RISK-AWARE LINEAR BANDITS WITH CONVEX LOSS

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

[1]Univ. Lille, Inria, CNRS, UMR 9198 - CRIStAL

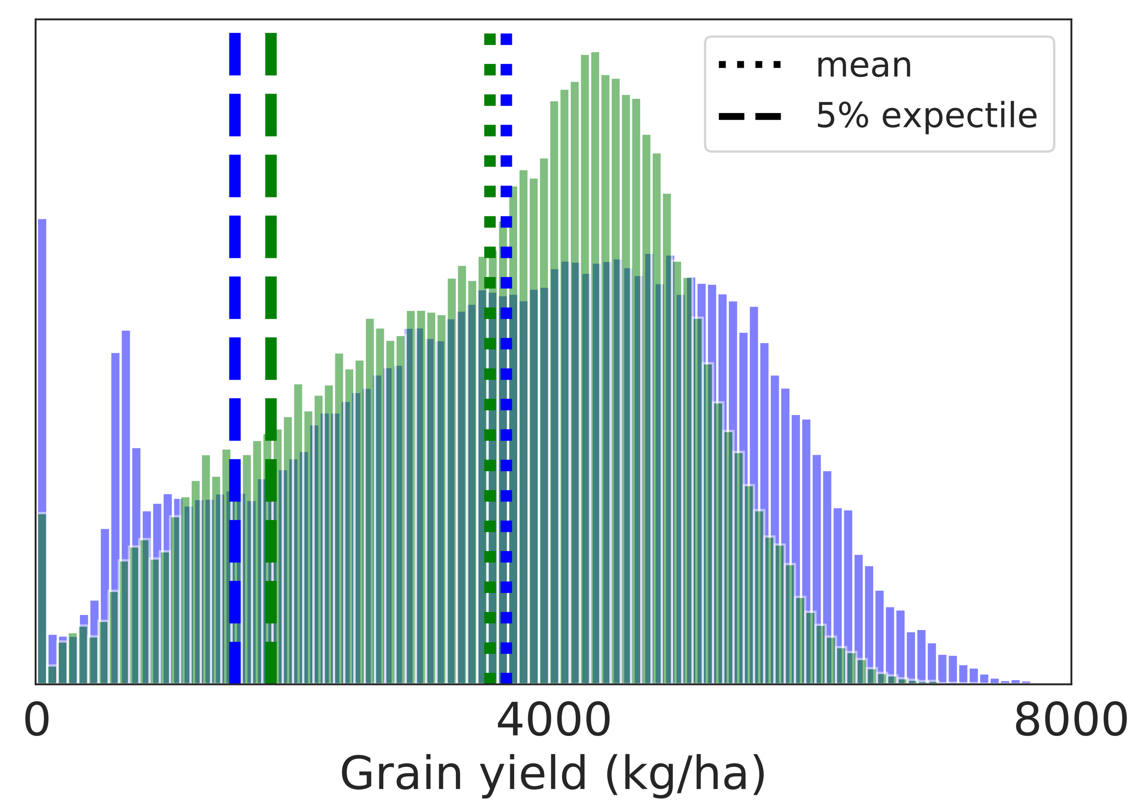
Setting

Linear bandits

- Play action X_t from a decision set $\mathcal{X}_t \subset \mathbb{R}^d$.
- Receive reward $Y_t \sim p_{\langle\theta^*, X_t\rangle}$, where $\{p_\varphi\}$ is a statistical model.
- **Goal**: minimise regret $\mathcal{R}_T = \sum_{t=1}^T \max_{x\in\mathcal{X}_t} \rho\left(p_{\langle\theta^*,x\rangle}\right) - \rho\left(p_{\langle\theta^*,X_t\rangle}\right)$, where ρ is a certain **risk measure**.

