

Distribution-Dependent Rates for Multi-Distribution Learning

Rafael Hanashiro * Patrick Jaillet †

December 21, 2023

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 [AHZ23] 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.

Contents

1	Introduction	2
1.1	Main results	3
1.2	Related work	4
2	Preliminaries	5
2.1	Multi-distribution learning	5
2.2	Complexity measures	6
2.3	Algorithmic tools	6
3	Non-adaptive strategies	7
3.1	Uniform exploration (UE)	7
3.2	Non-uniform exploration (NUE)	7
4	Optimism	9
4.1	Comparison with UE	12
5	Discussion	12

*MIT, rafah@mit.edu

†MIT, jaillet@mit.edu

A	Expectaion of empirical process maximum	16
A.1	Symmetrization	16
A.2	Bounds on the Rademacher complexity	18
B	Empirical process concentration inequalities	23
B.1	McDiarmid	23
B.2	Bernstein	25
B.2.1	Preliminaries	25
B.2.2	Bounding V^+	28
B.2.3	Efron-Stein	30
B.2.4	Self-boundedness	30
B.2.5	Sub-gamma tail	31
B.2.6	Original setting	32
C	Proofs of Section 3	33
C.1	Proof of Theorem 3.1	34
C.2	Proof of Corollary 3.1	35
C.3	Proof of Theorem 3.2	37
D	Modified UCB-E	38
E	Proof of Theorem 4.1	42
E.1	Concentration inequality	42
E.2	Modified UCB-E analysis	43
E.3	LCB-DR correctness	44

1 Introduction

Classical statistical learning operates under the assumption that data comes from a single source [HTF09]. 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)* [RM22], 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* $\mathcal{U} \subset \mathcal{P}(\mathcal{X})$. Then, under a target reward function $r : \mathcal{A} \times \mathcal{X} \rightarrow \mathbb{R}$, DRO focuses on the problem

$$\max_{a \in \mathcal{A}} \left\{ \mu_{\text{DR}}(a) := \min_{Q \in \mathcal{U}} \mathbb{E}_{X \sim Q} [r(a, X)] \right\} \tag{DR}$$

Recent works [BHPQ17, SKHL20, HJZ22] 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 [AHZ23] 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 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)* [BMS11] 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 $O\left(\frac{1}{\sqrt{T}}\right)$. From a novel 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 [ABM10] 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 $\Delta_{\text{DR}}(a) := \max_{a' \in \mathcal{A}} \mu_{\text{DR}}(a') - \mu_{\text{DR}}(a)$ be the suboptimality gap of an action from a finite set \mathcal{A} . Given an algorithm, we denote its output after T sampling rounds by A_T° . In short, we make the following contributions:

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

$$\mathbb{E} [\Delta_{\text{DR}}(A_T^{\circ})] \lesssim \sum_{a \in \mathcal{A}: \Delta_{\text{DR}}(a) > 0} \Delta_{\text{DR}}(a) \exp(-n \Delta_{\text{DR}}^2(a))$$

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

- With $n_Q \in \mathbb{N}$ samples from distribution $Q \in \mathcal{U}$ over real-valued data, we show that NUE attains the rate

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

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 .

- Appealing to the principle of optimism, we devise the LCB-DR algorithm that, in a pre-specified permutation of the arms $(a_1, a_2, \dots, a_{|\mathcal{A}|})$, for $j = 1, \dots, |\mathcal{A}|$, sequentially performs a modified version of UCB-E, for T_j rounds, on “losses” $\{r(a_j, X_Q)\}_{Q \in \mathcal{U}}$, where $X_Q \sim Q$, as a means of identifying the worst-case distribution for a_j . We show that this guarantees an error probability of

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

Potentially of independent interest, this bound results from an analysis of UCB-E under a learner with previously acquired data (see Appendix D). This produces the additional \tilde{T}_j term, denoting the count of pre-collected samples from a subset $\mathcal{U}_j \subset \mathcal{U}$ of the distributions. It also enhances the complexity measure H_j , which aggregates the distribution suboptimality gaps $\Delta_{a_j}(Q) := \mathbb{E}[r(a_j, X_Q)] - \min_{Q' \in \mathcal{U}} \mathbb{E}[r(a_j, X_{Q'})]$, with $\Delta_{a_j, \min}$ being the smallest positive one, only over $Q \in \mathcal{U}_j$ rather than all of \mathcal{U} , as is found in the standard analysis. Mirroring the MAB literature, we contrast this bound with that of UE, showing that the comparison is captured by C_a , defined in Section 2.2.

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 appropriate 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 [BDBC⁺09, MMR09], where training data may not necessarily be sampled from the target distribution. To tackle this, several works [VNS⁺18, ZMV⁺21, SKK21] have applied the framework of DRO [Sca58, DY10, BTdHW⁺13] 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 [SDR21, RM22] 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 [MSS19] 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 [BHPQ17] focuses on binary classification, with the aim of guaranteeing $\mathbb{P}(\Delta_{\text{DR}}(A_T^\circ) \leq \epsilon) \geq 1 - \delta$ under a minimal number of samples T . The original work of [BHPQ17] assumes realizability and subsequent studies [CZZ18, NZ18, CH22, HJZ22] extended results to the agnostic case and gave improved rates, along with sample-complexity lower bounds. [AHZ23] later solidified the theory and posed several open problems, some of which were recently addressed in [Pen23, ZZC⁺23] via optimal algorithms.

In this work, we turn our attention to the simple regret $\mathbb{E}[\Delta_{\text{DR}}(A_T^\circ)]$. For finite decision sets \mathcal{A} , an integration of the tails reveals that the regret achieved by [HJZ22] 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 [SKHL20] ensures a rate of $O\left(k\sqrt{\frac{B^2+\log k}{T}}\right)$ and, in the fairness context, [AAK⁺22] obtains $O\left(\frac{B}{\sqrt{T}}\right)$ plus a term that uniformly bounds the generalization gap with high-probability. Subsequently, [SGJ22] was able to attain $O\left(\sqrt{\frac{B^2+k}{T}}\right)$, showing a matching lower bound, and [ZZYZ23] devised strategies with $O\left(\sqrt{\frac{B^2+k\log k}{T}}\right)$ regret 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 [BMS11] 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 [LS20, Ch. 33] for an in-depth discussion. In PE-MAB, [ABM10] 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 $r : \mathcal{A} \times \mathcal{X} \rightarrow \mathbb{R}$ 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 (DR) problem

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

In particular, 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 $[0, 1]$.

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 $A_T^\circ \in \mathcal{A}$ with the goal of minimizing the *simple regret* $\mathbb{E} [\Delta_{\text{DR}} (A_T^\circ)]$ or *error probability* $\mathbb{P} (\Delta_{\text{DR}} (A_T^\circ) > 0)$. The strategies described in this work are of the form $A_T^\circ = \operatorname{argmax}_{a \in \mathcal{A}} \mu_T^\circ (a)$ for an appropriately constructed proxy $\mu_T^\circ : \mathcal{A} \rightarrow \mathbb{R}$.

