Optimal Best Arm Identification under Differential Privacy

Marc Jourdan*
EPFL, Lausanne, Switzerland
marc.jourdan@epfl.ch

Achraf Azize*
FairPlay Joint Team, CREST, ENSAE Paris
achraf.azize@ensae.fr

Abstract

Best Arm Identification (BAI) algorithms are deployed in data-sensitive applications, such as adaptive clinical trials or user studies. Driven by the privacy concerns of these applications, we study the problem of fixed-confidence BAI under global Differential Privacy (DP) for Bernoulli distributions. While numerous asymptotically optimal BAI algorithms exist in the non-private setting, a significant gap remains between the best lower and upper bounds in the global DP setting. This work reduces this gap to a small multiplicative constant, for any privacy budget ϵ. First, we provide a tighter lower bound on the expected sample complexity of any δ-correct and ϵ-global DP strategy. Our lower bound replaces the Kullback–Leibler (KL) divergence in the transportation cost used by the non-private characteristic time with a new information-theoretic quantity that optimally trades off between the KL divergence and the Total Variation distance scaled by ϵ. Second, we introduce a stopping rule based on these transportation costs and a private estimator of the means computed using an arm-dependent geometric batching. En route to proving the correctness of our stopping rule, we derive concentration results of independent interest for the Laplace distribution and for the sum of Bernoulli and Laplace distributions. Third, we propose a Top Two sampling rule based on these transportation costs. For any budget ϵ, we show an asymptotic upper bound on its expected sample complexity that matches our lower bound to a multiplicative constant smaller than 8. Our algorithm outperforms existing δ-correct and ϵ-global DP BAI algorithms for different values of ϵ.

1 Introduction

The stochastic Multi-Armed Bandit (MAB) is an interactive sequential decision-making model [Bubeck et al., 2012; Lattimore and Szepesvári, 2020], introduced by William R. Thompson [Thompson, 1933]. Thompson's motivation for studying MABs is to design clinical trials that adapt treatment allocations on the fly as the medicines appear more or less effective. Specifically, in MABs, a learner interacts with $K \in \mathbb{N}$ unknown probability distributions, referred to as *arms*. In clinical trials, the arms are the candidate medicines, while the observations are patient reactions, 1 if the patient is cured and 0 otherwise. The learner aims to identify the arm with the highest average efficiency, i.e., the medicine that cures most patients in expectation. Given a fixed error $\delta \in (0, 1)$, Best Arm Identification (BAI) [Audibert and Bubeck, 2010b; Jamieson and Nowak, 2014] algorithms in the fixed confidence setting [Even-Dar et al., 2006; Gabillon et al., 2012; Garivier and Kaufmann, 2016] suggest a candidate answer that coincides with the optimal arm with probability more than $1 - \delta$, while using as few samples as possible.

BAI algorithms have been increasingly deployed in data-sensitive applications, such as adaptive clinical trials [Thompson, 1933; Robbins, 1952; Aziz et al., 2021], pandemic mitigation [Libin et al., 2019],

*Equal Contribution

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

user studies [Losada et al., 2022], crowdsourcing [Zhou et al., 2014], online advertisement [Chen et al., 2014], hyperparameter tuning [Li et al., 2017], and communication networks [Lindståhl et al., 2022], to name a few. Due to the adaptive nature of these procedures, critical data privacy concerns are raised [Tucker et al., 2016], as exemplified by the adaptive dose finding trial. For each new patient n, a physician chooses a dose level $a_n \in [K] := \{1, \cdots, K\}$ based on previous observations, and collects a binary observation measuring the effect of the selected dose on the patient. Crucially, the patients' reactions might reveal information regarding their health. Subsequently, these outcomes will guide the physician's decision for future patients. Eventually, the physician adaptively decides to stop the trial and recommends a dose $\hat{a}_{\tau_\delta}$ after collecting τ_δ samples, referred to as *sample complexity*. Even if those outcomes are kept secret, the experimental findings and protocol are detailed thoroughly to the health authorities. This report contains the sequence of chosen dose levels $(a_n)_{n \leq \tau_\delta}$ and the recommended dose level $\hat{a}_{\tau_\delta}$, both indirectly leaking information regarding the patients involved in the trial. This example underscores the need for privacy-preserving fixed-confidence BAI algorithms.

We adopt the Differential Privacy (DP) framework [Dwork and Roth, 2014], which bounds the influence of any single data point. Given a privacy budget ϵ, we consider the ϵ-global DP constraint that assumes the existence of a trusted curator (e.g., the physician running the clinical trial), who observes the outcomes and ensures privacy when publishing these findings. While ϵ-global DP is well-studied in regret minimization [Mishra and Thakurta, 2015; Azize and Basu, 2022; Azize et al., 2025], its impact on fixed-confidence BAI is less understood [Sajed and Sheffet, 2019; Kalogerias et al., 2021]. A significant gap remains between the existing lower and upper bounds [Azize et al., 2023, 2024]. This paper reduces this gap to a small constant for any privacy budget ϵ. Appendix C.1 contains a detailed literature review.

Contributions. Our contributions for fixed-confidence BAI under ϵ-global DP are threefold.

1. Lower bound under global DP. We derive a novel information-theoretic lower bound on the expected sample complexity of any δ-correct and ϵ-global DP BAI algorithms (Theorem 2). Our lower bound replaces the Kullback-Leibler (KL) divergence in the transportation cost of the non-private characteristic time with an information-theoretic quantity d_ϵ (Eq. (1)) that smoothly interpolates between the KL divergence and the Total Variation (TV) distance scaled by ϵ.

2. Private estimator and Generalized Likelihood Ratio (GLR) stopping rule. We introduce a private estimator using arm-dependent geometric batching without forgetting and a GLR stopping rule based on the d_ϵ refined transportation costs. Its correctness (Theorem 6) required novel tails concentration results for Laplace distributions and the sum of Bernoulli and Laplace distributions, which could be of independent interest.

3. Asymptotically optimal algorithm. We propose a new Top Two sampling rule (DP-TT, Algorithm 1) based on the d_ϵ-transportation costs suggested by our lower bound. We show that the asymptotic expected sample complexity of DP-TT matches our lower bound for any privacy budget ϵ up to a constant smaller than 8 (Theorem 7). DP-TT outperforms all the other δ-correct ϵ-global DP BAI algorithms on all tested instances and all ϵ.

2 Background: Best Arm Identification under Differential Privacy

In this section, we present the Best Arm Identification (BAI) under fixed confidence problem [Garivier and Kaufmann, 2016], introduce the Differential Privacy (DP) [Dwork et al., 2006] constraint, and finally extend DP to BAI algorithms.

BAI under Fixed Confidence. Let $\mathcal{F} := \{\mathrm{Ber}(p) \mid p \in (0,1)\}$ be the set of Bernoulli distributions. A bandit *instance* $\boldsymbol{\nu} := (\nu_a)_{a \in [K]} \in \mathcal{F}^K$ is characterized by its means $\mu := (\mu_a)_{a \in [K]} \in (0,1)^K$. The *best* (optimal) arm $a^\star$ is assumed to be unique, i.e., $a^\star(\boldsymbol{\nu}) = a^\star(\mu) := \arg\max_{a \in [K]} \mu_a = \{a^\star\}$. Let $\delta \in (0,1)$ be the risk parameter. A fixed confidence BAI algorithm π specifies three rules that rely on previously observed samples and some exogenous randomness. The *sampling rule* determines the next arm to pull $a_n \in [K]$ for which $X_{n,a_n} \sim \nu_{a_n}$ is observed. The *recommendation rule* recommends a *candidate* arm $\tilde{a} \in [K]$. The *stopping rule* decides when to stop collecting additional samples and output the current candidate arm. The stopping time $\tau_{\epsilon,\delta}$ is the *sample complexity*. Let $\mathbb{P}_{\boldsymbol{\nu}\pi}$ and $\mathbb{E}_{\boldsymbol{\nu}\pi}$ denote the probability and expectation taken over the randomness of the observations from $\boldsymbol{\nu}$ and the algorithm π (e.g., due to its privacy mechanism). A fixed-confidence BAI algorithm π is δ-*correct* when $\mathbb{P}_{\boldsymbol{\nu}\pi}(\tau_{\epsilon,\delta} < +\infty,\ \tilde{a} \notin a^\star(\boldsymbol{\nu})) \leq \delta$ for all $\boldsymbol{\nu} \in \mathcal{F}^K$.

Differential Privacy (DP). An algorithm satisfies the Differential Privacy constraint if the algorithm's outputs are "essentially" equally likely to occur, for any two input datasets that only differ in one individual's data. An adversary only observing the mechanism's output cannot distinguish whether any individual's data was included. A privacy budget ϵ captures the closeness of the output distributions. Smaller ϵ means stronger privacy.

Definition 1 (ϵ-DP [Dwork et al., 2006])**.** *A mechanism $\mathcal{M}$ satisfies ϵ-DP for a given $\epsilon \geq 0$, if, for all neighboring datasets $D \sim D'$, where $D \sim D'$ if and only if $d_{Ham}(D, D') := \sum_{t=1}^{T} \mathbb{1}\{D_t \neq D'_t\} \leq 1$,* i.e.*, D and D' differ by at most one record, and for all sets of output $\mathcal{O} \subseteq \mathrm{Range}(\mathcal{M})$, $\Pr[\mathcal{M}(\mathcal{D}) \in \mathcal{O}] \leq e^{\epsilon} \Pr[\mathcal{M}(\mathcal{D}') \in \mathcal{O}]$ where the probability space is over the coin flips of the mechanism $\mathcal{M}$.*

To ensure ϵ-DP, the Laplace mechanism [Dwork et al., 2010; Dwork and Roth, 2014] adds calibrated Laplacian noise to the algorithm's output. Let $\mathrm{Lap}(b)$ be the Laplace distribution with mean/variance $(0, 2b^2)$.

Theorem 1 (Laplace mechanism, Theorem 3.6 [Dwork and Roth, 2014])**.** *Let $f : \mathcal{X} \to \mathbb{R}^d$ be an algorithm with sensitivity $s(f) \triangleq \max_{\mathcal{D},\mathcal{D}' \text{ s.t } d_{Ham}(\mathcal{D},\mathcal{D}')=1} \|f(\mathcal{D}) - f(\mathcal{D}')\|_1$, where $\|\cdot\|_1$ is the ℓ_1 norm. Let $(Z_i)_{i\in[d]}$ be i.i.d. from Lap$(s(f)/\epsilon)$, then the noisy output $f(\mathcal{D}) + (Z_i)_{i\in[d]}$ satisfies ϵ-DP.*

To be consistent with the literature on private bandits, we use global DP to denote the central DP model with a trusted central decision maker. While the notation δ is standard in the (ϵ, δ)-DP relaxation of the $(\epsilon, 0)$-DP constraint considered in this paper, we use δ for the confidence parameter to be consistent with the literature on pure exploration problems.

DP for BAI. In BAI algorithms, the private input is the observation dataset and the output is the recommended candidate arm $\tilde{a}$ and the sequence of sampled actions $(a_n)_{n<\tau_{\epsilon,\delta}}$ until stopping at $\tau_{\epsilon,\delta}$. Let $R = \{r_1, \dots\}$ be a sequence of private observations. Given a fixed sequence of observations R, we denote by $\Pr[\pi(R) = (T+1, \tilde{a}, (a_1, \dots, a_T))]$ the probability that the BAI algorithm π stops at step $T+1$, recommending action $\tilde{a}$ and sampling actions $(a_1, \dots, a_T)$ when interacting with R. The randomisation in this probability comes only from the BAI algorithm's sampling, recommendation and stopping rules, whereas the observations are fixed. Then, a BAI algorithm π is said to be *ϵ-global DP* if, for every two neighboring sequences of observations R and R', and for every possible stopping time, recommendation and sampled actions $(T+1, \tilde{a}, (a_1, \dots, a_T))$, we have that

$$\Pr[\pi(R) = (T+1, \tilde{a}, (a_1, \dots, a_T))] \leq e^{\epsilon} \Pr[\pi(R') = (T+1, \tilde{a}, (a_1, \dots, a_T))] .$$

Main Goal: Design ϵ-global DP δ-correct BAI algorithms, with the smallest sample complexity $\tau_{\epsilon,\delta}$.

Notation. Let $[x]_0^1 := \max\{0, \min\{1, x\}\}$ be the clipping operator to $[0, 1]$. Let $\mathbb{1}(\cdot)$ be the indicator function. For two probability distributions $\mathbb{P}$ and $\mathbb{Q}$ on the measurable space $(\Omega, \mathcal{G})$, the Total Variation (TV) distance is $\mathrm{TV}(\mathbb{P} \,\|\, \mathbb{Q}) := \sup_{A\in\mathcal{G}}\{\mathbb{P}(A) - \mathbb{Q}(A)\}$ and the Kullback-Leibler (KL) divergence is $\mathrm{KL}(\mathbb{P} \,\|\, \mathbb{Q}) := \int \log\left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}(\omega)\right) \mathrm{d}\mathbb{P}(\omega)$, when $\mathbb{P} \ll \mathbb{Q}$, and $+\infty$ otherwise. The KL divergence and TV distance between two Bernoulli distributions with means $(p, q) \in (0, 1)^2$ are the relative entropy denoted by kl, i.e., $\mathrm{KL}(\mathrm{Ber}(p) \,\|\, \mathrm{Ber}(q)) = \mathrm{kl}(p, q) := p\log(p/q) + (1-p)\log((1-p)/(1-q))$, and the absolute mean difference, i.e., $\mathrm{TV}(\mathrm{Ber}(p) \,\|\, \mathrm{Ber}(q)) = |p - q|$. Let $\triangle_K := \{w \in \mathbb{R}^K \mid w \geq 0,\ \sum_{a\in[K]} w_a = 1\}$ be the probability simplex of dimension $K-1$. For all $a \in [K]$, let $N_{n,a} := \sum_{t\in[n-1]} \mathbb{1}(a_t = a)$ be the *global* pulling count of arm a before time n.

3 Lower Bound on the Expected Sample Complexity

In order to be δ-correct, an algorithm has to be able to distinguish $\boldsymbol{\nu}$ from *alternative* instances with different best arms, i.e., an instance $\boldsymbol{\kappa} \in \mathrm{Alt}(\boldsymbol{\nu}) := \{\boldsymbol{\kappa} \in \mathcal{F}^K \mid a^\star(\boldsymbol{\kappa}) \neq a^\star(\boldsymbol{\nu})\}$. On the other hand, being ϵ-global DP forces an algorithm to have similar behaviour on similar instances. The interplay between the stochasticity of the bandit instance, controlled with the KL divergence, and the stochasticity of the privacy mechanism, controlled with the TV distance, is smoothly captured by

$$\mathrm{d}_\epsilon(\nu, \kappa) := \inf_{\varphi\in\mathcal{F}} \{\epsilon \cdot \mathrm{TV}(\nu \,\|\, \varphi) + \mathrm{KL}(\varphi \,\|\, \kappa)\} , \tag{1}$$

recently introduced by Azize et al. [2025] for ϵ-global DP regret minimization. The d_ϵ divergence measures the shortest two-parts path between the two distributions ν and κ, by finding the best

intermediate distribution $\varphi \in \mathcal{F}$. The cost of moving from ν to φ is measured with the TV distance rescaled by ϵ, while it is measured with the KL divergence when moving from φ to κ.

The tension between the δ-correct and ϵ-global DP constraints yields the following problem-dependent non-asymptotic lower bound on the expected sample complexity $\mathbb{E}_{\boldsymbol{\nu}\pi}[\tau_{\epsilon,\delta}]$ for any algorithm π on any instance $\boldsymbol{\nu}$, holding for any values of ϵ and δ.

Theorem 2. *Let $(\epsilon, \delta) \in \mathbb{R}_+^\star \times (0,1)$. For any algorithm π that is δ-correct and ϵ-global DP on $\mathcal{F}^K$,*

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

for all $\boldsymbol{\nu} \in \mathcal{F}^K$ with unique best arm. The inverse of the characteristic time *$T_\epsilon^\star(\boldsymbol{\nu})$ is defined as*

$$T_\epsilon^\star(\boldsymbol{\nu})^{-1} := \sup_{w \in \triangle_K} \inf_{\boldsymbol{\kappa} \in \mathrm{Alt}(\boldsymbol{\nu})} \sum_{a=1}^K w_a \mathrm{d}_\epsilon(\nu_a, \kappa_a) \,. \tag{2}$$

Comments. (a) The characteristic time in the lower bound is the value of a two-player zero-sum game between a MIN player, who plays instances $\boldsymbol{\kappa}$ close of $\boldsymbol{\nu}$ is order to confuse the MAX player, who in order plays an arm allocation $w \in \triangle_K$ to distinguish between $\boldsymbol{\nu}$ and $\boldsymbol{\kappa}$.

(b) The crucial part in characteristic times similar to Eq. (2) is finding the "right" measure capturing the "distinguishability" between instances. In the non-private lower bounds, this is captured by the KL divergence for parametric distributions [Garivier and Kaufmann, 2016] and by the Kinf (i.e., $\inf \mathrm{KL}$ under mean constraint) for non-parametric distributions [Agrawal et al., 2020]. In the DP lower bounds of Azize et al. [2023], it is captured by $\min\{\mathrm{KL}, \epsilon\mathrm{TV}\}$. In Theorem 2, it is captured by d_ϵ (as in Eq. (1)) that smoothly interpolates between KL and TV. Azize et al. [2025] recently introduced d_ϵ for ϵ-global DP regret minimization. Our results show that d_ϵ also tightly captures the hardness of fixed-confidence BAI under ϵ-global DP. Namely, our DP-TT algorithm achieves a matching upper bound when $\delta \to 0$ (up to a constant smaller than 8), for all instances with distinct means and all values of ϵ.

(c) Azize et al. [2023, Theorem 2] provides a lower bound on the sample complexity of any ϵ-global δ-correct algorithm, where the inverse characteristic time is $\sup_{w \in \triangle_K} \inf_{\boldsymbol{\kappa} \in \mathrm{Alt}(\boldsymbol{\nu})} \min\{\sum_{a=1}^K w_a \mathrm{KL}\left(\nu_a \parallel \kappa_a\right), 6\epsilon \sum_{a=1}^K w_a \mathrm{TV}\left(\nu_a \parallel \kappa_a\right)\}$. The lower bound of Theorem 2 is strictly tighter than that of Theorem 2 in [Azize et al., 2023], for all instances $\boldsymbol{\nu}$ and values of ϵ. The reason is that $d_\epsilon(\mathbb{P}, \mathbb{Q}) \leq \min\{\mathrm{KL}(\mathbb{P}, \mathbb{Q}), \epsilon\mathrm{TV}(\mathbb{P}, \mathbb{Q})\}$ for any two distributions.

(d) The lower bound of Theorem 2 suggests the existence of two privacy regimes, depending on the value of ϵ and the instance $\boldsymbol{\nu}$. Specifically, when ϵ is big, d_ϵ reduces to the KL, and we retrieve the classic non-private lower bound. On the other hand, as $\epsilon \to 0$, d_ϵ reduces to ϵ TV, and the characteristic time reduces to $\frac{1}{\epsilon} T_{\mathrm{TV}}^\star(\boldsymbol{\nu}) := \frac{1}{\epsilon} \left(\sup_{w \in \triangle_K} \inf_{\boldsymbol{\kappa} \in \mathrm{Alt}(\boldsymbol{\nu})} \sum_{a=1}^K w_a \mathrm{TV}\left(\nu_a \parallel \kappa_a\right) \} \right)^{-1} = \frac{1}{\epsilon} \sum_{a=1}^k \frac{1}{\Delta_a}$, where $\Delta_a = \mu^\star - \mu_a$ for $a \neq a^\star$ and $\Delta_{a^\star} = \min_{a \neq a^\star} \Delta_a$. This improves the high privacy regime lower bound of prior work by a factor 6. Also, the value of ϵ at which the privacy regimes change can be tightly specified, which we quantify for Bernoulli instances in the following.

Proof Sketch and Techniques. The proof uses the standard reduction to hypothesis testing [Garivier and Kaufmann, 2016], using the data-processing inequality. The asymptotic techniques used by Azize et al. [2025] for regret cannot be adapted for our non-asymptotic lower bound. Thus, new techniques are needed. The main technical novelty of the proof is a tighter quantification of the "similar" behaviour of a DP mechanism when applied to stochastic datasets. Specifically, let $\mathcal{M}$ be an ϵ-DP mechanism. Given two data-generating distributions $\mathbb{P}$ and $\mathbb{Q}$, letting $\mathbb{M}_{\mathbb{P},\mathcal{M}}$ (resp. $\mathbb{M}_{\mathbb{Q},\mathcal{M}}$) be the marginal over outputs of the mechanism when the input dataset is generated through $\mathbb{P}$ (resp. $\mathbb{Q}$), then we show that

$$\mathrm{KL}\left(\mathbb{M}_{\mathbb{P},\mathcal{M}} \parallel \mathbb{M}_{\mathbb{Q},\mathcal{M}}\right) \leq \inf_{\mathbb{L}} \left\{ \epsilon \inf_{\mathbb{C}_{\mathbb{P},\mathbb{L}}} \left\{ \mathbb{E}_{D,D' \sim \mathbb{C}_{\mathbb{P},\mathbb{L}}} \left[d_{\mathrm{Ham}}(D, D') \right] \right\} + \mathrm{KL}\left(\mathbb{L} \parallel \mathbb{Q}\right) \right\} ,$$

where the first infimum is over all distributions $\mathbb{L}$ on the input space, and the second infimum is an optimal transport problem over all couplings between $\mathbb{P}$ and $\mathbb{L}$, where the cost is the Hamming distance (introduced in Definition 1). This bound of general interest could be applied to get tighter lower bounds in any DP application using stochastic inputs. For product and bandit distributions, we solve the optimal transport using maximal couplings, where the Total Variation naturally appears,

while keeping the first infimum unchanged, giving rise to the d_ϵ quantity. Finally, plugging the new upper bound on the KL in the hypothesis reduction concludes the sample complexity lower bound proof. A detailed proof and discussion of all these claims is given in Appendix D.

Transportation Costs Based on Signed Divergences. In non-parametric BAI, the ordering between the mean parameters is captured by a signed Kinf [Jourdan et al., 2022]. We adopt this convention by introducing a signed divergences: d_ϵ^+ and d_ϵ^- defined in Lemma 3 on means rather than probability distributions, where $d_\epsilon^\pm$ refers to both of them. Given $(\kappa, \nu) \in \mathcal{F}^2$ with means $(\lambda, \mu) \in (0,1)^2$, they satisfy $\mathrm{d}_\epsilon(\kappa,\nu) = d_\epsilon^+(\lambda,\mu)$ when $\mu > \lambda$, and $\mathrm{d}_\epsilon(\kappa,\nu) = d_\epsilon^-(\lambda,\mu)$ otherwise (Lemma 22).

Lemma 3. *For all $x \in [0,1]$, let us define $g_\epsilon^-(x) := \frac{xe^\epsilon}{x(e^\epsilon-1)+1}$ and $g_\epsilon^+(x) := 1 - g_\epsilon^-(1-x) = (g_\epsilon^-)^{-1}(x)$. For all $(\lambda,\mu) \in \mathbb{R}\times[0,1]$, the* signed divergences *are defined as*

$$d_\epsilon^+(\lambda,\mu) := \mathbb{1}\left(\mu > [\lambda]_0^1\right) \inf_{z\in[[\lambda]_0^1,\mu]} \left\{\mathrm{kl}(z,\mu) + \epsilon(z - [\lambda]_0^1)\right\}$$

$$= \begin{cases} 0 & \text{if } \mu \in [0, [\lambda]_0^1] \\ -\log\left(1-\mu(1-e^{-\epsilon})\right) - \epsilon[\lambda]_0^1 & \text{if } \mu \in (g_\epsilon^-([\lambda]_0^1), 1] \\ \mathrm{kl}\left(\lambda,\mu\right) & \text{if } \lambda \in (0,1) \text{ and } \mu \in ([\lambda]_0^1, g_\epsilon^-([\lambda]_0^1)] \end{cases},$$

$$d_\epsilon^-(\lambda,\mu) := \mathbb{1}\left(\mu < [\lambda]_0^1\right) \inf_{z\in[\mu,[\lambda]_0^1]} \left\{\mathrm{kl}(z,\mu) + \epsilon([\lambda]_0^1 - z)\right\} = d_\epsilon^+(1-\lambda, 1-\mu)\,. \tag{3}$$

When two distributions are close enough compared to the privacy ϵ, the signed divergences reduce to the KL divergence: the indistinguishability due to the stochasticity of the instance is stronger than the indistinguishability due to DP. The function g_ϵ^- (resp. g_ϵ^+) represents the maximal (resp. minimal) mean for which this property hold for d_ϵ^+ (resp. d_ϵ^-).

For $(\mu, w) \in \mathbb{R}^K \times \mathbb{R}_+^K$, the *transportation cost* of the pair of arms $(a,b) \in [K]^2$ is defined as

$$W_{\epsilon,a,b}(\mu,w) := \mathbb{1}\left([\mu_a]_0^1 > [\mu_b]_0^1\right) \inf_{u\in[0,1]} \left\{w_a d_\epsilon^-(\mu_a,u) + w_b d_\epsilon^+(\mu_b,u)\right\}\,. \tag{4}$$

The signed divergences $d_\epsilon^\pm$ and the transportation costs $(W_{\epsilon,a,b})_{(a,b)\in[K]^2}$ satisfy all the desired properties required to study BAI algorithms based on the empirical version of $W_{\epsilon,a,b}$ (see Lemmas 23, 24, 25 and 26, as well as Lemmas 35, 36, 37 and 38), e.g., symmetry, explicit formula, monotonicity, strict convexity, etc.

Properties of the Characteristic Time and Optimal Allocation. The set $w_\epsilon^\star(\boldsymbol{\nu})$ of *optimal allocations* is the maximizer of the outer supremum on $\triangle_K$ that defines $T_\epsilon^\star(\boldsymbol{\nu})^{-1}$ in Eq. (2). Theorem 4 gathers key properties satisfied by $T_\epsilon^\star(\boldsymbol{\nu})$ and $w_\epsilon^\star(\boldsymbol{\nu})$, for Bernoulli distributions. See lemmas proven in Appendix G, i.e., Lemmas 36, 43 and 47.

Theorem 4. *Let $\boldsymbol{\nu} \in \mathcal{F}^K$ having means $\mu \in (0,1)^K$ with unique best arm $a^\star$. Then, we have*

$$T_\epsilon^\star(\boldsymbol{\nu})^{-1} = \max_{w\in\triangle_K} \min_{a\neq a^\star} W_{\epsilon,a^\star,a}(\mu,w) \quad \text{and} \quad T_\epsilon^\star(\boldsymbol{\nu}) \geq \sum\nolimits_{a\in[K]} \Delta_{\epsilon,a}^{-1}\,. \tag{5}$$

where $\Delta_{\epsilon,a^\star} := \min_{a\neq a^\star} d_\epsilon^-(\mu_{a^\star},\mu_a)$ and $\Delta_{\epsilon,a} := d_\epsilon^+(\mu_a,\mu_{a^\star})$ for all $a \neq a^\star$. The optimal allocation is unique, has dense support and ensures the equality of the transportation costs with $T_\epsilon^\star(\boldsymbol{\nu})^{-1}$ (i.e., information balance equation), namely $w_\epsilon^\star(\boldsymbol{\nu}) = \{w_\epsilon^\star\}$, $\min_{a\in[K]} w_{\epsilon,a}^\star > 0$ and $W_{\epsilon,a^\star,a}(\mu,w_\epsilon^\star) = T_\epsilon^\star(\boldsymbol{\nu})^{-1}$ for all $a \neq a^\star$.

In Garivier and Kaufmann [2016], the characteristic time and its optimal allocation can be computed with a simpler optimisation problem. A simpler optimization problem can also be solved to compute $T_\epsilon^\star(\boldsymbol{\nu})$ and $w_\epsilon^\star(\boldsymbol{\nu})$ explicitly (Lemma 47).

Allocation Dependent Low Privacy Regime. Let $(\mu, w, a, b) \in (0,1)^K \times \mathbb{R}_+^K \times [K]^2$ such that $\mu_a > \mu_b$ and $\min\{w_a, w_b\} > 0$. The non-private Bernoulli transportation costs [Garivier and Kaufmann, 2016] are defined as

$$W_{a,b}(\mu,w) := w_a \mathrm{kl}(\mu_a, \mu_{a,b}^w) + w_b \mathrm{kl}(\mu_b, \mu_{a,b}^w) \quad \text{with} \quad \mu_{a,b}^w := \frac{w_a\mu_a + w_b\mu_b}{w_a + w_b}\,.$$

We provide an *allocation-dependent* low-privacy condition that depends on (ϵ,μ,w) (Lemma 45), i.e., $W_{\epsilon,a,b}(\mu,w) = W_{a,b}(\mu,w)$ is implied by

$$\mu_a - \mu_b \leq (1-e^{-\epsilon}) \min\left\{(1 + w_a/w_b)\,\mu_a g_\epsilon^-(1-\mu_a), (1+w_b/w_a)\,(1-\mu_b) g_\epsilon^-(\mu_b)\right\}\,. \tag{6}$$

Plugging $w^\star_\epsilon$ from Theorem 4 in Eq. (6) would give an implicit condition on (ϵ, μ) under which the non-private characteristic time $T^\star(\boldsymbol{\nu})$ for Bernoulli distributions is recovered, i.e., $T^\star_\epsilon(\boldsymbol{\nu}) = T^\star(\boldsymbol{\nu})$. From a privacy-utility tradeoff perspective, the choice of ϵ that provides the highest privacy protection while maintaining a low sample complexity is exactly this change-of-regime ϵ, depending on the unknown μ. A weaker (yet explicit) allocation-independent sufficient condition for $T^\star_\epsilon(\boldsymbol{\nu}) = T^\star(\boldsymbol{\nu})$ is $\epsilon \geq \max_{a \neq a^\star} \epsilon_{a^\star,a}$ where $\epsilon_{a,b} := \log\left(\frac{\mu_a(1-\mu_b)}{\mu_b(1-\mu_a)}\right)$.

4 Generalized Likelihood Ratio Stopping Rule

Designing appropriate recommendation and stopping rules for the BAI problem can be framed as a sequential hypothesis testing task with multiple hypotheses $\{\mu_a = \max_{b\in[K]} \mu_b\}$. One of the earliest approaches to active hypothesis testing—where data collection is also optimized—was introduced by Chernoff [1959], who advocated for the use of Generalized Likelihood Ratio (GLR) tests for stopping decisions. This methodology is also popular in the context of BAI [Garivier and Kaufmann, 2016]. Despite its relevance, fewer works attempted to extend it for private sequential hypothesis testing, see, e.g., Zhang et al. [2022] under Rényi DP and Azize et al. [2024] under ϵ-local and ϵ-global DP.

Mean Estimator. Three rules need to be specified to define a BAI algorithm: recommendation, sampling, and stopping rules. An important remark in designing BAI algorithms is that the dependence of these rules on the private input observation dataset comes solely through the sequence of mean estimators. Thus, designing a sequence of mean estimators that satisfy DP is crucial when defining a ϵ-global DP BAI algorithm. To estimate the sequence of means, defined in Lines 5-8 of Algorithm 1, we rely on two ingredients: adaptive arm-dependent episodes with a geometric update grid and the Laplace mechanism. We call this mechanism estimating the sequence of means the Geometric Private Estimator, i.e., GPE$_\eta(\epsilon)$. Most notably, we eliminate "observation forgetting" from GPE$_\eta(\epsilon)$, an important design choice made in all past BAI algorithms Sajed and Sheffet [2019]; Azize et al. [2023, 2024]. Specifically, for some $\eta > 0$ called the geometric grid parameter, GPE$_\eta(\epsilon)$ estimates the noisy means in arm-dependent phases: a phase changes when the counts of an arm has increased multiplicatively by $1 + \eta$ (Line 5). Then, GPE$_\eta(\epsilon)$ only updates the mean of the arm that changed phases, by accumulating the observations collected from its last phase and adding Laplace noise (Line 7). Due to this accumulation step, we *do not forget* the observations from past phases. Thus, each estimated noisy mean $\tilde{\mu}_{n,a}$ in Line 7 contains $\tilde{N}_{n,a}$ i.i.d. observations from ν_a and $k_{n,a} \approx \log_{1+\eta} \tilde{N}_{n,a}$ i.i.d. observations from Lap$(1/\epsilon)$. In contrast, using forgetting produces a noisy mean that contains fewer i.i.d. observations from ν_a (e.g. $\tilde{N}_{n,a}/2$ samples for forgetting with $\eta = 1$), but only *one* Laplace noise. While removing forgetting allows us to keep more signal, i.e., more i.i.d samples from ν_a, we need more noise, i.e., the cumulative sum of Lap$(1/\epsilon)$, which is logarithmic in the number of samples from ν_a. Tighter concentration inequalities allow controlling the cumulative sum of Laplace noise. See below for a detailed discussion about our novel concentration results. As long as the number of samples from the Lap$(1/\epsilon)$ is logarithmic in the number of samples from ν_a, the effect of noise on the sample complexity is similar to having only *one* additional Laplace noise.

Privacy Analysis. By adaptive post-processing, the following lemma is proved naturally for any ϵ.

Lemma 5. *Any BAI algorithm using only GPE$_\eta(\epsilon)$ to access observations is ϵ-global DP on* $[0, 1]$.

Proof Sketch. The proof combines two steps. First, we show that the sequence of mean estimators produced by GPE$_\eta(\epsilon)$ is ϵ-DP. The crucial observation is that a change in one observation *only* affects the partial sum collected in just *one* arm-phase. By the Laplace mechanism, adding one Lap$(1/\epsilon)$ to the partial sum is enough to make it ϵ-DP. Then, by post-processing, the sequence of accumulated partial sums $(\tilde{S}_{k_{n,a},a})$ and noisy means $(\mu_{n,a})$ (Line 7) are also ϵ-DP. The second step shows how to use the sequential nature of the process and adaptive post-processing to conclude that BAI algorithms using only GPE$_\eta(\epsilon)$ are ϵ-global DP. The detailed proof is in Appendix E.

Recommendation Rule. The recommendation rule $\tilde{a}_n$ is defined as the arm with the highest clipped noisy empirical mean, i.e., $\tilde{a}_n \in \arg\max_{a\in[K]} [\tilde{\mu}_{n,a}]_0^1$ where ties are broken uniformly at random.

GLR Stopping Rule. The GLR stopping rule runs K sequential GLR tests in parallel, and stops as soon as one of these tests can reject the null hypothesis. When comparing the recommendation $\tilde{a}_n$ with an alternative arm a, the GLR statistic is defined as the transportation cost $W_{\epsilon,\tilde{a}_n,a}$ evaluated

empirically at $(\tilde{\mu}_n, \tilde{N}_n)$ (see Eq. (4)). Intuitively, $W_{\epsilon,\tilde{a}_n,a}(\tilde{\mu}_n, \tilde{N}_n)$ represents the amount of empirical evidence to reject the hypothesis that arm a has a higher mean than $\tilde{a}_n$. One can stop and recommend $\tilde{a}_n$ when all these statistics exceed a given stopping threshold. Given a privacy budget and risk $(\epsilon, \delta) \in \mathbb{R}_+^\star \times (0,1)$ and a stopping threshold $c : \mathbb{N} \times \mathbb{R}_+^\star \times (0,1) \to \mathbb{R}_+$, we define

$$\tau_{\epsilon,\delta} = \inf\{\ n \ \mid\ \forall a \neq \tilde{a}_n,\ \ W_{\epsilon,\tilde{a}_n,a}(\tilde{\mu}_n, \tilde{N}_n) > c(\tilde{N}_{n,\tilde{a}_n}, \epsilon, \delta) + c(\tilde{N}_{n,a}, \epsilon, \delta)\ \}\ . \tag{7}$$

Given its proximity to the characteristic time $T_\epsilon^\star(\boldsymbol{\nu})$, see Eq. (5) (Theorem 4), the GLR stopping rule is a good candidate to match the lower bound, i.e., if one could sample arms according to $w_\epsilon^\star(\boldsymbol{\nu})$ and use the stopping threshold $\log(1/\delta)$. Unfortunately, this threshold is too good to be δ-correct and $w_\epsilon^\star(\boldsymbol{\nu})$ should be estimated as it is unknown (Section 5).

Calibration of the Stopping Threshold. Regardless of the sampling rule, the stopping threshold should ensure δ-correctness of the GLR stopping rule for any pair (ϵ, δ), see Theorem 6.

Theorem 6. *Let $(\epsilon, \delta, \eta) \in \mathbb{R}_+^\star \times (0,1) \times \mathbb{R}_+^\star$. Let $s > 1$, ζ be the Riemann ζ function and $\overline{W}_{-1}(x) = -W_{-1}(-e^{-x})$ for all $x \geq 1$, where W_{-1} is the negative branch of the Lambert W function, satisfying $\overline{W}_{-1}(x) \approx x + \log x$ (Lemma 52). Given any sampling rule using the GPE$_\eta(\epsilon)$, using the GLR stopping rule as in Eq.* (7) *with the GPE$_\eta(\epsilon)$ and the stopping threshold $c(n, \epsilon, \delta) := c_1(n, \delta) + c_2(n, \epsilon)$ where*

$$\begin{aligned} c_1(n,\delta) &= \overline{W}_{-1}\left(\log\left(K\zeta(s)/\delta\right) + s\log(k_\eta(n)) + 3 - \log 2\right) - 3 + \log 2\ , \\ c_2(n,\epsilon) &= k_\eta(n)\left(\log\left(1 + 2\epsilon n/k_\eta(n)\right) + 1\right) \quad \textit{with} \quad k_\eta(x) := 1 + \log_{1+\eta} x\ , \end{aligned} \tag{8}$$

yields a δ-correct and ϵ-global DP algorithm for all Bernoulli instances with a unique best arm.

The proof of Theorem 6 builds on novel concentration results of independent interest (Appendix F.2). Our explicit instance-independent upper bounds are pivotal to derive the stopping threshold in Eq. (8), which avoids the large instance-dependent constants used in the regret minimisation literature [Azize et al., 2025].

Concentration Results. First, we give tail bounds for the cumulative sum of i.i.d. Laplace observations (Lemma 16). We use Chernoff's method with the convex conjugate of the moment generating function of Lap$(1/\epsilon)$, hence improving on Azize et al. [2025, Lemma 18] that approximates it. Second, we derive tail bounds for the sum between independent cumulative sums of t i.i.d. Bernoulli and n_t i.i.d. Laplace observations (Lemmas 18 and 19). They involve the *modified* signed divergences $\widetilde{d}_\epsilon^{\pm}$ that better capture the non-asymptotic tails behaviour, and are equivalent to $d_\epsilon^{\pm}$ to an additive term $\Theta(\log(1 + 2\epsilon r_t)/r_t)$ where $r_t := t/n_t$ (Lemma 30). Whenever $r_t \to +\infty$, we recover the same noise effect as adding only one Lap$(1/\epsilon)$ observation. For $x > 0$, the exponential decrease of the probability of exceeding $\mu + x$ (resp. being lower than $\mu - x$) scales as $t\widetilde{d}_\epsilon^{-}(\mu + x, \mu, r_t)$ (resp. $t\widetilde{d}_\epsilon^{+}(\mu - x, \mu, r_t)$). The proof builds on fine-grained tail bounds of the sum of two independent random variables, i.e., we bound those probabilities by the maximal product between their respective survival functions (Lemma 10). While Azize et al. [2025, Lemma 19] directly integrates their tail bounds, Lemma 12 can be used with any tail bounds. Third, we obtain time-uniform upper tail bounds for $\tilde{N}_{n,a}\widetilde{d}_\epsilon^{\pm}(\tilde{\mu}_{n,a}, \mu_a, \tilde{N}_{n,a}/k_{n,a})$ by exploiting the geometric-grid update of $(\tilde{\mu}_n, \tilde{N}_n, k_n)$.

Threshold Scaling. The threshold c_1 in Eq. (8) ensures δ-correctness of the *modified* GLR stopping rule, defined in Appendix F.1 with the *modified* transportation costs $\widetilde{W}_{\epsilon,a,b}$ and divergences $\widetilde{d}_\epsilon^{\pm}$ (Appendix F.1). Independent of ϵ, it scales as $\log(1/\delta) + \Theta(\log\log(1/\delta))$ when $\delta \to 0$ and $\Theta(\log\log(n))$ when $n \to +\infty$. The threshold c_2 in Eq. (8) is an upper bound on $\tilde{N}_{n,a}(d_\epsilon^{\pm}(\tilde{\mu}_{n,a}, \mu_a) - \widetilde{d}_\epsilon^{\pm}(\tilde{\mu}_{n,a}, \mu_a, \tilde{N}_{n,a}/k_{n,a}))$ (Lemma 30) that scales as $\Theta(\epsilon n)$ when $\epsilon \to 0$ and as $\Theta((\log n)^2)$ when $n \to +\infty$. Both c_1 and c_2 scales as $\Theta(1/\log(1+\eta))$ when $\eta \to 0$.

Limitation. As the threshold in Eq. (7) is the sum of per-arm thresholds, it scales as $2\log(1/\delta)$ when $\delta \to 0$, hence incurs a suboptimal factor 2 asymptotically. Obtaining a threshold in $\log(1/\delta)$ is left for future work. It requires controlling the re-weighted sum of modified divergences $\widetilde{d}_\epsilon^{\pm}$. Azize et al. [2023, Theorem 4] has a suboptimal factor 2 for the same reason and incurs an additive factor $\frac{1}{n\epsilon^2}\log(1/\delta)^2$ due to the separate control of the Laplace and the Bernoulli observations (based on sub-Gaussian concentration results). Azize et al. [2024, Lemma 18] alleviates this factor 2 in their low privacy regime, yet it also pays $\frac{1}{n\epsilon^2}\log(1/\delta)^2$.

Algorithm 1 Differentially Private Top Two (DP-TT) Algorithm.

Input: setting parameters $(\epsilon, \delta) \in \mathbb{R}_+^\star \times (0,1)$, algorithmic hyperparameters $(\eta, \beta) \in \mathbb{R}_+^\star \times (0,1)$ and threshold c, e.g., $(\eta, \beta) = (1, 1/2)$ and c as in Eq. (8). $(W_{\epsilon,a,b})_{(a,b)\in[K]}$ as in Eq. (4).
Output: Stopping time $\tau_{\epsilon,\delta}$, recommendation $\tilde{a}_{\tau_{\epsilon,\delta}}$ and pulling history $(a_n)_{n<\tau_{\epsilon,\delta}}$.
Initialization: For all $a \in [K]$, pull arm a, observe $X_{a,a} \sim \nu_a$ and draw $Y_{1,a} \sim \mathrm{Lap}(1/\epsilon)$. Set $n = K+1$. For all $a \in [K]$, set $\tilde{S}_{n,a} = X_{a,a} + Y_{1,a}$, $k_{n,a} = 1$, $T_1(a) = n$, $N_{n,a} = \tilde{N}_{n,a} = 1$, $\tilde{\mu}_{n,a} = \tilde{S}_{n,a}/\tilde{N}_{n,a}$, $L_{n,a} = 0$ and $N_{n,a}^a = 0$.
for $n \geq K+1$ **do**
  if there exists $a \in [K]$ such that $N_{n,a} \geq (1+\eta)^{k_{n,a}}$ **then** ▷ Per-arm geometric update grid
    For this arm a, change phase $k_{n,a} \leftarrow\!\shortmid k_{n,a} + 1$, and $(T_{k_{n,a}}(a), \tilde{N}_{n,a}) = (n, N_{T_{k_{n,a}}(a),a})$;
    Set $\tilde{S}_{k_{n,a},a} = \sum_{t=T_{k_{n,a}-1}(a)}^{T_{k_{n,a}}(a)-1} X_{t,a}\mathbb{1}(a_t = a) + Y_{k_{n,a},a} + \tilde{S}_{k_{n,a}-1,a}$ with $Y_{k_{n,a},a} \sim \mathrm{Lap}(1/\epsilon)$, and update the mean $\tilde{\mu}_{n,a} = \tilde{S}_{k_{n,a},a}/\tilde{N}_{n,a}$;
  end if
  Set $\tilde{a}_n \in \arg\max_{a\in[K]}[\tilde{\mu}_{n,a}]_0^1$; ▷ Recommendation rule
  if $W_{\epsilon,\tilde{a}_n,a}(\tilde{\mu}_n, \tilde{N}_n) > \sum_{b\in\{\tilde{a}_n,a\}} c(\tilde{N}_{n,b}, \epsilon, \delta)$ for all $a \neq \hat{a}_n$ **then** ▷ GLR stopping rule
    return $(n, \tilde{a}_n, (a_t)_{t<n})$.
  end if
  Set $B_n = \tilde{a}_n$ and $C_n \in \arg\min_{a\neq B_n}\{W_{\epsilon,B_n,a}(\tilde{\mu}_n, N_n) + \log N_{n,a}\}$; ▷ EB-TCI
  Set $a_n = B_n$ if $N_{n,B_n}^{B_n} \leq \beta L_{n+1,B_n}$, and $a_n = C_n$ otherwise; ▷ β-tracking
  Pull a_n, observe and store $X_{n,a_n} \sim \nu_{a_n}$;
  Update $(N_{n+1,a_n}, L_{n+1,B_n}, N_{n+1,B_n}^{B_n}) = (N_{n,a_n}, L_{n,B_n}, N_{n,B_n}^{B_n}) + (1, 1, \mathbb{1}(B_n = a_n))$;
end for

5 Top Two Sampling Rule

Equipped with a recommendation and stopping rules, we define a sampling rule using the $\mathrm{GPE}_\eta(\epsilon)$. Within the fixed-confidence BAI literature, we adopt the Top Two approach [Russo, 2016; Qin et al., 2017; Shang et al., 2020; Jourdan et al., 2022] that recently received increased scrutiny due to its good theoretical guarantees [Jourdan and Degenne, 2024; You et al., 2023; Jourdan et al., 2024; Bandyopadhyay et al., 2024], competitive empirical performance, and low computational cost. The Differentially Private Top Two (DP-TT) algorithm (Algorithm 1) uses the EB-TCI-β sampling rule [Jourdan et al., 2022]. In Appendix I, we introduce the Track-and-Stop [Garivier and Kaufmann, 2016] and LUCB [Kalyanakrishnan et al., 2012] sampling rules for fixed-confidence BAI under ϵ-global DP.

After initialization, a Top Two sampling rule specifies four choices [Jourdan, 2024]: a leader arm $B_n \in [K]$, a challenger arm $C_n \in [K] \setminus \{B_n\}$, a target allocation $\beta_n(B_n, C_n) \in [0,1]$ and a mechanism to choose the next arm to sample from, i.e., $a_n \in \{B_n, C_n\}$ by using $\beta_n(B_n, C_n)$. The leader should select a good estimator of the best arm $a^\star$. We use the empirical best (EB) leader that coincides with our recommendation rule, i.e., $B_n := \tilde{a}_n$. The challenger should be a confusing alternative arm, for which the empirical evidence that the leader has a better mean is low. We use the TCI challenger [Jourdan et al., 2022] that penalizes oversampled challenger to foster implicit exploration, i.e., $C_n \in \arg\min_{a\neq B_n}\{W_{\epsilon,B_n,a}(\tilde{\mu}_n, N_n) + \log N_{n,a}\}$ where ties are broken uniformly at random. Crucially, we leverage our novel transportation costs $(W_{\epsilon,B_n,a})_{a\neq B_n}$ featuring the signed divergences $d_\epsilon^\pm$ that are evaluated empirically at $(\tilde{\mu}_n, N_n)$, see Eq. (3) and (4). The target should be chosen to balance the allocation between the leader and the challenger arms. Let $\beta \in (0,1)$, e.g., $\beta = 1/2$. We use a fixed β-design $\beta_n(B_n, C_n) := \beta$. The mechanism to choose the next arm to sample should enforce that this target is reached on average. We use K independent β-tracking procedures (one per leader), i.e., $a_n = B_n$ if $N_{n,B_n}^{B_n} \leq \beta L_{n+1,B_n}$ and $a_n = B_n$ otherwise, where $N_{n,a}^a = \sum_{t\in[n-1]} \mathbb{1}((B_t, a_t) = (a,a))$ and $L_{n,a} = \sum_{t\in[n-1]} \mathbb{1}(B_t = a)$. Using Degenne et al. [2020b, Theorem 6] for each tracking procedure yields $-1/2 \leq N_{n,a}^a - \beta L_{n,a} \leq 1$ for all $a \in [K]$.

Computational and Memory Cost. The $\text{GPE}_\eta(\epsilon)$ sums the observations, and the recommendation and GLR stopping rules are updated when an arm is updated. Using the closed-form formula for $W_{\epsilon,a,b}$ (Lemma 38), the per-iteration computational and global memory costs of DP-TT are $\mathcal{O}(K)$.

Asymptotic Upper Bound on the Expected Sample Complexity. Given a fixed target β, the empirical allocation of $a^\star$ converges towards β, that differs from $w^\star_{\epsilon,a^\star}$. At best, we can estimate the β-optimal allocation $w^\star_{\epsilon,\beta}(\boldsymbol{\nu})$, i.e., maximizer of the inverse β-characteristic time $T^\star_{\epsilon,\beta}(\boldsymbol{\nu})^{-1}$ defined as in Eq. (5) with the constraint $w_{a^\star} = \beta$. While being only nearly asymptotic optimal, i.e., $T^\star_\epsilon(\boldsymbol{\nu}) = \min_{\beta\in(0,1)} T^\star_{\epsilon,\beta}(\boldsymbol{\nu})$, it satisfies $T^\star_{\epsilon,1/2}(\boldsymbol{\nu}) \leq 2T^\star_\epsilon(\boldsymbol{\nu})$ (Lemma 44).

DP-TT is ϵ-global DP, δ-correct and matches $T^\star_\epsilon(\boldsymbol{\nu})$ to a small constant, for any privacy budget ϵ. While Theorem 7 is an asymptotic upper bound for $\delta \to 0$, it holds for any ϵ.

Theorem 7. *Let $(\epsilon, \delta, \eta, \beta) \in \mathbb{R}^\star_+ \times (0,1) \times \mathbb{R}^\star_+ \times (0,1)$ and c as in Eq.* (8)*. The DP-TT algorithm is ϵ-global DP, δ-correct and satisfies that, for any Bernoulli instance $\boldsymbol{\nu}$ with distinct means $\mu \in (0,1)^K$,*

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_{\boldsymbol{\nu}\pi}\left[\tau_{\epsilon,\delta}\right]}{\log(1/\delta)} \leq 2(1+\eta) T^\star_{\epsilon,\beta}(\boldsymbol{\nu}) \,.$$

Proof. The proof (Appendix H) builds on the unified analysis of Jourdan et al. [2022] and relies heavily on the derived regularity properties for $d^\pm_\epsilon$, $(W_{\epsilon,a,b})_{(a,b)}$, $T^\star_{\epsilon,\beta}(\boldsymbol{\nu})$ and $w^\star_{\epsilon,\beta}(\boldsymbol{\nu})$ (Appendix G). □

For $(\eta, \beta) = (1, 1/2)$, the asymptotic upper bound is $4T^\star_{\epsilon,1/2}(\boldsymbol{\nu}) \leq 8T^\star_\epsilon(\boldsymbol{\nu})$. For any privacy budget ϵ, we reduced the gap between known lower and upper bounds for fixed-confidence BAI under ϵ-global DP to a constant lower than 8, hence closing the open problem in Azize et al. [2024]. A discussion on how to improve this constant is deferred to Appendix C.2. Since DP-TT enjoys good empirical performance for moderate value of δ (Figure 1), adapting the non-asymptotic upper bound in Jourdan et al. [2024] to the private setting is an interesting direction for future work.

Comparison with Azize et al. [2023, 2024]. AdaP-TT and AdaP-TT* use the $\text{DAF}(\epsilon)$ estimator, GLR-inspired recommendation/stopping rules and the TTUCB [Jourdan and Degenne, 2024] sampling rule (i.e., UCB-TC-β [Jourdan, 2024]), all based on arm-dependent doubling, forgetting and unclipped estimators. While AdaP-TT relies on the non-private *Gaussian* transportation costs, AdaP-TT* accounts for a high privacy regime by clipping the mean gap, i.e., $(\mu_a - \mu_b)_+ \min\{3\epsilon, (\mu_a - \mu_b)_+\}$ instead of $(\mu_a - \mu_b)^2$. The AdaP-TT and AdaP-TT* algorithms are ϵ-global DP and δ-correct. The sample complexity of AdaP-TT only matches the *high privacy* lower bound for instances where the means are of similar order. AdaP-TT* improves on AdaP-TT by matching the *high privacy* lower bound for all instances with distinct means. However, both AdaP-TT and AdaP-TT* fail to match the lower bound beyond the high-privacy regime, due to the use of non-adapted transportation costs. In contrast, DP-TT uses the d_ϵ-inspired transportation costs, matching the lower bound up to a small constant for all values of ϵ.

Comparison with Sajed and Sheffet [2019]. DP-SE [Sajed and Sheffet, 2019] is an ϵ-global DP version of the Successive Elimination algorithm introduced for the regret minimisation setting, modified by Azize et al. [2023] into a ϵ-global and δ-correct BAI algorithm. Compared to DP-TT, DP-SE is less adaptive and not anytime, since it relies on uniform sampling within each phase and the phase length depends explicitly on the risk δ. The high probability upper bound on the sample complexity scales as $\mathcal{O}(\sum_{a\neq a^\star}(\Delta_a \min\{\epsilon, \Delta_a\})^{-1})$ with $\Delta_a = \mu_{a^\star} - \mu_a$. This matches the lower bound when $\epsilon \to 0$ (to a constant), but fails to recover the sample-complexity lower bound beyond this regime.

6 Experiments

The empirical performance of DP-TT with $(\eta, \beta) = (1, 1/2)$ is compared to AdaP-TT [Azize et al., 2023], AdaP-TT* [Azize et al., 2024], and DP-SE [Sajed and Sheffet, 2019] on different Bernoulli instances for varying privacy budget, ranging from 0.001 to 125 (see Appendix J.1). The first instance has means $\mu_1 = (0.95, 0.9, 0.9, 0.9, 0.5)$ and the second instance has means $\mu_2 = (0.75, 0.7, 0.7, 0.7, 0.7)$. As a benchmark, we also compare to the non-private EB-TCI-β

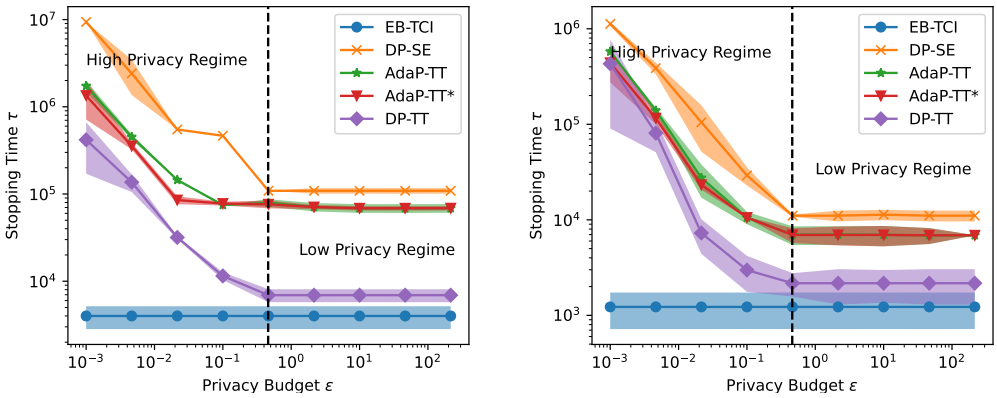

Figure 1: Empirical stopping time $\tau_{\epsilon,\delta}$ (mean ± 2 std) for $\delta = 10^{-2}$ with respect to the privacy budget ϵ on Bernoulli instances (a) μ_1 and (b) μ_2. The vertical line separates the two privacy regimes.

algorithm with $\beta = 1/2$. For $\delta = 10^{-2}$, we run each algorithm 1000 times, and plot the averaged empirical stopping times in Figure 1. Additional experiments are in Appendix J.

Figure 1 shows that DP-TT outperforms all the other δ-correct and ϵ-global DP BAI algorithms, for different values of ϵ and in all the instances tested. The empirical performance of DP-TT demonstrates two regimes. A high-privacy regime, where the stopping time depends on the privacy budget ϵ, and a low privacy regime, where the performance of DP-TT is independent of ϵ, and requires twice the number of samples used by the non-private EB-TCI-β. In Figure 1, the vertical lines are both at $\epsilon = 0.45$. We compare with our allocation-independent condition, i.e., $\max_{a \neq a^\star} \epsilon_{a^\star,a}$ where $\epsilon_{a,b} = \log\left(\frac{\mu_{a^\star}(1-\mu_a)}{\mu_a(1-\mu_{a^\star})}\right)$. For μ_2, we have $\epsilon_{1,a} = 0.25$ for all $a \neq 1$ that is close to the empirical separation. For μ_1, we have $\epsilon_{1,5} = 2.94$ and $\epsilon_{1,a} = 0.75$ for all $a \in \{2,3,4\}$. When an arm has a significantly lower mean, an allocation-dependent condition, such as Eq. (6), is required to reflect the empirical separation. Intuitively, an arm with a significantly lower mean will also be sampled significantly less, hence there is a compounding effect.

7 Conclusion

Motivated by the privacy requirements of sensitive applications of BAI, we address the problem of fixed-confidence BAI under ϵ-global DP. We narrow the gap between the lower and upper bounds on the expected sample complexity to a multiplicative constant smaller than 8, for all ϵ values. Our novel lower bound incorporates d_ϵ an information-theoretic quantity smoothly balancing KL divergence and TV distance, scaled by ϵ. We design a private, arm-dependent geometric grid estimator *without forgetting* and a GLR stopping rule based on the d_ϵ-transportation costs, whose correctness requires novel concentration results for Laplace and mixed distributions. Finally, we proposed a Top Two sampling rule that achieves an asymptotic upper bound matching our lower bound to a small constant.

We detailed research directions to further reduce the constant gap between the lower and upper bounds, by improving both the calibration of the stopping threshold and the analysis of the sampling rule. The most exciting direction for future work is to extend our results to other classes of distributions (e.g., Gaussian or bounded distributions), structured settings (e.g., linear or unimodal), or other identification problems (e.g., approximate BAI or Good Arm Identification). Another interesting research direction is to extend the proposed technique to other variants of pure DP (e.g., (ϵ, δ)-DP or Rényi DP [Mironov, 2017]) or other trust models (e.g., shuffle DP [Cheu, 2021; Girgis et al., 2021]).

Acknowledgments and Disclosure of Funding

We thank Junya Honda for the interesting discussions. Marc Jourdan was supported by the Swiss National Science Foundation (grant number 212111) and by an unrestricted gift from Google. Achraf Azize thanks the support of the FairPlay Joint Team.

References

Naman Agarwal and Karan Singh. The price of differential privacy for online learning. In *International Conference on Machine Learning*, pages 32–40. PMLR, 2017.

Shubhada Agrawal, Sandeep Juneja, and Peter W. Glynn. Optimal δ-correct best-arm selection for heavy-tailed distributions. In *Algorithmic Learning Theory (ALT)*, 2020.

J-Y. Audibert and S. Bubeck. Regret Bounds and Minimax Policies under Partial Monitoring. *Journal of Machine Learning Research*, 2010a.

J-Y. Audibert, S. Bubeck, and R. Munos. Best Arm Identification in Multi-armed Bandits. In *Conference on Learning Theory*, 2010.

Jean-Yves Audibert and Sébastien Bubeck. Regret bounds and minimax policies under partial monitoring. *The Journal of Machine Learning Research*, 11:2785–2836, 2010b.

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

Peter Auer, Chao-Kai Chiang, Ronald Ortner, and Madalina Drugan. Pareto front identification from stochastic bandit feedback. In *Artificial intelligence and statistics*, pages 939–947. PMLR, 2016.

Maryam Aziz, Emilie Kaufmann, and Marie-Karelle Riviere. On multi-armed bandit designs for dose-finding clinical trials. *The Journal of Machine Learning Research*, 22(1):686–723, 2021.

Achraf Azize and Debabrota Basu. When privacy meets partial information: A refined analysis of differentially private bandits. *Advances in Neural Information Processing Systems*, 35:32199–32210, 2022.

Achraf Azize and Debabrota Basu. Concentrated differential privacy for bandits. In *2nd IEEE Conference on Secure and Trustworthy Machine Learning*, 2024.

Achraf Azize, Marc Jourdan, Aymen Al Marjani, and Debabrota Basu. On the complexity of differentially private best-arm identification with fixed confidence. *Thirty-Seventh Conference on Neural Information Processing Systems*, 2023.

Achraf Azize, Marc Jourdan, Aymen Al Marjani, and Debabrota Basu. Differentially private best-arm identification. *arXiv preprint arXiv:2406.06408*, 2024.

Achraf Azize, Yulian Wu, Junya Honda, Francesco Orabona, Shinji Ito, and Debabrota Basu. Optimal regret of bernoulli bandits under global differential privacy. *arXiv preprint arXiv:2505.05613*, 2025.

Agniv Bandyopadhyay, Sandeep Juneja, and Shubhada Agrawal. Optimal top-two method for best arm identification and fluid analysis. *Thirty-Eighth Conference on Neural Information Processing Systems*, 2024.

