots N} \sum_{t=1}^{T} \ell_{kt}$

The losses ℓ_{jt} may depend on the past, ie, on $J_1, \dots J_{t-1}$

Methodology: We denote by $p_t = (p_{1t} \dots p_{Nt}) \in \mathcal{X}$ the probability distribution used to draw J_t, conditionally to the past

Regret: $R_T = \sum_{t=1}^T \ell_{J_t,t} - \min_k \sum_{t=1}^T \ell_{kt} = \left[\sum_{t=1}^T \ell_{J_t,t} - \sum_{t=1}^T \sum_{j=1}^N p_{jt}\ell_{jt}\right] + \left[\sum_{t=1}^T \sum_j p_{jt}\ell_{jt} - \min_k \sum_{t=1}^T \ell_{kt}\right]$

This can be controlled independently of the probability distributions chosen

We already learned how to control this term! we denote it by $\overline{R}_T$ below

The information available at the begining of round t is $(\ell_s, p_s, J_s)_{s\leq t-1}$

We denote $\mathcal{F}_{t-1} = \sigma\{(\ell_s, p_s, J_s)_{s\leq t-1}\}$: ℓ_t and p_t are $\mathcal{F}_{t-1}$-measurable while J_t is drawn at random using an auxiliary randomization $U_t \sim \mathcal{U}_{[0,1]}$, independent from $\mathcal{F}_{t-1}$.

Then: $\mathbb{E}[\ell_{J_t,t} \mid \mathcal{F}_{t-1}] = \sum_{j=1}^N p_{jt}\ell_{jt}$ (J_t is not fixed by the conditioning, only its distribution p_t is so.)

↳ Expected regret (conditionally expected regret) $\overline{R}_T = \sum_{t,j} p_{jt}\ell_{jt} - \min_k \sum_{t=1}^T \ell_{kt}$

We already saw that we could ensure $\overline{R}_T \leq O\left((M-m)\sqrt{T\ln N}\right)$ if $\ell_{jt} \in [m,M]$ $\forall j,t$

↳ Martingale $S_T = \sum_{t=1}^T \ell_{J_t,t} - \sum_{t=1}^T \sum_j p_{jt}\ell_{jt}$

The Hoeffding-Azuma inequality ensures that if $\ell_{jt} \in [m,M]$ $\forall j \forall t$, then, no matter which p_t were selected

with $X_t = \ell_{J_t,t}$, $b_t = M$ and $a_t = m$, $\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] = \sum_j p_{jt}\ell_{jt}$

$$\mathbb{P}\left\{ S_T \leq (M-m)\sqrt{\tfrac{T}{2}\ln 1/\delta} \right\} \geq 1-\delta$$

Conclusion: $\forall T$, with probability at least $1-\delta$, $R_T \leq \overline{R}_T + (M-m)\sqrt{\frac{T}{2}\ln\frac{1}{\delta}}$

E.g. with the fully adaptive version of EWA:

$\forall T$, $\forall \delta \in (0,1)$, with probability at least $1-\delta$, $R_T \leqslant (M-m)\sqrt{T}\left(\sqrt{\ln N} + \sqrt{\frac{1}{2}\ln\frac{1}{\delta}}\right) + (M-m)\left(2 + \frac{4}{3}\ln N\right)$

This is called a high probability bound; it is non-asymptotic → **Exercise:** Can you get a high probability bound of the form: $\forall \delta \in (0,1)$, with proba $\geqslant 1-\delta$, $\forall T$, $R_T \leqslant$ ... ?

Consequence: Asymptotic almost-sure bound.

