Risk-aware linear bandits with convex loss

AISTATS 2023

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

[1] Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9198 - CRIStAL, F-59000, Lille, France

Linear bandits

At round t:

- Observe action set $\mathcal{X}_t \subset \mathbb{R}^d$ and play action $X_t \in \mathcal{X}_t$.
- Receive $Y_t \sim p_{\langle \theta^*, X_t \rangle}$ where $\{p_\varphi\}$ is a statistical model.
- **Goal**: minimise regret

$$\mathcal{R}_T = \sum_{t=1}^{T} \max_{x \in \mathcal{X}_t} \mathbb{E}_{p_{\langle \theta^*, x \rangle}}[Y_t] - \mathbb{E}_{p_{\langle \theta^*, X_t \rangle}}[Y_t]$$

Linear bandits

At round t:

- Observe action set $\mathcal{X}_t \subset \mathbb{R}^d$ and play action $X_t \in \mathcal{X}_t$.
- Receive $Y_t \sim p_{\langle \theta^*, X_t \rangle}$ where $\{p_\varphi\}$ is a statistical model.
- **Goal**: minimise regret

$$\mathcal{R}_T = \sum_{t=1}^{T} \max_{x \in \mathcal{X}_t} \mathbb{E}_{p_{\langle \theta^*, x \rangle}} [Y_t] - \mathbb{E}_{p_{\langle \theta^*, X_t \rangle}} [Y_t]$$
$$= \sum_{t=1}^{T} \max_{x \in \mathcal{X}_t} \langle \theta^*, x \rangle - \langle \theta^*, X_t \rangle ,$$

if the bandit is mean-linear $\mathbb{E}_{p_\varphi} [Y_t] = \varphi$.

Example: $p_\varphi = \mathcal{N}(\varphi, 1)$.

Risk-aware linear bandits

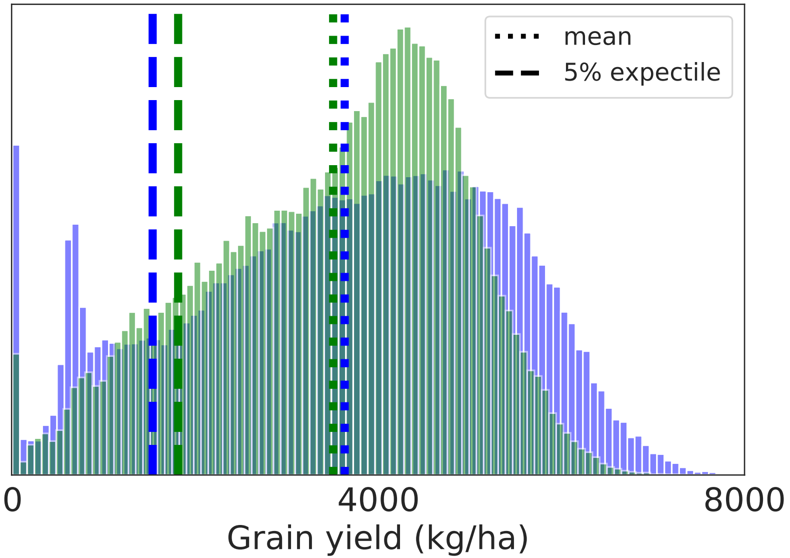
At round t:

- Observe action set $\mathcal{X}_t \subset \mathbb{R}^d$ and play action $X_t \in \mathcal{X}_t$.
- Receive $Y_t \sim p_{\langle \theta^*, X_t \rangle}$ 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 $\rho \colon \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ is a risk measure.

Motivation: real-world recommendations



Elicitable risk measures

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

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

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

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

Elicitable risk measures

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

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

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

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

Same regret as for mean-linear bandits:

$$\mathcal{R}_T = \sum_{t=1}^{T} \max_{x \in \mathcal{X}_t} \langle \theta^*, x \rangle - \langle \theta^*, X_t \rangle ,$$

but $\langle \theta^*, X_t \rangle$ represents a different risk measure than the expectation.

Examples of elicitable risk measures

Name	$\rho_{\mathcal{L}}$	Associated loss $\mathcal{L}(y, \xi)$
Mean	$\mathbb{E}[Y]$	$\frac{1}{2}(y-\xi)^2$
$\mathfrak{p}$-expectile	$\operatorname{argmin}_{\xi \in \mathbb{R}} \mathbb{E}[\psi_{\mathfrak{p}}(Y-\xi)]$ $\psi_{\mathfrak{p}}(z) = \lvert\mathfrak{p} - \mathbb{1}_{z<0}\rvert z^2$	$\psi_{\mathfrak{p}}(y-\xi)$

Example: $p_{\varphi}(y) = \frac{\sqrt{2\mathfrak{p}(1-\mathfrak{p})}}{\sqrt{\pi}\sqrt{\mathfrak{p}}+\sqrt{1-\mathfrak{p}}} \exp\left(-\frac{\psi_{\mathfrak{p}}(y-\varphi)}{2}\right)$

☞ $\rho_{\mathfrak{p}-\text{expectile}}(p_{\varphi}) = \varphi$.

