

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 (i.e. $\forall t \geq 1$, 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 »)

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

or (« statistical version », totally equivalent)

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

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 $\mathbb{E}[X_t | \mathcal{F}_{t-1}] = \mathbb{E}[X_t]$.

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 $s \in \mathbb{R}$,

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

(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):

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

$\uparrow$ (martingale) = $\uparrow$ sum of $\uparrow$ martingale increments or martingale differences

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

$$\mathbb{P}\{S_T > \varepsilon\} \stackrel{\forall \lambda > 0}{=} \mathbb{P}\{e^{\lambda S_T} > e^{\lambda \varepsilon}\} \leq e^{-\lambda \varepsilon} \mathbb{E}[e^{\lambda S_T}]$$

↑
Markov's inequality

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

- 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 $T-1 \rightarrow T$, where $T \geq 2$:

+ taking expectations by tower rule: $\mathbb{E} = \mathbb{E}[\mathbb{E}[\cdot | \mathcal{F}_0]]$

The extension of Hoeffding's lemma ensures that

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

so that

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

Substituting above:

$$\mathbb{P}\{S_T > \varepsilon\} \leq \inf_{\lambda > 0} \exp\left(-\lambda \varepsilon + \frac{\lambda^2}{8} \sum_{t=1}^T (b_t - a_t)^2\right)$$

strictly convex function to minimize in the exponent: minimum achieved at λ^*

such that $\lambda^* \sum_{t=1}^T (b_t - a_t)^2/4 = \varepsilon$ (gradient vanishes)

ie $\lambda^* = 4\varepsilon / \sum_{t=1}^T (b_t - a_t)^2$

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

But first (reminder) $\downarrow$ unconditional version

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

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

Proof (most elegant one I know of):

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

ψ is differentiable at each $s \in \mathbb{R}$: c.f. X bounded, thus $\eta \mapsto X e^{\eta X}$ locally dominated around s by an integrable r.v. independent of η
 $\hookrightarrow$ 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:

$$\psi''(s) = \frac{\mathbb{E}[X^2 e^{sX}] \mathbb{E}[e^{sX}] - (\mathbb{E}[X e^{sX}])^2}{(\mathbb{E}[e^{sX}])^2} = \text{Var}_Q(X)$$

under the probability Q defined by

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

$$X \in [a, b]: \quad \text{Var}_Q(X) = \inf_{\mu \in \mathbb{R}} \mathbb{E}_Q[(X-\mu)^2] \leq \mathbb{E}_Q\left[\left(X - \frac{a+b}{2}\right)^2\right] = \frac{(b-a)^2}{4}$$

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

i.e.,

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

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 E by $E[\cdot|G]$?

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

$$\text{a.s., } \Psi''(s) \text{ exists and equals } \Psi''(s) = \frac{E[X^2 e^{sx}|G] E[e^{sx}|G] - (E[X e^{sx}|G])^2}{(E[e^{sx}|G])^2}$$

= 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:

$$Y = X - E[X|G] \in [A, B] \quad \text{where } A = a - E[X|G] \leq 0, \quad B = b - E[X|G] \geq 0$$

one both G -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 $E[\cdot|G]$: using $E[Y|G] = 0$ and A, B G -measurable:

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

← 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:

$$\forall u, v \in \mathbb{R}, \quad \forall p \in [0, 1], \quad \ln(p e^{su} + (1-p) e^{sv}) \leftarrow \ln \text{ of expected value of } e^{sZ} \text{ where } Z = \begin{cases} u & \text{w.p. } p \\ v & \text{w.p. } 1-p \end{cases}$$

$$\leq s(pu + (1-p)v) + \frac{s^2}{8}(v-u)^2$$

↖ expected value of Z
↖ range is [u, v]

In particular,

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

$$= \exp\left(\frac{s^2}{8}(b-a)^2\right) \quad \leftarrow \text{recall that a.s., } B-A = b-a$$

Summarizing:

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

$$= \mathbb{E}[e^{sX} | \mathcal{G}] \times \exp(-s \mathbb{E}[X | \mathcal{G}])$$

Proof 3

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

The unconditional version of Hoeffding's lemma ensures that

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

Why do we consider the $\mathbb{E}_A$?

random variable such that

or, equivalently,

Because $\mathbb{E}[X | \mathcal{G}]$ is the unique $\mathcal{G}$ -measurable

$$\forall A \in \mathcal{G}, \quad \mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X | \mathcal{G}] \mathbb{1}_A]$$

