Part 1: $\sqrt{KT}$ distribution-free regret bounds for stochastic bandits

 Lecture 7 — Spring 2021 term Lecture 7 — Spring 2021 term

An exercise of the homework is about proving:

Distribution-free (ie uniform) lower bounds.

We prove: For all $K \geqslant 2$ and $T \geqslant K/5$

$$\inf_{\substack{\text{strategies}\\ \psi}} \ \sup_{\substack{\nu_1,\ldots,\nu_K\\ \text{in } \mathcal{P}([0,1])}} R_T \geqslant \frac{1}{20}\sqrt{TK}.$$

and even: sup over $\nu_1 \ldots \nu_K$ being Bernoulli distributions

We saw that UCB enjoyed a distribution-free regret bound of order $\sqrt{TK \ln T}$, but the $\sqrt{\ln T}$ is unnecessary. The optimal (minmax) distribution-free regret bound for bounded stochastic bandits is of order $\sqrt{TK}$.

We now discuss an algorithm achieving this optimal order of magnitude; it is called MOSS and is a variation on UCB, with a smaller / more careful exploration bonus.

 Lecture 7 — Spring 2021 term Lecture 7 — Spring 2021 term

The MOSS strategy (Minimax Optimal Strategy in the Stochastic case of bandit problems)

Index policy relying on

$$U_a(t) = \hat{\mu}_a(t) + \sqrt{\frac{1}{2N_a(t)} \ln_+\left(\frac{t}{K N_a(t)}\right)}$$

for $t \geq K$,

and where $\ln_+ = \max\{\ln, 0\}$

That is: For $t = 1, \ldots, K$: pull arm $A_t = t$

For $t \geq K+1$: pull arm $A_t \in \underset{a=1..K}{\arg\max}\, U_a(t-1)$

Difference to UCB: we replace the exploration bonus $\sqrt{2\ln t / N_a(t)}$ by $\sqrt{\ln_+\left(t/KN_a(t)\right) / (2N_a(t))}$

↳ no exploration after a was pulled sufficiently often (t/K times)

We prove a distribution-free bound:

Theorem: MOSS is such that $\displaystyle\sup_{\substack{\nu_1, \ldots, \nu_K \\ \text{distributions} \\ \text{over } [0,1]}} \overline{R}_T \leq K - 1 + 45\sqrt{KT}$

(the constant 45 can be improved)

↓

but indeed minimax optimal as its name indicates!

Open question: Take inspiration from the MOSS proof to write a better (more direct) proof for UCB!

 Lecture 7 — Spring 2021 term Lecture 7 — Spring 2021 term

Proof. First step: $U_{a^*}(t-1) \leqslant U_{A_t}(t-1)$ by definition of A_t as an argmax

thus
$$R_T = \sum_{t=1}^{T} \mathbb{E}\big[\mu^* - \mu_{A_t}\big]$$
$$\leqslant \underbrace{K-1}_{\substack{\text{if we played each arm once in the} \\ \text{first } K \text{ steps, and} \\ \text{at most } K-1 \text{ were suboptimal}}} + \sum_{t=K+1}^{T} \mathbb{E}\big[\mu^* - U_{a^*}(t-1)\big] + \underbrace{\sum_{t=K+1}^{T} \mathbb{E}\big[U_{A_t}(t-1) - \mu_{A_t}\big]}_{\leqslant \sqrt{KT} + \sum_{t=K+1}^{T} \mathbb{E}\left[\left(U_{A_t}(t-1) - \mu_{A_t} - \sqrt{K/T}\right)^+\right]}$$

Second step: Control of each $\mathbb{E}\big[\mu^* - U_{a^*}(t)\big]$ term by $20\sqrt{K/t}$

We write for $t \geqslant K$
$$\mathbb{E}\big[\mu^* - U_{a^*}(t)\big]$$
$$\leqslant \mathbb{E}\Big[\big(\mu^* - U_{a^*}(t)\big)^+\Big]$$
$$\leqslant \sum_{\ell=0}^{+\infty} \mathbb{E}\Big[\big(\mu^* - U_{a^*}(t)\big)^+ \mathbb{1}_{\{N_{a^*}(t) \in [x_{\ell+1}, x_\ell]\}}\Big]$$
where $x_\ell = \beta^{-\ell}\, t/K$ for some fixed $\beta > 1$ and $\ell = 0, 1, 2, \ldots$
$$+ \mathbb{E}\Big[\big(\mu^* - U_{a^*}(t)\big)^+ \mathbb{1}_{\{N_{a^*}(t) \geqslant t/K\}}\Big]$$

