Distribution-Dependent Rates for Multi-Distribution Learning

Rafael Hanashiro RAFAH@MIT.EDU
MIT
Cambridge, MA 02139

Patrick Jaillet JAILLET@MIT.EDU
MIT
Cambridge, MA 02139

Editors: Matus Telgarsky and Jonathan Ullman

Abstract

To address the needs of modeling uncertainty in sensitive machine learning applications, the setup of distributionally robust optimization (DRO) seeks good performance uniformly across a variety of tasks. The recent multi-distribution learning (MDL) framework Awasthi et al. (2023) tackles this objective in a dynamic interaction with the environment, where the learner has sampling access to each target distribution. Drawing inspiration from the field of pure-exploration multi-armed bandits, we provide *distribution-dependent* guarantees in the MDL regime, that scale with suboptimality gaps and result in superior dependence on the sample size when compared to the existing distribution-independent analyses. We investigate two non-adaptive strategies, uniform and non-uniform exploration, and present non-asymptotic regret bounds using novel tools from empirical process theory. Furthermore, we devise an adaptive optimistic algorithm, LCB-DR, that showcases enhanced dependence on the gaps, mirroring the contrast between uniform and optimistic allocation in the multi-armed bandit literature. We also conduct a small synthetic experiment illustrating the comparative strengths of each strategy.

Keywords: multi-distribution learning, multi-armed bandits

1. Introduction

Classical statistical learning operates under the assumption that data comes from a single source Hastie et al. (2009). However, the growing use of machine learning in safety-critical applications has brought forth the demand for more robust models that address stochastic heterogeneity. One well-established paradigm is *distributionally robust optimization (DRO)* Rahimian and Mehrotra (2022), which seeks good performance uniformly across a collection of distributions. Concretely, let $\mathcal{A}$ and $\mathcal{X}$ be decision and data spaces, respectively, and suppose that data is sampled from a distribution within some *uncertainty set* $\mathcal{U} \subset \mathcal{P}(\mathcal{X})$. Under a target reward function $r : \mathcal{A} \times \mathcal{X} \to \mathbb{R}$ and distribution $Q \in \mathcal{U}$, an action $a \in \mathcal{A}$ yields expected reward $\mu(a; Q) := \mathbb{E}_{X_Q \sim Q}[r(a, X_Q)]$. DRO then focuses on the problem

$$\max_{a \in \mathcal{A}} \left\{ \mu_{\mathrm{DR}}(a) := \min_{Q \in \mathcal{U}} \mu(a; Q) \right\} \tag{DR}$$

Recent works Blum et al. (2017); Sagawa* et al. (2020); Haghtalab et al. (2022) have studied the setting of finite $\mathcal{U}$ and tackle it via interactive dynamics with the environment. More precisely, the emergent *multi-distribution learning (MDL)* framework Awasthi et al. (2023) assumes sampling access to $\mathcal{U}$, where a learning agent sequentially selects which distributions to sample from given a fixed sampling budget.

The current literature (e.g., see Awasthi et al. (2023)) is populated with *distribution-independent* rates; i.e., bounds that are independent of problem parameters. While broad in its applicability, this approach falls short in capturing the nuances of the underlying environment. Oftentimes, it is more intuitive to analyze the learner's performance in a fixed setting, as opposed to considering a worst-case instance for each sample size. When domain knowledge is available, a "one-size-fits-all" rate does not provide any insight on how to take advantage of this information.

To address these drawbacks, in this work, we study *distribution-dependent* guarantees for the MDL problem. Motivated by its close ties to the well-studied *pure exploration multi-armed bandits (PE-MAB)* Bubeck et al. (2011) paradigm, we analyze the simple strategies of uniform and non-uniform exploration, as well as their optimistic counterpart, ensuring regret guarantees that scale with suboptimality gaps and decay much faster with the sampling budget.

1.1. Main results

We place MDL algorithms into one of two categories: non-adaptive and adaptive. In the former, data is collected without any interaction with the environment and, in the latter, the learner sequentially selects distributions based on previously acquired samples. We introduce two strategies of the non-adaptive type: uniform (UE) and non-uniform (NUE) exploration (Section 3). As the names suggest, UE gathers the same number of samples from each distribution, while NUE can benefit from varied sample sizes. Using tools from empirical process theory, we provide non-asymptotic regret guarantees that scale with the suboptimality gaps of the problem and decay exponentially with the sampling budget T (Section 3.1). This stands in contrast to the distribution-independent rates found in the recent literature, which hold under a worst-case environment and, thus, only scale with $O\left(1/\sqrt{T}\right)$. From a Bernstein-type concentration inequality, we then show how NUE can exploit distributional variability to allocate samples more effectively (Section 3.2).

While the non-adaptive methods already display exponentially decreasing regret, adaptivity can further improve the dependence on instance-specific variables. Motivated by the enhancements of UCB-E Audibert et al. (2010) over uniform exploration in the PE-MAB literature, we introduce the analogous LCB-DR algorithm (Section 4) and showcase how optimism can result in superior dependence on the suboptimality gaps when compared to UE (Section 4.1).

Let $\Delta_{\mathrm{DR}}(a) := \max_{a^* \in \mathcal{A}} \mu_{\mathrm{DR}}(a^*) - \mu_{\mathrm{DR}}(a)$ be the suboptimality gap of an action from a finite set $\mathcal{A}$, and suppose that rewards are bounded in $[0, M]$. Given an algorithm, we denote its output after T sampling rounds by A_T^{o}. In short, we make the following contributions:

(i) With $n \in \mathbb{N}$ samples from each distribution, we show in Section 3.1 that UE has a simple regret decay of order

$$\mathbb{E}\left[\Delta_{\mathrm{DR}}\left(A_T^{\mathrm{o}}\right)\right] \leq \sum_{a \in \mathcal{A}: \Delta_{\mathrm{DR}}(a) > 0} \Delta_{\mathrm{DR}}(a) \exp\left(-\frac{n \Delta_{\mathrm{DR}}^2(a)}{M^2}\right)$$

Moreover, we present the distribution-independent rate $\mathbb{E}\left[\Delta_{\mathrm{DR}}\left(A_T^{\mathrm{o}}\right)\right] \lesssim \sqrt{\frac{|\mathcal{U}| \log(|\mathcal{U}||\mathcal{A}|)}{T}}$.

(ii) With $n_Q \in \mathbb{N}$ samples from distribution $Q \in \mathcal{U}$ over real-valued data and bounded reward $r \in [0, M]$, we show in Section 3.2 that NUE attains the rate

$$\mathbb{E}\left[\Delta_{\mathrm{DR}}\left(A_T^{\mathrm{o}}\right)\right] \leq \sum_{a \in \mathcal{A}: \Delta_{\mathrm{DR}}(a)>0} \Delta_{\mathrm{DR}}(a) \exp\left(-\frac{\Delta_{\mathrm{DR}}^2(a)}{\sigma_T^2+\Sigma_T^2+V_T+\frac{M\,\Delta_{\mathrm{DR}}(a)}{\min_{Q \in \mathcal{U}} n_Q}}\right)$$

where σ_T^2, Σ_T^2 and V_T are empirical process variance quantities that scale with the variances of each $Q \in \mathcal{U}$ and decrease with the n_Q.

(iii) Appealing to the principle of optimism, we devise the LCB-DR algorithm that, in a pre-specified permutation of the arms $\left(a_1, a_2, \ldots, a_{|\mathcal{A}|}\right)$, for $j = 1, \ldots, |\mathcal{A}|$, sequentially performs a modified version of UCB-E, for T_j rounds, on "losses" $\left\{\mu\left(a_j, Q\right)\right\}_{Q \in \mathcal{U}}$ as a means of identifying the worst-case distribution for a_j. In Section 4, we show that, under rewards bounded in $[0, 1]$, this guarantees an error probability of

$$\mathbb{P}\left(\Delta_{\mathrm{DR}}\left(A_T^{\mathrm{o}}\right)>0\right) \leq \sum_{j=1}^{|\mathcal{A}|} \exp\left(-\frac{\left(C_{a_j}^2 \wedge 1\right)\left(T_j+\tilde{T}_j-|\mathcal{U}_j|\right)}{H_j}\right)$$

This bound, which may be of independent interest, results from an analysis of UCB-E under a learner with previously acquired data (see Appendix D). Since the learner has already accumulated samples from previous iterations in each UCB-E batch, some "arms" can be identified as suboptimal a priori. We show that the algorithm essentially operates on a subset $\mathcal{U}_j \subset \mathcal{U}$ of the arms, whose total number $\tilde{T}_j$ of pre-collected samples contributes to the regret decay. Furthermore, while the standard analysis scales with the sum of the reciprocals of *all* suboptimality gaps, in this case, the quantity H_j sum only over the smaller set $\mathcal{U}_j$. The quantity C_a is a newly introduced complexity measure that captures the difference in difficulty between the two tasks we face: identifying a as suboptimal and finding its worst-case distribution. Drawing parallels with the MAB literature, we compare this bound to that of UE, showing that the contrast is characterized by C_a.

(iv) In Appendix F, we briefly discuss how the results can be extended to infinite decision sets.

For ease of exposition, we removed constants and terms decreasing with T inside the exponential, as well as any quantities outside of it. The formal statements are deferred to the corresponding sections.

1.2. Related work

The predominance of machine learning in society has highlighted the need for robust models that maintain high-quality performance in a multitude of scenarios. Given the inherent uncertainty in identifying the environment, much attention has been given to the problem of learning under distribution shifts Ben-David et al. (2009); Mansour et al. (2009), where training data may not necessarily be sampled from the target distribution. To tackle this, several works Volpi et al. (2018); Zhang et al. (2021); Sutter et al. (2021) have applied the framework of DRO Scarf (1958); Delage and Ye (2010); Ben-Tal et al. (2013) by assuming that the shift occurs within a neighborhood $\mathcal{U}$ of some nominal distribution, typically generating data, and solving (DR). There are many ways to construct $\mathcal{U}$ and optimize the objective, and we refer to Shapiro et al. (2021); Rahimian and Mehrotra (2022) for a thorough review.

A more recent line of work has specialized to finite and unstructured $\mathcal{U} = \{Q_1, \dots, Q_k\}$, under sampling access to each distribution. Agnostic federated learning Mohri et al. (2019) solves (DR) under mixtures of $\mathcal{U}$, providing high-probability bounds on the generalization gap of non-uniform exploration and an algorithm with empirical optimization guarantees. Collaborative PAC learning Blum et al. (2017) focuses on binary classification, with the aim of guaranteeing $\mathbb{P}\left(\Delta_{\text{DR}}\left(A_T^{\text{o}}\right) \leq \epsilon\right) \geq 1 - \delta$ under a minimal number of samples T. The original work of Blum et al. (2017) assumes realizability and subsequent studies Chen et al. (2018); Nguyen and Zakynthinou (2018); Carmon and Hausler (2022); Haghtalab et al. (2022) extended results to the agnostic case and gave improved rates, along with sample-complexity lower bounds. Awasthi et al. (2023) later solidified the theory and posed several open problems, some of which were recently addressed in Peng (2024); Zhang et al. (2024) via optimal algorithms.

In this work, we turn our attention to the simple regret $\mathbb{E}\left[\Delta_{\text{DR}}\left(A_T^{\text{o}}\right)\right]$. For finite decision sets $\mathcal{A}$, an integration of the tails reveals that the regret achieved by Haghtalab et al. (2022) is $O\left(\sqrt{\frac{\log|\mathcal{A}| + k\log k}{T}}\right)$. When $\mathcal{A} \subset \mathbb{R}^d$ has Euclidean diameter at most $B > 0$ and, for each $x \in \mathcal{X}$, the function $r(\cdot, x)$ is both convex and Lipschitz, several studies have proposed comparable rates using game dynamics. Group DRO Sagawa* et al. (2020) ensures a rate of $O\left(k\sqrt{\frac{B^2 + \log k}{T}}\right)$ and, in the fairness context, Abernethy et al. (2022) obtains $O\left(\frac{B}{\sqrt{T}}\right)$ plus a term that uniformly bounds the generalization gap with high-probability. Subsequently, Zhang et al. (2023) devised strategies with $O\left(\sqrt{\frac{B^2 + k\log k}{T}}\right)$ regret, matching the lower bound of Soma et al. (2022) up to log factors, and additionally studied the setting with distribution-specific sampling budget constraints.

Since the learner does not incur any costs when gathering data, MDL closely resembles PE-MAB Bubeck et al. (2011) under the *fixed budget* regime, where distributions represent the arms. It is standard in the MAB literature to distinguish between distribution-dependent and independent rates. The former typically depends on the suboptimality gaps and scales much faster with T. In contrast, the latter holds for worst-case environments for each T, resulting in slower regret decay. See (Lattimore and Szepesvari, 2020, Ch. 33) for an in-depth discussion. In PE-MAB, Audibert et al. (2010) introduced the UCB-E strategy, which improves performance relative to the gaps when compared to uniform exploration. Motivated by these results, we demonstrate analogous faster distribution-dependent rates in the MDL setting and explore a similar contrast between UE and LCB-DR.

2. Preliminaries

Notation. We frequently use the notation $[k] := \{1, \dots, k\}$, where $k \in \mathbb{N}$. For a measurable space $\mathcal{X}$ (we will omit the σ-algebras), we let $\mathcal{P}(\mathcal{X})$ denote the set of all distributions over it. For two real-valued functions f and g, we let $f \lesssim g$ and $f \gtrsim g$ denote inequalities up to universal constants. Given values $a, b \in \mathbb{R}$, we define $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$.

2.1. Multi-distribution learning

Let $\mathcal{X}$ be the space where our data lives in and $\mathcal{A}$ the space where we make decisions. Given data $X_Q \sim Q \in \mathcal{P}(\mathcal{X})$, statistical learning aims to maximize the stochastic objective $\mu(a;Q) = \mathbb{E}[r(a, X_Q)]$ with respect to $a \in \mathcal{A}$, where $r : \mathcal{A} \times \mathcal{X} \to \mathbb{R}$ is an underlying reward function. In the MDL paradigm, we capture distributional uncertainty by assuming that the distributions come from some uncertainty set $\mathcal{U} \subset \mathcal{P}(\mathcal{X})$ and instead aim to solve the distributionally robust problem (DR), where our goal is to maximize $\mu_{\text{DR}}(a) = \min_{Q \in \mathcal{U}} \mu(a;Q)$. We measure the performance of a decision $a \in \mathcal{A}$ via its *suboptimality gap* $\Delta_{\text{DR}}(a) := \mu^*_{\text{DR}} - \mu_{\text{DR}}(a)$, where $\mu^*_{\text{DR}} := \max_{a \in \mathcal{A}} \mu_{\text{DR}}(a)$ is the optimal objective value. Throughout this work, we operate under the following assumptions.

Assumption 1 (Finite decision/uncertainty sets) *$|\mathcal{A}| = l$ and $|\mathcal{U}| = k$, where $2 \leq l, k < \infty$.*

Assumption 2 (Bounded rewards) *The reward function r is bounded in $[0, M]$, for some $M > 0$.*

To solve (DR), we interact with the environment for a total of $T \in \mathbb{N}$ rounds. In each round $t \in [T]$, we (i) select a distribution $Q_t \in \mathcal{U}$ and (ii) receive independent data point $X_t \sim Q_t$. After the T rounds, we output a decision $A^{\text{o}}_T \in \mathcal{A}$ with the goal of minimizing the *simple regret* $\mathbb{E}[\Delta_{\text{DR}}(A^{\text{o}}_T)]$ or *error probability* $\mathbb{P}(\Delta_{\text{DR}}(A^{\text{o}}_T) > 0)$. The strategies described in this work are of the form $A^{\text{o}}_T = \operatorname{argmax}_{a \in \mathcal{A}} \mu^{\text{o}}_T(a)$ for an appropriately constructed proxy $\mu^{\text{o}}_T : \mathcal{A} \to \mathbb{R}$.

Remark 3 (Simple regret v.s. error probability) *Note that both performance measures are closely related: since $r \in [0, M]$, we have that $\Delta_{\text{DR}} \in [0, M]$ and, thus,*

$$\Delta_{\text{DR,min}} \, \mathbb{P}(\Delta_{\text{DR}}(A^o_T) > 0) \leq \mathbb{E}[\Delta_{\text{DR}}(A^o_T)] \leq M \mathbb{P}(\Delta_{\text{DR}}(A^o_T) > 0)$$

where $\Delta_{\text{DR,min}}$ is the minimal positive gap (see Section 2.2).

2.2. Complexity measures

For each decision $a \in \mathcal{A}$, we define its worst performing distribution $Q^*_a := \operatorname{argmin}_{Q \in \mathcal{U}} \mu(a;Q)$ and the suboptimality gaps $\Delta_a(Q) := \mu(a;Q) - \mu_{\text{DR}}(a)$. Much of the analysis that follows is characterized by the minimal positive gaps

$$\Delta_{\text{DR,min}} := \min\{\Delta_{\text{DR}}(a) > 0 : a \in \mathcal{A}\} \quad \text{and} \quad \Delta_{a,\text{min}} := \min\{\Delta_a(Q) > 0 : Q \in \mathcal{U}\}$$

These quantities are additionally used to define complexity measures

$$H_a := \sum_{Q \in \mathcal{U} : \Delta_a(Q) > 0} \Delta_a^{-2}(Q) \quad \text{and} \quad C_a := \begin{cases} \dfrac{\Delta_{\text{DR}}(a)}{\Delta_{a,\text{min}}}, & a \notin \operatorname*{argmax}_{a \in \mathcal{A}} \mu_{\text{DR}}(a) \\ \dfrac{\Delta_{\text{DR,min}}}{\Delta_{a,\text{min}}}, & \text{otherwise} \end{cases}$$

for each $a \in \mathcal{A}$. In pure exploration bandits, H_a is commonly used to characterize the complexity of identifying the optimal arm (e.g., Audibert et al. (2010)), which in our setting translates to identifying Q^*_a. The intuition behind C_a is that it compares the difficulty of the two tasks we face: when $C_a \leq 1$ for some $a \notin \operatorname{argmax}_{a \in \mathcal{A}} \mu_{\text{DR}}(a)$, or $\Delta_{\text{DR}}(a) \leq \Delta_{a,\text{min}}$, it is more challenging to rule out a as suboptimal than it is to identify Q^*_a.

2.3. Algorithmic tools

For each distribution $Q \in \mathcal{U}$, let $X_Q, \left\{X_Q^{(i)}\right\}_{i=1}^{\infty} \overset{iid}{\sim} Q$ be a sequence of independent data points. For each $(t, a, Q) \in \mathbb{N} \times \mathcal{A} \times \mathcal{U}$, we define the empirical mean $\hat{\mu}_t (a; Q) \coloneqq \frac{1}{t} \sum_{i=1}^{t} r \left(a, X_Q^{(i)}\right)$. Under a fixed sampling algorithm, let $n_t (Q) \coloneqq \sum_{s=1}^{t} \mathbb{I} \{Q_s = Q\}$ denote the number of times that Q is played up to time t. The data received is then given by $X_t = X_{Q_t}^{(n_t(Q_t))}$.

3. Non-adaptive strategies

We begin by describing two simple non-adaptive strategies. In essence, both sample a fixed number of times from each distribution in $\mathcal{U}$ and construct a proxy μ_T^{o} that is the natural empirical version of μ_{DR}. Proofs of the results are deferred to Appendix C.

3.1. Uniform exploration (UE)

The most straight-forward strategy is the idea of *uniform exploration (UE)* (Algorithm 1). As the name suggests, we sample the same number $n \in \mathbb{N}$ of times from each distribution, for a total of $T = nk$ samples, and form the empirical proxy $\mu_T^{\mathrm{o}} (a) = \min_{Q \in \mathcal{U}} \hat{\mu}_n (a; Q)$.

Algorithm 1: Uniform Exploration (UE)

Input: Number of samples $n \in \mathbb{N}$
1 Sample n times from each distribution $Q \in \mathcal{U}$
2 Construct $\mu_T^{\mathrm{o}}(a) = \min_{Q \in \mathcal{U}} \hat{\mu}_n(a; Q)$
Output: $A_T^{\mathrm{o}} = \arg\max_{a \in \mathcal{A}} \mu_T^{\mathrm{o}}(a)$

Theorem 4 (UE regret) *Suppose that* $n \geq \left(\frac{8M}{\Delta_{\mathrm{DR,min}}}\right)^2 \log k$. *Then, the UE algorithm attains the following simple regret bound:*

$$\mathbb{E} \left[\Delta_{\mathrm{DR}} \left(A_T^o\right)\right] \leq \sum_{a \in \mathcal{A} : \Delta_{\mathrm{DR}}(a) > 0} \Delta_{\mathrm{DR}} (a) \exp \left(-\frac{n}{2M^2} \left[\Delta_{\mathrm{DR}} (a) - 8M \sqrt{\frac{\log k}{n}} \right]^2 \right)$$

Since our empirical proxy μ_T^{o} is not an unbiased estimate of μ_{DR}, we end up with an approximation error bounded by $M \sqrt{\frac{\log k}{n}}$. The lower bound on n is then required to apply tail bounds by ensuring that $\Delta_{\mathrm{DR}} (a) - 8M \sqrt{\frac{\log k}{n}} \geq 0$ for all $a \in \mathcal{A}$ (see Appendix C.1).

Remark 5 (Small gaps) *When* $\Delta_{\mathrm{DR,min}}$ *is really small, the lower bound condition on* n *may be difficult to attain. This may be counterintuitive, for example, when all gaps are small, as we expect the problem to be easy. In such situations, an alternative guarantee is*

$$\mathbb{E} \left[\Delta_{\mathrm{DR}} \left(A_T^o\right)\right] \leq \Delta + \sum_{a \in \mathcal{A} : \Delta_{\mathrm{DR}}(a) > \Delta} \Delta_{\mathrm{DR}} (a) \exp \left(-\frac{n}{2M^2} \left[\Delta_{\mathrm{DR}} (a) - 8M \sqrt{\frac{\log k}{n}} \right]^2 \right)$$

for any $\Delta > 0$, *provided that* $n \geq \left(\frac{8M}{\Delta}\right)^2 \log k$.

With some further manipulation, we can additionally obtain a distribution-independent regret bound.

Corollary 6 (UE distribution-independent regret) *Suppose that* $n \geq \left(\frac{8}{\Delta_{\mathrm{DR,min}}}\right)^2 \log k$ *and* $M \geq 1$. *Then, UE attains the following distribution-independent regret:* $\mathbb{E}\left[\Delta_{\mathrm{DR}}\left(A_T^o\right)\right] \lesssim M\sqrt{\frac{k\log(kl)}{T}}$.

