RISK-AWARE LINEAR BANDITS WITH CONVEX LOSS

Patrick Saux[1], Odalric-Ambrym Maillard [1]

[1]Inria, Univ. Lille, CNRS, UMR 918f9 - CRIStAL

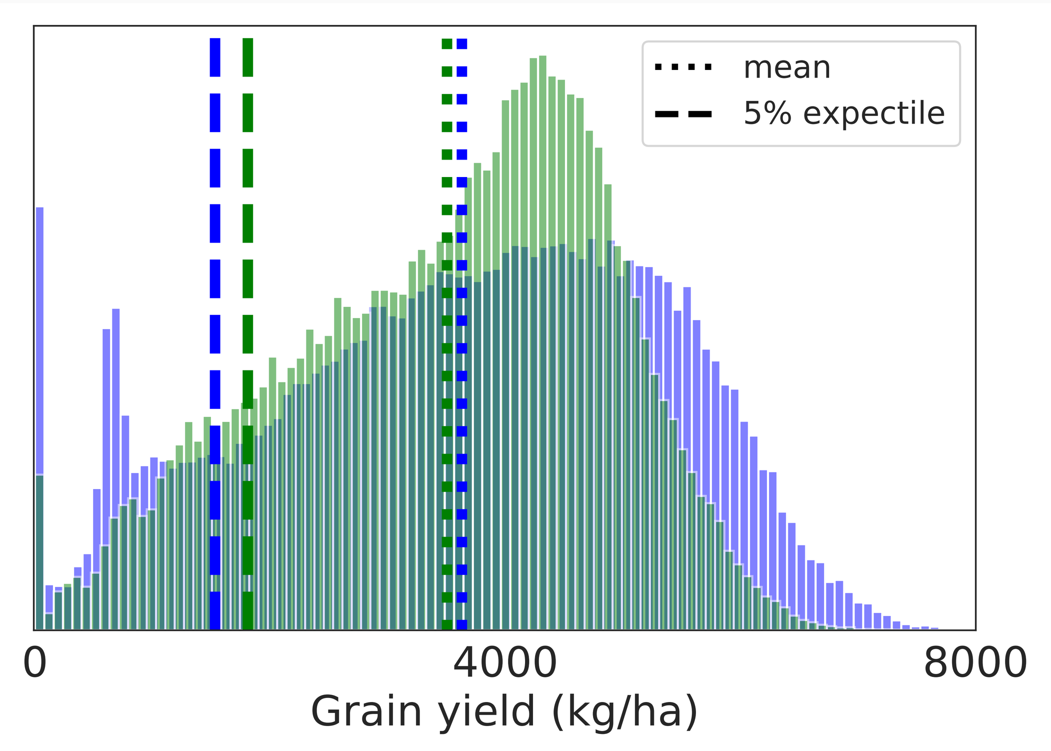
Setting

At time t:

- Observe action set $\mathcal{X}_t \subset \mathbb{R}^d$ and select action X_t,
- Receive reward $Y_t \sim \Phi(X_t)$ where $\Phi : \mathbb{R}^d \to \mathcal{P}(\mathbb{R})$,
- Linear model: $\Phi = \underbrace{\varphi}_{\text{known}} \circ \underbrace{\langle \theta^*, \cdot \rangle}_{\text{unknown, linear}}$,
- **Goal**: minimize regret $\mathcal{R}_T = \sum_{t=1}^T \max_{x \in \mathcal{X}_t} \rho(\varphi \circ \langle \theta^*, x \rangle) - \rho(\varphi \circ \langle \theta^*, X_t \rangle)$, where ρ is a certain **risk measure**.

$\hookrightarrow \neq$ existing settings: $\mathbb{E}[Y_t \mid X_t] = \mu(\langle \theta^*, X_t \rangle)$ (generalized mean-linear).

Example: risk-aversion in agriculture



Elicitable risk measures

Convex loss: $\mathcal{L} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$.

Definitions

- Risk measure elicited by $\mathcal{L}$:

$$\rho_{\mathcal{L}} : \nu \in \mathcal{P}(\mathbb{R}) \mapsto \min_{\xi \in \mathbb{R}} \mathbb{E}_{Y \sim \nu}[\mathcal{L}(Y, \xi)] \,.$$

- Adapted loss to a linear bandit (φ, θ^*):

$$\rho_{\mathcal{L}}\left(\varphi \circ \langle \theta^*, X_t \rangle\right) = \langle \theta^*, X_t \rangle \,.$$

