Lower Bounds for Differentially Private Bandits and RL

Xingyu Zhou*

Abstract

Here is the abstract...

*Wayne State University, Detroit, USA. Email: xingyu.zhou@wayne.edu

Contents

1 Preliminaries

We now formally introduce the necessary notations and backgrounds for our main results on private MABs. The standard MAB protocol is given by Algorithm 1.

Algorithm 1 MAB Protocol

1: **Input:** A learning agent $\mathcal{M}$ and users $u_1, \ldots, u_T$.
2: **for** $t = 1, \ldots, T$ **do**
3: User u_t arrives
4: Agent prescribes action a_t to u_t
5: User u_t sends reward feedback r_t to agent
6: **end for**

Each user u_t in the user sequences $U = \{u_1, \ldots, u_T\}$ is identified by its K potential rewards denoted by vector x_t – one for each arm, although the learner can only observe one particular reward in this vector. Specifically, each element $x_{t,a}$ is an i.i.d sample from distribution P_a, and one gets to see that $r_t = x_{t,a_t}$. Let $X = [x_1, x_2, \ldots, x_T] \in \mathbb{R}^{K \times T}$ and $X' = [x'_1, x'_2, \ldots, x'_T] \in \mathbb{R}^{K \times T}$ be the corresponding dataset for user sequences U and U', respectively. We say U and U' are neighboring user sequences X if $\sum_{t=1}^{T} \sum_{k=1}^{K} \mathbb{1}_{\{x_{t,k} \neq x'_{t,k}\}} = 1$. Based on this, we have the following definitions of (central) (ε, δ)-DP and ρ-zCDP for private bandits.

Definition 1.1 ((ε, δ)-DP)**.** For any $\varepsilon, \delta \geq 0$, a randomized MAB algorithm $\mathcal{M}$ is (ε, δ)-DP if for any neighboring user sequences U and U' with corresponding datasets X and X' such that for all events $E \subset \mathcal{A}^T$, we have

$$\mathbb{P}\left[\mathcal{M}(U) \in E\right] \leq e^{\varepsilon} \mathbb{P}\left[\mathcal{M}(U') \in E\right] + \delta.$$

Definition 1.2 (ρ-zCDP)**.** For $\rho \geq 0$, a randomized MAB algorithm $\mathcal{M}$ is ρ-zCDP if for any neighboring user sequences U and U' with corresponding datasets X and X', we have we have for all $\alpha \in (1, \infty)$

$$\mathrm{D}_{\alpha}\left(\mathcal{M}(U) \middle\| \mathcal{M}(U')\right) \leq \rho\alpha,$$

where $\mathrm{D}_{\alpha}\left(P \| Q\right)$ is the Rényi divergence (of order α) of the distribution P from the distribution Q, and is given by $\mathrm{D}_{\alpha}\left(P \| Q\right) := \frac{1}{\alpha - 1} \log\left(\mathbb{E}_{x \sim Q}\left[\left(\frac{P(x)}{Q(x)}\right)^{\alpha}\right]\right)$.

The following facts on group privacy will be one main tool for us.

Lemma 1.3 ((ε, δ)-DP Group Privacy [DR14])**.** *For $\varepsilon, \delta \geq 0$, if a mechanism $\mathcal{M}$ is (ε, δ)-DP, then for all datasets D, D' with hamming distance $d_{\mathrm{ham}}\left(D, D'\right)$ and all measurable $\mathcal{S} \in$ Range$(\mathcal{M})$, we have*

$$\mathbb{P}\left[\mathcal{M}(D) \in \mathcal{S}\right] \leq e^{\varepsilon d_{\mathrm{ham}}(D, D')} \mathbb{P}\left[\mathcal{M}(D') \in E\right] + \delta d_{\mathrm{ham}}\left(D, D'\right) e^{\varepsilon (d_{\mathrm{ham}}(D, D') - 1)}.$$

