Improved Regret in Stochastic Decision-Theoretic Online Learning under Differential Privacy

Ruihan Wu RUW076@UCSD.EDU
University of California, San Diego

Yu-Xiang Wang YUXIANGW@UCSD.EDU
University of California, San Diego

Editors: Matus Telgarsky and Jonathan Ullman

Abstract

Hu and Mehta (2024) posed an open problem: *what is the optimal instance-dependent rate for the stochastic decision-theoretic online learning (with K actions and T rounds) under ε-differential privacy?* Before, the best known upper bound and lower bound are $O\left(\frac{\log K}{\Delta_{\min}} + \frac{\log K \log T}{\varepsilon}\right)$ and $\Omega\left(\frac{\log K}{\Delta_{\min}} + \frac{\log K}{\varepsilon}\right)$ (where $\Delta_{\min}$ is the gap between the optimal and the second actions). In this paper, we partially address this open problem by having two new results. First, we provide an improved upper bound for this problem $O\left(\frac{\log K}{\Delta_{\min}} + \frac{\log^2 K}{\varepsilon}\right)$, which is T-independent and only has a log dependency in K. Second, to further understand the gap, we introduce the *deterministic setting*, a weaker setting of this open problem, where the received loss vector is deterministic. At this weaker setting, a direct application of the analysis and algorithms from the original setting still leads to an extra log factor. We conduct a novel analysis which proves upper and lower bounds that match at $\Theta(\frac{\log K}{\varepsilon})$.

Keywords: Differential privacy, online learning

1. Introduction

Differential privacy (DP; Dwork et al. (2014)) provides a formal guarantee of data privacy, requiring that the outputs from any two datasets differing in a single individual's data do not differ significantly. In sequential decision-making settings, where the dataset consists of a sequence of observed losses or rewards, DP is extended to compare outputs from two sequences that differ at a single time step. DP has been extensively studied in two key sequential decision-making frameworks (online learning (Cesa-Bianchi and Lugosi, 2006; Arora et al., 2012) and multi-arm bandit (Lai and Robbins, 1985)) under various settings (Jain et al., 2012; Smith and Thakurta, 2013; Jain and Thakurta, 2014; Agarwal and Singh, 2017; Tossou and Dimitrakakis, 2017; Sajed and Sheffet, 2019; Hu et al., 2021; Asi et al., 2023b).

In this paper, we study stochastic decision-theoretic online learning (Freund and Schapire, 1997) under *pure* differential privacy, a setting identified as an *open problem* by Hu and Mehta (2024). In this problem, there are K actions, each associated with an unknown distribution of loss. At each time step, the learner selects an action and observes a stochastic loss drawn from its distribution. The problem is under the full-information setting, where the learner observes the stochastic losses of all actions after picking one, with the goal of minimizing the expected cumulative loss over time. A realistic instance of this problem arises in online product recommendation, where the platform selects one article to recommend to a user and observes their engagement (e.g., click or skip) as the

Table 1: A summarization of the previous existing results and our new results for the problem *stochastic decision-theoretic online learning under differential privacy*.

Settings	Lower bound	Upper bound
Instance-dependent bound for the *original setting*	$\frac{\log K}{\Delta_{\min}} + \frac{\log K}{\varepsilon}$ (Hu et al., 2021)	$\frac{\log K}{\Delta_{\min}} + \frac{\log K \log T}{\varepsilon}$ (Hu et al., 2021)
		$\frac{\log K}{\Delta_{\min}} + \frac{\log^2 K}{\varepsilon}$ (**This work**)
Instance-*in*dependent bound for the *original setting*	$\sqrt{T \log K} + \frac{\log K}{\varepsilon}$ (Hu et al., 2021)	$\sqrt{T \log K} + \frac{K \log K \log^2 T}{\varepsilon}$ (Jain and Thakurta, 2014)
		$\sqrt{T \log K} + \frac{\log K \log T}{\varepsilon}$ (Asi et al., 2023b; Hu et al., 2021)
		$\sqrt{T \log K} + \frac{\log^2 K}{\varepsilon}$ (**This work**)
The *deterministic setting*	$\frac{\log K}{\varepsilon}$ (**This work**)	$\frac{\log^2 K}{\varepsilon}$ (extended from our result in the original setting)
		$\frac{\log K}{\varepsilon}$ (**This work**)

Table 2: Detailed specifications in Algorithm 1 that achieve the exising result and our new results.

	Bernoulli resampling or not (B)	Noise distribution in report-noisy-max ($\mathcal{Q}_\varepsilon$)
Theorem 2 (Hu et al., 2021)	no	Laplace distribution
Theorem 3 (**This work**)	yes	Laplace distribution, Exponential distribution, Gumbel distribution
Theorem 11 (**This work**) (for *deterministic setting*)	no	Exponential distribution, Gumbel distribution

incurred loss. Concurrently, it logs feedback from other users who are independently exposed to the remaining products through alternative channels such as the general homepage, resulting in a full loss vector with one observed outcome per product. This full-information feedback supports efficient learning of recommendation policies, but also introduces privacy risks, as each loss vector reflects sensitive behavioral data from multiple individuals.

Jain and Thakurta (2014) provides an instance-independent bound $O\left(\sqrt{T \log K} + \frac{K \log K \log^2 T}{\varepsilon}\right)$ for general online linear optimization, which can be adapted as an upper bound for this problem. The best instance-independent bound so far for this problem is $O\left(\sqrt{T \log K} + \frac{\log K \log T}{\varepsilon}\right)$, achieved by Asi et al. (2023b) and Hu et al. (2021), where the lower bound is $O\left(\sqrt{T \log K} + \frac{\log K}{\varepsilon}\right)$. Particularly, the open problem (Hu and Mehta, 2024) asked for the instance-dependent bound in terms of $K, T, \varepsilon, \Delta_{\min}$, where $\Delta_{\min}$ is the gap of expected losses between the optimal and the second actions. The best existing instance-dependent bound is $O\left(\frac{\log K}{\Delta_{\min}} + \frac{\log K \log T}{\varepsilon}\right)$ (Hu et al., 2021) and the proved lower bound is $\Omega\left(\frac{\log K}{\Delta_{\min}} + \frac{\log K}{\varepsilon}\right)$. The algorithm in Hu et al. (2021) for these two bounds is quite standard: the algorithm applies a doubling metric to divide the time dimensions into epochs. At each epoch, it accumulates the observed loss vectors first, and uses a standard DP mechanism, report-noisy-max (Dwork et al., 2014) with Laplace noise, to pick an action for the whole next epoch. The algorithm is presented in Algorithm 1.

We propose a variant of the algorithm from Hu et al. (2021) that first resamples the stochastic loss vectors into Bernoulli variables before accumulating them. This modification enables an analysis of a new instance-dependent upper bound of $O\left(\frac{\log K}{\Delta_{\min}} + \frac{\log^2 K}{\varepsilon}\right)$. Compared to the best existing bound stated in Theorem 2, our result improves performance when $T > K$ (a small burn-in period). Notably,

Algorithm 1 Variants of RNM-FTNL(B, $\mathcal{Q}_\varepsilon$)

Specifying the variant: a bit $B \in \{0,1\}$ for indicating whether the loss vector is resampled or not; a noise distribution $\mathcal{Q}_\varepsilon$ parametrized by ε. \\ The original RNM-FTNL (Hu et al., 2021) can be recovered by setting $B = 0$ and $\mathcal{Q}_\varepsilon$ as the laplace distribution $\mathrm{Lap}(\frac{2}{\varepsilon})$.
Input: Action set $[K]$ and privacy parameter ε
Draw J_0 from a uniform distribution over $[K]$.
for $r = 1, \cdots, \lceil \log_2(T-1) \rceil + 1$ **do**
  Set $G_r = (0, \cdots, 0) \in \mathbb{R}^K$
  for $t = 2^{r-1}, \cdots, 2^r - 1$ **do**
    Play the action $I_t \leftarrow J_{r-1}$.
    Receive the loss vector $\ell^{(t)} = (\ell_1^{(t)}, \cdots, \ell_K^{(t)}) \sim \mathcal{P}_1 \times \cdots \times \mathcal{P}_K$.
    if $B = 0$ **then**
      $\tilde{\ell}^{(t)} \leftarrow \ell^{(t)}$
    else
      $\tilde{\ell}^{(t)} \leftarrow (\tilde{\ell}_1^{(t)}, \cdots, \tilde{\ell}_K^{(t)}) \sim \mathcal{B}(\ell_1^{(t)}) \times \cdots \times \mathcal{B}(\ell_K^{(t)})$, where $\mathcal{B}(p)$ is the Bernoulli distribution with mean p. \\ Bernoulli resampling
    end if
    $G_r \leftarrow G_r + \tilde{\ell}^{(t)}$
  end for
  $J_r \leftarrow \arg\max_{j \in K} -G_{r,j} + Q_{r,j}$ where $Q_{r,j} \sim \mathcal{Q}_\varepsilon$
end for

we are the first to demonstrate that the instance-dependent regret remains constant in T and depends only logarithmically on K, as the lower bound predicts. As a simple corollary, it also provides a new instance-independent upper bound $O\left(\sqrt{T \log K} + \frac{\log^2 K}{\varepsilon}\right)$ with a similar improvement.

By comparing the upper and lower bound, the extra factor appears alongside the DP parameter ε. To better understand the existence of the extra factor, we consider a strictly weaker setting in which the received loss vectors are deterministic, allowing us to isolate differential privacy (DP) from the stochasticity of observed losses. Although this setting is not realistic, directly applying the analysis and algorithms developed for the original open problem still yields the same extra factor. In this weaker setting, we introduce a new variant of the algorithm from Hu et al. (2021), replacing the Laplace noise in the report-noisy-max mechanism with either exponential or Gumbel noise. We prove a lower bound for this deterministic setting and show that the upper bound achieved by our new algorithm matches it, attaining a rate of $\frac{\log K}{\varepsilon}$. Finally, we discuss how these findings in the simplified setting offer insights into the original open problem.

We summarize both prior results and our new contributions in Table 1. Additionally, Table 2 outlines the specific variants of Algorithm 1 that achieve the existing and proposed results. **The organization of this paper**: Section 2 introduces the problem setting, the existing results, and the related work; Section 3 presents our new results for the open problem; Section 4 describes our new results for the deterministic setting of the open problem and discusses how these findings may inform further progress; The appendix contains additional proofs and technical details.

2. Preliminaries

2.1. Problem setting

In this paper, we focus on the open problem posed by Hu and Mehta (2024) and we begin by reviewing the problem setting in this section. The stochastic variant of decision-theoretic online learning (Freund and Schapire, 1997) assumes a finite set of K actions. Each action $i \in [K]$ is associated with an unknown loss distribution $\mathcal{P}_i$ that is unknown to the learner, whose support lies within $[0, 1]$ – this support assumption follows the same set-up as the open problem in Hu and Mehta (2024).