Examples of elicitable risk measures

Name	$\rho_{\mathcal{L}}(\nu)$	Associated loss $\mathcal{L}(y,\xi)$
Mean	$\mathbb{E}_{Y \sim \nu}[Y]$	$\frac{1}{2}(y-\xi)^2$
p-expectile	$\operatorname{argmin}_{\xi \in \mathbb{R}} \mathbb{E}_{Y \sim \nu}[\psi(Y-\xi)]$ $\psi(z) = \lvert p - \mathbb{1}_{z<0} \rvert z^2$	$\psi(y-\xi)$
Entropic risk $\gamma \neq 0$	$\frac{1}{\gamma} \log \mathbb{E}_{Y \sim \nu}[e^{\gamma Y}]$	$\xi + \frac{1}{\gamma}(e^{\gamma(y-\xi)} - 1)$

Remark: variance and CVaR are *not* (first-order) elicitable.

LinUCB with convex loss

Input: regularisation parameter α, projection Π, exploration bonus sequence $(\gamma_t)_{t \in \mathbb{N}}$.
Initialization: Observe $\mathcal{X}_1$.
for $t = 1, \dots, T$ **do**
  $\widehat{\theta}_t \in \operatorname{argmin}_{\mathbb{R}^d} \sum_{s=1}^{t-1} \mathcal{L}(Y_s, \langle \theta, X_s \rangle) + \frac{\alpha}{2} \|\theta\|_2^2$; ▷ `Empirical risk minimization`
  $\bar{\theta}_t = \Pi(\widehat{\theta}_t)$; ▷ `Projection`
  $X_t = \operatorname{argmax}_{x \in \mathcal{X}_t} \langle \bar{\theta}_t, x \rangle + \gamma_t(x)$; ▷ `Play arm`
  Observe Y_t and $\mathcal{X}_{t+1}$.

Numerical computation of $\widehat{\theta}_t$ at each step!
$\neq$ mean-linear case: $\widehat{\theta}_t = \left(\sum_{s=1}^{t-1} X_s X_s^\top + \alpha I_d\right)^{-1} \sum_{s=1}^{t-1} Y_s X_s$.

Analysis

Notations and assumptions

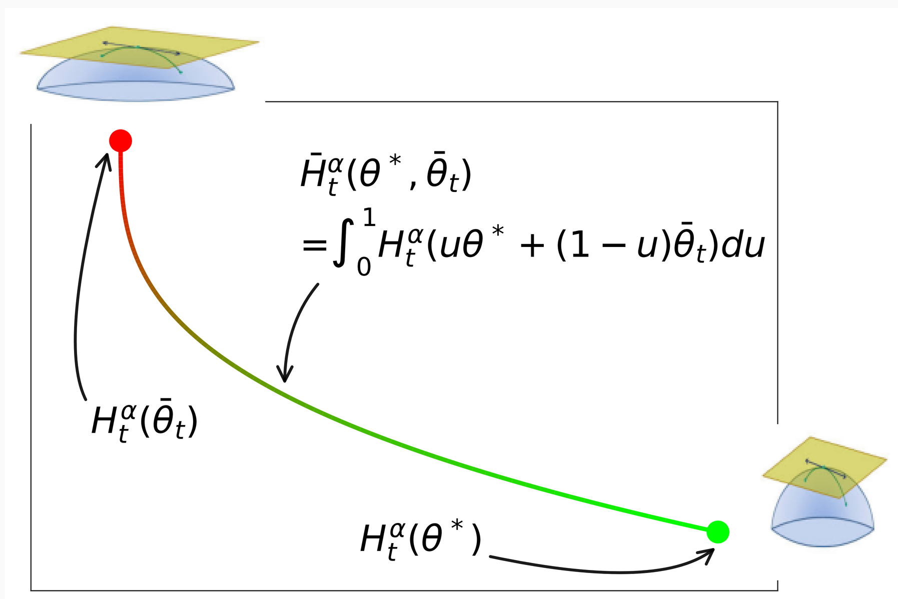
- $\partial \mathcal{L}(y, \xi) = \frac{\partial \mathcal{L}}{\partial \xi}(y, \xi)$,
- $V_t^\alpha = \sum_{s=1}^{t-1} X_s X_s^\top + \alpha I_d$,
- $H_t^\alpha(\theta) = \sum_{s=1}^{t-1} \partial^2 \mathcal{L}(Y_s, \langle \theta, X_s \rangle) X_s X_s^\top + \alpha I_d$,
- $\theta^* \in \Theta \subseteq \mathcal{B}_{\|\cdot\|_2}(0, S)$ convex and $\forall t \in \mathbb{N}$, $\mathcal{X}_t \subseteq \mathcal{B}_{\|\cdot\|_2}(0, L)$.