Lemma 1.4 (ρ-zCDP Group Privacy [BS16])**.** *For $\rho \geq 0$, if a mechanism $\mathcal{M}$ is ρ-zCDP, then for all datasets D, D' with hamming distance $d_{\mathrm{ham}}\left(D, D'\right)$ and all all $\alpha \in (1, \infty)$, we have*

$$\mathrm{D}_{\alpha}\left(\mathcal{M}(D) \middle\| \mathcal{M}(D')\right) \leq \rho\alpha \left(d_{\mathrm{ham}}\left(D, D'\right)\right)^2$$

Minimax regret. Consider a given policy $\pi = \{\pi_t\}_{t\in[T]}$ where π_t maps history to a distribution over $[K]$ at time t and a given problem instance (environment) $\nu = \{P_a\}_{a\in[K]}$ where P_a is the distribution for arm a with mean μ_a. We then define the (expected pseudo) regret as

$$\mathcal{R}_{\nu,\pi}(T) := T\mu^* - \mathbb{E}_{\nu,\pi}\left[\sum_{t=1}^{T} r_t\right] = T\mu^* - \mathbb{E}_{\nu,\pi}\left[\sum_{t=1}^{T} \mu_{a_t}\right],$$

where $\mu^* = \max_{a\in[K]} \mu_a$. Minimax regret refers to taking sup over all MAB instances while taking inf over a class of policies. In this paper, we let $\Pi_{\mathcal{C}}$ be the class of MAB policies that satisfy a privacy constraint $\mathcal{C}$, i.e., (ε, δ)-DP or ρ-zCDP, and $\mathcal{E}_K$ be all MAB problem instances. The minimax regret is defined as follows.

$$\mathcal{R}_{\mathcal{C}}^{\text{minimax}}(T) := \inf_{\pi\in\Pi_{\mathcal{C}}} \sup_{\nu\in\mathcal{E}_K} \mathcal{R}_{\nu,\pi}(T).$$

2 Warm Up: Two-Arm Case

In this section, we first consider a simple two-arm case to illustrate the key ideas and main steps one needs to take to derive the minimax lower bounds in the private case.

Our main inspiration is recent advances in DP Le Cam [LGG22; ASZ21]. This is not a surprise since Le Cam can be easily utilized to show non-private minimax lower bounds for MABs[1]. Hence, one may consider directly applying results in [LGG22; ASZ21] to the settings of MABs and obtain corresponding private bounds. However, one key challenge is to handle the possible non-iid samples in MABs due to adaptive action selection. It is worth noting that although the main results in [LGG22; ASZ21] do not necessarily require i.i.d samples, it is difficult to arrive at a meaningful result for MABs without i.i.d samples, which is due to the difficulty in bounding the total variation (TV) distance. Another challenge is to handle the adaptive selection of the second test point for Le Cam so as to deal with $K \geq 2$, which will be discussed in the next section. For the two-arm case, we only need to handle the first challenge, i.e., possible dependence across reward samples.

To this end, our main technique is to resort to a different probability space. That is, instead of using the

Proof. As in Le Cam, we consider two test points (environments): $\nu_1 = (\text{Ber}(1/2+\Delta), \text{Ber}(1/2))$ and $\nu_2 = (\text{Ber}(1/2), \text{Ber}(1/2+\Delta))$ for some $\Delta \in (0, 1/2)$ chosen later. That is, arm i is the optimal arm for ν_i. Then, we have for any $\pi \in \Pi_{\mathcal{C}}$,

$$\begin{aligned}\sup_{\nu\in\mathcal{E}_K} \mathcal{R}_{\nu,\pi}(T) &\geq \frac{1}{2}\mathcal{R}_{\nu_1,\pi}(T) + \frac{1}{2}\mathcal{R}_{\nu_2,\pi}(T) \\ &= \frac{\Delta}{2}\sum_{t=1}^{T} \mathbb{E}_{\nu_1,\pi}\left[\mathbb{1}(a_t = 2)\right] + \mathbb{E}_{\nu_2,\pi}\left[\mathbb{1}(a_t = 1)\right] \qquad (1)\end{aligned}$$