The Borel-Cantelli lemma, using $\delta_T = 1/T^2$, ensures that

$$\mathbb{P}\Big(\limsup \Big\{ R_T > \underbrace{(M-m)\sqrt{T}\left(\sqrt{\ln N} + \sqrt{\ln T}\right) + (M-m)\left(2 + \tfrac{4}{3}\ln N\right)}_{} \Big\}\Big) = 0$$

limsup of events

We denote $\rho(T)$ this quantity: $\rho(T) \sim (M-m)\sqrt{T \ln T}$

That is, almost-surely

$R_T / \rho(T) > 1$ for finitely many T

thus, $\displaystyle \limsup_{T\to+\infty} \frac{R_T}{\rho(T)} \leqslant 1$ a.s. or equivalently,

limsup of a sequence of numbers

$$\limsup_{T\to+\infty} \frac{R_T}{(M-m)\sqrt{T \ln T}} \leqslant 1 \quad \text{a.s.}$$

Exercise: [To be stated in a more detailed way on the next page]

Show that we actually have $\displaystyle \limsup_{T\to+\infty} \frac{R_T}{(M-m)\sqrt{T\ln(\ln T)}} \leqslant C$ a.s. where C is a constant

(a rate which should be reminiscient of the law of the iterated logarithm.)

and I should have started with that...

Note: Of course, since $\mathbb{E}[S_T] = 0$, we have $\mathbb{E}[R_T] = \mathbb{E}[\overline{R}_T]$

Because we have deterministic bounds on $\overline{R}_T$, we get bounds on $\mathbb{E}[R_T]$. But this doesn't tell us much on R_T, this is why we prefer our high-probability bounds.

Exercise [Full Statement]

(1) Remind yourself of Doob's martingale inequality (actually: inequalities – there are two of them, but we'll need only the most famous one).

(2) Show the following MAXIMAL version of the Hoeffding-Azuma inequality:

$\forall \delta \in (0,1)$, with probability at least $1-\delta$,

$$\max_{t \leqslant T} \left\{ \sum_{s=1}^{t} X_s - \sum_{s=1}^{t} \mathbb{E}\left[X_s \mid \mathcal{F}_{s-1}\right] \right\} \leqslant \sqrt{\frac{\sum_{t=1}^{T}(b_t-a_t)^2}{2} \ln \frac{1}{\delta}}$$

(3) Show that for any algorithm with expected regret $\overline{R}_T$ less than something of order $(M-m)\sqrt{T \ln N}$, the corresponding randomized algorithm has a regret R_T such that

For all strategies of the opponent picking losses $\ell_{j,t} \in [m, M]$,

$$\limsup_{T \to +\infty} \frac{R_T}{(M-m)\sqrt{T \ln(\ln T)}} \leqslant C \quad \text{a.s.}$$

where C is a universal constant (propose a numerical value).

(4) Is this C optimal? (Consider the law of the iterated logarithm as a basis for your discussion.)

Hint for (3): Consider the regimes $\{2^r+1, \ldots, 2^{r+1}\}$ for $r = 1, 2, \ldots$ and pick $\delta_r = 1/r^2$ for the application of the Borel-Cantelli lemma. (Cf. doubling trick!)

Part 3: Introduction to stochastic bandits

Stochastic bandits. Finitely many arms.

Setting: K arms indexed by 1, 2, ... K

With each arm j is associated a probability distribution ν_j (over $\mathbb{R}$) with an expectation.

At each round $t = 1, 2, \ldots$

- The decision-maker picks $I_t \in \{1, \ldots K\}$, possibly at random
- She gets a reward Y_t drawn at random according to ν_{I_t} (given I_t)
- This is the only feedback she gets / the only observation she has access to.

Aim: We denote by $\mu_i = E(\nu_i)$ the expectation of ν_i

(note: operator E vs. expectation $\mathbb{E}$ of an expression involving random variables.)

Pseudo-regret $\overline{R}_T = T\mu^* - \mathbb{E}\left[\sum_{t=1}^{T} Y_t\right]$ to be controlled

where $\mu^* = \max_{j \leq K} \mu_j$

Useful notation:

$\Delta_a = \mu^* - \mu_a$ gap of arm a

$\Delta_a = 0$: a is an optimal arm (there can be several of them)

$\Delta_a > 0$: a is a suboptimal arm

$N_a(T) = \sum_{t=1}^{T} \mathbb{1}_{\{I_t = a\}}$ total number of times that a is pulled.