Martingale lemma
With respect to the natural bandit filtration,

- $S_t = \sum_{s=1}^{t-1} \partial \mathcal{L}(Y_s, \langle \theta^*, X_s \rangle) X_s$ defines a martingale.
- $M_t^\lambda = \exp\left(\lambda^\top S_t - \frac{\sigma^2}{2} \|\lambda\|^2_{H_t^\alpha(\theta^*)}\right)$ defines a supermartingale for each $\lambda \in \mathbb{R}^d$ (under mild assumptions).

Useful for time-uniform concentration of $\bar{\theta}_t$ around θ^*!

Geometric sufficient condition for optimism
Parameter space Θ is a Hessian manifold equipped with the metric $g_\theta = H_t^\alpha(\theta)$.
Local metric (depends on θ, except if ρ =mean).



Linear optimism works if $\exists \kappa, \beta > 0$ s.t.

$$\kappa \tilde{H}_t^\alpha(\theta^*, \bar{\theta}_t) \succcurlyeq H_t^\beta(\theta^*) \,,$$
$$\kappa \tilde{H}_t^\alpha(\theta^*, \bar{\theta}_t) \succcurlyeq H_t^\beta(\bar{\theta}_t) \,.$$

This is satisfied with $\kappa = \frac{M}{m}$ and $\beta = \kappa \alpha$ if $\forall y, \xi \in \mathbb{R}$, $m \leqslant \partial^2 \mathcal{L}(y, \xi) \leqslant M$.

Regret of LinUCB with convex loss

With probability at least $1-\delta$, $\mathcal{R}_T^{\text{LinUCB}} = \mathcal{O}\left(\frac{\kappa \sigma d}{\sqrt{m}} \sqrt{T} \log \frac{TL^2}{d}\right)$.

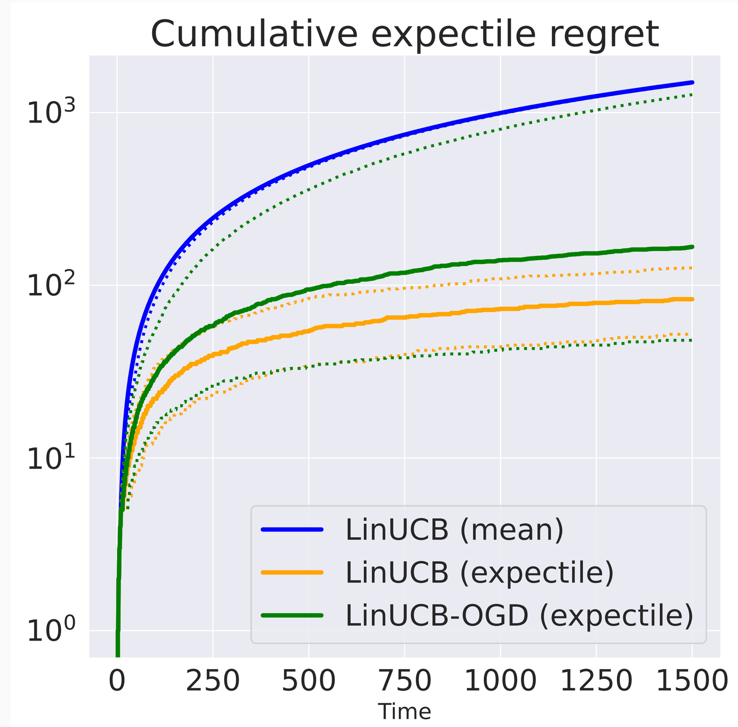
A faster approximate algorithm: LinUCB-OGD

Input: horizon T, regularisation parameter α, projection Π, exploration bonus sequence $(\gamma_{t,T}^{\text{OGD}})_{t \leqslant T}$, gradient descent step sequence $(\varepsilon_t)_{t \in \mathbb{N}}$, episode length $h > 0$.
Initialization: Observe $\mathcal{X}_1$, set $\widehat{\theta}_0^{\text{OGD}}$, $t = 1$, $n = 1$.
for $t = 1, \dots, T$ **do**
  if $t = nh + 1$ **then**
    $\widehat{\theta}_n^{\text{OGD}} = \widehat{\theta}_{n-1}^{\text{OGD}} - \varepsilon_{n-1}\left(\sum_{k=1}^h \partial \mathcal{L}(Y_{(n-1)h+k}, \langle \widehat{\theta}_{n-1}^{\text{OGD}}, X_{(n-1)h+k} \rangle) + \alpha \widehat{\theta}_{n-1}^{\text{OGD}}\right)$; ▷ `OGD`
    $\bar{\theta}_n^{\text{OGD}} = \frac{1}{n} \sum_{j=1}^n \Pi(\widehat{\theta}_j^{\text{OGD}})$; ▷ `Average over previous OGD steps`
    $n \leftarrow n + 1$
  $X_t = \operatorname{argmax}_{x \in \mathcal{X}_t} \langle \bar{\theta}_n^{\text{OGD}}, x \rangle + \gamma_{t,T}^{\text{OGD}}(x)$; ▷ `Play with same` $\bar{\theta}_n^{\text{OGD}}$ `for` h `steps`
  Observe Y_t and $\mathcal{X}_{t+1}$,
  $t \leftarrow t + 1$.

