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

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

[1]School of Computer Science
Shandong University

[2]Division of Computer, Electrical, and Mathematical Sciences and Engineering
King Abdullah University of Science and Technology

[3]Department of Computer Science
Nanjing University

March 2022

Table of Contents

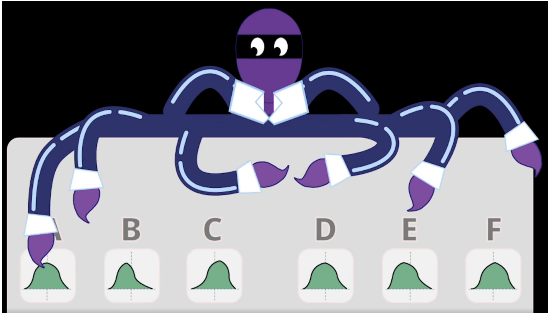
Motivation

- Bandits:
  exploration-exploitation dilemma in decision-making with uncertainty.
- Differential Privacy (DP):
  privacy issue in bandit: rewards.
- Previous assumptions:
  bounded/ sub-Gaussian distributions for rewards.
- The rewards in real world:
  heavy-tailed distributions.
  - modeling stock prices
  - preferential attachment in social networks
  - online behavior on websites
- Problem:
  multi-armed bandits (MAB) with heavy-tailed rewards in both central and local DP models.

Table of Contents

MAB with Heavy-tailed Rewards

- T time steps, K arms
- Unknown heavy-tailed $x_t \overset{i.i.d}{\sim} \mathcal{X}_{a_t} : \mathbb{E}_{X\sim\mathcal{X}_a}[|X|^{1+v}] \leq u, \quad v \in (0,1]$



- Bounded: $X \in [0,1]$
- Sub-gaussian: $\mathbb{E}e^{\lambda(X-\mathbb{E}X)} \leq e^{\frac{\sigma^2\lambda^2}{2}}, \mathbb{E}e^{\lambda(\mathbb{E}X-X)} \leq e^{\frac{\sigma^2\lambda^2}{2}}, \sigma \in [0,1]$

[1].https://multithreaded.stitchfix.com/blog/2020/08/05/bandits/

Performance Criteria

Definition (Regret)

The learner aims to maximize her/his expected cumulative reward over time, *i.e.*, to minimize the (expected) cumulative *regret*, defined as

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

where $\mu^* = \max_{a\in[K]} \mu_a$ and μ_a is the mean of distribution $\mathcal{X}_a$ for $a \in [K]$.

Differential Privacy and Local Differential Privacy

- Challenge: in online(bandit) learning settings, the algorithm might not see all of the data before making a decision.
- Strategy: define differential privacy (DP) in the **stream setting** since rewards are released continually.

Definition (Differential Privacy)

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 time step), and any measurable subset $\mathcal{O}$ of the output space of $\mathcal{M}$, we have

$$\mathbb{P}\left[\mathcal{M}(\sigma) \in \mathcal{O}\right] \leq e^{\epsilon} \cdot \mathbb{P}\left[\mathcal{M}(\sigma') \in \mathcal{O}\right].$$

Definition (Local Differential Privacy)

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 $\mathbb{P}\left[\mathcal{M}(x) \in \mathcal{O}\right] \leq e^{\epsilon} \cdot \mathbb{P}\left[\mathcal{M}(x') \in \mathcal{O}\right]$.

Differential Privacy and Local Differential Privacy

- Differential Privacy: a trusted curator collects all the data and then preserves the privacy.
- Local Differential Privacy: data providers only trust their local single devices and privatize their individual data before sending to the collector .

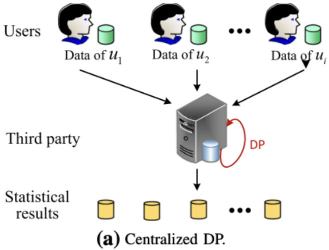


Figure: DP and LDP

[1].Zhao, Ping, et al. "A survey of local differential privacy for securing internet of vehicles." The Journal of Supercomputing 76.11 (2020): 8391-8412.

Table of Contents

Contributions

- ϵ-DP model
  - **DP Robust Upper Confidence Bound (UCB)** algorithm
    - Instance-dependent regret upper bound
  - **DP Robust Successive Elimination (SE)** algorithm
    - Instance-dependent regret upper and lower bounds (optimal)
    - Instance-independent regret upper bound
- ϵ-LDP model,
  - **LDP Robust SE** algorithm
    - Instance-dependent regret upper and lower bounds (optimal)
    - Instance-independent regret upper and lower bounds (near-optimal)

Contributions

Summary of our contributions

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

- Here $\Delta_a \triangleq \mu^* - \mu_a$ is the mean reward gap of arm a.