At each time step $t = 1, \cdots, T$:

1. The learner picks any action $I_t \in [K]$ according to any (randomized) algorithm $\mathcal{M}$.
2. The learning algorithm suffers loss $\ell_{I_t}^{(t)} \sim \mathcal{P}_{I_t}$.
3. The learner observes the losses of all the actions, a loss vector $\ell^{(t)} := (\ell_1^{(t)}, \cdots, \ell_K^{(t)}) \sim \mathcal{P}_1 \times \cdots \times \mathcal{P}_K$.

The goal is to minimize the pseudoregret $\mathrm{PseudoRegret}(\mathcal{A}; T, \mathcal{P}_1, \cdots, \mathcal{P}_K)$, which is the gap between the expectation of accumulated suffered losses and the minimum expectation of accumulated loss among K actions:

$$\mathbb{E}\left[\sum_{t=1}^{T} \ell_{I_t}^{(t)}\right] - \min_{i \in [K]} \mathbb{E}\left[\sum_{t=1}^{T} \ell_i^{(t)}\right],$$

where the randomness in the expectation is contributed by both the loss vector $\ell^{(t)}$ and the randomized algorithm $\mathcal{M}$. We further denote μ_i as the expectation of the loss from action i, $\mathbb{E}_{\ell_i \in \mathcal{P}_i}[\ell_i]$. Without the loss of generality, we assume $\mu^* = \mu_1 < \mu_2 \leq \cdots \mu_K$. Furthermore, we denote the gaps $\Delta_i := \mu_i - \mu_1$ and specifically, we denote the gap between the optimal and second optimal by $\Delta_{\min} := \mu_2 - \mu_1$. With the notations of gaps, the pseudoregret can be rewritten:

$$\mathrm{PseudoRegret}(\mathcal{A}; T, \mathcal{P}_1, \cdots, \mathcal{P}_K) = \mathbb{E}\left[\sum_{t=1}^{T} \mu_{I_t}\right] - T \cdot \mu_1 = \sum_{t=1}^{T} \mathbb{E}\left[\Delta_{I_t}\right]. \tag{1}$$

The optimal rate for the pseudoregret at this non-private setting is $\frac{\log(K)}{\Delta_{\min}}$, given by Kotłowski (2018); Mourtada and Gaïffas (2019).

In this paper, we study the problem under the framework of differential privacy (DP; Dwork et al. (2014)), a standard definition of privacy that requires the outcome distribution from the given randomized algorithm would not be changed too much if only one individual in the dataset has been changed. Particularly, differential privacy in online learning (Dwork et al., 2010) is *event-level*, which assumes the individual is the loss vector at a single time step t and the formal definition is as follow; also in this paper we only consider the *pure* DP rather than *approximate* DP, as set in the open prolem (Hu and Mehta, 2024).

Definition 1 (Differential privacy in online learning) *A randomized online learning algorithm $\mathcal{M}$ is ε-differentially private if for any two loss vector sequences $\ell^{(1:t)} = (\ell^{(\tau)})_{\tau \in [t]}$ and $(\ell')^{(1:t)}$ differing in at most one vector and any decision set $\mathcal{D}_{1:t} \subseteq [K]^t$, we have $\mathbb{P}[\mathcal{M}(\ell^{(1:t)}) \in \mathcal{D}_{1:t}] \leq e^{\varepsilon} \cdot \mathbb{P}[\mathcal{M}((\ell')^{(1:t)}) \in \mathcal{D}_{1:t}]$ for all $t \leq T$.*

We now state the open problem posed by Hu and Mehta (2024): for the stochastic variant of decision-theoretic online learning,

what is the optimal instance-dependent rate for the pseudoregret under ε-DP?

Or equivalently, what is the optimal rate in terms of $\varepsilon, \Delta_{\min}, K, T$ for the pseudoregret (Equation 1) that can be achieved by any algorithm?

2.2. Best existing results

The best lower bound for this open problem so far, proved by Hu et al. (2021), is

$$\Omega\left(\frac{\log K}{\Delta_{\min}} + \frac{\log K}{\varepsilon}\right). \tag{2}$$

The lower bound means that the pseudoregret of any ε-DP algorithm cannot have a better rate than this lower bound for all problem instances $(T, \mathcal{P}_1, \cdots, \mathcal{P}_k)$. Hu et al. (2021) also introduces the algorithm FNM-FTNL, which achieves the best rate so far for upper bounding the pseudoregret,

$$O\left(\frac{\log K}{\Delta_{\min}} + \frac{\log K \log T}{\varepsilon}\right).$$

We present their algorithm FNM-FTNL in Algorithm 1, by specifying $B = 0$ and the noise distribution $\mathcal{Q}_\varepsilon$ as the laplace distribution $\mathrm{Lap}(\frac{2}{\varepsilon})$; $\mathrm{Lap}(\beta)$ has the probability density function $f(x) = \frac{\beta}{2} e^{-\frac{|x|}{\beta}}$ for $x \in \mathbb{R}$. The algorithm applies a doubling trick to divide the time dimensions into epochs. At each epoch r, it accumulates the received loss vectors first and uses the report-noisy-max DP mechanism (Dwork et al., 2014) (with the laplace noise) to pick an action J_r for the next epoch $r+1$ while preserving the ε-DP guarantee. We formally state their results in the following theorem.

Theorem 2 (Best existing result; (Hu et al., 2021).) *When specifying $B = 0$ and $\mathcal{Q}_\varepsilon$ as the laplace distribution $\mathrm{Lap}(\frac{2}{\varepsilon})$, Algorithm 1 is ε-differentially private and satisfies the gaurantee*

$$\mathrm{PseudoRegret}(\textit{RNM-FTNL}(B, \mathcal{Q}_\varepsilon); T, \mathcal{P}_1, \cdots, \mathcal{P}_K) = O\left(\frac{\log K}{\Delta_{\min}} + \frac{\log K \log T}{\varepsilon}\right). \tag{3}$$

2.3. Related work

The open problem is considering one specific setting in private online prediction from experts (Asi et al., 2023b). Private online prediction from expert advice can have bandit setting and full information setting (Smith and Thakurta, 2013), based on assuming the learner observes the reward or loss only from the selected action at the time or from all actions. There are three models of adversaries at the full information setting from the strongest to the weakest: *adaptive* adversaries, who can decide the loss (distribution) upon from the picked action from the last time step (Jain et al., 2012; Smith and Thakurta, 2013; Jain and Thakurta, 2014; Agarwal and Singh, 2017; Asi et al., 2023b); *oblivious* adversaries, who decide a sequence of loss distributions before the online procedure (Asi et al., 2023b); *stochastic* adversaries, who pick one loss distribution and at each time step sample the loss i.i.d. from this distribution (Kairouz et al., 2021; Hu et al., 2021; Asi et al., 2023b). The open problem studied in this paper is at the full information setting with the stochastic adversary.

Private online prediction from experts is a special case of private online linear optimization (OLO) and private online convex optimization (OCO) (Smith and Thakurta, 2013; Agarwal and Singh, 2017; Kairouz et al., 2021; Agarwal et al., 2023; Asi et al., 2023a; Agarwal et al., 2024), where the optimization constraint is as an L1-sphere. Private OLO has been studied with different constraints too, such as the L2-ball or the cube, at both full-information setting and bandit setting.

3. Main Result for the Open Problem

3.1. An improved T-independent upper bound for the open problem.

New algorithm with the *Bernoulli resampling*. As introduced in Section 2.2, the original RNM-FTNL algorithm achieves the best known upper bound. Our improved regret rate is obtained by modifying RNM-FTNL with a key enhancement. While retaining the same doubling trick and accumulation of loss vectors, we introduce an additional step called *Bernoulli resampling*: each observed loss vector is resampled through a joint distribution of Bernoulli random variables, preserving the original expectation for each coordinate. That is, each coordinate in the resampled vector is a Bernoulli variable with mean equal to its corresponding observed loss. This modified algorithm is presented in Algorithm 1, with the resampling step specified by setting $B = 1$.

First of all, notice that Bernoulli resampling reduces all problem instances to a proper subset of instances with Bernoulli loss distributions. This means that the worst case of our new algorithm is *not worse* than the one of the original RNM-FTNL[1]. Moreover, the step of Bernoulli resampling strictly reduces the regret for some problem instances. Here is one example of such problem instance: Suppose there are two actions and their loss probabilities respectively are

$$\mathbb{P}[\ell_1 = 0.3] = 1; \mathbb{P}[\ell_2 = 0.4] = 0.8, \mathbb{P}[\ell_2 = 0] = 0.2;$$

In this example, action 1 is the optimal action with the lower expected loss. However, the action J_1 to take decided by the original RNM-FTNL (see line 16 in Algorithm 1) will weigh more on the suboptimal action 2: at an extreme case when there is no DP guarantee ($\varepsilon = \infty$), $P[J_1 = 2] = 0.8$. With the Bernoulli resampling in our new algorithm, a lemma introduced later states that the suboptimal action 2 will always have less chance than the optimal action 1 to be selected. This means that our algorithm will force $P[J_1 = 2] \leq 0.5 \leq P[J_1 = 1]$ to happen and as consequence, our algorithm will suffer smaller loss from this action.

Extending the algorithm by different DP mechanisms. In the original algorithm, the Laplace noise is used to satisfy DP. Our main result also shows that the same results will hold for other two DP mechanisms: adding Exponential noise or the Gumbel noise (known as Exponential mechanism). Denote the Exponential distribution by $\mathrm{Exp}(\beta)$ with the probability density function $f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$ for $x \geq 0$, and denote the Gumbel distribution by $\mathrm{Gumbel}(\beta)$ with the probability density function $f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta} - e^{-\frac{x}{\beta}}}$ for $x \in \mathbb{R}$. When $\mathcal{Q}_\varepsilon$ is specified as $\mathrm{Exp}(\frac{1}{\varepsilon})$ and $\mathrm{Gumbel}(\frac{2}{\varepsilon})$, similar to the analysis for $\mathcal{Q}_\varepsilon = \mathrm{Lap}(\frac{2}{\varepsilon})$, it is also proved that Algorithm 1 is ε-DP. This is because each J_r is ε-DP w.r.t. the received loss vectors in the last epoch (Dwork et al., 2014; Qiao et al., 2021) and the sets of loss vectors in each epoch are disjoint.

In a later section, we will show that using these two alternative noise distributions allows for a tighter analysis in a strictly weaker setting—an analysis that we are not able to conduct with the

1. It remains unknown if Bernoulli resampling strictly improves the worst case.

Laplace noise used in the original RNM-FTNL. This raises the possibility that these alternative noise distributions could also enable tighter analysis for the original open problem. For completeness, we will demonstrate that these two additional DP mechanisms achieve the same regret bounds as those obtained using Laplace noise.