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

Now, consider the random variable

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

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

H is $\mathcal{G}$ -measurable & this suffices to show that for all $A \in \mathcal{G}$ with $P(A) > 0$,
 $E_A[H] \geq 0$ that is, $E[H 1_A] \geq 0$

$$\begin{aligned} \text{Indeed, } E_A[H] &= e^{s^2(b-a)/8} E_A[e^{sEX|G}] - E_A[E[e^{sX}|G]] \\ &= e^{s^2(b-a)/8} E_A[e^{sEX|G}] - E_A[e^{sX}] \\ &\stackrel{(\text{Jensen})}{\geq} e^{s^2(b-a)/8} e^{sE_A[X]} - E_A[e^{sX}] \geq 0 \end{aligned}$$

(unconditional version of Hoeffding's lemma)

Proof 1

Let's get back to it. In what follows all expectations relative to the original probability distribution P will be denoted by E , and expectations under alternative distributions Q will be denoted by E_Q .

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

Since $E[L_s] = E[E[L_s|G]] = 1$, L_s is a density

We define the probability Q_s as: $\frac{dQ_s}{dP} = L_s$

(2) We show that $\psi(s) = \frac{E[Xe^{sX}|G]}{E[e^{sX}|G]}$ also equals $E_{Q_s}[X|G]$

$\psi'(s) = \frac{E[X^2 e^{sX}|G]}{E[e^{sX}|G]}$ [see expression some pages ago] $\text{Var}_{Q_s}(X|G)$

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:

$$E[Z L_s | G] = E_{Q_s}[Z | G]$$

or equivalently, that $\forall A \in \mathcal{G}$, $E[E[Z L_s | G] 1_A] = E[E_{Q_s}[Z | G] 1_A]$
 (since both sides are $\mathcal{G}$ -measurable)

But: $E[E[Z L_s | G] 1_A] = E[Z L_s 1_A] \stackrel{\substack{\uparrow \\ \text{one characterization of } E[\cdot|G]}}{=} E[Z 1_A] \stackrel{\substack{\uparrow \\ L_s \text{ is } \frac{dQ_s}{dP}}}{=} E_{Q_s}[Z 1_A]$

and on the other end

$$E[E_{Q_S}[Z|G] 1_A]$$

$$= E[E_{Q_S}[Z|G] E[L_S|G] 1_A]$$

$$= E[E[L_S E_{Q_S}[Z|G] 1_A | G]]$$

$$= E[L_S E_{Q_S}[Z|G] 1_A]$$

$$= E_{Q_S}[E_Q[Z|G] 1_A]$$

$$= E_{Q_S}[E_Q[Z 1_A | G]] = E_{Q_S}[Z 1_A]$$

as $E[L_S | G] = 1$ as
by definition of L_S

$E_{Q_S}[Z|G]$ and 1_A
can go inside $E[|G]$

"tower rule"

$$L_S = \frac{dQ_S}{dP}$$

1_A G -measurable
"tower rule"

which concludes the proof of (2).

$$(3) \quad \psi'(0) = E[X|G] \quad (\text{clear})$$

and for all x ,

$$\psi''(x) = \text{Var}_{Q_x}(X|G) \leq \frac{(b-a)^2}{4}$$

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

(we prove that below)

$$\text{Indeed, } \text{Var}_{Q_x}(X|G) \leq \text{Var}_Q(X|G) + E[(Y - E[X|G])^2 | G]$$

for all bounded Y

=

$$E[(Y - X)^2 | G]$$

and we take $Y = \frac{b+a}{2}$

and use $X \in [a, b]$ to
conclude as in the unconditional case.

conditional
bias/variance
decomposition

A final remark:

A useful ~~app~~ improvement

(useful to get slightly better constants)

Hoeffding's lemma:

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:

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

Rk: In my first statement I only considered $G \equiv 0$

Proof: Apply the first statement to $X-G$

Consequence: some remark for 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

$\forall t, \quad \exists G_t \quad \mathcal{F}_{t-1}$ -measurable and bounded,
 $\exists a_t, b_t$ real numbers s.t.

$$a_t + G_t \leq X_t \leq b_t + G_t$$

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

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

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: $\rightarrow \leftarrow \uparrow \downarrow$)

(binary if accept/reject actions)

