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")

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

or ("statistical version", totally equivalent)

$$\forall \delta \in (0,1), \quad \text{with probability at least } 1-\delta, \quad \sum_{t=1}^{T} X_t - \sum_{t=1}^{T} \mathbb{E}[X_t \mid \mathcal{F}_{t-1}] \leq \sqrt{\frac{\sum_{t=1}^{T} (b_t - a_t)^2}{2} \ln \frac{1}{\delta}}$$

Note: Hoeffding's inequality is the special case when all X_t are independent and $\mathcal{F}_{t-1} = \sigma(X_1 \dots X_{t-1})$, 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 $s \in \mathbb{R}$,

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

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

Proof (of the theorem):

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

We denote $S_T = \sum_{t=1}^{T} X_t - \mathbb{E}[X_t \mid \mathcal{F}_{t-1}]$

(martingale = sum of martingale increments or martingale differences)

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

$$\mathbb{P}\{S_T > \varepsilon\} = \mathbb{P}\{e^{sS_T} > e^{s\varepsilon}\} \leq e^{-s\varepsilon}\mathbb{E}[e^{sS_T}]$$

($\forall s > 0$; Markov's inequality)

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

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

The extension of Hoeffding's lemma ensures that

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

so that

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

Substituting above:

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

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

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

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

But first (reminder) ↓ unconditional version

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

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

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

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 $\eta \mapsto Xe^{\eta X}$ 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 $\mathbb{E}[Xe^{sX}]$

with

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

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

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

under the probability Q defined by

$$\frac{dQ}{dP}(w) = \frac{e^{sX(w)}}{\mathbb{E}[e^{sX}]}$$

$X \in [a,b]$: $\mathrm{Var}_Q(X) = \inf_{\mu \in \mathbb{R}} \mathbb{E}_Q\left[(X-\mu)^2\right]$

$$\leq \mathbb{E}_Q\left[\left(X - \frac{a+b}{2}\right)^2\right] = \frac{(b-a)^2}{4}$$

Taylor: $\exists x$ s.t. $\psi(s) = \underbrace{\psi(0)}_{=0} + s\underbrace{\psi'(0)}_{=0} + \frac{s^2}{2}\underbrace{\psi''(x)}_{\leq (b-a)^2/4}$

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

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 $\mathbb{E}[\,\cdot\,|\mathcal{G}_j]$?

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

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

= some conditional variance under a different probability measure?

Yes, using the notion of « regular conditional probability », which always exists in our case, we could perform a change of measure again and identify a conditional variance.

But **PLEASE!** this is extremely heavy maths... I can't inflict that to you....

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 $A = a - \mathbb{E}[X|\mathcal{G}_j] \leq 0$, $B = b - \mathbb{E}[X|\mathcal{G}_j] \geq 0$ 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:

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

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

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

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

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:

$$\forall u,v \in \mathbb{R},\ \forall p \in [0,1],\ \forall s \in \mathbb{R},\quad \ln\left(pe^{su} + (1-p)e^{sv}\right) \leq s\left(pu + (1-p)v\right) + \frac{s^2}{8}(v-u)^2$$

($\ln$ of expected value of e^{sZ} where $Z = \begin{cases} u & \text{w.p. } p \\ v & \text{w.p. } 1-p \end{cases}$; the expected value of Z; range is $[u,v]$)

In particular,

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

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

Summarizing:

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

Proof 3 My preferred (not only because I found 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 $A \in \mathcal{G}_j$ 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

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

Why do we consider the $\mathbb{E}_A$? Because $\mathbb{E}[X \mid \mathcal{G}_j]$ is the unique $\mathcal{G}_j$-measurable random variable such that

$$\forall A \in \mathcal{G}_j,\quad \mathbb{E}[X \mathbb{1}_A] = \mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}_j]\,\mathbb{1}_A\right]$$

or, equivalently,

$$\forall A \in \mathcal{G}_j \text{ s.t. } \mathbb{P}(A) > 0,\quad \mathbb{E}_A[X] = \mathbb{E}_A\left[\mathbb{E}[X \mid \mathcal{G}_j]\right].$$

