On the Complexity of Differentially Private Best-Arm Identification with Fixed Confidence

Achraf Azize, Marc Jourdan, Aymen Al Marjani, Debabrota Basu

Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France.

Setting

FC-BAI with ϵ-global Differential Privacy

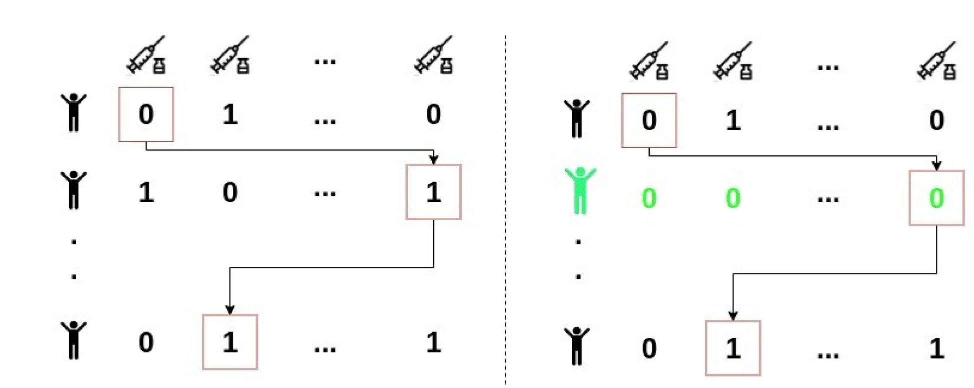
A BAI strategy π interacts with a set of users $\{u_1, \dots, u_T\}$ using the protocol

Algorithm 1 Interaction protocol

1. **Input:** A BAI strategy $\pi = (S_t, \mathrm{Rec}_t)_{t \geq 1}$ and Users $\{u_t\}_{t \geq 1}$ represented by the table $\underline{\mathbf{d}}$
2. **Output:** A stopping time τ, a sequence of samples actions $\underline{a}^\tau = (a_1, \dots, a_\tau)$ and a recommendation $\hat{a}$ satisfying ϵ-global DP
3. **for** $t = 1, \dots$ **do**
4. $\quad$ π samples $a_t \sim S_t(. \mid a_1, r_1, \dots, a_{t-1}, r_{t-1})$
5. $\quad$ **if** $a_t = \top$ **then**
6. $\quad\quad$ Halt.
7. $\quad\quad$ Return $\hat{a} \sim \mathrm{Rec}_t(. \mid a_1, r_1, \dots, a_{t-1}, r_{t-1})$ and $\tau = t$
8. $\quad$ **else**
9. $\quad\quad$ u_t sends the **sensitive** reward $r_t \triangleq \underline{\mathbf{d}}_{t, a_t}$
10. $\quad$ **end if**
11. **end for**

Goal: Protect the privacy of the users by designing a Differentially Private (DP) BAI strategy π, that is δ correct, with $\mathbb{E}[\tau]$ as small as possible.

Illustration: K medicine testing with T patients



The private input: Each user u_t is represented by $\mathbf{x}_t \triangleq (x_{t,1}, \dots, x_{t,K})$, but only x_{t,a_t} is observed. We represent a set of users $\{u_t\}_{t=1}^T$ until T by *the table of potential rewards* $\underline{\mathbf{d}}^T \triangleq \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$.

The mechanism: A BAI strategy π interacts with the private table $\underline{\mathbf{d}}^T$, following the interaction of Algorithm 1, to halt at time T, produce a sequence of action $\underline{a}^T$ and recommend the action $\hat{a}$, with probability $\pi(\underline{a}^T, \hat{a}, T \mid \underline{\mathbf{d}}^T) \triangleq \mathrm{Rec}_{T+1}(\hat{a} \mid \mathcal{H}_T) \mathrm{S}_{T+1}(\top \mid \mathcal{H}_T) \prod_{t=1}^T \mathrm{S}_t(a_t \mid \mathcal{H}_{t-1})$.

Neighboring tables: Two reward tables $\underline{\mathbf{d}}^T$ and $\underline{\mathbf{d}}'^T$ are neighbouring if they only differ in one row, i.e. $d_{\mathrm{Ham}}(\underline{\mathbf{d}}^T, \underline{\mathbf{d}}'^T) = 1$.