Debabrota Basu, Christos Dimitrakakis, and Aristide Tossou. Differential privacy for multi-armed bandits: What is it and what is its cost? *arXiv preprint arXiv:1905.12298*, 2019.

Claude Berge. *Topological Spaces: including a treatment of multi-valued functions, vector spaces, and convexity*. Courier Corporation, 1997.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.

Alexandra Carpentier and Andrea Locatelli. Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Conference on Learning Theory*, pages 590–604. PMLR, 2016.

T.-H. Hubert Chan, Elaine Shi, and Dawn Song. Private and continual release of statistics. *ACM Trans. Inf. Syst. Secur.*, 14(3), nov 2011. ISSN 1094-9224. doi: 10.1145/2043621.2043626.

Lijie Chen, Anupam Gupta, Jian Li, Mingda Qiao, and Ruosong Wang. Nearly optimal sampling algorithms for combinatorial pure exploration. In *Proceedings of the 30th Conference on Learning Theory (COLT)*, 2017.

Shouyuan Chen, Tian Lin, Irwin King, Michael R Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. *Advances in neural information processing systems*, 27, 2014.

Wei Chen, Yajun Wang, and Yang Yuan. Combinatorial multi-armed bandit: General framework and applications. In *International Conference on Machine Learning*, 2013.

Zhirui Chen, P. N. Karthik, Yeow Meng Chee, and Vincent Tan. Fixed-budget differentially private best arm identification. In *The Twelfth International Conference on Learning Representations*, 2024.

Herman Chernoff. Sequential Design of Experiments. *The Annals of Mathematical Statistics*, 30: 755–770, 1959.

Albert Cheu. Differential privacy in the shuffle model: A survey of separations. *arXiv preprint arXiv:2107.11839*, 2021.

Sayak Ray Chowdhury and Xingyu Zhou. Shuffle private linear contextual bandits. In *International Conference on Machine Learning*, pages 3984–4009. PMLR, 2022.

R. Combes and A. Proutière. Unimodal bandits: Regret lower bounds and optimal algorithms. In *International Conference on Machine Learning (ICML)*, 2014.

Rémy Degenne and Wouter M Koolen. Pure exploration with multiple correct answers. *Advances in Neural Information Processing Systems*, 32, 2019.

Rémy Degenne, Pierre Ménard, Xuedong Shang, and Michal Valko. Gamification of pure exploration for linear bandits. In *International Conference on Machine Learning*, pages 2432–2442. PMLR, 2020a.

Rémy Degenne, Han Shao, and Wouter Koolen. Structure adaptive algorithms for stochastic bandits. In *International Conference on Machine Learning*, pages 2443–2452. PMLR, 2020b.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Conference on Theory of Cryptography*, TCC'06, pages 265–284, Berlin, Heidelberg, 2006. Springer-Verlag.

Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N Rothblum. Differential privacy under continual observation. In *ACM symposium on Theory of computing*, pages 715–724. ACM, 2010.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. PAC bounds for multi-armed bandit and Markov decision processes. In *Conference on Computational Learning Theory*, COLT '02, page 255–270, Berlin, Heidelberg, 2002. Springer-Verlag. ISBN 354043836X.

Eyal Even-Dar, Shie Mannor, Yishay Mansour, and Sridhar Mahadevan. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(6), 2006.

S. Filippi, O. Cappé, A. Garivier, and C. Szepesvári. Parametric Bandits : The Generalized Linear case. In *Advances in Neural Information Processing Systems*, 2010.

Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. *Advances in Neural Information Processing Systems*, 25, 2012.

Evrard Garcelon, Kamalika Chaudhuri, Vianney Perchet, and Matteo Pirotta. Privacy amplification via shuffling for linear contextual bandits. In *International Conference on Algorithmic Learning Theory*, pages 381–407. PMLR, 2022.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027. PMLR, 2016.

Antonious M Girgis, Deepesh Data, Suhas Diggavi, Ananda Theertha Suresh, and Peter Kairouz. On the renyi differential privacy of the shuffle model. In *ACM SIGSAC Conference on Computer and Communications Security*, pages 2321–2341, 2021.

Yuxuan Han, Zhipeng Liang, Yang Wang, and Jiheng Zhang. Generalized linear bandits with local differential privacy. *Advances in Neural Information Processing Systems*, 34:26511–26522, 2021.

Osama Hanna, Antonious M Girgis, Christina Fragouli, and Suhas Diggavi. Differentially private stochastic linear bandits:(almost) for free. *IEEE Journal on Selected Areas in Information Theory*, 2024.

Junya Honda and Akimichi Takemura. Non-asymptotic analysis of a new bandit algorithm for semi-bounded rewards. *Journal of Machine Learning Research (JMLR)*, 16:3721–3756, 2015.

Bingshan Hu and Nidhi Hegde. Near-optimal thompson sampling-based algorithms for differentially private stochastic bandits. In *Uncertainty in Artificial Intelligence*, pages 844–852. PMLR, 2022.

Bingshan Hu, Zhiming Huang, and Nishant A Mehta. Near-optimal algorithms for private online learning in a stochastic environment. *arXiv e-prints*, pages arXiv–2102, 2021.

Bingshan Hu, Zhiming Huang, Tianyue H Zhang, Mathias Lécuyer, and Nidhi Hegde. Connecting thompson sampling and ucb: Towards more efficient trade-offs between privacy and regret. *arXiv preprint arXiv:2505.02383*, 2025.

K. Jamieson, S. Katariya, A. Deshpande, and R. Nowak. Sparse Dueling Bandits. In *Proceedings of the 18th Conference on Artificial Intelligence and Statistics*, 2015.

Kevin Jamieson and Robert Nowak. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2014.

Marc Jourdan. *Solving Pure Exploration Problems with the Top Two Approach*. PhD thesis, Université de Lille, 2024.

Marc Jourdan and Rémy Degenne. Non-asymptotic analysis of a ucb-based top two algorithm. *Advances in Neural Information Processing Systems*, 36, 2024.

Marc Jourdan, Rémy Degenne, Dorian Baudry, Rianne de Heide, and Emilie Kaufmann. Top two algorithms revisited. *Advances in Neural Information Processing Systems*, 35:26791–26803, 2022.

Marc Jourdan, Rémy Degenne, and Emilie Kaufmann. Dealing with unknown variances in best-arm identification. *International Conference on Algorithmic Learning Theory*, 2023.

Marc Jourdan, Rémy Degenne, and Emilie Kaufmann. An ε-best-arm identification algorithm for fixed-confidence and beyond. *Advances in Neural Information Processing Systems*, 36, 2024.

Kwang-Sung Jun, Lalit Jain, Houssam Nassif, and Blake Mason. Improved Confidence Bounds for the Linear Logistic Model and Applications to Bandits. In *International Conference on Machine Learning*, 2021.

Dionysios S Kalogerias, Kontantinos E Nikolakakis, Anand D Sarwate, and Or Sheffet. Quantile multi-armed bandits: Optimal best-arm identification and a differentially private scheme. *IEEE Journal on Selected Areas in Information Theory*, 2(2):534–548, 2021.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. Pac subset selection in stochastic multi-armed bandits. In *International Conference on Machine Learning*, volume 12, pages 655–662, 2012.

Julian Katz-Samuels and Clayton Scott. Top Feasible Arm Identification. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019.

Tomás Kocák and Aurélien Garivier. Epsilon best arm identification in spectral bandits. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2636–2642, 2021.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Fengjiao Li, Xingyu Zhou, and Bo Ji. Differentially private linear bandits with partial distributed feedback. In *2022 20th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt)*, pages 41–48. IEEE, 2022.

Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. *The Journal of Machine Learning Research*, 18(1):6765–6816, 2017.

Pieter JK Libin, Timothy Verstraeten, Diederik M Roijers, Jelena Grujic, Kristof Theys, Philippe Lemey, and Ann Nowé. Bayesian best-arm identification for selecting influenza mitigation strategies. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018*, 2019.

Simon Lindståhl, Alexandre Proutiere, and Andreas Johnsson. Measurement-based admission control in sliced networks: A best arm identification approach. In *GLOBECOM 2022-2022 IEEE Global Communications Conference*, pages 1484–1490. IEEE, 2022.

David E Losada, David Elsweiler, Morgan Harvey, and Christoph Trattner. A day at the races: using best arm identification algorithms to reduce the cost of information retrieval user studies. *Applied Intelligence*, 52(5):5617–5632, 2022.

S. Magureanu, R. Combes, and A. Proutière. Lipschitz Bandits: Regret lower bounds and optimal algorithms. In *Proceedings on the 27th Conference On Learning Theory*, 2014.

Shie Mannor and John N Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research*, 5(Jun):623–648, 2004.

Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*, pages 263–275. IEEE, 2017.

Nikita Mishra and Abhradeep Thakurta. (Nearly) optimal differentially private stochastic multi-arm bandits. In *Conference on Uncertainty in Artificial Intelligence*, 2015.

Seth Neel and Aaron Roth. Mitigating bias in adaptive data gathering via differential privacy. In *International Conference on Machine Learning*, pages 3720–3729. PMLR, 2018.

Nikola Pavlovic, Sudeep Salgia, and Qing Zhao. Differentially private kernelized contextual bandits. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.

Chao Qin, Diego Klabjan, and Daniel Russo. Improving the expected improvement algorithm. *Advances in Neural Information Processing Systems*, 30, 2017.

H. Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

Daniel Russo. Simple bayesian algorithms for best arm identification. In *Conference on Learning Theory*, pages 1417–1418. PMLR, 2016.

Touqir Sajed and Or Sheffet. An optimal private stochastic-mab algorithm based on optimal private stopping rule. In *International Conference on Machine Learning*, pages 5579–5588. PMLR, 2019.

Xuedong Shang, Rianne Heide, Pierre Menard, Emilie Kaufmann, and Michal Valko. Fixed-confidence guarantees for bayesian best-arm identification. In *International Conference on Artificial Intelligence and Statistics*, pages 1823–1832. PMLR, 2020.

Roshan Shariff and Or Sheffet. Differentially private contextual linear bandits. In *Advances in Neural Information Processing Systems*, pages 4296–4306, 2018.

Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 27, 2014.

Rangarajan K Sundaram. *A first course in optimization theory*. Cambridge university press, 1996.

Jay Tenenbaum, Haim Kaplan, Yishay Mansour, and Uri Stemmer. Differentially private multi-armed bandits in the shuffle model. *Advances in Neural Information Processing Systems*, 34: 24956–24967, 2021.

Abhradeep Guha Thakurta and Adam Smith. (Nearly) optimal algorithms for private online learning in full-information and bandit settings. *Advances in Neural Information Processing Systems*, 26, 2013.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.

Aristide CY Tossou and Christos Dimitrakakis. Achieving privacy in the adversarial multi-armed bandit. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

Katherine Tucker, Janice Branson, Maria Dilleen, Sally Hollis, Paul Loughlin, Mark J Nixon, and Zoë Williams. Protecting patient privacy when sharing patient-level data from clinical trials. *BMC medical research methodology*, 16(1):5–14, 2016.

J. Ville. *Étude Critique de la Notion de Collectif*. Collection des monographies des probabilités. Gauthier-Villars, 1939.

Siwei Wang and Jun Zhu. Optimal learning policies for differential privacy in multi-armed bandits. *Journal of Machine Learning Research*, 25(314):1–52, 2024.

Wei You, Chao Qin, Zihao Wang, and Shuoguang Yang. Information-directed selection for top-two algorithms. In *Conference on Learning Theory*, pages 2850–2851. PMLR, 2023.

Wanrong Zhang, Yajun Mei, and Rachel Cummings. Private sequential hypothesis testing for statisticians: Privacy, error rates, and sample size. In *International Conference on Artificial Intelligence and Statistics*, pages 11356–11373. PMLR, 2022.

Yao Zhao, Connor James Stephens, Csaba Szepesvári, and Kwang-Sung Jun. Revisiting simple regret: Fast rates for returning a good arm. In *40th International Conference on Machine Learning (ICML)*, 2023.

Kai Zheng, Tianle Cai, Weiran Huang, Zhenguo Li, and Liwei Wang. Locally differentially private (contextual) bandits learning. In *Advances in Neural Information Processing Systems*, volume 33, pages 12300–12310, 2020.

Yuan Zhou, Xi Chen, and Jian Li. Optimal pac multiple arm identification with applications to crowdsourcing. In *International Conference on Machine Learning*, pages 217–225. PMLR, 2014.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims made in the abstract and introduction reflect the paper's contributions and scope. Those claims are substantiated with details in the Sections 3, 4, 5 and 6 of the main paper. They are proven theoretically in Appendices D, F, G, H, I and J.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations of Theorems 6 and 7 are detailed in Sections 4 and Appendix C.2. The low computational and memory costs of the DP-TT algorithm is discussed in Sections 4 and 5. Our work addresses the ϵ-global Differential Privacy constraint for fixed-confidence Best Arm Identification. Our algorithm can be deployed in data-sensitive applications, such as adaptive clinical trials or user studies.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: Theorem 2 is proven in Appendix D. Theorem 4 is based on numerous results proven in Appendix G. Lemma 5 is proven in Appendix E. Theorem 6 is proven in Appendix F. Theorem 7 is proven in Appendix H. All the other results mentioned in the main paper are proven theoretically in Appendices D, F, G, H and I.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Implementation details on our experiments are given in Section 6 and Appendix J. The DP-TT algorithm is summarized in Algorithm 1, and includes default choices of hyper-parameters. Our experiments are reproducible and the code is provided as supplementary material.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The code for our experiments is provided as supplementary material, and we provide sufficient instructions to faithfully reproduce our experimental results. As we rely on synthetic experiments on Bernoulli distributions, there is no dataset per se.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The code for our experiments is provided as supplementary material. We provide sufficient implementation details in Section 6 and Appendix J. The DP-TT algorithm is summarized in Algorithm 1, and includes default choices of hyper-parameters.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

Justification: The figures in Section 6 and Appendix J report the empirical means over 1000 runs for each algorithm. The 2-sigma error bars are based on the standard deviation of the means over those runs.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix J contains detailed information as regards the computational ressources used for our experiments. The specific institutional cluster is omitted for the sake of anonymity. Overall, our experiments are relatively computationally inexpensive, and they could be run on a professional laptop.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: After reading it, we confirm that the research conducted in our paper complies with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.

- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Our work addresses the ϵ-global Differential Privacy constraint for fixed-confidence Best Arm Identification. Our algorithm can be deployed in data-sensitive applications, such as adaptive clinical trials or user studies. Therefore, our paper can have a significant positive societal impact.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: To the best of our understanding, our paper doesn't necessitate safeguards. Our experiments use synthetic Bernoulli data and our algorithm satisfy the ϵ-global Differential Privacy constraint.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: Our paper does not rely on existing assets. Our experiments use synthetic Bernoulli data and our algorithm was implemented from scratch.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: Our paper does not introduce new assets. The code for our experiments is provided as supplementary material, yet it does not constitute an asset per se.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: Our paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: Our paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: Our theoretical and empirical contributions do not involve LLMs.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.

A Outline

The appendices are organized as follows:

- Notation are summarized in Appendix B.
- A detailed related work and an extended discussion is given in Appendix C.
- The lower bound on the expected sample complexity under ϵ-global DP (Theorem 2) is proven in Appendix D.
- The proof of Lemma 5 is given in Appendix E.
- The proof of our concentration results are detailed in Appendix F. In particular, this includes the proof of Theorem 6.
- Appendix G gathers key properties on the (resp. modified) divergence $d_\epsilon^\pm$ (resp. $\widetilde{d}_\epsilon^\pm$), the (resp. modified) transportation costs $W_{\epsilon,a,b}$ (resp. $\widetilde{W}_{\epsilon,a,b}$) and (resp. β-)characteristic times $T_\epsilon^\star(\boldsymbol{\nu})$ (resp. $T_{\epsilon,\beta}^\star(\boldsymbol{\nu})$) and their (resp. β-)optimal allocation $w_\epsilon^\star(\boldsymbol{\nu})$ (resp. $w_{\epsilon,\beta}^\star(\boldsymbol{\nu})$). In particular, this includes the proof of Theorem 4 based on Lemmas 43 and 47.
- The proof of the upper bound on the asymptotic expected sample complexity of DP-TT (Theorem 7) is given in Appendix H.
- In Appendix I, we propose variants of algorithms to tackle ϵ-global DP BAI. We aim at providing several choices for the interested practitioners.
- Implementation details and additional experiments are presented in Appendix J.

Table 1: Notation for the setting.

Notation	Type	Description
K	$\mathbb{N}$	Number of arms
$\mathcal{F}$	$\subseteq \mathcal{P}([0,1])$	Class of Bernoulli distributions
ν_a	$\mathcal{F}$	Bernoulli distribution of arm $a \in [K]$
$\boldsymbol{\nu}$	$\mathcal{F}^K$	Vector of Bernoulli distributions, $\nu := (\nu_a)_{a\in[K]}$
μ_a	$(0,1)$	Mean of arm $a \in [K]$
μ	$(0,1)^K$	Vector of means, $\mu := (\mu_a)_{a\in[K]}$
$a^\star(\mu), a^\star(\boldsymbol{\nu})$	$\subseteq [K]$	Set of best arms, $a^\star(\boldsymbol{\nu}) = a^\star(\mu) := \arg\max_{a\in[K]} \mu_a$
$a^\star$	$[K]$	Unique best arm, i.e., $a^\star(\mu) = \{a^\star\}$
ϵ	$\mathbb{R}_+^\star$	Privacy budget for ϵ-global DP
δ	$(0,1)$	Risk for δ-correctness
$\mathrm{Alt}(\boldsymbol{\nu})$	$\subseteq \mathcal{F}^K$	Alternative instances with different best arms

B Notation

We recall some commonly used notation: the set of integers $[n] := \{1, \cdots, n\}$, the complement $X^\complement$ and interior $\mathring{X}$ of a set X, the indicator function $\mathbb{1}(X)$ of an event, the probability $\mathbb{P}_{\boldsymbol{\nu}\pi}$ and the expectation $\mathbb{E}_{\boldsymbol{\nu}\pi}$ taken over the randomness of the observations from $\boldsymbol{\nu}$ and the algorithm π, Landau's notation o, $\mathcal{O}$, Ω and Θ, the $(K-1)$-dimensional probability simplex $\triangle_K := \left\{w \in \mathbb{R}_+^K \mid w \geq 0,\ \sum_{i\in[K]} w_i = 1\right\}$. The functions $[x]_0^1 := \max\{0, \min\{1, x\}\}$, $k_\eta(x) := 1 + \log_{1+\eta} x$, $\overline{W}_{-1}$ in Lemma 52, h in Eq. (31), r in Eq. (33), ζ is the Riemann ζ function. Moreover, we recall the definitions: $d_\epsilon^\pm$ in Eq. (3), d_ϵ in Eq. (1), $\widetilde{d}_\epsilon^\pm$ in Eq. (32), $W_{\epsilon,a,b}^\pm$ in Eq. (4), $\widetilde{W}_{\epsilon,a,b}^\pm$ in Eq. (34), $(T_\epsilon^\star(\boldsymbol{\nu}), T_{\epsilon,\beta}^\star(\boldsymbol{\nu}), w_\epsilon^\star(\boldsymbol{\nu}), w_{\epsilon,\beta}^\star(\boldsymbol{\nu}))$ in Eq.35. While Table 1 gathers problem-specific notation, Table 2 groups notation for the algorithms.

Table 2: Notation for the algorithm.

Notation	Type	Description
B_n	$[K]$	(EB) Leader at time n
C_n	$[K]$	(TC) Challenger at time n
a_n	$[K]$	Arm sampled at time n
X_{n,a_n}	$\{0,1\}$	Sample observed at the end of time n, i.e. $X_{n,a_n} \sim \nu_{a_n}$
$Y_{k_{n,a},a}$	$\mathbb{R}$	Noisy perturbation drawn at the beginning of phase $k_{n,a}$ for arm a, i.e. $Y_{k_{n,a},a} \sim \text{Lap}(1/\epsilon)$
$\mathcal{F}_n$		History before time n
$\tilde{a}_n$	$[K]$	Arm recommended before time n
$\tau_{\epsilon,\delta}$	$\mathbb{N}$	Sample complexity (stopping time)
$c(n,\epsilon,\delta)$	$\mathbb{N} \times \mathbb{R}_+^\star \times (0,1) \to \mathbb{R}_+^\star$	Stopping threshold function
$c_1(n,\delta)$	$\mathbb{N} \times (0,1) \to \mathbb{R}_+^\star$	Stopping threshold function
$c_2(n,\epsilon)$	$\mathbb{N} \times \mathbb{R}_+^\star \to \mathbb{R}_+^\star$	Approximation threshold function
$N_{n,a}$	$\mathbb{N}$	Number of pulls of arm a before time n
$k_{n,a}$	$\mathbb{N}$	Current phase of arm a at time n
$T_k(a)$	$\mathbb{N}$	Time n where the arm a changes to phase k
$\tilde{S}_{k,a}$	$\mathbb{R}$	Private sum of observations for arm a at phase k
$\tilde{N}_{n,a}$	$\mathbb{N}$	Number of pulls of arm a at the beginning of phase $k_{n,a}$
$\tilde{\mu}_{n,a}$	$\mathbb{R}$	Private estimator of the empirical mean of arm a at the beginning of phase $k_{n,a}$
$L_{n,a}$	$\mathbb{N}$	Counts of $B_t = a$ before time n
$N_{n,a}^a$	$\mathbb{N}$	Counts of $(B_t, a_t) = (a,a)$ before time n
β	$(0,1)$	Fixed proportion

C Related Work and Extended Discussion

We provide a more detailed literature review in Appendix C.1, and an extended discussion in Appendix C.2.

C.1 Related Work

Structured Bandits. While we consider unstructured bandits [Auer et al., 2002], numerous structural assumptions have been studied: linear bandits [Soare et al., 2014], generalized linear bandits [Filippi et al., 2010] such as logistic bandits [Jun et al., 2021], combinatorial bandits [Chen et al., 2013], sparse bandits [Jamieson et al., 2015], spectral bandits [Kocák and Garivier, 2021], unimodal bandits [Combes and Proutière, 2014], Lipschitz [Magureanu et al., 2014], partial monitoring [Audibert and Bubeck, 2010a], etc. Coping for the structural assumption while preserving ϵ-global DP is an interesting direction for future works.

Pure Exploration Problems. While we consider only BAI [Even-Dar et al., 2002], other pure exploration problems have been studied in the literature: ϵ-BAI [Mannor and Tsitsiklis, 2004], thresholding bandits [Carpentier and Locatelli, 2016], Top-k identification [Katz-Samuels and Scott, 2019], Pareto set identification [Auer et al., 2016], best partition identification [Chen et al., 2017], etc. Extending our ϵ-global DP results to answer these identification problems is an interesting research direction.

Performance Metrics. In pure exploration problems, the two major theoretical frameworks are the *fixed-confidence* setting [Even-Dar et al., 2006; Jamieson and Nowak, 2014; Garivier and Kaufmann, 2016], which is the focus of this paper, and the *fixed-budget* setting [Audibert et al., 2010; Gabillon et al., 2012]. In the fixed-budget setting, the objective is to minimize the probability of misidentifying a correct answer with a fixed number of samples T. Recent works have also considered the anytime setting, in which the agent aims at achieving a low probability of error at any deterministic time [Zhao et al., 2023; Jourdan et al., 2024]. Extending our findings to support ϵ-global DP in the fixed-budget or the anytime setting is an interesting direction for future works, see e.g., Chen et al. [2024].

DP in Bandits. DP has been studied for multi-armed bandits under different bandit settings: finite-armed stochastic [Mishra and Thakurta, 2015; Sajed and Sheffet, 2019; Zheng et al., 2020; Hu et al., 2021; Azize and Basu, 2022; Hu and Hegde, 2022; Azize and Basu, 2024; Wang and Zhu, 2024; Hu et al., 2025], adversarial [Thakurta and Smith, 2013; Agarwal and Singh, 2017; Tossou and Dimitrakakis, 2017], linear [Hanna et al., 2024; Li et al., 2022; Azize and Basu, 2024], contextual linear [Shariff and Sheffet, 2018; Neel and Roth, 2018; Zheng et al., 2020; Azize and Basu, 2024], and kernel bandits [Pavlovic et al., 2025], among others. Most of these works were for regret minimisation, but the problem has also been explored for best-arm identification, with fixed confidence [Azize et al., 2023, 2024] and fixed budget [Chen et al., 2024]. The problem has also been studied under three different DP trust models: (a) global DP where the users trust the centralised decision maker [Mishra and Thakurta, 2015; Shariff and Sheffet, 2018; Sajed and Sheffet, 2019; Azize and Basu, 2022; Hu and Hegde, 2022], (b) local DP where each user deploys a local perturbation mechanism to send a "noisy" version of the rewards to the policy [Basu et al., 2019; Zheng et al., 2020; Han et al., 2021], and (c) shuffle DP where users still feed their data to a local perturbation, but now they trust an intermediary to apply a uniformly random permutation on all users' data before sending to the central servers [Tenenbaum et al., 2021; Garcelon et al., 2022; Chowdhury and Zhou, 2022].

In the first papers on DP for bandits, the tree-based mechanism [Dwork et al., 2010; Chan et al., 2011] was used to compute the sum of rewards privately. However, this mechanism was proven to be sub-optimal, matching the lower bounds up to logarithmic factors. Then, forgetting was first proposed by Sajed and Sheffet [2019] to get rid of the tree-based mechanism, then adapted to UCB in [Hu et al., 2021; Azize and Basu, 2022]. Finally, if the KL is the divergence that controls the complexity of bandits without privacy [Lai and Robbins, 1985; Garivier and Kaufmann, 2016], then Azize and Basu [2022] were the first to show that the TV controls the complexity of private bandits, in the high privacy regime.

In this paper, *we focus on ϵ-pure DP, under a global trust model, in stochastic finite-armed bandits, for best arm identification under fixed confidence*.

Gap in the Literature. This problem setting is first studied by Azize et al. [2023], who proposed the first problem-dependent sample complexity lower bound, and introduced AdaP-TT, an ϵ-global DP version of the Top Algorithm. However, the sample complexity upper bound of AdaP-TT only matches the lower bound in *the high privacy regime* $\epsilon \to 0$, and for instances where the means have similar order (see Condition 1 in [Azize et al., 2023] in the discussion after Theorem 5 in [Azize et al., 2023]).

Azize et al. [2024] proposes AdaP-TT*, an improved version of AdaP-TT. The improvement is achieved by using a transport inspired by the sample complexity lower bound from [Azize et al., 2023]. Using the new transport, AdaP-TT* gets rid of Condition 1 needed by AdaP-TT, and achieves the high privacy lower bound for all instances up to a multiplicative factor 48.

However, both AdaP-TT and AdaP-TT* do not match the lower bound, beyond the high privacy regime, i.e. for both the low privacy regime and transitional regimes.

C.2 Extended Discussions

Limitations of Theorem 7 Using adaptive targets $\beta_n(B_n, C_n)$ in DP-TT could replace $T^\star_{\epsilon,\beta}(\boldsymbol{\nu})$ by $T^\star_\epsilon(\boldsymbol{\nu})$, which is the optimal scaling asymptotically (Theorem 2). While we propose two adaptive choices of target based on IDS [You et al., 2023] or BOLD [Bandyopadhyay et al., 2024] (Appendix I), we leave their analysis for future work. Based on finer concentration results, the stopping threshold could scale asymptotically as $\log(1/\delta)$ instead of $2\log(1/\delta)$, hence improving by a factor 2. The assumption that the means are distinct is used to prove sufficient exploration; it can be removed by using forced exploration or a fine-grained analysis [Jourdan and Degenne, 2024; Jourdan et al., 2024]. Provided these improvements are proven, the multiplicative factor would be reduced to $1+\eta$, where η is the parameter that controls the geometrically increasing phases. While it improves the asymptotic upper bound, choosing η too close to 0 negatively impacts the performance of DP-TT, due to the dependency in $\mathcal{O}(1/\log(1+\eta))$ of the stopping threshold.

Beyond Bernoulli Distributions The non-asymptotic lower bound on the expected sample complexity (Theorem 2) holds for any class of distributions $\mathcal{F}$, and thus is already true beyond Bernoullis (e.g. exponential families, sub-Gaussians, etc). However, by going beyond Bernoullis, d_ϵ might not

admit a closed-form solution, e.g., the TV between Gaussian distributions has no simple formula. We conjecture that most regularity properties used to derive Theorem 4, and needed for the upper bound proofs, also hold for other classes of distributions. From a theoretical perspective, this might render the proof of the sufficient regularity properties more challenging. From a practical perspective, this increases the computational cost of the stopping rule and the challenger arm, as they require computing an empirical transportation cost based on the d_ϵ divergence. Finally, since the objective function inside the d_ϵ is continuous and convex over a compact interval, using off-the-shelf convex optimisation solvers to compute d_ϵ is still straightforward.

Our privacy analysis (Lemma 8) holds for any distribution with support in $[0, 1]$, and thus is already true beyond Bernoullis. It is straightforward to extend this to any distribution with a support in $[a, b]$, by multiplying the noise terms by the range $b - a$. Also, all prior works in bandits with DP are either for Bernoulli or bounded distributions, for the simple reason that this assumption makes estimating the empirical mean privately using the Laplace mechanism straightforward, as the sensitivity is controlled. This helps focus on the more interesting tradeoffs between DP, exploration and exploitation, without any additional technical overheads that may be introduced by estimating privately the mean of unbounded distributions.

The concentration results (Theorem 6) used to derive a stopping threshold ensuring δ-correctness are highly specific to Bernoulli distributions. However, the proof builds on fine-grained tail bounds of the sum of two independent random variables, i.e., we bound those probabilities by the maximal product between their respective survival functions (Lemma 10). Therefore, the technical tools used for this intermediate result can be also used to study other one-parameter exponential families, e.g., Gaussian observation, with other noise mechanisms, e.g., Gaussian noise.

The proof of Theorem 7 (Appendix H) builds on the unified analysis of the Top Two algorithm Jourdan et al. [2022], coping with many classes of distributions, e.g., one-parameter exponential families and non-parametric bounded distributions. It relies heavily on the derived regularity properties (Appendix G) for the signed divergences, transportation costs, characteristic times, and optimal allocations. Based on the non-private BAI literature, we conjecture that most regularity properties used to derive Theorem 7 also hold for other classes of distributions. Due to the lack of explicit formulae, the proofs are challenging. The DP-TT algorithm becomes computationally costly due to the intertwined optimisation procedure when computing the transportation costs numerically.

Beyond BAI The technical arguments used in the proof of Theorem 2 allow obtaining a similar lower bound for (i) one-parameter exponential families, e.g., Gaussian, (ii) pure exploration problems with a unique correct answer, e.g., top-m best arm identification, and (iii) structured instances without arm correlation, e.g., unimodal bandits. This is straightforwardly done by adapting the definition of $\mathrm{Alt}(\boldsymbol{\nu})$, $a^\star(\boldsymbol{\nu})$, and $\mathcal{F}$ in Eq. (2). For settings that do not fall into the above three categories, the characteristic time requires more subtle modifications; yet our intermediate technical results can still be used. Based on the non-private fixed-confidence literature, we conjecture the changes of characteristic times described below.

- For bounded distribution (non-parametric) or Gaussian with unknown variance (two-parameters exponential family), the KL should be replaced by an infimum over KL Agrawal et al. [2020]; Jourdan et al. [2023]. The plurality of distributions having the same mean implies the existence of a distribution that is the most confusing given a specific constraint on the mean, captured by the Kinf.
- For γ-Best Arm Identification (multiple correct answers), an outer maximization over all the correct answers should be added Degenne and Koolen [2019]. The plurality of correct answers implies the existence of a correct answer that is the easiest to verify.
- For linear bandits (correlated means), the $\inf$ in the definition of d_ϵ per-arm should be replaced by a joint infimum over all the arms and moved outside the sum over arms Degenne et al. [2020a].

Beyond ϵ-DP Our lower bound technique (Theorem 2) can be seen as a generalisation of the packing argument, and thus depends on the group privacy property. Specifically, in Appendix D, just before Eq. (11), we used the group privacy of ϵ-DP to bound $\mathrm{KL}(\mathcal{M}_D, \mathcal{M}_{D'}) \leq \epsilon \mathrm{dham}(D, D')$.

- For ρ-zCDP, there is a similar group privacy property that can directly be plugged here, stating that $\mathrm{KL}(\mathcal{M}_D, \mathcal{M}_{D'}) \leq \rho \mathrm{dham}(D, D')^2$. This means that for ρ-zCDP, $\epsilon \times \mathrm{dham}$ is replaced by $\rho \times \mathrm{dham}^2$ in the fundamental optimal transport inequality of Eq. (11). We leave solving tightly this new optimal transport for future work.
- For (ϵ, δ)-DP, the group privacy property is not tight. This is a classic issue in (ϵ, δ)-DP lower bounds, where other techniques are used (e.g., fingerprinting). Adapting these techniques for bandits is an interesting open problem (even for bandits with regret minimisation).

It is straightforward to make DP-TT achieve either ρ-zCDP or (ϵ, δ)-DP, by replacing the Laplace mechanism with the Gaussian mechanism. To design a correct stopping rule with the Gaussian mechanism, it is possible to use the same fine-grained tail bounds of the sum of two independent random variables (Lemma 10) by only replacing the concentration of Laplace random variables with the concentration of Gaussian random variables. In the Top Two family of algorithms, an important design choice is the transportation cost, used for stopping and sampling the challenger. DP-TT uses a d_ϵ based transport, inspired by the lower bound of Theorem 2. Thus, to go beyond ϵ-DP for algorithm design, it is important to derive a tight lower bound that suggests the use of a new transportation cost.

Behavior with Near-optimal Arms The unique best arm assumption is standard in the fixed-confidence BAI setting. Even in the non-private setting, this assumption is necessary to obtain a δ-correct algorithm with finite expected stopping time, as the characteristic time becomes infinite otherwise. In the private setting, this phenomenon persists as can be seen by the lower bound in Eq. (5), i.e., $T^\star_\epsilon(\nu) \to +\infty$ when $\Delta_{\epsilon,a^\star} \to 0$. In near-tie scenarios, DP-TT will perform as badly as any other δ-correct and ϵ-global DP algorithms due to the fundamental lower bound. Empirically, this will be reflected by empirical transportation costs that are too small for the stopping condition to be met. In applications where the near-tie scenario is common, practitioners consider γ-Best Arm Identification, i.e., identify an arm that is γ-close to the best arm. For the non-private fixed-confidence γ-BAI setting, we refer to Jourdan et al. [2024] for references and guarantees satisfied by Top Two algorithms. To tackle γ-BAI, DP-TT should be modified by using the appropriate transportation costs in both the stopping rule and the definition of the challenger, i.e., $W_{\epsilon,\gamma,a,b}$ instead of $W_{\epsilon,a,b}$.

Implications of Not Forgetting DP-TT has the same memory complexity as the forgetting algorithms. The reason is that DP-TT only needs to store one accumulated noisy sum of rewards across phases, while the forgetting ones store the noisy sum of rewards of only the last phase. For privacy considerations, under our threat model where the adversary only observes the output of the algorithm (sequence of actions, final recommendation and stopping time), both forgetting and non-forgetting algorithms provide the same DP guarantees thanks to post-processing.

One can imagine other threat models where forgetting the rewards provides better privacy guarantees. One possible example of this is the pan-private threat model Dwork et al. [2010], where the adversary can also intrude into the internal states of the algorithm during its execution. In this threat model, if an adversary observes the internal states of the execution of DP-TT at some phase $\ell > 1$, they can see the full sum of rewards and maybe infer the membership of, say, the first user (up to tradeoffs guaranteed by the ϵ-DP constraint, since the sum is noisy). However, for the forgetting algorithms, if the adversary observes the internal states of the execution of forgetting algorithms at some phase $\ell > 1$, the adversary can infer nothing about the reward of the first patient, since its reward has been deleted completely.

D Lower Bound

Let $\mathcal{M} : \mathcal{X}^n \to \mathcal{O}$ be an ϵ-DP mechanism. For $D \in \mathcal{X}^n$ an input dataset, we denote by $\mathcal{M}_D$ the distribution over outputs, when the input is D, and $\mathcal{M}_D(E)$ the probability of observing output E when the input is D.

Let $\mathbb{P}$ and $\mathbb{Q}$ be two data-generating distributions over $\mathcal{X}^n$. We denote by $\mathbb{M}_{\mathbb{P},\mathcal{M}}$ the marginal over outputs of the mechanism $\mathcal{M}$ when the input dataset is generated through $\mathbb{P}$, i.e.

$$\mathbb{M}_{\mathbb{P},\mathcal{M}}(A) := \int_{D \in \mathcal{X}^n} \mathcal{M}_D(A) \, \mathrm{d}\mathbb{P}(D) \, , \tag{9}$$

for any event A in the output space. We define similarly $\mathbb{M}_{\mathbb{Q},\mathcal{M}}$ the marginal over outputs of the mechanism $\mathcal{M}$ when the input dataset is generated through $\mathbb{Q}$.

The main question is to control the divergence $\mathrm{KL}\left(\mathbb{M}_{\mathbb{P},\mathcal{M}} \parallel \mathbb{M}_{\mathbb{Q},\mathcal{M}}\right)$ when the mechanism $\mathcal{M}$ satisfies DP. In general, for any mechanism $\mathcal{M}$, the data-processing inequality provides the following bound

$$\mathrm{KL}\left(\mathbb{M}_{\mathbb{P},\mathcal{M}} \parallel \mathbb{M}_{\mathbb{Q},\mathcal{M}}\right) \leq \mathrm{KL}\left(\mathbb{P} \parallel \mathbb{Q}\right) . \tag{10}$$

Now, for ϵ-DP mechanisms, we want to translate the DP constraint to a tight bound on the divergence $\mathrm{KL}\left(\mathbb{M}_{\mathbb{P},\mathcal{M}} \parallel \mathbb{M}_{\mathbb{Q},\mathcal{M}}\right)$. To do so, let $\mathbb{L}$ be any other distribution on $\mathcal{X}^n$. Let $\mathbb{C}_{\mathbb{P},\mathbb{L}}$ be a coupling of $(\mathbb{P},\mathbb{L})$, i.e., the marginals of $\mathbb{C}_{\mathbb{P},\mathbb{L}}$ are $\mathbb{P}$ and $\mathbb{L}$. We can now rewrite our the marginals using the definition of couplings. For $\mathbb{M}_{\mathbb{P},\mathcal{M}}$, we have

$$\mathbb{M}_{\mathbb{P},\mathcal{M}}(A) := \int_{D\in\mathcal{X}^n} \mathcal{M}_D\left(A\right) \,\mathrm{d}\mathbb{P}\left(D\right) = \int_{D,D'\in\mathcal{X}^n} \mathcal{M}_D\left(A\right) \,\mathrm{d}\mathbb{C}_{\mathbb{P},\mathbb{L}}\left(D,D'\right) ,$$

and for $\mathbb{Q}$ we get

$$\begin{aligned}\mathbb{M}_{\mathbb{Q},\mathcal{M}}(A) := \int_{D'\in\mathcal{X}^n} \mathcal{M}_{D'}\left(A\right) \,\mathrm{d}\mathbb{Q}\left(D'\right) &= \int_{D'\in\mathcal{X}^n} \mathcal{M}_{D'}\left(A\right) \frac{\mathrm{d}\mathbb{Q}\left(D'\right)}{\mathrm{d}\mathbb{L}\left(D'\right)} \,\mathrm{d}\mathbb{L}\left(D'\right) \\ &= \int_{D,D'\in\mathcal{X}^n} \mathcal{M}_{D'}\left(A\right) \frac{\mathrm{d}\mathbb{Q}\left(D'\right)}{\mathrm{d}\mathbb{L}\left(D'\right)} \,\mathrm{d}\mathbb{C}_{\mathbb{P},\mathbb{L}}\left(D,D'\right) .\end{aligned}$$

Using the data-processing inequality, we get

$$\begin{aligned}\mathrm{KL}\left(\mathbb{M}_{\mathbb{P},\mathcal{M}} \parallel \mathbb{M}_{\mathbb{Q},\mathcal{M}}\right) &\leq \int_{D,D'\in\mathcal{X}^n}\int_{o\in\mathcal{O}} \log\left(\frac{\mathcal{M}_D\left(o\right)}{\mathcal{M}_{D'}\left(o\right)\frac{\mathrm{d}\mathbb{Q}(D')}{\mathrm{d}\mathbb{L}(D')}}\right) \mathcal{M}_D\left(o\right) \,\mathrm{d}o\,\mathrm{d}\mathbb{C}_{\mathbb{P},\mathbb{L}}\left(D,D'\right) \\ &= \int_{D,D'\in\mathcal{X}^n} \left(\mathrm{KL}\left(\mathcal{M}_D \parallel \mathcal{M}_{D'}\right) + \log\left(\frac{\mathrm{d}\mathbb{L}\left(D'\right)}{\mathrm{d}\mathbb{Q}\left(D'\right)}\right)\right) \mathrm{d}\mathbb{C}_{\mathbb{P},\mathbb{L}}\left(D,D'\right) \\ &= \mathbb{E}_{D,D'\sim\mathbb{C}_{\mathbb{P},\mathbb{L}}}\left[\mathrm{KL}\left(\mathcal{M}_D \parallel \mathcal{M}_{D'}\right)\right] + \mathrm{KL}\left(\mathbb{L} \parallel \mathbb{Q}\right) .\end{aligned}$$

Since this is true for any coupling $\mathbb{C}_{\mathbb{P},\mathbb{L}}$ and any distribution $\mathbb{L}$, we get the final bound

$$\mathrm{KL}\left(\mathbb{M}_{\mathbb{P},\mathcal{M}} \parallel \mathbb{M}_{\mathbb{Q},\mathcal{M}}\right) \leq \inf_{\mathbb{L}\in\mathcal{P}(\mathcal{X}^n)} \left\{ \inf_{\mathbb{C}_{\mathbb{P},\mathbb{L}}\in\mathcal{C}(\mathbb{P},\mathbb{L})} \left\{\mathbb{E}_{D,D'\sim\mathbb{C}_{\mathbb{P},\mathbb{L}}}\left[\mathrm{KL}\left(\mathcal{M}_D \parallel \mathcal{M}_{D'}\right)\right]\right\} + \mathrm{KL}\left(\mathbb{L} \parallel \mathbb{Q}\right)\right\}$$

where $\mathcal{P}(\mathcal{X}^n)$ is the set of all distributions over $\mathcal{X}^n$ and $\mathcal{C}\left(\mathbb{P},\mathbb{L}\right)$ is the set of all couplings between $\mathbb{P}$ and $\mathbb{L}$. Using that the $\mathcal{M}$ is ϵ-DP, we can use the simple bound $\mathrm{KL}\left(\mathcal{M}_D \parallel \mathcal{M}_{D'}\right) \leq \epsilon d_{\mathrm{Ham}}(D,D')$ which gives

$$\mathrm{KL}\left(\mathbb{M}_{\mathbb{P},\mathcal{M}} \parallel \mathbb{M}_{\mathbb{Q},\mathcal{M}}\right) \leq \inf_{\mathbb{L}\in\mathcal{P}(\mathcal{X}^n)} \left\{ \epsilon \inf_{\mathbb{C}_{\mathbb{P},\mathbb{L}}\in\mathcal{C}(\mathbb{P},\mathbb{L})} \left\{\mathbb{E}_{D,D'\sim\mathbb{C}_{\mathbb{P},\mathbb{L}}}\left[d_{\mathrm{Ham}}(D,D')\right]\right\} + \mathrm{KL}\left(\mathbb{L} \parallel \mathbb{Q}\right)\right\} . \tag{11}$$

D.1 Product Distributions

Suppose that $\mathbb{P} := \bigotimes_{i=1}^n \mathbb{P}_i$ and $\mathbb{Q} := \bigotimes_{i=1}^n \mathbb{Q}_i$ are product distributions. Consider the subset of product distributions $\mathbb{L} := \bigotimes_{i=1}^n \mathbb{L}_i$, and the maximal coupling $\mathbb{C}_\infty(\mathbb{P},\mathbb{L}) := \prod_{i=1}^n \mathbb{C}_\infty(\mathbb{P}_i,\mathbb{L}_i)$. Plugging these in Equation (11), we get

$$\begin{aligned}\mathrm{KL}\left(\mathbb{M}_{\mathbb{P},\mathcal{M}} \parallel \mathbb{M}_{\mathbb{Q},\mathcal{M}}\right) &\leq \inf_{\mathbb{L}_1,\ldots,\mathbb{L}_n} \left\{ \epsilon \sum_{i=1}^n \mathbb{E}_{D_i,D_i'\sim\mathbb{C}_\infty(\mathbb{P}_i,\mathbb{L}_i)}\left[\mathbb{1}\left\{D_i \neq D_i'\right\}\right] + \sum_{i=1}^n \mathrm{KL}\left(\mathbb{L}_i \parallel \mathbb{Q}_i\right)\right\} \\ &= \inf_{\mathbb{L}_1,\ldots,\mathbb{L}_n} \left\{ \sum_{i=1}^n \left(\epsilon\,\mathrm{TV}\left(\mathbb{P}_i \parallel \mathbb{L}_i\right) + \mathrm{KL}\left(\mathbb{L}_i \parallel \mathbb{Q}_i\right)\right)\right\} \\ &= \sum_{i=1}^n \inf_{\mathbb{L}_i\in\mathcal{P}(\mathcal{X})} \left\{\epsilon\,\mathrm{TV}\left(\mathbb{P}_i \parallel \mathbb{L}_i\right) + \mathrm{KL}\left(\mathbb{L}_i \parallel \mathbb{Q}_i\right)\right\} = \sum_{i=1}^n \mathrm{d}_\epsilon(\mathbb{P}_i,\mathbb{Q}_i) ,\end{aligned}$$

where

$$\mathrm{d}_\epsilon(\mathbb{P},\mathbb{Q}) := \inf_{\mathbb{L}\in\mathcal{P}(\mathcal{X})} \left\{\epsilon\,\mathrm{TV}\left(\mathbb{P} \parallel \mathbb{L}\right) + \mathrm{KL}\left(\mathbb{L} \parallel \mathbb{Q}\right)\right\} . \tag{12}$$

Algorithm 2 Bandit interaction between a policy and an environment

Input: A policy π and an environment $\nu \triangleq (P_a : a \in [K])$
for $t = 1, \dots$ **do**
    The policy samples an action $a_t \sim \pi_t(. \mid a_1, r_1, \dots, a_{t-1}, r_{t-1})$
    The policy observes a reward $r_t \sim P_{a_t}$
end for
if Regret minimisation **then**
    The interaction ends after T steps
else FC-BAI
    The policy decides to stop the interaction at step $\tau_{\epsilon,\delta}$ and recomends the final guess $\hat{a}$
end if

D.2 Sequential KL decomposition for bandits under DP

In this section, we adapt the techniques from product distributions to bandit marginals.

First, we introduce the bandit canonical model.

The bandit canonical model. A stochastic bandit (or environment) is a collection of distributions $\nu \triangleq (P_a : a \in [K])$, where $[K]$ is the set of available K actions. The learner and the environment interact sequentially over T rounds. In each round $t \in 1, \dots, T$, the learner chooses an action $a_t \in [K]$, which is fed to the environment. The environment then samples a reward $r_t \in \mathbb{R}$ from distribution P_{a_t} and reveals r_t to the learner. The interaction between the learner (or policy) and environment induces a probability measure on the sequence of outcomes $H_T \triangleq (a_1, r_1, a_2, r_2, \dots, a_T, r_T)$. In the following, we construct the probability space that carries these random variables.

Let $T \in \mathbb{N}^\star$ be the horizon. Let $\nu = (P_a : a \in [K])$ a bandit instance with $K \in \mathbb{N}^\star$ finite arms, and each P_a is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with $\mathfrak{B}$ being the Borel set. For each $t \in [T]$, let $\Omega_t = ([K] \times \mathbb{R})^t \subset \mathbb{R}^{2t}$ and $\mathcal{F}_t = \mathfrak{B}(\Omega_t)$. We first formalise the definition of a policy.

Definition 2 (The policy)**.** *A policy π is a sequence $(\pi_t)_{t=1}^T$, where π_t is a probability kernel from $(\Omega_t, \mathcal{F}_t)$ to $([K], 2^{[K]})$. Since $[K]$ is discrete, we adopt the convention that for $a \in [K]$,*

$$\pi_t(a \mid a_1, r_1, \dots, a_{t-1}, r_{t-1}) = \pi_t(\{a\} \mid a_1, r_1, \dots, a_{t-1}, r_{t-1})$$

We want to define a probability measure on $(\Omega_T, \mathcal{F}_T)$ that respects our understanding of the sequential nature of the interaction between the learner and a stationary stochastic bandit. Specifically, the sequence of outcomes should satisfy the following two assumptions:

(a) The conditional distribution of action a_t given $a_1, r_1, \dots, a_{t-1}, r_{t-1}$ is $\pi(a_t \mid H_{t-1})$ almost surely.

(b) The conditional distribution of reward r_t given $a_1, r_1, \dots, a_{t-1}, r_{t-1}, a_t$ is P_{a_t} almost surely.

The probability measure on $(\Omega_T, \mathcal{F}_T)$ depends on both the environment ν and the policy π. To construct this probability, let λ be a σ-finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ for which P_a is absolutely continuous with respect to λ for all $a \in [K]$. Let $p_a = dP_a/d\lambda$ be the Radon–Nikodym derivative of P_a with respect to λ. Letting ρ be the counting measure with $\rho(B) = |B|$, the density $p_{\nu\pi} : \Omega_T \to \mathbb{R}$ can now be defined with respect to the product measure $(\rho \times \lambda)^T$ by

$$p_{\nu\pi}(a_1, r_1, \dots, a_T, r_T) \triangleq \prod_{t=1}^{T} \pi_t(a_t \mid a_1, r_1, \dots, a_{t-1}, r_{t-1}) p_{a_t}(r_t)$$

and $\mathcal{P}_{\nu\pi}$ is defined as

$$\mathbb{P}_{\nu\pi}(B) \triangleq \int_B p_{\nu\pi}(\omega)(\rho \times \lambda)^T(\,\mathrm{d}\omega) \quad \text{forall} B \in \mathcal{F}_T$$

Hence $(\Omega_T, \mathcal{F}_T, \mathbb{P}_{\nu\pi})$ is a probability space over histories induced by the interaction between π and ν. We define also a marginal distribution over the sequence of actions by

$$m_{\nu\pi}(a_1, \dots, a_T) \triangleq \int_{r_1, \dots, r_T} p_{\nu\pi}(a_1, r_1, \dots, a_T, r_T)\,\mathrm{d}r_1 \dots\,\mathrm{d}r_T,$$

and forall$C \in \mathcal{P}([K]^T)$,

$$\mathbb{M}_{\nu\pi}(C) \triangleq \sum_{(a_1,\ldots,a_T)\in C} m_{\nu\pi}(a_1, a_2, \ldots, a_T).$$

Finally, $([K]^T, \mathcal{P}([K]^T), \mathbb{M}_{\nu\pi})$ is a probability space over sequence of actions produced when π interacts with ν for T time-steps.

The KL upper bound. Now, we adapt the techniques for the bandit marginals. Let $\nu = \{P_a, a \in [K]\}$ and $\nu' = \{P'_a, a \in [K]\}$ be two bandit instances in $\mathcal{F}^K$. We recall that, when the policy π interacts with the bandit instance ν, it induces a marginal distribution $\mathbb{M}_{\nu\pi}$ over the sequence of actions. We define $\mathbb{M}_{\nu'\pi}$ similarly.

The goal is to upper bound the quantity $\mathrm{KL}\left(\mathbb{M}_{\nu\pi} \parallel \mathbb{M}_{\nu'\pi}\right)$. The marginals $\mathbb{M}_{\nu\pi}$ and $\mathbb{M}_{\nu\pi}$ in the sequential setting "look like" marginals generated by "product distributions". However, the hardness of the sequential setting lies in the fact that the data-generating distributions depend on the stochastic sequential actions chosen. Thus, the results of the previous section cannot be directly applied. To adapt the proof ideas of the previous section to the bandit case, we introduce the idea of a coupled bandit instance.

Let $\nu'' = \{P''_a : a \in [K]\}$ be any "intermediary" bandit instance from $\mathcal{F}^K$. Define c_a as the maximal coupling between P_a and P''_a, i.e., $c_a := \mathbb{C}_\infty(P_a, P''_a)$. Fix a policy $\pi = \{\pi_t\}_{t=1}^T$.

Here, we build a coupled environment γ of ν and ν''. The policy π interacts with the coupled environment γ up to a given time horizon T to produce an augmented history $\{(a_t, r_t, r''_t)\}_{t=1}^T$. The iterative steps of this interaction process are:

1. The probability of choosing an action $a_t = a$ at time t is dictated only by the policy π_t and $a_1, r_1, a_2, r_2, \ldots, a_{t-1}, r_{t-1}$, *i.e.*the policy ignores $\{r''_s\}_{s=1}^{t-1}$.
2. The distribution of rewards (r_t, r''_t) is c_{a_t} and is conditionally independent of the previous observed history $\{(a_s, r_s, r''_s)\}_{t=1}^{t-1}$.

This interaction is similar to the interaction process of policy π with the first bandit instance ν, with the addition of sampling an extra r''_t from the coupling of P_{a_t} and P''_{a_t}.

The distribution of the augmented history induced by the interaction of π and the coupled environment can be defined as

$$p_{\gamma\pi}(a_1, r_1, r''_1 \ldots, a_T, r_T, r''_T) := \prod_{t=1}^{T} \pi_t(a_t \mid a_1, r_1, \ldots, a_{t-1}, r_{t-1}) c_{a_t}(r_t, r''_t)\,.$$

To simplify the notation, let $\mathbf{a} := (a_1, \ldots, a_T)$, $\mathbf{r} := (r_1, \ldots, r_T)$, $\mathbf{r'} := (r'_1, \ldots, r'_T)$ and $\mathbf{r''} := (r'_1, \ldots, r'_T)$. Also, let $c_{\mathbf{a}}(\mathbf{r}, \mathbf{r''}) := \prod_{t=1}^T c_{a_t}(r_t, r''_t)$ and $\pi(\mathbf{a} \mid \mathbf{r}) := \prod_{t=1}^T \pi_t(a_t \mid a_1, r_1, \ldots, a_{t-1}, r_{t-1})$. We put $\mathbf{h} := (\mathbf{a}, \mathbf{r}, \mathbf{r''})$. With the new notation

$$p_{\gamma\pi}(\mathbf{a}, \mathbf{r}, \mathbf{r''}) := \pi(\mathbf{a} \mid \mathbf{r}) c_{\mathbf{a}}(\mathbf{r}, \mathbf{r''})\,.$$

By the definition of the couplings, we have that $m_{\nu\pi}$ is the marginal of $p_{\gamma\pi}$ when integrated over $(\mathbf{r}, \mathbf{r''})$, i.e.,

$$m_{\nu\pi}(\mathbf{a}) = \int_{\mathbf{r},\mathbf{r''}} p_{\gamma\pi}(\mathbf{a}, \mathbf{r}, \mathbf{r''})\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r''}\,.$$

Now, we define a new joint distribution $q_{\gamma\pi}$, inspired by the techniques used for product distributions:

$$q_{\gamma\pi}(\mathbf{a}, \mathbf{r}, \mathbf{r''}) := \pi(\mathbf{a} \mid \mathbf{r''}) \frac{p'_{\mathbf{a}}(\mathbf{r''})}{p''_{\mathbf{a}}(\mathbf{r''})} c_{\mathbf{a}}(\mathbf{r}, \mathbf{r''})\,,$$

where $p'_{\mathbf{a}}(\mathbf{r''}) := \prod_{t=1}^T p'_{a_t}(r''_t)$, and similarly, $p''_{\mathbf{a}}(\mathbf{r''}) := \prod_{t=1}^T p''_{a_t}(r''_t)$.

First, observe that it is indeed a valid joint distribution, i.e.

$$\sum_{\mathbf{a}} \int_{\mathbf{r},\mathbf{r''}} q_{\gamma\pi}(\mathbf{a}, \mathbf{r}, \mathbf{r''})\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r''} = \sum_{\mathbf{a}} \int_{\mathbf{r},\mathbf{r''}} \pi(\mathbf{a} \mid \mathbf{r''}) \frac{p'_{\mathbf{a}}(\mathbf{r''})}{p''_{\mathbf{a}}(\mathbf{r''})} c_{\mathbf{a}}(\mathbf{r}, \mathbf{r''})\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r''}$$

$$= \sum_{\mathbf{a}} \int_{\mathbf{r''}} \pi(\mathbf{a} \mid \mathbf{r''}) p'_{\mathbf{a}}(\mathbf{r''}) \, \mathrm{d}\mathbf{r''} = \int_{\mathbf{r''}} p'_{\mathbf{a}}(\mathbf{r''}) \, \mathrm{d}\mathbf{r''} = 1 \, ,$$

and that $m_{\nu'\pi}$ is the marginal of $q_{\gamma\pi}$ when integrated over $(\mathbf{r}, \mathbf{r''})$, i.e.,

$$\begin{aligned}\int_{\mathbf{r},\mathbf{r''}} q_{\gamma\pi}(\mathbf{a}, \mathbf{r}, \mathbf{r''}) \, \mathrm{d}\mathbf{r} \, \mathrm{d}\mathbf{r''} &= \int_{\mathbf{r},\mathbf{r''}} \pi(\mathbf{a} \mid \mathbf{r''}) \frac{p'_{\mathbf{a}}(\mathbf{r''})}{p''_{\mathbf{a}}(\mathbf{r''})} c_{\mathbf{a}}(\mathbf{r}, \mathbf{r''}) \, \mathrm{d}\mathbf{r} \, \mathrm{d}\mathbf{r''} \\ &= \int_{\mathbf{r''}} \pi(\mathbf{a} \mid \mathbf{r''}) p'_{\mathbf{a}}(\mathbf{r''}) \, \mathrm{d}\mathbf{r''} = m_{\nu'\pi}(\mathbf{a}) \, .\end{aligned}$$

Using the data-processing inequality, we get

$$\mathrm{KL}\left(\mathbb{M}_{\nu\pi} \parallel \mathbb{M}_{\nu'\pi}\right) \leq \mathrm{KL}\left(p_{\gamma\pi} \parallel q_{\gamma\pi}\right) \, . \tag{13}$$

Now, we compute

$$\begin{aligned}\mathrm{KL}\left(p_{\gamma\pi} \parallel q_{\gamma\pi}\right) &\overset{(a)}{=} \mathbb{E}_{\mathbf{h}:=(\mathbf{a},\mathbf{r},\mathbf{r''})\sim p_{\gamma\pi}} \left[\log\left(\frac{\pi(\mathbf{a} \mid \mathbf{r})c_{\mathbf{a}}(\mathbf{r}, \mathbf{r''})}{\pi(\mathbf{a} \mid \mathbf{r''}) \frac{p'_{\mathbf{a}}(\mathbf{r''})}{p''_{\mathbf{a}}(\mathbf{r''})} c_{\mathbf{a}}(\mathbf{r}, \mathbf{r''})}\right)\right] \\ &\overset{(b)}{\leq} \mathbb{E}_{\mathbf{h}:=(\mathbf{a},\mathbf{r},\mathbf{r''})\sim p_{\gamma\pi}} \left[\epsilon d_{\mathrm{Ham}}(\mathbf{r}, \mathbf{r''}) + \log\left(\frac{p''_{\mathbf{a}}(\mathbf{r''})}{p'_{\mathbf{a}}(\mathbf{r''})}\right)\right] \\ &\overset{(c)}{=} \sum_{t=1}^{T} \mathbb{E}_{\mathbf{h}\sim p_{\gamma\pi}} \left[\epsilon \mathbb{1}\left\{r_t \neq r''_t\right\} + \log\left(\frac{p''_{a_t}(r''_t)}{p'_{a_t}(r''_t)}\right)\right] \\ &\overset{(d)}{=} \sum_{t=1}^{T} \mathbb{E}_{\mathbf{h}\sim p_{\gamma\pi}} \left[\mathbb{E}_{\mathbf{h}\sim p_{\gamma\pi}}[\epsilon \mathbb{1}\left\{r_t \neq r''_t\right\} + \log\left(\frac{p''_{a_t}(r''_t)}{p'_{a_t}(r''_t)}\right) \mid a_t]\right] \\ &\overset{(e)}{=} \sum_{t=1}^{T} \mathbb{E}_{\mathbf{h}\sim p_{\gamma\pi}} \left[\epsilon \mathrm{TV}\left(p_{a_t} \parallel p''_{a_t}\right) + \mathrm{KL}\left(p''_{a_t} \parallel p'_{a_t}\right)\right] \\ &\overset{(f)}{=} \mathbb{E}_{\nu\pi} \left[\sum_{t=1}^{T} \epsilon \mathrm{TV}\left(p_{a_t} \parallel p''_{a_t}\right) + \mathrm{KL}\left(p''_{a_t} \parallel p'_{a_t}\right)\right] \, .\end{aligned}$$

where:

(a) by the definition of the KL

(b) the group privacy property, applied to the ϵ-global DP policy, we have

$$\pi(\mathbf{a} \mid \mathbf{r}) \leq e^{\epsilon d_{\mathrm{Ham}}(\mathbf{r},\mathbf{r''}} \pi(\mathbf{a} \mid \mathbf{r''})$$

(c) by the definition of dham

(d) by the towering property of conditional expectations

(e) given a_t, we have $r_t \sim p_{a_t}$, $r'_t \sim p'_{a_t}$ and $r" \sim p''_{a_t}$

(f) by linearity of the expectation, and the fact that the expression inside the expectation only depends on the actions a_t

Since this is true for any "intermediary" bandit instance $\nu'' \in \mathcal{F}^K$, we take $\nu''_\star$ to be the environment where the infinimum of the $d_\epsilon(P_a, P'_a)$ is attained for each arm $a \in [K]$. Specifically, let $\nu''_\star = (p^\star_a, a \in [K])$ where

$$p^\star_a = \operatorname*{arg\,min}_{\mathbb{L}\in\mathcal{F}} \left\{\epsilon \, \mathrm{TV}\left(p_a \parallel \mathbb{L}\right) + \mathrm{KL}\left(\mathbb{L} \parallel p'_a\right)\right\}$$

Plugging $\nu''_\star$ gives

$$\mathrm{KL}\left(\mathbb{M}_{\nu\pi} \parallel \mathbb{M}_{\nu'\pi}\right) \leq \mathbb{E}_{\nu\pi} \left[\sum_{t=1}^{T} \mathrm{d}_\epsilon(p_{a_t}, p'_{a_t})\right] \tag{14}$$

Let $N_{t,a} = \sum_{s<t} \mathbb{1}\{a_s = a\}$ be the counts of arm a before step t. Then, we can rewrite the bound as

$$\mathrm{KL}\left(\mathbb{M}_{\nu\pi} \parallel \mathbb{M}_{\nu'\pi}\right) \leq \sum_{a=1}^{K} \mathbb{E}_{\nu\pi}[N_{T+1,a}] \mathrm{d}_\epsilon(p_a, p'_a)\,, \tag{15}$$

Stopping time version of the KL decomposition for BAI under DP. Let π be an ϵ-DP BAI strategy. Let ν and λ be two bandit instances. Denote by $\mathbb{M}_{\nu\pi}$ the marginal distribution of the output of the BAI strategy when π interacts with ν. By using Wald's lemma in the proof technique seen before for the canonical bandit setting under FC-BAI, we get that

$$\mathrm{KL}\left(\mathbb{M}_{\nu\pi} \parallel \mathbb{M}_{\lambda\pi}\right) \leq \mathbb{E}_{\nu\pi}\left(\sum_{t=1}^{\tau_{\epsilon,\delta}} \mathrm{d}_\epsilon(\nu_{a_t}, \lambda_{a_t})\right) = \sum_{a=1}^{K} \mathbb{E}_{\nu\pi}[N_{\tau_{\epsilon,\delta}+1,a}] \mathrm{d}_\epsilon(\nu_a, \lambda_a)\,, \tag{16}$$

where τ is the stopping time.

