Fixed-Budget Differentially Private Best Arm Identification

Zhirui Chen, P. N. Karthik, Yeow Meng Chee, and **Vincent Y. F. Tan**

National University of Singapore

Feb 20th, 2024

Best arm identification in linear bandits

- Arm set $[K] = \{1, \ldots, K\}$ and their feature vectors $\{\boldsymbol{a}_i\}_{i=1}^K \subseteq \mathbb{R}^d$.

Best arm identification in linear bandits

- Arm set $[K] = \{1, \ldots, K\}$ and their feature vectors $\{\boldsymbol{a}_i\}_{i=1}^K \subseteq \mathbb{R}^d$.
- Arm $i \in [K]$ is associated with reward distribution ν_i supported on $[0, 1]$, and its mean is

$$\mu_i = \langle \boldsymbol{a}_i, \boldsymbol{\theta}^* \rangle,$$

  where $\boldsymbol{\theta}^* \in \mathbb{R}^d$ is unknown parameter.

Best arm identification in linear bandits

- Arm set $[K] = \{1, \ldots, K\}$ and their feature vectors $\{\boldsymbol{a}_i\}_{i=1}^K \subseteq \mathbb{R}^d$.
- Arm $i \in [K]$ is associated with reward distribution ν_i supported on $[0, 1]$, and its mean is

$$\mu_i = \langle \boldsymbol{a}_i, \boldsymbol{\theta}^* \rangle,$$

  where $\boldsymbol{\theta}^* \in \mathbb{R}^d$ is unknown parameter.
- Given fixed-budget $T > 0$, for each time step $t = 1, \ldots, T$, the agent pulls arm $A_t \in [K]$ and obtains reward $X_t := X_{A_t, N_{A_t,t}}$, where

$$N_{i,t} = \sum_{s=1}^{t} 1_{\{A_s = i\}}$$

  is the number of times arm i is pulled up to time t, and $X_{i,n} \sim \nu_i$ denotes the reward obtained on the n^{th} pull of arm i.

Best arm identification in linear bandits

- After T time steps, the agent identifies $\hat{I}_T \in [K]$ as the best arm.

Best arm identification in linear bandits

- After T time steps, the agent identifies $\hat{I}_T \in [K]$ as the best arm.
- The objective is to identify the best arm with probability as high as possible, i.e., $\mathbb{P}(\hat{I}_T = i^*)$ is as large as possible, where

$$i^* = \arg\max_{i \in [K]} \mu_i$$

is denoted as the best arm and $\mu_i \in \mathbb{R}$ is the mean of distribution ν_i.

Best arm identification in linear bandits

- After T time steps, the agent identifies $\hat{I}_T \in [K]$ as the best arm.
- The objective is to identify the best arm with probability as high as possible, i.e., $\mathbb{P}(\hat{I}_T = i^*)$ is as large as possible, where

$$i^* = \arg\max_{i \in [K]} \mu_i$$

  is denoted as the best arm and $\mu_i \in \mathbb{R}$ is the mean of distribution ν_i.
- We assume the best arm is unique.

Problem Statement: Differential Privacy

- Let $\mathcal{X} := \{\boldsymbol{x} = (x_{i,t})_{i\in[K],t\in[T]}\} \subseteq [0,1]^{KT}$ denote the collection of all possible rewards outcomes from the arms.

Problem Statement: Differential Privacy

- Let $\mathcal{X} := \{\boldsymbol{x} = (x_{i,t})_{i\in[K],t\in[T]}\} \subseteq [0,1]^{KT}$ denote the collection of all possible rewards outcomes from the arms.
- Any sequential arm selection *policy* of the decision maker takes inputs from $\mathcal{X}$ and produces $(A_1, \ldots, A_T, \hat{I}_T) \in [K]^{T+1}$ as outputs in the following manner: for an input $\boldsymbol{x} = (x_{i,t}) \in \mathcal{X}$,

$$
\begin{aligned}
\text{Output at time } t = 1 &: \ A_1 = A_1, \\
\text{Output at time } t = 2 &: \ A_2 = A_2(A_1, x_{A_1, N_{A_1,1}}) \\
\text{Output at time } t = 3 &: \ A_3 = A_3(A_1, x_{A_1, N_{A_1,1}}, A_2, x_{A_2, N_{A_2,2}}) \\
&\ \vdots \\
\text{Output at time } t = T &: \ A_T = A_T(A_1, x_{A_1, N_{A_1,1}}, \ldots, A_{T-1}, x_{N_{A_{T-1}}, T-1}) \\
\text{Terminal output} &: \ \hat{I}_T = \hat{I}_T(A_1, x_{A_1, N_{A_1,1}}, \ldots, A_T, x_{N_{A_{T-1}}, T}).
\end{aligned}
$$