Now we need to specify the probability space for the expectation. If the policy π is deterministic, one can follow tape-based construction in . In particular, we define a collection of mutually independent Bernoulli random variables $\{X_{i,a}\}_{i\in[T],a\in[K]}$ where $\mathbb{E}[X_{i,a}] = \mu_a$. We view $X_{i,a}$ as the reward received by the learning agent for the i-th time it chooses arm a. Given a deterministic policy, then any event of the algorithm can be captured by this probability space where the probability measure is a product distribution. However, for

[1] In fact, the text-book proof in [LS20] can be viewed as a variant of Le Cam where the second test point is chosen based on the first test point.

a differentially private algorithm, it must be randomized. Moreover, its randomness may depend on the samples. Thus, we model it as a probability kernel K between the space of reward tape and the space of the algorithm's outputs and let $\mathbb{P}_{\pi(X)} := K(\cdot|X)$. With this notation, we have $\mathbb{P}_{\nu,\pi} = \mathbb{P}_\nu \circ \mathbb{P}_{\pi(X)}$.

With the above definition of our probability space and its corresponding measure, each summand in (1) can be rewritten as

$$\begin{aligned}
\mathbb{E}_{\nu_1,\pi}\left[\mathbb{1}(a_t=2)\right] + \mathbb{E}_{\nu_2,\pi}\left[\mathbb{1}(a_t=1)\right] &\overset{(a)}{=} \mathbb{E}_{X\sim\mathbb{P}_{\nu_1}}\left[\mathbb{E}_{\pi(X)}\left[\mathbb{1}(a_t=2)\right]\right] + \mathbb{E}_{X\sim\mathbb{P}_{\nu_2}}\left[\mathbb{E}_{\pi(X)}\left[\mathbb{1}(a_t=1)\right]\right] \\
&\overset{(b)}{=} \mathbb{E}_{(X_1,X_2)\sim\mathbb{Q}(\mathbb{P}_{\nu_1},\mathbb{P}_{\nu_2})}\left[\mathbb{E}_{\pi(X_1)}\left[\mathbb{1}(a_t=2)\right] + \mathbb{E}_{\pi(X_2)}\left[\mathbb{1}(a_t=1)\right]\right] \\
&\overset{(c)}{=} \mathbb{E}_{(X_1,X_2)\sim\mathbb{Q}(\mathbb{P}_{\nu_1},\mathbb{P}_{\nu_2})}\left[\mathbb{P}_{\pi(X_1)}\left[a_t=2\right] + \mathbb{P}_{\pi(X_2)}\left[a_t=1\right]\right], \qquad (2)
\end{aligned}$$

where (a) holds by definition of $\mathbb{P}_{\nu,\pi}$; in (b), $\mathbb{Q}(\mathbb{P}_{\nu_1},\mathbb{P}_{\nu_2})$ is the maximal coupling between two product measures $\mathbb{P}_{\nu_1}$ and $\mathbb{P}_{\nu_2}$, which in fact can be easily constructed given the fact that each element in the product measure is a Bernoulli distribution. Specifically, we let $\{U_{i,a}\}_{i\in[T],a\in[K]}$ be mutually independent and identically distributed uniform random variables on $[0,1]$. Then, we have any (i,a)-th element of X_1 and X_2 are given by $\mathbb{1}(U_{i,a}\le\mu_{\nu_1,a})$ and $\mathbb{1}(U_{i,a}\le\mu_{\nu_2,a})$, respectively; in (c) we replace expectation of indicator by probability.

(ε,δ)-DP Case. We first look at the case (ε,δ)-DP. By the definition of $(\varepsilon,0)$-DP and group privacy in Lemma 1.3, we have for all X_1, X_2

$$\begin{aligned}
\mathbb{P}_{\pi(X_1)}\left[a_t=2\right] + \mathbb{P}_{\pi(X_2)}\left[a_t=1\right] &\ge e^{-\varepsilon d_{\text{ham}}(X_1,X_2)}\mathbb{P}_{\pi(X_2)}\left[a_t=2\right] - e^{-\varepsilon}\delta d_{\text{ham}}\left(D,D'\right) + \mathbb{P}_{\pi(X_2)}\left[a_t=1\right] \\
&\overset{(a)}{\ge} e^{-\varepsilon d_{\text{ham}}(X_1,X_2)}\left(\mathbb{P}_{\pi(X_2)}\left[a_t=2\right] + \mathbb{P}_{\pi(X_2)}\left[a_t=1\right]\right) - e^{-\varepsilon}\delta d_{\text{ham}}\left(D,D'\right) \\
&= e^{-\varepsilon d_{\text{ham}}(X_1,X_2)} - e^{-\varepsilon}\delta d_{\text{ham}}\left(D,D'\right), \qquad (3)
\end{aligned}$$