D.3 Sample Complexity Lower Bound Proof

Theorem 2 (Sample complexity lower bound for BAI under ϵ-DP)**.** *Let $(\epsilon, \delta) \in \mathbb{R}_+^\star \times (0,1)$. For any algorithm π that is δ-correct and ϵ-global DP on $\mathcal{F}^K$,*

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

for all $\boldsymbol{\nu} \in \mathcal{F}^K$ with unique best arm. The inverse of the characteristic time *$T_\epsilon^\star(\boldsymbol{\nu})$ is defined as*

$$T_\epsilon^\star(\boldsymbol{\nu})^{-1} := \sup_{w \in \triangle_K} \inf_{\boldsymbol{\kappa} \in \mathrm{Alt}(\boldsymbol{\nu})} \sum_{a=1}^{K} w_a \mathrm{d}_\epsilon(\nu_a, \kappa_a)\,, \tag{17}$$

$$\mathrm{d}_\epsilon(\nu_a, \kappa_a) := \inf_{\varphi_a \in \mathcal{F}} \left\{\mathrm{KL}\left(\varphi_a \parallel \kappa_a\right) + \epsilon \cdot \mathrm{TV}\left(\nu_a \parallel \varphi_a\right)\right\}\,. \tag{18}$$

Proof. Let π be an ϵ-global DP δ-correct BAI strategy. Let ν be a bandit instance and $\lambda \in \mathrm{Alt}(\nu)$.

Let $\mathbb{M}_{\nu\pi}$ denote the probability distribution of the output when the BAI strategy π interacts with ν. For any alternative instance $\lambda \in \mathrm{Alt}(\nu)$, the data-processing inequality gives that

$$\mathrm{KL}\left(\mathbb{M}_{\nu\pi} \parallel \mathbb{M}_{\lambda,\pi}\right) \geq \mathrm{kl}\left(\mathbb{M}_{\nu\pi}\left(\tilde{a} = a^\star(\boldsymbol{\nu})\right), \mathbb{M}_{\lambda,\pi}\left(\tilde{a} = a^\star(\boldsymbol{\nu})\right)\right) \geq \mathrm{kl}(1-\delta, \delta)\,, \tag{19}$$

where the second inequality is because π is δ-correct, i.e., $\mathbb{M}_{\nu\pi}\left(\tilde{a} = a^\star(\boldsymbol{\nu})\right) \geq 1-\delta$ and $\mathbb{M}_{\lambda,\pi}\left(\tilde{a} = a^\star(\boldsymbol{\nu})\right) \leq \delta$, and the monotonicity of the kl. Now, using the stopping time version of the KL decomposition for FC-BAI, we get that

$$\mathrm{kl}(1-\delta, \delta) \leq \mathrm{KL}\left(\mathbb{M}_{\nu,\pi} \parallel \mathbb{M}_{\lambda,\pi}\right) \leq \sum_{a=1}^{K} \mathbb{E}_{\nu\pi}[N_{\tau_{\epsilon,\delta}+1,a}] \mathrm{d}_\epsilon(\nu_a, \lambda_a)\,.$$

Since this is true for all $\lambda \in \mathrm{Alt}(\nu)$, we get

$$\begin{aligned}
\mathrm{kl}(1-\delta, \delta) &\leq \inf_{\lambda \in \mathrm{Alt}(\nu)} \sum_{a=1}^{K} \mathbb{E}_{\nu\pi}[N_{\tau_{\epsilon,\delta}+1,a}] \mathrm{d}_\epsilon(\nu_a, \lambda_a) \\
&\overset{(a)}{=} \mathbb{E}[\tau_{\epsilon,\delta}] \inf_{\lambda \in \mathrm{Alt}(\nu)} \sum_{a=1}^{K} \frac{\mathbb{E}\left[N_{\tau_{\epsilon,\delta}+1,a}\right]}{\mathbb{E}[\tau_{\epsilon,\delta}]} \mathrm{d}_\epsilon(\nu_a, \lambda_a) \\
&\overset{(b)}{\leq} \mathbb{E}[\tau_{\epsilon,\delta}] \left(\sup_{\omega \in \triangle_K} \inf_{\lambda \in \mathrm{Alt}(\nu)} \sum_{a=1}^{K} \omega_a \mathrm{d}_\epsilon(\nu_a, \lambda_a)\right)\,.
\end{aligned}$$

(a) is due to the fact that $\mathbb{E}[\tau_{\epsilon,\delta}]$ does not depend on λ. (b) is obtained by noting that the vector $(\omega_a)_{a\in[K]} \triangleq \left(\frac{\mathbb{E}_{\nu,\pi}\left[N_{\tau_{\epsilon,\delta}+1,a}\right]}{\mathbb{E}_{\nu,\pi}[\tau_{\epsilon,\delta}]}\right)_{a\in[K]}$ belongs to the simplex $\triangle_K$. The theorem follows by noting that for $\delta \in (0,1), \mathrm{kl}(1-\delta, \delta) \geq \log(1/3\delta)$. □

E Privacy Analysis

In this section, we prove Lemma 5. First, we justify using a geometric grid for updating the means (Lemma 8). Second, we obtain Lemma 5 as a combination of Lemma 8 and the post-processing property of DP (Proposition 1).

E.1 Releasing partial sums privately

First, the following lemma justifies the use of geometric grids, and provides that the price of getting rid of forgetting is summing the Laplace noise from previous phases.

Lemma 8 (Privacy of our grid-based mean estimator)**.** *Let $T \in \{1, \dots\}$, $\ell < T$ and $t_1, \dots t_\ell, t_{\ell+1}$ be in $[1, T]$ such that $1 = t_1 < \cdots < t_\ell < t_{\ell+1} - 1 = T$.*

Let $\mathcal{M}$ be the following mechanism:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{pmatrix} \overset{\mathcal{M}}{\to} \begin{pmatrix} (x_1 + \cdots + x_{t_2-1}) + (Y_1) \\ (x_1 + \cdots + x_{t_3-1}) + (Y_1 + Y_2) \\ \vdots \\ (x_1 + \cdots + x_T) + (Y_1 + Y_2 + \cdots + Y_{\ell-1}) \end{pmatrix}$$

where $(Y_1, \dots, Y_\ell) \sim^{iid} \mathrm{Lap}(1/\epsilon)$.

Then, for any $\{x_1, \dots, x_T\} \in [0,1]^T$, $\mathcal{M}$ is ϵ-DP.

Proof. First, consider the following mechanism, that only computes the partial sums:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{pmatrix} \to \begin{pmatrix} x_1 + \cdots + x_{t_2-1} \\ x_{t_2} + \cdots + x_{t_3-1} \\ \vdots \\ x_{t_{\ell-1}} + \cdots + x_T \end{pmatrix}.$$

Because $x_t \in [0,1]$, the sensitivity of each partial sum is 1. Since each partial sum is computed over non-overlapping sequences, combining the Laplace mechanism (Theorem 1) with the parallel composition property of DP (Lemma 3) gives that the following mechanism:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{pmatrix} \overset{\mathcal{P}}{\to} \begin{pmatrix} x_1 + \cdots + x_{t_2-1} + Y_1 \\ x_{t_2} + \cdots + x_{t_3-1} + Y_2 \\ \vdots \\ x_{t_{\ell-1}} + \cdots + x_T + Y_{\ell-1} \end{pmatrix}$$

is ϵ-DP, where $(Y_1, \dots, Y_{\ell-1}) \sim^{\text{iid}} \mathrm{Lap}(1/\epsilon)$.

Consider the post-processing function $f : (x_1, \dots x_{\ell-1}) \to (x_1, x_1 + x_2, \dots, x_1 + x_2 + \cdots + x_{\ell-1})$. Then, we have that that $\mathcal{M} = f \circ \mathcal{P}$. So, by the post-processing property of DP, $\mathcal{M}$ is ϵ-DP. □

Remark 1. *Mechanism $\mathcal{P}$, defined in the proof of Lemma 8, is the fundamental mechanism used by all previous bandit algorithms [Sajed and Sheffet, 2019; Azize and Basu, 2022; Hu and Hegde, 2022; Azize et al., 2024] to justify the use of forgetting. Our mechanism $\mathcal{M}$ is just summing over the partial sums computed on each phase, and thus the price of having sums of x_i that start from the beginning (i.e. do not forget) is that we have to sum now the noise from all previous phases too.*

E.2 Proof of Lemma 5

We are now ready to prove Lemma 5, i.e. that any BAI algorithm based solely on using $\mathrm{GPE}_\eta(\epsilon)$ to access observations is ϵ-global DP on $[0,1]$.

Proof. Let π be a BAI algorithm using only $\mathrm{GPE}_\eta(\epsilon)$ to access observations. Let $R = \{x_1, \dots\}$ and $R' = \{x'_1, \dots\}$ be two neighbouring sequences of private observations, i.e. there exists a $t^\star \in \{1, \dots\}$ such that $x_t = x'_t$ for all $t \neq t^\star$, i.e. that R and R' only differ at $t^\star$.

Fix a stopping time, recommendation and sampled actions $(T+1, \tilde{a}, (a_1, \dots, a_T))$, we want to show that

$$\Pr[\pi(R) = (T+1, \tilde{a}, (a_1, \dots, a_T))] \leq e^{\epsilon} \Pr[\pi(R') = (T+1, \tilde{a}, (a_1, \dots, a_T))] .$$

Step 1: Probability decompositions: First, let us denote by τ, $\tilde{A}$ and $A_1, \dots, A_\tau$ the random variables of stopping, recommendation and sampled actions, when π interacts with R. Similarly, let τ', $\tilde{A}'$ and $A'_1, \dots, A'_\tau$ the random variables of stopping, recommendation and sampled actions, when π interacts with R'.

We have

$$\Pr[\pi(R) = (T+1, \tilde{a}, (a_1, \dots, a_T))] = \Pr[\tau = T+1, \tilde{A} = \tilde{a}, A_1 = a_1, \dots, A_T = a_T]$$
$$\Pr[\pi(R') = (T+1, \tilde{a}, (a_1, \dots, a_T))] = \Pr[\tau' = T+1, \tilde{A}' = \tilde{a}, A'_1 = a_1, \dots, A'_T = a_T]$$

Since for all $t < t^\star$, $x_t = x'_t$, the policy samples the same actions, up to step $t^\star$, i.e.

$$\Pr[A_1 = a_1, \dots, A_{t^\star} = a_{t^\star}] = \Pr[A'_1 = a_1, \dots, A'_{t^\star} = a_{t^\star}]$$

And thus

$$\frac{\Pr[\pi(R) = (T+1, \tilde{a}, (a_1, \dots, a_T))]}{\Pr[\pi(R') = (T+1, \tilde{a}, (a_1, \dots, a_T))]}$$
$$= \frac{\Pr[\tau = T+1, \tilde{A} = \tilde{a}, A_{t^\star+1} = a_{t^\star+1}, \dots, A_T = a_T \mid A_1 = a_1, \dots, A_{t^\star} = a_{t^\star}]}{\Pr[\tau' = T+1, \tilde{A}' = \tilde{a}, A'_{t^\star+1} = a_{t^\star+1}, \dots, A'_T = a_T \mid A'_1 = a_1, \dots, A'_{t^\star} = a_{t^\star}]}$$

Let us denote by $t_1, \dots, t_\ell$ the time step corresponding to the beginning of the phases when π interacts with R, and $t'_1, \dots, t'_{\ell'}$ the the time step corresponding to the beginning of the phases π interacts with **R'**.

Also, let $t_{k_\star}$ be the beginning of the phase for which $t^\star$ belongs in R phases. Similarly, let $t'_{k'_\star}$ be the beginning of the phase for which $t^\star$ belongs in R' phases.

Since the actions $a_1, \dots, a_T$ are fixed, and $r_t = r'_t$ for $t < t^\star$, $t^\star$ falls in the same phase under both R and R'. Thus, $t_{k_\star} = t'_{k'_\star}$ and $k_\star = k'_\star$.

Step 2: Using the structure of GPE$_\eta(\epsilon)$

Let $\tilde{S}^p_{k^\star} = \sum_{s=t_{k^\star}}^{t_{k^\star+1}-1} x_s + Y_{k_\star}$ be the noisy partial sum of observations collected at phase $k^\star$ for **r**, where $Y_{k^\star} \sim \mathrm{Lap}(1/\epsilon)$. Similarly, let $\tilde{S'}^p_{k^\star} = \sum_{s=t_{k^\star}}^{t_{k^\star+1}-1} x'_s + Y'_{k_\star}$ be the noisy partial sum of observations collected at phase $k^\star$ for **r'**, where $Y'_{k^\star} \sim \mathrm{Lap}(1/\epsilon)$. Using the structure of GPE$_\eta(\epsilon)$, we have that:

(a) If the value of the noisy partial sums at phase $k^\star$ is exactly the same between the neighbouring R and R', then the BAI algorithm π will sample the same sequence of actions from step $t^\star$ onward, recommend the same final guess and stop at the same time, with the same probability under R and $\mathbb{R}'$. Thus, for any $s \in \mathbb{R}$:

$$\Pr[\tau = T+1, \tilde{A} = \tilde{a}, A_{t^\star+1} = a_{t^\star+1}, \dots, A_T = a_T \mid A_1 = a_1, \dots, A_{t^\star} = a_{t^\star}, \tilde{S}^p_{k^\star} = s]$$
$$= \Pr[\tau' = T+1, \tilde{A}' = \tilde{a}, A'_{t^\star+1} = a_{t^\star+1}, \dots, A'_T = a_T \mid A'_1 = a_1, \dots, A'_{t^\star} = a_{t^\star}, \tilde{S'}^p_{k^\star} = s] \tag{20}$$

This is due to the fact that, in GPE$_\eta(\epsilon)$, the reward at step $t^\star$ only affects the statistic $\tilde{S}^p_{k^\star}$, and nothing else.

(b) Since rewards are $[0, 1]$, using the Laplace mechanism, we have that

$$\Pr[\tilde{S}^p_{k^\star} = s \mid A_1 = a_1, \dots, A_{t^\star} = a_{t^\star}] \leq e^{\epsilon} \Pr(\tilde{S'}^p_{k^\star} = s \mid A_1 = a_1, \dots, A'_{t^\star} = a_{t^\star}) . \tag{21}$$

Step 3: Combining Eq. 20 and Eq. 21, aka post-processing:

We have

$$\begin{aligned}
&\Pr[\tau = T+1, \tilde{A} = \tilde{a}, A_{t^\star+1} = a_{t^\star+1}, \dots, A_T = a_T \mid A_1 = a_1, \dots, A_{t^\star} = a_{t^\star}] \\
&= \int_{s\in\mathbb{R}} \Pr[\tau = T+1, \tilde{A} = \tilde{a}, A_{t^\star+1} = a_{t^\star+1}, \dots, A_T = a_T \mid A_1 = a_1, \dots, A_{t^\star} = a_{t^\star}, \tilde{S}^p_{k^\star} = s] \\
&\Pr[\tilde{S}^p_{k^\star} = s \mid A_1 = a_1, \dots, A_{t^\star} = a_{t^\star}] \\
&\leq \int_{s\in\mathbb{R}} e^\epsilon \Pr[\tau' = T+1, \tilde{A}' = \tilde{a}, A'_{t^\star+1} = a_{t^\star+1}, \dots, A'_T = a_T \\
&\qquad\qquad \mid A_1 = a_1, \dots, A'_{t^\star} = a_{t^\star}, \tilde{S'}^p_{k^\star} = s] \\
&\Pr(\tilde{S'}^p_{k^\star} = s \mid A_1 = a_1, \dots, A'_{t^\star} = a_{t^\star}) \\
&= e^\epsilon \Pr[\tau' = T+1, \tilde{A}' = \tilde{a}, A'_{t^\star+1} = a_{t^\star+1}, \dots, A'_T = a_T \mid A'_1 = a_1, \dots, A'_{t^\star} = a_{t^\star}] \,.
\end{aligned}$$

This concludes the proof:

$$\frac{\Pr[\pi(R) = (T+1, \tilde{a}, (a_1, \dots, a_T))]}{\Pr[\pi(R') = (T+1, \tilde{a}, (a_1, \dots, a_T))]} \leq e^\epsilon \,.$$

□

E.3 Recalling the post-processing and composition properties of DP

Proposition 1 (Post-processing [Dwork and Roth, 2014])**.** *Let $\mathcal{M}$ be a mechanism and f be an arbitrary randomised function defined on $\mathcal{M}$'s output. If $\mathcal{M}$ is ϵ-DP, then $f \circ \mathcal{M}$ is ϵ-DP.*

The post-processing property ensures that any quantity constructed only from a private output is still private, with the same privacy budget. This is a consequence of the data processing inequality.

Proposition 2 (Simple Composition)**.** *Let $\mathcal{M}^1, \dots, \mathcal{M}^k$ be k mechanisms. We define the mechanism*

$$\mathcal{G} : D \to \bigotimes_{i=1}^{k} \mathcal{M}^i_D$$

as the k composition of the mechanisms $\mathcal{M}^1, \dots, \mathcal{M}^k$.

If each $\mathcal{M}^i$ is ϵ_i-DP, then $\mathcal{G}$ is $\sum_{i=1}^k \epsilon_i$-DP.

Proposition 3 (Parallel Composition)**.** *Let $\mathcal{M}^1, \dots, \mathcal{M}^k$ be k mechanisms, such that $k < n$, where n is the size of the input dataset. Let $t_1, \dots t_k, t_{k+1}$ be indexes in $[1, n]$ such that $1 = t_1 < \dots < t_k < t_{k+1} - 1 = n$.*
Let's define the following mechanism

$$\mathcal{G} : \{x_1, \dots, x_n\} \to \bigotimes_{i=1}^{k} \mathcal{M}^i_{\{x_{t_i}, \dots, x_{t_{i+1}-1}\}}$$

$\mathcal{G}$ is the mechanism that we get by applying each $\mathcal{M}^i$ to the i-th partition of the input dataset $\{x_1, \dots, x_n\}$ according to the indexes $t_1 < \dots < t_k < t_{k+1}$.

If each $\mathcal{M}^i$ is ϵ-DP, then $\mathcal{G}$ is ϵ-DP.

In parallel composition, the k mechanisms are applied to different "non-overlapping" parts of the input dataset. If each mechanism is DP, then the parallel composition of the k mechanisms is DP, *with the same privacy budget*. This property will be the basis for designing private bandit algorithms.

F Concentration Results

In Appendix F, we detail the proof of all our concentration results. In Appendix F.1, we start by introducing a variant of GLR-based stopping rule using the modified transportation costs $\widetilde{W}_{\epsilon,a,b}$

(see Appendix G.2.1 for details) which are defined based on the modified divergences $\widetilde{d}_\epsilon^{\pm}$ (see Appendix G.1.1 for details). The proof of Theorem 6 is given in Appendix F.2. In Appendix F.3, we show tail bounds for a sum between independent Bernoulli and Laplace observations that feature the product of the tail bounds of each process. We prove time-uniform and fixed-time tails concentration for Laplace distribution in Appendix F.4, and recall existing results for Bernoulli in Appendix F.5. In Appendix F.6, we provide tail bounds for a sum between independent Bernoulli and Laplace observations that feature the modified divergence $\widetilde{d}_\epsilon$ defined in Eq. (32). In Appendix F.7, we give geometric grid time uniform tails concentration for the reweighted modified divergence.

F.1 Modified GLR Stopping Rule

The modified GLR stopping rule is defined as

$$\tau_{\epsilon,\delta}^{\text{MGLR}} = \inf\left\{ n \mid \forall a \neq \tilde{a}_n,\ \widetilde{W}_{\epsilon,\tilde{a}_n,a}(\tilde{\mu}_n, \tilde{N}_n) > \sum_{b\in\{\tilde{a}_n,a\}} \widetilde{c}(k_{n,b},\delta) \right\} \text{ with } \tilde{a}_n \in \operatorname*{arg\,max}_{a\in[K]} [\tilde{\mu}_{n,a}]_0^1\,, \tag{22}$$

where $(\tilde{\mu}_n, \tilde{N}_n)$ are the outputs of $\text{GPE}_\eta(\epsilon)$. The modified transportation costs $(\widetilde{W}_{\epsilon,a,b})_{(a,b)\in[K]^2}$ are defined in Eq. (34), i.e., for all $(\mu, w) \in \mathbb{R}^K \times \mathbb{R}_+^K$ and all $(a,b) \in [K]^2$ such that $a \neq b$,

$$\widetilde{W}_{\epsilon,a,b}(\mu, w) := \mathbb{1}\left([\mu_a]_0^1 > [\mu_b]_0^1\right) \inf_{u\in(0,1)} \left\{ w_a \widetilde{d}_\epsilon^-(\mu_a, u, r(w_a)) + w_b \widetilde{d}_\epsilon^+(\mu_b, u, r(w_b)) \right\}\,,$$

where $r(x) := \frac{x}{1+\log_{1+\eta} x}$ is defined in Eq. (33) for all $x \geq 1$. The modified divergence $\widetilde{d}_\epsilon^{\pm}$ are defined in Eq. (32), i.e., for all $(\lambda, \mu, r) \in \mathbb{R} \times (0,1) \times \mathbb{R}_+^\star$,

$$\widetilde{d}_\epsilon^-(\lambda,\mu,r) := \mathbb{1}\left(\mu < [\lambda]_0^1\right) \inf_{z\in(\mu,[\lambda]_0^1)} \left\{ \mathrm{kl}(z,\mu) + \frac{1}{r} h(r\epsilon(\lambda - z)) \right\}\,,$$
$$\widetilde{d}_\epsilon^+(\lambda,\mu,r) := \mathbb{1}\left(\mu > [\lambda]_0^1\right) \inf_{z\in([\lambda]_0^1,\mu)} \left\{ \mathrm{kl}(z,\mu) + \frac{1}{r} h(r\epsilon(z - \lambda)) \right\}\,,$$

where $h(x) := \sqrt{1+x^2} - 1 + \log\left(\frac{2}{x^2}\left(\sqrt{1+x^2}-1\right)\right)$ is defined in Eq. (31) for all $x > 0$.

Lemma 9 gives a stopping threshold under which the modified GLR stopping rule is δ-correct.

Lemma 9. *Let $\delta \in (0,1)$ and $\epsilon > 0$. Let $\eta > 0$. Let $s > 1$ and ζ be the Riemann ζ function. Let $\overline{W}_{-1}(x) = -W_{-1}(-e^{-x})$ for all $x \geq 1$, where W_{-1} is the negative branch of the Lambert W function. It satisfies $\overline{W}_{-1}(x) \approx x + \log x$, see Lemma 52. Given any sampling rule using the $\text{GPE}_\eta(\epsilon)$, combining $\text{GPE}_\eta(\epsilon)$ with the modified GLR stopping rule as in Eq.* (22) *with the stopping threshold*

$$\widetilde{c}(k,\delta) = \overline{W}_{-1}\left(\log\left(\frac{K\zeta(s)}{\delta}\right) + s\log(k) + 3 - \log 2\right) - 3 + \log 2\,. \tag{23}$$

yields a δ-correct and ϵ-global DP algorithm for Bernoulli instances with unique best arm.

Proof. Lemma 5 yields the ϵ-global DP. Let $\mathcal{E}_\delta = \mathcal{E}_{\delta,a^\star,+} \cap \bigcap_{a\neq a^\star} \mathcal{E}_{\delta,a,-}$ with

$$\mathcal{E}_{\delta,a^\star,+} = \left\{ \forall n \in \mathbb{N},\ \tilde{N}_{n,a^\star} \widetilde{d}_\epsilon^+(\tilde{\mu}_{n,a^\star}, \mu_{a^\star}, \tilde{N}_{n,a^\star}/k_{n,a^\star}) \leq \widetilde{c}(k_{n,a^\star},\delta) \right\}\,,$$
$$\forall a \neq a^\star, \quad \mathcal{E}_{\delta,a,-} = \left\{ \forall n \in \mathbb{N},\ \tilde{N}_{n,a} \widetilde{d}_\epsilon^-(\tilde{\mu}_{n,a}, \mu_a, \tilde{N}_{n,a}/k_{n,a}) \leq \widetilde{c}(k_{n,a},\delta) \right\}\,,$$

where $(\tilde{\mu}_n, \tilde{N}_n, k_n)$ are given by $\text{GPE}_\eta(\epsilon)$, $\widetilde{c}$ as in Eq. (23) and $\widetilde{d}_\epsilon^{\pm}$ as in Eq. (32).

Using Lemmas 20 and 21, we have $\mathbb{P}_{\boldsymbol{\nu}\pi}(\mathcal{E}_{\delta,a,-}^{\complement}) \leq \delta/K$ for all $a \neq a^\star$, and $\mathbb{P}_{\boldsymbol{\nu}\pi}(\mathcal{E}_{\delta,a^\star,+}^{\complement}) \leq \delta/K$. By union bound over $a \in [K]$, we obtain $\mathbb{P}_{\boldsymbol{\nu}\pi}(\mathcal{E}_\delta^{\complement}) \leq \delta$.

Let $\tau_{\epsilon,\delta}^{\text{MGLR}}$ as in Eq. (22) and $\tilde{a}_n \in \arg\max_{a\in[K]} [\tilde{\mu}_{n,a}]_0^1$. Then, we directly have that

$$\mathbb{P}_{\boldsymbol{\nu}\pi}\left(\tau_{\epsilon,\delta}^{\text{MGLR}} < +\infty,\ \tilde{a}_{\tau_{\epsilon,\delta}^{\text{MGLR}}} \neq a^\star\right) \leq \mathbb{P}_{\boldsymbol{\nu}\pi}\left(\mathcal{E}_\delta^{\complement}\right) + \mathbb{P}_{\boldsymbol{\nu}\pi}\left(\mathcal{E}_\delta \cap \{\tau_{\epsilon,\delta}^{\text{MGLR}} < +\infty,\ \tilde{a}_{\tau_{\epsilon,\delta}^{\text{MGLR}}} \neq a^\star\}\right)$$

$$\leq \delta + \mathbb{P}_{\boldsymbol{\nu}\pi}\left(\mathcal{E}_\delta \cap \{\tau^{\text{MGLR}}_{\epsilon,\delta} < +\infty,\ \tilde{a}_{\tau^{\text{MGLR}}_{\epsilon,\delta}} \neq a^\star\}\right) .$$

Under $\mathcal{E}_\delta \cap \{\tau^{\text{MGLR}}_{\epsilon,\delta} < +\infty,\ \tilde{a}_{\tau^{\text{MGLR}}_{\epsilon,\delta}} \neq a^\star\}$, by definition of the stopping rule as in Eq. (7) and the stopping threshold in Eq. (8), we obtain that there exists $a \neq a^\star$ and $n \in \mathbb{N}$ such that $[\tilde{\mu}_{n,a}]_0^1 > [\tilde{\mu}_{n,a^\star}]_0^1$ and

$$\begin{aligned}
&\sum_{b\in\{a,a^\star\}} \widetilde{c}(k_{n,b},\delta) < \widetilde{W}_{\epsilon,a,a^\star}(\tilde{\mu}_n,\tilde{N}_n) \\
&= \inf_{u\in(0,1)} \left\{ \tilde{N}_{n,a}\widetilde{d}^-_\epsilon(\tilde{\mu}_{n,a},u,r(\tilde{N}_{n,a})) + \tilde{N}_{n,a^\star}\widetilde{d}^+_\epsilon(\tilde{\mu}_{n,a^\star},u,r(\tilde{N}_{n,a^\star})) \right\} \\
&= \inf_{(u_a,u_{a^\star})\in(0,1)^2,\ u_a\leq u_{a^\star}} \{ \tilde{N}_{n,a}\widetilde{d}^-_\epsilon(\tilde{\mu}_{n,a},u_a,r(\tilde{N}_{n,a})) + \tilde{N}_{n,a^\star}\widetilde{d}^+_\epsilon(\tilde{\mu}_{n,a^\star},u_{a^\star},r(\tilde{N}_{n,a^\star})) \} \\
&\leq \tilde{N}_{n,a}\widetilde{d}^-_\epsilon(\tilde{\mu}_{n,a},\mu_a,r(\tilde{N}_{n,a})) + \tilde{N}_{n,a^\star}\widetilde{d}^+_\epsilon(\tilde{\mu}_{n,a^\star},\mu_{a^\star},r(\tilde{N}_{n,a^\star})) \\
&\leq \tilde{N}_{n,a}\widetilde{d}^-_\epsilon(\tilde{\mu}_{n,a},\mu_a,\tilde{N}_{n,a}/k_{n,a}) + \tilde{N}_{n,a^\star}\widetilde{d}^+_\epsilon(\tilde{\mu}_{n,a^\star},\mu_{a^\star},\tilde{N}_{n,a^\star}/k_{n,a^\star}) \leq \sum_{b\in\{a,a^\star\}} \widetilde{c}(k_{n,b},\delta) ,
\end{aligned}$$

where we used the definition of $\widetilde{W}_{\epsilon,a,a^\star}$ in Eq. (34) and Lemma 40 in the two equalities and $\mu_{a^\star} > \mu_a$ in the following inequality. The second to last inequality uses that $r(\tilde{N}_{n,a}) \leq \tilde{N}_{n,a}/k_{n,a}$ for all $a \in [K]$ by definition of $(k_n, \tilde{N}_n)$, i.e., $k_{n,a} \leq 1 + \log_{1+\eta} \tilde{N}_{n,a} \leq k_{n,a} + 1$, and that $r \mapsto \widetilde{d}^\pm_\epsilon(\lambda,u,r)$ is non-decreasing, see Lemmas 31 and 32. The last inequality is obtained by the concentration event $\mathcal{E}_\delta$. Since this yields a contradiction, we obtain $\mathcal{E}_\delta \cap \{\tau^{\text{MGLR}}_{\epsilon,\delta} < +\infty,\ \tilde{a}_{\tau^{\text{MGLR}}_{\epsilon,\delta}} \neq a^\star\} = \emptyset$. This concludes the proof, i.e., $\mathbb{P}_{\boldsymbol{\nu}\pi}\left(\tau^{\text{MGLR}}_{\epsilon,\delta} < +\infty,\ \tilde{a}_{\tau^{\text{MGLR}}_{\epsilon,\delta}} \neq a^\star\right) \leq \delta$. □

F.2 Proof of Theorem 6

Lemma 5 yields the ϵ-global DP. The proof of δ-correctness is the same as the one of Lemma 9 detailed above. In particular, we use the same concentration event $\mathcal{E}_\delta = \mathcal{E}_{\delta,a^\star,+} \cap \bigcap_{a\neq a^\star} \mathcal{E}_{\delta,a,-}$ that satisfies $\mathbb{P}_{\boldsymbol{\nu}\pi}(\mathcal{E}_\delta^\complement) \leq \delta$.

Under $\mathcal{E}_\delta \cap \{\tau_{\epsilon,\delta} < +\infty,\ \tilde{a}_{\tau_{\epsilon,\delta}} \neq a^\star\}$, by definition of the GLR stopping rule as in Eq. (7) and the stopping threshold in Eq. (8), we obtain that there exists $a \neq a^\star$ and $n \in \mathbb{N}$ such that $[\tilde{\mu}_{n,a}]_0^1 > [\tilde{\mu}_{n,a^\star}]_0^1$,

$$\sum_{b\in\{a,a^\star\}} \Big(c_1(\tilde{N}_{n,b},\delta) + c_2(\tilde{N}_{n,b},\epsilon) \Big) = \sum_{b\in\{a,a^\star\}} c(k_{n,b},\epsilon,\delta) < W_{\epsilon,a,a^\star}(\tilde{\mu}_n,\tilde{N}_n) .$$

Then, we obtain

$$\begin{aligned}
W_{\epsilon,a,a^\star}(\tilde{\mu}_n,\tilde{N}_n) &= \inf_{u\in[0,1]} \left\{ \tilde{N}_{n,a} d^-_\epsilon(\tilde{\mu}_{n,a},u) + \tilde{N}_{n,a^\star} d^+_\epsilon(\tilde{\mu}_{n,a^\star},u) \right\} \\
&= \inf_{(u_a,u_{a^\star})\in[0,1]^2,\ u_a\leq u_{a^\star}} \{ \tilde{N}_{n,a} d^-_\epsilon(\tilde{\mu}_{n,a},u_a) + \tilde{N}_{n,a^\star} d^+_\epsilon(\tilde{\mu}_{n,a^\star},u_{a^\star}) \} \\
&\leq \tilde{N}_{n,a} d^-_\epsilon(\tilde{\mu}_{n,a},\mu_a) + \tilde{N}_{n,a^\star} d^+_\epsilon(\tilde{\mu}_{n,a^\star},\mu_{a^\star}) ,
\end{aligned}$$

where we used the definition of $W_{\epsilon,a,a^\star}$ in Eq. (4) and Lemma 35 in the two equalities, and $(u_{a^\star}, u_a) = (\mu_{a^\star}, \mu_a) \in [0,1]^2$ that satisfies $u_{a^\star} > u_a$ in the following inequality.

Using Lemma 39 and initialization yields $\min\{r(\tilde{N}_{n,a^\star}), r(\tilde{N}_{n,a})\} > 0$ by . When $[\tilde{\mu}_{n,a}]_0^1 > \mu_a$ and $\mu_{a^\star} > [\tilde{\mu}_{n,a^\star}]_0^1$, Lemma 30 yields

$$\begin{aligned}
&\tilde{N}_{n,a^\star}(d^+_\epsilon(\tilde{\mu}_{n,a^\star},\mu_{a^\star}) - \widetilde{d}^+_\epsilon(\tilde{\mu}_{n,a^\star},\mu_{a^\star},r(\tilde{N}_{n,a^\star}))) \leq k_\eta(\tilde{N}_{n,a^\star})(\log(1+2\epsilon\frac{\tilde{N}_{n,a^\star}}{k_\eta(\tilde{N}_{n,a^\star})}) + 1) \\
&\tilde{N}_{n,a}(d^-_\epsilon(\tilde{\mu}_{n,a},\mu_a) - \widetilde{d}^-_\epsilon(\tilde{\mu}_{n,a},\mu_a,r(\tilde{N}_{n,a}))) \leq k_\eta(\tilde{N}_{n,a}) \left(\log\left(1+2\epsilon\frac{\tilde{N}_{n,a}}{k_\eta(\tilde{N}_{n,a})}\right) + 1 \right) ,
\end{aligned}$$

where we used that $(\mu_a, \mu_{a^\star}) \in (0,1)^2$ and $x/r(x) = 1+\log_{1+\eta} x = k_\eta(x)$. When $[\tilde{\mu}_{n,a}]_0^1 \leq \mu_a$, we have $d_\epsilon^-(\tilde{\mu}_{n,a}, \mu_a) = 0 = \widetilde{d}_\epsilon^-(\tilde{\mu}_{n,a}, \mu_a, r(\tilde{N}_{n,a}))$. When $\mu_{a^\star} \leq [\tilde{\mu}_{n,a^\star}]_0^1$, we have $d_\epsilon^+(\tilde{\mu}_{n,a^\star}, \mu_{a^\star}) = 0 = \widetilde{d}_\epsilon^+(\tilde{\mu}_{n,a^\star}, \mu_{a^\star}, r(\tilde{N}_{n,a^\star}))$. In either case, the above inequalities are still valid since the left hand side is null and the right hand side is positive. Therefore, we have

$$\begin{aligned}
&\tilde{N}_{n,a} d_\epsilon^-(\tilde{\mu}_{n,a}, \mu_a) + \tilde{N}_{n,a^\star} d_\epsilon^+(\tilde{\mu}_{n,a^\star}, \mu_{a^\star}) \\
&\leq \tilde{N}_{n,a} \widetilde{d}_\epsilon^-(\tilde{\mu}_{n,a}, \mu_a, r(\tilde{N}_{n,a})) + \tilde{N}_{n,a^\star} \widetilde{d}_\epsilon^+(\tilde{\mu}_{n,a^\star}, \mu_{a^\star}, r(\tilde{N}_{n,a^\star})) + \sum_{b \in \{a, a^\star\}} c_2(\tilde{N}_{n,b}, \epsilon) \\
&\leq \sum_{b \in \{a, a^\star\}} \widetilde{c}(k_{n,b}, \delta) + \sum_{b \in \{a, a^\star\}} c_2(\tilde{N}_{n,b}, \epsilon) \leq \sum_{b \in \{a, a^\star\}} \Big(c_1(\tilde{N}_{n,b}, \delta) + c_2(\tilde{N}_{n,b}, \epsilon) \Big) \,,
\end{aligned}$$

where the second inequality uses the proof of Lemma 9, and third leverages that

$$\widetilde{c}(k_{n,a}, \delta) \leq \overline{W}_{-1}\left(\log\left(\frac{K\zeta(s)}{\delta} \right) + s \log(k_\eta(\tilde{N}_{n,a})) + 3 - \log 2 \right) - 3 + \log 2 \,,$$

by using that $\overline{W}_{-1}$ is increasing (Lemma 52) and $k_{n,a} \leq 1 + \log_{1+\eta} \tilde{N}_{n,a} = k_\eta(\tilde{N}_{n,a})$ for all $a \in [K]$, as well as $r(x) = x/k_\eta(x)$. Combining all the above inequalities, we have shown that

$$\sum_{b \in \{a, a^\star\}} (c_1(\tilde{N}_{n,b}, \delta) + c_2(\tilde{N}_{n,b}, \epsilon)) < W_{\epsilon, a, a^\star}(\tilde{\mu}_n, \tilde{N}_n) \leq \sum_{b \in \{a, a^\star\}} (c_1(\tilde{N}_{n,b}, \delta) + c_2(\tilde{N}_{n,b}, \epsilon)) \,.$$

This yields a contradiction, hence we have $\mathcal{E}_\delta \cap \{\tau_{\epsilon,\delta} < +\infty,\ \tilde{a}_{\tau_{\epsilon,\delta}} \neq a^\star\} = \emptyset$. This concludes the proof, i.e., $\mathbb{P}_{\boldsymbol{\nu}\pi}\left(\tau_{\epsilon,\delta} < +\infty,\ \tilde{a}_{\tau_{\epsilon,\delta}} \neq a^\star\right) \leq \delta$.

F.3 Fixed Time Tails Bounds for a Convolution of Probability Distributions

We derive general upper and lower bounds on the upper and lower tails of the convolution (i.e., sum) between two independent random variables (Lemma 10). We provide upper (Lemma 12) and lower (Lemma 12) tail bounds for a sum (i.e., convolution) between independent Bernoulli and Laplace i.i.d. observations for a fixed time. The bounds are expressed as a function of the infimum over a bounded interval of a $-\frac{1}{t}\log(\cdot)$ transform of the product between the (upper or lower) tail bounds of each process. Therefore, in Lemmas 12 and 12, we can plug any bounds on the (upper or lower) tail concentration of each process. While those bounds are standard for Bernoulli distribution (Lemma 17 in Appendix F.5), we propose new bounds for Laplace distribution (Lemma 16 in Appendix F.4).

Sketch of Proof of Lemma 10 The main difficulty when studying the sum of two random variables lies in the fact that it involves the integral of the convolution of their probability measures. In all generality, it is difficult to upper bound such a quantity. The main idea behind our proof technique is to split the event of interest into a partition of carefully chosen events. Then, on each of those smaller events, we derive a "tight" upper bound on the integral of the convolution of their probability measures. It is reasonable to wonder how one could choose those events such that the upper bound is easier to obtain. When the event is defined as the intersection of two independent events, then we obtain a straightforward upper bound by the product of their respective probablities. When the event truly mixes the distributions, we need to use a smarter approach to control the integrated function. First, we upper bound a sub-component of this function by a maximum of the product of their respective probablities (on a small interval that is defined by the smaller event). Second, after this upper bound, the integrated function coincides with the hazard function, whose integral is the cumulative hazard function. To conclude the proof, it only remains to merge together the different upper bounds.

To the best of our knownledge, the proof technique closest to ours is the one used to prove Lemma 64 in Jourdan et al. [2022]. They control the probability that two random variables have an unexpected empirical ranking as a function of the boundary crossing probabilities of each random variable. While tackling a distinct problem, they adopt the same proof structure. They decompose the event into carefully chosen events on which they can upper bound the integral of the convolution of their probability distributions. The upper bounds are obtained similarly as ours, with fewer events to consider.

Lemma 10 gives upper and lower bounds on the upper and lower tails of the sum of two independent random variables. This result is of independent interest.

Lemma 10. *Let θ and λ be two independent real random variables such that (i) θ has bounded support included in $[\alpha, \beta]$ and mean $\mu \in (\alpha, \beta)$ and (ii) λ has zero mean. Let*

$$\forall u \in [0,1],\ \forall v \in (0,1], \quad p(u,v) := u(1 - \log(u) + \log(v)) .$$

Then, for all $x > 0$, we have

$$\begin{aligned}
&\mathbb{P}(\theta + \lambda \geq \mu + x) \leq \mathbb{P}(\lambda \geq x)\mathbb{P}(\theta \in [\alpha, \mu]) + \mathbb{P}(\lambda \leq 0)\mathbb{P}(\theta \in [\min\{\beta, \mu + x\}, \beta]) \\
&\qquad + p\left(\sup_{z \in (\mu, \min\{\beta, \mu + x\})} \{\mathbb{P}(\theta \in [z, \beta])\mathbb{P}(\lambda \geq \mu + x - z)\}\, ,\ \mathbb{P}(\theta \in [\mu, \beta])\mathbb{P}(\lambda \geq 0) \right) , \\
&\mathbb{P}(\theta + \lambda \geq \mu + x) \geq \sup_{z \in (\mu, \min\{\beta, \mu + x\})} \{\mathbb{P}(\theta \in [z, \beta])\mathbb{P}(\lambda \geq \mu + x - z)\} , \\
&\mathbb{P}(\theta + \lambda \leq \mu - x) \leq \mathbb{P}(\lambda \leq -x)\mathbb{P}(\theta \in [\mu, \beta]) + \mathbb{P}(\lambda \geq 0)\mathbb{P}(\theta \in [\alpha, \max\{\alpha, \mu - x\}]) \\
&\qquad p\left(\sup_{z \in (\max\{\alpha, \mu - x\}, \mu)} \{\mathbb{P}(\theta \in [\alpha, z])\mathbb{P}(\lambda \leq \mu - x - z)\}\, ,\ \mathbb{P}(\theta \in [\alpha, \mu])\mathbb{P}(\lambda \leq 0) \right) , \\
&\mathbb{P}(\theta + \lambda \leq \mu - x) \geq \sup_{z \in (\max\{\alpha, \mu - x\}, \mu)} \{\mathbb{P}(\theta \in [\alpha, z])\mathbb{P}(\lambda \leq \mu - x - z)\} .
\end{aligned}$$

Proof. **I. Upper Bound on Upper Tail.** We start by studying $\mathbb{P}(\theta + \lambda \geq \mu + x)$ where $x > 0$. We can suppose that there exists $y_1 \in (\max\{x + \mu - \beta, 0\}, x)$ such that $\mathbb{P}(\theta \geq x + \mu - y_1)\mathbb{P}(\lambda \geq y_1) > 0$. Otherwise, the probability of $\{\theta + \lambda \geq \mu + x\}$ is 0, and both bounds are 0 as well. Let y_1 be such a value, and

$$y_3 \in [x, x + \mu) \quad \text{and} \quad y_2 \in (\min\{x + \mu - \beta, 0\}, 0] .$$

First, we note that $-\log \mathbb{P}(\theta \geq x + \mu - y_1)$ and $-\log \mathbb{P}(\lambda \geq y_1)$ are finite, since $\mathbb{P}(\theta \geq x + \mu - y_1)\mathbb{P}(\lambda \geq y_1) > 0$ implies that $\min\{\mathbb{P}(\theta \geq x + \mu - y_1), \mathbb{P}(\lambda \geq y_1)\} > 0$. Second, we note that y_2 only exists when $x + \mu < \beta$, i.e., $(\min\{x + \mu - \beta, 0\}, 0] \neq \emptyset$. In order to study the cases $x + \mu < \beta$ and $x + \mu \geq \beta$ simultaneously, we adopt the convention that the maximum of a positive quantity on an empty set is defined as zero. Note that the situation $x + \mu < \beta$ has more subcases.

We partition of the event $\{\theta + \lambda \geq \mu + x\}$ into eight sets, namely

$$\begin{aligned}
\{\theta + \lambda \geq \mu + x,\ \theta \in [\alpha, \beta],\ \lambda \in \mathbb{R}\} = &\{\lambda \in (\max\{x + \mu - \beta, 0\}, y_1),\ \theta \in [x + \mu - \lambda, \beta]\} \\
&\cup \{\theta \in [x + \mu - y_1, \beta],\ \lambda \geq y_1\} \\
&\cup \{\theta \in (\mu, x + \mu - y_1),\ \lambda \geq x + \mu - \theta\} \\
&\cup \{\theta \in [x + \mu - y_3, \mu],\ \lambda \geq x + \mu - \theta\} \\
&\cup \{\theta \in [\alpha, x + \mu - y_3), \lambda \geq x + \mu\} \\
&\cup \{\lambda \in [y_3, x + \mu),\ \theta \in [x + \mu - \lambda, x + \mu - y_3)\} \\
&\cup \{\lambda \in [y_2, 0],\ \theta \in [x + \mu - \lambda, \beta]\} \\
&\cup \{\lambda \in [x + \mu - \theta, y_2),\ \theta \in [x + \mu - y_2, \beta]\} .
\end{aligned}$$

First, it is direct to see that

$$\begin{aligned}
&\{\lambda \in [y_2, 0],\ \theta \in [x + \mu - \lambda, \beta]\} \cup \{\lambda \in [x + \mu - \theta, y_2),\ \theta \in [x + \mu - y_2, \beta]\} \\
&\quad \subseteq \{\lambda \leq 0, \theta \in [\min\{\beta, \mu + x\}, \beta]\} , \\
&\{\theta \in [x + \mu - y_3, \mu],\ \lambda \geq x + \mu - \theta\} \cup \{\theta \in [\alpha, x + \mu - y_3), \lambda \geq x + \mu\} \\
&\quad \cup \{\lambda \in [y_3, x + \mu),\ \theta \in [x + \mu - \lambda, x + \mu - y_3)\} \subseteq \{\lambda \geq x,\ \theta \in [\alpha, \mu]\} .
\end{aligned}$$

By union bound, the probability of the union of those five events is upper bounded by the sum of the probability of those two events, i.e., $\mathbb{P}(\lambda \geq x,\ \theta \in [\alpha, \mu]) + \mathbb{P}(\lambda \leq 0, \theta \in [\min\{\beta, \mu + x\}, \beta])$.

A. Separate Conditions. Those two events and one of the three remaining do not require to control (θ, λ) simultaneously, as they separate the conditions on (θ, λ). Thanks to the independence of (θ, λ), the probability of those events can be simply upper bounded by the product of the respective probability of those conditions. Therefore, we obtain

$$\begin{aligned}
&\mathbb{P}(\lambda \geq x,\ \theta \in [\alpha, \mu]) + \mathbb{P}(\lambda \leq 0, \theta \in [\min\{\beta, \mu + x\}, \beta]) + \mathbb{P}\left(\theta \in [x + \mu - y_1, \beta],\ \lambda \geq y_1\right) \\
&= \mathbb{P}(\lambda \geq x)\mathbb{P}(\theta \in [\alpha, \mu]) + \mathbb{P}(\lambda \leq 0)\mathbb{P}(\theta \in [\min\{\beta, \mu + x\}, \beta])
\end{aligned}$$

$$+ \mathbb{P}(\theta \in [x+\mu-y_1, \beta])\mathbb{P}(\lambda \geq y_1) \, .$$

B. Mixed Conditions. The two remaining events truly require to control (θ, λ) simultaneously, i.e., consider their convolution. The proof idea is the following: (1) we integrate one integral to obtain one survival function, (2) we make appear the other survival function artificially, (3) we upper bound the product of their survival functions on the whole set and (4) we integrate the remaining hazard function, whose integral is the cumulative hazard function. Let $\mathrm{d}G$ and $\mathrm{d}F$ be the probability measures of θ and λ on $\mathbb{R}$.

For all $s \in (\max\{x+\mu-\beta, 0\}, y_1)$, we have $\mathbb{P}(\lambda \geq s) \geq \mathbb{P}(\lambda \geq y_1) > 0$. Then, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\lambda \in (\max\{x+\mu-\beta, 0\}, y_1),\ \theta \in [x+\mu-\lambda, \beta]\right) \\
&= \int_{s \in (\max\{x+\mu-\beta,0\}, y_1)} \mathbb{P}(\theta \in [x+\mu-s, \beta]) \mathrm{d}F(s) \\
&= \int_{s \in (\max\{x+\mu-\beta,0\}, y_1)} \mathbb{P}(\lambda \geq s)\mathbb{P}(\theta \in [x+\mu-s, \beta]) \frac{1}{\mathbb{P}(\lambda \geq s)} \mathrm{d}F(s) \\
&\leq \sup_{s \in (\max\{x+\mu-\beta,0\}, y_1)} \{\mathbb{P}(\lambda \geq s)\mathbb{P}(\theta \in [x+\mu-s, \beta])\} \int_{s \in (\max\{x+\mu-\beta,0\}, y_1)} \frac{1}{\mathbb{P}(\lambda \geq s)} \mathrm{d}F(s) \\
&\leq \sup_{s \in (\max\{x+\mu-\beta,0\}, y_1)} \{\mathbb{P}(\lambda \geq s)\mathbb{P}(\theta \in [x+\mu-s, \beta])\} \left(-\log(\mathbb{P}(\lambda \geq y_1)) + \log(\mathbb{P}(\lambda \geq 0))\right) \, ,
\end{aligned}$$

where we used that $\mathbb{P}(\lambda \geq \max\{x+\mu-\beta, 0\}) \leq \mathbb{P}(\lambda \geq 0)$.

For all $z \in (\mu, x+\mu-y_1)$, we have $\mathbb{P}(\theta \in [z, \beta]) \geq \mathbb{P}(\theta \in [x+\mu-y_1, \beta]) > 0$. Then, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\theta \in (\mu, x+\mu-y_1),\ \lambda \geq x+\mu-\theta\right) \\
&= \int_{z \in (\mu, x+\mu-y_1)} \mathbb{P}(\lambda \geq x+\mu-z) \mathrm{d}G(z) \\
&= \int_{z \in (\mu, x+\mu-y_1)} \mathbb{P}(\lambda \geq x+\mu-z)\mathbb{P}(\theta \in [z, \beta]) \frac{1}{\mathbb{P}(\theta \in [z, \beta])} \mathrm{d}G(z) \\
&\leq \sup_{z \in (\mu, x+\mu-y_1)} \{\mathbb{P}(\lambda \geq x+\mu-z)\mathbb{P}(\theta \in [z, \beta])\} \int_{z \in (\mu, x+\mu-y_1)} \frac{1}{\mathbb{P}(\theta \in [z, \beta])} \mathrm{d}G(z) \\
&\leq \sup_{s \in (y_1, x)} \{\mathbb{P}(\lambda \geq s)\mathbb{P}(\theta \in [x+\mu-s, \beta])\} \\
&\qquad \cdot \left(-\log(\mathbb{P}(\theta \in [x+\mu-y_1, \beta])) + \log(\mathbb{P}(\theta \in [\mu, \beta]))\right) \, .
\end{aligned}$$

C. Combining Results. Putting everything together, we have, for all $y_1 \in (\max\{x+\mu-\beta, 0\}, x)$,

$$\begin{aligned}
&\mathbb{P}\left(\theta + \lambda \geq \mu + x\right) \leq \mathbb{P}(\lambda \geq x)\mathbb{P}(\theta \in [\alpha, \mu]) + \mathbb{P}(\lambda \leq 0)\mathbb{P}(\theta \in [\min\{\beta, \mu+x\}, \beta]) \\
&+ \mathbb{P}(\theta \in [x+\mu-y_1, \beta])\mathbb{P}(\lambda \geq y_1) \\
&+ \sup_{s \in (\max\{x+\mu-\beta,0\}, y_1)} \{\mathbb{P}(\lambda \geq s)\mathbb{P}(\theta \in [x+\mu-s, \beta])\} \left(-\log(\mathbb{P}(\lambda \geq y_1)) + \log(\mathbb{P}(\lambda \geq 0))\right) \\
&+ \sup_{s \in (y_1, x)} \{\mathbb{P}(\lambda \geq s)\mathbb{P}(\theta \in [x+\mu-s, \beta])\} \left(-\log(\mathbb{P}(\theta \in [x+\mu-y_1, \beta])) + \log(\mathbb{P}(\theta \in [\mu, \beta]))\right) \\
&\leq \mathbb{P}(\lambda \geq x)\mathbb{P}(\theta \in [\alpha, \mu]) + \mathbb{P}(\lambda \leq 0)\mathbb{P}(\theta \in [\min\{\beta, \mu+x\}, \beta]) \\
&+ \mathbb{P}(\theta \in [x+\mu-y_1, \beta])\mathbb{P}(\lambda \geq y_1) \\
&+ \sup_{s \in (\max\{x+\mu-\beta,0\}, x)} \{\mathbb{P}(\lambda \geq s)\mathbb{P}(\theta \in [x+\mu-s, \beta])\} \\
&\qquad \cdot \left(-\log(\mathbb{P}(\lambda \geq y_1)\mathbb{P}(\theta \in [x+\mu-y_1, \beta])) + \log(\mathbb{P}(\theta \in [\mu, \beta])\mathbb{P}(\lambda \geq 0))\right) \, ,
\end{aligned}$$

where the second inequality is obtained by extending the two suprema to $(\max\{x+\mu-\beta, 0\}, x)$, which is possible since multiplied by a positive value, and factorizing them together. Taking

$$y_1^\star \in \operatorname*{arg\,max}_{s \in (\max\{x+\mu-\beta,0\}, x)} \{\mathbb{P}(\lambda \geq s)\mathbb{P}(\theta \in [x+\mu-s, \beta])\}\} \, ,$$

and the change of variable $z = x+\mu-s$, i.e.,

$$\sup_{s \in (\max\{x+\mu-\beta,0\}, x)} \{\mathbb{P}(\lambda \geq s)\mathbb{P}(\theta \in [x+\mu-s, \beta])\}$$

$$= \sup_{z \in (\mu, \min\{\beta, x+\mu\})} \{\mathbb{P}(\theta \in [z, \beta])\mathbb{P}(\lambda \geq x + \mu - z)\} ,$$

concludes the proof of the upper bound on the upper tail.

II. Lower Bound on Upper Tail. Let $z \in (\mu, \min\{\beta, \mu + x\})$. Then, we have directly that

$$\{\theta \in [z, \beta],\ \lambda \geq \mu + x - z\} \subseteq \{\theta + \lambda \geq \mu + x\} .$$

Using independence, we obtain

$$\mathbb{P}(\theta \in [z, \beta])\mathbb{P}(\lambda \geq \mu + x - z) = \mathbb{P}(\theta \in [z, \beta], \lambda \geq \mu + x - z) \leq \mathbb{P}\left(\theta + \lambda \geq \mu + x\right) .$$

Taking the supremum over $z \in (\mu, \min\{\beta, \mu + x\})$ on the left hand side concludes the proof of the lower bound on the upper tail.

III. Upper/Lower Bound on Lower Tail. The third and forth inequalities are a direct consequence of the first and second inequalities applied to the two independent real random variables $-\theta$ and $-\lambda$ since (1) $-\theta$ has bounded support included in $[-\beta, -\alpha]$ and mean $-\mu \in (-\beta, -\alpha)$, and (2) $-\lambda$ has zero mean. Namely,

$$\begin{aligned}
&\mathbb{P}(\theta + \lambda \leq \mu - x) = \mathbb{P}(-\theta - \lambda \geq -\mu + x) ,\\
&\mathbb{P}(\lambda \leq -x)\mathbb{P}(\theta \in [\mu, \beta]) = \mathbb{P}(-\lambda \geq x)\mathbb{P}(-\theta \in [-\beta, -\mu]) ,\\
&\mathbb{P}(\lambda \geq 0)\mathbb{P}(\theta \in [\alpha, \max\{\alpha, \mu - x\}]) = \mathbb{P}(-\lambda \leq 0)\mathbb{P}(-\theta \in [\min\{-\alpha, x - \mu\}, -\alpha]) ,\\
&\mathbb{P}(\theta \in [\alpha, \mu])\mathbb{P}(\lambda \leq 0) = \mathbb{P}(-\theta \in [-\mu, -\alpha])\mathbb{P}(-\lambda \geq 0) ,\\
&\sup_{z \in (\max\{\alpha, \mu - x\}, \mu)} \{\mathbb{P}(\theta \in [\alpha, z])\mathbb{P}(\lambda \leq \mu - x - z)\}\\
&\quad = \sup_{\tilde{z} \in (-\mu, \min\{-\alpha, -\mu + x\})} \{\mathbb{P}(-\theta \in [\tilde{z}, -\alpha])\mathbb{P}(-\lambda \geq -\mu + x - \tilde{z})\} ,
\end{aligned}$$

where we used the change of variable $\tilde{z} = -z$. □

Properties on the Rate Function Lemma 11 gathers properties on the rate function f in Lemmas 12 and 12.

Lemma 11. *Let us define*

$$\forall x \geq 0, \quad f(x) := (x + 3 - \log 2)\exp(-x) . \tag{24}$$

On $\mathbb{R}_+$, the function f is twice continuously differentiable, positive, decreasing and strictly convex. It satisfies $f(0) > 1$, $\lim_{x \to +\infty} f(x) = 0$ and

$$f(x) \leq \delta \quad \iff \quad x \geq \overline{W}_{-1}(\log(1/\delta) + 3 - \log 2) - 3 + \log 2 ,$$

where $\overline{W}_{-1}$ is defined in Lemma 52.

Proof. Direct manipulation yields $f(0) = 3 - \log 2 > 1$, $\lim_{x \to +\infty} f(x) = 0$,
$\forall x \geq 0, \quad f'(x) = -(x + 2 - \log 2)\exp(-x) < 0 \quad$ and $\quad f''(x) = (x + 1 - \log 2)\exp(-x) > 0 .$
Using that $f(x) = e^{3 - \log 2}\exp(-h(x + 3 - \log 2))$ where $h(x) = x - \log(x)$, Lemma 52 yields

$$\begin{aligned}
f(x) \leq \delta \quad &\iff \quad h(x + 3 - \log 2) \geq \log\left(e^{3 - \log 2}/\delta\right)\\
&\iff \quad \overline{W}_{-1}(\log(1/\delta) + 3 - \log 2) - 3 + \log 2 \leq x .
\end{aligned}$$

□

Fixed Time Upper and Lower Tails Concentration Lemma 12 gives an upper and lower tails bound for a sum between independent Bernoulli and Laplace i.i.d. observations for a fixed time.