3.2. Non-uniform exploration (NUE)

A natural extension of the UE strategy is to sample a different number of times from each distribution. To address this, *non-uniform exploration (NUE)* (Algorithm 2) samples $n_Q \in \mathbb{N}$ times from each distribution $Q \in \mathcal{U}$, for a total of $T = \sum_{Q\in\mathcal{U}} n_Q$ samples. Similarly, we define the proxy $\mu_T^{\mathrm{o}}(a) = \min_{Q\in\mathcal{U}} \hat{\mu}_{n_Q}(a; Q)$.

Since our goal is to showcase the interplay between sample size and variance, here we focus on real-valued data in $\mathcal{X} \subset \mathbb{R}$ and Lipschitz reward functions r. Let us define mean $\mu_Q \coloneqq \mathbb{E}\left[X_Q\right]$ and variance $\sigma_Q^2 \coloneqq \mathrm{Var}\left(X_Q\right)$ for each $Q \in \mathcal{U}$. Additionally, let us sort the sample sizes in increasing order as follows: $0 =: n_{(0)} \leq n_{(1)} \leq \cdots \leq n_{(k)}$ and let $Q_{(j)}$ denote the corresponding distribution in the jth position: $n_{Q_{(j)}} = n_{(j)}$. The regret bound presented will rely on the following variance quantities:

$$V_T \coloneqq \sum_{j=1}^{k} \left(n_{(j)} - n_{(j-1)}\right) \mathbb{E}\left[\max_{r\in\{j,\ldots,k\}} \frac{1}{n_{(r)}^2}\left[X_{Q_{(r)}} - \mu_{Q_{(r)}}\right]^2\right]$$

$$\Sigma_T^2 \coloneqq \mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \left(X_Q^{(i)} - \mu_Q\right)^2\right]$$

$$\sigma_T^2 \coloneqq \max_{Q\in\mathcal{U}} \frac{\sigma_Q^2}{n_Q}$$

Lastly, we make use of the quantity $G_T \coloneqq \frac{32M\log k}{\min_{Q\in\mathcal{U}} n_Q} + 8L\sigma_T\sqrt{2\log k}$, which we note decreases with the $\{n_Q\}$. This is the analogue of $8M\sqrt{\frac{\log k}{n}}$ found in the UE analysis.

Algorithm 2: Non-uniform exploration (NUE)

Input: Number of samples $\{n_Q\}_{Q\in\mathcal{U}} \subset \mathbb{N}$ allocated to each distribution

Sample n_Q times from each distribution $Q \in \mathcal{U}$

Construct $\mu_T^{\mathrm{o}}(a) = \min_{Q\in\mathcal{U}} \hat{\mu}_{n_Q}(a; Q)$

Output: $A_T^{\mathrm{o}} = \mathrm{argmax}_{a\in\mathcal{A}}\, \mu_T^{\mathrm{o}}(a)$

Theorem 7 (NUE regret) *Suppose that* $r(a, \cdot)$ *is L-Lipschitz for each* $a \in \mathcal{A}$, *and that* $\Delta_{\mathrm{DR,min}} \geq G_T$. *Then, the NUE algorithm attains the following simple regret bound:*

$$\mathbb{E}\left[\Delta_{\mathrm{DR}}\left(A_T^o\right)\right] \leq \sum_{a\in\mathcal{A}:\Delta_{\mathrm{DR}}(a)>0} \Delta_{\mathrm{DR}}(a) \exp\left(-\frac{\left[\Delta_{\mathrm{DR}}(a) - G_T\right]^2}{16L^2\left(2\sigma_T^2 + \Sigma_T^2 + 6V_T\right) + \frac{2\sqrt{6}M}{\min_{Q\in\mathcal{U}} n_Q}\left[\Delta_{\mathrm{DR}}(a) - G_T\right]}\right)$$

As intuition suggests, the definitions imply that sampling more from distributions with higher variance yields better rates. On the other hand, due to the presence of $\min_{Q\in\mathcal{U}} n_Q$ in the bound, it may also be favorable to balance this principle with ensuring that no distribution is significantly undersampled.

3.3. Uniform v.s. non-uniform exploration

We briefly note that the NUE bound recovers the UE one. Suppose that the sample sizes are uniform and equal to n, and let us make the mild assumption that $n \geq \log k$. Instead of exploiting the Lipschitzness of r in Section B.2.6, we can exploit its boundedness to readily conclude that $V_T \lesssim M^2/n$, which also serves as an upper bound on σ_T^2 and Σ_T^2. Since $\Delta_{\mathrm{DR}}(a) \leq M$, the denominator in the exponential term of the NUE regret is then upper bounded by $\sim M^2/n$. By additionally bounding $G_T \lesssim M\sqrt{\frac{\log k}{n}}$, we recover the UE bound.

The benefit of using non-uniform sampling, however, is revealed by examining how variance plays a role in the NUE bound. Let us more generally express the probability of selecting a suboptimal arm $a \in \mathcal{A}$ for UE and NUE as follows (see Appendix C):

$$\underbrace{\exp\left(-\frac{n}{M^2}\left[\Delta_{\mathrm{DR}}(a) - B_T\right]^2\right)}_{\text{UE}} \quad \text{v.s.} \quad \underbrace{\exp\left(-\frac{\left[\Delta_{\mathrm{DR}}(a) - B_T\right]^2}{\sigma_T^2 + \Sigma_T^2 + V_T + \frac{M}{\min_Q n_Q}\left[\Delta_{\mathrm{DR}}(a) - B_T\right]}\right)}_{\text{NUE}}$$

where we have omitted constants. Here, B_T is a quantity that decreases with the sample size and is the same in both rates. In Theorems 4 and 7, we set it to $M\sqrt{\frac{\log k}{n}}$ and G_T, respectively, but either choice applies. To mirror the standard Hoeffding v.s. Bernstein discussion, consider a small-sample regime where $\Delta_{\mathrm{DR}}(a) - B_T$ is suitably small. The comparison then reduces to $\frac{M^2}{n}$ (for UE) v.s. $\sigma_T^2 + \Sigma_T^2 + V_T$ (for NUE), where the smaller term is better. Note that M captures the range of the reward function r, while σ_T^2, Σ_T^2 and V_T capture the variance of the distributions in $\mathcal{U}$. The latter is more nuanced and can be favorable when the reward takes large values but the data concentrates in a small region. This shows that NUE can be better when the learner allocates more samples to distributions with higher variance.

3.4. Bounds on variance quantities

While the variance quantities introduced seem hard to control and lack interpretability, here we highlight some strategies and examples to mitigate this issue. Proofs of all results below are deferred to Appendix G.

3.4.1. CRUDE BOUND

Note the variance hierarchy $\sigma_T^2 \leq \Sigma_T^2 \leq V_T$. To unify them, we can bound the max with a sum to get $V_T \leq \sum_{Q\in\mathcal{U}} \frac{\sigma_Q^2}{n_Q}$, which we can then substitute all three terms with. However, this results in a linear dependence on k that we aim to avoid.

3.4.2. BOUNDING Σ_T^2

Suppose that our data is bounded: $X_Q \in [0,1]$ for each $Q \in \mathcal{U}$. Then we can establish the following upper bound: $\Sigma_T^2 \lesssim \sqrt{\frac{\log k}{\min_{Q\in\mathcal{U}} n_Q^3}} + \sigma_T^2$. Since the first term on the right-hand side decays faster than $O\left(\frac{1}{\min_{Q\in\mathcal{U}} n_Q}\right)$, we can focus our attention on σ_T^2, which is a more interpretable quantity.

3.4.3. BOUNDING V_T

The most formidable quantity is V_T, but we can readily relate it to Σ_T^2 via the relationship $V_T \leq \min\left\{\max_{Q\in\mathcal{U}} n_Q, k\right\} \Sigma_T^2$. In a setting where k is not too large, this result shows that control over Σ_T^2 also ensures control over V_T.

For a more concrete example, suppose that $\mathcal{U} = \{Q_1, \ldots, Q_k\}$, where $Q_1, \ldots, Q_{k-1}$ share a common small variance σ^2 and Q_k has a much larger variance $\nu^2 \gg \sigma^2$. In addition, suppose that $Q_1, \ldots, Q_{k-1}$ are supported in $[0,1]$. Consider the NUE procedure with n samples from each $Q_1, \ldots, Q_{k-1}$ and $m = T - n(k-1) \geq n$ samples (where $T \geq nk$ is the total number of samples) from Q_k. Intuitively, we would like for $m \gg n$ since Q_k is harder to learn (i.e., has more variability). This can be reflected in the strong variance: $V_T \lesssim \frac{\sqrt{\log k}+\sigma^2}{n} + \frac{\nu^2}{T-nk}$. Comparing with the UE counterpart (and ignoring σ_T^2 and Σ_T^2 since V_T is the dominating term) M^2k/T, we note that NUE can decay much faster when ν^2, M^2, k and T are large relative to σ^2 and n.

For example, consider $\sigma^2 = 1/4$ and $\nu^2 = M = k = C > 1$. Suppose that n sample are allocated to $Q_1, \ldots, Q_{k-1}$ and $n(C+1)$ samples to Q_k. Let us use the first bound of Theorem 27: $B_T \asymp M\sqrt{\frac{\log k}{\min_{Q\in\mathcal{U}} n_Q}} \asymp C\sqrt{\frac{\log C}{n}}$. Note that this also holds for UE since it allocates $2n$ samples to each distribution. Using the fact that $V_T \lesssim \frac{\sqrt{\log C}}{n}$, the exact rates for both strategies are then

$$\underbrace{\exp\left(-\frac{n}{C^2}\left[\Delta_{\mathrm{DR}}(a) - C\sqrt{\frac{\log C}{n}}\right]^2\right)}_{\text{UE}} \quad \text{v.s.} \quad \underbrace{\exp\left(-\frac{n\left[\Delta_{\mathrm{DR}}(a) - C\sqrt{\frac{\log C}{n}}\right]^2}{\sqrt{\log C} + C\left[\Delta_{\mathrm{DR}}(a) - C\sqrt{\frac{\log C}{n}}\right]}\right)}_{\text{NUE}}$$

Introduce $\epsilon_a := \Delta_{\mathrm{DR}}(a) - C\sqrt{\frac{\log C}{n}}$. Then the comparison becomes C^2 (for UE) v.s. $\sqrt{\log C} + C\epsilon_a$ (for NUE), where the smaller term is better. Since $\epsilon_a \leq C$, we see that the latter is always smaller and, when $\epsilon_a \ll C$, NUE provides a significantly sharper bound for this action. We highlight that, in this example, the quantity $\min_{Q\in\mathcal{U}} n_Q$ in the NUE rate did not pose an issue, as it is equal, up to constants, to the UE allocations (i.e., n v.s. $2n$). In addition, applying the Bernstein bound to UE does not help since the variance quantities are still polynomial in C.

4. Optimism

As opposed to the non-adaptive strategies covered thus far, the next algorithm we present makes sampling decisions as it interacts with the environment. For this analysis, we additionally operate under the following uniqueness assumption.

Assumption 8 (Reward bound and unique optima) *We assume that* $r \in [0,1]$*; i.e., we specialize to* $M = 1$*. Moreover, assume that* $a^* = \operatorname{argmax}_{a \in \mathcal{A}} \mu_{\mathrm{DR}}(a)$ *and* $Q_a^* = \operatorname{argmin}_{Q \in \mathcal{U}} \mu(a; Q)$ *are the unique optimal decision and the unique worst-case distribution for* a*, respectively.*

As is standard in UCB-style algorithms, for some choice of parameter $\epsilon > 0$, we define *index*

$$\mathrm{LCB}_t(Q; a, \epsilon) \coloneqq \hat{\mu}_{n_t(Q)}(a; Q) - \sqrt{\frac{\epsilon}{n_t(Q)}} \quad \forall (t, a, Q) \in \mathbb{N} \times \mathcal{A} \times \mathcal{U}$$

which represents a *lower confidence bound (LCB)* on the true mean $\mu(a; Q)$. At a high-level, the *LCB-DR* strategy (Algorithm 3) iterates through each decision $a \in \mathcal{A}$ and performs a modified version of UCB-E Audibert et al. (2010) to identify Q_a^*. The modification takes advantage of the fact that data sampled in a previous round can be reused for the current one. In essence, we analyze UCB-E when each distribution starts the game with a certain number of pulls. Intuitively, if some distribution has already been played sufficiently many times, it will not be played again in this round, yielding an improved sample complexity.

For completeness, we initiate the procedure by sampling from each distribution once; that is, $n_k(Q) \coloneqq 1$ for each $Q \in \mathcal{U}$. As a result, we define $T_0 \coloneqq \bar{T}_0 \coloneqq k$ to be the total number of samples gathered before the game starts. The inputs to the algorithm are a permutation $(a_1, \ldots, a_l)$ of $\mathcal{A}$, dictating the order in which decisions are iterated through, and non-negative index parameters $(\epsilon_1, \ldots, \epsilon_l)$. The procedure then works as follows: at each round $j \in [l]$,

1. Since we reuse samples from previous rounds, some distributions may already have enough samples by the start of the current round and, thus, may not be sampled from at all. We define $\mathcal{U}_j \coloneqq \left\{ Q \in \mathcal{U} : n_{\bar{T}_{j-1}}(Q) < \frac{36}{25} \epsilon_j \Delta_{a_j}^{-2}(Q) \right\}$ as a proxy for the arms that will be played in this round, where we use the convention $\Delta_{a_j}\left(Q_{a_j}^*\right) \coloneqq \Delta_{a_j,\min}$. We additionally define $k_j \coloneqq |\mathcal{U}_j| \, \mathbb{I}\left\{Q_{a_j}^* \in \mathcal{U}_j\right\}$, $\tilde{T}_j \coloneqq \sum_{Q \in \mathcal{U}_j} n_{\bar{T}_{j-1}}(Q)$ and $H_j \coloneqq \sum_{Q \in \mathcal{U}_j} \Delta_{a_j}^{-2}(Q)$.

2. Let $\bar{T}_{j-1} \coloneqq \sum_{r=0}^{j-1} T_r$ denote the total number of samples obtained up to and including round $j-1$. Allocate

$$T_j \geq \frac{36}{25} \epsilon_j H_j - \tilde{T}_j + k_j \tag{1}$$

samples to this round such that $\bar{T}_{j-1} + T_j \leq u_j$ for some deterministic quantity u_j. This simply ensures that we don't choose T_j arbitrarily large, but the specific value of u_j is less important as the regret decays logarithmically (relative to sample size) with it.

3. For each $t = \bar{T}_{j-1} + 1, \ldots, \bar{T}_j$, sample $X_t \sim Q_t \coloneqq \operatorname{argmin}_{Q \in \mathcal{U}} \mathrm{LCB}_{t-1}(Q; a_j, \epsilon_j)$. In essence, we play the modified UCB-E for T_j rounds on expected rewards $\{\mu(a_j; Q)\}_{Q \in \mathcal{U}}$. We emphasize that the learner *does not* need to know $\mathcal{U}_j$, since the minimization is over all of $\mathcal{U}$.

4. Define $\hat{Q}_j \coloneqq \operatorname{argmin}_{Q \in \mathcal{U}} \hat{\mu}_{n_{\bar{T}_j}(Q)}(a_j; Q)$ and $\mu_T^{\mathrm{o}}(a_j) \coloneqq \hat{\mu}_{n_{\bar{T}_j}(\hat{Q}_j)}\left(a_j; \hat{Q}_j\right)$. Intuitively, $\hat{Q}_j$ and μ_T^{o} are proxies for $Q_{a_j}^*$ and μ_{DR}, respectively.

Algorithm 3: LCB-DR

Input: Initial number of samples $T_0 = \bar{T}_0 = k$, permutation $(a_1, \dots, a_l)$ of $\mathcal{A}$ and index parameters $(\epsilon_1, \dots, \epsilon_l)$.

```
for j = 1 to l do
    Define proxy set U_j and quantities k_j, T̃_j and H_j.
    Allocate T_j samples to this round.
    for t = T̄_{j-1} + 1 to T̄_j do
        Sample data point X_t ~ Q_t.
    end
    Define proxies Q̂_j and μ^o_T(a_j).
end
```

Output: $A^{\mathrm{o}}_T = \mathrm{argmax}_{a\in\mathcal{A}}\, \mu^{\mathrm{o}}_T(a)$.

Finally, after gathering $T := \sum_{j=0}^{l} T_j$ total samples, we maximize the proxy objective: $A^{\mathrm{o}}_T := \mathrm{argmax}_{a\in\mathcal{A}}\, \mu^{\mathrm{o}}_T(a)$. By analyzing the optimality of the modified UCB-E algorithm (see Appendix D), we can then reach the following conclusion.

Theorem 9 (LCB-DR error probability) *Under Assumption 8, the LCB-DR algorithm attains the following error probability:*

$$\mathbb{P}\left(A^o_T \neq a^*\right) \leq 2k \sum_{j=1}^{l} u_j \exp\left(-\frac{2\left(C^2_{a_j} \wedge 1\right)\epsilon_j}{25}\right)$$

Note that regret decays with ϵ_j, so to ensure good dependence on the sample sizes, our goal is to make the lower bound (1) as tight as possible. When it holds with equality, we can set $u_0 = k$ and $u_j = k\,(j+1) + \frac{72}{25}\sum_{r=1}^{j} \epsilon_r H_{a_r}$ (see Appendix E.2). In addition, under equality, we have that $\epsilon_j = \frac{25}{36}\frac{T_j + \tilde{T}_j - k_j}{H_j}$, so that the decay of each term scales with $O\left(\frac{\left(C^2_{a_j}\wedge 1\right)\left(T_j+\tilde{T}_j-k_j\right)}{H_j}\right)$. Intuitively, at each round $j \in [l]$, the sample complexity depends on the difficulty of identifying the worst-case distribution $Q^*_{a_j}$, which, as in PE-MAB, is controlled by the suboptimality gaps $\left\{\Delta_{a_j}(Q)\right\}_{Q\in\mathcal{U}}$.

Remark 10 (Improvement over UCB-E) *We highlight the importance of using samples obtained in previous rounds: as opposed to the standard UCB-E analysis, we have the additional $\tilde{T}_j$ contribution, we only offset by $k_j \leq k$, and the complexity measure H_j improves upon H_{a_j} by only summing over a subset of $\mathcal{U}$.*

In practice, to tighten the lower bound (1), we must deal with two sources of uncertainty: (i) the proxy set $\mathcal{U}_j$ and (ii) the suboptimality gaps $\Delta_{a_j}(Q)$. Let us address these separately:

(i) Recall that $\mathcal{U}_j$ is a proxy for the set of arms $\mathcal{U}'_j$ that will be played in round j. In Appendix D, we show that with high-probability, $\mathcal{U}'_j \subset \mathcal{U}_j$. In fact, they will be equal except for arms satisfying $\frac{4}{25}\epsilon_j\Delta^{-2}_{a_j}(Q) < n_{\bar{T}_{j-1}}(Q) < \frac{36}{25}\epsilon_j\Delta^{-2}_{a_j}(Q)$. As a result, $\tilde{T}_j$ is approximately the number of previously collected samples from arms that *will be played* in this round. We can thus keep track of an evolving lower bound as we collect data. Note that we should choose

ϵ_j large enough so that we uncover $\mathcal{U}_j'$ before (1) becomes loose. In addition, we can trivially bound $k_j \leq k$, since it will typically be negligible relative to $\tilde{T}_j$ and H_j.

(ii) To handle H_j, we can estimate the $\Delta_{a_j}(Q)$ online as data is collected in round j, which has been empirically shown in Audibert et al. (2010) to perform well. Note that the sum being over the unknown $\mathcal{U}_j$ is not much of an issue since, as discussed above, it is approximately the same as the true set of arms played.

In Section 5, we provide an experiment on synthetic data that concretely implements these ideas.

4.1. Adaptive v.s. non-adaptive

We saw in Section 3.3 that NUE can boost performance by leveraging the interplay between sample size and variance. In a similar spirit, LCB-DR provides a complementary advantage over UE by adaptively allocating samples.

Focusing on the dominating terms, the probability of selecting a suboptimal arm $a_j \in \mathcal{A}$, that is in the jth permutation position for LCB-DR, is $\approx \exp\left(-\frac{T\Delta^2_{\text{DR}}(a_j)}{k}\right)$ for UE and $\approx \exp\left(-\frac{\left(C^2_{a_j} \wedge 1\right)\left(T_j + \tilde{T}_j\right)}{H_j}\right)$ for LCB-DR when the lower bound (1) is tight. Extracting the quantity inside the exponential, we break it down into two cases:

- $\Delta_{\text{DR}}(a_j) \leq \Delta_{a_j,\min}$ (or $C_{a_j} \leq 1$): intuitively, this means that it is more difficult to rule out a_j as suboptimal than to identify $Q^*_{a_j}$. Then, the comparison reduces to $\frac{T}{k\Delta^{-2}_{a_j,\min}}$ (for UE) v.s. $\frac{T_j+\tilde{T}_j}{H_j}$ (for LCB-DR).

- $\Delta_{a_j,\min} \leq \Delta_{\text{DR}}(a_j)$ (or $C_{a_j} \geq 1$): intuitively, this means that it is more difficult to identify $Q^*_{a_j}$ than to rule out a_j as suboptimal. Then, the comparison is between $\frac{T}{k\Delta^{-2}_{\text{DR}}(a_j)}$ (for UE) v.s. $\frac{T_j+\tilde{T}_j}{H_j}$ (for LCB-DR).

These yield $\frac{T}{k\min\left\{\Delta^{-2}_{a_j,\min}, \Delta^{-2}_{\text{DR}}(a_j)\right\}}$ (for UE) v.s. $\frac{T_j+\tilde{T}_j}{H_j}$ (for LCB-DR), where the larger term is the better rate. When sample sizes are large relative to l, so that $T \approx T_j + \tilde{T}_j$, optimism is favorable when $H_j \leq k\min\left\{\Delta^{-2}_{a_j,\min}, \Delta^{-2}_{\text{DR}}(a_j)\right\}$. As in MAB, this is always the case when $\Delta^{-2}_{a_j,\min}$ is the smaller term; otherwise, it depends on the problem instance. Note that H_j can be much smaller when $|\mathcal{U}_j| \ll k$.

5. Experiments

We conduct a small synthetic experiment to illustrate the power of our different techniques and demonstrate important practical considerations. Suppose that our distributions are Bernoulli's $P_i = \text{Ber}(p_i)$ and our decisions live in the set $\mathcal{A} \subset [0,1]$. Instead of a reward function r, here we will work the quadratic loss function $\ell(a,x) := (a-x)^2$. That is, our objective is now of the form $\min_{a\in\mathcal{A}} \max_{i\in[k]} \mathbb{E}_{X\sim\text{Ber}(p_i)}\left[(a-X)^2\right]$.