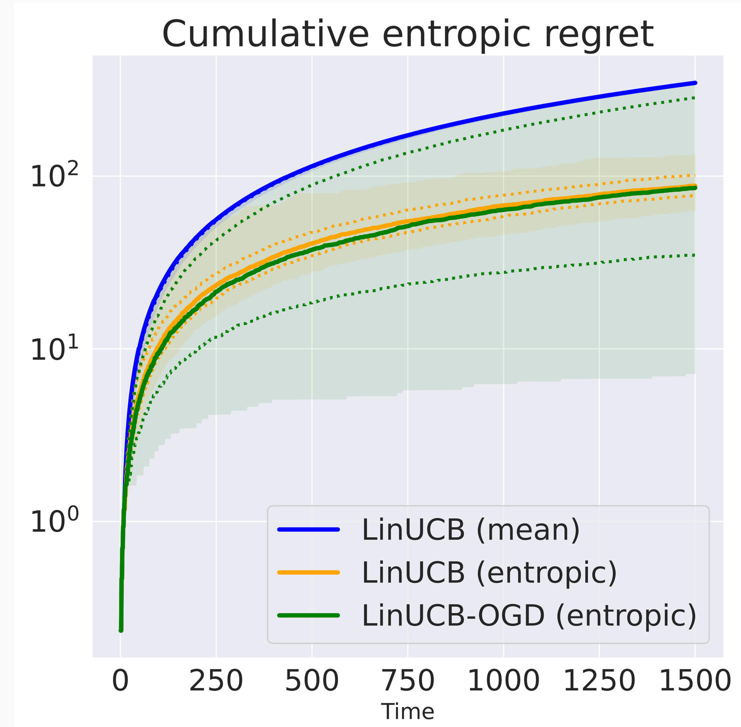
Regret of LinUCB-OGD with convex loss

With probability at least $1-\delta$, $\mathcal{R}_T^{\text{LinUCB-OGD}} = \mathcal{O}\left(\sqrt{T} \times \text{Polylog}(T)\right)$.

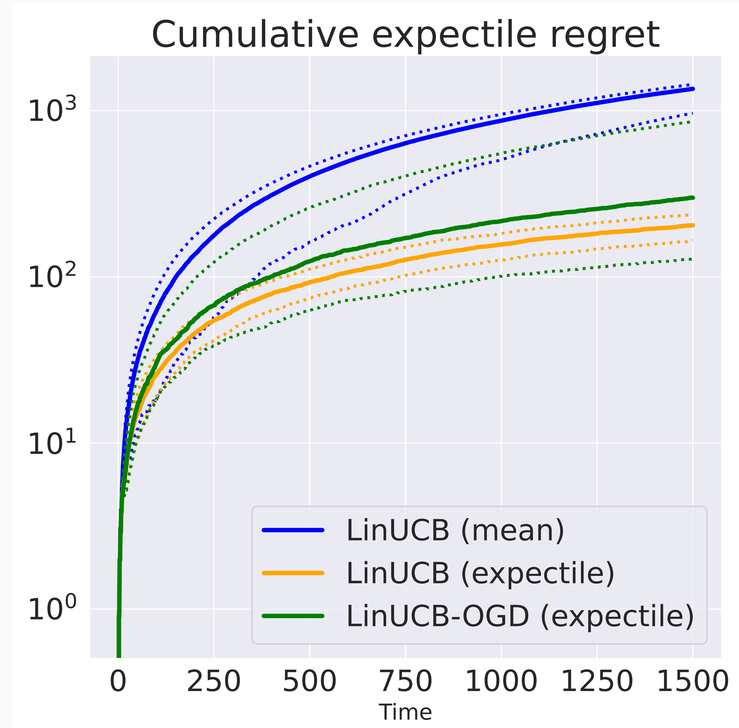
Numerical experiments



Gaussian expectile bandit.



Bernoulli entropic risk bandit.



Linear expectile bandit with expectile-based asymmetric noises.

Aknowledgments: The authors acknowledge the funding of the French National Research Agency, the French Ministry of Higher Education and Research, Inria, the MEL and the I-Site ULNE regarding project R-PILOTE-19-004-APPRENF and Bandits For Health (B4H).