Table of Contents

DP Robust UCB

Previous private MAB methods under different settings:

- bounded setting: Tree-based mechanism to privately calculate the sum of rewards and then modify UCB algorithm
- heavy-tailed setting: reward is unbounded so we first preprocess the rewards to make them bounded.

Non-private MAB with heavy-tailed rewards

- robust-UCB (Bubeck et al. 2013): combining the UCB algorithm with several robust mean estimators.

DP Robust UCB: Truncation technique, Tree-based mechanism, UCB and Laplacian mechanism

DP Robust UCB

Algorithm 1 DP Robust Upper Confidence Bound

Input: time horizon T, parameters ϵ, v, u.

Create an empty tree Tree_a for each arm $a \in [K]$.
Initialize pull number $n_a \leftarrow 0$ for each arm $a \in [K]$.
Denote B_n as $(\frac{\epsilon u n}{\log^{1.5} T})^{1/(1+v)}$ for any $n \in \mathbb{N}^+$.
for $t = 1, \dots, K$ **do**
    Pull arm t and observe a reward x_t.
    Update the pull number $n_t \leftarrow n_t + 1$.
    Truncate the reward by $\widetilde{x}_t \leftarrow x_t \cdot \mathbb{I}_{|x_t| \le B_{n_t}}$.
    Insert $\widetilde{x}_t$ into Tree_t.
end for

$Tree_a$ for each arm

for $t = K+1, \dots, T$ **do**
    Obtain $\widehat{S}_a(t)$ for each $a \in [K]$ via Tree-based Mechanism. → Private sum of truncated rewards
    Pull arm

$$a_t = \arg\max_a \frac{\widehat{S}_a(t)}{n_a} + 18 u^{\frac{1}{1+v}} \Big(\frac{\log(2t^4) \log^{1.5+\frac{1}{v}} T}{n_a \epsilon}\Big)^{\frac{v}{1+v}}$$

→ Robust UCB

    and observe the reward x_t.
    Update the pull number $n_{a_t} \leftarrow n_{a_t} + 1$.
    Truncate the reward by $\widetilde{x}_t \leftarrow x_t \cdot \mathbb{I}_{|x_t| \le B_{n_{a_t}}}$. → Truncate reward
    Insert $\widetilde{x}_t$ into Tree_{a_t}.
end for

Theorem (Upper Bound of DP Robust UCB)

Under our assumptions, for any $0 < \epsilon \leq 1$ *the instance-dependent expected regret of DP Robust UCB algorithm satisfies*

$$\mathcal{R}_T \leq O\left(\sum_{a:\Delta_a>0}\left(\frac{\log^{2.5+\frac{1}{v}} T}{\epsilon}\left(\frac{u}{\Delta_a}\right)^{\frac{1}{v}} + \Delta_a\right)\right). \quad (2)$$

- Optimal rate of the regret in non-private version(Bubeck et al.,2013): $O(\sum_{a:\Delta_a>0}[\log T(\frac{u}{\Delta_a})^{\frac{1}{v}} + \Delta_a])$
- There is an additional factor of $\frac{\log^{1.5+\frac{1}{v}} T}{\epsilon}$.
- Whether it is possible to further improve the regret?

DP Robust SE

Algorithm 3 DP Robust Successive Elimination

Input: confidence β, parameters ϵ, v, u.