5.1. Non-adaptive comparisons

We begin by comparing the non-adaptive algorithms. Consider the action set $\{0, 1/15, \dots, 14/15\}$ and the biases $\{0.4, 0.1, 0.11, 0.12, 0.13\}$. The optimal action is $6/15 = 0.4$. Notably, one distribution has much higher variance than the rest. This adheres to the example given in Section 3.4.2. We are given a budget of $T = 1000$ samples and test the following strategies:

- UE: sample 200 times from each distribution. This yielded an error probability of 0.36.
- NUE: compute the *exact* variance of each distribution and allocate proportional to it. This yielded an error of 0.22.
- NUE with estimation: use the first 200 samples to estimate the variances and allocate proportional to these estimates. The resulting error was 0.24

The third approach is key when we do not have a priori knowledge of the distributions, which is often the case. We plot the performances obtained in Figure 1(*a*). As expected, sampling more from high-variance distributions yielded a smaller error. In particular, we see that even with a "burn-in" estimation phase, NUE outperformed UE.

5.2. Comparing all methods

Next, we run all strategies in the following setup: we define action set $\mathcal{A} = \{0.37, 0.61\}$ and biases $\{0.2, 0.5, 0.5, 0.5, 0.5, 0.5, 0.75\}$. These values were chosen so that the distributionally robust loss of both actions are very similar; 0.61 is slightly better than 0.37. In addition, the middle distributions are less informative for the robust performance. We test the following strategies:

- LCB-DR: with full knowledge of the distributions, and setting ϵ_j's to the fixed constant 5, we sampled a total of approximately 11500 times (averaged over 500 runs) and obtained an error probability of 0.045.
- UE: we ran it on the same sample size obtained by LCB-DR (i.e., 11500 samples) and obtained an error of 0.16.
- NUE with estimation: using $1/8$ of the 11500 samples for variance estimation, NUE yielded an error of 0.19. Notably, allocating more samples to high-variance distributions hurt performance in this setup.
- LCB-DR with estimation: we ran a variant of LCB-DR that estimates the suboptimality gaps online; we describe it in detail below. This yielded an average sample size of approximately 10000 and error of 0.1.

A plot of the performances is given in Figure 1(*b*). Evidently, LCB-DR significantly outperformed its non-adaptive counterparts, even when we estimate the gaps online.

5.3. LCB-DR with estimation

On round j, recall that there are two sources of uncertainty in the LCB-DR algorithm: our guess $\mathcal{U}_j$ of the distributions that will be played and the suboptimality gaps $\Delta_{a_j}(Q)$. As discussed in

Section 4, the former is approximately the set of distributions that are indeed played, which we can learn along the way.

To handle the latter, we follow the recipe of (Audibert et al., 2010, Figure 4): consider UCB-E on k arms over T rounds. To tune the exploration parameter ϵ optimally, one needs to know $H = \sum_i \Delta_i^{-2}$. The idea is to split T into k phases of equal size. In phase m, we can estimate the gaps empirically, and it turns out that the m largest ones, call them $\hat{\Delta}_{(k-m)}, \ldots, \hat{\Delta}_{(k)}$, are well-approximated (see (Audibert et al., 2010, Theorem 3)). We can then use them to compute an approximation of the complexity measure of interest, $H = \tilde{\Theta}\left(\max_i i\Delta_{(i)}^{-2}\right) \approx \max_{k-m \le i \le k} i\hat{\Delta}_{(i)}^{-2}$, and tune ϵ_j accordingly. The first equality is a well-known fact, and we refer to (Audibert et al., 2010, Section 6.1) for a proof. As we progress through the phases and collect more data, this approximation improves.

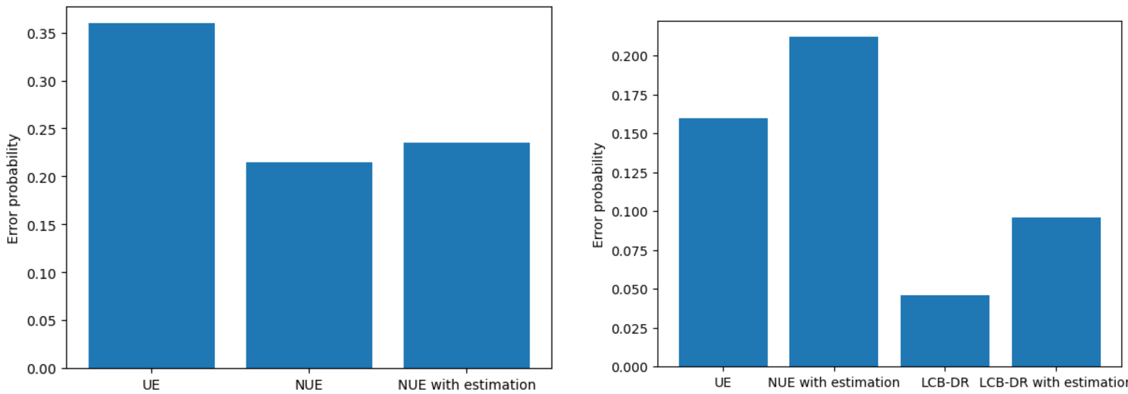
A key difference in LCB-DR is that we use H to tune T instead of ϵ. For this reason, we implement it as follows:

- Start with a large value of T_j and initialize $\mathcal{U}_j$ to be the singleton set of the first distribution played.
- Run UCB-E normally and update $\mathcal{U}_j$ with the sampled distributions. However, do not yet update T_j when doing so.
- Once we get past some fixed fraction of T_j, say 10%, check if we need to update it as follows:
  - Estimate all of the gaps empirically.
  - Take the largest gap, call it $\hat{\Delta}_{(k_j)}$, for which the distribution is in $\mathcal{U}_j$ (since H_j only sums over $\mathcal{U}_j$). We start in phase $m = 0$ and construct the estimate $\hat{H}_j^{(m)} = k_j \hat{\Delta}_{(k_j)}^{-2} = \max_{k_j - m \le i \le k_j} i\hat{\Delta}_{(i)}^{-2}$. We then define a temporary $\hat{T}_j^{(m)}$ based on it (using (1)) and check whether the current iteration is past $(m+1)\,\hat{T}_j^{(m)}/k_j$. If it is, then the $\hat{H}_j^{(m)}$ estimate should be sound, due to the reasoning above, in which case we set T_j to $\hat{T}_j^{(m)}$ and continue the process: take the two largest gaps from $\mathcal{U}_j$ (now $m = 1$), and so forth. When we reach a point where the current iteration is not large enough, then we go back to running UCB-E.

6. Discussion

In this work, we delve into the problem of DRO within the MDL framework, an area of growing popularity in high-stakes machine learning applications. Rooted in empirical process theory and inspired by the PE-MAB literature, we offer novel insight into the key strategies of uniform and non-uniform exploration via distribution-dependent bounds. By scaling with instance-specific quantities, our proposed bounds decay much faster, with respect to sample sizes, than existing ones. We additionally devise an optimistic method, LCB-DR, that shows improvements over its non-adaptive counterparts, paralleling classical findings in the MAB setting.

While LCB-DR exhibits favorable rates, we reiterate that tuning certain parameters involves estimating unknown quantities. This raises the question of whether there exists a more astute way to select such quantities with minimal prior information. Moreover, the procedure requires specifying



(*a*) Non-adaptive methods (*b*) All methods

Figure 1: Error probabilities achieved by the different methods proposed.

the order to play the actions in. Although the absence of any problem knowledge might preclude exploiting this sequence effectively, perhaps some preliminary understanding of the distributions allows potential advantages (e.g., start with actions that explore as much as possible, so that $\mathcal{U}_j$ is small in future iterations).

References

Jacob D Abernethy, Pranjal Awasthi, Matthäus Kleindessner, Jamie Morgenstern, Chris Russell, and Jie Zhang. Active sampling for min-max fairness. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 53–65. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/abernethy22a.html.

Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best arm identification in multi-armed bandits. In *Annual Conference Computational Learning Theory*, 2010. URL https://api.semanticscholar.org/CorpusID:216050617.

Pranjal Awasthi, Nika Haghtalab, and Eric Zhao. Open problem: The sample complexity of multi-distribution learning for VC classes. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 5943–5949. PMLR, 12–15 Jul 2023. URL https://proceedings.mlr.press/v195/awasthi23a.html.

Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine Learning*, 79(1-2):151–175, October 2009. doi: 10.1007/s10994-009-5152-4. URL https://doi.org/10.1007/s10994-0o09-5152-4.

Aharon Ben-Tal, Dick den Hertog, Anja De Waegenaere, Bertrand Melenberg, and Gijs Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Manage-*

ment Science, 59(2):341–357, February 2013. doi: 10.1287/mnsc.1120.1641. URL https://doi.org/10.1287/mnsc.1120.1641.

Avrim Blum, Nika Haghtalab, Ariel D Procaccia, and Mingda Qiao. Collaborative PAC learning. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/186a157b2992e7daed3677ce8e9fe40f-Paper.pdf.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities : a nonasymptotic theory of independence*. Oxford University Press, Oxford, United Kingdom, first edition. edition, 2013 - 2013. ISBN 9780199535255.

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412(19):1832–1852, April 2011. doi: 10.1016/j.tcs.2010.12.059. URL https://doi.org/10.1016/j.tcs.2010.12.059.

Yair Carmon and Danielle Hausler. Distributionally robust optimization via ball oracle acceleration. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, page 35866–35879. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/e90b00adc3ba130eb2510d93ba3ff250-Paper-Conference.pdf.

Jiecao Chen, Qin Zhang, and Yuan Zhou. Tight bounds for collaborative PAC learning via multiplicative weights. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/ed519dacc89b2bead3f453b0b05a4a8b-Paper.pdf.

Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, June 2010. doi: 10.1287/opre.1090.0741. URL https://doi.org/10.1287/opre.1090.0741.

Evarist Giné and Richard Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge, revised edition. edition, 2021. ISBN 1-009-02278-4.

Nika Haghtalab, Michael Jordan, and Eric Zhao. On-demand sampling: Learning optimally from multiple distributions. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, page 406–419. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/02917acec264a52a729b99d9bc857909-Paper-Conference.pdf.

Trevor. Hastie, Robert. Tibshirani, and J. H. (Jerome H.) Friedman. *The elements of statistical learning : data mining, inference, and prediction*. Springer series in statistics. Springer, New York, 2nd ed. edition, 2009. ISBN 9780387848570.

Tor Lattimore and Csaba Szepesvari. *Bandit Algorithms*. Cambridge University Press (Virtual Publishing), Cambridge, England, July 2020.

Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. Domain adaptation: Learning bounds and algorithms. In *COLT 2009 - The 22nd Conference on Learning Theory, Montreal, Quebec, Canada, June 18-21, 2009*, 2009. URL http://www.cs.mcgill.ca/%7Ecolt2009/papers/003.pdf#page=1.

Mehryar Mohri, Gary Sivek, and Ananda Theertha Suresh. Agnostic federated learning. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 4615–4625. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/mohri19a.html.

Huy Nguyen and Lydia Zakynthinou. Improved algorithms for collaborative PAC learning. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/3569df159ec477451530c4455b2a9e86-Paper.pdf.

Binghui Peng. The sample complexity of multi-distribution learning. In Shipra Agrawal and Aaron Roth, editors, *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 4185–4204. PMLR, 30 Jun–03 Jul 2024. URL https://proceedings.mlr.press/v247/peng24b.html.

Hamed Rahimian and Sanjay Mehrotra. Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization*, 3:1–85, July 2022. doi: 10.5802/ojmo.15. URL https://doi.org/10.5802/ojmo.15.

Shiori Sagawa*, Pang Wei Koh*, Tatsunori B. Hashimoto, and Percy Liang. Distributionally robust neural networks. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=ryxGuJrFvS.

Herbert Scarf. A min-max solution of an inventory problem. In *Studies in the mathematical theory of inventory and production*, page 201–209. Stanford University Press, Palo Alto, CA, 1958.

Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and theory, Third Edition*. SIAM, August 2021.

Tasuku Soma, Khashayar Gatmiry, and Stefanie Jegelka. Optimal algorithms for group distributionally robust optimization and beyond. *arXiv preprint arXiv:2212.13669*, 2022.

Tobias Sutter, Andreas Krause, and Daniel Kuhn. Robust generalization despite distribution shift via minimum discriminating information. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, page 29754–29767. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/f86890095c957e9b949d11d15f0d0cd5-Paper.pdf.

Ramon van Handel. Probability in High Dimension:. Technical report, Defense Technical Information Center, Fort Belvoir, VA, June 2014. URL http://www.dtic.mil/docs/citations/ADA623999.

Riccardo Volpi, Hongseok Namkoong, Ozan Sener, John C Duchi, Vittorio Murino, and Silvio Savarese. Generalizing to unseen domains via adversarial data augmentation. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/1d94108e907bb8311d8802b48fd54b4a-Paper.pdf.

Martin (Martin J.) Wainwright. *High-dimensional statistics : a non-asymptotic viewpoint*. Cambridge series in statistical and probabilistic mathematics ; 48. Cambridge University Press, Cambridge, United Kingdom, 2019. ISBN 9781108498029.

Jingzhao Zhang, Aditya Krishna Menon, Andreas Veit, Srinadh Bhojanapalli, Sanjiv Kumar, and Suvrit Sra. Coping with label shift via distributionally robust optimisation. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=BtZhsSGNRNi.

Lijun Zhang, Peng Zhao, Zhenhua Zhuang, Tianbao Yang, and Zhi-Hua Zhou. Stochastic approximation approaches to group distributionally robust optimization. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=IcIQbCWoFj.

Zihan Zhang, Wenhao Zhan, Yuxin Chen, Simon S Du, and Jason D Lee. Optimal multi-distribution learning. In Shipra Agrawal and Aaron Roth, editors, *Proceedings of Thirty Seventh Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pages 5220–5223. PMLR, 30 Jun–03 Jul 2024. URL https://proceedings.mlr.press/v247/zhang24b.html.

Appendix A. Expectation of empirical process maximum

Let $\mathcal{U} \subset \mathcal{P}(\mathcal{X})$ be a finite set of distributions over a data space $\mathcal{X}$, with $2 \leq k := |\mathcal{U}| < \infty$. For each distribution $Q \in \mathcal{U}$, we have an associated sample size $n_Q \in \mathbb{N}$ and define $T := \sum_{Q\in\mathcal{U}} n_Q$. When $\mathcal{X} \subset \mathbb{R}$, we additionally denote the variance of each distribution by $\sigma_Q^2 := \mathrm{Var}(Q)$ and define $\sigma_T := \max_{Q\in\mathcal{U}} \frac{\sigma_Q}{\sqrt{n_Q}}$.

In the development that follows, we will work with independent $\mathcal{X}$-valued random variables $(X_Q)_{Q\in\mathcal{U}}, X := \left(X_Q^{(i)}\right)_{Q\in\mathcal{U}, i\in[n_Q]}$, where $X_Q, \left(X_Q^{(i)}\right)_{i\in[n_Q]} \overset{iid}{\sim} Q$ for each $Q \in \mathcal{U}$. For a collection of functions $\{f_Q : \mathcal{X} \to [-1,1]\}_{Q\in\mathcal{U}}$, such that each $f_Q(X_Q)$ is centered, our primary goal will be to bound the following quantity:

$$\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q} f_Q\left(X_Q^{(i)}\right)\right|\right]$$

In particular, we will show the following bounds.

Theorem 11 *Let $\{f_Q : \mathcal{X} \to [-M, M]\}_{Q\in\mathcal{U}}$ be a collection of functions such that $\mathbb{E}[f_Q(X_Q)] = 0$ for each $Q \in \mathcal{U}$. Then,*

$$\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q} f_Q\left(X_Q^{(i)}\right)\right|\right] \leq 4M\sqrt{\frac{\log k}{\min_{Q\in\mathcal{U}} n_Q}}$$

Moreover, if $\mathcal{X} \subset \mathbb{R}$ and each function f_Q is L-Lipschitz, then

$$\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q} f_Q\left(X_Q^{(i)}\right)\right|\right] \leq \frac{16M\log k}{\min_{Q\in\mathcal{U}} n_Q} + 4L\sigma_T\sqrt{2\log k}$$

We note that the first bound can be directly obtained by a high-probability bound via Hoeffding's inequality, along with a union bound, and a subsequent integration of the tails. The second bound (Theorem 18) requires a more careful analysis and, in the process of deriving it, we additionally show the first result (Corollary 15).

The proof will follow in two parts: first, in Section A.1, we use symmetrization to bound the quantity of interest with a notion of Rademacher complexity, and subsequently derive bounds on this complexity in Section A.2.

A.1. Symmetrization

A standard approach to bound empirical process maxima is via symmetrization. We begin by defining the Rademacher complexity variant of a class of functions $\{h_Q : \mathcal{X} \to \mathbb{R}\}_{Q\in\mathcal{U}}$:

$$\mathfrak{R}_T\left(\{h_Q\}_{Q\in\mathcal{U}}\right) := \mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q} \epsilon_i h_Q\left(X_Q^{(i)}\right)\right|\right]$$

where $\epsilon_1, \ldots, \epsilon_{\max_{Q\in\mathcal{U}} n_Q} \overset{iid}{\sim} \text{Rad}$ (i.e., they are each uniform on $\{-1, 1\}$) are independent from X. Note that we place no assumptions on $h_Q(X_Q)$ being centered. The following result employs the standard symmetrization trick (see, e.g., (Wainwright, 2019, Theorem 4.10)) with different sample sizes, and we prove it here for completeness.

Theorem 12 (Symmetrization) *For any collection of functions $\{h_Q : \mathcal{X} \to \mathbb{R}\}_{Q\in\mathcal{U}}$, we have that*

$$\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q}\left\{h_Q\left(X_Q^{(i)}\right) - \mathbb{E}[h_Q(X_Q)]\right\}\right|\right] \leq 2\,\mathfrak{R}_T\left(\{h_Q\}_{Q\in\mathcal{U}}\right)$$

Proof Let $Y := \left(Y_Q^{(i)}\right)_{Q\in\mathcal{U}, i\in[n_Q]}$ be an independent copy of X and let P denote their common distribution. In addition, define Rademacher variables $\epsilon^n \overset{iid}{\sim} \text{Rad}$ that are independent from X and

Y. Then,

$$\begin{aligned}
&\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q}\left\{h_Q\left(X_Q^{(i)}\right)-\mathbb{E}\left[h_Q\left(X_Q\right)\right]\right\}\right|\right]\\
&\quad=\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q}\left\{h_Q\left(X_Q^{(i)}\right)-\mathbb{E}\left[h_Q\left(Y_Q^{(i)}\right)\right]\right\}\right|\right]\\
&\quad=\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\mathbb{E}\left[\frac{1}{n_Q}\sum_{i=1}^{n_Q}\left\{h_Q\left(X_Q^{(i)}\right)-h_Q\left(Y_Q^{(i)}\right)\right\}\middle|X\right]\right|\right]\\
&\quad\overset{(1)}{\leq}\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q}\left[h_Q\left(X_Q^{(i)}\right)-h_Q\left(Y_Q^{(i)}\right)\right]\right|\right]\\
&\quad=\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q}\epsilon_i\left[h_Q\left(X_Q^{(i)}\right)-h_Q\left(Y_Q^{(i)}\right)\right]\right|\right]\\
&\quad\leq\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left\{\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q}\epsilon_i h_Q\left(X_Q^{(i)}\right)\right|+\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q}\epsilon_i h_Q\left(Y_Q^{(i)}\right)\right|\right\}\right]\\
&\quad\leq\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left\{\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q}\epsilon_i h_Q\left(X_Q^{(i)}\right)\right|\right\}+\max_{Q\in\mathcal{U}}\left\{\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q}\epsilon_i h_Q\left(Y_Q^{(i)}\right)\right|\right\}\right]\\
&\quad=2\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q}\epsilon_i h_Q\left(X_Q^{(i)}\right)\right|\right]\\
&\quad=2\,\mathfrak{R}_T\left(\{h_Q\}_{Q\in\mathcal{U}}\right)
\end{aligned}$$

where inequality (1) follows from Jensen. ■

A.2. Bounds on the Rademacher complexity

For the symmetrization trick to be useful, we need to bound $\mathfrak{R}_T\left(\{h_Q\}_{Q\in\mathcal{U}}\right)$. To this end, we begin by defining the Rademacher complexity of a set $\Theta\subset\mathbb{R}^n$:

$$\hat{\mathfrak{R}}\left(\Theta\right):=\mathbb{E}\left[\sup_{\theta\in\Theta}\left|\langle\epsilon^n,\theta\rangle\right|\right]$$

where $\epsilon^n=(\epsilon_1,\ldots,\epsilon_n)\overset{iid}{\sim}\text{Rad}$. The process $\{\langle\epsilon^n,\theta\rangle\}_{\theta\in\Theta}$ is sub-Gaussian and, for finite Θ, the Rademacher complexity admits a particularly simple bound, shown next. For a deeper dive into the field, see, e.g., (Wainwright, 2019, Chapter 5).

Lemma 13 (Bounding the Rademacher complexity of a finite set) *Let $\Theta\subset\mathbb{R}^n$ satisfy $2\leq|\Theta|<\infty$. Then,*

$$\hat{\mathfrak{R}}\left(\Theta\right)\leq 2D_\Theta\sqrt{\log|\Theta|}$$

where $D_\Theta:=\max_{\theta\in\Theta}\|\theta\|_2$.

Proof Note that since each ϵ_i is 1-sub-Gaussian,

$$\mathbb{E}\left[e^{\lambda\langle\epsilon^n,\theta\rangle}\right] = \prod_{i=1}^{n}\mathbb{E}\left[e^{\lambda\epsilon_i\theta_i}\right] \leq \prod_{i=1}^{n} e^{\frac{\lambda^2\theta_i^2}{2}} = e^{\frac{\lambda^2\|\theta\|_2^2}{2}} \leq e^{\frac{\lambda^2 D_\Theta^2}{2}}$$

for any $\theta \in \Theta$ and $\lambda \in \mathbb{R}$. That is, $\langle\epsilon^n,\theta\rangle$ is a centered D_Θ-sub-Gaussian variable and we can, thus, apply the standard maximal inequality (e.g., (Boucheron et al., 2013 - 2013, Theorem 2.5)) to obtain the claim. ∎

We can relate both notions of Rademacher complexity introduced thus far to conclude the following result.