We say that $\boldsymbol{x} = (x_{i,n})$ and $\boldsymbol{x}' = (x'_{i,n})$ are *neighbouring* if they differ in exactly one location, i.e., there exists a unique (exactly one) $(i, n) \in [K] \times [T]$ such that

$$x_{i,n} \neq x'_{i,n} \quad \text{and} \quad x_{j,s} = x'_{j,s} \quad \text{for all} \quad (j, s) \neq (i, n).$$

Differential Privacy

We say that $\boldsymbol{x} = (x_{i,n})$ and $\boldsymbol{x}' = (x'_{i,n})$ are *neighbouring* if they differ in exactly one location, i.e., there exists a unique (exactly one) $(i, n) \in [K] \times [T]$ such that

$$x_{i,n} \neq x'_{i,n} \quad \text{and} \quad x_{j,s} = x'_{j,s} \quad \text{for all} \quad (j, s) \neq (i, n).$$

Definition: Differential Privacy

Given any $\varepsilon > 0$, a randomised policy $\mathcal{M} : \mathcal{X} \to [K]^{T+1}$ satisfies *ε-differential privacy* if, for any pair of neighbouring $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$,

$$\mathbb{P}^{\mathcal{M}}(\mathcal{M}(\boldsymbol{x}) \in \mathcal{S}) \leq e^{\varepsilon}\, \mathbb{P}^{\mathcal{M}}(\mathcal{M}(\boldsymbol{x}') \in \mathcal{S}) \quad \forall\, \mathcal{S} \subset [K]^{T+1}.$$

Methodology: Overview

- To meet the ε-DP guarantee, our approach is to add Laplacian noise to the empirical mean reward of each arm.

Methodology: Overview

- To meet the ε-DP guarantee, our approach is to add Laplacian noise to the empirical mean reward of each arm.
- The magnitude of the noise is inversely proportional to the product of ε and the number of times the arm is pulled. In particular, we choose

$$\widetilde{\xi}_i^{(p)} \sim \mathrm{Lap}\left(\frac{1}{(N_{i,T_p} - N_{i,T_{p-1}})\varepsilon}\right)$$

where T_p is the time step at the start of phase p.

Methodology: Overview

- To meet the ε-DP guarantee, our approach is to add Laplacian noise to the empirical mean reward of each arm.
- The magnitude of the noise is inversely proportional to the product of ε and the number of times the arm is pulled. In particular, we choose

$$\widetilde{\xi}_i^{(p)} \sim \mathrm{Lap}\left(\frac{1}{(N_{i,T_p} - N_{i,T_{p-1}})\varepsilon}\right)$$

  where T_p is the time step at the start of phase p.

- Intuitively, to minimize the maximum Laplacian noise that is added (so as to minimize the failure probability of identifying the best arm), we aim to balance the number of pulls for each arm.

Fix $d' \in \mathbb{N}$. For any set $\mathcal{S} \subset \mathbb{R}^{d'}$ with $|\mathcal{S}| = d'$ vectors, each of length d', let $\textsc{Det}(\mathcal{S})$ to denote the absolute value of the determinant of the $d' \times d'$ matrix formed by stacking the vectors in $\mathcal{S}$ as the columns of the matrix.

Fix $d' \in \mathbb{N}$. For any set $\mathcal{S} \subset \mathbb{R}^{d'}$ with $|\mathcal{S}| = d'$ vectors, each of length d', let $\textsc{Det}(\mathcal{S})$ to denote the absolute value of the determinant of the $d' \times d'$ matrix formed by stacking the vectors in $\mathcal{S}$ as the columns of the matrix.