1. Opponent picks state of the world y_t
2. Statistician picks action $j_t \in \{1, \dots, N\}$ } simultaneously
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 $y_t \in Y$
2. Simultaneously, experts provide forecasts $f_{j,t} \in Y$, $j \in \{1, \dots, N\}$
and statistician picks forecast $\hat{j}_t \in Y$
3. y_t and $\hat{j}_t$ are revealed, losses $\ell(y_t, y_t)$ and $\ell(f_{j_t}, y_t)$ are suffered

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

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

Solution:

(at least, an easy solution, there might be others)

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

and pick action J_t (in Example 1)

forecast $\hat{j}_t = f_{J_t, t}$ (in Example 2)

↳

General setting:

Simultaneously

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

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

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

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

Methodology:

We denote by $p_t = (p_{t1} \dots p_{tN}) \in \mathcal{X}$ the probability distribution used to draw J_t , conditionally to the past

$$\text{Regret: } R_T = \sum_{t=1}^T l_{J_t, t} - \min_k \sum_{t=1}^T l_{k, t} = \sum_{t=1}^T l_{J_t, t} - \sum_{t=1}^T \sum_{j=1}^N p_{jt} l_{jt} + \sum_{t=1}^T \sum_{j=1}^N p_{jt} l_{jt} - \min_k \sum_{t=1}^T l_{k, t}$$

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 $(l_s, p_s, J_s)_{s \leq t-1}$

We denote $\mathcal{F}_{t-1} = \sigma\{(l_s, p_s, J_s)_{s \leq t-1}\}$: 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: $E[l_{J_t, t} | \mathcal{F}_{t-1}] = \sum_{j=1}^N p_{jt} l_{jt}$

(J_t is not fixed by the conditioning, only its distribution p_t is so.)

↳ Expected regret (conditionally expected regret) $\bar{R}_T = \sum_{t=1}^T \sum_{j=1}^N p_{jt} l_{jt} - \min_k \sum_{t=1}^T l_{k, t}$

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

↳ Martingale $S_T = \sum_{t=1}^T l_{J_t, t} - \sum_{t=1}^T \sum_{j=1}^N p_{jt} l_{jt}$

The Hoeffding-Azuma inequality

ensures that

if $l_{jt} \in [m, M] \forall j, t$, then, no matter which p_t were selected

with $X_t = l_{J_t, t}$ and $Y_t = \sum_{j=1}^N p_{jt} l_{jt}$, $X_t \in [m, M]$ and $Y_t \in [m, M]$

$$P\{S_T \leq (M-m)\sqrt{\frac{T}{2} \ln \frac{1}{\delta}}\} \geq 1-\delta$$

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

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

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

This is called a high probability bound; it is non-asymptotic

→ Exercise: Can you get a high probability bound of the form: $\forall \varepsilon \in (0,1), \text{ with prob. } \geq 1-\delta, \forall T, R_T \leq \dots$?

Consequence: Asymptotic almost-sure bound.

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

$$\mathbb{P} \left(\limsup_{T \rightarrow \infty} \left\{ R_T > (M-m)\sqrt{T} \left(\sqrt{\ln N} + \sqrt{\frac{1}{2} \ln T} \right) + (M-m)(2 + \frac{4}{3} \ln N) \right\} \right) = 0$$

↑
limsup of events

That is, almost-surely

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

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

$$\text{thus } \limsup_{T \rightarrow \infty} \frac{R_T}{\rho(T)} \leq 1 \text{ a.s.}$$

↑
limsup of a sequence of numbers

or equivalently,

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

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

$$\text{Show that we actually have } \limsup_{T \rightarrow \infty} \frac{R_T}{(M-m)\sqrt{T \ln(\ln T)}} \leq C \text{ a.s.}$$

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

where C is a constant

← and I should have started with that...

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

Because we have deterministic bounds on $\bar{R}_T$, 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:

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

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

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

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

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

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 $\{2^{r+1}, \dots, 2^{r+1}\}$ for $r=1,2,\dots$ and pick $S_r = 1/r^2$ 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 $\mathcal{P}_j$
(over $\mathcal{R}$)
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 $\mathcal{P}_{I_t}$ (given I_t)
- This is the only feedback she gets / the only observation she has access to.

Aim:We denote by $\mu_i = E(\mathcal{P}_i)$ the expectation of $\mathcal{P}_i$ (note: operator E vs. expectation E of an expression involving random variables.)Pseudo-regret $\bar{R}_T = T\mu^* - E\left[\sum_{t=1}^T Y_t\right]$ to be controlledwhere $\mu^* = \max_{j \in K} \mu_j$ 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

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

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

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

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

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

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

always ≥ 1 since each arm was tried sequentially during rounds $1, 2, \dots, K$ - Get a reward Y_t Theorem:

If the distributions $\tilde{\nu}_j$ 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 $\tilde{\nu}_j$ 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 $\tilde{\nu}_i$ at hand (via the gaps $\Delta_i = \mu^* - \mu_i$).