Lemma 12. *Let $\mu \in (0, 1)$ and $\epsilon > 0$. Let $Z_t = \sum_{s \in [t]} X_s$ where $X_s \sim Ber(\mu)$ are i.i.d. observations. Let $S_t = \sum_{s \in [n_t]} Y_s$ where $Y_s \sim Lap(1/\epsilon)$ are i.i.d. observations where $(n_t)_{t \in \mathbb{N}}$ be a piece-wise constant increasing function from $\mathbb{N}$ to $\mathbb{N}$. Let f as in Eq.* (24)*. Then, for all $t \in \mathbb{N}$ and all $x > 0$,*

$$\mathbb{P}(Z_t + S_t \geq t(x + \mu)) \leq f\left(t \inf_{z \in (\mu, \min\{1, x+\mu\})} \left\{-\frac{1}{t}\log\left(\mathbb{P}(Z_t \geq tz)\mathbb{P}(S_t \geq t(x + \mu - z))\right)\right\}\right)$$

$$\mathbb{P}(Z_t + S_t \leq t(\mu - x)) \leq f\left(t \inf_{z \in (\max\{0, \mu - x\}, \mu)} \left\{-\frac{1}{t}\log\left(\mathbb{P}(Z_t \leq tz)\mathbb{P}(S_t \leq t(\mu - x - z))\right)\right\}\right)$$

Proof. Let $t \in \mathbb{N}$ and $x > 0$. Then, Z_t and S_t are two independent real random variables such that (1) Z_t has bounded support included in $[0, t]$ and mean $t\mu \in (0, t)$ and (ii) S_t has zero mean. By symmetry of $\mathrm{Lap}(1/\epsilon)$ around 0, the cumulative sum of n_t observations (i.e., S_t) is also symmetric around 0. However, Z_t follows $\mathrm{Bin}(t, \mu)$ which can be skewed. Therefore, we have

$$\mathbb{P}(S_t \geq 0) = 1/2 = \mathbb{P}(S_t \leq 0) \quad \text{and} \quad \max\{\mathbb{P}(Z_t \in [t\mu, t]), \mathbb{P}(Z_t \in [0, t\mu])\} \leq 1 \,,$$
$$\forall z \in [0,1], \quad \mathbb{P}(Z_t \in [tz, t]) = \mathbb{P}(Z_t \geq tz) \quad \text{and} \quad \mathbb{P}(Z_t \in [0, tz]) = \mathbb{P}(Z_t \leq tz) \,.$$

Using that $z \mapsto \mathbb{P}(Z_t \geq tz)$ is decreasing on $(\mu, \min\{1, x + \mu\})$ and $z \mapsto \mathbb{P}(S_t \geq t(\mu - x - z)$ is increasing on $(\mu, \min\{1, x + \mu\})$, we obtain

$$\begin{aligned}
&\max\{\mathbb{P}(Z_t \geq t \min\{1, x + \mu\}))\mathbb{P}(S_t \leq 0), \mathbb{P}(S_t \geq tx)\mathbb{P}(Z_t \leq t\mu)\} \\
&\leq \sup_{z \in (\mu, \min\{1, x+\mu\})} \{\mathbb{P}(Z_t \geq tz)\mathbb{P}(S_t \geq t(x + \mu - z))\} \,.
\end{aligned}$$

Let us define $g(x) := x(3 - \log(2) - \log(x))$. Using Lemma 10 for (Z_t, S_t) and considering $tx > 0$ and $z \in (\mu, \min\{\beta, \mu + x\})$ (i.e., $tz \in (t\mu, t\min\{\beta, \mu + x\})$), we obtain

$$\mathbb{P}(Z_t + S_t \geq t(x + \mu)) \leq g\left(\sup_{z \in (\mu, \min\{1, x+\mu\})} \{\mathbb{P}(Z_t \geq tz)\mathbb{P}(S_t \geq t(x + \mu - z))\}\right) \,.$$

Let f as in Eq. (24) of Lemma 11. Then, we have $f(x) = g(\exp(-x))$. This concludes the proof of the upper bound on the upper tail. The second result is obtained similarly based on Lemma 10 and the above results. □

F.4 Tails Concentration of Cumulative Laplace Distributions

We derive time-uniform (Lemma 15) and fixed-time (Lemma 16) tails concentration for the cumulative sum of i.i.d. Laplace observations. Our proof technique is based on the Chernoff method and Ville's inequality as in Eq. (26). Therefore, we need to derive the convex conjuguate of the moment generating function of a Laplace distribution (Lemma 13). While the time-uniform result requires using the peeling method, the proof of the fixed-time concentration is simpler. To use the peeling method, we need to control the deviation of the process on slices of time (Lemma 14).

Convex Conjuguate of the Moment Generating Function of Laplace Distribution Let $\epsilon > 0$. The moment generating function of the Laplace distribution $\mathrm{Lap}(1/\epsilon)$ is defined as

$$\forall \lambda \in (0, \epsilon), \quad \psi_{\mathrm{Lap},\epsilon}(\lambda) = \log \mathbb{E}_{X \sim \mathrm{Lap}(1/\epsilon)}\left[\exp(\lambda X)\right] = -\log(1 - \lambda^2/\epsilon^2) \,. \tag{25}$$

Lemma 13 explicits the convex conjuguate of $\psi_{\mathrm{Lap},\epsilon}$ and its associated maximizer.

Lemma 13. *Let $\psi_{Lap,\epsilon}$ as in Eq.* (25)*. Let us define*

$$\forall x > 0, \quad \psi^\star_{Lap,\epsilon}(x) := \max_{\lambda \in (0,\epsilon)} \{\lambda x - \psi_{Lap,\epsilon}(\lambda)\} \quad \textit{and} \quad \lambda(x) := \arg\max_{\lambda \in (0,\epsilon)} \{\lambda x - \psi_{Lap,\epsilon}(\lambda)\} \,.$$

Then, for all $x > 0$, we have

$$\lambda(x) = \frac{1}{x}\left(\sqrt{1 + (x\epsilon)^2} - 1\right) \in (0, \epsilon) \quad \textit{and} \quad \psi^\star_{Lap,\epsilon}(x) = h(\epsilon x) > 0 \,.$$

where h is defined in Eq. (31)*.*

Proof. Let $f(\lambda) = \lambda x - \psi_{\mathrm{Lap},\epsilon}(\lambda)$ for all $\lambda \in (0, \epsilon)$. Direct manipulation yields that

$$\forall \lambda \in (0, \epsilon), \quad f'(\lambda) = x - \frac{2\lambda}{\epsilon^2 - \lambda^2} \quad \text{and} \quad f''(\lambda) = -2\frac{\epsilon^2 + \lambda^2}{(\epsilon^2 - \lambda^2)^2} < 0 \,.$$

Moreover, for all $\lambda \in (0, \epsilon)$, we have

$$f'(\lambda) = 0 \quad \iff \quad \lambda^2 + 2\lambda/x - \epsilon^2 = 0 \quad \iff \quad \lambda = \frac{1}{x}\left(\sqrt{1 + (x\epsilon)^2} - 1\right) \,.$$

We used that the second solution to the second order polynomial equation is negative, hence not in $(0, \epsilon)$. Moreover, it is direct to see that $\frac{1}{x}\left(\sqrt{1 + (x\epsilon)^2} - 1\right) \in (0, \epsilon)$ since $\sqrt{1 + x^2} - 1 \leq x$, as it

is equivalent to $1 + x^2 \leq (x+1)^2$ which is true when $x > 0$. Since f is strictly concave, the above computation gives its unique maximizer on $(0, \epsilon)$, namely we have $\lambda(x) = \frac{1}{x}\left(\sqrt{1+(x\epsilon)^2} - 1\right)$. Moreover, the convex conjuguate of $\psi_{\text{Lap},\epsilon}$ is

$$\begin{aligned}\psi^\star_{\text{Lap},\epsilon}(x) = f(\lambda(x)) &= \sqrt{1+(x\epsilon)^2} - 1 + \log\left(1 - \frac{1}{(x\epsilon)^2}\left(\sqrt{1+(x\epsilon)^2} - 1\right)^2\right) \\ &= \sqrt{1+(x\epsilon)^2} - 1 + \log\left(\frac{2}{(x\epsilon)^2}\left(\sqrt{1+(x\epsilon)^2} - 1\right)\right) .\end{aligned}$$

This concludes the proof. □

Test Martingale for Cumulative Laplace Observations Let $\epsilon > 0$ and $S_t = \sum_{s\in[t]} Y_s$ where $Y_s \sim \text{Lap}(1/\epsilon)$ are i.i.d. observations. Let us define

$$\forall \lambda \in (0, \epsilon), \quad M_t(\lambda) := \exp(\lambda S_t - t\psi_{\text{Lap},\epsilon}(\lambda)) .$$

It is direct to see that $M_0(\lambda) = 0$ almost surely and

$$\mathbb{E}[M_t(\lambda) \mid \mathcal{F}_{t-1}] = M_{t-1}(\lambda)\mathbb{E}_{X\sim\text{Lap}(1/\epsilon)}[\exp(\lambda X - \psi_{\text{Lap},\epsilon}(\lambda))] = M_{t-1}(\lambda) .$$

Therefore, $M_t(\lambda)$ is a test martingale. Using Ville's inequality [Ville, 1939] yields that

$$\forall \delta \in (0,1),\ \forall \lambda \in (0, \epsilon), \quad \mathbb{P}\left(\exists t \in \mathbb{N},\ \lambda S_t - t\psi_{\text{Lap},\epsilon}(\lambda) \geq \log(1/\delta)\right) \leq \delta . \tag{26}$$

Time Uniform Tails Concentration Lemma 14 controls the deviation of the process on slices of time.

Lemma 14. *Let $\epsilon > 0$ and $S_t = \sum_{s\in[t]} Y_s$ where $Y_s \sim$ Lap$(1/\epsilon)$ are i.i.d. observations. Let $N > 0$. For all $x > 0$, there exists $\lambda(x)$ such that for all $t \geq N$,*

$$\{S_t \geq tx\} \subseteq \{\lambda(x)S_t - t\psi_{\text{Lap},\epsilon}(\lambda(x)) \geq Nh(\epsilon x)\} ,$$

where $\lambda(x)$ as in Lemma 13 and h as in Eq. (31)*.*

Proof. Using Lemma 13, we obtain $\lambda(x) \in (0, \epsilon)$ and $\psi^\star_{\text{Lap},\epsilon}(x) = h(\epsilon x) > 0$, hence $t\psi^\star_{\text{Lap},\epsilon}(x) \geq N\psi^\star_{\text{Lap},\epsilon}(x)$ for $t \geq N$. Then, direct computations yield

$$\begin{aligned}S_t \geq tx &\implies \lambda(x)S_t - t\psi_{\text{Lap},\epsilon}(\lambda(x)) \geq t\left(x\lambda(x) - \psi_{\text{Lap},\epsilon}(\lambda(x))\right) = t\psi^\star_{\text{Lap},\epsilon}(x) \\ &\implies \lambda S_t - t\psi_{\text{Lap},\epsilon}(\lambda) \geq N\psi^\star_{\text{Lap},\epsilon}(x) = Nh(\epsilon x) .\end{aligned}$$

This concludes the proof. □

Lemma 15 gives time-uniform tails concentration for the cumulative sum of i.i.d. Laplace observations. It is obtained by applying Lemma 14 on slices of time with geometric growth rate.

Lemma 15. *Let $\delta \in (0,1)$. Let $\gamma > 0$, $s > 1$ and ζ be the Riemann ζ function. Let h^{-1} be the inverse of h defined as in Eq.* (31)*, which is well-defined by Lemma 28. Let $\epsilon > 0$ and $S_t = \sum_{s\in[t]} Y_s$ where $Y_s \sim$ Lap$(1/\epsilon)$ are i.i.d. observations. Then,*

$$\begin{aligned}&\mathbb{P}\left(\exists t \in \mathbb{N},\ S_t \geq \frac{t}{\epsilon}h^{-1}\left(\frac{1+\gamma}{t}\left(\log\left(\frac{\zeta(s)}{\delta}\right) + s\log\left(1 + \log_{1+\gamma} t\right)\right)\right)\right) \leq \delta , \\ &\mathbb{P}\left(\exists t \in \mathbb{N},\ S_t \leq -\frac{t}{\epsilon}h^{-1}\left(\frac{1+\gamma}{t}\left(\log\left(\frac{\zeta(s)}{\delta}\right) + s\log\left(1 + \log_{1+\gamma} t\right)\right)\right)\right) \leq \delta .\end{aligned}$$

Proof. Let us define the geometric grid $N_i = (1+\gamma)^{i-1}$, hence we have $\mathbb{N} = \bigcup_{i\in\mathbb{N}}[N_i, N_{i+1})$. For all $i \in \mathbb{N}$, let $x_i(\delta) > 0$ to be defined later, and $\lambda(x_i(\delta))$ as in Lemma 14. For all $t \in \mathbb{N}$, let $g(t,\delta)$ to be defined later such that $g(t,\delta) \geq x_i(\delta)$ for $t \in [N_i, N_{i+1})$. Using Lemma 14 with $x_i(\delta) > 0$ and $g(t,\delta) \geq x_i(\delta)$ for $t \in [N_i, N_{i+1})$, a union bound yields that

$$\mathbb{P}\left(\exists t \in \mathbb{N},\ S_t \geq tg(t,\delta)\right)$$

$$
\begin{aligned}
&\leq \sum_{i\in\mathbb{N}} \mathbb{P}\left(\exists t \in [N_i, N_{i+1}) : S_t \geq t x_i(\delta)\right) \\
&\leq \sum_{i\in\mathbb{N}} \mathbb{P}\left(\exists t \in [N_i, N_{i+1}) : \lambda(x_i(\delta)) S_t - t\psi_{\text{Lap},\epsilon}(\lambda(x_i(\delta))) \geq N_i h(\epsilon x_i(\delta))\right) \\
&\leq \sum_{i\in\mathbb{N}} e^{-N_i h(\epsilon x_i(\delta))} \,,
\end{aligned}
$$

where the last inequality uses Ville's inequality as in Eq. (26) for all $i \in \mathbb{N}$. Let us define

$$
\begin{aligned}
g(t,\delta) &= \frac{1}{\epsilon} h^{-1}\left(\frac{1+\gamma}{t}\left(\log\left(\frac{\zeta(s)}{\delta}\right) + s\log\left(1+\log_{1+\gamma}(t)\right)\right)\right) \,, \\
x_i(\delta) &= \frac{1}{\epsilon} h^{-1}\left(\frac{1}{N_i}\log\left(\frac{i^s\zeta(s)}{\delta}\right)\right) \,.
\end{aligned}
$$

Using Lemma 28, we obtain that $x_i(\delta) > 0$ and that h^{-1} is increasing on $\mathbb{R}_+^\star$. Using $t \in [N_i, N_{i+1})$ and $i = 1 + \log_{1+\gamma} N_i$, we obtain

$$
\begin{aligned}
g(t,\delta) &\geq \frac{1}{\epsilon} h^{-1}\left(\frac{1}{N_i}\left(\log\left(\frac{\zeta(s)}{\delta}\right) + s\log\left(1+\log_{1+\gamma}(t)\right)\right)\right) \\
&\geq \frac{1}{\epsilon} h^{-1}\left(\frac{1}{N_i}\log\left(\frac{i^s\zeta(s)}{\delta}\right)\right) = x_i(\delta) \,.
\end{aligned}
$$

Therefore, we have

$$
\mathbb{P}\left(\exists t \in \mathbb{N},\ S_t \geq t g(t,\delta)\right) \leq \sum_{i\in\mathbb{N}} e^{-N_i h(\epsilon x_i(\delta))} \leq \frac{\delta}{\zeta(s)} \sum_{i\in\mathbb{N}} \frac{1}{i^s} = \delta \,.
$$

This concludes the proof of the first result.

By symmetry of the Lap$(1/\epsilon)$ around zero, the cumulative sum of i.i.d. observations is symmetric around zero. Combining the first result with the symmetry around zero yields the second result. □

Fixed Time Tails Concentration When the time is fixed and not random, there is no need to consider slices of time and we can directly control the deviation of the process.

Lemma 16. *Let $\epsilon > 0$ and $S_t = \sum_{s\in[t]} Y_s$ where $Y_s \sim$ Lap$(1/\epsilon)$ are i.i.d. observations. Let h as in Eq.* (31)*. Then,*

$$
\begin{aligned}
&\forall t \in \mathbb{N}, \forall x > 0, \quad \mathbb{P}(S_t \geq tx) \leq \exp(-th(\epsilon x)) \,, \\
&\forall t \in \mathbb{N}, \forall x > 0, \quad \mathbb{P}(S_t \leq -tx) \leq \exp(-th(\epsilon x)) \,.
\end{aligned}
$$

Proof. The first result can be obtained with the same manipulation as in the proof of Lemma 15, i.e., combining Ville's inequality in Eq. (26) with Lemma 14 at $N = t$.

By symmetry of the Lap$(1/\epsilon)$ around zero, the cumulative sum of i.i.d. observations is symmetric around zero. Combining the first result with the symmetry around zero yields the second result. □

F.5 Fixed Time Tails Concentration of Cumulative Bernoulli Distributions

The fixed time upper and lower tail concentration of cumulative Bernoulli distributions are well-studied. Using the Chernoff method yields Lemma 17, whose proof is omitted since it is a classic result.

Lemma 17 (Chernoff Tail Bound for Bernoulli Distributions [Boucheron et al., 2013])**.** *Let $\mu \in (0,1)$ and $Z_t = \sum_{s\in[t]} X_s$ where $X_s \sim$ Ber(μ) are i.i.d. observations. Then,*

$$
\begin{aligned}
&\forall t \in \mathbb{N}, \forall x \in (\mu, 1), \quad \mathbb{P}(Z_t \geq tx) \leq \exp(-t\mathrm{kl}(x,\mu)) \,, \\
&\forall t \in \mathbb{N}, \forall x \in (0, \mu), \quad \mathbb{P}(Z_t \leq tx) \leq \exp(-t\mathrm{kl}(x,\mu)) \,.
\end{aligned}
$$

F.6 Fixed Time Tails Concentration for a Convolution between Bernoulli and Laplace Distributions

We provide upper (Lemma 18) and lower (Lemma 19) tail concentrations for a sum (i.e., convolution) between independent Bernoulli and Laplace i.i.d. observations for a fixed time.

Fixed Time Upper Tail Concentration Lemma 18 shows an upper tail concentration on the sum (i.e., convolution) between independent Bernoulli and Laplace i.i.d. observations.

Lemma 18. *Let $\mu \in (0,1)$ and $Z_t = \sum_{s\in[t]} X_s$ where $X_s \sim Ber(\mu)$ are i.i.d. observations. Let $(n_t)_{t\in\mathbb{N}}$ be a piece-wise constant increasing function from $\mathbb{N}$ to $\mathbb{N}$. Let $\epsilon > 0$ and $S_t = \sum_{s\in[n_t]} Y_s$ where $Y_s \sim Lap(1/\epsilon)$ are i.i.d. observations. Then,*

$$\forall t \in \mathbb{N},\ \forall x > 0, \quad \mathbb{P}(Z_t + S_t \geq t(\mu + x)) \leq f\left(t\widetilde{d}_\epsilon^-(\mu + x, \mu, t/n_t)\right) ,$$

where f is defined in Eq. (24) *and $\widetilde{d}_\epsilon^-$ is defined in Eq.* (32)*.*

Proof. Let $t \in \mathbb{N}$ and $x > 0$. Combining Lemmas 16 and 17, we obtain, for all $x > 0$ and all $z \in (\mu, \min\{1, x + \mu\})$,

$$-\frac{1}{t}\log\left(\bar{G}_t(tz)\bar{F}_{n_t}(t(x + \mu - z))\right) \geq \mathrm{kl}(z, \mu) + \frac{n_t}{t}h\left(\frac{t}{n_t}\epsilon(x + \mu - z)\right) ,$$

where we used that $x + \mu - z > 0$. Taking the infimum on $(\mu, \min\{1, x + \mu\})$ on both sides and using that $[x + \mu]_0^1 = \min\{1, x + \mu\} > \mu$, we obtain

$$\inf_{z\in(\mu,\min\{1,x+\mu\})}\left\{-\frac{1}{t}\log\left(\bar{G}_t(tz)\bar{F}_{n_t}(t(x + \mu - z))\right)\right\} \geq \widetilde{d}_\epsilon^-(\mu + x, \mu, t/n_t) ,$$

where $\widetilde{d}_\epsilon^-$ is defined in Eq. (32). Since f is decreasing on $\mathbb{R}_+$ (Lemma 11), using Lemma 12 yields

$$\mathbb{P}(Z_t + S_t \geq t(x + \mu)) \leq f\left(t\widetilde{d}_\epsilon^-(\mu + x, \mu, t/n_t)\right) .$$

which concludes the proof. □

Fixed Time Lower Tail Concentration Lemma 19 shows a lower tail concentration on the sum (i.e., convolution) between independent Bernoulli and Laplace i.i.d. observations.

Lemma 19. *Let $\mu \in (0,1)$ and $Z_t = \sum_{s\in[t]} X_s$ where $X_s \sim Ber(\mu)$ are i.i.d. observations. Let $(n_t)_{t\in\mathbb{N}}$ be a piece-wise constant increasing function from $\mathbb{N}$ to $\mathbb{N}$. Let $\epsilon > 0$ and $S_t = \sum_{s\in[n_t]} Y_s$ where $Y_s \sim Lap(1/\epsilon)$ are i.i.d. observations. Then,*

$$\forall t \in \mathbb{N},\ \forall x > 0, \quad \mathbb{P}(Z_t + S_t \leq t(\mu - x)) \leq f\left(t\widetilde{d}_\epsilon^+(\mu - x, \mu, t/n_t)\right) ,$$

where f is defined in Eq. (24) *and $\widetilde{d}_\epsilon^+$ is defined in Eq.* (32)*.*

Proof. Let $t \in \mathbb{N}$ and $x > 0$. Combining Lemmas 16 and 17, we obtain, for all $x > 0$ and all $z \in (\max\{0, \mu - x\}, \mu)$,

$$-\frac{1}{t}\log\left(G_t(tz)F_{n_t}(t(\mu - x - z))\right) \geq \mathrm{kl}(z, \mu) + \frac{n_t}{t}h\left(\frac{t}{n_t}\epsilon(z + x - \mu)\right) ,$$

where we used that $\mu - x - z < 0$. Taking the infimum on $z \in (\max\{0, \mu - x\}, \mu)$ on both sides and using that $[\mu - x]_0^1 = \max\{0, \mu - x\} < \mu$, we obtain

$$\inf_{z\in(\max\{0,\mu-x\},\mu)}\left\{-\frac{1}{t}\log\left(G_t(tz)F_{n_t}(t(\mu - x - z))\right)\right\} \geq \widetilde{d}_\epsilon^+(\mu - x, \mu, t/n_t) ,$$

where $\widetilde{d}_\epsilon^+$ is defined in Eq. (32). Since f is decreasing on $\mathbb{R}_+$ (Lemma 11), using Lemma 12 yields

$$\mathbb{P}(Z_t + S_t \geq t(\mu - x)) \leq f\left(t\widetilde{d}_\epsilon^+(\mu - x, \mu, t/n_t)\right) ,$$

which concludes the proof. □

F.7 Geometric Grid Time Uniform Tails Concentration for a Convolution between Bernoulli and Laplace Distributions

We provide upper (Lemma 20) and lower (Lemma 21) tail concentrations for a sum (i.e., convolution) between independent Bernoulli and Laplace i.i.d. observations that holds time uniformly on a geometric grid.

Geometric Grid Time Uniform Upper Tail Concentration Lemma 20 gives a threshold ensuring that a geometric grid time uniform upper tail concentration holds with probability at least $1-\delta$.

Lemma 20. *Let $\delta \in (0,1)$. Let $(\tilde{\mu}_n, \tilde{N}_n, k_n)$ are given by GPE$_\eta(\epsilon)$. Let $s > 1$ and ζ be the Riemann ζ function. Let $\overline{W}_{-1}(x) = -W_{-1}(-e^{-x})$ for all $x \geq 1$, where W_{-1} is the negative branch of the Lambert W function. It satisfies $\overline{W}_{-1}(x) \approx x + \log x$, see Lemma 52. Let us define*

$$c(k,\delta) = \overline{W}_{-1}\left(\log\left(1/\delta\right) + s\log(k) + \log(\zeta(s)) + 3 - 2\log 2\right) - 3 + 2\log 2\,. \tag{27}$$

For all $a \in [K]$, let us define

$$\mathcal{E}_{\delta,a,-} = \left\{\forall n \in \mathbb{N},\ \tilde{N}_{n,a}\widetilde{d}_\epsilon^-(\tilde{\mu}_{n,a}, \mu_a, \tilde{N}_{n,a}/k_{n,a}) \leq c(k_{n,a},\delta)\right\}\,, \tag{28}$$

where $\widetilde{d}_\epsilon^-$ is defined in Eq. (32)*. Then, we have $\mathbb{P}_{\boldsymbol{\nu}\pi}(\mathcal{E}_{\delta,a,-}^{\complement}) \leq \delta$ for all $a \in [K]$.*

Proof. Let us define the geometric grid $N_i = (1+\eta)^{i-1}$, hence we have $\mathbb{N} = \bigcup_{i\in\mathbb{N}}[N_i, N_{i+1})$. Let $a \in [K]$. If $\tilde{N}_{n,a} \in [N_i, N_{i+1})$, then we have $\tilde{N}_{n,a} = \lceil N_i \rceil$ and $k_{n,a} = i$. By union bound, we obtain

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\nu}\pi}(\mathcal{E}_{\delta,a,-}^{\complement}) = \mathbb{P}_{\boldsymbol{\nu}\pi}\left(\exists n \in \mathbb{N},\ \tilde{N}_{n,a}\widetilde{d}_\epsilon^-(\tilde{\mu}_{n,a}, \mu_a, \tilde{N}_{n,a}/k_{n,a}) \geq c(k_{n,a},\delta)\right)\\
&\leq \sum_{i\in\mathbb{N}} \mathbb{P}_{\boldsymbol{\nu}\pi}\left(\exists i \in \mathbb{N},\ (\tilde{N}_{n,a}, k_{n,a}) = (\lceil N_i \rceil, i)\ \wedge\ \tilde{N}_{n,a}\widetilde{d}_\epsilon^-(\tilde{\mu}_{n,a}, \mu_a, \tilde{N}_{n,a}/k_{n,a}) \geq c(k_{n,a},\delta)\right)\\
&= \sum_{i\in\mathbb{N}} \mathbb{P}\left(\lceil N_i \rceil \widetilde{d}_\epsilon^-((Z_{\lceil N_i \rceil} + S_i)/\lceil N_i \rceil, \mu_a, \lceil N_i \rceil/i) \geq c(i,\delta)\right)\,,
\end{aligned}$$

where $Z_{\lceil N_i \rceil}$ is the cumulative sum of $\lceil N_i \rceil$ i.i.d. observations from Ber(μ_a) and S_i is the cumulative sum of i i.i.d. observations from Lap$(1/\epsilon)$.

For all $i \in \mathbb{N}$, let $x_i > 0$ be the unique solution of $\lceil N_i \rceil \widetilde{d}_\epsilon^-(x + \mu_a, \mu_a, \lceil N_i \rceil/i) = c(i,\delta)$, which exists by Lemma 33. Then, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\lceil N_i \rceil \widetilde{d}_\epsilon^-((Z_{\lceil N_i \rceil} + S_i)/\lceil N_i \rceil, \mu_a, \lceil N_i \rceil/i) \geq c(i,\delta)\right)\\
&= \mathbb{P}\left(\widetilde{d}_\epsilon^-((Z_{\lceil N_i \rceil} + S_i)/\lceil N_i \rceil, \mu_a, \lceil N_i \rceil/i) \geq \widetilde{d}_\epsilon^-(x_i + \mu_a, \mu_a, \lceil N_i \rceil/i)\right)\\
&\leq \mathbb{P}(Z_{\lceil N_i \rceil} + S_i \geq \lceil N_i \rceil(x_i + \mu_a)) \leq f\left(\lceil N_i \rceil \widetilde{d}_\epsilon^-(x_i + \mu_a, \mu_a, \lceil N_i \rceil/i)\right) = f\left(c(i,\delta)\right)\,,
\end{aligned}$$

where $f(x) := (x + 3 - \log 2)\exp(-x)$ for all $x \geq 0$. The first and the last equalities are otained by definition of x_i, i.e., $\lceil N_i \rceil \widetilde{d}_\epsilon^-(x + \mu_a, \mu_a, \lceil N_i \rceil/i) = c(i,\delta)$. The first inequality is obtained by using Lemma 34, and the second inequality is obtained by using Lemma 18. Using Lemma 11 yields

$$f(x) \leq \delta \quad \Longleftrightarrow \quad \overline{W}_{-1}(\log\left(1/\delta\right) + 3 - \log 2) - 3 + \log 2 \leq x\,.$$

Taking

$$c(i,\delta) = \overline{W}_{-1}(\log\left(i^s\zeta(s)/\delta\right) + 3 - \log 2) - 3 + \log 2\,,$$

we can conclude the proof since $\mathbb{P}_{\boldsymbol{\nu}\pi}(\mathcal{E}_{\delta,a,-}^{\complement}) \leq \sum_{i\in\mathbb{N}} f\left(c(i,\delta)\right) \leq \sum_{i\in\mathbb{N}} \frac{\delta}{\zeta(s)i^s} \leq \delta$. □

Geometric Grid Time Uniform Lower Tail Concentration Lemma 21 gives a threshold ensuring that a geometric grid time uniform lower tail concentration holds with probability at least $1-\delta$.

Lemma 21. *Let $\delta \in (0,1)$. Let $(\tilde{\mu}_n, \tilde{N}_n, k_n)$ are given by GPE$_\eta(\epsilon)$. Let c as in Eq.* (27)*. For all $a \in [K]$, let us define*

$$\mathcal{E}_{\delta,a,+} = \left\{ \forall n \in \mathbb{N},\ \tilde{N}_{n,a} \widetilde{d}_\epsilon^+(\tilde{\mu}_{n,a}, \mu_a, \tilde{N}_{n,a}/k_{n,a}) \leq c(k_{n,a}, \delta) \right\} , \tag{29}$$

where $\widetilde{d}_\epsilon^+$ is defined in Eq. (32)*. Then, we have $\mathbb{P}_{\boldsymbol{\nu}\pi}(\mathcal{E}_{\delta,a,+}^{\complement}) \leq \delta$ for all $a \in [K]$.*

Proof. Let us define the geometric grid $N_i = (1+\eta)^{i-1}$, hence we have $\mathbb{N} = \bigcup_{i\in\mathbb{N}} [N_i, N_{i+1})$. Let $a \in [K]$. If $\tilde{N}_{n,a} \in [N_i, N_{i+1})$, then we have $\tilde{N}_{n,a} = \lceil N_i \rceil$ and $k_{n,a} = i$. By union bound, we obtain

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\nu}\pi}(\mathcal{E}_{\delta,a,+}^{\complement}) = \mathbb{P}_{\boldsymbol{\nu}\pi}\left(\exists n \in \mathbb{N},\ \tilde{N}_{n,a} \widetilde{d}_\epsilon^+(\tilde{\mu}_{n,a}, \mu_a, \tilde{N}_{n,a}/k_{n,a}) \geq c(k_{n,a}, \delta) \right) \\
&\leq \sum_{i\in\mathbb{N}} \mathbb{P}_{\boldsymbol{\nu}\pi}\left(\exists i \in \mathbb{N},\ (\tilde{N}_{n,a}, k_{n,a}) = (\lceil N_i \rceil, i) \ \wedge\ \tilde{N}_{n,a} \widetilde{d}_\epsilon^+(\tilde{\mu}_{n,a}, \mu_a, \tilde{N}_{n,a}/k_{n,a}) \geq c(k_{n,a}, \delta) \right) \\
&= \sum_{i\in\mathbb{N}} \mathbb{P}\left(\lceil N_i \rceil \widetilde{d}_\epsilon^+((Z_{\lceil N_i \rceil} + S_i)/\lceil N_i \rceil, \mu_a, \lceil N_i \rceil / i) \geq c(i,\delta) \right) ,
\end{aligned}$$

where $Z_{\lceil N_i \rceil}$ is the cumulative sum of $\lceil N_i \rceil$ i.i.d. observations from Ber(μ_a) and S_i is the cumulative sum of i i.i.d. observations from Lap$(1/\epsilon)$.

For all $i \in \mathbb{N}$, let $x_i > 0$ be the unique solution of $\lceil N_i \rceil \widetilde{d}_\epsilon^+(\mu_a - x, \mu_a, \lceil N_i \rceil / i) = c(i,\delta)$, which exists by Lemma 33. Then, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\lceil N_i \rceil \widetilde{d}_\epsilon^+((Z_{\lceil N_i \rceil} + S_i)/\lceil N_i \rceil, \mu_a, \lceil N_i \rceil / i) \geq c(i,\delta) \right) \\
&= \mathbb{P}\left(\widetilde{d}_\epsilon^+((Z_{\lceil N_i \rceil} + S_i)/\lceil N_i \rceil, \mu_a, \lceil N_i \rceil / i) \geq \widetilde{d}_\epsilon^+(\mu_a - x_i, \mu_a, \lceil N_i \rceil / i) \right) \\
&\leq \mathbb{P}(Z_{\lceil N_i \rceil} + S_i \leq \lceil N_i \rceil (\mu_a - x_i)) \leq f\left(\lceil N_i \rceil \widetilde{d}_\epsilon^+(\mu_a - x_i, \mu_a, \lceil N_i \rceil / i) \right) = f\left(c(i,\delta)\right) \leq \frac{\delta}{\zeta(s) i^s}
\end{aligned}$$

where $f(x) := (x + 3 - \log 2)\exp(-x)$ for all $x \geq 0$. The first and the last equalities are otained by definition of x_i, i.e., $\lceil N_i \rceil \widetilde{d}_\epsilon^+(\mu_a - x, \mu_a, \lceil N_i \rceil / i) = c(i,\delta)$. The first inequality is obtained by using Lemma 34, and the second inequality is obtained by using Lemma 19. The last inequality uses the same derivations based on Lemma 11 as in the proof of Lemma 20 by taking

$$c(i,\delta) = \overline{W}_{-1}(\log(i^s \zeta(s)/\delta) + 3 - \log 2) - 3 + \log 2 .$$

This concludes the proof since $\mathbb{P}_{\boldsymbol{\nu}\pi}(\mathcal{E}_{\delta,a,-}^{\complement}) \leq \sum_{i\in\mathbb{N}} \frac{\delta}{\zeta(s) i^s} \leq \delta$. □

G Divergence, Transportation Cost and Characteristic Time

Appendix G is organized as follow. First, we derive regularity properties for the signed (modified) divergences $d_\epsilon^\pm$ (Appendix G.1) and $\widetilde{d}_\epsilon^\pm$ (Appendix G.1.1). Second, we derive regularity properties the (modified) transportation costs $W_{\epsilon,a,b}$ (Appendix G.2) and $\widetilde{W}_{\epsilon,a,b}$ (Appendix G.2.1) for a pair of arms (a,b). Third, we study the characteristic time for ϵ-global DP BAI (Appendix G.3).

G.1 Signed Divergence

Recall $[x]_0^1 := \max\{0, \min\{1, \lambda\}\}$ and

$$\forall (\lambda,\mu) \in (0,1)^2, \quad \mathrm{kl}(\lambda,\mu) := \lambda \log\left(\frac{\lambda}{\mu}\right) + (1-\lambda)\log\left(\frac{1-\lambda}{1-\mu}\right)$$

where kl is infinity when $\{\mu,\lambda\} \cap \{0,1\} \neq \emptyset$. The signed divergences $d_\epsilon^\pm$ are defined in Eq. (3), i.e.,

$$\begin{aligned}
\forall (\lambda,\mu) \in \mathbb{R}\times[0,1], \quad d_\epsilon^-(\lambda,\mu) &:= \mathbb{1}\left(\mu < [\lambda]_0^1\right) \inf_{z\in[\mu,[\lambda]_0^1]} \left\{ \mathrm{kl}(z,\mu) + \epsilon([\lambda]_0^1 - z) \right\} , \\
d_\epsilon^+(\lambda,\mu) &:= \mathbb{1}\left(\mu > [\lambda]_0^1\right) \inf_{z\in[[\lambda]_0^1,\mu]} \left\{ \mathrm{kl}(z,\mu) + \epsilon(z - [\lambda]_0^1) \right\} .
\end{aligned}$$

Lemma 22 relates d_ϵ and $d_\epsilon^\pm$.

Lemma 22. *Let $d_\epsilon^\pm$ and d_ϵ as in Eq.* (3) *and* (1)*. Let $(\kappa,\nu)\in\mathcal{F}^2$ with means $(\lambda,\mu)\in(0,1)^2$. Then,*

$$\mathrm{d}_\epsilon(\kappa,\nu)=\begin{cases}0 & \text{if } \lambda=\mu\\ d_\epsilon^-(\lambda,\mu) & \text{if } \mu<\lambda\\ d_\epsilon^+(\lambda,\mu) & \text{if } \mu>\lambda\end{cases}.$$

Proof. When $\lambda=\mu$, we have $\mathrm{d}_\epsilon(\kappa,\nu)=0$ by taking $\varphi=\nu$ and using the non-negativity of d_ϵ.

Let $\varphi\in\mathcal{F}$ with mean $z\in(0,1)$. When $\mu<\lambda$, we have

$$\begin{aligned}\mathrm{d}_\epsilon(\kappa,\nu)&=\min\{\inf_{z\in(0,\mu)}\{\mathrm{kl}(z,\mu)+\epsilon(\lambda-z)\},\inf_{z\in[\mu,\lambda]}\{\mathrm{kl}(z,\mu)+\epsilon(\lambda-z)\},\\&\qquad\inf_{z\in(\lambda,1)}\{\mathrm{kl}(z,\mu)+\epsilon(z-\lambda)\}\}\\&=\inf_{z\in[\mu,\lambda]}\{\mathrm{kl}(z,\mu)+\epsilon(\lambda-z)\}=d_\epsilon^-(\lambda,\mu),\end{aligned}$$

where we partitioned $(0,1)$ and used that (1) $z\mapsto\mathrm{kl}(z,\mu)+\epsilon(z-\lambda)$ is increasing on $(\lambda,1)$, hence the infimum on this interval is achieved at λ, and (2) $z\mapsto\mathrm{kl}(z,\mu)+\epsilon(\lambda-z)$, is decreasing on $(0,\mu)$, hence the infimum on this interval is achieved at μ.

When $\mu>\lambda$, we have

$$\begin{aligned}\mathrm{d}_\epsilon(\kappa,\nu)&=\min\{\inf_{z\in(0,\lambda)}\{\mathrm{kl}(z,\mu)+\epsilon(\lambda-z)\},\inf_{z\in[\lambda,\mu]}\{\mathrm{kl}(z,\mu)+\epsilon(z-\lambda)\},\\&\qquad\inf_{z\in(\mu,1)}\{\mathrm{kl}(z,\mu)+\epsilon(z-\lambda)\}\}\\&=\inf_{z\in[\lambda,\mu]}\{\mathrm{kl}(z,\mu)+\epsilon(z-\lambda)\}=d_\epsilon^+(\lambda,\mu),\end{aligned}$$

where we partitioned $(0,1)$ and used that (1) $z\mapsto\mathrm{kl}(z,\mu)+\epsilon(z-\lambda)$ is increasing on $(\mu,1)$, hence the infimum on this interval is achieved at μ, and (2) $z\mapsto\mathrm{kl}(z,\mu)+\epsilon(\lambda-z)$, is decreasing on $(0,\lambda)$, hence the infimum on this interval is achieved at λ. □

Lemma 23 shows a strong link between $d_\epsilon^\pm$. This symmetry property can be used to carry regularity properties from d_ϵ^+ to d_ϵ^-, and vice versa.

Lemma 23. *Let $d_\epsilon^\pm$ as in Eq. 3. For all $\mu\in[0,1]$ and all $\lambda\in\mathbb{R}$,*

$$d_\epsilon^+(1-\lambda,1-\mu)=d_\epsilon^-(\lambda,\mu)\quad\text{and}\quad d_\epsilon^-(1-\lambda,1-\mu)=d_\epsilon^+(\lambda,\mu).$$

Proof. By definitions and change of variable $\tilde{z}=1-z$ and $\mathrm{kl}(1-\tilde{z},1-\mu)=\mathrm{kl}(\tilde{z},\mu)$, we obtain

$$\begin{aligned}d_\epsilon^+(1-\lambda,1-\mu)&=\mathbb{1}\left(\mu<[\lambda]_0^1\right)\inf_{z\in[1-[\lambda]_0^1,1-\mu]}\{\mathrm{kl}(z,1-\mu)+\epsilon(\max\{0,\min\{1,\lambda\}-(1-z))\}\\&=\mathbb{1}\left(\mu<[\lambda]_0^1\right)\inf_{\tilde{z}\in[\mu,[\lambda]_0^1]}\{\mathrm{kl}(1-\tilde{z},1-\mu)+\epsilon(\max\{0,\min\{1,\lambda\}-\tilde{z})\}\\&=\mathbb{1}\left(\mu<[\lambda]_0^1\right)\inf_{\tilde{z}\in[\mu,[\lambda]_0^1]}\{\mathrm{kl}(\tilde{z},\mu)+\epsilon(\max\{0,\min\{1,\lambda\}-\tilde{z})\}=d_\epsilon^-(\lambda,\mu).\end{aligned}$$

The second equality is a consequence of the first. □

Lemma 24 gathers regularity properties on the functions $g_\epsilon^\pm$ that appear in the explicit solutions of $d_\epsilon^\pm$, as shown below. Intuitively, those functionals govern locally the separation between the low privacy regime where $d_\epsilon^\pm$ is equals to the kl and the high privacy regime where the divergence has to be modified to account for the privacy budget ϵ.

Lemma 24. *Let $\epsilon>0$. Let $g_\epsilon^\pm$ defined as*

$$\forall x\in[0,1],\quad g_\epsilon^+(x):=\frac{x}{x(1-e^\epsilon)+e^\epsilon}\quad\text{and}\quad g_\epsilon^-(x):=\frac{xe^\epsilon}{x(e^\epsilon-1)+1}.\tag{30}$$

On $[0,1]$, the function g_ϵ^+ is twice continuously differentiable, increasing and strictly convex. It satisfies $g_\epsilon^+(0)=0$, $g_\epsilon^+(1)=1$ and $g_\epsilon^+(x)<x$ on $(0,1)$. On $[0,1]$, the function g_ϵ^- is twice continuously differentiable, increasing and strictly concave. It satisfies $g_\epsilon^-(0)=0$, $g_\epsilon^-(1)=1$ and $g_\epsilon^-(x)>x$ on $(0,1)$. For all $x\in[0,1]$, we have $g_\epsilon^+(g_\epsilon^-(x))=x$ and $g_\epsilon^-(1-x)+g_\epsilon^+(x)=1$. For all $x\in[0,1]$, we have $\lim_{\epsilon\to0}g_\epsilon^+(x)=\lim_{\epsilon\to0}g_\epsilon^-(x)=x$; it satisfies $\lim_{\epsilon\to+\infty}g_\epsilon^-(x)=1$ if $x\neq0$ and $\lim_{\epsilon\to+\infty}g_\epsilon^+(x)=0$ if $x\neq1$.

Proof. Using that $e^\epsilon > 1$, direct computations yield that, for all $x \in [0,1]$,

$$(g_\epsilon^+)'(x) = \frac{e^\epsilon}{(x(1-e^\epsilon)+e^\epsilon)^2} > 0 \quad \text{and} \quad (g_\epsilon^+)''(x) = -2\frac{e^\epsilon(1-e^\epsilon)}{(x(1-e^\epsilon)+e^\epsilon)^2} > 0\,,$$
$$(g_\epsilon^-)'(x) = \frac{e^\epsilon}{(x(e^\epsilon-1)+1)^2} > 0 \quad \text{and} \quad (g_\epsilon^-)''(x) = -2\frac{e^\epsilon(e^\epsilon-1)}{(x(e^\epsilon-1)+1)^3} < 0\,.$$

Therefore, g_ϵ^+ is twice continuously differentiable, increasing and strictly convex on $[0,1]$ and g_ϵ^- is twice continuously differentiable, increasing and strictly concave on $[0,1]$. It is direct to see that $g_\epsilon^+(0) = g_\epsilon^-(0) = 0$ and $g_\epsilon^+(1) = g_\epsilon^-(1) = 1$. Since they are strictly convex and strictly concave, we obtain $g_\epsilon^+(x) < x$ and $g_\epsilon^-(x) > x$ for all $x \in (0,1)$. It is direct to see that, for all $x \in [0,1]$, we have $g_\epsilon^+(g_\epsilon^-(x)) = x$ and $1 - g_\epsilon^+(x) = g_\epsilon^-(1-x)$. It is direct to see that, $\lim_{\epsilon\to 0} g_\epsilon^+(x) = \lim_{\epsilon\to 0} g_\epsilon^-(x) = x$ for all $x \in [0,1]$, and $\lim_{\epsilon\to+\infty} g_\epsilon^+(x) = 0$ if $x \neq 1$ and $\lim_{\epsilon\to+\infty} g_\epsilon^-(x) = 1$ if $x \neq 0$. □

Lemma 25 gathers regularity properties of d_ϵ^+. In particular, it gives a closed-form solution, which is a key property used in our implementation to reduce the computational cost.

Lemma 25. *Let d_ϵ^+ as in Eq.* (3)*, and $g_\epsilon^\pm$ as in Eq.* (30)*. For all $\mu \in [0,1]$ and $\lambda \in \mathbb{R}$, we have*

$$d_\epsilon^+(\lambda,\mu) = \begin{cases} 0 & \text{if } \mu \in [0, [\lambda]_0^1] \\ -\log\left(1-\mu(1-e^{-\epsilon})\right) - \epsilon[\lambda]_0^1 & \text{if } \mu \in (g_\epsilon^-([\lambda]_0^1), 1] \\ \mathrm{kl}\left(\lambda,\mu\right) & \text{if } \lambda \in (0,1) \ \wedge \ \mu \in ([\lambda]_0^1, g_\epsilon^-([\lambda]_0^1)] \end{cases}.$$

The function $(\lambda,\mu) \mapsto d_\epsilon^+(\lambda,\mu)$ is jointly continuous on $\mathbb{R}\times[0,1]$. For all $\mu \in [0,1]$, the function $\lambda \mapsto d_\epsilon^+(\lambda,\mu)$ is constant on $(-\infty,0]$ and on $[1,+\infty)$. Then,

$$\forall \lambda \in (0,1), \forall \mu \in [0,1], \quad d_\epsilon^+(\lambda,\mu) = \begin{cases} 0 & \text{if } \mu \in [0,\lambda] \\ \mathrm{kl}\left(\lambda,\mu\right) & \mu \in (\lambda, g_\epsilon^-(\lambda)] \\ -\log\left(1-\mu(1-e^{-\epsilon})\right) - \epsilon\lambda & \text{if } \mu \in (g_\epsilon^-(\lambda), 1] \end{cases}.$$

For all $\mu \in [0,1]$, the function $\lambda \mapsto d_\epsilon^+(\lambda,\mu)$ is continuously differentiable, positive, decreasing and convex on $(0,\mu)$; it is affine with negative slope $-\epsilon$ on $(0, g_\epsilon^+(\mu))$ and twice continuously differentiable and strictly convex on $(g_\epsilon^+(\mu),\mu)$.

For all $\lambda \in (0,1)$, the function $\mu \mapsto d_\epsilon^+(\lambda,\mu)$ is positive, three times differentiable with continuous first derivative, increasing and strictly convex on $(\lambda,1]$; its second derivative is discontinuous at $g_\epsilon^-(\lambda)$ with gap $\frac{\partial^2 d_\epsilon^+}{\partial\mu^2}(\lambda, g_\epsilon^-(\lambda)) - \lim_{\mu\to g_\epsilon^-(\lambda)^+} \frac{\partial^2 d_\epsilon^+}{\partial\mu^2}(\lambda,\mu) > 0$. Moreover, we have

$$\forall \mu \in (\lambda,1], \quad \frac{\partial d_\epsilon^+}{\partial\mu}(\lambda,\mu) = \begin{cases} \frac{1-e^{-\epsilon}}{1-\mu(1-e^{-\epsilon})} & \text{if } \mu \in (g_\epsilon^-(\lambda), 1] \\ \frac{\mu-\lambda}{\mu(1-\mu)} & \text{if } \mu \in (\lambda, g_\epsilon^-(\lambda)] \end{cases}.$$

The function d_ϵ^+ is jointly convex on $(0,1)\times[0,1]$.

Proof. Recall that $d_\epsilon^+(\lambda,\mu) = \mathbb{1}\left(\mu > [\lambda]_0^1\right) \inf_{z\in[[\lambda]_0^1,\mu]} f_\epsilon^+([\lambda]_0^1,\mu,z)$ where $f_\epsilon^+(\lambda,\mu,z) = \mathrm{kl}(z,\mu) + \epsilon(z-\lambda)$. Direct computations yield that, for all $z \in ([\lambda]_0^1,\mu)$,

$$\frac{\partial f_\epsilon^+}{\partial z}(\lambda,\mu,z) = \log\left(\frac{z(1-\mu)}{(1-z)\mu}\right) + \epsilon \quad \text{and} \quad \frac{\partial f_\epsilon^+}{\partial z}(\lambda,\mu,z) = 0 \iff z = g_\epsilon^+(\mu)\,,$$
$$\frac{\partial^2 f_\epsilon^+}{\partial z^2}(\lambda,\mu,z) = \frac{1}{z(1-z)} > 0\,.$$

Therefore, $z \to f_\epsilon^+(\lambda,\mu,z)$ is twice continuously differentiable, positive and strictly convex on $([\lambda]_0^1,\mu)$. Moreover, $z \to f_\epsilon^+(\lambda,\mu,z)$ is decreasing on $([\lambda]_0^1, \max\{g_\epsilon^+(\mu),\lambda\})$ and increasing on $(\max\{g_\epsilon^+(\mu),\lambda\},\mu)$. Using Lemma 24, we obtain

$$f_\epsilon^+(\lambda,\mu,\lambda) = \mathrm{kl}\left(\lambda,\mu\right)\,,$$
$$\mathrm{kl}(g_\epsilon^+(\mu),\mu) = -(g_\epsilon^+(\mu) + g_\epsilon^-(1-\mu))\log\left(\mu(1-e^\epsilon)+e^\epsilon\right) + \epsilon g_\epsilon^-(1-\mu)$$

$$= -\log\left(1-\mu(1-e^{-\epsilon})\right) - \epsilon g_\epsilon^+(\mu)\,,$$
$$f_\epsilon^+(\lambda,\mu,g_\epsilon^+(\mu)) = -\log\left(1-\mu(1-e^{-\epsilon})\right) - \epsilon\lambda\,.$$

By definition of the indicator function, we have $d_\epsilon^+(\lambda,\mu)=0$ if $\mu\in[0,[\lambda]_0^1]$. When $\lambda\le 0$, for all $\mu\in(0,1)$, we have

$$\forall \mu\in(0,1),\quad d_\epsilon^+(\lambda,\mu) = f_\epsilon^+([\lambda]_0^1,\mu,g_\epsilon^+(\mu)) = -\log\left(1-\mu(1-e^{-\epsilon})\right) - \epsilon[\lambda]_0^1\,,$$

by using the properties of $z\to f_\epsilon^+(\lambda,\mu,z)$ on $(0,1)=(g_\epsilon^-([\lambda]_0^1),1)$ by Lemma 24. This function can be extended by continuity to $\mu=0=g_\epsilon^-([\lambda]_0^1)$ with value $d_\epsilon^+(\lambda,0)=0$. When $\lambda\in(0,1)$ and $\mu\in(g_\epsilon^-(\lambda),1)$, we have

$$\forall \mu\in(0,1),\quad d_\epsilon^+(\lambda,\mu) = f_\epsilon^+([\lambda]_0^1,\mu,g_\epsilon^+(\mu)) = -\log\left(1-\mu(1-e^{-\epsilon})\right) - \epsilon[\lambda]_0^1\,,$$

by using the properties of $z\to f_\epsilon^+(\lambda,\mu,z)$ on $(g_\epsilon^-(\lambda),1)=(g_\epsilon^-([\lambda]_0^1),1)$ by Lemma 24. This function can be extended by continuity to $(\lambda,\mu)=(0,0)=\lim_{\lambda\to 0^+}(\lambda,g_\epsilon^-([\lambda]_0^1))$ with value $d_\epsilon^+(0,0)=0$. In both cases, this function can be extended by continuity to $\mu=1$ with value $d_\epsilon^+(\lambda,1)=\epsilon(1-[\lambda]_0^1)$.

When $\lambda\in(0,1)$, i.e., $[\lambda]_0^1=\lambda$, and $\mu\in(\lambda,g_\epsilon^-(\lambda))\subseteq(0,1)$ by Lemma 24, we have

$$d_\epsilon^+(\lambda,\mu) = f_\epsilon^+(\lambda,\mu,\lambda) = \mathrm{kl}\,(\lambda,\mu)\ .$$

This function can be extended by continuity to $\mu=\lambda$ with value $d_\epsilon^+(\lambda,\lambda)=0$ since $\mathrm{kl}\,(\lambda,\lambda)=0$. Using Lemma 24, this function can be extended by continuity to $\mu=g_\epsilon^-(\lambda)$ (i.e., $\lambda=g_\epsilon^+(\mu)$) with value

$$d_\epsilon^+(\lambda,g_\epsilon^-(\lambda)) = \mathrm{kl}\left(\lambda,g_\epsilon^-(\lambda)\right) = \mathrm{kl}\left(g_\epsilon^+(\mu),\mu\right) = -\log\left(1-\mu(1-e^{-\epsilon})\right) - \epsilon g_\epsilon^+(\mu)\,.$$

Therefore, we have

$$\forall\lambda\in(0,1),\ \forall\mu\in[\lambda,g_\epsilon^-(\lambda)],\quad d_\epsilon^+(\lambda,\mu)=\mathrm{kl}\,(\lambda,\mu)\ .$$

Using that $\lim_{\lambda\to 0^+}[\lambda,g_\epsilon^-(\lambda)]=\{0\}$, this function can be extended by continuity to $\lambda=0$ with value 0. Using that $\lim_{\lambda\to 1^-}[\lambda,g_\epsilon^-(\lambda)]=\{1\}$, this function can be extended by continuity to $\lambda=1$ with value $0=\lim_{(\mu,\lambda)\to 1^-}-\log\left(1-\mu(1-e^{-\epsilon})\right)-\epsilon[\lambda]_0^1$.

Putting all the continuity arguments together, we have shown that $(\lambda,\mu)\to d_\epsilon^+(\lambda,\mu)$ is jointly continuous on $\mathbb{R}\times[0,1]$. Moreover, it is direct to see that, for all $\mu\in[0,1]$, the function $\lambda\to d_\epsilon^+(\lambda,\mu)$ is constant on $(-\infty,0]$ and on $[1,+\infty)$. Then,

$$\forall\lambda\in(0,1),\forall\mu\in[0,1],\quad d_\epsilon^+(\lambda,\mu)=\begin{cases}0 & \text{if } \mu\in[0,\lambda]\\ \mathrm{kl}\,(\lambda,\mu) & \mu\in(\lambda,g_\epsilon^-(\lambda)]\\ -\log\left(1-\mu(1-e^{-\epsilon})\right)-\epsilon\lambda & \text{if } \mu\in(g_\epsilon^-(\lambda),1]\end{cases}\,.$$

Let $\mu\in[0,1]$ and $\lambda\in(0,\mu)$. Using that $\mu\in(g_\epsilon^-(\lambda),1]$ if and only if $\lambda\in(0,g_\epsilon^+(\mu))$. For all $\mu\in[0,1]$, the function $\lambda\to d_\epsilon^+(\lambda,\mu)$ is positive and affine with negative slope $-\epsilon$ on $(0,g_\epsilon^+(\mu))$. Let $\lambda\in(g_\epsilon^+(\mu),\mu)$. Direct computation yields that

$$\frac{\partial d_\epsilon^+}{\partial\lambda}(\lambda,\mu)=\frac{\partial\mathrm{kl}}{\partial\lambda}(\lambda,\mu)=\log\left(\frac{\lambda(1-\mu)}{(1-\lambda)\mu}\right)<0\,,$$
$$\lim_{\lambda\to g_\epsilon^+(\mu)^+}\frac{\partial d_\epsilon^+}{\partial\lambda}(\lambda,\mu)=-\epsilon=\lim_{\lambda\to g_\epsilon^+(\mu)^-}\frac{\partial d_\epsilon^+}{\partial\lambda}(\lambda,\mu)\,,$$
$$\frac{\partial^2 d_\epsilon^+}{\partial\lambda^2}(\lambda,\mu)=\frac{\partial^2\mathrm{kl}}{\partial\lambda^2}(\lambda,\mu)=\frac{1}{\lambda(1-\lambda)}>0\,.$$

For all $\mu\in[0,1]$, the function $\lambda\to d_\epsilon^+(\lambda,\mu)$ is continuously differentiable, positive, decreasing and convex on $(0,\mu)$. For all $\mu\in[0,1]$, the function $\lambda\to d_\epsilon^+(\lambda,\mu)$ is twice continuously differentiable, positive and strictly convex on $(g_\epsilon^+(\mu),\mu)$. Combining the above results concludes the part of $\lambda\mapsto d_\epsilon^+(\lambda,\mu)$ on $(0,\mu)$.

Let $\lambda\in(0,1)$. Let $a>0$ and $k\in\mathbb{N}$. The k-th derivative of $u(x)=a(1-ax)^{-1}$ on $[0,1]$ is $u^{(k)}(x)=(k-1)!a^{k+1}(1-ax)^{-(k+1)}$. Then,

$$\forall\mu\in(g_\epsilon^-(\lambda),1],\ \forall k\in\mathbb{N},\quad \frac{\partial^k d_\epsilon^+}{\partial\mu^k}(\lambda,\mu)=\frac{(1-e^{-\epsilon})^k(k-1)!}{(1-\mu(1-e^{-\epsilon}))^k}>0\,,$$

$$\forall \mu \in (\lambda, g_\epsilon^-(\lambda)], \quad \frac{\partial d_\epsilon^+}{\partial \mu}(\lambda,\mu) = \frac{\mu-\lambda}{\mu(1-\mu)} > 0\,,$$
$$\frac{\partial^2 d_\epsilon^+}{\partial \mu^2}(\lambda,\mu) = \frac{(\mu-\lambda)^2+\lambda(1-\lambda)}{\mu^2(1-\mu)^2} > 0\,,$$
$$\frac{\partial^3 d_\epsilon^+}{\partial \mu^3}(\lambda,\mu) > 0\,.$$

Direct computation yields

$$\lim_{\mu \to g_\epsilon^-(\lambda)} \frac{\mu-\lambda}{\mu(1-\mu)} = (1-e^{-\epsilon})(1+\lambda(e^\epsilon-1))\,,$$
$$\lim_{\mu \to g_\epsilon^-(\lambda)} \frac{1-e^{-\epsilon}}{1-\mu(1-e^{-\epsilon})} = (1-e^{-\epsilon})(1+\lambda(e^\epsilon-1))\,,$$
$$\lim_{\mu \to g_\epsilon^-(\lambda)} \left\{ \frac{(\mu-\lambda)^2+\lambda(1-\lambda)}{\mu^2(1-\mu)^2} - \frac{(1-e^{-\epsilon})^2}{(1-\mu(1-e^{-\epsilon}))^2} \right\} = \frac{\lambda(1-\lambda)}{g_\epsilon^-(\lambda)^2(1-g_\epsilon^-(\lambda))^2} > 0\,.$$

For all $\lambda \in (0,1)$, the function $\mu \to d_\epsilon^+(\lambda,\mu)$ is positive, three times differentiable with continuous first derivative and increasing on $(\lambda,1]$. For all $\lambda \in (0,1)$, the function $\mu \to d_\epsilon^+(\lambda,\mu)$ is strictly convex on $(\lambda, g_\epsilon^-(\lambda)]$ and $(g_\epsilon^-(\lambda),1]$. The second derivative is discontinuous at $g_\epsilon^-(\lambda)$ with gap $\frac{\partial^2 d_\epsilon^+}{\partial \mu^2}(\lambda, g_\epsilon^-(\lambda)) - \lim_{\mu \to g_\epsilon^-(\lambda)^+} \frac{\partial^2 d_\epsilon^+}{\partial \mu^2}(\lambda,\mu) > 0$. Thanks to the continuity of the first derivative and the sign of the second derivative, the function $\mu \to d_\epsilon^+(\lambda,\mu)$ is strict convexity on $(\lambda,1]$.