ϵ-global DP for BAI: A BAI strategy π is **ϵ-global DP**, if for all $T \geq 1$, all neighbouring table of rewards $\underline{\mathbf{d}}^T$ and $\underline{\mathbf{d}}'^T$, all sequences of sampled actions $\underline{a}^T$ and recommended actions $\widehat{a} \in [K]$ we have that

$$\pi(\underline{a}^T, \hat{a}, T \mid \underline{\mathbf{d}}^T) \leq e^\epsilon \pi(\underline{a}^T, \hat{a}, T \mid \underline{\mathbf{d}}'^T)$$

Correctness: A BAI strategy π is **δ-correct** for a class $\mathcal{M}$, if for every instance $\boldsymbol{\nu} \in \mathcal{M}$, π recommends the optimal action $a^\star(\boldsymbol{\nu}) = \arg\max_{a \in [K]} \mu_a$ with probability at least $1 - \delta$, i.e. $\mathbb{P}_{\boldsymbol{\nu},\pi}(\tau < \infty, \widehat{a} = a^\star(\boldsymbol{\nu})) \geq 1 - \delta$.

Contributions

1. We derive the first lower bound on sample complexity of any δ-correct ϵ-global DP BAI strategy.
2. We design an ϵ-global DP variant of Top Two algorithms, named AdaP-TT, based on two simple design techniques, i.e. adaptive episodes for each arm and Laplacian mechanism.
3. We derive an asymptotic upper bound on the sample complexity of AdaP-TT. We show that AdaP-TT enjoys both theoretical near-optimality and good experimental performance.

Algorithm Design

Main Ingredients:
1. Per-arm doubling (Line 5).
2. Forgetting, i.e. the private empirical estimate of arm a is only computed using the rewards collected in the last phase of arm a (Line 8).
3. Each empirical mean (Line 9) is made ϵ-DP by adding Laplace noise.

Algorithm 2 AdaP-TT

1: **Input:** $\beta \in (0,1)$, risk $\delta \in (0,1)$, privacy budget ϵ, thresholds $c_{\epsilon,k_1,k_2} : \mathbb{N}^2 \times (0,1) \to \mathbb{R}^+$
2: **Output:** Recommendation $\hat{a}$ and Stopping time τ satisfying ϵ-global DP
3: **Initialization:** $\forall a \in [K]$, pull arm a, set $k_a = 1$, $T_1(a) = K + 1$, $L_{n,a} = 0$, $N_{n,a} = 1$, $n = K + 1$.
4: **for** $n > K$ **do**
5: $\quad$ **if** there exists $a \in [K]$ such that $N_{n,a} \geq 2N_{T_{k_a}(a),a}$ **then**
6: $\quad\quad$ Change phase $k_a \leftarrow k_a + 1$ for this arm a
7: $\quad\quad$ Set $T_{k_a}(a) = n$ and $\tilde{N}_{k_a,a} = N_{T_{k_a}(a),a} - N_{T_{k_a-1}(a),a}$
8: $\quad\quad$ Set $\hat{\mu}_{k_a,a} = \tilde{N}_{k_a,a}^{-1} \sum_{s=T_{k_a-1}(a)}^{T_{k_a}(a)-1} X_s \mathbb{1}\{I_s = a\}$
9: $\quad\quad$ Set $\tilde{\mu}_{k_a,a} = \hat{\mu}_{k_a,a} + Y_{k_a,a}$ where $Y_{k_a,a} \sim \mathrm{Lap}((\epsilon \tilde{N}_{k_a,a})^{-1})$
10: $\quad$ **end if**
11: $\quad$ Set $\hat{a}_n = \arg\max_{b \in [K]} \tilde{\mu}_{k_b,b}$
12: $\quad$ **if** $\frac{(\tilde{\mu}_{k_{\hat{a}_n},\hat{a}_n} - \tilde{\mu}_{k_b,b})^2}{1/\tilde{N}_{k_{\hat{a}_n},\hat{a}_n} + 1/\tilde{N}_{k_b,b}} \geq 2c_{\epsilon,k_{\hat{a}_n},k_b}(\tilde{N}_{k_{\hat{a}_n},\hat{a}_n}, \tilde{N}_{k_b,b}, \delta)$ for all $b \neq \hat{a}_n$ **then**
13: $\quad\quad$ **return** $(\hat{a}_n, n)$
14: $\quad$ **end if**
15: $\quad$ Set $B_n = \arg\max_{a \in [K]} \{\tilde{\mu}_{k_a,a} + \sqrt{k_a/\tilde{N}_{k_a,a}} + k_a/(\epsilon \tilde{N}_{k_a,a})\}$
16: $\quad$ Set $C_n = \arg\min_{a \neq B_n} \frac{\tilde{\mu}_{k_{B_n},B_n} - \tilde{\mu}_{k_a,a}}{\sqrt{1/N_{n,B_n} + 1/N_{n,a}}}$
17: $\quad$ Set $I_n = B_n$ if $N_{n,B_n}^{B_n} \leq \beta L_{n+1,B_n}$, else $I_n = C_n$
18: $\quad$ Pull I_n and observe $X_n \sim \nu_{I_n}$
19: $\quad$ Set $N_{n+1,I_n} \leftarrow N_{n,I_n} + 1$, $N_{n+1,I_n}^{B_n} \leftarrow N_{n,I_n}^{B_n} + 1$ and $L_{n+1,B_n} \leftarrow L_{n,B_n} + 1$. Set $n \leftarrow n + 1$
20: **end for**