Main result: new T-independent rate for the open problem. In summary, our main result will be built on our new algorithm, which extends the original FNM-FTNL by adding the Bernoulli resampling and other DP mechanisms. The exact specifications of B and $\mathcal{Q}_\varepsilon$ in Algorithm 1 for attaining our main result are listed in Table 2. We now formally state our main result and the proof idea will be introduced in the next subsection.

Theorem 3 (Main result: new rate for the open problem.) *When specifying $B = 1$ and $\mathcal{Q}_\varepsilon$ as the Laplace distribution $\mathrm{Lap}(\frac{2}{\varepsilon})$, the Exponential distribution $\mathrm{Exp}(\frac{1}{\varepsilon})$, or the Gumbel distribution $\mathrm{Gumbel}(\frac{2}{\varepsilon})$, Algorithm 1 is ε-differentially private and satisfies the guarantee*

$$\mathrm{PseudoRegret}(\textit{RNM-FTNL}(B, \mathcal{Q}_\varepsilon); T, \mathcal{P}_1, \cdots, \mathcal{P}_K) = O\left(\frac{\log K}{\Delta_{\min}} + \frac{\log^2 K}{\varepsilon}\right). \tag{4}$$

Remark 4 *First, by comparing our new upper bound with the best existing upper bound in Theorem 2, it improves over existing results when $T > K$ (a small burn-in period). Notably, we are the first to show that the instant-dependent regret remains constant in T and is only $\log K$ dependent, as the lower bound (Equation 2) predicts. From the analysis of Hu et al. (2021), we observe that it was possible to have a upper bound of $O\left(\frac{\log K}{\Delta_{\min}} + \frac{K \log K}{\varepsilon}\right)$ by bounding the number of groups in K, which is also T-independent. However, our bound is much tighter than their analysis, reduces the T-independent gap from K to $\log K$.*

The instance-dependent bound from Theorem 3 further imply a new instance-independent bound. We present the result in the following theorem; the proof follows the same steps as a similar corollary in Hu et al. (2021) and we put the proof in Appendix D.

Corollary 5 (New rate for the instance-independent bound) *When specifying $B = 1$ and $\mathcal{Q}_\varepsilon$ as the Laplace distribution $\mathrm{Lap}(\frac{2}{\varepsilon})$, the Exponential distribution $\mathrm{Exp}(\frac{1}{\varepsilon})$, or the Gumbel distribution $\mathrm{Gumbel}(\frac{2}{\varepsilon})$, Algorithm 1 is ε-differentially private and satisfies the guarantee*

$$\mathrm{PseudoRegret}(\textit{RNM-FTNL}(B, \mathcal{Q}_\varepsilon); T, \mathcal{P}_1, \cdots, \mathcal{P}_K) = O\left(\sqrt{T \log K} + \frac{\log^2 K}{\varepsilon}\right).$$

3.2. Proof of Theorem 3

A lemma given by Bernoulli resampling. We first introduce an important property given by the *Bernoulli resampling*. For any $j_1 < j_2$, i.e. the action j_1 has the smaller loss in expectation, suppose J_r is the output of report-noisy-max mechanism (line 16 in Algorithm 1) and the Bernoulli resampling will enforce the truth of $\mathbb{P}[J_r = j_1] \geq \mathbb{P}[J_r = j_2]$. In simple words, the better action always has larger chance to be selected as the action for the next epoch $r + 1$. Moreover, since $\mathbb{P}[J_r = j] \leq \mathbb{P}[J_r = j - 1] \leq \cdots \leq \mathbb{P}[J_r = 1]$, we have the upper bound $\mathbb{P}[J_r = j] \leq \frac{1}{j}$, which helps us derive a more fine-grained analysis. We formally state this conclusion in the next lemma.

Lemma 6 (Monotonicity for bionomial distributions) *Suppose J_r is the output from report-noisy-max, as defined at line 16 in Algorithm 1. When we specify Algorithm 1 by $B = 1$ and the noise distribution $\mathcal{Q}_\varepsilon$ is* $\mathrm{Lap}(\frac{2}{\varepsilon})$, $\mathrm{Exp}(\frac{1}{\varepsilon})$ *or* $\mathrm{Gumbel}(\frac{2}{\varepsilon})$. *For any $r \geq 1$ and $j_1 < j_2$, $\mathbb{P}[J_r = j_1] \geq \mathbb{P}[J_r = j_2]$. Moreover, $\mathbb{P}[J_r = j] \leq \frac{1}{j}$.*

The proof follows the steps

1. With the Bernoulli resampling, the accumulated loss $G_{r,j}$ has the binomial distribution $\mathcal{B}(2^{r-1}, \mu_j)$. Binomial distribution has the property (Wadsworth and Bryan (1960); Appendix A)
$$\mu_{j_1} \leq \mu_{j_2} \Rightarrow \forall x, F_{G_{r,j_1}}(x) \geq F_{G_{r,j_2}}(x).$$
where $F_A(x)$ is denoted as the cumulative density function for any random variable $\mathbb{P}[A \leq x]$.
2. Denote $N_{r,j} := -G_{r,j} + Q_{r,j}$. Because $Q_{r,j}$ share the same distribution $\mathcal{Q}_\varepsilon$ and $F_{G_{r,j_1}}(x) \geq F_{G_{r,j_2}}(x)$ from the step 1, we can prove that $F_{N_{r,j_1}}(x) \leq F_{N_{r,j_2}}(x)$.
3. Let $H = \max_{j \neq j_1, j_2} N_{r,j}$. By applying $F_{N_{r,j_1}}(x) \leq F_{N_{r,j_2}}(x)$ from the step 2, we can prove
$$\mathbb{P}[J_r = j_1] = \mathbb{P}[N_{r,j_1} > \max\{N_{r,j_2}, H\}] \geq \mathbb{P}[N_{r,j_2} > \max\{N_{r,j_1}, H\}] = \mathbb{P}[J_r = j_2].$$

We put the full proof in the Appendix B.

Proof sketch of Theorem 3. Now we show the proof sketch for Theorem 3 by omitting some calculations that are similar to the proof in Hu and Mehta (2024); the complete proof is in Appendix C.
Proof [Proof sketch of Theorem 3.] The Algorithm 3 is ε-differentially private as discussed at the beginning of this section. Next, we are going to bound the pseudoregret. If we can prove Equation 4 for any $T := 2^R - 1$ where R is any non-negative integer, Equation 4 would also hold for arbitrary T, because Algorithm 1 is independent of the T and the regret of Algorithm 1 is non-decreasing in T. Therefore, we can assume $T := 2^{R+1} - 1$ for some non-negative integer R and can rewrite the pseudoregret (defined in Equation 1) according to the Algorithm 1:

$$\sum_{t=1}^{T} \mathbb{E}\left[\Delta_{I_t}\right] = \sum_{r=1}^{R} \sum_{t=2^{r-1}}^{2^r-1} \sum_{j=1}^{K} \Delta_j \mathbb{P}[J_{r-1} = j] = \sum_{j=1}^{K} \Delta_j \sum_{r=1}^{R} 2^{r-1} \mathbb{P}[J_{r-1} = j]$$
$$= \underbrace{\sum_{j:\Delta_j \leq \varepsilon} \Delta_j \sum_{r=1}^{R} 2^{r-1} \mathbb{P}[J_{r-1} = j]}_{\text{Regret}_\uparrow} + \underbrace{\sum_{j:\Delta_j > \varepsilon} \Delta_j \sum_{r=1}^{R} 2^{r-1} \mathbb{P}[J_{r-1} = j]}_{\text{Regret}_\downarrow}$$

$\text{Regret}_\uparrow \leq O\left(\frac{\log K}{\Delta_{\min}}\right)$ can be shown in the same way as a part of proof for Theorem 9 in Hu et al. (2021). The remaining is to focus on bounding $\text{Regret}_\downarrow$, where Lemma 6 is applied.

According to the Lemma 9 in Hu et al. (2021) and Lemma 16 (in Appendix), for all three noise distributions $\mathcal{Q}_\varepsilon$, there exists universal constants $c_1, c_2 > 0$ s.t.

$$\mathbb{P}\left[J_r = j\right] \leq c_1 \cdot \exp(-2^{r+1} \Delta_j \min\{\Delta_j, \varepsilon\}/c_2). \tag{5}$$

With this property and a similar proof for Theorem 24 in Hu et al. (2021), for $\Delta_j, \varepsilon > 0$ and $r(j) \in \mathbb{N}$,

$$\sum_{r=r(j)+1}^{R} 2^{r-1} \mathbb{P}\left[J_{r-1} = j\right] < \frac{c_1 c_2}{\Delta_j \min\{\Delta_j, \varepsilon\}} \cdot \exp(-2^{r(j)} \Delta_j \min\{\Delta_j, \varepsilon\}/c_2). \tag{6}$$

Let $r(j) = \left\lceil \log_2 \left(\frac{c_2 (\log K)}{\Delta_j \varepsilon} \right) \right\rceil$. $\forall j$ such that $\Delta_j > \varepsilon$, $\sum_{r=1}^{R} 2^{r-1} \mathbb{P}[J_{r-1} = j]$ can be bounded as

$$\sum_{r=1}^{r(j)} 2^{r-1} \mathbb{P}[J_{r-1} = j] + \sum_{r=r(j)+1}^{R} 2^{r-1} \mathbb{P}[J_{r-1} = j] < \left(\sum_{r=1}^{r(j)} 2^{r-1} \frac{1}{j} \right) + \frac{c_1 c_2}{\Delta_j \varepsilon} \cdot \exp(-2^{r(j)} \Delta_j \varepsilon / c_2)$$

$$< 2^{r(j)} \frac{1}{j} + \frac{c_1 c_2}{\Delta_j \varepsilon} \cdot \exp(-2^{r(j)} \Delta_j \varepsilon / c_2) \leq \frac{2 c_2}{\Delta_j \varepsilon} \cdot \frac{\log K}{j} + \frac{c_1 c_2}{\Delta_j \varepsilon} \cdot \frac{1}{K},$$

where the first inequality holds by Lemma 6 (*since it is assumed that $B = 1$ for the Algorithm 1 in the theorem statement*) and Equation 6, the second inequality holds by $\sum_{r=1}^{r(j)} 2^{r-1} = 2^{r(j)} - 1 < 2^{r(j)}$, and the third inequality holds by taking the value of $r(j)$. Therefore,

$$\text{Regret}_{\downarrow} < \sum_{j:\Delta_j > \varepsilon}^{K} \Delta_j \left(\frac{2 c_2}{\Delta_j \varepsilon} \frac{\log K}{j} + \frac{c_1 c_2}{\Delta_j \varepsilon} \frac{1}{K} \right) \leq \frac{2 c_2}{\varepsilon} \sum_{j:\Delta_j > \varepsilon}^{K} \left(\frac{\log K}{j} + \frac{c_1}{K} \right) = O\left(\frac{\log^2 K}{\varepsilon} \right).$$

By putting the analysis for $\text{Regret}_{\uparrow}$ and $\text{Regret}_{\downarrow}$ together, we have proved that the pseudoregret is bounded by $O\left(\frac{\log K}{\Delta_{\min}} + \frac{\log^2 K}{\varepsilon} \right)$. ■

4. Optimal Rate for a Weaker Deterministic Setting of the Open Problem

We notice that for both the existing result and our new result, the gap the the known lower bound appears with the DP factor ε rather than the $\Delta_{\min}$. This motivates us to study a weaker setting of the open problem to focus on differential privacy regardless of the sampling error in the observed losses. Specifically, we study the same open problem but with the assumption that the distributions $\mathcal{P}_j$ ($j \in [K]$) concentrate on the single value μ_j, i.e. $\mathbb{P}_{\ell_j \sim \mathcal{P}_j}[\ell_j = \mu_j] = 1$, and we call this weaker setting as the *deterministic setting*.

