On Private and Robust Bandits

Yulian Wu*[1], Xingyu Zhou*[2], Youming Tao[3] and Di Wang[1]

[1] KAUST, [2] Wayne State University, [3] Shandong University

NeurIPS 2023

Table of Contents

Multi-armed Bandits

- The agent interacts with the environment for T rounds.
- In each round t, the agent chooses an action $a_t \in [K]$
- Standard reward r_t is generated independently from inlier distribution.
- After contamination, the agent observes contaminated reward x_t.

Robustness:

Two Classes of Heavy-tailed Reward Distributions

Definition (Finite k-th raw moment)

A distribution over $\mathbb{R}$ is said to have a finite k-th raw moment if it is within

$$\mathcal{P}_k = \left\{ P : \mathbb{E}_{X\sim P}\left[|X|^k \right] \leq 1 \right\}, \quad k \geq 2,$$

where k is considered fixed but arbitrary.

Definition (Finite k-th central moment)

A distribution over $\mathbb{R}$ is said to have a finite k-th central moment if it is within

$$\mathcal{P}_k^c = \left\{ P : \mathbb{E}_{X\sim P}\left[|X-\mu|^k \right] \leq 1 \right\}, \quad k \geq 2,$$

where $\mu := \mathbb{E}_{X\sim P}[X] \in [-D, D]$ and $D \geq 1$.

Definition (Heavy-tailed MABs with Huber contamination)

Given the corruption level $\alpha \in [0, 1/2)$. For each round $t \in [T]$, the observed reward x_t for action a_t, is sampled independently from the true distribution $P_{a_t} \in \mathcal{P}_k$ (or $P_{a_t} \in \mathcal{P}_k^c$) with probability $1 - \alpha$; otherwise is sampled from some arbitrary and unknown contamination distribution $G_{a_t} \in \mathcal{G}$.

Privacy

Definition (Differential Privacy for MABs)

For any $\epsilon > 0$, a learning algorithm $\mathcal{M} : \mathbb{R}^T \to [K]^T$ is ϵ-DP if for all sequences $\mathcal{D}_T, \mathcal{D}'_T \in \mathbb{R}^T$ differing only in a single element and for all events $E \subset [K]^T$, we have

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

Table of Contents

Regrets

- μ_a: the mean of the inlier distribution of arm $a \in [K]$;
- $\mu^* = \max_{a \in [K]} \mu_a$;
- Π^ϵ: the set of all ϵ-DP MAB algorithms;
- $\mathcal{E}_{\alpha,k}$: the set of all instances of heavy-tailed MABs (with parameter k) with Huber contamination (of level α).

Definition (Clean Regret)

Fix an algorithm $\pi \in \Pi^\epsilon$ and an instance $\nu \in \mathcal{E}_{\alpha,k}$. Then, the clean regret of π under ν is given by $\mathcal{R}_T(\pi, \nu) := \mathbb{E}_{\pi,\nu}[T\mu^* - \sum_{t=1}^{T} \mu_{a_t}]$.

To capture the intrinsic difficulty of the private and robust MAB problem, we are also interested in its minimax regret.

Definition (Minimax Regret)

The minimax regret of our private and robust MAB problem is defined as $\mathcal{R}^{\text{minimax}}_{\epsilon,\alpha,k} := \inf_{\pi \in \Pi^\epsilon} \sup_{\nu \in \mathcal{E}_{\alpha,k}} \mathbb{E}_{\pi,\nu}[T\mu^* - \sum_{t=1}^{T} \mu_{a_t}]$.