Now,
$$U_{a^*}(t) = \hat{\mu}_{a^*}(t) + \begin{cases} 0 & \text{if } N_{a^*}(t) \geqslant t/K \\ \sqrt{\dfrac{1}{2N_{a^*}(t)} \ln\left(\dfrac{t}{K N_{a^*}(t)}\right)} \geqslant \underbrace{\sqrt{\dfrac{1}{2x_\ell} \ln\left(\dfrac{t}{K x_\ell}\right)}}_{\text{denoted by } \varepsilon_\ell} & \text{if } N_{a^*}(t) \in [x_{\ell+1}, x_\ell] \end{cases}$$

Therefore, (*)
$$\mathbb{E}\big[\mu^* - U_{a^*}(t)\big]$$
$$\leqslant \mathbb{E}\Big[\big(\mu^* - \hat{\mu}_{a^*}(t)\big)^+ \mathbb{1}_{\{N_{a^*}(t) \geqslant t/K\}}\Big] + \sum_{\ell=0}^{+\infty} \mathbb{E}\Big[\big(\mu^* - \hat{\mu}_{a^*}(t) - \varepsilon_\ell\big)^+ \mathbb{1}_{\{N_{a^*}(t) \in [x_{\ell+1}, x_\ell]\}}\Big]$$

Lemma:
$$\mathbb{E}\Big[\big(\mu^* - \hat{\mu}_{a^*}(t) - \varepsilon\big)^+ \mathbb{1}_{\{N_{a^*}(t) \geqslant n_0\}}\Big] \leqslant \frac{1}{\sqrt{n_0}}\, e^{-2n_0\varepsilon^2}$$

Proof of the lemma:
$$\mathbb{E}\Big[\big(\mu^* - \hat{\mu}_{a^*}(t) - \varepsilon\big)^+ \mathbb{1}_{\{N_{a^*}(t) \geqslant n_0\}}\Big]$$
$$= \int_0^{+\infty} \mathbb{P}\Big\{\mu^* - \hat{\mu}_{a^*}(t) - \varepsilon \geqslant u \ \ \& \ \ N_{a^*}(t) \geqslant n_0\Big\}\, du$$

$$= \int_0^{+\infty} \mathbb{P}\big\{ Z^*_t \geqslant (\varepsilon+u) N_{a^*}(t) \ \ \& \ \ N_{a^*}(t) \geqslant n_0 \big\} \, du$$

where $Z^*_t = N_{a^*}(t)\big(\mu^* - \hat{\mu}_{a^*}(t)\big) = \sum_{s=1}^{t} (\mu^* - Y_s) \mathbb{1}_{\{A_t = a^*\}}$ is a martingale,

and for all $x \in \mathbb{R}$, $S_{x,t} = e^{x Z^*_t - x^2/8 \, N_{a^*}(t)}$ is a supermartingale.

we did something similar for/with UCB

Thus, by Markov-Chernoff, we continue the bounding as, for $x > 0$:

$$= \int_0^{+\infty} \mathbb{P}\Big\{ e^{x Z^*_t - x^2/8 \, N_{a^*}(t)} \geqslant \exp\Big(N_{a^*}(t) \Big(x(\varepsilon+u) - \frac{x^2}{8} \Big)\Big) \ \ \& \ \ N_{a^*}(t) \geqslant n_0 \Big\} \, du$$

$$\leqslant \int_0^{+\infty} \sum_{\ell=n_0}^{+\infty} e^{-2\ell(\varepsilon+u)^2} \, \mathbb{E}\Big[S_{4(\varepsilon+u),t} \mathbb{1}_{\{N_{a^*}(t) = \ell\}} \Big] \, du$$

we pick $x = 4(\varepsilon+u)$ so that $x(\varepsilon+u) - x^2/8 = 2(\varepsilon+u)^2$

independent of ℓ, which will be useful in other proofs!

$$\leqslant e^{-2n_0\varepsilon^2} \int_0^{+\infty} e^{-2n_0 u^2} \underbrace{\mathbb{E}\Big[S_{4(\varepsilon+u),t} \mathbb{1}_{\{N_{a^*}(t) \geqslant n_0\}} \Big]}_{\text{where } \mathbb{E}[S_{4(\varepsilon+u),t}] \leqslant 1} \, du$$

we use $e^{-2\ell(\varepsilon+u)^2} \leqslant e^{-2n_0(\varepsilon+u)^2} \leqslant e^{-2n_0\varepsilon^2} e^{-2n_0 u^2}$