Corollary 14 *For a collection of functions $\{h_Q : \mathcal{X} \to \mathbb{R}\}_{Q\in\mathcal{U}}$, define the random variable*

$$D\left(\{h_Q\}_{Q\in\mathcal{U}}\right) := \max_{Q\in\mathcal{U}} \sqrt{\sum_{i=1}^{n_Q}\left(\frac{h_Q\left(X_Q^{(i)}\right)}{n_Q}\right)^2}$$

Then, we have that

$$\mathfrak{R}_T\left(\{h_Q\}_{Q\in\mathcal{U}}\right) \leq 2\sqrt{\log k}\,\mathbb{E}\left[D\left(\{h_Q\}_{Q\in\mathcal{U}}\right)\right]$$

Proof Fix $\mathbf{x} := \left(x_Q^i\right)_{Q\in\mathcal{U},i\in[n_Q]} \in \mathcal{X}^T$. Let $n := \max_{Q\in\mathcal{U}} n_Q$ and define vectors $\theta_Q^{\mathbf{x}} \in \mathbb{R}^n$ by

$$\left[\theta_Q^{\mathbf{x}}\right]_i := \begin{cases} \frac{h_Q\left(x_Q^i\right)}{n_Q} & i \leq n_Q \\ 0 & \text{otherwise} \end{cases} \qquad \forall i \in [n]\,, Q \in \mathcal{U}$$

and define the set of all such vectors $\Theta^{\mathbf{x}} := \left\{\theta_Q^{\mathbf{x}} : Q \in \mathcal{U}\right\}$, so that $|\Theta^{\mathbf{x}}| = k \geq 2$. Then, note that

$$\hat{\mathfrak{R}}\left(\Theta^{\mathbf{x}}\right) = \mathbb{E}\left[\max_{\theta\in\Theta^{\mathbf{x}}}|\langle\epsilon^n,\theta\rangle|\right] = \mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q}\epsilon_i h_Q\left(x_Q^i\right)\right|\right]$$

Moreover, since $D_{\Theta^{\mathbf{x}}} = \max_{Q\in\mathcal{U}}\sqrt{\sum_{i=1}^{n_Q}\left(\frac{h_Q\left(x_Q^i\right)}{n_Q}\right)^2}$, Lemma 13 yields

$$\mathfrak{R}_T\left(\{h_Q\}_{Q\in\mathcal{U}}\right) = \mathbb{E}\left[\hat{\mathfrak{R}}\left(\Theta^X\right)\right] \leq \mathbb{E}\left[2D_{\Theta^X}\sqrt{\log|\Theta^X|}\right] = 2\sqrt{\log k}\,\mathbb{E}\left[D\left(\{h_Q\}_{Q\in\mathcal{U}}\right)\right]$$

∎

We can then readily obtain the first inequality of interest after scaling both sides of the bound below by M.

Corollary 15 *Let* $\{f_Q : \mathcal{X} \to [-1,1]\}_{Q\in\mathcal{U}}$ *be a collection of functions such that* $\mathbb{E}[f_Q(X_Q)] = 0$ *for each* $Q \in \mathcal{U}$. *Then,*

$$\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q} f_Q\left(X_Q^{(i)}\right)\right|\right] \le 4\sqrt{\frac{\log k}{\min_{Q\in\mathcal{U}} n_Q}}$$

Proof Since each $f_Q \in [-1,1]$, we have that

$$D\left(\{f_Q\}_{Q\in\mathcal{U}}\right) \le \sqrt{\max_{Q\in\mathcal{U}} \frac{1}{n_Q}} = \sqrt{\frac{1}{\min_{Q\in\mathcal{U}} n_Q}}$$

Hence, combining Theorem 12 and Corollary 14 yields

$$\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q} f_Q\left(X_Q^{(i)}\right)\right|\right] \le 2\,\mathfrak{R}_T\left(\{f_Q\}_{Q\in\mathcal{U}}\right) \le 4\sqrt{\log k}\,\mathbb{E}\left[D\left(\{f_Q\}_{Q\in\mathcal{U}}\right)\right] \le 4\sqrt{\frac{\log k}{\min_{Q\in\mathcal{U}} n_Q}}$$

■

To obtain the second bound, we require a more refined analysis. We begin by introducing two simple auxiliary lemmas.

Lemma 16 *Let* $b, c > 0$ *and suppose that* $x^2 \le bx + c$. *Then,* $x \le b + \sqrt{c}$.

Proof Define quadratic $p(z) := z^2 - bz - c$, so that $p(x) \le 0$. Since $p(0) = -c < 0$, consider its roots $r_1 < 0 < r_2$. Then, p is positive on (r_2, ∞) and, thus,

$$x \le r_2 = \frac{b + \sqrt{b^2 + 4c}}{2} \le b + \sqrt{c}$$

■

Lemma 17 (Variance of Lipschitz functions) *Let* $Z \in \mathcal{Z} \subset \mathbb{R}$ *be a random variable, and suppose that* $f : \mathcal{Z} \to \mathbb{R}$ *is L-Lipschitz. Then,*

$$\operatorname{Var}(f(Z)) \le 2L^2 \operatorname{Var}(Z)$$

Proof Let Z' be an independent copy of Z. Then,

$$\begin{aligned}
\operatorname{Var}(f(Z)) &= \mathbb{E}\left[\left(f(Z) - \mathbb{E}\left[f(Z')\right]\right)^2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[f(Z) - f(Z') \middle| Z\right]^2\right] \\
&\le \mathbb{E}\left[\left(f(Z) - f(Z')\right)^2\right] && \text{Jensen's} \\
&\le L^2\mathbb{E}\left[\left(Z - Z'\right)^2\right] && \text{Lipschitzness} \\
&= 2L^2\left\{\operatorname{Var}(Z) + \mathbb{E}\left[\left(Z - \mathbb{E}[Z]\right)\left(\mathbb{E}[Z] - Z'\right)\right]\right\} && Z \overset{(d)}{=} Z' \\
&= 2L^2 \operatorname{Var}(Z) && Z \perp\!\!\!\perp Z'
\end{aligned}$$

■

Borrowing ideas from (Giné and Nickl, 2021, Corollary 3.5.7), we then conclude the second target bound.

Theorem 18 *Suppose that $\mathcal{X} \subset \mathbb{R}$. Let $\{f_Q : \mathcal{X} \to [-M, M]\}_{Q\in\mathcal{U}}$ be a collection of functions such that $\mathbb{E}[f_Q(X_Q)] = 0$ and f_Q is L-Lipschitz for each $Q \in \mathcal{U}$. Then,*

$$\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q} f_Q\left(X_Q^{(i)}\right)\right|\right] \leq \frac{16M\log k}{\min_{Q\in\mathcal{U}} n_Q} + 4L\sigma_T\sqrt{2\log k}$$

Proof We begin with the following observation: from Jensen's, we obtain

$$C := \sqrt{\log k}\mathbb{E}\left[D\left(\{f_Q\}_{Q\in\mathcal{U}}\right)\right] \leq \sqrt{(\log k)\,\mathbb{E}\left[D\left(\{f_Q\}_{Q\in\mathcal{U}}\right)^2\right]}$$

Next, we bound the expectation on the right-hand side:

$$\begin{aligned}
\mathbb{E}\left[D\left(\{f_Q\}_{Q\in\mathcal{U}}\right)^2\right] &= \mathbb{E}\left[\max_{Q\in\mathcal{U}}\sum_{i=1}^{n_Q}\left(\frac{f_Q\left(X_Q^{(i)}\right)}{n_Q}\right)^2\right] \\
&= \mathbb{E}\left[\max_{Q\in\mathcal{U}}\sum_{i=1}^{n_Q}\left\{\left(\frac{f_Q\left(X_Q^{(i)}\right)}{n_Q}\right)^2 - \mathbb{E}\left[\left(\frac{f_Q(X_Q)}{n_Q}\right)^2\right] + \mathbb{E}\left[\left(\frac{f_Q(X_Q)}{n_Q}\right)^2\right]\right\}\right] \\
&\leq \underbrace{\max_{Q\in\mathcal{U}}\left\{\frac{\mathbb{E}\left[f_Q^2(X_Q)\right]}{n_Q}\right\}}_{=:(*_1)} + \underbrace{\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q}\left\{\frac{f_Q^2\left(X_Q^{(i)}\right)}{n_Q} - \mathbb{E}\left[\frac{f_Q^2(X_Q)}{n_Q}\right]\right\}\right|\right]}_{=:(*_2)}
\end{aligned}$$

From Lemma 17 and the fact that $\mathbb{E}[f_Q(X_Q)] = 0$, we know that

$$(*_1) = \max_{Q\in\mathcal{U}}\frac{\mathrm{Var}(f_Q(X_Q))}{n_Q} \leq 2L^2\max_{Q\in\mathcal{U}}\frac{\sigma_Q^2}{n_Q} = 2L^2\sigma_T^2$$

As for $(*_2)$, we can apply Theorem 12 on functions $h_Q(x) := \frac{f_Q^2(x)}{n_Q}$ to conclude that

$$\begin{aligned}
(*_2) &\leq 2\,\mathfrak{R}_T\left(\{h_Q\}_{Q\in\mathcal{U}}\right) && \text{Thm. 12}\\
&\leq 4\sqrt{\log k}\,\mathbb{E}\left[D\left(\{h_Q\}_{Q\in\mathcal{U}}\right)\right] && \text{Cor. 14}\\
&= 4\sqrt{\log k}\,\mathbb{E}\left[\max_{Q\in\mathcal{U}}\sqrt{\sum_{i=1}^{n_Q}\left(\frac{f_Q\left(X_Q^{(i)}\right)}{n_Q}\right)^4}\right]\\
&\leq 4M\sqrt{\log k}\,\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left\{\frac{1}{n_Q}\sqrt{\sum_{i=1}^{n_Q}\left(\frac{f_Q\left(X_Q^{(i)}\right)}{n_Q}\right)^2}\right\}\right] && f_Q^4 = M^4\left(\frac{f_Q}{M}\right)^4 \leq M^2 f_Q^2\\
&\leq 4M\sqrt{\log k}\max_{Q\in\mathcal{U}}\left\{\frac{1}{n_Q}\right\}\mathbb{E}\left[D\left(\{f_Q\}_{Q\in\mathcal{U}}\right)\right]\\
&= \frac{4M}{\min_{Q\in\mathcal{U}} n_Q}C
\end{aligned}$$

In other words, we have that

$$C^2 \leq (\log k)\,\mathbb{E}\left[D\left(\{f_Q\}_{Q\in\mathcal{U}}\right)^2\right] \leq \frac{4M\log k}{\min_{Q\in\mathcal{U}} n_Q}C + 2L^2\sigma_T^2\log k$$

Then, Lemma 16 implies that

$$C \leq \frac{4M\log k}{\min_{Q\in\mathcal{U}} n_Q} + L\sigma_T\sqrt{2\log k}$$

Combining this with Theorem 12 and Corollary 14, we conclude that

$$\mathbb{E}\left[\max_{Q\in\mathcal{U}}\left|\frac{1}{n_Q}\sum_{i=1}^{n_Q} f_Q\left(X_Q^{(i)}\right)\right|\right] \leq 2\,\mathfrak{R}_T\left(\{f_Q\}_{Q\in\mathcal{U}}\right) \leq 4C \leq \frac{16M\log k}{\min_{Q\in\mathcal{U}} n_Q} + 4L\sigma_T\sqrt{2\log k}$$

■

Appendix B. Empirical process concentration inequalities

Again, suppose that $\mathcal{U} \subset \mathcal{P}(\mathcal{X})$ is a collection of k distributions, and define independent variables $X := \left(X_Q^{(i)}\right)_{Q\in\mathcal{U}, i\in[n_Q]}$, where $n_Q \in \mathbb{N}$ and $\left(X_Q^{(i)}\right)_{i\in[n_Q]} \overset{iid}{\sim} Q$ for each $Q \in \mathcal{U}$. Our object of interest in this section is the random variable

$$Z_f := \min_{Q\in\mathcal{U}} \frac{1}{n_Q}\sum_{i=1}^{n_Q} f\left(X_Q^{(i)}\right)$$

for a function $f : \mathcal{X} \to \mathbb{R}$. As will become clear later, our primary goal will be to obtain concentration inequalities on $Z_{f,g} := Z_f - Z_g$.

B.1. McDiarmid

To obtain the UE regret bound, we will apply a very simple concentration inequality, called McDiarmid's inequality (e.g., see (Boucheron et al., 2013 - 2013, Theorem 6.2)). Here, we specialize to

$$Z_f = \min_{Q\in\mathcal{U}} \frac{1}{n}\sum_{i=1}^{n} f\left(X_Q^{(i)}\right)$$

Let us define the function $\Phi_f : \left(\mathcal{X}^k\right)^n \to [0,1]$ by $\Phi_f\left(\mathbf{x}_1,\ldots,\mathbf{x}_n\right) := \min_{Q\in\mathcal{U}} \frac{1}{n}\sum_{i=1}^{n} f\left(x_Q^i\right)$, where each $\mathbf{x}_i = \left(x_Q^i\right)_{Q\in\mathcal{U}} \in \mathcal{X}^k$. Then, we can write $Z_f = \Phi_f(X)$, where we view X as n vectors of dimension k. Next, we show that Φ_f satisfies the bounded differences property when f is bounded.

Proposition 19 (Bounded differences) *Suppose that $f : \mathcal{X} \to [0,1]$. Then,*

$$\max_{i\in[n]} \sup_{\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{y}\in\mathcal{X}^k} \left|\Phi_f\left(\mathbf{x}_1,\ldots,\mathbf{x}_n\right) - \Phi_f\left(\mathbf{x}_1,\ldots,\mathbf{x}_{i-1},\mathbf{y},\mathbf{x}_{i+1},\ldots,\mathbf{x}_n\right)\right| \le \frac{1}{n}$$

Proof Let us begin with a simple observation: for real-valued functions $g, h : \mathcal{Z} \to \mathbb{R}$, where $\mathcal{Z}$ is any domain, we have that

$$\inf_{z'\in\mathcal{Z}} g\left(z'\right) - \inf_{z\in\mathcal{Z}} h(z) = \sup_{z\in\mathcal{Z}}\left\{\inf_{z'\in\mathcal{Z}} g\left(z'\right) - h(z)\right\} \le \sup_{z\in\mathcal{Z}}\left\{g(z) - h(z)\right\} \le \sup_{z\in\mathcal{Z}}\left|g(z) - h(z)\right|$$

By symmetry, it then follows that $\left|\inf_{z'\in\mathcal{Z}} g\left(z'\right) - \inf_{z\in\mathcal{Z}} h(z)\right| \le \sup_{z\in\mathcal{Z}}\left|g(z) - h(z)\right|$. Next, fix any index $i \in [n]$ and inputs $\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{y} := \left(y_Q\right)_{Q\in\mathcal{U}} \in \mathcal{X}^k$, and define vectors $\mathbf{x} := \left(\mathbf{x}_1,\ldots,\mathbf{x}_n\right)$ and $\mathbf{x}' := \left(\mathbf{x}_1,\ldots,\mathbf{x}_{i-1},\mathbf{y},\mathbf{x}_{i+1},\ldots,\mathbf{x}_n\right)$. Then, from our initial observation, we know that

$$\begin{aligned}
\left|\Phi_f(\mathbf{x}) - \Phi_f\left(\mathbf{x}'\right)\right| &= \frac{1}{n}\left|\min_{Q'\in\mathcal{U}}\left\{\sum_{j=1}^{n} f\left(x_{Q'}^j\right)\right\} - \min_{Q\in\mathcal{U}}\left\{f\left(y_Q\right) + \sum_{j\in[n]:j\neq i} f\left(x_Q^j\right)\right\}\right| \\
&\le \frac{1}{n}\max_{Q\in\mathcal{U}}\left|\sum_{j=1}^{n} f\left(x_Q^j\right) - \left[f\left(y_Q\right) + \sum_{j\in[n]:j\neq i} f\left(x_Q^j\right)\right]\right| \\
&\le \frac{1}{n}\max_{Q\in\mathcal{U}}\left|f\left(x_Q^i\right) - f\left(y_Q\right)\right| \\
&\le \frac{1}{n}
\end{aligned}$$

■

When the inequality in Proposition 19 holds, we say that Φ_f satisfies the *bounded differences property* with constant parameter $\frac{1}{n}$. This immediately implies the next claim.

Corollary 20 *For any two functions $f, g : \mathcal{X} \to [0,1]$, the function $\Phi_f - \Phi_g$ satisfies the bounded differences property with constant parameter $\frac{2}{n}$.*

Proof Using the same variables $\mathbf{x}$ and $\mathbf{x}'$ as in the proof of Proposition 19, we obtain

$$\left|[\Phi_f(\mathbf{x}) - \Phi_g(\mathbf{x})] - \left[\Phi_f\left(\mathbf{x}'\right) - \Phi_g\left(\mathbf{x}'\right)\right]\right| \leq \left|\Phi_f(\mathbf{x}) - \Phi_f\left(\mathbf{x}'\right)\right| + \left|\Phi_g(\mathbf{x}) - \Phi_g\left(\mathbf{x}'\right)\right| \leq \frac{2}{n}$$

■

Via McDiarmid's, this property then directly yields the following concentration result.

Corollary 21 *Let* $f, g : \mathcal{X} \to [0,1]$. *Then,*

$$\mathbb{P}\left(Z_{f,g} - \mathbb{E}\left[Z_{f,g}\right] \geq t\right) \leq \exp\left(-\frac{nt^2}{2}\right) \quad \forall t \geq 0$$

Proof Since $Z_{f,g} = \left(\Phi_f - \Phi_g\right)(X)$ and X has independent components, we simply apply Corollary 20 and McDiarmid's. ■

B.2. Bernstein

In contrast to McDiarmid's inequality, our next goal is to derive a more involved bound that additionally scales with the variance. To this end, we sort the sample sizes: $0 =: n_{(0)} \leq n_{(1)} \leq \cdots \leq n_{(k)}$ and let $Q_{(j)} \in \mathcal{U}$ be such that $n_{Q_{(j)}} = n_{(j)}$. Our analysis then relies on the following:

$$V_T := \sum_{j=1}^{k} \left(n_{(j)} - n_{(j-1)}\right) \mathbb{E}\left[\max_{r \in \{j,\ldots,k\}} \frac{1}{n_{(r)}^2}\left[X_{Q_{(r)}} - \mu_{Q_{(r)}}\right]^2\right]$$

$$\Sigma_T^2 := \mathbb{E}\left[\max_{Q \in \mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \left(X_Q^{(i)} - \mu_Q\right)^2\right]$$

$$\sigma_T^2 := \max_{Q \in \mathcal{U}} \frac{\sigma_Q^2}{n_Q}$$

Theorem 22 *Suppose that* $\mathcal{X} \subset \mathbb{R}$ *and* $f, g : \mathcal{X} \to [0, M]$ *are L-Lipschitz. Then,*

$$\mathbb{P}\left(Z_{f,g} - \mathbb{E}\left[Z_{f,g}\right] \geq t\right) \leq \exp\left(-\frac{t^2}{16L^2\left(2\sigma_T^2 + \Sigma_T^2 + 6V_T\right) + \frac{2\sqrt{6}Mt}{\min_{Q \in \mathcal{U}} n_Q}}\right) \quad \forall t \geq 0$$

B.2.1. PRELIMINARIES

To prove Theorem 22, we must first state some standard results and definitions from the theory of concentration of measure. We do not prove most results stated, and refer to Boucheron et al. (2013 - 2013) for further reference.

We say that a random variable $X \in \mathbb{R}$ is *sub-gamma on the right tail* with parameters $\nu, c > 0$ if

$$\log \mathbb{E}\left[e^{\lambda(X - \mathbb{E}[X])}\right] \leq \frac{\nu^2\lambda^2}{2(1 - c\lambda)} \quad \forall \lambda \in \left[0, \frac{1}{c}\right)$$

We denote the class of such variables by $\Gamma_+(\nu, c)$. Due to the decaying tail, we get the following concentration bound.

Proposition 23 (Sub-gamma concentration) *Let $X \in \Gamma_+(\nu, c)$. Then,*

$$\mathbb{P}\left(X - \mathbb{E}\left[X\right] \geq t\right) \leq \exp\left(-\frac{t^2}{2\left(\nu^2 + ct\right)}\right) \quad \forall t \geq 0$$

Proof See (Boucheron et al., 2013 - 2013, Section 2.4). ■

Next, we introduce the notion of self-bounding functions: we say that a nonnegative function $f : \mathcal{X}^n \to \mathbb{R}_+$ has the *self-bounding property* if there exists functions $\left\{f_i : \mathcal{X}^{n-1} \to \mathbb{R}\right\}_{i\in[n]}$ such that

$$f\left(\mathbf{x}\right) - f_i\left(\mathbf{x}_{\setminus i}\right) \in [0, 1] \quad \text{and} \quad \sum_{i=1}^{n}\left[f\left(\mathbf{x}\right) - f_i\left(\mathbf{x}_{\setminus i}\right)\right] \leq f\left(\mathbf{x}\right)$$

for all $i \in [n]$ and $\mathbf{x} \in \mathcal{X}^n$, where we define $\mathbf{x}_{\setminus i} \coloneqq (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$. A simple observation about such functions is that they are closed under convex combinations.

Lemma 24 (Convex combination of self-bounding functions) *Suppose that f and g satisfy the self-bounding property and let $\alpha \in [0, 1]$. Then, $\alpha f + (1 - \alpha) g$ also satisfies the self-bounding property.*

Proof Let $\{f_i\}$ and $\{g_i\}$ be the functions satisfying the self-bounding property, and define $h \coloneqq \alpha f + (1 - \alpha) g$ and $h_i \coloneqq \alpha f_i + (1 - \alpha) g_i$. Then, for any $i \in [n]$ and $\mathbf{x} \in \mathcal{X}^n$,

$$h\left(\mathbf{x}\right) - h_i\left(\mathbf{x}_{\setminus i}\right) = \alpha\left[f\left(\mathbf{x}\right) - f_i\left(\mathbf{x}_{\setminus i}\right)\right] + (1 - \alpha)\left[g\left(\mathbf{x}\right) - g_i\left(\mathbf{x}_{\setminus i}\right)\right] \in [0, 1]$$

and

$$\begin{aligned}\sum_{i=1}^{n}\left[h\left(\mathbf{x}\right) - h_i\left(\mathbf{x}_{\setminus i}\right)\right] &= \alpha\sum_{i=1}^{n}\left[f\left(\mathbf{x}\right) - f_i\left(\mathbf{x}_{\setminus i}\right)\right] + (1 - \alpha)\sum_{i=1}^{n}\left[g\left(\mathbf{x}\right) - g_i\left(\mathbf{x}_{\setminus i}\right)\right] \\ &\leq \alpha f\left(\mathbf{x}\right) + (1 - \alpha) g\left(\mathbf{x}\right) \\ &= h\left(\mathbf{x}\right)\end{aligned}$$

■

The reason for introducing such functions is that they possess a favorable bound on their cumulant-generating function (cgf).