Although the deterministic setting may seem limited in practical relevance, it provides a simplified environment to isolate and examine key challenges in the original problem. Interestingly, a direct application of existing analyses and algorithms from the original setting still results in the same extra $\log K$ factor, suggesting that the underlying complexity persists. Motivated by this, we undertake a deeper investigation of this weaker setting.

4.1. Lower bound for the deterministic setting

Although Hu et al. (2021) has shown the lower bound $\Omega\left(\frac{\log K}{\varepsilon} \right)$ for the original open problem, their result does not directly apply to the deterministic setting, as the worst-case instance they construct falls outside this weaker setting. Their construction relies on stochasticity and does not satisfy the assumptions of the deterministic case. Instead, we develop a new lower bound instance that achieves the same rate of $\frac{\log K}{\varepsilon}$, and it applies to both the deterministic setting and the original problem. The lower bound result is stated in the following theorem.

Theorem 7 (Lower bound for the deterministic setting.) *For any ε-differentially private online learning algorithm $\mathcal{M}$, $K \in \mathbb{N}$ and $\Delta_{\min}$, $\exists (u_1, \cdots, u_K) \in [0,1]^K$ s.t. at the deterministic setting,*

$$\text{PseudoRegret}(\mathcal{M}; T, \mathcal{P}_1, \cdots, \mathcal{P}_K) \geq c_1 \frac{\log K}{\varepsilon},$$

where c_1 is a universal constant independent of K,ε and $(\mu_1, \cdots, \mu_k)$. Moreover, the sorted $(u_{(1)}, \cdots, u_{(K)})$ in the worst instance construction is $(0, \Delta_{\min}, \Delta_{\min}, 1, \cdots, 1)$

Remark 8 *While the lower bound rate remains the same as in prior work, the proof is simpler and more self-contained, without needing to use an interesting DP version of Fanos' method (Acharya et al., 2021) that apparently overkill the problem.*

Remark 9 (Connection between Theorem 7 and open problem) *A key mystery in prior work is that the worst-case instance for the non-private version of the problem is given by $(u_{(1)}, \cdots, u_{(K)}) = (0, \Delta_{\min}, \cdots, \Delta_{\min})$, where all suboptimal actions share the same loss distributions. However, this instance does not correspond to the hardest case (to come up with an analysis) for the private version of the problem. – prior analysis in Hu et al. (2021) has proved the tightest rate $\Theta\left(\frac{\log K}{\Delta_{\min}} + \frac{\log K}{\varepsilon}\right)$ for this particular instance, but not able to prove this rate for all instances. Our new lower bound construction sheds light on this discrepancy: the set of instances exhibiting regret of at least $\Omega(\frac{\log K}{\varepsilon})$ under privacy constraints is broader than the hardest instances in the non-private setting. As a result, establishing the $O(\frac{\log K}{\varepsilon})$ term in the upper bound for all instances is inherently more difficult than analyzing just the worst-case instance in the non-private regime.*

4.2. Upper bound for the deterministic setting

Suboptimal results by extending the analysis for the original open problem. First, we will see how the algorithm together with the similar analysis in Section 3 works for the deterministic setting. We can repeat the analysis in Theorem 3 without considering the sampling errors, and get the following rate as a corollary; the detailed argument is in Appendix F.

Corollary 10 (Extension from Theorem 3.) *When $\mathcal{Q}_\varepsilon$ is the Laplace distribution $\mathrm{Lap}(\frac{2}{\varepsilon})$, the Exponential distribution $\mathrm{Exp}(\frac{1}{\varepsilon})$, or the Gumbel distribution $\mathrm{Gumbel}(\frac{2}{\varepsilon})$, Algorithm 1 is ε-differentially private and satisfies the guarantee for the deterministic setting:*

$$\mathrm{PseudoRegret}(\textit{RNM-FTNL}(B, \mathcal{Q}_\varepsilon); T, \mathcal{P}_1, \cdots, \mathcal{P}_K) = O\left(\frac{\log^2 K}{\varepsilon}\right)$$

Unfortunately, by comparing the rate with the lower bound, there is still an extra log factor in K, same as what it is for the original open problem.

The tight upper bound for the deterministic setting. We first provide an algorithm in the same framework of variants of RNM-FTNL. The algorithm is with a new specification of $B = 0$ and Q_ε as Exponential distribution or Gumbel distribution in Algorithm 1. Notice that we are not sticking with the Laplace distribution that is used in the original RNM-FTNL (Hu et al., 2021) for this setting. This is because the report-noisy-max mechanism with Gumbel noise is known as Exponential mechanism (Qiao et al., 2021), which has explicit forms for the probability of each action as an output. The tractable expression of the regret is soft-max like and let us to derive our tight analysis. In addition, we can make a similar conclusion for the Exponential distribution by a reduction since the previous study (McKenna and Sheldon, 2020) shows it is consistently better than the Gumbel distribution. Nevertheless, we are not able to prove the same rate for the Laplace distribution, and whether it brings the same rate remains unknown.

With this new algorithm, we are able to prove the optimal rate for the deterministic setting, as stated in the following theorem.

Theorem 11 (Main result 2: optimal rate for the deterministic setting.) *When specifying* $B = 0$ *and* $\mathcal{Q}_\varepsilon$ *as the Exponential distribution* $\mathrm{Exp}(\frac{1}{\varepsilon})$ *or the Gumbel distribution* $\mathrm{Gumbel}(\frac{2}{\varepsilon})$*, Algorithm 1 is* ε*-differentially private and satisfies the guarantee for the deterministic setting*

$$\mathrm{PseudoRegret}(\textit{RNM-FTNL}(B, \mathcal{Q}_\varepsilon); T, \mathcal{P}_1, \cdots, \mathcal{P}_K) = O\left(\frac{\log K}{\varepsilon}\right).$$

Moreover, this rate is optimal for the deterministic setting.

We show two lemmas for the soft-max like function $f(x) = \frac{\sum_{i=1}^K 2^x a_i e^{-2^x a_i}}{\sum_{i=1}^K e^{-2^x a_i}}$; proofs are done by some calculas in Section G.

Lemma 12 *For any* $i \in [K], a_i \in \mathbb{R}$*,* $f(x) = \frac{\sum_{i=1}^K 2^x a_i e^{-2^x a_i}}{\sum_{i=1}^K e^{-2^x a_i}}$ *has the property* $f'(x) \leq \log 2 \cdot f(x)$*.*

Lemma 13 *For any* $0 = a_1 < a_2 \leq, \cdots, \leq a_K$*,* $\sum_{r=1}^\infty \frac{\sum_{i=1}^K 2^r a_i \exp(-2^r a_i)}{\sum_{i=1}^K \exp(-2^r a_i)} \leq O(\log K)$*.*

Proof [Proof sketch for Theorem 11] Report-noisy-max mechanism with Gumbel noise is equivalent to Exponential mechanism (Gumbel, 1954; Qiao et al., 2021), so we can have a tractable expression for $\mathbb{P}[J_r = j]$: $\mathbb{P}\left[J_r = j | \forall i \in [K], G_{r,i}\right] = \frac{\exp(\varepsilon \cdot (-G_{r,j}))}{\sum_{i=1}^K \exp(\varepsilon \cdot (-G_{r,i}))}$. Moreover, at the deterministic setting, $\mathbb{P}[J_r = j] = \mathbb{P}\left[J_r = j | \forall i \in [K], G_{r,i} = 2^{r-2}\mu_i\right]$. Therefore the regret has this tractable expression:

$$O(1) + \sum_{r=3}^R \sum_{j=1}^K 2^{r-1} \Delta_j \frac{\exp\left(-2^{r-2}\Delta_j \varepsilon\right)}{\sum_{i=1}^K \exp\left(-2^{r-2}\Delta_i \varepsilon\right)}.$$

If we define this soft-max like function $f(x) := \frac{\sum_{i=1}^K 2^x a_i e^{-2^x a_i}}{\sum_{i=1}^K e^{-2^x a_i}}$, Lemma 13 proves $\forall 0 = a_1 < a_2 \leq \cdots \leq a_k$, $\sum_{x=1}^\infty f(x) = O(\log K)$. We can specify $a_i = \Delta_i \varepsilon$ and finish proving that the above regret is $O(\frac{\log K}{\varepsilon})$. The full proof with the detailed calculation and for the Exponential noise is in Appendix H. ■

Remark 14 *We connect Theorem 11 back to the original open problem by the points below.*

1. *We provide a tight analysis for this deterministic setting, a special setting of the original open problem, while the analysis from the stochastic setting implies a suboptimal result.*
2. *Comparing the analyses across the two settings, we hypothesize that the current analysis for the original open problem may be loose due to its simplification – specifically, it considers losses only in relation to each action and the optimal one (Equation 5). In contrast, the tighter analysis for the deterministic setting jointly considers all actions through a softmax-like function* $f(x)$ *when analyzing the regret. This suggests that a more holistic treatment of the action set may be necessary to achieve tighter bounds in the original setting.*
3. *Motivated by the above discussion, we propose the following conjecture for the original open problem: for all* j *s.t.* $\varepsilon < \Delta_j$*,* $\mathbb{P}\left[J_r = j\right] \leq c_1 \cdot \frac{\exp\left(-2^{r+1}\Delta_j \varepsilon / c_2\right)}{1 + \sum_{j:\Delta_j > \varepsilon} \exp(-2^{r+1}\Delta_i \varepsilon / c_2)}$*. This conjecture is strictly stronger than Equation 5. If it holds, then the open problem could be resolved with a regret bound of* $\Theta\left(\frac{\log K}{\Delta_{\min}} + \frac{\log K}{\varepsilon}\right)$*, using an argument similar to our analysis for the deterministic setting. At present, however, we are unable to prove or disprove the conjecture.*

5. Conclusion

In this paper, we propose a new upper bound for the open problem that is independent of T and depends only logarithmically on K. In addition, we focus on a weaker variant of the problem, where the losses are assumed to be deterministic. We present a new analysis for the deterministic setting and establish a tight bound, offering insights that may inform future progress on the original open problem.