All in all, $\mathbb{E}\Big[\big(\mu^* - \hat{\mu}_{a^*}(t) - \varepsilon\big)^+ \mathbb{1}_{\{N_{a^*}(t) \geqslant n_0\}} \Big]$

$$\leqslant e^{-2n_0\varepsilon^2} \underbrace{\int_0^{+\infty} e^{-2n_0 u^2} \, du}_{\substack{\text{integral of} \\ \text{a Gaussian density,} \\ \text{up to the normalization factor}}} = e^{-2n_0\varepsilon^2} \sqrt{\pi/8n_0} \leqslant e^{-2n_0\varepsilon^2} / \sqrt{n_0}$$

where $\sqrt{\pi/8} \leqslant 1$

$\hookrightarrow$ $\sigma^2 = \dfrac{n_0}{4}$ in $\displaystyle\int_0^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-u^2/2\sigma^2} \, du = 1/2$

Substituting the lemma in $(*)$:

$$\mathbb{E}\big[\mu^* - U_{a^*}(t)\big] \leqslant \sqrt{\frac{K}{t}} + \sum_{\ell=0}^{+\infty} \underbrace{\frac{1}{\sqrt{x_{\ell+1}}}\, e^{-2x_{\ell+1}\varepsilon_\ell^2}}$$

$$\underbrace{\frac{1}{\sqrt{x_{\ell+1}}} \exp\Big(-2\underbrace{x_{\ell+1}\frac{1}{2x_\ell}}_{1/\beta} \ln\Big(\underbrace{\frac{t}{Kx_\ell}}_{}\Big)\Big)}$$

$\ln \beta^\ell = \ell \ln \beta$

$$= \sqrt{\frac{K}{t}}\, \beta^{(\ell+1)/2} \exp\Big(-\frac{\ell}{\beta}\ln\beta\Big)$$

$$= \sqrt{\frac{K}{t}}\, \beta^{1/2 + \ell(1/2 - 1/\beta)}$$

where $\displaystyle\sum_{\ell \geqslant 0} \beta^{\ell(1/2 - 1/\beta)}$ is $< +\infty$ as soon as $\beta \in (1,2)$

E.g., for $\beta = 3/2$,

$$\sum_{\ell=0}^{+\infty} \Big(\frac{3}{2}\Big)^{1/2 + \ell(1/2 - 2/3)} = \sqrt{\frac{3}{2}} \sum_{\ell=0}^{+\infty} \alpha^{+\ell} = \frac{1}{1-\alpha}\sqrt{\frac{3}{2}} \leqslant 19$$

where $\alpha = \Big(\frac{3}{2}\Big)^{1/2 - 2/3} \in (0,1)$

All in all: we obtain a $\sqrt{K/t} + 19\sqrt{K/t} = 20\sqrt{K/t}$ bound, as claimed.

Third step: $\displaystyle\sum_{t=K+1}^{T} \mathbb{E}\Big[\big(U_{A_t}(t-1) - \mu_{A_t} - \sqrt{K/T}\big)^+\Big]$ is $\leqslant 4\sqrt{KT}$

$$= \sum_{t=K}^{T-1} \mathbb{E}\Big[\big(U_{A_{t+1}}(t) - \mu_{A_{t+1}} - \sqrt{K/T}\big)^+\Big]$$

We decompose the expectations of interest according to the $\{A_{t+1}=a\}$ and $\{N_a(t)=\ell\}$:

$$\sum_{t=K}^{T-1} \mathbb{E}\Big[\big(U_{A_{t+1}}(t) - \mu_{A_{t+1}} - \sqrt{K/T}\big)^+\Big] = \sum_{a=1}^{K}\sum_{\ell=1}^{T}\sum_{t=K}^{T-1} \mathbb{E}\Big[\big(U_a(t) - \mu_a - \sqrt{K/T}\big)^+ \mathbb{1}_{\{A_{t+1}=a\}} \mathbb{1}_{\{N_a(t)=\ell\}}\Big]$$