Proposition 25 (Cgf of self-bounding functions) *Suppose that $f : \mathcal{X}^n \to \mathbb{R}_+$ has the self-bounding property and let $X^n = (X_1, \ldots, X_n)$ be independent random variables. Then,*

$$\log\mathbb{E}\left[e^{\lambda f(X^n)}\right] \leq \left(e^{\lambda} - 1\right)\mathbb{E}\left[f\left(X^n\right)\right] \quad \forall \lambda \in \mathbb{R}$$

Proof See (Boucheron et al., 2013 - 2013, Theorem 6.12). ■

The last tool we need employs symmetrization once again. For the next result and the development that follows, we omit the parentheses in $a_+^2 \coloneqq (a_+)^2$; that is, we take the positive part before squaring.

Proposition 26 (Exponential Efron-Stein) *Suppose that $X^n = (X_1, \dots, X_n)$ are independent random variables and let $W^n = (W_1, \dots, W_n)$ be independent copies of them. Given a nonnegative function $f : \mathcal{X}^n \to \mathbb{R}_+$, define variables $Z := f(X^n)$ and its symmetrized counterpart*

$$Z_i' := f(X_1, \dots, X_{i-1}, W_i, X_{i+1}, \dots, X_n) \quad \forall i \in [n]$$

Additionally, let

$$V^+ := \sum_{i=1}^{n} \mathbb{E}\left[(Z - Z_i')_+^2 \middle| X^n \right]$$

Then, we have that

$$\log \mathbb{E}\left[e^{\lambda(Z - \mathbb{E}[Z])} \right] \leq \frac{\theta\lambda}{1 - \theta\lambda} \log \mathbb{E}\left[e^{\frac{\lambda V^+}{\theta}} \right]$$

for any $\theta, \lambda > 0$ such that $\theta\lambda < 1$.

Proof See (Boucheron et al., 2013 - 2013, Theorem 6.16). ∎

Proof [Proof of Theorem 22] To conclude our main result, we begin with a more general setup: let $X := \left(X_Q^{(i)}\right)_{Q \in \mathcal{U}, i \in [n]}$, where $n \in \mathbb{N}$, be a collection of independent $\mathcal{X}$-valued random variables, and let $X^{(i)} := \left(X_Q^{(i)}\right)_{Q \in \mathcal{U}}$ for each $i \in [n]$. We de not impose any assumptions on their distributions. Our random variables of interest will be

$$Z_f := \min_{Q \in \mathcal{U}} \sum_{i=1}^{n} f_Q\left(X_Q^{(i)}\right) \quad \text{and} \quad Z_{f,g} := Z_f - Z_g$$

for collections of functions $f = \left\{ f_Q : \mathcal{X} \to \left[0, \frac{b}{\sqrt{6}}\right] \right\}_{Q \in \mathcal{U}}$ and $g = \left\{ g_Q : \mathcal{X} \to \left[0, \frac{b}{\sqrt{6}}\right] \right\}_{Q \in \mathcal{U}}$, where $b > 0$. Define

$$\mu_{f,i,Q} := \mathbb{E}\left[f_Q\left(X_Q^{(i)}\right) \right] \quad \text{and} \quad \sigma_{f,i,Q}^2 := \mathrm{Var}\left(f_Q\left(X_Q^{(i)}\right) \right)$$

Similarly, consider the variance variants:

$$V_f := \sum_{i=1}^{n} \mathbb{E}\left[\max_{Q \in \mathcal{U}} \left[f_Q\left(X_Q^{(i)}\right) - \mu_{f,i,Q} \right]^2 \right]$$

$$\Sigma_f^2 := \mathbb{E}\left[\max_{Q \in \mathcal{U}} \sum_{i=1}^{n} \left[f_Q\left(X_Q^{(i)}\right) - \mu_{f,i,Q} \right]^2 \right]$$

$$\sigma_f^2 := \max_{Q \in \mathcal{U}} \sum_{i=1}^{n} \sigma_{f,i,Q}^2$$

Following the analysis of (Boucheron et al., 2013 - 2013, Theorem 12.2), we will use the tools provided and proceed in 5 steps:

1. Upper bound V^+.
2. Apply exponential Efron-Stein along with the bound on V^+.
3. Show the self-boundedness of certain functions and apply the cgf bound.
4. Show that $Z_{f,g}$ is sub-gamma and apply the tail bound.
5. Specialize the analysis to the original setting.

B.2.2. BOUNDING V^+

For each pair $(i, Q) \in [n] \times \mathcal{U}$, let $W_Q^{(i)}$ be an independent copy of $X_Q^{(i)}$ and define $W^{(i)} := \left(W_Q^{(i)}\right)_{Q\in\mathcal{U}}$. Moreover, define

$$Y_i := \left(X^{(1)}, \ldots, X^{(i-1)}, W^{(i)}, X^{(i+1)}, \ldots, X^{(n)}\right) \quad \forall i \in [n]$$

and function $\Phi_{f,g} : \left(\mathcal{X}^k\right)^n \to \mathbb{R}$ by

$$\Phi_{f,g}\left(\mathbf{x}_1, \ldots, \mathbf{x}_n\right) := \min_{Q\in\mathcal{U}} \sum_{i=1}^{n} f_Q\left(x_Q^i\right) - \min_{Q'\in\mathcal{U}} \sum_{i=1}^{n} g_{Q'}\left(x_{Q'}^i\right)$$

where $\mathbf{x}_i = \left(x_Q^i\right)_{Q\in\mathcal{U}} \in \mathcal{X}^k$ for each $i \in [n]$. In what follows, we will use the more compact notation $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. Note that $Z_{f,g} = \Phi_{f,g}(X)$ and

$$\begin{aligned} Z_i' &:= \Phi_{f,g}\left(Y_i\right) \\ &= \min_{Q\in\mathcal{U}} \left\{ f_Q\left(W_Q^{(i)}\right) + \sum_{j\in[n]:j\neq i} f_Q\left(X_Q^{(j)}\right) \right\} - \min_{Q'\in\mathcal{U}} \left\{ g_{Q'}\left(W_{Q'}^{(i)}\right) + \sum_{j\in[n]:j\neq i} g_{Q'}\left(X_{Q'}^{(j)}\right) \right\} \end{aligned}$$

Given functions $h = \{h_Q : \mathcal{X} \to \mathbb{R}\}_{Q\in\mathcal{U}}$, define minimizer $\hat{Q}_h : \left(\mathcal{X}^k\right)^n \to \mathcal{U}$ by

$$\hat{Q}_h\left(\mathbf{x}\right) := \operatorname*{argmin}_{Q\in\mathcal{U}} \sum_{i=1}^{n} h_Q\left(x_Q^i\right)$$

so that

$$\Phi_{f,g}\left(\mathbf{x}\right) = \sum_{i=1}^{n} f_{\hat{Q}_f(\mathbf{x})}\left(x_{\hat{Q}_f(\mathbf{x})}^i\right) - \sum_{i=1}^{n} g_{\hat{Q}_g(\mathbf{x})}\left(x_{\hat{Q}_g(\mathbf{x})}^i\right)$$

and

$$\sum_{i=1}^{n} f_{\hat{Q}_f(\mathbf{x})}\left(x_{\hat{Q}_f(\mathbf{x})}^i\right) - \sum_{i=1}^{n} g_Q\left(x_Q^i\right) \leq \Phi_{f,g}\left(\mathbf{x}\right) \leq \sum_{i=1}^{n} f_{Q'}\left(x_{Q'}^i\right) - \sum_{i=1}^{n} g_{\hat{Q}_g(\mathbf{x})}\left(x_{\hat{Q}_g(\mathbf{x})}^i\right)$$

for any $\mathbf{x} \in \left(\mathcal{X}^k\right)^n$ and $Q, Q' \in \mathcal{U}$. Choosing $Q = \hat{Q}_g(X)$ and $Q' = \hat{Q}_f(Y_i)$ below then yields

$$\begin{aligned}
Z_{f,g} - Z_i' &= \Phi_{f,g}(X) - \Phi_{f,g}(Y_i) \\
&\le \sum_{j=1}^n f_{\hat{Q}_f(Y_i)}\left(X^{(j)}_{\hat{Q}_f(Y_i)}\right) - \sum_{j=1}^n g_{\hat{Q}_g(X)}\left(X^{(j)}_{\hat{Q}_g(X)}\right) \\
&\qquad - \left[f_{\hat{Q}_f(Y_i)}\left(W^{(i)}_{\hat{Q}_f(Y_i)}\right) + \sum_{j\in[n]:j\neq i} f_{\hat{Q}_f(Y_i)}\left(X^{(j)}_{\hat{Q}_f(Y_i)}\right)\right] \\
&\qquad\qquad + \left[g_{\hat{Q}_g(X)}\left(W^{(i)}_{\hat{Q}_g(X)}\right) + \sum_{j\in[n]:j\neq i} g_{\hat{Q}_g(X)}\left(X^{(j)}_{\hat{Q}_g(X)}\right)\right] \\
&= f_{\hat{Q}_f(Y_i)}\left(X^{(i)}_{\hat{Q}_f(Y_i)}\right) - f_{\hat{Q}_f(Y_i)}\left(W^{(i)}_{\hat{Q}_f(Y_i)}\right) + g_{\hat{Q}_g(X)}\left(W^{(i)}_{\hat{Q}_g(X)}\right) - g_{\hat{Q}_g(X)}\left(X^{(i)}_{\hat{Q}_g(X)}\right)
\end{aligned}$$

Then,

$$\begin{aligned}
&\left(Z_{f,g} - Z_i'\right)_+^2 \\
&\quad \le \left[f_{\hat{Q}_f(Y_i)}\left(X^{(i)}_{\hat{Q}_f(Y_i)}\right) - f_{\hat{Q}_f(Y_i)}\left(W^{(i)}_{\hat{Q}_f(Y_i)}\right) + g_{\hat{Q}_g(X)}\left(W^{(i)}_{\hat{Q}_g(X)}\right) - g_{\hat{Q}_g(X)}\left(X^{(i)}_{\hat{Q}_g(X)}\right)\right]^2 \\
&\quad \le 2\left[f_{\hat{Q}_f(Y_i)}\left(X^{(i)}_{\hat{Q}_f(Y_i)}\right) - f_{\hat{Q}_f(Y_i)}\left(W^{(i)}_{\hat{Q}_f(Y_i)}\right)\right]^2 + 2\left[g_{\hat{Q}_g(X)}\left(X^{(i)}_{\hat{Q}_g(X)}\right) - g_{\hat{Q}_g(X)}\left(W^{(i)}_{\hat{Q}_g(X)}\right)\right]^2
\end{aligned} \tag{2}$$

Recall that our goal is to bound $V^+ = \sum_{i=1}^n \mathbb{E}\left[(Z_{f,g} - Z_i')_+^2 \middle| X\right]$. We begin with the second term: by adding and subtracting $\mu_{g,i,\hat{Q}_g(X)}$, expanding the square and noting that the cross term is 0 under the conditional expectation, we get that

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{E}\left[\left[g_{\hat{Q}_g(X)}\left(X^{(i)}_{\hat{Q}_g(X)}\right) - g_{\hat{Q}_g(X)}\left(W^{(i)}_{\hat{Q}_g(X)}\right)\right]^2 \middle| X\right] \\
&\qquad = \sum_{i=1}^n \left\{\left[g_{\hat{Q}_g(X)}\left(X^{(i)}_{\hat{Q}_g(X)}\right) - \mu_{g,i,\hat{Q}_g(X)}\right]^2 + \mathbb{E}\left[\left[g_{\hat{Q}_g(X)}\left(W^{(i)}_{\hat{Q}_g(X)}\right) - \mu_{g,i,\hat{Q}_g(X)}\right]^2 \middle| X\right]\right\} \\
&\qquad \le \underbrace{\max_{Q\in\mathcal{U}}\left\{\sum_{i=1}^n \left[g_Q\left(X^{(i)}_Q\right) - \mu_{g,i,Q}\right]^2\right\}}_{=:\Gamma_g} + \underbrace{\max_{Q\in\mathcal{U}}\left\{\sum_{i=1}^n \mathbb{E}\left[\left[g_Q\left(W^{(i)}_Q\right) - \mu_{g,i,Q}\right]^2\right]\right\}}_{=\sigma_g^2}
\end{aligned}$$

Note that we were able to upper bound via a maximization outside of the sum since the Q indices were fixed w.r.t. i. The first term in (2) is not so readily bounded due to the dependence of Y_i on i. Hence, we rely on a weaker approach: for each $i \in [n]$, we have that

$$\begin{aligned}
&\left[f_{\hat{Q}_f(Y_i)}\left(X^{(i)}_{\hat{Q}_f(Y_i)}\right) - f_{\hat{Q}_f(Y_i)}\left(W^{(i)}_{\hat{Q}_f(Y_i)}\right)\right]^2 \\
&\qquad \le 2\left[f_{\hat{Q}_f(Y_i)}\left(X^{(i)}_{\hat{Q}_f(Y_i)}\right) - \mu_{f,i,\hat{Q}_f(Y_i)}\right]^2 + 2\left[f_{\hat{Q}_f(Y_i)}\left(W^{(i)}_{\hat{Q}_f(Y_i)}\right) - \mu_{f,i,\hat{Q}_f(Y_i)}\right]^2 \\
&\qquad \le 2\max_{Q\in\mathcal{U}}\left\{\left[f_Q\left(X^{(i)}_Q\right) - \mu_{f,i,Q}\right]^2\right\} + 2\max_{Q\in\mathcal{U}}\left\{\left[f_Q\left(W^{(i)}_Q\right) - \mu_{f,i,Q}\right]^2\right\}
\end{aligned}$$

Summing and taking conditional expectations then yields

$$\sum_{i=1}^{n} \mathbb{E}\left[\left[f_{\hat{Q}_f(Y_i)}\left(X^{(i)}_{\hat{Q}_f(Y_i)}\right) - f_{\hat{Q}_f(Y_i)}\left(W^{(i)}_{\hat{Q}_f(Y_i)}\right)\right]^2 \middle| X\right]$$
$$\leq 2 \underbrace{\sum_{i=1}^{n} \max_{Q\in\mathcal{U}} \left\{\left[f_Q\left(X^{(i)}_Q\right) - \mu_{f,i,Q}\right]^2\right\}}_{=:T_f} + 2 \underbrace{\sum_{i=1}^{n} \mathbb{E}\left[\max_{Q\in\mathcal{U}} \left\{\left[f_Q\left(W^{(i)}_Q\right) - \mu_{f,i,Q}\right]^2\right\}\right]}_{=V_f}$$

Finally, by putting everything together, we can obtain the upper bound

$$V^+ \leq 2\left(\Gamma_g + \sigma_g^2\right) + 4\left(T_f + V_f\right)$$

where $\mathbb{E}\left[\Gamma_g\right] = \Sigma_g^2$ and $\mathbb{E}\left[T_f\right] = V_f$.

B.2.3. EFRON-STEIN

Next, we apply exponential Efron-Stein (Proposition 26): for $\lambda \in \left[0, b^{-1}\right)$, we have that

$$\begin{aligned}
\log \mathbb{E}\left[e^{\lambda\left(Z_{f,g} - \mathbb{E}\left[Z_{f,g}\right]\right)}\right] &\leq \frac{b\lambda}{1-b\lambda} \log \mathbb{E}\left[e^{\lambda b^{-1} V^+}\right] \\
&\leq \frac{b\lambda}{1-b\lambda} \log \mathbb{E}\left[e^{\lambda b^{-1}\left[2\left(\Gamma_g + \sigma_g^2\right) + 4\left(T_f + V_f\right)\right]}\right] \\
&= \frac{b\lambda}{1-b\lambda} \left\{\log \mathbb{E}\left[e^{b\lambda\left[\frac{1}{3}\left(6b^{-2}\Gamma_g\right) + \frac{2}{3}\left(6b^{-2}T_f\right)\right]}\right] + \lambda b^{-1}\left(2\sigma_g^2 + 4V_f\right)\right\}
\end{aligned} \tag{3}$$

B.2.4. SELF-BOUNDEDNESS

To bound the cgf of $\frac{1}{3}\left(6b^{-2}\Gamma_g\right) + \frac{2}{3}\left(6b^{-2}T_f\right)$, we will show the self-boundedness of

$$h^{(1)}(\mathbf{x}) := 6b^{-2} \max_{Q\in\mathcal{U}} \sum_{i=1}^{n} \left[g_Q\left(x_Q^i\right) - \mu_{g,i,Q}\right]^2 \quad \text{and} \quad h^{(2)}(\mathbf{x}) := 6b^{-2} \sum_{i=1}^{n} \max_{Q\in\mathcal{U}} \left[f_Q\left(x_Q^i\right) - \mu_{f,i,Q}\right]^2$$

so that the function $\frac{1}{3}h^{(1)} + \frac{2}{3}h^{(2)}$ is also self-bounded by Lemma 24 and we can thus bound the cgf of $\left(\frac{1}{3}h^{(1)} + \frac{2}{3}h^{(2)}\right)(X) = \frac{1}{3}\left(6b^{-2}\Gamma_g\right) + \frac{2}{3}\left(6b^{-2}T_f\right)$ using Proposition 25. We begin by showing that $h^{(1)}$ is self-bounded: let

$$h_i^{(1)}\left(\mathbf{x}_{\setminus i}\right) := 6b^{-2} \max_{Q\in\mathcal{U}} \sum_{j\in[n]:j\neq i} \left[g_Q\left(x_Q^j\right) - \mu_{g,j,Q}\right]^2 \quad \forall i \in [n]$$

and define the maximizing distribution in $h^{(1)}$:

$$\tilde{Q}(\mathbf{x}) := \operatorname*{argmax}_{Q\in\mathcal{U}} \sum_{i=1}^{n} \left[g_Q\left(x_Q^i\right) - \mu_{g,i,Q}\right]^2$$

Fix some $\mathbf{x} \in \left(\mathcal{X}^k\right)^n$ and $i \in [n]$. Clearly, we have that $h^{(1)}(\mathbf{x}) \geq h_i^{(1)}\left(\mathbf{x}_{\setminus i}\right)$. Moreover,

$$\begin{aligned} h^{(1)}(\mathbf{x}) - h_i^{(1)}\left(\mathbf{x}_{\setminus i}\right) &= 6b^{-2}\left[\sum_{j=1}^{n}\left[g_{\tilde{Q}(\mathbf{x})}\left(x^j_{\tilde{Q}(\mathbf{x})}\right) - \mu_{g,i,\tilde{Q}(\mathbf{x})}\right]^2 - \max_{Q\in\mathcal{U}}\left\{\sum_{j\in[n]:j\neq i}\left[g_Q\left(x^j_Q\right) - \mu_{g,j,Q}\right]^2\right\}\right] \\ &\leq 6b^{-2}\left[g_{\tilde{Q}(\mathbf{x})}\left(x^i_{\tilde{Q}(\mathbf{x})}\right) - \mu_{g,i,\tilde{Q}(\mathbf{x})}\right]^2 \\ &\leq 1 \end{aligned}$$

where the last line follows from our assumption that $g_Q \in \left[0, \frac{b}{\sqrt{6}}\right]$. We can add up the bounds to get

$$\sum_{i=1}^{n}\left[h^{(1)}(\mathbf{x}) - h_i^{(1)}\left(\mathbf{x}_{\setminus i}\right)\right] \leq 6b^{-2}\sum_{i=1}^{n}\left[g_{\tilde{Q}(\mathbf{x})}\left(x^i_{\tilde{Q}(\mathbf{x})}\right) - \mu_{g,i,\tilde{Q}(\mathbf{x})}\right]^2 = h^{(1)}(\mathbf{x})$$

Together, these show that $h^{(1)}$ is self-bounded. To show the same for $h^{(2)}$, consider the functions

$$h_i^{(2)}\left(\mathbf{x}_{\setminus i}\right) := 6b^{-2}\sum_{j\in[n]:j\neq i}\max_{Q\in\mathcal{U}}\left[f_Q\left(x^j_Q\right) - \mu_{f,j,Q}\right]^2$$

Again, we have that $h^{(2)}(\mathbf{x}) \geq h_i^{(2)}\left(\mathbf{x}_{\setminus i}\right)$ and

$$h^{(2)}(\mathbf{x}) - h_i^{(2)}\left(\mathbf{x}_{\setminus i}\right) = 6b^{-2}\max_{Q\in\mathcal{U}}\left[f_Q\left(x^i_Q\right) - \mu_{f,i,Q}\right]^2 \leq 1$$

$$\sum_{i=1}^{n}\left[h^{(2)}(\mathbf{x}) - h_i^{(2)}\left(\mathbf{x}_{\setminus i}\right)\right] = h^{(2)}(\mathbf{x})$$

That is, $h^{(2)}$ is also self-bounded. As a result, Proposition 25 implies that

$$\begin{aligned} \log\mathbb{E}\left[e^{b\lambda\left[\frac{1}{3}\left(6b^{-2}\Gamma_g\right)+\frac{2}{3}\left(6b^{-2}T_f\right)\right]}\right] &\leq \left(e^{b\lambda}-1\right)\mathbb{E}\left[\frac{1}{3}\left(6b^{-2}\Gamma_g\right)+\frac{2}{3}\left(6b^{-2}T_f\right)\right] \\ &= \left(e^{b\lambda}-1\right)b^{-2}\left(2\Sigma_g^2+4V_f\right) \\ &\leq \lambda b^{-1}\left(4\Sigma_g^2+8V_f\right) \end{aligned} \tag{4}$$

provided that $\lambda \in \left[0, b^{-1}\right)$, where in the last line we have used the inequality $e^x \leq 1+2x$ for $x \leq 1$.