Remark 2.1: Simple regret v.s. error probability

Note that both performance measures are closely related: since $r \in [0, 1]$, we have that $\Delta_{\text{DR}} \in [0, 1]$ and, thus,

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

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

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

These quantities are additionally used to define complexity measures

$$H_a := \sum_{Q \in \mathcal{U} : \Delta_a(Q) > 0} \Delta_a^{-2} (Q) \quad \text{and} \quad C_a := \begin{cases} \frac{\Delta_{\text{DR}} (a)}{\Delta_{a, \min}} & a \neq a^* \\ \frac{\Delta_{\text{DR}, \min}}{\Delta_{a^*, \min}} & a = a^* \end{cases}$$

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., [ABM10]), 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 $a \neq a^*$, or $\Delta_{\text{DR}} (a) \leq \Delta_{a, \min}$, 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 $X_Q, \{X_Q^{(i)}\}_{i=1}^\infty \stackrel{iid}{\sim} Q$ 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

$$\hat{\mu}_t (a; Q) := \frac{1}{t} \sum_{i=1}^t r (a, X_Q^{(i)})$$

Under a fixed sampling algorithm, let $n_t (Q) := \sum_{s=1}^t \mathbb{I} \{Q_s = Q\}$ 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° 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

$$\mu_T^\circ(a) = \min_{Q \in \mathcal{U}} \hat{\mu}_n(a; Q)$$

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 $\mu_T^\circ(a) = \min_{Q \in \mathcal{U}} \hat{\mu}_n(a; Q)$

Output: $A_T^\circ = \operatorname{argmax}_{a \in \mathcal{A}} \mu_T^\circ(a)$

Theorem 3.1: UE regret

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

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

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

Corollary 3.1: UE distribution-independent regret

Suppose that $n \geq \left(\frac{8}{\Delta_{\text{DR}, \min}}\right)^2 \log k$. Then, the UE algorithm attains the following distribution-independent simple regret bound:

$$\mathbb{E} [\Delta_{\text{DR}}(A_T^\circ)] \lesssim \sqrt{\frac{k \log(kl)}{T}}$$

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

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

Here, we consider real-valued data in $\mathcal{X} \subset \mathbb{R}$ and define mean $\mu_Q := \mathbb{E}[X_Q]$ and variance $\sigma_Q^2 := \text{Var}(X_Q)$ for each $Q \in \mathcal{U}$. Additionally, let us sort the sample sizes in increasing order as follows: $0 =: n_{(0)} \leq n_{(1)} \leq \dots \leq n_{(k)}$ and let $Q_{(j)}$ denote the corresponding distribution in the j th position: $n_{Q_{(j)}} = n_{(j)}$. The regret bound presented will rely on the following variance quantities:

$$\begin{aligned} V_T &:= \sum_{j=1}^k (n_{(j)} - n_{(j-1)}) \mathbb{E} \left[\max_{r \in \{j, \dots, k\}} \frac{1}{n_{(r)}} [X_{Q_{(r)}} - \mu_{Q_{(r)}}]^2 \right] \\ \Sigma_T^2 &:= \mathbb{E} \left[\max_{Q \in \mathcal{U}} \frac{1}{n_Q} \sum_{i=1}^{n_Q} (X_Q^{(i)} - \mu_Q)^2 \right] \\ \sigma_T^2 &:= \max_{Q \in \mathcal{U}} \frac{\sigma_Q^2}{n_Q} \end{aligned}$$

Lastly, we make use of the quantity $G_T := 8 \left(\frac{4 \log k}{\min_{Q \in \mathcal{U}} n_Q} + L \sigma_T \sqrt{2 \log k} \right)$, which we note decreases with the $\{n_Q\}$.

Algorithm 2 Non-uniform exploration (NUE)

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

- 1: Sample n_Q times from each distribution $Q \in \mathcal{U}$
- 2: Construct $\mu_T^\circ(a) = \min_{Q \in \mathcal{U}} \hat{\mu}_{n_Q}(a; Q)$

Output: $A_T^\circ = \text{argmax}_{a \in \mathcal{A}} \mu_T^\circ(a)$

Theorem 3.2: NUE regret

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

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

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.

Remark 3.1: Crude bound

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

Remark 3.2: Uniform v.s. non-uniform exploration

As more data is collected and the terms $\sqrt{\frac{\log k}{n}}$ and G_T become negligibly small, we can more easily compare the regret decay of each strategy: ignoring constants, the probability of selecting a suboptimal arm $a \in \mathcal{A}$ for UE is approximately $\approx \exp(-n\Delta_{\text{DR}}^2(a))$, while, for NUE, it is $\approx \exp\left(-\frac{\Delta_{\text{DR}}^2(a)}{\sigma_T^2 + \Sigma_T^2 + V_T + \frac{\Delta_{\text{DR}}(a)}{\min_{Q \in \mathcal{U}} n_Q}}\right)$. Hence, a non-uniform allocation can prove beneficial when $\sigma_T^2 + \Sigma_T^2 + V_T + \frac{\Delta_{\text{DR}}(a)}{\min_{Q \in \mathcal{U}} n_Q} \ll \frac{1}{n}$.

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 3 (Unique optima). a^* and Q_a^* are the *unique* optimal decision and the *unique* worst-case distribution for $a \in \mathcal{A}$, respectively.

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

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

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 [ABM10] 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, $n_k(Q) := 1$ for each $Q \in \mathcal{U}$. As a result, we define $T_0 := \bar{T}_0 := k$ to be the total number of samples gathered before the game starts. The inputs to the algorithm are a permutation (a_1, \dots, a_l) of \mathcal{A} , dictating the order in which decisions are iterated through, and index parameters $(\epsilon_1, \dots, \epsilon_l)$ satisfying

$$\epsilon_j \geq \frac{25}{36} \Delta_{a_j, \min}^2 (u_{j-1} - 1) \quad (1)$$

where

$$u_0 := k \quad \text{and} \quad u_j := k(j+1) + \frac{36}{25} \sum_{r=1}^j \epsilon_r \left(H_{a_r} + \Delta_{a_r, \min}^{-2} \right)$$

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 the following set as a proxy for the arms that will be played in this round:

$$\mathcal{U}_j := \left\{ Q \in \mathcal{U} \setminus \{Q_{a_j}^*\} : n_{\bar{T}_{j-1}}(Q) < \frac{36}{25} \epsilon_j \Delta_{a_j}^{-2}(Q) \right\} \cup \{Q_{a_j}^*\}$$

Additionally, define

$$k_j := |\mathcal{U}_j|, \quad \tilde{T}_j := \sum_{Q \in \mathcal{U}_j} n_{\bar{T}_{j-1}}(Q) \quad \text{and} \quad H_j := \sum_{Q \in \mathcal{U}_j \setminus \{Q_{a_j}^*\}} \Delta_{a_j}^{-2}(Q)$$

2. Allocate

$$T_j := \frac{36}{25} \epsilon_j \left(H_j + \Delta_{a_j, \min}^{-2} \right) - \tilde{T}_j + k_j$$

samples to this round and let $\bar{T}_j := \sum_{r=0}^j T_r$ denote the total number of samples obtained up to and including round $j \in [l]$.

3. For each $t = \bar{T}_{j-1} + 1, \dots, \bar{T}_j$, sample

$$X_t \sim Q_t := \underset{Q \in \mathcal{U}}{\operatorname{argmin}} \operatorname{LCB}_{t-1}(Q; a_j, \epsilon_j)$$

In essence, we play the modified UCB-E for T_j rounds on expected rewards $\{\mu(a_j; Q)\}_{Q \in \mathcal{U}}$.