We now use

$$\big(U_a(t) - \mu_a - \sqrt{K/T}\big)^+ \leqslant \big(\hat\mu_a(t) - \mu_a - \sqrt{K/T}\big)^+ + \begin{cases} 0 & \text{if } N_a(t) \geqslant \frac{t}{K} \\ \sqrt{\frac{1}{2N_a(t)}\ln\Big(\frac{t}{K N_a(t)}\Big)} & \end{cases}$$

also smaller than $\sqrt{\frac{1}{2N_a(t)}\ln\Big(\frac{T}{K N_a(t)}\Big)}$ if $N_a(t) < \frac{t}{K}$

and get therefore the upper bound

$$\sum_{a=1}^{K}\sum_{\ell=1}^{T}\sum_{t=K}^{T-1} \mathbb{E}\Big[\big(\hat\mu_a(t) - \mu_a - \sqrt{K/T}\big)^+ \mathbb{1}_{\{A_{t+1}=a\}} \mathbb{1}_{\{N_a(t)=\ell\}}\Big]$$

$$+ \sum_{a=1}^{K}\sum_{\ell=1}^{\lfloor T/K \rfloor} \sqrt{\frac{1}{2\ell}\ln\Big(\frac{T}{K\ell}\Big)}\; \mathbb{E}\Big[\sum_{t=K}^{T-1} \mathbb{1}_{\{N_a(t)=\ell\}} \mathbb{1}_{\{A_{t+1}=a\}}\Big]$$

 Lecture 7 — Spring 2021 term Lecture 7 — Spring 2021 term

We will repeatedly use that

$$\forall a, \ell, \qquad \sum_{t=K}^{T-1} \mathbb{1}_{\{N_a(t)=\ell\}} \mathbb{1}_{\{A_{t+1}=a\}} \leq 1$$

(ie, disjoint union)

cf. $N_a(t)$ increases by 1 whenever a is played

Also,
$$\sum_{\ell=1}^{\lfloor T/K \rfloor} \sqrt{\frac{1}{2\ell}\ln\left(\frac{T}{K\ell}\right)} \leq \int_0^{\lfloor T/K \rfloor} \sqrt{\frac{1}{2x}\ln\left(\frac{T}{Kx}\right)}\,dx$$

change of variable $u = T/(Kx)$

$$\leq \sqrt{\frac{T}{2K}} \int_1^{+\infty} u^{-3/2}\sqrt{\ln u}\,du$$

by the change of variable $u = e^{v^2}$

$$= \sqrt{\frac{T}{2K}} \int_0^{+\infty} 2 v^2 e^{-v^2/2}\,dv = \sqrt{\pi}\sqrt{\frac{T}{K}}$$

Summarizing what we proved so far:

$$\sum_{t=K}^{T-1} \mathbb{E}\left[\left(U_{A_{t+1}}(t) - \mu_{A_{t+1}} - \sqrt{K/t}\right)^+\right] \leq \sqrt{\pi}\sqrt{KT} + \sum_{a=1}^{K} \underbrace{\sum_{\ell=1}^{T}\sum_{t=K}^{T-1} \mathbb{E}\left[\left(\hat\mu_a(t) - \mu_a - \sqrt{K/T}\right)^+ \mathbb{1}_{\{A_{t+1}=a\}}\mathbb{1}_{\{N_a(t)=\ell\}}\right]}_{\text{will show that } \leq \sqrt{\pi/2}\,\sqrt{T/K} \text{ for each } a}$$

We resort again to $Z_{a,t} = N_a(t)\left(\hat\mu_a(t) - \mu_a\right)$ martingale

and $S^{(a)}_{x,t} = e^{xZ_{a,t} - \frac{x^2}{8}N_a(t)}$ supermartingale

where $x = 4\left(\sqrt{K/T} + u\right)$

For each a,

$$\sum_{\ell=1}^{T}\sum_{t=K}^{T-1} \mathbb{E}\left[\left(\hat\mu_a(t) - \mu_a - \sqrt{K/T}\right)^+ \mathbb{1}_{\{A_{t+1}=a\}}\mathbb{1}_{\{N_a(t)=\ell\}}\right]$$

$$= \sum_{\ell=1}^{T}\sum_{t=K}^{T-1} \int_0^{+\infty} \mathbb{P}\left\{xZ_{a,t} - \frac{x^2}{8}N_a(t) \geq N_a(t)\left(x\left(u+\sqrt{K/T}\right) - \frac{x^2}{8}\right) \;\&\; A_{t+1}=a \;\&\; N_a(t)=\ell\right\}du$$

$x = 4\left(\sqrt{K/T}+u\right) > 0$

$$\leq \sum_{\ell=1}^{T}\sum_{t=K}^{T-1} \int_0^{+\infty} \underbrace{e^{-2\ell\left(u+\sqrt{K/T}\right)^2}}_{\leq e^{-2\ell u^2}\, e^{-2\ell K/T}} \mathbb{E}\Big[S^{(a)}_{x,t} \underbrace{\mathbb{1}_{\{A_{t+1}=a\}}\mathbb{1}_{\{N_a(t)=\ell\}}}_{\text{the sum over } t \text{ of these will be } \leq 1}\Big]\,du$$

by Markov-Chernoff

issue: this depends on t... but can be replaced, in some sense, by $S^{(a)}_{x,0} = 1$