Let $(\mu_1,\mu_2) \in [0,1]^2$ and $(\lambda_1,\lambda_2) \in (0,1)^2$. On the convex set $\mathcal{F}_0 = \{(\lambda,\mu) \in (0,1)\times[0,1] \mid \mu \in [0,\lambda]\}$, the function d_ϵ^- is null hence jointly convex. Let $((\mu_1,\lambda_1),(\mu_2,\lambda_2)) \in (((0,1)\times[0,1])\setminus\mathcal{F}_0)^2$. Let $(z_1,z_2) \in [\lambda_1,\mu_1]\times[\lambda_2,\mu_2]$ be the minimizers realizing $d_\epsilon^+(\lambda_1,\mu_1)$ and $d_\epsilon^+(\lambda_2,\mu_2)$. Since it is a convex set, we have $(\alpha\lambda_1+(1-\alpha)\lambda_2, \alpha\mu_1+(1-\alpha)\mu_2) \in ((0,1)\times[0,1])\setminus\mathcal{F}_0$ for all $\alpha \in [0,1]$. Moreover, we have $\alpha z_1+(1-\alpha)z_2 \in [\alpha\lambda_1+(1-\alpha)\lambda_2, \alpha\mu_1+(1-\alpha)\mu_2]$ for all $\alpha \in [0,1]$. Using the definition of d_ϵ^+ as an infimum, we obtain

$$\begin{aligned}
&d_\epsilon^+(\alpha\lambda_1+(1-\alpha)\lambda_2, \alpha\mu_1+(1-\alpha)\mu_2)\\
&\le \mathrm{kl}(\alpha z_1+(1-\alpha)z_2, \alpha\mu_1+(1-\alpha)\mu_2) + \epsilon(\alpha z_1+(1-\alpha)z_2 - (\alpha\lambda_1+(1-\alpha)\lambda_2))\\
&\le \alpha\left(\mathrm{kl}(z_1,\mu_1)+\epsilon(z_1-\lambda_1)\right) + (1-\alpha)\left(\mathrm{kl}(z_2,\mu_2)+\epsilon(z_2-\lambda_2)\right)\\
&= \alpha d_\epsilon^+(\lambda_1,\mu_1) + (1-\alpha) d_\epsilon^+(\lambda_2,\mu_2)
\end{aligned}$$

where the second inequality comes from the joint convexity of the Kullback-Leibler divergence. Combining both results, we have shown that the function d_ϵ^+ is jointly convex on $(0,1)\times[0,1]$. □

Lemma 26 gather regularity properties of d_ϵ^-. In particular, it gives a closed-form solution, which is a key property used in our implementation to reduce the computational cost.

Lemma 26. *Let d_ϵ^- as in Eq.* (3)*, and $g_\epsilon^\pm$ as in Eq.* (30)*. For all $\mu \in [0,1]$ and all $\lambda \in \mathbb{R}$, we have*

$$d_\epsilon^-(\lambda,\mu) = \begin{cases} 0 & \text{if } \mu \in [[\lambda]_0^1, 1] \\ -\log\left(1+\mu(e^\epsilon-1)\right) + \epsilon[\lambda]_0^1 & \text{if } \mu \in [0, g_\epsilon^+([\lambda]_0^1)) \\ \mathrm{kl}(\lambda,\mu) & \text{if } \lambda \in (0,1) \text{ and } \mu \in [g_\epsilon^+([\lambda]_0^1), [\lambda]_0^1) \end{cases}.$$

The function $(\lambda,\mu) \mapsto d_\epsilon^-(\lambda,\mu)$ is jointly continuous on $\mathbb{R}\times[0,1]$. For all $\mu \in [0,1]$, the function $\lambda \mapsto d_\epsilon^-(\lambda,\mu)$ is constant on $(-\infty,0]$ and on $[1,+\infty)$. Then,

$$\forall \lambda \in (0,1), \forall \mu \in [0,1], \quad d_\epsilon^-(\lambda,\mu) = \begin{cases} 0 & \text{if } \mu \in [\lambda,1] \\ \mathrm{kl}\left(\lambda,\mu\right) & \text{if } \mu \in [g_\epsilon^+(\lambda),\lambda) \\ -\log\left(1+\mu(e^\epsilon-1)\right)+\epsilon\lambda & \text{if } \mu \in [0, g_\epsilon^+(\lambda)) \end{cases}.$$

For all $\mu \in [0,1]$, the function $\lambda \mapsto d_\epsilon^-(\lambda,\mu)$ is continuously differentiable, positive, increasing and convex on $(\mu,1)$; it is affine with positive slope ϵ on $(g_\epsilon^-(\mu),1)$ and twice continuously differentiable and strictly convex on $(\mu, g_\epsilon^-(\mu))$.

For all $\lambda \in (0,1)$, the function $\mu \mapsto d_\epsilon^-(\lambda,\mu)$ is positive, three times differentiable with continuous first derivative, decreasing and strictly convex on $[0,\lambda)$; its second derivative is discontinuous at $g_\epsilon^+(\lambda)$ with gap $\lim_{\mu\to g_\epsilon^+(\lambda)^-} \frac{\partial^2 d_\epsilon^-}{\partial \mu^2}(\lambda,\mu) - \frac{\partial^2 d_\epsilon^-}{\partial \mu^2}(\lambda, g_\epsilon^+(\lambda)) < 0$. Moreover, we have

$$\forall \mu \in [0,\lambda), \quad \frac{\partial d_\epsilon^-}{\partial \mu}(\lambda,\mu) = \begin{cases} -\frac{e^\epsilon - 1}{1+\mu(e^\epsilon-1)} & \text{if } \mu \in [0, g_\epsilon^+(\lambda)) \\ -\frac{\lambda-\mu}{\mu(1-\mu)} & \text{if } \mu \in [g_\epsilon^+(\lambda), \lambda) \end{cases} .$$

The function d_ϵ^- is jointly convex on $(0,1)\times[0,1]$.

Proof. Using Lemmas 23 and 24, we have

$$\begin{aligned} &d_\epsilon^-(\lambda,\mu) = d_\epsilon^+(1-\lambda,1-\mu) \quad \text{and} \quad g_\epsilon^+(\lambda) = 1 - g_\epsilon^-(1-\lambda)\,, \\ &\frac{\partial d_\epsilon^-}{\partial \mu}(\lambda,\mu) = -\frac{\partial d_\epsilon^+}{\partial \mu}(1-\lambda,1-\mu) \quad \text{and} \quad \frac{\partial^2 d_\epsilon^-}{\partial \mu^2}(\lambda,\mu) = \frac{\partial^2 d_\epsilon^+}{\partial \mu^2}(1-\lambda,1-\mu)\,. \end{aligned}$$

Moreover, we have $\mathrm{kl}(\lambda,\mu) = \mathrm{kl}(1-\lambda,1-\mu)$ and

$$-\log\left(1+\mu(e^\epsilon-1)\right) + \epsilon[\lambda]_0^1 = -\log\left(1-(1-\mu)(1-e^{-\epsilon})\right) - \epsilon[1-\lambda]_0^1\,.$$

Combining the above with properties of d_ϵ^+ in Lemma 25 concludes the proof. □

G.1.1 Modified Divergence

Let us define

$$\forall x > 0, \quad h(x) := \sqrt{1+x^2} - 1 + \log\left(\frac{2}{x^2}\left(\sqrt{1+x^2}-1\right)\right)\,. \tag{31}$$

For all $(\lambda,\mu,r) \in \mathbb{R}\times(0,1)\times\mathbb{R}_+^\star$, we define

$$\begin{aligned} \widetilde{d}_\epsilon^-(\lambda,\mu,r) &:= \mathbb{1}\left(\mu < [\lambda]_0^1\right) \inf_{z\in(\mu,[\lambda]_0^1)} \left\{ \mathrm{kl}(z,\mu) + \frac{1}{r} h(r\epsilon(\lambda - z)) \right\}\,, \\ \widetilde{d}_\epsilon^+(\lambda,\mu,r) &:= \mathbb{1}\left(\mu > [\lambda]_0^1\right) \inf_{z\in([\lambda]_0^1,\mu)} \left\{ \mathrm{kl}(z,\mu) + \frac{1}{r} h(r\epsilon(z - \lambda)) \right\}\,. \end{aligned} \tag{32}$$

Lemma 27 shows a strong link between $\widetilde{d}_\epsilon^\pm$. This symmetry property can be used to carry regularity properties from $\widetilde{d}_\epsilon^+$ to $\widetilde{d}_\epsilon^-$, and vice versa.

Lemma 27. *Let $\widetilde{d}_\epsilon^\pm$ as in Eq. (32). For all $(\lambda,\mu)\in\mathbb{R}\times[0,1]$, we have*

$$\widetilde{d}_\epsilon^+(1-\lambda,1-\mu,r) = \widetilde{d}_\epsilon^-(\lambda,\mu,r) \quad \textit{and} \quad \widetilde{d}_\epsilon^-(1-\lambda,1-\mu,r) = \widetilde{d}_\epsilon^+(\lambda,\mu,r)\,.$$

Proof. Using the definitions, the change of variable $\tilde{z} = 1-z$ and $\mathrm{kl}(1-\tilde{z},1-\mu) = \mathrm{kl}(\tilde{z},\mu)$, we obtain

$$\begin{aligned} \widetilde{d}_\epsilon^+(1-\lambda,1-\mu,r) &= \mathbb{1}\left(\mu < [\lambda]_0^1\right) \inf_{z\in[1-[\lambda]_0^1,1-\mu]} \left\{ \mathrm{kl}(z,1-\mu) + \frac{1}{r} h\left(r\epsilon(\lambda-(1-z))\right) \right\} \\ &= \mathbb{1}\left(\mu < [\lambda]_0^1\right) \inf_{\tilde{z}\in[\mu,[\lambda]_0^1]} \left\{ \mathrm{kl}(1-\tilde{z},1-\mu) + \frac{1}{r} h\left(r\epsilon(\lambda-\tilde{z})\right) \right\} \\ &= \mathbb{1}\left(\mu < [\lambda]_0^1\right) \inf_{\tilde{z}\in[\mu,[\lambda]_0^1]} \left\{ \mathrm{kl}(\tilde{z},\mu) + \frac{1}{r} h\left(r\epsilon(\lambda-\tilde{z})\right) \right\} = \widetilde{d}_\epsilon^-(\lambda,\mu,r)\,. \end{aligned}$$

The second equality is a consequence of the first. □

Lemma 28 gathers regularity properties of the function h defined in Eq. (31).

Lemma 28. *Let h as in Eq.* (31)*. Then,*

$$\forall x > 0, \quad h'(x) = \frac{x}{\sqrt{x^2+1}+1} > 0 \quad \textit{and} \quad h''(x) = \frac{1}{1+x^2+\sqrt{1+x^2}} > 0\,.$$

On $\mathbb{R}_+^\star$, the function h is twice continuously differentiable, increasing and strictly convex. Moreover, it satisfies

$$h(x) =_{x\to 0} x^2/4 + \mathcal{O}(x^4) \quad \textit{and} \quad h(x) =_{x\to+\infty} x - \mathcal{O}(\log(x))\,.$$

Proof. For all $x > 0$, $h_1(x) = x + \log(x)$, $h_2(x) = \sqrt{1+x^2} - 1$ and $h_3(x) = \sqrt{1+x^2} - x$. Then

$$h_1'(x) = 1 + \frac{1}{x} \quad , \quad h_2'(x) = \frac{x}{\sqrt{1+x^2}} \quad \text{and} \quad h_3'(x) = \frac{x}{\sqrt{1+x^2}} - 1\,,$$

Then, we have

$$\forall x > 0, \quad h(x) = h_1(h_2(x)) - 2\log(x) + \log 2\,.$$

Therefore, we have

$$\begin{aligned} h'(x) = h_2'(x)h_1'(h_2(x)) - \frac{2}{x} &= \frac{x}{\sqrt{1+x^2}}\left(1 + \frac{1}{\sqrt{1+x^2}-1}\right) - \frac{2}{x} \\ &= \frac{x}{\sqrt{1+x^2}-1} - \frac{2}{x} = \sqrt{1+\frac{1}{x^2}} - \frac{1}{x} = h_3(1/x)\,. \end{aligned}$$

Note that

$$\sqrt{1+\frac{1}{x^2}} - \frac{1}{x} = \frac{x}{\sqrt{x^2+1}+1}\,.$$

Moreover, we have w

$$h''(x) = -\frac{1}{x^2}h_3'(1/x) = -\frac{1}{x^2}\left(\frac{1/x}{\sqrt{1+(1/x)^2}} - 1\right) = \frac{1}{1+x^2+\sqrt{1+x^2}}\,.$$

By taking the limit, we have $\lim_{x\to 0^+} h(x) = 0$. Moreover, we see that $\lim_{x\to 0^+} h'(x) = 0$ and $\lim_{x\to 0^+} h''(x) = 1/2$. Therefore, one can conclude that $h(x) =_{x\to 0} x^2/4 + \mathcal{O}(x^4)$ by Taylor expansion. The second result is obtained directly by limit. □

Lemma 29 provides upper and lower bound on the function $r \mapsto h(rx)/r$ involved in the definition of $\widetilde{d}_\epsilon^{\pm}$.

Lemma 29. *Let h as in Eq.* (31)*. Let $\kappa(r,x) = h(rx)/r - x$ for all $r > 0$ and all $x \in \mathbb{R}_+^\star$. Then, we have*

$$\forall r > 0, \quad \frac{\partial \kappa}{\partial r}(r,x) = \frac{rxh'(rx) - h(rx)}{r^2} = \log\left(\frac{1}{2}(\sqrt{1+(rx)^2}+1)\right) > 0\,.$$

On $\mathbb{R}_+^\star$, the function $r \mapsto \kappa(r,x)$ is increasing. Moreover, we have

$$\forall r > 0, \forall x \in \mathbb{R}_+, \quad 0 \le r\kappa(r,x) + \log(1+2xr) + 1 \le 1 + \log 4\,,$$

Proof. Using Lemma 28 and the definition in Eq. (31), we obtain that

$$\forall x > 0, \quad xh'(x) - h(x) = -\log\left(\frac{2}{x^2}\left(\sqrt{1+x^2}-1\right)\right) = \log\left(\frac{1}{2}(\sqrt{1+x^2}+1)\right) > 0\,,$$

where we used that $\sqrt{1+x^2}+1 > 2$ for the last inequality. Let us define

$$\forall x \in \mathbb{R}_+, \quad g_1(x) = \frac{2(1+2x)}{\sqrt{1+x^2}+1}\,.$$

Then, we obtain $g_1(0) = 1$, $\lim_{x\to+\infty} g_1(x) = 4$ and

$$g_1'(x) = 2\frac{2+2\sqrt{1+x^2}-x}{\sqrt{1+x^2}(\sqrt{1+x^2}+1)^2} > 2\frac{2+x}{\sqrt{1+x^2}(\sqrt{1+x^2}+1)^2} > 0\,.$$

Since g_1 is strictly increasing on $\mathbb{R}_+^\star$, we obtain $\log g_1(x) \geq \log g_1(0) = 0$ and $\log g_1(x) \leq \log 4$ for all $x \in \mathbb{R}_+$.

By definition, we obtain

$$\begin{aligned} r\kappa(r,x) + \log(1+2xr) + 1 &= h(rx) - rx + 1 + \log(1+2xr) \\ &= \sqrt{1+(rx)^2} - rx + \log\left(\frac{2(1+2xr)}{\sqrt{1+r^2x^2}+1}\right) . \end{aligned}$$

Using that $0 \leq \sqrt{1+x^2} - x \leq 1$ on $\mathbb{R}_+$, we obtain

$$\begin{aligned} &r\kappa(r,x) + \log(1+2xr) + 1 \geq \log\left(\frac{2(1+2xr)}{\sqrt{1+r^2x^2}+1}\right) = \log(g_1(rx)) \geq 0 , \\ &r\kappa(r,x) + \log(1+2xr) + 1 \leq 1 + \log(g_1(rx)) \leq 1 + \log 4 . \end{aligned}$$

This concludes the proof. □

Lemma 30 provides lower and upper bounds on the gap between $\widetilde{d}_\epsilon^{\pm}$ and $d_\epsilon^{\pm}$.

Lemma 30. *Let $d_\epsilon^{\pm}$ and $\widetilde{d}_\epsilon^{\pm}$ as in Eq.* (3) *and* (32)*. For all $(\lambda,\mu,r) \in \mathbb{R} \times (0,1) \times \mathbb{R}_+^\star$ such that $[\lambda]_0^1 < \mu$. Then,*

$$d_\epsilon^+(\lambda,\mu) \leq \widetilde{d}_\epsilon^+(\lambda,\mu,r) + \frac{\log(1+2\epsilon r)+1}{r} .$$

For all $(\lambda,\mu,r) \in \mathbb{R} \times (0,1) \times \mathbb{R}_+^\star$ such that $[\lambda]_0^1 > \mu$. Then,

$$d_\epsilon^-(\lambda,\mu) \leq \widetilde{d}_\epsilon^-(\lambda,\mu,r) + \frac{\log(1+2\epsilon r)+1}{r} .$$

For all $(\lambda,\mu,r) \in [0,1] \times (0,1) \times \mathbb{R}_+^\star$ such that $\lambda < \mu$. Then,

$$d_\epsilon^+(\lambda,\mu) \geq \widetilde{d}_\epsilon^+(\lambda,\mu,r) - \frac{\log 4}{r} .$$

For all $(\mu,\lambda,r) \in [0,1] \times \mathbb{R}_+^\star$ such that $\lambda > \mu$. Then,

$$d_\epsilon^-(\lambda,\mu) \geq \widetilde{d}_\epsilon^-(\lambda,\mu,r) - \frac{\log 4}{r} .$$

Proof. Since $\mu \in (0,1)$, we have $[\lambda]_0^1 = \max\{0,\lambda\}$. Therefore, we have $z - \lambda \geq z - [\lambda]_0^1$ and $z - [\lambda]_0^1 \in (0, \mu - [\lambda]_0^1) \subset (0,1)$ for all $z \in ([\lambda]_0^1, \mu)$. Using Lemmas 28 and 29 and $\epsilon > 0$, we obtain, for all $r > 0$ and all $z \in ([\lambda]_0^1, \mu)$,

$$\begin{aligned} \epsilon(z - [\lambda]_0^1) &\leq \frac{1}{r} h(r\epsilon(z - [\lambda]_0^1)) + \frac{\log(1+2\epsilon(z-[\lambda]_0^1)r)+1}{r} \\ &\leq \frac{1}{r} h(r\epsilon(z-\lambda)) + \frac{\log(1+2\epsilon r)+1}{r} . \end{aligned}$$

Therefore, for all $z \in ([\lambda]_0^1, \mu)$, we obtain that

$$\mathrm{kl}(z,\mu) + \epsilon(z - [\lambda]_0^1) \leq \mathrm{kl}(z,\mu) + \frac{1}{r} h(r\epsilon(z-\lambda)) + \frac{\log(1+2\epsilon r)+1}{r} .$$

Taking the infimum over $z \in ([\lambda]_0^1, \mu)$ on both sides of both inequalities and using that

$$\begin{aligned} d_\epsilon^+(\lambda,\mu) &= \inf_{z \in [[\lambda]_0^1, \mu]} \left\{\mathrm{kl}(z,\mu) + \epsilon(z - [\lambda]_0^1)\right\} = \inf_{z \in ([\lambda]_0^1, \mu)} \left\{\mathrm{kl}(z,\mu) + \epsilon(z - [\lambda]_0^1)\right\} , \\ \widetilde{d}_\epsilon^+(\lambda,\mu,r) &= \inf_{z \in ([\lambda]_0^1, \mu)} \left\{\mathrm{kl}(z,\mu) + \frac{1}{r} h(r\epsilon(z-\lambda))\right\} , \end{aligned}$$

we obtain

$$d_\epsilon^+(\lambda,\mu) \leq \widetilde{d}_\epsilon^+(\lambda,\mu,r) + \frac{\log(1+2\epsilon r)+1}{r} .$$

This concludes the proof of the first result. Using Lemmas 23 and 27 yields the second result.

Suppose that $\lambda \in [0,1]$, hence $\lambda = [\lambda]_0^1$. Using Lemmas 28 and 29 and $\epsilon > 0$, we obtain, for all $r > 0$ and all $z \in ([\lambda]_0^1, \mu)$,

$$\frac{1}{r} h(r\epsilon(z-\lambda)) \leq \epsilon(z-\lambda) + \frac{\log 4 - \log(1 + 2\epsilon(z-\lambda)r)}{r} \leq \epsilon(z - [\lambda]_0^1) + \frac{\log 4}{r} \, .$$

Adding $\mathrm{kl}(z,\mu)$ on both sides and taking the infimum over $z \in ([\lambda]_0^1, \mu)$ on both sides of both inequalities yields the proof of third result. Using Lemmas 23 and 27 yields the forth result. □

Lemma 31 gathers regularity properties on the modified divergences $\widetilde{d}_\epsilon^+$. In particular, it gives a closed-form solution based on an implicit solution of a fixed-point equation. This is a key property used in our implementation to reduce the computational cost.

Lemma 31. *Let $\widetilde{d}_\epsilon^+$ as in Eq.* (32)*, and $g_\epsilon^\pm$ as in Eq.* (30)*. For all $\mu \in (0,1)$, $\lambda \in \mathbb{R}$ and $r > 0$, we have*

$$\widetilde{d}_\epsilon^+(\lambda,\mu,r) = \begin{cases} 0 & \text{if } \mu \in (0, [\lambda]_0^1] \\ \mathrm{kl}(x_\epsilon^+(\lambda,\mu,r) + g_\epsilon^+(\mu), \mu) + \frac{1}{r} h(r\epsilon(x_\epsilon^+(\lambda,\mu,r) + g_\epsilon^+(\mu) - \lambda)) & \text{if } \mu \in ([\lambda]_0^1, 1) \end{cases} ,$$

where $x_\epsilon^+(\lambda,\mu,r) \in (\max\{0, \lambda - g_\epsilon^+(\mu)\}, \mu - g_\epsilon^+(\mu))$ is the unique solution for $x \in (\max\{0, \lambda - g_\epsilon^+(\mu)\}, \mu - g_\epsilon^+(\mu))$ of the equation

$$\log\left(1 + \frac{x}{g_\epsilon^+(\mu)(1 - x - g_\epsilon^+(\mu))}\right) + \epsilon \left(\frac{r\epsilon(x + g_\epsilon^+(\mu) - \lambda)}{\sqrt{(r\epsilon(x + g_\epsilon^+(\mu) - \lambda))^2 + 1} + 1} - 1 \right) = 0 \, .$$

For all $(\mu, r) \in (0,1) \times \mathbb{R}_+^\star$, the function $\lambda \mapsto \widetilde{d}_\epsilon^+(\lambda,\mu,r)$ is positive, twice continuously differentiable, decreasing and strictly convex on $(-\infty, \mu)$; it satisfies $\lim_{\lambda \to \mu^-} \widetilde{d}_\epsilon^+(\lambda,\mu,r) = 0$ and $\lim_{\lambda \to -\infty} \widetilde{d}_\epsilon^+(\lambda,\mu,r) = +\infty$.

For all $(\lambda, r) \in \mathbb{R} \times \mathbb{R}_+^\star$, the function $\mu \mapsto \widetilde{d}_\epsilon^+(\lambda,\mu,r)$ is positive, twice continuously differentiable, increasing and strictly convex on $([\lambda]_0^1, 1)$. Moreover, we have

$$\forall \mu \in ([\lambda]_0^1, 1), \quad \frac{\partial \widetilde{d}_\epsilon^+}{\partial \mu}(\lambda,\mu,r) = \frac{\mu - g_\epsilon^+(\mu) - x_\epsilon^+(\lambda,\mu,r)}{\mu(1-\mu)} \, .$$

For all $(\lambda, \mu) \in \mathbb{R} \times (0,1)$ such that $\mu \in (0, [\lambda]_0^1]$, the function $r \mapsto \widetilde{d}_\epsilon^+(\lambda,\mu,r)$ is the zero function. For all $(\lambda, \mu) \in \mathbb{R} \times (0,1)$ such that $\mu \in ([\lambda]_0^1, 1)$, the function $r \mapsto \widetilde{d}_\epsilon^+(\lambda,\mu,r)$ is positive, continuously differentiable and increasing on $\mathbb{R}_+$.

Proof. By definition of the indicator function, we have $\widetilde{d}_\epsilon^+(\lambda,\mu,r) = 0$ if $\mu \in (0, [\lambda]_0^1]$. Let (λ, μ) such that $\mu \notin (0, [\lambda]_0^1]$, i.e., $([\lambda]_0^1, \mu)$ is non-empty. Since $\mu \in (0,1)$, this implies that $\lambda \in (-\infty, 1)$ necessarily, i.e., $[\lambda]_0^1 = \max\{0, \lambda\}$.

Recall that $\widetilde{d}_\epsilon^+(\lambda,\mu,r) = \mathbb{1}\left(\mu > [\lambda]_0^1\right) \inf_{z \in ([\lambda]_0^1, \mu)} \widetilde{f}_\epsilon^+(\lambda,\mu,r,z)$ where $\widetilde{f}_\epsilon^+(\lambda,\mu,r,z) = \mathrm{kl}(z,\mu) + \frac{1}{r} h(r\epsilon(z-\lambda))$. Using Lemma 28, direct computations yield that, for all $z \in ([\lambda]_0^1, \mu)$,

$$\begin{aligned} \frac{\partial \widetilde{f}_\epsilon^+}{\partial z}(\lambda,\mu,r,z) &= \log\left(\frac{z(1-\mu)}{(1-z)\mu}\right) + \epsilon h'(r\epsilon(z-\lambda)) \\ &= \log\left(\frac{z(1-\mu)}{(1-z)\mu}\right) + \epsilon \frac{r\epsilon(z-\lambda)}{\sqrt{(r\epsilon(z-\lambda))^2 + 1} + 1} \, , \\ \frac{\partial^2 \widetilde{f}_\epsilon^+}{\partial z^2}(\lambda,\mu,r,z) &= \frac{1}{z(1-z)} + r\epsilon^2 h''(r\epsilon(z-\lambda)) > 0 \, . \end{aligned}$$

Therefore, $z \to \widetilde{f}_\epsilon^+(\lambda,\mu,r,z)$ is twice continuously differentiable, positive and strictly convex on $([\lambda]_0^1,\mu)$. Moreover, we have

$$\lim_{z\to\mu}\frac{\partial\widetilde{f}_\epsilon^+}{\partial z}(\lambda,\mu,r,z) = \epsilon h'(r\epsilon(\mu-\lambda)) > 0\,,$$
$$\frac{\partial\widetilde{f}_\epsilon^+}{\partial z}(\lambda,\mu,r,g_\epsilon^+(\mu)) = -\epsilon\left(1-\frac{r\epsilon(z-\lambda)}{\sqrt{(r\epsilon(z-\lambda))^2+1}+1}\right) < 0\,,$$
$$\text{When } [\lambda]_0^1 > g_\epsilon^+(\mu),\quad \frac{\partial\widetilde{f}_\epsilon^+}{\partial z}(\lambda,\mu,r,\lambda) = \log\left(\frac{\lambda(1-\mu)}{(1-\lambda)\mu}\right) < 0\,.$$

Note that $\max\{[\lambda]_0^1, g_\epsilon^+(\mu)\} = \max\{\lambda, g_\epsilon^+(\mu)\}$ since $\mu \in (0,1)$. Using that $z \to \frac{\partial\widetilde{f}_\epsilon^+}{\partial z}(\lambda,\mu,r,z)$ is continuously differentiable and increasing on $([\lambda]_0^1,\mu)$, with negative value at $\max\{\lambda, g_\epsilon^+(\mu)\}$ and finite positive limit at μ, $z \mapsto \widetilde{f}_\epsilon^+(\lambda,\mu,r,z)$ admit a unique minimizer on $(\max\{\lambda, g_\epsilon^+(\mu)\},\mu)$. Let $\widetilde{g}_\epsilon^+(\lambda,\mu,r) \in (\max\{\lambda, g_\epsilon^+(\mu)\},\mu)$ be defined as this unique minimizer, defined implicitly as solution for $z \in (\max\{\lambda, g_\epsilon^+(\mu)\},\mu)$ of the equation

$$\log\left(\frac{z(1-\mu)}{(1-z)\mu}\right) + \epsilon\frac{r\epsilon(z-\lambda)}{\sqrt{(r\epsilon(z-\lambda))^2+1}+1} = 0\,.$$

Then, we have $\frac{\partial\widetilde{f}_\epsilon^+}{\partial z}(\lambda,\mu,r,z) = 0$ if and only if $z = \widetilde{g}_\epsilon^+(\lambda,\mu,r)$. Moreover, $z \mapsto \widetilde{f}_\epsilon^+(\lambda,\mu,r,z)$ is decreasing on $([\lambda]_0^1, \widetilde{g}_\epsilon^+(\lambda,\mu,r))$ and increasing on $(\widetilde{g}_\epsilon^+(\lambda,\mu,r),\mu)$.

Let us define $z = g_\epsilon^+(\mu) + x$ where $x \in (\max\{0, \lambda - g_\epsilon^+(\mu)\}, \mu - g_\epsilon^+(\mu))$. Then, we have

$$\frac{\partial\widetilde{f}_\epsilon^+}{\partial z}(\lambda,\mu,r,g_\epsilon^+(\mu)+x)$$
$$= \log\left(1+\frac{x}{g_\epsilon^+(\mu)(1-x-g_\epsilon^+(\mu))}\right) + \epsilon\left(\frac{r\epsilon(x+g_\epsilon^+(\mu)-\lambda)}{\sqrt{(r\epsilon(x+g_\epsilon^+(\mu)-\lambda))^2+1}+1} - 1\right)$$

Therefore, we have $\widetilde{g}_\epsilon^+(\lambda,\mu,r) = g_\epsilon^+(\mu) + x_\epsilon^+(\lambda,\mu,r)$ where $x_\epsilon^+(\lambda,\mu,r) \in (\max\{0,\lambda - g_\epsilon^+(\mu)\}, \mu - g_\epsilon^+(\mu))$ is the solution for $x \in (\max\{0,\lambda - g_\epsilon^+(\mu)\}, \mu - g_\epsilon^+(\mu))$ of the equation

$$\log\left(1+\frac{x}{g_\epsilon^+(\mu)(1-x-g_\epsilon^+(\mu))}\right) + \epsilon\left(\frac{r\epsilon(x+g_\epsilon^+(\mu)-\lambda)}{\sqrt{(r\epsilon(x+g_\epsilon^+(\mu)-\lambda))^2+1}+1} - 1\right) = 0\,.$$

When $\lambda \in (0,1)$ and $\mu \to \lambda = [\lambda]_0^1$, it is direct to see that $\widetilde{g}_\epsilon^+(\lambda,\mu,r) \to \lambda$. Then, we have

$$\lim_{\mu\to\lambda^+}\widetilde{d}_\epsilon^+(\lambda,\mu,r) = \lim_{\mu\to\lambda^+}\{\mathrm{kl}(\widetilde{g}_\epsilon^+(\lambda,\mu,r),\mu)\} + \frac{1}{r}\lim_{\widetilde{g}_\epsilon^+(\lambda,\mu,r)\to\lambda^+}\{h(r\epsilon(\widetilde{g}_\epsilon^+(\lambda,\mu,r)-\lambda))\} = 0\,.$$

Direct computation yields that, for $z \in ([\lambda]_0^1,\mu) \subset (0,1)$,

$$\frac{\partial\widetilde{f}_\epsilon^+}{\partial\mu}(\lambda,\mu,r,z) = \frac{\mu-z}{\mu(1-\mu)} > 0\,,$$
$$\frac{\partial^2\widetilde{f}_\epsilon^+}{\partial\mu^2}(\lambda,\mu,r,z) = \frac{(\mu-z)^2+z(1-z)}{\mu^2(1-\mu)^2} > 0\,,$$
$$\frac{\partial^2\widetilde{f}_\epsilon^+}{\partial z^2}(\lambda,\mu,r,z) = \frac{1}{z(1-z)} + r\epsilon^2 h''(r\epsilon(z-\lambda)) > 0\,,$$
$$\frac{\partial^2\widetilde{f}_\epsilon^+}{\partial\mu\partial z}(\lambda,\mu,r,z) = -\frac{1}{\mu(1-\mu)} < 0\,.$$

Since $\frac{\partial\widetilde{f}_\epsilon^+}{\partial z}(\lambda,\mu,r,g_\epsilon^+(\lambda,\mu,r)) = 0$, the implicit function theorem yields that

$$\frac{\partial\widetilde{g}_\epsilon^+}{\partial\mu}(\lambda,\mu,r) = -\frac{\frac{\partial^2 f_\epsilon^+}{\partial\mu\partial z}(\lambda,\mu,r,g_\epsilon^+(\lambda,\mu,r))}{\frac{\partial^2\widetilde{f}_\epsilon^+}{\partial z^2}(\lambda,\mu,r,g_\epsilon^+(\lambda,\mu,r))} > 0\,.$$

Moreover, for $\mu \in ([\lambda]_0^1, 1)$,

$$\begin{aligned}
\frac{\partial \widetilde{d}_\epsilon^+}{\partial \mu}(\lambda,\mu,r) &= \frac{\partial \widetilde{f}_\epsilon^+}{\partial \mu}(\lambda,\mu,r,\widetilde{g}_\epsilon^+(\lambda,\mu,r)) + \frac{\partial \widetilde{g}_\epsilon^+}{\partial \mu}(\lambda,\mu,r)\frac{\partial \widetilde{f}_\epsilon^+}{\partial z}(\lambda,\mu,r,\widetilde{g}_\epsilon^+(\lambda,\mu,r)) \\
&= \frac{\partial \widetilde{f}_\epsilon^+}{\partial \mu}(\lambda,\mu,r,\widetilde{g}_\epsilon^+(\lambda,\mu,r)) = \frac{\mu - g_\epsilon^+(\mu) - x_\epsilon^+(\lambda,\mu,r)}{\mu(1-\mu)} > 0\,, \\
\frac{\partial^2 \widetilde{d}_\epsilon^+}{\partial \mu^2}(\lambda,\mu,r) &= \frac{\partial^2 \widetilde{f}_\epsilon^+}{\partial \mu^2}(\lambda,\mu,r,\widetilde{g}_\epsilon^+(\lambda,\mu,r))\frac{\partial \widetilde{g}_\epsilon^+}{\partial \mu}(\lambda,\mu,r) > 0\,.
\end{aligned}$$

Therefore, for all $(\lambda, r) \in \mathbb{R} \times \mathbb{R}_+^\star$, the function $\mu \mapsto \widetilde{d}_\epsilon^+(\lambda,\mu,r)$ is positive, twice continuously differentiable, increasing and strictly convex on $([\lambda]_0^1, 1)$.

Let $(\mu, r) \in (0,1) \times \mathbb{R}_+^\star$. Direct computation yields that, for $z \in ([\lambda]_0^1, \mu) \subset (0,1)$,

$$\begin{aligned}
\frac{\partial \widetilde{f}_\epsilon^+}{\partial \lambda}(\lambda,\mu,r,z) &= -\epsilon h'(r\epsilon(z-\lambda)) < 0\,, \\
\frac{\partial^2 \widetilde{f}_\epsilon^+}{\partial \lambda \partial z}(\lambda,\mu,r,z) &= -r\epsilon^2 h''(r\epsilon(z-\lambda)) < 0\,.
\end{aligned}$$

Since $\frac{\partial \widetilde{f}_\epsilon^+}{\partial z}(\lambda,\mu,r,g_\epsilon^+(\lambda,\mu,r)) = 0$, the implicit function theorem yields that

$$\frac{\partial \widetilde{g}_\epsilon^+}{\partial \lambda}(\lambda,\mu,r) = -\frac{\frac{\partial^2 f_\epsilon^+}{\partial \lambda \partial z}(\lambda,\mu,r,g_\epsilon^+(\lambda,\mu,r))}{\frac{\partial^2 \widetilde{f}_\epsilon^+}{\partial z^2}(\lambda,\mu,r,g_\epsilon^+(\lambda,\mu,r))} = \frac{r\epsilon^2 h''(r\epsilon(g_\epsilon^+(\lambda,\mu,r)-\lambda))}{\frac{1}{z(1-z)} + r\epsilon^2 h''(r\epsilon(g_\epsilon^+(\lambda,\mu,r)-\lambda))} < 1\,.$$

Direct computation yields that, for $\lambda \in (-\infty, \mu)$,

$$\begin{aligned}
\frac{\partial \widetilde{d}_\epsilon^+}{\partial \lambda}(\lambda,\mu,r) &= \frac{\partial \widetilde{f}_\epsilon^+}{\partial \lambda}(\lambda,\mu,r,\widetilde{g}_\epsilon^+(\lambda,\mu,r)) + \frac{\partial \widetilde{g}_\epsilon^+}{\partial \lambda}(\lambda,\mu,r)\frac{\partial \widetilde{f}_\epsilon^+}{\partial z}(\lambda,\mu,r,\widetilde{g}_\epsilon^+(\lambda,\mu,r)) \\
&= \frac{\partial \widetilde{f}_\epsilon^+}{\partial \lambda}(\lambda,\mu,r,\widetilde{g}_\epsilon^+(\lambda,\mu,r)) = -\epsilon h'(r\epsilon(\widetilde{g}_\epsilon^+(\lambda,\mu,r)-\lambda)) < 0\,, \\
\frac{\partial^2 \widetilde{d}_\epsilon^+}{\partial \lambda^2}(\lambda,\mu,r) &= r\epsilon^2\left(1 - \frac{\partial \widetilde{g}_\epsilon^+}{\partial \lambda}(\lambda,\mu,r)\right) h''(r\epsilon(\widetilde{g}_\epsilon^+(\lambda,\mu,r)-\lambda)) > 0\,.
\end{aligned}$$

Therefore, for all $(\mu, r) \in (0,1) \times \mathbb{R}_+^\star$, the function $\lambda \mapsto \widetilde{d}_\epsilon^+(\lambda,\mu,r)$ is positive, twice continuously differentiable, decreasing and strictly convex on $(-\infty, \mu)$. Similarly as above, it is direct to see that $\lim_{\lambda\to\mu^-} \widetilde{d}_\epsilon^+(\lambda,\mu,r) = 0$ and $\lim_{\lambda\to-\infty} \widetilde{d}_\epsilon^+(\lambda,\mu,r) = +\infty$.

Let $(\lambda, \mu) \in \mathbb{R} \times (0,1)$. When $\mu \in (0, [\lambda]_0^1]$, we have $\widetilde{d}_\epsilon^+(\lambda,\mu,r) = 0$ for all $r \in [1, +\infty)$, hence $r \mapsto \widetilde{d}_\epsilon^+(\lambda,\mu,r)$ is non-decreasing. Let κ as in Lemma 29. Using Lemma 29, we have

$$\forall z > \lambda, \quad \frac{\partial \widetilde{f}_\epsilon^+}{\partial r}(\lambda,\mu,r,z) = \frac{\partial \kappa}{\partial r}(r, \epsilon(z-\lambda)) > 0\,.$$

When $\mu \in ([\lambda]_0^1, 1)$, we have $\widetilde{g}_\epsilon^+(\lambda,\mu,r) \in (\max\{\lambda, g_\epsilon^+(\mu)\}, \mu)$ and, for all $r > 0$,

$$\begin{aligned}
\frac{\partial \widetilde{d}_\epsilon^+}{\partial r}(\lambda,\mu,r) &= \frac{\partial \widetilde{f}_\epsilon^+}{\partial r}(\lambda,\mu,r,\widetilde{g}_\epsilon^+(\lambda,\mu,r)) + \frac{\partial \widetilde{g}_\epsilon^+}{\partial r}(\lambda,\mu,r)\frac{\partial \widetilde{f}_\epsilon^+}{\partial z}(\lambda,\mu,r,\widetilde{g}_\epsilon^+(\lambda,\mu,r)) \\
&= \frac{\partial \widetilde{f}_\epsilon^+}{\partial r}(\lambda,\mu,r,\widetilde{g}_\epsilon^+(\lambda,\mu,r)) > 0\,,
\end{aligned}$$

where we used that $\widetilde{g}_\epsilon^+(\lambda,\mu,r) > \lambda$. This concludes the last part of the proof. □

Lemma 32 gathers regularity properties on the modified divergences $\widetilde{d}_\epsilon^-$. In particular, it gives a closed-form solution based on an implicit solution of a fixed-point equation. This is a key property used in our implementation to reduce the computational cost.

Lemma 32. *Let $\widetilde{d}_\epsilon^-$ as in Eq.* (32)*, and $g_\epsilon^\pm$ as in Eq.* (30)*. For all $\mu \in (0,1)$, $\lambda \in \mathbb{R}$ and $r > 0$, we have*

$$\widetilde{d}_\epsilon^-(\lambda,\mu,r) = \begin{cases} 0 & \text{if } \mu \in [[\lambda]_0^1, 1) \\ \mathrm{kl}(g_\epsilon^-(\mu) - x_\epsilon^-(\lambda,\mu,r), \mu) + \frac{1}{r} h(r\epsilon(x_\epsilon^-(\lambda,\mu,r) + \lambda - g_\epsilon^-(\mu))) & \text{if } \mu \in (0, [\lambda]_0^1) \end{cases},$$

where $x_\epsilon^-(\lambda,\mu,r) := x_\epsilon^+(1-\lambda, 1-\mu, r) \in (\max\{g_\epsilon^-(\mu) - \lambda, 0\}, g_\epsilon^-(\mu) - \mu)$ is the solution for $x \in (\max\{g_\epsilon^-(\mu) - \lambda, 0\}, g_\epsilon^-(\mu) - \mu)$ of the equation

$$\log\left(1 + \frac{x}{(1 - g_\epsilon^-(\mu))(g_\epsilon^-(\mu) - x)}\right) + \epsilon\left(\frac{r\epsilon(x - g_\epsilon^-(\mu) + \lambda)}{\sqrt{(r\epsilon(x - g_\epsilon^-(\mu) + \lambda))^2 + 1} + 1} - 1\right) = 0\,.$$

For all $(\mu, r) \in (0,1) \times \mathbb{R}_+^\star$, the function $\lambda \mapsto \widetilde{d}_\epsilon^-(\lambda,\mu,r)$ is positive, twice continuously differentiable, increasing and strictly convex on $(\mu, +\infty)$; it satisfies $\lim_{\lambda \to \mu^+} \widetilde{d}_\epsilon^+(\lambda,\mu,r) = 0$ and $\lim_{\lambda \to +\infty} \widetilde{d}_\epsilon^-(\lambda,\mu,r) = +\infty$.

For all $(\lambda, r) \in \mathbb{R} \times \mathbb{R}_+^\star$, the function $\mu \mapsto \widetilde{d}_\epsilon^-(\lambda,\mu,r)$ is positive, twice continuously differentiable, decreasing and strictly convex on $(0, [\lambda]_0^1)$. Moreover, we have

$$\forall \mu \in (0, [\lambda]_0^1), \quad \frac{\partial \widetilde{d}_\epsilon^-}{\partial \mu}(\lambda,\mu,r) = \frac{\mu - g_\epsilon^-(\mu) + x_\epsilon^-(\lambda,\mu,r)}{\mu(1-\mu)}\,.$$

For all $(\lambda, \mu) \in \mathbb{R} \times (0,1)$ such that $\mu \in (0, [\lambda]_0^1]$, the function $r \mapsto \widetilde{d}_\epsilon^-(\lambda,\mu,r)$ is the zero function. For all $(\lambda, \mu) \in \mathbb{R} \times (0,1)$ such that $\mu \in ([\lambda]_0^1, 1)$, the function $r \mapsto \widetilde{d}_\epsilon^-(\lambda,\mu,r)$ is positive, continuously differentiable and increasing on $\mathbb{R}_+$.

Proof. Using Lemmas 27 and 24, we have

$$\widetilde{d}_\epsilon^-(\lambda,\mu,r) = \widetilde{d}_\epsilon^+(1-\lambda, 1-\mu, r) \quad \text{and} \quad g_\epsilon^+(\lambda) = 1 - g_\epsilon^-(1-\lambda)\,,$$
$$\frac{\partial \widetilde{d}_\epsilon^-}{\partial \mu}(\lambda,\mu,r) = -\frac{\partial \widetilde{d}_\epsilon^+}{\partial \mu}(1-\lambda, 1-\mu, r) \quad \text{and} \quad \frac{\partial^2 \widetilde{d}_\epsilon^-}{\partial \mu^2}(\lambda,\mu,r) = \frac{\partial^2 \widetilde{d}_\epsilon^+}{\partial \mu^2}(1-\lambda, 1-\mu, r)\,.$$

Let $x_\epsilon^+(1-\lambda, 1-\mu, r) \in (\max\{0, g_\epsilon^-(\mu) - \lambda\}, g_\epsilon^-(\mu) - \mu)$ be the unique solution for $x \in (\max\{0, g_\epsilon^-(\mu) - \lambda\}, g_\epsilon^-(\mu) - \mu)$ of the equation

$$\log\left(1 + \frac{x}{(1 - g_\epsilon^-(\mu))(g_\epsilon^-(\mu) - x)}\right) + \epsilon\left(\frac{r\epsilon(x - g_\epsilon^-(\mu) + \lambda)}{\sqrt{(r\epsilon(x - g_\epsilon^-(\mu) + \lambda))^2 + 1} + 1} - 1\right) = 0\,,$$

where we used $g_\epsilon^+(1-\mu) = 1 - g_\epsilon^-(\mu)$ to simplify the formula given in Lemma 31. Therefore, we define $x_\epsilon^-(\lambda,\mu,r) = x_\epsilon^+(1-\lambda, 1-\mu, r)$. Then, we have

$$\mathrm{kl}(g_\epsilon^-(\mu) - x_\epsilon^-(\lambda,\mu,r), \mu) + \frac{1}{r} h(r\epsilon(x_\epsilon^-(\lambda,\mu,r) + \lambda - g_\epsilon^-(\mu))) =$$
$$\mathrm{kl}(x_\epsilon^+(1-\lambda, 1-\mu, r) + g_\epsilon^+(1-\mu), 1-\mu) + \frac{h(r\epsilon(x_\epsilon^+(1-\lambda, 1-\mu, r) + g_\epsilon^+(1-\mu) - 1 + \lambda))}{r}$$

where we used that $\mathrm{kl}(g_\epsilon^-(\mu) - x_\epsilon^-(\lambda,\mu,r), \mu) = \mathrm{kl}(1 - g_\epsilon^-(\mu) + x_\epsilon^-(\lambda,\mu,r), 1-\mu)$. Combining the above with the properties on $\widetilde{d}_\epsilon^+$ in Lemma 31 concludes the proof. □

Lemma 33 shows that we can invert $\widetilde{d}_\epsilon^\pm$ with respect to their first argument, which is a key property used in Appendix F.

Lemma 33. *Let $\widetilde{d}_\epsilon^\pm$ as in Eq.* (32)*. For all $(\mu, r, c) \in (0,1) \times \mathbb{R}_+^\star \times \mathbb{R}_+^\star$, there exists $x > 0$ such that $\widetilde{d}_\epsilon^+(\mu - x, \mu, r) = c$. For all $(\mu, r, c) \in (0,1) \times \mathbb{R}_+^\star \times \mathbb{R}_+^\star$, there exists $x > 0$ such that $\widetilde{d}_\epsilon^-(\mu + x, \mu, r) = c$.*

Proof. Let us define $f(x) = \widetilde{d}_\epsilon^+(\mu - x, \mu, r)$ for all $x > 0$. Using Lemma 31, we know that f is continuous and increasing on $\mathbb{R}_+^\star$ and it satisfies $\lim_{x\to 0^+} f(x) = 0$ and $\lim_{x\to+\infty} f(x) = +\infty$. Therefore, there exists a unique $x > 0$ such that $\widetilde{d}_\epsilon^+(\mu - x, \mu, r) = c$. Using Lemma 32, we can conclude similarly for $\widetilde{d}_\epsilon^-$. □

Lemma 33 shows that $\widetilde{d}_\epsilon^\pm$ is non-decreasing with respect to their first argument, which is a key property used in Appendix F.

Lemma 34. *Let $\widetilde{d}_\epsilon^\pm$ as in Eq.* (32)*. For all $(\mu, r) \in (0,1) \times \mathbb{R}_+^\star$ and all $(\lambda_1, \lambda_2) \in \mathbb{R} \times (-\infty, \mu)$,*

$$\widetilde{d}_\epsilon^+(\lambda_1, \mu, r) \geq \widetilde{d}_\epsilon^+(\lambda_2, \mu, r) \quad \Longrightarrow \quad \lambda_1 \leq \lambda_2$$

For all $(\mu, r) \in (0,1) \times \mathbb{R}_+^\star$ and all $(\lambda_1, \lambda_2) \in \mathbb{R} \times (\mu, +\infty)$,

$$\widetilde{d}_\epsilon^-(\lambda_1, \mu, r) \geq \widetilde{d}_\epsilon^-(\lambda_2, \mu, r) \quad \Longrightarrow \quad \lambda_1 \geq \lambda_2$$

Proof. Using Lemma 31, we known that $\lambda \mapsto \widetilde{d}_\epsilon^+(\lambda, \mu, r)$ is decreasing on $(-\infty, \mu)$. Let $(\lambda_1, \lambda_2) \in \mathbb{R} \times (-\infty, \mu)$. Then, we have

$$\lambda_1 > \lambda_2 \quad \Longrightarrow \quad \widetilde{d}_\epsilon^+(\lambda_1, \mu, r) < \widetilde{d}_\epsilon^+(\lambda_2, \mu, r)\,,$$

which is equivalent to the statement of the lemma by contraposition. Using Lemma 32, we can conclude similarly for $\widetilde{d}_\epsilon^-$. □

G.2 Transportation Cost

Recall that $W_{\epsilon,a,b}$ is defined in Eq. (50), i.e., for all $(\mu, w) \in \mathbb{R}^K \times \mathbb{R}_+^K$,

$$\forall (a,b) \in [K]^2, \quad W_{\epsilon,a,b}(\mu, w) := \mathbb{1}\left([\mu_a]_0^1 > [\mu_b]_0^1\right) \inf_{u\in[0,1]} \left\{ w_a d_\epsilon^-(\mu_a, u) + w_b d_\epsilon^+(\mu_b, u) \right\}\,,$$

where $d_\epsilon^\pm$ are defined in Eq. (3).

Lemma 35 gathers regularity properties on the transportation costs.

Lemma 35. *Let $d_\epsilon^\pm$ as in Eq.* (3)*. For all $(\lambda, \mu) \in (0,1)^2$ such that $\lambda \geq \mu$ and $w \in \mathbb{R}_+^2$.*

- *The function $u \mapsto w_1 d_\epsilon^-(\lambda, u) + w_2 d_\epsilon^+(\mu, u)$ is strictly convex on $[\mu, \lambda]$ when $\max\{w_1, w_2\} > 0$ and on $[0,1]$ when $\min\{w_1, w_2\} > 0$. Then,*

$$\inf_{u\in[0,1]} \{w_1 d_\epsilon^-(\lambda, u) + w_2 d_\epsilon^+(\mu, u)\} = \inf_{u\in[\mu,\lambda]} \{w_1 d_\epsilon^-(\lambda, u) + w_2 d_\epsilon^+(\mu, u)\}\,.$$

- *The function $(\lambda, \mu, w) \mapsto \inf_{u\in[0,1]} \{w_1 d_\epsilon^-(\lambda, u) + w_2 d_\epsilon^+(\mu, u)\}$ is continuous on $(0,1) \times (0,1) \times \mathbb{R}_+^2$.*
- *If $\max\{w_1, w_2\} > 0$, $u_\star(\lambda, \mu, w) = \arg\min_{u\in[0,1]} \{w_1 d_\epsilon^-(\lambda, u) + w_2 d_\epsilon^+(\mu, u)\}$ is unique and continuous on $(0,1) \times (0,1) \times \mathbb{R}_+^2$.*
- *If $\min\{w_1, w_2\} > 0$ and $\lambda > \mu$, $u_\star(\lambda, \mu, w) \in (\mu, \lambda)$ and $\min\{d_\epsilon^-(\lambda, u_\star(\lambda, \mu, w)), d_\epsilon^+(\mu, u_\star(\lambda, \mu, w))\} > 0$.*

Moreover,

$$\inf_{u\in[0,1]} \{w_1 d_\epsilon^-(\lambda, u) + w_2 d_\epsilon^+(\mu, u)\} = \inf_{(u_1,u_2)\in[0,1]^2\,:\,u_1\leq u_2} \{w_1 d_\epsilon^-(\lambda, u_1) + w_2 d_\epsilon^+(\mu, u_2)\}\,.$$

Proof. These results are obtained by leveraging Lemmas 25 and 26 at each step.

For $u \leq \mu$, the function is equal to $w_1 d_\epsilon^-(\lambda, u)$, which is decreasing and strictly convex on $[0, \lambda)$ unless $w_1 = 0$ since $u \leq \mu \leq \lambda$. Therefore, the minimum over that interval is attained at μ. For $u \geq \lambda$, the function is equal to $w_2 d_\epsilon^+(\mu, u)$, which is increasing and strictly convex on $(\mu, 1]$ unless $w_2 = 0$ since $u \geq \lambda \geq \mu$. Therefore, the minimum over that interval is attained at λ. On the interval (μ, λ), the function is equal to $w_1 d_\epsilon^-(\lambda, u) + w_2 d_\epsilon^+(\mu, u)$, hence it is the sum of two convex functions,

one of which is strictly convex. Furthermore, the function is continuous at μ and λ. This concludes the first part of the proof.

As we have just shown, we can restrict the infimum to $[\mu, \lambda]$. We apply Berge's Maximum theorem [Berge, 1997, page 116]. Let

$$\begin{aligned}
\phi(u, \lambda, \mu, w) &= -w_1 d_\epsilon^-(\lambda, u) - w_2 d_\epsilon^+(\mu, u) \,, \\
\Gamma(\lambda, \mu, w) &= [\mu, \lambda] \,, \\
M(\lambda, \mu, w) &= \max\{\phi(u, \lambda, \mu, w) \mid u \in \Gamma(\lambda, \mu, w)\} \,, \\
\Phi(\lambda, \mu, w) &= \arg\max\{\phi(u, \lambda, \mu, w) \mid u \in \Gamma(\lambda, \mu, w)\} \,.
\end{aligned}$$

We verify the hypotheses of the theorem:

- ϕ is continuous on $[\mu, \lambda] \times (0,1) \times (0,1) \times \mathbb{R}_+^2$, by using the properties in Lemmas 25 and 26 since $(\lambda, \mu) \in (0,1)^2$.
- Γ is nonempty, compact-valued and continuous (since constant).

We obtain that M is continuous on $(0,1) \times (0,1) \times \mathbb{R}_+^2$ and that Φ is upper hemicontinuous. This concludes the second part of the proof.

When $\max\{w_1, w_2\} > 0$, we have just shown that ϕ is a strictly concave function of u. Combining this with the fact that Γ is convex, we can argue as in [Sundaram, 1996, Theorem 9.17] to prove that Φ is a single-valued upper hemicontinuous correspondence, hence a continuous function. This concludes the third part of the proof.

Suppose that $\min\{w_1, w_2\} > 0$ and $\lambda > \mu$. Using Lemmas 25 and 26, the function $u \mapsto w_1 d_\epsilon^-(\lambda, u) + w_2 d_\epsilon^+(\mu, u)$ is continuously differentiable on (μ, λ) with derivative $w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda, u) + w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu, u)$ where

$$\begin{aligned}
\forall u \in (\mu, 1], \quad \frac{\partial d_\epsilon^+}{\partial u}(\mu, u) &= \begin{cases} \frac{1-e^{-\epsilon}}{1-u(1-e^{-\epsilon})} & \text{if } u \in (g_\epsilon^-(\mu), 1] \\ \frac{u-\mu}{u(1-u)} & \text{if } u \in (\mu, g_\epsilon^-(\mu)] \end{cases} \,, \\
\forall u \in [0, \lambda), \quad \frac{\partial d_\epsilon^-}{\partial u}(\lambda, u) &= \begin{cases} -\frac{e^\epsilon - 1}{1+u(e^\epsilon - 1)} & \text{if } u \in [0, g_\epsilon^+(\lambda)) \\ -\frac{\lambda - u}{u(1-u)} & \text{if } u \in [g_\epsilon^+(\lambda), \lambda) \end{cases} \,.
\end{aligned}$$

Since $\frac{\partial d_\epsilon^-}{\partial u}(\lambda, u) \to_{u \to \lambda^-} 0$ and $\frac{\partial d_\epsilon^+}{\partial u}(\mu, u) \to_{u \to \mu^+} 0$, we obtain

$$\begin{aligned}
\lim_{u \to \lambda^-} \left\{ w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda, u) + w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu, u) \right\} &= w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu, \lambda) > 0 \,, \\
\lim_{u \to \mu^+} \left\{ w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda, u) + w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu, u) \right\} &= w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda, \mu) < 0 \,.
\end{aligned}$$

Therefore, the infimum is attained inside the open interval. Using Lemmas 25 and 26, we can conclude the proof of the first part of the fourth property.

Using the strict convexity of $u_1 \mapsto w_1 d_\epsilon^-(\lambda, u_1)$ and $u_2 \mapsto w_2 d_\epsilon^+(\mu, u_2)$ on (μ, λ), we obtain that

$$\inf_{u \in (\mu, \lambda)} \{w_1 d_\epsilon^-(\lambda, u) + w_2 d_\epsilon^+(\mu, u)\} = \inf_{(u_1, u_2) \,:\, \mu < u_1 \leq u_2 < \lambda} \{w_1 d_\epsilon^-(\lambda, u_1) + w_2 d_\epsilon^+(\mu, u_2)\} \,.$$

Re-using the same arguments as above, we obtain that

$$\begin{aligned}
&\inf_{(u_1, u_2) \,:\, \mu < u_1 \leq u_2 < \lambda} \{w_1 d_\epsilon^-(\lambda, u_1) + w_2 d_\epsilon^+(\mu, u_2)\} \\
&= \inf_{(u_1, u_2) \in [0,1]^2 \,:\, u_1 \leq u_2} \{w_1 d_\epsilon^-(\lambda, u_1) + w_2 d_\epsilon^+(\mu, u_2)\} \,.
\end{aligned}$$

This concludes the proof of the second part of the fourth property. □

Lemma 36 relates the transportation costs $W_{\epsilon, a^\star, a}$ with the transportation costs used in Eq. (2) to define the characteristic time. Crucially, this shows the equivalence with the definitions in Eq. (35).