4. Define

$$\hat{Q}_j := \underset{Q \in \mathcal{U}}{\operatorname{argmin}} \hat{\mu}_{n_{\bar{T}_j}(Q)}(a_j; Q) \quad \text{and} \quad \mu_T^\circ(a_j) := \hat{\mu}_{n_{\bar{T}_j}(\hat{Q}_j)}(a_j; \hat{Q}_j)$$

Intuitively, \hat{Q}_j and μ_T° are proxies for $Q_{a_j}^*$ and μ_{DR} , respectively.

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

Theorem 4.1: LCB-DR error probability

Under Assumption 3 and the parameter lower bound (1), the LCB-DR algorithm attains

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 $(\epsilon_1, \dots, \epsilon_l)$.

1: **for** $j = 1, \dots, l$ **do**

2: Define proxy set $\mathcal{U}_j = \left\{ Q \in \mathcal{U} \setminus \{Q_{a_j}^*\} : n_{\bar{T}_{j-1}}(Q) < \frac{36}{25} \epsilon_j \Delta_{a_j}^{-2}(Q) \right\} \cup \{Q_{a_j}^*\}$ and quantities $k_j = |\mathcal{U}_j|$, $\tilde{T}_j = \sum_{Q \in \mathcal{U}_j} n_{\bar{T}_{j-1}}(Q)$ and $H_j = \sum_{Q \in \mathcal{U}_j \setminus \{Q_{a_j}^*\}} \Delta_{a_j}^{-2}(Q)$

3: Allocate $T_j = \frac{36}{25} \epsilon_j \left(H_j + \Delta_{a_j, \min}^{-2} \right) - \tilde{T}_j + k_j$ samples to this round and let $\bar{T}_j = \sum_{r=0}^j T_r$

4: **for** $t = \bar{T}_{j-1} + 1, \dots, \bar{T}_j$ **do**

5: Sample $X_t \sim Q_t := \operatorname{argmin}_{Q \in \mathcal{U}} \operatorname{LCB}_{t-1}(Q; a_j, \epsilon_j)$

6: **end for**

7: Define $\hat{Q}_j = \operatorname{argmin}_{Q \in \mathcal{U}} \hat{\mu}_{n_{\bar{T}_j}}(a_j; Q)$ and $\mu_T^o(a_j) = \hat{\mu}_{n_{\bar{T}_j}}(\hat{Q}_j)(a_j; \hat{Q}_j)$

8: **end for**

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

the following error probability:

$$\mathbb{P}(A_T^o \neq a^*) \leq 2k \sum_{j=1}^l u_j \exp \left(- \frac{(C_{a_j}^2 \wedge 1) (T_j + \tilde{T}_j - k_j)}{18 (H_j + \Delta_{a_j, \min}^{-2})} \right)$$

Intuitively, at each round $j \in [l]$, the sample complexity scales with the difficulty of identifying the worst-case distribution $Q_{a_j}^*$, which, as in PE-MAB, is controlled by the suboptimality gaps $\{\Delta_{a_j}(Q)\}_{Q \in \mathcal{U}}$.

Remark 4.1: Random T

Note that the bound of Theorem 4.1 is deterministic, as $\frac{T_j + \tilde{T}_j - k_j}{H_j + \Delta_{a_j, \min}^{-2}} = \frac{36}{25} \epsilon_j$, but the total number of samples $T = k + \sum_{j=1}^l \left[\frac{36}{25} \epsilon_j \left(H_j + \Delta_{a_j, \min}^{-2} \right) - \tilde{T}_j + k_j \right]$ is random.

Remark 4.2: 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} .

Remark 4.3: Selecting T_j

Note that the choice of T_j requires knowledge of unknown quantities, such as \mathcal{U}_j and $Q_{a_j}^*$. However, as shown in the statement of Theorem D.1, for the analysis to hold, all we need is $T_j \geq \frac{36}{25}\epsilon_j \left(H_j + \Delta_{a_j, \min}^{-2} \right) - \tilde{T}_j + k_j$, although we obtain a different concentration inequality when substituting ϵ_j into Hoeffding's.

4.1 Comparison with UE

Focusing on the dominating terms, the probability of selecting a suboptimal arm $a \in \mathcal{A}$ is approximately $\approx \exp\left(-\frac{T\Delta_{\text{DR}}^2(a)}{k}\right)$ for UE and, further bounding for simplicity, $\approx \exp\left(-\frac{(C_a^2 \wedge 1)(T_j + \tilde{T}_j - k_j)}{H_a}\right)$ for LCB-DR with a in the j th permutation position. Assuming large enough sample sizes relative to l , the comparison between the two strategies essentially boils down to $k\Delta_{\text{DR}}^{-2}(a)$ (for UE) v.s. $H_a(C_a^{-2} \vee 1)$ (for LCB-DR), where the smaller terms yields the better rate. We break it down into two cases:

- $\Delta_{\text{DR}}(a) \leq \Delta_{a, \min}$ (or $C_a \leq 1$): intuitively, this means that it is more difficult to rule out a as suboptimal than to identify Q_a^* . Then, LCB-DR's expression becomes $H_a\Delta_{\text{DR}}^{-2}(a)\Delta_{a, \min}^2$, and the comparison reduces to that of MAB: $k\Delta_{a, \min}^{-2}$ (for UE) v.s. H_a (for LCB-DR). In this case, the latter is always favorable.
- $\Delta_{a, \min} \leq \Delta_{\text{DR}}(a)$ (or $C_a \geq 1$): intuitively, this means that it is more difficult to identify Q_a^* than to rule out a as suboptimal. Then, the comparison is between $k\Delta_{\text{DR}}^{-2}(a)$ (for UE) v.s. H_a (for LCB-DR), and either bound can be superior depending on the problem instance.

In essence, optimism is favorable when $\frac{H_a}{k}$ is smaller than the easier task's complexity, $\Delta_{a, \min}^{-2}$ or $\Delta_{\text{DR}}^{-2}(a)$.

5 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 the order to play the arms in. Although the absence of any problem knowledge might preclude exploiting this sequence effectively, perhaps some preliminary understanding of the distributions allows potential advantages.

References

- [AAK⁺22] 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.
- [ABM10] Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best arm identification in multi-armed bandits. In *Annual Conference Computational Learning Theory*, 2010.
- [AHZ23] 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.
- [BDBC⁺09] 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.
- [BHPQ17] 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.
- [BLM13] 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.
- [BMS11] 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.
- [BTdHW⁺13] Aharon Ben-Tal, Dick den Hertog, Anja De Waegenare, Bertrand Melenberg, and Gijs Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, February 2013.
- [CH22] 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.
- [CZZ18] 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.

- [DY10] 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.
- [GN21] 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.
- [HJZ22] 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.
- [HTF09] 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.
- [LS20] Tor Lattimore and Csaba Szepesvari. *Bandit Algorithms*. Cambridge University Press (Virtual Publishing), Cambridge, England, July 2020.
- [MMR09] 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.
- [MSS19] 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.
- [NZ18] Huy Nguyen and Lydia Zakyntinou. 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.
- [Pen23] Binghui Peng. The sample complexity of multi-distribution learning. *arXiv preprint arXiv:2312.04027*, 2023.
- [RM22] Hamed Rahimian and Sanjay Mehrotra. Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization*, 3:1–85, July 2022.
- [Sca58] 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.
- [SDR21] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and theory, Third Edition*. SIAM, August 2021.
- [SGJ22] Tasuku Soma, Khashayar Gatmiry, and Stefanie Jegelka. Optimal algorithms for group distributionally robust optimization and beyond. *arXiv preprint arXiv:2212.13669*, 2022.