Definition: Max-Det collection

Fix $d' \in \mathbb{N}$. Given any finite set $\mathcal{A} \subset \mathbb{R}^{d'}$ with $|\mathcal{A}| \geq d'$, we say $\mathcal{B} \subset \mathcal{A}$ with $|\mathcal{B}| = d'$ is a $\textsc{Max-Det}$ *collection of* $\mathcal{A}$ if

$$\textsc{Det}(\mathcal{B}) \geq \textsc{Det}(\mathcal{B}') \quad \text{for all } \mathcal{B}' \subset \mathcal{A} \text{ with } |\mathcal{B}'| = d'.$$

Methodology: Max-Det collection

- Example: Let $d' = 2$ and $\mathcal{S}$ be the set of vectors $\left\{ \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \end{bmatrix} \right\}$

Methodology: Max-Det collection

- Example: Let $d' = 2$ and $\mathcal{S}$ be the set of vectors $\left\{ \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \end{bmatrix} \right\}$
- The subsets of vectors of size $d = 2$ are

$$\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix} \right\}, \quad \mathcal{B}_2 = \left\{ \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \end{bmatrix} \right\}, \quad \mathcal{B}_3 = \left\{ \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \end{bmatrix} \right\}.$$

Methodology: Max-Det collection

- Example: Let $d' = 2$ and $\mathcal{S}$ be the set of vectors $\left\{ \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \end{bmatrix} \right\}$
- The subsets of vectors of size $d = 2$ are

$$\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix} \right\}, \quad \mathcal{B}_2 = \left\{ \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \end{bmatrix} \right\}, \quad \mathcal{B}_3 = \left\{ \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \end{bmatrix} \right\}.$$

- The absolute values of the determinants are

$$\mathrm{Det}(\mathcal{B}_1) = \left| \det \left(\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} \right) \right| = 3, \quad \mathrm{Det}(\mathcal{B}_2) = 2, \quad \mathrm{Det}(\mathcal{B}_3) = 2.$$

Methodology: Max-Det collection

- Example: Let $d' = 2$ and $\mathcal{S}$ be the set of vectors $\left\{ \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \end{bmatrix} \right\}$
- The subsets of vectors of size $d = 2$ are

$$\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix} \right\}, \quad \mathcal{B}_2 = \left\{ \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \end{bmatrix} \right\}, \quad \mathcal{B}_3 = \left\{ \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \end{bmatrix} \right\}.$$

- The absolute values of the determinants are

$$\mathrm{Det}(\mathcal{B}_1) = \left| \det\left(\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} \right) \right| = 3, \quad \mathrm{Det}(\mathcal{B}_2) = 2, \quad \mathrm{Det}(\mathcal{B}_3) = 2.$$

- So, the $\textsc{Max-Det}$ collection is $\mathcal{B}_1$.

Methodology: DP-BAI Policy

- Our policy for Differentially Private Best Arm Identification, called DP-BAI, based on the idea of successive elimination (SE) of arms, operates over a total of M *phases*, where $M = \Theta(\log d)$.

Methodology: DP-BAI Policy

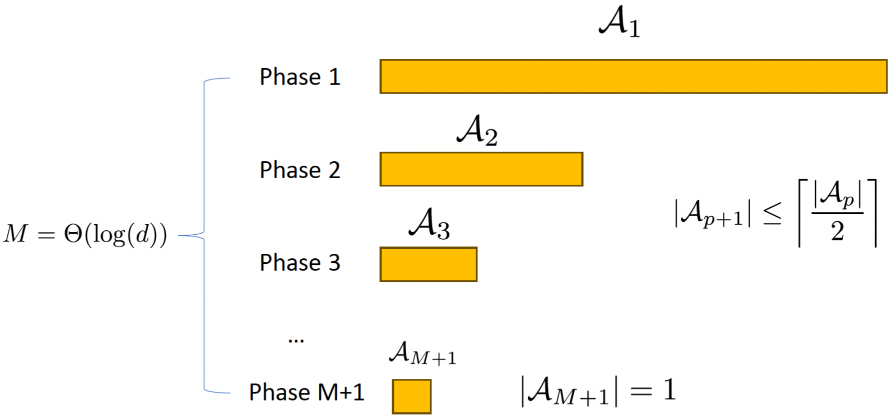
- Our policy for Differentially Private Best Arm Identification, called DP-BAI, based on the idea of successive elimination (SE) of arms, operates over a total of M *phases*, where $M = \Theta(\log d)$.
- In each phase $p \in [M]$, the agent maintains an *active set* $\mathcal{A}_p$ of arms which are potential contenders for emerging as the best arm. The policy ensures that with high probability, the true best arm lies within the active set in each phase.