B.2.5. SUB-GAMMA TAIL

Finally, we can combine Equations (3) and (4) to get that

$$\log\mathbb{E}\left[e^{\lambda\left(Z_{f,g}-\mathbb{E}\left[Z_{f,g}\right]\right)}\right] \leq \frac{\lambda^2}{1-b\lambda}\left(2\sigma_g^2+4\Sigma_g^2+12V_f\right) = \frac{\left(4\sigma_g^2+8\Sigma_g^2+24V_f\right)\lambda^2}{2\left(1-b\lambda\right)}$$

for all $\lambda \in \left[0, b^{-1}\right)$. That is, $Z_{f,g} \in \Gamma_+\left(\sqrt{4\sigma_g^2+8\Sigma_g^2+24V_f}, b\right)$, which we know from Proposition 23 yields the tail bound

$$\mathbb{P}\left(Z_{f,g}-\mathbb{E}\left[Z_{f,g}\right]\geq t\right) \leq \exp\left(-\frac{t^2}{2\left(4\sigma_g^2+8\Sigma_g^2+24V_f+bt\right)}\right) \quad \forall t\geq 0 \tag{5}$$

B.2.6. ORIGINAL SETTING

Recall that our original variables of interest live in some set $\mathcal{X}_0 \subset \mathbb{R}$, and that sample sizes n_Q may vary. Let $n \coloneqq \max_{Q\in\mathcal{U}} n_Q$ and consider the space $\mathcal{X} = \mathcal{X}_0 \cup \{x_0\}$ for the setup of this proof, where $x_0 \notin \mathcal{X}_0$. Suppose that $\left(X_Q^{(i)}\right)_{i\in[n_Q]} \overset{iid}{\sim} Q$ and $X_Q^{(n_Q+1)} = \cdots = X_Q^{(n)} = x_0$ almost surely. Let $f : \mathcal{X}_0 \to \mathbb{R}$ be the L-Lipschitz function from the statement of Theorem 22, and consider its extension $\tilde{f} : \mathcal{X} \to \mathbb{R}$ given by

$$\tilde{f}\left(x\right) \coloneqq \begin{cases} f\left(x\right) & x \in \mathcal{X}_0 \\ 0 & x = x_0 \end{cases}$$

We apply the analysis above to the functions $f_Q \coloneqq \frac{\tilde{f}}{n_Q}$, ensuring that

$$Z_f = \min_{Q\in\mathcal{U}} \frac{1}{n_Q} \sum_{i=1}^{n_Q} f\left(X_Q^{(i)}\right)$$

where the variables follow the appropriate distributions, as in the original goal. Note that $f_Q \in \left[0, \frac{M}{n_Q}\right]$, so that we can set $b = \frac{\sqrt{6}M}{\min_{Q\in\mathcal{U}} n_Q}$. We analogously define everything for g. Next, we apply Lemma 17 under the Lipschitzness assumption to obtain

$$\sigma_g^2 = \max_{Q\in\mathcal{U}} \left\{ n_Q \operatorname{Var}\left(\frac{g\left(X_Q\right)}{n_Q}\right)\right\} \le 2L^2 \max_{Q\in\mathcal{U}} \frac{\sigma_Q^2}{n_Q} = 2L^2\sigma_T^2$$

For each $Q \in \mathcal{U}$, let $X_Q \sim Q$ be independent from $\left(X_Q^{(i)}\right)_{i\in[n_Q]}$. Then,

$$\begin{aligned}
\Sigma_g^2 &= \mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \left[g\left(X_Q^{(i)}\right) - \mathbb{E}\left[g\left(X_Q\right)\right]\right]^2\right] \\
&= \mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \mathbb{E}\left[g\left(X_Q^{(i)}\right) - g\left(X_Q\right)\middle| X_Q^{(i)}\right]^2\right] \\
&\le L^2\mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \mathbb{E}\left[\left(X_Q^{(i)} - X_Q\right)^2\middle| X_Q^{(i)}\right]\right] && \text{Lipschitzness + Jensen's} \\
&= L^2\mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \left[\left(X_Q^{(i)} - \mu_Q\right)^2 + \sigma_Q^2\right]\right] && \mathbb{E}\left[\left(X_Q^{(i)} - \mu_Q\right)\left(\mu_Q - X_Q\right)\middle| X_Q^{(i)}\right] = 0 \\
&\le L^2\left\{\mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \left(X_Q^{(i)} - \mu_Q\right)^2\right] + \max_{Q\in\mathcal{U}} \frac{\sigma_Q^2}{n_Q}\right\} \\
&= L^2\left(\Sigma_T^2 + \sigma_T^2\right)
\end{aligned}$$

It remains to bound V_f: recall that $0 = n_{(0)} \leq n_{(1)} \leq \cdots \leq n_{(k)}$ and $n_{(j)} = n_{Q_{(j)}}$, so that

$$\begin{aligned} V_f &= \sum_{i=1}^{n} \mathbb{E}\left[\max_{Q\in\mathcal{U}} \left[f_Q\left(X_Q^{(i)}\right) - \mu_{f,i,Q}\right]^2\right] \\ &= \sum_{j=1}^{k} \left(n_{(j)} - n_{(j-1)}\right) \mathbb{E}\left[\max_{r\in\{j,\ldots,k\}} \frac{1}{n_{(r)}^2} \left[f\left(X_{Q_{(r)}}\right) - \mathbb{E}\left[f\left(X_{Q_{(r)}}\right)\right]\right]^2\right] \end{aligned}$$

With a similar symmetrization trick, we can further bound each expectation in the sum: let X'_Q be an independent copy of X_Q. Then,

$$\begin{aligned} \mathbb{E}\left[\max_{r\in\{j,\ldots,k\}} \frac{1}{n_{(r)}^2} \left[f\left(X_{Q_{(r)}}\right) - \mathbb{E}\left[f\left(X_{Q_{(r)}}\right)\right]\right]^2\right] &= \mathbb{E}\left[\max_{r\in\{j,\ldots,k\}} \frac{1}{n_{(r)}^2} \mathbb{E}\left[f\left(X_{Q_{(r)}}\right) - f\left(X'_{Q_{(r)}}\right)\middle| X_{Q_{(r)}}\right]^2\right] \\ &\overset{(1)}{\leq} L^2 \mathbb{E}\left[\max_{r\in\{j,\ldots,k\}} \frac{1}{n_{(r)}^2} \mathbb{E}\left[\left(X_{Q_{(r)}} - X'_{Q_{(r)}}\right)^2 \middle| X_{Q_{(r)}}\right]\right] \\ &\overset{(2)}{\leq} L^2 \mathbb{E}\left[\max_{r\in\{j,\ldots,k\}} \frac{1}{n_{(r)}^2} \left\{\left[X_{Q_{(r)}} - \mu_{Q_{(r)}}\right]^2 + \sigma_{Q_{(r)}}^2\right\}\right] \\ &\leq 2L^2 \mathbb{E}\left[\max_{r\in\{j,\ldots,k\}} \frac{1}{n_{(r)}^2} \left[X_{Q_{(r)}} - \mu_{Q_{(r)}}\right]^2\right] \end{aligned}$$

where, in (1), we have applied Lipschitzness and Jensen's and, in (2), we note again that the cross term cancels when expanding the square. Hence, we get that

$$V_f \leq 2L^2 \sum_{j=1}^{k} \left(n_{(j)} - n_{(j-1)}\right) \mathbb{E}\left[\max_{r\in\{j,\ldots,k\}} \frac{1}{n_{(r)}^2} \left[X_{Q_{(r)}} - \mu_{Q_{(r)}}\right]^2\right] = 2L^2 V_T$$

Plugging these values back into the bound (5) then yields the claim. ■

Appendix C. Proofs of Section 3

Recall our non-adaptive proxy objective

$$\mu_T^{\circ}(a) = \min_{Q\in\mathcal{U}} \frac{1}{n_Q} \sum_{i=1}^{n_Q} r\left(a, X_Q^{(i)}\right)$$

where, for UE, $n_Q = n$ for all $Q \in \mathcal{U}$. For $a \in \mathcal{A}$, define generalization gaps

$$D_a \coloneqq \mu_{\mathrm{DR}}(a) - \mu_T^{\circ}(a) = \min_{Q\in\mathcal{U}} \mu(a;Q) - \min_{Q'\in\mathcal{U}} \hat{\mu}_{n_{Q'}}(a;Q')$$

Using the same argument as in the proof of Proposition 19, we note that

$$|D_a| \leq \max_{Q\in\mathcal{U}} \left|\mu(a;Q) - \hat{\mu}_{n_Q}(a;Q)\right| = \max_{Q\in\mathcal{U}} \left|\frac{1}{n_Q} \sum_{i=1}^{n_Q} \left[\mathbb{E}\left[r\left(a, X_Q\right)\right] - r\left(a, X_Q^{(i)}\right)\right]\right| =: U_a$$

Then from the theory of Appendix A, we can conclude the following bounds.

Theorem 27 *For rewards bounded in $[0, M]$, we have that for any $a \in \mathcal{A}$,*

$$\mathbb{E}[U_a] \leq 4M\sqrt{\frac{\log k}{\min_{Q\in\mathcal{U}} n_Q}}$$

Additionally, when $\mathcal{X} \subset \mathbb{R}$ and $r(a, \cdot)$ is L-Lipschitz for each $a \in \mathcal{A}$, it follows that

$$\mathbb{E}[U_a] \leq \frac{16M\log k}{\min_{Q\in\mathcal{U}} n_Q} + 4L\sigma_T\sqrt{2\log k}$$

Proof We apply Theorem 11 on functions $f_Q(x) := \mathbb{E}[r(a, X_Q)] - r(a, x)$. Note that $f_Q \in [-M, M]$ when $r \in [0, M]$. Moreover, if $r(a, \cdot)$ is L-Lipschitz, then so is f_Q, as we only add a constant to it. ■

Let $\mathbb{E}[U_a] \leq B$ be any of the bounds from Theorem 27. Then, we get that

$$\begin{aligned}
\mathbb{E}[\mu_T^{\mathrm{o}}(a^*) - \mu_T^{\mathrm{o}}(a)] &= \Delta_{\mathrm{DR}}(a) + \mathbb{E}[\mu_{\mathrm{DR}}(a) - \mu_T^{\mathrm{o}}(a)] - \mathbb{E}\left[\overset{*}{\mu}_{\mathrm{DR}} - \mu_T^{\mathrm{o}}(a^*)\right] \\
&= \Delta_{\mathrm{DR}}(a) + \mathbb{E}[D_a] - \mathbb{E}[D_{a^*}] \\
&\geq \Delta_{\mathrm{DR}}(a) - |\mathbb{E}[D_a]| - |\mathbb{E}[D_{a^*}]| \\
&\geq \Delta_{\mathrm{DR}}(a) - \mathbb{E}[|D_a|] - \mathbb{E}[|D_{a^*}|] \\
&\geq \Delta_{\mathrm{DR}}(a) - 2\mathbb{E}[U_a] \\
&\geq \Delta_{\mathrm{DR}}(a) - 2B
\end{aligned}$$

for all $a \in \mathcal{A}$. Hence,

$$\begin{aligned}
\mathbb{P}(A_T^{\mathrm{o}} = a) &\leq \mathbb{P}(\mu_T^{\mathrm{o}}(a) \geq \mu_T^{\mathrm{o}}(a^*)) \\
&= \mathbb{P}(\mu_T^{\mathrm{o}}(a) - \mu_T^{\mathrm{o}}(a^*) - \mathbb{E}[\mu_T^{\mathrm{o}}(a) - \mu_T^{\mathrm{o}}(a^*)] \geq \mathbb{E}[\mu_T^{\mathrm{o}}(a^*) - \mu_T^{\mathrm{o}}(a)]) \\
&\leq \mathbb{P}(\mu_T^{\mathrm{o}}(a) - \mu_T^{\mathrm{o}}(a^*) - \mathbb{E}[\mu_T^{\mathrm{o}}(a) - \mu_T^{\mathrm{o}}(a^*)] \geq \Delta_{\mathrm{DR}}(a) - 2B) \qquad (6)
\end{aligned}$$

What remains is to apply the concentration inequalities of Appendix B.

C.1. Proof of Theorem 4

Here, we use the UE proxy $\mu_T^{\mathrm{o}}(a) = \min_{Q\in\mathcal{U}} \frac{1}{n}\sum_{i=1}^{n} r\left(a, X_Q^{(i)}\right)$. We can then obtain the following concentration inequality.

Corollary 28 (UE concentration inequality) *We have that*

$$\mathbb{P}\left(\mu_T^{o}(a) - \mu_T^{o}(a') - \mathbb{E}\left[\mu_T^{o}(a) - \mu_T^{o}(a')\right] \geq t\right) \leq \exp\left(-\frac{nt^2}{2M^2}\right)$$

for all $t \geq 0$ and $a, a' \in \mathcal{A}$.

Proof Note that in the notation of Appendix B.1, $Z_{r(a,\cdot)/M} = \frac{\mu_T^{\mathrm{o}}(a)}{M}$. Since $r(a, \cdot) \in [0, M]$ for each $a \in \mathcal{A}$, the claim follows by applying Corollary 21. ■

Next, note that under the assumption $n \geq \left(\frac{8M}{\Delta_{\mathrm{DR,min}}}\right)^2 \log k$, we get that $\Delta_{\mathrm{DR}}(a) \geq 8M\sqrt{\frac{\log k}{n}}$ for all $a \in \mathcal{A}$ with a positive gap. Hence, for all such a, plugging in the bound $B = 4M\sqrt{\frac{\log k}{n}}$ into Equation (6) yields

$$\mathbb{P}\left(A_T^{\mathrm{o}} = a\right) \leq \mathbb{P}\left(\mu_T^{\mathrm{o}}(a) - \mu_T^{\mathrm{o}}(a^*) - \mathbb{E}\left[\mu_T^{\mathrm{o}}(a) - \mu_T^{\mathrm{o}}(a^*)\right] \geq \Delta_{\mathrm{DR}}(a) - 8M\sqrt{\frac{\log k}{n}}\right) \quad \text{Eq. (6)}$$
$$\leq \exp\left(-\frac{n}{2M^2}\left[\Delta_{\mathrm{DR}}(a) - 8M\sqrt{\frac{\log k}{n}}\right]^2\right) \quad \text{Cor. 28}$$

This directly yields the desired regret bound:

$$\mathbb{E}\left[\Delta_{\mathrm{DR}}\left(A_T^{\mathrm{o}}\right)\right] = \sum_{a\in\mathcal{A}:\Delta_{\mathrm{DR}}(a)>0} \Delta_{\mathrm{DR}}(a)\, \mathbb{P}\left(A_T^{\mathrm{o}} = a\right)$$
$$\leq \sum_{a\in\mathcal{A}:\Delta_{\mathrm{DR}}(a)>0} \Delta_{\mathrm{DR}}(a) \exp\left(-\frac{n}{2M^2}\left[\Delta_{\mathrm{DR}}(a) - 8M\sqrt{\frac{\log k}{n}}\right]^2\right)$$

C.2. Proof of Corollary 6

An alternative way of writing the UE regret bound is as follows:

$$\mathbb{E}\left[\Delta_{\mathrm{DR}}\left(A_T^{\mathrm{o}}\right)\right] = \sum_{a\in\mathcal{A}:\Delta_{\mathrm{DR}}(a)\leq\Delta} \Delta_{\mathrm{DR}}(a)\, \mathbb{P}\left(A_T^{\mathrm{o}} = a\right) + \sum_{a\in\mathcal{A}:\Delta_{\mathrm{DR}}(a)>\Delta} \Delta_{\mathrm{DR}}(a)\, \mathbb{P}\left(A_T^{\mathrm{o}} = a\right)$$
$$\leq \Delta + \sum_{a\in\mathcal{A}:\Delta_{\mathrm{DR}}(a)>\Delta} \Delta_{\mathrm{DR}}(a) \exp\left(-\frac{n}{2M^2}\left[\Delta_{\mathrm{DR}}(a) - 8M\sqrt{\frac{\log k}{n}}\right]^2\right)$$

for any $\Delta \geq 0$. In other words,

$$\mathbb{E}\left[\Delta_{\mathrm{DR}}\left(A_T^{\mathrm{o}}\right)\right] \leq \inf_{\Delta\geq 0}\left\{\Delta + \sum_{a\in\mathcal{A}:\Delta_{\mathrm{DR}}(a)>\Delta} \Delta_{\mathrm{DR}}(a) \exp\left(-\frac{n}{2M^2}\left[\Delta_{\mathrm{DR}}(a) - 8M\sqrt{\frac{\log k}{n}}\right]^2\right)\right\} \tag{7}$$

Next, we introduce a simple technical lemma.

Lemma 29 *Let $\alpha, \beta > 0$. Then, the function $f(x) := x\exp\left(-\alpha(x-\beta)^2\right)$ is decreasing for $x \geq \frac{1}{2}\left(\beta + \sqrt{\beta^2 + \frac{2}{\alpha}}\right)$.*

Proof Notice that

$$f'(x) = \exp\left(-\alpha(x-\beta)^2\right) - 2\alpha x(x-\beta)\exp\left(-\alpha(x-\beta)^2\right)$$
$$= \left[1 - 2\alpha x(x-\beta)\right]\exp\left(-\alpha(x-\beta)^2\right)$$

Now, note that the function $x \mapsto 2\alpha x(x-\beta)-1$ is quadratic, convex and has roots $\frac{1}{2}\left(\beta + \sqrt{\beta^2 + \frac{2}{\alpha}}\right)$ and $\frac{1}{2}\left(\beta - \sqrt{\beta^2 + \frac{2}{\alpha}}\right)$. Since the former is larger, it follows that the quadratic is nonnegative for larger values. In other words, $f'(x) \leq 0$ whenever $x \geq \frac{1}{2}\left(\beta + \sqrt{\beta^2 + \frac{2}{\alpha}}\right)$. ■

As a result, we can show the following inequality.

Lemma 30 *Provided that $l \geq 2$ and $\Delta_{\text{DR}}(a) \geq \frac{8M\sqrt{\log k}+M\sqrt{2\log l}}{\sqrt{n}}$, we have that*

$$\Delta_{\text{DR}}(a) \exp\left(-\frac{n}{2M^2}\left[\Delta_{\text{DR}}(a) - 8M\sqrt{\frac{\log k}{n}}\right]^2\right) \leq \frac{8M\sqrt{\log k} + M\sqrt{2\log l}}{l\sqrt{n}}$$

Proof Note that the left-hand side of the claim is of the form $f(\Delta_{\text{DR}}(a))$, where f is defined as in Lemma 29 with $\alpha := \frac{n}{2M^2}$ and $\beta := 8M\sqrt{\frac{\log k}{n}}$, so that we know it is decreasing for $x \geq K$, where

$$\begin{aligned}
K &:= \frac{1}{2}\left(\beta + \sqrt{\beta^2 + \frac{2}{\alpha}}\right) \\
&= \frac{1}{2}\left[8M\sqrt{\frac{\log k}{n}} + \sqrt{\frac{64M^2\log k}{n} + \frac{4}{n}}\right] \\
&= \frac{8M\sqrt{\log k} + \sqrt{64M^2\log k + 4}}{2\sqrt{n}} \\
&\leq \frac{8M\sqrt{\log k} + 1}{\sqrt{n}} && \sqrt{a+b} \leq \sqrt{a} + \sqrt{b} \\
&\leq \frac{8M\sqrt{\log k} + M\sqrt{2\log l}}{\sqrt{n}} && M\sqrt{2\log l} \geq 1
\end{aligned}$$

The result then follows by plugging in $\frac{8M\sqrt{\log k}+M\sqrt{2\log l}}{\sqrt{n}}$ into f to get the right-hand side of the claim. ■

Finally, we can set $\Delta := \frac{8M\sqrt{\log k}+M\sqrt{2\log l}}{\sqrt{n}}$ in Equation (7) and apply Lemma 30 to obtain

$$\begin{aligned}
\mathbb{E}\left[\Delta_{\text{DR}}(A_T^{\text{o}})\right] &\leq \frac{8M\sqrt{\log k} + M\sqrt{2\log l}}{\sqrt{n}} + \left|\{a \in \mathcal{A} : \Delta_{\text{DR}}(a) > \Delta\}\right| \frac{8M\sqrt{\log k} + M\sqrt{2\log l}}{l\sqrt{n}} \\
&\leq \frac{16M\sqrt{\log k} + 2M\sqrt{2\log l}}{\sqrt{n}} \\
&\lesssim M\sqrt{\frac{\log(kl)}{n}}
\end{aligned}$$

where in the last line we have used the fact that $\sqrt{a} + \sqrt{b} \leq \sqrt{2(a+b)}$. Substituting $n = T/k$ then yields the result.

C.3. Proof of Theorem 7

Returning to the general NUE proxy $\mu^{\text{o}}_T(a) = \min_{Q\in\mathcal{U}} \frac{1}{n_Q}\sum_{i=1}^{n_Q} r\left(a, X_Q^{(i)}\right)$, let us further assume that $\mathcal{X} \subset \mathbb{R}$. Then, we conclude the following result.

Corollary 31 (NUE concentration inequality) *Suppose that $r(a,\cdot) : \mathcal{X} \to [0, M]$ is L-Lipschitz for each $a \in \mathcal{A}$. Then, we have that*

$$\mathbb{P}\left(\mu^{\text{o}}_T(a) - \mu^{\text{o}}_T(a') - \mathbb{E}\left[\mu^{\text{o}}_T(a) - \mu^{\text{o}}_T(a')\right] \geq t\right) \leq \exp\left(-\frac{t^2}{16L^2\left(2\sigma_T^2 + \Sigma_T^2 + 6V_T\right) + \frac{2\sqrt{6}Mt}{\min_{Q\in\mathcal{U}} n_Q}}\right)$$

for all $t \geq 0$ and $a, a' \in \mathcal{A}$.