- [SKHL20] Shiori Sagawa*, Pang Wei Koh*, Tatsunori B. Hashimoto, and Percy Liang. Distributionally robust neural networks. In *International Conference on Learning Representations*, 2020.
- [SKK21] 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.
- [VNS⁺18] 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.
- [Wai19] 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.
- [ZMV⁺21] Jingzhao Zhang, Aditya Krishna Menon, Andreas Veit, Srinadh Bhojanapalli, Sanjiv Kumar, and Suvrit Sra. Coping with label shift via distributionally robust optimization. In *International Conference on Learning Representations*, 2021.
- [ZZC⁺23] Zihan Zhang, Wenhao Zhan, Yuxin Chen, Simon S Du, and Jason D Lee. Optimal multi-distribution learning. *arXiv preprint arXiv:2312.05134*, 2023.
- [ZZYZ23] Lijun Zhang, Peng Zhao, Tianbao Yang, and Zhi-Hua Zhou. Stochastic approximation approaches to group distributionally robust optimization. *arXiv preprint arXiv:2302.09267*, 2023.

A Expectaion 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 $\sigma_Q^2 := \text{Var}(Q)$ 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 $X_Q, \left(X_Q^{(i)} \right)_{i \in [n_Q]} \stackrel{iid}{\sim} Q$ for each $Q \in \mathcal{U}$. For a collection of functions $\{f_Q : \mathcal{X} \rightarrow [-1, 1]\}_{Q \in \mathcal{U}}$, 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 A.1

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

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

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

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

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 A.3) requires a more careful analysis and, in the process of deriving it, we additionally show the first result (Corollary A.2).

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 $\{h_Q : \mathcal{X} \rightarrow [-1, 1]\}_{Q \in \mathcal{U}}$:

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

where $\epsilon_1, \dots, \epsilon_{\max_{Q \in \mathcal{U}} n_Q} \stackrel{iid}{\sim} \text{Rad}$ are independent from X . Note that we place no assumptions on $h_Q(X_Q)$ being centered. We begin by stating an auxiliary lemma.

Lemma A.1

For random variable $Z \in \mathcal{Z}$ and function class $\mathcal{F} \subset \mathbb{R}^{\mathcal{Z}}$, we have that

$$\sup_{f \in \mathcal{F}} |\mathbb{E}[f(Z)]| \leq \mathbb{E} \left[\sup_{f \in \mathcal{F}} |f(Z)| \right]$$

Proof. For any $f \in \mathcal{F}$,

$$|\mathbb{E}[f(Z)]| \leq \mathbb{E}[|f(Z)|] \leq \mathbb{E} \left[\sup_{g \in \mathcal{F}} |g(Z)| \right]$$

The claim then follows by taking the supremum over $f \in \mathcal{F}$ on the left-hand side. \blacksquare

The proof of the following result is virtually the same as that of [Wai19, Theorem 4.10], with minor modifications, and we present it here for completeness.

Theorem A.2: Symmetrization

For any collection of functions $\{h_Q : \mathcal{X} \rightarrow [-1, 1]\}_{Q \in \mathcal{U}}$, we have that

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

Proof. Let $Y := (Y_Q^{(i)})_{Q \in \mathcal{U}, i \in [n_Q]}$ be an independent copy of X and let P denote their common distribution. Then,

$$\begin{aligned} \mathbb{E} \left[\max_{Q \in \mathcal{U}} \left| \frac{1}{n_Q} \sum_{i=1}^{n_Q} \{h_Q(X_Q^{(i)}) - \mathbb{E}[h_Q(X_Q)]\} \right| \right] &= \mathbb{E} \left[\max_{Q \in \mathcal{U}} \left| \frac{1}{n_Q} \sum_{i=1}^{n_Q} \{h_Q(X_Q^{(i)}) - \mathbb{E}[h_Q(Y_Q^{(i)})]\} \right| \right] \\ &= \int_{\mathcal{X}^T} \max_{Q \in \mathcal{U}} \left| \mathbb{E} \left[\frac{1}{n_Q} \sum_{i=1}^{n_Q} [h_Q(x_Q^i) - h_Q(Y_Q^{(i)})] \right] \right| dP(\mathbf{x}) \\ &=: (*_1) \end{aligned}$$

Here, we view $\mathbf{x} := (x_Q^i)_{Q \in \mathcal{U}, i \in [n_Q]} \in \mathcal{X}^T$ as a T -dimensional vector. For each such vector, define function class

$$\mathcal{F}_{\mathbf{x}} := \left\{ \mathbf{y} \mapsto \frac{1}{n_Q} \sum_{i=1}^{n_Q} [h_Q(x_Q^i) - h_Q(y_Q^i)] : Q \in \mathcal{U} \right\} \subset \{\mathcal{X}^T \rightarrow \mathbb{R}\}$$

We can then apply Lemma A.1 to obtain

$$\begin{aligned}
(*_1) &= \int_{\mathcal{X}^T} \max_{f \in \mathcal{F}_x} |\mathbb{E}[f(Y)]| dP(\mathbf{x}) \\
&\leq \int_{\mathcal{X}^T} \mathbb{E} \left[\max_{f \in \mathcal{F}_x} |f(Y)| \right] dP(\mathbf{x}) && \text{Lem. A.1} \\
&= \int_{\mathcal{X}^T} \mathbb{E} \left[\max_{Q \in \mathcal{U}} \left| \frac{1}{n_Q} \sum_{i=1}^{n_Q} [h_Q(x_Q^i) - h_Q(Y_Q^{(i)})] \right| \right] dP(\mathbf{x}) \\
&= \mathbb{E} \left[\max_{Q \in \mathcal{U}} \left| \frac{1}{n_Q} \sum_{i=1}^{n_Q} [h_Q(X_Q^{(i)}) - h_Q(Y_Q^{(i)})] \right| \right] \\
&=: (*_2)
\end{aligned}$$

Next, define $n := \max_{Q \in \mathcal{U}} n_Q$ and let $\tilde{\epsilon}_1, \dots, \tilde{\epsilon}_n \in \{-1, 1\}$ be fixed quantities. From symmetry and independence, we have that

$$\left(h_Q(X_Q^{(i)}) - h_Q(Y_Q^{(i)}) \right)_{Q \in \mathcal{U}, i \in [n_Q]} \stackrel{d}{=} \left(\tilde{\epsilon}_i [h_Q(X_Q^{(i)}) - h_Q(Y_Q^{(i)})] \right)_{Q \in \mathcal{U}, i \in [n_Q]}$$