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

Example: risk-aversion in agriculture



Elicitable risk measures

Definitions

- Risk measure elicited by a convex loss $\mathcal{L}: \mathbb{R}\times\mathbb{R} \to \mathbb{R}_+$:

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

- Adapted loss to the linear bandit if $\rho_{\mathcal{L}}$ **is linear** on the statistical model $\{p_\varphi\}$:

$$\rho_{\mathcal{L}}(p_\varphi) = \varphi\,.$$

Examples of elicitable risk measures

Name	$\rho_{\mathcal{L}}$	$\mathcal{L}(y,\xi)$	Example of adapted statistical model
Mean	$\mathbb{E}[Y]$	$\frac{1}{2}(y-\xi)^2$	$p_\varphi(y) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(y-\varphi)^2}{2}\right)$
p-expectile	$\operatorname*{argmin}_{\xi\in\mathbb{R}} \mathbb{E}[\psi_p(Y-\xi)]$ $\psi_p(z) = \lvert p - \mathbb{1}_{z<0}\rvert z^2$	$\psi_p(y-\xi)$	$p_\varphi(y) = \frac{\sqrt{2p(1-p)}}{\sqrt{\pi}\sqrt{p}+\sqrt{1-p}}\exp\left(-\frac{\psi_p(y-\varphi)}{2}\right)$

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

LinUCB-CR (Convex Risk)

Input: regularisation parameter α, projection operator Π, sequence of exploration bonus functions $(\gamma_t)_{t\in\mathbb{N}}$.

for $t = 1,\dots,T$ **do**

$\quad \widehat{\theta}_t \in \operatorname*{argmin}_{\theta\in\mathbb{R}^d} \sum_{s=1}^{t-1} \mathcal{L}(Y_s, \langle\theta, X_s\rangle) + \frac{\alpha}{2}\|\theta\|_2^2$; ▷ `ERM`

$\quad \bar{\theta}_t = \Pi(\widehat{\theta}_t)$; ▷ `Projection`

$\quad X_t = \operatorname*{argmax}_{x\in\mathcal{X}_t} \langle\bar{\theta}_t, x\rangle + \gamma_t(x)$; ▷ `Play arm`

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

Bounded loss curvature: $\forall y,\xi\in\mathbb{R},\ 0 < m \leqslant \frac{\partial^2\mathcal{L}}{\partial\xi^2}(y,\xi) \leqslant M$.

Supermartingale lemma

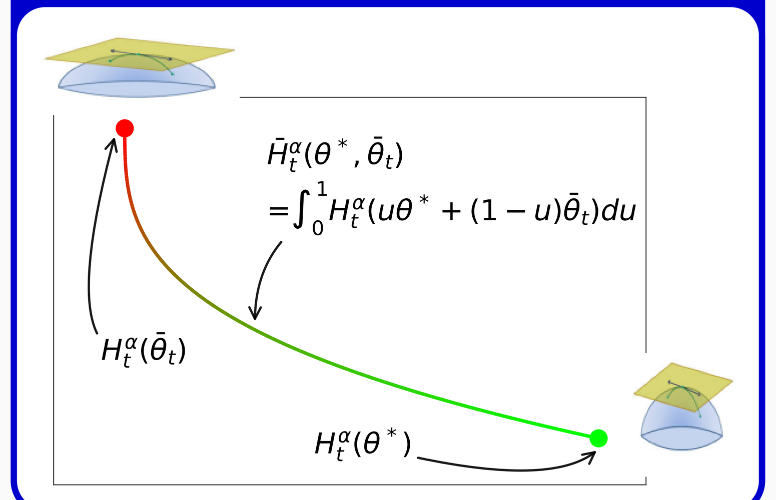
Self-normalised concentration bound with the **local** Hessian

$$H_t^\alpha(\theta) = \sum_{s=1}^{t-1} \partial^2\mathcal{L}(Y_s, \langle\theta, X_s\rangle) X_s X_s^\top + \alpha I_d$$

instead of the global Hessian

$$V_t^\alpha = \sum_{s=1}^{t-1} X_s X_s^\top + \alpha I_d\,.$$

Transportation of metrics



Regret of LinUCB-CR

With probability at least $1-\delta$,

$$\mathcal{R}_T^{\text{LinUCB-CR}} = \mathcal{O}\left(\frac{\sigma\kappa d}{\sqrt{m}}\sqrt{T}\log\frac{TL^2}{d}\right).$$

- σ: $\approx$ variance of $\partial\mathcal{L}(Y_t, \langle\theta^*, X_t\rangle)$
- $\kappa = \frac{M}{m}$†
- d: dimension of actions
- L: upper bound on $\|X_t\|_2$
- m: lower bound on $\partial^2\mathcal{L}$

† conjecture: $\kappa \approx$ constant in certain cases.

