Gilles Stoltz

Researcher, CNRS
Affiliate professor, HEC Paris

Paris-Saclay master in mathematics (tracks: Optimization / Data science / ALEA)

Sequential learning, sequential optimization

Final exam – Wednesday April 13, 2016

EXERCISE #1 (α,ψ) - UCB

Consider $\left\{ \begin{array}{l} \text{a (convex) function } \psi \\ \text{a bandit model } \mathcal{D} \end{array} \right\}$ such that all distributions $\nu_j \in \mathcal{D}$ satisfy

$$\forall \lambda \geq 0, \qquad \max\left\{ \ln \mathbb{E}\left[e^{\lambda(X-\mathbb{E}X)}\right], \ln \mathbb{E}\left[e^{\lambda(\mathbb{E}X - X)}\right] \right\} \leq \psi(\lambda)$$

where $X \sim \nu_j$.

For example, Hoeffding's lemma shows that for $\mathcal{D} = \mathcal{P}([0,1])$, $\psi(\lambda) = \lambda^2/8$ is a suitable choice.

For all $x \geq 0$, we define $\psi^*(x) = \sup_{\lambda \geq 0}\{\lambda x - \psi(\lambda)\}$ and assume that ψ^* is invertible, with inverse denoted by $(\psi^*)^{-1}$.

We generalize UCB in the following way:

Algorithm (α,ψ) - UCB : for a bandit problem with K arms

Parameters $\alpha > 0$ and $\psi : [0,+\infty) \longrightarrow \mathbb{R}$

For $t = 1, 2, \dots K$: Pull each arm once, $I_t = j$, and get a reward $Y_t \sim \nu_{I_t}$

For $t \geq K+1$: – Pull $I_t \in \operatorname*{argmax}_{j \in \{1,\dots K\}} \left\{ \hat{\mu}_{j,t-1} + (\psi^*)^{-1}\left(\frac{\alpha \ln t}{N_j(t-1)}\right) \right\}$ (ties broken arbitrarily)

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

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

– Get a reward Y_t drawn at random according to ν_{I_t} (conditionally to I_t)

We want to bound the pseudo-regret of (α,ψ)-UCB:

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

where $\mu_j = E(\nu_j)$ is the expectation of $\nu_j \in \mathcal{D}$

$$\mu^* = \max_{j=1,\dots K} \mu_j$$

(1) We first establish that

$$\hat{\mu}_{j,t-1} + (\psi^*)^{-1}\left(\frac{\alpha \ln t}{N_j(t-1)}\right)$$

is an upper confidence bound on μ_j (for all j).

(1.1) Show that for all $\lambda \in \mathbb{R}$,

$$M_t(\lambda) = \exp\Big(\lambda \sum_{s=1}^{t} (Y_s - \mu_j)\, \mathbb{1}_{\{I_s = j\}} - \psi(\lambda)\, N_j(t)\Big)$$

is a supermartingale, ie, that

$$\mathbb{E}\big[M_t(\lambda) \mid \mathcal{F}_{t-1}\big] \leqslant M_{t-1}(\lambda)$$

for a filtration $(\mathcal{F}_s)$ to indicate.

Hint: $$e^{\lambda(Y_t - \mu_j)\mathbb{1}_{\{I_t=j\}}} = \mathbb{1}_{\{I_t \neq j\}} + \mathbb{1}_{\{I_t = j\}}\, e^{\lambda(Y_t - \mu_j)}$$

(1.2) Prove that $\forall \varepsilon > 0$, $\forall \ell \geqslant 1$, $\forall t \geqslant K+1$,

$$\mathbb{P}\big\{\hat{\mu}_{j,t-1} + \varepsilon \leqslant \mu_j \text{ and } N_j(t-1) = \ell\big\} \leqslant e^{-\ell\,\psi^*(\varepsilon)}$$

(1.3) Deduce that $$\mathbb{P}\left\{\hat{\mu}_{j,t-1} + (\psi^*)^{-1}\left(\frac{\alpha \ln t}{N_j(t-1)}\right) \leqslant \mu_j\right\} \leqslant \frac{1}{t^{\alpha-1}}.$$

(2) We now establish the regret bound. Fix a suboptimal arm j.