Hence, if we define Rademacher variables $\epsilon^n \stackrel{iid}{\sim} \text{Rad}$ that are independent from X and Y , we can conclude that

$$\begin{aligned}
(*_2) &= \frac{1}{2^n} \sum_{\tilde{\epsilon}^n \in \{-1, 1\}^n} \mathbb{E} \left[\max_{Q \in \mathcal{U}} \left| \frac{1}{n_Q} \sum_{i=1}^{n_Q} \tilde{\epsilon}_i [h_Q(X_Q^{(i)}) - h_Q(Y_Q^{(i)})] \right| \right] \\
&= \mathbb{E} \left[\max_{Q \in \mathcal{U}} \left| \frac{1}{n_Q} \sum_{i=1}^{n_Q} \epsilon_i [h_Q(X_Q^{(i)}) - h_Q(Y_Q^{(i)})] \right| \right] \\
&\leq \mathbb{E} \left[\max_{Q \in \mathcal{U}} \left\{ \left| \frac{1}{n_Q} \sum_{i=1}^{n_Q} \epsilon_i h_Q(X_Q^{(i)}) \right| + \left| \frac{1}{n_Q} \sum_{i=1}^{n_Q} \epsilon_i h_Q(Y_Q^{(i)}) \right| \right\} \right] \\
&\leq \mathbb{E} \left[\max_{Q \in \mathcal{U}} \left\{ \left| \frac{1}{n_Q} \sum_{i=1}^{n_Q} \epsilon_i h_Q(X_Q^{(i)}) \right| \right\} + \max_{Q \in \mathcal{U}} \left\{ \left| \frac{1}{n_Q} \sum_{i=1}^{n_Q} \epsilon_i h_Q(Y_Q^{(i)}) \right| \right\} \right] \\
&= 2 \mathbb{E} \left[\max_{Q \in \mathcal{U}} \left| \frac{1}{n_Q} \sum_{i=1}^{n_Q} \epsilon_i h_Q(X_Q^{(i)}) \right| \right] \\
&= 2 \mathfrak{R}_T(\{h_Q\}_{Q \in \mathcal{U}})
\end{aligned}$$

■

A.2 Bounds on the Rademacher complexity

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

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

where $\epsilon^n = (\epsilon_1, \dots, \epsilon_n) \stackrel{iid}{\sim} \text{Rad}$. 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., [Wai19, Chapter 5].

Lemma A.2: Bounding the Rademacher complexity of a finite set

Let $\Theta \subset \mathbb{R}^n$ satisfy $2 \leq |\Theta| < \infty$. Then,

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

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., [BLM13, Theorem 2.5]) to obtain the claim. ■

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

Corollary A.1

For a collection of functions $\{h_Q : \mathcal{X} \rightarrow [-1, 1]\}_{Q \in \mathcal{U}}$, define the random variable

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

Then, we have that

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

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

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

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

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

Moreover, since $D_{\Theta^x} = \max_{Q \in \mathcal{U}} \sqrt{\sum_{i=1}^{n_Q} \left(\frac{h_Q(x_Q^i)}{n_Q} \right)^2}$, Lemma A.2 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 bound of interest.

Corollary A.2

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

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

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

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

Hence, combining Theorem A.2 and Corollary A.1 yields

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

■

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

Lemma A.3

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

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

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

■

Lemma A.4: Variance of Lipschitz functions

Let $Z \in \mathcal{Z} \subset \mathbb{R}$ be a random variable, and suppose that $f : \mathcal{Z} \rightarrow \mathbb{R}$ is L -Lipschitz. Then,

$$\text{Var}(f(Z)) \leq 2L^2 \text{Var}(Z)$$

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

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

■

Borrowing ideas from [GN21, Corollary 3.5.7], we then conclude the second target bound.

Theorem A.3

Suppose that $\mathcal{X} \subset \mathbb{R}$. Let $\{f_Q : \mathcal{X} \rightarrow [-1, 1]\}_{Q \in \mathcal{U}}$ 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,

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

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:

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

From Lemma A.4 and the fact that $\mathbb{E} [f_Q \left(X_Q \right)] = 0$, we know that

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

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

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

In other words, we have that

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

Then, Lemma A.3 implies that

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

Combining this with Theorem A.2 and Corollary A.1, we conclude that

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

■

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 $\left(X_Q^{(i)} \right)_{i \in [n_Q]} \stackrel{iid}{\sim} Q$ 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 $f : \mathcal{X} \rightarrow \mathbb{R}$. 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 [BLM13, 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 $\Phi_f : (\mathcal{X}^k)^n \rightarrow [0, 1]$ by $\Phi_f(\mathbf{x}_1, \dots, \mathbf{x}_n) := \min_{Q \in \mathcal{U}} \frac{1}{n} \sum_{i=1}^n f(x_Q^i)$, where each $\mathbf{x}_i = (x_Q^i)_{Q \in \mathcal{U}} \in \mathcal{X}^k$. 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 B.1: Bounded differences

Suppose that $f : \mathcal{X} \rightarrow [0, 1]$. Then,

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

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

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

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

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

■

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

Corollary B.1

For any two functions $f, g : \mathcal{X} \rightarrow [0, 1]$, 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 B.1, we obtain

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

■

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

Corollary B.2

Let $f, g : \mathcal{X} \rightarrow [0, 1]$. Then,

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

Proof. Since $Z_{f,g} = (\Phi_f - \Phi_g)(X)$ and X has independent components, we simply apply Corollary B.1 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: $0 =: n_{(0)} \leq n_{(1)} \leq \dots \leq n_{(k)}$ and let $Q_{(j)} \in \mathcal{U}$ be such that $n_{Q_{(j)}} = n_{(j)}$. Our analysis then relies on the following:

$$\begin{aligned} V_T &:= \sum_{j=1}^k (n_{(j)} - n_{(j-1)}) \mathbb{E} \left[\max_{r \in \{j, \dots, k\}} \frac{1}{n_{(r)}^2} [X_{Q_{(r)}} - \mu_{Q_{(r)}}]^2 \right] \\ \Sigma_T^2 &:= \mathbb{E} \left[\max_{Q \in \mathcal{U}} \frac{1}{n_Q^2} \sum_{i=1}^{n_Q} (X_Q^{(i)} - \mu_Q)^2 \right] \\ \sigma_T^2 &:= \max_{Q \in \mathcal{U}} \frac{\sigma_Q^2}{n_Q} \end{aligned}$$

Theorem B.1

Suppose that $\mathcal{X} \subset \mathbb{R}$ and $f, g : \mathcal{X} \rightarrow [0, 1]$ are L -Lipschitz. Then,

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

B.2.1 Preliminaries

To prove Theorem B.1, 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 [BLM13] 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 B.2: Sub-gamma concentration

Let $X \in \Gamma_+(\nu, c)$. Then,

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

Proof. See [BLM13, Section 2.4]. ■

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

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

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

Lemma B.1: 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 $h := \alpha f + (1 - \alpha)g$ and $h_i := \alpha f_i + (1 - \alpha)g_i$. Then, for any $i \in [n]$ and $\mathbf{x} \in \mathcal{X}^n$,

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

and

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

■

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

Proposition B.3: Cgf of self-bounding functions

Suppose that $f : \mathcal{X}^n \rightarrow \mathbb{R}_+$ has the self-bounding property and let $X^n = (X_1, \dots, X_n)$ be independent random variables. Then,

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

Proof. See [BLM13, 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 $a_+^2 := (a_+)^2$; that is, we take the positive part before squaring.