References

Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Differentially private assouad, fano, and le cam. In *Algorithmic Learning Theory*, pages 48–78. PMLR, 2021.

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

Naman Agarwal, Satyen Kale, Karan Singh, and Abhradeep Thakurta. Differentially private and lazy online convex optimization. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 4599–4632. PMLR, 2023.

Naman Agarwal, Satyen Kale, Karan Singh, and Abhradeep Guha Thakurta. Improved differentially private and lazy online convex optimization: Lower regret without smoothness requirements. In *Proceedings of the 41st International Conference on Machine Learning*, pages 343–361. PMLR, 2024.

Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of computing*, 8(1):121–164, 2012.

Hilal Asi, Vitaly Feldman, Tomer Koren, and Kunal Talwar. Near-optimal algorithms for private online optimization in the realizable regime. In *International Conference on Machine Learning*, pages 1107–1120. PMLR, 2023a.

Hilal Asi, Vitaly Feldman, Tomer Koren, and Kunal Talwar. Private online prediction from experts: Separations and faster rates. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 674–699. PMLR, 2023b.

Jean-Yves Audibert and Sébastien Bubeck. Minimax policies for adversarial and stochastic bandits. In *COLT*, pages 217–226, 2009.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

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

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

Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

Emil Julius Gumbel. Statistical theory of extreme valuse and some practical applications. *Nat. Bur. Standards Appl. Math. Ser. 33*, 1954.

Bingshan Hu and Nishant A Mehta. Open problem: Optimal rates for stochastic decision-theoretic online learning under differentially privacy. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 5330–5334. PMLR, 2024.

Bingshan Hu, Zhiming Huang, and Nishant A Mehta. Near-optimal algorithms for private online learning in a stochastic environment. *arXiv preprint arXiv:2102.07929*, 2021.

Prateek Jain and Abhradeep Guha Thakurta. (near) dimension independent risk bounds for differentially private learning. In *International Conference on Machine Learning*, pages 476–484. PMLR, 2014.

Prateek Jain, Pravesh Kothari, and Abhradeep Thakurta. Differentially private online learning. In *Conference on Learning Theory*, pages 24–1. JMLR Workshop and Conference Proceedings, 2012.

Peter Kairouz, Brendan McMahan, Shuang Song, Om Thakkar, Abhradeep Thakurta, and Zheng Xu. Practical and private (deep) learning without sampling or shuffling. In *International Conference on Machine Learning*, pages 5213–5225. PMLR, 2021.

Wojciech Kotłowski. On minimaxity of follow the leader strategy in the stochastic setting. *Theoretical Computer Science*, 742:50–65, 2018.

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

Ryan McKenna and Daniel R Sheldon. Permute-and-flip: A new mechanism for differentially private selection. *Advances in Neural Information Processing Systems*, 33:193–203, 2020.

Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 94–103. IEEE, 2007.

Jaouad Mourtada and Stéphane Gaïffas. On the optimality of the hedge algorithm in the stochastic regime. *Journal of Machine Learning Research*, 20(83):1–28, 2019.

Gang Qiao, Weijie Su, and Li Zhang. Oneshot differentially private top-k selection. In *International Conference on Machine Learning*, pages 8672–8681. PMLR, 2021.

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.

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

Aristide Tossou and Christos Dimitrakakis. Achieving privacy in the adversarial multi-armed bandit. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

George Proctor Wadsworth and Joseph G Bryan. *Introduction to probability and random variables*, volume 7. McGraw-Hill New York, 1960.

Appendix A. Proof of the Property of Binomial Distribution

Lemma 15 *Suppose $F(k;n,p)$ is the cumulative density function (CDF) of the binomial distribution $\mathcal{B}(n,p)$. For any $0 \leq p_1 < p_2 \leq 1$, $F(k;n,p_1) \geq F(k;n,p_2)$.*

Proof [Proof of Lemma 15.] Suppose $F_{\text{beta}-\text{dist}}(x;\alpha,\beta)$ is the CDF of beta-distribution. It has been proved the equivalence between the two CDFs (Wadsworth and Bryan, 1960):

$$F(k;n,p) = F_{\text{beta}-\text{dist}}(1-p;n-k,k+1).$$

Therefore, for any $p_1 < p_2$,

$$F(k;n,p_1) = F_{\text{beta}-\text{dist}}(1-p_1;n-k,k+1) \geq F_{\text{beta}-\text{dist}}(1-p_2;n-k,k+1) = F(k;n,p_2)$$

■

Appendix B. Proof of Lemma 6

Proof [Proof of Lemma 6.] Let $N_{r,j} = -G_{r,j} + Q_{r,j}$ and denote $F_A(x)$ as the cumulative density function for any random variable $\mathbb{P}[A \leq x]$. We can first prove for any $j_1 < j_2$ and $x \in \mathbb{R}$, $F_{N_{r,j_1}}(x) \leq F_{N_{r,j_2}}(x)$. To see its correctness, we can decompose $F_{N_{r,j_1}}(x)$ as

$$F_{N_{r,j_2}}(x) = \int_{-\infty}^{\infty} \mathbb{P}[-G_{r,j_1} \leq x - s] f_{Q_{r,j_1}}(s) ds = \int_{-\infty}^{\infty} (1 - F_{G_{r,j_1}}(s-x)) f_{Q_{r,j_1}}(s) ds$$

and similarly $F_{N_{r,j_2}}(x) = \int_{-\infty}^{\infty} (1 - F_{G_{r,j_2}}(s-x)) f_{Q_{r,j_2}}(s) ds$.

Moreover, *because $B = 1$ is specified for the algorithm*, $G_{r,j}$ is from the binomial distribution $\mathcal{B}(2^{r-1}, \mu_j)$. Binomial distribution has the property (Wadsworth and Bryan (1960); Appendix A)

$$\mu_{j_1} \leq \mu_{j_2} \Rightarrow F_{G_{r,j_1}}(x) \geq F_{G_{r,j_2}}(x).$$

With this property, we can show $F_{N_{r,j_1}}(x) \leq F_{N_{r,j_2}}(x)$ by

$$\int_{-\infty}^{\infty} (1 - F_{G_{r,j_1}}(s-x)) f_{Q_{r,j_1}}(s) ds \leq \int_{-\infty}^{\infty} (1 - F_{G_{r,j_2}}(s-x)) f_{Q_{r,j_2}}(s) ds$$

Now we turn to prove $\mathbb{P}[J_r = j_1] \geq \mathbb{P}[J_r = j_2]$ for $j_1 < j_2$. Let $H = \max_{j \neq j_1, j_2} N_{r,j}$ and let N'_{r,j_2} be a random variable which is independent of N_{r,j_2} but has the same distribution as N_{r,j_2}. By applying $F_{N_{r,j_1}}(x) \leq F_{N_{r,j_2}}(x)$ proved above, we have

$$\begin{aligned}
\mathbb{P}[J_r = j_1] &= \mathbb{P}[N_{r,j_1} > \max\{N_{r,j_2}, H\}] = \int_{-\infty}^{\infty} (1 - F_{N_{r,j_1}}(s)) f_{\max\{N_{r,j_2},H\}}(s) ds \\
&\geq \int_{-\infty}^{\infty} (1 - F_{N'_{r,j_2}}(s)) f_{\max\{N_{r,j_2},H\}}(s) ds \\
&= \mathbb{P}[N'_{r,j_2} > \max\{N_{r,j_2}, H\}] = \mathbb{P}[N_{r,j_2} > \max\{N'_{r,j_2}, H\}].
\end{aligned}$$

Because H and N'_{r,j_2} are independent, by applying $F_{N_{r,j_1}}(x) \leq F_{N_{r,j_2}}(x) = F_{N'_{r,j_2}}(x)$ again, $F_{\max\{N'_{r,j_2},H\}}(x) = F_{N'_{r,j_2}}(x) \cdot F_H(x) \geq F_{N_{r,j_1}}(x) \cdot F_H(x) = F_{\max\{N_{r,j_1},H\}}(x)$. Therefore

$$\begin{aligned}\mathbb{P}[J_r = j_1] &\geq \mathbb{P}[N_{r,j_2} > \max\{N'_{r,j_2}, H\}] = \int_{-\infty}^{\infty} F_{\max\{N'_{r,j_2},H\}}(s) f_{N_{r,j_2}}(s) ds \\ &\geq \int_{-\infty}^{\infty} F_{\max\{N_{r,j_1},H\}}(s) f_{N_{r,j_2}}(s) ds = \mathbb{P}[N_{r,j_2} > \max\{N_{r,j_1}, H\}] = \mathbb{P}[J_r = j_2].\end{aligned}$$

Finaly, we are going to show $\mathbb{P}[J_r = j] \leq \frac{1}{j}$. This can be derived by $1 = \sum_{i=1}^{K} \mathbb{P}[J_r = i] \geq \sum_{i=1}^{j} \mathbb{P}[J_r = i] \geq \sum_{i \leq j} \mathbb{P}[J_r = j] = j \cdot \mathbb{P}[J_r = j]$. ■

Appendix C. Full Proof of Theorem 3

We first prove a lemma for the other two noise distributions $\mathcal{Q}_\varepsilon$

Lemma 16 *$\mathcal{Q}_\varepsilon$ is $\mathrm{Exp}(\frac{1}{\varepsilon})$ or $\mathrm{Gumbel}(\frac{2}{\varepsilon})$, there exists universal constants $c_1, c_2 > 0$ such that*

$$\mathbb{P}\left[J_r = j\right] \leq c_1 \cdot \exp(-2^r \Delta_j \min\{\Delta_j, \varepsilon\}/c_2).$$

Proof [Proof of Lemma 16.] The proof for $\mathcal{Q}_\varepsilon = \mathrm{Exp}(\frac{1}{\varepsilon})$ is almost the same as their proof for $\mathcal{Q}_\varepsilon = \mathrm{Lap}(\frac{2}{\varepsilon})$:

$$\begin{aligned}\mathbb{P}\left[J_r = j\right] &\leq \mathbb{P}\left[-G_{r,j} + Q_{r,j} > -G_{r,1} + Q_{r,1}\right] \\ &\leq \mathbb{P}\left[G_{r,j} - G_{r,1} \leq 2^{r-1}\frac{\Delta_j}{2}\right] + \mathbb{P}\left[Q_{r,j} - Q_{r,1} \geq 2^{r-1}\frac{\Delta_j}{2}\right].\end{aligned}$$

From the Hoeffding inequality,

$$\mathbb{P}\left[G_{r,j} - G_{r,1} \leq 2^{r-1}\frac{\Delta_j}{2}\right] = \mathbb{P}\left[G_{r,j} - G_{r,1} - 2^{r-1}\Delta_j \leq -2^{r-1}\frac{\Delta_j}{2}\right] \leq \exp\left(-2^{r-1}\frac{\Delta_j^2}{4}\right).$$

By the cdf of any eponential distribution,

$$\mathbb{P}\left[Q_{r,j} - Q_{r,1} \geq 2^{r-1}\frac{\Delta_j}{2}\right] \leq \mathbb{P}\left[Q_{r,j} \geq 2^{r-1}\frac{\Delta_j}{2}\right] \leq \exp\left(-\varepsilon 2^{r-1}\frac{\Delta_j}{2}\right).$$

Therefore, for $\mathcal{Q}_\varepsilon = \mathrm{Exp}(\frac{1}{\varepsilon})$, $\mathbb{P}\left[J_r = j\right] \leq 2 \cdot \exp(-2^r \Delta_j \min\{\Delta_j, \varepsilon\}/8)$.