where (a) holds by $d_{\text{ham}}(\cdot,\cdot)\ge 0$. Plugging (3) into (2), yields that

$$\begin{aligned}
&\mathbb{E}_{\nu_1,\pi}\left[\mathbb{1}(a_t=2)\right] + \mathbb{E}_{\nu_2,\pi}\left[\mathbb{1}(a_t=1)\right] \\
\ge&\mathbb{E}_{(X_1,X_2)\sim\mathbb{Q}(\mathbb{P}_{\nu_1},\mathbb{P}_{\nu_2})}\left[e^{-\varepsilon d_{\text{ham}}(X_1,X_2)} - e^{-\varepsilon}\delta d_{\text{ham}}\left(D,D'\right)\right] \\
\overset{(a)}{\ge}& e^{-\varepsilon\mathbb{E}_{(X_1,X_2)\sim\mathbb{Q}(\mathbb{P}_{\nu_1},\mathbb{P}_{\nu_2})}[d_{\text{ham}}(X_1,X_2)]} - \mathbb{E}_{(X_1,X_2)\sim\mathbb{Q}(\mathbb{P}_{\nu_1},\mathbb{P}_{\nu_2})}\left[e^{-\varepsilon}\delta d_{\text{ham}}\left(D,D'\right)\right] \\
\overset{(b)}{=}& e^{-2\Delta T\varepsilon} - 2\Delta T e^{-\varepsilon}\delta \\
\overset{(c)}{\ge}& 1 - 2\Delta T\varepsilon - 2\Delta T e^{-\varepsilon}\delta \\
\overset{(d)}{\ge}& 1 - 2\Delta T(\varepsilon+\delta)
\end{aligned}$$

where (a) holds by Jensen's inequality; (b) holds by our construction of coupling and two considered environments ν_1 and ν_2; (c) holds from $e^x\ge 1+x$; (d) is true since $\varepsilon\ge 0$.

Plugging the above bound into (1), yields that for any π that is (ε,δ)-DP,

$$\sup_{\nu\in\mathcal{E}_K}\mathcal{R}_{\nu,\pi}(T) \ge \frac{\Delta T}{2}\cdot\left(1-2\Delta T(\varepsilon+\delta)\right).$$

Thus, choosing $\Delta = \frac{1}{4T(\varepsilon+\delta)}$, which satisfies $\Delta\in(0,1/2)$ if $T\ge\frac{1}{2(\varepsilon+\delta)}$. With this choice of Δ, we have that the minimax regret is $\Omega(\frac{1}{\varepsilon+\delta})$.

ρ-zCDP Case. We again continue from (2). In particular, we have

$$\begin{aligned}\mathbb{P}_{\pi(X_1)}\left[a_t=2\right]+\mathbb{P}_{\pi(X_2)}\left[a_t=1\right] &= 1-\left(\mathbb{P}_{\pi(X_1)}\left[a_t=1\right]-\mathbb{P}_{\pi(X_2)}\left[a_t=1\right]\right)\\ &\overset{(a)}{\geq} 1-\mathrm{TV}\left(\mathbb{P}_{\pi(X_1)},\mathbb{P}_{\pi(X_2)}\right)\\ &\overset{(b)}{\geq} 1-\sqrt{\mathrm{KL}\left(\mathbb{P}_{\pi(X_1)}\middle\|\,\mathbb{P}_{\pi(X_2)}\right)/2}\\ &\overset{(c)}{=} 1-\sqrt{\rho(d_{\mathrm{ham}}\left(X_1,X_2\right))^2/2}, \end{aligned} \tag{4}$$