Proof Once again, using the definitions of Appendix B, we get that $Z_{r(a,\cdot)} = \mu^{\text{o}}_T(a)$. Since $r(a,\cdot) \in [0, M]$ is L-Lipschitz for each $a \in \mathcal{A}$, the claim follows by applying Theorem 22. ■

As in the UE analysis, provided that $\Delta_{\text{DR,min}} \geq G_T = \frac{32M\log k}{\min_{Q\in\mathcal{U}} n_Q} + 8L\sigma_T\sqrt{2\log k}$, we can plug $B = \frac{16M\log k}{\min_{Q\in\mathcal{U}} n_Q} + 4L\sigma_T\sqrt{2\log k}$ into Equation (6) to conclude that

$$\begin{aligned}
\mathbb{P}\left(A^{\text{o}}_T = a\right) &\leq \mathbb{P}\left(\mu^{\text{o}}_T(a) - \mu^{\text{o}}_T(a^*) - \mathbb{E}\left[\mu^{\text{o}}_T(a) - \mu^{\text{o}}_T(a^*)\right] \geq \Delta_{\text{DR}}(a) - G_T\right) && \text{Eq. (6)}\\
&\leq \exp\left(-\frac{\left[\Delta_{\text{DR}}(a) - G_T\right]^2}{16L^2\left(2\sigma_T^2 + \Sigma_T^2 + 6V_T\right) + \frac{2\sqrt{6}M}{\min_{Q\in\mathcal{U}} n_Q}\left[\Delta_{\text{DR}}(a) - G_T\right]}\right) && \text{Cor. 28}
\end{aligned}$$

for all $a \in \mathcal{A}$ with positive gap. This in turn yields the regret bound

$$\begin{aligned}
\mathbb{E}\left[\Delta_{\text{DR}}\left(A^{\text{o}}_T\right)\right] &= \sum_{a\in\mathcal{A}:\Delta_{\text{DR}}(a)>0} \Delta_{\text{DR}}(a)\,\mathbb{P}\left(A^{\text{o}}_T = a\right)\\
&\leq \sum_{a\in\mathcal{A}:\Delta_{\text{DR}}(a)>0} \Delta_{\text{DR}}(a)\exp\left(-\frac{\left[\Delta_{\text{DR}}(a) - G_T\right]^2}{16L^2\left(2\sigma_T^2 + \Sigma_T^2 + 6V_T\right) + \frac{2\sqrt{6}M}{\min_{Q\in\mathcal{U}} n_Q}\left[\Delta_{\text{DR}}(a) - G_T\right]}\right)
\end{aligned}$$

Appendix D. Modified UCB-E

Our goal is to perform a minimization variant of UCB-E Audibert et al. (2010) for T rounds on the set of "arms" $\mathcal{U}$. Since we will analyze all random variables under a fixed high-probability event, we treat all quantities here as deterministic. In particular, we work with $\mu(Q), \hat{\mu}_t(Q) \in [0,1]$ for each $Q \in \mathcal{U}$ and $t \in \{n_0(Q), \ldots, n_0(Q) + T\}$, where $n_0(Q) \geq 1$ is the number of pulls from arm $Q \in \mathcal{U}$ that we start the game with. We assume a unique optimal arm $Q^* := \operatorname{argmin}_{Q\in\mathcal{U}} \mu(Q)$, with $\mu^* := \mu(Q^*)$, and define suboptimality gaps $\Delta(Q) := \mu(Q) - \mu^*$ and $\Delta_{\min} := \min_{Q\in\mathcal{U}\setminus\{Q^*\}} \Delta(Q)$. For some choice of plays $\{Q_t\}_{t=1}^T$, let

$$n_t(Q) := n_0(Q) + \sum_{s=1}^{t} \mathbb{I}\{Q_s = Q\}$$

denote the number of times distribution Q has been played at time $t \in [T]$. Additionally, we define the following subset of arms:

$$\mathcal{U}_0 \coloneqq \left\{ Q \in \mathcal{U} \setminus \{Q^*\} : n_0(Q) < \frac{36}{25} \epsilon \Delta^{-2}(Q) \right\} \cup \left\{ Q^* : n_0(Q^*) < \frac{36}{25} \epsilon \Delta_{\text{min}}^{-2} \right\}$$

along with its cardinality (provided that it contains Q^*) $k_0 \coloneqq |\mathcal{U}_0| \, \mathbb{I}\{Q^* \in \mathcal{U}_0\}$, total initial sample size $\tilde{T}_0 \coloneqq \sum_{Q \in \mathcal{U}_0} n_0(Q)$ and the complexity notion it defines:

$$H_0 \coloneqq \Delta_{\text{min}}^{-2} \mathbb{I}\{Q^* \in \mathcal{U}_0\} + \sum_{Q \in \mathcal{U}_0 \setminus \{Q^*\}} \Delta^{-2}(Q)$$

The intuition is that $\mathcal{U}_0$ is a proxy for the set of arms played:

$$\mathcal{U}' \coloneqq \{ Q \in \mathcal{U} : n_T(Q) > n_0(Q) \}$$

The UCB-E algorithm works by defining indices (adjusted here for lower confidence bounds)

$$\text{LCB}_t(Q; \epsilon) \coloneqq \hat{\mu}_{n_t(Q)}(Q) - \sqrt{\frac{\epsilon}{n_t(Q)}} \quad \forall Q \in \mathcal{U}$$

given a parameter $\epsilon > 0$ and, at each time step $t \in [T]$, playing

$$Q_t \coloneqq \operatorname*{argmin}_{Q \in \mathcal{U}} \text{LCB}_{t-1}(Q; \epsilon)$$

After T rounds, we output

$$\hat{Q} \coloneqq \operatorname*{argmin}_{Q \in \mathcal{U}} \hat{\mu}_{n_T(Q)}(Q)$$

Theorem 32 (Modified UCB-E optimality) *Suppose that*

$$|\mu(Q) - \hat{\mu}_t(Q)| < \frac{1}{5} \sqrt{\frac{\epsilon}{t}}$$

for all $Q \in \mathcal{U}$ and $t \in \{n_0(Q), \ldots, n_0(Q) + T\}$, and that

$$T \geq \frac{36}{25} \epsilon H_0 - \tilde{T}_0 + k_0$$

Then, it follows that $\hat{Q} = Q^*$ and

$$\frac{1}{5} \sqrt{\frac{\epsilon}{n_T(Q^*)}} \leq \frac{\Delta_{min}}{2}$$

Proof

First, notice that for any $t \in \{0, \ldots, T\}$ and $Q \in \mathcal{U}$, we have by assumption that

$$\left| \mu(Q) - \hat{\mu}_{n_t(Q)}(Q) \right| < \frac{1}{5} \sqrt{\frac{\epsilon}{n_t(Q)}} \tag{8}$$

since $n_t(Q) \in \{n_0(Q), \ldots, n_0(Q) + T\}$. All we need to do is show that, for any $Q \in \mathcal{U} \setminus \{Q^*\}$,

$$n_T(Q) \geq \frac{4}{25} \epsilon \Delta^{-2}(Q) \quad \text{and} \quad n_T(Q^*) \geq \frac{4}{25} \epsilon \Delta_{\text{min}}^{-2} \tag{9}$$

since this implies that

$$\frac{1}{5}\sqrt{\frac{\epsilon}{n_T(Q)}} \leq \frac{\Delta(Q)}{2} \quad \text{and} \quad \frac{1}{5}\sqrt{\frac{\epsilon}{n_T(Q^*)}} \leq \frac{\Delta_{\text{min}}}{2} \leq \frac{\Delta(Q)}{2}$$

The second inequality is one of our desired results. To obtain the other, we observe that

$$\begin{aligned}
\hat{\mu}_{n_T(Q)}(Q) - \hat{\mu}_{n_T(Q^*)}(Q^*) &= \hat{\mu}_{n_T(Q)}(Q) - \mu(Q) + \Delta(Q) + \mu^* - \hat{\mu}_{n_T(Q^*)}(Q^*) \\
&> \Delta(Q) - \frac{1}{5}\sqrt{\frac{\epsilon}{n_T(Q)}} - \frac{1}{5}\sqrt{\frac{\epsilon}{n_T(Q^*)}} && \text{Eq. (8)} \\
&\geq \Delta(Q) - \frac{\Delta(Q)}{2} - \frac{\Delta(Q)}{2} \\
&= 0
\end{aligned}$$

Since this holds for all $Q \in \mathcal{U} \setminus \{Q^*\}$, it follows that $\hat{Q} = Q^*$. To show (9), we break into two cases.

D.1. Case 1: $Q^* \notin \mathcal{U}_0$

First, suppose that $Q^* \notin \mathcal{U}_0$ and note that

$$n_T(Q) \geq \frac{36}{25} \epsilon \Delta^{-2}(Q) > \frac{4}{25} \epsilon \Delta^{-2}(Q) \quad \text{and} \quad n_T(Q^*) \geq n_0(Q^*) \geq \frac{36}{25} \epsilon \Delta_{\text{min}}^{-2} > \frac{4}{25} \epsilon \Delta_{\text{min}}^{-2}$$

for any $Q \notin \mathcal{U}_0 \cup \{Q^*\}$ by definition. To show the first inequality for $\mathcal{U}_0$, we observe that $k_0 = 0$ and $H_0 = \sum_{Q \in \mathcal{U}_0} \Delta^{-2}(Q)$ and make the following claim, that also applies in Case 2 (D.2).

Lemma 33 *Fix $t \in [T]$. If $Q_t = Q \neq Q^*$, then*

$$n_{t-1}(Q) < \frac{36}{25} \epsilon \Delta^{-2}(Q)$$

Proof We have that

$$\begin{aligned}
\mu^* &> \hat{\mu}_{n_{t-1}(Q^*)}(Q^*) - \frac{1}{5}\sqrt{\frac{\epsilon}{n_{t-1}(Q^*)}} && \text{Eq. (8)} \\
&\geq \text{LCB}_{t-1}(Q^*; \epsilon) \\
&\geq \text{LCB}_{t-1}(Q; \epsilon) && Q_t = Q \\
&= \hat{\mu}_{n_{t-1}(Q)}(Q) - \sqrt{\frac{\epsilon}{n_{t-1}(Q)}} \\
&> \mu(Q) - \frac{6}{5}\sqrt{\frac{\epsilon}{n_{t-1}(Q)}} && \text{Eq. (8)}
\end{aligned}$$

Rearranging then yields the claim. ■

In other words, once $n_t(Q) \geq \frac{36}{25}\epsilon\Delta^{-2}(Q)$, arm $Q \neq Q^*$ will no longer be played after round t. This means that any arm outside of $\mathcal{U}_0 \cup \{Q^*\}$ will not be played at all. That is, $\mathcal{U}' \subset \mathcal{U}_0 \cup \{Q^*\}$. In addition, if Q^* is not played in the first

$$T' \coloneqq \sum_{Q \in \mathcal{U}_0} \left[\frac{36}{25}\epsilon\Delta^{-2}(Q) - n_0(Q)\right] = \frac{36}{25}\epsilon H_0 - \tilde{T}_0 + k_0$$

rounds, then the plays will be distributed within $\mathcal{U}_0$, resulting in

$$n_T(Q) \geq n_{T'}(Q) \geq \frac{36}{25}\epsilon\Delta^{-2}(Q) > \frac{4}{25}\epsilon\Delta^{-2}(Q) \quad \forall Q \in \mathcal{U}_0$$

where the first inequality uses the assumption that $T \geq T'$. When Q^* is played, we get the following result.

Proposition 34 *Suppose that Q^* is played in some round. Then,*

$$n_T(Q) \geq \frac{4}{25}\epsilon\Delta^{-2}(Q) \quad \forall Q \in \mathcal{U}_0$$

Proof Let $Q \in \mathcal{U}_0$ and let $t \in [T]$ be any round such that $Q_t = Q^*$. Then,

$$\begin{aligned}
\mu(Q) - \frac{4}{5}\sqrt{\frac{\epsilon}{n_T(Q)}} &\geq \mu(Q) - \frac{4}{5}\sqrt{\frac{\epsilon}{n_{t-1}(Q)}} \\
&> \mathrm{LCB}_{t-1}(Q;\epsilon) && \text{Eq. (8)} \\
&\geq \mathrm{LCB}_{t-1}(Q^*;\epsilon) \\
&> \mu^* - \frac{6}{5}\sqrt{\frac{\epsilon}{n_{t-1}(Q^*)}} && \text{Eq. (8)} \\
&\geq \mu^* - \frac{6}{5}\sqrt{\frac{\epsilon}{n_0(Q^*)}} \\
&\geq \mu^* - \Delta_{\text{min}} && n_0(Q^*) \geq \frac{36}{25}\epsilon\Delta_{\text{min}}^{-2} \\
&\geq \mu^* - \Delta(Q)
\end{aligned}$$

The claim then follows by rearranging the terms. ■

D.2. Case 2: $Q^* \in \mathcal{U}_0$

Next, we note that

$$k_0 = |\mathcal{U}_0| \quad \text{and} \quad H_0 = \Delta_{\text{min}}^{-2} + \sum_{Q \in \mathcal{U}_0 \setminus \{Q^*\}} \Delta^{-2}(Q)$$

As a direct consequence of Lemma 33, we can conclude that our proxy set $\mathcal{U}_0$ indeed contains the arms played.

Corollary 35 $\mathcal{U}' \subset \mathcal{U}_0$.

Proof Fix $Q \in \mathcal{U}' \setminus \{Q^*\}$ and let $t \in [T]$ denote any round in which $Q_t = Q$. From Lemma 33 we then get that $n_0(Q) \leq n_{t-1}(Q) < \frac{36}{25}\epsilon\Delta^{-2}(Q)$. ■

Next, we show that suboptimal arms in the proxy set do not have too many samples by the end of the procedure.

Proposition 36

$$n_T(Q) < \frac{36}{25}\epsilon\Delta^{-2}(Q) + 1 \quad \forall Q \in \mathcal{U}_0 \setminus \{Q^*\}$$

Proof If $Q \in \mathcal{U}_0 \setminus (\mathcal{U}' \cup \{Q^*\})$, then

$$n_T(Q) = n_0(Q) < \frac{36}{25}\epsilon\Delta^{-2}(Q) < \frac{36}{25}\epsilon\Delta^{-2}(Q) + 1$$

Otherwise, fix any $Q \in \mathcal{U}' \setminus \{Q^*\}$ and let $t \in [T]$ be the largest time step such that $Q_t = Q$ (i.e., the last round in which Q is played). Lemma 33 then implies that

$$n_T(Q) = n_{T-1}(Q) = \cdots = n_t(Q) = n_{t-1}(Q) + 1 < \frac{36}{25}\epsilon\Delta^{-2}(Q) + 1$$

■

This, in turn, implies that the optimal arm has sufficiently many samples and, in fact, is in $\mathcal{U}'$.

Proposition 37

$$n_T(Q^*) > \frac{36}{25}\epsilon\Delta_{min}^{-2} + 1$$

Proof We have that

$$\begin{aligned}
n_T(Q^*) &= T + n_0(Q^*) - \sum_{Q \in \mathcal{U}' \setminus \{Q^*\}} [n_T(Q) - n_0(Q)] \\
&= T + n_0(Q^*) - \sum_{Q \in \mathcal{U}_0 \setminus \{Q^*\}} [n_T(Q) - n_0(Q)] && \text{Cor. 35} \\
&= T + \tilde{T}_0 - \sum_{Q \in \mathcal{U}_0 \setminus \{Q^*\}} n_T(Q) \\
&> T + \tilde{T}_0 - \sum_{Q \in \mathcal{U}_0 \setminus \{Q^*\}} \left[\frac{36}{25}\epsilon\Delta^{-2}(Q) + 1\right] && \text{Prop. 36} \\
&= T + \tilde{T}_0 - \frac{36}{25}\epsilon\left(H_0 - \Delta_{\text{min}}^{-2}\right) - k_0 + 1 \\
&\geq \frac{36}{25}\epsilon\Delta_{\text{min}}^{-2} + 1
\end{aligned}$$

where the last line follows from our lower bound assumption on T. ■

Corollary 38 *We have that* $Q^* \in \mathcal{U}'$.

Proof This immediately follows from Proposition 37 and the fact that $Q^* \in \mathcal{U}_0$:

$$n_T(Q^*) > \frac{36}{25}\epsilon\Delta_{\text{min}}^{-2} + 1 \geq n_0(Q^*) + 1$$

■

We are then able to show that, by the end of the game, every arm has sufficiently many samples.

Proposition 39

$$n_T(Q) \geq \frac{4}{25}\epsilon\Delta^{-2}(Q) \quad \forall Q \in \mathcal{U}\setminus\{Q^*\}$$

Proof Let $Q \in \mathcal{U}\setminus\{Q^*\}$. Since $Q^* \in \mathcal{U}'$ by Corollary 38, let $t \in [T]$ be the last round such that $Q_t = Q^*$. Then,

$$\begin{aligned}
\mu(Q) - \frac{4}{5}\sqrt{\frac{\epsilon}{n_T(Q)}} &\geq \mu(Q) - \frac{4}{5}\sqrt{\frac{\epsilon}{n_{t-1}(Q)}} \\
&> \text{LCB}_{t-1}(Q;\epsilon) && \text{Eq. (8)} \\
&\geq \text{LCB}_{t-1}(Q^*;\epsilon) \\
&> \mu^* - \frac{6}{5}\sqrt{\frac{\epsilon}{n_{t-1}(Q^*)}} && \text{Eq. (8)} \\
&= \mu^* - \frac{6}{5}\sqrt{\frac{\epsilon}{n_T(Q^*) - 1}} && n_T(Q^*) = n_t(Q^*) = n_{t-1}(Q^*) + 1 \\
&> \mu^* - \Delta(Q) && \text{Prop. 37 and } \Delta_{\text{min}} \leq \Delta(Q)
\end{aligned}$$

The claim then follows by rearranging the terms. ■

Let $Q \in \mathcal{U}\setminus\{Q^*\}$. From Propositions 37 and 39, we can thus conclude inequalities (9)

$$n_T(Q) \geq \frac{4}{25}\epsilon\Delta^{-2}(Q) \quad \text{and} \quad n_T(Q^*) \geq \frac{36}{25}\epsilon\Delta_{\text{min}}^{-2} + 1 > \frac{4}{25}\epsilon\Delta_{\text{min}}^{-2}$$

■

Appendix E. Proof of Theorem 9

Suppose that we are operating under permutation $(a_1, \ldots, a_l)$ and parameters $(\epsilon_1, \ldots, \epsilon_l)$. To show our desired result, we will define a high-probability event, under which the modified UCB-E analysis ensures the correctness of LCB-DR's decision.

E.1. Concentration inequality

From the boundedness of $r \in [0, 1]$, Hoeffding's inequality implies that

$$\mathbb{P}\left(|\mu(a;Q) - \hat{\mu}_t(a;Q)| < \frac{1}{5}\sqrt{\frac{\epsilon}{t}}\right) \geq 1 - 2\exp\left(-\frac{2\epsilon}{25}\right)$$

for all $a \in \mathcal{A}, Q \in \mathcal{U}, t \in \mathbb{N}$ and $\epsilon \geq 0$. Fix some $j \in [l]$. Then, taking union bounds yields

$$\mathbb{P}\left(\bigcap_{Q \in \mathcal{U}} \bigcap_{t \in [u_j]} \left\{|\mu(a_j;Q) - \hat{\mu}_t(a_j;Q)| < \frac{C_{a_j} \wedge 1}{5}\sqrt{\frac{\epsilon_j}{t}}\right\}\right) \geq 1 - 2ku_j \exp\left(-\frac{2\left(C_{a_j}^2 \wedge 1\right)\epsilon_j}{25}\right)$$

We then define the high-probability event of interest:

$$A_j := \bigcap_{Q \in \mathcal{U}} \bigcap_{t \in [u_j]} \left\{|\mu(a_j;Q) - \hat{\mu}_t(a_j;Q)| < \frac{C_{a_j} \wedge 1}{5}\sqrt{\frac{\epsilon_j}{t}}\right\}$$

E.2. Modified UCB-E analysis

Here, we apply the UCB-E analysis of Appendix D. By assumption, recall that $\bar{T}_j \leq u_j$, so that $n_{\bar{T}_{j-1}}(Q) + T_j \leq u_j$ and, thus, under event A_j,

$$|\mu(a_j;Q) - \hat{\mu}_t(a_j;Q)| < \frac{C_{a_j} \wedge 1}{5}\sqrt{\frac{\epsilon_j}{t}} \leq \frac{1}{5}\sqrt{\frac{\epsilon_j}{t}}$$

for all $Q \in \mathcal{U}$ and $t \in \left\{n_{\bar{T}_{j-1}}(Q), \ldots, n_{\bar{T}_{j-1}}(Q) + T_j\right\}$. We can then conclude the following result.

Theorem 40 *For any $j \in [l]$, under event A_j, it follows that $\hat{Q}_j = Q^*_{a_j}$ and*

$$|\mu_{\text{DR}}(a_j) - \mu^o_T(a_j)| < \begin{cases} \dfrac{\Delta_{\text{DR}}(a_j)}{2} & a_j \neq a^* \\ \dfrac{\Delta_{\text{DR,min}}}{2} & a_j = a^* \end{cases}$$

Proof If we set $T = T_j, \epsilon = \epsilon_j, n_0 = n_{\bar{T}_{j-1}}, \mu = \mu(a_j;\cdot)$ and $\hat{\mu}_t = \hat{\mu}_t(a_j;\cdot)$ in the setup of Appendix D, then we can immediately see that $\hat{Q}_j = Q^*_{a_j}$ by Theorem 32, as its assumptions are satisfied under A_j. Moreover, we have that

$$\begin{aligned}
\left|\mu_{\mathrm{DR}}(a_j) - \mu^{\mathrm{o}}_T(a_j)\right| &= \left|\mu\left(a_j, Q^*_{a_j}\right) - \hat{\mu}_{n_{\bar{T}_j}\left(Q^*_{a_j}\right)}\left(a_j, Q^*_{a_j}\right)\right| && \hat{Q}_j = Q^*_{a_j} \\
&< \frac{C_{a_j} \wedge 1}{5}\sqrt{\frac{\epsilon_j}{n_{\bar{T}_j}\left(Q^*_{a_j}\right)}} && \text{event } A_j \text{ and } n_{\bar{T}_j} \le u_j \\
&\le C_{a_j}\frac{\Delta_{a_j,\min}}{2} && \text{Thm. 32} \\
&= \begin{cases} \frac{\Delta_{\mathrm{DR}}(a_j)}{2} & a_j \neq a^* \\ \frac{\Delta_{\mathrm{DR},\min}}{2} & a_j = a^* \end{cases}
\end{aligned}$$

■

Let us also remark that when $T_j = \frac{36}{25}\epsilon_j H_j - \tilde{T}_j + k_j$ (i.e., the lower bound (1) holds with equality), then

$$\begin{aligned}
\bar{T}_j &= \sum_{r=0}^{j} T_r \\
&= k + \sum_{r=1}^{j}\left[\frac{36}{25}\epsilon_r \underbrace{H_r}_{\le 2H_{a_r}} \underbrace{-\tilde{T}_r}_{\le 0} + \underbrace{k_r}_{\le k}\right] \\
&\le k(j+1) + \frac{72}{25}\sum_{r=1}^{j}\epsilon_r H_{a_r}
\end{aligned}$$

so that we can set u_j to the last expression.

E.3. LCB-DR correctness

Under the event $\bigcap_{j=1}^{l} A_j$, we know that

$$\begin{aligned}
\mu^{\mathrm{o}}_T(a^*) - \mu^{\mathrm{o}}_T(a) &= \mu^{\mathrm{o}}_T(a^*) - \overset{*}{\mu}_{\mathrm{DR}} + \Delta_{\mathrm{DR}}(a) + \mu_{\mathrm{DR}}(a) - \mu^{\mathrm{o}}_T(a) \\
&> \Delta_{\mathrm{DR}}(a) - \frac{\Delta_{\mathrm{DR},\min}}{2} - \frac{\Delta_{\mathrm{DR}}(a)}{2} && \text{Thm. 40} \\
&\ge 0 && \Delta_{\mathrm{DR},\min} \le \Delta_{\mathrm{DR}}(a)
\end{aligned}$$

for every $a \neq a^*$. That is, $A^{\mathrm{o}}_T = \operatorname{argmax}_{a\in\mathcal{A}} \mu^{\mathrm{o}}_T(a) = a^*$ and, thus, $\mathbb{P}(A^{\mathrm{o}}_T = a^*) \ge \mathbb{P}\left(\bigcap_{j=1}^{l} A_j\right)$. The result then follows from a union bound on the high-probability events $\{A_j\}_{j=1}^{l}$.

