Optimal Rates of (Locally) Differentially Private Heavy-tailed Multi-Armed Bandits

Youming Tao*[1], Yulian Wu*[2], Peng Zhao[3], Di Wang#[2]

[1] Shandong University, [2]King Abdullah University of Science and Technology, [3]Nanjing University

* Equal Contribution, # Corresponding Author

Problem Formulation

We study the heavy-tailed MAB under the constraint of DP and LDP respectively.

I. Heavy-tailed Multi-Armed Bandits (MAB)

- Arm: $\{1, \cdots, K\}$.
- Action: The learner selects an arm $a \in [K]$ to pull at each time t and obtains a reward x_t.
- Reward: Each arm $a \in [K]$ is associated with a fixed but unknown reward distribution $\mathcal{X}_a$. We consider that $\mathcal{X}_a$'s are **heavy-tailed** such that they have only $(1+v)$-th moment with some $v \in (0, 1]$, *i.e.*,

$$\mathbb{E}_{x \sim \mathcal{X}_a}[|x|^{1+v}] \leq u, \tag{1}$$

  where v and u are known constants.
- Goal: To minimize the (expected) cumulative **regret** $\mathcal{R}_T$ over the time horizon T:

$$\mathcal{R}_T \triangleq T\mu^* - \mathbb{E}\left[\sum_{t=1}^{T} x_t\right], \tag{2}$$

II. (Local) Differential Privacy (DP/LDP)

- DP: An algorithm $\mathcal{M}$ is ϵ-differentially private (DP) if for any adjacent streams σ and σ' (*i.e.*, σ and σ' differ at only one item), and any measurable subset $\mathcal{O}$ of the output space of $\mathcal{M}$, we have $\mathbb{P}[\mathcal{M}(\sigma) \in \mathcal{O}] \leq e^{\epsilon} \cdot \mathbb{P}[\mathcal{M}(\sigma') \in \mathcal{O}]$.
- LDP: An algorithm $\mathcal{M}: \mathcal{X} \to \mathcal{Y}$ is said to be ϵ-locally differentially private (LDP) if for any $x, x' \in \mathcal{X}$, and any measurable subset $\mathcal{O} \subset \mathcal{Y}$, it holds that $\mathbb{P}[\mathcal{M}(x) \in \mathcal{O}] \leq e^{\epsilon} \cdot \mathbb{P}[\mathcal{M}(x') \in \mathcal{O}]$.

Our Contributions

We reveal the differences between the private MAB with light-tailed and heavy-tailed rewards.

- We establish lower bounds for both DP and LDP models and devise optimal algorithms with matching upper bounds.
- New hard instances, mechanisms and private robust estimators are developed as byproducts.

Summary of Results

Problem	Model	Upper Bound	Lower Bound
Heavy-tailed Reward (Instance-dependent Bound)	ϵ-DP	$\mathrm{O}\left(\frac{\log T}{\epsilon}\sum_{\Delta_a>0}\left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}} + \max_a \Delta_a\right)$	$\Omega\left(\frac{\log T}{\epsilon}\sum_{\Delta_a>0}\left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}}\right)$
	ϵ-LDP	$\mathrm{O}\left(\frac{\log T}{\epsilon^2}\sum_{\Delta_a>0}\left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}} + \max_a \Delta_a\right)$	$\Omega\left(\frac{\log T}{\epsilon^2}\sum_{\Delta_a>0}\left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}}\right)$
Bounded/sub-Gaussian Reward (Instance-dependent Bound)	ϵ-DP	$\mathrm{O}\left(\frac{K\log T}{\epsilon} + \sum_{\Delta_a>0}\frac{\log T}{\Delta_a}\right)$	$\Omega\left(\frac{K\log T}{\epsilon} + \sum_{\Delta_a>0}\frac{\log T}{\Delta_a}\right)$
	ϵ-LDP	$\mathrm{O}\left(\frac{1}{\epsilon^2}\sum_{\Delta_a>0}\frac{\log T}{\Delta_a} + \Delta_a\right)$	$\Omega\left(\frac{1}{\epsilon^2}\sum_{\Delta_a>0}\frac{\log T}{\Delta_a}\right)$
Heavy-tailed Reward (Instance-independent Bound)	ϵ-DP	$\mathrm{O}\left(\left(\frac{K\log T}{\epsilon}\right)^{\frac{v}{1+v}} T^{\frac{1}{1+v}}\right)$	—
	ϵ-LDP	$\mathrm{O}\left(\left(\frac{K\log T}{\epsilon^2}\right)^{\frac{v}{1+v}} T^{\frac{1}{1+v}}\right)$	$\Omega\left(\left(\frac{K}{\epsilon^2}\right)^{\frac{v}{1+v}} T^{\frac{1}{1+v}}\right)$
Bounded/sub-Gaussian Reward (Instance-independent Bound)	ϵ-DP	$\mathrm{O}\left(\sqrt{KT\log T} + \frac{K\log T}{\epsilon}\right)$	$\Omega\left(\sqrt{KT} + \frac{K\log T}{\epsilon}\right)$
	ϵ-LDP	$\mathrm{O}\left(\frac{\sqrt{KT\log T}}{\epsilon}\right)$	$\Omega(\frac{\sqrt{KT}}{\epsilon})$

Methods Overview

I. Central DP Model

- **DP Robust Upper Confidence Bound (UCB)**:
  - We first develop an adaptive tree mechanism to privately and continuously calculate the sum of truncated rewards for each arm.
  - Then, we adopt a UCB based strategy with a carefully designed confidence bound for per-time arm selection.

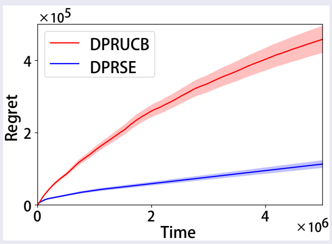
  However, we show that the UCB based DP algorithm is sub-optimal.
- **DP Robust Successive Elimination (SE)**:
  To further improve the regret, we proposed an SE based DP algorithm.
  In each iteration, the algorithm
  1. sets all the remaining arms as *viable* options
  2. pulls all the viable arms to get the same private confidence interval around their empirical rewards
  3. eliminates the arms with lower empirical rewards from the viable options if they are sub-optimal compared with other viable arms.

II. Local DP Model

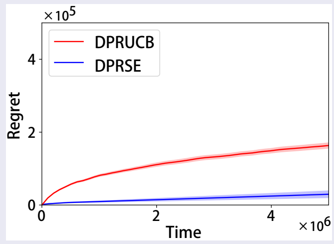
- **UCB is not unsatisfactory**
  - In local DP model, to use UCB strategy, each reward will be shrunken to a certain range and added Laplacian noise, which introduces enormous error to the reward mean estimation.
  - The reason is that, the truncation threshold depends on the pulling number, *i.e.*, the Laplacian noise added for each reward is proportional to its pulling number.
- **A better solution with SE**
  - To achieve a better utility, we propose an ϵ-LDP version of the SE algorithm.
  - The algorithm maintains a (private) confidence interval for each arm via the perturbed rewards instead of the noisy average.
  - Here the Laplacian noise added to each reward is **independent** on its pulling number, which is much smaller than the noise added in the LDP version of UCB strategy when the time horizon is sufficiently large.

Experiments

Synthetic data from Pareto distributions to generate reward, eg: arms means are $\{0.9, 0.85, 0.7, 0.45, 0.1\}$.



(a) $v = 0.5, \epsilon = 0.5$



(b) $v = 0.9, \epsilon = 1.0$