Now, consider the event

$$H = \left\{ \mathbb{E}[e^{sX} \mid \mathcal{G}_j] - e^{s\mathbb{E}[X \mid \mathcal{G}_j]} e^{s^2(b-a)^2/8} > 0 \right\} \in \mathcal{G}_j$$

We want to prove that $\mathbb{P}(H) = 0$. We proceed by contradiction:

If we had $\mathbb{P}(H) > 0$, then $\mathbb{P}_H$ would be defined and

$$\mathbb{E}_H[e^{sX}] = \mathbb{E}_H[\mathbb{E}[e^{sX}|\mathcal{G}]] > \mathbb{E}_H[e^{s\mathbb{E}[X|\mathcal{G}]}]\, e^{s^2(b-a)^2/8}$$

(= : a property of conditional expectations; > : by definition of H and because a function >0 on a set with probability >0 has an expectation >0)

$$\geq e^{s\,\mathbb{E}_H[\mathbb{E}[X|\mathcal{G}]]}\, e^{s^2(b-a)^2/8}$$

(unconditional) Jensen's inequality

$$= e^{s\,\mathbb{E}_H[X]}\, e^{s^2(b-a)^2/8}$$

which would be in contradiction with the unconditional Hoeffding's lemma:

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

We conclude that $\mathbb{P}(H) = 0$.

A final remark: A better version of Hoeffding's lemma in its conditional form is the following one:

> **Hoeffding's lemma:** X random variable s.t. there exists G $\mathcal{G}$-measurable and integrable, as well as $a, b \in \mathbb{R}$ with: $G + a \leq X \leq G + b$ a.s.
>
> Then $\forall s \in \mathbb{R}$, $\ln \mathbb{E}[e^{sX}|\mathcal{G}] - s\mathbb{E}[X|\mathcal{G}] \leq \frac{s^2}{8}(b-a)^2$

→ In my first statement, I only considered $G \equiv 0$.

Why can we consider general G's? See next page.

Then we get the following extension of the Hoeffding-Azuma inequality:

Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration and $(X_t)_{t\geq 1}$ be a sequence of adapted r.v. such that

$\forall t$, $\exists G_t$ $\mathcal{F}_{t-1}$-measurable and integrable, $\exists a_t, b_t$ | $a_t + G_t \leq X_t \leq b_t + G_t$

then, $\forall \delta \in (0,1)$, w.p. $1-\delta$,

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

Why does Hoeffding's lemma hold for more general G's ?

We consider $a \leq X - G \leq b$ as

Hoeffding's lemma indicates that

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

We have $G \in L^1$ thus $X \in L^1$ and therefore,

$$s\, \mathbb{E}[X-G|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}] - G \in L^1$$

Now, we also have $\mathbb{E}\left[e^{s(X-G)} \,|\, \mathcal{G}\right] = e^{-sG}\, \mathbb{E}\left[e^{sX} \,|\, \mathcal{G}\right]$

Hence the result by cancelling out the $-sG$ terms.

(I don't think we need here that $e^{sG} \in L^1$ or equivalently $e^{sX} \in L^1$; do we ?)

↳ maybe we do / you all seemed to say that we did... anyway in most of our applications, G will be bounded as well !

What can we do when no convexity assumption holds?