Table of Contents

Lower Bound

Theorem

Consider a private and robust MAB problem where inlier distributions have finite k-th raw (or central) moments ($k \geq 2$). Then, its minimax regret satisifes

$$\mathcal{R}^{minimax}_{\epsilon,\alpha,k} = \Omega\left(\sqrt{KT} + (K/\epsilon)^{1-\frac{1}{k}} T^{\frac{1}{k}} + T\alpha^{1-\frac{1}{k}}\right).$$

Table of Contents

A Meta Algorithm

Algorithm 1 Private and Robust Arm Elimination

Input: Number of arms K, time horizon T, privacy budget ϵ, Huber parameter $\alpha \in (0, 1/2)$, error probability $\delta \in (0, 1]$, inliner distribution parameters i.e., k and optional D.
Initialize: $\tau = 0$, active set of arms $\mathcal{S} = \{1, \cdots, K\}$.
for batch $\tau = 1, 2, \ldots$ **do**
  Set batch size $B_\tau = 2^\tau$.
  if $B_\tau < \mathcal{T}$ **then**
    Randomly select an action $a \in [K]$.
    Play action a for B_τ times.
  else
    for each active arm $a \in \mathcal{S}$ **do**
      for i from 1 to B_τ **do**
        Pull arm a, observe contaminated reward x_i^a.
        If total number of pulls reaches T, **exit**.
      end for
      Set truncation threshold M_τ.
      Set additional parameters Φ.
      Compute estimate $\widetilde{\mu}_a = \texttt{PRM}(\{x_i^a\}_{i=1}^{B_\tau}, M_\tau, \Phi)$.
    end for
    Set confidence radius β_τ.
    Let $\widetilde{\mu}_{\max} = \max_{a \in \mathcal{S}} \widetilde{\mu}_a$.
    Remove all arms a from $\mathcal{S}$ s.t. $\widetilde{\mu}_{\max} - \widetilde{\mu}_a > 2\beta_\tau$.
  end if
end for

Table of Contents

Finite Raw Moment Case

Algorithm 2 PRM for the finite raw moment case

1: **Input:** A collection of data $\{x_i\}_{i=1}^n$, truncation parameter M, additional parameters $\Phi = \{\epsilon\}$.
2: **for** $i = 1, 2, \ldots, n$ **do**
3: Truncate data $\bar{x}_i = x_i \cdot \mathbb{1}_{\{|x_i| \leq M\}}$.
4: **end for**
5: Return private estimate $\widetilde{\mu} = \frac{\sum_{i=1}^n \bar{x}_i}{n} + \mathrm{Lap}(\frac{2M}{n\epsilon})$.

Theorem (Performance Guarantees)

Consider a private and robust MAB with inlier distributions satisfying Definition 1 and $0 < \alpha \leq \alpha_1 \in (0, 1/2)$. Let Algorithm 1 be instantiated with Algorithm 2 . Set $\mathcal{T} = \Omega(\frac{\log(1/\delta)}{\alpha_1})$ and $\delta = 1/T$. Then Algorithm 1 is ϵ-DP with its regret upper bound

$$\mathcal{R}_T = O\left(\sqrt{KT\log T} + \left(\frac{K\log T}{\epsilon}\right)^{\frac{k-1}{k}} T^{\frac{1}{k}} + T\alpha_1^{1-\frac{1}{k}} + \frac{K\log T}{\alpha_1}\right).$$

Finite Central Moment Case

Algorithm 3 PRM for the finite central moment case

1: **Input:** A collection of data $\{x_i\}_{i=1}^{2n}$, truncation parameter M, additional parameters $\Phi = \{\epsilon, D, r\}, r \in \mathbb{R}$.
2: // First step: initial estimate
3: $B_j = [j, j+r), j \in \mathcal{J} = \{-D, -D+r, \dots, D-r\}$.
4: Compute private histogram using the first fold of data: $\widetilde{p}_j = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{X_i \in B_j\}}}{n} + \text{Lap}\left(\frac{2}{n\epsilon}\right)$.
5: Get the initial estimate $J = \arg\max_{j \in \mathcal{J}} \widetilde{p}_j$.
6: // Second step: final estimate
7: Get final estimator using the second fold of data: $\widetilde{\mu} = J + \frac{1}{n}\sum_{i=n+1}^{2n}(X_i - J)\mathbb{1}_{\{|X_i - J| \leq M\}} + \text{Lap}\left(\frac{2M}{n\epsilon}\right)$.