As for $\mathcal{Q}_\varepsilon = \mathrm{Gumbel}(\frac{2}{\varepsilon})$, it is known that the report-noisy-max with gumbel noise is equivalent to Exponential Mechanism (McSherry and Talwar, 2007; Qiao et al., 2021), which is

$$\mathbb{P}\left[J_r = j | \forall i \in [K], G_{r,i}\right] = \frac{\exp\left(\varepsilon \cdot (-G_{r,j})\right)}{\sum_{i=1}^{K} \exp\left(\varepsilon \cdot (-G_{r,i})\right)}.$$

We bound $\mathbb{P}[J_r = j]$ as

$$\begin{aligned}\mathbb{P}[J_r = j] &= \mathbb{E}_{\forall i\in[K],G_{r,i}}[\mathbb{P}[J_r = j|\forall i \in [K], G_{r,i}]] \\ &= \mathbb{E}_{\forall i\in[K],G_{r,i}}\left[\frac{\exp(-\varepsilon\cdot(-G_{r,j}))}{\sum_{i=1}^{K}\exp(-\varepsilon\cdot(-G_{r,i}))}\right] \\ &\le \mathbb{E}_{\forall i\in[K],G_{r,i}}\left[\frac{\exp(\varepsilon\cdot(-G_{r,j}))}{\exp(\varepsilon\cdot(-G_{r,1}))+\exp(\varepsilon\cdot(-G_{r,j}))}\right] \\ &= \mathbb{E}\left[\frac{1}{\exp(\varepsilon\cdot(G_{r,j}-G_{r,1}))+1}\right].\end{aligned}$$

Denote the event $\mathcal{E}$ as $G_{r,j} - G_{r,1} \ge \frac{1}{2}2^{r-1}\Delta_j$, because $\frac{1}{\exp(\varepsilon\cdot(G_{r,j}-G_{r,1}))+1} \le 1$ is always true,

$$\mathbb{P}[J_r = j] \le \mathbb{E}\left[\frac{1}{\exp(\varepsilon\cdot(G_{r,j}-G_{r,1}))+1}|\mathcal{E}\right] + (1-\mathbb{P}[\mathcal{E}])$$

$$\le \frac{1}{\exp\left(\frac{1}{2}2^{r-1}\Delta_j\varepsilon\right))+1} + (1-\mathbb{P}[\mathcal{E}]) \le \exp\left(-2^{r-1}\Delta_j\varepsilon/2)\right) + (1-\mathbb{P}[\mathcal{E}]).$$

The bound for $1-\mathbb{P}[\mathcal{E}] = \mathbb{P}[(G_{r,j}-G_{r,1}) < 2^{r-1}\Delta_j/2]$ is

$$\mathbb{P}[G_{r,j}-G_{r,1}] \le \exp\left(-2^{r-1}\Delta_j^2/4\right)$$

where the inequality is held by the Hoeffding inequality. Therefore,

$$\mathbb{P}[J_r = j] \le \exp\left(-2^{r-1}\Delta_j\varepsilon/2)\right) + \exp\left(-2^{r-1}\Delta_j^2/4\right).$$

Our proof for the case $\mathcal{Q}_\varepsilon = \text{Gumbel}(\frac{2}{\varepsilon})$ is complete. ■

Now we show the full proof for Theorem 3.

Proof [Proof of Theorem 3.] If we can prove Equation 4 for any $T := 2^R - 1$ where R is any non-negative integer, Equation 4 would also hold for arbitrary T, because Algorithm 1 is independent of the T and the regret of Algorithm 1 is non-decreasing in T. Therefore, we can assume $T := 2^{R+1}-1$ for some non-negative integer R and can rewrite the pseudoregret (defined in Eqeation 1) according to the Algorithm 1:

$$\sum_{t=1}^{T}\mathbb{E}[\Delta_{I_t}] = \sum_{r=1}^{R}\sum_{t=2^{r-1}}^{2^r-1}\mathbb{E}[\Delta_{I_t}] = \sum_{r=1}^{R}2^{r-1}\mathbb{E}\left[\Delta_{J_{r-1}}\right] = \sum_{r=1}^{R}2^{r-1}\sum_{j=1}^{K}\Delta_j\mathbb{P}[J_{r-1}=j]$$

$$= \sum_{j=1}^{K}\Delta_j\sum_{r=1}^{R}2^{r-1}\mathbb{P}[J_{r-1}=j] = \underbrace{\sum_{j:\Delta_j\le\varepsilon}\Delta_j\sum_{r=1}^{R}2^{r-1}\mathbb{P}[J_{r-1}=j]}_{\text{Regret}_\uparrow} + \underbrace{\sum_{j:\Delta_j>\varepsilon}^{K}\Delta_j\sum_{r=1}^{R}2^{r-1}\mathbb{P}[J_{r-1}=j]}_{\text{Regret}_\downarrow}$$

According to the Lemma 9 in Hu et al. (2021) and Lemma 16, for all three noise distributions $\mathcal{Q}_\varepsilon$, there exists universal constants $c_1, c_2 > 0$ such that

$$\mathbb{P}[J_r = j] \le c_1\cdot\exp(-2^{r+1}\Delta_j\min\{\Delta_j,\varepsilon\}/c_2), \tag{7}$$

and similar to the proof for theorem 24 in Hu et al. (2021), for $\Delta_j, \varepsilon > 0$, we can calculate

$$\begin{aligned}
\sum_{r=r(j)+1}^{R} 2^{r-1}\mathbb{P}\left[J_{r-1}=j\right] &\leq \sum_{r=r(j)+1}^{R} 2^{r-1} c_1 \cdot \exp(-2^r \Delta_j \min\{\Delta_j, \varepsilon\}/c_2) \\
&< c_1 \sum_{r=r(j)+1}^{R} \sum_{t=2^{r-1}+1}^{2^r} \exp(-t\Delta_j \min\{\Delta_j, \varepsilon\}/c_2) \\
&< c_1 \sum_{t=2^{r(j)}+1}^{\infty} \cdot \exp(-t\Delta_j \min\{\Delta_j, \varepsilon\}/c_2) \\
&< c_1 \int_{2^{r(j)}}^{\infty} \cdot \exp(-t\Delta_j \min\{\Delta_j, \varepsilon\}/c_2) dt \\
&= \frac{c_1 c_2}{\Delta_j \min\{\Delta_j, \varepsilon\}} \cdot \exp(-2^{r(j)}\Delta_j \min\{\Delta_j, \varepsilon\}/c_2) \qquad (8)
\end{aligned}$$

We first bound $\text{Regret}_{\downarrow}$. Let $r(j) = \left\lceil \log_2 \left(\frac{c_2 (\ln K)}{\Delta_j \varepsilon} \right) \right\rceil$. Then for any j s.t. $\Delta_j > \varepsilon$,

$$\begin{aligned}
\sum_{r=1}^{R} 2^{r-1}\mathbb{P}[J_{r-1}=j] &= \sum_{r=1}^{r(j)} 2^{r-1}\mathbb{P}[J_{r-1}=j] + \sum_{r=r(j)+1}^{R} 2^{r-1}\mathbb{P}[J_{r-1}=j] \\
&< \left(\sum_{r=1}^{r(j)} 2^{r-1} \frac{1}{j} \right) + \frac{c_1 c_2}{\Delta_j \varepsilon} \cdot \exp(-2^{r(j)}\Delta_j \varepsilon / c_2) \\
&< 2^{r(j)} \frac{1}{j} + \frac{c_1 c_2}{\Delta_j \varepsilon} \cdot \exp(-2^{r(j)}\Delta_j \varepsilon / c_2) \\
&\leq \frac{2c_2}{\Delta_j \varepsilon} \cdot \frac{\ln K}{j} + \frac{c_1 c_2}{\Delta_j \varepsilon} \cdot \frac{1}{K},
\end{aligned}$$

where the first inequality holds by Lemma 6 (*since it is assumed that $B = 1$ for the Algorithm 1 in the theorem statement*) and Equation 8, the second inequality holds by $\sum_{r=1}^{r(j)} 2^{r-1} = 2^{r(j)} - 1 < 2^{r(j)}$, and the third inequality holds by taking the value of $r(j)$. Therefore,

$$\text{Regret}_{\downarrow} = \sum_{j:\Delta_j > \varepsilon}^{K} \Delta_j \sum_{r=1}^{R} 2^{r-1}\mathbb{P}[J_{r-1}=j] < \sum_{j:\Delta_j > \varepsilon}^{K} \Delta_j \left(\frac{2c_2}{\Delta_j \varepsilon} \cdot \frac{\ln K}{j} + \frac{c_1 c_2}{\Delta_j \varepsilon} \cdot \frac{1}{K} \right)$$

$$\leq \frac{2c_2}{\varepsilon} \cdot \sum_{j:\Delta_j > \varepsilon}^{K} \frac{\ln K}{j} + \frac{c_1 c_2}{\varepsilon} \cdot \sum_{j:\Delta_j > \varepsilon}^{K} \frac{1}{K} = O\left(\frac{(\ln K)^2}{\varepsilon} \right).$$

The remaining is to bound $\text{Regret}_{\uparrow}$, which is the same as a part of proof for Theorem 9 in Hu et al. (2021). For completeness, we illustrate the details here. The idea is to group j. Define $\Delta_{(l)} := 2^{l-1}\Delta_{\min}$ and denote $H_l := \{j : \Delta_{(l)} \leq \Delta_j < \Delta_{(l+1)}\} \cap \{j : \Delta_j < \varepsilon, j \geq 2\}$. Then for

any $j \in H_l$, we pick $r(j) := \tau_l = \left\lceil \frac{c_2 \ln(|H_l|)}{\Delta_{(l)}^2} \right\rceil$.