↳ Non-convex aggregation via randomization

Example 1: 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

Example 2: 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 $f_{jt} \in \mathcal{Y}$, $j \in \{1 \dots N\}$
   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 $\ell(f_{jt}, y_t)$ 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]
e.g., $\mathcal{Y} = \{1 \dots M\}$ in M-ary classification

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

Solution: (at least, an easy solution, these might be others)

Draw $J_t \in \{1 \dots N\}$ at random and pick { action J_t (in Example 1); forecast $\hat{y}_t = f_{J_t, t}$ (in Example 2) }

⇓

General setting: Simultaneously { 1. Opponent picks $\ell_t = (\ell_{1t}, \dots \ell_{Nt}) \in \mathbb{R}^N$; 2. Statistician draws $J_t \in \{1, \dots N\}$ }

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

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

⚠ 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: $R_T = \sum_{t=1}^T \ell_{J_t,t} - \min_k \sum_{t=1}^T \ell_{kt} = \sum_{t=1}^T \ell_{J_t,t} - \sum_{t=1}^T \sum_{j=1}^N p_{jt} \ell_{jt}$

$+ \sum_{t=1}^T \sum_j p_{jt} \ell_{jt} - \min_k \sum_{t=1}^T \ell_{kt}$

This can be controlled independently of the probability distributions chosen

We already learned how to control this term!

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

We denote $\mathcal{F}_{t-1} = \sigma\{(\ell_s, p_s, J_s)_{s \le t-1}\}$: 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 $\overline{R}_T \le O\left((M-m)\sqrt{T \ln N}\right)$ 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 $\ell_{jt} \in [m, M]$ $\forall j,t$, then, no matter which p_t were selected

with $X_t = \ell_{J_t,t}$ and $i_{X_t} = M$ and $a_{X_t} = m$ and $G_X = \sum p_{jt} \ell_{jt}$

$$\mathbb{P}\left\{ S_T \le (M-m)\sqrt{\frac{T}{2} \ln \frac{1}{\delta}} \right\} \ge 1 - \delta$$

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

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

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

$$R_T \leq (M-m)\sqrt{T}\left(\sqrt{\ln N} + \sqrt{\tfrac{1}{2}\ln\tfrac{1}{\delta}}\right) + (M-m)\left(2 + \tfrac{4}{3}\ln N\right)$$

This is called a high probability bound; it is non-asymptotic → Exercise: Can you get a high probability bound of the form: $\forall \varepsilon \in (0,1)$, with probability $\geq 1-\delta$, $\forall T$, $R_T \leq \ldots$?

Consequence: Asymptotic almost-sure bound.

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

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

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

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

(limsup of a sequence of numbers)

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

Show that we actually have $\limsup_{T \to +\infty} \frac{R_T}{(M-m)\sqrt{T \ln(\ln T)}} \leq C$ 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 $1-\delta$,

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

(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 $\ell_t \in [m, M]$,

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

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!)

Note: If you can solve (3) and have interesting things to say for (4), please send me your work.

Stochastic bandits. Finitely many arms.

Setting:

K arms indexed by $1, 2, \dots K$

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

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

- The decision-maker picks $I_t \in \{1, \dots K\}$, 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 $\bar{R}_T = T\mu^* - \mathbb{E}\left[\sum_{t=1}^{T} Y_t\right]$ to be controlled

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

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 $\bar{R}_T$ is a very "expected" notion of regret

$\bar{R}_T \le$ probably $< \mathbb{E}\left[\max_{a=1,\dots K} \sum_{t}^{T} \dots - \sum_{t=1}^{T} Y_t\right]$

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

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

(the 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

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

This regret bound is obtained via the following proposition :

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

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

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 $[a,1]$, not on the specific distributions ν_i at hand) : for the UCB algorithm,

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

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] :

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

where by definition of the bandit model, $\mathbb{E}[Y_t | I_t] = \mu_{I_t}$ ← Given I_t, Y_t is drawn at random according to ν_{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 (in view of $T = \sum_j \mathbb{E}[N_j(T)]$)

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

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

↓ it suffices to consider the suboptimal arms...