Proposition B.4: 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 $f : \mathcal{X}^n \rightarrow \mathbb{R}_+$, 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

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

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 [BLM13, Theorem 6.16]. ■

Proof of Theorem B.1. To conclude our main result, we begin with a more general setup: let $X := (X_Q^{(i)})_{Q \in \mathcal{U}, i \in [n]}$, where $n \in \mathbb{N}$, be a collection of independent \mathcal{X} -valued random variables, and let $X^{(i)} := (X_Q^{(i)})_{Q \in \mathcal{U}}$ for each $i \in [n]$. We do 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 $f = \left\{ f_Q : \mathcal{X} \rightarrow \left[0, \frac{b}{\sqrt{6}} \right] \right\}_{Q \in \mathcal{U}}$ and $g = \left\{ g_Q : \mathcal{X} \rightarrow \left[0, \frac{b}{\sqrt{6}} \right] \right\}_{Q \in \mathcal{U}}$, where $b > 0$. Define

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

Similarly, consider the variance variants:

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

Following the analysis of [BLM13, 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

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

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

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

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

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

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

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

so that

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

and

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

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

$$\begin{aligned}
Z_{f,g} - Z'_i &= \Phi_{f,g}(X) - \Phi_{f,g}(Y_i) \\
&\leq \sum_{j=1}^n f_{\hat{Q}_f(Y_i)}(X_{\hat{Q}_f(Y_i)}^{(j)}) - \sum_{j=1}^n g_{\hat{Q}_g(X)}(X_{\hat{Q}_g(X)}^{(j)}) \\
&\quad - \left[f_{\hat{Q}_f(Y_i)}(W_{\hat{Q}_f(Y_i)}^{(i)}) + \sum_{j \in [n]: j \neq i} f_{\hat{Q}_f(Y_i)}(X_{\hat{Q}_f(Y_i)}^{(j)}) \right] \\
&\quad + \left[g_{\hat{Q}_g(X)}(W_{\hat{Q}_g(X)}^{(i)}) + \sum_{j \in [n]: j \neq i} g_{\hat{Q}_g(X)}(X_{\hat{Q}_g(X)}^{(j)}) \right] \\
&= f_{\hat{Q}_f(Y_i)}(X_{\hat{Q}_f(Y_i)}^{(i)}) - f_{\hat{Q}_f(Y_i)}(W_{\hat{Q}_f(Y_i)}^{(i)}) + g_{\hat{Q}_g(X)}(W_{\hat{Q}_g(X)}^{(i)}) - g_{\hat{Q}_g(X)}(X_{\hat{Q}_g(X)}^{(i)})
\end{aligned}$$

Then,

$$\begin{aligned}
&(Z_{f,g} - Z'_i)_+^2 \\
&\leq \left[f_{\hat{Q}_f(Y_i)}(X_{\hat{Q}_f(Y_i)}^{(i)}) - f_{\hat{Q}_f(Y_i)}(W_{\hat{Q}_f(Y_i)}^{(i)}) + g_{\hat{Q}_g(X)}(W_{\hat{Q}_g(X)}^{(i)}) - g_{\hat{Q}_g(X)}(X_{\hat{Q}_g(X)}^{(i)}) \right]^2 \\
&\leq 2 \left[f_{\hat{Q}_f(Y_i)}(X_{\hat{Q}_f(Y_i)}^{(i)}) - f_{\hat{Q}_f(Y_i)}(W_{\hat{Q}_f(Y_i)}^{(i)}) \right]^2 + 2 \left[g_{\hat{Q}_g(X)}(X_{\hat{Q}_g(X)}^{(i)}) - g_{\hat{Q}_g(X)}(W_{\hat{Q}_g(X)}^{(i)}) \right]^2
\end{aligned} \tag{B.1}$$

Recall that our goal is to bound $V^+ = \sum_{i=1}^n \mathbb{E} \left[(Z_{f,g} - Z'_i)_+^2 \mid X \right]$. 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

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{E} \left[\left[g_{\hat{Q}_g(X)}(X_{\hat{Q}_g(X)}^{(i)}) - g_{\hat{Q}_g(X)}(W_{\hat{Q}_g(X)}^{(i)}) \right]^2 \mid X \right] \\
&= \sum_{i=1}^n \left\{ \left[g_{\hat{Q}_g(X)}(X_{\hat{Q}_g(X)}^{(i)}) - \mu_{g,i,\hat{Q}_g(X)} \right]^2 + \mathbb{E} \left[\left[g_{\hat{Q}_g(X)}(W_{\hat{Q}_g(X)}^{(i)}) - \mu_{g,i,\hat{Q}_g(X)} \right]^2 \mid X \right] \right\} \\
&\leq \underbrace{\max_{Q \in \mathcal{U}} \left\{ \sum_{i=1}^n \left[g_Q(X_Q^{(i)}) - \mu_{g,i,Q} \right]^2 \right\}}_{=: \Gamma_g} + \underbrace{\max_{Q \in \mathcal{U}} \left\{ \sum_{i=1}^n \mathbb{E} \left[\left[g_Q(W_Q^{(i)}) - \mu_{g,i,Q} \right]^2 \right] \right\}}_{=: \sigma_g^2}
\end{aligned}$$

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 (B.1) 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

$$\begin{aligned}
&\left[f_{\hat{Q}_f(Y_i)}(X_{\hat{Q}_f(Y_i)}^{(i)}) - f_{\hat{Q}_f(Y_i)}(W_{\hat{Q}_f(Y_i)}^{(i)}) \right]^2 \\
&\leq 2 \left[f_{\hat{Q}_f(Y_i)}(X_{\hat{Q}_f(Y_i)}^{(i)}) - \mu_{f,i,\hat{Q}_f(Y_i)} \right]^2 + 2 \left[f_{\hat{Q}_f(Y_i)}(W_{\hat{Q}_f(Y_i)}^{(i)}) - \mu_{f,i,\hat{Q}_f(Y_i)} \right]^2 \\
&\leq 2 \max_{Q \in \mathcal{U}} \left\{ \left[f_Q(X_Q^{(i)}) - \mu_{f,i,Q} \right]^2 \right\} + 2 \max_{Q \in \mathcal{U}} \left\{ \left[f_Q(W_Q^{(i)}) - \mu_{f,i,Q} \right]^2 \right\}
\end{aligned}$$

Summing and taking conditional expectations then yields

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

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

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

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

B.2.3 Efron-Stein

Next, we apply exponential Efron-Stein (Proposition B.4): for $\lambda \in [0, b^{-1}]$, we have that

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

B.2.4 Self-boundedness

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

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

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

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

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

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

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

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

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

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

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

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

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

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

That is, $h^{(2)}$ is also self-bounded. As a result, Proposition B.3 implies that

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

provided that $\lambda \in [0, b^{-1}]$, 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 (B.2) and (B.3) to get that

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

for all $\lambda \in [0, b^{-1}]$. 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 B.2 yields the tail bound

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

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 $n := \max_{Q \in \mathcal{U}} n_Q$ 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 $(X_Q^{(i)})_{i \in [n_Q]} \stackrel{iid}{\sim} Q$ and $X_Q^{(n_Q+1)} = \dots = X_Q^{(n)} = x_0$ almost surely. Let $f : \mathcal{X}_0 \rightarrow \mathbb{R}$ be the L -Lipschitz function from the statement of Theorem B.1, and consider its extension $\tilde{f} : \mathcal{X} \rightarrow \mathbb{R}$ given by

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

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

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

where the variables follow the appropriate distributions, as in the original goal. Note that $f_Q \in [0, \frac{1}{n_Q}]$, so that we can set $b = \frac{\sqrt{6}}{\min_{Q \in \mathcal{U}} n_Q}$. We analogously define everything for g . Next, we apply Lemma A.4 under the Lipschitzness assumption to obtain

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

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

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

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

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

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,

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

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

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

Plugging these values back into the bound (B.4) then yields the claim. \blacksquare