$$\begin{aligned}
\text{Regret}_{\uparrow} &= \sum_{j:\Delta_j \le \varepsilon} \Delta_j \cdot \left(\sum_{r=1}^{r(j)} 2^{r-1} \mathbb{P}[J_{r-1} = j] + \sum_{r=r(j)+1}^{R} 2^{r-1} \mathbb{P}[J_{r-1} = j] \right) \\
&= \sum_{l=1}^{\infty} \sum_{j \in H_l} \Delta_j \cdot \left(\sum_{r=1}^{\tau_l} 2^{r-1} \mathbb{P}[J_{r-1} = j] + \sum_{r=\tau_l+1}^{R} 2^{r-1} \mathbb{P}[J_{r-1} = j] \right) \\
&= \sum_{l=1}^{\infty} \left(\sum_{r=1}^{\tau_l} 2^{r-1} \sum_{j \in H_l} \Delta_j \cdot \mathbb{P}[J_{r-1} = j] \right) + \sum_{l=1}^{\infty} \left(\sum_{j \in H_l} \Delta_j \cdot \sum_{r=\tau_l+1}^{R} 2^{r-1} \mathbb{P}[J_{r-1} = j] \right) \\
&\le \sum_{l=1}^{\infty} \left(\sum_{r=1}^{\tau_l} 2^{r-1} \right) \cdot 2\Delta_{(l)} + \sum_{l=1}^{\infty} \left(\sum_{j \in H_l} \Delta_j \cdot \frac{c_1 c_2}{\Delta_j^2} \cdot \exp(-2^{r(j)} \Delta_j^2 / c_2) \right) \\
&< \sum_{l=1}^{\infty} 2^{\tau_l + 2} \Delta_{(l)} + \sum_{l=1}^{\infty} \left(|H_l| \cdot \frac{c_1 c_2}{\Delta_{(l)}} \cdot \exp(-2^{r(j)} \Delta_{(l)}^2 / c_2) \right) \\
&\le \sum_{l=1}^{\infty} \frac{8 c_2 \ln(|H_l|)}{\Delta_{(l)}} + \sum_{l=1}^{\infty} \frac{c_1 c_2}{\Delta_{(l)}} \\
&\le \frac{8 c_2 \ln K + c_1 c_2}{\Delta_{\min}} \sum_{l=1}^{\infty} \frac{1}{2^{l-1}} \\
&= \frac{8 c_2 \ln K + c_1 c_2}{\Delta_{\min}},
\end{aligned}$$

The first inequality is because Equation 6 and the fact that for $j \in H_l$, $\sum_{j \in H_l} \Delta_j \cdot \mathbb{P}[J_{r-1} = j] \le 2\Delta_{(l)} \sum_{j \in H_l} \mathbb{P}[J_{r-1} = j] \le 2\Delta_{(l)}$; the second inequality holds by $\sum_{r=1}^{\tau_l} 2^{r-1} < 2^{\tau_l}$ and the fact that for $j \in H_l$, $\Delta_j \ge \Delta_{(l)}$; the third inequality holds by taking the value of τ_l; the fourth inequality holds by the definition of $\Delta_{(l)}$ and the fact $|H_l| \le K$.

Putting the analysis for $\text{Regret}_{\uparrow}$ and $\text{Regret}_{\downarrow}$ together, we have proved that the pseudoregret is bounded by $O\left(\frac{\log(K)}{\Delta_{\min}} + \frac{(\log K)^2}{\varepsilon}\right)$. ■

Appendix D. Proof of Corollary 5

The proof follows the exact same steps as the proof for Corollary 11 in Hu et al. (2021), which is also well-known as early as Audibert and Bubeck (2009). For completeness, we repeat the exact steps here.

Proof [Proof of Corollary 5] Let $\Delta^* := \sqrt{\log K / T}$ be the critical gap. Then, for all actions j that $\Delta_j < \Delta^*$, the can contribute the regret at most $T \cdot \Delta^* = \sqrt{T \log K}$. To bound the contributions for actions j that $\Delta_j \ge \Delta^*$, we can simply adapt the proof of our Theorem 3 and Corollary **??** for only these actions, and the effective $\Delta_{\min}$ becomes Δ^*.Therefore, the bound for the overall regret becomes

$$O\left(\sqrt{T \log K} + \frac{\log K}{\Delta^*} + \frac{(\log K)^2}{\varepsilon}\right) = O\left(\sqrt{T \log K} + \frac{(\log K)^2}{\varepsilon}\right)$$

■

Appendix E. Proof of Theorem 7

The lower bound for the original setting, that is $\Omega\left(\frac{\log(K)}{\Delta_{\min}}+\frac{\log(K)}{\varepsilon}\right)$, is an application of Corollary 4 in Acharya et al. (2021). However, Corollary 4 in Acharya et al. (2021) requires a bounded KL divergence, while at our deterministic setting where each $\mathcal{P}_i$ has probability 0 on all values except μ_i, the KL divergence between $\mathcal{P}=\mathcal{P}_1\times\cdots\times\mathcal{P}_k$ and $\mathcal{P}'$ is infinity when $\mathcal{P}\neq\mathcal{P}'$. Therefore, we show an easy and standard construction for our setting.

Proof [Proof of Theorem 7.] For any $l\in[K]$, define $\mathcal{P}^{(l)}:=\mathcal{P}_1^{(l)}\times\cdots\times\mathcal{P}_K^{(l)}$, where $\mathbb{P}_{\ell_i\sim\mathcal{P}_i^{(l)}}[\ell_i=\mu_i^{(l)}]=1$, $\mu_l^{(l)}=0$, $\mu_{l-1}^{(l)}=\mu_{l+1}^{(l)}=\Delta_{\min}$ and $\mu_i^{(l)}=1$ for all $i\notin[l-1,l+1]$. Here the subscript index is cyclic in K: $\mu_0^{(l)}=\mu_K^{(l)}$ and $\mu_{K+1}^{(l)}=\mu_1^{(l)}$. Suppose $\mathcal{A}$ is any online algorithm that is ε-differentially private. When K actions have the loss from $\mathcal{P}^{(l)}$, denote $I_t^{(l)}$ is the action from $\mathcal{A}$ and further for any length of the online procedure T, let $R^{(l)}(T)$ is the pseudoregret. Therefore

$$R^{(l)}(T)\geq\sum_{t=1}^{T}\mathbb{P}[I_t^{(l)}\notin[l-1,l+1]]$$

One the other hand, because $\mathcal{A}$ is differentially private, for any $l,l'\in[K]$, any action i, and any $t\geq T$,

$$\mathbb{P}[I_t^{(l)}=i]\leq e^{t\cdot\varepsilon}\cdot\mathbb{P}[I_t^{(l')}=i].$$

Therefore,

$$\mathbb{P}[I_t^{(l)}\notin[l-1,l+1]]=1-\sum_{\hat{l}=l-1}^{l+1}\mathbb{P}[I_t^{(l)}=\hat{l}]\geq 1-\sum_{\hat{l}=l-1}^{l+1}\frac{e^{t\cdot\varepsilon}}{K-5}\sum_{l'\notin[l-2,l+2]}\mathbb{P}[I_t^{(l')}=\hat{l}].$$

We take a sum of all $l\in[K]$:

$$\begin{aligned}
\sum_{l=1}^{K}\mathbb{P}[I_t^{(l)}\notin[l-1,l+1]]&\geq K-\frac{e^{t\cdot\varepsilon}}{K-5}\sum_{l=1}^{K}\sum_{\hat{l}=l-1}^{l+1}\sum_{l'\notin[l-2,l+2]}\mathbb{P}[I_t^{(l')}=\hat{l}]\\
&=K-\frac{e^{t\cdot\varepsilon}}{K-5}\sum_{l'=1}^{K}\sum_{l\notin[l'-2,l'+2]}\sum_{\hat{l}=l-1}^{l+1}\mathbb{P}[I_t^{(l')}=\hat{l}]\\
&\geq K-\frac{e^{t\cdot\varepsilon}}{K-5}\sum_{l'=1}^{K}3\cdot\mathbb{P}[I_t^{(l')}\notin[l'-1,l'+1]]\\
&=K-\frac{3e^{t\cdot\varepsilon}}{K-5}\sum_{l=1}^{K}\mathbb{P}[I_t^{(l)}\notin[l-1,l+1]]
\end{aligned}$$

where the first equality holds by swiping the order of summations of l and l'. This gives us $\sum_{l=1}^{K}\mathbb{P}[I_t^{(l)}\notin[l-1,l+1]]\geq\frac{K(K-5)}{3e^{t\cdot\varepsilon}+K-5}$. Thus,

$$\frac{1}{K}\sum_{l=1}^{K}R^{(l)}(T)\geq\sum_{t=1}^{T}\frac{K-5}{3e^{t\cdot\varepsilon}+K-5}\geq\sum_{t=1}^{T}\int_t^{t+1}\frac{K-5}{3e^{\tau\cdot\varepsilon}+K-5}d\tau=\int_1^{T+1}\frac{K-5}{3e^{t\cdot\varepsilon}+K-5}dt$$

where the second inequality holds because $\frac{K-5}{3e^{t\cdot\varepsilon}+K-5}$ is monotonically decreasing. The antiderivatives for $g(t)=\frac{K-5}{3e^{t\cdot\varepsilon}+K-5}$ are $\frac{\ln\left(\frac{e^{t\varepsilon}}{3e^{t\varepsilon}+K-5}\right)}{\varepsilon}+C$ for any constant C, which implies:

$$\frac{1}{K}\sum_{l=1}^{K}R^{(l)}(T)\geq\int_{1}^{T+1}\frac{K-5}{3e^{t\cdot\varepsilon}+K-5}dt=\frac{\ln\left(\frac{e^{(T+1)\varepsilon}}{3e^{(T+1)\varepsilon}+K-5}\cdot\frac{3e^{\varepsilon}+K-5}{e^{\varepsilon}}\right)}{\varepsilon}.$$

From here, it implies that there exists l_T^* s.t.

$$R^{(l_T^*)}(T)\geq\frac{\ln\left(\frac{e^{(T+1)\varepsilon}}{3e^{(T+1)\varepsilon}+K-5}\cdot\frac{3e^{\varepsilon}+K-5}{e^{\varepsilon}}\right)}{\varepsilon}.$$

When $T\to\infty$,

$$\lim_{T\to\infty}R^{(l_T^*)}(T)\geq\frac{\ln\left(\frac{e^{\varepsilon}+(K-5)/3}{e^{\varepsilon}}\right)}{\varepsilon}=\frac{\ln\left(e^{\varepsilon}+(K-5)/3\right)}{\varepsilon}-1\geq\frac{\ln((K-5)/3)}{\varepsilon}-1=\Omega\left(\frac{\ln K}{\varepsilon}\right).$$

■

Appendix F. Proof of Corollary 10

The proof for the deterministic setting is a straightforward extension from the proof for Theorem 3 (the result at the original setting).