Privacy analysis: For rewards in $[0,1]$, AdaP-TT is ϵ-global DP. A change in one user *only affects* the empirical mean at one episode of an arm, which is made private using the Laplace Mechanism.

Correctness: AdaP-TT is δ-correct for thresholds $\tilde{c}_{\epsilon,k_1,k_2}(n, m, \delta)$ which verify $\tilde{c}_{\epsilon,k_1,k_2}(n, m, \delta) \approx 2\log(1/\delta) + (1/n + 1/m)\log(1/\delta)^2/\epsilon^2$.

Upper bound on expected sample complexity: AdaP-TT with thresholds $\tilde{c}_{\epsilon,k_1,k_2}$ satisfies that, for all $\mu \in \mathbb{R}^K$ such that $\min_{a \neq b} |\mu_a - \mu_b| > 0$,

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_\mu[\tau_\delta]}{\log(1/\delta)} \leq 4T^\star_{\mathrm{kl},\beta}(\boldsymbol{\mu}) \left(1 + \sqrt{1 + \frac{\Delta_{\max}^2}{2\epsilon^2}}\right)$$

Comparison to lower bound For instances where gaps have the same order of magnitude, i.e. there exists a constant $C \geq 1$ such that $\Delta_{\max}/\Delta_{\min} \leq C$, there exists a universal constant c, such that

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_\mu[\tau_\delta]}{\log(1/\delta)} \leq c \max\left\{T^\star_{\mathrm{kl},1/2}(\boldsymbol{\mu}), C\epsilon^{-1} \sum_{a \neq a^\star} \Delta_a^{-1}\right\}$$

Comparison to DP-SE. DP-SE is a ϵ-global DP version of the successive elimination algorithm, with a sample complexity $\mathcal{O}(\sum_{a \neq a^\star} \Delta_a^{-2} + \sum_{a \neq a^\star} (\epsilon \Delta_a)^{-1})$. DP-SE too achieves (to constants) the high-privacy lower bound $T^\star_{\mathrm{TV}}(\boldsymbol{\mu})/\epsilon$, but has two drawbacks:
1. DP-SE is less adaptive than AdaP-TT, i.e. in a phase, DP-SE continues to sample arms that might already be known to be bad.
2. AdaP-TT is anytime, i.e. its sampling strategy does not depend on the risk δ.

Sample complexity lower bound

The lower bound: Let $\delta \in (0,1)$ and $\epsilon > 0$. For any δ-correct ϵ-global DP BAI strategy, we have that

$$\mathbb{E}_{\boldsymbol{\nu}}[\tau] \geq T^\star(\boldsymbol{\nu}, \epsilon) \log(1/3\delta)$$

where $(T^\star(\boldsymbol{\nu}, \epsilon))^{-1} \triangleq \sup_{\omega \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\nu})} \min \left(\sum_{a=1}^K \omega_a D_{\mathrm{KL}}(\nu_a \,\|\, \lambda_a), 6\epsilon \sum_{a=1}^K \omega_a \mathrm{TV}(\nu_a \,\|\, \lambda_a)\right)$.

Simplification:

$$T^\star(\boldsymbol{\nu}, \epsilon) \geq \max\left(T^\star_{\mathrm{KL}}(\boldsymbol{\nu}), \frac{1}{6\epsilon} T^\star_{\mathrm{TV}}(\boldsymbol{\nu})\right),$$

where $(T^\star_{\mathbf{d}}(\boldsymbol{\nu}))^{-1} \triangleq \sup_{\omega \in \Sigma_K} \inf_{\boldsymbol{\lambda} \in \mathrm{Alt}(\boldsymbol{\nu})} \sum_{a=1}^K \omega_a \mathbf{d}(\nu_a, \lambda_a)$, and $\mathbf{d}$ is either KL or TV.