Methodology: DP-BAI Policy

- Our policy for Differentially Private Best Arm Identification, called DP-BAI, based on the idea of successive elimination (SE) of arms, operates over a total of M *phases*, where $M = \Theta(\log d)$.
- In each phase $p \in [M]$, the agent maintains an *active set* $\mathcal{A}_p$ of arms which are potential contenders for emerging as the best arm. The policy ensures that with high probability, the true best arm lies within the active set in each phase.
- The cardinality $|\mathcal{A}_p|$ is set to s_p, where s_p is determined in the initialisation stage, and the policy ensures that

$$s_{p+1} \leq \left\lceil \frac{s_p}{2} \right\rceil \quad \text{and} \quad s_{M+1} = 1.$$

Methodology: DP-BAI Policy



Dimensionality Reduction:

- At the beginning of each phase p, suppose that $d_p := \dim(\mathrm{span}\{\boldsymbol{a}_i^{(p-1)} : i \in \mathcal{A}_p\})$, where $\boldsymbol{a}_i^{(0)}$ is initialised to be $\boldsymbol{a}_i$.

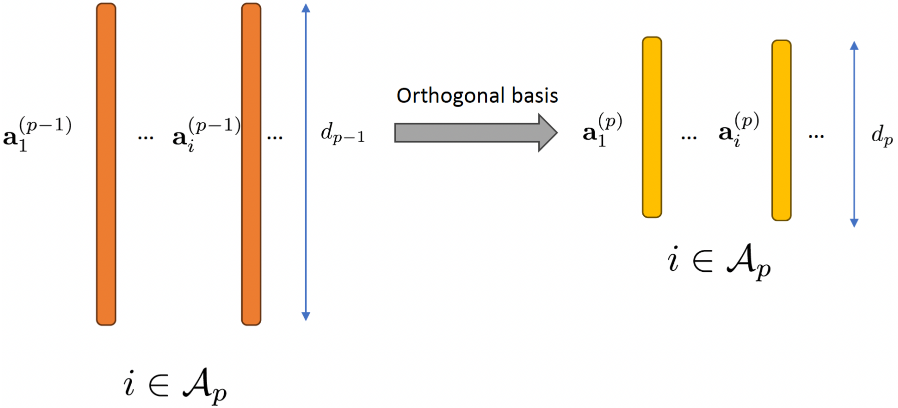
Dimensionality Reduction:

- At the beginning of each phase p, suppose that $d_p := \dim(\mathrm{span}\{\boldsymbol{a}_i^{(p-1)} : i \in \mathcal{A}_p\})$, where $\boldsymbol{a}_i^{(0)}$ is initialised to be $\boldsymbol{a}_i$.
- The agent chooses an arbitrary orthogonal basis $\mathcal{U}_p = (\boldsymbol{u}_1^{(p)}, \ldots, \boldsymbol{u}_{d_p}^{(p)})$ for $\mathrm{span}\{\boldsymbol{a}_i^{(p-1)} : i \in \mathcal{A}_p\}$, and obtains a new set of vectors $\{\boldsymbol{a}_i^{(p)} : i \in \mathcal{A}_p\}$ via

$$\boldsymbol{a}_i^{(p)} := [\boldsymbol{a}_i^{(p-1)}]_{\mathcal{U}_p},$$

where $[\boldsymbol{v}]_{\mathcal{U}_p}$ denotes the coordinates of $\boldsymbol{v}$ with respect to $\mathcal{U}_p$.

Methodology: DP-BAI Policy



Methodology: DP-BAI Policy

Sampling strategy:

- There are two cases in our sampling strategy. Recall that $s_p = |\mathcal{A}_p|$.

Methodology: DP-BAI Policy

Sampling strategy:

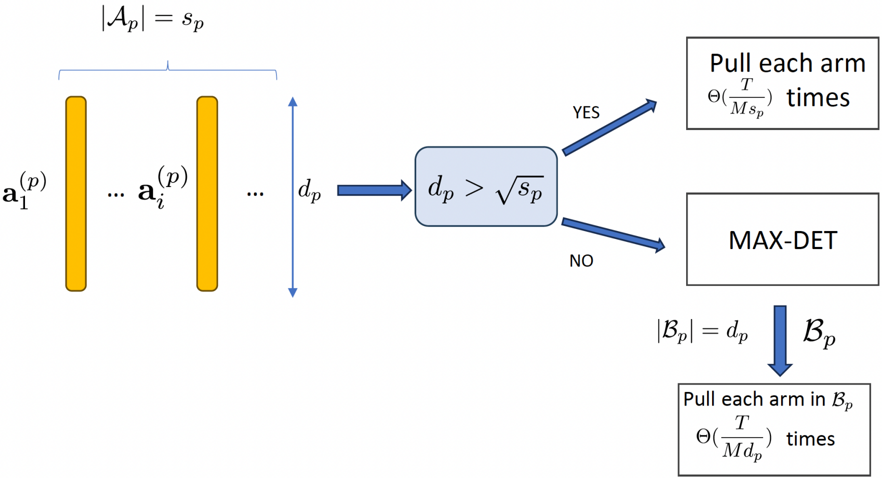
- There are two cases in our sampling strategy. Recall that $s_p = |\mathcal{A}_p|$.
- In case of $d_p > \sqrt{s_p}$ (number of arms remaining is small), the agent pulls each arm in $\mathcal{A}_p$ uniformly randomly for $\Theta\left(\frac{T}{Ms_p}\right)$ times.

Methodology: DP-BAI Policy

Sampling strategy:

- There are two cases in our sampling strategy. Recall that $s_p = |\mathcal{A}_p|$.
- In case of $d_p > \sqrt{s_p}$ (number of arms remaining is small), the agent pulls each arm in $\mathcal{A}_p$ uniformly randomly for $\Theta\left(\frac{T}{Ms_p}\right)$ times.
- In the case of $d_p \le \sqrt{s_p}$ (number of arms remaining is large), the agent constructs a Max-Det collection $\mathcal{B}_p \subset \mathcal{A}_p$ consisting of $|\mathcal{B}_p| = d_p$ arms, and pulls each arm $i \in \mathcal{B}_p$ for $\Theta\left(\frac{T}{Md_p}\right)$ many times.

Methodology: DP-BAI Policy



Methodology: DP-BAI Policy

Private empirical mean:

- For each arm $i \in \mathcal{A}_p$ that was pulled at least once in phase p, the agent computes the empirical means via

$$\hat{\mu}_i^{(p)} = \frac{1}{N_{i,T_p} - N_{i,T_{p-1}}} \sum_{s=N_{i,T_{p-1}}+1}^{N_{i,T_p}} X_{i,s},$$

Private empirical mean:

- For each arm $i \in \mathcal{A}_p$ that was pulled at least once in phase p, the agent computes the empirical means via

$$\hat{\mu}_i^{(p)} = \frac{1}{N_{i,T_p} - N_{i,T_{p-1}}} \sum_{s=N_{i,T_{p-1}}+1}^{N_{i,T_p}} X_{i,s},$$

- Subsequently the agent generates its private empirical mean $\widetilde{\mu}_i^{(p)}$ via

$$\widetilde{\mu}_i^{(p)} = \hat{\mu}_i^{(p)} + \widetilde{\xi}_i^{(p)},$$

where $\widetilde{\xi}_i^{(p)} \sim \mathrm{Lap}\left(\frac{1}{(N_{i,T_p} - N_{i,T_{p-1}})\varepsilon}\right)$ is independent of the arm pulls and arm rewards.

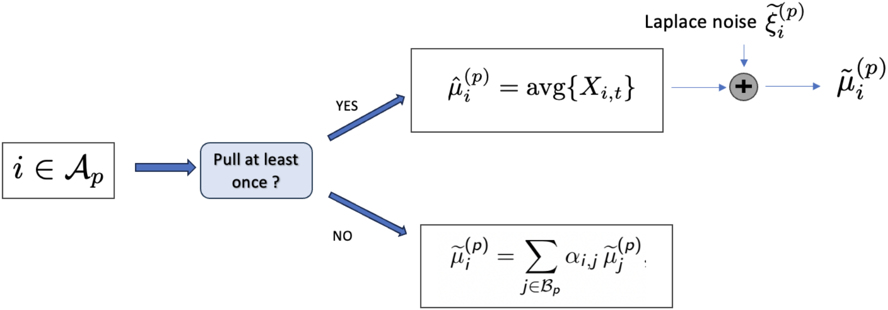
Private empirical mean:
For $i \in \mathcal{A}_p$ that was not pulled in phase p, the agent computes its corresponding private empirical mean via

$$\widetilde{\mu}_i^{(p)} = \sum_{j \in \mathcal{B}_p} \alpha_{i,j} \, \widetilde{\mu}_j^{(p)},$$

where $(\alpha_{i,j})_{j \in \mathcal{B}_p}$ is the unique set of coefficients such that

$$\boldsymbol{a}_i^{(p)} = \sum_{j \in \mathcal{B}_p} \alpha_{i,j} \, \boldsymbol{a}_j^{(p)}.$$

Methodology: DP-BAI Policy



Methodology: DP-BAI Policy

Recommendation rule:

- At the end of phase p, the policy retains only the top s_{p+1} arms with the largest private empirical means.