 Lecture 7 — Spring 2021 term Lecture 7 — Spring 2021 term

But: remember Doob's maximal inequality for non-negative supermartingales:

$$\mathbb{P}\Big\{ \sup_{t \geq 0} S^{(a)}_{x,t} \geq C \Big\} \leq \frac{\mathbb{E}\big[S^{(a)}_{x,0}\big]}{C} = \frac{1}{C}$$

↳ see also an alternative treatment on the next page

Then,
$$\sum_{\ell=1}^{T} \sum_{t=K}^{T-1} \mathbb{E}\Big[\big(\hat{\mu}_a(t) - \mu_a - \sqrt{K/T}\big)^+ \mathbb{1}_{\{A_{t+1}=a\}} \mathbb{1}_{\{N_a(t)=\ell\}} \Big]$$

as before! ↳
$$= \sum_{\ell=1}^{+\infty} \sum_{t=K}^{T-1} \int_0^{+\infty} \mathbb{P}\Big\{ S^{(a)}_{x,t} \geq e^{2\ell(u+\sqrt{K/T})^2} \text{ and } A_{t+1}=a \text{ and } N_a(t)=\ell \Big\}\, du$$

$$\leq \sum_{\ell=1}^{T} \int_0^{+\infty} \sum_{t=K}^{T-1} \mathbb{P}\Big\{ \Big(\sup_{\tau\geq 0} S^{(a)}_{x,\tau}\Big) \geq e^{2\ell(u+\sqrt{K/T})^2} \text{ and } A_{t+1}=a \text{ and } N_a(t)=\ell \Big\}\, du$$

cf. disjoint union!
$$\leq \sum_{\ell=1}^{T} \int_0^{+\infty} \mathbb{P}\Big\{ \Big(\sup_{\tau\geq 0} S^{(a)}_{x,\tau}\Big) \geq e^{2\ell(u+\sqrt{K/T})^2} \Big\}\, du$$

Doob's maximal inequality
$$\leq \sum_{\ell=1}^{T} \int_0^{+\infty} \underbrace{e^{-2\ell(u+\sqrt{K/T})^2}}_{\leq e^{-2\ell u^2} \times e^{-2\ell K/T}} du \leq \sum_{\ell=1}^{+\infty} \frac{1}{\sqrt{\ell}} e^{-2\ell K/T}$$

and same treatment as in the lemma of the first part of the proof

This step is concluded by calculations:

$$\sum_{\ell=1}^{T} \frac{1}{\sqrt{\ell}} e^{-2\ell K/T} \leq \int_0^T \frac{1}{\sqrt{x}} e^{-2xK/T} dx$$
$$= \sqrt{\frac{T}{2K}} \int_0^{+\infty} \frac{e^{-u}}{\sqrt{u}} du = \sqrt{\frac{T}{2K}} \int_0^{+\infty} e^{-v^2} dv = \sqrt{\frac{\pi}{2}} \sqrt{\frac{T}{K}}$$

Final bound is: $\sqrt{\pi}\sqrt{KT} + K\sqrt{\frac{\pi}{2}}\sqrt{T/K} \leq 4\sqrt{KT}$

<u>General conclusion</u>: Final bound given by

$$K-1 + \Big(\sum_{t=K+1}^{T} 20\sqrt{\frac{K}{t}}\Big) + \sqrt{KT} + 4\sqrt{KT} \leq K-1 + 5\sqrt{KT} + 20\int_0^T \sqrt{K/t}\, dt$$
$$= K-1 + 45\sqrt{KT}$$

Alternative treatment (credits to Enzo Miller) of the end of Step #3:

We were stuck at

$$\sum_{\ell=1}^{T} \sum_{t=K}^{T-1} \int_0^{+\infty} e^{-2\ell u^2} e^{-2\ell K/T} \, \mathbb{E}\Big[S^{(a)}_{x,t} \, \mathbb{1}_{\{A_{t+1}=a\}} \, \mathbb{1}_{\{N_a(t)=\ell\}} \Big] \, du$$

$$= \sum_{\ell=1}^{T} \int_0^{+\infty} e^{-2\ell u^2} e^{-2\ell K/T} \, \mathbb{E}\Bigg[\underbrace{\sum_{t=K}^{T-1} S^{(a)}_{x,t} \, \mathbb{1}_{\{A_{t+1}=a\}} \, \mathbb{1}_{\{N_a(t)=\ell\}}}_{= S^{(a)}_{x,T_\ell}} \Bigg] \, du$$

where T_ℓ is given by:

$T_\ell = \inf \big\{ t \in \{1,\dots,T\} : \ A_{t+1} = a$ and $N_a(t) = \ell \big\}$

We should get $\mathbb{E}\big[S^{(a)}_{x,T_\ell} \big] \leq \mathbb{E}\big[S^{(a)}_{x,0} \big] = 1$ from the optional stopping theorem (« théorème d'arrêt de Doob ») provided some verifications. (T_ℓ should be a bounded stopping time.)

Part 2: Adversarial bandits

Adversarial bandits. (Rather stated in terms of losses than rewards!)

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

1. The opponent and the decision-maker simultaneously choose $\ell_t = (\ell_{j,t})_{j \in \{1,\dots,N\}}$ and $I_t \sim p_t$, where $p_t \in \mathcal{P}(\{1,\dots,N\})$

2. The opponent gets to see p_t and I_t; the decision-maker only observes $\ell_{I_t,t}$ (her own loss).

Regret:

$$R_T = \sum_{t=1}^{T} \ell_{I_t,t} - \min_{j=1\dots N} \sum_{t=1}^{T} \ell_{j,t}$$

vs. Pseudo-regret:

$$\overline{R}_T = \mathbb{E}\left[\sum_{t=1}^{T} \ell_{I_t,t}\right] - \min_{j=1\dots N} \mathbb{E}\left[\sum_{t=1}^{T} \ell_{j,t}\right]$$

same definition as for stochastic bandits, up to the conversion of losses $\ell_{j,t}$ into rewards $M - \ell_{j,t}$ (for a well-chosen bound M)

Why $\mathbb{E}$? Cf. $\ell_{j,t}$ are random variables, as they depend on the past, and in particular on $I_1, \dots, I_{t-1}$

We have $\overline{R}_T \leqslant \mathbb{E}[R_T]$.

We actually rather shoot for high-probability bounds on R_T, but studying $\overline{R}_T$ will be a good warm-up!

In these lecture notes, I'll take $N = K$ as the number of components

↳ we used N for individual sequences

↳ K stochastic bandits

and I alternatively took N and K in the next pgs... (My bad...)

Adversarial bandits: bound on $\overline{R}_T$ via exponential weights.

Key: Estimators of the losses (the unseen and the seen ones):

$$\hat{\ell}_{jt} = \frac{\ell_{I_t,t}}{p_{jt}} \mathbb{1}_{\{I_t = j\}} \quad \text{if } p_{jt} > 0 \text{ (which we will assume)}$$

They are (conditionally) unbiased: denoting by $\mathcal{F}_{t-1} = \sigma(U_1, \ldots, U_{t-1}, \ell_1, \ldots, \ell_{t-1}, p_1, \ldots, p_{t-1}, I_1, \ldots, I_{t-1})$ (auxiliary randomizations of opponent + decision-maker) the total information available at the beginning of round t (of course, the decision-maker does not have that much information!), we have:

- ℓ_t and p_t are $\mathcal{F}_{t-1}$-measurable; the only randomness comes from the random draw of I_t according to p_t thanks to U_t
- $\hat{\ell}_{jt}$ can be rewritten $\hat{\ell}_{jt} = \dfrac{\ell_{jt}}{p_{jt}} \mathbb{1}_{\{I_t = j\}}$

so that

$$\mathbb{E}\left[\hat{\ell}_{jt} \,\middle|\, \mathcal{F}_{t-1}\right] = \frac{\ell_{jt}}{p_{jt}} \underbrace{\mathbb{E}\left[\mathbb{1}_{\{I_t=j\}} \,\middle|\, \mathcal{F}_{t-1}\right]}_{= p_{jt}} = \frac{\ell_{jt}}{p_{jt}} p_{jt} = \ell_{jt}$$

since we assumed $p_{jt} > 0$.

Algorithm: $p_1 = (1/N, \ldots, 1/N)$ and for $t \geq 2$, $p_t = (p_{jt})_{j=1\ldots N}$ is defined as

$$p_{jt} = \exp\left(-\eta_t \sum_{s=1}^{t-1} \hat{\ell}_{js}\right) \Big/ \sum_{k=1}^{N} \exp\left(-\eta_t \sum_{s=1}^{t-1} \hat{\ell}_{ks}\right)$$

for a non-increasing sequence $(\eta_t)_{t \geq 2}$

↳ ensures indeed that $p_{jt} > 0$.