(2.1) Explain why $I_t = j$ for $t \geq K+1$ entails one of the following, where a^* is a fixed optimal arm:

$$\hat{\mu}_{a^*,t-1} + (\psi^*)^{-1}\left(\frac{\alpha \ln t}{N_{a^*}(t-1)}\right) \leqslant \mu^*$$

$$\hat{\mu}_{j,t-1} - (\psi^*)^{-1}\left(\frac{\alpha \ln t}{N_j(t-1)}\right) > \mu_j$$

$$N_j(t-1) < \frac{\alpha \ln T}{\psi^*(\Delta_j/2)} \qquad \text{where } \Delta_j = \mu^* - \mu_j$$

(2.2) Establish a regret bound of the form:

$$\overline{R}_T \leqslant \sum_{j:\Delta_j>0} \left(\frac{\alpha \Delta_j}{\psi^*(\Delta_j/2)} \ln T + \frac{\alpha}{\alpha-2}\right).$$

(3) Can this bound be related to the one we proved for UCB on $\mathcal{D} = \mathcal{P}([0,1])$? The latter was equal to

$$\sum_{j:\Delta_j>0} \left(\frac{8}{\Delta_j} \ln T + 2\right).$$

Exercise #2. Budgeted prediction.

With fixed and known: horizon T & loss range $[0,1]$
budget m

For each round $t = 1, 2, \ldots T$:

- the decision-maker and the opponent player simultaneously pick $I_t \in \{1, \ldots N\}$ (possibly at random according to some $p_t \in \mathcal{P}\{1, \ldots N\}$) and $\underline{\ell}_t = (\ell_{1t}, \ldots \ell_{Nt}) \in [0,1]^N$
- the opponent player observes I_t and p_t
- the decision-maker decides whether or not she wants to observe the ℓ_{jt}; she can do so only if she hasn't requested to perform such an observation more than $m-1$ times in total so far.

Thus the decision-maker can observe at most m vectors among the $\underline{\ell}_1, \ldots \underline{\ell}_T$; she decides which she wants to observe.

Strategy / Let's construct it step by step. We fix $\delta \in (0,1)$.

The decision-maker uses a sequence $Z_1 \ldots Z_T$ of iid random variables distributed according to a Bernoulli distribution with parameter $\varepsilon > 0$ to decide whether she request the observation at round t ($Z_t = 1$) or not ($Z_t = 0$).

(1) To respect the budget constraint we want to pick ε s.t.

$$\mathbb{P}\{ Z_1 + \cdots Z_T \leq m \} \geq 1 - \delta.$$

Show that $\varepsilon_1 = \frac{m}{T} - \frac{1}{T}\sqrt{\frac{m}{\delta}}$ is a suitable choice when $\delta \geq 1/m$.

(Resort to Chebychev's inequality: $\forall \varepsilon > 0,\ \mathbb{P}\{X - \mathbb{E}X > \varepsilon\} \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}$.)

(2) We define $\hat{\ell}_{jt} = \frac{\ell_{jt}}{\varepsilon} Z_t$ as our estimator for ℓ_{jt}.

Show that for a well-chosen filtration $(\mathcal{F}_s)_{s \geq 0}$ we have:

$$\forall j, \forall t, \qquad \mathbb{E}\left[\hat{\ell}_{jt} \mid \mathcal{F}_{t-1}\right] = \ell_{jt}.$$

(3) A lemma: shows that for all $\eta > 0$ and all non-negative numbers $u_{jt} \geq 0$, $j \in \{1, \dots N\}$ and $t \in \{1, \dots T\}$, we have

$$\sum_{t=1}^{T} \sum_{i=1}^{N} \underbrace{\frac{e^{-\eta \sum_{s=1}^{t-1} u_{is}}}{\sum_{k=1,\dots N} e^{-\eta \sum_{s=1}^{t-1} u_{ks}}}}_{\text{denoted by } q_{it}} u_{it} - \min_{k=1..N} \sum_{t=1}^{T} u_{kt} \leq \frac{\ln N}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{i=1}^{N} q_{it}\, u_{it}^2$$

<u>Hint</u>: use $e^{-x} \leq 1 - x + \frac{x^2}{2}$.