$\mathcal{R}_T^{\text{LinUCB-CR}} = \mathcal{O}\left(\frac{\sigma\kappa d}{\sqrt{m}}\sqrt{T\log\frac{TL^2}{d}}\right)$ under stochastic arrival of action sets.

⟶ Take-home message ⟵

The analysis of LinUCB can be lifted from mean-linear to elicitable convex loss with essentially the same regret bound.

A faster approximate algorithm: LinUCB-OGD-CR

Input: T, α, Π, $(\gamma_{t,T}^{\text{OGD}})_{t\leqslant T}$, OGD steps $(\varepsilon_t)_{t\leqslant T}$, episode length $h>0$.

Initialization: Set $\widehat{\theta}_0^{\text{OGD}}$, $t=1$, $n=1$.

for $t = 1,\dots,T$ **do**

$\quad$ **if** $t = nh+1$ **then**

$\quad\quad \widehat{\theta}_n^{\text{OGD}} = \widehat{\theta}_{n-1}^{\text{OGD}} - \varepsilon_{n-1}\left(\sum_{k=1}^h \partial\mathcal{L}(Y_{(n-1)h+k}, \langle\widehat{\theta}_{n-1}^{\text{OGD}}, X_{(n-1)h+k}\rangle) + \alpha\widehat{\theta}_{n-1}^{\text{OGD}}\right)$

$\quad\quad \bar{\theta}_n^{\text{OGD}} = \frac{1}{n}\sum_{j=1}^n \Pi(\widehat{\theta}_j^{\text{OGD}})$; ▷ `Average previous OGD steps`

$\quad\quad n \leftarrow n+1$

$\quad X_t = \operatorname*{argmax}_{x\in\mathcal{X}_t} \langle\bar{\theta}_n^{\text{OGD}}, x\rangle + \gamma_{t,T}^{\text{OGD}}(x)\,.$; ▷ `Freeze` $\bar{\theta}_n^{\text{OGD}}$ `for` h `steps`