Theorem: The strategy above, tuned with $\eta_t = \dfrac{1}{M}\sqrt{\dfrac{\ln N}{Nt}}$, is such that:

(the range $[0,M]$ is assumed to be known...)

for all opponents picking losses $\ell_{jt} \in [0,M]$,

$$\overline{R}_T = \mathbb{E}\left[\sum_{t=1}^{T} \ell_{I_t,t}\right] - \min_{i=1\ldots N} \mathbb{E}\left[\sum_{t=1}^{T} \ell_{it}\right] \leq 2M\sqrt{TN\ln N}$$

 Lecture 7 — Spring 2021 term Lecture 7 — Spring 2021 term

The proof is based on the following lemma.

! UNBOUNDED and $\geqslant 0$

<u>Lemma</u>: The exponentially weighted average strategy on losses $\tilde{\ell}_{jt} \in [0, +\infty[$, ie,

... which is why we develop a new upper bound

$$\tilde{p}_{jt} = \exp\Big(-\eta_t \sum_{s=1}^{t-1} \tilde{\ell}_{js}\Big) \Big/ \sum_{k=1}^{N} \exp\Big(-\eta_t \sum_{s=1}^{t-1} \tilde{\ell}_{ks}\Big),$$

with $\eta_t \downarrow$,

is such that

$$\sum_{t=1}^{T} \sum_{j=1}^{N} \tilde{p}_{jt}\,\tilde{\ell}_{jt} - \min_{i=1\ldots N} \sum_{t=1}^{T} \tilde{\ell}_{it} \leqslant \frac{\ln N}{\eta_T} + \sum_{t=1}^{T} \frac{\eta_t}{2} \sum_j \tilde{p}_{jt}\,\tilde{\ell}_{jt}^{\,2}.$$

<u>Proof</u>: We saw earlier in this series of lectures that the EWA strategy (with $\eta_t \downarrow$) is such that

$$\forall \tilde{\ell}_{jt} \in \mathbb{R}, \qquad \sum_{t,j} \tilde{p}_{jt}\,\tilde{\ell}_{jt} - \min_{i=1\ldots N} \sum_{t=1}^{T} \tilde{\ell}_{it} \leqslant \frac{\ln N}{\eta_T} + \sum_{t=1}^{T} \tilde{\delta}_t$$

where
$$\tilde{\delta}_t = \sum_{j=1}^{N} \tilde{p}_{jt}\,\tilde{\ell}_{jt} + \frac{1}{\eta_t} \ln \sum_{j=1}^{N} \tilde{p}_{jt}\, e^{-\eta_t \tilde{\ell}_{jt}}$$

We use here $e^{-x} \leqslant 1 - x + \dfrac{x^2}{2} \quad \forall x \geqslant 0$

So that
$$\ln \sum_j \tilde{p}_{jt}\, e^{-\eta_t \tilde{\ell}_{jt}} \leqslant \ln\Big(1 - \eta_t \sum_j \tilde{p}_{jt}\,\tilde{\ell}_{jt} + \frac{\eta_t^2}{2} \sum_j \tilde{p}_{jt}\,\tilde{\ell}_{jt}^{\,2}\Big)$$
$$\underset{\substack{\ln(1+u) \leqslant u \\ \forall u > -1}}{\leqslant} -\eta_t \sum_j \tilde{p}_{jt}\,\tilde{\ell}_{jt} + \frac{\eta_t^2}{2} \sum_j \tilde{p}_{jt}\,\tilde{\ell}_{jt}^{\,2}$$

hence the stated bound.

<u>Proof</u> (of the theorem): We have no control on how large the $\hat{\ell}_{jt}$ can be, and they could be very large! So, we would not be ready to apply any bound with a remainder $M_T \ln N$ term, where M_T is such that $\hat{\ell}_{jt} \in [0, M_T]$ $\forall j, t$ ... as this M_T could be even super-linear. That's why we go back to the beginning of the

proof for the fully adaptive algorithm .... The η_t can be picked as $\ln N / \sum_{s=1}^{t-1} \delta_s$ or in terms of the upper bounds on the δ_s (we choose the latter version for the sake of concreteness). ↳ see below!