where (a) holds by definition of TV distance; (b) holds by Pinsker's inequality; (c) holds by continuity as $\alpha\to 1$ and group privacy in Lemma 1.4. Note that here we use data processing inequality. Plugging (4) into (2), yields that

$$\begin{aligned}\mathbb{E}_{\nu_1,\pi}\left[\mathbb{1}(a_t=2)\right]+\mathbb{E}_{\nu_2,\pi}\left[\mathbb{1}(a_t=1)\right] &\geq \mathbb{E}_{(X_1,X_2)\sim\mathbb{Q}(\mathbb{P}_{\nu_1},\mathbb{P}_{\nu_2})}\left[1-d_{\mathrm{ham}}\left(X_1,X_2\right))\sqrt{\rho/2}\right]\\ &= 1-2\Delta T\sqrt{\rho/2}.\end{aligned}$$

Plugging this result into (1), yields that for any π that is ρ-zCDP,

$$\sup_{\nu\in\mathcal{E}_K}\mathcal{R}_{\nu,\pi}(T)\geq\frac{\Delta T}{2}\cdot\left(1-2\Delta T\sqrt{\rho/2}\right).$$

Thus, choosing $\Delta=\frac{1}{2\sqrt{2\rho T}}$, which is smaller than $1/2$ if $T>\frac{1}{\sqrt{2\rho}}$. With these choices, we have the minimax regret bound $\Omega\left(\frac{1}{\sqrt{\rho}}\right)$.

3 General Case

Proof. We consider $K+1$ problem instances: for $i\in[K]$, ν_i has a special arm i with a slightly larger mean, i.e., $\mathrm{Ber}(1/2)+\Delta$ and all other arms have distributions $\mathrm{Ber}(1/2)$. Moreover, we also consider one baseline instance ν_0, in which all arms have distributions $\mathrm{Ber}(1/2)$.

$$\sup_{\nu\in\mathcal{E}_K}\mathcal{R}_{\nu,\pi}(T)\geq\frac{1}{K}\sum_{i=1}^{K}\mathcal{R}_{\nu_i,\pi}(T)\geq\frac{\Delta}{K}\sum_{i=1}^{K}\sum_{t=1}^{T}\mathbb{P}_{\nu_i,\pi}\left[\mathbb{1}(a_t\neq i)\right] \tag{5}$$

We assume $K\geq 24$ since for any $K<24$, the lower bound established via two-arm case implies a valid bound (with a worse constant). The main idea is to show that for any policy π, with a proper choice of Δ, there always exists at least $K/3$ number of i in (5) such that for all $t\in[T]$, $\mathbb{P}_{\nu_i,\pi}\left[a_t\neq i\right]\geq 1/2$. To this end, we first identify such a special group of i by considering the baseline instance ν_0. In particular, we will establish the following claim.

Claim 3.1. *For any algorithm π, there are at least $K/3$ arms i such that for all $t\in[T]$, $\mathbb{P}_{\nu_0,\pi}\left[N_i(t)\leq\frac{24t}{K}\right]\geq 7/8$ and $\mathbb{P}_{\nu_0,\pi}\left[a_t=i\right]\leq 3/K$, where $N_i(t)$ is the number of pulls for arm i by time t.*

To show this, we first note that: (i) there are at least $2K/3$ arms such that $\mathbb{E}_{\nu_0,\pi}\left[N_i(t)\right]\leq 3t/K$ *and* (ii) there are at least $2K/3$ arms such that $\mathbb{P}_{\nu_0,\pi}\left[a_t=i\right]\leq 3/K$. Both can be shown via contradiction. That is, if there are at least $K/3$ arms such that $\mathbb{E}_{\nu_0,\pi}\left[N_i(t)\right]>3t/K$, then the total number of pulls is already larger than t. Similarly, if there are at least $K/3$ arms such that $\mathbb{P}_{\nu_0,\pi}\left[a_t=i\right]>3/K$, then the total probability is

larger than one. By Markov inequality, (i) implies that $\mathbb{P}_{\nu_0,\pi}\left[N_i(t) \le 24t/K\right] \ge 7/8$ for at least $K/3$ arms i. Since there are only K unique arms, there exist at least $K/3$ arms that satisfy both (i) and (ii), and hence the claim holds.