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

LinUCB-CR

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

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

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

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

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

Numerical computation of $\hat{\theta}_t$ at each step!

$\neq$ mean-linear case: $\hat{\theta}_t = \left(\sum_{s=1}^{t-1} X_s X_s^\top + \alpha I_d\right)^{-1} \sum_{s=1}^{t-1} Y_s X_s$.

Analysis

Assumption (Bounded loss curvature)

$$\exists m, M \in \mathbb{R}_+^*,\ \forall y, \xi \in \mathbb{R},\ 0 < m \leq \frac{\partial^2 \mathcal{L}}{\partial \xi^2}(y, \xi) \leq M\,.$$

Definition (Global and local Hessian)

Global Hessian:

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

Local Hessian:

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

Analysis

Proposition (Self-normalised time-uniform concentration)

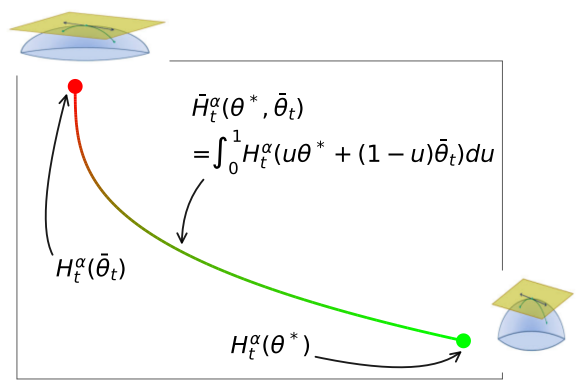
With probability $\geq 1 - \delta$, *for all* $t \in \mathbb{N}$,

$$\left\| \sum_{s=1}^{t-1} \partial \mathcal{L}(Y_s, \langle \theta^*, X_s \rangle) \right\|^2_{H_t^\alpha(\theta^*)^{-1}} \leq \sigma^2 \left(2 \log \frac{1}{\delta} + \log \frac{\det H_t^\alpha(\theta^*)}{\det \alpha I_d} \right) .$$

- Time-uniform confidence sets for θ^*.

Analysis

Transportation of Hessian metrics



Analysis

Theorem (Regret of LinUCB-CR)

With probability at least $1-\delta$,

$\kappa = \frac{M}{m}^{\dagger}$

dimension of actions

$\approx$ *variance of* $\partial\mathcal{L}(Y_t, \langle\theta^*, X_t\rangle)$

upper bound on $\|X_t\|_2$

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

lower bound on $\partial^2\mathcal{L}$

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

A faster approximate algorithm: LinUCB-OGD-CR

- Replace $\hat{\theta}_t \in \arg\min_{\theta \in \mathbb{R}^d} \sum_{s=1}^{t-1} \mathcal{L}(Y_s, \langle \theta, X_s \rangle) + \frac{\alpha}{2} \|\theta\|_2^2$ with episodic online gradient descent (OGD) with batch size h.

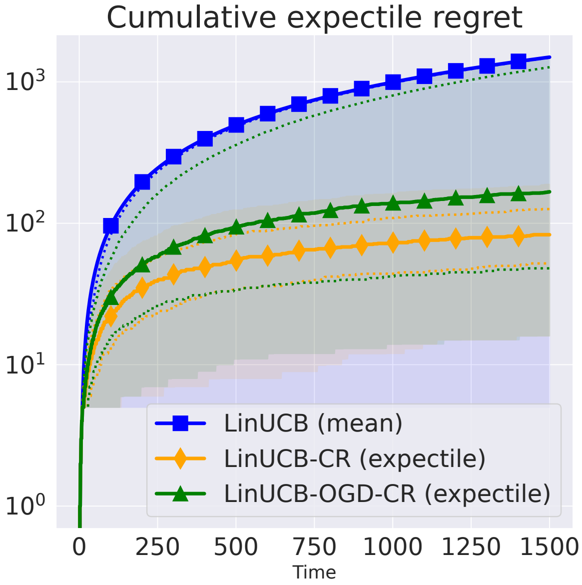
Theorem (Regret of LinUCB-OGD-CR)

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

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

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

Experiments




RISK
AHEAD