Recommendation rule:

- At the end of phase p, the policy retains only the top s_{p+1} arms with the largest private empirical means.
- At the end of the Mth phase, the policy returns the only arm left in $\mathcal{A}_{M+1}$ as the best arm.

Theoretical Result: DP Constraint

Privacy Guarantee for $\mathrm{DP\text{-}BAI}$

The $\mathrm{DP\text{-}BAI}$ policy with privacy and budget parameters (ε, T) satisfies the ε-DP constraint, i.e., for any pair of neighbouring $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$,

$$\mathbb{P}^{\Pi_{\mathrm{DP\text{-}BAI}}}(\Pi_{\mathrm{DP\text{-}BAI}}(\boldsymbol{x}) \in \mathcal{S}) \leq e^{\varepsilon}\, \mathbb{P}^{\Pi_{\mathrm{DP\text{-}BAI}}}(\Pi_{\mathrm{DP\text{-}BAI}}(\boldsymbol{x}') \in \mathcal{S}) \quad \forall\, \mathcal{S} \subset [K]^{T+1}.$$

Theoretical Result: Hardness Quantity

- Let $\Delta_i := \mu_{i^*(v)} - \mu_i$ denote the sub-optimality gap of arm $i \in [K]$.

Theoretical Result: Hardness Quantity

- Let $\Delta_i := \mu_{i^*(v)} - \mu_i$ denote the sub-optimality gap of arm $i \in [K]$.
- Let $(l_1, \ldots, l_K)$ be a permutation of $[K]$ such that

$$\Delta_{l_1} \leq \Delta_{l_2} \leq \ldots \leq \Delta_{l_K},$$

and let $\Delta_{(i)} := \Delta_{l_i}$ for all $i \in [K]$ be the ordered gaps.

Theoretical Result: Hardness Quantity

- Let $\Delta_i := \mu_{i^*(v)} - \mu_i$ denote the sub-optimality gap of arm $i \in [K]$.
- Let $(l_1, \ldots, l_K)$ be a permutation of $[K]$ such that

$$\Delta_{l_1} \leq \Delta_{l_2} \leq \ldots \leq \Delta_{l_K},$$

  and let $\Delta_{(i)} := \Delta_{l_i}$ for all $i \in [K]$ be the ordered gaps.
- The *hardness* of an instance $v = ((\boldsymbol{a}_i)_{i\in[K]}, (\nu_i)_{i\in[K]}, \boldsymbol{\theta}^*, \varepsilon)$ is defined as

$$H(v) := H_{\mathrm{BAI}}(v) + H_{\mathrm{pri}}(v),$$

  where

$$H_{\mathrm{BAI}}(v) := \max_{2\leq i\leq (d^2\wedge K)} \frac{i}{\Delta_{(i)}^2} \quad \text{and} \quad H_{\mathrm{pri}}(v) := \frac{1}{\varepsilon} \cdot \max_{2\leq i\leq (d^2\wedge K)} \frac{i}{\Delta_{(i)}}.$$

Theoretical Result: Upper Bound

Error Probability Guarantee for DP-BAI

Fix instance v and let $i^*(v)$ denote the unique best arm. For all sufficiently large T, the error probability of $\Pi_{\text{DP-BAI}}$ with budget T and privacy parameter ε satisfies

$$\mathbb{P}_v^{\Pi_{\text{DP-BAI}}}(\hat{I}_T \neq i^*(v)) \leq \exp\left(-\frac{T}{65\,M\,H}\right),$$

where $\mathbb{P}_v^{\Pi_{\text{DP-BAI}}}$ denotes the probability measure induced by $\Pi_{\text{DP-BAI}}$ under the instance v.