Proof [Proof sketch of Corollary 10.] With the additional assumption at the deterministic setting that $\mathbb{P}_{\ell_j\sim\mathcal{P}_j}[\ell_j=\mu_j]=1$, $\mathbb{P}[J_r=j]$ can be bounded in the form when $\mathcal{Q}_\varepsilon$ is laplace distribution, exponential distribution, or gumbel distribution:

$$\mathbb{P}\left[J_r=j\right]\leq c_1\cdot\exp(-2^r\Delta_j\varepsilon/c_2)\tag{9}$$

for some universal constants $c_1,c_2>0$, a slight improvement from the bound $\mathbb{P}\left[J_r=j\right]\leq c_1\cdot\exp(-2^r\Delta_j\min\{\Delta_j,\varepsilon\}/c_2)$ (Equation 7) at the original setting. Then by extending the similar derivation in the proof of Theorem 3 (Section C), we can prove that the pseudo regret is bounded by $O\left(\frac{(\log K)^2}{\varepsilon}\right)$ ■

Appendix G. Proof of the Properties for the Soft-Max Like Function

Proof [Proof of Lemma 12.] This can be proved by calculating the derivatives $f'(x)$:

$$\begin{aligned}
f'(x)&=\frac{\left(\sum_{i=1}^{K}2^x a_i e^{-2^x a_i}\right)'}{\sum_{i=1}^{K}e^{-2^x a_i}}-\frac{\left(\sum_{i=1}^{K}2^x a_i e^{-2^x a_i}\right)\cdot\left(\sum_{i=1}^{K}e^{-2^x a_i}\right)'}{\left(\sum_{i=1}^{K}e^{-2^x a_i}\right)^2}\\
&=\left(\log 2\cdot\frac{\sum_{i=1}^{K}2^x a_i\cdot e^{-2^x a_i}}{\sum_{i=1}^{K}e^{-2^x a_i}}-\log 2\cdot\frac{\sum_{i=1}^{K}(2^x a_i)^2\cdot e^{-2^x a_i}}{\sum_{i=1}^{K}e^{-2^x a_i}}\right)+\log 2\cdot\frac{(\sum_{i=1}^{K}2^x a_i e^{-2^x a_i})^2}{(\sum_{i=1}^{K}e^{-2^x a_i})^2}\\
&=(\log 2)f(x)-(\log 2)\frac{(\sum_{i=1}^{K}(2^x a_i)^2 e^{-2^x a_i})(\sum_{i=1}^{K}e^{-2^x a_i})-(\sum_{i=1}^{K}2^x a_i e^{-2^x a_i})^2}{(\sum_{i=1}^{K}e^{-2^x a_i})^2}\\
&\leq(\log 2)f(x),
\end{aligned}$$

where the last inequality is held by Cauchy Schwarz Inequality. ■

Proof [Proof of Lemma 13.] Let $f(x) = \frac{\sum_{i=1}^K 2^x a_i e^{-2^x a_i}}{\sum_{i=1}^K e^{-2^x a_i}}$. Then,

$$\sum_{r=1}^{\infty} \frac{\sum_{i=1}^K 2^r a_i \exp\left(-2^r a_i\right)}{\sum_{i=1}^K \exp\left(-2^r a_i\right)} = \sum_{r=1}^{\infty} f(r) = \sum_{r=1}^{\infty} \left[\left(f(r) - \int_{r-1}^{r} f(x)dx \right) + \int_{r-1}^{r} f(x)dx \right].$$

From the Lagrange's mean value theorem, $\int_{r-1}^{r} f(x)dx = f(x_r)$ for some $x_r \in [r-1, r]$. Therefore

$$f(r) - \int_{r-1}^{r} f(x)dx = f(r) - f(x_r) = \int_{x_r}^{r} f'(x)dx \leq \int_{x_r}^{r} \log 2 f(x)dx \leq \log 2 \int_{r-1}^{r} f(x)dx, \tag{10}$$

where the first inequality holds by $f'(x) \leq \log 2 \cdot f(x)$ that we just proved and the second inequality is true because $f(x) \geq 0$ for all x. With the Equation 10, we now have

$$\sum_{r=1}^{\infty} \frac{\sum_{i=1}^K 2^r a_i \exp\left(-2^r a_i\right)}{\sum_{i=1}^K \exp\left(-2^r a_i\right)} \leq (\log 2 + 1) \sum_{r=1}^{\infty} \int_{r-1}^{r} f(x)dx = (\log 2 + 1) \int_{0}^{\infty} f(x)dx. \tag{11}$$

The last thing is to bound $\int_0^\infty f(x)dx$. Notice that the antiderivatives for $f(x) = \frac{\sum_{i=1}^K 2^x a_i e^{-2^x a_i}}{\sum_{i=1}^K e^{-2^x a_i}}$ is $F(x) = -\frac{1}{\log 2} \log\left(\sum_{i=1}^K e^{-2^x a_i}\right) + C$ for any constant C. Moreover, because $0{=}a_1 < a_2 \leq , \cdots, \leq a_K$,

$$F(0) = -\frac{1}{\log 2} \log\left(\sum_{i=1}^{K} e^{-a_i}\right) + C \geq -\frac{1}{\log 2} \log(K) + C; \lim_{x \infty} F(x) = -\frac{1}{\log 2} \log(1) + C = C.$$

Therefore $\int_0^\infty f(x)dx = \lim_{x \to +\infty} F(x) - F(0) = \frac{2}{\log 2} \log(K)$. Taking this equality to Equation 11, our proof is complete. ■

Appendix H. Proof of Theorem 11

Proof [Proof of Theorem 11.] We first prove for the gumbel distribution $\text{Gumbel}(\frac{2}{\varepsilon})$. It is known that the report-noisy-max with gumbel noise is equivalent to Exponential Mechanism (McSherry and Talwar, 2007; Qiao et al., 2021), which is

$$\mathbb{P}\left[J_r = j | \forall i \in [K], G_{r,i}\right] = \frac{\exp\left(\varepsilon \cdot (-G_{r,j})\right)}{\sum_{i=1}^K \exp\left(\varepsilon \cdot (-G_{r,i})\right)}.$$

Because we are considering the deterministic setting, $G_{r,i} = 2^{r-1}\mu_i$ with probability 1. Therefore,

$$\mathbb{P}\left[J_r = j\right] = \frac{\exp\left(-2^{r-1}\mu_i \varepsilon\right)}{\sum_{i=1}^K \exp\left(-2^{r-1}\mu_j \varepsilon\right)} = \frac{\exp\left(-2^{r-1}\Delta_i \varepsilon\right)}{\sum_{i=1}^K \exp\left(-2^{r-1}\Delta_j \varepsilon\right)}.$$

Then let $a_i = \Delta_i \varepsilon$ in Lemma 13 and we can show the upper bound for pseudoregret:

$$\sum_{t=1}^{T} \mathbb{E}\left[\Delta_{I_t}\right] \leq 3 + \sum_{r=3}^{\infty} 2^{r-1} \sum_{j=1}^{K} \Delta_j \mathbb{P}[J_{r-1} = j] \leq 3 + 2 \cdot \sum_{r=1}^{\infty} \frac{\sum_{i=1}^{K} 2^r \Delta_i \exp\left(-2^r \Delta_i \varepsilon\right)}{\sum_{i=1}^{K} \exp\left(-2^r \Delta_j \varepsilon\right)}$$
$$= 3 + \frac{2}{\varepsilon} \cdot \sum_{r=1}^{\infty} \frac{\sum_{i=1}^{K} 2^r \Delta_i \varepsilon \exp\left(-2^r \Delta_i \varepsilon\right)}{\sum_{i=1}^{K} \exp\left(-2^r \Delta_j \varepsilon\right)} \leq O\left(\frac{\log K}{\varepsilon}\right)$$

We have proved the upper bound for $\mathcal{Q}_\varepsilon = \text{Gumbel}(\frac{2}{\varepsilon})$ and now we can prove the upper bound for the exponential distribution $\mathcal{Q}_\varepsilon = \text{Exp}(\frac{1}{\varepsilon})$. To distinguish, I_t is still the action from $\mathcal{Q}_\varepsilon = \text{Gumbel}(\frac{2}{\varepsilon})$, and we denote I_t^{exp} as the action from $\mathcal{Q}_\varepsilon = \text{Exp}(\frac{1}{\varepsilon})$ and J_r^{exp} as the output from report-noisy-max with the exponential noise. McKenna and Sheldon (2020) has proved (in their Theorem 2) that the report-noisy-max with exponential noise is consistently better than the exponential mechanism, which is equivalent to the report-noisy-max with gumbel noise (Gumbel, 1954; Qiao et al., 2021).

$$\sum_{j=1}^{K}(-2^{r-1}\mu_j)\cdot\mathbb{P}[J_r^{\text{exp}} = j] \geq \sum_{j=1}^{K}(-2^{r-1}\mu_j)\cdot\mathbb{P}[J_r = j] \Rightarrow \sum_{j=1}^{K}\mu_j\cdot\mathbb{P}[J_r^{\text{exp}} = j] \leq \sum_{j=1}^{K}\mu_j\cdot\mathbb{P}[J_r = j].$$

Subtract μ_1 from both sides, we have $\sum_{j=1}^{K} \Delta_j \cdot \mathbb{P}[J_r^{\text{exp}} = j] \leq \sum_{j=1}^{K} \Delta_j \cdot \mathbb{P}[J_r = j]$, and then the pseudoregret when $\mathcal{Q}_\varepsilon = \text{Exp}(\frac{1}{\varepsilon})$ can be bounded by

$$\sum_{t=1}^{T} \mathbb{E}\left[\Delta_{I_t^{\text{exp}}}\right] \leq 3 + \sum_{r=1}^{R-2} 2^{r+1} \sum_{j=1}^{K} \Delta_j \mathbb{P}[J_{r+1}^{\text{exp}} = j] \leq 3 + \sum_{r=1}^{R-2} 2^{r+1} \sum_{j=1}^{K} \Delta_j \mathbb{P}[J_{r+1} = j],$$

which now is the case of $\mathcal{Q}_\varepsilon = \text{Gumbel}(\frac{2}{\varepsilon})$ and bounded by $O\left(\frac{\log K}{\varepsilon}\right)$.

We have proved the pseudoregret can be bounded by $O\left(\frac{\log K}{\varepsilon}\right)$ when specifying $B = 0$ and $\mathcal{Q}_\varepsilon = \text{Gumbel}(\frac{2}{\varepsilon})$ or $\mathcal{Q}_\varepsilon = \text{Exp}(\frac{1}{\varepsilon})$. On the other hand, the lower bound is proved in Theorem 7. This means that our algorithm with the analyzed upper bound $O\left(\frac{\log K}{\varepsilon}\right)$ is optimal. ■