Lemma 36. *Let $W_{\epsilon,a^\star,a}$ and d_ϵ as in Eq.* (4) *and* (1)*. Let $\mu \in (0,1)^K$ such that $a^\star(\mu) = \{a^\star\}$. Let Alt$(\mu) = \{\lambda \in (0,1)^K \mid a^\star(\lambda) \neq \{a^\star\}\}$. Then,*

$$\forall w \in \triangle_K, \quad \inf_{\lambda \in \mathit{Alt}(\mu)} \sum_{a \in [K]} w_a \mathrm{d}_\epsilon(\mu_a, \lambda_a) = \min_{a \neq a^\star} W_{\epsilon,a^\star,a}(\mu, w) \, .$$

Proof. It is direct to see that $\mathrm{Alt}(\mu) = \bigcup_{a \neq a^\star} \mathcal{C}_a$ where $\mathcal{C}_a = \{\lambda \in (0,1)^K \mid \lambda_a \geq \lambda_{a^\star}\}$. Then,

$$\forall w \in \triangle_K, \quad \inf_{\lambda \in \mathrm{Alt}(\mu)} \sum_{a \in [K]} w_a d_\epsilon(\mu_a, \lambda_a) = \min_{a \neq a^\star} \inf_{\lambda \in \mathcal{C}_a} \sum_{c \in [K]} w_c d_\epsilon(\mu_c, \lambda_c) \, .$$

By non-negativity of $d_\epsilon(\mu_a, \lambda_a)$ for all $a \in [K]$, we obtain

$$\begin{aligned} \inf_{\lambda \in \mathcal{C}_a} \sum_{c \in [K]} w_c d_\epsilon(\mu_c, \lambda_c) &= \inf_{\lambda \in \mathcal{C}_a} \sum_{c \in \{a, a^\star\}} w_c d_\epsilon(\mu_c, \lambda_c) \\ &= \inf_{(\lambda_a, \lambda_{a^\star}) \in (0,1)^2, \, \lambda_a \geq \lambda_{a^\star}} \sum_{c \in \{a, a^\star\}} w_c d_\epsilon(\mu_c, \lambda_c) \, , \end{aligned}$$

where the two equalities are obtained by choosing $\lambda(a) \in (0,1)^K$ such that $\lambda(a)_b = \mu_b$ for all $b \notin \{a, a^\star\}$ with the two other coordinates choosen freely such that $\lambda(a)_a \geq \lambda(a)_{a^\star}$. Using that $\mu_{a^\star} > \mu_a$, we can partition this set as follows

$$\begin{aligned} \mathcal{C}_{a,a^\star} = \{(\lambda_a, \lambda_{a^\star}) \in (0,1)^2 \mid \lambda_a \geq \lambda_{a^\star}\} &= \{(\lambda_a, \lambda_{a^\star}) \in (0, \mu_a)^2 \mid \lambda_a \geq \lambda_{a^\star}\} \\ &\cup \{(\lambda_a, \lambda_{a^\star}) \in [\mu_a, \mu_{a^\star}] \times (0, \mu_a)\} \\ &\cup \{(\lambda_a, \lambda_{a^\star}) \in (\mu_{a^\star}, 1)^2 \mid \lambda_a \geq \lambda_{a^\star}\} \\ &\cup \{(\lambda_a, \lambda_{a^\star}) \in (\mu_{a^\star}, 1) \times [\mu_a, \mu_{a^\star}]\} \\ &\cup \{(\lambda_a, \lambda_{a^\star}) \in [\mu_a, \mu_{a^\star}]^2 \mid \lambda_a \geq \lambda_{a^\star}\} \, . \end{aligned}$$

Using Lemma 22, $\mu_{a^\star} > \mu_a$ and Lemmas 25 and 26, we obtain

$$\begin{aligned} &\inf_{(\lambda_a, \lambda_{a^\star}) \in (0,\mu_a)^2 \mid \lambda_a \geq \lambda_{a^\star}} \{w_{a^\star} d_\epsilon(\mu_{a^\star}, \lambda_{a^\star}) + w_a d_\epsilon(\mu_a, \lambda_a)\} \\ &= \inf_{(\lambda_a, \lambda_{a^\star}) \in (0,\mu_a)^2 \mid \lambda_a \geq \lambda_{a^\star}} \{w_{a^\star} d_\epsilon^-(\mu_{a^\star}, \lambda_{a^\star}) + w_a d_\epsilon^-(\mu_a, \lambda_a)\} = w_{a^\star} d_\epsilon^-(\mu_{a^\star}, \mu_a) \, , \\ &\inf_{(\lambda_a, \lambda_{a^\star}) \in [\mu_a,\mu_{a^\star}] \times (0,\mu_a)} \{w_{a^\star} d_\epsilon(\mu_{a^\star}, \lambda_{a^\star}) + w_a d_\epsilon(\mu_a, \lambda_a)\} \\ &= \inf_{(\lambda_a, \lambda_{a^\star}) \in [\mu_a,\mu_{a^\star}] \times (0,\mu_a)} \{w_{a^\star} d_\epsilon^-(\mu_{a^\star}, \lambda_{a^\star}) + w_a d_\epsilon^+(\mu_a, \lambda_a)\} = w_{a^\star} d_\epsilon^-(\mu_{a^\star}, \mu_a) \, , \\ &\inf_{(\lambda_a, \lambda_{a^\star}) \in (\mu_{a^\star},1)^2 \mid \lambda_a \geq \lambda_{a^\star}} \{w_{a^\star} d_\epsilon(\mu_{a^\star}, \lambda_{a^\star}) + w_a d_\epsilon(\mu_a, \lambda_a)\} \\ &= \inf_{(\lambda_a, \lambda_{a^\star}) \in (\mu_{a^\star},1)^2 \mid \lambda_a \geq \lambda_{a^\star}} \{w_{a^\star} d_\epsilon^+(\mu_{a^\star}, \lambda_{a^\star}) + w_a d_\epsilon^+(\mu_a, \lambda_a)\} = w_a d_\epsilon^+(\mu_a, \mu_{a^\star}) \, , \\ &\inf_{(\lambda_a, \lambda_{a^\star}) \in (\mu_{a^\star},1) \times [\mu_a,\mu_{a^\star}]} \{w_{a^\star} d_\epsilon(\mu_{a^\star}, \lambda_{a^\star}) + w_a d_\epsilon(\mu_a, \lambda_a)\} \\ &= \inf_{(\lambda_a, \lambda_{a^\star}) \in (\mu_{a^\star},1) \times [\mu_a,\mu_{a^\star}]} \{w_{a^\star} d_\epsilon^-(\mu_{a^\star}, \lambda_{a^\star}) + w_a d_\epsilon^+(\mu_a, \lambda_a)\} = w_a d_\epsilon^+(\mu_a, \mu_{a^\star}) \, , \\ &\inf_{(\lambda_a, \lambda_{a^\star}) \in [\mu_a,\mu_{a^\star}]^2 \mid \lambda_a \geq \lambda_{a^\star}} \{w_{a^\star} d_\epsilon(\mu_{a^\star}, \lambda_{a^\star}) + w_a d_\epsilon(\mu_a, \lambda_a)\} \\ &= \inf_{(\lambda_a, \lambda_{a^\star}) \in [\mu_a,\mu_{a^\star}]^2 \mid \lambda_a \geq \lambda_{a^\star}} \{w_{a^\star} d_\epsilon^-(\mu_{a^\star}, \lambda_{a^\star}) + w_a d_\epsilon^+(\mu_a, \lambda_a)\} \, . \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} &\inf_{(\lambda_a, \lambda_{a^\star}) \in \mathcal{C}_{a,a^\star}} \{w_{a^\star} d_\epsilon(\mu_{a^\star}, \lambda_{a^\star}) + w_a d_\epsilon(\mu_a, \lambda_a)\} \\ &= \inf_{(\lambda_a, \lambda_{a^\star}) \in [\mu_a,\mu_{a^\star}]^2 \mid \lambda_a \geq \lambda_{a^\star}} \{w_{a^\star} d_\epsilon^-(\mu_{a^\star}, \lambda_{a^\star}) + w_a d_\epsilon^+(\mu_a, \lambda_a)\} \\ &= \inf_{u \in [\mu_a,\mu_{a^\star}]^2} \{w_{a^\star} d_\epsilon^-(\mu_{a^\star}, u) + w_a d_\epsilon^+(\mu_a, u)\} \\ &= \inf_{u \in [0,1]} \{w_{a^\star} d_\epsilon^-(\mu_{a^\star}, u) + w_a d_\epsilon^+(\mu_a, u)\} = W_{\epsilon,a^\star,a}(\mu, w) \, , \end{aligned}$$

where the second equality is obtained similarly as in Lemma 35 by leveraging the strict convexity of $d_\epsilon^\pm$ in their second argument (see Lemmas 25 and 26). We used Lemma 35 and the definition of $W_{\epsilon,a^\star,a}(\mu,w)$ for the last two equalities. This concludes the proof. □

Lemma 37 gathers additional properties on the transportation costs.

Lemma 37. *Let $d_\epsilon^\pm$ as in Eq.* (3)*.*

- *Let $(\lambda,\mu)\in(0,1)^2$ such that $\lambda>\mu$. When $w_2>0$, the function $w_1\mapsto \min_{u\in[0,1]}\{w_1 d_\epsilon^-(\lambda,u)+w_2 d_\epsilon^+(\mu,u)\}$ is increasing on $\mathbb{R}_+$. When $w_1>0$, the function $w_2\mapsto \min_{u\in[0,1]}\{w_1 d_\epsilon^-(\lambda,u)+w_2 d_\epsilon^+(\mu,u)\}$ is increasing on $\mathbb{R}_+$.*
- *Let $(\lambda,\mu)\in(0,1)^2$ and $\mu\in(0,1)^K$. The function $w\mapsto \min_{u\in[0,1]}\{w_1 d_\epsilon^-(\lambda,u)+w_2 d_\epsilon^+(\mu,u)\}$ is concave on $\mathbb{R}_+^2$. The function $w\mapsto \min_{a\in[K]\setminus\{1\}}\min_{u\in[0,1]}\{w_1 d_\epsilon^-(\mu_1,u)+w_a d_\epsilon^+(\mu_a,u)\}$ is concave on $\mathbb{R}_+^K$.*

Proof. Let $w_2>0$ and $w_1'>w_1\geq 0$. Using Lemma 35, since $w_1'>0$, there exists $u'\in[0,1]$ with $d_\epsilon^-(\lambda,u')>0$ such that

$$\begin{aligned}\min_{u\in[0,1]}\{w_1' d_\epsilon^-(\lambda,u)+w_2 d_\epsilon^+(\mu,u)\} &= w_1' d_\epsilon^-(\lambda,u')+w_2 d_\epsilon^+(\mu,u')\\ &> w_1 d_\epsilon^-(\lambda,u')+w_2 d_\epsilon^+(\mu,u')\\ &\geq \min_{u\in[0,1]}\{w_1 d_\epsilon^-(\lambda,u)+w_2 d_\epsilon^+(\mu,u)\}\,.\end{aligned}$$

Let $w_1>0$ and $w_2'>w_2\geq 0$. Then, we can show similarly by using Lemma 35 that

$$\min_{u\in[0,1]}\{w_1 d_\epsilon^-(\lambda,u)+w_2' d_\epsilon^+(\mu,u)\} > \min_{u\in[0,1]}\{w_1 d_\epsilon^-(\lambda,u)+w_2 d_\epsilon^+(\mu,u)\}\,.$$

This concludes the first part of the proof. The proof of the second part is direct since those functions are minimum of linear functions, hence concave. □

Lemma 38 gives a closed-form solution for the transportation costs. This is a key property used in our implementation to reduce the computational cost.

Lemma 38. *Let $d_\epsilon^\pm$ and $g_\epsilon^\pm$ as in Eq.* (3) *and* (30)*. For all $(a,c)\in\mathbb{R}_+^2$ and $b\in\mathbb{R}$, let $r_{1,+}(a,b,c):=\frac{\sqrt{b^2+4ac}-b}{2a}$. For all $(\lambda,\mu)\in(0,1)^2$ and $w\in\mathbb{R}_+^2$ such that $\min\{w_1,w_2\}>0$ and $\lambda>\mu$.*

• When (1) $g_\epsilon^-(\mu)\geq\lambda$, or (2) $g_\epsilon^-(\mu)<\lambda$, $g_\epsilon^+(\lambda)\leq g_\epsilon^-(\mu)$ and $\frac{w_2\mu+w_1\lambda}{w_2+w_1}\in[g_\epsilon^+(\lambda),g_\epsilon^-(\mu)]$, we have $u_\star(\lambda,\mu,w)=\frac{w_2\mu+w_1\lambda}{w_2+w_1}$ and

$$\min_{u\in[0,1]}\{w_1 d_\epsilon^-(\lambda,u)+w_2 d_\epsilon^+(\mu,u)\} = w_1\mathrm{kl}(\lambda,u_\star(\lambda,\mu,w))+w_2\mathrm{kl}(\mu,u_\star(\lambda,\mu,w))\,.$$

• When (3) $g_\epsilon^-(\mu)<\lambda$, $g_\epsilon^+(\lambda)>g_\epsilon^-(\mu)$ and $u_{3,\star}(w)\in[g_\epsilon^-(\mu),g_\epsilon^+(\lambda)]$ where

$$u_{3,\star}(w):=\frac{w_1(e^\epsilon-1)-w_2(1-e^{-\epsilon})}{(w_2+w_1)(1-e^{-\epsilon})(e^\epsilon-1)}\,,$$

we have $u_\star(\lambda,\mu,w)=u_{3,\star}(w)$ and

$$\begin{aligned}&\min_{u\in[0,1]}\{w_1 d_\epsilon^-(\lambda,u)+w_2 d_\epsilon^+(\mu,u)\}\\ &= w_1\left(-\log\left(1+u_{3,\star}(w)(e^\epsilon-1)\right)+\epsilon\lambda\right)+w_2\left(-\log\left(1-u_{3,\star}(w)(1-e^{-\epsilon})\right)-\epsilon\mu\right)\,.\end{aligned}$$

• When (4) $g_\epsilon^-(\mu)<\lambda$, $g_\epsilon^+(\lambda)\leq g_\epsilon^-(\mu)$ and $\frac{w_2\mu+w_1\lambda}{w_2+w_1}\in(\mu,g_\epsilon^+(\lambda))$, or (5) $g_\epsilon^-(\mu)<\lambda$, $g_\epsilon^+(\lambda)>g_\epsilon^-(\mu)$ and $u_{3,\star}(w)<g_\epsilon^-(\mu)$, we have $u_\star(\lambda,\mu,w)=u_{1,\star}(\mu,w)$ and

$$\begin{aligned}&u_{1,\star}(\mu,w):=r_{1,+}\left((w_2+w_1)(e^\epsilon-1),(w_2-(w_2\mu+w_1)(e^\epsilon-1)),w_2\mu\right)\,,\\ &\min_{u\in[0,1]}\{w_1 d_\epsilon^-(\lambda,u)+w_2 d_\epsilon^+(\mu,u)\}\end{aligned}$$

$$= w_1 \left(-\log\left(1 + u_{1,\star}(\mu, w)(e^\epsilon - 1)\right) + \epsilon\lambda\right) + w_2 \mathrm{kl}(\mu, u_{1,\star}(\mu, w)) \,.$$

• *When (6)* $g_\epsilon^-(\mu) < \lambda$, $g_\epsilon^+(\lambda) \le g_\epsilon^-(\mu)$ *and* $\frac{w_2\mu + w_1\lambda}{w_2 + w_1} \in (g_\epsilon^-(\mu), \lambda)$, *or (7)* $g_\epsilon^-(\mu) < \lambda$, $g_\epsilon^+(\lambda) > g_\epsilon^-(\mu)$ *and* $u_{3,\star}(w) > g_\epsilon^+(\lambda)$, *we have* $u_\star(\lambda, \mu, w) = u_{2,\star}(\lambda, w)$ *and*

$$u_{2,\star}(\lambda, w) := 1 - r_{1,+}\left((w_2 + w_1)(e^\epsilon - 1), (w_1 - (w_1(1-\lambda) + w_2)(e^\epsilon - 1)), w_1(1-\lambda)\right) \,,$$

$$\min_{u \in [0,1]} \{w_1 d_\epsilon^-(\lambda, u) + w_2 d_\epsilon^+(\mu, u)\}$$
$$= w_1 \mathrm{kl}(\lambda, u_{2,\star}(\lambda, w)) + w_2 \left(-\log\left(1 - u_{2,\star}(\lambda, w)(1 - e^{-\epsilon})\right) - \epsilon\mu\right) \,.$$

Proof. Suppose that $g_\epsilon^-(\mu) \ge \lambda$. Using Lemma 24, we know that $g_\epsilon^-(\mu) \ge \lambda$ if and only if $\mu \ge g_\epsilon^+(\lambda)$. Therefore, for all $u \in (\mu, \lambda)$, we have

$$w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda, u) + w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu, u) = -w_1 \frac{\lambda - u}{u(1-u)} + w_2 \frac{u - \mu}{u(1-u)} = \frac{(w_2 + w_1)u - (w_2\mu + w_1\lambda)}{u(1-u)} \,.$$

Therefore, we have

$$u_\star(\lambda, \mu, w) = \frac{w_2\mu + w_1\lambda}{w_2 + w_1} \in (\mu, \lambda) \,.$$

Suppose that $g_\epsilon^-(\mu) < \lambda$. Using Lemma 24, we know that $g_\epsilon^-(\mu) < \lambda$ if and only if $\mu < g_\epsilon^+(\lambda)$. Using strict convexity of the function on (μ, λ), it is enough to exhibit one local minimum to obtain a global minimum on (μ, λ).

Suppose that $g_\epsilon^+(\lambda) \le g_\epsilon^-(\mu)$. Similarly as above, we obtain, for all $u \in [g_\epsilon^+(\lambda), g_\epsilon^-(\mu)]$,

$$w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda, u) + w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu, u) = \frac{(w_2 + w_1)u - (w_2\mu + w_1\lambda)}{u(1-u)} \,.$$

Suppose that $\frac{w_2\mu + w_1\lambda}{w_2 + w_1} \in [g_\epsilon^+(\lambda), g_\epsilon^-(\mu)]$. Then, we can conclude as above that

$$u_\star(\lambda, \mu, w) = \frac{w_2\mu + w_1\lambda}{w_2 + w_1} \in [g_\epsilon^+(\lambda), g_\epsilon^-(\mu)] \,,$$

since it is a local minimum of a strictly convex function.

Suppose that $\frac{w_2\mu + w_1\lambda}{w_2 + w_1} < g_\epsilon^+(\lambda)$. Since the gradient is positive on $[g_\epsilon^+(\lambda), g_\epsilon^-(\mu)]$, we know that the minimum on (μ, λ) is achieved on $(\mu, g_\epsilon^+(\lambda))$, i.e., $u_\star(\lambda, \mu, w) \in (\mu, g_\epsilon^+(\lambda))$. Then, for all $u \in (\mu, g_\epsilon^+(\lambda))$,

$$w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda, u) + w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu, u) = -w_1 \frac{e^\epsilon - 1}{1 + u(e^\epsilon - 1)} + w_2 \frac{u - \mu}{u(1-u)} \,.$$

Using Lemma 24, direct computation yields

$$w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda, u) + w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu, u) > 0 \quad \iff \quad \mu < u\left(1 + \frac{w_1}{w_2}\left(1 - \frac{g_\epsilon^-(u)}{u}\right)\right) \,,$$

$$\lim_{u \to \mu^+} u\left(1 + \frac{w_1}{w_2}\left(1 - \frac{g_\epsilon^-(u)}{u}\right)\right) = \mu\left(1 + \frac{w_1}{w_2}\left(1 - \frac{g_\epsilon^-(\mu)}{\mu}\right)\right) < \mu \,,$$

$$\lim_{u \to g_\epsilon^+(\lambda)^-} u\left(1 + \frac{w_1}{w_2}\left(1 - \frac{g_\epsilon^-(u)}{u}\right)\right) = g_\epsilon^+(\lambda)\left(1 + \frac{w_1}{w_2}\left(1 - \frac{\lambda}{g_\epsilon^+(\lambda)}\right)\right) > \mu \,,$$

where the second result uses that $u < g_\epsilon^-(u)$ and the last result is obtained by continuity of the differentials (Lemmas 25 and 26) and the positivity on $[g_\epsilon^+(\lambda), g_\epsilon^-(\mu)]$. For all $(a, c) \in \mathbb{R}_+^2$ and $b \in \mathbb{R}$, we define $r_{1,+}(a, b, c) = \frac{\sqrt{b^2 + 4ac} - b}{2a}$. Therefore, we have

$$\begin{aligned}
& w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda, u) + w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu, u) = 0 \\
\iff \quad & (w_2 + w_1)(e^\epsilon - 1)u^2 + (w_2 - (w_2\mu + w_1)(e^\epsilon - 1))\, u - w_2\mu = 0 \\
\iff \quad & u_\star(\lambda, \mu, w) = r_{1,+}\left((w_2 + w_1)(e^\epsilon - 1), (w_2 - (w_2\mu + w_1)(e^\epsilon - 1)), w_2\mu\right) \in (\mu, g_\epsilon^+(\lambda))
\end{aligned}$$

where we used that $u_\star(\lambda,\mu,w) \in (\mu, g_\epsilon^+(\lambda))$ is unique for the last equivalence, and that the second root of the second order polynomial equation is negative. Notice that $u_\star(\lambda,\mu,w)$ is independent of λ.

Suppose that $\frac{w_2\mu+w_1\lambda}{w_2+w_1} > g_\epsilon^-(\mu)$. Since the gradient is negative on $[g_\epsilon^+(\lambda), g_\epsilon^-(\mu)]$, we know that the minimum on (μ,λ) is achieved on $(g_\epsilon^-(\mu),\lambda)$, i.e., $u_\star(\lambda,\mu,w) \in (g_\epsilon^-(\mu),\lambda)$. Then, for all $u \in (g_\epsilon^-(\mu),\lambda)$,

$$w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda,u) + w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu,u) = -w_1 \frac{\lambda-u}{u(1-u)} + w_2 \frac{1-e^{-\epsilon}}{1-u(1-e^{-\epsilon})} \,.$$

Using Lemma 24, direct computation yields

$$\begin{aligned}
&w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda,u) + w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu,u) < 0 \quad \iff \quad \lambda > u\left(1 + \frac{w_2}{w_1}\left(1 - \frac{g_\epsilon^+(u)}{u}\right)\right) \,, \\
&\lim_{u\to\lambda^-} u\left(1 + \frac{w_2}{w_1}\left(1 - \frac{g_\epsilon^+(u)}{u}\right)\right) = \lambda\left(1 + \frac{w_2}{w_1}\left(1 - \frac{g_\epsilon^+(\lambda)}{\lambda}\right)\right) > \lambda \,, \\
&\lim_{u\to g_\epsilon^-(\mu)^+} u\left(1 + \frac{w_2}{w_1}\left(1 - \frac{g_\epsilon^+(u)}{u}\right)\right) = g_\epsilon^-(\mu)\left(1 + \frac{w_2}{w_1}\left(1 - \frac{\mu}{g_\epsilon^-(\mu)}\right)\right) < \lambda \,,
\end{aligned}$$

where the second result uses that $u > g_\epsilon^+(u)$ and the last result is obtained by continuity of the differentials (Lemmas 25 and 26) and the negativity on $[g_\epsilon^+(\lambda), g_\epsilon^-(\mu)]$.

Using Lemma 23, we obtain

$$\arg\min_{u\in[0,1]}\{w_1 d_\epsilon^-(\lambda,u) + w_2 d_\epsilon^+(\mu,u)\} = 1 - \arg\min_{u\in[0,1]}\{w_1 d_\epsilon^+(1-\lambda,u) + w_2 d_\epsilon^-(1-\mu,u)\} \,.$$

Using Lemma 24, we obtain

$$\begin{aligned}
&g_\epsilon^-(\mu) < \lambda \quad \iff \quad g_\epsilon^-(1-\lambda) < 1-\mu \,, \\
&\frac{w_2\mu + w_1\lambda}{w_2+w_1} \in (g_\epsilon^-(\mu),\lambda) \quad \iff \quad \frac{w_2(1-\mu)+w_1(1-\lambda)}{w_2+w_1} \in (1-\lambda, g_\epsilon^+(1-\mu)) \,.
\end{aligned}$$

Therefore, we can leverage the above case to obtain $u_\star(\lambda,\mu,w) = u_{2,\star}(\lambda,w)$ where

$$u_{2,\star}(\lambda,w) = 1 - r_{1,+}\left((w_2+w_1)(e^\epsilon-1), (w_1 - (w_1(1-\lambda)+w_2)(e^\epsilon-1)), w_1(1-\lambda)\right)$$

Notice that $u_\star(\lambda,\mu,w)$ is independent of μ.

Suppose that $g_\epsilon^-(\mu) < \lambda$ and $g_\epsilon^+(\lambda) > g_\epsilon^-(\mu)$. Similarly as above, we obtain, for all $u \in [g_\epsilon^-(\mu), g_\epsilon^+(\lambda)]$,

$$w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda,u) + w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu,u) = -w_1 \frac{e^\epsilon-1}{1+u(e^\epsilon-1)} + w_2 \frac{1-e^{-\epsilon}}{1-u(1-e^{-\epsilon})} \,.$$

Therefore, we obtain

$$\begin{aligned}
& w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda,u) + w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu,u) > 0 \\
\iff \quad & w_2(1-e^{-\epsilon})(1+u(e^\epsilon-1)) - w_1(e^\epsilon-1)(1-u(1-e^{-\epsilon})) > 0 \\
\iff \quad & u > u_{3,\star}(w) := \frac{w_1(e^\epsilon-1) - w_2(1-e^{-\epsilon})}{(w_2+w_1)(1-e^{-\epsilon})(e^\epsilon-1)} \,.
\end{aligned}$$

Suppose that $u_{3,\star}(w) \in [g_\epsilon^-(\mu), g_\epsilon^+(\lambda)]$. Then, we can conclude as above that

$$u_\star(\lambda,\mu,w) = u_{3,\star}(w) \in [g_\epsilon^-(\mu), g_\epsilon^+(\lambda)] \,,$$

since it is a local minimum of a strictly convex function. Notice that $u_{3,\star}(w)$ is independent of (λ,μ).

Suppose that $u_{3,\star}(w) > g_\epsilon^+(\lambda)$. Since the gradient is negative on $[g_\epsilon^-(\mu), g_\epsilon^+(\lambda)]$, we know that the minimum on (μ,λ) is achieved on $(g_\epsilon^+(\lambda),\lambda)$, i.e., $u_\star(\lambda,\mu,w) \in (g_\epsilon^+(\lambda),\lambda)$. Then, for all $u \in (g_\epsilon^+(\lambda),\lambda)$,

$$w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda,u) + w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu,u) = -w_1 \frac{\lambda-u}{u(1-u)} + w_2 \frac{1-e^{-\epsilon}}{1-u(1-e^{-\epsilon})} \,.$$

This recovers the condition solved above. As we know that $u_\star(\lambda,\mu,w) \in (g_\epsilon^+(\lambda),\lambda)$, we obtain $u_\star(\lambda,\mu,w) = u_{2,\star}(\lambda,w)$ where

$$u_{2,\star}(\lambda,w) = 1 - r_{1,+}\left((w_2+w_1)(e^\epsilon-1), (w_1 - (w_1(1-\lambda)+w_2)(e^\epsilon-1)), w_1(1-\lambda)\right)$$

Suppose that $u_{3,\star}(w) < g_\epsilon^-(\mu)$. Since the gradient is positive on $[g_\epsilon^-(\mu), g_\epsilon^+(\lambda)]$, we know that the minimum on (μ,λ) is achieved on $(\mu, g_\epsilon^-(\mu))$, i.e., $u_\star(\lambda,\mu,w) \in (\mu, g_\epsilon^-(\mu))$. Then, for all $u \in (\mu, g_\epsilon^-(\mu))$,

$$w_1 \frac{\partial d_\epsilon^-}{\partial u}(\lambda,u) + w_2 \frac{\partial d_\epsilon^+}{\partial u}(\mu,u) = -w_1 \frac{e^\epsilon - 1}{1+u(e^\epsilon-1)} + w_2 \frac{u-\mu}{u(1-u)} \,.$$

This recovers the condition solved above. As we know that $u_\star(\lambda,\mu,w) \in (\mu, g_\epsilon^-(\mu))$, we obtain $u_\star(\lambda,\mu,w) = u_{1,\star}(\mu,w)$ where

$$u_{1,\star}(\mu,w) = r_{1,+}\left((w_2+w_1)(e^\epsilon-1), (w_2 - (w_2\mu+w_1)(e^\epsilon-1)), w_2\mu\right) \,.$$

This concludes the proof.

□

G.2.1 Modified Transportation Cost

Let $\eta > 0$ be the geometric parameter used for the geometric grid update of our private empirical mean estimator. Let us define

$$\forall x \geq 1, \quad r(x) := \frac{x}{1+\log_{1+\eta} x} \,, \tag{33}$$

which is increasing if and only if $x > \frac{e}{1+\eta}$. For all $(\mu,w) \in \mathbb{R}^K \times \mathbb{R}_+^K$ and all $(a,b) \in [K]^2$ such that $a \neq b$, we define

$$\widetilde{W}_{\epsilon,a,b}(\mu,w) := \mathbb{1}\left([\mu_a]_0^1 > [\mu_b]_0^1\right) \inf_{u\in(0,1)} \left\{ w_a \widetilde{d}_\epsilon^-(\mu_a,u,r(w_a)) + w_b \widetilde{d}_\epsilon^+(\mu_b,u,r(w_b)) \right\} \,, \tag{34}$$

where $\widetilde{d}_\epsilon^\pm$ are defined in Eq. (32).

Lemma 39 gathers regularity properties of the function r defined in Eq. (33).

Lemma 39. *Let r as in Eq.* (33)*. Then,*

$$\forall x \geq 1, \quad r'(x) = \frac{\log(x(1+\eta)/e)}{\log(1+\eta)(1+\log_{1+\eta} x)^2} \,,$$
$$r''(x) = -\frac{1}{x(\log(1+\eta))^2} \frac{\log((1+\eta)xe^{-2})}{(1+\log_{1+\eta} x)^3} \,.$$

On $[1,+\infty)$, the function r is twice continuously differentiable. It is decreasing on $[1, e/(1+\eta))$ and increasing on $(e/(1+\eta), +\infty)$; its minium is $r(e/(1+\eta)) \in (0,1)$. It is strictly convex on $[1, e^2/(1+\eta))$ and strictly concave on $(e^2/(1+\eta), +\infty)$.

Proof. The proof is obtained by direct differentiation and manipulation. We have

$$\forall \eta > 0, \quad r(e/(1+\eta)) = \frac{e\log(1+\eta)}{1+\eta} \in (0,1) \,.$$

□

Lemma 40 shows that the modified transportation costs can be rewritten differently, which is a key property used in Appendix F.

Lemma 40. *Let $\widetilde{d}_\epsilon^\pm$ as in Eq.* (32)*, and r as in Eq.* (33)*. For all $(\lambda,\mu) \in \mathbb{R}^2$ such that $[\lambda]_0^1 > [\mu]_0^1$ and $(w_1,w_2) \in [1,+\infty)^2$. Then,*

$$\begin{aligned}
&\inf_{u\in(0,1)} \{ w_1 \widetilde{d}_\epsilon^-(\lambda,u,r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu,u,r(w_2)) \} \\
&= \inf_{u\in([\mu]_0^1,[\lambda]_0^1)} \{ w_1 \widetilde{d}_\epsilon^-(\lambda,u,r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu,u,r(w_2)) \} \\
&= \inf_{(u_1,u_2)\in(0,1)^2 :\, u_1\leq u_2} \{ w_1 \widetilde{d}_\epsilon^-(\lambda,u_1,r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu,u_2,r(w_2)) \} \,.
\end{aligned}$$

Proof. These results are obtained by leveraging Lemmas 31 and 32.

Note that the condition $[\lambda]_0^1 > [\mu]_0^1$ implies that $\mu \in (-\infty, 1)$ and $\lambda \in (0, +\infty)$, i.e., $[\mu]_0^1 = \max\{0, \mu\}$ and $[\lambda]_0^1 = \min\{1, \lambda\}$.

Suppose that $\mu \leq 0$ and $\lambda \geq 1$. Then, we have $[\mu]_0^1 = 0$ and $[\lambda]_0^1 = 1$. Therefore, the first part of the result holds by definition.

Suppose that $\mu \leq 0$ and $\lambda \in (0, 1)$. Then, we have $[\mu]_0^1 = 0$ and $[\lambda]_0^1 = \lambda$. For $u \in [\lambda, 1)$, the function is equal to $w_2 \widetilde{d}_\epsilon^+(\mu, u, r(w_2))$, which is increasing and strictly convex on $(0, 1)$. Therefore, the minimum over that interval is attained at λ. For $u \in (0, \lambda)$, the function is equal to $w_1 \widetilde{d}_\epsilon^-(\lambda, u, r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu, u, r(w_2))$. Since it is the sum of two strictly convex function, the minimum over that interval is achieved in $(0, \lambda)$. This concludes the proof of the first part of the result for this case.

Suppose that $\mu \in (0, 1)$ and $\lambda \geq 1$. Then, we have $[\mu]_0^1 = \mu$ and $[\lambda]_0^1 = 1$. For $u \in (0, \mu]$, the function is equal to $w_1 \widetilde{d}_\epsilon^-(\lambda, u, r(w_2))$, which is decreasing and strictly convex on $(0, 1)$. Therefore, the minimum over that interval is attained at μ. For $u \in (\mu, 1)$, the function is equal to $w_1 \widetilde{d}_\epsilon^-(\lambda, u, r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu, u, r(w_2))$. Since it is the sum of two strictly convex function, the minimum over that interval is achieved in $(\mu, 1)$. This concludes the proof of the first part of the result for this case.

Suppose that $(\mu, \lambda) \in (0, 1)^2$. Then, we have $[\mu]_0^1 = \mu$ and $[\lambda]_0^1 = \lambda$. For $u \in [\lambda, 1)$, the function is equal to $w_2 \widetilde{d}_\epsilon^+(\mu, u, r(w_2))$, which is increasing and strictly convex on $(0, 1)$. Therefore, the minimum over that interval is attained at λ. For $u \in (0, \mu]$, the function is equal to $w_1 \widetilde{d}_\epsilon^-(\lambda, u, r(w_2))$, which is decreasing and strictly convex on $(0, 1)$. Therefore, the minimum over that interval is attained at μ. For $u \in (\mu, \lambda)$, the function is equal to $w_1 \widetilde{d}_\epsilon^-(\lambda, u, r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu, u, r(w_2))$. Since it is the sum of two strictly convex function, the minimum over that interval is achieved in (μ, λ). This concludes the proof of the first part of the result for this case.

In summary, we have shown that

$$\begin{aligned}
&\inf_{u \in (0,1)} \{w_1 \widetilde{d}_\epsilon^-(\lambda, u, r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu, u, r(w_2))\} \\
&= \inf_{u \in ([\mu]_0^1, [\lambda]_0^1)} \{w_1 \widetilde{d}_\epsilon^-(\lambda, u, r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu, u, r(w_2))\} .
\end{aligned}$$

Using the strict convexity of $u_1 \mapsto w_1 \widetilde{d}_\epsilon^-(\lambda, u_1, r(w_1))$ and $u_2 \mapsto w_2 \widetilde{d}_\epsilon^+(\mu, u_2, r(w_2))$ on $([\mu]_0^1, [\lambda]_0^1)$, we obtain that

$$\begin{aligned}
&\inf_{u \in ([\mu]_0^1, [\lambda]_0^1)} \{w_1 \widetilde{d}_\epsilon^-(\lambda, u, r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu, u, r(w_2))\} \\
&= \inf_{(u_1, u_2) \,:\, [\mu]_0^1 < u_1 \leq u_2 < [\lambda]_0^1} \{w_1 \widetilde{d}_\epsilon^-(\lambda, u_1, r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu, u_2, r(w_2))\} .
\end{aligned}$$

Re-using the same arguments as above, we obtain that

$$\begin{aligned}
&\inf_{(u_1, u_2) \,:\, [\mu]_0^1 < u_1 \leq u_2 < [\lambda]_0^1} \{w_1 \widetilde{d}_\epsilon^-(\lambda, u_1, r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu, u_2, r(w_2))\} = \\
&= \inf_{(u_1, u_2) \in (0,1)^2 \colon u_1 \leq u_2} \{w_1 \widetilde{d}_\epsilon^-(\lambda, u_1, r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu, u_2, r(w_2))\} .
\end{aligned}$$

This concludes the proof. □

Lemma 41 gives a closed-form solution for the modified transportation costs based on an implicit solution of a fixed-point equation. This is a key property used in our implementation to reduce the computational cost.

Lemma 41. *Let $\widetilde{d}_\epsilon^\pm$ as in Eq.* (32)*, $x_\epsilon^\pm$ as in Lemmas 31 and 32, and r as in Eq.* (33)*. For all $(\lambda, \mu) \in \mathbb{R}^2$ such that $[\lambda]_0^1 > [\mu]_0^1$ and $w \in [1, +\infty)^2$. Then,*

$$\begin{aligned}
&\inf_{u \in (0,1)} \{w_1 \widetilde{d}_\epsilon^-(\lambda, u, r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu, u, r(w_2))\} \\
&= w_1 \widetilde{d}_\epsilon^-(\lambda, u^\star(\lambda, \mu, w), r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu, u^\star(\lambda, \mu, w), r(w_2)) ,
\end{aligned}$$

where $u^\star(\lambda,\mu,w) \in ([\mu]_0^1,[\lambda]_0^1)$ *is the unique solution for* $u \in ([\mu]_0^1,[\lambda]_0^1)$ *of the equation*

$$u(w_1+w_2) - w_1 g_\epsilon^-(u) - w_2 g_\epsilon^+(u) + w_1 x_\epsilon^-(\lambda,u,r(w_1)) - w_2 x_\epsilon^+(\mu,u,r(w_2)) = 0\,.$$

Proof. Using Lemma 40, we have

$$\begin{aligned}&\inf_{u\in(0,1)} \{w_1 \widetilde{d}_\epsilon^-(\lambda,u,r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu,u,r(w_2))\}\\ &= \inf_{u\in([\mu]_0^1,[\lambda]_0^1)} \{w_1 \widetilde{d}_\epsilon^-(\lambda,u,r(w_1)) + w_2 \widetilde{d}_\epsilon^+(\mu,u,r(w_2))\}\,.\end{aligned}$$

Using Lemmas 32 and 31, we obtain

$$\begin{aligned}\forall u \in (0,[\lambda]_0^1), \quad \frac{\partial \widetilde{d}_\epsilon^-}{\partial u}(\lambda,u,r(w_1)) &= \frac{u - g_\epsilon^-(u) + x_\epsilon^-(\lambda,u,r(w_1))}{u(1-u)}\,,\\ \forall u \in ([\mu]_0^1,1), \quad \frac{\partial \widetilde{d}_\epsilon^+}{\partial u}(\mu,u,r(w_2)) &= \frac{u - g_\epsilon^+(u) - x_\epsilon^+(\mu,u,r(w_2))}{u(1-u)}\,.\end{aligned}$$

Therefore, for all $u \in ([\mu]_0^1,[\lambda]_0^1)$,

$$\begin{aligned}&w_1 \frac{\partial \widetilde{d}_\epsilon^-}{\partial u}(\lambda,u,r(w_1)) + w_2 \frac{\partial \widetilde{d}_\epsilon^+}{\partial u}(\mu,u,r(w_2))\\ &= \frac{w_1(u - g_\epsilon^-(u) + x_\epsilon^-(\lambda,u,r(w_1))) + w_2(u - g_\epsilon^+(u) - x_\epsilon^+(\mu,u,r(w_2)))}{u(1-u)}\\ &= \frac{u(w_1+w_2) - (w_1 g_\epsilon^-(u) + w_2 g_\epsilon^+(u)) + w_1 x_\epsilon^-(\lambda,u,r(w_1)) - w_2 x_\epsilon^+(\mu,u,r(w_2))}{u(1-u)}\,.\end{aligned}$$

For $u \in ([\mu]_0^1,[\lambda]_0^1)$, let us define

$$g_1(u) := u(w_1+w_2) - w_1 g_\epsilon^-(u) - w_2 g_\epsilon^+(u) + w_1 x_\epsilon^-(\lambda,u,r(w_1)) - w_2 x_\epsilon^+(\mu,u,r(w_2))\,.$$

Using the proof of Lemmas 32 and 31, we know that

$$\begin{aligned}&\lim_{u\to[\mu]_0^1} \frac{\partial \widetilde{d}_\epsilon^+}{\partial u}(\mu,u,r(w_1)) = 0 \quad \text{and} \quad \lim_{u\to[\lambda]_0^1} \frac{\partial \widetilde{d}_\epsilon^-}{\partial u}(\lambda,u,r(w_1)) = 0\,,\\ &\forall u \in (0,[\lambda]_0^1), \quad \frac{\partial \widetilde{d}_\epsilon^-}{\partial u}(\lambda,u,r(w_1)) < 0 \quad \text{and} \quad \forall u \in ([\mu]_0^1,1), \quad \frac{\partial \widetilde{d}_\epsilon^+}{\partial u}(\mu,u,r(w_1)) > 0\,.\end{aligned}$$

Combined with the strict convexity of $\widetilde{d}_\epsilon^\pm$ in their second argument, the equation $g_1(u) = 0$ admits a unique solution on $([\mu]_0^1,[\lambda]_0^1)$. Since $u(1-u) > 0$, we obtain the implicit equation defining $u^\star(\lambda,\mu,w)$ as above. □

G.3 Characteristic Time

Let $\boldsymbol{\nu}$ be a Bernoulli instance with means $\mu \in (0,1)^2$ and unique best arm $a^\star \in [K]$, i.e., $\arg\max_{a\in[K]} \mu_a = \{a^\star\}$. For all $\beta \in (0,1)$, we define

$$T_\epsilon^\star(\boldsymbol{\nu})^{-1} = \sup_{w\in\triangle_K} \min_{a\neq a^\star} W_{\epsilon,a^\star,b}(\mu,w) \quad \text{and} \quad w_\epsilon^\star(\boldsymbol{\nu}) = \arg\max_{w\in\triangle_K} \min_{a\neq a^\star} W_{\epsilon,a^\star,b}(\mu,w)\,, \tag{35}$$

$$T_{\epsilon,\beta}^\star(\boldsymbol{\nu})^{-1} = \sup_{w\in\triangle_K, w_{a^\star}=\beta} \min_{a\neq a^\star} W_{\epsilon,a^\star,b}(\mu,w) \quad \text{and} \quad w_{\epsilon,\beta}^\star(\boldsymbol{\nu}) = \arg\max_{w\in\triangle_K, w_{a^\star}=\beta} \min_{a\neq a^\star} W_{\epsilon,a^\star,b}(\mu,w)$$

where $W_{\epsilon,a,b}$ are defined in Eq. (4).

Lemma 42 gathers regularity properties on the characteristic times and their optimal allocations.

Lemma 42. *Let* $W_{\epsilon,a,b}$ *as in Eq.* (4)*. Let* $(T_\epsilon^\star, T_{\epsilon,\beta}^\star)$ *and* $(w_\epsilon^\star, w_{\epsilon,\beta}^\star)$ *as in Eq.* (35)*. The function* $(\mu,w) \mapsto \min_{a\neq a^\star(\mu)} W_{\epsilon,a^\star(\mu),a}(\mu,w)$ *is continuous on* $(0,1)^K \times \triangle_K$*. The functions* $\boldsymbol{\nu} \mapsto T_\epsilon^\star(\boldsymbol{\nu})^{-1}$ *and* $\boldsymbol{\nu} \mapsto T_{\epsilon,\beta}^\star(\boldsymbol{\nu})^{-1}$ *are continuous on* $\mathcal{F}^K$*. The correspondences* $\boldsymbol{\nu} \mapsto w_\epsilon^\star(\boldsymbol{\nu})$ *and* $\boldsymbol{\nu} \mapsto w_{\epsilon,\beta}^\star(\boldsymbol{\nu})$ *are upper hemicontinuous on* $\mathcal{F}^K$ *with compact convex values.*

Proof. Let $\mathcal{F}_a^K = \{\boldsymbol{\nu} \in \mathcal{F}^K \mid a \in a^\star(\boldsymbol{\nu})\}$. Since $\bigcup_{a\in[K]} \mathcal{F}_a^K = \mathcal{F}^K$, it is enough to show the property for all $\mathcal{F}_a^K$ for $a \in [K]$. Let $a^\star \in [K]$.

First, the function $(w, \boldsymbol{\nu}) \mapsto \min_{a\neq a^\star} \inf_{u\in[0,1]} \{w_{a^\star} d_\epsilon^-(\mu_{a^\star}, u) + w_a d_\epsilon^+(\mu_{a^\star}, u)\}$ is continuous on $\triangle_K \times \mathcal{F}^K$ by Lemma 35 and the fact that a minimum of continuous functions is continuous. It is concave in w by Lemma 37.

The correspondence $(w, \boldsymbol{\nu}) \mapsto \triangle_K$ is nonempty compact-valued and continuous (since constant). By Berge's maximum theorem, we get that $\boldsymbol{\nu} \mapsto T_\epsilon^\star(\boldsymbol{\nu})^{-1}$ is continuous on $\mathcal{F}_{a^\star}^K$ and that $\boldsymbol{\nu} \mapsto w_\epsilon^\star(\boldsymbol{\nu})$ is upper hemicontinuous with compact values. By [Sundaram, 1996, Theorem 9.17], the concavity of the function being maximized implies that $\boldsymbol{\nu} \mapsto w_\epsilon^\star(\boldsymbol{\nu})$ is convex-valued.

The correspondence $(w, \boldsymbol{\nu}) \mapsto \triangle_K \cap \{w_{a^\star} = \beta\}$ is nonempty compact-valued and continuous (since constant). By Berge's maximum theorem, we get that $\boldsymbol{\nu} \mapsto T_{\epsilon,\beta}^\star(\boldsymbol{\nu})^{-1}$ is continuous on $\mathcal{F}_{a^\star}^K$ and that $w_\beta^\star(\boldsymbol{\nu})$ is upper hemicontinuous with compact values. By [Sundaram, 1996, Theorem 9.17], the concavity of the function being maximized implies that $\boldsymbol{\nu} \mapsto w_{\epsilon,\beta}^\star(\boldsymbol{\nu})$ is convex-valued. □

Lemma 43 provides additional properties on the characteristic times and their optimal allocations. In particular, this results show that the (β-)optimal allocations is unique, has positive allocation for each arm and that the transportation costs are equal at equilibrium. Those properties are key in the analysis of a sampling rule.

Lemma 43. *Let $W_{\epsilon,a,b}$ as in Eq.* (4)*. Let $(T_\epsilon^\star, T_{\epsilon,\beta}^\star)$ and $(w_\epsilon^\star, w_{\epsilon,\beta}^\star)$ as in Eq.* (35)*. Let $\beta \in (0,1)$ and $\boldsymbol{\nu} \in \mathcal{F}^K$ such that $a^\star(\boldsymbol{\nu}) = \{a^\star\}$ is a singleton.*

- *$T_\epsilon^\star(\boldsymbol{\nu})^{-1} > 0$ and $T_{\epsilon,\beta}^\star(\boldsymbol{\nu})^{-1} > 0$.*
- *$\min_{a\in[K]} w_a^\star > 0$ and $\min_{a\in[K]} w_{\beta,a}^\star > 0$ for all $w^\star \in w_\epsilon^\star(\boldsymbol{\nu})$ and $w_\beta^\star \in w_{\epsilon,\beta}^\star(\boldsymbol{\nu})$.*
- *the (β-)optimal allocations are unique and the transportation costs are all equals at equilibrium*

$$w_\epsilon^\star(\boldsymbol{\nu}) = \{w_\epsilon^\star\} \quad \text{and} \quad \forall a \neq a^\star,\ \inf_{u\in[0,1]} \left\{w_{\epsilon,a^\star}^\star d_\epsilon^-(\mu_{a^\star}, u) + w_{\epsilon,a}^\star d_\epsilon^+(\mu_a, u)\right\} = T_\epsilon^\star(\boldsymbol{\nu})^{-1}\,,$$

$$w_{\epsilon,\beta}^\star(\boldsymbol{\nu}) = \{w_{\epsilon,\beta}^\star\} \quad \text{and} \quad \forall a \neq a^\star,\ \inf_{u\in[0,1]} \left\{w_{\epsilon,\beta,a^\star}^\star d_\epsilon^-(\mu_{a^\star}, u) + w_{\epsilon,\beta,a}^\star d_\epsilon^+(\mu_a, u)\right\} = T_{\epsilon,\beta}^\star(\boldsymbol{\nu})^{-1}$$

Proof. Using the definition of the supremum with $1_K/K \in \triangle_K$ and Lemma 35, we obtain

$$\begin{aligned} T_\epsilon^\star(\boldsymbol{\nu})^{-1} &= \sup_{w\in\triangle_K} \min_{a\neq a^\star} \inf_{u\in[0,1]} \left\{w_{a^\star} d_\epsilon^-(\mu_{a^\star}, u) + w_a d_\epsilon^+(\mu_a, u)\right\} \\ &\geq \frac{1}{K} \min_{a\neq a^\star} \inf_{u\in[0,1]} \left\{d_\epsilon^-(\mu_{a^\star}, u) + d_\epsilon^+(\mu_a, u)\right\} > 0\,, \end{aligned}$$

where the last inequality strict uses Lemma 35 and $\mu_a < \mu_{a^\star}$ for all $a \neq a^\star$. Similarly, we can prove that $T_{\epsilon,\beta}^\star(\boldsymbol{\nu})^{-1} > 0$. This concludes the first part of the proof.

We proceed towards contradiction. Suppose that there exists $w^\star \in w_\epsilon^\star(\boldsymbol{\nu})$ and b with $w_b^\star = 0$. Then, we will show $T_\epsilon^\star(\boldsymbol{\nu})^{-1} = 0$, which is a contradiction with the above result. If $b = a^\star$ we have

$$T_\epsilon^\star(\boldsymbol{\nu})^{-1} = \min_{a\neq a^\star} \inf_{u\in[0,1]} w_a^\star d_\epsilon^+(\mu_a, u) \leq \min_{a\neq a^\star} w_a^\star d_\epsilon^+(\mu_a, \mu_a) = 0\,.$$

If $b \neq a^\star$, we have

$$\begin{aligned} T_\epsilon^\star(\boldsymbol{\nu})^{-1} &= \min_{a\neq a^\star} \inf_{u\in[0,1]} \left\{w_{a^\star}^\star d_\epsilon^-(\mu_{a^\star}, u) + w_a^\star d_\epsilon^+(\mu_a, u)\right\} \\ &\leq \inf_{u\in[0,1]} \left\{w_{a^\star}^\star d_\epsilon^-(\mu_{a^\star}, u) + w_b^\star d_\epsilon^+(\mu_b, u)\right\} = \inf_{u\in[0,1]} w_{a^\star}^\star d_\epsilon^-(\mu_{a^\star}, u) = 0\,. \end{aligned}$$

A similar proof allows to show the result for $w_{\epsilon,\beta}^\star(\boldsymbol{\nu})$ by reasoning on $T_{\epsilon,\beta}^\star(\boldsymbol{\nu})^{-1}$. This concludes the second part of the proof.

For notational simplicity, we assume without loss of generality that $a^\star = 1$ is the best arm. At the optimal allocations, all w_a are positive. Let us define $G_b(x) = \inf_{u\in[0,1]} \{d_\epsilon^-(\mu_1, u) + x d_\epsilon^+(\mu_b, u)\}$ for all $b \neq 1$. Let $w^\star \in w_\epsilon^\star(\boldsymbol{\nu})$. Then, we have

$$T_\epsilon^\star(\boldsymbol{\nu})^{-1} = \max_{w\in\triangle_K, w_1>0} w_1 \min_{b\neq 1} G_b\left(\frac{w_b}{w_1}\right) \quad \text{and} \quad w^\star \in \arg\max_{w\in\triangle_K} w_1 \min_{b\neq 1} G_b\left(\frac{w_b}{w_1}\right)\,.$$

Introducing $x_b^\star = \frac{w_b^\star}{w_1^\star}$ for all $b \neq 1$, using that $\sum_{b\in[K]} w_b^\star = 1$, one has

$$w_1^\star = \frac{1}{1+\sum_{c\neq 1} x_c^\star} \quad \text{and} \quad \forall b \neq 1,\ w_b^\star = \frac{x_b^\star}{1+\sum_{c\neq 1} x_c^\star}\ .$$

If $x^\star$ is unique, then so is $w^\star$. Since it is optimal, $\{x_b^\star\}_{b=2}^K \in \mathbb{R}^{K-1}$ belongs to

$$\operatorname*{arg\,max}_{\{x_b\}_{b=2}^K \in \mathbb{R}^{K-1}} \frac{\min_{b\neq 1} G_b\left(x_b\right)}{1+\sum_{c=2}^K x_c}\ . \tag{36}$$

Let's show that all the $G_b\left(x_b^\star\right)$ have to be equal. Let $\mathcal{O} = \{a \in [K]\setminus\{1\} \mid G_a\left(x_a^\star\right) = \min_{b\neq 1} G_b\left(x_b^\star\right)\}$ and $\mathcal{A} = [K]\setminus(\{1\}\cup\mathcal{O})$. Assume that $\mathcal{A}\neq\emptyset$. For all $a\in\mathcal{A}$ and $b\in\mathcal{O}$, one has $G_b\left(x_b^\star\right) > G_a\left(x_a^\star\right)$. Using the continuity of the G_b functions and the fact that they are increasing (Lemma 37), there exists $\epsilon>0$ such that

$$\forall b\in\mathcal{A}, a\in\mathcal{O},\quad G_b\left(x_b^\star - \epsilon/|\mathcal{A}|\right) > G_a\left(x_a^\star + \epsilon/|\mathcal{O}|\right) > G_a\left(x_a^\star\right)\ .$$

We introduce $\bar{x}_b = x_b^\star - \epsilon/|\mathcal{A}|$ for all $b\in\mathcal{A}$ and $\bar{x}_a = x_a^\star + \epsilon/|\mathcal{O}|$ for all $a\in\mathcal{O}$, hence $\sum_{b=2}^K \bar{x}_b = \sum_{b=2}^K x_b^\star$. There exists $a\in\mathcal{O}$ such that $\min_{b\neq 1} G_b\left(\bar{x}_b\right) = G_a\left(x_a^\star+\epsilon/|\mathcal{O}|\right)$, hence

$$\frac{\min_{b\neq 1} G_b\left(\bar{x}_b\right)}{1+\bar{x}_2+\ldots\bar{x}_K} = \frac{G_a\left(x_a^\star+\epsilon/|\mathcal{O}|\right)}{1+x_2^\star+\cdots+x_K^\star} > \frac{G_a\left(x_a^\star\right)}{1+x_2^\star+\cdots+x_K^\star} = \frac{\min_{b\neq 1} G_b\left(x_b^\star\right)}{1+x_2^\star+\cdots+x_K^\star}\ .$$

This is a contradiction with the fact that $x^\star$ belongs to (36). Therefore, we have $\mathcal{A}=\emptyset$.

We have proved that there is a unique value by $y^\star\in\mathbb{R}_+$, such that for all $b\neq 1$, $G_b\left(x_b^\star\right) = y^\star$. Now since G_b is increasing, this defines a unique value for $x_b^\star$, equal to $G_b^{-1}(y^\star)$.

For y in the intersection of the ranges of all G_b, let $x_b(y) = G_b^{-1}(y)$. Then, $y^\star$ belongs to

$$\operatorname*{arg\,max}_{y\in[0,\min_{b\neq 1}\lim_{+\infty} G_b(x))} \frac{y}{1+\sum_{b\neq 1} x_b(y)}\ . \tag{37}$$

For $\beta\in(0,1)$, the same results (and proof) hold for $w_{\epsilon,\beta}^\star(\boldsymbol{\nu})$ by noting that

$$T_{\epsilon,\beta}^\star(\boldsymbol{\nu})^{-1} = \max_{w\in\triangle_K : w_1=\beta} \beta \min_{b\neq 1} G_b\left(w_b/\beta\right)\ .$$

Let $w_{\epsilon,\beta}^\star \in w_{\epsilon,\beta}^\star(\boldsymbol{\nu})$, since we have equality at the equilibrium, we obtain $\beta G_b\left(w_{\epsilon,\beta,b}^\star/\beta\right) = T_{\epsilon,\beta}^\star(\boldsymbol{\nu})^{-1}$ for all $b\neq 1$. Using the inverse mapping x_b, we obtain $w_{\epsilon,\beta,b}^\star = \beta x_b\left(T_{\epsilon,\beta}^\star(\boldsymbol{\nu})^{-1}/\beta\right)$ for all $b\neq 1$. This concludes the third part of the proof. □

Lemma 44 shows that an asymptotically $1/2$-optimal algorithm has an asymptotic expected sample complexity which is at worse twice the asymptotic expected sample complexity of an asymptotically optimal algorithm. This result motivates the recommendation to the practitioner of using $\beta = 1/2$ when no prior information is available on the true instance $\boldsymbol{\nu}$.

Lemma 44. *Let $(T_\epsilon^\star, T_{\epsilon,\beta}^\star, w_\epsilon^\star)$ as in Eq. (35). Let $\beta\in(0,1)$ and $\boldsymbol{\nu}\in\mathcal{F}^K$ such that $a^\star(\boldsymbol{\nu}) = \{a^\star\}$ is a singleton. Then,*

$$T_{\epsilon,1/2}^\star(\boldsymbol{\nu}) \leq 2T_\epsilon^\star(\boldsymbol{\nu}) \quad \textit{and} \quad \frac{T_\epsilon^\star(\boldsymbol{\nu})^{-1}}{T_{\epsilon,\beta}^\star(\boldsymbol{\nu})^{-1}} \leq \max\left\{\frac{\beta^\star}{\beta}, \frac{1-\beta^\star}{1-\beta}\right\} \quad \textit{with} \quad \beta^\star = w_{\epsilon,a^\star}^\star\ .$$

Proof. Define for each non-negative vector $\boldsymbol{\psi}\in\mathbb{R}_+^K$,

$$f(\boldsymbol{\psi}) := \min_{a\neq a^\star} \inf_{u\in[0,1]} \left\{\psi_{a^\star} d_\epsilon^-(\mu_{a^\star}, u) + \psi_a d_\epsilon^+(\mu_a, u)\right\}\ .$$

$T_\epsilon^\star(\boldsymbol{\nu})^{-1}$ is the maximum of $f(\boldsymbol{\psi})$ over probability vectors $\boldsymbol{\psi}\in\triangle_K$. Here, we instead define f for all non-negative vectors, and proceed by varying the total budget of measurement effort available

$\sum_{a\in[K]}\psi_a$. Using Lemma 37, f is non-decreasing in ψ_a for all a. f is homogeneous of degree 1. That is $f(c\boldsymbol{\psi}) = cf(\boldsymbol{\psi})$ for all $c \geq 1$. For each $c_1, c_2 > 0$ define

$$g\left(c_1, c_2\right) = \max\left\{ f(\boldsymbol{\psi}) \mid \boldsymbol{\psi} \in \mathbb{R}_+^K,\ \psi_{a^\star(\boldsymbol{\nu})} = c_1,\ \sum_{a\neq a^\star(\boldsymbol{\nu})} \psi_a \leq c_2,\right\} .$$

The function g inherits key properties of f; it is also non-decreasing and homogeneous of degree 1. We have

$$\begin{aligned} T^\star_{\epsilon,\beta}(\boldsymbol{\nu})^{-1} &= \max\left\{ f(\boldsymbol{\psi}) \mid \boldsymbol{\psi} \in \mathbb{R}_+^K,\ \psi_{a^\star} = \beta,\ \sum_{a\in[K]} \psi_a = 1\right\} \\ &= \max\left\{ f(\boldsymbol{\psi}) \mid \boldsymbol{\psi} \in \mathbb{R}_+^K,\ \psi_{a^\star} = \beta,\ \sum_{a\neq a^\star} \psi_a \leq 1-\beta\right\} = g(\beta, 1-\beta)\,, \end{aligned}$$

where the second equality uses that f is non-decreasing. Similarly, $T^\star_\epsilon(\boldsymbol{\nu})^{-1} = g\left(\beta^\star, 1-\beta^\star\right)$ where $\beta^\star = w^\star_{\epsilon,a^\star}$. Setting $r := \max\left\{\frac{\beta^\star}{\beta}, \frac{1-\beta^\star}{1-\beta}\right\}$ implies $r\beta \geq \beta^\star$ and $r(1-\beta) \geq 1-\beta^\star$. Therefore

$$rT^\star_{\epsilon,\beta}(\boldsymbol{\nu})^{-1} = rg(\beta, 1-\beta) = g(r\beta, r(1-\beta)) \geq g\left(\beta^\star, 1-\beta^\star\right) = T^\star_\epsilon(\boldsymbol{\nu})^{-1}\,.$$

Taking $\beta = \frac{1}{2}$, yields that $T^\star_\epsilon(\boldsymbol{\nu})^{-1} \leq 2\max\{\beta^\star, 1-\beta^\star\}T^\star_{1/2}(\boldsymbol{\nu})^{-1} \leq 2T^\star_{1/2}(\boldsymbol{\nu})^{-1}$. □

Lemma 45 gives sufficient conditions on the means and allocations in order for the transportation costs to be equals to the non-private transportation costs. Moreover, it gives sufficient conditions on the means in order for this equality to hold irrespective of the considered allocation. Taken together, this result allows to have fine and coarse understanding of the separation between the high privacy regime and the low privacy regime for ϵ-global DP BAI.

Lemma 45. *Let $W_{\epsilon,a,b}$ as in Eq.* (4)*. Let $\mu \in (0,1)^K$ such that $a^\star = \arg\max_{a\in[K]} \mu_a$ is unique. Let $w \in (\mathbb{R}_+^\star)^K$. Let $\epsilon > 0$. For all $x \in (0,1)$, we define $f_\epsilon(x) := (1-x)\left(1 - \frac{1}{1+x(e^\epsilon-1)}\right) = (1-x)g_\epsilon^-(x)(1-e^{-\epsilon})$. Let us define $\mu^w_{a^\star,a} := \frac{w_{a^\star}\mu_{a^\star} + w_a\mu_a}{w_{a^\star}+w_a}$ for all $a \neq a^\star$. For all $a \neq a^\star$, we have*

$$\begin{aligned} &\mu_{a^\star} - \mu_a \leq \min\left\{\left(1+\frac{w_{a^\star}}{w_a}\right) f_\epsilon(1-\mu_{a^\star}), \left(1+\frac{w_a}{w_{a^\star}}\right) f_\epsilon(\mu_a)\right\} \\ \implies\quad &W_{\epsilon,a^\star,a}(\mu, w) = w_{a^\star}\mathrm{kl}(\mu_{a^\star}, \mu^w_{a^\star,a}) + w_a\mathrm{kl}(\mu_a, \mu^w_{a^\star,a})\,. \end{aligned}$$

Moreover, we have

$$\max_{a^\star\in[K],\, \mu\in(0,1)^K,\, a^\star(\mu)=\{a^\star\},\, w\in(\mathbb{R}_+^\star)^K} \min\left\{\left(1+\frac{w_{a^\star}}{w_a}\right) f_\epsilon(1-\mu_{a^\star}), \left(1+\frac{w_a}{w_{a^\star}}\right) f_\epsilon(\mu_a)\right\} \leq \epsilon/2$$

and, for all $a \neq a^\star$, we have

$$\begin{aligned} &\epsilon \geq \log\left(\frac{\mu_{a^\star}(1-\mu_a)}{\mu_a(1-\mu_{a^\star})}\right) = \frac{\partial \mathrm{kl}}{\partial x_1}(\mu_{a^\star}, \mu_a) = \frac{\partial \mathrm{kl}}{\partial x_1}(\mu_a, \mu_{a^\star}) \\ \implies\quad &\forall w \in (\mathbb{R}_+^\star)^K,\quad W_{\epsilon,a^\star,a}(\mu, w) = w_{a^\star}\mathrm{kl}(\mu_{a^\star}, \mu^w_{a^\star,a}) + w_a\mathrm{kl}(\mu_a, \mu^w_{a^\star,a})\,. \end{aligned}$$

Proof. Let us define $f_\epsilon(x) = (1-x)\left(1 - \frac{1}{1+x(e^\epsilon-1)}\right)$ for all $x \in (0,1)$. Then, we have

$$\begin{aligned} \frac{\mu_a(1-\mu_a)(e^\epsilon-1)}{1+\mu_a(e^\epsilon-1)} &= (1-\mu_a)\left(1 - \frac{1}{1+\mu_a(e^\epsilon-1)}\right) = f_\epsilon(\mu_a)\,, \\ \frac{\mu_{a^\star}(1-\mu_{a^\star})(e^\epsilon-1)}{e^\epsilon-\mu_{a^\star}(e^\epsilon-1)} &= \mu_{a^\star}\left(1 - \frac{1}{1+(1-\mu_{a^\star})(e^\epsilon-1)}\right) = f_\epsilon(1-\mu_{a^\star})\,. \end{aligned}$$

Using Lemma 24, direct manipulation yields that

$$f_\epsilon(1-\mu_{a^\star}) < \mu_{a^\star} - \mu_a \iff g^+_\epsilon(\mu_{a^\star}) > \mu_a \iff f_\epsilon(\mu_a) < \mu_{a^\star} - \mu_a$$

$$g_\epsilon^-(\mu_a) < \mu^w_{a^\star,a} \quad \iff \quad f_\epsilon(\mu_a) < \frac{w_{a^\star}}{w_{a^\star}+w_a}(\mu_{a^\star}-\mu_a)\,,$$
$$g_\epsilon^+(\mu_{a^\star}) > \mu^w_{a^\star,a} \quad \iff \quad f_\epsilon(1-\mu_{a^\star}) < \frac{w_a}{w_{a^\star}+w_a}(\mu_{a^\star}-\mu_a)\,.$$

Using that $\max\left\{\frac{w_a}{w_{a^\star}+w_a}, \frac{w_{a^\star}}{w_{a^\star}+w_a}\right\} \le 1$, we obtain that

$$\begin{aligned}
&\left(g_\epsilon^-(\mu_a) < \mu_{a^\star} \ \wedge\ g_\epsilon^-(\mu_a) < \mu^w_{a^\star,a}\right) \quad \iff \quad \left(1+\frac{w_a}{w_{a^\star}}\right) f_\epsilon(\mu_a) < \mu_{a^\star}-\mu_a\,,\\
&\left(g_\epsilon^-(\mu_a) < \mu_{a^\star} \ \wedge\ g_\epsilon^+(\mu_{a^\star}) > \mu^w_{a^\star,a}\right) \quad \iff \quad \left(1+\frac{w_{a^\star}}{w_a}\right) f_\epsilon(1-\mu_{a^\star}) < \mu_{a^\star}-\mu_a\,,\\
&\left(g_\epsilon^-(\mu_a) \ge \mu_{a^\star} \ \vee\ \left(g_\epsilon^-(\mu_a) < \mu_{a^\star} \ \wedge\ \mu^w_{a^\star,a} \in [g_\epsilon^+(\mu_{a^\star}), g_\epsilon^-(\mu_a)]\right)\right)\\
&\iff \quad \left(g_\epsilon^-(\mu_a) \ge \mu_{a^\star} \ \vee\ \mu^w_{a^\star,a} \in [g_\epsilon^+(\mu_{a^\star}), g_\epsilon^-(\mu_a)]\right)\\
&\iff \quad \left(\min\{f_\epsilon(\mu_a), f_\epsilon(1-\mu_{a^\star})\} \ge \mu_{a^\star}-\mu_a\right.\\
&\qquad\qquad \left.\vee\, \mu_{a^\star}-\mu_a \le \min\left\{\left(1+\frac{w_{a^\star}}{w_a}\right) f_\epsilon(1-\mu_{a^\star}), \left(1+\frac{w_a}{w_{a^\star}}\right) f_\epsilon(\mu_a)\right\}\right)\\
&\iff \quad \mu_{a^\star}-\mu_a \le \max\left\{\min\{f_\epsilon(\mu_a), f_\epsilon(1-\mu_{a^\star})\},\right.\\
&\qquad\qquad \left.\min\left\{\left(1+\frac{w_{a^\star}}{w_a}\right) f_\epsilon(1-\mu_{a^\star}), \left(1+\frac{w_a}{w_{a^\star}}\right) f_\epsilon(\mu_a)\right\}\right\}\\
&\iff \quad \mu_{a^\star}-\mu_a \le \min\left\{\left(1+\frac{w_{a^\star}}{w_a}\right) f_\epsilon(1-\mu_{a^\star}), \left(1+\frac{w_a}{w_{a^\star}}\right) f_\epsilon(\mu_a)\right\}\,.
\end{aligned}$$

Combining those conditions with Lemma 38 concludes the first part of the proof.