$\mathcal{S} \leftarrow \{1, \cdots, K\}$ → Set all the arms as viable options
Initialize: $t \leftarrow 0$, $\tau \leftarrow 0$.
repeat
  $\tau \leftarrow \tau + 1$.
  Set $\bar{\mu}_a = 0$ for all $a \in \mathcal{S}$.
  $r \leftarrow 0$, $D_\tau \leftarrow 2^{-\tau}$.
  $R_\tau \leftarrow \left\lceil u^{\frac{1}{v}} \left(\frac{24^{(1+v)/v} \log(4|\mathcal{S}|\tau^2/\beta)}{\epsilon D_\tau^{(1+v)/v}}\right) + 1 \right\rceil$.
  $B_\tau \leftarrow \left(\frac{u R_\tau \epsilon}{\log(4|\mathcal{S}|\tau^2/\beta)}\right)^{1/(1+v)}$.
  while $r < R_\tau$ **do**
    $r \leftarrow r + 1$.
    for $a \in \mathcal{S}$ **do**
      $t \leftarrow t + 1$.
      Sample a reward $x_{a,r}$.
      $\widetilde{x}_{a,r} \leftarrow x_{a,r} \cdot \mathbb{I}_{\{|x_{a,r}| \le B_\tau\}}$.
    end for
  end while
  For each $a \in \mathcal{S}$, compute $\bar{\mu}_a \leftarrow (\sum_{l=1}^{R_\tau} \widetilde{x}_{a,l})/R_\tau$.
  Set $\widetilde{\mu}_a \leftarrow \bar{\mu}_a + \mathrm{Lap}(\frac{2B_\tau}{R_\tau \epsilon})$ for all $a \in \mathcal{S}$.
  $\widetilde{\mu}_{\max} \leftarrow \max_{a \in \mathcal{S}} \widetilde{\mu}_a$.
  $err_\tau \leftarrow u^{1/(1+v)} \left(\frac{\log(4|\mathcal{S}|\tau^2/\beta)}{R_\tau \epsilon}\right)^{v/(1+v)}$.
  for all viable arm a **do**
    if $\widetilde{\mu}_{\max} - \widetilde{\mu}_a > 12 err_\tau$ **then**
      Remove arm a from $\mathcal{S}$.
    end if
  end for
until $|\mathcal{S}| = 1$
Pull the arm in $\mathcal{S}$ in all remaining $T - t$ rounds.

Pull all the viable arms to get the same private confidence interval around empirical rewards

Eliminate the arms with sub-optimal empirical rewards

DP Robust SE

Theorem (DP Upper Bound)

In DP Robust SE algorithm, for sufficiently large T and any $\epsilon \in (0,1]$, the instance-dependent and instance-independent expected regret satisfies

$$\mathcal{R}_T \leq O\left(\frac{u^{\frac{1}{1+v}} \log T}{\epsilon} \sum_{\Delta_a > 0} \left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}} + \max_a \Delta_a\right), \mathcal{R}_T \leq O\left(u^{\frac{v}{(1+v)^2}} \left(\frac{K \log T}{\epsilon}\right)^{\frac{v}{1+v}} T^{\frac{1}{1+v}}\right)$$

respectively.

Theorem (DP Instance-dependent Lower Bound)

There exists a heavy-tailed K-armed bandit instance with $u \leq 1$, $\mu_a \leq \frac{1}{6}$ and $\Delta_a \in (0, \frac{1}{12})$, such that for any ϵ-DP ($0 < \epsilon \leq 1$) algorithm $\mathcal{A}$ whose expected regret is at most $T^{\frac{3}{4}}$, we have

$$\mathcal{R}_T \geq \Omega\left(\frac{\log T}{\epsilon} \sum_{\Delta_a > 0} \left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}}\right). \tag{3}$$

LDP Robust SE

The basic idea is similar to DP Robust SE, while the algorithm now maintains private confidence interval for **each arm** via the perturbed rewards instead of the noisy average.

Theorem (LDP Upper Bound)

In LDP Robust SE algorithm. For any $\epsilon \in (0,1]$ *and sufficiently large* T*, the instance-dependent expected regret satisfies*

$$\mathcal{R}_T \leq O\left(\frac{u^{\frac{2}{v}} \log T}{\epsilon^2} \sum_{\Delta_a > 0} \left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}} + \max_a \Delta_a\right). \quad (4)$$

Moreover, the instance-independent expected regret satisfies

$$\mathcal{R}_T \leq O\left(u^{\frac{2}{1+v}} \left(\frac{K \log T}{\epsilon^2}\right)^{\frac{v}{1+v}} T^{\frac{1}{1+v}}\right), \quad (5)$$

where the $O(\cdot)$*-notations omit* $\log\log\frac{1}{\Delta_a}$ *terms.*

Theorem (LDP Instance-dependent Lower Bound)

There exists a heavy-tailed K-armed bandit instance with $u \leq 1$ and $\Delta_a \triangleq \mu_1 - \mu_a \in (0, \frac{1}{5})$, such that for any ϵ-LDP ($0 < \epsilon \leq 1$) algorithm whose regret $\leq o(T^\alpha)$ for any $\alpha > 0$, the regret satisfies

$$\liminf_{T \to \infty} \frac{\mathcal{R}_T}{\log T} \geq \Omega\left(\frac{1}{\epsilon^2} \sum_{\Delta_a > 0} (\frac{1}{\Delta_a})^{\frac{1}{v}}\right).$$