To continue, we will only focus on the special set of arms that satisfy Claim 3.1. This set of arm indexes may be different under different policies, but the size of the set is always larger than $K/3$. Our goal is to show that for each arm index i in this set, we have $\mathbb{P}_{\nu_i,\pi}\left[\mathbb{1}(a_t \neq i)\right] \ge 1/2$ for all $t \in [T]$.

Fix one index i in the set and for this arm i, the reward tape is only of size $m := 24T/K$ ($K \ge 24$) while all other arms still have size T. The law for this "truncated" sample space $\bar{\Omega}$ is denoted by $\bar{\mathbb{P}}$. For any event $A \in \bar{\Omega}$, we aim to establish

$$\bar{\mathbb{P}}_{\nu_i,\pi}(A) - \bar{\mathbb{P}}_{\nu_0,\pi}(A) \le 1/8 \tag{6}$$

under different privacy constraints of π by choosing Δ properly. Let us first assume that we have this result already and see how we can show that $\mathbb{P}_{\nu_i,\pi}\left[a_t \neq i\right] \ge 1/2$. Before we proceed, we also note that for any $A \subset \bar{\Omega} \subset \Omega$, we have $\bar{\mathbb{P}}_{\nu,\pi}\left[A\right] = \mathbb{P}_{\nu,\pi}\left[A\right]$.

Here, we cannot choose $A_t = \{a_t = i\}$ and then try to use (6). This is because $A_t \not\subset \bar{\Omega}$. To overcome this, we consider two events that are indeed in $\bar{\Omega}$ for all $t \in [T]$.

$$A_t = \{a_t = i \text{ and } N_i(t) \le m\} \text{ and } A_t' = \{N_i(t) > m\}.$$

For $t \le m = 24T/K$, it is obvious that $A_t, A_t' \subset \bar{\Omega}$; for $t > m = 24T/K$, we note that whether $N_i(t) > m$ is determined by the reward tape for arm i (of size m) and other reward tapes (of size T), hence $A_t' \in \bar{\Omega}$. Then, we have

$$\begin{aligned}
\mathbb{P}_{\nu_i,\pi}\left[a_t = i\right] &\le \mathbb{P}_{\nu_i,\pi}\left[A_t\right] + \mathbb{P}_{\nu_i,\pi}\left[A_t'\right] \\
&= \bar{\mathbb{P}}_{\nu_i,\pi}\left[A_t\right] + \bar{\mathbb{P}}_{\nu_i,\pi}\left[A_t'\right] \\
&\overset{(a)}{\le} \bar{\mathbb{P}}_{\nu_0,\pi}\left[A_t\right] + 1/8 + \bar{\mathbb{P}}_{\nu_0,\pi}\left[A_t'\right] + 1/8 \\
&\overset{(b)}{\le} 1/8 + 1/8 + 1/8 + 1/8 \\
&= 1/2
\end{aligned}$$

where (a) comes from (6); (b) holds by the choice of i, i.e., satisfying Claim 3.1. Hence, we have shown that for all the arms in the special set, $\mathbb{P}_{\nu_i,\pi}\left[a_t \neq i\right] \ge 1/2$ for all $t \in [T]$. Since the size of the set is at least $K/3$, we have

$$\sup_{\nu \in \mathcal{E}_K} \mathcal{R}_{\nu,\pi}(T) \ge \frac{1}{K}\sum_{i=1}^{K} \mathcal{R}_{\nu_i,\pi}(T) \ge \frac{\Delta}{K}\sum_{i=1}^{K}\sum_{t=1}^{T} \mathbb{P}_{\nu_i,\pi}\left[\mathbb{1}(a_t \neq i)\right] \ge \Omega(\Delta T). \tag{7}$$

We are only left to determine the choice of Δ such that (6) is true under different privacy constraints.