Show the following distribution-free bound (that only

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

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

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, $\leftarrow$ Given I_t , y_t is drawn at random according to ν_{I_t}

$$\mathbb{E}[y_t | I_t] = \mu_{I_t}$$

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

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

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 $T = \sum_j \mathbb{E}[N_j(T)]$)

$$\begin{aligned} \bar{R}_T &= \sum_j (\mu^* - \mu_j) \mathbb{E}[N_j(T)] = \sum_{j=1}^K \Delta_j \mathbb{E}[N_j(T)] \\ &= \sum_{j: \Delta_j > 0} \Delta_j \mathbb{E}[N_j(T)] \end{aligned}$$

it suffices to consider the suboptimal arms...

We conclude by substituting and by bounding $2\Delta_j \leq 2$.

$$\mathbb{E}[N_j(T)] \leq \frac{8 \ln T}{\Delta_j^2} + 2$$

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

$$\begin{aligned} \bar{R}_T &= T\mu^* - \mathbb{E}\left[\sum_{t=1}^T y_t\right] \\ &= \sum_{j=1}^K \Delta_j \mathbb{E}[N_j(T)] \end{aligned}$$

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

→ It will show why this algorithm is called UCB:

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

estimate based on
the raw performance
↔ exploitation of the
results

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
state w.r.t. the
Hoeffding - Azuma inequality

→ [1-7] $\sim \sum_{i=1}^K \mu_i = 1 - \mu_{a^*}$
then $\mu_{a^*} = 1 - \mu_{a^*}$
as well
and μ_{a^*} supported on $[0,1]$

LEMMA:

$\forall j, \forall t \geq j$ (so that $N_j(t) \geq 1$)

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

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

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

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

→ Application: $N_i(T) = 1 + \sum_{t=K+1}^T \mathbb{1}_{\frac{1}{2} \mathbb{I}_t = i}$

We show below that $t \geq K+1$ and $\mathbb{I}_t = i$ entails one of the following:

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

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

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

Indeed, we would otherwise have

$$\hat{\mu}_{a^*, t-1} + \sqrt{\frac{2 \ln t}{N_{a^*}(t-1)}} \geq \mu^*$$

negation of (ii)

$$= \mu_i + \Delta_i$$

definition of Δ_i

$$> \mu_i + 2 \sqrt{\frac{2 \ln t}{N_i(t-1)}}$$

the negation of (iii)
is $\Delta_i^2 > 8 \ln t / N_i(t-1)$

$$\geq \hat{\mu}_{i, t-1} + \sqrt{\frac{2 \ln t}{N_i(t-1)}}$$

negation of (i)

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

Thus,

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

each $\leq t \delta$
where $\delta = \frac{3}{4} \epsilon^4$

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

$$\leq 1 + 2 \sum_{t=K+1}^T t^{-3} + \mathbb{E}\left[\sum_{t=K+1}^T \mathbb{1}_{N_i(t-1) \leq 8 \ln t / \Delta_i^2 \text{ and } I_t = i}\right]$$

$8 \ln t \leq 8 \ln T$

$$\leq 1 + 2 \sum_{t=K+1}^T t^{-3} + \left(\frac{8 \ln T}{\Delta_i^2} + 1\right) - 1$$

deterministically upper bounded by

$$\leq 2 \int_1^{+\infty} t^{-3} dt = \left[-\frac{1}{2} t^{-2}\right]_1^{+\infty} = 1$$

as $I_t = i$ only if $N_i(t-1) \leq \frac{8 \ln T}{\Delta_i^2} + 1$
thus only if $N_i(t) \leq \frac{8 \ln T}{\Delta_i^2} + 1$

so that the total sum

$$\sum_{s=1}^t \mathbb{1}_{I_s = i} = N_i(t)$$

is controlled by this number

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

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):

at

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

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

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

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

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

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

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

y the bandit model,
 t is drawn independently 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 $x \in \mathbb{R}$, $(M_t)_{t \geq 0} = \left(\exp\left(x Z_t - \frac{x^2}{8} N_t(t)\right) \right)_{t \geq 0}$ is an adapted supermartingale $(\mathcal{F}_t)_{t \geq 0}$

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

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

(3) From these we will conclude.