$T^\star_{\mathrm{TV}}$ for Bernoulli instances: ν_a = Bernoulli(μ_a) and $\mu_1 > \mu_2 \geq \dots \geq \mu_K$. Let $\Delta_a \triangleq \mu_1 - \mu_a$ and $\Delta_{\min} \triangleq \min_{a \neq 1} \Delta_a$.

$$T^\star_{\mathrm{TV}}(\boldsymbol{\nu}) = \frac{1}{\Delta_{\min}} + \sum_{a=2}^K \frac{1}{\Delta_a}$$

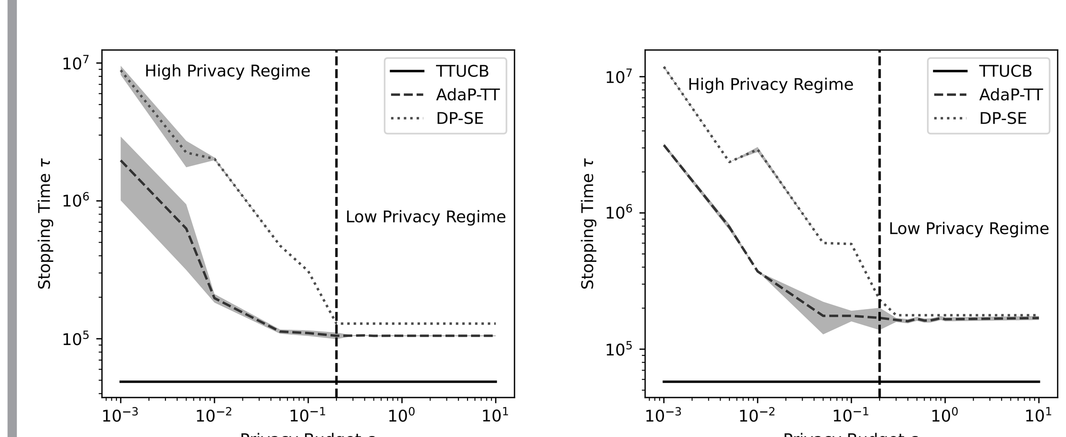
Pinsker inequality: $T^\star_{\mathrm{TV}}(\boldsymbol{\nu}) \geq \sqrt{2T^\star_{\mathrm{KL}}(\boldsymbol{\nu})}$.

Technical result of interest: Transportation lemma under ϵ-global DP. Let $\delta \in (0,1)$ and $\epsilon > 0$. Let $\boldsymbol{\nu}$ be a bandit instance and $\lambda \in \mathrm{Alt}(\boldsymbol{\nu})$. For any δ-correct ϵ-global DP BAI strategy,

$$6\epsilon \sum_{a=1}^K \mathbb{E}_{\boldsymbol{\nu},\pi}[N_a(\tau)] \,\mathrm{TV}(\nu_a \,\|\, \lambda_a) \geq \mathrm{kl}(1 - \delta, \delta),$$

$\mathrm{kl}(1 - \delta, \delta) \triangleq x \log \frac{x}{y} + (1 - x) \log \frac{1-x}{1-y}$ for $x, y \in (0,1)$.

Experimental analysis



1. AdaP-TT outperforms DP-SE.
2. The performance of AdaP-TT has two regimes: a high-privacy regime (for $\epsilon < 0.2$) and a low privacy regime (for $\epsilon > 0.2$).

Conclusion and future works

What do we achieve?

- The hardness of a BAI bandit problem with ϵ-global DP depends on a coupled effect of the privacy budget ϵ and the TV and KL characteristic times.
- In the low-privacy regime, bandits with ϵ-global DP are not harder than non-private bandits.
- Adaptive episodes with doubling, coupled with forgetting, allows adding less noise to the empirical means.
- AdaP-TT is near-optimal and enjoys good empirical performance.

What remains to be done?

- Closing the gap between the lower and upper bounds with a tighter theoretical analysis.
- Extending the analysis to other DP settings, like (ϵ, δ)-DP and Rényi-DP, or other trust models, namely local DP and shuffle DP.