Finite Central Moment Case

Theorem (Performance Guarantees, $\alpha = 0$)

Let Algorithm 1 be instantiated with Algorithm 3. Set $\mathcal{T} = \Omega(\frac{\log(D/\delta)}{\epsilon})$ and $\delta = 1/T$. Then, Algorithm 1 is ϵ-DP with its regret upper bound

$$\mathcal{R}_T = O(\sqrt{KT \log T} + (K \log T/\epsilon)^{\frac{k-1}{k}} \; T^{\frac{1}{k}} + \gamma),$$

where $\gamma := O\left(KD \log(DT)/\epsilon\right)$.

Theorem (Performance Guarantees, $\alpha > 0$)

For $\alpha \leq \alpha_1 \in (0, 0.133)$, let Algorithm 1 be instantiated with Algorithm 3. Set $\delta = 1/T$, then Algorithm 1 is ϵ-DP with its regret upper bound

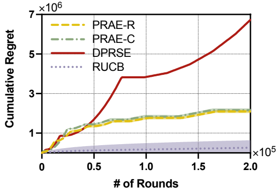
$$\mathcal{R}_T = O(\sqrt{KT \log T} + (K \log T/\epsilon)^{\frac{k-1}{k}} \; T^{\frac{1}{k}} + T\alpha_1^{1-\frac{1}{k}} + \hat{\gamma}),$$

where $\hat{\gamma} := O\left(\frac{DK \log T}{\alpha_1^2} + \frac{\iota DK \log T}{\epsilon} + \frac{DK \log(DT)}{\epsilon}\right)$ *and* $\iota = \frac{1-\alpha}{0.249-\alpha}$.

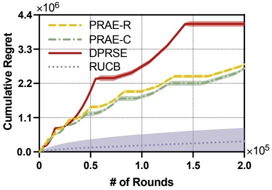
Table of Contents

Experiments

- PRAE-R: Our Algorithm for Finite Raw Moment Case
- PRAE-C: Our Algorithm for Finite Central Moment Case
- DPRSE [Tao et al., 2021]: DP heavy-tailed MAB
- RUCB [Kapoor et al., 2019]: non-private robust algorithm



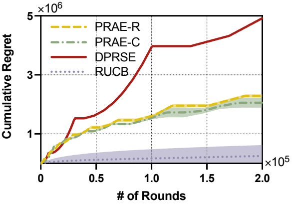
(a) $\alpha = 2\%$, $\epsilon = 0.2$



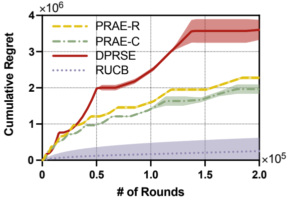
(b) $\alpha = 10\%$, $\epsilon = 1$

Figure: Experimental results under Pareto distribution

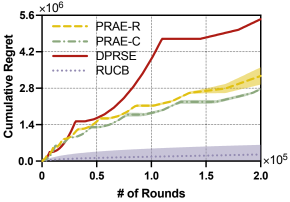
Experiments



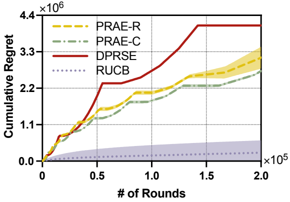
(a) $\alpha = 5\%$, $\epsilon = 0.5$



(b) $\alpha = 5\%$, $\epsilon = 1$



(c) $\alpha = 10\%$, $\epsilon = 0.5$



(d) $\alpha = 10\%$, $\epsilon = 1$

Figure: Experimental results under Student's *t* reward

References

Youming Tao, Yulian Wu, Peng Zhao, and Di Wang. Optimal rates of (locally) differentially private heavy-tailed multi-armed bandits. *arXiv preprint arXiv:2106.02575*, 2021.

Sayash Kapoor, Kumar Kshitij Patel, and Purushottam Kar. Corruption-tolerant bandit learning. *Machine Learning*, 108(4): 687–715, 2019.

Thank you!