Appendix F. Extending to infinite decision sets

While the results discussed thus far only apply to finite decision sets $\mathcal{A}$, it is possible to extend to larger (possibly infinite) sets via standard covering arguments. Suppose that we have access to a finite ϵ-cover $\mathcal{A}_\epsilon$ of $\{r(a,\cdot)\}_{a\in\mathcal{A}}$ in the following sense: for all $a \in \mathcal{A}$, there exists a $\phi_a \in \mathcal{A}_\epsilon$ such that

$$\max_{Q\in\mathcal{U}} \mathbb{E}_{X\sim Q}\left[|r(a,X) - r(\phi_a, X)|\right] \le \epsilon$$

The idea is that the regret under the finite set $\mathcal{A}_\epsilon$ is close to the regret under $\mathcal{A}$, so that a learner can play the game dynamics on the former.

Lemma 41 (Controlling regret using a cover) *Let* $\Delta_{\mathrm{DR}}(a;\mathcal{A}_\epsilon) := \max_{a^*\in\mathcal{A}_\epsilon} \mu_{\mathrm{DR}}(a^*) - \mu_{\mathrm{DR}}(a)$ *denote the suboptimality gap with respect to* $\mathcal{A}_\epsilon$*. Then,* $\Delta_{\mathrm{DR}}(a) \le \Delta_{\mathrm{DR}}(a;\mathcal{A}_\epsilon) + \epsilon$ *for all* $a \in \mathcal{A}$*.*

Proof We can relate this new gap to the quantity of interest by noting that for any $a \in \mathcal{A}$,

$$\begin{aligned}
\Delta_{\mathrm{DR}}(a;\mathcal{A}) &= \max_{a^*\in\mathcal{A}} \mu_{\mathrm{DR}}(a^*) - \mu_{\mathrm{DR}}(a) \\
&= \max_{a^*\in\mathcal{A}} \mu_{\mathrm{DR}}(a^*) - \max_{a^*_\epsilon\in\mathcal{A}_\epsilon} \mu_{\mathrm{DR}}(a^*_\epsilon) + \Delta_{\mathrm{DR}}(a;\mathcal{A}_\epsilon) \\
&= \max_{a^*\in\mathcal{A}} \left\{\mu_{\mathrm{DR}}(a^*) - \max_{a^*_\epsilon\in\mathcal{A}_\epsilon} \mu_{\mathrm{DR}}(a^*_\epsilon)\right\} + \Delta_{\mathrm{DR}}(a;\mathcal{A}_\epsilon)
\end{aligned}$$

We can bound the error term as follows: for any $a \in \mathcal{A}$,

$$\begin{aligned}
\mu_{\mathrm{DR}}(a) - \max_{a^*_\epsilon\in\mathcal{A}_\epsilon} \mu_{\mathrm{DR}}(a^*_\epsilon) &\le \mu_{\mathrm{DR}}(a) - \mu_{\mathrm{DR}}(\phi_a) \\
&= \min_{Q\in\mathcal{U}} \mathbb{E}_{X\sim Q}\left[r(a,X)\right] - \min_{Q\in\mathcal{U}} \mathbb{E}_{X\sim Q}\left[r(\phi_a,X)\right] \\
&\le \max_{Q\in\mathcal{U}} \mathbb{E}_{X\sim Q}\left[r(a,X) - r(\phi_a,X)\right] \\
&\le \epsilon
\end{aligned}$$

That is,

$$\Delta_{\mathrm{DR}}(a;\mathcal{A}) \le \Delta_{\mathrm{DR}}(a;\mathcal{A}_\epsilon) + \epsilon \quad \forall a \in \mathcal{A}$$

■

F.1. Binary classification

A special case is the binary classification setting:

- The data are pairs $(X,Y) \in \mathcal{X} \times \{0,1\}$.
- Decisions are binary-valued functions $a : \mathcal{X} \to \{0,1\}$ and $\mathrm{VC}(\mathcal{A}) = d < \infty$.
- The reward function is $r(a,(x,y)) = \mathbb{I}\{a(x) = y\}$, so that

$$\mathbb{E}_{(X,Y)\sim Q}\left[r(a,(X,Y))\right] = \mathbb{P}_{(X,Y)\sim Q}\left(a(X) = Y\right)$$

Suppose that we have a finite ϵ-cover $\mathcal{A}_\epsilon$ of $\mathcal{A}$ in the following sense: for any $a \in \mathcal{A}$, there exists a $\phi_a \in \mathcal{A}_\epsilon$ such that

$$\max_{Q \in \mathcal{U}} Q_{\mathcal{X}} \left(a \neq \phi_a\right) \leq \epsilon$$

where $Q_{\mathcal{X}}$ is the marginal distribution of Q over $\mathcal{X}$ (recall that now the distributions are over pairs $(X, Y) \in \mathcal{X} \times \{0, 1\}$. To see why this yields a cover in the more general definition, we note that for any $(X, Y) \sim Q$,

$$\begin{aligned} \mathbb{E}\left[\left|r\left(a, (X, Y)\right) - r\left(\phi_a, (X, Y)\right)\right|\right] &= \mathbb{E}\left[\left|\mathbb{I}\left\{a\left(X\right) = Y\right\} - \mathbb{I}\left\{\phi_a\left(X\right) = Y\right\}\right|\right] \\ &= \mathbb{E}\left[\left|\mathbb{I}\left\{\phi_a\left(X\right) \neq Y\right\} - \mathbb{I}\left\{a\left(X\right) \neq Y\right\}\right|\right] \\ &= \mathbb{E}\left[\left|\left(\phi_a\left(X\right) - Y\right)^2 - \left(a\left(X\right) - Y\right)^2\right|\right] \end{aligned}$$

Next, we make the following elementary observation: for $x, y, z \in \{0, 1\}$, we have that

$$(x - z)^2 - (y - z)^2 = (x - y)^2 - 2 \underbrace{(x - y)(z - y)}_{\geq 0} \leq (x - y)^2$$

By symmetry, it then follows that $\left|(x - z)^2 - (y - z)^2\right| \leq (x - y)^2$. Applying this to the above then yields

$$\mathbb{E}\left[\left|\left(\phi_a\left(X\right) - Y\right)^2 - \left(a\left(X\right) - Y\right)^2\right|\right] \leq \mathbb{E}\left[\left(\phi_a\left(X\right) - a\left(X\right)\right)^2\right] \leq \epsilon$$

That is, we ensure the condition $\Delta_{\mathrm{DR}}\left(a; \mathcal{A}\right) \leq \Delta_{\mathrm{DR}}\left(a; \mathcal{A}_\epsilon\right) + \epsilon$ for all $a \in \mathcal{A}$.

Finally, let us briefly discuss how to construct such a cover from samples. Suppose that we independently sample $O\left(\frac{d \log(1/\epsilon) + \log(k/\delta)}{\epsilon}\right)$ times from each distribution $Q \in \mathcal{U}$, and let S denote the aggregated sample. Then, let $\mathcal{A}_\epsilon \subset \mathcal{A}$ be the result of selecting one representative from each of the sets in

$$\left\{\left\{a \in \mathcal{A} : a|_S = I\right\} : I \in \{0, 1\}^{|S|}\right\}$$

That is, for each $a \in \mathcal{A}$, there exists some $a_\epsilon \in \mathcal{A}_\epsilon$ such that a and a_ϵ agree on S. The set $\mathcal{A}_\epsilon$ is an ϵ-cover as defined previously. In addition, from the Sauer-Shelah lemma (van Handel, 2014, Lemma 7.12), we know that

$$|\mathcal{A}_\epsilon| \lesssim \left(\frac{k\left(d \log\left(1/\epsilon\right) + \log\left(k/\delta\right)\right)}{d\epsilon}\right)^d$$

Appendix G. UE v.s. NUE

Here, we will prove the bounds stated in Section 3.4. For convenience, we present the variance quantities again below:

$$V_T = \sum_{j=1}^{k} \left(n_{(j)} - n_{(j-1)}\right) \mathbb{E}\left[\max_{r\in\{j,\ldots,k\}} \frac{1}{n_{(r)}^2} \left[X_{Q_{(r)}} - \mu_{Q_{(r)}}\right]^2\right]$$

$$\Sigma_T^2 = \mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \left(X_Q^{(i)} - \mu_Q\right)^2\right]$$

$$\sigma_T^2 = \max_{Q\in\mathcal{U}} \frac{\sigma_Q^2}{n_Q}$$

We begin by proving the bound on Σ_T^2.

Lemma 42 *Suppose that our data is bounded: $X_Q \in [0,1]$. Then,*

$$\Sigma_T^2 \leq 8\sqrt{\frac{2\log(2k)}{\min_{Q\in\mathcal{U}} n_Q^3}} + \sigma_T^2$$

Proof Recall that

$$\Sigma_T^2 = \mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} Y_{i,Q}^2\right]$$

where we define $Y_{i,Q} \coloneqq X_Q^{(i)} - \mu_Q \in [-1,1]$ and note that $\mathbb{E}\left[Y_{i,Q}^2\right] = \sigma_Q^2$. Let us begin by noting that

$$\Sigma_T^2 \leq \mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \left(Y_{i,Q}^2 - \mathbb{E}\left[Y_{i,Q}^2\right]\right)\right] + \sigma_T^2$$

For a one-sided symmetrization argument, let $Z_{i,Q}$ be independent copies of the $Y_{i,Q}$ and let $\epsilon^n \overset{iid}{\sim}$ Rad be independent from them, where $n \coloneqq \max_{Q\in\mathcal{U}} n_Q$. Then, we can bound the first quantity in the upper bound as follows:

$$\begin{aligned}
\mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \left(Y_{i,Q}^2 - \mathbb{E}\left[Y_{i,Q}^2\right]\right)\right] &= \mathbb{E}\left[\max_{Q\in\mathcal{U}} \mathbb{E}\left[\frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \left(Y_{i,Q}^2 - Z_{i,Q}^2\right)\middle| Y\right]\right] \\
&\leq \mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \left(Y_{i,Q}^2 - Z_{i,Q}^2\right)\right] \\
&= \mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \epsilon_i \left(Y_{i,Q}^2 - Z_{i,Q}^2\right)\right] \\
&\leq \mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \epsilon_i Y_{i,Q}^2\right] + \mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} -\epsilon_i Z_{i,Q}^2\right] \\
&= 2\mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \epsilon_i Y_{i,Q}^2\right]
\end{aligned}$$

where Y denotes the collection of all $Y_{i,Q}$'s. In the next lemma, we bound the last quantity above.

Lemma 43 (Contraction) *We have that*

$$\mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \epsilon_i Y_{i,Q}^2\right] \leq \mathbb{E}\left[\max_{Q\in\mathcal{U}} C_Q \sum_{i=1}^{n_Q} \epsilon_i Y_{i,Q}\right]$$

where $C_Q := \frac{2}{n_Q \cdot \min_{Q'\in\mathcal{U}} n_{Q'}}$.

Proof [Proof of Lemma 43] Fix an index $j \in [n]$, where $n := \max_{Q\in\mathcal{U}} n_Q$. For each $Q \in \mathcal{U}$, let us additionally define dummy variables $Y_{n_Q+1,Q}, \ldots, Y_{n,Q} := 0$, so that

$$\mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} \epsilon_i Y_{i,Q}^2\right] = \mathbb{E}\left[\max_{Q\in\mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n} \epsilon_i Y_{i,Q}^2\right]$$

In what follows, we use $\mathbb{E}_{\epsilon_j}$ to denote an expectation only w.r.t. ϵ_j, while all other random variables remain fixed (that is, conditioned on all other variables due to independence). Note that

$$\begin{aligned}
\mathbb{E}_{\epsilon_j} &\left[\max_{Q\in\mathcal{U}} \left\{ \frac{1}{n_Q^2} \sum_{i=1}^{j} \epsilon_i Y_{i,Q}^2 + C_Q \sum_{i=j+1}^{n} \epsilon_i Y_{i,Q} \right\}\right] \\
&= \frac{1}{2} \max_{Q,Q'\in\mathcal{U}} \left\{ \frac{1}{n_Q^2} \sum_{i=1}^{j-1} \epsilon_i Y_{i,Q}^2 + \frac{Y_{j,Q}^2}{n_Q^2} + C_Q \sum_{i=j+1}^{n} \epsilon_i Y_{i,Q} \right. \\
&\qquad \left. + \frac{1}{n_{Q'}^2} \sum_{i=1}^{j-1} \epsilon_i Y_{i,Q'}^2 - \frac{Y_{j,Q'}^2}{n_{Q'}^2} + C_{Q'} \sum_{i=j+1}^{n} \epsilon_i Y_{i,Q'} \right\}
\end{aligned}$$

Next, note that

$$\begin{aligned}
\frac{Y_{j,Q}^2}{n_Q^2} - \frac{Y_{j,Q'}^2}{n_{Q'}^2} &= \left(\frac{Y_{j,Q}}{n_Q} + \frac{Y_{j,Q'}}{n_{Q'}}\right)\left(\frac{Y_{j,Q}}{n_Q} - \frac{Y_{j,Q'}}{n_{Q'}}\right) \\
&\leq \left(\frac{1}{n_Q} + \frac{1}{n_{Q'}}\right)\left|\frac{Y_{j,Q}}{n_Q} - \frac{Y_{j,Q'}}{n_{Q'}}\right| \\
&\leq \left|C_Q Y_{j,Q} - C_{Q'} Y_{j,Q'}\right|
\end{aligned}$$

Hence,

$$
\begin{aligned}
\mathbb{E}_{\epsilon_j}&\left[\max_{Q\in\mathcal{U}}\left\{\frac{1}{n_Q^2}\sum_{i=1}^{j}\epsilon_i Y_{i,Q}^2 + C_Q\sum_{i=j+1}^{n}\epsilon_i Y_{i,Q}\right\}\right] \\
&\le \frac{1}{2}\max_{Q,Q'\in\mathcal{U}}\left\{\frac{1}{n_Q^2}\sum_{i=1}^{j-1}\epsilon_i Y_{i,Q}^2 + C_Q\sum_{i=j+1}^{n}\epsilon_i Y_{i,Q}\right. \\
&\qquad\left.+\frac{1}{n_{Q'}^2}\sum_{i=1}^{j-1}\epsilon_i Y_{i,Q'}^2 + C_{Q'}\sum_{i=j+1}^{n}\epsilon_i Y_{i,Q'} + \left|C_Q Y_{j,Q} - C_{Q'} Y_{j,Q'}\right|\right\} \\
&= \frac{1}{2}\max_{Q,Q'\in\mathcal{U}}\left\{\frac{1}{n_Q^2}\sum_{i=1}^{j-1}\epsilon_i Y_{i,Q}^2 + C_Q\sum_{i=j+1}^{n}\epsilon_i Y_{i,Q}\right. \\
&\qquad\left.+\frac{1}{n_{Q'}^2}\sum_{i=1}^{j-1}\epsilon_i Y_{i,Q'}^2 + C_{Q'}\sum_{i=j+1}^{n}\epsilon_i Y_{i,Q'} + C_Q Y_{j,Q} - C_{Q'} Y_{j,Q'}\right\} \\
&= \mathbb{E}_{\epsilon_j}\left[\max_{Q\in\mathcal{U}}\left\{\frac{1}{n_Q^2}\sum_{i=1}^{j-1}\epsilon_i Y_{i,Q}^2 + C_Q\sum_{i=j}^{n}\epsilon_i Y_{i,Q}\right\}\right]
\end{aligned}
$$

From independence, we can thus integrate iteratively starting at $j = n$ to conclude that

$$
\mathbb{E}\left[\max_{Q\in\mathcal{U}}\frac{1}{n_Q^2}\sum_{i=1}^{n}\epsilon_i Y_{i,Q}^2\right] \le \mathbb{E}\left[\max_{Q\in\mathcal{U}} C_Q\sum_{i=1}^{n}\epsilon_i Y_{i,Q}\right] = \mathbb{E}\left[\max_{Q\in\mathcal{U}} C_Q\sum_{i=1}^{n_Q}\epsilon_i Y_{i,Q}\right]
$$

■

Again using symmetrization, let $Z_{i,Q}$ be independent copies of the $Y_{i,Q}$ and independent from ϵ^n. Since $Y_{i,Q}$ are centered, we have that

$$
\begin{aligned}
\mathbb{E}\left[\max_{Q\in\mathcal{U}} C_Q\sum_{i=1}^{n_Q}\epsilon_i Y_{i,Q}\right] &= \mathbb{E}\left[\max_{Q\in\mathcal{U}}\mathbb{E}\left[C_Q\sum_{i=1}^{n_Q}\epsilon_i\left(Y_{i,Q} - Z_{i,Q}\right)\middle|\epsilon^n, Y\right]\right] \\
&\le \mathbb{E}\left[\max_{Q\in\mathcal{U}} C_Q\sum_{i=1}^{n_Q}\epsilon_i\left(Y_{i,Q} - Z_{i,Q}\right)\right] \\
&= \mathbb{E}\left[\max_{Q\in\mathcal{U}} C_Q\sum_{i=1}^{n_Q}\left(Y_{i,Q} - Z_{i,Q}\right)\right] \\
&\le 2\mathbb{E}\left[\max_{Q\in\mathcal{U}} C_Q\left|\sum_{i=1}^{n_Q} Y_{i,Q}\right|\right]
\end{aligned}
$$

Next, we bound this expectation using Hoeffding's inequality. We begin with a high-probability bound:

$$\begin{aligned}
\mathbb{P}\left(\max_{Q\in\mathcal{U}} C_Q \left|\sum_{i=1}^{n_Q} Y_{i,Q}\right| \geq t\right) &\leq \sum_{Q\in\mathcal{U}} \mathbb{P}\left(C_Q \left|\sum_{i=1}^{n_Q} Y_{i,Q}\right| \geq t\right) \\
&\leq 2\sum_{Q\in\mathcal{U}} \exp\left(-\frac{2t^2}{C_Q^2 n_Q}\right) \\
&= 2\sum_{Q\in\mathcal{U}} \exp\left(-\frac{t^2 n_Q \min_{Q'\in\mathcal{U}} n_{Q'}^2}{2}\right) \\
&\leq 2k\exp\left(-\frac{t^2 \min_{Q\in\mathcal{U}} n_Q^3}{2}\right)
\end{aligned}$$

We can subsequently integrate the tails to obtain the in-expectation bound

$$\mathbb{E}\left[\max_{Q\in\mathcal{U}} C_Q \left|\sum_{i=1}^{n_Q} Y_{i,Q}\right|\right] \leq 2\sqrt{\frac{2\log(2k)}{\min_{Q\in\mathcal{U}} n_Q^3}}$$

Combining all bounds presented thus far finally yields

$$\Sigma_T^2 \leq 8\sqrt{\frac{2\log(2k)}{\min_{Q\in\mathcal{U}} n_Q^3}} + \sigma_T^2$$

■

Next, we show how V_T relates to Σ_T^2.

Lemma 44 *We have that*

$$V_T \leq \min\left\{\max_{Q\in\mathcal{U}} n_Q, k\right\} \Sigma_T^2$$

Proof Let $n := \max_{Q\in\mathcal{U}} n_Q$ and note that we can equivalently express

$$V_T = \sum_{i=1}^{n} \mathbb{E}\left[\max_{Q\in\mathcal{U}: n_Q\geq i} \frac{1}{n_Q^2}\left(X_Q^{(i)} - \mu_Q\right)^2\right]$$

From this, we see that

$$\begin{aligned}
V_T &\leq n\mathbb{E}\left[\max_{i\in[n]} \max_{Q\in\mathcal{U}: n_Q\geq i} \frac{1}{n_Q^2}\left(X_Q^{(i)} - \mu_Q\right)^2\right] \\
&= n\mathbb{E}\left[\max_{Q\in\mathcal{U}} \max_{i\in[n_Q]} \frac{1}{n_Q^2}\left(X_Q^{(i)} - \mu_Q\right)^2\right] \\
&\leq n\mathbb{E}\left[\max_{Q\in\mathcal{U}} \sum_{i=1}^{n_Q} \frac{1}{n_Q^2}\left(X_Q^{(i)} - \mu_Q\right)^2\right] \\
&= n\Sigma_T^2
\end{aligned}$$

Alternatively, we can begin by bounding the max by a sum in V_T:

$$\begin{aligned} V_T &\le \mathbb{E}\left[\sum_{i=1}^{n} \sum_{Q\in\mathcal{U}:n_Q\ge i} \frac{1}{n_Q^2}\left(X_Q^{(i)}-\mu_Q\right)^2\right] \\ &= \mathbb{E}\left[\sum_{Q\in\mathcal{U}} \sum_{i=1}^{n_Q} \frac{1}{n_Q^2}\left(X_Q^{(i)}-\mu_Q\right)^2\right] \\ &\le k\mathbb{E}\left[\max_{Q\in\mathcal{U}} \sum_{i=1}^{n_Q} \frac{1}{n_Q^2}\left(X_Q^{(i)}-\mu_Q\right)^2\right] \\ &= k\Sigma_T^2 \end{aligned}$$

■

Finally, we prove the upper bound on V_T stated in the example of Section 3.4.3.

Lemma 45 *Let $\mathcal{U} = \{Q_1, \ldots, Q_k\}$, where*

- *$Q_1, \ldots, Q_{k-1}$ share a common variance σ^2 and are supported in $[0,1]$.*
- *Q_k has variance ν^2.*
- *We sample n times from each $Q_1, \ldots, Q_{k-1}$ and $m = T - n(k-1) \ge n$ times from Q_k, for a total of $T \ge nk$ samples.*

Then,

$$V_T \le \frac{\sqrt{2\log(k-1)} + \sigma^2}{n} + \frac{\nu^2}{T - n(k-1)}$$

Proof Note that

$$\begin{aligned} V_T &= n\mathbb{E}\left[\max\left\{\max_{j\in[k-1]}\left\{\frac{1}{n^2}\left(X_{Q_j}-\mu_{Q_j}\right)^2\right\}, \frac{1}{m^2}\left(X_{Q_k}-\mu_{Q_k}\right)^2\right\}\right] + \frac{(m-n)\,\nu^2}{m^2} \\ &\le \frac{1}{n}\mathbb{E}\left[\max_{j\in[k-1]}\left\{\left(X_{Q_j}-\mu_{Q_j}\right)^2\right\}\right] + \frac{n\nu^2}{m^2} + \frac{(m-n)\,\nu^2}{m^2} \\ &= \frac{1}{n}\underbrace{\mathbb{E}\left[\max_{j\in[k-1]}\left\{\left(X_{Q_j}-\mu_{Q_j}\right)^2\right\}\right]}_{(*)} + \frac{\nu^2}{m} \end{aligned}$$

Since $\left|\left(X_{Q_j}-\mu_{Q_j}\right)^2 - \sigma^2\right| \le 1$ for all $j \in [k-1]$, we then have that

$$(*) = \mathbb{E}\left[\max_{j\in[k-1]}\left\{\left(X_{Q_j}-\mu_{Q_j}\right)^2 - \sigma^2\right\}\right] + \sigma^2 \le \sqrt{2\log(k-1)} + \sigma^2$$

■