Theorem (LDP Instance-independent Lower Bound)

There exists a heavy-tailed K-armed bandit instance with the $(1+v)$-th bounded moment of each reward distribution is bounded by 1. Moreover, if T is large enough, for any the ϵ-LDP algorithm $\mathcal{A}$ with $\epsilon \in (0, 1]$, the expected regret must satisfy

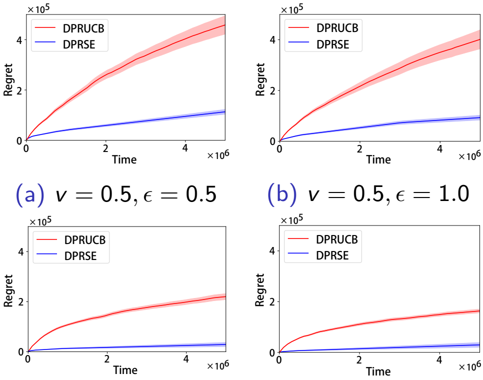
$$\mathcal{R}_T \geq \Omega\left(\left(\frac{K}{\epsilon^2}\right)^{\frac{v}{1+v}} T^{\frac{1}{1+v}}\right).$$

Table of Contents

Experimental Results in DP Model

- DPRUCB and DPRSE for an instance of 5 arms.
- Pareto distributions as the reward distributions with means being $0.9, 0.7, 0.5, 0.3, 0.1$ in setting 1.

Figure: DP Model Setting 1



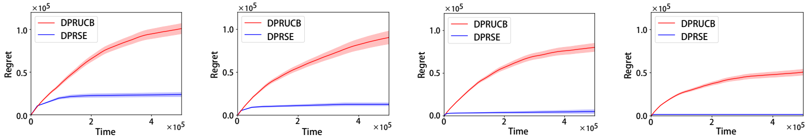
(a) $v = 0.5, \epsilon = 0.5$ (b) $v = 0.5, \epsilon = 1.0$

(c) $v = 0.9, \epsilon = 0.5$ (d) $v = 0.9, \epsilon = 1.0$

Experimental Results in DP Model

In setting 2, the means of each arm a are $\{0.9, 0.55, 0.3, 0.15, 0.1\}$.

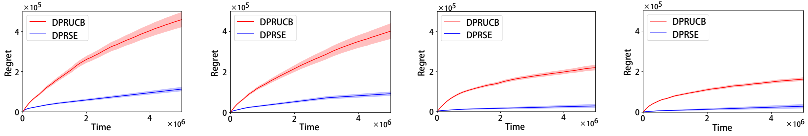
Figure: DP Model Setting 2



(a) $v = 0.5, \epsilon = 0.5$ (b) $v = 0.5, \epsilon = 1.0$ (c) $v = 0.9, \epsilon = 0.5$ (d) $v = 0.9, \epsilon = 1.0$

In setting 3, the means of each arm a are $\{0.9, 0.85, 0.7, 0.45, 0.1\}$.

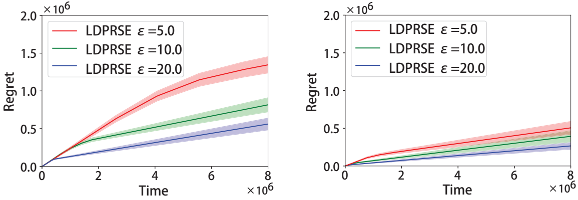
Figure: DP Model Setting 3



(a) $v = 0.5, \epsilon = 0.5$ (b) $v = 0.5, \epsilon = 1.0$ (c) $v = 0.9, \epsilon = 0.5$ (d) $v = 0.9, \epsilon = 1.0$

We evaluate LDPRSE for the local DP model in the Setting 3 as the central DP model.

Figure: LDP Model



(a) $v = 0.5$ (b) $v = 0.9$

Open problems

1. Throughout the whole paper we need to assume both u and v are known. How to address a more practical case where they are unknown?
2. For the setting of MAB with bounded reward, an UCB-based private algorithm can also attain an optimal regret guarantee. Whether it is possible to get an optimal DP variant of UCB algorithm for our problem.

Thank You!