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

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

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

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

→ It will show why this algorithm is called UCB :

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

- $\hat{\mu}_{j,t-1}$: estimate based on the raw performance ↔ exploitation of the results
- $\sqrt{\dfrac{2 \ln t}{N_j(t-1)}}$: larger for arms j not much sampled so far ↔ 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)

→ **LEMMA:** $\forall j, \ \forall t \geq j$ (so that $N_j(t) \geq 1$)

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

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

Or

→ By symmetry: $\forall \delta \in (0,1)$,

$$\mathbb{P}\left\{ \mu_j < \hat{\mu}_{j,t} + \sqrt{\frac{\ln(1/\delta)}{2 N_j(t)}} \right\} \geq 1 - t\delta$$

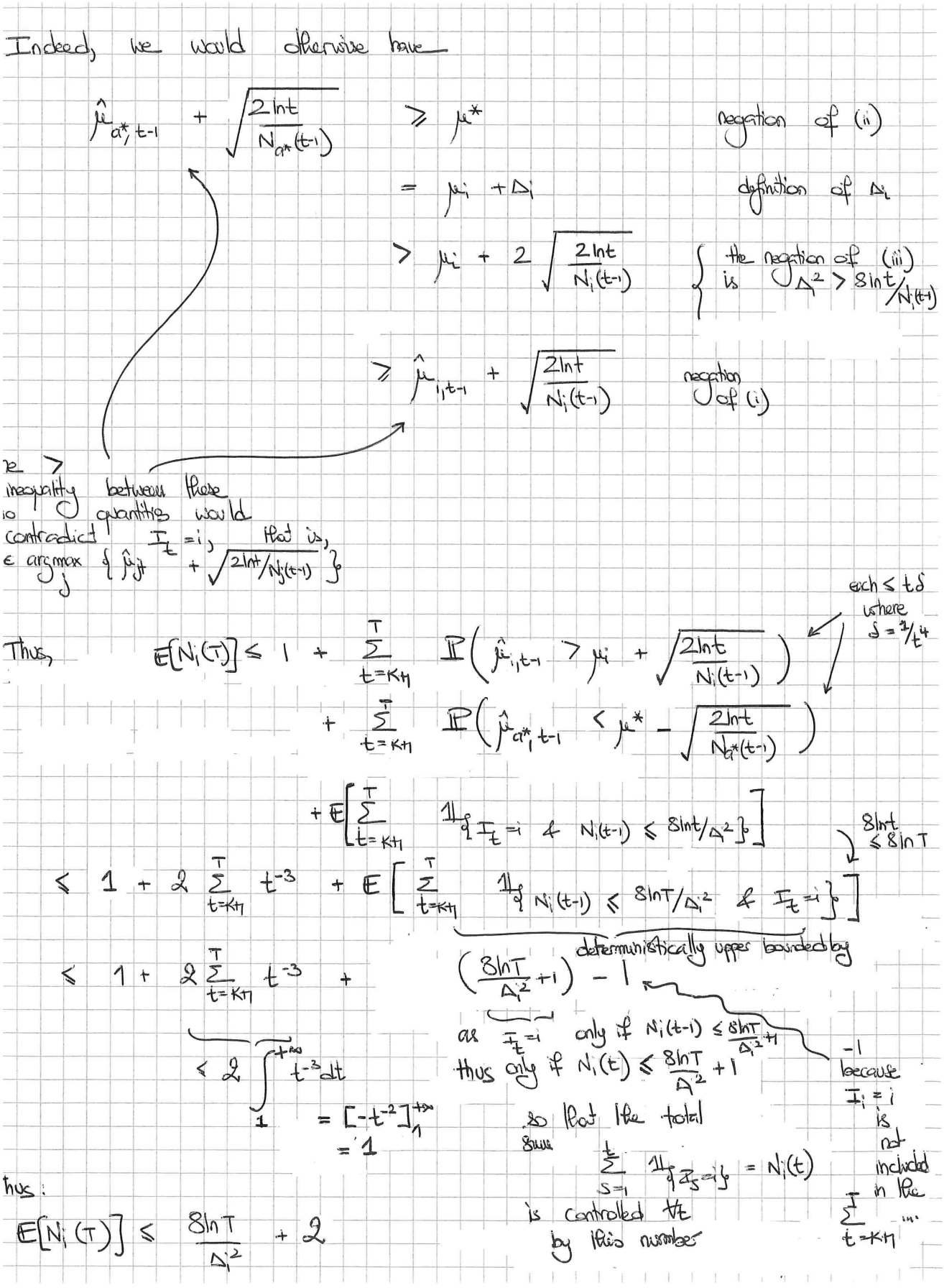
→ Application : $N_i(T) = 1 + \displaystyle\sum_{t=K+1}^{T} \mathbb{1}_{\{I_t = i\}}$

