Open Problem: Optimal Rates for Stochastic Decision-Theoretic Online Learning Under Differentially Privacy

Bingshan Hu BINGSHA1@CS.UBC.CA
University of British Columbia

Nishant A. Mehta NMEHTA@UVIC.CA
University of Victoria

Editors: Shipra Agrawal and Aaron Roth

Abstract

For the stochastic variant of decision-theoretic online learning with K actions, T rounds, and minimum gap $\Delta_{\min}$, the optimal, gap-dependent rate of the pseudo-regret is known to be $O\left(\frac{\log K}{\Delta_{\min}}\right)$. We ask to settle the optimal gap-dependent rate for the problem under ε-differential privacy.

1. Introduction

The stochastic variant of decision-theoretic online learning (Freund and Schapire, 1997) may be the simplest non-worst-case online learning setting. Given K actions, in each round $t = 1, 2, \ldots, T$:

1. a possibly randomized learning algorithm selects an action $I_t \in [K] = \{1, 2, \ldots, K\}$;
2. each action j is assigned a loss $\ell_{j,t}$ drawn independently from an unknown distribution P_j;
3. the learning algorithm suffers loss $\ell_{I_t,t}$ and observes the losses of all the actions.

Due to the stochasticity, a natural goal is to obtain low *pseudo-regret*, the difference of the learning algorithm's expected cumulative loss and the minimum expected cumulative loss among the actions:

$$\mathsf{E}\left[\sum_{t=1}^{T} \ell_{I_t,t}\right] - \min_{j\in[K]} \mathsf{E}\left[\sum_{t=1}^{T} \ell_{j,t}\right].$$

Assume for simplicity that the losses lie in $[0, 1]$ and the optimal action is unique with mean loss μ^*. For convenience, let the actions (denoted by j) be indexed in order of non-decreasing expected loss (μ_j), so that $\mu^* = \mu_1 < \mu_2 \leq \ldots \leq \mu_K$. In this setting and its bandit cousin, pseudo-regret bounds often are parameterized via gaps of the form $\Delta_j := \mu_j - \mu^*$, giving "gap-dependent" bounds. For this problem, Follow the Leader (FTL) (Kotłowski, 2018) achieves $O\left(\frac{\log K}{\Delta_{\min}}\right)$ gap-dependent pseudo-regret (Kotłowski, 2019), which is optimal (Mourtada and Gaïffas, 2019).

The learning theory community has given significant attention to the interplay between differential privacy (Dwork, 2006) and learning. Yet, fundamental questions remain open in differentially private learning (defined shortly). We wish to highlight the following question:

> **Under ε-differential privacy, what is the optimal gap-dependent rate for the pseudo-regret for the stochastic variant of decision-theoretic online learning?**

We now present a definition of differential privacy for online learning (Sajed and Sheffet, 2019).

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

Asi et al. (2023) showed that the worst-case (over all settings of the gaps) pseudo-regret is

$$O\left(\sqrt{T\log d} + \frac{\log(K)\log(T)}{\varepsilon}\right),$$

Using a strategy called RNM-FTNL, Hu et al. (2021) showed the same worst-case bound[1] as well as a gap-dependent pseudo-regret upper bound that in some cases can be as large as

$$O\left(\frac{\log K}{\Delta_{\min}} + \frac{\log(K)\log(T)}{\varepsilon}\right). \tag{1}$$

RNM-FTNL runs in epochs that double in length. In each successive epoch, the algorithm uses Follow the Perturbed Leader (Kalai and Vempala, 2005) with independent Laplace noise (at scale $1/\varepsilon$) but with two modifications: *(i)* only data from the previous epoch is used; *(ii)* the same "noisy leader" is played for all rounds of the current epoch. In Section 2, we will present and discuss a refinement of the above bound. Hu et al. (2021) also showed the lower bound

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

Until 2023, we believed this lower bound was the optimal gap-dependent rate. At this point, we lack sufficient confidence to conjecture the optimality nor suboptimality of the lower bound. The next section presents the full version of (1), followed by our intuition for why the lower bound is the correct rate as well as our intuition to the contrary for why the lower bound is *not* the correct rate.