Note: * Pseudo regret $\overline{R}_T$ is a very "expected" notion of regret

$$\overline{R}_T \underset{\text{probably } <}{\leq} \mathbb{E}\left[\max_{a=1,\ldots K} \sum_{t} \ldots - \sum_{t=1}^{T} Y_t\right]$$

* Can be rewritten (see later) as $\overline{R}_T = \sum_{a=1}^{K} \Delta_a \, \mathbb{E}[N_a(T)]$

 Lecture 4 — Spring 2021 term Lecture 4 — Spring 2021 term

Upper confidence bound [UCB] algorithm: Very popular !

For $t = 1, 2, \dots K$

- Pull arm $I_t = t$, get a reward Y_t

For $t = K+1, K+2, \dots$

- Pull an arm $I_t \in \underset{j \in \{1, \dots K\}}{\operatorname{argmax}} \left\{ \hat{\mu}_{j,t-1} + \sqrt{\frac{2 \ln t}{N_j(t-1)}} \right\}$

(tie-breaking rule: pick the element with smallest index)

where $N_j(t-1) = \sum_{s=1}^{t-1} \mathbb{1}_{\{I_s = j\}}$

and where $\hat{\mu}_{j,t-1} = \frac{1}{N_j(t-1)} \sum_{s=1}^{t-1} Y_s \mathbb{1}_{\{I_s = j\}}$

always ≥ 1 since each arm was tried sequentially during rounds $1, 2, \dots K$

- Get a reward Y_t

Theorem: If the distributions ν_j have supports all included in $[0,1]$, then the pseudo-regret of UCB is smaller than

$$\bar{R}_T \leq \sum_{i : \Delta_i > 0} \left(\frac{8 \ln T}{\Delta_i} + 2 \right)$$

This regret bound is obtained via the following proposition:

Proposition: If the distributions ν_j have supports all included in $[0,1]$, then

$$\forall i \text{ s.t. } \Delta_i > 0, \qquad \mathbb{E}\big[N_i(T) \big] \leq \frac{8 \ln T}{\Delta_i^2} + 2.$$

Exercise: The bounds above are called distribution-dependent because they depend heavily on the distributions ν_i at hand (via the gaps $\Delta_i = \mu^* - \mu_i$).

Show the following distribution-free bound (that only depends on the support $[0,1]$, not on the specific distributions ν_i at hand) : for the UCB algorithm,

$$\sup_{\substack{\nu_1, \dots, \nu_K \text{ with} \\ \text{supports in } [0,1]}} \overline{R}_T \leqslant O\left(\sqrt{TK \ln T}\right).$$

Hint: For small values of Δ_i, the bound of the Proposition can be worse than the trivial T bound...

Proof [of the theorem based on the Proposition] :

$$\overline{R}_T = T\mu^* - \mathbb{E}\left[\sum_{t=1}^{T} Y_t\right]$$

where by definition of the bandit model, ← Given I_t, Y_t is drawn at random according to ν_{I_t}

$$\mathbb{E}[Y_t \mid I_t] = \mu_{I_t}$$

thus (by the tower rule) $\mathbb{E}[Y_t] = \mathbb{E}[\mu_{I_t}]$

$$= \sum_j \mu_j \, \mathbb{E}\left[\mathbb{1}_{\{I_t = j\}}\right]$$

Summing over t: $\mathbb{E}\left[\sum_{t=1}^{T} Y_t\right] = \sum_{j=1}^{K} \mu_j \, \mathbb{E}[N_j(T)]$

and $\left(\text{in view of } T = \sum_j \mathbb{E}[N_j(T)]\right)$

$$\overline{R}_T = \sum_j (\mu^* - \mu_j)\, \mathbb{E}[N_j(T)] = \sum_{j=1}^{K} \Delta_j \, \mathbb{E}[N_j(T)]$$

$$= \sum_{j : \Delta_j > 0} \Delta_j \, \mathbb{E}[N_j(T)]$$

it suffices to consider the suboptimal arms...

We conclude by substituting $\mathbb{E}[N_j(T)] \leqslant \dfrac{8 \ln T}{\Delta_j^2} + 2$ and by bounding $2\Delta_j \leqslant 2$.