The lemma yields for the $\hat{\ell}_{jt}$:

$$\underbrace{\sum_{t,j} p_{jt}\,\hat{\ell}_{jt}} - \min_{i=1,\dots,N} \sum_{t=1}^{T} \hat{\ell}_{it} \leqslant \frac{\ln N}{\eta_T} + \sum_{t=1}^{T} \frac{\eta_t}{2} \underbrace{\sum_j p_{jt}\,\hat{\ell}_{jt}^{\,2}}$$

by definition of the $\hat{\ell}_{jt}$, this

$$= \sum_j \ell_{I_t,t} \frac{p_{jt}}{p_{jt}} \mathbb{1}_{\{I_t=j\}} = \ell_{I_t,t}$$

similar treatment:

$$= \sum_j \ell^2_{I_t,t} \frac{1}{p_{jt}} \mathbb{1}_{\{I_t=j\}}$$

$$\leqslant M^2 \sum_j \mathbb{1}_{\{I_t=j\}} / p_{jt}$$

To simplify even further the choice of the η_t, we first take $\mathbb{E}$ of both sides:

$$\mathbb{E}\Big[\sum_t \ell_{I_t,t}\Big] - \underbrace{\mathbb{E}\Big[\min \sum_t \hat{\ell}_{it}\Big]}_{\leqslant \min \sum_{t=1}^{T} \mathbb{E}[\hat{\ell}_{it}]} \leqslant \frac{\ln N}{\eta_T} + \frac{M^2}{2} \sum_{t=1}^{T} \eta_t \sum_{j=1}^{N} \underbrace{\mathbb{E}\Big[\frac{\mathbb{1}_{\{I_t=j\}}}{p_{jt}}\Big]}_{=1}$$

$$\underbrace{\mathbb{E}[\hat{\ell}_{it}]}_{=\mathbb{E}[\ell_{it}]}$$ by the tower rule and the fact that $\hat{\ell}_{it}$ is conditionally unbiased.

Thus,

$$\overline{R}_T = \mathbb{E}\Big[\sum_{t=1}^{T} \ell_{I_t,t}\Big] - \min_{i=1,\dots,N} \mathbb{E}\Big[\sum_{t=1}^{T} \ell_{it}\Big] \leqslant \underbrace{\frac{\ln N}{\eta_T} + \frac{M^2 N}{2} \sum_{t=1}^{T} \eta_t}$$

the only adaptation to be made is wrt T (as M is assumed to be known)

The optimal constant η would be s.t. $\ln N / \eta = \frac{M^2 N}{2} T \eta$, that is,

η proportional to $\dfrac{1}{M}\sqrt{\dfrac{\ln N}{NT}}$

↳ try $\eta_t = \dfrac{\gamma}{M}\sqrt{\dfrac{\ln N}{Nt}}$ where γ is to be optimized.

The final bound is $M\sqrt{N \ln N}\left(\frac{\sqrt{T}}{\gamma} + \frac{\gamma}{2}\underbrace{\sum_{t=1}^{T}\frac{1}{\sqrt{t}}}_{\leq \int_0^T \frac{1}{\sqrt{t}}\,dt \,\leq\, 2\sqrt{T}}\right)$

$$\leq \sqrt{T}\left(\frac{1}{\gamma}+\gamma\right) = 2\sqrt{T} \quad \text{for } \gamma = 1$$

Remarks / insights

* We heavily used above that the range $[0,M]$ is known

- 0 to apply safely $e^{-x} \leq 1 - x + x^2/2$
- M to bound $\ell_{I_t t}^2 \leq M^2$
- 0 and M known because the η_t are set based on the bounds obtained based on the previous two inequalities

* When the range $[m,M]$ is known, with $m \in \mathbb{R}$

↳ Translate all losses by $-m$, eg, consider

$$\hat{\ell}_{jt} = \frac{\ell_{I_t t} - m}{p_{jt}} \mathbb{1}_{\{I_t = j\}} \qquad \text{and} \qquad \eta_t = \frac{1}{M-m}\sqrt{\frac{\ln N}{Nt}}$$

to get $\qquad \overline{R}_T \leq 2(M-m)\sqrt{TN \ln N}$

* What about an unknown range $[m,M]$? ⇝ See the homework #2!

Additional elements ⇝ See the extra lecture notes posted on the website

* With exponential weights, one can get high-probability bounds on the true regret, of the same order of magnitude:

$$\text{wp } 1-\delta, \qquad R_T \leq \square\,(M-m)\sqrt{TN \ln(N/\delta)}$$

* There exists an algorithm (called INF) s.t. R_T is controlled (in $\mathbb{E}$ or wp $1-\delta$) by $O(\sqrt{TN})$, ie, no $\sqrt{\ln N}$ term, against oblivious individual seq. ℓ_{jt}