2. Known Results in Detail, and Intuition

For any $m \in \{1, 2, \ldots\}$, define the m^{th} action bracket as

$$\mathcal{J}_m := \left\{j \in K \colon 2^{m-1} \cdot \Delta_{\min} \leq \Delta_j < 2^m \cdot \Delta_{\min}\right\} \tag{3}$$

and let the m^{th} within-bracket-minimum-gap be defined as $\Delta_{(m)} := 2^{m-1}\Delta_{\min}$. It follows that for any $j \in \mathcal{J}_m$, we have $\Delta_{(m)} \leq \Delta_j < 2\Delta_{(m)}$.

Hu et al. (2021) showed that RNM-FTNL's pseudo-regret is at most

$$O\left(\frac{\log K}{\Delta_{\min}} + \sum_{m\geq 1\colon |\mathcal{J}_m|>0, \Delta_{(m)}>\varepsilon} \frac{1+\log|\mathcal{J}_m|}{\varepsilon}\right) \tag{4}$$

Surprisingly (to us), the bound (4) is *worse* when some suboptimal actions have gaps that exceed the minimum gap. To see this, and to build intuition later, we first introduce a special problem instance that we call the *doubling gaps* construction. In this construction:

- We set $K = 1 + n$, where n is a perfect square (so $n = x^2$ for a positive integer x).

1. After Asi et al. (2023), in June 2023 Hu et al. (2021) corrected their analysis to get the worst-case rate shown here.

- The gaps for actions 2 through K follow the structure

$$\underbrace{\Delta_{\min} \ \dots \ \Delta_{\min}}_{\sqrt{n} \text{ times}} \ \underbrace{2\Delta_{\min} \ \dots \ 2\Delta_{\min}}_{\sqrt{n} \text{ times}} \ \underbrace{2^2\Delta_{\min} \ \dots \ 2^2\Delta_{\min}}_{\sqrt{n} \text{ times}} \ \cdots \ \underbrace{2^{n-1}\Delta_{\min} \ \dots \ 2^{n-1}\Delta_{\min}}_{\sqrt{n} \text{ times}};$$

  so, these actions fall into $\Theta(\sqrt{K})$ groups, and in each successive group the gap doubles.

Consider the doubling gaps construction. If $\Delta_{\min} = O(1/T)$, then the actions in the smallest-gap group can contribute at most $O(1)$ to the pseudo-regret; hence, when reasoning about pseudo-regret we may assume (for this construction) that $\sqrt{n} = O(\log T)$. With this construction and constraint on n, the bound in (4) becomes

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

Yet, if we were to instead employ a construction in which all suboptimal actions have the same gap $\Delta_{\min}$ — that is, the construction that is usually used to show worst-case lower bounds in both non-private and private hypothesis testing, in non-private stochastic DTOL, and in non-private stochastic bandits — then (4) becomes

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

matching the lower bound (2).

According to information-theoretic intuition, one would expect that the problem only becomes easier as some actions' suboptimalities (expected loss minus μ^*) increase. That is, in moving from the typical worst-case construction to the doubling gaps construction, intuitively the pseudo-regret should not increase. Sure, the price paid for playing some actions becomes larger in the doubling gaps construction, but an action's increased suboptimality should only make it easier to avoid the action.

Towards a better algorithm or analysis. In what follows, we adopt a fast and loose analysis style to build mathematical intuition. For the below, we consider the doubling gaps construction. We say a suboptimal action is sufficiently sampled if the algorithm has collected enough observations of the action to statistically distinguish the action from the optimal action.

Fix one of the n groups of actions containing some action j. In round t, the amount by which the expectation of the minimum cumulative loss (where the minimum is taken over the actions in the group) falls below the expected cumulative loss of a fixed action in the group is of order $\sqrt{t \log \sqrt{n}} \asymp \sqrt{t \log K}$. On the other hand, the expected cumulative loss of a fixed action in the group exceeds that of the optimal action by $t\Delta_j$. Therefore, statistically distinguishing the "best-looking" action in the group from the optimal action requires order $\frac{\log K}{\Delta_j^2}$ samples. Now, suppose that FTL is as unlucky as possible. Then, ignoring constants, in the first $\frac{\log K}{(2^{n-1}\Delta_{\min})^2}$ rounds, FTL plays an action in the last group (group n, the one with the largest gaps), picking up pseudo-regret of order $\frac{\log K}{(2^{n-1}\Delta_{\min})^2} \cdot 2^{n-1}\Delta_{\min} = \frac{\log K}{2^{n-1}\Delta_{\min}}$. Similarly, when unlucky, for three fourths[2] of the first $\frac{\log K}{(2^{n-2}\Delta_{\min})^2}$ rounds, FTL plays an action in group $n-1$, picking up order $\frac{\log K}{2^{n-2}\Delta_{\min}}$ pseudo-regret.