(4) We consider the probability distributions

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

and $I_t \sim (\tilde{p}_{1t}, \dots \tilde{p}_{Nt})$.

Show that for this strategy the pseudo-regret is controlled as

$$\mathbb{E}\left[\sum_{t=1}^{T} \ell_{I_t,t}\right] - \min_{j=1..N} \mathbb{E}\left[\sum_{t=1}^{T} \ell_{jt}\right] \leq \frac{\ln N}{\eta} + \frac{\eta T}{2\varepsilon}.$$

(5) Conclusion: Construct a strategy that **NEVER** asks more than m observations and whose pseudo-regret is controlled as

$$\mathbb{E}\left[\sum_{t=1}^{T} \ell_{I_t,t}\right] - \min_{j=1..N} \mathbb{E}\left[\sum_{t=1}^{T} \ell_{j,t}\right] \leq \frac{2T}{\sqrt{m}}\sqrt{\ln N} + \frac{4T}{m}.$$

(Note that with $m = T$, we recover the classical $\sqrt{T}$ bound.)

Two <u>much more</u> difficult questions: [! please do exercise #3 <u>first</u>]

(6) Prove that for any forecaster

$$\sup_{\ell_{j,t}\in[0,1]} \left\{ \mathbb{E}\left[\sum_{t=1}^{T} \ell_{I_t,t}\right] - \min_{j} \sum_{t=1}^{T} \ell_{j,t} \right\} \geq c\,\frac{T}{\sqrt{m}}$$

for some numerical constant $c > 0$.

(7) Prove a high-probability bound on the regret

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

EXERCISE #3 — Approachability of a convex set.

A loss function $\ell : \{1, \dots N\} \times \{1, \dots M\} \to \mathbb{R}^d$ is given and known to all players.

A closed convex set $\mathcal{C} \subset \mathbb{R}^d$ is fixed.

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

- the decision-maker and the opponent simultaneously pick $I_t \in \{1, \dots N\}$ and $J_t \in \{1, \dots M\}$ possibly at random, according to distributions denoted by $p_t \in \mathcal{P}\{1, \dots N\}$ and $q_t \in \mathcal{P}\{1, \dots M\}$
- the decision-maker suffers a loss $\ell(I_t, J_t)$
- both players observe I_t and J_t

Aims: The decision-maker wants to ensure that

$$\frac{1}{T} \sum_{t=1}^{T} \ell(I_t, J_t) \longrightarrow \mathcal{C} \quad \text{a.s.}$$

that is,

$$\inf_{c \in \mathcal{C}} \left\| c - \frac{1}{T} \sum_{t=1}^{T} \ell(I_t, J_t) \right\|_2 \longrightarrow 0 \quad \text{a.s}$$

The opponent player wants to prevent this convergence.

Blackwell's condition: $\forall q \in \mathcal{P}\{1, \dots M\} \quad \exists p \in \mathcal{P}\{1, \dots N\} \;|\; \ell(p, q) \in \mathcal{C}$

where we use:

Short-hand notation:

$$\begin{cases} \ell(p, j) = \sum_i p_i \, \ell(i, j) \\ \ell(p, q) = \sum_i \sum_j p_i \, q_j \, \ell(i, j) \end{cases}$$

(1) Show that if Blackwell's condition does not hold, then the opponent player has a strategy such that $\exists \gamma > 0$ |

for all strategies of the decision-maker,

a.s., $\liminf_{T \to +\infty} \inf_{c \in \mathcal{C}} \left\| c - \frac{1}{T} \sum_{t=1}^{T} \ell(I_t, J_t) \right\|_2 \geqslant \gamma$