For all $x \in (0,1)$, we have

$$\begin{aligned}
f'_\epsilon(x) &= \frac{(1-x)(e^\epsilon-1) - x(e^\epsilon-1)(1+x(e^\epsilon-1))}{(1+x(e^\epsilon-1))^2} = -(e^\epsilon-1)\frac{x^2(e^\epsilon-1)+2x-1}{(1+x(e^\epsilon-1))^2}\,,\\
f'_\epsilon(x) = 0 \quad &\iff \quad x = \frac{e^{\epsilon/2}-1}{e^\epsilon-1}\,,\\
f''_\epsilon(x) &= -2(e^\epsilon-1)\frac{(1+x(e^\epsilon-1))^2 - (e^\epsilon-1)\left(x^2(e^\epsilon-1)+2x-1\right)}{(1+x(e^\epsilon-1))^3}\\
&= -\frac{2e^\epsilon(e^\epsilon-1)}{(1+x(e^\epsilon-1))^3} \le 0\,.
\end{aligned}$$

As f_ϵ is strictly concave, the maximum is achieved at $\frac{e^{\epsilon/2}-1}{e^\epsilon-1}$ with value

$$\max_{x\in(0,1)} f_\epsilon(x) = f_\epsilon\left(\frac{e^{\epsilon/2}-1}{e^\epsilon-1}\right) = \frac{e^\epsilon-e^{\epsilon/2}}{e^\epsilon-1}\left(1-e^{-\epsilon/2}\right) = \frac{(e^{\epsilon/2}-1)^2}{e^\epsilon-1}\,.$$

Let $\kappa_1(x) = x(e^x-1) - 4(e^{x/2}-1)^2$ for all $x > 0$. Then, we have

$$\frac{(e^{\epsilon/2}-1)^2}{e^\epsilon-1} \le \epsilon/4 \quad \iff \quad \kappa_1(\epsilon) \ge 0\,.$$

Then, we have $\kappa_1(0) = 0$ and

$$\kappa'_1(x) = 4e^{x/2} - 3e^x - 1 + xe^x \quad \text{and} \quad \kappa''_1(x) = e^x\left(2(e^{-x/2}-1)+x\right)\,.$$

Using that $e^{-x/2}-1 \ge -x/2$, we obtain $\kappa''_1(x) \ge 0$. Using that $\kappa'_1(0) = 0$, we obtain $\kappa'_1(x) \ge 0$. Using that $\kappa_1(0) = 0$, we obtain $\kappa_1(x) \ge 0$. Therefore, we have shown that

$$\forall \epsilon > 0, \quad \max_{x\in(0,1)} f_\epsilon(x) \le \epsilon/4\,.$$

Direct manipulation yields that

$$\min\left\{\left(1+\frac{w_{a^\star}}{w_a}\right) f_\epsilon(1-\mu_{a^\star}), \left(1+\frac{w_a}{w_{a^\star}}\right) f_\epsilon(\mu_a)\right\}$$

$$\leq \left(1 + \min\left\{\frac{w_{a^\star}}{w_a}, \frac{w_a}{w_{a^\star}}\right\}\right) \max_{x\in(0,1)} f_\epsilon(x) \leq \epsilon/2\,.$$

Taking the supremum over $w \in (\mathbb{R}_+^\star)^K$, $\mu \in (0,1)^K$ such that $a^\star = a^\star(\mu)$ and over $a^\star \in [K]$ concludes the second part of the proof.

Let $a \neq a^\star$. Direct manipulations yield that

$$\begin{aligned}
&\mu_{a^\star} - \mu_a \leq \min\{f_\epsilon(1-\mu_{a^\star}), f_\epsilon(\mu_a)\} \\
\implies \quad &\forall w \in (\mathbb{R}_+^\star)^K, \quad \mu_{a^\star} - \mu_a \leq \min\left\{\left(1 + \frac{w_{a^\star}}{w_a}\right) f_\epsilon(1-\mu_{a^\star}), \left(1 + \frac{w_a}{w_{a^\star}}\right) f_\epsilon(\mu_a)\right\} \\
\implies \quad &\forall w \in (\mathbb{R}_+^\star)^K, \quad W_{\epsilon,a^\star,a}(\mu, w) = w_{a^\star}\mathrm{kl}(\mu_{a^\star}, \mu^w_{a^\star,a}) + w_a\mathrm{kl}(\mu_a, \mu^w_{a^\star,a})\,.
\end{aligned}$$

Recall that $f_\epsilon(x) = (1-x)\left(1 - \frac{1}{1+x(e^\epsilon-1)}\right)$. Then, we have directly that

$$f_\epsilon(x) \geq y \quad \iff \quad \frac{1-x-y}{1-x} \geq \frac{1}{1+x(e^\epsilon-1)} \quad \iff \quad \frac{(y+x)(1-x)}{x(1-x-y)} \leq e^\epsilon\,.$$

Plugging this result, we obtain

$$\begin{aligned}
\mu_{a^\star} - \mu_a \leq \min\{f_\epsilon(1-\mu_{a^\star}), f_\epsilon(\mu_a)\} \quad &\iff \quad e^\epsilon \geq \frac{\mu_{a^\star}(1-\mu_a)}{\mu_a(1-\mu_{a^\star})} \\
&\iff \quad \epsilon \geq \log\left(\frac{\mu_{a^\star}(1-\mu_a)}{\mu_a(1-\mu_{a^\star})}\right)\,.
\end{aligned}$$

Recall that

$$\frac{\partial \mathrm{kl}}{\partial x_1}(\mu_{a^\star}, \mu_a) = \frac{\partial \mathrm{kl}}{\partial x_1}(\mu_a, \mu_{a^\star}) = \log\left(\frac{\mu_{a^\star}(1-\mu_a)}{\mu_a(1-\mu_{a^\star})}\right)\,.$$

This concludes the proof of the last part of the result. □

Lemma 46 shows that our lower bound is larger (hence better) than the one derived in Azize et al. [2024].

Lemma 46. *Let $T_g^\star(\boldsymbol{\nu}, \epsilon)$ as in Theorem 13 in Azize et al. [2024], and $T_\epsilon^\star(\boldsymbol{\nu})$ as in Eq.* (35)*. Then, we have $T_g^\star(\boldsymbol{\nu}, \epsilon) \leq T_\epsilon^\star(\boldsymbol{\nu})$.*

Proof. Let $T_g^\star(\boldsymbol{\nu}, \epsilon)$ as in Theorem 13 in Azize et al. [2024]. A sufficient condition to obtain $T_g^\star(\boldsymbol{\nu}, \epsilon) \leq T_\epsilon^\star(\boldsymbol{\nu})$ is to show that, for all $\lambda \in \mathrm{Alt}(\mu)$, we have

$$\sum_{a\in[K]} w_a d_\epsilon(\mu_a, \lambda_a) \leq \min\left\{\sum_{a\in[K]} w_a \mathrm{kl}(\mu_a, \lambda_a), 6\epsilon \sum_{a\in[K]} w_a|\mu_a - \lambda_a|\right\}\,,$$

since we can conclude by taking the infimum over $\lambda \in \mathrm{Alt}(\mu)$ and the supremum over $w \in \triangle_K$ on both sides of the inequalities. By definition of d_ϵ and evaluation the function at $z = \mu$ and $z = \lambda$ respectively, we obtain

$$d_\epsilon(\lambda, \mu) = \inf_{z\in(0,1)} \{\mathrm{kl}(z, \mu) + \epsilon|\lambda - z|\} \leq \min\{\mathrm{kl}(\lambda, \mu), \epsilon|\lambda - \mu|\}\,.$$

By summing those inequalities over arms $a \in [K]$, we obtain

$$\begin{aligned}
\sum_{a\in[K]} w_a d_\epsilon(\mu_a, \lambda_a) &\leq \sum_{a\in[K]} w_a \min\{\mathrm{kl}(\mu_a, \lambda_a), \epsilon|\mu_a - \lambda_a|\} \\
&\leq \min\left\{\sum_{a\in[K]} w_a \mathrm{kl}(\mu_a, \lambda_a), \epsilon \sum_{a\in[K]} w_a|\mu_a - \lambda_a|\right\}\,.
\end{aligned}$$

Using that $\sum_{a\in[K]} w_a|\mu_a - \lambda_a| \geq 0$ and $6\epsilon \geq \epsilon$, this concludes the proof. □

In Garivier and Kaufmann [2016], the authors show how to rewrite the optimization problem underlying the characteristic time and its optimal allocation as a simpler optimization problem. Lemma 47 shows that similar properties holds for ϵ-global DP BAI. In particular, it shows that computing the characteristic time $T^\star_\epsilon(\boldsymbol{\nu})$ and their optimal allocation $w^\star_\epsilon(\boldsymbol{\nu})$ can be done explicitly based on solving nested fixed-point equations. This result is key to implement computationally tractable Track-and-Stop algorithms. Additionally, Lemma 47 gives an explicit lower bound on the characteristic time $T^\star_\epsilon(\boldsymbol{\nu})$.

Lemma 47. *Let $d^\pm_\epsilon$ as in Eq.* (3)*, and $(T^\star_\epsilon, w^\star_\epsilon)$ as in Eq.* (35)*. Let $a \neq a^\star$. For $x \in [0, +\infty)$, let*

$$G_a(x) := \inf_{u\in[0,1]} \left\{ d^-_\epsilon(\mu_{a^\star}, u) + x d^+_\epsilon(\mu_a, u) \right\} \text{ and } u_a(x) := \operatorname*{arg\,min}_{u\in[\mu_a,\mu_{a^\star}]} \left\{ d^-_\epsilon(\mu_{a^\star}, u) + x d^+_\epsilon(\mu_a, u) \right\}.$$

- *The function G_a is an increasing and strictly concave one-to-one mapping from $[0, +\infty)$ to $[0, d^-_\epsilon(\mu_{a^\star}, \mu_a))$; it satisfies that $G_a(0) = 0$ and $\lim_{x\to+\infty} G_a(x) = d^-_\epsilon(\mu_{a^\star}, \mu_a)$.*
- *The function u_a is a decreasing one-to-one mapping from $[0, +\infty)$ to $(\mu_a, \mu_{a^\star}]$; it satisfies that $u_a(0) = \mu_{a^\star}$ and $\lim_{x\to+\infty} u_a(x) = \mu_a$.*
- *Let $x_a(y)$ be defined as the unique solution of $G_a(x) = y$ for all $y \in [0, d^-_\epsilon(\mu_{a^\star}, \mu_a))$. The function x_a is an increasing and strictly convex one-to-one mapping from $[0, d^-_\epsilon(\mu_{a^\star}, \mu_a))$ to $[0, +\infty)$; it satisfies that $x_a(0) = 0$ and $\lim_{y\to d^-_\epsilon(\mu_{a^\star},\mu_a)} x_a(y) = +\infty$.*

For all $y \in [0, \min_{a\neq a^\star} d^-_\epsilon(\mu_{a^\star}, \mu_a))$, let us define

$$G(y) := \frac{y}{1 + \sum_{a\neq a^\star} x_a(y)} \quad \text{and} \quad F(y) := \sum_{a\neq a^\star} \frac{d^-_\epsilon(\mu_{a^\star}, u_a(x_a(y)))}{d^+_\epsilon(\mu_a, u_a(x_a(y)))} \,.$$

- *The function F is an increasing one-to-one mapping from $[0, \min_{a\neq a^\star} d^-_\epsilon(\mu_{a^\star}, \mu_a))$ to $[0, +\infty)$; it satisfies that $F(0) = 0$ and $\lim_{y\to\min_{a\neq a^\star} d^-_\epsilon(\mu_{a^\star},\mu_a)} F(y) = +\infty$.*
- *On $[0, \min_{a\neq a^\star} d^-_\epsilon(\mu_{a^\star}, \mu_a))$, the function G is maximized at the unique $y^\star$ solution in $[0, \min_{a\neq a^\star} d^-_\epsilon(\mu_{a^\star}, \mu_a))$ of the fixed-point equation $F(y) = 1$. Moreover, we have $w^\star_\epsilon(\boldsymbol{\nu})_a = w^\star_\epsilon(\boldsymbol{\nu})_{a^\star} x_a(y^\star)$ for all $a \neq a^\star$,*

$$w^\star_\epsilon(\boldsymbol{\nu})_{a^\star} = \frac{1}{1 + \sum_{a\neq a^\star} x_a(y^\star)} \quad \text{and} \quad T^\star_\epsilon(\boldsymbol{\nu})^{-1} = \frac{y^\star}{1 + \sum_{a\neq a^\star} x_a(y^\star)} \,.$$

- *Moreover, we have*

$$T^\star_\epsilon(\boldsymbol{\nu}) \geq \frac{1}{\min_{a\neq a^\star} d^-_\epsilon(\mu_{a^\star}, \mu_a)} + \sum_{a\neq a^\star} \frac{1}{d^+_\epsilon(\mu_a, \mu_{a^\star})} \,.$$

If $\epsilon < \log\left(\frac{\mu_a(1-\mu_b)}{\mu_b(1-\mu_a)}\right)$, we have $d^+_\epsilon(\mu_a, \mu_{a^\star}) = -\log\left(1 - \mu_{a^\star}(1 - e^{-\epsilon})\right) - \epsilon\mu_a$ and $d^-_\epsilon(\mu_{a^\star}, \mu_a) = -\log\left(1 + \mu_a(e^\epsilon - 1)\right) + \epsilon\mu_{a^\star}$.

Proof. Using Lemma 37, we know that G_a is concave. Let $u_a(x) \in \arg\min_{u\in[0,1]} \{d^-_\epsilon(\mu_{a^\star}, u) + x d^+_\epsilon(\mu_a, u)\}$ for all $x \in [0, +\infty)$, whose explicit formula is given in Lemma 45. It is direct to see that $G_a(0) = 0$ and $u_a(0) = \mu_{a^\star}$. Using the optimality condition of $u_a(x)$, we obtain, for all $x \in [0, +\infty)$,

$$\begin{aligned} G'_a(x) &= u'_a(x) \left(\frac{\partial d^-_\epsilon}{\partial u}(\mu_{a^\star}, u_a(x)) + x \frac{\partial d^+_\epsilon}{\partial u}(\mu_a, u_a(x)) \right) + d^+_\epsilon(\mu_a, u_a(x)) \\ &= d^+_\epsilon(\mu_a, u_a(x)) > 0 \,, \end{aligned}$$

where the last inequality is obtained by Lemma 35 and using that $d^+_\epsilon(\mu_a, u_a(0)) = d^+_\epsilon(\mu_a, \mu_{a^\star}) > 0$. Therefore, G_a is an increasing one-to-one mapping from $[0, +\infty)$ to $[0, \lim_{x\to+\infty} G_a(x))$.

Let $\mu^x_{a^\star,a} = \frac{\mu_{a^\star} + x\mu_a}{1+x}$ for all $x \in [0, +\infty)$. It is easy to see that $G_a(0) = 0$, $u_a(0) = \mu_{a^\star}$ and $\lim_{x\to+\infty} \mu^x_{a^\star,a} = \mu_a$. Using Lemma 45, we obtain that

$$\lim_{x\to+\infty} \min\left\{ (1 + 1/x)\, f_\epsilon(1 - \mu_{a^\star}), (1 + x)\, f_\epsilon(\mu_a) \right\} = f_\epsilon(1 - \mu_{a^\star}) \,.$$

When $\mu_{a^\star} - \mu_a \leq f_\epsilon(1-\mu_{a^\star})$, we obtain

$$\begin{aligned}\lim_{x\to+\infty} G_a(x) &= \lim_{x\to+\infty} \left\{\mathrm{kl}(\mu_{a^\star}, \mu^x_{a^\star,a}) + x\mathrm{kl}(\mu_a, \mu^x_{a^\star,a})\right\} \\ &= \mathrm{kl}(\mu_{a^\star}, \mu_a) + \lim_{x\to+\infty} \left\{x\mathrm{kl}(\mu_a, \mu^x_{a^\star,a})\right\} = \mathrm{kl}(\mu_{a^\star}, \mu_a)\,,\end{aligned}$$

where we used that

$$x\mathrm{kl}(\mu_a, \mu^x_{a^\star,a}) = x\left(\mu_a \log\left(1 - \frac{\mu_{a^\star}-\mu_a}{\mu_a(1-\mu_a)}\frac{1}{\frac{\mu_{a^\star}}{\mu_a}+x}\right) + \log\left(1 + \frac{\mu_{a^\star}-\mu_a}{1-\mu_a}\frac{1}{\frac{1-\mu_{a^\star}}{1-\mu_a}+x}\right)\right)\,,$$

$$\lim_{x\to+\infty}\left\{x\mathrm{kl}(\mu_a, \mu^x_{a^\star,a})\right\} = \frac{(\mu_{a^\star}-\mu_a)^2}{\mu_a(1-\mu_a)^2}\lim_{x\to+\infty}\left\{\frac{x}{\left(\frac{\mu_{a^\star}}{\mu_a}+x\right)\left(\frac{1-\mu_{a^\star}}{1-\mu_a}+x\right)}\right\} = 0\,,$$

where we used that $\log(1+x) =_{x\to 0} x + \mathcal{O}(x^2)$. Using Lemma 26 and the proof of Lemma 45, we know that $\mu_{a^\star} - \mu_a \leq f_\epsilon(1-\mu_{a^\star})$ if and only if $\mu_a \in [g^+_\epsilon(\mu_{a^\star}), \mu_{a^\star})$, hence we have $\mathrm{kl}(\mu_{a^\star}, \mu_a) = d^-_\epsilon(\mu_{a^\star}, \mu_a)$. This concludes the proof in the first case.

When $\mu_{a^\star} - \mu_a > f_\epsilon(1-\mu_{a^\star})$, we obtain

$$\lim_{x\to+\infty} G_a(x) = \lim_{x\to+\infty}\left\{-\log\left(1 + u_{1,\star}(\mu_a, x)(e^\epsilon - 1)\right) + \epsilon\mu_{a^\star} + x\mathrm{kl}(\mu_a, u_{1,\star}(\mu_a, x)\right\}\,,$$

where

$$u_{1,\star}(\mu_a, x) = \frac{\sqrt{\left(x(1-\mu_a(e^\epsilon-1)) - (e^\epsilon-1)\right)^2 + 4(1+x)(e^\epsilon-1)x\mu_a} - \left(x(1-\mu_a(e^\epsilon-1)) - (e^\epsilon-1)\right)}{2(1+x)(e^\epsilon-1)}\,.$$

Direct manipulation yields that $\lim_{x\to+\infty} u_{1,\star}(\mu_a, x) = \mu_a$, hence

$$\lim_{x\to+\infty} G_a(x) = -\log\left(1+\mu_a(e^\epsilon-1)\right) + \epsilon\mu_{a^\star} + \lim_{x\to+\infty}\left\{x\mathrm{kl}(\mu_a, u_{1,\star}(\mu_a, x)\right\}\,.$$

Let us denote $v_{1,\star}(\mu_a, x) = u_{1,\star}(\mu_a, x) - \mu_a \geq 0$, i.e., $\lim_{x\to+\infty} v_{1,\star}(\mu_a, x) = 0$. Direct manipulation yields that

$$\begin{aligned}&v_{1,\star}(\mu_a, x) \\ &= \frac{1+\mu_a(e^\epsilon-1)}{2(e^\epsilon-1)}\left(1 - \frac{1}{x+1}\right) \\ &\qquad \left(\sqrt{1 - \frac{2x(1-\mu_a(e^\epsilon+1))(e^\epsilon-1) - (e^\epsilon-1)^2}{x^2(1+\mu_a(e^\epsilon-1))^2}} - 1 + \frac{(e^\epsilon-1)(1-2\mu_a)}{x(1+\mu_a(e^\epsilon-1))}\right) \\ &= \sqrt{1 - \frac{2x(1-\mu_a(e^\epsilon+1))(e^\epsilon-1) - (e^\epsilon-1)^2}{x^2(1+\mu_a(e^\epsilon-1))^2}} - 1 + \frac{(e^\epsilon-1)(1-2\mu_a)}{x(1+\mu_a(e^\epsilon-1))} \\ &=_{x\to+\infty} \frac{(e^\epsilon-1)(1-2\mu_a)}{x(1+\mu_a(e^\epsilon-1))} - \frac{2x(1-\mu_a(e^\epsilon+1))(e^\epsilon-1) - (e^\epsilon-1)^2}{2x^2(1+\mu_a(e^\epsilon-1))^2} + \mathcal{O}(1/x^2) \\ &=_{x\to+\infty} \frac{2x(e^\epsilon-1)2(1-\mu_a)\mu_a(e^\epsilon-1) + (e^\epsilon-1)^2}{2x^2(1+\mu_a(e^\epsilon-1))^2} + \mathcal{O}(1/x^2)\,,\end{aligned}$$

$$\text{hence}\quad v_{1,\star}(\mu_a, x) =_{x\to+\infty} \frac{2(1-\mu_a)\mu_a(e^\epsilon-1)^2}{x(1+\mu_a(e^\epsilon-1))^2} + \mathcal{O}(1/x^2)\,.$$

where we used that $\sqrt{1-x} - 1 =_{x\to 0} -x/2 + \mathcal{O}(x^2)$ to obtain the last result. Similarly as before, we derive

$$\begin{aligned}x\mathrm{kl}(\mu_a, u_{1,\star}(\mu_a, x)) = x\Bigg(&\mu_a \log\left(1 - \frac{1}{\mu_a(1-\mu_a)}\frac{v_{1,\star}(\mu_a, x)}{1 + v_{1,\star}(\mu_a, x)/\mu_a}\right) \\ &+ \log\left(1 + \frac{1}{1-\mu_a}\frac{v_{1,\star}(\mu_a, x)}{1 - v_{1,\star}(\mu_a, x)/(1-\mu_a)}\right)\Bigg)\end{aligned}$$

$$\lim_{x\to+\infty}\left\{x\mathrm{kl}(\mu_a,\mu^x_{a^\star,a})\right\} = \frac{1}{\mu_a(1-\mu_a)^2}\lim_{x\to+\infty}\left\{\frac{xv_{1,\star}(\mu_a,x)^2}{\left(1-\frac{v_{1,\star}(\mu_a,x)}{1-\mu_a}\right)\left(1+\frac{v_{1,\star}(\mu_a,x)}{\mu_a}\right)}\right\}$$
$$= \frac{\lim_{x\to+\infty} xv_{1,\star}(\mu_a,x)^2}{\mu_a(1-\mu_a)^2} = 0\,,$$

where we used that $v_{1,\star}(\mu_a,x) =_{x\to+\infty} \mathcal{O}(1/x)$ to conclude. Therefore, we have shown that $\lim_{x\to+\infty} G_a(x) = -\log\left(1+\mu_a(e^\epsilon-1)\right)+\epsilon\mu_{a^\star}$. Using Lemma 26 and the proof of Lemma 45, we know that $\mu_{a^\star}-\mu_a > f_\epsilon(1-\mu_{a^\star})$ if and only if $\mu_a \in [0, g^+_\epsilon(\mu_{a^\star}))$, hence we have $-\log\left(1+\mu_a(e^\epsilon-1)\right)+\epsilon\mu_{a^\star} = d^-_\epsilon(\mu_{a^\star},\mu_a)$. This concludes the proof in the second case.

Therefore, G_a is a strictly increasing one-to-one mapping from $[0,+\infty)$ to $[0, d^-_\epsilon(\mu_{a^\star},\mu_a))$. Using the implicit function theorem, we obtain

$$\forall x\in[0,+\infty),\quad u'_a(x) = -\frac{\frac{\partial d^+_\epsilon}{\partial u}(\mu_a,u_a(x))}{\frac{\partial^2 d^-_\epsilon}{\partial u^2}(\mu_{a^\star},u_a(x)) + x\frac{\partial^2 d^+_\epsilon}{\partial u^2}(\mu_a,u_a(x))} < 0\,,$$

where the strict inequality is obtained by using properties in Lemmas 25 and 26, since $u_a(x)\in(\mu_a,\mu_{a^\star})$ by Lemmas 35 and 25. Similarly, we obtain

$$\forall x\in[0,+\infty),\quad G''_a(x) = u'_a(x)\frac{\partial d^+_\epsilon}{\partial u}(\mu_a,u_a(x)) > 0\,,$$

Therefore, we have shown that G_a is strictly concave and that u_a is decreasing.

Let us define $x_a(y)$ as the unique solution of $G_a(x)=y$, which is well-defined based on our above computations. Therefore, we have

$$y = d^-_\epsilon(\mu_{a^\star},u_a(x_a(y))) + x_a(y)d^+_\epsilon(\mu_a,u_a(x_a(y)))\,.$$

Using the derivative of the inverse function, we obtain

$$\forall y\in[0,d^-_\epsilon(\mu_{a^\star},\mu_a)),\quad x'_a(y) = \frac{1}{G'_a(x_a(y))} = \frac{1}{d^+_\epsilon(\mu_a,u_a(x_a(y)))} > 0\,,$$

hence x_a is increasing on $[0,d^-_\epsilon(\mu_{a^\star},\mu_a))$. Moreover, we have

$$\forall y\in[0,d^-_\epsilon(\mu_{a^\star},\mu_a)),\quad x''_a(y) = -\frac{u'_a(x_a(y))}{d^+_\epsilon(\mu_a,u_a(x_a(y)))^3}\frac{\partial d^+_\epsilon}{\partial u}(\mu_a,u_a(x_a(y)) > 0\,,$$

hence x_a is strictly convex on $[0,d^-_\epsilon(\mu_{a^\star},\mu_a))$.

For all $y\in[0,\min_{a\neq a^\star} d^-_\epsilon(\mu_{a^\star},\mu_a))$, let us define $G(y) = \frac{y}{1+\sum_{a\neq a^\star}x_a(y)}$ and $F(y) = \sum_{a\neq a^\star}\frac{d^-_\epsilon(\mu_{a^\star},u_a(x_a(y)))}{d^+_\epsilon(\mu_a,u_a(x_a(y)))}$. Using the above results, direct manipulations yield that, for all $y\in[0,\min_{a\neq a^\star} d^-_\epsilon(\mu_{a^\star},\mu_a))$,

$$G'(y) = \frac{1+\sum_{a\neq a^\star}x_a(y) - y\sum_{a\neq a^\star}x'_a(y)}{(1+\sum_{a\neq a^\star}x_a(y))^2} = \frac{1+\sum_{a\neq a^\star}x_a(y) - \sum_{a\neq a^\star}\frac{y}{d^+_\epsilon(\mu_a,u_a(x_a(y)))}}{(1+\sum_{a\neq a^\star}x_a(y))^2}$$
$$= \frac{1-F(y)}{(1+\sum_{a\neq a^\star}x_a(y))^2}\,,$$

hence we obtain that $G'(y)=0$ if and only if $F(y)=1$. Using that $x_a(0)=0$, $u_a(0)=\mu_{a^\star}$ and $d^-_\epsilon(\mu_{a^\star},\mu_{a^\star})=0$, we obtain that $F(0)=0$.

Using that $\lim_{y\to d^-_\epsilon(\mu_{a^\star},\mu_a)}x_a(y) = +\infty$, $\lim_{x\to+\infty}u_a(x)=\mu_a$, $d^-_\epsilon(\mu_{a^\star},\mu_a)>0$ and $d^+_\epsilon(\mu_a,\mu_a)=0$, we obtain that $\lim_{y\to\min_{a\neq a^\star}d^-_\epsilon(\mu_{a^\star},\mu_a)}F(y)=+\infty$.

Let $H(y)=\sum_{a\neq a^\star}\frac{1}{d^+_\epsilon(\mu_a,u_a(x_a(y)))}$ for all $y\in[0,\min_{a\neq a^\star}d^-_\epsilon(\mu_{a^\star},\mu_a))$. Then, we have

$$\sum_{a\neq a^\star}x_a(y) = yH(y)-F(y)\quad,\quad \sum_{a\neq a^\star}x'_a(y) = H(y)\,,$$

$$\frac{\partial d_\epsilon^-}{\partial u}(\mu_{a^\star}, u_a(x)) + x\frac{\partial d_\epsilon^+}{\partial u}(\mu_a, u_a(x)) = 0\,,$$
$$d_\epsilon^-(\mu_{a^\star}, u_a(x_a(y))) + x_a(y) d_\epsilon^+(\mu_a, u_a(x_a(y))) = y\,.$$

Then, for all $y \in [0, \min_{a\neq a^\star} d_\epsilon^-(\mu_{a^\star}, \mu_a))$, we have

$$\begin{aligned}
H'(y) &= -\sum_{a\neq a^\star} \frac{u_a'(x_a(y))x_a'(y)}{(d_\epsilon^+(\mu_a, u_a(x_a(y))))^2}\frac{\partial d_\epsilon^+}{\partial u}(\mu_a, u_a(x_a(y))) \\
&= -\sum_{a\neq a^\star} \frac{u_a'(x_a(y))}{(d_\epsilon^+(\mu_a, u_a(x_a(y))))^3}\frac{\partial d_\epsilon^+}{\partial u}(\mu_a, u_a(x_a(y)))\,, \\
F'(y) &= \sum_{a\neq a^\star} \frac{u_a'(x_a(y))x_a'(y)}{(d_\epsilon^+(\mu_a, u_a(x_a(y))))^2} \\
&\quad \left(d_\epsilon^+(\mu_a, u_a(x_a(y)))\frac{\partial d_\epsilon^-}{\partial u}(\mu_{a^\star}, u_a(x_a(y))) - d_\epsilon^-(\mu_{a^\star}, u_a(x_a(y)))\frac{\partial d_\epsilon^+}{\partial u}(\mu_a, u_a(x_a(y)))\right) \\
&= -\sum_{a\neq a^\star} \frac{u_a'(x_a(y))x_a'(y)}{(d_\epsilon^+(\mu_a, u_a(x_a(y))))^2}\frac{\partial d_\epsilon^+}{\partial u}(\mu_a, u_a(x_a(y))) \\
&\qquad \big(x_a(y) d_\epsilon^+(\mu_a, u_a(x_a(y))) + d_\epsilon^-(\mu_{a^\star}, u_a(x_a(y)))\big) \\
&= -y\sum_{a\neq a^\star} \frac{u_a'(x_a(y))x_a'(y)}{(d_\epsilon^+(\mu_a, u_a(x_a(y))))^2}\frac{\partial d_\epsilon^+}{\partial u}(\mu_a, u_a(x_a(y))) = yH'(y)\,,
\end{aligned}$$

Therefore, showing that H is increasing is a sufficient condition to show that F is increasing. Using the above results, we have, for all $a \neq a^\star$ and all $y \in [0, \min_{a\neq a^\star} d_\epsilon^-(\mu_{a^\star}, \mu_a))$, we have

$$\frac{1}{(d_\epsilon^+(\mu_a, u_a(x_a(y))))^3}\frac{\partial d_\epsilon^+}{\partial u}(\mu_a, u_a(x_a(y))) > 0 \quad \text{and} \quad u_a'(x_a(y)) < 0\,.$$

Therefore, H is increasing as a summation of increasing function, hence F is increasing.

Let $y^\star$ such that $F(y^\star) = 1$. Reusing the above manipulation, we obtain

$$\begin{aligned}
G''(y) &= -\frac{F'(y)(1 + \sum_{a\neq a^\star} x_a(y)) + 2(1 - F(y))\sum_{a\neq a^\star} x_a'(y)}{(1 + \sum_{a\neq a^\star} x_a(y))^3} \\
&= -\frac{yH'(y)(1 + yH(y) - F(y)) + 2(1 - F(y))H(y)}{(1 + yH(y) - F(y))^3}\,, \\
G''(y^\star) &= -\frac{H'(y^\star)}{y^\star H(y^\star)^2} < 0\,,
\end{aligned}$$

Therefore, $y^\star$ is the unique maximum of G. We conclude this part of the proof by using the intermediate results in the proof of Lemma 43.

By strict convexity of x_a and using its properties proven above, we obtain

$$x_a(y) \geq x_a(0) + yx_a'(0) = \frac{y}{d_\epsilon^+(\mu_a, \mu_{a^\star})}\,.$$

Summing those inequalities, we obtain

$$\forall y \in [0, \min_{a\neq a^\star} d_\epsilon^-(\mu_{a^\star}, \mu_a)), \quad G(y) = \frac{y}{1 + \sum_{a\neq a^\star} x_a(y)} \leq \frac{1}{\frac{1}{y} + \sum_{a\neq a^\star} \frac{1}{d_\epsilon^+(\mu_a, \mu_{a^\star})}}\,.$$

Using that $y \mapsto 1/(1/y + \alpha)$ is increasing for $\alpha > 0$, we obtain that

$$T_\epsilon^\star(\boldsymbol{\nu})^{-1} = \max_{y\in[0, \min_{a\neq a^\star} d_\epsilon^-(\mu_{a^\star}, \mu_a))} G(y) \leq \frac{1}{\frac{1}{\min_{a\neq a^\star} d_\epsilon^-(\mu_{a^\star}, \mu_a))} + \sum_{a\neq a^\star} \frac{1}{d_\epsilon^+(\mu_a, \mu_{a^\star})}}\,.$$

This concludes the proof of the second to last result. The last result is obtained by combining Lemmas 26 and 25 and the derivation in the proof of Lemma 45. □

Lemma 48 is a technical result used in the proof of sufficient exploration of our sampling rule.

Lemma 48. *Let $d_\epsilon^\pm$ as in Eq.* (3)*. Let $\mu \in (0,1)^K$. There exists $\alpha > 0$ such that*

$$C_\mu := \min_{(a,b):\mu_a > \mu_b} \inf_{\substack{\lambda_a,\lambda_b: \\ \max_{c\in\{a,b\}} |\mu_c - \lambda_c| \leq \alpha}} \inf_{u\in[0,1]} \left\{ d_\epsilon^-(\lambda_a, u) + d_\epsilon^+(\lambda_b, u) \right\} > 0 \,. \tag{38}$$

Proof. Using Lemma 35 for $w_1 = w_2 = 1$, the function $\mu \mapsto \inf_{u\in[0,1]} \left\{ d_\epsilon^-(\mu_a, u) + d_\epsilon^+(\mu_b, u) \right\}$ is continuous on $\mathcal{F}^K$. Since it has strictly positive values when $\mu_a > \mu_b$ (Lemma 35), there exists α such that

$$\inf_{\substack{\lambda_a,\lambda_b: \\ \max_{c\in\{a,b\}} |\mu_c - \lambda_c| \leq \alpha}} \inf_{u\in[0,1]} \left\{ d_\epsilon^-(\lambda_a, u) + d_\epsilon^+(\lambda_b, u) \right\} > 0 \,.$$

Further lower bounding by a finite number of strictly positive constants yields the result. □

Lemma 48 is a technical result used in the proof of convergence towards the optimal allocation of our sampling rule.

Lemma 49. *Let $d_\epsilon^\pm$ as in Eq.* (3)*. Let $(\phi_1, \phi_2) \in (0,1)^2$. Let $\mathcal{I}_a := \left\{ \mu \in (0,1)^K \mid a \in a^\star(\mu) \right\}$ for all $a \in [K]$. For all $a^\star \in [K]$, all $\mu \in \mathcal{I}_{a^\star}$, all $(a,b) \in ([K] \setminus \{a^\star\})^2$ such that $a \neq b$, and all $\beta \in [0,1]$, define*

$$G_{a,b}(\mu, \beta) := \inf_{u\in[0,1]} \left\{ \beta d_\epsilon^-(\mu_{a^\star}, u) + \phi_1 d_\epsilon^+(\mu_a, u) \right\} - \inf_{u\in(0,1)} \left\{ \beta d_\epsilon^-(\mu_{a^\star}, u) + \phi_2 d_\epsilon^+(\mu_b, u) \right\} \,.$$

The function $(\mu, \beta) \mapsto G_{a,b}(\mu, \beta)$ is continuous on $(0,1)^K \times [0,1]$. For all $\xi > 0$, the function $(\mu, \beta) \mapsto \inf_{\tilde{\beta}:|\beta - \tilde{\beta}| \leq \xi} G_{a,b}(\mu, \beta)$ is continuous on $(0,1)^K$.

Proof. Since $\bigcup_{a\in[K]} \mathcal{I}_a^K = (0,1)^K$, it is enough to show the property for all $a \in [K]$. Let $a^\star \in [K]$, $\mu \in \mathcal{I}_{a^\star}$, $(a,b) \in ([K] \setminus \{a^\star\})^2$ such that $a \neq b$. As done in Lemma 42 by using Lemma 35, we obtain that the function $(\mu, \beta) \mapsto G_{a,b}(\mu, \beta)$ is continuous on $\mathcal{I}_{a^\star} \times [0,1]$ for all $a^\star \in [K]$, hence on $(0,1)^K \times [0,1]$. Let $\Phi : \mu \mapsto \left\{ \tilde{\beta} : |\beta - \tilde{\beta}| \leq \xi \right\}$, it is a continuous (constant), compact valued and non-empty correspondence. Using the above continuity, Berge's theorem yields that $\mu \mapsto \inf_{\tilde{\beta}:|\beta - \tilde{\beta}| \leq \xi} G_{a,b}(\mu, \tilde{\beta})$ is continuous on $(0,1)^K$. □

H Asymptotic Upper Bound on the Expected Sample complexity

Let $\boldsymbol{\nu}$ be a Bernoulli instance with means $\mu \in (0,1)^2$ and unique best arm $a^\star \in [K]$, i.e., $\arg\max_{a\in[K]} \mu_a = \{a^\star\}$. Let $\beta \in (0,1)$. Let $w^\star_{\epsilon,\beta}(\boldsymbol{\nu}) = \{w^\star_{\epsilon,\beta}\}$ be the unique β-optimal allocation defined in Eq. (35), which satisfies $\min_{a\in[K]} w^\star_{\epsilon,\beta,a} > 0$ by Lemma 43. At equilibrium, we have equality of the transportation costs by Lemma 43, namely

$$\forall a \neq a^\star, \quad W_{\epsilon,a^\star,a}(\mu, w^\star_{\epsilon,\beta}) = T^\star_{\epsilon,\beta}(\boldsymbol{\nu})^{-1} \,, \tag{39}$$

where $W_{\epsilon,a,b}$ is defined in Eq. (4) and $T^\star_{\epsilon,\beta}$ is defined in Eq. (35).

Let $\gamma > 0$. Let $\omega \in \triangle_K$ be any allocation over arms such that $\min_a \omega_a > 0$. We denote by $T_\gamma(\omega)$ the *convergence time* towards ω, which is a random variable quantifying the number of samples required for the global empirical allocations $N_n/(n-1)$ to be γ-close to ω for any subsequent time, namely

$$T_\gamma(\omega) := \inf \left\{ T \geq 1 \mid \forall n \geq T, \ \left\| \frac{N_n}{n-1} - \omega \right\|_\infty \leq \gamma \right\} \,. \tag{40}$$

The proof of Theorem 7 follows the same analysis as the unified analysis of Top Two algorithms, see, e.g., Jourdan et al. [2022]. Appendix H is organised as follows. After recalling some technical results (Appendix H.1), we prove sufficient exploration of our sampling rule (Appendix H.2). Second, we prove that convergence time towards the β-optimal allocation of our sampling rule (Appendix H.3) has finite expectation. Finally, we conclude the proof of Theorems 7 (Appendix H.4).

H.1 Technical Results from the Literature

Lemma 50 relates the global counts $(N_{n,a})_{a\in[K]}$ and the local counts $(\tilde{N}_{n,a})_{a\in[K]}$.

Lemma 50. *Let $\eta > 0$ be the geometric parameter used for the geometric grid update of our private empirical mean estimator. For all $(a,k) \in [K]\times\mathbb{N}$ s.t. $\mathbb{E}_{\boldsymbol{\nu}\pi}[T_k(a)] < +\infty$, $N_{T_k(a),a} = \tilde{N}_{k,a} = \lceil(1+\eta)^{k-1}\rceil$. For all $a\in[K]$ and all $n\in\mathbb{N}$, $N_{n,a} \geq \tilde{N}_{n,a} \geq N_{n,a}/(1+\eta)$.*

Proof. Let $a \in [K]$. After initialisation, we have $k = 1$, $T_1(a) = K+1$ and $N_{T_1(a),a} = 1$. Using the definition of the phase switch, it is direct to see that $N_{T_2(a),a} = \tilde{N}_{2,a} = \lceil 1+\eta\rceil$ when $\mathbb{E}_{\boldsymbol{\nu}\pi}[T_2(a)] < +\infty$. Similarly, we obtain $N_{T_k(a),a} = \tilde{N}_{k,a} = \lceil(1+\eta)^{k-1}\rceil$ when $\mathbb{E}_{\boldsymbol{\nu}\pi}[T_k(a)] < +\infty$. The last result is a direct consequence of the definition of the per-arm geometric update grid. □

Lemma 51 controls the deviation $N^a_{n,a} - \beta L_{n,a}$ enforced by the tracking procedure.

Lemma 51 (Lemma 2.2 in Jourdan and Degenne [2024])**.** *For all $n > K$ and all $a\in[K]$, $-1/2 \leq N^a_{n,a} - \beta L_{n,a} \leq 1$.*

Lemma 52 gathers properties on the $\overline{W}_{-1}$ function used in the stopping threshold.

Lemma 52 (Jourdan et al. [2023])**.** *Let $\overline{W}_{-1}(x) := -W_{-1}(-e^{-x})$ for all $x \geq 1$, where W_{-1} is the negative branch of the Lambert W function. The function $\overline{W}_{-1}$ is increasing on $(1,+\infty)$ and strictly concave on $(1,+\infty)$. In particular, $\overline{W}'_{-1}(x) = \left(1 - \frac{1}{\overline{W}_{-1}(x)}\right)^{-1}$ for all $x > 1$. Then, for all $y \geq 1$ and $x \geq 1$,*

$$\overline{W}_{-1}(y) \leq x \quad \iff \quad y \leq x - \log(x)\,.$$

Moreover, for all $x > 1$,

$$x + \log(x) \leq \overline{W}_{-1}(x) \leq x + \log(x) + \min\left\{\frac{1}{2}, \frac{1}{\sqrt{x}}\right\}\,.$$

Lemma 53 gives an upper bound on a time define implicit as a function of $\overline{W}_{-1}$, namely it is an inversion result.

Lemma 53 (Lemma 32 in Azize et al. [2024])**.** *Let $\overline{W}_{-1}$ defined in Lemma 52. Let $A > 0$, $B > 0$ such that $B/A + \log A > 1$ and $C(A,B) = \sup\{x \mid x < A\log x + B\}$. Then, $C(A,B) < h_1(A,B)$ with $h_1(z,y) = z\overline{W}_{-1}(y/z + \log z)$.*

Lemma 54 shows that upon correction the supremum of sub-exponential random variables is also a sub-exponential random variable.

Lemma 54 (Lemma 72 in Jourdan et al. [2022])**.** *Suppose that $(X_n)_{n\geq1}$ are sub-exponential random variables with constants (C_n), such that $c := \inf_n C_n > 0$. Then $\sup_n(X_n/\log(e+n))$ is sub-exponential.*

Lemma 55 gives a coarse convergence rate of the private empirical estimators of the means towards their true means.

Lemma 55. *There exist sub-exponential random variable W_μ such that almost surely, for all $a\in[K]$ and all n such that $\tilde{N}_{n,a} \geq 1$,*

$$\tilde{N}_{n,a}|\tilde{\mu}_{n,a} - \mu_a| \leq W_\mu \log(e + \tilde{N}_{n,a})\,.$$

In particular, any random variable which is polynomial in W_μ has a finite expectation.

Proof. Let us define

$$W_\mu = \max_{a\in[K]} \sup_{n\in\mathbb{N}} \frac{\tilde{N}_{n,a}|\tilde{\mu}_{n,a} - \mu_a|}{\log(e+\tilde{N}_{n,a})}\,.$$

Let $a\in[K]$. Let us define the geometric grid $N_k = \lceil(1+\eta)^{k-1}\rceil$ for all $k\in\mathbb{N}$, on which we effectively need to control the concentration. The maximum of a finite number of sub-exponential

random variables is sub-exponential. Therefore, using the geometric update grid, it suffices to show that

$$\sup_{k\in\mathbb{N}} \frac{N_k|(Z_{N_k}+S_k)/N_k-\mu_a|}{\log(e+N_k)}$$

is sub-exponential, where Z_{N_k} is the cumulative sum of N_k i.i.d. observations from $\mathrm{Ber}(\mu_a)$ and S_k is the cumulative sum of k i.i.d. observations from $\mathrm{Lap}(1/\epsilon)$.

Using that $Z_{N_k}-N_k\mu_a$ is sub-Gaussian and S_k is sub-exponential, for a fixed k, $|Z_{N_k}-N_k\mu_a+S_k|$ is sub-exponential. Applying Lemma 54, we obtain that

$$\sup_{k\in\mathbb{N}} \frac{N_k|(Z_{N_k}+S_k)/N_k-\mu_a|}{\log(e+N_k)}$$

is sub-exponential. We finally obtain that the maximum over the finitely many arms has the same property. □

H.2 Sufficient Exploration

The first step of in the generic analysis of Top Two algorithms [Jourdan et al., 2022] consists in showing sufficient exploration. The main idea is that, if there are still undersampled arms, either the leader or the challenger will be among them. Therefore, after a long enough time, no arm can still be undersampled. We emphasise that there are multiple ways to select the leader/challenger pair in order to ensure sufficient exploration. Therefore, other choices of leader/challenger pair would yield similar results.

Given an arbitrary phase $p\in\mathbb{N}$, we define the sampled enough set, i.e., the arms having reached phase p, and the arm with highest mean in this set (when not empty) as

$$S_n^p=\{a\in[K]\mid N_{n,a}\geq(1+\eta)^{p-1}\}\quad\text{and}\quad a_n^\star=\operatorname*{arg\,max}_{a\in S_n^p}\mu_a\,. \tag{41}$$

Since $\min_{a\neq b}|\mu_a-\mu_b|>0$, $a_n^\star$ is unique. Let $p\in\mathbb{N}$ such that $(p-1)/4\in\mathbb{N}$. We define the highly and the mildly under-sampled sets as

$$U_n^p:=\{a\in[K]\mid N_{n,a}<(1+\eta)^{(p-1)/2}\}\quad\text{and}\quad V_n^p:=\{a\in[K]\mid N_{n,a}<(1+\eta)^{3(p-1)/4}\}\,. \tag{42}$$

Those arms have not reached phase $(p-1)/2$ and phase $3(p-1)/4$, respectively.

Lemma 56 shows that, when the leader is sampled enough, it is the arm with highest true mean among the sampled enough arms.

Lemma 56. *Let S_n^p and $a_n^\star$ as in (41). There exists p_0 with $\mathbb{E}_{\boldsymbol{\nu}\pi}[\exp(\alpha p_0)]<+\infty$ for all $\alpha>0$ such that if $p\geq p_0$, for all n such that $S_n^p\neq\emptyset$, we have*

- *For all $a\in S_n^p$, we have $\tilde{\mu}_{n,a}\in(0,1)$ and $a_n^\star=\arg\max_{a\in S_n^p}\tilde{\mu}_{n,a}$.*
- *If $B_n\in S_n^p$, then $B_n=a_n^\star$.*

Proof. Let p_0 to be specified later. Let $p\geq p_0$. Let $n\in\mathbb{N}$ such that $S_n^p\neq\emptyset$, where S_n^p and $a_n^\star$ as in Equation (41). Since $N_{n,a}\geq(1+\eta)^{p-1}$ for all $a\in S_n^p$, we have $N_{n,a}\geq(1+\eta)^{p-1}$. Using Lemma 55 and $x\to\log(e+x)/x$ is decreasing, we obtain that

$$\tilde{\mu}_{n,a_n^\star}\geq\mu_{a_n^\star}-W_\mu\frac{\log(e+(1+\eta)^{p-1})}{(1+\eta)^{p-1}}\,,$$

$$\forall a\in S_n^p\setminus\{a_n^\star\},\quad\tilde{\mu}_{n,a}\leq\mu_a+W_\mu\frac{\log(e+(1+\eta)^{p-1})}{(1+\eta)^{p-1}}\,.$$

Let $\overline{\Delta}_{\min}=\min_{a\neq b}|\mu_a-\mu_b|$ and $\Delta_0=\min_{a\in[K]}\min\{\mu_a,1-\mu_a\}>0$. By assumption on the considered instances, we know that $\overline{\Delta}_{\min}>0$. Let $p_1=\lceil\log_{1+\eta}(X_1-e)\rceil+1$ with

$$\begin{aligned}X_1&=\sup\left\{x>1\mid\ x\leq 4(\min\{\overline{\Delta}_{\min},\Delta_0\})^{-1}W_\mu\log x+e\right\}\\&\leq h_1(4(\min\{\overline{\Delta}_{\min},\Delta_0\})^{-1}W_\mu,\ e)\,,\end{aligned}$$

where we used Lemma 53, and h_1 defined therein. Then, for all $p \in \mathbb{N}$ such that $p \geq p_1 + 1$ and all $n \in \mathbb{N}$ such that $S_n^p \neq \emptyset$, we have

$$\forall a \in S_n^p, \quad \mu_a - \min\{\overline{\Delta}_{\min}, \Delta_0\}/4 \leq \tilde{\mu}_{n,a} \leq \mu_a + \min\{\overline{\Delta}_{\min}, \Delta_0\}/4 \,.$$

Therefore, we have $\tilde{\mu}_{n,a} \in (0,1)$ for all $a \in S_n^p$. Since $\tilde{\mu}_{n,a_n^\star} \geq \mu_{a_n^\star} - \min\{\overline{\Delta}_{\min}, \Delta_0\}/4$ and $\tilde{\mu}_{n,a} \leq \mu_a + \min\{\overline{\Delta}_{\min}, \Delta_0\}/4$ for all $a \in S_n^p \setminus \{a_n^\star\}$, we obtain $a_n^\star = \arg\max_{a \in S_n^p} \tilde{\mu}_{n,a}$ since $\arg\max_{a \in S_n^p} \tilde{\mu}_{n,a}$ is unique. The leader is defined as $B_n = \arg\max_{a \in [K]} [\tilde{\mu}_{n,a}]_0^1$. If $B_n \in S_n^p$, we obtain

$$B_n = \arg\max_{a \in S_n^p} [\tilde{\mu}_{n,a}]_0^1 = \arg\max_{a \in S_n^p} \tilde{\mu}_{n,a} = a_n^\star \,.$$

For all $\alpha \in \mathbb{R}_+$, we have $\exp(\alpha p_1) \leq e^{3\alpha}(X_1 - e)^{\alpha/\log 2}$, hence $\mathbb{E}_{\boldsymbol{\nu}\pi}[\exp(\alpha p_1)] < +\infty$ by using Lemma 55 and $h_1(x,e) \sim_{x \to +\infty} x \log x$ to obtain that $\exp(\alpha p_1)$ is at most polynomial in W_μ. Taking $p_0 = p_1$ concludes the proof. □

Lemma 57 shows that the transportation costs between the sampled enough arms with largest true means and the other sampled enough arms are increasing fast enough.

Lemma 57. *Let S_n^p as in Eq. (41). There exists p_1 with $\mathbb{E}_{\boldsymbol{\nu}\pi}[\exp(\alpha p_1)] < +\infty$ for all $\alpha > 0$ such that if $p \geq p_1$, for all n such that $S_n^p \neq \emptyset$, for all $(a,b) \in (S_n^p)^2$ such that $\mu_a > \mu_b$, we have*

$$W_{\epsilon,a,b}(\tilde{\mu}_n, N_n) \geq (1+\eta)^{p-1} C_\mu \,,$$

where $C_\mu > 0$ is a problem dependent constant.

Proof. Let p_0 as in Lemma 56. Let $p \geq p_0$. Let $n \in \mathbb{N}$ such that $S_n^p \neq \emptyset$, where S_n^p as in Eq. (41). Since $N_{n,a} \geq (1+\eta)^{p-1}$ for all $a \in S_n^p$, we have $\tilde{N}_{n,a} \geq (1+\eta)^{p-1}$ by using Lemma 50. Let $(a,b) \in (S_n^p)^2$ such that $\mu_a > \mu_b$. Using Lemma 48, there exists $\alpha_\mu > 0$ such that

$$C_\mu = \min_{(a,b):\mu_a > \mu_b} \inf_{\substack{\lambda_a, \lambda_b: \\ \max_{c \in \{a,b\}} |\mu_c - \lambda_c| \leq \alpha_\mu}} \inf_{u \in [0,1]} \left\{ d_\epsilon^-(\lambda_a, u) + d_\epsilon^+(\lambda_b, u) \right\} > 0 \,.$$

Let $\eta > 0$ s.t. $\eta < \frac{1}{4}\min\{\overline{\Delta}_{\min}, \Delta_0, \alpha_\mu\}$ where $\overline{\Delta}_{\min} = \min_{a \neq b} |\mu_a - \mu_b|$ and $\Delta_0 = \min_{a \in [K]} \min\{\mu_a, 1 - \mu_a\}$. Similarly as in the proof of Lemma 56, we can construct p_2 with $\mathbb{E}_{\boldsymbol{\nu}\pi}[\exp(\alpha p_2)] < +\infty$ for all $\alpha > 0$ such that if $p \geq p_2$, for all n such that $S_n^p \neq \emptyset$, we have $|\tilde{\mu}_{n,a} - \mu_a| \leq \eta$ for all $a \in S_n^p$. Therefore, we have $\tilde{\mu}_{n,a} = [\tilde{\mu}_{n,a}]_0^1$ and $[\tilde{\mu}_{n,b}]_0^1 = \tilde{\mu}_{n,b}$. Moreover, we have $\tilde{\mu}_{n,a} \geq \mu_a - \eta > \mu_b + \eta \geq \tilde{\mu}_{n,b}$. Then, we obtain

$$\begin{aligned}
W_{\epsilon,a,b}(\tilde{\mu}_n, N_n) &= \inf_{u \in [0,1]} \left\{ N_{n,a} d_\epsilon^-(\tilde{\mu}_{n,a}, u) + N_{n,b} d_\epsilon^+(\tilde{\mu}_{n,b}, u) \right\} \\
&\geq (1+\eta)^{p-1} \inf_{u \in [0,1]} \left\{ d_\epsilon^-(\tilde{\mu}_{n,a}, u) + d_\epsilon^+(\tilde{\mu}_{n,b}, u) \right\} \\
&\geq (1+\eta)^{p-1} \inf_{\substack{\lambda_a, \lambda_b: \\ \max_{c \in \{a,b\}} |\mu_c - \lambda_c| \leq \alpha_\mu}} \inf_{u \in [0,1]} \left\{ d_\epsilon^-(\lambda_a, u) + d_\epsilon^+(\lambda_b, u) \right\} \geq (1+\eta)^{p-1} C_\mu \,.
\end{aligned}$$

This concludes the proof. □

Lemma 58 shows that the transportation costs between sampled enough arms and undersampled arms are not increasing too fast.

Lemma 58. *Let S_n^p be as in Eq. (41). There exists p_2 with $\mathbb{E}_{\boldsymbol{\nu}\pi}[\exp(\alpha p_2)] < +\infty$ for all $\alpha > 0$ such that if $p \geq p_2$, for all n such that $S_n^p \neq \emptyset$, For all $p \geq p_2$ and all n such that $S_n^p \neq \emptyset$, for all $a \in S_n^p$ and $b \notin S_n^p$,*

$$W_{\epsilon,a,b}(\tilde{\mu}_n, N_n) \leq (1+\eta)^{p-1} D_\mu \,,$$

where $D_\mu \in (0, +\infty)$ is a problem dependent constant.

Proof. Let $n \in \mathbb{N}$ such that $S_n^p \neq \emptyset$, where S_n^p as in Eq. (41). Since $N_{n,a} \geq (1+\eta)^{p-1}$ for all $a \in S_n^p$, we have $\tilde{N}_{n,a} \geq (1+\eta)^{p-1}$ by using Lemma 50. Likewise, $N_{n,b} < (1+\eta)^{p-1}$ for all

$b \notin S_n^p$, we have $\tilde{N}_{n,b} < (1+\eta)^{p-1}$. Let $a \in S_n^p$ and $b \notin S_n^p$. Since the result is direct when $[\tilde{\mu}_{n,a}]_0^1 \leq [\tilde{\mu}_{n,b}]_0^1$, we assume $[\tilde{\mu}_{n,a}]_0^1 > [\tilde{\mu}_{n,b}]_0^1$ in the following.

Let $\eta > 0$ s.t. $\eta < \frac{1}{4}\min\{\overline{\Delta}_{\min}, \Delta_0\}$ where $\overline{\Delta}_{\min} = \min_{a \neq b} |\mu_a - \mu_b|$ and $\Delta_0 = \min_{a \in [K]} \min\{\mu_a, 1-\mu_a\} > 0$. Similarly as in the proof of Lemma 56, we can construct p_2 with $\mathbb{E}_{\boldsymbol{\nu}\pi}[\exp(\alpha p_2)] < +\infty$ for all $\alpha > 0$ such that if $p \geq p_2$, for all n such that $S_n^p \neq \emptyset$, we have $|\tilde{\mu}_{n,a} - \mu_a| \leq \eta$ for all $a \in S_n^p$. Let $g_\epsilon^+(x) = \frac{x}{x(1-e^\epsilon)+e^\epsilon}$ as in Lemma 24. Using Lemma 24, for all $a \in S_n^p$, we have

$$1 > \mu_a + \min\{\overline{\Delta}_{\min}, \Delta_0\}/4 \geq \tilde{\mu}_{n,a} > g_\epsilon^+(\tilde{\mu}_{n,a}) \geq g_\epsilon^+(\mu_a - \min\{\overline{\Delta}_{\min}, \Delta_0\}/4) > 0\,.$$

Taking $u = \tilde{\mu}_{n,a} \in [0,1]$ and using that $d_\epsilon^-(\tilde{\mu}_{n,a}, \tilde{\mu}_{n,a}) = 0$, we obtain

$$\begin{aligned} W_{\epsilon,a,b}(\tilde{\mu}_n, N_n) &= \inf_{u \in [0,1]} \left\{ N_{n,a} d_\epsilon^-(\tilde{\mu}_{n,a}, u) + N_{n,b} d_\epsilon^+(\tilde{\mu}_{n,b}, u) \right\} \\ &\leq N_{n,b} d_\epsilon^+(\tilde{\mu}_{n,b}, \tilde{\mu}_{n,a}) \leq (1+\eta)^{p-1} d_\epsilon^+(\tilde{\mu}_{n,b}, \tilde{\mu}_{n,a})\,, \end{aligned}$$

where the last term is positive since $\tilde{\mu}_{n,a} > [\tilde{\mu}_{n,b}]_0^1$ and $\tilde{\mu}_{n,a} \in (0,1)$ by Lemma 25.

