Jointly Efficient and Optimal Algorithms for Logistic Bandits

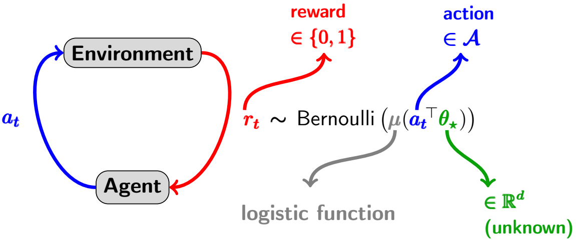
Louis Faury[1], Marc Abeille[1], Kwang-Sung Jun[2], Clément Calauzènes[1]

[1] Criteo AI Lab

[2] University of Arizona

Logistic Bandits

- Repeated game with **structured binary** feedback.

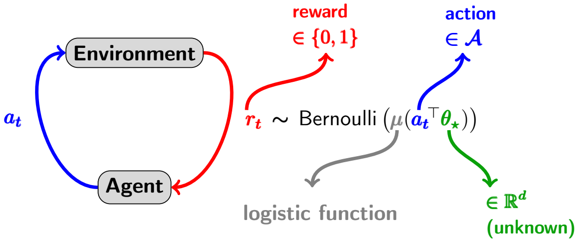


- The agent tries to minimize its cumulative pseudo-regret:

$$\text{Regret}(T) := T \max_{a\in\mathcal{A}} \mu(a^\top \boldsymbol{\theta_\star}) - \sum_{t=1}^{T} \mu(\boldsymbol{a_t}^\top \boldsymbol{\theta_\star}) \ .$$

Logistic Bandits

- Repeated game with **structured binary** feedback.



- The agent tries to minimize its cumulative pseudo-regret:

$$\text{Regret}(T) := T \max_{a\in\mathcal{A}} \mu(a^\top \theta_\star) - \sum_{t=1}^{T} \mu(a_t^\top \theta_\star) \ .$$

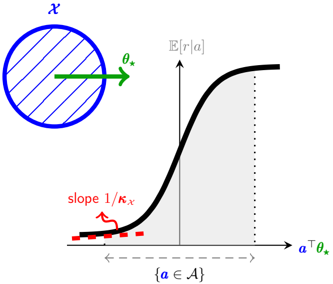
⇝ Highly relevant for practical applications,

⇝ Neat study of non-linearity in parametric bandits.

Statistical efficiency

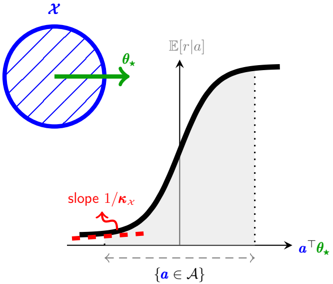
$$\kappa_\star = \frac{1}{\min_{a \in \mathcal{A}} \dot{\mu}(a^\top \theta_\star)} \; .$$

- ⇝ problem-dependent,
- ⇝ "distance" to linear model,
- ⇝ typically large ($\approx 10^3, 10^4$).



Statistical efficiency

$$\kappa_\star = \frac{1}{\min_{a\in\mathcal{A}} \dot{\mu}(a^\top \theta_\star)} \;.$$



⇝ problem-dependent,

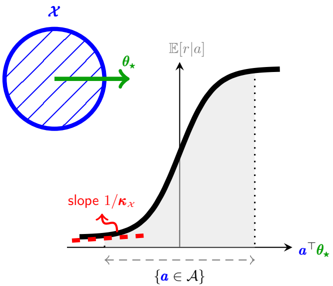
⇝ “distance” to linear model,

⇝ typically large ($\approx 10^3, 10^4$).

- **Towards minimax optimality** with respect to d, T and $\kappa_\star$:
  - [Filippi et al. 2010]: Regret$(T) = \tilde{\mathcal{O}}(\kappa_\star d\sqrt{T})$,
  - [Faury et al. 2020]: Regret$(T) = \tilde{\mathcal{O}}(\ d\sqrt{T})$,
  - [Abeille et al. 2021]: Regret$(T) = \tilde{\Omega}(d\sqrt{T/\kappa_\star}) \leftarrow$ **minimax-optimal**.

Statistical efficiency

$$\kappa_\star = \frac{1}{\min_{a\in\mathcal{A}} \dot{\mu}(a^\top \theta_\star)} \ .$$



⇝ problem-dependent,

⇝ "distance" to linear model,

⇝ typically large ($\approx 10^3, 10^4$).