Hints: Show that $\exists q_0 \in \mathcal{P}\{1, \dots M\}$ | $\inf_{p \in \mathcal{P}\{1, \dots N\}} \inf_{c \in \mathcal{C}} \| c - \ell(p, q_0) \|_2 > 0$

Explain also why $\left\| \frac{1}{T} \sum_{t=1}^{T} \ell(I_t, J_t) - \frac{1}{T} \sum_{t=1}^{T} \ell(p_t, q_t) \right\|_2 \longrightarrow 0$ a.s.

(↳ A list of useful results is in appendix.)

(2) We now assume that Blackwell's condition holds and design a strategy for the decision-maker:

- Play $p_1 = (1/N, \dots 1/N)$
- For $t = 2, 3, \dots$
  - Compute $\overline{m}_{t-1} = \frac{1}{t-1} \sum_{s=1}^{t-1} \ell(p_s, J_s)$
  - Project onto $\mathcal{C}$: $\overline{c}_{t-1} = \Pi_{\mathcal{C}}(\overline{m}_{t-1})$
  - Pick $p_t \in \operatorname{argmin}_{p \in \mathcal{P}\{1..N\}} \max_{q \in \mathcal{P}\{1..M\}} \langle \overline{m}_{t-1} - \overline{c}_{t-1}, \ell(p, q) \rangle$

where $\langle u, v \rangle = \sum_i u_i v_i$ denotes the inner product of $\mathbb{R}^d$

(2.1) Recall on a picture why

$\forall c \in \mathcal{C} \quad \langle \overline{m}_{t-1} - \overline{c}_{t-1}, c - \overline{c}_{t-1} \rangle \leqslant 0$

(2.2) Prove that $\forall q \in \mathcal{P}\{1, \dots M\}$, $\langle \overline{m}_{t-1} - \overline{c}_{t-1}, \ell(p_t, q) - \overline{c}_{t-1} \rangle \leqslant 0$

Hint: A consequence of Sion's lemma is that

$$\max_q \min_p \langle \overline{m}_{t-1} - \overline{c}_{t-1}, \ \ell(p,q) - \overline{c}_{t-1} \rangle = \min_p \max_q \langle \overline{m}_{t-1} - \overline{c}_{t-1}, \ell(p,q) \rangle$$

(2.3) Denote by $d_t = \inf_{c \in \mathcal{C}} \| c - \overline{m}_t \|_2$

and by $L = \max_{i,j} |\ell(i,j)|$.

Show that $d_{t+1}^2 \leqslant \left(1 - \frac{2}{t+1}\right) d_t^2 + \frac{4M^2}{(t+1)^2}$

(2.4) Prove $\forall t \geqslant 1$, $\inf_{c \in \mathcal{C}} \left\| c - \frac{1}{t} \sum_{s \leqslant t} \ell(p_s, J_s) \right\|_2 \leqslant \frac{2M}{\sqrt{t}}$.

(2.5) Conclude to the desired convergence.

Note: see a list of useful results in appendix.

Useful results.

Hoeffding–Azuma inequality: Let $(\mathcal{F}_t)_{t \geq 0}$ be a filtration and let $(X_t)_{t \geq 1}$ be a sequence of adapted random variables such that

$\forall t \geq 1$, $\exists Z_{t-1}$ $\mathcal{F}_{t-1}$-measurable, $\exists (a_t, b_t) \in \mathbb{R}^2$ $\Big|$ $F_{t-1} + a_t \leqslant X_t \leqslant F_{t-1} + b_t$ a.s.

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

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

Sion's lemma (as seen in the exercices): Let $\mathcal{X}, \mathcal{Y}$ be two convex sets with $\mathcal{X}$ metric and compact, and let $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$ be

- bounded,
- uniformly continuous,
- concave in the second argument,
- convex in the first argument,

then

$$\inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} f(x,y) = \sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} f(x,y).$$