When $\tilde{\mu}_{n,b} \leq 0$, Lemma 25 yields that

$$d_\epsilon^+(\tilde{\mu}_{n,b}, \tilde{\mu}_{n,a}) = -\log\left(1 - \tilde{\mu}_{n,a}(1-e^{-\epsilon})\right) \leq \epsilon\,,$$

where we used that $x \to -\log(1 - x(1-e^{-\epsilon}))$ is increasing on $(0,1)$. When $\tilde{\mu}_{n,b} \in (0, g_\epsilon^+(\tilde{\mu}_{n,a}))$, Lemma 25 yields that

$$d_\epsilon^+(\tilde{\mu}_{n,b}, \tilde{\mu}_{n,a}) = -\log\left(1 - \tilde{\mu}_{n,a}(1-e^{-\epsilon})\right) - \epsilon\tilde{\mu}_{n,b} \leq \epsilon\,.$$

When $\tilde{\mu}_{n,b} \in [g_\epsilon^+(\tilde{\mu}_{n,a}), \tilde{\mu}_{n,a})$, Lemma 25 yields that

$$\begin{aligned} d_\epsilon^+(\tilde{\mu}_{n,b}, \tilde{\mu}_{n,a}) &= \mathrm{kl}(\tilde{\mu}_{n,b}, \tilde{\mu}_{n,a}) \leq -\log\min\{\tilde{\mu}_{n,a}, 1-\tilde{\mu}_{n,a}\} \\ &\leq -\log\min\{\mu_a - \min\{\overline{\Delta}_{\min}, \Delta_0\}/4, 1 - \mu_a - \min\{\overline{\Delta}_{\min}, \Delta_0\}/4\}\,, \end{aligned}$$

where we used the classical result that $\mathrm{kl}(q,p) \leq -\log\min\{p, 1-p\}$. Let us define

$$D_\mu = \epsilon + \max_{a \in [K]} \left\{ -\log\min\{\mu_a - \min\{\overline{\Delta}_{\min}, \Delta_0\}/4, 1 - \mu_a - \min\{\overline{\Delta}_{\min}, \Delta_0\}/4\} \right\}\,.$$

Then, we have shown that $d_\epsilon^+(\tilde{\mu}_{n,b}, \tilde{\mu}_{n,a}) \leq D_\mu$ where $D_\mu \in (0, +\infty)$. This yields the result. □

Lemma 59 shows that the challenger is mildly undersampled if the leader is not mildly undersampled.

Lemma 59. *Let V_n^p be as in Equation* (42)*. There exists p_3 with $\mathbb{E}_{\boldsymbol{\nu}\pi}[\exp(\alpha p_3)] < +\infty$ for all $\alpha > 0$ such that if $p \geq p_3$, for all n such that $U_n^p \neq \emptyset$, $B_n \notin V_n^p$ implies $C_n \in V_n^p$.*

Proof. Let p_3 to be specified later. Let $p \geq p_3$. Let $n \in \mathbb{N}$ such that $U_n^p \neq \emptyset$ and $V_n^p \neq [K]$, where $U_n^p \subseteq V_n^p$ are defined in Eq. (42). Since the statement holds when $B_n \in V_n^p$, we suppose that $B_n \notin V_n^p$ in the following.

Let p_0 as in Lemma 56, p_1 and C_μ as in Lemma 57, and p_2 and D_μ as in Lemma 58. Let $p_4 = \max\{2p_2 - 1, \frac{4}{3}\max\{p_0, p_1\} - 1/3\}$, which satisfied that $\mathbb{E}_{\boldsymbol{\nu}\pi}[\exp(\alpha p_4)] < +\infty$ for all $\alpha > 0$ by using Lemmas 56, 57 and 58. Then, for all $p \geq p_4 = \max\{2p_2 - 1, \frac{4}{3}\max\{p_0, p_1\} - 1/3\}$ and all n such that $B_n \notin V_n^p$, we have $\tilde{\mu}_{n,a} \in (0,1)$ for all $a \notin V_n^p$, $B_n = b_n^\star := \arg\max_{a \notin V_n^p} \mu_a$, $B_n \notin U_n^p$ and

$$\begin{aligned} &\forall b \notin \{b_n^\star\} \cup V_n^p, \quad W_{\epsilon,b_n^\star,b}(\tilde{\mu}_n, N_n) + \log N_{n,b} \geq (1+\eta)^{3(p-1)/4} C_\mu + \frac{3(p-1)}{4}\log(1+\eta)\,, \\ &\forall b \in U_n^p, \quad W_{\epsilon,b_n^\star,b}(\tilde{\mu}_n, N_n) + \log N_{n,b} \leq (1+\eta)^{(p-1)/2} D_\mu + \frac{p-1}{2}\log(1+\eta)\,, \end{aligned}$$

where we used Lemmas 56, 57 and 58. Direct manipulations yield that

$$(1+\eta)^{3(p-1)/4} C_\mu + \frac{3(p-1)}{4}\log(1+\eta) \geq (1+\eta)^{(p-1)/2} D_\mu + \frac{p-1}{2}\log(1+\eta)$$

$$\Longleftarrow \quad p \geq p_5 = 4\lceil \log_{1+\eta}(D_\mu/C_\mu)\rceil + 1\,,$$

where $\mathbb{E}_{\boldsymbol{\nu}\pi}[\exp(\alpha p_5)] < +\infty$ for all $\alpha > 0$ since it is a deterministic constant. Let $p_3 = \max\{p_4, p_5\}$ which satisfies $\mathbb{E}_{\boldsymbol{\nu}\pi}[\exp(\alpha p_3)] < +\infty$ for all $\alpha > 0$. Then, we have shown that for all $p \geq p_3$, for all n such that $B_n \notin V_n^p$, we have $B_n = b_n^\star$ and

$$\min_{b \notin \{b_n^\star\} \cup V_n^p} \{W_{\epsilon, b_n^\star, b}(\tilde{\mu}_n, N_n) + \log N_{n,b}\} > \max_{b \in U_n^p} \{W_{\epsilon, b_n^\star, b}(\tilde{\mu}_n, N_n) + \log N_{n,b}\}\,.$$

By definition of the TC challenger, i.e., $C_n \in \arg\min_{b \neq B_n} \{W_{\epsilon, B_n, b}(\tilde{\mu}_n, N_n) + \log N_{n,b}\}$, we obtain that $C_n \in V_n^p$. Otherwise, there would be a contradiction since we assumed $U_n^p \neq \emptyset$. This concludes the proof. □

Lemma 60 shows that all the arms are sufficient explored for large enough n.

Lemma 60. *There exists N_0 with $\mathbb{E}_{\boldsymbol{\nu}\pi}[N_0] < +\infty$ such that, for all $n \geq N_0$ and all $a \in [K]$, $N_{n,a} \geq \sqrt{n/K}$.*

Proof. Let p_0 and p_3 as in Lemmas 56 and 59. Combining Lemmas 56 and 59 yields that, for all $p \geq p_4 = \max\{p_3, 4p_0/3 - 1/3\}$ and all n such that $U_n^p \neq \emptyset$, we have $B_n \in V_n^p$ or $C_n \in V_n^p$. We have $\mathbb{E}_{\boldsymbol{\nu}\pi}[(1+\eta)^{p_2}] < +\infty$. We have $(1+\eta)^{p-1} \geq K(1+\eta)^{3(p-1)/4}$ for all $p \geq p_5 = 4\lceil \log_{1+\eta} K\rceil + 1$. Let $p \geq \max\{p_5, p_4\}$. For notational simplicity, we conduct the proof as if that $k(1+\eta)^{p-1} \in \mathbb{N}$ for all $k \in [K]$. It is direct to adapt the proof by using the operator $\lceil \cdot \rceil$.

Suppose towards contradiction that $U^p_{K(1+\eta)^{p-1}}$ is not empty. Then, for any $1 \leq t \leq K(1+\eta)^{p-1}$, U_t^p and V_t^p are non empty as well. Using the pigeonhole principle, there exists some $a \in [K]$ such that $N_{(1+\eta)^{p-1},a} \geq (1+\eta)^{3(p-1)/4}$. Thus, we have $\left|V^p_{(1+\eta)^{p-1}}\right| \leq K-1$. Our goal is to show that $\left|V^p_{2(1+\eta)^{p-1}}\right| \leq K-2$. A sufficient condition is that one arm in $V^p_{(1+\eta)^{p-1}}$ is pulled at least $(1+\eta)^{3(p-1)/4}$ times between $(1+\eta)^{p-1}$ and $2(1+\eta)^{p-1}-1$.

Case 1. Suppose there exists $a \in V^p_{(1+\eta)^{p-1}}$ such that $L_{2(1+\eta)^{p-1},a} - L_{(1+\eta)^{p-1},a} \geq \beta^{-1}\left((1+\eta)^{3(p-1)/4} + 3/2\right)$. Using Lemma 51, we obtain

$$N^a_{2(1+\eta)^{p-1},a} - N^a_{(1+\eta)^{p-1},a} \geq \beta(L_{2(1+\eta)^{p-1},a} - L_{(1+\eta)^{p-1},a}) - 3/2 \geq (1+\eta)^{3(p-1)/4}\,,$$

hence a is sampled $(1+\eta)^{3(p-1)/4}$ times between $(1+\eta)^{p-1}$ and $2(1+\eta)^{p-1}-1$.

Case 2. Suppose that for all $a \in V^p_{(1+\eta)^{p-1}}$, we have $L_{2(1+\eta)^{p-1},a} - L_{(1+\eta)^{p-1},a} < \beta^{-1}\left((1+\eta)^{3(p-1)/4} + 3/2\right)$. Then,

$$\sum_{a \notin V^p_{(1+\eta)^{p-1}}} (L_{2(1+\eta)^{p-1},a} - L_{(1+\eta)^{p-1},a}) \geq (1+\eta)^{p-1} - K\beta^{-1}\left((1+\eta)^{3(p-1)/4} + 3/2\right)\,.$$

Using Lemma 51, we obtain

$$\left|\sum_{a \notin V^p_{(1+\eta)^{p-1}}} (N^a_{2(1+\eta)^{p-1},a} - N^a_{(1+\eta)^{p-1},a}) - \beta \sum_{a \notin V^p_{(1+\eta)^{p-1}}} (L_{2(1+\eta)^{p-1},a} - L_{(1+\eta)^{p-1},a})\right| \leq 3(K-1)/2\,.$$

Combining all the above, we obtain

$$\begin{aligned}
&\sum_{a \notin V^p_{(1+\eta)^{p-1}}} (L_{2(1+\eta)^{p-1},a} - L_{(1+\eta)^{p-1},a}) - \sum_{a \notin V^p_{(1+\eta)^{p-1}}} (N^a_{2(1+\eta)^{p-1},a} - N^a_{(1+\eta)^{p-1},a}) \\
&\geq (1-\beta) \sum_{a \notin V^p_{(1+\eta)^{p-1}}} (L_{2(1+\eta)^{p-1},a} - L_{(1+\eta)^{p-1},a}) - 3(K-1)/2 \\
&\geq (1-\beta)\left((1+\eta)^{p-1} - K\beta^{-1}\left((1+\eta)^{3(p-1)/4} + 3/2\right)\right) - 3(K-1)/2 \geq K(1+\eta)^{3(p-1)/4}
\end{aligned}$$

where the last inequality is obtained for $p \geq p_6 + 1$ with

$$p_6 = \sup \left\{ p \in \mathbb{N} \mid (1-\beta) \left((1+\eta)^{p-1} - K\beta^{-1} \left((1+\eta)^{3(p-1)/4} + 3/2 \right) \right) - \frac{3}{2}(K-1) \right.$$
$$\left. < K(1+\eta)^{3(p-1)/4} \right\} .$$

The left hand side summation is exactly the number of times where an arm $a \notin V^p_{(1+\eta)^{p-1}}$ was leader but wasn't sampled, hence we have shown that

$$\sum_{t=(1+\eta)^{p-1}}^{2(1+\eta)^{p-1}-1} \mathbb{1}\left(B_t \notin V^p_{(1+\eta)^{p-1}},\, a_t = C_t \right) \geq K(1+\eta)^{3(p-1)/4} .$$

For any $(1+\eta)^{p-1} \leq t \leq 2(1+\eta)^{p-1} - 1$, U^p_t is non-empty, hence we have $B_t \notin V^p_{(1+\eta)^{p-1}}$ (hence $B_t \notin V^p_t$) implies $C_t \in V^p_t \subseteq V^p_{(1+\eta)^{p-1}}$. Therefore, we have shown that

$$\sum_{t=(1+\eta)^{p-1}}^{2(1+\eta)^{p-1}-1} \mathbb{1}\left(a_t \in V^p_{(1+\eta)^{p-1}} \right) \geq \sum_{t=(1+\eta)^{p-1}}^{2(1+\eta)^{p-1}-1} \mathbb{1}\left(B_t \notin V^p_{(1+\eta)^{p-1}},\, a_t = C_t \right) \geq K(1+\eta)^{3(p-1)/4}$$

Therefore, there is at least one arm in $V^p_{(1+\eta)^{p-1}}$ that is sampled $(1+\eta)^{3(p-1)/4}$ times between $(1+\eta)^{p-1}$ and $2(1+\eta)^{p-1} - 1$.

In summary, we have shown $\left| V^p_{2(1+\eta)^{p-1}} \right| \leq K - 2$ for all $p \geq p_7 = \max\{p_6, p_4, p_5\}$. By induction, for any $1 \leq k \leq K$, we have $\left| V^p_{k(1+\eta)^{p-1}} \right| \leq K - k$, and finally $U^p_{K(1+\eta)^{p-1}} = \emptyset$ for all $p \geq p_7$. Defining $N_0 = K(1+\eta)^{p_7 - 1}$, we have $\mathbb{E}_{\boldsymbol{\nu}\pi}[N_0] < +\infty$ by using Lemmas 56 and 59 for $p_4 = \max\{p_3, 4p_0/3 - 1/3\}$ and p_6 and p_5 are deterministic. For all $n \geq N_0$, we let $(1+\eta)^{p-1} = \frac{n}{K}$. Then, by applying the above, we have $U^p_{K(1+\eta)^{p-1}} = U^{\log_{1+\eta}(n/K)+1}_n$ is empty, which shows that $N_{n,a} \geq \sqrt{n/K}$ for all $a \in [K]$. □

H.3 Convergence Towards β-Optimal Allocation

The second step of in the generic analysis of Top Two algorithms Jourdan et al. [2022] is to show the convergence of the empirical proportions towards the β-optimal allocation. First, we show that the leader coincides with the best arm. Hence, the tracking procedure will ensure that the empirical proportion of time we sample it is exactly β. Second, we show that a sub-optimal arm whose empirical proportion overshoots its β-optimal allocation will not be sampled next as challenger. Therefore, this "overshoots implies not sampled" mechanism will ensure the convergence towards the β-optimal allocation. We emphasise that there are multiple ways to select the leader/challenger pair in order to ensure convergence towards the β-optimal allocation. Therefore, other choices of leader/challenger pair would yield similar results. Note that our results heavily rely on having obtained sufficient exploration first.

Lemma 61 shows the leader and the candidate answer are equal to the best arm for large enough n.

Lemma 61. *Let N_0 be as in Lemma 60. There exists $N_1 \geq N_0$ with $\mathbb{E}_{\boldsymbol{\nu}\pi}[N_1] < +\infty$ such that, for all $n \geq N_1$, we have $\tilde{\mu}_n \in (0,1)^K$ and $\tilde{a}_n = B_n = a^\star$.*

Proof. Let $\Delta_{\min} = \min_{a \neq a^\star}(\mu_{a^\star} - \mu_a)$ and $\Delta_0 = \min_{a \in [K]} \min\{\mu_a, 1 - \mu_a\} > 0$. Using Lemma 55, we obtain, for all $n \geq N_0$,

$$\tilde{\mu}_{n,a^\star} \geq \mu_{a^\star} - W_\mu \frac{\log(e + \sqrt{n/K}/(1+\eta))}{\sqrt{n/K}/(1+\eta)}$$
$$\forall a \neq a^\star, \quad \tilde{\mu}_{n,a} \leq \mu_a + W_\mu \frac{\log(e + \sqrt{n/K}/(1+\eta))}{\sqrt{n/K}/(1+\eta)} ,$$

where we used that $x \to \log(e+x)/x$ is decreasing and $\tilde{N}_{n,a} \geq N_{n,a}/(1+\eta) \geq \sqrt{n/K}/(1+\eta)$. Let $N_1 = \max\{N_0, \lceil K(1+\eta)^2 X_1^2 \rceil\}$ where

$$X_1 = \sup\left\{x > 1 \mid \; x \leq 4(\Delta_{\min}, \Delta_0)^{-1} W_\mu \log x + e\right\} \leq h_1(4(\Delta_{\min}, \Delta_0)^{-1} W_\mu, \; e)\,,$$

where we used Lemma 53, and h_1 defined therein. Using Lemmas 55 and 60, we obtain $\mathbb{E}_{\boldsymbol{\nu}\pi}[N_1] < +\infty$. Then, we have $0 < \mu_a - \Delta_0/4 \leq \tilde{\mu}_{n,a} \leq \mu_a + \Delta_0/4 < 1$ for all $a \in [K]$. Moreover, for all $n \geq N_1$, we have $\tilde{\mu}_{n,a^\star} \geq \mu_{a^\star} - \Delta_{\min}/4$ and $\tilde{\mu}_{n,a} \leq \mu_a + \Delta_{\min}/4$ for all $a \neq a^\star$, hence

$$a^\star = \arg\max_{a \in [K]} \tilde{\mu}_{n,a} = \arg\max_{a \in [K]} [\tilde{\mu}_{n,a}]_0^1 = \tilde{a}_n = B_n\,.$$

This concludes the proof. □

Lemma 62 shows that that the pulling proportion of the best arm converges towards β. It is a direct consequence of Lemma 61 by using the same proof as Lemma 39 in Azize et al. [2024], hence we omit the proof.

Lemma 62 (Lemma 39 in Azize et al. [2024])**.** *Let $\gamma > 0$, and N_1 be as in Lemma 61. There exists a deterministic constant $C_0 \geq 1$ such that, for all $n \geq C_0 N_1$, we have $\left|\frac{N_{n,a^\star}}{n-1} - \beta\right| \leq \gamma$.*

Lemma 63 shows that if a sub-optimal arm overshoots its β-optimal allocation then it cannot be selected as challenger for large enough n.

Lemma 63. *Let $\gamma \in (0, \gamma_\mu)$ where γ_μ is a problem dependent constant. Let N_1 and C_0 be as in Lemma 61 and 62. There exists $N_2 \geq C_0 N_1$ with $\mathbb{E}_{\boldsymbol{\nu}\pi}[N_2] < +\infty$ such that, for all $n \geq N_2$,*

$$\exists a \neq a^\star, \quad \frac{N_{n,a}}{n-1} \geq \gamma + \omega^\star_{\epsilon,\beta,a} \quad \implies \quad C_n \neq a\,.$$

Proof. Let $\eta > 0$ and $\gamma > 0$ be small enough, which we will specify below. Let $\tilde{\gamma} \in (0, \gamma)$. Let N_1 as in Lemma 61 and C_0 as in Lemma 62 for $\tilde{\gamma}$. Let $n \geq C_0 N_1$. Therefore, we have $\tilde{\mu}_n \in (0,1)^K$ and $\tilde{a}_n = B_n = a^\star$ and $\left|\frac{N_{n,a^\star}}{n-1} - \beta\right| \leq \tilde{\gamma}$. Using the same proof as in Lemma 61, there exists N_3 with $\mathbb{E}_{\boldsymbol{\nu}\pi}[N_3] < +\infty$ such that, for all $n \geq N_3$, we have $\|\tilde{\mu}_n - \mu\|_\infty \leq \eta$. Let $n \geq \max\{C_0 N_1, N_3\}$.

Let $a \neq a^\star$ such that $\frac{N_{n,a}}{n-1} \geq \omega^\star_{\epsilon,\beta,a} + \gamma$. Suppose towards contradiction that $\frac{N_{n,b}}{n-1} > \omega^\star_{\epsilon,\beta,b}$ for all $b \notin \{a^\star, a\}$. Then, for all $n \geq C_0 N_1$, we have

$$1 - \beta + \tilde{\gamma} \geq 1 - \frac{N_{n,a^\star}}{n-1} = \sum_{b \neq a^\star} \frac{N_{n,b}}{n-1} > \gamma + \sum_{b \neq a^\star} \omega^\star_{\epsilon,\beta,b} = 1 - \beta + \gamma\,,$$

which yields a contradiction since $\tilde{\gamma} < \gamma$. Therefore, for all $n \geq C_0 N_1$, we have

$$\exists a \neq a^\star, \quad \frac{N_{n,a}}{n-1} \geq \omega^\star_{\epsilon,\beta,a} + \gamma \quad \implies \quad \exists b \notin \{a^\star, a\}, \quad \frac{N_{n,b}}{n-1} \leq \omega^\star_{\epsilon,\beta,b}\,.$$

Let $b \notin \{a^\star, a\}$ such that $\frac{N_{n,b}}{n-1} \leq \omega^\star_{\epsilon,\beta,b}$. By definition of the TC challenger, we obtain

$$\begin{aligned}
C_n \neq a \; &\impliedby \; W_{\epsilon,a^\star,a}(\tilde{\mu}_n, N_n) + \log N_{n,a} > W_{\epsilon,a^\star,b}(\tilde{\mu}_n, N_n) + \log N_{n,b} \\
&\impliedby \; \frac{1}{n-1}\left(W_{\epsilon,a^\star,a}(\tilde{\mu}_n, N_n) - W_{\epsilon,a^\star,b}(\tilde{\mu}_n, N_n)\right) > \frac{1}{n-1} \log \frac{\omega^\star_{\epsilon,\beta,b}}{\omega^\star_{\epsilon,\beta,a} + \gamma} \\
&\impliedby \; \frac{1}{n-1}\left(W_{\epsilon,a^\star,a}(\tilde{\mu}_n, N_n) - W_{\epsilon,a^\star,b}(\tilde{\mu}_n, N_n)\right) > \frac{1}{n-1} \max_{a \neq b} \left|\log \frac{\omega^\star_{\epsilon,\beta,b}}{\omega^\star_{\epsilon,\beta,a}}\right|\,,
\end{aligned}$$

where we used the positivity of the β-optimal allocation (Lemma 43) to ensure that $\max_{a \neq b} \left|\log \frac{\omega^\star_{\epsilon,\beta,b}}{\omega^\star_{\epsilon,\beta,a}}\right| \in (0, +\infty)$. Using that $\tilde{\mu}_{n,a^\star} > \max\{\tilde{\mu}_{n,a}, \tilde{\mu}_{n,b}\}$, we obtain

$$\begin{aligned}
&\frac{1}{n-1}\left(W_{\epsilon,a^\star,a}(\tilde{\mu}_n, N_n) - W_{\epsilon,a^\star,b}(\tilde{\mu}_n, N_n)\right) \\
&\geq \inf_{u \in [0,1]} \left\{\frac{N_{n,a^\star}}{n-1} d_\epsilon^-(\tilde{\mu}_{n,a^\star}, u) + (\omega^\star_{\epsilon,\beta,a} + \gamma) d_\epsilon^+(\tilde{\mu}_{n,a}, u)\right\}
\end{aligned}$$

$$- \inf_{u\in[0,1]} \left\{ \frac{N_{n,a^\star}}{n-1} d_\epsilon^-(\tilde{\mu}_{n,a^\star}, u) + \omega^\star_{\epsilon,\beta,b} d_\epsilon^+(\tilde{\mu}_{n,b}, u) \right\}$$
$$\geq \inf_{\tilde{\beta}:|\tilde{\beta}-\beta|\leq\tilde{\gamma}} G_{a,b}(\tilde{\mu}_n, \tilde{\beta}) \geq \inf_{\lambda:\|\lambda-\mu\|_\infty\leq\eta} \inf_{\tilde{\beta}:|\tilde{\beta}-\beta|\leq\tilde{\gamma}} G_{a,b}(\lambda, \tilde{\beta}) \,,$$

where, for all $(a,b) \in ([K] \setminus \{a^\star\})^2$ such that $a \neq b$,

$$G_{a,b}(\lambda, \tilde{\beta}) = \inf_{u\in[0,1]} \left\{ \tilde{\beta} d_\epsilon^-(\lambda_{a^\star}, u) + (\omega^\star_{\epsilon,\beta,a} + \gamma) d_\epsilon^+(\lambda_a, u) \right\}$$
$$- \inf_{u\in[0,1]} \left\{ \tilde{\beta} d_\epsilon^-(\lambda_{a^\star}, u) + \omega^\star_{\epsilon,\beta,b} d_\epsilon^+(\lambda_b, u) \right\} .$$

Using the equality at equilibrium from (39) (see Lemma 43) and the fact that the transportation costs are increasing in their allocation argument (see Lemma 37), we obtain $G_{a,b}(\mu, \beta) > 0$ for all $(a,b) \in ([K] \setminus \{a^\star\})^2$ such that $a \neq b$, since

$$\inf_{u\in[0,1]} \left\{ \beta d_\epsilon^-(\mu_{a^\star}, u) + (\omega^\star_{\epsilon,\beta,a} + \gamma) d_\epsilon^+(\mu_a, u) \right\} > W_{\epsilon,a^\star,a}(\mu, w^\star_{\epsilon,\beta}) = W_{\epsilon,a^\star,b}(\mu, w^\star_{\epsilon,\beta}) \,.$$

By Lemma 49, the functions $(\lambda, \tilde{\beta}) \to G_{a,b}(\lambda, \tilde{\beta})$ and $\lambda \to \inf_{\tilde{\beta}:|\tilde{\beta}-\beta|\leq\tilde{\gamma}} G_{a,b}(\lambda, \tilde{\beta})$ are continuous. Therefore, there exists η_μ and γ_μ small enough such that

$$\inf_{\lambda:\|\lambda-\mu\|_\infty\leq\eta} \inf_{\tilde{\beta}:|\tilde{\beta}-\beta|\leq\tilde{\gamma}} G_{a,b}(\lambda, \tilde{\beta}) \geq G_{a,b}(\mu, \beta)/2 \geq \frac{1}{2} \min_{a\neq b, a\neq a^\star, b\neq a^\star} G_{a,b}(\mu, \beta) > 0 \,,$$

where the last strict inequality uses that the minimum of a finite number of positive constants is also positive. Considering such (η_μ, γ_μ) at the beginning of the proof and taking $N_2 = \max\{C_0 N_1, N_3, \kappa_\mu\}$ where

$$\kappa_\mu = 2 + \frac{2 \max_{a\neq b} \left| \log \frac{\omega^\star_{\epsilon,\beta,b}}{\omega^\star_{\epsilon,\beta,a}} \right|}{\min_{a\neq b, a\neq a^\star, b\neq a^\star} G_{a,b}(\mu, \beta)} < +\infty \,,$$

As it satisfies $\mathbb{E}_{\boldsymbol{\nu}\pi}[N_2] < +\infty$, this concludes the proof. □

Lemma 64 shows that the convergence time towards the β-optimal allocation has finite expectation. It is a direct consequence of Lemmas 61, 62 and 63 by using the same proof as Lemma 41 in Azize et al. [2024], hence we omit the proof.

Lemma 64 (Lemma 41 in Azize et al. [2024])**.** *Let $\gamma \in (0, \gamma_\mu)$ where γ_μ is a problem dependent constant, and $T_\gamma(w)$ as in Eq.* (40)*. Then, we have $\mathbb{E}_{\boldsymbol{\nu}\pi}[T_\gamma(\omega^\star_{\epsilon,\beta})] < +\infty$.*

H.4 Asymptotic Upper Bound

The final step of the generic analysis of Top Two algorithms [Jourdan et al., 2022] is to invert the GLR stopping rule in Eq. (7) by leveraging the convergence of the empirical proportions towards the β-optimal allocation. Provided this convergence is shown, the asymptotic upper bound on the expected sample complexity only depends on the dependence in $\log(1/\delta)$ of the threshold that ensures δ-correctness. Compared to the non-private GLR stopping rule, the GLR stopping rule in Eq. (7) pay an extra cost to ensure privacy.

Lemma 65. *Let $\epsilon > 0$, $\eta > 0$ and $(\delta, \beta) \in (0,1)^2$. Let $T^\star_{\epsilon,\beta}(\boldsymbol{\nu})$ as in Eq.* (35) *and $\omega^\star_{\epsilon,\beta}$ be its associated β-optimal allocation. Assume that there exists $\gamma_\mu > 0$ such that $\mathbb{E}_{\boldsymbol{\nu}\pi}[T_\gamma(\omega^\star_{\epsilon,\beta})] < +\infty$ for all $\gamma \in (0, \gamma_\mu)$, where $T_\gamma(w)$ is defined in Eq.* (40)*. Combining such a sampling rule, using the GPE$_\eta(\epsilon)$ update, with the GLR stopping rule as in Eq.* (7) *and the stopping threshold c as in Eq.* (8) *yields an ϵ-global DP and δ-correct algorithm which satisfies that, for all $\boldsymbol{\nu}$ with mean μ such that $|a^\star(\mu)| = 1$,*

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_{\boldsymbol{\nu}\pi}[\tau_{\epsilon,\delta}]}{\log(1/\delta)} \leq 2(1+\eta) T^\star_{\epsilon,\beta}(\boldsymbol{\nu}) \,.$$

Proof. Lemma 5 yields the ϵ-global DP. Theorem 6 yields the δ-correctness.

Let $\zeta > 0$ and $a^\star$ be the unique best arm. Using the equality at equilibrium from (39) (see Lemma 43) and the continuity of $(\mu, w) \mapsto \min_{a \neq a^\star(\mu)} W_{\epsilon, a^\star(\mu), a}(\mu, w)$ (see Lemma 42), there exists $\gamma_\zeta > 0$ such that $\left\| \frac{N_n}{n-1} - \omega^\star_{\epsilon,\beta} \right\|_\infty \leq \gamma_\zeta$ and $\|\tilde{\mu}_n - \mu\|_\infty \leq \gamma_\zeta$ implies that

$$\forall a \neq a^\star, \quad W_{\epsilon, a^\star, a}(\tilde{\mu}_n, N_n/(n-1)) \geq \frac{(1-\zeta)}{T^\star_{\epsilon,\beta}(\boldsymbol{\nu})} \, .$$

We choose such a γ_ζ. Let $\gamma_\mu > 0$ be such that for $\mathbb{E}_{\boldsymbol{\nu}\pi}[T_\gamma(\omega^\star_{\epsilon,\beta})] < +\infty$ for all $\gamma \in (0, \gamma_\mu)$, where $T_\gamma(\omega)$ is defined in Eq. (40). Let $\gamma \in (0, \min\{\gamma_\mu, \gamma_\zeta, \min_{a \in [K]} \omega^\star_{\epsilon,\beta,a}/4, \Delta_{\min}/4, \Delta_0/4\})$ where $\Delta_{\min} = \min_{a \neq a^\star}(\mu_{a^\star} - \mu_a)$ and $\Delta_0 = \min_{a \in [K]} \min\{\mu_a, 1 - \mu_a\}$. For all $n \geq T_\gamma(\omega^\star_{\epsilon,\beta})$, we have

$$\tilde{N}_{n,a} \geq N_{n,a}/(1+\eta) \geq (n-1)(\omega^\star_{\epsilon,\beta,a} - \gamma)/(1+\eta) \geq (n-1)\frac{3}{4(1+\eta)} \min_{a \in [K]} \omega^\star_{\epsilon,\beta,a} > 0 \, ,$$

where the last inequality used the positivity of the β-optimal allocation (Lemma 43). Since arms are sampled linearly, it is direct to construct $N_3 \geq T_\gamma(\omega^\star_{\epsilon,\beta})$ with $\mathbb{E}_{\boldsymbol{\nu}\pi}[N_3] < +\infty$ such that $\|\tilde{\mu}_n - \mu\|_\infty \leq \gamma$ and $\left\| \frac{N_n}{n-1} - \omega^\star_{\epsilon,\beta} \right\|_\infty \leq \gamma$ (as well as $\min_{a \in [K]} N_{n,a} > e$ trivially).

Recall that $c(n, \epsilon, \delta) = c_1(n, \delta) + c_2(n, \epsilon)$ where $n \mapsto c_1(n, \delta)$ and $n \mapsto c_1(n, \delta)$ are increasing (see Lemmas 52 and 39). Since $\tilde{N}_{n,a} \leq N_{n,a} \leq n$, we obtain

$$\sum_{b \in \{a^\star, a\}} c(\tilde{N}_n, \epsilon, \delta) \leq 2(c_1(n, \delta) + c_2(n, \epsilon)) \, .$$

Using Lemma 37 and $\tilde{N}_{n,a} \geq N_{n,a}/(1+\eta)$ for all $a \in [K]$ (Lemma 50), we obtain

$$W_{\epsilon, a^\star, a}(\tilde{\mu}_n, \tilde{N}_n) \geq \frac{n-1}{1+\eta} W_{\epsilon, a^\star, a}\left(\tilde{\mu}_n, \frac{N_n}{n-1}\right) \, .$$

Let $\kappa \in (0,1)$ and $T > N_3/\kappa$. For all $n \in [\kappa T, T]$, we have $\tilde{a}_n = a^\star$ and, for all $a \neq a^\star$,

$$\begin{aligned}
& \tau_{\epsilon,\delta} > n \\
\implies \quad & \exists a \neq a^\star, \quad W_{\epsilon, a^\star, a}(\tilde{\mu}_n, \tilde{N}_n) \leq \sum_{b \in \{a^\star, a\}} c(\tilde{N}_n, \epsilon, \delta) \\
\implies \quad & \exists a \neq a^\star, \quad \frac{n-1}{1+\eta} W_{\epsilon, a^\star, a}\left(\tilde{\mu}_n, \frac{N_n}{n-1}\right) \leq 2(c_1(n, \delta) + c_2(n, \epsilon)) \\
\implies \quad & \exists a \neq a^\star, \quad \frac{n-1}{1+\eta} \frac{(1-\zeta)}{T^\star_{\epsilon,\beta}(\boldsymbol{\nu})} \leq 2c_1(T, \delta) + 2c_2(T, \epsilon) \, ,
\end{aligned}$$

where we used that $n \mapsto c_1(n, \delta)$ and $n \mapsto c_2(n, \epsilon)$ are increasing and $n \leq T$. Therefore, we obtain

$$\begin{aligned}
\min\{\tau_{\epsilon,\delta}, T\} &\leq \kappa T + \sum_{n=\kappa T}^{T} \mathbb{1}(\tau_\delta > n) \\
&\leq \kappa T + \sum_{n=\kappa T}^{T} \mathbb{1}\left(\frac{n-1}{1+\eta} \frac{(1-\zeta)}{T^\star_{\epsilon,\beta}(\boldsymbol{\nu})} \leq 2c_1(T, \delta) + 2c_2(T, \epsilon) \right) \\
&\leq \kappa T + 1 + \frac{2(1+\eta) T^\star_{\epsilon,\beta}(\boldsymbol{\nu})}{1-\zeta} (c_1(T, \delta) + c_2(T, \epsilon)) \, .
\end{aligned}$$

Let $T_\zeta(\delta)$ defined as

$$T_\zeta(\delta) := \inf\left\{ T \geq 1 \mid \frac{1}{1-\kappa} \left(1 + \frac{2(1+\eta) T^\star_{\epsilon,\beta}(\boldsymbol{\nu})}{1-\zeta} (c_1(T, \delta) + c_2(T, \epsilon)) \right) \leq T \right\} \, .$$

Using Lemma 52, we know that $\overline{W}_{-1}(x) =_{x \to \infty} x + \log x$, hence we have $\limsup_{\delta \to 0} c_1(T, \delta)/\log(1/\delta) \leq 1$. Since $\lim_{\delta \to 0} c_2(T, \epsilon)/\log(1/\delta) = 0$, we obtain

$\limsup_{\delta\to 0} \frac{T_\zeta(\delta)}{\log(1/\delta)} \le \frac{2(1+\eta)T^\star_{\epsilon,\beta}(\boldsymbol{\nu})}{(1-\zeta)(1-\kappa)}$. For every $T \ge \max\{T_\zeta(\delta), N_3/\kappa\}$, we have $\tau_{\epsilon,\delta} \le T$, hence $\mathbb{E}_{\boldsymbol{\nu}\pi}[\tau_{\epsilon,\delta}] \le T_\zeta(\delta) + \mathbb{E}_{\boldsymbol{\nu}\pi}[N_3]/\kappa < +\infty$. Therefore, for all $\zeta, \kappa > 0$, we obtain

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_{\boldsymbol{\nu}\pi}[\tau_{\epsilon,\delta}]}{\log(1/\delta)} \le \limsup_{\delta\to 0} \frac{T_\zeta(\delta)}{\log(1/\delta)} \le \frac{2(1+\eta)T^\star_{\epsilon,\beta}(\boldsymbol{\nu})}{(1-\zeta)(1-\kappa)}\,.$$

Letting ζ and κ go to zero concludes the proof. □

Proof of Theorem 7 The proof is obtained by combining Theorem 6 and Lemmas 5, 60, 64 and 65.

I Variants of Algorithms

In Appendix I, we propose several variants of the algorithmic components used in our algorithm. The objective is to give freedom of choice for the practitioners interested in solving ϵ-global DP BAI. Given the rich literature on BAI, it is unreasonable to provide details for the ϵ-global DP version of all the existing BAI algorithms. Therefore, we settle for a few instances that has received increased scrutiny in the BAI literature.

First, we adapt the Track-and-Stop sampling rule [Garivier and Kaufmann, 2016] to solve ϵ-global DP BAI (Appendix I.1). This leverages the computational tractable procedure to compute the optimal allocation $w^\star_\epsilon$ derived in Lemma 47. Second, we explore some alternative choices of components of the Top Two sampling rule for ϵ-global DP BAI (Appendix I.2). This includes adaptive choice of target for the leader, hence aiming at achieving $T^\star_\epsilon(\boldsymbol{\nu})$ instead of $T^\star_{\epsilon,\beta}(\boldsymbol{\nu})$. Third, we adapt the LUCB sampling rule [Kalyanakrishnan et al., 2012] for ϵ-global DP BAI (Appendix I.3).

I.1 Track-and-Stop Sampling Rule

The Track-and-Stop (TaS) sampling rule was introduced in the seminal paper [Garivier and Kaufmann, 2016]. At each time n, it solves the optimization problem defining the characteristic time for the current empirical estimator $\tilde{\mu}_n$. When $\tilde{\mu}_n \in (0,1)^K$, we define $\widetilde{w}_n = w^\star_\epsilon(\tilde{\nu}_n)$ where $\tilde{\nu}_n$ is the Bernoulli instance with means $\tilde{\mu}_n$. When $\tilde{\mu}_n \notin (0,1)^K$, $[\tilde{\mu}_n]_0^1$ corresponds to a degenerate Bernoulli instance, hence we define $\widetilde{w}_n = 1_K/K$. Since $\tilde{\mu}_n$ is updated on a per-arm geometric grid governed by η, the optimal allocation $\widetilde{w}_n$ is updated on the same per-arm geometric grid. Therefore, choosing a larger η yields lower computational cost of TaS at the cost of larger expected sampled complexity, i.e., asymptotic multiplicative factor $1+\eta$ due to the update grid.

Given the vector $\widetilde{w}_n \in \triangle_K$, the next arm a_n to sample is obtained by using C-Tracking [Garivier and Kaufmann, 2016] with forced exploration in order to ensure that sufficient exploration holds. This is done here by projecting on $\triangle_K^\epsilon = \{w \in [\epsilon,1]^K \mid \sum_{a\in[K]} w_a = 1\}$ for a well chosen $\epsilon \in (0, 1/K]$. Let $\widetilde{w}_n^{\epsilon_n}$ be the ℓ_∞ projection of $\widetilde{w}_n$ on $\triangle_K^{\epsilon_n}$ with $\epsilon_n = (K^2+n)^{-1/2}/2$. While we consider a projection that changes at each time n (due to ϵ_n), $\widetilde{w}_n^{\epsilon_n}$ could also be updated on a per-arm geometric grid, i.e., when $\widetilde{w}_n$ is updated itself. For all $n \ge K+1$, the TaS sampling rule defines

$$a_n \in \operatorname*{arg\,max}_{a\in[K]} \left\{ \sum_{t\in[n]} \widetilde{w}_{t,a}^{\epsilon_t} - N_{n,a} \right\}. \tag{43}$$

In summary, our proposed Track-and-Stop algorithm is defined as in DP-TT with the sole modification that Lines 13-14 are replaced by the sampling rule defined in Eq. (43).

Optimal Allocation Oracle In Lemma 47, we show that $w^\star_\epsilon(\boldsymbol{\nu})$ can be computed explicitly based on the unique fixed-point solution $F_\mu(y) = 1$ for $y \in [0, \min_{a\neq a^\star(\mu)} d^-_\epsilon(\mu_{a^\star(\mu)}, \mu_a))$, where F_μ is an increasing one-to-one mapping from $[0, \min_{a\neq a^\star(\mu)} d^-_\epsilon(\mu_{a^\star(\mu)}, \mu_a))$ to $[0, +\infty)$ defined as

$$F_\mu(y) = \sum_{a\neq a^\star(\mu)} \frac{d^-_\epsilon(\mu_{a^\star(\mu)}, u_a(x_a(y)))}{d^+_\epsilon(\mu_a, u_a(x_a(y)))}\,. \tag{44}$$

The definitions of u_a and x_a is defered to Lemma 47, u_a is decreasing and x_a is increasing and strictly convex.

Asymptotic Expected Sample Complexity Combining the TaS sampling rule a_n as in Eq. (43) with the $\text{GPE}_\eta(\epsilon)$ update and the GLR stopping rule as in Eq. (7) for the stopping threshold as in Eq. (8) yields a δ-correct and ϵ-global DP algorithm (see Lemma 5 and Theorem 6). Moreover, we conjecture that its satisfies that, for all $\boldsymbol{\nu} \in \mathcal{F}^K$ with unique best arm,

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\nu}\pi}\left[\tau_{\epsilon,\delta}\right]}{\log(1/\delta)} \leq 2(1+\eta) T^\star_\epsilon(\boldsymbol{\nu}) \, .$$

The multiplicative factor $1+\eta$ comes from the per-arm geometric update grid, and the factor 2 comes from the asymptotic scaling in $2\log(1/\delta)$ of the stopping threshold. Using Theorem 7 for $\beta = 1/2$ and $T^\star_{\epsilon,1/2}(\boldsymbol{\nu}) \leq 2T^\star_\epsilon(\boldsymbol{\nu})$ (Lemma 44), proving this conjecture would only yield an asymptotic improvement by a factor of at most 2. However, this would come at the price of a significantly higher computational cost.

Proof Sketch of Conjecture While the detailed proof of this conjecture is beyond the scope of this work, an astute reader could notice that all the necessary steps were proven to derive Theorem 7 for DP-TT. At a high level, it is intuitive that the asymptotic analysis of Track-and-Stop is simpler than the one of DP-TT.

First, the forced exploration is enforced algorithmically, hence an equivalent of Lemma 60 can be shown for the Track-and-Stop sampling rule. In contrast, the proof of sufficient exploration for DP-TT is more challenging and involves a subbtle reasoning towards contradiction, see Appendix H.2 for more details.

Second, the convergence towards the optimal allocation is also enforced algorithmically. Thanks to the forced exploration and due to the continuity of $\boldsymbol{\nu} \mapsto w^\star_\epsilon(\boldsymbol{\nu})$ (Lemma 42) and the convergence $\tilde{\mu}_n \to_{n\to+\infty} \mu$, the empirical optimal allocation $\widetilde{w}_n$ converges towards the true optimal allocation $w^\star_\epsilon(\boldsymbol{\nu})$. Therefore, an equivalent of Lemma 64 can be shown for the Track-and-Stop sampling rule. In contrast, the proof of convergence towards β-optimal allocation for DP-TT is more challenging and leverage subbtle regularity properties of the β characteristic time and its optimal allocation, e.g., the equality at equilibrium of all the transportations costs in Eq. (39), see Appendix H.3 for more details.

Third, the invertion of the GLR stopping rule can be done similarly as for DP-TT. The sole modification lies in using our derived regularity properties for $w^\star_\epsilon(\boldsymbol{\nu})$ instead of $w^\star_{\epsilon,\beta}(\boldsymbol{\nu})$, e.g., the equality at equilibrium of all the transportations costs in Lemma 43. Therefore, an equivalent of Lemma 65 can be shown for the Track-and-Stop sampling rule with $2(1+\eta)T^\star_\epsilon(\boldsymbol{\nu})$ instead of $2(1+\eta)T^\star_{\epsilon,\beta}(\boldsymbol{\nu})$, see Appendix H.4 for more details.

I.2 Top Two Sampling Rule

As detailed in Chapter 2.2 in Jourdan [2024], a Top Two sampling rule is defined by four choices: a leader arm $B_n \in [K]$, a challenger arm $C_n \in [K] \setminus \{B_n\}$, a target $\beta_n(B_n, C_n) \in [0,1]$ and a mechanism to reach the target, i.e., $a_n \in \{B_n, C_n\}$ by using $\beta_n(B_n, C_n)$. For instance, the sampling rule in DP-TT uses the EB leader, the TCI challenger, a fixed target $\beta \in (0,1)$ and K independent β-tracking procedures (one per leader). We propose adaptive choice of target (Appendix I.2.1), as well as leader fostering implicit exploration (Appendix I.2.2).

I.2.1 Adaptive Target

When the target is fixed to β beforehand, the Top Two sampling rule can achieve $T^\star_{\epsilon,\beta}(\boldsymbol{\nu})$ at best. We propose adaptive choices of the target inspired by the recent literature on asymptotically optimal Top Two algorithms [You et al., 2023; Bandyopadhyay et al., 2024].

BOLD Target Given the EB-TCI leader/challenger pair (B_n, C_n) defined in DP-TT, we adapt the BOLD target from Bandyopadhyay et al. [2024]. Let us define

$$u_{\epsilon,B_n,a}(\tilde{\mu}_n, N_n) = \operatorname*{arg\,min}_{u \in [0,1]} \left\{ N_{n,B_n} d^-_\epsilon(\tilde{\mu}_{n,B_n}, u) + N_{n,a} d^+_\epsilon(\tilde{\mu}_{n,b}, u) \right\} \, , \tag{45}$$

whose closed-form solution is given in Lemma 45. Then, the deterministic BOLD target defines the next arm to pull as

$$a_n = B_n \quad \text{if} \quad \sum_{a \neq B_n} \frac{d_\epsilon^-(\tilde{\mu}_{n,B_n}, u_{\epsilon,B_n,a}(\tilde{\mu}_n, N_n))}{d_\epsilon^+(\tilde{\mu}_{n,a}, u_{\epsilon,B_n,a}(\tilde{\mu}_n, N_n))} > 1 \quad \text{and} \quad a_n = C_n \quad \text{otherwise.} \tag{46}$$

In summary, the sole modification in DP-TT is Line 14 that is replaced by the sampling rule defined in Eq. (46).

For any single-parameter exponential family of distributions, Bandyopadhyay et al. [2024] shows that the BOLD target allows to reach asymptotic optimality. Forced exploration is added by Bandyopadhyay et al. [2024] to ensure that sufficient exploration holds. Showing that the BOLD target can achieve asymptotic optimality without forced exploration, i.e., meaning that it ensures sufficient exploration on its own, is an open problem.

IDS Target Given the EB-TCI leader/challenger pair (B_n, C_n) defined in DP-TT, we adapt the IDS target from You et al. [2023]. Namely, the randomized IDS target defines the next arm to pull from as

$$a_n = \begin{cases} B_n & \text{with proba } \beta_n(B_n, C_n) \\ C_n & \text{otherwise} \end{cases} \text{ where } \beta_n(B_n, C_n) = \frac{N_{n,B_n} d_\epsilon^-(\tilde{\mu}_{n,B_n}, u_{\epsilon,B_n,C_n}(\tilde{\mu}_n, N_n)}{W_{\epsilon,B_n,C_n}(\tilde{\mu}_n, N_n)}\,, \tag{47}$$

where $u_{\epsilon,B_n,C_n}(\tilde{\mu}_n, N_n)$ is defined in Eq. (45). In summary, the sole modification in DP-TT is Line 14 that is replaced by the sampling rule defined in Eq. (47).

While we could use $K(K-1)$ tracking procedures to select $a_n \in \{B_n, C_n\}$, we use randomization above for the sake of simplicity. For Gaussian distributions with known variance, You et al. [2023] shows that the IDS target allows to reach asymptotic optimality. Showing that the IDS target can achieve optimality for other classes of distributions is an open problem.

Asymptotic Expected Sample Complexity Sampling a_n as in Eq. (46) or (47) for the EB-TCI leader/challenger pair (B_n, C_n) defined in DP-TT based on the $\text{GPE}_\eta(\epsilon)$ update and the GLR stopping rule as in Eq. (7) for the stopping threshold as in Eq. (8) yields a δ-correct and ϵ-global DP algorithm (see Lemma 5 and Theorem 6).

While we conjecture that their asymptotic expected sample complexities $\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\nu}\pi}[\tau_{\epsilon,\delta}]}{\log(1/\delta)}$ are both upper bounded by $2(1+\eta)T_\epsilon^\star(\boldsymbol{\nu})$, we emphasize that our analysis doesn't provide the necessary steps for this result to hold. This is an interesting research direction left for future work.

I.2.2 Implicit Exploring Leaders and TC Chalenger

The empirical best (EB) leader is a greedy choice of leader that doesn't foster implicit exploration. Without additional exploration mechanism, it can suffer from large empirical stopping time despite being enough for an asymptotic analysis, see [Jourdan et al., 2022]. This motivated the choice of the TCI challenger for DP-TT, since it fosters additional implicit exploration by penalizing over sampled challengers with the $\log N_{n,a}$ term. We propose other choices of leaders that foster implicit exploration, and define the TC challenger that removes this penalization.

The UCB leader is defined as

$$B_n^{\text{UCB}} \in \operatorname*{arg\,max}_{a \in [K]} U_{n,a} \quad \text{where} \quad U_{n,a} = \max\left\{u \in [0,1] \mid N_{n,a} d_\epsilon^+([\tilde{\mu}_{n,a}]_0^1, u) \leq \log(n)\right\}\,. \tag{48}$$

By adding a bonus to the empirical mean, we are optimistic since we consider that the means are better than suggested by our observations.

The IMED leader builds on the IMED algorithm [Honda and Takemura, 2015] is defined as

$$B_n^{\text{IMED}} \in \operatorname*{arg\,min}_{a \in [K]} \left\{N_{n,a} d_\epsilon^+([\tilde{\mu}_{n,a}]_0^1, \tilde{\mu}_n^\star) + \log N_{n,a}\right\} \quad \text{where} \quad \tilde{\mu}_n^\star = \max_{a \in [K]} [\tilde{\mu}_{n,a}]_0^1\,. \tag{49}$$

The TC challenger is defined as

$$C_n^{\text{TC}} \in \operatorname*{arg\,min}_{a \neq B_n} W_{\epsilon,B_n,b}(\tilde{\mu}_n, N_n)\,, \tag{50}$$

where $W_{\epsilon,a,b}$ is defined as in Eq. (4).

In summary, the sole modification in DP-TT is Line 13 which can be replaced by choosing the leader as in Eq. (48) or Eq. (49), or choosing the challenger as in Eq. (50).

Asymptotic Expected Sample Complexity Choosing the leader as in Eq. (48) or Eq. (49) or the challenger as in Eq. (50) based on the β-tracking as in DP-TT, the $\text{GPE}_\eta(\epsilon)$ update and the GLR stopping rule as in Eq. (7) for the stopping threshold as in Eq. (8) yields a δ-correct and ϵ-global DP algorithm (see Lemma 5 and Theorem 6). Moreover, we conjecture that its satisfies that, for all $\boldsymbol{\nu} \in \mathcal{F}^K$ with distinct means,

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_{\boldsymbol{\nu}\pi}\left[\tau_{\epsilon,\delta}\right]}{\log(1/\delta)} \leq 2(1+\eta) T^\star_{\epsilon,\beta}(\boldsymbol{\nu}) \, .$$

While the detailed proof of this conjecture is beyond the scope of this work, an astute reader could notice that all the necessary steps were proven to derive Theorem 7 for DP-TT. When using the TC challenger as in Eq. (50), the proofs of Lemmas 59 and 63 can be readily adapted. When using the UCB leader as in Eq. (48) or the IMED leader as in Eq. (49), the proofs of Lemmas 56 and 61 could also be adapted.

I.3 LUCB Sampling Rule

While the Top Two terminology was introduced in Russo [2016], the first sampling rule having a Top Two structure is the greedy sampling strategy in LUCB1 introduced by Kalyanakrishnan et al. [2012]. At each time n, it selects the EB leader $B_n^{\text{EB}} = \tilde{a}_n$ and the UCB challenger defined as

$$C_n^{\text{UCB}} \in \operatorname*{arg\,max}_{a \neq B_n^{\text{EB}}} U_{n,a} \quad \text{where} \quad U_{n,a} \quad \text{as in Eq. (48)} \, . \tag{51}$$

Then, it samples both B_n^{EB} and C_n^{UCB}. Instead of using the GLR stopping rule as in Eq.(7), LUCB1 stops when the LCB (lower confidence bound) of the leader exceeds the UCB of the challenger, i.e.,

$$\tau_{\epsilon,\delta}^{\text{LUCB1}} = \inf\left\{ n \mid \widetilde{L}_{n,B_n^{\text{EB}}} > U_{n,C_n^{\text{UCB}}} \right\} \, , \tag{52}$$

where

$$\widetilde{L}_{n,a} = \max\left\{ u \in [0,1] \mid N_{n,a} d_\epsilon^-([\tilde{\mu}_{n,a}]_0^1, u) \leq \log(n) \right\} \, . \tag{53}$$

In summary, the modifications in DP-TT are: (1) the sampling rule in Lines 13-15 is replaced by sampling both B_n^{EB} and C_n^{UCB}, and (2) the stopping rule in Line 10 is replaced by Eq. (52). While studying this algorithm is beyond the scope of this work, we emphasize that LUCB is known to not reach asymptotic (β-)optimality.

J Implementation Details and Supplementary Experiments

Appendix J is organized as follows. First, we provide additional detail on the implementation details for our algorithm (Appendix J.1). Second, we provide supplementary experiments to illustrate the good performance of our algorithm (Appendix J.2).

J.1 Implementation Details

We present additional experiments comparing the algorithms in different bandit instances with Bernoulli distributions, as defined by Sajed and Sheffet [2019], namely

$\mu_1 = (0.95, 0.9, 0.9, 0.9, 0.5)$, $\mu_2 = (0.75, 0.7, 0.7, 0.7, 0.7)$,
$\mu_3 = (0.1, 0.3, 0.5, 0.7, 0.9)$, $\mu_4 = (0.75, 0.625, 0.5, 0.375, 0.25)\}$,
$\mu_5 = (0.75, 0.53125, 0.375, 0.28125, 0.25)$, $\mu_6 = (0.75, 0.71875, 0.625, 0.46875, 0.25)\}$.

For each Bernoulli instance, we implement the algorithms with

$$\epsilon \in \{0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 10, 100, 125\} \, .$$

The risk level is set at $\delta = 0.01$. We verify empirically that the algorithms are δ-correct by running each algorithm 1000 times.

We implement all the algorithms in Python (version 3.8) and on an 8 core 64-bits Intel i5@1.6 GHz CPU.

Choice of Hyperparameters The default choice $\beta = 1/2$ is motivated by the worst-case inequality $T^\star_{\epsilon,1/2}(\nu) \leq 2T^\star_\epsilon(\nu)$ proven in Lemma 44. It implies that an asymptotically $1/2$-optimal algorithm has an expected sample complexity that is at worst twice as high as an asymptotically optimal algorithm. Moreover, this choice ensures that we sample the leader arm (i.e., the empirical estimate for $a^\star$) half of the time, and spend the remaining samples to explore all the other suboptimal arms (i.e., the challengers) until we are confident enough about $\tilde{a}_n$ being $a^\star$.

The default choice $\eta = 1$ is motivated by the trade-off existing between theoretical guarantees and empirical performance. Based on Theorem 7, smaller η yields "better" theoretical asymptotic guarantees, i.e., smaller upper bound on the expected sample complexity of DP-TT. However, the asymptotic regime of $\delta \to 0$ fails to capture the empirical trade-off for moderate values of δ. Choosing δ arbitrarily close to 0 will result in a large stopping threshold as k_η scales as $1/\log(1+\eta)$, see Eq. (8) in Theorem 6. Smaller η also implies that a larger cumulative Laplacian noise is added to the cumulative Bernoulli signal. The limit $\eta = 0$ coincides with adding one Laplace noise per Bernoulli observation. The concentration results in Appendix F.6 (Lemmas 18 and 19) show that the rate is governed by $\tilde{d}^\pm_\epsilon(\mu \pm x, \mu, 1)$, which is not equivalent to $d^\pm_\epsilon(\mu \pm x, \mu)$ asymptotically.

Remark 2. *To implement the thresholds of* AdaP-TT*,* AdaP-TT* *and DP-TT, we use empirical thresholds that we get by approximating the theoretical thresholds. The expressions of the empirical thresholds used can be found in the code in the supplementary material.*

J.2 Supplementary Experiments

Figure 2 confirms our experimental findings from Section 6. DP-TT outperforms all the other δ-correct and ϵ-global DP BAI algorithms, for different values of ϵ and in all the instances tested. The empirical performance of DP-TT demonstrates two regimes. A high-privacy regime, where the stopping time depends on the privacy budget ϵ, and a low privacy regime, where the performance of DP-TT is independent of ϵ, and requires twice the number of samples used by the non-private EB-TCI-β.

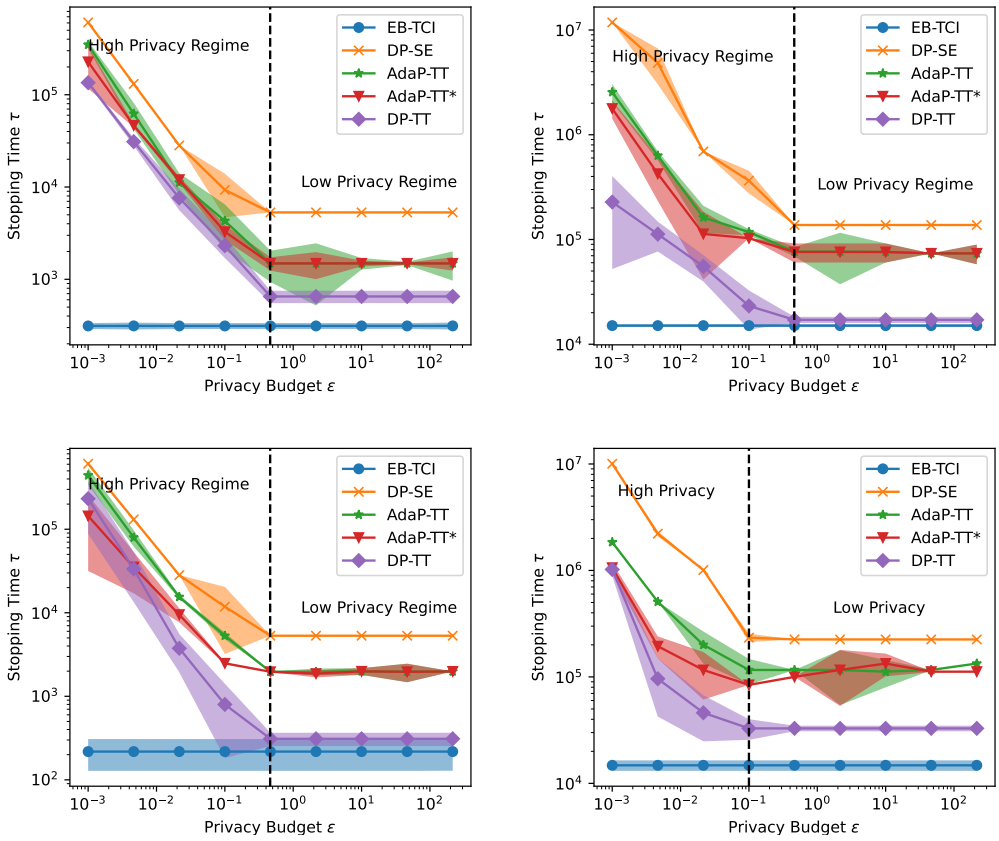

Figure 2: Empirical stopping time $\tau_{\epsilon,\delta}$ (mean ± 2 std. over 1000 runs) for $\delta = 10^{-2}$ with respect to the privacy budget ϵ for ϵ-global DP on Bernoulli instances μ_3, μ_4, μ_5 and μ_6 (top left to bottom right). The shaded vertical line separates the two privacy regimes.