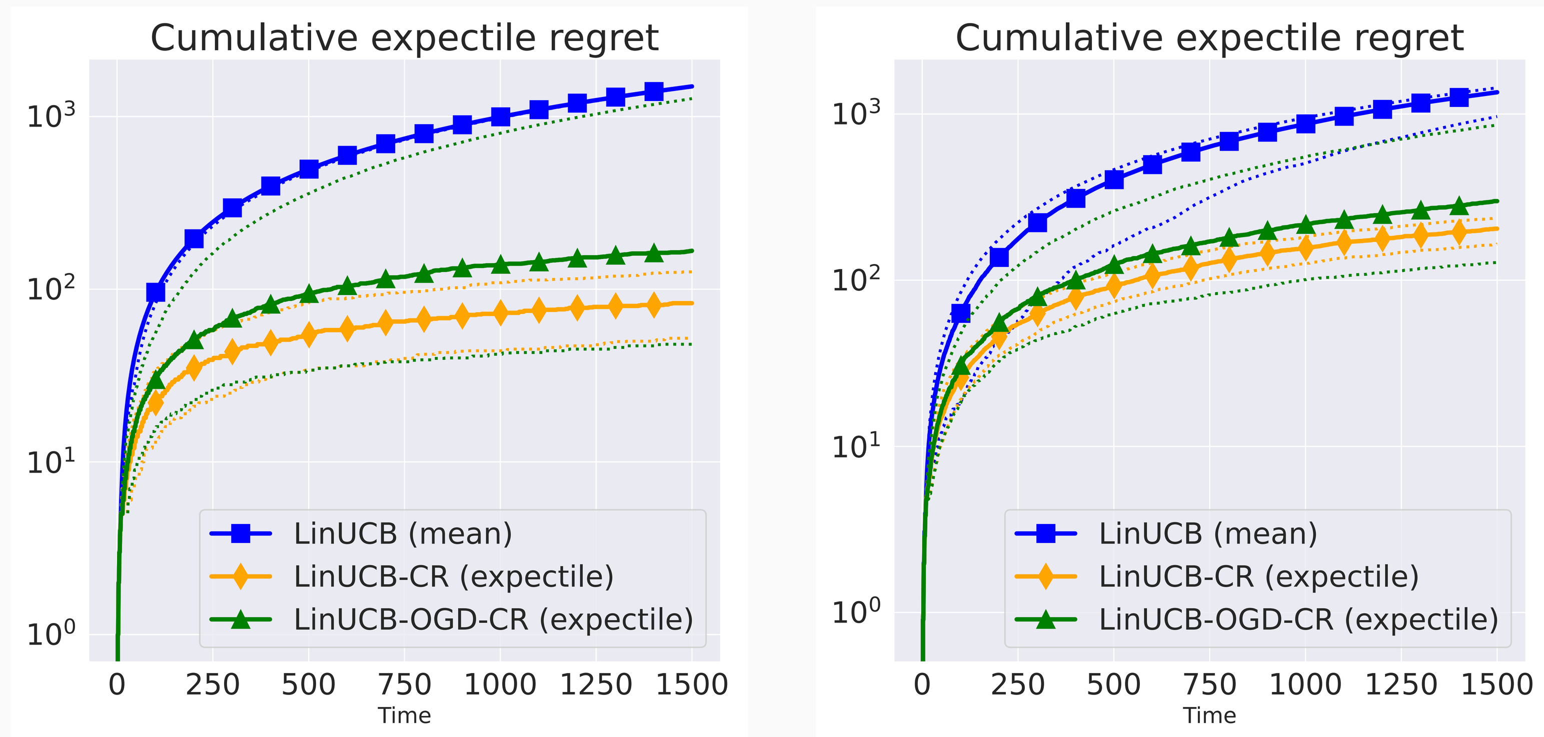
Regret of LinUCB-OGD-CR

With probability at least $1-\delta$, *under stochastic arrival of action sets*,

$$\mathcal{R}_T^{\text{LinUCB-OGD}} = \mathcal{O}\left(\sqrt{T}\times\text{Polylog}(T)\right)$$

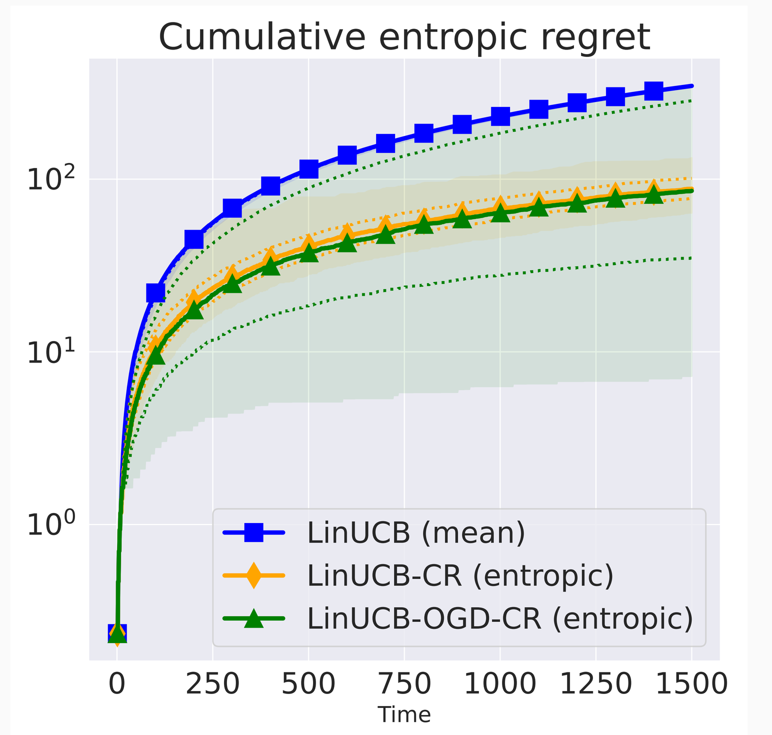
if episode length satisfies $h = \Omega\left(d^2\log\frac{1}{\delta}\right)$.

Numerical experiments



Gaussian expectile bandit.

Asymmetric Gaussian expectile bandit.



Bernoulli entropic risk bandit.

Full paper.

Acknowledgments: 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).