C Proofs of Section 3

Recall our non-adaptive proxy objective

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

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

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

Using the same argument as in the proof of Proposition B.1, we note that

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

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

Theorem C.1

For any $a \in \mathcal{A}$, we have that

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

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

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

Proof. We apply Theorem A.1 on functions $f_Q(x) := \mathbb{E}[r(a, X_Q)] - r(a, x)$. Note that $f_Q \in [-1, 1]$ since $r \in [0, 1]$. 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 C.1. Then, we get that

$$\begin{aligned} \mathbb{E}[\mu_T^\circ(a^*) - \mu_T^\circ(a)] &= \Delta_{\text{DR}}(a) + \mathbb{E}[\mu_{\text{DR}}(a) - \mu_T^\circ(a)] - \mathbb{E}[\mu_{\text{DR}}^* - \mu_T^\circ(a^*)] \\ &= \Delta_{\text{DR}}(a) + \mathbb{E}[D_a] - \mathbb{E}[D_{a^*}] \\ &\geq \Delta_{\text{DR}}(a) - |\mathbb{E}[D_a]| - |\mathbb{E}[D_{a^*}]| \\ &\geq \Delta_{\text{DR}}(a) - \mathbb{E}[|D_a|] - \mathbb{E}[|D_{a^*}|] \\ &\geq \Delta_{\text{DR}}(a) - 2\mathbb{E}[U_a] \\ &\geq \Delta_{\text{DR}}(a) - 2B \end{aligned}$$

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

$$\begin{aligned} \mathbb{P}(A_T^\circ = a) &\leq \mathbb{P}(\mu_T^\circ(a) \geq \mu_T^\circ(a^*)) \\ &= \mathbb{P}(\mu_T^\circ(a) - \mu_T^\circ(a^*) - \mathbb{E}[\mu_T^\circ(a) - \mu_T^\circ(a^*)] \geq \mathbb{E}[\mu_T^\circ(a^*) - \mu_T^\circ(a)]) \\ &\leq \mathbb{P}(\mu_T^\circ(a) - \mu_T^\circ(a^*) - \mathbb{E}[\mu_T^\circ(a) - \mu_T^\circ(a^*)] \geq \Delta_{\text{DR}}(a) - 2B) \end{aligned} \quad (\text{C.1})$$

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

C.1 Proof of Theorem 3.1

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

Corollary C.1: UE concentration inequality

We have that

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

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

Proof. Note that in the notation of Appendix B.1, $Z_{r(a,\cdot)} = \mu_T^\circ(a)$. Since $r(a,\cdot) \in [0,1]$ for each $a \in \mathcal{A}$, the claim follows by applying Corollary B.2. \blacksquare

Next, note that under the assumption $n \geq \left(\frac{8}{\Delta_{\text{DR},\min}}\right)^2 \log k$, we get that $\Delta_{\text{DR}}(a) \geq 8\sqrt{\frac{\log k}{n}}$ for all $a \in \mathcal{A}$ with a positive gap. Hence, for all such a , plugging in the bound $B = 4\sqrt{\frac{\log k}{n}}$ into Equation (C.1) yields

$$\begin{aligned} \mathbb{P}(A_T^\circ = a) &\leq \mathbb{P}\left(\mu_T^\circ(a) - \mu_T^\circ(a^*) - \mathbb{E}[\mu_T^\circ(a) - \mu_T^\circ(a^*)] \geq \Delta_{\text{DR}}(a) - 8\sqrt{\frac{\log k}{n}}\right) && \text{Eq. (C.1)} \\ &\leq \exp\left(-\frac{n}{2} \left[\Delta_{\text{DR}}(a) - 8\sqrt{\frac{\log k}{n}}\right]^2\right) && \text{Cor. C.1} \end{aligned}$$

This directly yields the desired regret bound:

$$\begin{aligned} \mathbb{E}[\Delta_{\text{DR}}(A_T^\circ)] &= \sum_{a \in \mathcal{A}: \Delta_{\text{DR}}(a) > 0} \Delta_{\text{DR}}(a) \mathbb{P}(A_T^\circ = a) \\ &\leq \sum_{a \in \mathcal{A}: \Delta_{\text{DR}}(a) > 0} \Delta_{\text{DR}}(a) \exp\left(-\frac{n}{2} \left[\Delta_{\text{DR}}(a) - 8\sqrt{\frac{\log k}{n}}\right]^2\right) \end{aligned}$$

C.2 Proof of Corollary 3.1

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

$$\begin{aligned} \mathbb{E}[\Delta_{\text{DR}}(A_T^\circ)] &= \sum_{a \in \mathcal{A}: \Delta_{\text{DR}}(a) \leq \Delta} \Delta_{\text{DR}}(a) \mathbb{P}(A_T^\circ = a) + \sum_{a \in \mathcal{A}: \Delta_{\text{DR}}(a) > \Delta} \Delta_{\text{DR}}(a) \mathbb{P}(A_T^\circ = a) \\ &\leq \Delta + \sum_{a \in \mathcal{A}: \Delta_{\text{DR}}(a) > \Delta} \Delta_{\text{DR}}(a) \exp\left(-\frac{n}{2} \left[\Delta_{\text{DR}}(a) - 8\sqrt{\frac{\log k}{n}}\right]^2\right) \end{aligned}$$

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

$$\mathbb{E}[\Delta_{\text{DR}}(A_T^\circ)] \leq \inf_{\Delta \geq 0} \left\{ \Delta + \sum_{a \in \mathcal{A}: \Delta_{\text{DR}}(a) > \Delta} \Delta_{\text{DR}}(a) \exp\left(-\frac{n}{2} \left[\Delta_{\text{DR}}(a) - 8\sqrt{\frac{\log k}{n}}\right]^2\right) \right\} \quad (\text{C.2})$$

Next, we introduce a simple technical lemma.

Lemma C.1

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

$$\begin{aligned} f'(x) &= \exp\left(-\alpha(x-\beta)^2\right) - 2\alpha x(x-\beta)\exp\left(-\alpha(x-\beta)^2\right) \\ &= [1 - 2\alpha x(x-\beta)]\exp\left(-\alpha(x-\beta)^2\right) \end{aligned}$$

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 $x \geq \frac{1}{2}\left(\beta + \sqrt{\beta^2 + \frac{2}{\alpha}}\right)$. \blacksquare

As a result, we can show the following inequality.

Lemma C.2

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

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

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 C.1 with $\alpha := \frac{n}{2}$ and $\beta := 8\sqrt{\frac{\log k}{n}}$, so that we know it is decreasing for $x \geq K$, where

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

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

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

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

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

C.3 Proof of Theorem 3.2

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

Corollary C.2: NUE concentration inequality

Suppose that $r(a, \cdot)$ is L -Lipschitz for each $a \in \mathcal{A}$. Then, we have that

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

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

Proof. Once again, using the definitions of Appendix B, we get that $Z_{r(a, \cdot)} = \mu_T^\circ(a)$. Since $r(a, \cdot) \in [0, 1]$ is L -Lipschitz for each $a \in \mathcal{A}$, the claim follows by applying Theorem B.1. \blacksquare