NOTE: Keep in mind the rewriting as we will often use it!

$$\overline{R}_T = T\mu^* - \mathbb{E}\left[\sum_{t=1}^{T} Y_t\right] = \sum_{a=1}^{K} \Delta_a \, \mathbb{E}[N_a(T)]$$

 Lecture 4 — Spring 2021 term Lecture 4 — Spring 2021 term

<u>Proof</u> [of the Proposition] : We fix an optimal arm $a^* \in \{1, \dots, K\}$, i.e. s.t. $\mu_{a^*} = \mu^*$.

$\longrightarrow$ It will show why this algorithm is called UCB :

<u>Because</u> $\hat{\mu}_{j,t-1} + \sqrt{\dfrac{2 \ln t}{N_j(t-1)}}$ will indeed appear as an upper confidence bound on μ_j

estimate based on the raw performance $\leftrightarrow$ exploitation of the results

larger for arms j not much sampled so far $\leftrightarrow$ forces some exploration

The UCB algorithm realizes some compromise / trade off between exploitation & exploration.

Later on we compare these statements to the Hoeffding - Azuma inequality

$\longrightarrow$ <u>LEMMA</u> : $\forall j,\ \forall t \geqslant j$ (so that $N_j(t) \geqslant 1$)

$$\forall \delta \in (0,1), \qquad \mathbb{P}\left\{ \mu_j > \hat{\mu}_{j,t} - \sqrt{\frac{\ln(1/\delta)}{2 N_j(t)}} \right\} \geqslant 1 - t\delta.$$

replace $Y \sim \nu_a$ by $1-Y \sim \nu'_a$; then $\mu'_a = 1-\mu_a$ as well and ν'_a supported on $[0,1]$

Or By symmetry: $\forall \delta \in (0,1)$, $\quad \mathbb{P}\left\{ \mu_j < \hat{\mu}_{j,t} + \sqrt{\dfrac{\ln(1/\delta)}{2 N_j(t)}} \right\} \geqslant 1 - t\delta$

$\longrightarrow$ Application : $\quad N_i(T) = 1 + \displaystyle\sum_{t=K+1}^{T} \mathbb{1}_{\{I_t = i\}}$

We show below that $t \geqslant K+1$ and $I_t = i$ entails one of the following :

(i) $\quad \hat{\mu}_{i,t-1} > \mu_i + \sqrt{\dfrac{2 \ln t}{N_i(t-1)}}$ $\quad$ [$\mu_i <$ lower confidence bound]

(ii) $\quad \hat{\mu}_{a^*,t-1} < \mu^* - \sqrt{\dfrac{2 \ln t}{N_{a^*}(t-1)}}$ $\quad$ [$\mu^* >$ upper confidence bound]

(iii) $\quad N_i(t-1) \leqslant \dfrac{8 \ln t}{\Delta_i^2}$ $\quad$ [i not played often enough yet]

Indeed, we would otherwise have

$$\hat{\mu}_{a^*, t-1} + \sqrt{\frac{2\ln t}{N_{a^*}(t-1)}} \geqslant \mu^* \qquad \text{negation of (ii)}$$

$$= \mu_i + \Delta_i \qquad \text{definition of } \Delta_i$$

$$> \mu_i + 2\sqrt{\frac{2\ln t}{N_i(t-1)}} \qquad \left\{ \begin{array}{l} \text{the negation of (iii)} \\ \text{is } \Delta_i^2 > 8\ln t / N_i(t-1) \end{array} \right.$$

$$\geqslant \hat{\mu}_{i,t-1} + \sqrt{\frac{2\ln t}{N_i(t-1)}} \qquad \text{negation of (i)}$$

the $>$ inequality between these two quantities would contradict $I_t = i$, that is, $i \in \arg\max_j \left\{ \hat{\mu}_{j,t} + \sqrt{2\ln t / N_j(t-1)} \right\}$

Thus,

$$\mathbb{E}[N_i(T)] \leqslant 1 + \sum_{t=K+1}^{T} \mathbb{P}\left(\hat{\mu}_{i,t-1} > \mu_i + \sqrt{\frac{2\ln t}{N_i(t-1)}} \right)$$

$$+ \sum_{t=K+1}^{T} \mathbb{P}\left(\hat{\mu}_{a^*, t-1} < \mu^* - \sqrt{\frac{2\ln t}{N_{a^*}(t-1)}} \right)$$

each $\leqslant t\delta$ where $\delta = 1/t^4$

$$+ \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{1}_{\{ I_t = i \;\&\; N_i(t-1) \leqslant 8\ln t / \Delta_i^2 \}} \right]$$

