Open Problem: Tight Bounds for Bernoulli Rewards in Kernelized Multi-Armed Bandits

Simone Drago and Marco Mussi
simone.drago@polimi.it, marco.mussi@polimi.it

Setting

Interaction Protocol: At every round $t \in [\![T]\!]$, we choose an action $\mathbf{x}_t \in \mathcal{X}$ and observes $y_t \sim \text{Ber}(f(\mathbf{x}_t))$

Goal: Maximize a *fixed unknown* function $f : \mathcal{X} \to [0, 1]$ over a decision set $\mathcal{X} \subseteq \mathbb{R}^d$

Regularity Conditions: f belongs to a *reproducing kernel Hilbert space* (RKHS) $\mathcal{H}_k$ with bounded kernel k

Learning Problem: Minimize the *regret* $R_T(\mathfrak{A}) := T\, f(\mathbf{x}^\star) - \sum_{t \in [\![T]\!]} f(\mathbf{x}_t)$ where $\mathbf{x}^\star \in \arg\max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$

State of the Art

	Subgaussian	Bernoulli
No Structure	Lattimore and Szepesvári 2020 (Corollary 5.5)	Garivier and Cappé 2011 (Theorem 10)
Linear	Abbasi-Yadkori et al. 2011 (Theorem 2)	Faury et al. 2022 (Proposition 3)
Metric Space	Kleinberg et al. 2008 (Theorem 4.2)	Magureanu et al. 2014 (Theorem 2)
RKHS	Chowdhury and Gopalan 2017 (Theorem 2)	Open Problem

Open Problems

1 *Can we effectively estimate $f(\mathbf{x})$ in a new point $\mathbf{x} \in \mathcal{X}$ based on the history of past observations $\mathcal{G}_t := \{(\mathbf{x}_s, y_s)\}_{s=1}^{t-1}$ where y_s are Bernoulli samples?*

2 *Can we derive concentration guarantees for the deviation $|f(\mathbf{x}) - \mu_t(\mathbf{x})|$ (being $\mu_t(\mathbf{x})$ a suitable estimator of $f(\mathbf{x})$) which is tight for the Bernoulli observations?*

3 *Can we design regret minimization algorithms which achieve a $\log T$ regret guarantee, highlighting the dependence on $D_{\text{KL}}(f(\mathbf{x}), f(\mathbf{x}^\star))$ when $\mathcal{X}$ is finite?*

References

Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári. Improved algorithms for linear stochastic bandits. In *NeurIPS*, 2011.
S. Chowdhury and A. Gopalan. On kernelized multi-armed bandits. In *ICML*, 2017.
L. Faury, M. Abeille, K. Jun, and C. Calauzènes. Jointly efficient and optimal algorithms for logistic bandits. In *AISTATS*, 2022.
A. Garivier and O. Cappé. The KL-UCB algorithm for bounded stochastic bandits and beyond. In *COLT*, 2011.
R. Kleinberg, A. Slivkins, and E. Upfal. Multi-armed bandits in metric spaces. In *STOC*, 2008.
T. Lattimore and C. Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.
S. Magureanu, R. Combes, and A. Proutière. Lipschitz bandits: Regret lower bound and optimal algorithms. In *COLT*, 2014.
N. Srinivas, A. Krause, S. Kakade, and M. Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *ICML*, 2010.