Theoretical Result: Upper Bound

Error Probability Guarantee for DP-BAI

Fix instance v and let $i^*(v)$ denote the unique best arm. For all sufficiently large T, the error probability of $\Pi_{\text{DP-BAI}}$ with budget T and privacy parameter ε satisfies

$$\mathbb{P}_v^{\Pi_{\text{DP-BAI}}}(\hat{I}_T \neq i^*(v)) \leq \exp\left(-\frac{T}{65\, M\, H}\right),$$

where $\mathbb{P}_v^{\Pi_{\text{DP-BAI}}}$ denotes the probability measure induced by $\Pi_{\text{DP-BAI}}$ under the instance v.

Because $M = \Theta(\log d)$, the upper bound implies that

$$\mathbb{P}_v^{\Pi_{\text{DP-BAI}}}(\hat{I}_T \neq i^*(v)) = \exp\left(-\Omega\left(\frac{T}{H \log d}\right)\right).$$

Theoretical Result: Minimax Lower Bound

Defintion: Consistent policy

A policy π for fixed-budget BAI with the ε-DP constraint is said to be *consistent* if

$$\lim_{T \to +\infty} \mathbb{P}_v^\pi\big(\hat{I}_T \neq i^*(v)\big) = 0, \quad \forall\, v \in \mathcal{P}.$$

Theoretical Result: Minimax Lower Bound

Defintion: Consistent policy

A policy π for fixed-budget BAI with the ε-DP constraint is said to be *consistent* if

$$\lim_{T \to +\infty} \mathbb{P}_v^{\pi}\big(\hat{I}_T \neq i^*(v)\big) = 0, \quad \forall\, v \in \mathcal{P}.$$

Minimax Lower Bound

Fix any $\beta_1, \beta_2, \beta_3 \in [0,1]$ with $\beta_1 + \beta_2 + \beta_3 < 3$ and a consistent policy π.For all sufficiently large T, there exists an instance $v \in \mathcal{P}$ such that

$$\mathbb{P}_v^{\pi}\big(\hat{I}_T \neq i^*(v)\big) > \exp\left(- \Omega\left(\frac{T}{(\log d)^{\beta_1} \left(H_{\mathrm{BAI}}(v)^{\beta_2} + H_{\mathrm{pri}}(v)^{\beta_3}\right)} \right)\right).$$

Theoretical Result: Minimax Lower Bound

- Lower bound $\implies$ for any $\beta \in [0, 1)$, there does not exist a consistent policy π with an upper bound on its error probability assuming any one of the following forms for all instances $v \in \mathcal{P}$:

Theoretical Result: Minimax Lower Bound

- Lower bound $\implies$ for any $\beta \in [0, 1)$, there does not exist a consistent policy π with an upper bound on its error probability assuming any one of the following forms for all instances $v \in \mathcal{P}$:
  - $\exp\left(-\Omega\left(\frac{T}{(\log d)^{\beta}(H_{\mathrm{BAI}}(v) + H_{\mathrm{pri}}(v))}\right)\right)$,

Theoretical Result: Minimax Lower Bound

- Lower bound $\implies$ for any $\beta \in [0, 1)$, there does not exist a consistent policy π with an upper bound on its error probability assuming any one of the following forms for all instances $v \in \mathcal{P}$:
  - $\exp\left(-\Omega\left(\frac{T}{(\log d)^{\beta}(H_{\mathrm{BAI}}(v) + H_{\mathrm{pri}}(v))}\right)\right)$,
  - $\exp\left(-\Omega\left(\frac{T}{(\log d)(H_{\mathrm{BAI}}(v)^{\beta} + H_{\mathrm{pri}}(v))}\right)\right)$,

Theoretical Result: Minimax Lower Bound

- Lower bound $\implies$ for any $\beta \in [0, 1)$, there does not exist a consistent policy π with an upper bound on its error probability assuming any one of the following forms for all instances $v \in \mathcal{P}$:
  - $\exp\left(-\Omega\left(\frac{T}{(\log d)^{\beta}(H_{\mathrm{BAI}}(v) + H_{\mathrm{pri}}(v))}\right)\right)$,
  - $\exp\left(-\Omega\left(\frac{T}{(\log d)(H_{\mathrm{BAI}}(v)^{\beta} + H_{\mathrm{pri}}(v))}\right)\right)$,
  - $\exp\left(-\Omega\left(\frac{T}{(\log d)(H_{\mathrm{BAI}}(v) + H_{\mathrm{pri}}(v)^{\beta})}\right)\right)$.