As in the UE analysis, provided that $\Delta_{\text{DR}, \min} \geq G_T = 8\left(\frac{4\log k}{\min_{Q \in \mathcal{U}} n_Q} + L\sigma_T\sqrt{2\log k}\right)$, we can plug $B = \frac{16\log k}{\min_{Q \in \mathcal{U}} n_Q} + 4L\sigma_T\sqrt{2\log k}$ into Equation (C.1) to conclude that

$$\begin{aligned} \mathbb{P}(A_T^\circ = a) &\leq \mathbb{P}(\mu_T^\circ(a) - \mu_T^\circ(a^*) - \mathbb{E}[\mu_T^\circ(a) - \mu_T^\circ(a^*)] \geq \Delta_{\text{DR}}(a) - G_T) && \text{Eq. (C.1)} \\ &\leq \exp\left(-\frac{[\Delta_{\text{DR}}(a) - G_T]^2}{16L^2(2\sigma_T^2 + \Sigma_T^2 + 6V_T) + \frac{2\sqrt{6}}{\min_{Q \in \mathcal{U}} n_Q} [\Delta_{\text{DR}}(a) - G_T]}\right) && \text{Cor. C.1} \end{aligned}$$

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

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

D Modified UCB-E

Our goal is to perform a minimization variant of UCB-E [ABM10] 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 $t \in \{n_0(Q), \dots, n_0(Q) + T\}$, 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:

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

along with its cardinality $k_0 := |\mathcal{U}_0|$, total initial sample size $\tilde{T}_0 := \sum_{Q \in \mathcal{U}_0} n_0(Q)$ and the complexity notion it defines: $H_0 := \sum_{Q \in \mathcal{U}_0 \setminus \{Q^*\}} \Delta^{-2}(Q)$. The intuition is that \mathcal{U}_0 is a proxy for the set of arms played:

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

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

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

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

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

After T rounds, we output

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

Theorem D.1: 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 $t \in \{n_0(Q), \dots, n_0(Q) + T\}$, and that

$$\begin{aligned}\epsilon &\geq \frac{25}{36} \Delta_{\min}^2 [n_0(Q^*) - 1] \\ T &\geq \frac{36}{25} \epsilon (H_0 + \Delta_{\min}^{-2}) - \tilde{T}_0 + k_0\end{aligned}$$

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 $t \in \{0, \dots, T\}$ and $Q \in \mathcal{U}$, we have by assumption that

$$|\mu(Q) - \hat{\mu}_{n_t(Q)}(Q)| < \frac{1}{5} \sqrt{\frac{\epsilon}{n_t(Q)}} \quad (\text{D.1})$$

since $n_t(Q) \in \{n_0(Q), \dots, n_0(Q) + T\}$. We can then conclude the following lemma.

Lemma D.1

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

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

Rearranging then yields the claim. ■

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

Corollary D.1

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

Proof. Fix $Q \in \mathcal{U}' \setminus \{Q^*\}$ and let $t \in [T]$ denote any round in which $Q_t = Q$. From Lemma D.1 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 D.1

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

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 D.1 then implies that

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

■

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

Proposition D.2

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

Proof. We have that

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

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

Corollary D.2

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

Proof. This immediately follows from Proposition D.2 and our lower bound assumption on ϵ :

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

■

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

Proposition D.3

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

Proof. Let $Q \in \mathcal{U} \setminus \{Q^*\}$. Since $Q^* \in \mathcal{U}'$ by Corollary D.2, let $t \in [T]$ be the last round such that $Q_t = Q^*$. Then,

$$\begin{aligned} \mu(Q) - \frac{4}{5}\sqrt{\frac{\epsilon}{n_T(Q)}} &\geq \mu(Q) - \frac{4}{5}\sqrt{\frac{\epsilon}{n_{t-1}(Q)}} \\ &> \text{LCB}_{t-1}(Q; \epsilon) && \text{Eq. (D.1)} \\ &\geq \text{LCB}_{t-1}(Q^*; \epsilon) \\ &> \mu^* - \frac{6}{5}\sqrt{\frac{\epsilon}{n_{t-1}(Q^*)}} && \text{Eq. (D.1)} \\ &= \mu^* - \frac{6}{5}\sqrt{\frac{\epsilon}{n_T(Q^*) - 1}} && n_T(Q^*) = n_t(Q^*) = n_{t-1}(Q^*) + 1 \\ &> \mu^* - \Delta(Q) && \text{Prop. D.2 and } \Delta_{\min} \leq \Delta(Q) \end{aligned}$$

The claim then follows by rearranging the terms. ■

Let $Q \in \mathcal{U} \setminus \{Q^*\}$. From Propositions D.2 and D.3, we note that

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

Rearranging the terms then yields

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

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

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

Since this holds for all $Q \in \mathcal{U} \setminus \{Q^*\}$, it follows that $\hat{Q} = Q^*$. ■

E Proof of Theorem 4.1

Suppose that we are operating under permutation (a_1, \dots, a_l) and parameters $(\epsilon_1, \dots, \epsilon_l)$ satisfying the bound (1). 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

$$\begin{aligned}
&\mathbb{P}\left(\bigcap_{Q \in \mathcal{U}} \bigcap_{t \in [u_j]} \left\{|\mu(a_j; Q) - \hat{\mu}_t(a_j; Q)| < \frac{C_{a_j} \wedge 1}{5}\sqrt{\frac{\epsilon_j}{t}}\right\}\right) \\
&\geq 1 - 2ku_j \exp\left(-\frac{2(C_{a_j}^2 \wedge 1)\epsilon_j}{25}\right) \\
&= 1 - 2ku_j \exp\left(-\frac{(C_{a_j}^2 \wedge 1)(T_j + \tilde{T}_j - k_j)}{18(H_j + \Delta_{a_j, \min}^{-2})}\right)
\end{aligned}$$

where the last line follows by rearranging terms in the definition of T_j . 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. Note that

$$\bar{T}_j = \sum_{r=0}^j T_r = k + \sum_{r=1}^j \left[\frac{36}{25} \epsilon_r \underbrace{\left(\underbrace{H_r}_{\leq H_{a_r}} + \Delta_{a_r, \min}^{-2} \right)}_{\leq 0} \underbrace{-\tilde{T}_r}_{\leq 0} + \underbrace{k_r}_{\leq k} \right] \leq u_j$$

Hence, $n_{\bar{T}_{j-1}}(Q) + T_j \leq \bar{T}_j \leq u_j$, for any $Q \in \mathcal{U}$ 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 $t \in \{n_{\bar{T}_{j-1}}(Q), \dots, n_{\bar{T}_{j-1}}(Q) + T_j\}$. Moreover, since $\bar{T}_0 = u_0$, we have from the lower bound (1) on $(\epsilon_1, \dots, \epsilon_l)$ that

$$\epsilon_j \geq \frac{25}{36} \Delta_{a_j, \min}^2 (u_{j-1} - 1) \geq \frac{25}{36} \Delta_{a_j, \min}^2 (\bar{T}_{j-1} - 1) \geq \frac{25}{36} \Delta_{a_j, \min}^2 (n_{\bar{T}_{j-1}}(Q_{a_j}^*) - 1)$$

for all $j \in [l]$. We can then conclude the following result.

Theorem E.1

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

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

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 D.1, as its assumptions are satisfied under A_j . Moreover, we have that

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

■

E.3 LCB-DR correctness

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

$$\begin{aligned} \mu_T^\circ(a^*) - \mu_T^\circ(a) &= \mu_T^\circ(a^*) - \mu_{\text{DR}}^* + \Delta_{\text{DR}}(a) + \mu_{\text{DR}}(a) - \mu_T^\circ(a) \\ &> \Delta_{\text{DR}}(a) - \frac{\Delta_{\text{DR},\min}}{2} - \frac{\Delta_{\text{DR}}(a)}{2} && \text{Thm. E.1} \\ &\geq 0 && \Delta_{\text{DR},\min} \leq \Delta_{\text{DR}}(a) \end{aligned}$$

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