(ε, δ)-DP Case. Following the same argument as in (2), we have

$$\begin{aligned}
\bar{\mathbb{P}}_{\nu_i,\pi}(A) - \bar{\mathbb{P}}_{\nu_0,\pi}(A) &= \bar{\mathbb{E}}_{(X_i,X_0)\sim\mathbb{Q}(\mathbb{P}_{\nu_1},\mathbb{P}_{\nu_2})}\left[\bar{\mathbb{P}}_{\pi(X_i)}\left[A\right] - \bar{\mathbb{P}}_{\pi(X_0)}\left[A\right]\right] \\
&\overset{(a)}{=} \bar{\mathbb{E}}_{(X_i,X_0)\sim\mathbb{Q}(\mathbb{P}_{\nu_i},\mathbb{P}_{\nu_0})}\left[1 - \left(\bar{\mathbb{P}}_{\pi(X_i)}\left[A^c\right] + \bar{\mathbb{P}}_{\pi(X_0)}\left[A\right]\right)\right]
\end{aligned}$$

where in (a), A^c is the compliment of A. Following the same argument in (3), we have

$$\begin{aligned}\bar{\mathbb{P}}_{\nu_i,\pi}(A) - \bar{\mathbb{P}}_{\nu_0,\pi}(A) &\leq 1 - \left(e^{-\varepsilon\bar{\mathbb{E}}_{(X_i,X_0)\sim\bar{\mathbb{Q}}(\bar{\mathbb{P}}_{\nu_i},\bar{\mathbb{P}}_{\nu_0})}[d_{\text{ham}}(X_i,X_0)]} - \mathbb{E}_{(X_i,X_0)\sim\bar{\mathbb{Q}}(\bar{\mathbb{P}}_{\nu_i},\bar{\mathbb{P}}_{\nu_0})}\left[e^{-\varepsilon}\delta d_{\text{ham}}\left(D,D'\right)\right]\right) \\ &\overset{(a)}{=} 1 - \left(e^{-\varepsilon m\Delta} - \delta m\Delta e^{-\varepsilon}\right) \\ &\overset{(b)}{\leq} \varepsilon m\Delta + \delta m\Delta\end{aligned}$$

where (a) holds by maximal coupling and noting that now we are in the truncated sample space, hence m rather than T; (b) holds by $e^x \geq 1 + x$ and $\varepsilon \geq 0$. Recall that $m = 24T/K$, and hence choosing $\Delta = \frac{K}{192T(\varepsilon+\delta)}$ guarantees (6) holds. Plugging this Δ into (7), yields the lower bound $\Omega(K/(\varepsilon+\delta))$

ρ-zCDP Case. Following the similar argument as in the two-arm case, we have

□

$$
\begin{aligned}
&\mathbb{E}_{X\sim P_1,M}\left[f(M,X)\right]+\mathbb{E}_{X\sim P_2,M}\left[f(M,X)\right]\\
\overset{(a)}{=}&\mathbb{E}_{X\sim P_1}\left[\mathbb{E}_{M(X)}\left[f(M,X_1)\right]\right]+\mathbb{E}_{X\sim P_2}\left[\mathbb{E}_{M(X)}\left[f(M,X)\right]\right]\\
\overset{(b)}{=}&\mathbb{E}_{(X_1,X_2)\sim\mathbb{Q}}\left[\mathbb{E}_{M(X_1)}\left[f(M,X_1)\right]+\mathbb{E}_{M(X_2)}\left[f(M,X_2)\right]\right]
\end{aligned}
$$

□

References

[ASZ21] J. Acharya, Z. Sun, and H. Zhang. "Differentially private assouad, fano, and le cam". In: *Algorithmic Learning Theory*. PMLR. 2021, pp. 48–78.

[BS16] M. Bun and T. Steinke. "Concentrated differential privacy: Simplifications, extensions, and lower bounds". In: *Theory of Cryptography Conference*. Springer. 2016, pp. 635–658.

[DR14] C. Dwork and A. Roth. "The algorithmic foundations of differential privacy." In: *Found. Trends Theor. Comput. Sci.* 9.3-4 (2014), pp. 211–407.

[LGG22] C. Lalanne, A. Garivier, and R. Gribonval. "On the Statistical Complexity of Estimation and Testing under Privacy Constraints". In: *arXiv preprint arXiv:2210.02215* (2022).

[LS20] T. Lattimore and C. Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.