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

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

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

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


Indeed, we would otherwise have
negation of (ii)
definition of Δ_i
the negation of (iii) is $\Delta^2 > 8\ln t / N_i(t-1)$
negation of (i)
$\geq$ inequality between these quantities would contradict $I_t = i$, that is, $i \in \arg\max_j \{ \hat{\mu}_{j\,t} + \sqrt{2\ln t / N_j(t-1)} \}$
each $\leq t\delta$ where $\delta = 1/t^4$
$\mathbb{E}[N_i(T)] \leq 1 + \sum_{t=K+1}^{T} \mathbb{P}\left(\hat{\mu}_{i,t-1} > \mu_i + \sqrt{\frac{2\ln t}{N_i(t-1)}} \right)$
$+ \sum_{t=K+1}^{T} \mathbb{P}\left(\hat{\mu}_{a^*,t-1} < \mu^* - \sqrt{\frac{2\ln t}{N_{a^*}(t-1)}} \right)$
$+ \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{1}\{ I_t = i \text{ and } N_i(t-1) \leq 8\ln t/\Delta_i^2 \} \right]$
$8\ln t \leq 8\ln T$
$\leq 1 + 2\sum_{t=K+1}^{T} t^{-3} + \mathbb{E}\left[\sum_{t=K+1}^{T} \mathbb{1}\{ N_i(t-1) \leq 8\ln T/\Delta_i^2 \text{ and } I_t = i \} \right]$
deterministically upper bounded by
$\leq 1 + 2\sum_{t=K+1}^{T} t^{-3} +$
$\left(\frac{8\ln T}{\Delta_i^2} + 1 \right) - 1$
$\leq 2\int_1^{+\infty} t^{-3} dt = [-t^{-2}]_1^{+\infty} = 1$
as $I_t = i$ only if $N_i(t-1) \leq \frac{8\ln T}{\Delta_i^2} + 1$
thus only if $N_i(t) \leq \frac{8\ln T}{\Delta_i^2} + 1$
-1 because $I_t = i$ is not included in the $\sum_{t=K+1}^{T}$ ...
so that the total same $\sum_{s=1}^{t} \mathbb{1}\{ I_s = i \} = N_i(t)$ is controlled the by this number
thus:
$\mathbb{E}[N_i(T)] \leq \frac{8\ln T}{\Delta_i^2} + 2$

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

Let

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

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

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

Indeed: 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

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

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

$$\mathbb{E}[(Y_t - \mu_a) \mathbb{1}_{\{I_t = a\}} \mid Y_1 \dots Y_{t-1}, I_t]$$

$$= (\mathbb{E}[Y_t \mid I_t, Y_1 \dots Y_{t-1}] - \mu_a) \mathbb{1}_{\{I_t = a\}}$$

$$= (\mu_{I_t} - \mu_a) \mathbb{1}_{\{I_t = a\}} = 0 \text{ a.s.,} \quad \text{as desired}$$

(by the bandit model, I_t is drawn dependently at random given I_t, thus by the 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) $\forall \varepsilon > 0, \forall \ell \geq 1, \quad \mathbb{P}\{Z_t \geq \varepsilon \text{ and } N_a(t) = \ell\} \leq e^{-2\varepsilon^2/\ell}$

(3) From these we will conclude.