Theoretical Result: Minimax Lower Bound

- Lower bound $\implies$ for any $\beta \in [0,1)$, there does not exist a consistent policy π with an upper bound on its error probability assuming any one of the following forms for all instances $v \in \mathcal{P}$:
  - $\exp\left(-\Omega\left(\frac{T}{(\log d)^{\beta}(H_{\mathrm{BAI}}(v)+H_{\mathrm{pri}}(v))}\right)\right)$,
  - $\exp\left(-\Omega\left(\frac{T}{(\log d)(H_{\mathrm{BAI}}(v)^{\beta}+H_{\mathrm{pri}}(v))}\right)\right)$,
  - $\exp\left(-\Omega\left(\frac{T}{(\log d)(H_{\mathrm{BAI}}(v)+H_{\mathrm{pri}}(v)^{\beta})}\right)\right)$.
- In this sense, the dependencies of the upper bound on $\log d$, $H_{\mathrm{BAI}}(v)$, and $H_{\mathrm{pri}}(v)$ are "tight".

Numerical Study

- We conduct a numerical study on synthetic data, and compare DP-BAI with BASELINE, an algorithm which follows DP-BAI but does not utilize our MAX-DET collection idea.

Numerical Study

- We conduct a numerical study on synthetic data, and compare DP-BAI with BASELINE, an algorithm which follows DP-BAI but does not utilize our MAX-DET collection idea.
- In addition, we compare DP-BAI to the state-of-the-art OD-LINBAI (Yang and Tan, 2022) algorithm for fixed-budget best arm identification.

Numerical Study

- We conduct a numerical study on synthetic data, and compare DP-BAI with BASELINE, an algorithm which follows DP-BAI but does not utilize our MAX-DET collection idea.
- In addition, we compare DP-BAI to the state-of-the-art OD-LinBAI (Yang and Tan, 2022) algorithm for fixed-budget best arm identification.
- OD-LinBAI is a non-private algorithm and serves as an upper bound in performance (in terms of the error probability) of our algorithm.

Numerical Study

- We conduct a numerical study on synthetic data, and compare DP-BAI with BASELINE, an algorithm which follows DP-BAI but does not utilize our MAX-DET collection idea.
- In addition, we compare DP-BAI to the state-of-the-art OD-LINBAI (Yang and Tan, 2022) algorithm for fixed-budget best arm identification.
- OD-LINBAI is a non-private algorithm and serves as an upper bound in performance (in terms of the error probability) of our algorithm.
- Also, we consider an ε-DP version of OD-LINBAI which we call DP-OD by using a privatization idea of Shariff and Sheffet (2018).

Numerical Study

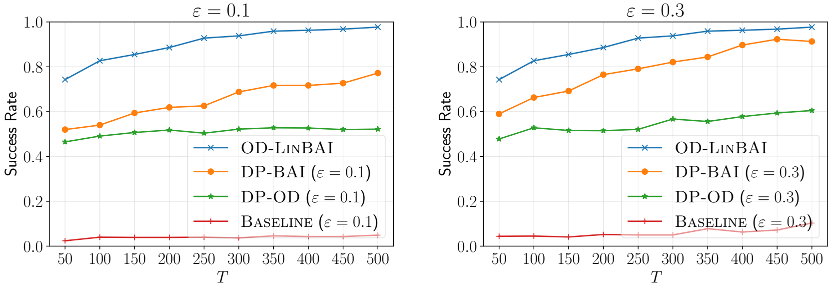


Figure 1: Comparison of DP-BAI to Baseline, OD-LinBAI and DP-OD for different values of T.

References

Audibert, J.-Y., Bubeck, S., and Munos, R. (2010). Best arm identification in multi-armed bandits. In *COLT*, pages 41–53.

Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256.

Carpentier, A. and Locatelli, A. (2016). Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Proceedings of the 29th Conference on Learning Theory*, pages 590–604.

Jamieson, K. and Nowak, R. (2014). Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *Proceedings of the 48th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6.

Kalyanakrishnan, S., Tewari, A., Auer, P., and Stone, P. (2012). Pac subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th International Conference on Machine Learning*, pages 655–662.

Shariff, R. and Sheffet, O. (2018). Differentially private contextual linear bandits. In *Advances in Neural Information Processing Systems*, volume 31, page 4301–4311.

Yang, J. and Tan, V. (2022). Minimax optimal fixed-budget best arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 35:12253–12266.