2. Since we already accounted for the first quarter.

Repeating this argument for all groups, FTL in total picks up $O\left(\sum_{m=0}^{n-1} \frac{\log K}{2^m \Delta_{\min}}\right) = O\left(\frac{\log K}{\Delta_{\min}}\right)$ pseudo-regret.

Next, consider the ε-differentially private setting for $\varepsilon \ll \Delta_{\min}$. Again fix one of the n groups of actions containing an action j. If using RNM-FTNL (which uses scale-$\frac{1}{\varepsilon}$ Laplace noise), after t rounds the expected deviations of the "best-looking" action in the group (relative to $t\Delta_j$) are of order $\max\left\{\frac{\log K}{\varepsilon}, \sqrt{t \log K}\right\}$. Therefore, until t is of order $\frac{\log K}{\varepsilon \cdot \Delta_j}$, RNM-FTNL cannot statistically distinguish this within-group "best-looking" action from the optimal action. Following an analysis analogous to the non-private case, RNM-FTNL picks up pseudo-regret of order $O\left(\sum_{m=0}^{n-1} \frac{\log K}{\varepsilon \cdot 2^m \Delta_{\min}} \cdot 2^m \Delta_{\min}\right) = O\left(\log(K) \sum_{m=0}^{n-1} \frac{1}{\varepsilon}\right)$ pseudo-regret, which is $O\left(\frac{\log(K)\log(T)}{\varepsilon}\right)$.

Better pseudo-regret under the uniform hypothesis. If it is possible to show that RNM-FTNL satisfies what we dub the *uniform hypothesis*, meaning that the algorithm (approximately) balances its mistakes across actions that have not yet been sufficiently sampled, then it seems that the algorithm would enjoy $O\left(\frac{\log^2 K}{\varepsilon}\right) = O\left(\frac{\log(K)\log\log(T)}{\varepsilon}\right)$ pseudo-regret. To see why, suppose in each successive epoch (an epoch ends when one more group is sufficiently sampled), plays of suboptimal actions are balanced (in expectation) across insufficiently sampled actions. Then in the doubling gaps construction, our sketch analysis implies that RNM-FTNL gets pseudo-regret of order

$$\sum_{m=0}^{n-1} \frac{\log K}{\varepsilon \cdot 2^m \Delta_{\min}} \sum_{k=0}^{m} \frac{1}{m+1} 2^k \Delta_{\min} \leq \frac{2 \log K}{\varepsilon} \sum_{m=0}^{n-1} \frac{1}{2^m \Delta_{\min}} \frac{1}{m+1} 2^m \Delta_{\min} = O\left(\frac{\log^2 K}{\varepsilon}\right).$$

We have been unable to show that RNM-FTNL satisfies the uniform hypothesis. We also tried to design a new algorithm that maintains lower and upper confidence bounds to try to enforce balancing pulls across all actions whose optimality statistically cannot be ruled out, but the use of the confidence bounds unfortunately led to unacceptable extra log factors (including $\log T$ factors).

3. Prizes

For the open problems we only consider the case of pure DP, i.e., ε-DP (not approximate DP). Also, all loss distributions are assumed to have support contained in $[0, 1]$.

Here are the prizes:

- Grand prize: \$300 USD for getting the first problem-dependent upper and lower bounds of matching order. Improving only the upper bound or only the lower bound is fine, as long as matching bounds are obtained. If improving the upper bound, it is fine to tighten the analysis of an existing algorithm or design a new algorithm. If the upper bound is not constructive (no algorithm is presented which achieves the bound), the prize is only \$200 USD; the first subsequent constructive result gets the remaining \$100 USD (not \$300 USD).

- Lower prize: \$100 USD for showing a lower bound of larger order but not satisfying the conditions for the grand prize (i.e., not showing upper and lower bounds of matching order).

References

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, 2023.

Cynthia Dwork. Differential privacy. In *International colloquium on automata, languages, and programming*, pages 1–12. Springer, 2006.

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.

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.

Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

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

Wojciech Kotłowski. Private communication, February 2019.

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.

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.