- **Towards minimax optimality** with respect to d, T and $\kappa_\star$:
  - [Filippi et al. 2010]: Regret$(T) = \tilde{\mathcal{O}}(\kappa_\star d\sqrt{T})$,
  - [Faury et al. 2020]: Regret$(T) = \tilde{\mathcal{O}}(\ d\sqrt{T})$,
  - [Abeille et al. 2021]: Regret$(T) = \tilde{\Omega}(d\sqrt{T/\kappa_\star})$ ← **minimax-optimal**.

⇝ drastically improved practical performances! but ..

Computational Efficiency

- **Computational inefficiency** of statistically optimal algorithms:
  - batch computation of maximum likelihood estimators (MLEs),
  - confidence sets are expensive to maintain (non-ellipsoidal),

Computational Efficiency

- **Computational inefficiency** of statistically optimal algorithms:
  - batch computation of maximum likelihood estimators (MLEs),
  - confidence sets are expensive to maintain (non-ellipsoidal),

can we achieve computational efficiency without sacrificing statistical tightness?

Computational Efficiency

- **Computational inefficiency** of statistically optimal algorithms:
  - batch computation of maximum likelihood estimators (MLEs),
  - confidence sets are expensive to maintain (non-ellipsoidal),

can we achieve computational efficiency without sacrificing statistical tightness?

- **Yes!** Introducing the **ECOLog**.
  - new efficient estimator for $\theta_\star$,
  - associated with tight confidence regions.

ECOLog

- **ECOLog** estimator:

$$\theta_{t+1} = \text{argmin}_{\Theta}\, \ell_{t+1}(\theta) + \frac{1}{2}\|\theta - \theta_t\|^2_{\mathbf{W}_t} \,.$$

  - current log-loss ℓ_{t+1} for freshest point.
  - local quadratic lower-bounds for past data:

$$\sum_{s=1}^{t} \ell_s(\theta) \approx \frac{1}{2}\|\theta - \theta_t\|^2_{\mathbf{W}_t} \,,$$

$$\text{with } \mathbf{W}_t \leftarrow \mathbf{W}_{t-1} + \dot{\mu}(a_{t-1}^\top \theta_t) a_{t-1} a_{t-1}^\top \,.$$

- Efficient procedure;

⇝ strongly convex objective with cheap $\mathcal{O}(d^2)$ gradient computations.

ECOLog: confidence sets

Theorem (Confidence Regions)

Let $\delta \in (0,1]$ *and:*

$$\mathcal{C}_t(\delta) := \left\{ \left\| \theta - \theta_t \right\|^2_{\mathbf{W}_t} \lessapprox d \log(t/\delta) \right\} \ ,$$

Under mild conditions:

$$\mathbb{P}(\forall t \geq 1,\ \theta_\star \in \mathcal{C}_t(\delta)) \geq 1 - \delta \ .$$

- ⇝ ellipsoidal confidence set,
- ⇝ sufficient statistics $(\theta_t, \mathbf{W}_{t+1})$ are cheap to update,
- ⇝ as tight as [Abeille et al. 2021],

Results

- Combining ECOLog with existing planning mechanism:

Algorithm	Regret Bound	Cost Per-Round	Minimax	Efficient
GLM-UCB [Filippi et al. 2010]	$\tilde{\mathcal{O}}\left(\kappa_\star d\sqrt{T}\right)$	$\mathcal{O}\left(d^2\lvert\mathcal{A}\rvert T\right)$	✘	✘
GLOC, OL2M [Jun et al. 2017] [Zhang et al. 2016]	$\tilde{\mathcal{O}}\left(\kappa_\star d\sqrt{T}\right)$	$\mathcal{O}\left(d^2\lvert\mathcal{A}\rvert\right)$	✘	✔
OFULog-r [Abeille et al. 2021]	$\tilde{\Theta}\left(d\sqrt{T/\kappa_\star}\right)$	$\mathcal{O}\left(d^2\lvert\mathcal{A}\rvert T\right)$	✔	✘
(ada-)OFU-ECOLog (this paper)	$\tilde{\Theta}\left(d\sqrt{T/\kappa_\star}\right)$	$\tilde{\mathcal{O}}\left(d^2\lvert\mathcal{A}\rvert\right)$	✔	✔

- ✔ joint statistical and computational efficiency,
- ✔ compatible with Thompson Sampling strategies.

See you at the poster!