$8\ln t \leqslant 8\ln T$

$$\leqslant 1 + 2\sum_{t=K+1}^{T} t^{-3} + \mathbb{E}\Bigg[\underbrace{\sum_{t=K+1}^{T} \mathbb{1}_{\{ N_i(t-1) \leqslant 8\ln T/\Delta_i^2 \;\&\; I_t = i \}}}_{\text{deterministically upper bounded by}} \Bigg]$$

$$\leqslant 1 + \underbrace{2\sum_{t=K+1}^{T} t^{-3}}_{\leqslant 2\int_1^{+\infty} t^{-3}\,dt \;=\; [-t^{-2}]_1^{+\infty} \;=\; 1} + \left(\frac{8\ln T}{\Delta_i^2} + 1 \right) - 1$$

as $I_t = i$ only if $N_i(t-1) \leqslant \frac{8\ln T}{\Delta_i^2}$
thus only if $N_i(t) \leqslant \frac{8\ln T}{\Delta_i^2} + 1$
so that the total sum $\sum_{s=1}^{t} \mathbb{1}_{\{I_s = i\}} = N_i(t)$ is controlled $\forall t$ by this number

-1 because $I_i = i$ is not included in the $\sum_{t=K+1}^{T} \cdots$

Thus:

$$\mathbb{E}[N_i(T)] \leqslant \frac{8\ln T}{\Delta_i^2} + 2$$

<u>Proof of the lemma</u> (Hoeffding-Azuma inequality with a random number of summands):

Let

$$Z_t = \sum_{s=1}^{t} (Y_s - \mu_a)\, \mathbb{1}_{\{I_s = a\}} \,;$$ we successively prove:

(0) $(Z_t)_{t \geq 0}$ is a martingale wrt. $(\mathcal{F}_t)_{t \geq 0} = (\sigma(Y_1, \ldots Y_t))$

where $\mathcal{F}_0 = \{\emptyset, \Omega\}$ trivial σ-algebra

<u>Indeed:</u> each I_t is $\mathcal{F}_{t-1}$-measurable (picked based only on past payoffs)

thus Z_t is $\mathcal{F}_t$-adapted

Showing that it is a martingale amounts to showing

$$\mathbb{E}\big[(Y_t - \mu_a)\, \mathbb{1}_{\{I_t = a\}} \mid Y_1 \ldots Y_{t-1} \big] \overset{?}{=} 0 \quad \text{a.s.}$$

but since I_t is $\mathcal{F}_{t-1}$-measurable, this quantity equals

$$\mathbb{E}\big[(Y_t - \mu_a)\, \mathbb{1}_{\{I_t = a\}} \mid Y_1 \ldots Y_{t-1}, I_t \big]$$
$$= \Big(\mathbb{E}\big[Y_t \mid I_t, Y_1 \ldots Y_{t-1} \big] - \mu_a \Big)\, \mathbb{1}_{\{I_t = a\}}$$
$$= (\mu_{I_t} - \mu_a)\, \mathbb{1}_{\{I_t = a\}} = 0 \text{ a.s., as desired}$$

by the bandit model, Y_t is drawn independently at random given I_t, thus by the very bandit model, this conditional expectation equals μ_{I_t}

Then: (try to prove these statements by yourself, as an exercise for the next session):

(1) For all $x \in \mathbb{R}$, $(M_t) = \left(\exp\left(x Z_t - \frac{x^2}{8} N_a(t) \right) \right)_{t \geq 0}$ is an $(\mathcal{F}_t)_{t \geq 0}$ adapted supermartingale

↳ in particular $\mathbb{E}[M_t] \leq 1$ for all t

(2) $\forall \varepsilon > 0$, $\forall \ell \geq 1$, $\mathbb{P}\{ Z_t \geq \varepsilon \text{ and } N_a(t) = \ell \} \leq e^{-2\varepsilon^2/\ell}$

(3) From these we will conclude.